CHARACTERISTICS OF GROUND WATER IN MAJHA REGION PUNJAB, INDIA USING SPECTROSCOPY AND MULTIVARIATE ANALYSIS

Thesis Submitted for the Award of the Degree of

DOCTOR OF PHILOSOPHY

in

Physics

by

Yogesh Kumar

41800593

Supervised by   Co-Supervised by

Dr. Neha Munjal   **Dr. Uma Kamboj**

Professor   Associate Professor
Department of Physics   Department of Physics
Lovely Professional University   Lovely Professional University
Phagwara, Punjab (India)   Phagwara, Punjab (India)

LOVELY PROFESSIONAL UNIVERSITY, PUNJAB

2025

DECLARATION

I, hereby declare that the presented work in the thesis entitled "Characteristics of Ground Water in Majha Region Punjab, India Using Spectroscopy and Multivariate Analysis" in fulfillment of degree of Doctor of Philosophy (Ph.D.) is outcome of research work carried out by me under the supervision of Dr. Neha Munjal working as Professor in the Department of Physics, School of Chemical Engineering and Physical Sciences of Lovely Professional University, Punjab, India and under the co-supervision of Dr. Uma Kamboj working as Associate Professor in the Department of Physics, School of Chemical Engineering and Physical Sciences of Lovely Professional University, Punjab, India. In keeping with general practice of reporting scientific observations, due acknowledgements have been made whenever work described here has been based on findings of other investigator. This work has not been submitted in part or full to any other University or Institute for the award of any degree.

(Signature of Scholar)

Name of the scholar   :   Yogesh Kumar

Registration No.   :   41800593

Department/school   :   Physics, School of Chemical Engineering and Physical Sciences

Lovely Professional University,

Punjab, India

CERTIFICATE

This is to certify that the work reported in the Ph.D. thesis entitled "Characteristics of Ground Water in Majha Region Punjab, India Using Spectroscopy and Multivariate Analysis" submitted in fulfillment of the requirement for the award of degree of Doctor of Philosophy (Ph.D.) in the Department of Physics, School of Chemical Engineering and Physical Sciences is a research work carried out by Mr. Yogesh Kumar Registration No. 41800593 is bonafide record of his/her original work carried out under my supervision and that no part of thesis has been submitted for any other degree, diploma or equivalent course.

(Signature of Supervisor)

Name of supervisor: Dr. Neha Munjal

Designation: Professor

Department/school:

Physics, School of Chemical

Engineering and Physical Sciences

Lovely Professional University,

Punjab, India

(Signature of Co-Supervisor)

Name of Co-Supervisor: Dr. Uma Kamboj

Designation: Associate Professor

Department/school:

Physics, School of Chemical

Engineering and Physical Sciences

Lovely Professional University,

Punjab, India

ABSTRACT

Near infrared spectroscopy (NIRS) has been in use for many years in various fields like pharmaceuticals, food and beverage industries, in the field of textile, in the field of agriculture etc. In the present work NIRS is used to develop a model to determine the physicochemical properties of groundwater from the four districts of Majha region of Punjab, India. The regression model is developed using the NIR data and reference data by applying multivariate analysis. The model can be used to examine the groundwater of Majha region of Punjab India qualitatively and quantitatively. One of the objectives of the work in this thesis was to analyze the contamination of the ground water samples due to urea and arsenic from the four districts viz. Amritsar, Pathankot, Gurdaspur and Tarn taran of Majha region of Punjab, India. Another aim was to determine the characteristics of the groundwater sample quantitatively by measuring the following physicochemical parameters: total hardness (TH), turbidity, pH value, chloride content, total dissolved solids (TDS) and electrical conductivity (EC). All the measurements for the physicochemical parameters were done at Kalpin Watertech Laboratories, Thane. Values of some of the parameters for some of the samples are not under the acceptable limits as prescribed by BIS and WHO.

MS excel software was used to determine correlation coefficient (r) between the various pairs of the physicochemical parameters and also to determine other descriptive statistical parameters like mean, median, standard deviation, variance etc. It was observed that the value of correlation coefficient between the following pairs viz. total dissolved solids-electrical conductivity, total hardness-total dissolved solids, total hardness-electrical conductivity, total dissolved solids-chlorine, chlorine-electrical conductivity, total hardness-chlorine is greater than or nearly equal to 0.6 and between total dissolved solids (TDS) and electrical conductivity (EC) it has value greater than 0.9. As there was good correlation observed between the above listed pairs of physicochemical parameters, bivariate regression analysis was done and six regression equations were developed. The developed regression equations from the reference physicochemical data help in determination of other parameters by knowing only one parameter. This will reduce the effective cost and save the time in the determination of other physicochemical parameters in question.

CAMO unscramble ×10.5 and MS Excel softwares were used for regression analysis and for getting graphs of NIR spectra. PCA was done on this basis of region (districts), source of groundwater (submersible or hand pumps) and depth of the source of groundwater by using the physicochemical data obtained in the laboratory and the NIR data. The softwares

used for performing PCA were clustvis online open software and CAMO unscramble ×10.5 software which was available at LPU, Phagwara.

The near infrared (NIR) spectra were obtained from CSIO-CSIR, Chandigarh. Fourier transforms infrared (FTIR) spectra and Fluorescent spectra for all the groundwater samples were obtained from spectrometer facility at CIF, LPU, Phagwara. Principal component analysis is done using NIR spectral data to classify the groundwater on the basis of region, source of water and depth of the source of water. The FTIR spectra of the groundwater samples was used to compare the peaks obtained with the peaks of pure urea and arsenic trioxide peaks to know about the contamination of urea and arsenic in the collected groundwater samples. It is observed that the FTIR spectra of groundwater samples did not show peaks for Arsenic oxide and urea, thus there are very feeble chances of arsenic and urea in the samples.

Partial least square regression was done to develop a model for the qualitative determination of the physicochemical parameters. The spectral data was preprocessed to reduce the noise in the data for this smoothing of the data, derivative, normalization, baseline corrections, multi scattering correction etc. were done. PLS model was developed in the range of 700-2500 nm. The relationship between physicochemical parameters and wavelengths were checked from the score and loading plots of the PLS model. It was observed that some wavelength ranges were not playing any role in the prediction of the parameters and the value of the R^2 was very small. To overcome this problem, the total wavelength range was split in the intervals of 300 nm and regression model was developed for each split range of wavelengths and the value of R^2 and RMSE values were determined for calibration and validation and were compared for each model. Only those models were considered for further analysis for calibration and validation for which R^2 had high value and RMSE had very small value. The regression model having higher value of R^2 for selected wavelength range were saved for further use in the industry and laboratory.

The presented work focused on the use of spectra to find the bonds in groundwater related to the contaminants present in the groundwater, classification of the data on the basis of source of water, region and depth of source of water. The work further aimed at developing a regression model with selected wavelength ranges for the prediction of physicochemical parameters of groundwater form the Majha region of Punjab. In the present work wavelength ranges were successfully selected which could predict the parameters of groundwater from the Majha region. The model will make the determination of the quality

parameters a cost effective and time saving process, the time consumed in the analysis and experiments will get reduced by the use of model. The cost of the instrumentation will also decrease since filters for only selected wavelengths are required instead of monochromators for whole NIR range from 700-2500 nm.

From the fluorescent data graphs are plotted and correlation is observed for physicochemical parameters at particular wavelengths or for a range of wavelengths. The correlation is observed for turbidity, pH value, chlorine content, electrical conductivity, total hardness and total dissolved solids. It can be concluded that fluorescent spectroscopy shows correlation for many of the parameters.

The parameters like pH value, total hardness, TDS, EC, turbidity, chlorine for the groundwater from Majha region of Punjab can be predicted using the developed regression model.

ACKNOWLEDGEMENT

understanding during my research work. Their prayers and good wishes for me was what sustained me this far. They supported and encouraged me at every step of my research work.

I would also bow with deep hearted respect and heartful thanks to my teachers: **Dr. R. Thangaraj** and **Dr. S. S. Sekhon**, both my teachers during my Post graduation days at Guru Nanak Dev University, Amritsar. This is all their teachings and blessings that I, still at 50+, dared to dive into the sea of knowledge.

Finally, I would like to thank all the persons whose names could not be given here but they have positively impacted my life and thinking. I would again like to thank The Supreme Power of God, and dedicate this work to The Supreme Energy and to my Late Parents.

YOGESH KUMAR (09/09/2025)

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BIS	Bureau of Indian Standards
Cl	chloride content
EC	Electrical conductivity
EM	Electromagnetic
FTIRS	Fourier Transform Infrared spectroscopy
IPLS	Interval Partial least Square
NIRS	Near Infrared Spectroscopy
NTA	Nephelometric turbidity unit
PCA	Principal Component Analysis
PCR	Principal Component Regression
PCs	Principal Components
PLS	Partial least Square
PLSR	Partial least Square Regression
r	Coefficient of correlation
R^2	Coefficient of determination
R^2_C	Coefficient of determination for calibration
R^2_V	Coefficient of determination for validation
$RMSE_C$	Root mean square error of calibration
$RMSE_V$	Root mean square error of validation
TDS	Total dissolves solids
TH	Total hardness
WHO	World Health Organization

··

CHAPTER 1

INTRODUCTION

··

CHAPTER 1 - INTRODUCTION

1.1 INTRODUCTION

Energy and matter are the main constituents of universe. Energy can exist in many forms like heat energy, sound energy, chemical energy, nuclear energy, mechanical energy etc. and one of the forms is electromagnetic radiation, in which we are interested here. The smallest form of matter is atom (in case of elements) and molecule (in case of compounds). When atoms and molecules were studied, a huge involvement of energy was found in the formation of these basic entities. It is found in due course of time that there are various types of phenomena which occur when energy and matter interact with each other. The simplest one is the splitting of white light (electromagnetic radiation or energy) into its constituent seven colours when this white light passes through a transparent triangular prism (matter). This phenomenon is known as dispersion of light and was first observed by Sir Isaac Newton in the year 1666 and he introduced the term spectrum for the band of seven colours so obtained. It is well known fact that red colour is least deviated whereas the deviation of the violet colour is maximum. Some of the other matter and electromagnetic energy interactions are- photoelectric effect which occurs when light of suitable frequency falls on suitable material, Compton Effect occurs when photon of light interacts with charged particle, ionization of the gaseous atoms or molecules take place when suitable energy interacts with the atoms or molecules.

The branch of physical sciences which deals with the interaction of matter with the electromagnetic radiations is known as spectroscopy. We observe and investigate the changes taking place in the nature and properties of the electromagnetic (EM) radiation after it has interacted with the matter. Before going into details of spectroscopy some facts about the EM radiations.

1.1.1 The Spectrum of Electromagnetic Radiations

The seven colours of the white light described by Newton form only a very small part of the known complete spectrum of electromagnetic radiations. To become well versed with the known complete spectrum, it becomes necessary to categorize the electromagnetic radiations on the basis of some property of the EM radiations. It is well known fact that speed (c) of EM radiations in vacuum is constant and is given by the equation

$$c = \nu\lambda \qquad\qquad\qquad (1.1)$$

Where, ν is frequency and λ is wavelength of the radiation. It was easy to categorize EM radiations on the basis of wavelength (or frequency[*]). The other parameter, which is used in place of wavelength (or frequency), to represent the electromagnetic radiation is, its wave number [1] which is represented by $\bar{\nu} = \dfrac{1}{\lambda_{vac}}$[†] and is measured in cm^{-1} or m^{-1}. The energy (E) associated with EM radiation is given by the formula

$$E = h\nu \qquad\qquad\qquad (1.2)$$

Where h is Planck's constant. From the above facts it is clear that

$$\text{Energy} \propto \text{frequency} \propto \text{wavenumber} \propto \frac{1}{\text{wavelength}} \qquad\qquad (1.3)$$

One can use any one out of the four above mentioned parameters to describe the EM radiations. The categorization of electromagnetic radiations on the basis of wavelength (or any other property) is known as the spectrum of the radiation. The known spectrum of EM radiations ranges from gamma radiations having wavelength less than 0.01 nm to radio waves having wavelength greater than 10^8 μm. The spectrum [2] is shown in Figure 1.1

While describing radiations falling in different regions of the EM spectrum, those units are chosen which are convenient to use in a particular situation. For example radiations in radiofrequency and microwave regions are described in terms of frequency in hertz (Hz) and for higher energy radiations wavelength in nm is used.

It is customary to describe radiations falling in the infrared region by wavenumbers in cm^{-1}. This becomes helpful in those cases where the peaks are very close to each other e.g. 36000 cm^{-1} and 36010 cm^{-1}, the difference can be conveniently written as 10 cm^{-1}. The corresponding wavelengths (frequencies) are 277.77 nm (1.0800×10^{15} Hz) and 277.70 nm (1.0803×10^{15} Hz) respectively and the difference is only 0.07 nm (0.0003 Hz) and the magnitude is too small to be considered and in many cases such small differences are usually ignored, but the corresponding difference 10 cm^{-1} (the numeric value) is not that small. So it is convenient to express EM radiations falling in infrared region in units of cm^{-1}.

---

[*] Frequency is given by $\nu = \dfrac{c/\mu}{\lambda/\mu}$, μ is absolute refractive index of the medium in which EM radiation is passing, frequency of given monochromatic EM radiation in any medium remains constant and wavelength changes in medium according to the relation $\lambda_{med} = \dfrac{\lambda}{\mu}$, velocity in medium also changes and becomes $\dfrac{c}{\mu}$.

[†] wavenumber is calculated for the given EM wave by considering the wavelength in vacuum i.e. λ_{vac}
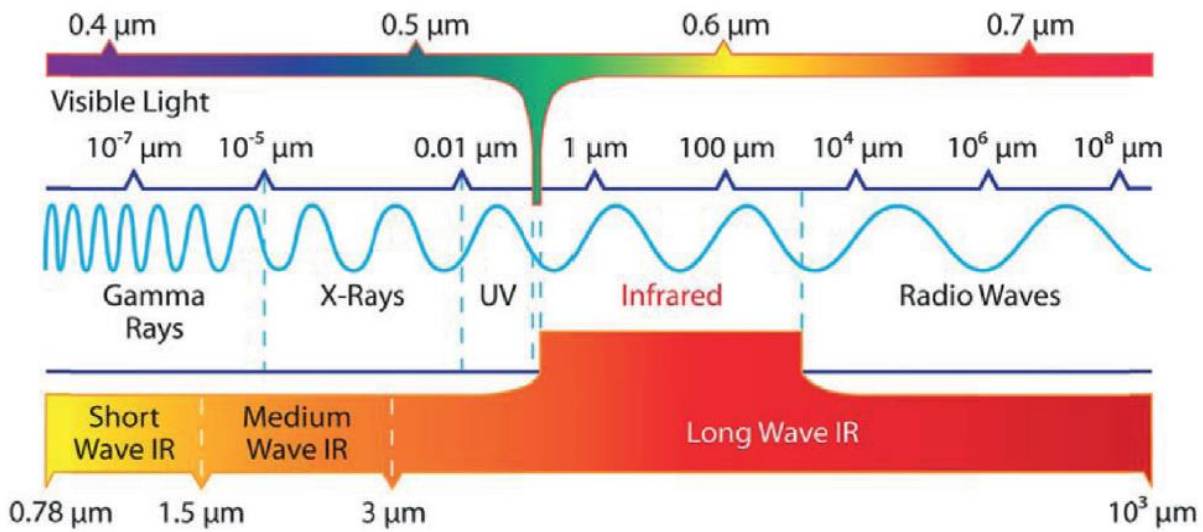
Figure 1.1 Electromagnetic spectrum (source Dr. Cathal Wilson and Dr. Gerard McGranaghan) [2]

The units can be converted into each other by the criteria given below:

$$1\,m = 100\,cm = 10^6\,\mu m = 10^9\,nm = 10^{10}\,Å$$

The equivalency among wavelength (randomly chosen 100 nm), wavenumber and frequency units is given below:

$$100\,nm \equiv 10^7\,m^{-1} \equiv 3{\times}10^{15}\,Hz \quad \text{(Showing equivalency and not equality)}$$

Corresponding energy units for 100 nm radiation are

$$100\,nm \equiv 2{\times}10^{-18}\,J = 12.4\,eV$$

As an example an infrared radiation of wavenumber 333 m^{-1} can be expressed in the equivalent frequency, wavelength and energy units as given below:

$$3{\times}10^6\,nm \equiv 333\,m^{-1} \equiv 1.00{\times}10^{11}\,Hz \equiv 0.0004\,eV$$

The EM spectrum is divided into various regions, which are briefly described here [3].

- *Gamma rays:* These rays have the highest energy (or frequency) and find their use in the medical field. These rays have wavelengths less than 0.01 nm.

- *X-rays:* These rays also have high energy and are used in medical fields and in other fields like engineering and by airport authorities to check the luggage of the passengers for any sort of illegal activity etc. The range of wavelengths lies between 0.01-10 nm

- *Ultraviolet region:* The range of wavelengths lies between 10 - 400 nm. These radiations find their use in sterilization in the medical field etc.

- *Visible region:* The range of wavelengths lies between 400 - 780 nm. This is the White light we get from the sun and find its uses in daily life.

- *Infrared region:* The range of wavelengths lies between 780 nm - 300 µm (and 13000 - 33 cm^{-1}). It is convenient to express this region in terms of wavenumbers. These radiations find their application in the medical field because these are responsible for the heat in the sun shine and objects at high temperatures also emit these radiations. These are also used in night vision instruments. This region is subdivided into the following regions (regions are described in terms of wavenumbers).

  Near infrared region (NIR): 3333 cm^{-1} to 13000 cm^{-1}(0.4 eV to 1.6 eV)

  Mid infrared region: 333 cm^{-1} to 3333 cm^{-1}(0.04 eV to 0.4 eV)

  Far infrared region: 33 cm^{-1} to 333 cm^{-1} (0.004 eV to 0.04 eV)

- *Microwave region:* The range of wavelengths lies between 300 µm to 10 cm. The range in terms of frequency is 1 THz to 3 GHz.

- *Radio frequency region:* The range of wavelengths lies between 10 cm to 100 m. The range in terms of frequency is 3 GHz to 3 MHz.

1.2 SPECTROSCOPY

Based on the basic constituents of the matter i.e. atoms and molecules, spectroscopy can be divided into two branches which are

1.2.1 Atomic spectroscopy

1.2.2 Molecular spectroscopy.

When electromagnetic (EM) radiations of suitable frequency (or energy) interact with the deserving atoms and or molecules, there may be shift in rotational energy levels or in rotational as well as vibrational energy levels or vibrational energy levels only and or there may be shift in electronic energy levels etc. in the atoms or molecules. Lower energy radiations (far infrared and microwave region) can produce only a shift in rotational energy levels, little higher energy radiations (near infrared region) can produce a shift in rotational as well as vibrational energy levels, and higher energy radiations (visible and ultra violet region of the EM spectrum) can produce a shift in electronic energy levels.

1.2.1 Atomic spectroscopy

When a suitable amount of energy is absorbed by an atom, its electrons gain energy and jump to higher energy level. The atom is said to be in excited state, but de-excitation happens in approximately 10 ns. In this process of de-excitation, electrons jump back to

lower energy level and release some energy in the form of EM radiations, this released energy is equal to the difference of the energies of the two energy levels involved. This energy in the form of EM radiations, forms the basis of atomic spectroscopy and is helpful in finding the structure of the atoms and is also helpful in the recognition of the atoms, e.g. elements in the solar atmosphere were found from this type of spectroscopy. It is observed and now well established fact that the atomic spectrum is in the form of lines.

1.2.2 Molecular Spectroscopy

Molecules may have two atoms or more atoms bound with each other chemically. These atoms may be of same element (e.g. H_2, Cl_2, O_3 etc.) or of different elements (e.g. CH_4, CO_2, NH_3, H_2O etc.). Depending on the amount of energy absorbed by the molecules they may show rotational, vibrational and electronic transitions. It is observed that the spectrum shown by molecules is band spectrum, each band is composed of many lines placed very-very close to each other, these lines in the bands can only be observed using instruments having very high resolving power.

When EM radiations are from the far-infrared region, pure rotational bands are observed, when the radiations are from the near-infrared (NIR) region then rotational-vibrational bands are observed (vibrational transitions are superposed by the rotational transitions) and when EM radiations are from the visible and ultra-violet region the electronic bands are observed which are superposed by the vibrational and rotational transitions.

Table 1.1 Types of spectroscopy across the electromagnetic spectrum [3 - 4]

Region	Spectroscopy
Radio frequency	Nuclear magnetic Resonance
Microwave	Electron spin Resonance
	Molecular (pure) Rotational
Infrared	Molecular Vibrational*
Ultraviolet-visible	Electronic
X-ray	Inner shell Electronics
Gamma ray	Mössbauer (nuclear internal structure)

*Rotational transitions commonly superimpose themselves onto vibrational

Dipole moment: Before further discussion, definition of dipole moment is given here, because it has important role to play in the spectroscopy. Dipole is an entity consisting of two equal and opposite charges bound together but separated by a small distance. Some of the

bonds in a molecule may have permanent dipole, and some of the bonds start showing dipole moment due to some vibrational mode like stretching or bending.

Dipole moment =one of the charge × distance between the two charges.

The direction of dipole moment is from positive to negative in case we are dealing with chemistry and opposite while dealing with Physics. The SI unit of dipole moment is coulomb meter (Cm).

1.3 TYPES OF MOLECULAR SPECTRA

Molecular spectra can be of three types depending on the energies of the electromagnetic radiations involved.

1.3.1 Pure Rotational Spectra

The energies associated with the radiations from the microwave and far infrared region are of the order of 0.005 eV. These radiations can cause transitions in rotational energy levels of the molecules and thus in this case we get pure rotational spectra. Only those molecules which have permanent dipole moment can show pure rotational spectra. Homonuclear molecules like H_2, N_2 etc. and symmetric molecules like CO_2, CH_4 etc. do not show this type of spectra because they do not possess permanent dipole moment, only the heteronuclear polar molecules like HCl, H_2O, NH_3 etc. can show this type of spectra.

1.3.2 Vibrational-Rotational Spectra

This is mainly absorption type spectra and is observed when the EM radiations are from NIR region. The energy of the NIR region is not high enough to excite the electrons to higher energy level. The molecule, while remaining in the same electronic state, will only show the vibrational and rotational transitions if it absorbs the near infrared EM energy. Energy changes involved are of the order of 0.01 eV. This energy is very large as compared to the energies required for molecules to show rotational spectra i.e. why rotational spectra are also obtained along with vibrational spectra. This type of spectra is obtained when there is a change in the dipole moment of molecule due to some mode of vibration like stretching or bending. Some hetero-nuclear polar molecules (like HCl, H_2O etc.) which have permanent dipole moment can show this type of spectra. Some polyatomic molecules (like CO_2, H_2O etc.) also show vibrational spectra, because the dipole moment changes, when these molecules vibrate.

The vibrations can occur by stretching (angle between the atoms do not change but the bond length changes) and bending (angle between atoms may change) of a molecule. The stretching can be symmetrical (e.g. in case of water molecule, oxygen atom remains almost fixed and both hydrogen atoms move away (or come close) from the oxygen atom simultaneously) and anti-symmetrical (one of the hydrogen atom moves away while the other hydrogen atom move close to the oxygen atom simultaneously). Bending can occur in four ways viz. scissoring (both hydrogen atoms come close to each other and move away from each other while remaining in the same plane, oxygen atom remains almost fixed), rocking (both hydrogen atoms bend in same direction), wagging (both hydrogen atoms move out, in same direction from their equilibrium plane) and twisting (both hydrogen atoms move out, in opposite direction from their equilibrium plane) [5].

Figure 1.2 Showing transitions in a diatomic molecule, the spacings shown are not real (source: John R. Ferraro and Kazuo Nakamoto) [6]
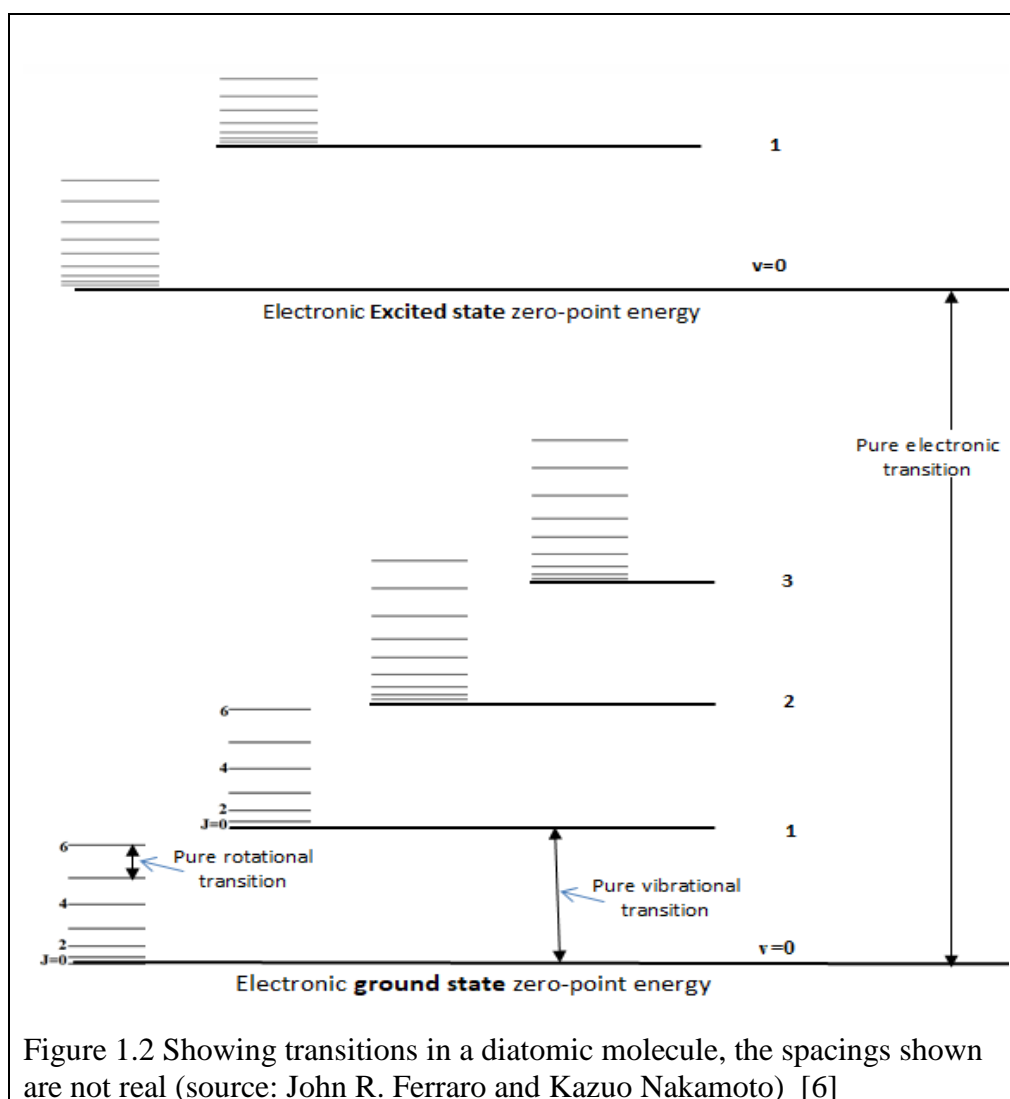
1.3.3 Electronic Spectra

The electronic spectra of a molecule arise when the energy involved is of the order of 1-10 eV. This energy is associated with the EM radiations falling in the ultra violet and visible region of the EM spectrum. When this energy is absorbed by molecule the electronic transitions take place from the lower energy level to the higher energy level. The energies involved are very large as compared to the energies required for rotational and vibrational transitions, therefore vibrational and rotational transitions accompany the electronic transitions.

1.4 THEORY OF NEAR-INFRARED SPECTROSCOPY

As already mentioned, vibrational and rotational spectra of molecules arise, when energy from NIR region of the EM spectrum is absorbed by the molecules. The energy is sufficient for the molecules (which have permanent dipole moments) to make transitions from low vibrational energy levels to higher vibrational energy levels. These molecules exhibit vibrational absorption spectra. According to classical Physics, if a vibrating molecule possesses an oscillating dipole moment, it can emit EM radiations falling in the near-infrared region, and the reverse can also happen. When the EM radiations from the near-infrared region interact with a molecule, only those vibrating bonds of the molecule will absorb specific frequencies which have dipole moment. It is observed that the spectrum obtained in this case has an intense band known as fundamental band (having wave number $\bar{\nu}$) and there are other less intense or weak bands which are known as overtones. The overtones are found at wavenumbers $2\bar{\nu}$, $3\bar{\nu}$ …… etc.

1.4.1 Molecule as Harmonic Oscillator

The vibrations in a molecule are considered as similar to stretching and compression of a spring [7]. Consider a diatomic molecule with atoms of masses m_1 and m_2 respectively, bound with each other through a chemical bond, the bond is assumed to act as a spring (Figure 1.3).

The equilibrium bond length is r_e and let r be the distorted bond length of the oscillating molecule. The oscillating molecule obeys Hook's law, which can be written as

$$f = -k(r - r_e)$$

<div align="right">(1.4)</div>

Where f is known as restoring force and k is known as the force constant (also known as spring constant). This law is obeyed only for small oscillations. The potential energy for such an oscillator is parabolic in nature and the formula for potential energy $V(r)$ is written as

Figure 1.3 Oscillating diatomic molecule

$$V(r) = \frac{1}{2}k(r - r_e)^2 = \frac{1}{2}kx^2 \qquad (1.5)$$

Here, $x = (r - r_e)$ is displacement coordinate. The oscillation frequency according to classical physics is given by

$$v = \frac{1}{2\pi}\sqrt{\frac{k}{m'}} \qquad (1.6)$$

Here m' is the reduced mass given by the relation

$$m' = m_1 m_2 / (m_1 + m_2) \qquad (1.7)$$

The energy of such an oscillator according to quantum physics is given by

$$E_v = \left(v + \frac{1}{2}\right)hv = \left(v + \frac{1}{2}\right)hc\bar{v} \qquad (1.8)$$

Where v is vibrational quantum number and can have values = 0, 1, 2, 3…….etc. and 'v' is frequency and '$\bar{v}$' is wave number, 'h' is plank's constant and c is speed of light. Equation (1.8) will give energy value in units of joule and it can be expressed in units of cm^{-1} and the expression is known as term value '$G(v)$' and can be written as [9]

$$G(v) = \frac{E_v}{hc} = \left(v + \frac{1}{2}\right)\bar{v} \qquad (1.9)$$

Figure 1.4 Potential energy curve and vibrational energy levels [8]

When the oscillator is in ground state of vibration, then the vibrational quantum number has a value v = 0, the energy corresponding to this value (equation (1.8)) is known as zero-point energy and is given by

$$E_0 = E_{v=0} = \frac{1}{2}hv \qquad (1.10)$$

The vibrational transitions in a harmonic oscillator can take place only if the selection rules

$$\Delta v = \pm 1 \qquad (1.11)$$
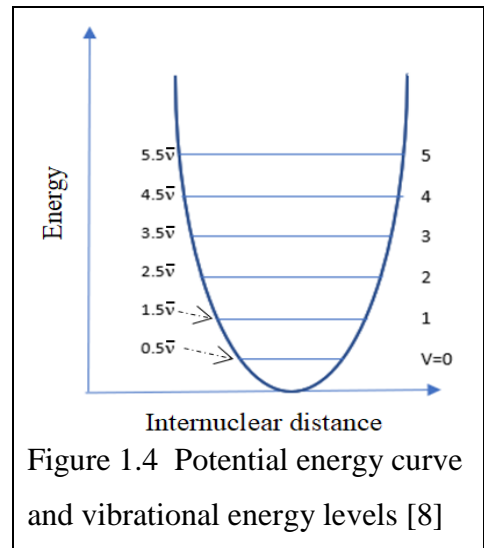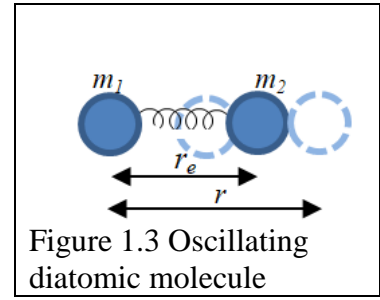
is followed. The molecule can absorb or emit energy so that there is a change in energy (associated with vibrational levels) only by the amount $\Delta E_v = h\nu$ at a time, i.e. all the vibrational energy levels are equally spaced, this is depicted in Figure 1.4. The transition from the level $v = 0$ to $v = 1$ is known as fundamental transition and other allowed transitions (following equation (1.8)) from vibrational level 1 to 2, 2 to 3,…. etc. are known as hot bands. It is observed that the frequency remains same in hot bands and is equal to the frequency of fundamental transition. It is also observed that intensity of the hot band transitions is normally low at room temperature as compared to the fundamental transitions, so more importance is given to the fundamental transitions.

1.4.2 Molecule as Anharmonic Oscillator

Any real molecule cannot be a simple harmonic oscillator and therefore real molecules do not obey Hook's law. If atoms in a molecule are stretched beyond a certain distance, the bond may break and the molecule dissociates (dissociation energy 'D_e' is shown in Figure 1.5 and loses its identity.

Figure 1.5 Anharmonic potential for a diatomic oscillator (shown with parabolic harmonic potential approximation for comparison)
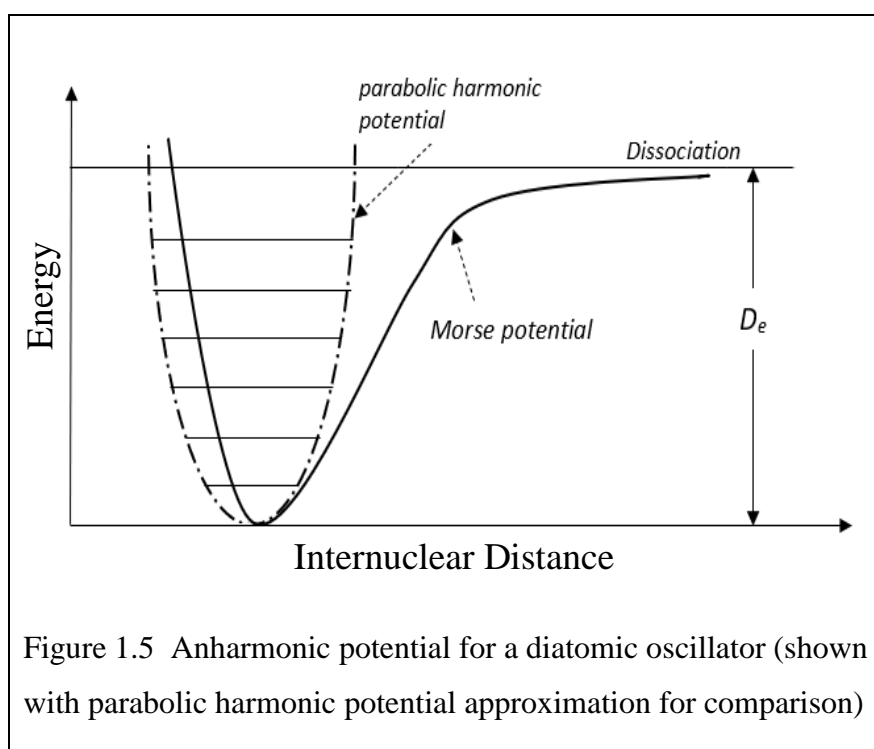
An anharmonic oscillator [9] is different from simple harmonic oscillator in the following ways:

- The vibrational energy levels are not equally spaced, therefore hot bands will have the different frequency from the fundamental band frequency.
- The selection rule given in equation (1.11) is not valid and transitions from $v = 0$ to $v = 1, 2, 3, 4\ldots$, are possible, i.e. overtone transitions are allowed, the selection rule can be $\Delta v = \pm 1, \pm 2, \pm 3 \ldots$ etc.

There are two effects which are responsible for anharmonicity of a real molecule, first one is mechanical and the other is electrical.

The potential energy term of a harmonic oscillator is quadratic (equation (1.5)) but the potential energy term of an anharmonic oscillator can have cubic, biquadratic and higher power terms as shown in the expression below (mechanical effect)

$$V = \frac{1}{2}kx^2 + k'x^3 + k''x^4 + \ldots \ldots \qquad k', k'', \ldots \ll k \qquad (1.12)$$

An empirical relation was derived by P.M. Morse which fits the anharmonic potential curve shown in Figure 1.5 and is called the Morse potential

$$V = D_e\left(1 - e^{-\beta x}\right)^2 \qquad (1.13)$$

Where, D_e is dissociation energy, x is displacement coordinate and β is a constant for the given molecule.

The electrical effect for anharmonicity arise from the fact that dipole moment (μ) of the molecule also has the square and higher power terms in its mathematical expression given by [9 - 10]

$$\mu = \mu_e + \left(\frac{d\mu}{dx}\right)_e x + \frac{1}{2!}\left(\frac{d^2\mu}{dx^2}\right)_e x^2 + \ldots \ldots \qquad (1.14)$$

Where, subscript 'e' is used for equilibrium configuration. This electrical anharmonicity is responsible for overtone transitions.

The term value $G(v)$ in case of anharmonic oscillator is given by

$$G(v) = \frac{E_v}{hc} = \bar{v}\left(v + \frac{1}{2}\right) - x_e\bar{v}\left(v + \frac{1}{2}\right)^2 \qquad (1.15)$$

Where $x_e\bar{v}$ is always positive and is known as anharmonic constant, it has a very small value as compared to $\bar{v}$.

1.5 INSTRUMENTATION FOR NIRS

Near Infrared Spectroscopy (NIRS) is an analytical technique used to measure the concentration of specific chemical compounds in a sample based on their near infrared (NIR) absorption or transmission characteristics. The instrumentation of NIRS typically consists of the following components:

a) *Light source*: NIRS uses a light source that emits NIR radiation, such as a tungsten halogen lamp or a super luminescent diode.

b) *Monochromator*: This component filters the NIR radiation and separates it into individual wavelengths, allowing for spectral analysis. The monochromator in NIRS instrumentation is typically a grating monochromator, which uses a diffraction grating to separate the light into individual wavelengths.

c) *Sample cell or cuvette:* The sample is placed in a cell or cuvette for analysis. The sample cell or cuvette can be made of materials such as quartz, plastic, or glass, depending on the sample and measurement requirements.

d) *Detector:* A photodiode or CCD detector is used to measure the intensity of the transmitted or absorbed NIR radiation.

e) *Data acquisition and processing software:* The NIR spectra are collected and processed using specialized software, which can perform a variety of data analysis functions, including peak detection, baseline correction, and spectral analysis. It can also perform advanced analysis and data processing functions, such as principal component analysis and partial least squares regression.

f) *Computer:* The NIR data is stored and analyzed on a computer system.

g) *Display screen:* The NIR spectra and results of the analysis are displayed on a computer screen.

Some NIRS instruments also include additional components, such as integrating spheres or diffuse reflectance accessories, to improve the measurement of the sample reflectance.

In summary, the instrumentation of NIRS typically consists of a light source, monochromator, sample cell, detector, data acquisition and processing software, computer, and display screen.

There are two basic instrument designs based on reflectance and transmittance for NIR spectroscopy, which are shown in Figure 1.6. NIRS can be performed using either transmission or reflectance modes.

In transmission mode, the NIR radiation passes through the sample and the intensity of the transmitted light is measured. In reflectance mode, the NIR radiation is reflected from the surface of the sample and intensity of the light reflected is measured.

In case of reflectance, it is observed that for a grounded sample the NIR light can only penetrate up to 1×10^{-3} m to 4×10^{-3} m of the sample surface. When reflectance technique is compared with the transmittance technique, a great variation is observed because of small penetration of NIR light into a non-homogeneous sample. When transmittance technique is used infrared light travels through whole of the length of the sample. Spectral measurement is done through whole of the sample, therefore errors, which may arise otherwise due to non-homogeneity of the sample, are reduced to large extent.

Figure 1.6 Block diagram of Near Infrared spectroscopy [11]
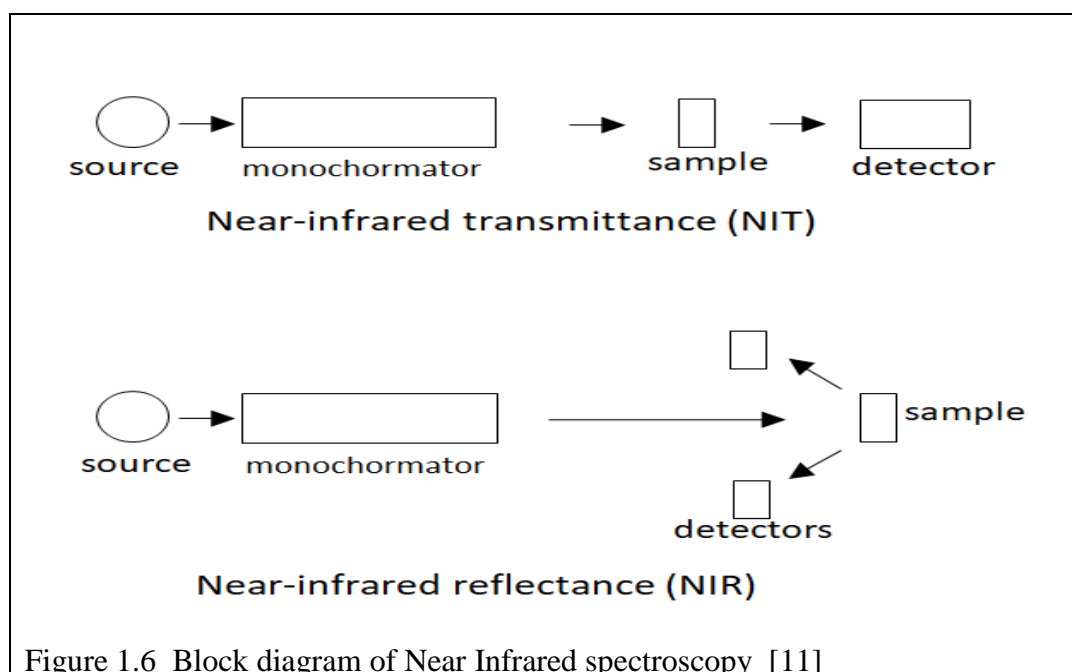
The transmittance technique is found to be very useful in those cases where sample consists of large sized particles. While using transmittance technique it is required to maintain a balance among the three viz. high frequency energy (800 - 1400 nm) used, path length of the sample and front surface scattering. When high frequency energy is used as compared to the low frequency energy it is observed that chances of scattering are high at the front surface for a sample consisting of small sized particles. For such a sample, detector may not be able to record a signal, because the amount of energy transmitted is very small in this case. To take into consideration both transmittance and reflectance, the instrument is designed in such a manner that it has both the capabilities.

1.6 FOURIER TRANSFORM INFRARED SPECTROSCOPY (FTIRS)

In earlier spectrometers, frequencies were selected using the phenomenon of dispersion of light by passing light through gratings, and this selected frequency is passed through the given sample to get spectrum. For a sample, consisting of atoms or molecules with only two energy levels E_1 and E_2, the frequency $(\nu = E/h)$ required for the transition
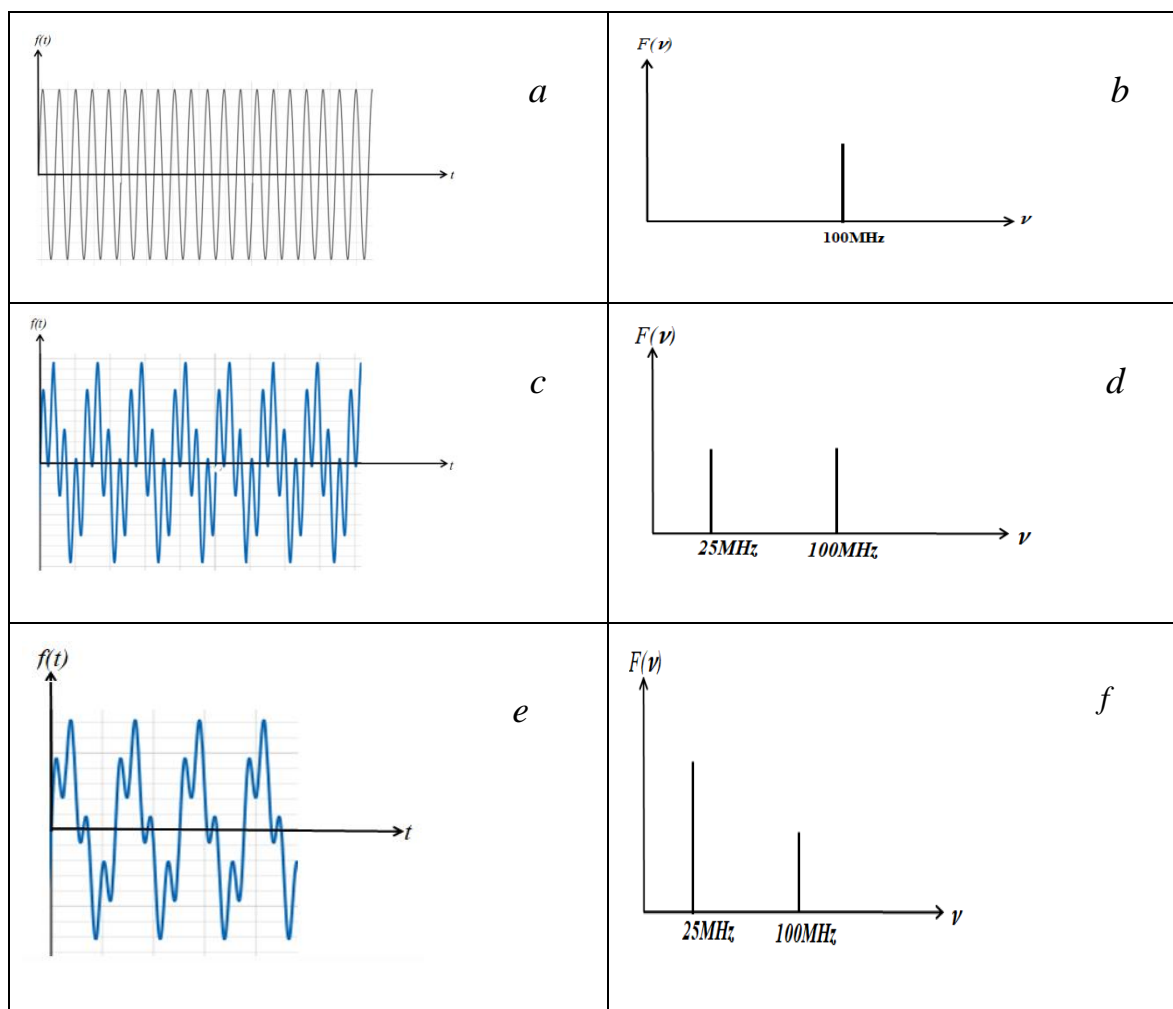
Figure 1.7 showing time domain (a,c,e) and transformed frequency domain (b,d,f), in fig (b), only one frequency is there and in fig (d) waves having two frequencies of same amplitude and in fig (f) the two frequencies have different amplitude. The numbers shown in frequency domains are arbitrarily chosen. (Note: The plots are not upto scale just indicative of the idea.)

between these levels is given by the formula, $\nu_t = \frac{|E_1 - E_2|}{h}$. This frequency is recorded by the detector and a peak is obtained in the spectrum, the output of the detector will show no

absorption for all other frequencies but shows a sudden absorption peak at this particular frequency ν_t.

In reality, the molecules or atoms have many energy levels and different frequencies are needed for the different transitions. The radiations interacting with the sample are also polychromatic and we have to make it monochromatic by some process so that a particular transition can take place. This is achieved by repeating the whole process again and again to get different frequencies one by one. The whole spectral data for all the frequencies are obtained after a time consuming tedious process which has to be repeated many times. The spectrum so obtained is in frequency domain. To make the whole process complete in very short span of time, spectrum is obtained in time domain and is converted into frequency domain using Fourier transforms, Fourier transforms are used to transform a function of x (where x is some variable) into a function of $1/x$ and vice versa. It is well established fact that frequency (ν) is equal to the reciprocal of time period ($\nu = 1/t$), so we can easily apply the Fourier transforms to convert time domain spectrum to frequency domain spectrum and vice versa. When this concept is applied to Infrared spectroscopy it is known as FTIR spectroscopy. (Figure 1.7 (a b c d e f)) [10].

light beam in

Interferometer splits beam into

two beams which will travel

paths D_1 and D_2

and then they interfere due to this path difference $|D_1 - D_2|$

light beam out
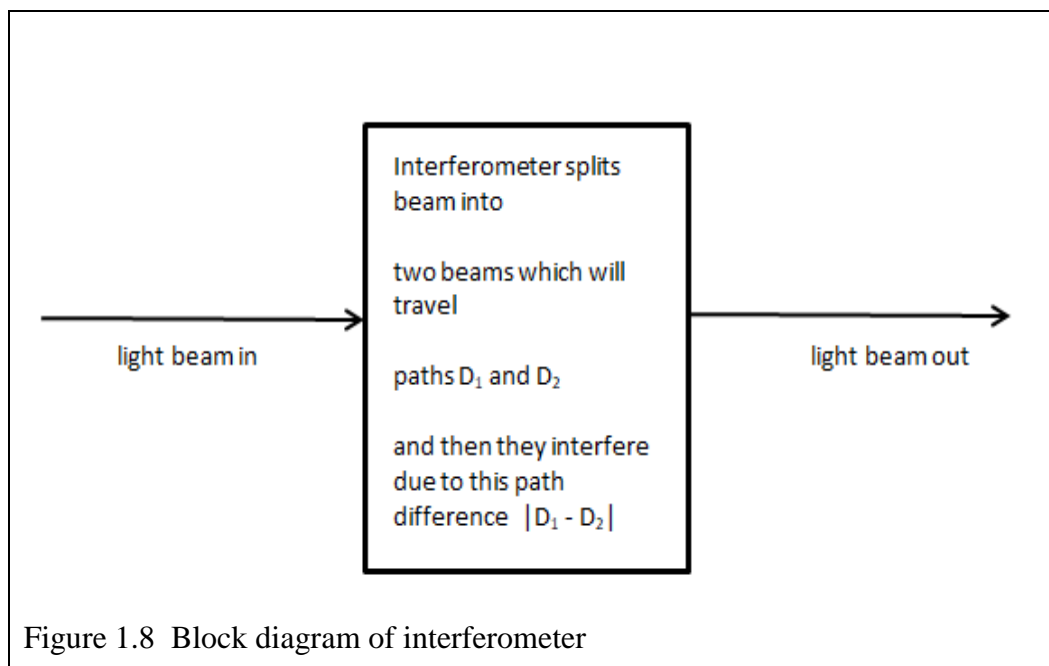
Figure 1.8 Block diagram of interferometer
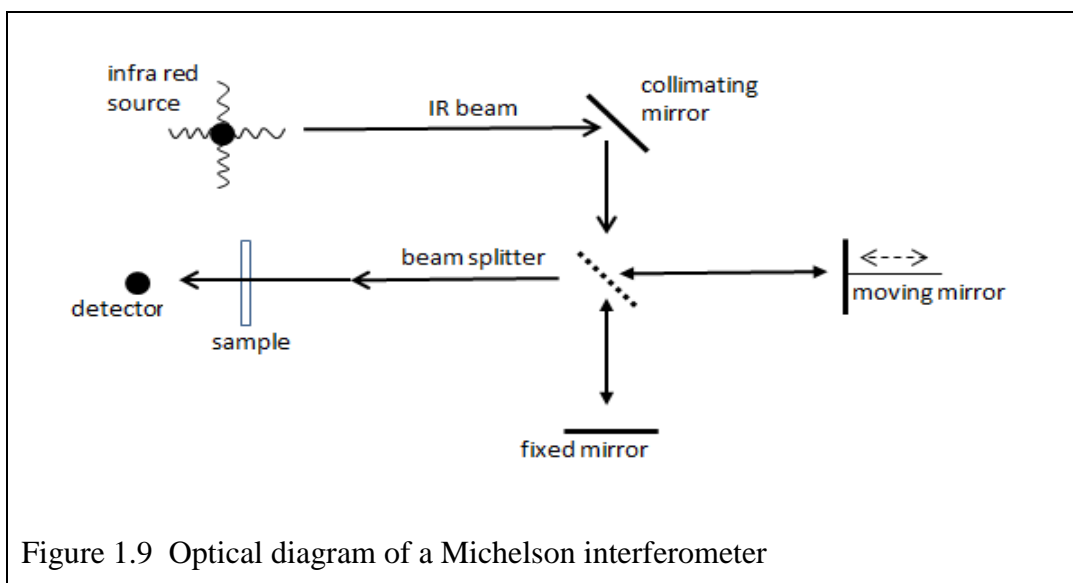
Figure 1.9 Optical diagram of a Michelson interferometer
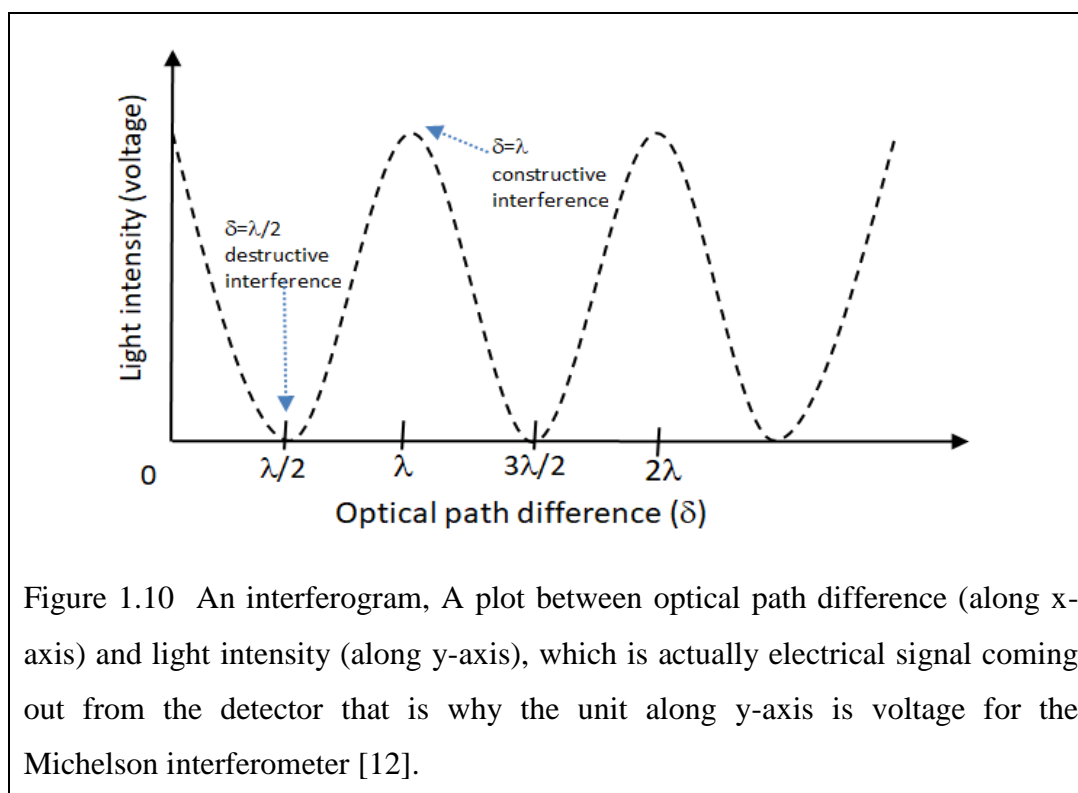
Figure 1.10 An interferogram, A plot between optical path difference (along x-axis) and light intensity (along y-axis), which is actually electrical signal coming out from the detector that is why the unit along y-axis is voltage for the Michelson interferometer [12].

FTIRS requires a device known as interferometer, the design of the interferometer is based on the Michelson's interferometer which he built in 1880s. A block diagram of the interferometer [12] is given in Figure 1.8 and internal design is given in Figure 1.9. It consists of a collimating mirror which collects light from the source, reflects it into the beam-splitter which further transmits and reflects light towards fixed mirror and moving mirror (the mirror can move left and right). These mirrors are perpendicular to each other. Beams get reflected back from these mirrors and after passing through the beam-splitter the beams recombine to

interact with the sample and pass through it and are detected by the detector. When the two beams recombine (before superimposing on each other they have travelled paths D_1 and D_2 and depending on the optical path difference $\delta = |D_1 - D_2|$ the interference takes place) in-phase (i.e. $\delta = n\lambda$, n = 0,1,2,3….; λ = wavelength) they give rise to constructive interference and when they recombine in out-of-phase (i.e. $\delta = (n+1/2)\lambda$, n = 0,1,2,3….) they give rise to destructive interference.

If path difference δ becomes zero i.e. the two interfering beams are identical, it is clear that constructive interference will take place and this condition is known as zero path difference (ZPD) condition. It is also known fact that intensity is proportional to the square of the amplitude.

The plot between optical path difference (along x-axis) and intensity (along y-axis) of the beam exiting the interferometer is known as interferogram, which happens to be the fundamental measurement, we obtain from an FTIR spectra (Figure 1.10). The light intensity is measured in voltage since it is the electrical signal obtained from the detector.

Using Michelson interferometer the interferogram is obtained by moving the mirror back and forth at least once. This is known as scan. The measured interferograms are then Fourier transformed to obtain the spectrum that is why it is named as Fourier Transform Infrared (FTIR) spectroscopy.

1.7 APPLICATIONS OF NIR:

NIR finds applications [13] in many disciplines like food, pharmaceutical, agriculture, in chemistry for finding various parameters of compounds, biomedical, astronomy [14] etc.

a) *Food Industry:* Karl Norris suggested in 1968 that NIR spectroscopy can be used to know the moisture and protein in grains and protein, moisture and oil in soya bean.

b) *Environmental Monitoring:* NIRS can be used in the field of environmental monitoring. It can be combined and used with the fiber optics to get knowledge of organic compounds which are volatile and present in various waste water sources like mud, sediments etc.

c) *To Know Makes of Products*: NIRS can be used to know the different makes of the same product, it can be used in quality control.

d) *Pharmaceutical Industry*: NIR can be used in pharmaceutical industry for the analysis of raw material, polymorphism of drugs and isomeric purity of optically active substances present in it.

e) *Polymer Industry*: NIRS can be used to analyze the packing material, pellets, laminates etc. It can be used to get information about reaction mechanism, crystallinity, water content etc.

f) *Biomedical Application:* NIRS can be used to monitor the growth of fertilized eggs and it can be combined with imaging technology to get visual information about egg development in a non-destructive way.

1.8 FLUORESCENCE SPECTROSCOPY

It is a spectroscopic technique [15] in which ultraviolet (UV) or visible light is made to interact with the matter. Electrons in the lower energy state absorb the incident EM radiations and jump to the higher energy state (excitation, it takes time of the order of 10^{-15} s). In a time of the order of micro second to nano second the electrons jump back to the lower energy state and in this process, some energy is emitted, which can be equal to the absorbed energy or can be of lesser energy than the absorbed energy (i.e. emitted energy may have higher wavelength than the incident radiation). This absorbed energy and emitted energy forms the basis of excitation and emission fluorescence spectroscopy.

The de-excitation which takes place by emitting light is termed as luminescence, which is further categorized into two types - fluorescence and phosphorescence. In this type of emission the emitted energy is of lower value (higher wavelength) than the value of the absorbed energy (lower wavelength). Fluorescence emission is very rapid as compared to phosphorescence. The time taken in fluorescence emission is of the order of 10 μs, whereas time taken in the process of phosphorescence may be in minutes or hours. The process of Fluorescence is more important than phosphorescence.

1.8.1 Atomic Fluorescence

When atoms in the gaseous form can absorb radiations of the same energy as required to electronically excite the atom in question, the atom may show fluorescence. When the emitted light has the same energy as the energy of the absorbed light then this is known as resonance fluorescence. This type of fluorescence is possible in sodium atoms.

1.8.2 Molecular Fluorescence

Excited molecules (by absorbing energy) can relax to their ground state by various mechanisms but the two important mechanisms are nonradiative relaxation and fluorescence emission.

The excitation takes from singlet state (S_0) to singlet state (S_1) as shown in Figure 1.11. The molecule when in excited state can show non-radiative relaxation by any of the process vibrational relaxation, internal conversion, intersystem crossing, quenching etc. It is observed that there is very small change in temperature due to these relaxations.

When relaxation takes place (in most of the cases) from higher excited state (but from the lower vibrational level in the excited state) to some higher vibrational level of the ground state then it is known as fluorescence. Vibrational relaxation and internal conversion takes place very rapidly as compared to fluorescence relaxation, in this case emission band spectrum is obtained with very closely spaced lines, which occur due to jumping back to different vibrational levels in the ground state from the higher energy state. The spectrum is plotted as fluorescence intensity versus wavelength or wavenumber.

Figure 1.11 (a) Jablonski diagram showing excitation and de-excitation process in a molecule, (b) showing electron spins for singlet and triplet states (c) showing absorbance and emission spectrum [16]
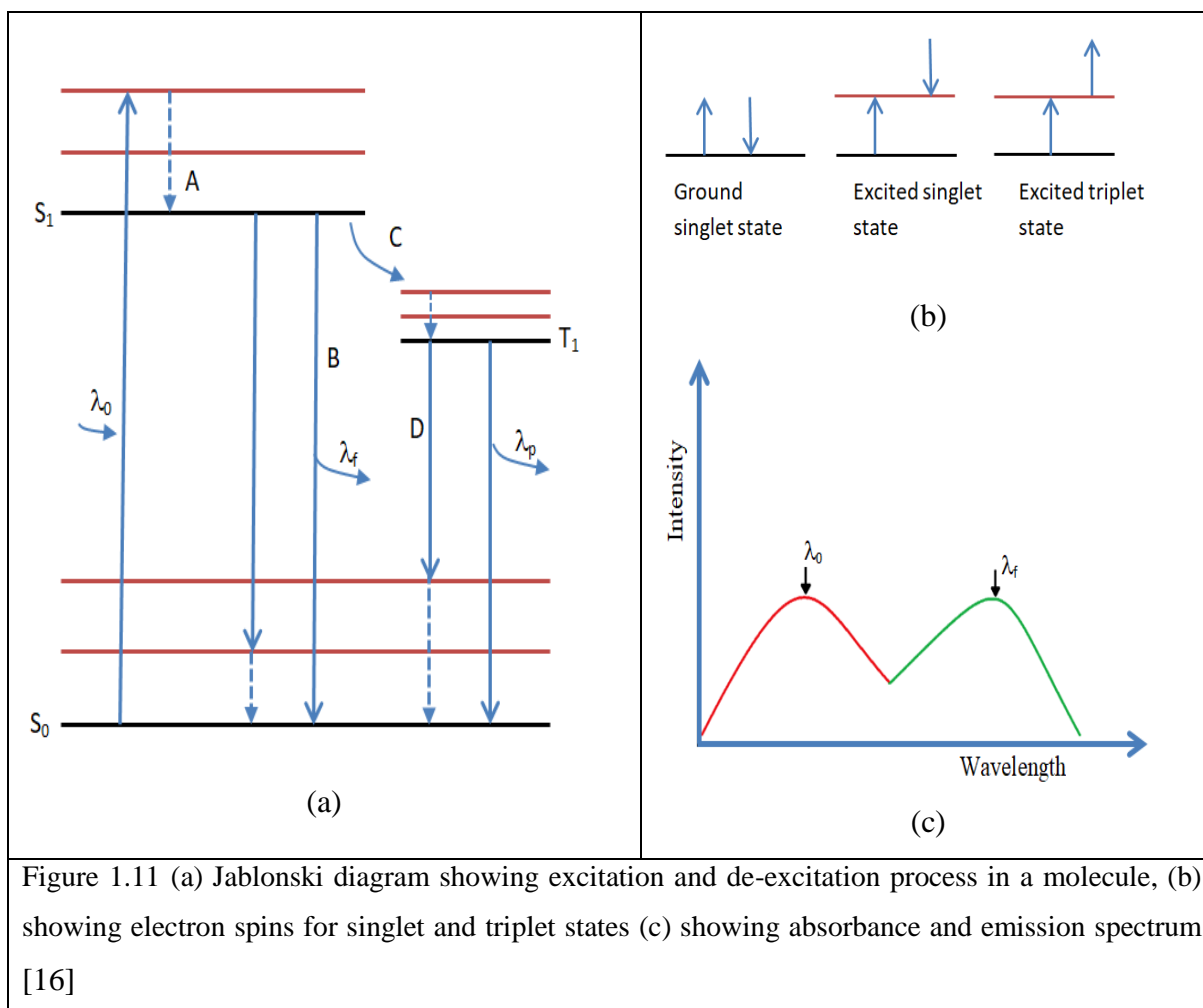
1.8.3 Applications of Fluorescence Analysis

There are various fields where fluorescence analysis finds application [17]. The various fields are– agriculture, drugs, food and food products, legal works, medical science, minerals and gems, museum work, organic chemistry, textiles, water and sewage etc.

1.8.4 Instrumentation for Fluorescence

The fluorescence spectroscopy consists of following instrumentations:

a) *Light source:* This can be Xenon arc lamp which can emit UV or visible radiations needed to excite the molecules in questions.

b) *Excitation monochromator:* This is needed to select the desired excitation wavelength from the source of light. It can scan a range of wavelengths also.

c) *Sample holder:* It holds the sample so that light can impinge on the sample, it is a special fluorescence cuvette having all the sides consisting of translucent quartz or glass.

d) *Emission monochromator:* It is helpful in analyzing the emitted light from the sample it selects the wavelength of interest or scans a range of emission wavelengths so that it can be analysed.

e) *Detector*: it is usually a photomultiplier tube (PMT), it act as detector and can measure the intensity of the emitted light.

f) *Filters:* Low band-pass filters, variable band-pass filters, or full-spectrum diode array spectrophotometers etc. can be used for the selection of desired emission wavelength.

1.9 GROUNDWATER

The water which is found beneath the Earth's surface is called groundwater it is found in spaces or pores within soil or rock, known as aquifers. Groundwater sources can be replenished by precipitation and surface water that infiltrates into the ground. Groundwater is often a critical source of drinking water and is also used for irrigation, industry, and energy production. Groundwater quality can be affected by human activities such as land use practices, waste disposal, and chemical spills, which can contaminate the water. Therefore, it becomes important to manage and protect groundwater resources. This is must to do for ensuring the sustainability of the resources for future generations.

Groundwater is a vital resource for many communities, especially in areas where surface water is scarce. It is one of the reliable resources of water because it is less

susceptible to the effects of drought and contamination compared to surface water sources. However, over-extraction of groundwater can lead to decreased water level and even the depletion of aquifers, which can take thousands of years to recharge.

To sustainably manage groundwater resources, it is important to understand the groundwater recharge rate, or how quickly the aquifer is being replenished. This information is used to set limits on how much water can be pumped from the aquifer without causing significant harm to the environment or future water supplies.

Groundwater is important in a way that it helps in lowering the impacts of floods by absorbing excess water and slowing down the rate of runoff. This can help to reduce the damage caused by flooding and improve the health of aquatic ecosystems.

Groundwater is a complex and dynamic resource, and its management requires a comprehensive approach that considers both the social and natural sciences [18]. For example, land use practices, such as deforestation or Urbanization, can alter the rate at which water infiltrates into the ground, affecting groundwater recharge rates and water quality. On the other hand, groundwater can also have impacts on surface water and the environment, including changes to stream flow and water levels, and the potential for subsidence or land collapse.

To sustainably manage groundwater resources, it is important to implement integrated water resources management approaches that consider the interconnections between groundwater, surface water, and the environment [19]. This includes conducting regular monitoring and assessment of groundwater quality and availability, as well as managing groundwater use through permit systems and other regulatory measures.

Additionally, community involvement and education are critical components of successful groundwater management. This includes educating communities on the importance of protecting groundwater resources and involving them in decision-making processes related to groundwater use and management.

Sustainable groundwater management requires a multi-disciplinary and collaborative approach that considers the needs of both human and natural systems, and balances the competing demands for this valuable resource.

In conclusion, groundwater is a valuable resource that requires careful management and protection to ensure its sustainability for future generations.

1.9.1 Physicochemical Properties of Water

Water is very important entity for sustaining life on earth and its consumption by living beings is as important as oxygen. If the groundwater or the drinkable water is polluted it can cause diseases in the humans and animals [20]. So it is important to test the water for quality parameters. Although there are many quality parameters specified by WHO and BIS but here only six of the physicochemical properties or parameters are considered for study due to cost and time factor.

1.9.1.1 Total Dissolved Solids (TDS)

Total dissolved solids (TDS) [21] refer to the salts (inorganic) and organic matter in small quantity present in water. These solids may be in the form of cations like calcium, sodium, magnesium, potassium etc. and in the form of anions like hydrogen carbonate, nitrate, chloride, sulphate etc. TDS is expressed in milligrams per liter (mg/L) or parts per million (ppm). TDS is an important indicator of water quality because it provides information about the presence of minerals, salts, and other substances that may affect the taste, color, and odor of water, as well as its suitability for various uses, like irrigation, drinking, and industrial processes.

The TDS level in water can depend on various types of factors like geology of the area, the presence of minerals and other substances and anthropogenic reasons like agriculture and waste disposal. High TDS levels can also indicate the presence of pollutants or contaminants that may be harmful to human health or the environment.

Table 1.2 TDS value of drinkable water and its rating or taste

TDS value	Rating/Taste
less than 300 mg/litre	Excellent
300 - 600 mg/litre	good
600 - 900 mg/litre	fair
900 - 1200 mg/litre	poor
greater than 1200 mg/litre	unacceptable

Taste of water may vary depending upon the quantity of dissolved solids in water. According to WHO guidelines, taste [22] of water is given in Table 1.2. The recommended TDS level in drinking water is typically less than 500 mg/L, although some agencies may set lower or higher limits based on local conditions. Irrigation water with a TDS level above

2,000 mg/L can be harmful to crops, while TDS levels in industrial water can vary depending on the specific use and requirements of the industry. If TDS value is very low, even then water is unacceptable due to its flat and insipid taste.

TDS levels can also vary greatly within a water source, depending on location and time of year. For example, TDS levels in groundwater may be highest during periods of drought, when the water table is low and the concentration of dissolved solids is higher. TDS levels in surface water may also vary seasonally, due to changes in precipitation, runoff, and evaporation.

It is important to regularly monitor TDS levels in water sources to detect changes and address any potential issues in a timely manner. In some cases, high TDS levels may indicate the need for treatment or further investigation to determine the source of the dissolved solids and address any potential health or environmental concerns.

In addition to its use as a water quality indicator, TDS can also provide important information about the geology and hydrology of an area. For example, TDS levels in groundwater can help to identify the presence of specific minerals or aquifer systems, and can be used to track changes in water levels and flow over time.

In conclusion, TDS is a useful water quality parameter and it provides information about the presence of substances that may affect the suitability and safety of water for various uses. TDS levels should be monitored regularly to ensure that water sources remain within safe and appropriate limits.

1.9.1.2 pH Value of Water

pH is related to the activity of hydrogen ions in a solution and is defined as negative of common logarithm of hydrogen ions present in the solution

$$pH = -\log[H^+] \tag{1.16}$$

The pH of a solution can have important implications for both health and the environment. WHO [23] has recommended that drinkable water should have pH values lying between 6.5 to 8.5, to ensure that it is safe for consumption and does not cause harm to plumbing or water treatment systems. Similarly, the pH value of surface water and groundwater can affect the health of aquatic ecosystems and the availability of nutrients for plants and other organisms. pH value of water does not show direct link to health because we also consume some foods which may have pH values in the range 2-3 like vinegar or citric juices etc. More acidic or basic water can easily corrode the pipes or vessels through which it passes or stored and can

help in adding impurities in the drinkable water. This way acidic or basic water can indirectly affect our health.

It's important to regularly monitor the pH value of water sources, including groundwater, to ensure that it is within a safe and appropriate range. If the pH value is outside of this range, it may indicate the presence of contaminants, such as acids or bases which could affect the water quality and safety. In these cases, additional testing and treatment may be necessary to ensure the water is safe for use.

In addition to its importance for human health and the environment, pH value also plays a crucial role in many chemical reactions and processes. For example, pH value can influence the solubility and toxicity of certain pollutants and contaminants, as well as the efficacy of treatment processes used to remove them from water. In some cases, changes in the pH value of water can also have significant impacts on infrastructure and industries. For example, low pH value levels can cause corrosion in pipelines and other metal structures, while high pH value levels can cause scaling in boilers and other industrial processes. It is also important to note that the pH of groundwater can vary depending on a variety of factors, such as the geology of the area, the presence of minerals and other substances, and human activities, such as agriculture and waste disposal.

In conclusion, pH value is an important indicator of water quality and has far-reaching implications for human health, the environment, and various industries. Regular monitoring of pH levels is necessary to ensure that water sources remain within a safe and appropriate range and to address any potential issues in a timely manner.

1.9.1.3 Chlorine in Drinking Water

Chloride (Cl) is an ion found in water, often present as chloride salts e.g. sodium chloride (NaCl). Chloride is an important water quality parameter because high levels can affect the taste, smell, and appearance of water and it also have environmental and health issues. Chlorine is a pungent smelling gas and is used as a disinfectant and also as bleach. In water it forms hypochlorous acid and hypochlorites [24]. Its acceptable limit in drinking water as stated by Bureau of Indian Standards is 250 mg/litre [25], although some agencies may set lower or higher limits based on local conditions. High levels of chloride in drinking water can have a salty taste and may contribute to the corrosion of plumbing and water treatment systems.

In the environment, high levels of chloride in surface water and groundwater can have negative impacts on aquatic ecosystems, altering the chemistry and decreasing the diversity

of plant and animal species. High chloride levels can also indicate the presence of pollutants or contaminants that may be harmful to human health or the environment.

Chloride levels in water are affected by a various types of natural and human activities, including the weathering of rocks and minerals, the use of salt for deicing roads, and the discharge of wastewater and other effluents.

In conclusion, chloride is an important water quality parameter that can affect the taste, odor, and appearance of water, as well as the health of aquatic ecosystems and the suitability of water for various uses. Regular monitoring of chloride levels is necessary to ensure that water sources remain within safe and appropriate limits and to address any potential issues in a timely manner.

1.9.1.4 Electrical Conductivity (EC)

Electrical Conductivity (EC) is ability of something to conduct an electrical current, and in case of water it is directly proportional to the concentration of salts dissolved in it and other ions in the solution. EC is expressed in units of conductivity, such as microsiemens per centimeter (μS/cm) or decisiemens per meter (dS/m).

EC is a widely used and versatile water quality parameter that provides important information about the presence of dissolved substances, such as salts, minerals, and pollutants, in water. The EC of water can vary greatly depending on the source and the presence of specific substances and can provide information that if water is suitable or not for various uses, such as drinking, irrigation, industrial processes etc.

In general, the EC of fresh water is low, typically less than 500 μS/cm, while the EC of saltwater can be much higher, ranging from 30,000 to 50,000 μS/cm or more. Pure water is good insulator and with increased ion concentrations, due to addition of some impurities, it becomes conducting. Electric conductivity therefore is a measure of ion concentration in water. According to WHO the acceptable limit of EC is 1500 μScm^{-1}.

The EC of groundwater and surface water can also vary depending on location, season, and other factors, such as the presence of pollutants or natural sources of minerals.

EC is commonly used in water resource management, agriculture, and environmental monitoring, as well as in various industrial processes, such as the production of semiconductors, pharmaceuticals, and food and beverages.

In conclusion, EC is an important and widely used water quality parameter that provides valuable information about the presence of dissolved substances, the suitability of water for various uses, and the overall health of aquatic ecosystems. For the assessment of

EC levels water should be regularly monitored so that the water sources remain within safe and appropriate limits and to address any potential issues in a timely manner.

1.9.1.5 Total Hardness (TH)

Total Hardness depends on the concentration of calcium (Ca) and magnesium (Mg) ions in water, which are commonly referred to as "hardness ions." Hardness ions are naturally occurring minerals that are present in most water sources and can affect the suitability of water for various uses. The direct implication of hardness is that more soap is required to get good lather.

Total hardness is typically categorized as: very hard, hard, moderately hard and soft if the total hardness has the values respectively over 180 mg/L, 121 to 180 mg/L, 61 to 120 mg/L, 0 to 60 mg/L.

Hardness is measured in terms of calcium carbonate equivalent in milligrams per litre of water. The hardness caused by carbonates is known as temporary and due to non carbonates is referred as permanent hardness. High levels of total hardness in drinking water can cause scaling in pipes, water heaters, and other appliances, as well as affect the performance of soaps etc. For these reasons, some water treatment processes are designed to reduce the hardness of water to make it more suitable for drinking and other uses.

Low levels of total hardness in water can cause corrosive conditions in pipes and appliances, which can lead to the leaching of heavy metals and other pollutants into the water. In these cases, water treatment processes may be used to raise the hardness of water to a safe and appropriate level.

In conclusion, total hardness is an important water quality parameter that provides information about the presence of hardness ions, the suitability of water for various uses, and the potential for scaling and corrosion in pipes and appliances. Regular monitoring and assessment of total hardness levels are necessary to ensure that water sources remain within safe and appropriate limits and to address any potential issues in a timely manner.

1.9.1.6 Turbidity

Turbidity is related to the appearance of water, it provides quantitative information about the cloudiness or haziness of water. It may have been caused by the presence of suspended solids, like silt, algae, clay, and other organic and inorganic matter. Turbidity is expressed in units of nephelometric turbidity units (NTU) or formazin turbidity units (FTU).

High levels of turbidity in water can indicate the presence of pollutants or other substances that can affect appearance of water as well as the taste and odor and its suitability for various uses. High turbidity can also interfere with the effectiveness of disinfectants and can reduce the penetration of light in aquatic ecosystems, affecting the growth and survival of aquatic plants and animals.

Low levels of turbidity are generally considered to be a sign of good water quality, as they indicate that the water is clear and free of suspended solids. However, low turbidity values do not necessarily indicate the absence of dissolved solids. Clear water may still have high TDS values such as salts and minerals that can affect the overall water quality and its suitability for various uses.

In conclusion, turbidity is an important water quality parameter that provides information about the presence of suspended solids and the overall clarity of water. Regular monitoring and assessment of turbidity levels are necessary to ensure that water sources remain within safe and appropriate limits and to address any potential issues in a timely manner.

1.10 REGIONS OF PUNJAB, INDIA

Punjab is a state in India, the word 'Punjab' is obtained from two words '*punj*' meaning 'five' and '*aab*' meaning water (i.e. rivers). There were five rivers flowing in the united Punjab (now two rivers are in Pakistan) [26].

Figure 1.12 Showing four districts of Majha region of Punjab[‡] India [26]
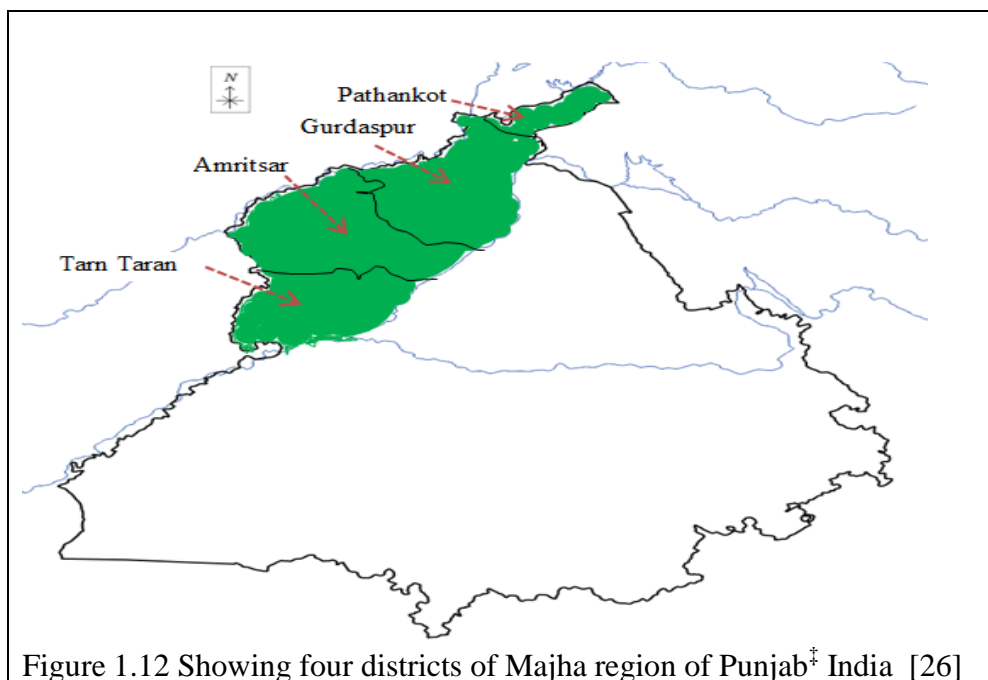
[‡] The map is not up to scale it is just indicative of the regions in consideration.

There are mainly three regions in Punjab named as Majha, Malwa and Doaba. Majha (meaning - at the center) is the region between Beas and Ravi, Malwa is on the South of Sutlej, and Doaba (two rivers) is the region between Beas and Sutlej. There are 22 districts in Punjab. Districts of Majha region are: Amritsar, Gurdaspur, Pathankot, Tarn Taran , districts of Malwa region are, Barnala, Bathinda, Fatehgarh Sahib, Faridkot, Fazilka, Firozpur, Ludhiana, Mansa, Moga, Mohali, Muktsar, Patiala, Ropar, Sangrur. The districts of Doaba region are Jalandhar, Kapurthala, Hoshiarpur, Shaheed Bhagat Singh Nagar.

1.10.1 Majha Region

The Majha region of Punjab, India is a culturally rich and historically significant area located in the northwestern part of the state. It is bordered by the rivers Beas and Ravi and encompasses the districts of Amritsar, Tarn taran, Pathankot and Gurdaspur.

The region is known for its fertile agricultural land and is an important center of the state's agricultural and food processing industries. Some of the major crops grown in the region include wheat, rice, sugarcane, and various fruits and vegetables.

The Majha region is also rich in cultural heritage and is home to several important historical sites and events, including the Golden Temple in Amritsar, the Jallianwala Bag Memorial, the region is associated with Sikhism, marriage of Shri Guru Nanak Dev ji took place in Batala therefore Majha region has a great historical importance. The region is also known for its rich tradition of Punjabi music and dance, and is considered to be the birthplace of many famous Punjabi folk songs and dances.

The Majha region of Punjab is rich in water resources and has several important sources of water, including rivers, canals, and groundwater. Some of the major water resources in the region include the river Beas, the river Ravi and some of the canals are there in this region. The Beas River is one of the major rivers in the region and provides water for irrigation, drinking, and industrial purposes. The river Beas merges with river Sutlej at Harike in Tarn taran. The Ravi River is another important source of water in the region and is used for irrigation, power generation, and other purposes. Groundwater is an important source of water in the region, and there are several hand pumps and tube wells in the region that provide water for drinking, irrigation, and other purposes.

In conclusion, the Majha region of Punjab is a culturally rich and economically important area that is known for its fertile agricultural land, rich cultural heritage, and important historical sites. It is a valuable part of the state's cultural and economic fabric and plays an important role in the overall development and growth of the state. The Majha region

of Punjab is well-endowed with water resources and has several major sources of water, including rivers, canals, and groundwater, which are crucial for the region's agricultural, industrial, and domestic needs. However, it is important to manage and protect these water resources effectively to ensure their long-term sustainability and availability.

1.11 STATISTICAL ANALYSIS

Statistical analysis is a process of extracting intelligent information from the data. It is an effective approach for the understanding of intricate data. For analyzing data it is necessary to collect data from the primary source and it becomes the first step. The collected data is organized in the tabular form and the process of analysis starts thereafter. From the tabulated data, information like mean, median, mode, range, variance, covariance, correlation and so on is extracted. The data can be presented in the form of graphs like histograms, bar graphs etc. The data between two variables can be represented graphically and the plot is known as scatter plot. The analysis of data helps in taking judicious decisions. There exist several statistical methods and techniques for the analysis of data. The statistical analysis can be of two types, which are discussed below.

1.11.1 Descriptive Statistics:

In this analysis summary and description of the basic features of a data are extracted. The basic features which are of importance are central tendencies like- mean, median, mode etc. There are other features belonging to the data set which particularly provide information about the spread and distribution of data like variance and standard deviation, there is a simple feature for spread and is known as range of the data. Main characteristics of the data in a compact and easy-to-understand format become available from the above mentioned features of descriptive statistics.

Mean

There are averages which are unique for a data set and one of them is the mean. It is a simple measure for getting information about the central tendency for a given numerical data set. It can be calculated by dividing the sum of all the observations with a number which is the count of all the observations. The formula for the mean is given by:

Mean = (sum of all the observations in the data set) / (count of observations)

For example, if we have a data set like {1, 2, 3, 4, 5}; the mean can be calculated as follows:

Mean = (1 + 2 + 3 + 4 + 5) / 5 = 15 / 5 = 3

In conclusion, the mean is a useful summary statistic. It provides a typical value unique for the data set. If there are outliers or extreme values then the mean value loses its importance. There arises need for other averages like median and mode.

Median

The median is another average and it measures the central tendency of the given numerical data set. It is the middle value in a data set. The data set is arranged in ascending order. It makes no difference even if the data values are arranged in descending order. If the count of the observations or data values is odd, then median is simply the middle value. If the data set has even count, then median is the mean of the two middle values.

The median is another useful summary statistic of the given data set and it provides a unique value for the given data set. The outliers and extreme values do not affect the median as they affect the mean. For this reason, the median is often preferred to the mean.

For example, if we have a data set containing the values {1, 2, 3, 4, 5}, the median value would be 3. And if data set contains even number of values like the set {1, 2, 3, 4, 5, 6}, the median would be the mean of the two middle values, i.e. (3 + 4) / 2 = 3.5.

In conclusion, the median is another useful measure of central tendency like mean. Its sensitivity towards outliers and extreme values is given no importance. It is often preferred against mean for the data set containing extreme values and outliers.

Variance

Variance measures the dispersion of data set having numerical values in it. It provides information about the deviation of the individual values or observations from the mean of the data set. Variance is a positive numerical value. It can be calculated using the following formula.

Variance = sum of squared deviations from the mean / (count of data values)

In conclusion, variance is a positive number that gives an indication about the spread of observations in the given data set from the mean of the data set. It is a commonly used summary statistic that provides insight into the distribution of data values or observations in the data set. A small numerical value of variance would mean that the observation in the data set is near to the mean, whereas a higher value of variance would mean that the observations are more dispersed.

Standard Deviation

The standard deviation is also a positive number and it provides information about the dispersion of observations in the numerical data set. It gives an indication of how much the

individual values in the data set deviate from the mean of the data set. The formula for the standard deviation of a data set is given by:

Standard deviation = square root of the variance

In conclusion, the standard deviation is a positive number that gives an indication about the spread of observations in the given data set from the mean of the data set. It is a commonly used summary statistic that provides insight into the distribution of data values or observations in the data set. A small numerical value of standard deviation would mean that the observation in the data set is near to the mean, whereas a higher value of standard deviation would mean that the observations are more dispersed.

Correlation

Correlation is a statistical term and is expressed by coefficient of correlation, which is a number lying between -1 and +1 including both ones. Higher the numerical value of coefficient of correlation greater is the chance that there is relation between the two variables in question. In other words it is a measure of the degree of association between two quantitative variables, often assumed to be linear. The correlation coefficient, quantifies this association between the two quantities. If the value of coefficient of correlation is positive one then it indicates a perfect positive correlation and if it is negative one then there is perfect negative correlation, the value zero indicates no correlation.

1.11.2 Inferential Statistics

This type of analysis uses a sample data set which subset of the main data set also known as population data set. Analysis of the sample data set is used to make inferences about the main data set i.e. about population. This involves estimating and making predictions about parameters like- mean, variance, standard deviation etc. of the population.

Hypothesis testing: This type of analysis is used to test a hypothesis for a population which is based on estimates obtained from the analysis of the sample data. The steps involved are– to formulate a null hypothesis and to formulate an alternative hypothesis. Next step is to use statistical tests which can be used to determine whether the null hypothesis can be rejected or not.

Regression analysis: This type of analysis is done to develop model which can find the relation between variables. The number of variables can be two or more. The developed mathematical model tends to fit the given data set. The developed model is then used to predict the values based on the independent variables.

Bivariate Linear Regression: It is a procedure used in statistics to examine the relation between two variables. In the two variables one is independent (X-axis) and second one is dependent (Y-axis). From the existing data an equation is obtained by the use of which dependent variable is predicted from some new value of independent variable. The linear equation obtained is a line which best fits the existing data. The line is chosen in such a way that the sum of squared differences between the observed and predicted values becomes minimum.

Multivariate analysis: This type of analysis is used to analyze data with multiple variables. It includes techniques such as principal component, factor and cluster analysis. These techniques help in the reduction of the dimensionality of the data. Patterns and relationships can also be identified among the variables.

It can be concluded that statistics and the analysis can be used to extract important information from the complex data. It helps in making judicious decisions based on the analysis. There are several statistical methods and techniques which are used depending on the type of data and the objectives. It is important to choose the best suitable method based on the data and the analysis objective.

1.11.3 Principal Component Analysis (PCA)

PCA is a multivariate data analysis technique and it is used to reduce the dimensions of a multidimensional data. By doing so the original information remains intact to large extent. PCA provides a new set of uncorrelated variables from the correlated variables in the original data set. The new uncorrelated variable sets are known as principal components which retain the most important patterns of the data. Principal component is projection of multidimensional data in a two dimensional plane. The Principal component which has high variance is known as the first principal component, the second principal component has the smallest variance.

PCA is a linear transformation method and is used in many fields, including machine learning, image processing, and pattern recognition. The steps involved in performing PCA are:

1) *Standardization of the data*: The variables in the data should be standardized. Each data value is subtracted from the mean. The value so obtained is divided by the standard deviation of the data set.

2) *Computing of the covariance matrix:* The covariance matrix is obtained.

3) *Computing the eigenvectors and eigenvalues:* Covariance matrix is used to obtain eigenvalues and eigenvector. The eigenvectors are used to define the new set of principal components, and the eigenvalues are used to measure the amount of variance. Each principal component explains variance.

4) *Selecting the principal components:* Based on eigenvalues, the number of principal components to retain can be selected. Typically, the principal components that explain the most variance are selected, and the rest are discarded.

5) *Transforming the data:* The last step is to transform the original data set into a new set of principal components.

The PCA results can be visualized by using score and loading plots. Score plot is used to visualize the relationship between observations in the reduced dimensional space of the PCA. This plot helps in visualization of the outliers, patterns, clusters in the data set. The loading helps in the visualization of the original variables and the principal components. Variables which are very close to each other are highly correlated. Loading plot helps in pointing out those variables which are driving the patterns in the score plot.

In conclusion, PCA is a technique used to reduce the dimensions of a large dataset. It also retains the information of the original data set. The goal of PCA is to transform a set of correlated variables into a new set of uncorrelated variables, known as principal components that capture the most important patterns in the data. The steps involved in performing PCA are standardizing the data, computing the covariance matrix, computing the eigenvectors and eigenvalues, selecting the principal components, and transforming the data.

1.11.4 Softwares Used in Data Analysis

Now-a-days there are various types of softwares available for the analysis of data. When the data is multidimensional and is very big manual analysis becomes very tedious and time consuming. The softwares which were used in the data analysis were MS Excel, Camo unscramble and clustvis webtool. MS excel is used carrying the data and also for basic statistics and to get pivot graphs and for simple regression. Information of other two softwares is given below.

CAMO: The Unscrambler ×10.5 (64-bit) is data analysis software developed by CAMO Software. It is a powerful and versatile tool for multivariate data analysis, modeling, and prediction. The Unscrambler ×10.5 (64-bit) offers a wide range of data analysis techniques, including principal component analysis (PCA) and partial least squares regression (PLS) etc.

The software is designed to help users make informed decisions based on their data by providing powerful and interactive visualizations and insights into their data. It also provides a user-friendly interface for creating, editing, and saving models, as well as for generating predictions based on these models. The Unscrambler ×10.5 (64-bit) is designed for a range of industries, including food and beverage, chemicals, pharmaceuticals, and biotechnology. It supports large data sets and can handle a wide range of file formats.

In conclusion, The Unscrambler ×10.5 (64-bit) is powerful and versatile data analysis software that offers a wide range of techniques for multivariate data analysis, modeling, and prediction. It is designed to help users make informed decisions based on their data and provides a user-friendly interface for creating, editing, and saving models, as well as for generating predictions based on these models.

Clustvis: Clustvis is a web-based tool for visualizing clustering of multivariate data. It provides a user-friendly interface for visualizing the results of clustering algorithms and helps in selecting the number of clusters and evaluating the quality of clustering results. Clustvis allows the user to visualize data using various dimensionality reduction techniques such as Principal Component Analysis (PCA) and also supports various other clustering methods.

In conclusion, Clustvis is a useful tool for data scientists and researchers who work with clustering of multivariate data, as it provides an interactive and visual way of understanding and evaluating clustering results.

1.12 REGRESSION AND MODEL BUILDING:

When two variables are related to each other and the value of coefficient of correlation between these is high, the regression analysis can be done. Coefficient of correlation (r) can be calculated using the following formula Karl Pearson's coefficient [27] of correlation for N observations

$$r = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{N\sigma_x \sigma_y} \qquad (1.17)$$

Where, $\bar{x}, \bar{y},$ represent the mean of the respective variable and

$$\sigma_x = \frac{\sqrt{\Sigma(x - \bar{x})^2}}{N}, \sigma_y = \frac{\sqrt{\Sigma(y - \bar{y})^2}}{N} \qquad (1.18)$$

represent the standard deviation for the respective variable.

Consider two arbitrary variables x and y with data set given in Table 1.3 having coefficient of correlation less than unity but with strong correlation

Table 1.3 Arbitrary data set for two variables

variable1= x (also known as predictor or regressor)	x_1	x_2	x_3	…………….	x_j	……	x_n
variable2= y (also known as response)	y_1	y_2	y_3	…………….	y_j	……	y_n

For this arbitrary data set a scatter plot can be obtained and a regression line [28] can also be obtained shown in Figure 1.13. Let the equation obtained be in the form of a straight line, given by equation (1.19)

$$y = \beta_0 + \beta_1 x + \varepsilon \qquad (1.19)$$

Where, β_0 and β_1 respectively are, the y intercept and the slope of the line obtained and also known as coefficients of regression and these are constants. In the Figure 1.13 all points do not lie on the line, for this an error term ε is added to the regression equation. This ε is statistical error term and is a random variable which can account for the failure of the regression model. Equation (1.19) is known as simple linear regression model because it only contains one regressor variable.
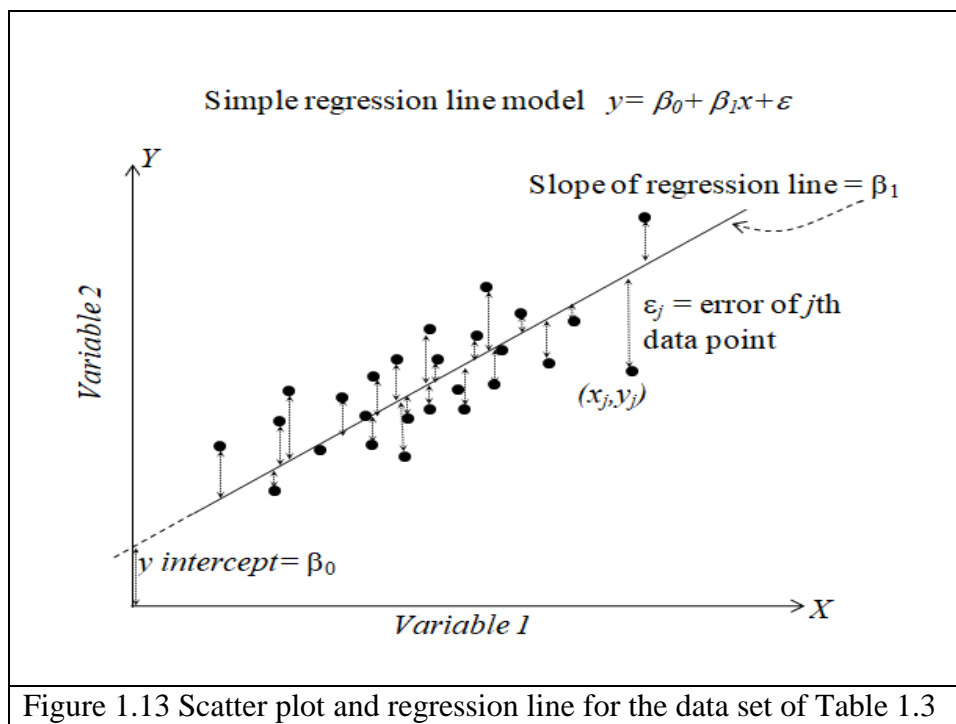
Figure 1.13 Scatter plot and regression line for the data set of Table 1.3

1.12.1 Partial least square regression

Partial Least Square Regression is a powerful tool for modeling complex relationships between datasets, particularly when dealing with high-dimensional, collinear, or noisy data.

This technique is developed by Herman Wold in 1950 in the field of economics. Now it is widely used in various scientific and engineering disciplines. It combines dimensionality reduction with regression, making it suitable for a wide range of applications across various fields.

In the context of spectral data, PLSR [29] is a powerful technique that simplifies complex, high-dimensional data into meaningful components. These components are then used to build a regression model that can accurately predict response variables, such as some chemical property. By focusing on the most relevant patterns in the spectral data, PLSR enhances the ability to make reliable predictions in various scientific and industrial applications.

Partial Least Squares Regression (PLSR) has the following key steps – Calibration, validation, and prediction.

Calibration involves creating a statistical model based on a set of known data. This is where the model "learns" the relationship between the predictors (independent variables) and the responses (dependent variables). The calibration dataset consists of both the predictors and the known responses. The model is fitted to this data, optimizing it to capture the relationship between the predictors and responses. In PLSR, a set of spectra (predictors) and the corresponding parameter data (responses) is used to calibrate the model.

Validation is the process of assessing how well the calibrated model performs on new, unseen data. This step helps to ensure that the model is not over-fitting (i.e., capturing noise rather than the true underlying relationship). The validation dataset is separate from the calibration dataset and includes both predictors and known responses. The model makes predictions on the validation data, and these predictions are compared to the actual known responses to evaluate model performance. Validation can be done either by cross validation or validation set. In the former data is split into several sets and the model is tested and in the latter case a portion of data is set aside as a validation set during the initial data split. The Model is accurate to how much degree is found from the $RMSE_C$ (root mean square error of the calibration and R^2_C (Coefficient of determination for calibration), $RMSE_V$ (root mean square error of validation) and R^2_V (Coefficient of determination for validation). The model is then used for prediction. Prediction refers to using the calibrated model to make predictions on new data where the responses are unknown.

Spectral preprocessing is very important step in the analysis of spectral data, it involves techniques like cleaning, normalizing, and enhancing the data etc. Proper

preprocessing improves the quality of the data and the performance of subsequent analyses. This is done prior to regression analysis.

Some of the common Spectral Preprocessing Techniques are given below:

Baseline Correction: It Removes background noise and baseline drifts which may have distorted spectral data.

Smoothing: Its aim is to reduce noise while preserving the important features of the spectra. One of the methods used in this is to replace each data point in the spectrum with the average of its neighboring points. A Window size (e.g., 3, 5, 7 points) is chosen and average of the points within the window is calculated and then the central data point is replaced with this average. It helps in the reduction of high-frequency noise but can smooth out sharp peaks if the window size is too large. There are other methods like exponential and Gaussian smoothening etc. which can be used.

Normalization: This is done to scale the spectra to a common range to eliminate differences which may arise due to varying sample sizes or instrument settings. There are various methods for this like Min-max scaling, vector normalization, and standard normal variate etc.

Derivatives: Enhances spectral features by removing baseline offsets because derivative of a constant is zero and it also helps in highlighting the peaks. The first derivative, second derivative etc. can be used.

Mean Centering and Scaling: In this the data is centered on zero and scales to unit variance to ensure that all variables contribute equally. The methods involved may be Mean centering, auto-scaling (standardization) etc. [30]

Multiplicative Scatter Correction (MSC): It is used to correct the multiplicative scatter effects and linear baseline shifts. MSC helps to enhance the quality of the spectral data, leading to more accurate and reliable analyses. It works by aligning spectrum in the dataset with a reference spectrum, usually the mean spectrum of all samples.

1.13 OUTLINE OF THE THESIS

Groundwater from the Majha region of Punjab is analysed for its physicochemical properties and spectroscopic analysis is done. In this thesis the work done on these samples is given. Thesis consist of six chapters, the chapter wise summary is as follows.

Chapter 1 provides introduction to the various techniques and basics of the theory of spectroscopy and about the statistical techniques used. Basics of NIR, FTIR and fluorescent spectroscopy are presented here. Applications, instrumentation of these techniques is given,

importance of groundwater, Regions of Punjab, contamination of the groundwater, physicochemical properties of water are given in brief.

Chapter 2 consists of literature review which is concerned with the previous work done in the related field which is in question here; it highlights the results, methods and techniques used by various researchers.

Chapter 3 covers the hypothesis, objectives of the present work, research gap, materials and methods used in the present research work. In this chapter the whole process from the sample collection to chemical procedures used, spectroscopic analysis and data collection, data analysis methods and statistical techniques employed in the research work are discussed.

Chapter 4 covers the FTIR and NIR analysis of collected groundwater samples for the presence of contaminants and covers the results.

Chapter 5 covers the correlations between pairs of physicochemical parameters and regression equation between the pairs showing some correlation. Descriptive statistics results are also given. PCA is also discussed on the basis of region, depth and source of groundwater sample.

Chapter 6 covers the use of fluorescence spectra, Fourier transfer infrared spectra and its correlation with the parameters and near infrared spectra with regression models using PLS and interval partial least square regression.

In conclusion, groundwater from the Majha region is assessed for contamination using spectroscopic techniques such as FTIR, NIR, and fluorescence, combined with statistical methods like PCA and PLSR.

The motivation behind this research is to provide an effective, cost-efficient way to monitor water contamination. The physicochemical data, spectroscopic techniques, and statistical analysis offer valuable insights into the groundwater quality in the Majha region. The aim of the work is to enhance groundwater monitoring processes and to address the critical issues of water contamination and pollution, which are major environmental concerns now-a-days.

CHAPTER 2

LITERATURE REVIEW

CHAPTER 2 - LITERATURE REVIEW

2.1 INTRODUCTION

Ground water is an important natural resource and is found under the surface of the Earth between rocks, sand, soil and gravel. The rural communities fulfill their 90 percent of water needs from underground water. Groundwater is used in various activities like drinking purpose, household, industry and agricultural activities etc. Survival is not possible without water. Water can be characterized by its physicochemical properties like pH, TDS, total hardness, Turbidity, chlorine content, nitrate, sulphate, fluoride, nitrate, electrical conductivity etc. World Health Organization (WHO) and other country based organizations like Bureau of Indian Standards (BIS) in India have specified the acceptable upper limits for these physicochemical parameters. The water quality is deteriorating due to human activities and natural activities generally known as anthropogenic and geogenic activities respectively. According to central ground water board India [31], in many regions of Punjab, Haryana, Rajasthan and Gujarat and in few regions in Delhi, Maharashtra, Uttar Pradesh, Madhya Pradesh etc. saline water is found with Electrical Conductivity greater than 4000 µS/cm. In many of the places in Rajasthan and southern Haryana the value of EC of ground water is found to be greater than 10000 µS /cm. Thus making this water unfit for drinking. There are other contaminants in water found in various parts of India like fluorine, arsenic, uranium etc. The ground waters in Punjab also found to be contaminated with uranium and other heavy metals. The researchers have found that in the Malwa belt of Punjab the contaminations of water are causing various types of diseases in the inhabitants. It has become a necessity to check the ground water on regular basis for the contaminations so that actions could be taken to treat the water for its purity to prevent the water borne diseases [32].

In this chapter the research carried by various researchers is reviewed concerned with the contamination of water, methods used for determining the physicochemical properties and the spectroscopic techniques or any other technique employed by the researchers for the analysis of the water. The NIR spectroscopy is in use for many years and employed in the field of water and other food products for analysis [33]. It helps in the study of samples by providing information about the molecular structure and the bonds in the samples and the peaks in the spectra can indicate the pollutants in the water [34]. The spectral data so obtained is very large and its analysis is done using multivariate analysis. NIR spectroscopy has been in use in various fields like medicine, food and beverages industry. It has been in

use for the qualitative and quantitative analysis of various food items like wheat, coriander, millets, mango and other food items [35 - 36]. It is clear from the literature review that FTIR is also used to get the spectra and its peaks are used to know about the bonds present in the sample [37 - 38]. The literature review is done chronologically form the earliest available literature to the most recent ones so to ensure that up-to-date information is included in the review.

2.2 LITERATURE REVIEW

Galapate R.P. et al (1998) [39] conducted this study to detect the domestic wastes in Kurose River in higashi-Hiroshima city Japan using synchronous fluorescence spectroscopy. The said technique was applied to differentiate natural organic matter and dissolved organic matter. Three samples of natural origin dissolved organic matter were obtained from commercial leaf mold, forest top soil and algal culture media. Two samples of Sewage effluent and gray water were taken from Kurose River as source of domestic waste. Fluorescence spectra were recorded and its analysis was also done. Peaks were obtained in the region from 300 to 400 nm, they obtained corrected peaks and concluded that Kurose river is polluted with domestic waste water.

Thomas, A. et al (1999) [40] presented hydro geomorphological mapping using IRS-1B II data in assessing ground water in Lehra gaga block of Sangrur district Punjab India. The Lehra gaga block is a part of the Indo-Gangetic alluvial plain. The data of the Indian remote sensing satellite was obtained in the form of false colour composites on two seasons in Oct. 92 and Mar. 93 and was interpreted using visual interpretation technique. The region has alluvial plain as its major geographic unit, 1.65 percent of the area has sand dunes. The south eastern area near Ghaggar River has paleochannels. The concentration of the soluble salts is within limits. It is observed that sodicity is more than salinity. When compared ground water quality with the hydro geomorphological map no direct relations were found. This study showed that remote sensing technology can be used for preparing the hydrogeomorphological map.

Virk H.S. et al (2001) [41] worked on radon monitoring in underground water of Gurdaspur and Bathinda districts of Punjab, India. Radon gas is produced due to the U/Th radioactive series as a decay product. Its monitoring is done using silver activated zinc sulphide phosphor scintillator. its concentrations were measured in various villages of Gurdaspur, Batala, Dinanagar, etc. The water which was drawn by hand pumps had high concentration of Radon than the water stored in tanks. It is concluded that radon

concentration shows an increasing trend while we move from Batala to Gurdaspur i.e. towards shivalik Himalaya. No relation between uranium and radon concentration found, its concentration is high in groundwater than in surface water. The bottled water is safe to drink as it does not contain Radon.

Howe J.K. et al (2002) [42] used attenuated total reflection Fourier transform infrared spectroscopy (ATR/FTIR) to study fouling of microfiltration membranes by natural waters. It is mentioned in the paper that dissolved organic matter, which is heterogeneous mixture of many chemicals produced by various biological activities taking place on land and in water, is main contributor for the fouling of the membrane. Surface waters from various water bodies like Medina River, Texas and Beaver Lake, Arkansas were taken and filtered from micro filtration membrane (made from 0.2µm polypropylene). The membrane so obtained is examined using ATR/FTIR spectroscopy. Samples of the size 45 by 10 mm were taken and ATR/FTIR spectra obtained. GRAMS/32 software was used to obtain difference spectra. The foulant material gives different vibrational spectra bands which are distinguishable from the bands of the membrane material. There were two types of foulants: organic (which contain compounds having carbonyl and amide groups) and inorganic (which contained aluminum silicates). With the change in the pH of the feed water from 4.0 to 9.0, the functional chemistry of the deposited material does not appear to change.

Farooqi et al (2009) [43] analyzed arsenic and fluoride in highly contaminated soils, which are causing groundwater contaminations in Punjab, Pakistan. The study area was 40-45 km south of Lahore along the East bank of river Ravi. 42 samples of surface soil, 41 samples from the depth of nearly 30 cm and 3 samples of surface sediment and 26 sediments from 500 cm deep from the surface of the river bank were collected. The samples were dried in oven, crushed and sieved to get particle size less than 2mm pH value is determined, X-ray fluorescence photometry was done and the samples were analysed for fluoride and arsenic concentration. Statistical analysis was done using statistical package for social sciences (SPSS). From surface to 30 cm deep the composition for all major elements is nearly same. Surface soil is found to be comparatively rich in arsenic as compared to deep soil whereas it is reverse for the fluoride concentration. From the statistical analysis of the data no positive correlation is found between the concentrations of arsenic and fluoride with the major element constituents of the soil, only the clay minerals show somewhat positive correlation with the concentrations of arsenic and fluoride. pH plays an important role in controlling the concentration of the fluoride. The pollution due to these elements is anthropogenic and the three sources which may be the cause are industrial waste, fertilizers and combustion of coal.

Sohn M et al (2009) [44] used fluorescence spectroscopy for the detection and differentiation of bacteria in foods and used multivariate analysis like principal component analysis. It was concluded that the use of fluorescence spectroscopy data and the multivariate analysis can be used for the detection and differentiation of bacteria in the food.

Prem singh etal (2010) [45] explored Sutlej River water (in Himachal Pradesh, India) to know the concentration of elements with lower atomic number in it. Four different locations were chosen to collect water samples. Water samples were taken in plastic containers which were cleaned using distilled water. The technique employed for the analysis was Energy Dispersive X-ray fluorescence. The water contained sulphur, iron, chlorine, calcium in unit of $\mu g/cm^2$ and some traces of titanium, chromium, manganese etc. were also found in the river water but less than $1\mu g/cm^2$ and also not in all the samples.

Li J. et al (2011) [46] used NIR diffused reflectance spectroscopy for the determination of mercury, lead and cadmium ions in water. Required reagents and apparatuses were collected and magnesium-mercaptopropyltri-methoxysilane (Mg-MTMS, thiol-modified magnesium phyllosilicate material) was prepared and characterized. Solutions of mercury, cadmium and lead (all nitrates) salt as stock solution were prepared. River water was mixed with these stock solutions in specific concentrations to get 38 mixture samples. Inductively coupled plasma-atomic emission spectroscopy was used to measure the water content in river water. It only showed the presence of only sodium, calcium, potassium and magnesium ions in the water. Mercury, lead and cadmium ions were not found. The above prepared samples were mixed with Mg-MTMS in specific concentration filtered and air dried. For developing partial least square (PLS) model, Spectral measurements were done and results were used for the calibration set. NIRS was used to scan the samples in the range of 4000-10000 cm^{-1}. PLS modeling was done using OPUS-QUANT program in the OPUS 6.0 software. PLS, coefficient of determination (R^2), root mean square error for cross validation ($RMSE_{CV}$), residual predictive deviation (RPD), standard deviation (SD), standard error of prediction (SEP) in cross validation, were used for the predictive accuracy of the models. It is concluded by the researchers that method proposed in this research paper may be good alternative for the determination of low concentration metal ions in environmental water samples.

Kumar A. et al (2011) [47] worked for creating geochemical modeling of uranium speciation in the subsurface aquatic environment of Punjab, India. Ground water samples were collected from varying depths in the districts of Mansa and Bathinda. They measured

various geochemical parameters during sample collection. Collected water samples were treated in lab and handled in oxygen free glove bag to maintain redox conditions. Cations, anions and uranium concentrations were measured using various relevant techniques and equipment. The statistical parameters like mean, median, minimum, maximum, quartiles, etc. are given in the paper in tabular form. They observed that concentration of dissolved uranium in ground water is controlled by the pH and oxidation reduction potential (E_h) conditions. Solubility of uranium increases depending on various other factors like TDS, alkalinity, salinity etc.

Srivastava et al (2013) [48] investigated the River Ganga water form Allahabad for metallic elements. Human activities are causing pollution of the river water. They used Laser induced breakdown spectroscopy technique to investigate the minerals present in the river water. The spectral region used was lying between 200-900 nm. From the spectrum it was found that various macronutrients sodium, hydrogen, calcium, potassium etc., micronutrients like titanium, chromium, silicon etc. and toxic elements like mercury and cadmium etc. are present in the river water. The study demonstrated that Laser induced breakdown spectroscopy technique is capable of elemental analysis.

Sidhu et al (2014) [49] aimed their work to analyze the arsenic metal in drinking water of Malwa region of Punjab, India. Water samples were collected randomly from eighteen villages of three districts lying in this region. The water samples were collected from hand pumps following a specific protocol. Physical properties of water were measured like dissolved oxygen and total dissolved solids. Arsenic was accurately measured by inductively coupled argon plasma-atomic emission spectroscopy. They concluded that arsenic level was not that much high but the pH and TDS values showed variation.

Chen Y. et al (2015) [50] applied surface enhanced Raman spectroscopy coupled with advanced chemometric method-multiplicative effects model to quantitatively analyze the presence of ametryn (a herbicide used to control weeds) in the sample of water taken from the Xiangjiang River in Changsha (China). Required chemicals like Ametryn, p-thiocresol, sodium citrate dehydrate and chlorauric acid hydrated etc. were procured form chemical companies, all these chemicals were of analytic grade and ultra-pure water was used in the process. Gold Nano particles were prepared. Twenty seven samples having different concentration of ametryn were prepared using ultrapure water and following the same procedure another twenty seven samples were prepared with the water taken from Xiangjiang River. Surface enhanced Raman spectroscopy was used to get spectrum of each sample. The samples prepared from the river water were analyzed using high performance liquid

chromatography for the purpose of comparison. It is concluded that surface enhanced raman spectroscopy combined with multiplicative effect model has great potential for quantitative analysis of Ametryn in water samples.

Saikia B.J. et al (2016) [51] did spectroscopic (raman and FTIR) evaluation of clay minerals and did estimation of metal contaminations in natural depositions of surface sediments from Brahmaputra river. Samples were collected from six different locations and from each location, five samples were collected. Raman and FTIR spectra of each sample were obtained. The result obtained indicated that sediments contained crystalline quartz. Various contaminants found present in the samples were copper, titanium, nickel, lead, manganese, magnesium, silicon, potassium, chromium etc. Infrared spectra peaks were near 695 cm^{-1} which point towards the presence of micro-crystalline quartz particles. There were other peaks form 1614 - 1620 cm^{-1} indicated the presence of silicate minerals. The Raman peaks were observed at 722 - 724 cm^{-1} and 486 - 488 cm^{-1}. This indicates the presence of geikielite ($MgTiO_3$) in the samples.

Virk, H. S. (2017) [52] prepared a preliminary report on the crisis of high concentration of uranium in ground waters in the state of Punjab India. It was found by Carin Smit in 2009 that in Malwa region of Punjab, the uranium is very high in the blood samples collected from the mentally retarded children. The quality affected habitations belong to Malwa belt. Hoshiarpur also has 15 habitations with highest uranium content. Punjab water supply and sanitation department has recommended the use of reverse osmosis system to be installed in rural areas where uranium concentration is higher than the permissible limit of 60 microgram per liter or ppb. In ground water, 539 habitations were having high content of Iron, in Patiala highest number of habitations were found to have high content of cadmium, Hero Kalan village of Mansa is found to have highest content of mercury, 150 habitations in Hoshiarpur have highest aluminum contamination, lead is found in many habitats in gurdaspur highest and followed by Jalandhar, selenium and arsenic is found in Malwa region. All this data is on the basis of survey conducted by PWSSD during 2009 to 2016. In some districts fluoride, chloride, nitrate, sulphate, calcium and magnesium contamination is also found. All the contaminants, if found more than the permissible limits are causing severe health issues. It is concluded in the report that 35% habitations in Punjab are affected by high concentration of uranium.

Wen Z. et al (2019) [53] investigated the composition and sources of dissolved organic carbon in Haihe River Basin, China. This study aimed to identify the source of dissolved organic carbon and the seasonal influence of concentration and composition of

dissolved organic carbon in the Haihe River basin. Samples from fifty nine locations were collected during spring and from twenty six locations during autumn. Various parameters like temperature, specific conductivity, salinity, and pH value were determined. Dissolved total carbon, total carbon, dissolved organic carbon, total dissolved nitrogen, total dissolved phosphorous were analyzed. Secchi disk depth was also measured at each sampling point to measure the transparency of water. Both spectroscopic and carbon stable isotope analysis was also done. Chromophoric dissolved organic matter absorption and fluorescence measurements were taken. Various other data analysis techniques were used, e.g. ANOVA was used to know the influence of seasons (spring vs. autumn) on dissolved organic matter concentration. Spectroscopic and stable carbon isotope analysis was found to be useful in identifying the end-member source of dissolved organic carbon.

Sharma T. et al [2019] [54] did a comprehensive study of seasonal variation of uranium in ground water in the Majha region of Punjab India. The districts which lie in this region are Amritsar, Pathankot, Tarn taran and Gurdaspur. For collecting samples the region was divided in 6×6 km^2 and from each grid a sample was collected. Various physicochemical parameters like pH, total dissolved solids, electrical conductivity and total hardness were determined. Samples were analysed for anions like chloride, nitrate, fluoride, and sulphate etc., this was done to determine a correlation between these parameters and uranium distribution. Uranium in water samples was estimated by LED-fluorimeter. Uranium content in samples collected from Amritsar was higher as compared to samples collected from Gurdaspur and Pathankot districts, but in all the three districts the uranium content was with in safe limits.

Quintelas C. et al (2020) [55] used Infrared spectroscopy to identify and quantify emerging pollutants in wastewater. Samples of emerging pollutants were prepared. The samples were analysed using ultra high performance liquid chromatography. Fourier transfer infrared spectroscopy spectra were recorded in the range from 200 cm^{-1} to 14000 cm^{-1}. Chemometric analysis which included k nearest neighbor, multiplicative scatter correction, first and second derivative and ordinary least square and partial least square regression was done. For all the analysis MATLAB software was used. It is concluded that Fourier transfer infrared spectroscopy, chemometric approach will be future replacement of ultrahigh performance liquid chromatography and gas chromatography for the emerging pollutants in waste water.

Mamera et al (2020) [56] applied Fourier transfer infrared spectroscopy for investigating the concentration of heavy metal ions in river and borehole water sources. They prepared calibration solutions of known concentration of heavy metal like copper, silver, cadmium, lead, zinc etc. The range of wavelength used was 400 - 4000 cm^{-1}. They used multivariate calibration method for establishing relation between heavy metal concentration and the spectral properties. Principal component regression (PCR) and partial least squares regression methods were used, root mean square error and root square error were applied in cross validation. Seven water samples from QwaQwa, South Africa taken and were analyzed using atomic absorption spectrometer. Fourier transfer infrared spectroscopy of water containing metal ions was taken and analysed. Concentration values in the models were calculated using partial least square and principal component analysis methods. These methods gave good regression fits for silver, cadmium, copper, lead and zinc ions. R^2 value for partial least square lied in between 0.95 to 1 and for principal component analysis lied from 0.98 to 0.99. It was concluded in the paper that Fourier transfer infrared spectroscopy has good potential and can be relied upon for the determination of heavy metals.

Scherer C. et al (2020) [57] used FTIR spectroscopic method, visual method and thermo analytical method to quantify micro-plastics in water and sediment of German river Elbe. Samples were collected from eleven sites out of which ten were taken using apstein plankton net and from one site only sediment was collected. Micro-plastic particles were extracted from the water and sediments. Visual (microscope attached with camera), Fourier transfer infrared spectroscopy and pyrolysis gas chromatography mass spectrometry analysis was done. Some methods were employed for validation and control. The results were analysed using graphpad prism 7.04 software. Correlation was used to compare the results obtained in visual and pyrolysis gas chromatography mass spectrometry. It was found that micro-plastic concentration was six lakh fold higher in sediments as compared to water phase which has only 5.57 particles per cubic meter.

Kumar R. et al (2020) [58] investigated the quality of the groundwater in the region of the Sutlej River Basin of Punjab, India for the purpose of drinking and agriculture. The study area was the district Fatehgarh sahib which falls under the Indo-Gangetic alluvial plain in the Sutlej river basin. Groundwater samples (from hand pumps and tube wells) of varying depths and agricultural soil samples were collected. The researchers analysed the water samples for thirteen physiochemical parameters like pH, total dissolved solids, electrical conductivity, dissolved Oxygen etc. Inductive plasma mass spectroscopy was used for the elemental analysis. Different indices like sodium adsorption ratio, sodium percentage, permeability

index, and corrosivity ratio etc. were used for groundwater quality check for irrigation. The result showed variation in parameters for pre and post monsoon period. The groundwater hydrochemistry (rock water interaction, rainfall and evaporation dominance) is demonstrated using Gibbs diagram. Multivariate statistical analyses like principal component analysis and Pearson correlation matrix analysis were done. It was applied to get understanding about the origin, geochemical characteristics and chemical interrelations between soil and water. It is also given in the research paper that toxic elements in the groundwater contain carcinogenic elements like chromium, arsenic etc. It is concluded that total dissolved solids, bicarbonate and total hardness and other physicochemical parameters of the groundwater were within permissible limits but the level of uranium and selenium is more than the permissible limits set by World Health Organization and Bureau of Indian Standards. The groundwater in most of the areas is not suitable for drinking purpose and should be purified before use.

Kumar A. et al (2020) [59] conducted a study to investigated the presence of arsenic and other trace metals in groundwater. The study was done in some part of the Indus basin, Punjab, India. The region of study was Bari doab which includes Amritsar and Tarn taran districts. If concentration of arsenic in ground water is higher than 10 µg/L, it can cause various types of diseases, it is a carcinogen also. Some of the parameters like pH, electrical conductivity and oxidation reduction potential were determined on-site. Various types of cations like potassium, sodium, magnesium, calcium and trace elements like cadmium, arsenic, zinc, chromium, iron etc. were analyzed in the laboratory. Anion analysis for nitrate was done using UV-Visible spectrophotometer. Sulphate concentration was estimated using turbidity meter and standard titrimetric method was used for determining the concentration of chloride and bicarbonate. In the study it was found that the arsenic and iron with low oxidation reduction potential was present mostly along the Ravi River. High concentration of sulphate and nitrate was found, zinc and copper was also found at shallow depth. High concentration of bicarbonate and calcium, magnesium and sodium was also observed. It is not clear about the enrichment of arsenic and a detailed study is needed in this regard.

Kumar R. et al (2021) [60] investigated the source, distribution and potential health risk assessment of groundwater quality in three districts Ludhiana, Bathinda and Barnala all lie in the southwestern region of Punjab, India. The samples were analysed for fifteen physicochemical parameters. Inductively coupled plasma mass spectrometer was used to determine the trace elements like cadmium, chromium, iron, uranium, arsenic, copper and lead etc. Maximum, minimum and mean value is calculated using Excel 2010 (Microsoft office). Multivariate analysis like correlational, principal component and cluster analysis

were also done using statistical package for social sciences software. Analysis of physicochemical parameters was also done. On the basis of total dissolved solids, alkalinity and hardness of the groundwater it was concluded that water in the three districts is not fit for drinking purposes. Statistical analysis showed that there is difference in the quality of the water pre and post monsoon seasons. It is concluded in the paper that uranium contamination is prevalent in groundwater in the Barnala and Bathinda whereas fluoride is found in all the three districts.

Rani R. et al (2022) [61] used different spectroscopic techniques like Fourier transform infrared and fluorescence spectroscopy to analyze the ground water from the Phagwara region falling in the district of Kapurthala, Punjab, India. FTIR peaks were observed at 3743.93 cm^{-1}, 3217.67 cm^{-1}, 1641.43 cm^{-1}, 449.43 cm^{-1}. Fluorescence data was collected between 190 to 400 nm for excitation and 300 to 800 nm range of emission. Peaks were observed at 446.49 nm, 482.98 nm and at 669.54 nm. It is concluded that FTIR and fluorescence spectroscopy can be used to identify the bonds in ground water.

D Gautam et al (2023) [62] did a study which aimed at analyzing the suitability of the sub-surface water of Una for drinking and agricultural purposes, the values of calcium, chlorine, fluoride, sulphate were below limits; pH, total dissolved solids, alkalinity, total hardness and nitrate were above the permissible limits given by BIS.

Kaptan Singh et al (2023) [63] did a study of ground water samples from the Bathinda region. Hydro-geochemical analysis of the ground water was done and concluded that water was sodium bicarbonate and sodium sulphate or chloride type. It was found that 10 percent of the samples were of very poor category for drinking purpose. After monsoon the water quality further deteriorate due to presence of uranium, arsenic, nickel, fluoride etc. Statistical analysis like mean, maximum, minimum, range, standard deviation, correlation etc. for various physicochemical parameters was done, the principal component analysis (PCA) was also done. PCA indicated that origin of nitrate is anthropogenic and origin of other contaminants is geogenic.

Javed Iqbal et al (2023) [64] collected 61 samples from Khushab district Punjab, Pakistan, and studied contamination due to fluoride and nitrate. It was found that the contamination was high according to World Health Organization guidelines. Principle component analysis was used for the analysis.

Gautam et al (2024) [65] attempted to characterize and interpret the groundwater quality (GWQ) from the Jakham River Basin in Southern Rajasthan. In the paper they used GIS environment and multivariate statistical approach. Various statistical indicators were

analysed like principal component analysis (PCA) along with others were used. It is suggested in the paper that PCA is a suitable tool in water quality analysis. Study shows that geological and human intervention has increased the levels of electrical conductivity, fluoride, bicarbonate, chloride, total dissolved solids, sodium and sulphate in potable water.

Tamanna et al (2024) [66] investigated the groundwater quality of Bathinda district of Punjab. The study aimed at determination of heavy metals in the groundwater. Various parameters like electrical conductivity, pH, salinity, total dissolved solid, temperature, chloride content, total hardness, calcium and heavy metals like cadmium, chromium, lead, nickel, and iron etc. were analysed. Some of the parameters were found to be above the permissible range as prescribed by the Bureau of Indian Standards for drinking purposes. 93.3% of the samples did not have acceptable limit for total dissolved solids and total hardness as prescribed by Bureau of Indian Standards. It was concluded that the groundwater of Bathinda district is mainly contaminated with lead and iron and its long-term use is not suitable for human consumption.

2.3 CONCLUSION

According to Bureau of Indian Standards there are allowed limits for various types of contaminations in the drinking water. Some of the Physical parameters with their permissible limits are pH value (6.5-8.5), TDS (500 mgL^{-1}), Turbidity (1 NTU), chloride (250 mgL^{-1}), total hardness (200 mgL^{-1}) [67] etc. Similar limits are also set by World Health Organization [68]. The use of fertilizers, herbicides, insecticides, fungicides etc. and with the development of industry (they are throwing industrial waste and chemicals into the water), the contamination of groundwater has increased to such an extent that the contaminants are causing various types of diseases in the humans. Some of the contaminants in the groundwater are causing various types of diseases like cancers, dental fluorosis, skin diseases, arsenicosis etc. [69 - 70]. In a previous study it is mentioned that concentration of arsenic is very high in groundwaters of various countries, mentioning few of them– Afghanistan, Bangladesh, China, Nepal, Pakistan, India, Sri Lanka etc., it is responsible for various types of diseases related to Skin, kidney, bladder, lungs including cancer etc. [71]. In another study radon monitoring was done in the underground water of Gurdaspur, and Bathinda districts of Punjab, India and it was found that the water drawn from hand pumps had more Radon as compared to water stored in tanks [72]. Physicochemical properties of water from Amritsar Punjab were correlated with the presence of uranium in the water and a positive correlation was observed [73]. A previous study mentioned that in USA 12% lung

cancer deaths are linked to the radon and uranium intake may damage the kidneys and it can be carcinogenic also [74]. In Tarn taran district a study was done on groundwater which revealed that 96% samples were found to be contaminated with arsenic and 51% of the samples had uranium to such an extent that consumption of such groundwater may cause cancer in the habitants of the area, it is also mentioned in the study that the presence of arsenic is more dangerous for children as compared to adults [75]. In another study physicochemical properties of groundwater like pH, phosphate, chloride, sulphate, potassium and nitrate etc. were determined in the Malwa region of Punjab, India which revealed that 92% sites showed the high level of nitrate from the prescribed limit of BIS [76]. Fluoride and nitrate contamination of groundwater is also of concern in many areas, high concentration of nitrate may cause serious health problems like gastric cancer, goiter, hypertension etc., when inhabitants are exposed to high concentration of fluoride through drinking water there are high chances of endemic diseases like dental and skeletal fluorosis [77].

There are many studies which used partial least square modeling for water parameter analysis, a review has been done [78], in another study use of partial least square modeling is said to be practical [79]. In another the use of partial least square in the field of chemistry and technology is discussed as a basic tool [80]. In a study [81] interval partial least squares regression is used on spectral data obtained using near infrared spectroscopy, this is related to the wavelength selection methods [82].

According to reported literature there are various factors [83] which affect the quality of ground water, some of the environmental factors like temperature and humidity etc. There are other geogenic [84] factors like the flowing rivers, mineral rocks in the region and mountains can also have effect on the quality parameters of water. The anthropogenic factors like waste from industrial, agricultural, and human activities also play an important role in the contamination and pollution of drinkable water. To evaluate the quality parameters and compare the results with the national and international standards provided by Bureau of Indian Standards and World Health Organization respectively, physical, chemical, and reference methods were employed. Studies have been conducted at the state, national and international levels.

The literature review enables the categorization of references based on methodology and findings.

Based on methodology: Based on methodology the following categorizations are observed Spectroscopic techniques have been used by Galapate et al. (1998), who employed fluorescence spectroscopy, and Li et al. (2011), who used NIR spectroscopy. Several studies

have applied statistical methods as well—for instance, Khatri et al. (2011) used partial least squares regression, while Sohn et al. (2009) combined fluorescence spectroscopy with statistical techniques. Additionally, Thomas et al. (1999) utilized remote sensing data for hydro-geochemical mapping.

Based on findings: Farooqi et al. (2009) and Kumar et al. (2020) identified that groundwater in regions of Punjab and Pakistan is contaminated with arsenic and fluoride due to both geogenic and anthropogenic sources. Virk et al. (2017) found contamination in Punjab's groundwater caused by the presence of uranium and heavy metals. Sidhu et al. (2014) examined the effects of arsenic in the drinking water of the Malwa region of Punjab, highlighting associated health risks. Tanu Sharma et al. (2019) and Singh et al. (2003) reported the presence of uranium in groundwater in the districts of the Majha region—namely, Amritsar, Gurdaspur, and Pathankot.

The present work differs from similar studies in the following ways: it focuses specifically on groundwater from the Majha region of Punjab and employs multivariate analysis, whereas other studies, such as Virk et al. (2017) and Farooqi et al. (2009), examined broader areas of Punjab in the context of groundwater contamination. Additionally, studies like Saikia et al. (2016) used Raman and FTIR spectroscopy, while the present study adopts a more comprehensive approach by using NIR, fluorescence, and FTIR spectroscopy, combined with PCA and PLSR. Khatri et al. (2021) applied partial least squares regression to assess water quality, whereas this work also employs interval partial least squares regression. Furthermore, while Kumar et al. (2020) focused on identifying groundwater contamination, the emphasis in the current study is on developing a predictive model using spectroscopic data. This model aims to enable real-time monitoring and offers a cost-effective alternative to traditional, tedious time consuming chemical methods, thereby distinguishing this work significantly from conventional approaches used by earlier researchers.

By going through literature it becomes clear that little research has been published on the application of spectroscopic methods and the physicochemical parameters of water samples along with multivariate analysis like principal component analysis, partial least square regression analysis to quantify water samples and developing models for the same. It is reported in the literature that predictions can be made using the model in this case physicochemical parameters of water can be predicted using regression models.

CHAPTER 3

HYPOTHESIS, OBJECTIVES AND METHODOLOGY

CHAPTER 3 - HYPOTHESIS, OBJECTIVES AND METHODOLOGY

3.1 RESEARCH GAP

Ground water plays an important role for the living beings. It is so important for humans and animals that they cannot live without it. Water is used in various activities also, like for drinking, for cooking food, for washing, for agriculture and in industry etc. There are various reasons which pollute and contaminate ground water these may be geogenic or anthropogenic. Some of the contaminants found in water are plastic particles, organic compounds like urea, minerals, heavy elements like arsenic, cadmium, radioactive elements like uranium etc. At national and international level lot of work has been done on river water, ground water and water from other water bodies to determine various properties of water, contaminants, pollutants and effluents present in water using various methods. The results are compared with the permissible values of the parameters provided by WHO (World Health Organization) and BIS (Bureau of Indian Standards). Researches have used various methods and techniques such as Gas chromatography, X-ray diffraction, ATR-FTIR, Ultra high performance liquid chromatography, inductively coupled plasma mass spectrometry etc. for the quality analysis. Some of the researchers have used near infrared spectroscopy in combination with partial least square regression, multiple linear regressions for the collected data. It is observed from the literature review that there is very less work done in Punjab, India for the estimation of pollutants in ground water and estimation of the parameters using regression model involving FTIR and NIR spectra and the physicochemical data collected by chemical methods.

Punjab is a very important state in India with great history of valour and has the blessings of the Sikh Gurus. Punjab is also famous for its role in the green revolution in India and is known as grain bowl or food basket of India. Punjab is divided into three regions known as Majha, Doaba and Malwa and extends from the latitudes 29.30° north to 32.32° north and longitudes 73.55° east to 76.50° east [85]. The region between river Ravi and Beas is Majha, the region between the river Beas and the Sutlej is Doaba and the region beyond Sutlej is Malwa. From the various studies done by various researchers, it is observed that the underground water in the Malwa region at various sites is not suitable for drinking. It is concluded on the basis of various factors related to water like total dissolved solids, presence of pollutants, uranium and other heavy metals. Research papers are available on Ground

water studies in this context from the Malwa region of Punjab, India. Majha region is also reported with adulteration in ground water, but there is scarcity of studies in the Majha region of Punjab. There is a need to study the ground water in this region also. In the present study NIR, FTIR and fluorescent spectroscopy and chemical data is used in the development of regression model, quality estimation, classification of the procured samples on the basis of data obtained from these samples.

3.2 RESEARCH HYPOTHESIS

Null Hypothesis – There is no significant relation between Near Infrared spectra, FTIR and fluorescent spectra of ground water samples and the physicochemical properties like Turbidity, total hardness, Total Dissolved Solids, pH, chloride content and Electrical Conductivity of the procured water samples.

Alternative Hypothesis – There is a significant relation between Near Infrared spectra, FTIR and fluorescent spectra of ground water samples and the ground water sample physicochemical properties – turbidity, total hardness, total dissolved solids, pH, chloride content and electrical conductivity.

3.3 RESEARCH OBJECTIVE

1. Determination of pollutants in ground water from districts of Majha Region of Punjab-Amritsar, Gurdaspur, Pathankot, Tarn taran using FTIR, fluorescence spectroscopy and chemical parameters.

2. Identification and quantity estimation of contaminants in water using spectroscopy and multivariate analysis.

3. Development of regression model using NIR, FTIR, fluorescence spectra and chemical data to predict the pollutants in the ground water samples.

3.4 SAMPLE COLLECTION

Samples were collected from the districts of Majha region (Figure 3.1) of Punjab [86], India viz. Amritsar (Coordinates: 31°35′N 74°59′E, with an average elevation of 234 m), Pathankot (Coordinates: 31°55′N 75°15′E, with an average elevation of 332 m), Gurdaspur (Coordinates: 31°55′N 75°15′E, with an elevation of 264 m) and Tarn taran (Coordinates: 31°28′N 74°56′E, with an elevation of 226 m). Total number of samples collected was sixty. This region is also known as Bari Doab because it lies between Beas and Ravi River. The confluence (merger) of the two rivers Beas and Sutlej occurs at Harike in Tarn taran and the area on the north of Sutlej extending up to the Ravi River also falls in the Majha region in India.

The groundwater samples were collected from hand pumps and submersibles. The depth range for hand pumps is between 20-220 feet and for submersibles is between 30-500 feet. The samples were collected in polythene vials having screw-on caps and these were kept at room temperature for further analysis.

Figure 3.1 Four districts of Majha region from where samples were collected are shown with names [87]
(*Note:- Map is not up to scale, it is just indicative*)
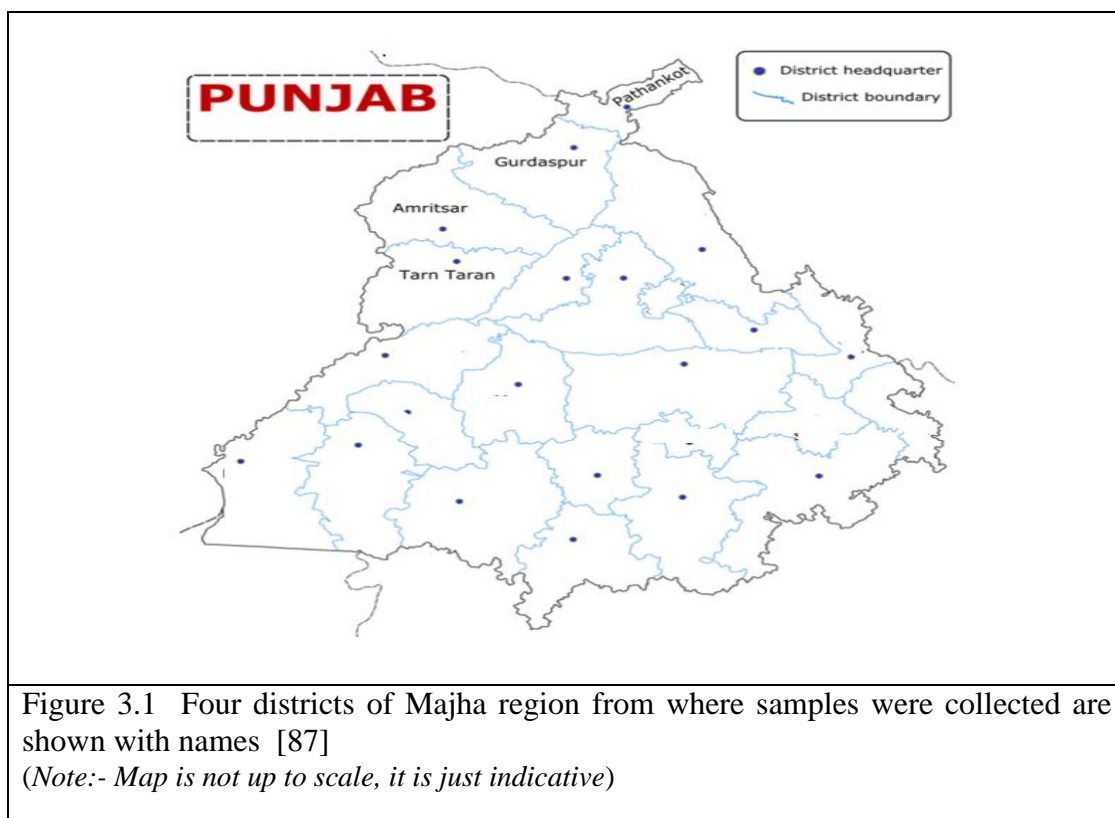
Each groundwater sample was analysed for pH, chlorine, total dissolved solids, total hardness, turbidity, electrical conductivity using standard procedures. Samples are given

codes starting with the alphabet, which is first letter of the name of the district like A for Amritsar, T for Tarn taran, P for Pathankot and G for Gurdaspur followed by count number, which are in continuity from 1 to 60. From every district, 15 samples were collected randomly from urban and rural areas from sources like hand pumps and submersibles (which are also known as motors or tube wells by the local people). Table 3.1 and Table 3.2 give a brief summary of the samples collected on the basis of source and region respectively. The Table 3.3 gives the details of the groundwater samples collected.

Table 3.1: Distribution of the collected samples on the basis of district and source of sample

District	Source of Sample	Sample Labels	Total Samples Collected	Wrongly Coded Samples
Pathankot		P1 To P14,P16	**15**	P15 Belongs to Gurdaspur
	Hand pump		7	
	Submersible		8	
Gurdaspur		P15, G17 To G30	**15**	
	Hand pump		7	
	Submersible		8	
Tarn taran		T31 To T44, A57	**15**	T45 Belongs to Amritsar
	Hand pump		2	
	Submersible		13	
Amritsar		T45, A46 To A56, A58-A60	**15**	A 57 Belongs to Tarn taran
	Hand pump		1	
	Submersible		14	

Table 3.2 Distribution of the collected samples on the basis of district and region

District	Region	Sample Labels	Total Samples Collected	Wrongly Coded Samples
Pathankot		P1 to P14,P16	**15**	P15 belongs to Gurdaspur
	Urban		2	
	Rural		13	
Gurdaspur		P15, G17 to G30	**15**	
	Urban		2	
	Rural		13	
Tarn taran		T31 to T44, A57	**15**	T45 belongs to Amritsar
	Urban		0	
	Rural		15	
Amritsar		T45, A46 to A56, A58-A60	**15**	A 57 belongs to Tarn taran
	Urban		0	
	Rural		15	

Table 3.3 Description of the samples- sample labels, depth, source, districts, region and name of place.

Sr. No.	Sample Label	Source of Sample	Depth in Feet	District	Region	Name of Place
1	P1	Submersible	110	Pathankot	Rural	Manwal
2	P2	Submersible	110	Pathankot	Rural	Anandpur Rara
3	P3	Hand Pump	20	Pathankot	Rural	Asabaan
4	P4	Submersible	35	Pathankot	Rural	Sunder Chak
5	P5	Hand Pump	30	Pathankot	Rural	Kotli
6	P6	Submersible	30	Pathankot	Rural	Dimko Dhlerian
7	P7	Hand Pump	30	Pathankot	Rural	Frida Nagar
8	P8	Submersible	30	Pathankot	Rural	Dhobra
9	P9	Hand Pump	25	Pathankot	Rural	Balsuya
10	P10	Submersible	120	Pathankot	Urban	Sarna
11	P11	Submersible	120	Pathankot	Urban	Dhangu Road
12	P12	Submersible	120	Pathankot	Rural	Rajprura
13	P13	Hand Pump	30	Pathankot	Rural	Rashpalma
14	P14	Hand Pump	35	Pathankot	Rural	Shahidpur
15	P15	Submersible	-	Gurdaspur	Rural	Nakki Gurdaspur
16	P16	Hand Pump	120	Pathankot	Rural	Parmanand
17	G17	Hand Pump	35	Gurdaspur	Rural	Dinangar
18	G18	Hand Pump	60	Gurdaspur	Urban	Prabodh Chander Nagar
19	G19	Hand Pump	70	Gurdaspur	Urban	Gurdaspur
20	G20	Submersible	105	Gurdaspur	Rural	Slumpr Koli

Table 3.3**:** – continue –

Sr. No.	Sample Label	Source of Sample	Depth in Feet	District	Region	Name of Place
21	G21	Hand Pump	40	Gurdaspur	Rural	Bkhshiwal
22	G22	Submersible	70	Gurdaspur	Rural	Kila Nath Singh
23	G23	Submersible	250	Gurdaspur	Rural	Kotla Mugla
24	G24	Hand Pump	60	Gurdaspur	Rural	Kalanaur
25	G25	Submersible	95	Gurdaspur	Rural	Hakim Pur
26	G26	Hand Pump	40	Gurdaspur	Rural	Bhagowal
27	G27	Hand Pump	55	Gurdaspur	Rural	Dharmkot
28	G28	Submersible	130	Gurdaspur	Rural	Shukrpura
29	G29	Hand Pump	35	Gurdaspur	Rural	Vadala Granthiyan
30	G30	Submersible	130	Gurdaspur	Rural	Cheema
31	T31	Submersible	250	Tarn taran	Rural	Bhorwal
32	T32	Submersible	180	Tarn taran	Rural	Vein Puin
33	T33	Submersible	120	Tarn taran	Rural	Lal Pura
34	T34	Submersible	280	Tarn taran	Rural	Rashiana
35	T35	Submersible	400	Tarn taran	Rural	Aladin Pur
36	T36	Submersible	250	Tarn taran	Rural	Kahlvan
37	T37	Submersible	120	Tarn taran	Rural	Gandivind
38	T38	Submersible	500	Tarn taran	Rural	Chambal
39	T39	Submersible	240	Tarn taran	Rural	Malia
40	T40	Submersible	240	Tarn taran	Rural	Kaerowal

Table 3.3: – continue –

Sr. No.	Sample Label	Source of Sample	Depth in Feet	District	Region	Name of Place
41	T41	Hand Pump	140	Tarn taran	Rural	Jhval
42	T42	Submersible	300	Tarn taran	Rural	Jhval
43	T43	Submersible	120	Tarn taran	Rural	Manan
44	T44	Submersible	120	Tarn taran	Rural	Lallu Ghuman
45	T45	Submersible	120	Amritsar	Rural	Mandiala
46	A46	Submersible	100	Amritsar	Rural	Ann Garh
47	A47	Submersible	130	Amritsar	Rural	Rajasansi
48	A48	Submersible	350	Amritsar	Rural	Harshe Shinna
49	A49	Submersible	120	Amritsar	Rural	Ftehgarh Churian
50	A50	Submersible	180	Amritsar	Rural	Bal Khurd
51	A51	Submersible	120	Amritsar	Rural	Verka
52	A52	Submersible	80	Amritsar	Rural	Vllah
53	A53	Submersible	120	Amritsar	Rural	Manawala Kalan
54	A54	Submersible	130	Amritsar	Rural	Manawala
55	A55	Submersible	150	Amritsar	Rural	Nijrpura
56	A56	Hand Pump	220	Amritsar	Rural	Bundala
57	A57	Hand Pump	40	Tarn taran	Rural	None Adda
58	A58	Submersible	120	Amritsar	Rural	Nangal Guru
59	A59	Submersible	120	Amritsar	Rural	Jandiala Guru
60	A60	Submersible	230	Amritsar	Rural	Mliaan

3.5 REFERENCE ANALYSIS

All the samples were tested for the following physicochemical properties- turbidity, total hardness, pH value of water, electrical conductivity, total dissolved solids and chlorine content. The pH value was measured using the pH digital meter. TDS was measured by filterable residual method. Electrical conductivity is measured using Wheatstone bridge method. To determine the chlorine content in the water a chemical method was used. Turbiditymeter was used to determine the turbidity of water in NTU unit. Total hardness is also calculated using standard procedures. All the procedures are defined by BIS under the IS 3025 with different part number for individual parameter. All the tests were done at Kalpin Watertech Laboratories, Thane (W). It is observed that these are the important physicochemical properties of water which are helpful in characterization of water from a particular region. For example it is observed from the literature that TDS is correlated with the presence of uranium in the water samples tested from the Malwa region of Punjab India. Higher concentration of the presence of ions (may be heavy metal ions, chloride, fluoride, nitrate etc.) in water is the indication of high value of electrical conductivity. If water is cloudy, translucent, or opaque then turbidity is high which indicates presence of undissolved impurities in water. If the pH value of water is very less than 7 (neutral value) then it is acidic and if value is greater than 7 then the water is alkaline. It is important to know the physicochemical properties of water for its characterization and quality assessment.

3.6 SPECTRAL ANALYSIS

Visible-Near infrared spectra, FTIR spectra and fluorescent spectra were obtained for all the groundwater samples. FTIR spectra were obtained using Perkin Elmer spectroscope and Fluorescence spectra were also obtained using Perkin Elmer spectroscope both available at central instrumentation facility, Lovely Professional University, Phagwara,, Punjab, India. NIR spectra were obtained using NIR DS2500 Spectrometer which is available at Council of Scientific and Industrial Research– Central Scientific Instrumental Organization. The range for NIR spectra is obtained for wavelengths ranging from 400 - 2500 nm with increments of 0.5 nm. Samples were placed in glass beakers with gold reflectors. There are two detectors in the spectrometer– one is silicon (Si) detector which was used for wavelengths ranging from 400 - 1100 nm and the other is lead sulphide (PbS) detector used for wavelengths ranging from 1100 - 2500 nm. Data were obtained in reflectance mode in triplicate which then was converted by the inbuilt VISION software of the instrument in to absorbance type. The

spectrometer is pre-dispersive in nature because polychromatic light is made to disperse in different wavelengths before illuminating the sample. The instrument is manufactured and designed in this way that it gives accurate results, because it is resistant to temperature changes, dust, vibrations and moisture etc.

The FTIR spectra were obtained using Perkin Elmer Diamond ATR Fourier Transform Infrared Spectrometer. It has a resolution of 0.5 cm^{-1}. The range of wavenumbers for the near infrared radiations used to obtain the spectral data is 350 - 8300 cm^{-1}. The instrument has the facility of atmospheric vapour compensation (AVC) to subtract the water and carbon dioxide absorption.

Fluorescence spectroscopy was performed with the help of Perkin Elmer spectrophotometer which is available at Central instrumentation facility center at LPU, Punjab, India. The source of light in the instrument is Xenon pulse lamp. The samples are illuminated with this light to obtain the fluorescence data. Wavelength range 200 - 900 nm is available for excitation and emission spectra. The instrument has a slit of variable width which can be varied from 1 - 20 nm.

Figure 3.2 NIR DS2500 Spectrometer	Figure 3.3 Sample holder for the NIR DS2500 instrument with gold reflector

Figure 3.4 FTIR Spectrophotometer (Perkin Elmer)	Figure 3.5 Fluorescence Spectrophotometer (Perkin Elmer)

3.7 STATISTICAL ANALYSIS

One of the branches of statistics is descriptive statistics which deals with the analysis and interpretation of quantitative data. It is used to summarize and present data in a meaningful way. The tools or measures involved are central tendency (e.g. mean, median, etc.), variability (e.g. range, variance, standard deviation), and correlation (e.g. correlation coefficient). Descriptive statistics provides a way to describe and analyze data in a clear and concise manner. It enables researcher to draw conclusions and make inferences based on the findings.

3.7.1 Measures of Central Tendency

Every numerical data is associated with three statistical terms namely Mean, Median and Mode which are considered as measures of central tendency. Whole of the data is scattered around these central values [27]. A common term Average is used for these three central measures. In the present work only Mean and Median are considered.

Mean: It is the most common average or measure of central tendency. It can be easily calculated by summing all the data values and dividing the sum by the total number of data points. It is also known as the arithmetical mean.

$$\text{Mean} = \bar{x} = \frac{x_1 + x_2 + x_3 + ... + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{3.1}$$

Here, x_i is the value of ith data point and n is the total number of data points.

Median: It is one of the measures of central tendency. To find the Median whole of the data is either arranged in ascending or descending order and then the data point at the middle of the this data is the median.

3.7.2 Measure of Dispersion

The central tendencies give single values around which data is scattered. But these values do not tell us about the spread of data.

To make it clear consider two data sets A and B.

$$A = \{4, 4, 4, 4, 4, 4, 4, 4, 4\}, \text{ mean= 4, median= 4}$$
$$B = \{1, 2, 3, 3, 4, 5, 5, 6, 7\}, \text{ mean= 4, median= 4}$$

It is clear from the data set A there is no dispersion or spreading of the data and all the data values are same equal to 4, but for data set B mean and median are also 4 but the data is varying from 1 to 7. Clearly the two data sets are different from each other but for someone who is not aware of the data set and only knows the mean and median values, it will not be possible for him or her to make any difference in the data sets on the basis of central tendencies. For understanding the spread of data, we have statistical metrics known as measures of dispersion, these are also known as measures of variability or measures of spread. These measures indicate the extent, the data values are scattered or spread out around the central tendency (mean, median, or mode) in a dataset. Some of the common measures of dispersion are:

Range: It is the simplest measure of the dispersion and it is also easy to calculate. It is simply the difference between the maximum and minimum values in a dataset.

$$\text{Range} = (\text{maximum value in data set}) - (\text{minimum value in data set})$$

Mean Absolute Deviation (MAD): It is the average of the absolute deviations from the mean. Consider the following data set

Table 3.4 Table showing random data set for deviation

Data values	Deviations from mean	Absolute deviation from mean
x_1	$x_1 - \bar{x}$	$\left| x_1 - \bar{x} \right|$
x_2	$x_2 - \bar{x}$	$\left| x_2 - \bar{x} \right|$
x_3	$x_3 - \bar{x}$	$\left| x_3 - \bar{x} \right|$
....	....	....
x_n	$x_n - \bar{x}$	$\left| x_n - \bar{x} \right|$

$$\text{MAD} = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + |x_3 - \bar{x}| + \dots\dots\dots + |x_n - \bar{x}|}{n} = \frac{\sum |x_i - \bar{x}|}{n} \qquad (3.2)$$

It is necessary to calculate it as it is used in the calculation of variance.

Variance: The average of the squared deviations from the mean.

$$\textit{Variance} = \frac{|x_1 - \bar{x}|^2 + |x_2 - \bar{x}|^2 + |x_3 - \bar{x}|^2 + \dots\dots\dots + |x_n - \bar{x}|^2}{n} = \frac{\sum |x_i - \bar{x}|^2}{n} \qquad (3.3)$$

The units of Variance are square of the units used to measure the data value *x*.

Standard Deviation: It is simply the square root of the variance. it is represented by σ

$$\sigma = \sqrt{\frac{\sum |x_i - \bar{x}|^2}{n}} \qquad (3.4)$$

The units of standard deviation are same as the units of data value *x*.

3.7.3 Measures of Strength of Relation

The measure of correlation refers to statistical methods used to quantify the strength and direction of relationships between two or more variables. Common measures of relation include correlation coefficient, regression analysis, coefficient of determination (R^2) etc. Simple linear regression (also known as bivariate regression) analysis is relation between two variables and a linear line is fitted into the data set, whereas multiple regressions include multiple variables. The correlation coefficient is denoted by *r* and its numerical value ranges from -1 to 1 without any units. The value of *r* = 1 indicates a perfect positive correlation which means the two variables in question increase or decrease together. The value of *r* = -1 indicates a perfect negative correlation which means the two variables in question move in opposite directions while one increases the other decreases and vice versa. The value of *r* = 0 indicates that there is no linear relationship between the variables. If the value is very close to 1 or -1 the stronger the correlation between the two variables. Values in between indicate the degree of linear dependence, with values closer to zero indicate a weaker relationship or no relationship. Correlation only measures the association between variables, not causation. Coefficient of correlation [88] is calculated using the following formula for two variable *x* and *y*

$$r = \frac{\text{cov}(x, y)}{st\,dev(x).st\,dev(y)} \qquad (3.5)$$

Correlation and Bivariate Regression Analysis

To know the correlation, the physicochemical data of the samples was fed to the MS Excel. The MS Excel provides a correlation table providing the correlation coefficient values

between all the pairs of physicochemical parameters. MS Excel is also used to obtain the regression analysis between two variables, the software has inbuilt application and it can provide the plot showing the best fit line along with the data points and also give the value of coefficient of determination.

3.7.4 Multivariate Analysis

Multivariate analysis [89 - 90] is a statistical method which is used when multiple variables are involved. It is used to find the relationship and patterns among these multiple variables simultaneously. The spectral data is complex in nature and it is not easy to directly know the wavelength sets responsible for the quantity and quality analysis of the ground water samples. In this situation there are multiple variables involved so multivariate analysis is required to be done to get regression equations. Multivariate analysis is used to identify, from the spectral data, the most informative wavelengths or spectral regions which show correlation with specific physicochemical property of the water sample. For the principal component analysis, partial least squares regression analysis and interval partial least square regression are used. These techniques are helpful in reducing the dimensionality of the data and to identify patterns and correlations and it also helps in the development of predictive models which can estimate water quality parameters like pH, turbidity, electrical conductivity, chlorine content, total dissolved solids and total hardness from spectral measurements. Multivariate analysis involving principal component, partial least square regression, interval partial least square regression is done using Camo unscramble×10.5 version [91] and clustvis webtool softwares.

3.7.5 Principal Component Analysis

Principal component analysis [92] is a statistical technique falls under the umbrella of multivariate analysis and is useful when the data set is high dimensional. This technique is used for the reduction in the dimensions i.e. number of variables is decreased while the variance from the original data set is preserved as much as possible. The new uncorrelated variables are obtained which are known as principle components. The data analysis becomes easy after the reduction in the dimensions and the data can be visualized easily because data can be seen as projected in 2D or 3D. It also helps in the reduction of the noise in the data. The data is standardized by making mean of every variable equal to zero and the standard deviation equal to 1, this is done so that all variable have same scale and have same importance in the calculations. The new variables are obtained from the linear combination of

the original variables and are known as first principle components. Second principal components can also be obtained which are orthogonal to the first PC. The first principle component captures the maximum possible variance and the second PC captures next maximum variance. The projected data in all PCs when represented in matrix form or linear equations the columns are known as loadings and the rows are known as scores. Loadings indicate how much each original variable (i.e. the weightage to be given) contributes to each principal component and scores are the coordinates of the original data in the new principal component space. Mainly PCA is used to reduce the dimensions of the large data set.

3.7.6 Partial Least Square Regression

While using Linear regression a model can be developed with the relationship between a dependent variable (response) and one or more independent variables (predictors). This method is suitable when simpler models with fewer predictors and low multi-collinearity are required. But if the dataset is complex, then linear regression (also known as least square regression) is to be replaced with another statistical method called Partial Least Square Regression (PLSR), which is more suitable for complex datasets with many predictors and high multi-collinearity. PLSR can handle the situations where the number of predictors is more than the number of observations. PLSR aims to find the latent variables that explain the maximum variance in both the predictors and the response variables. PLSR takes into consideration both principal component analysis (PCA) and multiple linear regression simultaneously. It splits the predictor variables into a set of orthogonal components as is done in PCA which are linear combinations of the original predictors. These components so obtained are known as latent variables or factors. These are then used to model the relationship between the predictors and the response variable. While dealing with water sample parameters and corresponding spectral data obtained from the water samples PLSR is used to describe the relationship between the physicochemical parameters (Y) and spectral data (X). A predictive model is obtained iteratively refining X and Y until a suitable model is reached. In this work, for doing PLSR, unscramble $\times 10.5$ software is used which is available at LPU Phagwara Punjab.

3.7.7 Interval Partial Least Squares

Interval partial least squares (iPLS) method is a modeling method used for spectral data analysis. The whole spectrum is split into equal disjoint intervals and PLS regression model is developed for each interval known as local regression model. In this way the

wavelengths which give good results are selected and are important wavelength intervals. The unimportant wavelength intervals are neglected. iPLS has been applied in different fields which include spectroscopy, data analysis etc. iPLS improves the prediction ability of NIR regression models and in identifying important wavelengths and intervals for spectral data analysis [93]. The regression models for all the intervals are then compared and the models showing high value of root mean square error (RMSE) or less value of coefficient of determination (R^2) are neglected. It can be concluded that interval partial least square regression is a powerful tool for spectral data analysis.

3.8 CONCLUSION

In the present work, the groundwater samples were collected from the districts: Amritsar, Tarantaran, Pathankot and Gurdaspur, which make the Majha region of Punjab, India. The samples were collected from the hand pumps and submersibles randomly. The samples were collected in polythene vials which had the screw-on caps. All the water samples were analysed for the physicochemical properties viz. pH, Chlorine, total dissolved solids, total hardness, turbidity and electrical conductivity. Spectroscopic analysis of all the samples was also done. The spectroscopic techniques used were NIR, FTIR and Fluorescence spectroscopy. The statistical analysis included: the mean, standard deviation, correlation coefficient, principal component analysis and partial least square and interval least square regression for developing the predictive model.

The sample collection, spectroscopic techniques and traditional chemical analysis helped in achieving the objectives of the research such as determination of the pollutants using spectroscopy and chemical analysis. The multivariate analysis helped in the estimation of the contaminants and then the advanced statistical techniques like PLSR helped in the development of the predictive model for the determination of the pollutants in the groundwater of the Majha region of Punjab, India.

The spectroscopic techniques like NIR, FTIR, and fluorescence spectroscopy are used to detect and identify various types of water pollutants using the obtained spectral data. The use of principal component analysis helps in the simplification and highlighting the key patterns in complex spectroscopic data. The PLSR is used to build predictive models to accurately relate these patterns to specific pollutant concentrations. The integration of all these spectroscopic methods and statistical techniques helps in uncovering the hidden trends, identify outliers, and improve quantification without going into the tedious and laborious chemical methods. This integrated approach is quick, reliable and can provide precise

measurements of contaminants in water samples. In the present work, the above said methods have been used for the collection of samples, spectral data collection, analysis for classification, correlation and regression analysis.

CHAPTER 4

RESULTS AND DISCUSSION

CHAPTER 4 - RESULTS AND DISCUSSION

4.1 INTRODUCTION

Contamination of Water: Water is one of the most important ingredients of life, without it we cannot imagine existence of life. If we consume water which is contaminated or polluted, it can badly affect our health and also the health of animals those who drink this water. The cause of contamination can be geogenic or anthropogenic. The contaminants which are found from the literature review include heavy metals like Cd, As, Hg, Pb, Cr etc. [94 - 95]. The heavy metal contamination and other contaminations of water bodies can be from industrial activities, from agricultural practices, like the use of the insecticides, pesticides, fertilizers etc. The toxic metals or chemicals leach into the underground water and makes it unfit for drinking purpose. We cannot even use contaminated water for irrigation purposes. Urea is one of the main fertilizers used in agriculture and there are chances that it may have contaminated the ground water, arsenic is one of the heavy metal contaminant which may have reached groundwater because of natural or human activities. In a study [96] it is mentioned that Majha region falls under a category where 60% of the habitation are affected by arsenic. In the present work focus is on the arsenic and urea contamination of the ground water from the Majha region of Punjab, India. Arsenic is very dangerous for humans. It can severely affect our health due to its long exposer through ground water. It can cause cancer, skin diseases, cardiovascular problems, diabetes etc. Urea is nowadays a very important fertilizer which is in use by the agriculturists and it can also pollute groundwater. Its consumption through water can be dangerous to human health. Spectroscopy is rapid, non-destructive, quantitatively very strong analytical technique and can be used to detect the contaminants in water. When electromagnetic radiations of suitable frequency interact with some suitable molecules then energy changes at different levels may occur known as transitions in the molecules these may be rotational, vibrational, or electronic transitions i.e. electrons jump to higher energy levels, a combination of transitions may occur depending on the amount of energy provided by the EM radiations. The transitions which take place in the molecule help in understanding the bonds present in it. The spectroscopic techniques which were employed in this work are Near Infra-Red, FTIR and fluorescent spectroscopy. In a study near infrared spectroscopy was used to predict the macronutrients in the soil [97] and there are many more studies where NIR and FTIR are used. Arsenic content was studied in

groundwater samples in Bihar India using UV spectroscopy [98]. Urea is also known as carbamide because it is diamide of carbonic acid, it contains nitrogen in it. If nitrogen can be measured in a sample it can help in knowing the water contaminants which have nitrogen. In a study soil nitrogen was detected using NIR [99]. Researchers have found that contaminated water show variations in its physicochemical properties like electrical conductivity, pH, TDS, chlorine content etc. [100]. In this chapter focus is on the study of pollutants in water samples collected from the Majha region of Punjab, India. The spectroscopic techniques used are NIR, FTIR, and Fluorescence spectroscopy.

4.2 RESULTS AND DISCUSSION

FTIR spectroscopy was employed on the collected groundwater samples for getting the peaks corresponding to the bonds of pollutants, in this case mainly for urea and arsenic. For reference FTIR was obtained for pure urea and arsenic trioxide. The ATR-FTIR spectrometer is available at CIF lab, Lovely Professional University, Phagwara. For the pure urea and arsenic trioxide, transmittance spectra peaks were obtained at characteristic wavenumber values. These two spectral plots for urea and arsenic trioxide are respectively shown in Figure 4.1 and Figure 4.2.

4.2.1 Analysis of FTIR Spectra of Urea

From the FTIR spectrum of urea it is observed that there are peaks corresponding to C-N bond, C-O bond, N-H bond and C=O bond. According to theory[§] [101] and literature review it is observed that N-H stretching bonds may be present at the wavenumbers (theoretical value 3012 cm^{-1}) ranging from 3300 - 3400 cm^{-1}. The theoretical values are calculated under simplified conditions and considering approximate values of the constants, without applying restrictions and environmental conditions around the bond in question, so there is always mismatch between theoretical and experimental values but it gives an idea of wavenumber close to the experimental value. There are two more peaks observed ranging from 3000-3300 cm^{-1} due to phase stretching vibrations. There are some peaks observed between 1500-2000 cm^{-1} due to C=O (theoretical value 1572 cm^{-1}) and H-N-H bonds due to bending (from literature value is near 1600 cm^{-1}). The peaks at wavenumbers near 1100 cm^{-1}

[§] stretching wavenumber can be calculated from the formula (using the values of Avogadro number (N) and speed of light)

$\bar{v}(in\, cm^{-1}) = \dfrac{1}{2\pi c}\sqrt{\dfrac{K.N}{\mu}} = 4.12\sqrt{\dfrac{K}{\mu}}$, spring constant K= n.(5x10^5)dyne/cm; n= 1 for single, 2 for double and 3 for triple bond,

for C=C, reduced mass μ =6, K=10×10^5 dyne/cm calculated wavenumber =1682 cm^{-1} and experimental value is approx.1650 cm^{-1}. Similarly for C-H single bond μ = 0.923 so calculated wavenumber is 3032 cm^{-1} and experimental value is approx. 3000 cm^{-1}.

(theoretical value 1145 cm^{-1}) are observed which may correspond to C-N stretching modes. All the peaks become the reference for comparison with the spectra of the groundwater samples for the presence of urea. In the NIR spectrum, the first harmonic corresponds to the fundamental wavelength observed in FTIR and the second harmonic in NIR may consist of overtones or combinations of these fundamental vibrations, from these consideration we can get additional spectral information. First overtone is obtained by multiplying the fundamental frequency (or wavenumber) by a factor of 2 and second overtone is thrice the value of fundamental frequency (or wavenumber). In terms of wavelength we have to divide the fundamental wavelength by 2 for getting the first overtone and so on.

4.2.2 Analysis of FTIR Spectra of Arsenic Trioxide

Structure of Arsenic trioxide is somewhat similar to O=As–O–As=O, it is given here just for the sake of bonds, there are double and single bonds like As=O (theoretical value 1133 cm^{-1}) and As-O (theoretical value 801 cm^{-1}) in the arsenic trioxide. In the transmittance spectra of arsenic trioxide shown in the Figure 4.2, the signature peaks are observed around 3500-4000 having high transmittance, 2300-2400 cm^{-1} with nearly 80 percent transmittance and 770 cm^{-1} with only 45 percent transmittance and there are some high transmittance peaks in the range from 1000 - 2000 cm^{-1}. These peaks act as reference and are used to compare with the spectra of the ground water samples for the presence of arsenic and their overtone wavelengths were compared with the NIR spectra of the ground water samples.

Four random FTIR spectra of ground water samples taken from four districts of Majha regions of Punjab India are given in Figure 4.3 to Figure 4.6. From the plots it is observed that there are peaks ranging from 500 - 600 cm^{-1} with transmittance of 80 - 85 percent, 1600 - 1700 cm^{-1} with transmittance around 70 - 75 percent, 2000 - 2400 cm^{-1} with transmittance around 95 percent and another peak around 3300 - 3500 cm^{-1} with transmittance around 50 percent. These peaks are similar to water peaks and the presence of arsenic and urea can be ruled out. All the samples are compared for urea and arsenic peaks and the results are given in the Table 4.1. When all the samples are analysed it is observed that corresponding peaks for arsenic and urea are not present, so it can be concluded that both the pollutants are not present in the water samples. In the Table 4.2 the possibility of bonds which may present in FTIR spectra of ground water samples is also given.
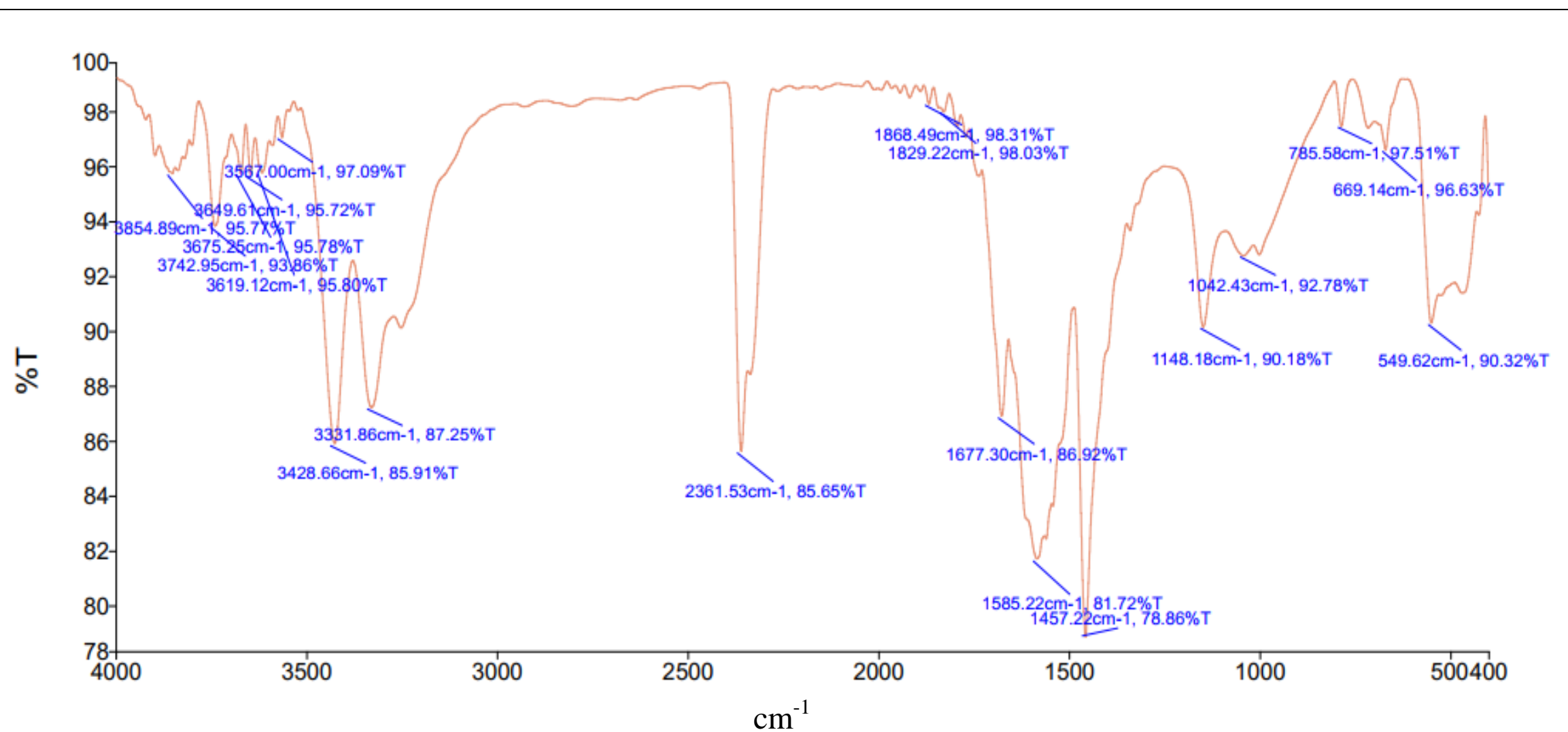
Figure 4.1 Fourier transform spectra of Urea using the ATR-FTIR spectrometer at CIF, LPU
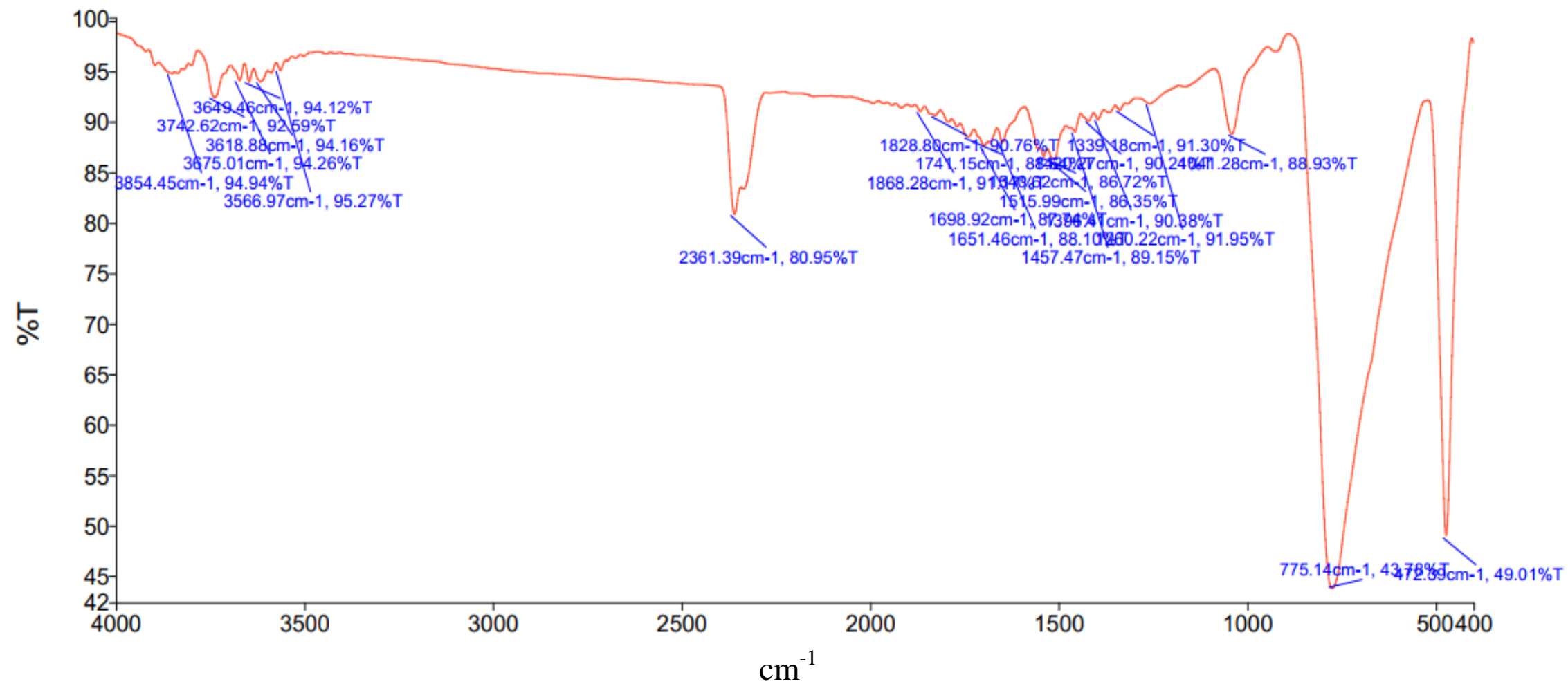
Figure 4.2 Fourier transform spectra of As₂O₃ using the ATR-FTIR spectrometer at CIF, LPU

Figure 4.3 FTIR spectra of   ground water Sample number 3
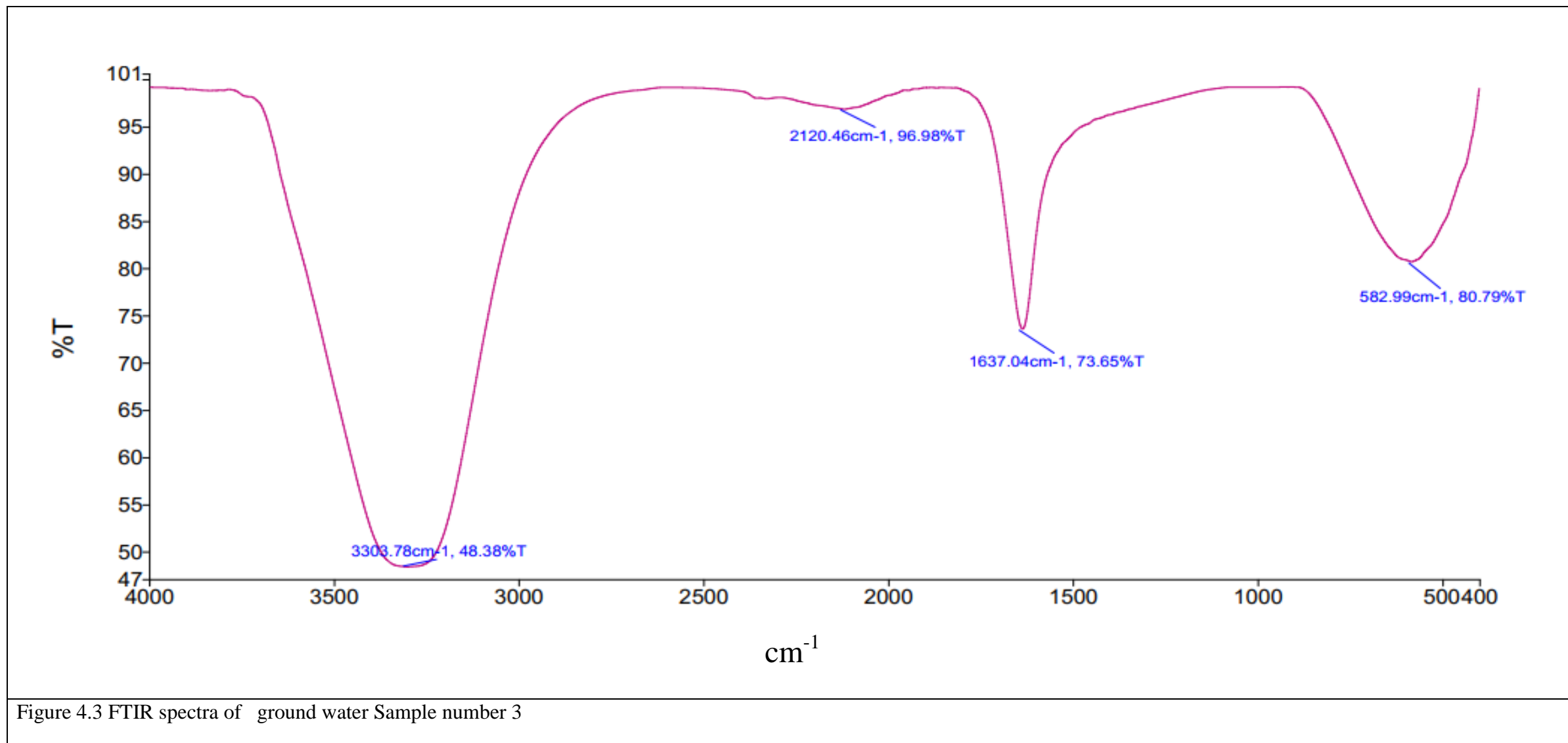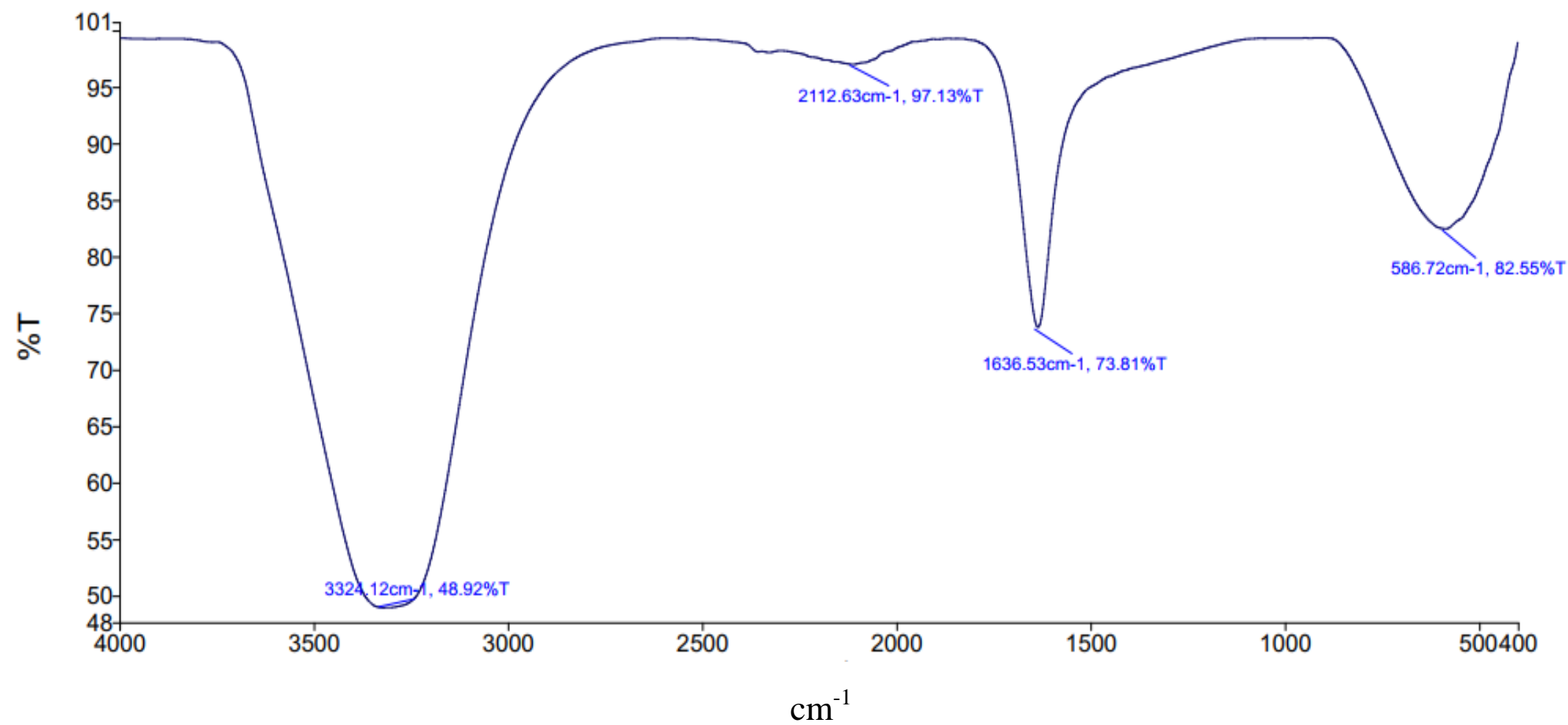
Figure 4.4 FTIR spectra of   ground water Sample number 19
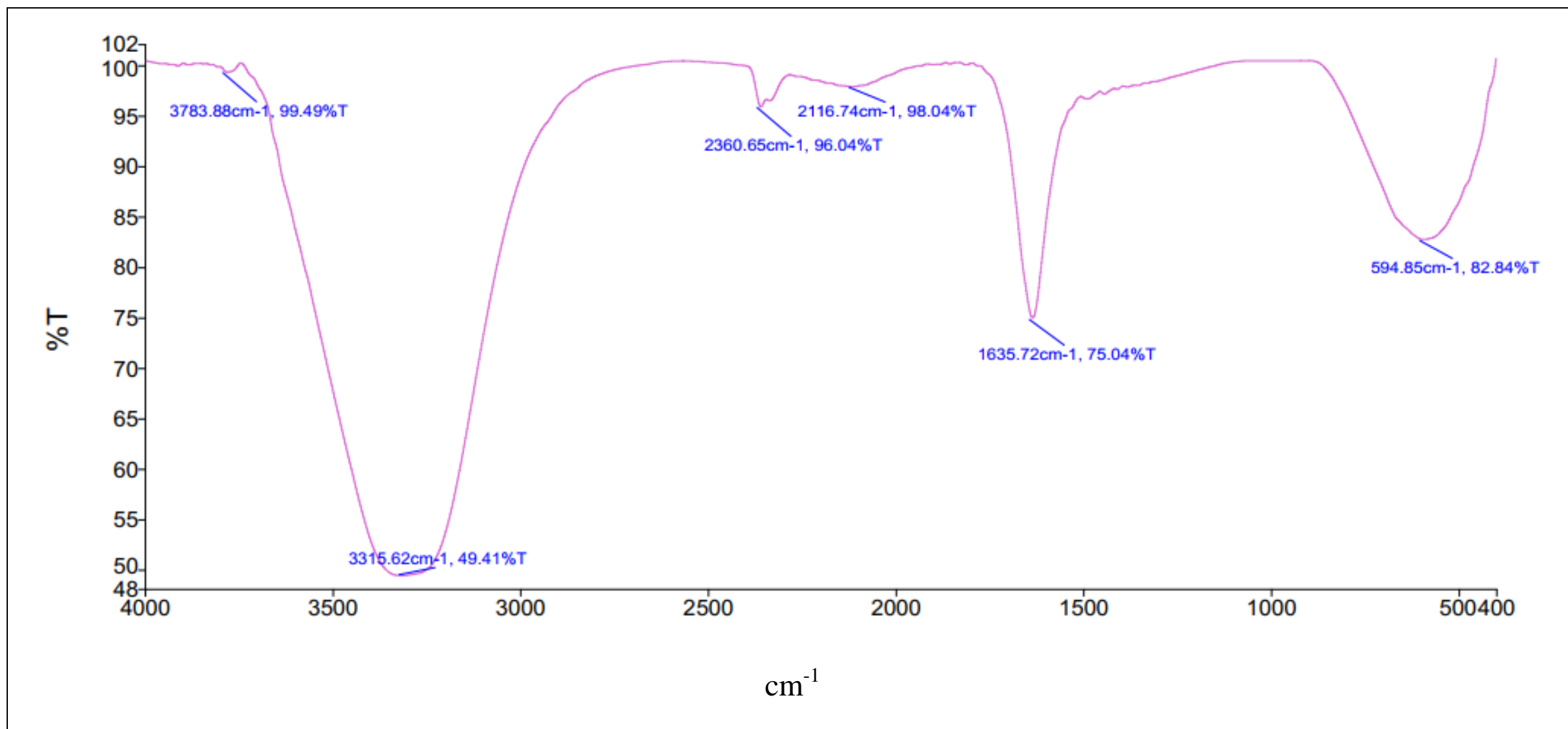
Figure 4.5 FTIR spectra of   ground water Sample number 35
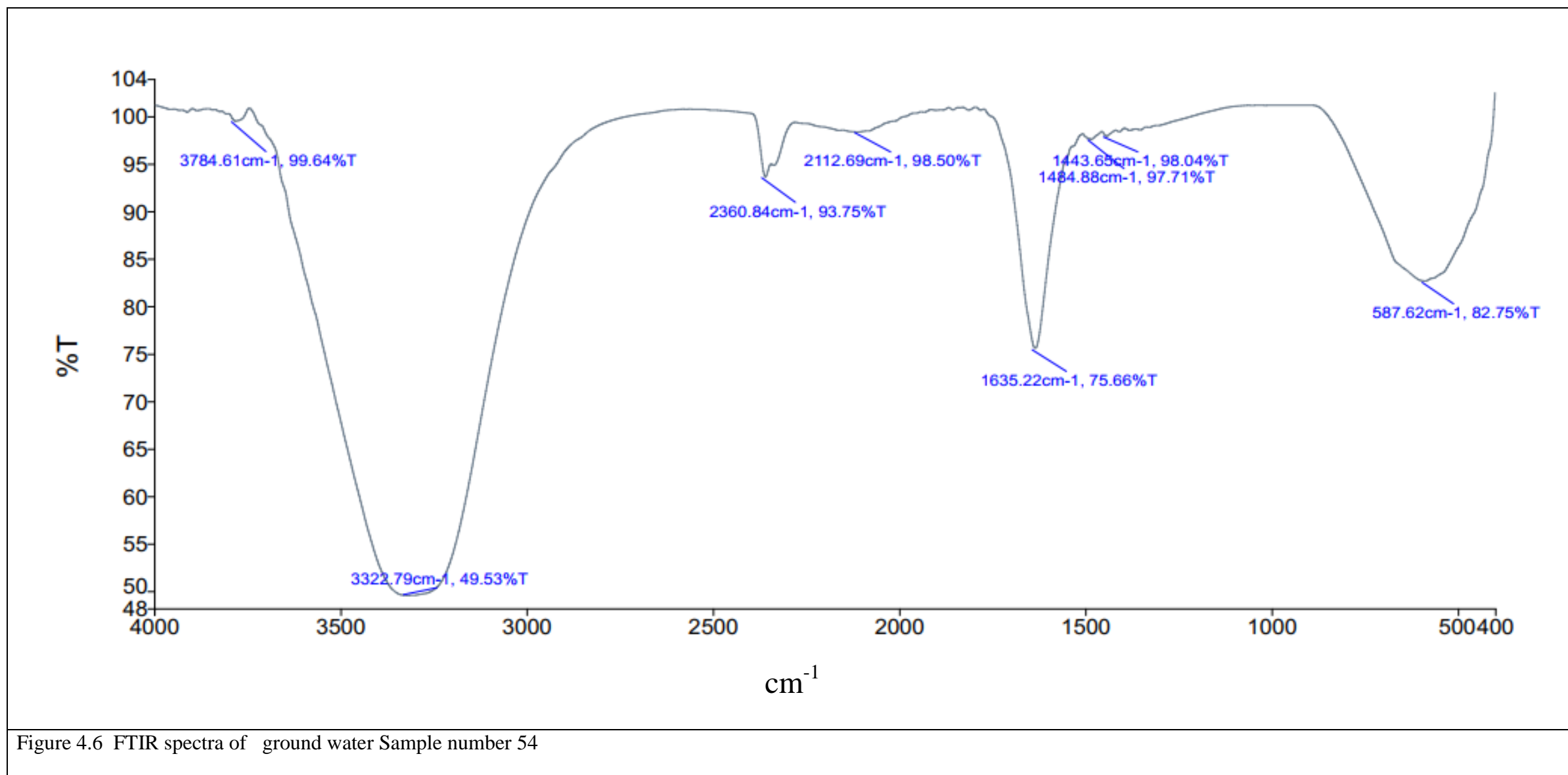
Figure 4.6 FTIR spectra of   ground water Sample number 54

Table 4.1 Description of the samples and presence of urea and arsenic trioxide

Sr. No.	Sample Label	Source of Sample	Depth in Feet	District	Region	Urea	Arsenic trioxide
1	P1	Submersible	110	Pathankot	Rural	Not Present	Not Present
2	P2	Submersible	110	Pathankot	Rural	Not Present	Not Present
3	P3	Hand Pump	20	Pathankot	Rural	Not Present	Not Present
4	P4	Submersible	35	Pathankot	Rural	Not Present	Not Present
5	P5	Hand Pump	30	Pathankot	Rural	Not Present	Not Present
6	P6	Submersible	30	Pathankot	Rural	Not Present	Not Present
7	P7	Hand Pump	30	Pathankot	Rural	Not Present	Not Present
8	P8	Submersible	30	Pathankot	Rural	Not Present	Not Present
9	P9	Hand Pump	25	Pathankot	Rural	Not Present	Not Present
10	P10	Submersible	120	Pathankot	Urban	Not Present	Not Present
11	P11	Submersible	120	Pathankot	Urban	Not Present	Not Present
12	P12	Submersible	120	Pathankot	Rural	Not Present	Not Present
13	P13	Hand Pump	30	Pathankot	Rural	Not Present	Not Present

14	P14	Hand Pump	35	Pathankot	Rural	Not Present	Not Present
15	P15	Submersible	-	Gurdaspur	Rural	Not Present	Not Present
16	P16	Hand Pump	120	Pathankot	Rural	Not Present	Not Present
17	G17	Hand Pump	35	Gurdaspur	Rural	Not Present	Not Present
18	G18	Hand Pump	60	Gurdaspur	Urban	Not Present	Not Present
19	G19	Hand Pump	70	Gurdaspur	Urban	Not Present	Not Present
20	G20	Submersible	105	Gurdaspur	Rural	Not Present	Not Present
21	G21	Hand Pump	40	Gurdaspur	Rural	Not Present	Not Present
22	G22	Submersible	70	Gurdaspur	Rural	Not Present	Not Present
23	G23	Submersible	250	Gurdaspur	Rural	Not Present	Not Present
24	G24	Hand Pump	60	Gurdaspur	Rural	Not Present	Not Present
25	G25	Submersible	95	Gurdaspur	Rural	Not Present	Not Present
26	G26	Hand Pump	40	Gurdaspur	Rural	Not Present	Not Present
27	G27	Hand Pump	55	Gurdaspur	Rural	Not Present	Not Present

28	G28	Submersible	130	Gurdaspur	Rural	Not Present	Not Present
29	G29	Hand Pump	35	Gurdaspur	Rural	Not Present	Not Present
30	G30	Submersible	130	Gurdaspur	Rural	Not Present	Not Present
31	T31	Submersible	250	Tarn taran	Rural	Not Present	Not Present
32	T32	Submersible	180	Tarn taran	Rural	Not Present	Not Present
33	T33	Submersible	120	Tarn taran	Rural	Not Present	Not Present
34	T34	Submersible	280	Tarn taran	Rural	Not Present	Not Present
35	T35	Submersible	400	Tarn taran	Rural	Not Present	Not Present
36	T36	Submersible	250	Tarn taran	Rural	Not Present	Not Present
37	T37	Submersible	120	Tarn taran	Rural	Not Present	Not Present
38	T38	Submersible	500	Tarn taran	Rural	Not Present	Not Present
39	T39	Submersible	240	Tarn taran	Rural	Not Present	Not Present
40	T40	Submersible	240	Tarn taran	Rural	Not Present	Not Present
41	T41	Hand Pump	140	Tarn taran	Rural	Not Present	Not Present

42	T42	Submersible	300	Tarn taran	Rural	Not Present	Not Present
43	T43	Submersible	120	Tarn taran	Rural	Not Present	Not Present
44	T44	Submersible	120	Tarn taran	Rural	Not Present	Not Present
45	T45	Submersible	120	Amritsar	Rural	Not Present	Not Present
46	A46	Submersible	100	Amritsar	Rural	Not Present	Not Present
47	A47	Submersible	130	Amritsar	Rural	Not Present	Not Present
48	A48	Submersible	350	Amritsar	Rural	Not Present	Not Present
49	A49	Submersible	120	Amritsar	Rural	Not Present	Not Present
50	A50	Submersible	180	Amritsar	Rural	Not Present	Not Present
51	A51	Submersible	120	Amritsar	Rural	Not Present	Not Present
52	A52	Submersible	80	Amritsar	Rural	Not Present	Not Present
53	A53	Submersible	120	Amritsar	Rural	Not Present	Not Present
54	A54	Submersible	130	Amritsar	Rural	Not Present	Not Present
55	A55	Submersible	150	Amritsar	Rural	Not Present	Not Present

56	A56	Hand Pump	220	Amritsar	Rural	Not Present	Not Present
57	A57	Hand Pump	40	Tarn taran	Rural	Not Present	Not Present
58	A58	Submersible	120	Amritsar	Rural	Not Present	Not Present
59	A59	Submersible	120	Amritsar	Rural	Not Present	Not Present
60	A60	Submersible	230	Amritsar	Rural	Not Present	Not Present

Table 4.2 Description of the samples and presence of bonds

Sr. No.	Sample Label	Source of Sample	Depth (Feet)	District	Region	Bond which may be Present
1	P1	Submersible	110	Pathankot	Rural	O–H, C=O, C=C, N–O, C–Cl
2	P2	Submersible	110	Pathankot	Rural	O–H, C=O, C=C, N–O, C–Cl
3	P3	Hand Pump	20	Pathankot	Rural	O–H, C=O, C=C, N–O, C–Cl
4	P4	Submersible	35	Pathankot	Rural	O–H, C=O, C=C, N–O, C–Cl
5	P5	Hand Pump	30	Pathankot	Rural	O–H, C=O, C=C, N–O, C–Cl
6	P6	Submersible	30	Pathankot	Rural	O–H, C=O, C=C, N–O, C–Cl
7	P7	Hand Pump	30	Pathankot	Rural	O–H, C=O, C=C, N–O, C–Cl
8	P8	Submersible	30	Pathankot	Rural	O–H, C=O, C=C, N–O, C–Cl
9	P9	Hand Pump	25	Pathankot	Rural	O–H, C=O, C=C, N–O, C–Cl
10	P10	Submersible	120	Pathankot	Urban	O–H, C=O, C=C, N–O, C–Cl
11	P11	Submersible	120	Pathankot	Urban	O–H, C=O, C=C, N–O, C–Cl
12	P12	Submersible	120	Pathankot	Rural	O–H, C=O, C=C, N–O, C–Cl
13	P13	Hand Pump	30	Pathankot	Rural	O–H, C=O, C=C, N–O, C–Cl

14	P14	Hand Pump	35	Pathankot	Rural	O–H, C=O, C=C, N–O, C–Cl
15	P15	Submersible	-	Gurdaspur	Rural	O–H, C=O, C=C, N–O, C–Cl
16	P16	Hand Pump	120	Pathankot	Rural	O–H, C=O, C=C, N–O, C–Cl
17	G17	Hand Pump	35	Gurdaspur	Rural	O–H, C=O, C=C, N–O, C–Cl
18	G18	Hand Pump	60	Gurdaspur	Urban	O–H, C=O, C=C, N–O, C–Cl
19	G19	Hand Pump	70	Gurdaspur	Urban	O–H, C=O, C=C, N–O, C–Cl
20	G20	Submersible	105	Gurdaspur	Rural	O–H, C=O, C=C, N–O, C–Cl
21	G21	Hand Pump	40	Gurdaspur	Rural	O–H, C=O, C=C, N–O, C–Cl
22	G22	Submersible	70	Gurdaspur	Rural	O–H, C=O, C=C, N–O, C–Cl
23	G23	Submersible	250	Gurdaspur	Rural	O–H, C=O, C=C, N–O, C–Cl
24	G24	Hand Pump	60	Gurdaspur	Rural	O–H, C=O, C=C, N–O, C–Cl
25	G25	Submersible	95	Gurdaspur	Rural	O–H, C=O, C=C, N–O, C–Cl
26	G26	Hand Pump	40	Gurdaspur	Rural	O–H, C=O, C=C, N–O, C–Cl
27	G27	Hand Pump	55	Gurdaspur	Rural	O–H, C=O, C=C, N–O, C–Cl

28	G28	Submersible	130	Gurdaspur	Rural	O–H, C=O, C=C, N–O, C–Cl
29	G29	Hand Pump	35	Gurdaspur	Rural	O–H, C=O, C=C, N–O, C–Cl
30	G30	Submersible	130	Gurdaspur	Rural	O–H, C=O, C=C, N–O, C–Cl
31	T31	Submersible	250	Tarn taran	Rural	O–H, C=O, C=C, N–O, C–Cl
32	T32	Submersible	180	Tarn taran	Rural	O–H, C=O, C=C, N–O, C–Cl
33	T33	Submersible	120	Tarn taran	Rural	O–H, C=O, C=C, N–O, C–Cl
34	T34	Submersible	280	Tarn taran	Rural	O–H, C=O, C=C, N–O, C–Cl
35	T35	Submersible	400	Tarn taran	Rural	O–H, C=O, C=C, N–O, C–Cl
36	T36	Submersible	250	Tarn taran	Rural	O–H, C=O, C=C, N–O, C–Cl
37	T37	Submersible	120	Tarn taran	Rural	O–H, C=O, C=C, N–O, C–Cl
38	T38	Submersible	500	Tarn taran	Rural	O–H, C=O, C=C, N–O, C–Cl
39	T39	Submersible	240	Tarn taran	Rural	O–H, C=O, C=C, N–O, C–Cl
40	T40	Submersible	240	Tarn taran	Rural	O–H, C=O, C=C, N–O, C–Cl
41	T41	Hand Pump	140	Tarn taran	Rural	O–H, C=O, C=C, N–O, C–Cl

42	T42	Submersible	300	Tarn taran	Rural	O–H, C=O, C=C, N–O, C–Cl
43	T43	Submersible	120	Tarn taran	Rural	O–H, C=O, C=C, N–O, C–Cl
44	T44	Submersible	120	Tarn taran	Rural	O–H, C=O, C=C, N–O, C–Cl
45	T45	Submersible	120	Amritsar	Rural	O–H, C=O, C=C, N–O, C–Cl
46	A46	Submersible	100	Amritsar	Rural	O–H, C=O, C=C, N–O, C–Cl
47	A47	Submersible	130	Amritsar	Rural	O–H, C=O, C=C, N–O, C–Cl
48	A48	Submersible	350	Amritsar	Rural	O–H, C=O, C=C, N–O, C–Cl
49	A49	Submersible	120	Amritsar	Rural	O–H, C=O, C=C, N–O, C–Cl
50	A50	Submersible	180	Amritsar	Rural	O–H, C=O, C=C, N–O, C–Cl
51	A51	Submersible	120	Amritsar	Rural	O–H, C=O, C=C, N–O, C–Cl
52	A52	Submersible	80	Amritsar	Rural	O–H, C=O, C=C, N–O, C–Cl
53	A53	Submersible	120	Amritsar	Rural	O–H, C=O, C=C, N–O, C–Cl
54	A54	Submersible	130	Amritsar	Rural	O–H, C=O, C=C, N–O, C–Cl
55	A55	Submersible	150	Amritsar	Rural	O–H, C=O, C=C, N–O, C–Cl

56	A56	Hand Pump	220	Amritsar	Rural	O–H, C=O, C=C, N–O, C–Cl
57	A57	Hand Pump	40	Tarn taran	Rural	O–H, C=O, C=C, N–O, C–Cl
58	A58	Submersible	120	Amritsar	Rural	O–H, C=O, C=C, N–O, C–Cl
59	A59	Submersible	120	Amritsar	Rural	O–H, C=O, C=C, N–O, C–Cl
60	A60	Submersible	230	Amritsar	Rural	O–H, C=O, C=C, N–O, C–Cl

4.3 SUMMARY

Ground water which is very important for all the life on the Earth may contains various types of contaminants these may come into water due to geogenic or anthropogenic activities. The present work focused on to identify the two pollutants urea and arsenic trioxide in the ground water samples. Comparison of the FTIR spectra of groundwater samples and the spectra of the urea and arsenic trioxide is done and the results are given in the Table 4.1. It is observed that bonds similar to urea and arsenic are not observed in the groundwater samples. The values of TDS and electrical conductivity indicate the presence of impurities in the ground water samples and some of the peaks observed may be due to the presence of other organic or inorganic matter.

The spectroscopic analysis for the groundwater samples of the Majha region of Punjab, India does not show detectable urea or arsenic in the samples. The water meets the safety norms in context with urea and arsenic and thus can be considered safe for drinking purposes. The spectroscopic data combined with the statistical techniques like PCA, PLSR and the multivariate analysis have significant implications for water safety standards. The results can be obtained quickly and economically, this helps in real time monitoring of the ground water in a particular area. As the method is economic and quick it can be used in greater number of locations.

CHAPTER 5

PHYSICOCHEMICAL PROPERTIES OF GROUND WATER SAMPLES AND CLASSIFICATION OF THE SAMPLES USING PCA

CHAPTER 5 - PHYSICOCHEMICAL PROPERTIES OF GROUND WATER SAMPLES AND CLASSIFICATION OF THE SAMPLES USING PCA

5.1 INTRODUCTION

Groundwater is an important and vital resource for sustaining life on the Earth. It is mentioned in the Groundwater Project (a registered charity in Ontario, Canada) [102], 2.5 billion people depend solely on groundwater resources to satisfy their basic daily water needs. Water is important for life as well as for the economy of a country since it is used in industry for various activities, some of the industrial products (like beverages) solely depend on the use of water and lot of water is used in agriculture for irrigation and other tasks. Besides the problem of water table is going down as it is drawn out from under the ground in abundance, the other important allied problem is its pollution. In the present work pollution of groundwater is to be taken into consideration. It becomes necessary to know the physical and chemical properties of groundwater so that groundwater can be characterized as contaminated or as pure [103 - 104]. The physicochemical properties/parameters which are investigated in this work are the following total dissolved solids, turbidity, pH value, total hardness, chlorine content and electrical conductivity. The values obtained are given in Table 5.1. The determination of parameters is done using the standard methods specified by Bureau of Indian Standards.

Turbidity is measured using nephelometric method, it is based on the comparison of scattered light from the water sample with some standard reference scattering light under same conditions. It is measured in nephelometric turbidity units (NTU) as prescribed by Bureau of Indian Standards (BIS) [105]. pH value is the indicator of H^+ ions in the solution and it is expressed as the logarithm of the reciprocal of hydrogen activity in moles per litre. The value is determined using the electrometrically as specified by the Bureau of Indian Standards (BIS) [106]. Electrical conductivity is the ability of water to allow the flow of current through it, and is determined using the method prescribed by BIS [107], the apparatus used in this method is wheatstone bridge, conductivity cell, thermometer etc. Total dissolved solids in water sample is determined by filtering and evaporating the filtrate to get residue and then applying the procedures as described by BIS [108]. The determination of the chlorine in the water sample is done by the method prescribed by BIS [109].

Total hardness is computed from the magnesium and calcium cations present in the water and sometimes water may have in it measurable quantity of aluminium, zinc, iron, manganese, barium and strontium. In such case total hardness is calculated from the total concentration of the cations present in water. Total hardness is calculated using Ethylenediamine tetraacetic acid (EDTA) method prescribed by BIS [110].

According to BIS and WHO the acceptable limit of pH value for drinking water should lie between 6.5 and 8.5 Table 5.1. Acidic or alkaline water is not fit for drinking purpose and may not suit for the irrigation purpose in agricultural. Water with pH value greater than seven tastes bitter and can cause skin irritation. The effectiveness of water disinfection with chlorine is reduced if pH of the water is low. The acidic water also leads to corrosion of metal pipes through which it flows and the metallic containers in which it is kept. The water having low pH value (less than seven), if used, it may corrode metals used in buildings. TDS stands for total dissolved solids in the water and the solids may be organic or inorganic in nature. These may be carbonates, bicarbonates, nitrates, chlorides, sulphate etc. of sodium, magnesium, potassium, calcium etc. The water which has high TDS value is somewhat bitter in taste. It is mentioned in a report by WHO that inverse relation exist between TDS and incidence of cancer, some of the heart diseases, also the mortality rates are also inversely related to the TDS levels in drinking water [111]. Higher TDS water may be beneficial due to the presence of important minerals like magnesium and calcium but not all dissolved solids are good and therefore composition of total dissolved solids is very important. The TDS value even if it is small or within permissible limits as prescribed by WHO or by BIS, but contains toxic metals like arsenic, fluoride etc. even in small quantities can be harmful to humans [112]. The water with high TDS values leave scales in washing machines, geysers etc. If water contains chlorine, the water gives a typical smell. It can cause corrosion in the pipes also. In the report by WHO, it is mentioned that bladder cancer appears to be linked with the consumption of chlorine [113]. The electrical conductivity indicates that ions are present in the water. Turbidity indicates the cloudiness of water and is due the presence of suspension of particles of clay, silts, iron, manganese, plant debris and or pathogens. Higher value of turbidity reduces the acceptance of drinking water and higher value of turbidity may cause an outbreak of a disease, but this is not an established fact [114]. Hardness is due to the presence of carbonates (temporary hardness) and non-carbonates (permanent hardness) of magnesium and or calcium. Such water on boiling can cause scaling in the container, boilers etc. Various types of diseases are associated with the inadequate and excess intake of calcium and magnesium like kidney stones, osteoporosis, insulin resistance

and obesity, calcium may interact with iron, magnesium, phosphorous etc. and inhibits their absorption in the body [115]. From the reports of WHO on different water quality parameters it is clear that there is need to monitor the quality of groundwater for the drinking purpose. The acceptable limits set by BIS [116] and WHO [117 - 118] and for the groundwater quality parameters are given in Table 5.1

Table 5.1 Acceptable limits for quality parameters for drinking water

PARAMETER	BIS (Acceptable Limit)	WHO (Acceptable Limit)
pH value	6.5-8.5	6.5-8.5
Turbidity (NTU)	1	1
Total dissolved solids, (TDS) mg/L	500	500
Chloride (as Cl), mg/L	250	250
Total Hardness (TH) (as CaCO$_3$), mg/L	200	200
Electrical conductivity (µS/cm)	-	1500

Table 5.2 Physicochemical parameters of groundwater samples

Sr. No.	Sample Name	Source of water	Depth (Feet)	Region	Turbidity (NTU)	Total Dissolved solids (TDS) (mg/L)	Electrical Conductivity (µS/cm)	pH Value	Total Hardness (mg/L)	Chloride as Cl (mg/L)
1	P1	Submersible	110	Rural	0	198	305	7.68	130	35
2	P2	Submersible	110	Rural	**2**	318.5	490	**9.55**	194	60
3	P3	Hand pump	20	Rural	0	303	467	7.88	**234**	40
4	P4	Submersible	35	Rural	1	242	373	**9.84**	**214**	60.5
5	P5	Hand pump	30	Rural	0	219.7	338	7.65	162	25
6	P6	Submersible	30	Rural	0	273	420	6.55	130	64
7	P7	Hand pump	30	Rural	0	248	382	7.44	150	29
8	P8	Submersible	30	Rural	1	361	554	7.54	**250**	32

9	P9	Hand Pump	25	Rural	0	407	628	7.63	**290**	50.5
10	P10	Submersible	120	Urban	0	407	627	7.34	**330**	64
11	P11	Submersible	120	Urban	0	423	650	7.55	**270**	53
12	P12	Submersible	120	Rural	0	352	543	7.64	138	33
13	P13	Hand Pump	30	Rural	**2**	469	720	**8.91**	**294**	68
14	P14	Hand Pump	35	Rural	**2**	421	646	7.86	170	50
15	P15	Submersible	not known	Rural	0	235	361	6.72	90	30
16	P16	Hand Pump	120	Rural	**3**	545	839	**8.74**	**240**	70
17	G17	Hand Pump	35	Rural	0	463	713	6.77	**270**	60
18	G18	Hand Pump	60	Urban	**2**	453	697	6.85	**250**	65
19	G19	Hand Pump	70	Urban	**1**	**508**	783	6.91	**300**	72
20	G20	Submersible	105	Rural	**3**	480	739	7.4	142	54
21	G21	Hand Pump	40	Rural	**4**	**711**	1094	7.59	**536**	150
22	G22	Submersible	70	Rural	0	445	686	7.42	**254**	38
23	G23	Submersible	250	Rural	0	296	458	6.88	90	31
24	G24	Hand Pump	60	Rural	**1**	**847**	1303	7.58	**298**	75
25	G25	Submersible	95	Rural	0	410	634	7.34	70	23
26	G26	Hand Pump	40	Rural	0	466	716	7.84	72	27.5
27	G27	Hand Pump	55	Rural	**1**	**721**	1107	7.82	**436**	135
28	G28	Submersible	130	Rural	0	424	657	6.87	198	58.4
29	G29	Hand Pump	35	Rural	0	**530**	816	6.92	**310**	58
30	G30	Submersible	130	Rural	0	406	626	6.81	**230**	40
31	T31	Submersible	250	Rural	0	**612**	942	7.57	**214**	66.4

32	T32	Submersible	180	Rural	0	**819**	1263	**8.58**	130	65
33	T33	Submersible	120	Rural	0	**648**	998	7.68	**205**	66
34	T34	Submersible	280	Rural	0	**623**	959	6.82	132	50.5
35	T35	Submersible	400	Rural	0	428	657	6.94	152	47.5
36	T36	Submersible	250	Rural	0	**783**	1205	6.88	**290**	76
37	T37	Submersible	120	Rural	0	**826**	1272	7.58	**374**	92.7
38	T38	Submersible	500	Rural	0	**447**	692	6.91	198	67.8
39	T39	Submersible	240	Rural	0	397	610	6.85	142	48
40	T40	Submersible	240	Rural	0	410	633	7.44	**226**	62.4
41	T41	Hand Pump	140	Rural	0	**624**	961	6.82	**232**	56
42	T42	Submersible	300	Rural	0	**771**	1187	6.94	**254**	86
43	T43	Submersible	120	Rural	0	**1557**	2390	7.47	**380**	104
44	T44	Submersible	120	Rural	0	**761**	1181	7.88	**226**	73
45	T45	Submersible	120	Rural	0	**655**	1011	6.71	**230**	80
46	A46	Submersible	100	Rural	0	**1043**	1603	7.64	**338**	104
47	A47	Submersible	130	Rural	0	**856**	1320	7.02	**390**	98
48	A48	Submersible	350	Rural	0	393	608	6.61	120	35
49	A49	Submersible	120	Rural	0	497	828	7.11	134	47.4
50	A50	Submersible	180	Rural	0	**542**	830	7.24	**205**	55
51	A51	Submersible	120	Rural	0	**591**	904	7.07	**270**	88
52	A52	Submersible	80	Rural	0	**542**	834	7.13	**290**	58
53	A53	Submersible	120	Rural	0	333	513	7.48	70	33
54	A54	Submersible	130	Rural	**2**	**1148**	1765	7.57	**290**	85

55	A55	Submersible	150	Rural	0	**1095**	1685	7.26	**458**	112
56	A56	Hand Pump	220	Rural	**3**	441	683	7.38	**240**	35
57	A57	Hand Pump	40	Rural	**5**	376	579	**6.48**	**270**	43
58	A58	Submersible	120	Rural	0	**551**	848	7.13	**310**	41
59	A59	Submersible	120	Rural	0	**827**	1273	7.08	**390**	148
60	A60	Submersible	230	Rural	0	494	760	7.62	160	25

5.2 DESCRIPTIVE STATISTICS

All the data obtained for physical and chemical properties is analysed for statistical parameters like mean, median, standard deviation, variance, correlation etc. using add in tab in Microsoft Excel software. It is observed from the Table 5.2 and Table 5.3 that the Turbidity values range from 0 to 5 NTU and its mean values comes out to be 0.5500 NTU

Table 5.3 Statistical Parameters of the samples

Statistical parameter	Turbidity (NTU)	Total dissolved solids (TDS) (mg/L)	Electrical conductivity (µS/cm)	pH value	Total hardness (mg/L)	Chloride as Cl (mg/L)
Mean	0.5500	544.5367	839.4333	7.4068	234.3667	61.6600
Standard Error	0.1454	32.6467	50.1582	0.0848	12.8374	3.7039
Median	0.0000	467.5000	718.0000	7.3900	231.0000	58.2000
Standard Deviation	1.1263	252.8799	388.5238	0.6572	99.4379	28.6907
Sample Variance	1.2686	63948.2281	150950.7582	0.4319	9887.8972	823.1540
Range	5.0000	1359.0000	2085.0000	3.3600	466.0000	127.0000
Minimum	0.0000	198.0000	305.0000	6.4800	70.0000	23.0000
Maximum	5.0000	1557.0000	2390.0000	9.8400	536.0000	150.0000

and standard deviation is 0.1454 NTU, the value of TDS ranges from 198 to 1557 mg/L, its mean value is 544.5367 mg/L, its standard deviation is 32.6467 mg/L. Electrical conductivity

ranges from 305 to 2390 µS/cm, its mean value is 839.4333 µS/cm and standard deviation is 50.1582 µS/cm, pH value ranges from 6.48 to 9.84, and mean value comes out to be 7.4068 and standard deviation is 0.0848, total hardness ranges from 70 to 536 mg/L, and mean value is 234.3667 mg/L and its standard deviation is 12.8374 mg/L, Chloride content range from 23 to 150 mg/L, the mean is 61.6600 mg/L and standard deviations is 3.7039 mg/L.

5.3 VARIATION OF THE PHYSICOCHEMICAL PARAMETERS WITH DEPTH

All the data is collected in MS Excel sheet and with the help of insert chart tab the graphs are plotted. All the parameters were normalized first in each column by dividing every entry in the same column with the maximum value of the parameter in the same column. All the normalized parameters were obtained and the whole data was sorted with respect to the depth column from minimum value to maximum. All the normalized graphs are separately given in Figure 5.1 to Figure 5.6.

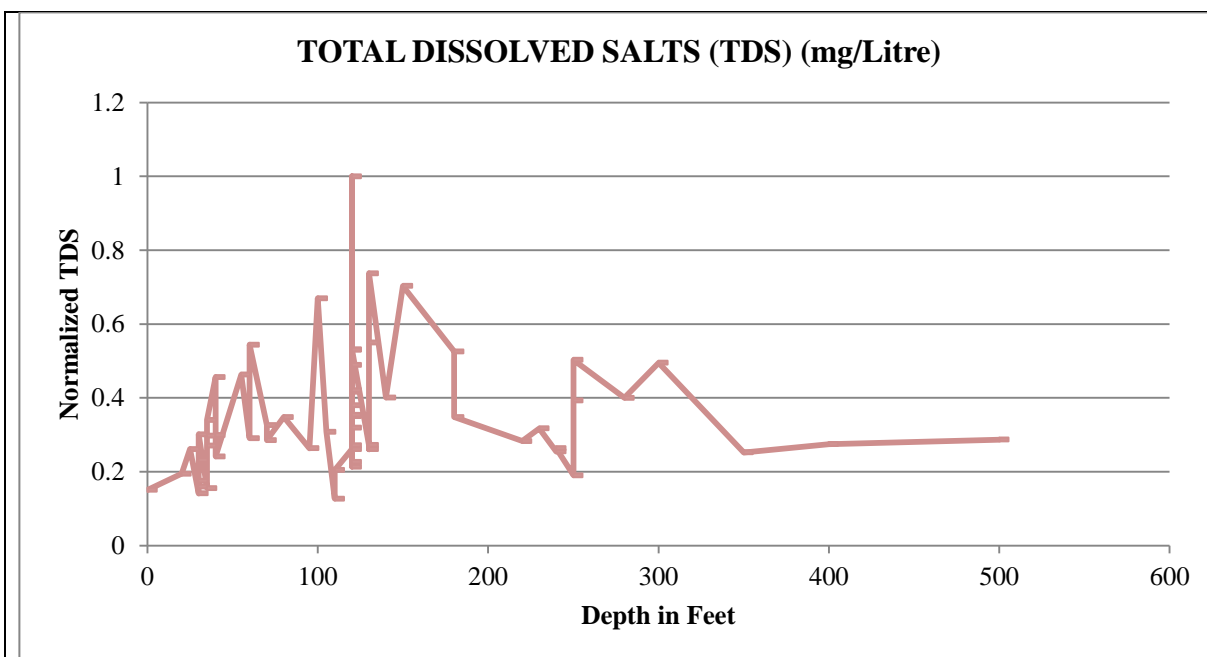Figure 5.1 Variation of normalized turbidity w.r.t. depth
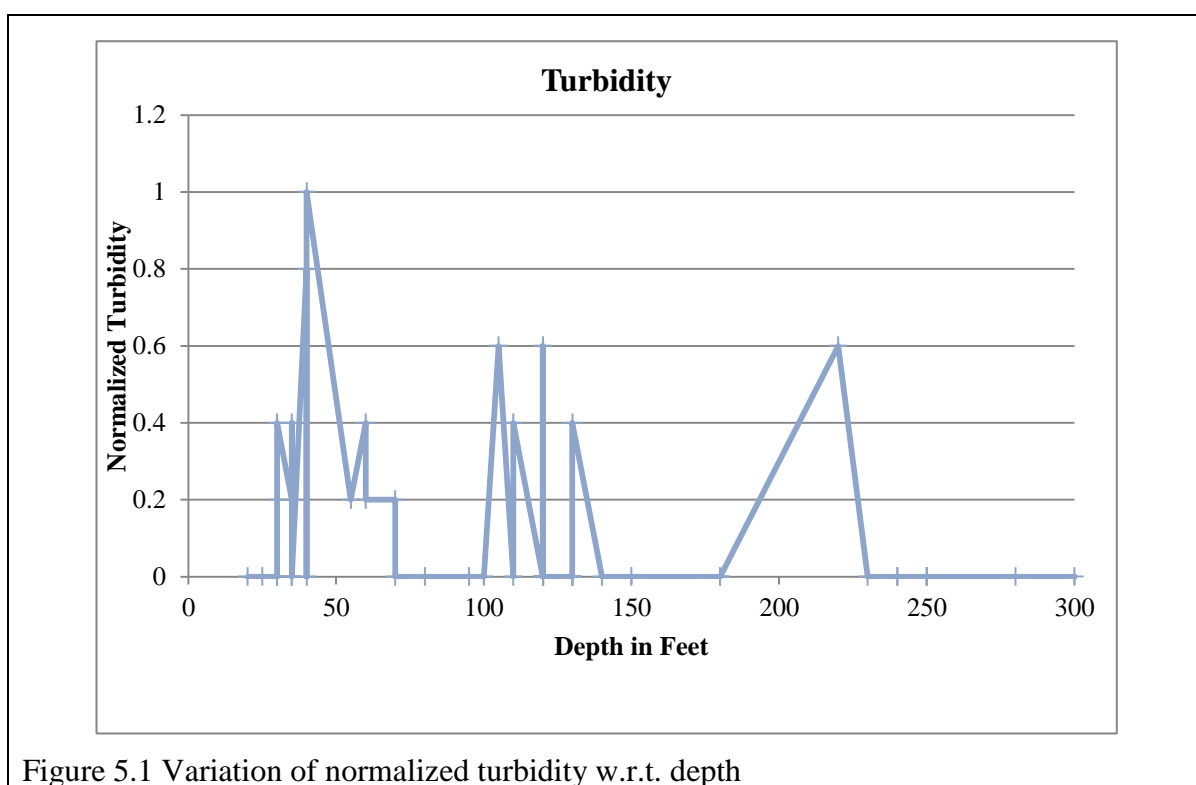
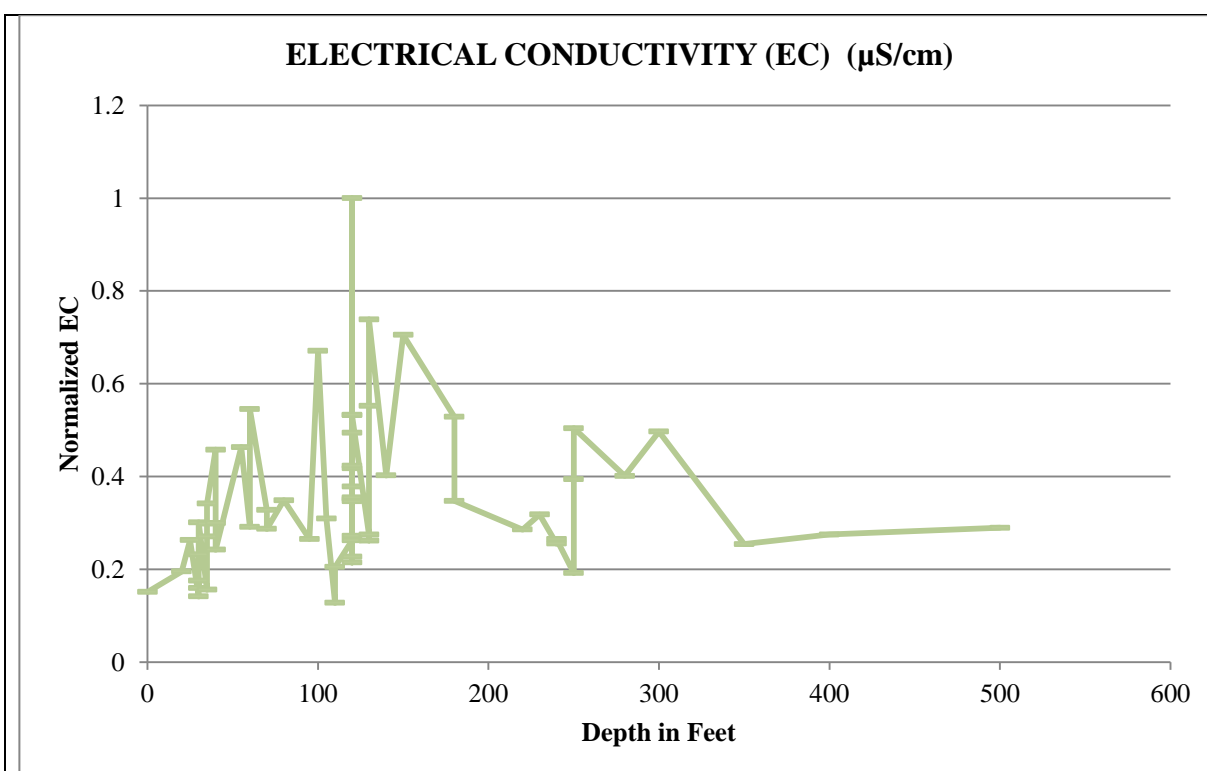Figure 5.2 Variation of normalized TDS w.r.t. depth

Figure 5.3 Variation of normalized Electrical Conductivity w.r.t. depth

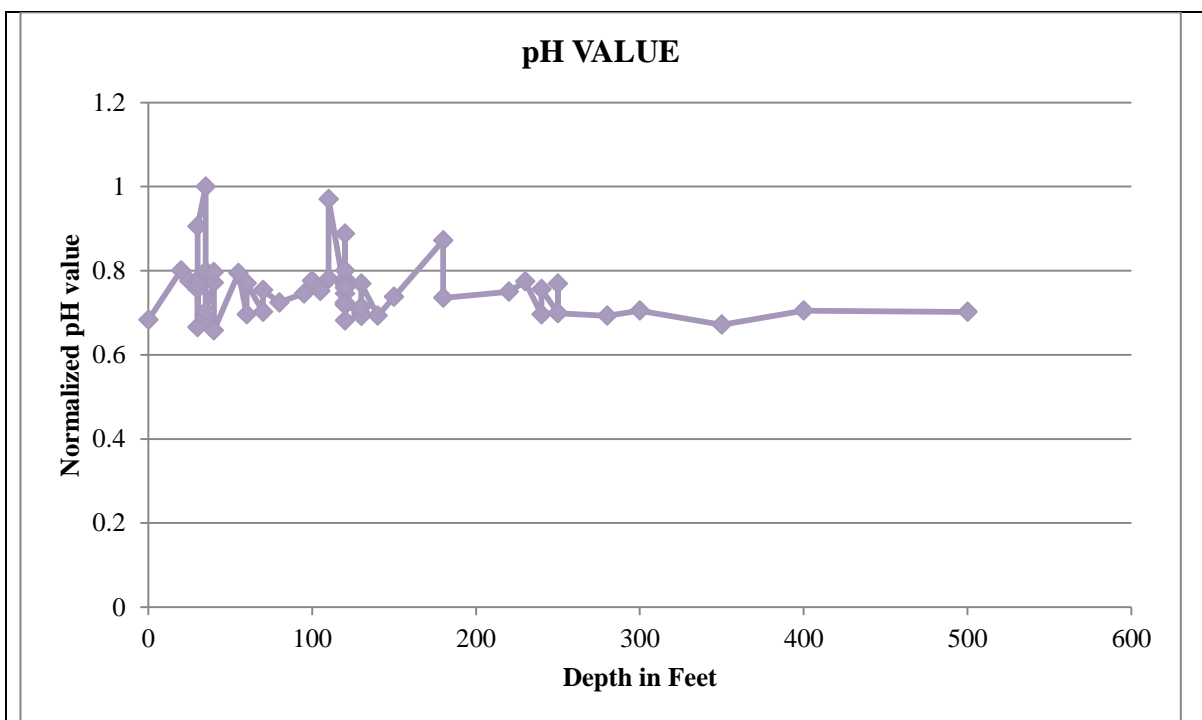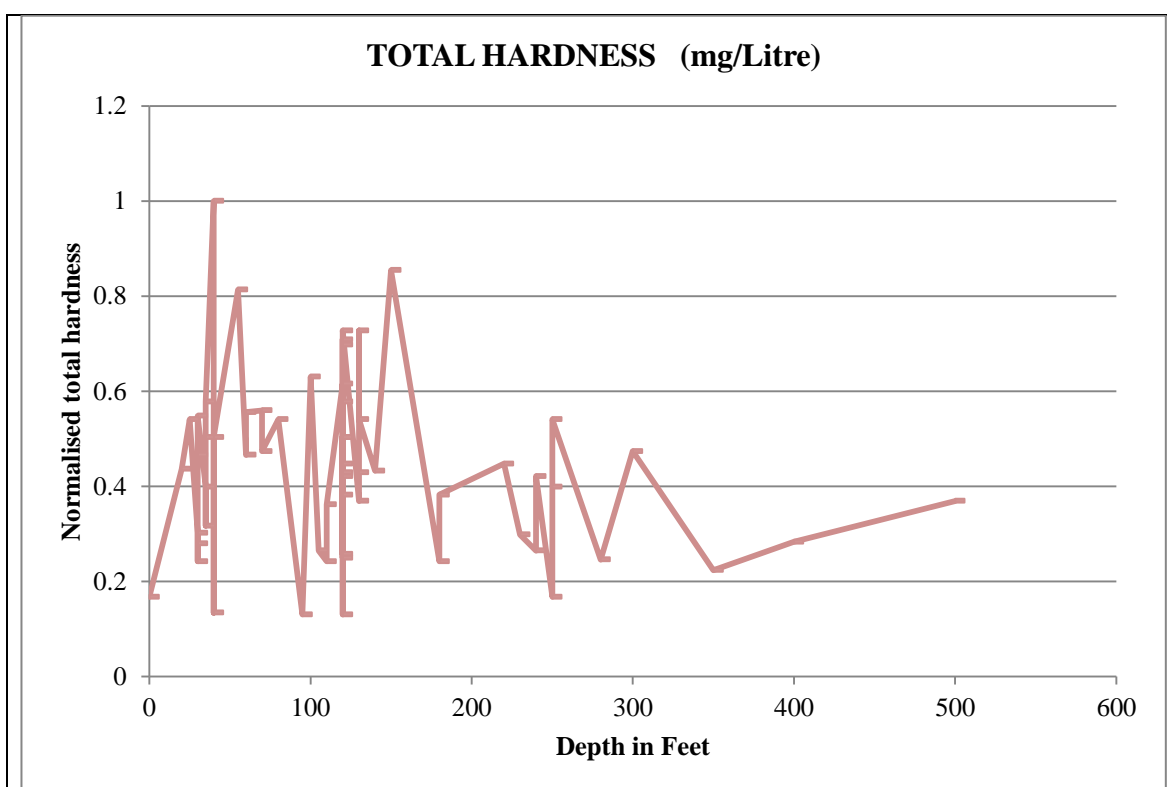Figure 5.4 Variation of normalized pH value w.r.t. depth

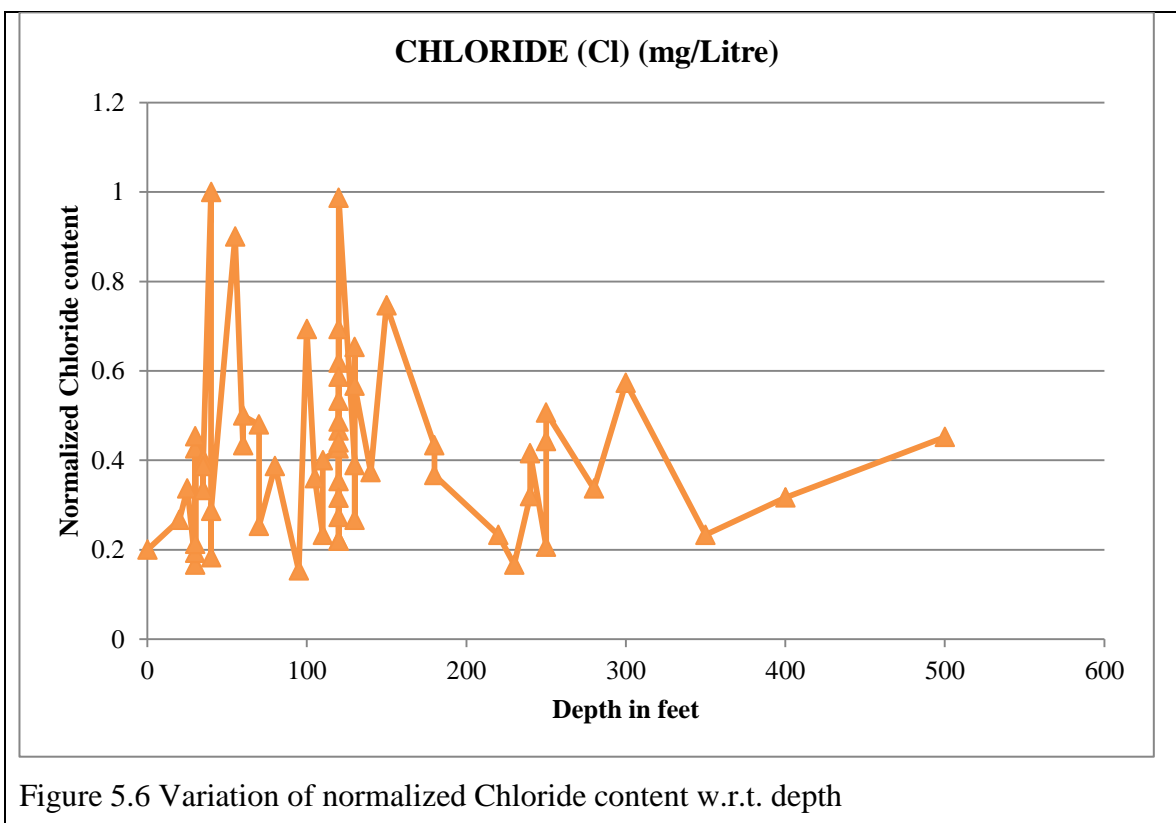Figure 5.5 Variation of normalized total hardness w.r.t. depth

Figure 5.6 Variation of normalized Chloride content w.r.t. depth

The plots of total dissolved solids and electrical conductivity vs. depth seem to be similar. It can be concluded that there is high correlation between both of them, it is also clear from the correlation Table 5.4.

Table 5.4 Correlations among various physical and chemical parameters

	Turbidity (NTU)	Total Dissolved solids (TDS) (mg/L	Electrical Conductivity (µS/cm)	pH Value	Total Hardness (mg/L)	Chloride as Cl (mg/L)
Turbidity (NTU)	1.0000					
Total Dissolved solids (TDS) (mg/L)	-0.0327	1.0000				
Electrical Conductivity (µS/cm)	-0.0345	0.9998	1.0000			
pH Value	0.2304	-0.0291	-0.0308	1.0000		
Total Hardness (mg/L)	0.2300	0.5997	0.5967	0.0222	1.0000	
Chloride as Cl (mg/L)	0.1509	0.7001	0.6987	0.0493	0.8064	1.0000

Chloride and total hardness also show somewhat similar trend and it is also clear from the correlation table. The graphs for the following pairs also show somewhat same variations for total hardness and total dissolved solids, chlorine and total dissolved solids, chlorine and electrical conductivity. It is also observed that after depth of nearly 200 feet the parameters show a decrease in their values.

5.4 REGRESSION ANALYSIS

For the pair of physicochemical properties which showed correlation coefficients greater than 0.5 a regression analysis was done to find the line of best fit. Equations (5.1) to (5.6) are the regression line equations obtained using MS Excel software.

$$EC = 1.5374\,(TDS) + 2.4461, \qquad R^2 = 0.9993 \qquad (5.1)$$

$$TH = 0.2526\,(TDS) + 94.327, \qquad R^2 = 0.3595 \qquad (5.2)$$

$$Cl = 0.0914\,(TDS) + 11.068, \qquad R^2 = 0.5586 \qquad (5.3)$$

$$TH = 0.1630\,(EC) + 94.959, \qquad R^2 = 0.3543 \qquad (5.4)$$

$$Cl = 0.0593\,(EC) + 11.079, \qquad R^2 = 0.5554 \qquad (5.5)$$

$$Cl = 0.2180\,(TH) + 9.9894, \qquad R^2 = 0.5643 \qquad (5.6)$$

Linear relations are given between EC-TDS, TH-TDS, Cl-TDS, Cl-TDS, TH-EC, Cl-EC and Cl-TH respectively.
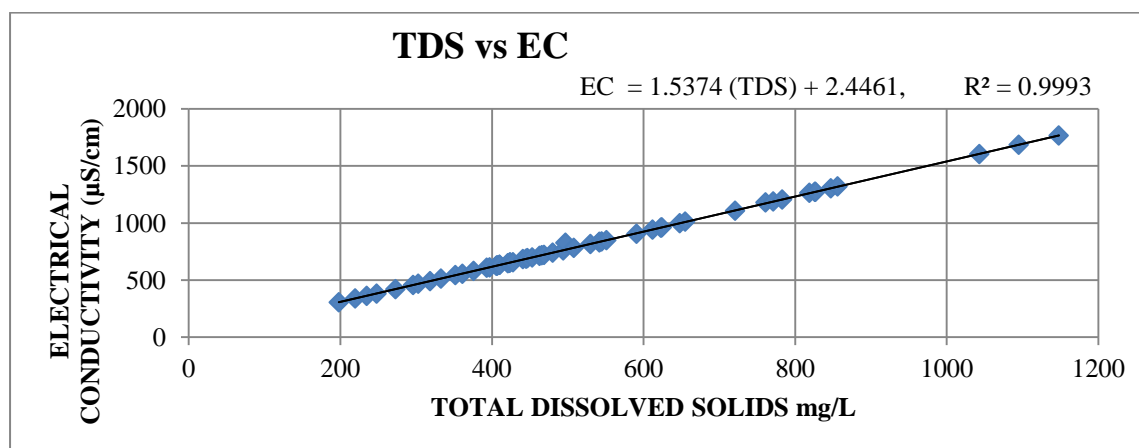

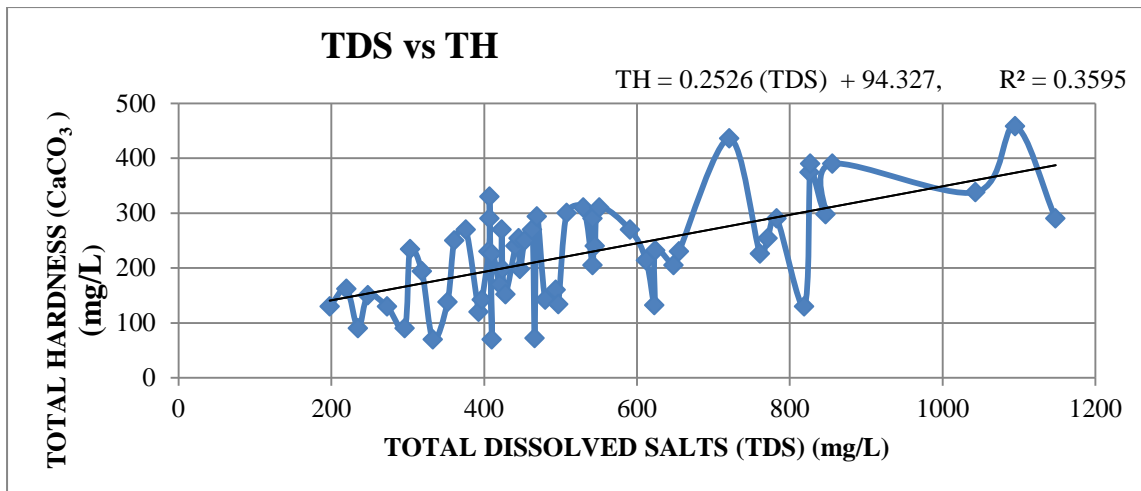Figure 5.7 Linear Regression Total Dissolved Solids Vs Electrical Conductivity

Figure 5.8 Linear Regression Total Dissolved Solids Vs Total Hardness
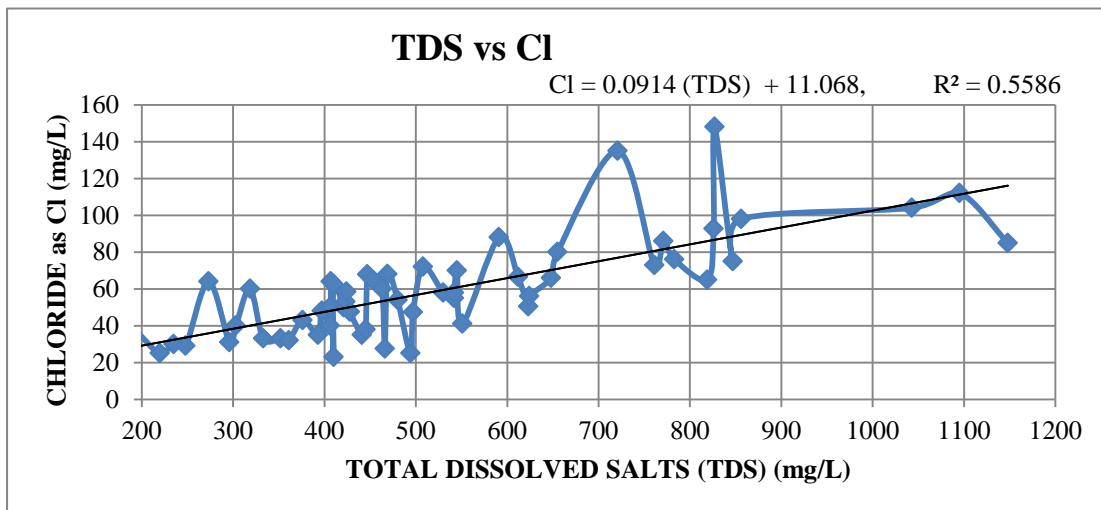

Figure 5.9 Linear Regression Total Dissolved Solids Vs Chlorine Content
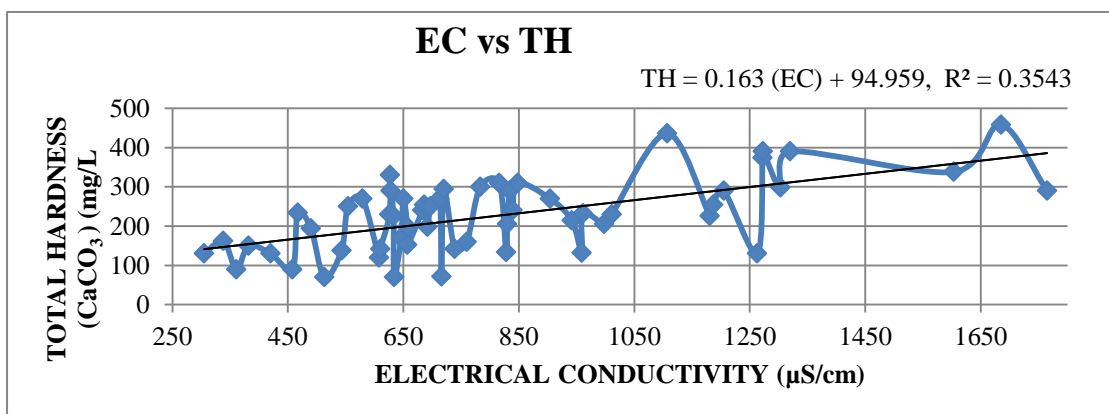

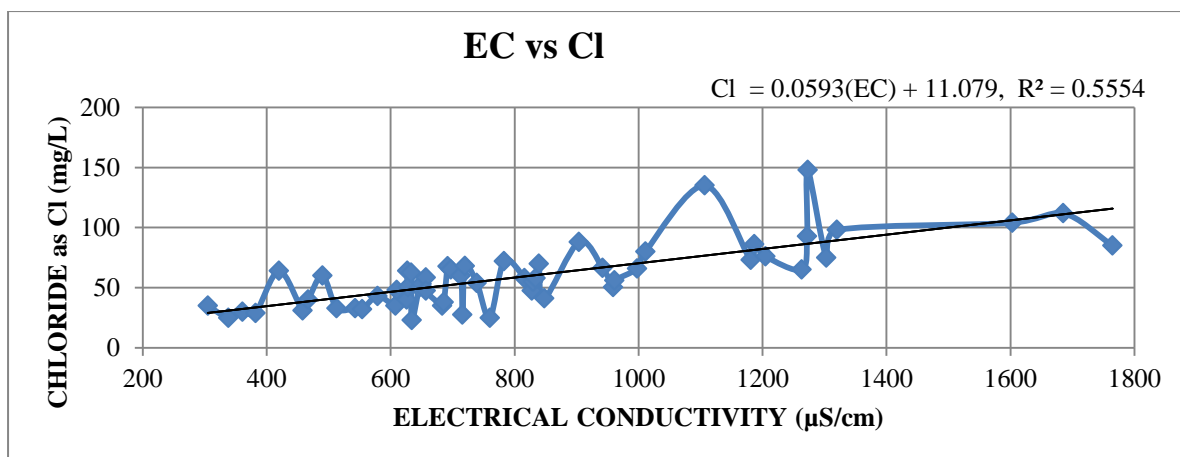Figure 5.10 Linear Regression Electrical Conductivity Vs Total Hardness

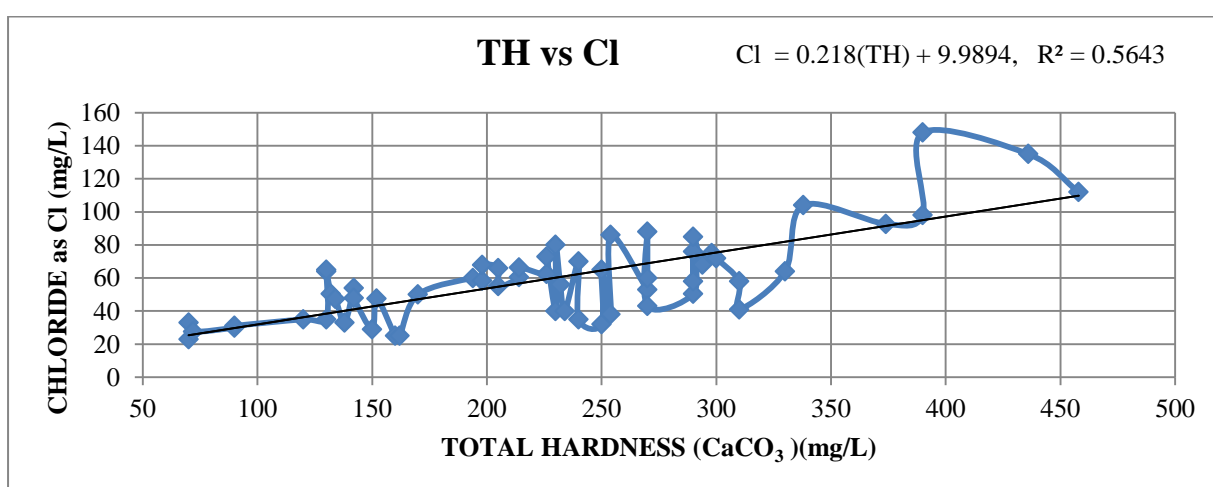Figure 5.11 Linear Regression Electrical Conductivity Vs Chlorine Content

Figure 5.12 Linear Regression Total Hardness Vs Chlorine Content
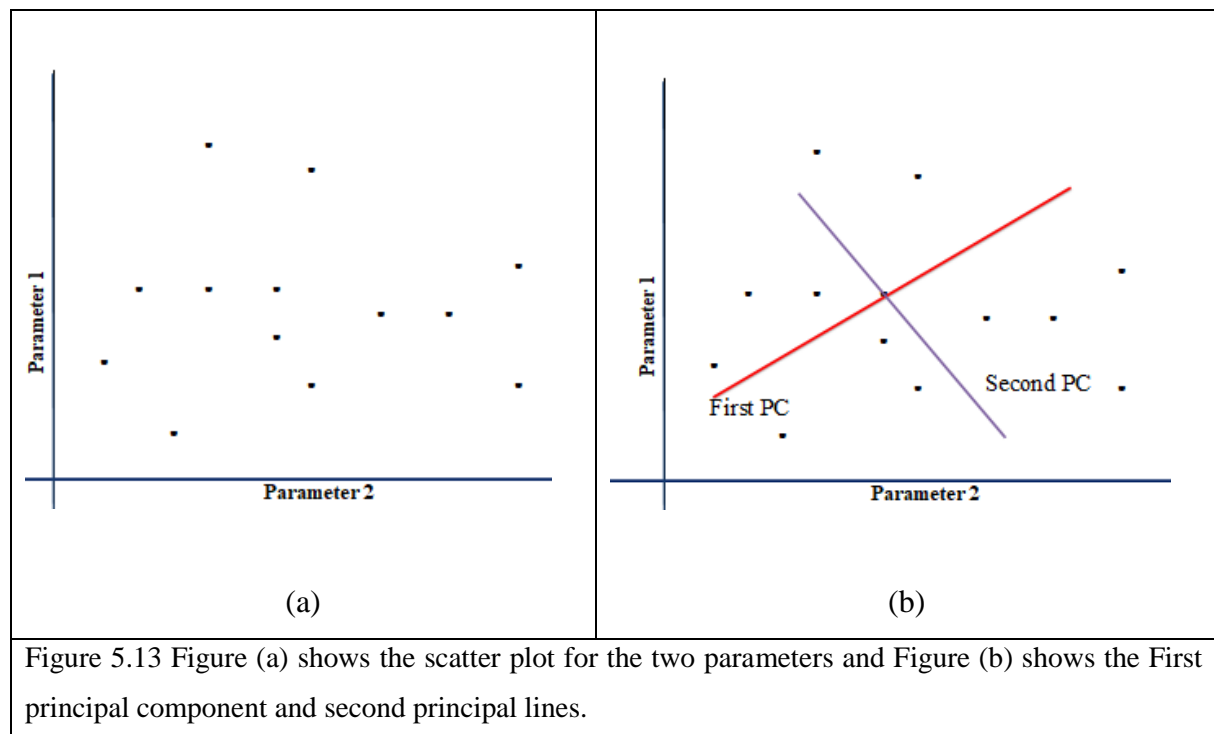
5.5 PRINCIPAL COMPONENT ANALYSIS

PCA stands for Principal Component Analysis has many applications like data analysis, data compression, removing noise and reduction in the dimension of a multidimensional data etc. PCA analysis becomes helpful in the sense that it removes or eliminates similar type of data which has no contribution in the decision making process. PCA is a wonderful technique in the sense that in one way it reduces the dimensions of the data and on the other way there is no loss of information. Principal Component Analysis helps in finding a new data set with reduced dimensions that makes it easy to analysis the result. Consider a data set having large number of variables and dimensions. PCA in general is a dimensionality-reduction technique which reduces the dimensions of the original data set as much as possible while keeping intact the information of the original data. Although PCA finds many applications in machine learning it is used here for noise reduction also and thus it

is a preprocessing technique also.   Principal Component Analysis minimizes the more significant variance of dimensions hence the noise can be reduced completely.

Consider the following the data set with two features parameter 1 and parameter 2 as given below

$$\text{Parameter 1} \quad : \quad x_1 \quad x_2 \quad x_3 \quad x_4 \quad ....$$
$$\text{Parameter 2} \quad : \quad y_1 \quad y_2 \quad y_3 \quad y_4 \quad ....$$

Let us consider the scatter plot for these two parameters as shown in Figure 5.13 (a) and in figure (b) the first principal component and second principal component are shown.

(a)   (b)

Figure 5.13 Figure (a) shows the scatter plot for the two parameters and Figure (b) shows the First principal component and second principal lines.

It is worth noting that the principal components are calculated using the variance of parameter 1 and of parameter 2. The variance of the two data sets is calculated and the one with higher or similar variance are grouped in first category and the data with smaller or varying variance are categorized as second group. The greater variance would mean that there is more number of data points spreading on this principal component. Therefore the principal component which has more data points becomes more important. This principal component with higher variance has more ability to explain the differences in the data. The data points along this principal component are also less correlated.

The above example has only two parameters and if there are hundreds or thousands of samples, (let us consider sixty samples and six parameters attached with each sample) then

computer programs or specially designed softwares are used. The basic steps involved in the process of finding the principal components can be summarized as below-

Data collection: The first step is to collect the data and create a matrix with sample number of rows i.e. sixty in this case and six columns for the parameters.

Standardization: The quality parameters are standardized by subtracting the mean from each data point and then dividing it by the standard deviation for each parameter.

Covariance matrix: Now from the standardized matrix find the covariance matrix.

Eigenvalues and eigenvectors: In this step the eigenvalues and eigenvectors are determined from the covariance matrix.

Component selection: From the previous step, top k eigenvectors are selected which correspond to the largest eigenvalues obtained. These are the components which can explain the spreading of the data.

Transformation: The eigenvectors selected in the previous steps are the principal components which define the new coordinate system. Now transform the original data into the new coordinate system.

Data reduction: The new transformed data will have sixty rows and k columns, which are less than the original data columns, that way data reduction takes place.

Although PCA has various advantages like reduction in dimension, reduction in noise etc. but sometimes there may be disadvantages of using PCA e.g. it may be difficult to find the covariance matrix, sometimes the PCA data is difficult to understand as compared to the original data. Only the linear relationships are considered though there is possibility of non-linear relationship in such type of complex data, the principal components are linear combinations of original variables and thus the interpretation of the result is complicated task, there is a possibility of high variance shown by one of the components but it may not be of much use, the data set required is very large to obtain reliable results. There is need of other analysis to further confirm the results.

Principal component analysis (PCA) was done using Clustvis-a web tool. PCA analysis was performed on basis of physicochemical parameters (Figure 5.14), it was observed that almost complete data was covered using two principal components. In the PCA loading plot, it was observed that as total hardness was lying near the center of the PCA, thus no specific relation for total hardness could be observed with other parameters. Further Turbidity, pH and chlorine content of the samples have similar values for PC1 and PC2 showing that these parameter values are interrelated for the collected set of samples. This shows that the three parameters are positively correlated to each other and if one is known,

other two can be easily calculated using the bivariate regression equations between them. The left side of PCA shows that electrical conductivity and total dissolved solids are positively related with each other with a lower correlation as compared to coefficient of correlation between turbidity, pH and chlorine content. Thus the PCA shows us that by calculating two parameters rest three parameters can be calculated. Total hardness showed no correlation with any of the other parameters.
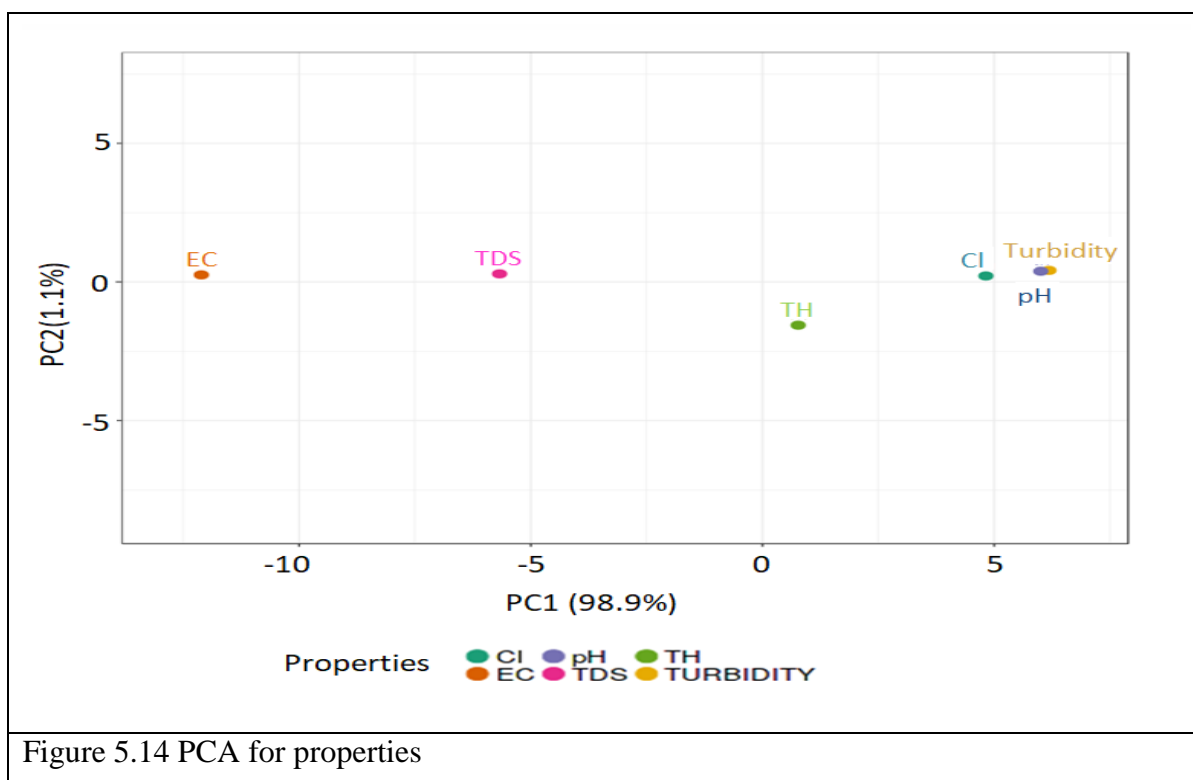

Figure 5.14 PCA for properties

Second PCA (Figure 5.15) was developed on the basis of the sources of water i.e. hand pump and submersible. First two PCs cover 73% of the complete data set. PCA shows that though the samples cannot be classified with good accuracy on basis of sources of water, but the samples collected by submersible source were present in the upper portion of PCA and that collected with hand pump were clustered on the south east part of PCA.

Figure 5.15 PCA on the basis of source of water

Figure 5.16 PCA on the basis of depth
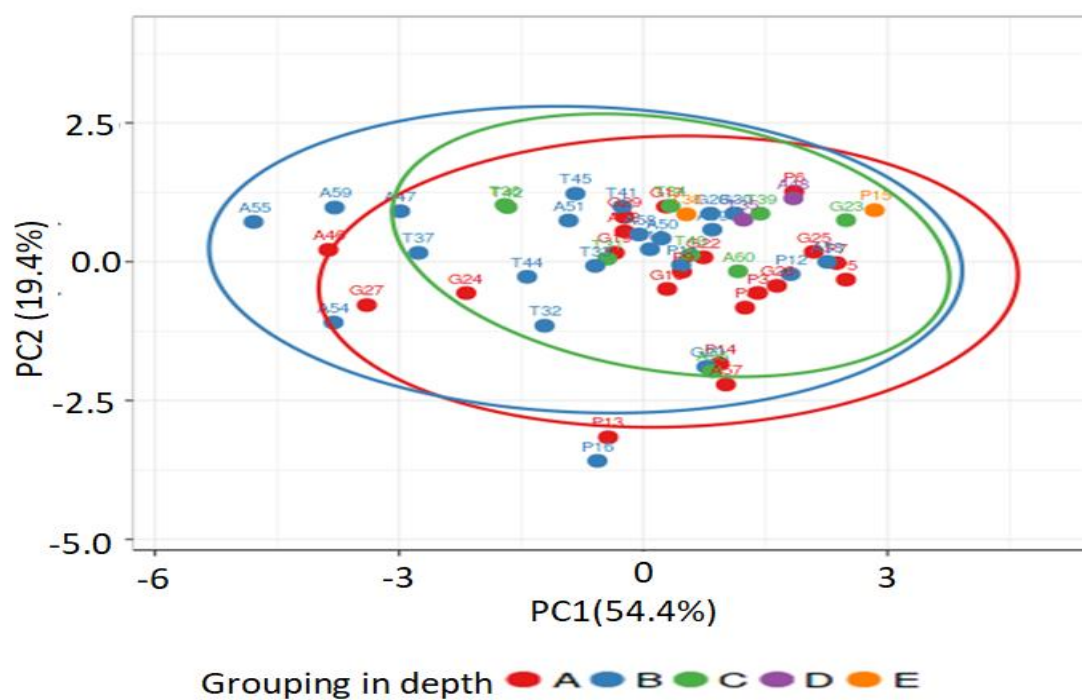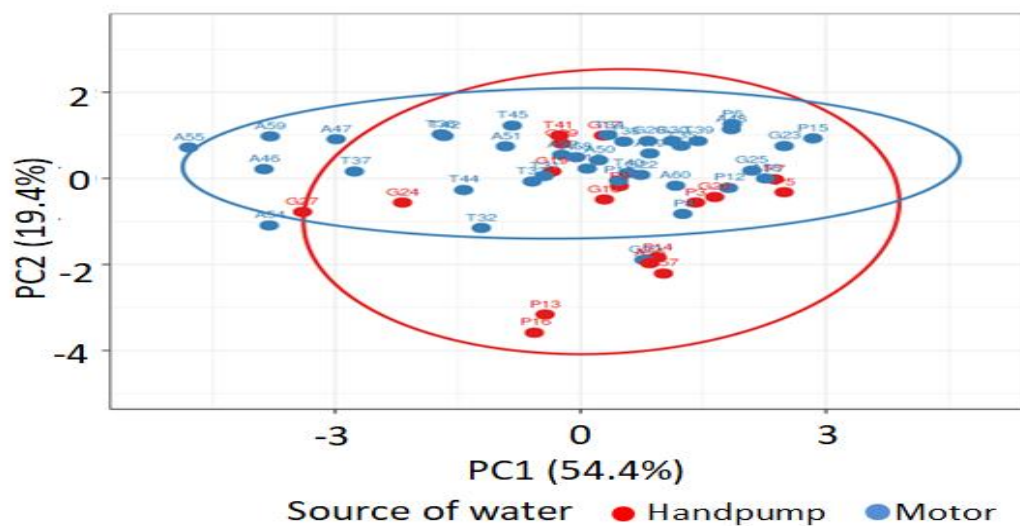
PCA was employed on all the samples on the basis of depth of the collected samples. The depth of the samples has been given codes A, B, C, D, E; where A is used for the range 0-100 ft., B for 100-200 ft, C for 200-300 ft., D for 300-400 ft and E for 400-500 ft. respectively. It can be observed that 73% data was covered using the two PCs and no specific

categorization and classification was possible on the basis of the depth of the collected ground water samples. On depth analysis it was observed that the samples collected with depth in range of 100-200 feet were lying on the left side of the PCA and could be considered in different groups from the other samples. Further samples of A and B group were on opposite sides of the PCA showing that the property of the samples was not similar for the different depth samples, though they could not be differentiated with good accuracy.

Figure 5.17 PCA on the basis of district
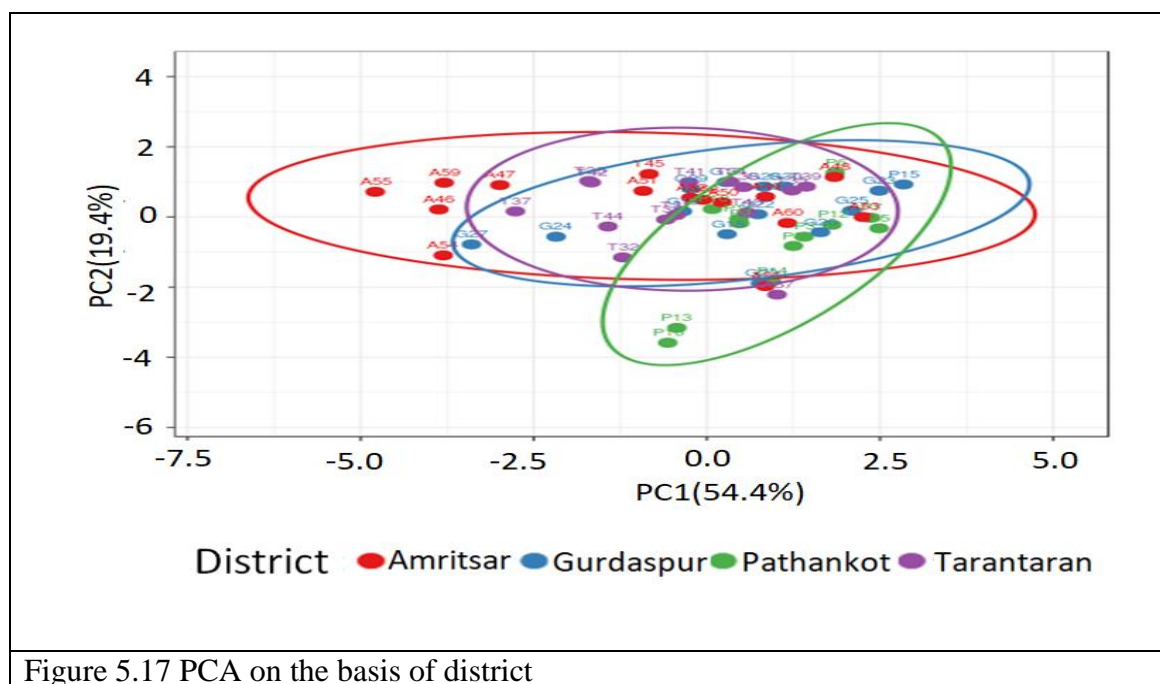
Last PCA was employed on the basis of district and it was also observed that no specific classification can be done on the basis of the district of the collected samples.

It is observed from the PCA graph in figure 5.14, that four parameters: turbidity, pH, chlorine and total hardness are on the positive side where as the other two parameters electrical conductivity and total dissolved solids are on the negative side of the plot. This suggests a correlation between the respective parameters. In the figure 5.15, 73% of the variance is covered by the two principal components. No perfect separation is seen in this PCA. In the figure 5.16 the variance covered is 73%, samples vary with depth but the overlapping is also seen. In the figure 5.17, PCA on the basis of districts, no clear classification is seen. Many samples are observed to exceed the WHO/BIS limits for turbidity, pH, TDS, total hardness and electrical conductivity. From the PCA graphs it is not possible to categorize the samples on the basis of source, depth and districts but these help in the understanding of the relationship between the parameters and also help in reducing the dimensions of the data. The high EC value and TDS levels which exceeds the WHO/BIS

limit of 500mg/L indicate salinity of the drinking water. The PCA graphs help in identifying the sources and also help in identifying the depths for which the samples failed the BIS/WHO standards. It makes it possible to monitor the groundwater samples and action can be taken accordingly.

5.6 CONCLUSION

To follow the WHO and BIS guidelines it was observed that 16.67% samples in case of Turbidity, 45% in case of TDS, 6.67% in case of electrical conductivity, 10% in case of pH value, 65% in case of Total hardness and 0% in case of chloride are out of range. The percentage of out of range values of properties in the same order as above in case of Amritsar are 13.33, 66.67, 20.00, 66.67, 73.33, 0; in case of Pathankot 26.67, 6.67, 0, 26.67, 53.33, 0; in case of Gurdaspur 15.38, 30.77, 0, 0, 53.33, 0 and in case of Tarn taran 6.67, 66.67, 6.67, 13.33, 66.67, 0.   When correlation among these properties is observed it was found that correlation between TDS and EC is 0.9997, TDS and TH is 0.5973, TDS and Cl is 0.7168; between EC and TH is 0.5933, between EC and Cl is 0.7174, between TH and Cl is 0.7989. Only six pairs showed some correlation. For these six pairs regression equations have also been obtained. Further Principal component analysis was performed on the basis of depth, source of water and district, which showed that categorization of the samples, is not possible in any of the case. All the physical and chemical data and NIR data were combined and the Clustvis webtool application was used to do the PCA. The NIR data was imported in this webtool in the form of a matrix. On the basis of this data, clustvis webtool application performed the Principal component analysis (PCA). This was done to classify the data on the basis of sources of water, districts, regions viz. rural and urban, depth of the source of water. The graphs obtained are shown in figures above.

CHAPTER 6

PREDICTION OF PARAMETERS OF GROUND WATER USING NEAR INFRARED AND FOURIER INFRARED SPECTROSCOPY AND MULTIVARIATE ANALYSIS

CHAPTER 6 - PREDICTION OF PARAMETERS OF GROUND WATER USING NEAR INFRARED AND FOURIER INFRARED SPECTROSCOPY AND MULTIVARIATE ANALYSIS

6.1 INTRODUCTION

Water is an important entity in the life of mankind and it is necessary that humans use pollution free water for their daily needs. Nearly one third of world's population has the reach of the groundwater for their daily needs like drinking cooking etc., and it is in the pipe line to provide groundwater to the 700 million population (not getting water) in the future [119]. Groundwater is an essential source of freshwater, it accounts for approximately 30.1% of the world's available freshwater [120]. It becomes necessary that the drinking water must remain pollution free. The reasons [121] which are responsible for its contamination and how to get rid of these contaminations must be known. The contaminated or polluted water if consumed can cause various types of diseases in humans as well as in animals [122]. Various causes of water contamination may include Urbanization, industrialization, use of chemicals in agriculture, use of plastic and micro-plastics etc. The contaminated water may sometimes be useful for drinking e.g. if it has some minerals within acceptable limits like calcium, magnesium etc. The hard water which may be useful for drinking purpose but it can harm the boilers. When the contamination is beyond acceptable limit it becomes unfit for drinking and sometimes becomes unfit in agriculture and even in industry. One of the anthropogenic causes of water pollution includes the use of pesticides, insecticides and fertilizers etc. Punjab is mainly agricultural land and these chemicals are used in huge quantities by the farmers. These chemicals mix up in the water and can leach into the groundwater thus making it polluted.

There is need of pollution free and clean water for all and to put checks on the contamination of water [123]. For this lot of testing and awareness among people is needed. It is also necessary and duty of all to use water in judicious manner and also to stop the wastage of water. There should be arrangement of water testing labs which are easily accessible to all the persons so that in case of any doubt water can be tested for pH, TDS, presence of metals, electrical conductivity, fluorides [124] and other contaminants. In this present work focus is mainly on the groundwater of all the four districts of Majha region of

Punjab. In this present work an effort has been made to check the contamination of water using spectroscopic technique, which is cost effective and non-destructive technique.

Electromagnetic radiations are made to fall on the groundwater samples and the changes in some of the property of the electromagnetic radiations like wavelength, frequency, wavenumber etc. is observed. In the study of groundwater samples near infrared region of the spectrum is used [125]. This region has energy less than the energy of the red light (which is at one end of visible radiation) and has least energy among all the seven colours. NIR region has further less energy than the red colour and lies approximately between 750 nm to 2500 nm. This energy is capable of making molecules (which follow certain conditions) shift between vibrational energy levels and rotational energy levels by absorbing radiations. NIR region is one among the three parts of infrared region other two parts are mid infrared region and far infrared region, NIR region has highest energy among these three parts. It is worth mentioning that infrared radiations were discovered by Hershel in 1800 [126]. The spectra obtained consist of two parts one is fingerprint region which lies approximately below 1500 cm^{-1} and is unique for a particular molecule, the remaining spectra above 1500 cm^{-1} shows peaks for the particular bonds present in the sample [127] and combination bands lie in the range from 2500 - 1900 cm^{-1}.

When radiations fall on a molecule, it may absorb some part of the energy and some part of it gets transmitted. The plots of spectra obtained can be either absorbance or emission spectra. In the NIR spectra four overtones are observed the fourth overtone has very low intensity, in the NIR spectra there can be combination bands also [128]. The absorption peaks can be due to C–H bond, N–H bonds, O–H bonds are present in alcohols and water and in carboxylic acid etc., S–H bonds are present in thiols etc., C=O bonds are present in ketones, carboxylic acids, aldehydes etc., C–H bonds are present in aromatic compounds, hydrocarbons etc. N–H bond may be present due to primary and secondary amines and some amides [129] etc. Organic and inorganic compounds may contain some of these bonds and these may be present in groundwater. NIR spectroscopy has been in use for many years and it has been used in food industry [130], pharmaceutical [131], textile [132], material science [133], wood industry [134] and is used in other fields also. In this present work physicochemical parameters of ground water are to be determined, the literature review reveals that researchers have used different kind of methods to determine these parameters, out of the methods mentioned in the literature are tedious, laborious, time consuming and are not economical. Here by the use of the developed regression model and using NIR spectroscopy the parameters are determined and this method will be non-destructive, will

take very less time and there will be no need to go to chemical laboratories and also will be cost effective in comparison to the methods used by other researchers.

6.2 FLUORESCENCE ANALYSIS

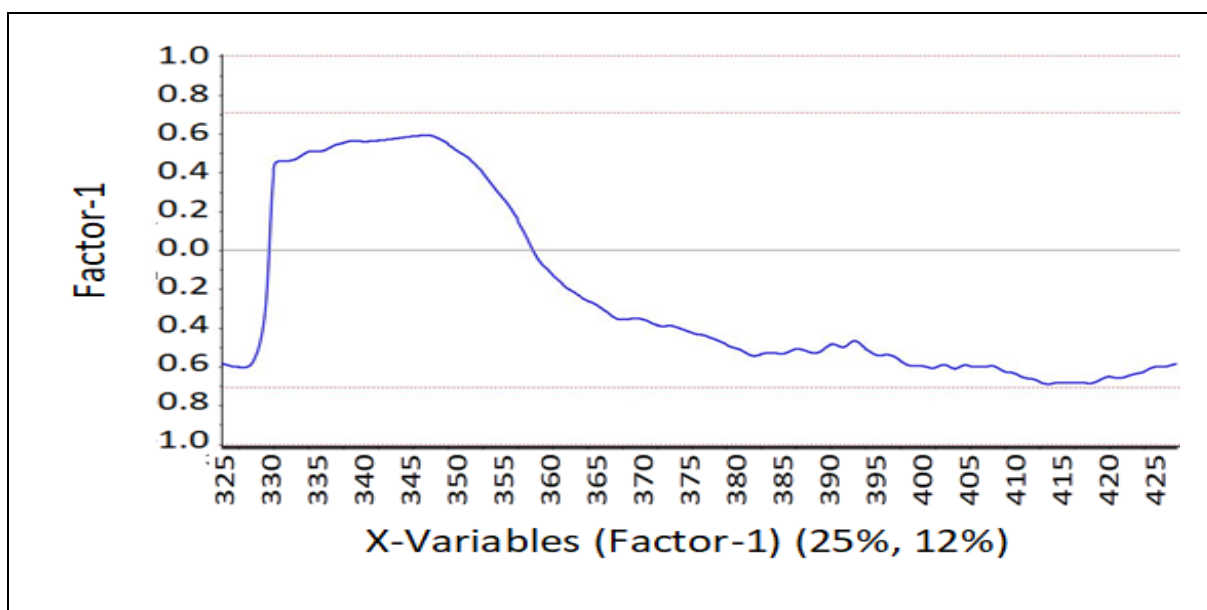The fluorescence graphs obtained for various parameters for correlation with the wavelengths are shown here.

Figure 6.1 Correlation from 325 nm to 425 nm for turbidity based on Fluorescence spectroscopy

It is observed from the Figure 6.1 that at 415 nm wavelength there is correlation near -0.7 and in the range 335 to 355 there is a weak correlation absorbance of fluorescence with turbidity.
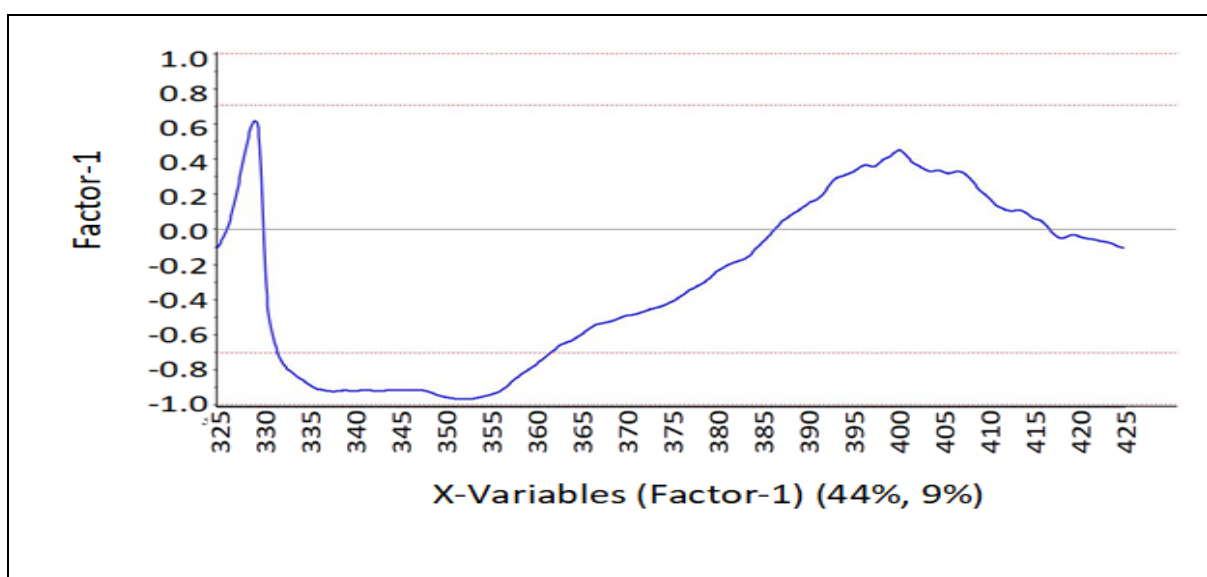

Figure 6.2 Correlation from 325 nm to 425 nm for total hardness based on Fluorescence spectroscopy
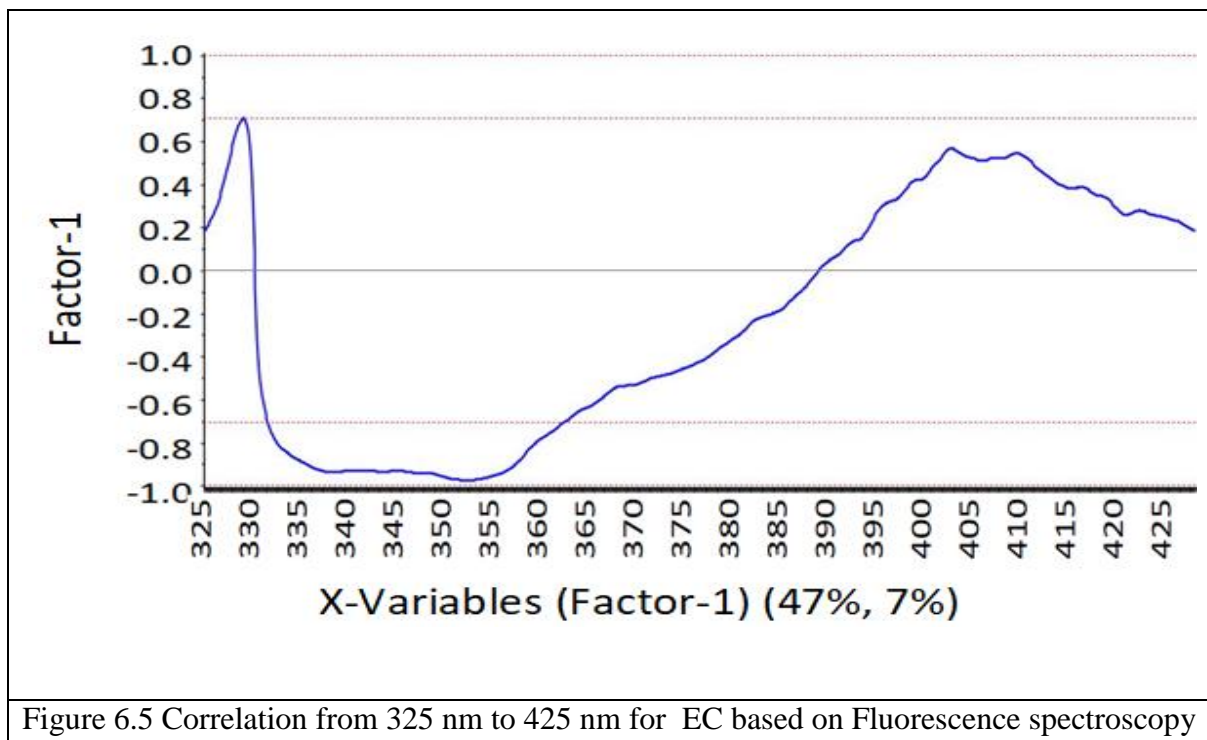
It is observed from the Figure 6.2 that there is a weak correlation in case of total hardness near 330 nm, in the positive side and in the range 335 nm to 365 nm a strong negative correlation is observed.

Figure 6.3 Correlation from 325 nm to 425 nm for TDS based on Fluorescence spectroscopy
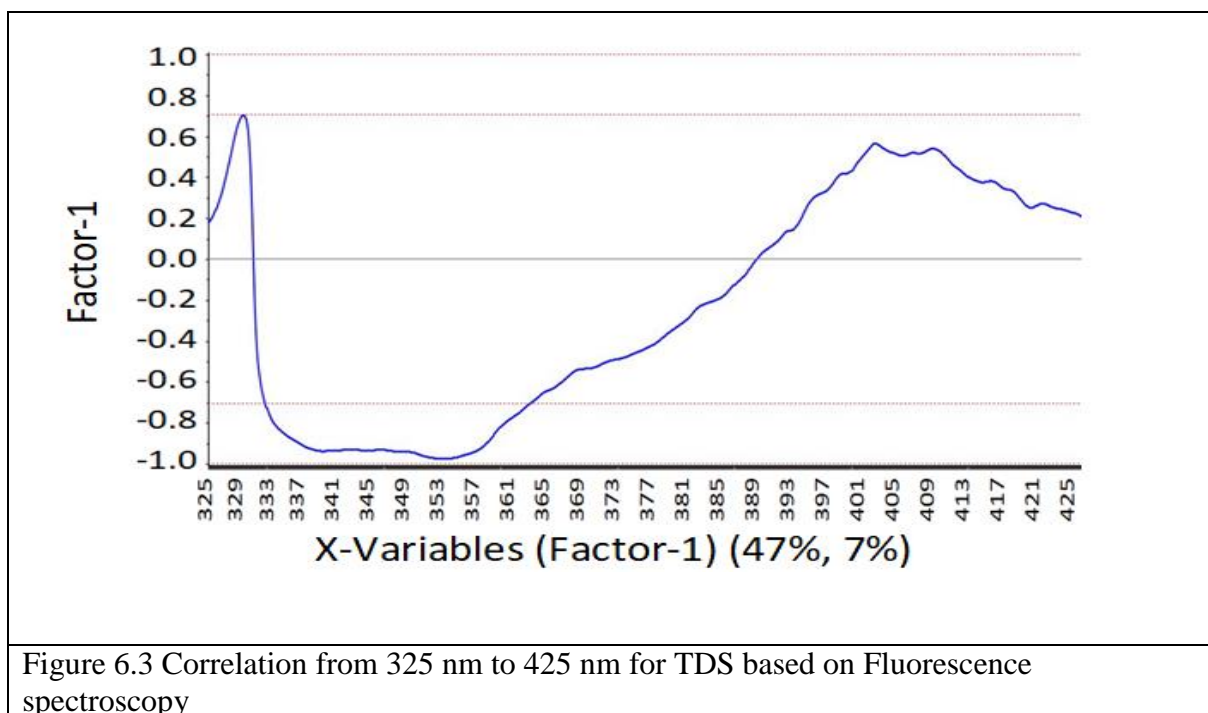
It is observed from the Figure 6.3 for TDS there is positive correlation at 330 nm and negative correlation in the range 330 nm to 360 nm.

Figure 6.4 Correlation from 325 nm to 425 nm for pH based on Fluorescence spectroscopy
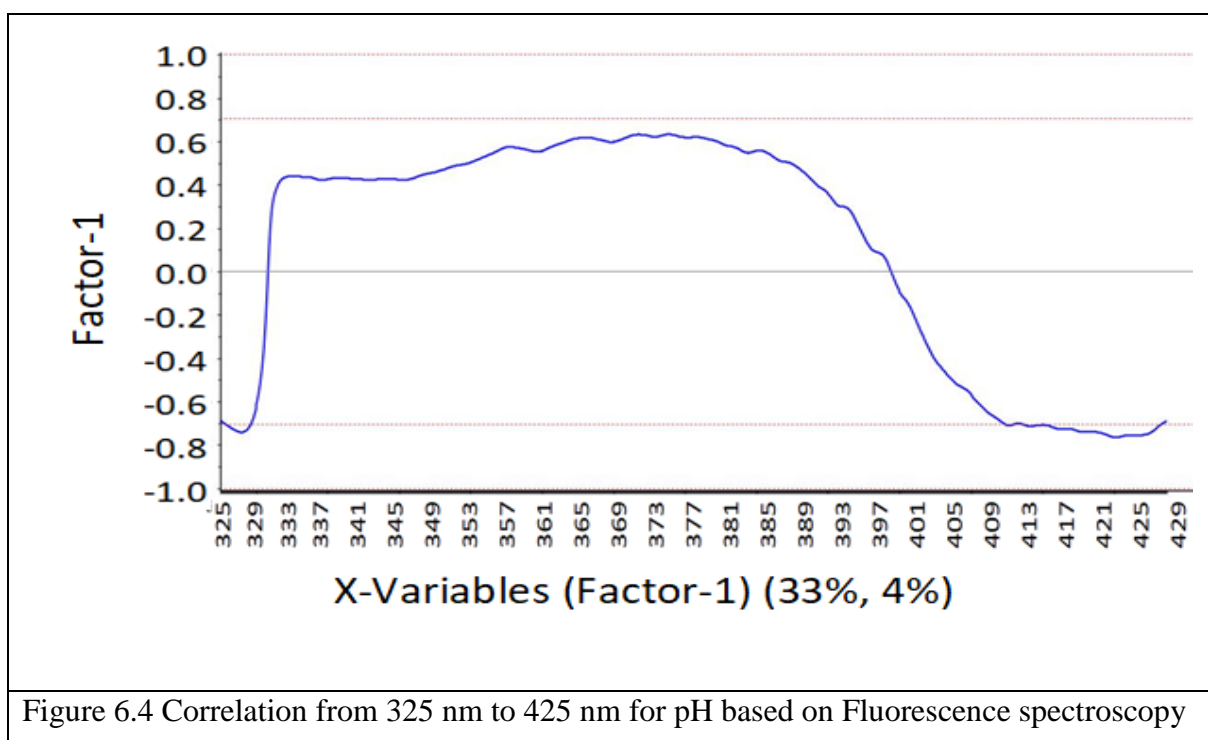
It is observed from the Figure 6.4 in case of pH value, there is negative correlation from 325-330 nm and from 410-430 nm and there is positive correlation from 370-380 nm.

Figure 6.5 Correlation from 325 nm to 425 nm for EC based on Fluorescence spectroscopy

It is observed from the Figure 6.5 that in case of EC there is positive correlation near 330 nm, and negative correlation in the range 335-365 nm.

Figure 6.6 Correlation from 325 nm to 425 nm for Cl based on Fluorescence spectroscopy
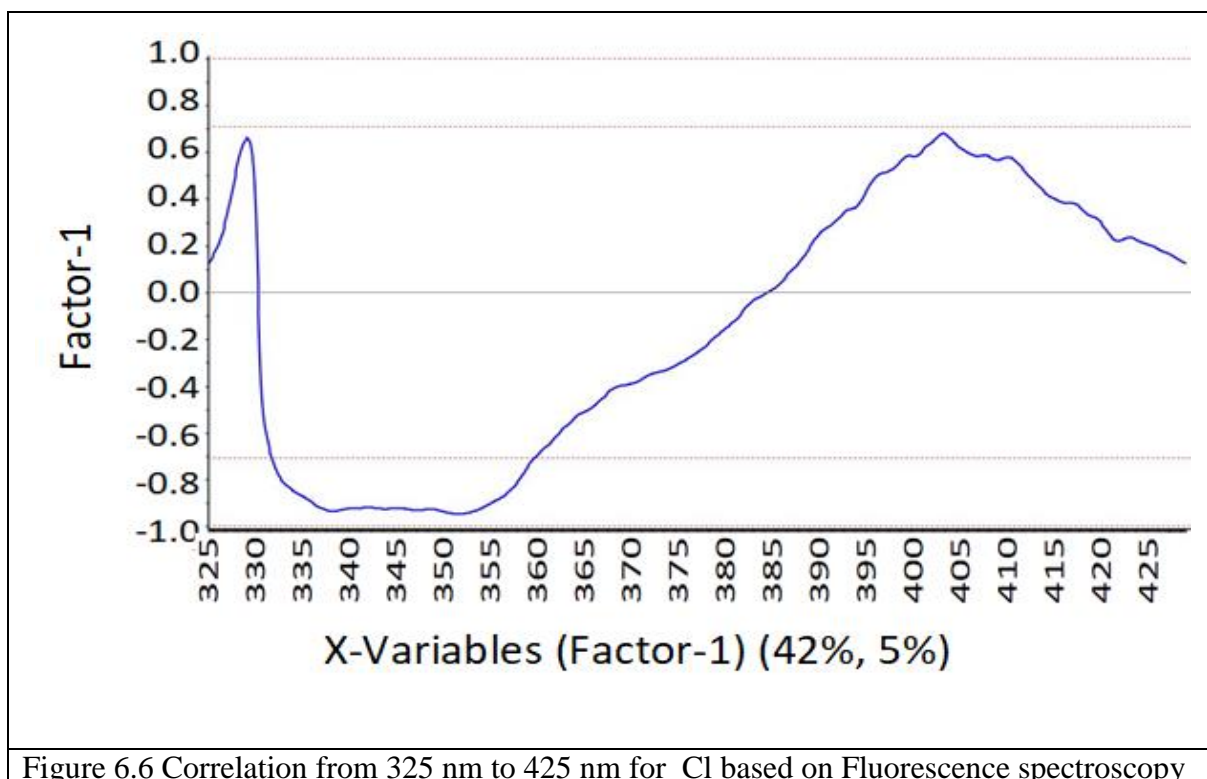
It is observed from the Figure 6.6 that in case of chlorine content there is positive correlation at 330 nm and near 400 nm. There is negative correlation in the range 335-360 nm.

Figure 6.7 Correlation loading TDS based on Fluorescence Spectroscopy
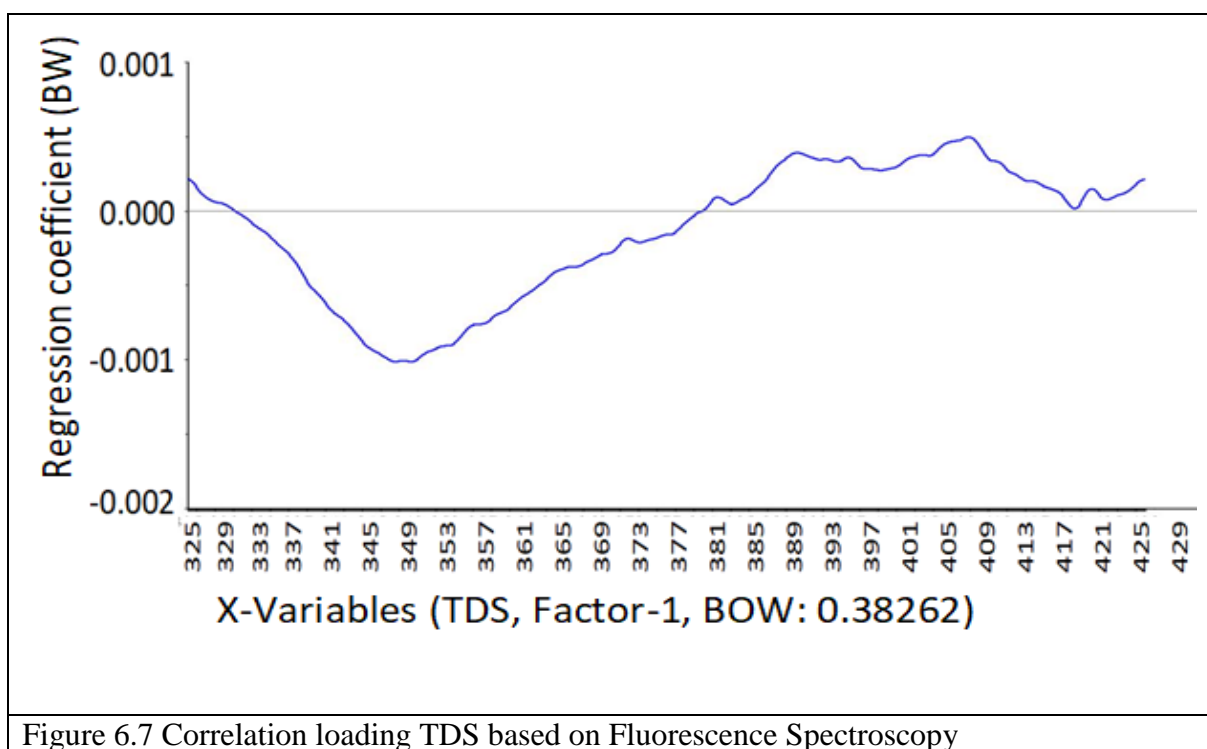
In the Figure 6.7, there is regression coefficient on Y axis and wavelengths from 325 to 430 nm on the X axis. This graph is used to get information about the wavelength selection by regression coefficients of the PLS model. The graph is for TDS correlation loading and it remains close to zero and thus there is no correlation observed.

Figure 6.8 Correlation loading total hardness based on Fluorescence spectroscopy
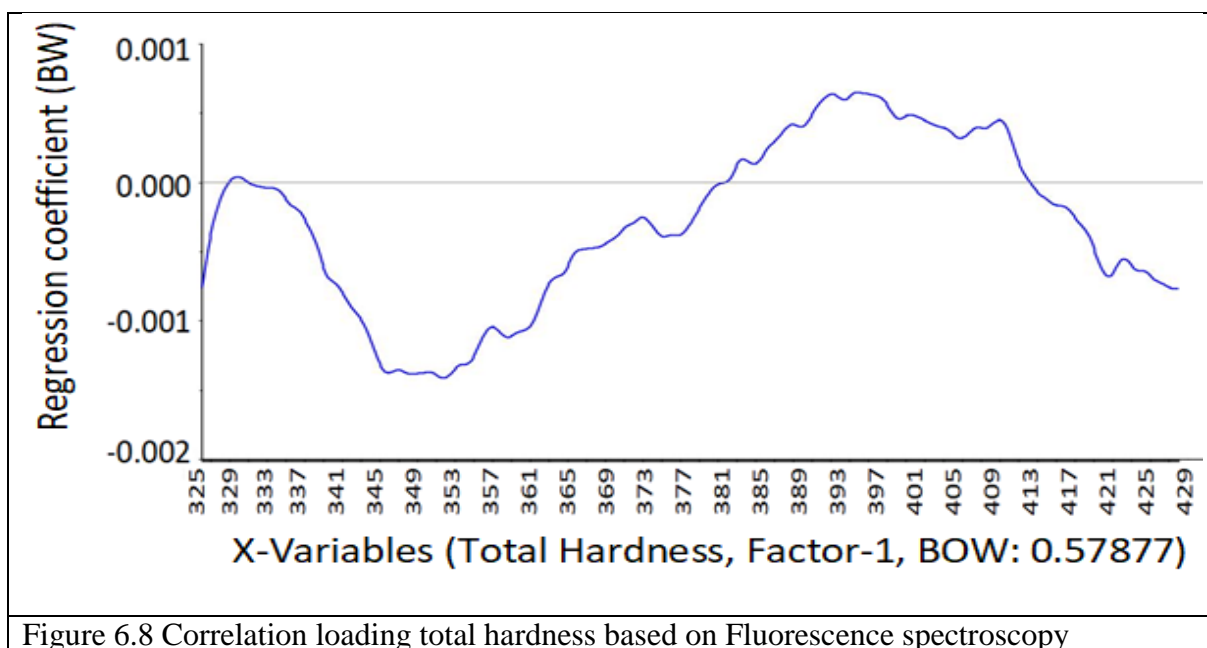
In the Figure 6.8, there is regression coefficient on Y axis and wavelengths from 325 to 430 nm on the X axis. This graph is used to get information about the wavelength selection by regression coefficients of the PLS model. The graph is for total hardness correlation loading and it remains close to zero and thus there is no correlation observed.
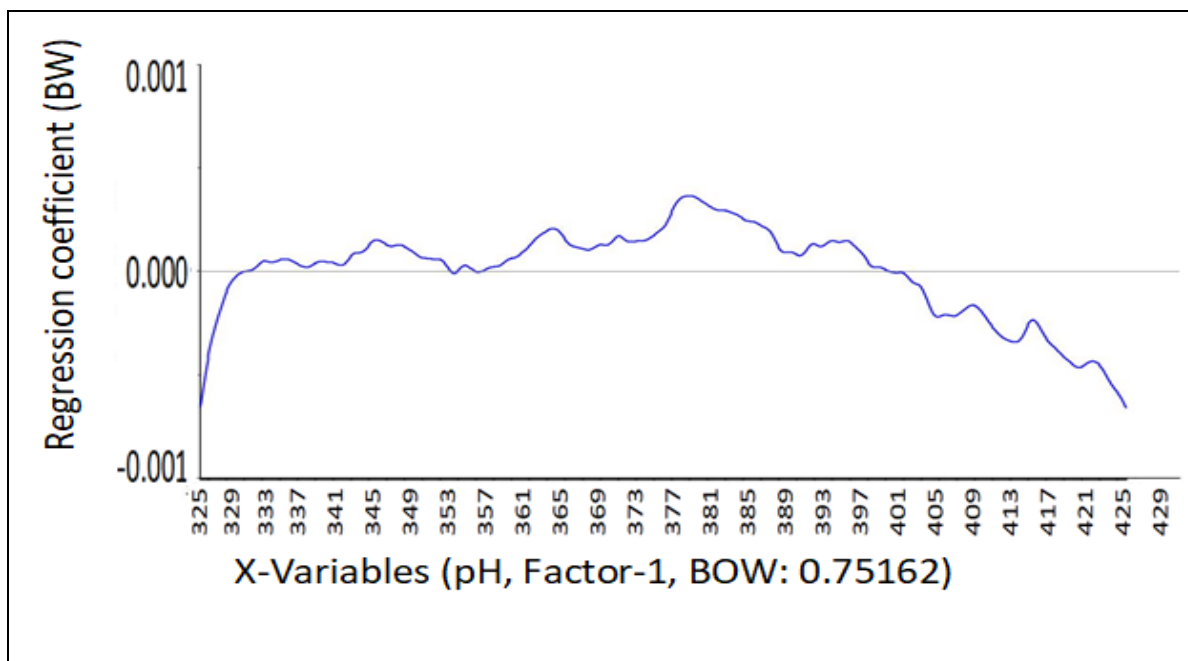

Figure 6.9 Correlation loading pH based on Fluorescence spectroscopy

In the Figure 6.9 there is regression coefficient on Y axis and wavelengths from 325 to 430 nm on the X axis. This graph is used to get information about the wavelength selection by regression coefficients of the PLS model. The graph is for correlation loading pH and it remains close to zero and thus there is no correlation observed.

Figure 6.10 Correlation loading EC based on Fluorescence spectroscopy

In the Figure 6.10 there is regression coefficient on Y axis and wavelengths from 325 to 430 nm on the X axis. This graph is used to get information about the wavelength selection by regression coefficients of the PLS model. The graph is for EC correlation loading and it remains close to zero and thus there is no correlation observed.
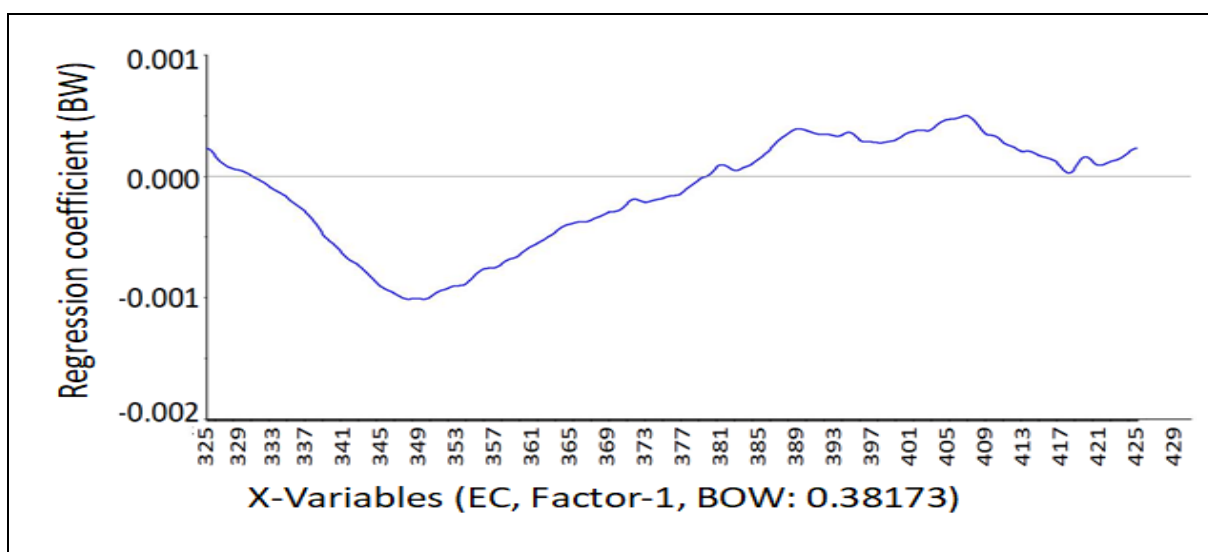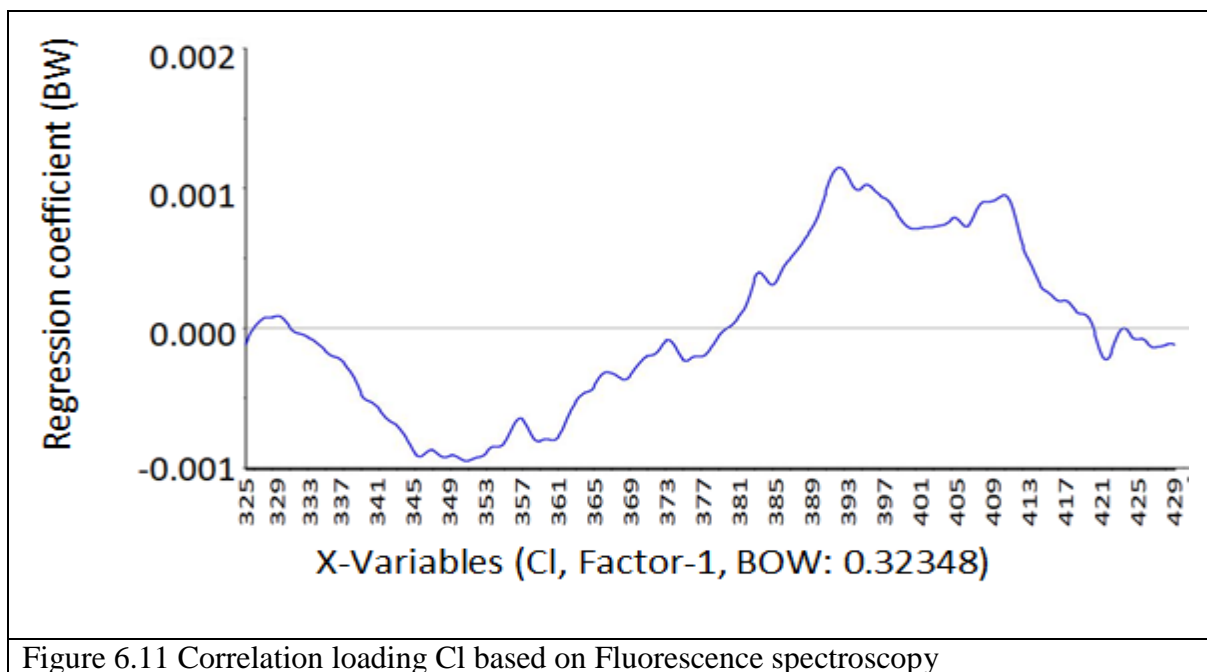

Figure 6.11 Correlation loading Cl based on Fluorescence spectroscopy

In the Figure 6.11 there is regression coefficient on Y axis and wavelengths from 325 to 430 nm on the X axis. This graph is used to get information about the wavelength selection by regression coefficients of the PLS model. The graph is for correlation loading chlorine and it remains close to zero and thus there is no correlation observed.

Figure 6.12 Fluorescent spectra of all the samples
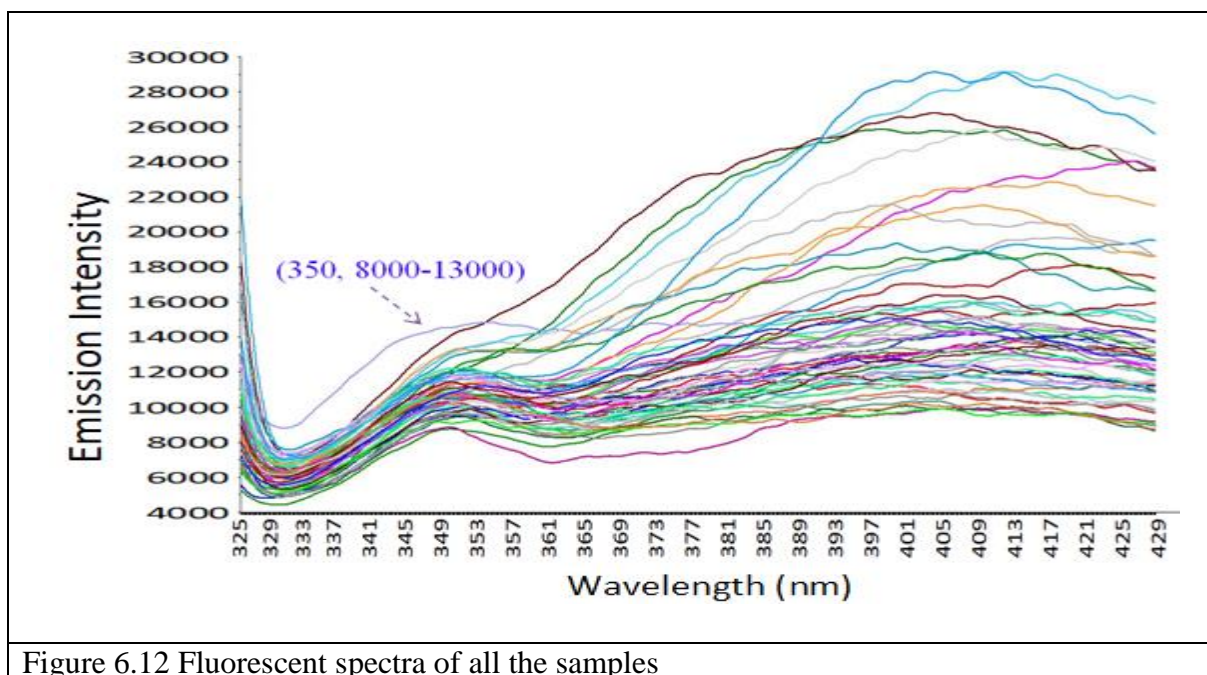
In the Figure 6.12 the fluorescent spectra of the samples is given, it shows that all the samples are behaving somewhat similarly. There are emission peak at 325 nm for all the samples and then the emission goes on decreasing for all the samples and then somewhat linearly goes on increasing from 330 to 400 nm and the emission are ranging from all the samples goes from 8000 to 28000.
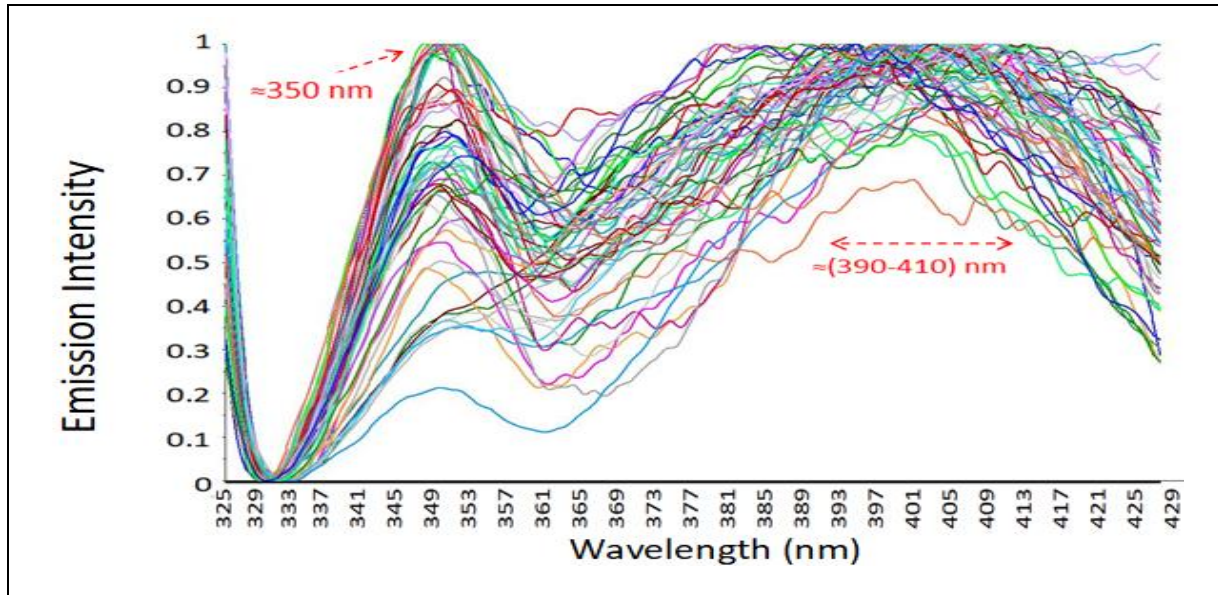

Figure 6.13 After preprocessing smoothing base line normalize fluorescence spectra for all the samples

In the Figure 6.13 the fluorescence spectra for all the samples is given after preprocessing smoothing base line normalized data. There is emission near 350 nm and from 390 nm to 410 nm.

Figure 6.14 Correlation with total hardness based on Fluorescence spectroscopy

From the Figure 6.14 it is observed that there is no correlation of total hardness with the wavelengths.

Figure 6.15 Correlation with TDS based on Fluorescence spectroscopy

From the Figure 6.15 it is observed that there is no correlation of TDS with the wavelengths.

Figure 6.16. Correlation with pH based on Fluorescence spectroscopy

From the Figure 6.16 it is observed that there is no correlation of pH with the wavelengths.

Figure 6.17 Correlation with EC based on Fluorescence spectroscopy
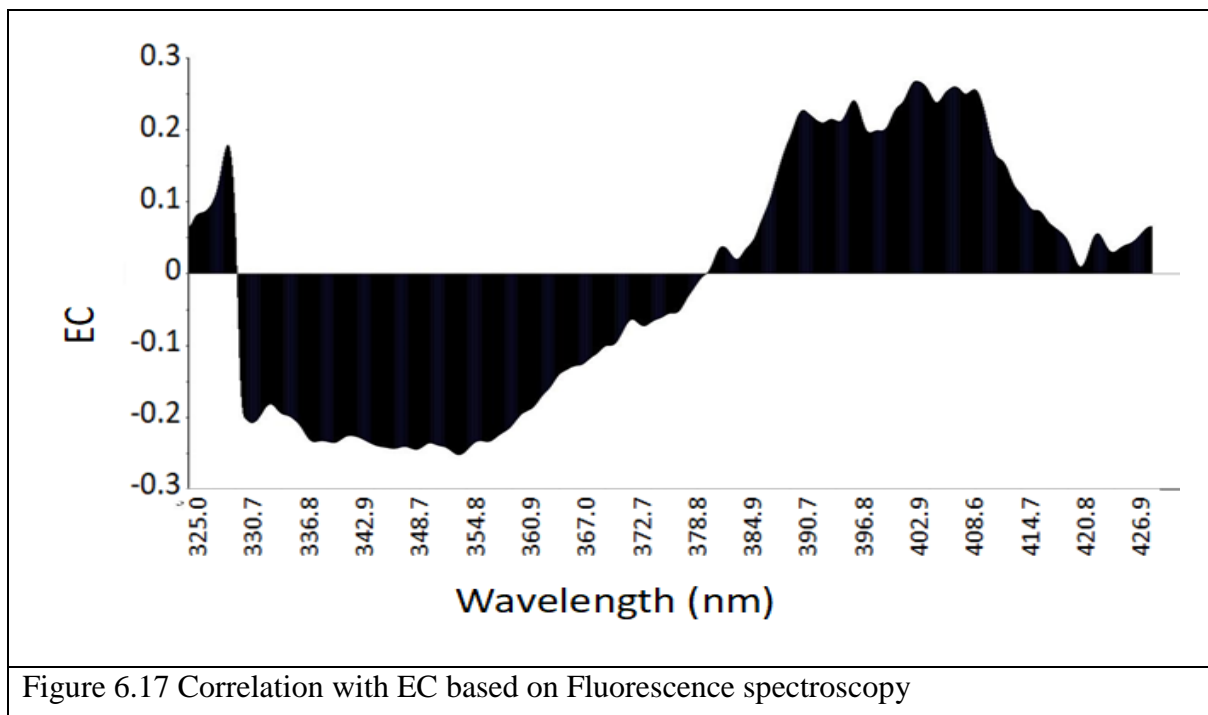
From the Figure 6.17 it is observed that there is no correlation of EC with the wavelengths.

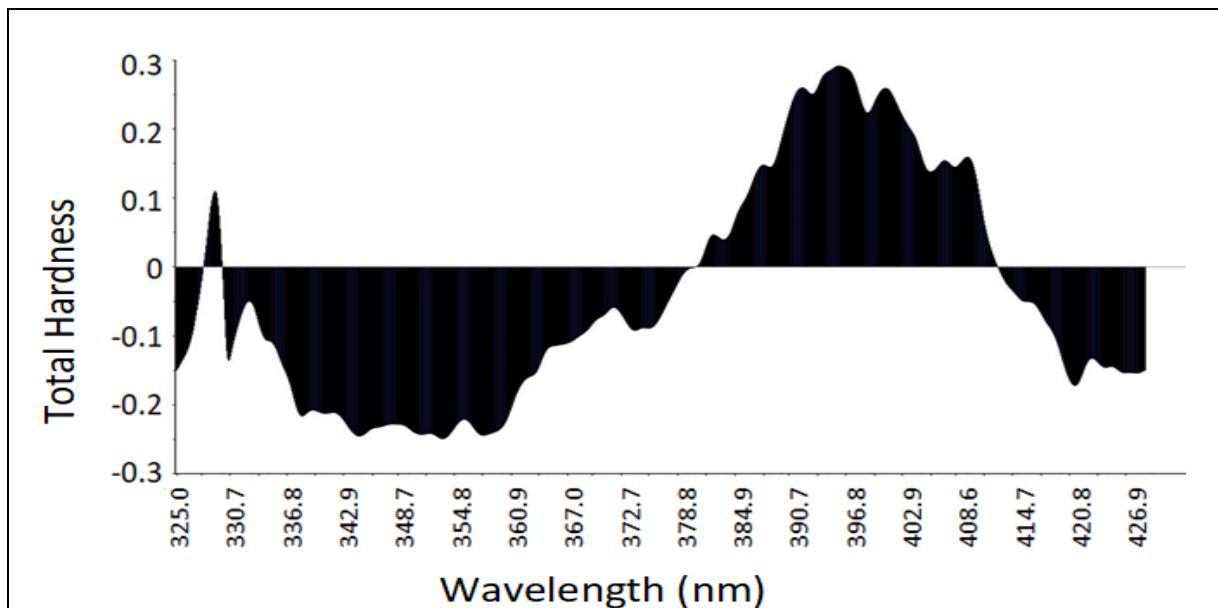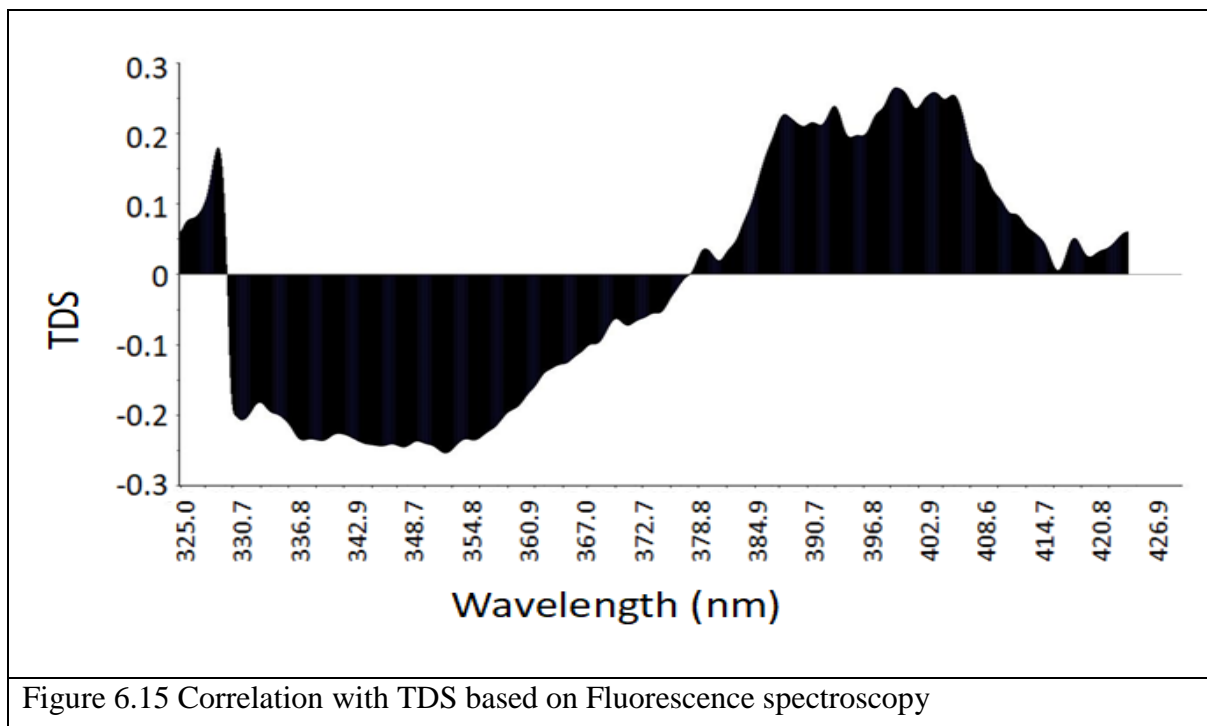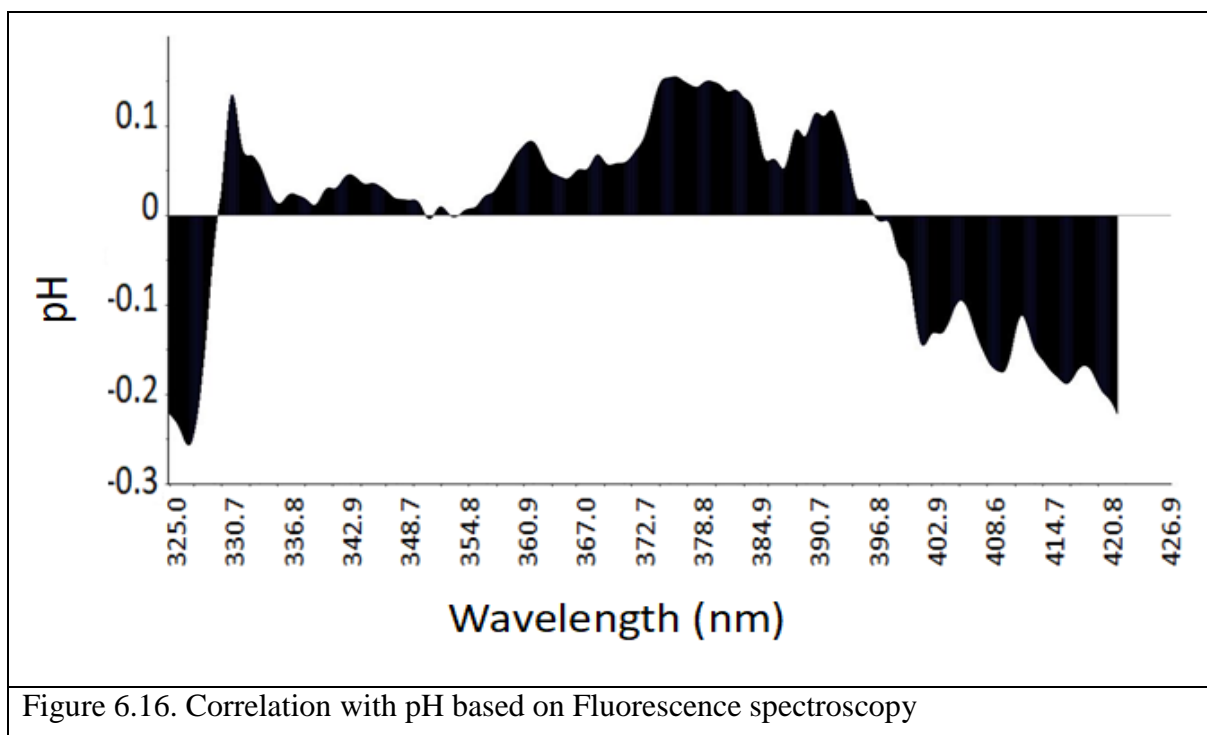Figure 6.18 Correlation with Cl based on Fluorescence spectroscopy
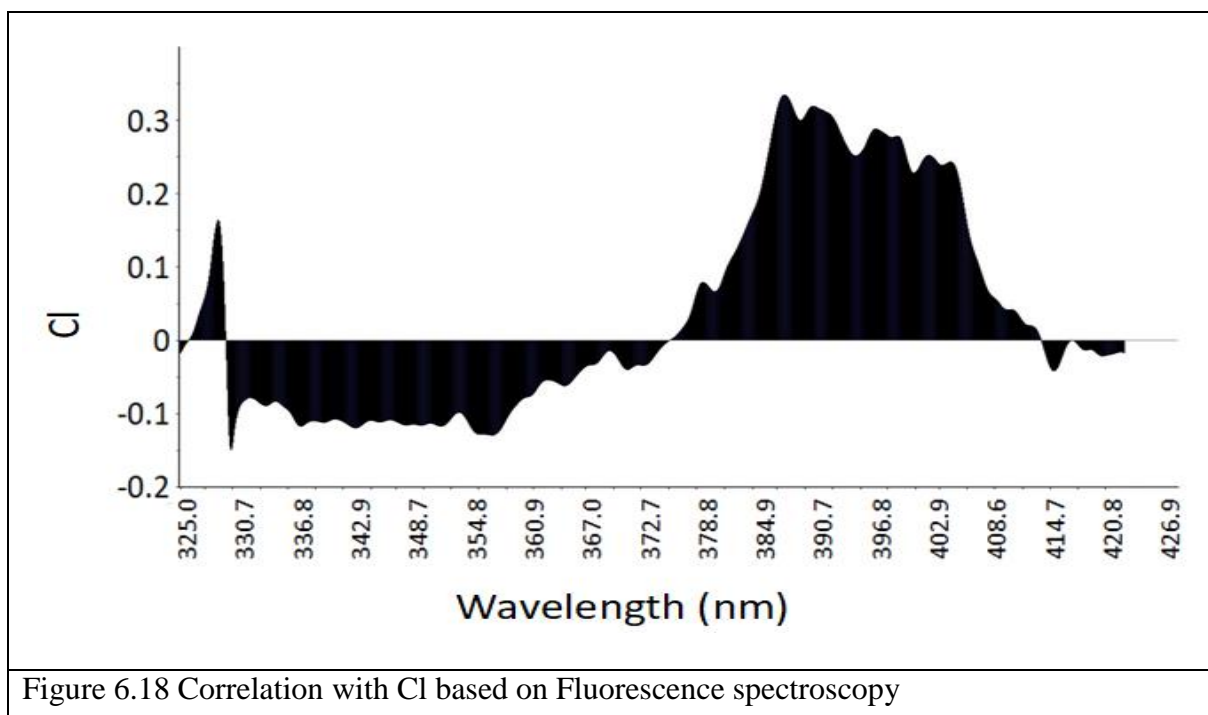
From the Figure 6.18 it is observed that there is no correlation of chlorine with the wavelengths.

Figure 6.19 Correlation with turbidity based on Fluorescence spectroscopy
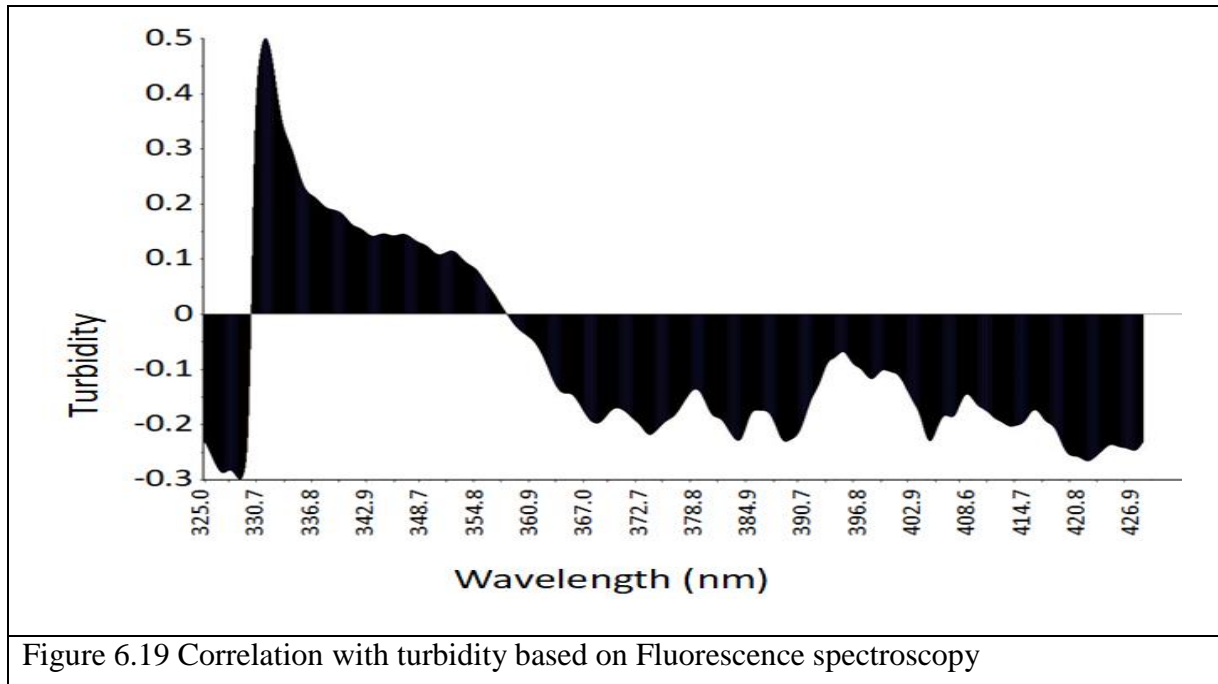
From the Figure 6.19 it is observed that there is no correlation of turbidity with the wavelengths

6.3 NIR SPECTRAL ANALYSIS:

Spectral data plots of groundwater samples are shown in the Figure 6.20 for the raw data and the baseline and baseline normalized graphs are given in Figure 6.21 and Figure 6.22 respectively. The plot shown in Figure 6.23 is obtained after applying second derivative.

Figure 6.20   NIR Graph for all the samples from 700-2500 nm (Raw data)
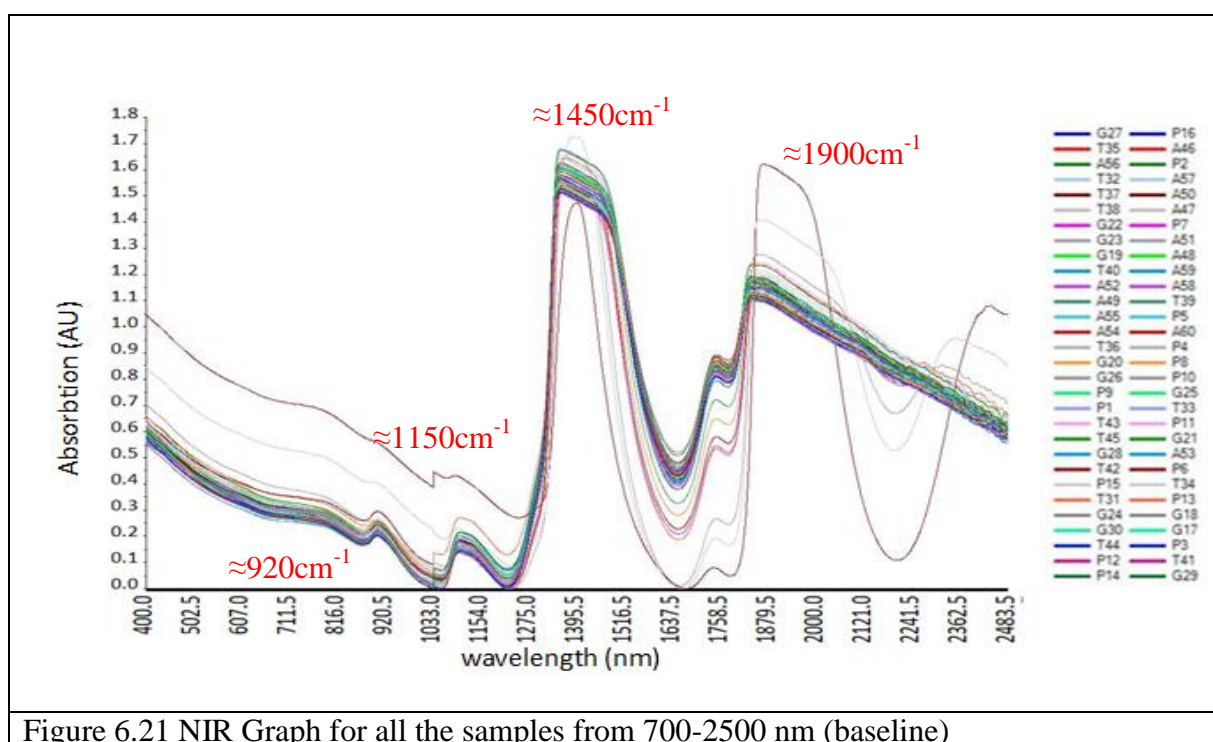
Figure 6.21 NIR Graph for all the samples from 700-2500 nm (baseline)
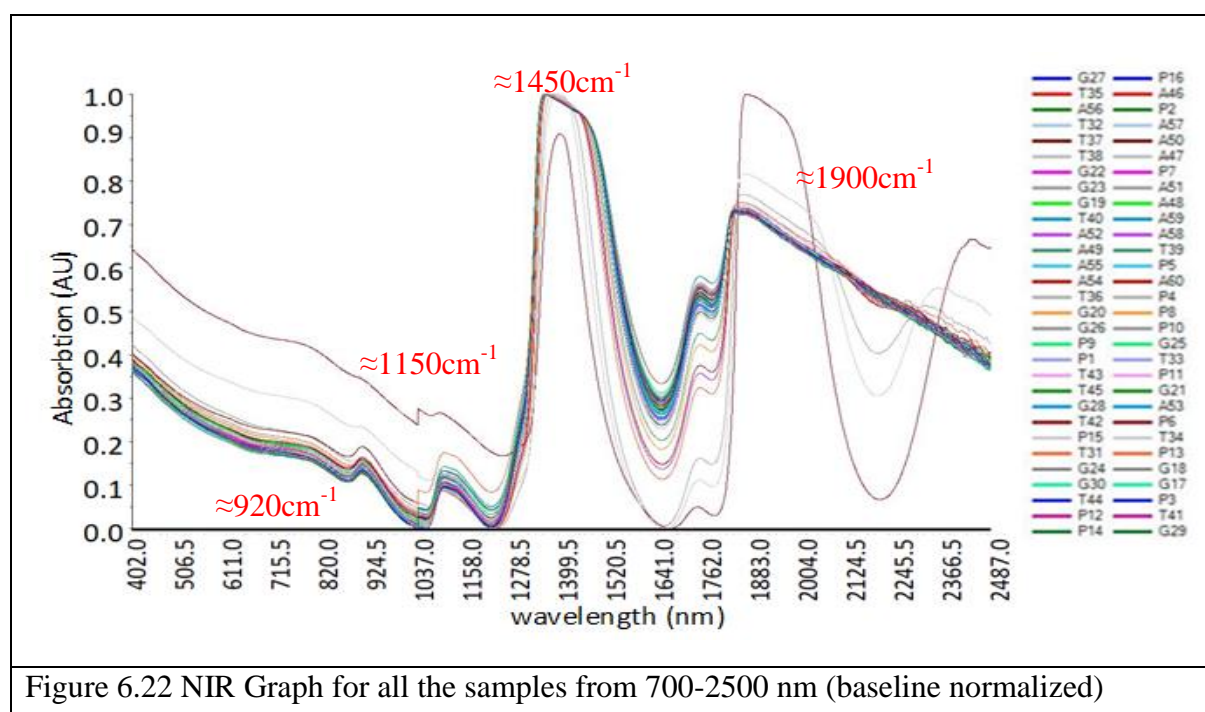
Figure 6.22 NIR Graph for all the samples from 700-2500 nm (baseline normalized)

Figure 6.23   NIR Graph for all the samples from 700-2500 nm (2nd derivative)
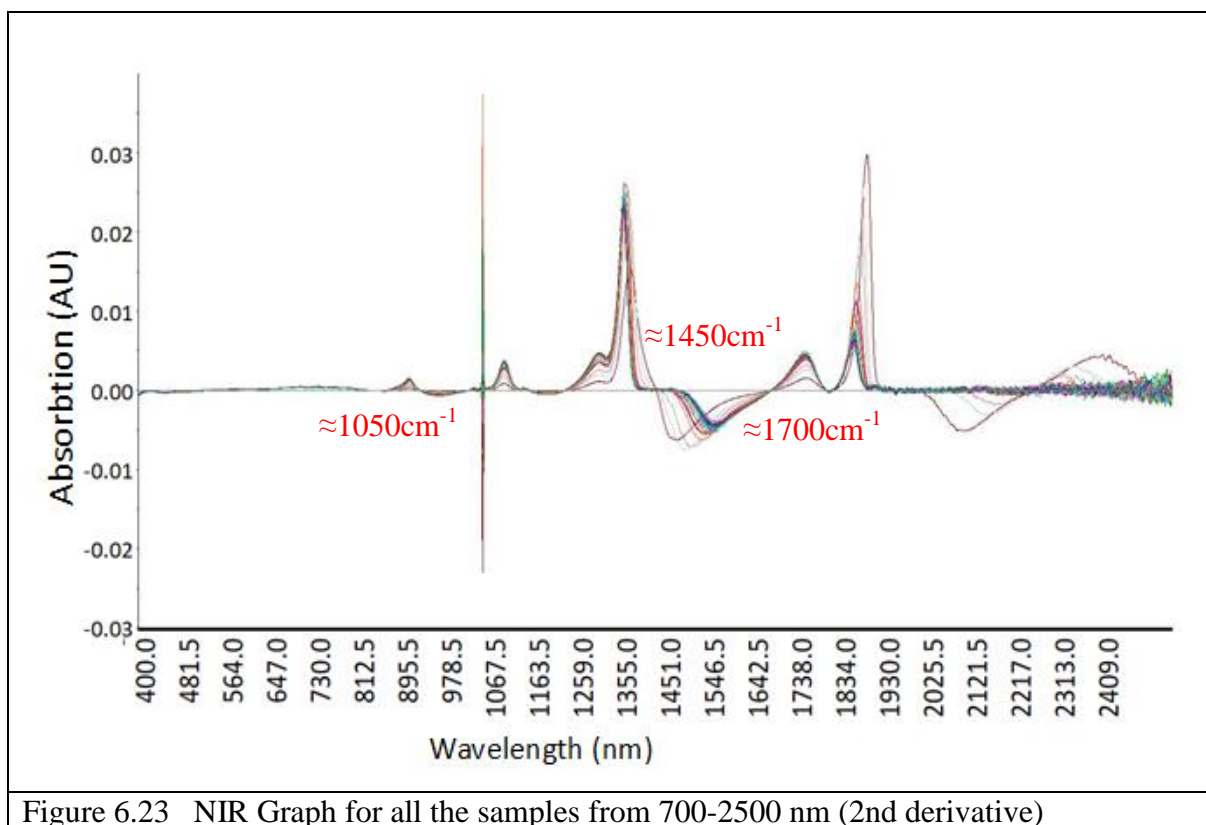
It is clear from the plots that there are very weak absorption peaks near 1100 nm, 1200 nm, and 1300 nm. There are peaks observed at the wavelengths 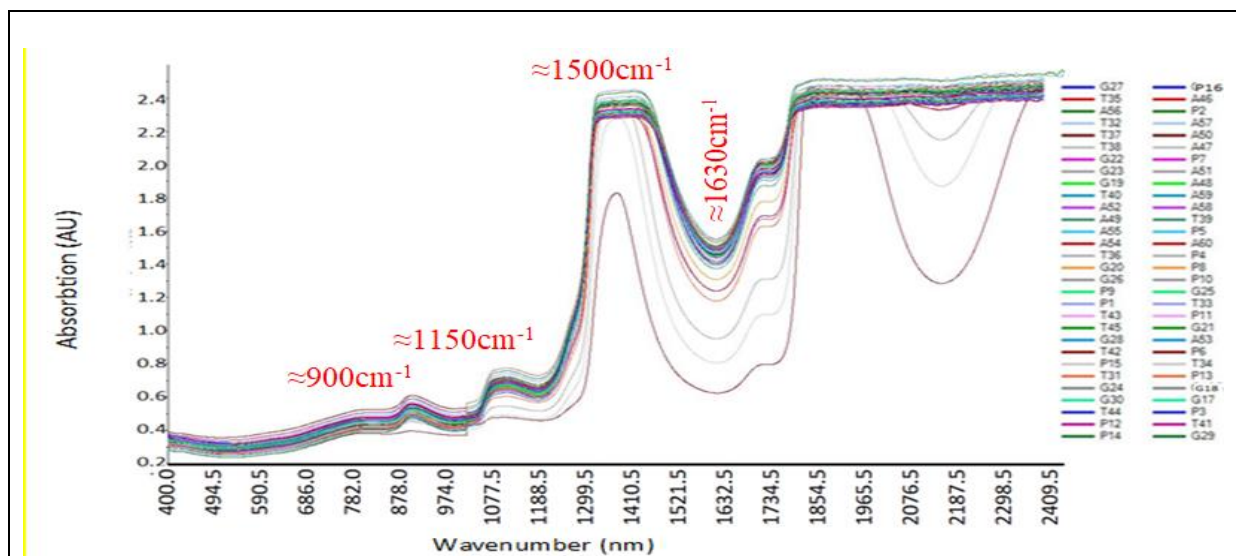near 1500 nm and 1700 nm indicating the presence of bonds at these wavelength values. The bonds which may be present are C-H, C=O overtones and may be combination bands. The similar is observed from the graph obtained after applying 2nd derivative.

6.4 PLS REGRESSION PLOTS

Scattered plots between reference value and predicted value for various physicochemical parameters with PLS regression model between 700 to 2500 nm range for near infrared spectroscopy are given in figures below. There are two important variables which are used to confirm about the acceptance of the regression model, these are root mean square error (RMSE) and coefficient of determination (R^2). It is worth mentioning that for a model to be acceptable the root mean square error should be very-very small, whereas coefficient of determination should have value near or greater than 0.7. In the plots measured values are on the X-axis and calculated values from the model are on the Y-axis.

Figure 6.24 Scattered plot between the ref and predicted chlorine value with PLS regression between 700 to 2500 nm range (MSC–1D)

In the Figure 6.24 plot is shown for reference and predicted values with multiscattering correction first derivative for Chlorine. The root mean square error for calculated value is 0.0698815 and coefficient of determination is 0.8638657, which shows it may be acceptable but for the validation the values for RMSE is high and it is not acceptable for R^2. It can be concluded that this model is not acceptable.

Figure 6.25 Scattered plot between the ref and predicted electrical conductivity value with PLS regression between 700 to 2500 nm range (MSC-1D)

In the Figure 6.25 plot is shown for reference and predicted values with multiscattering correction first derivative for electrical conductivity. The root mean square error for calculated value is 0.0547022 and coefficient of determination is 0.884102, which shows it may be acceptable but for the validation the values for RMSE is high and it is not acceptable for R^2. It can be concluded that this model is not acceptable.

Figure 6.26 Scattered plot between the ref and predicted total hardness value with PLS regression between 700 to 2500 nm range (MSC–1D)

In the Figure 6.26 plot is shown for reference and predicted values with multiscattering correction first derivative for total hardness. The root mean square error for calculated value is 0.0139907 and coefficient of determination is 0.9802571, which shows it to be acceptable and for validation the value for RMSE is less 0.047051 and it is 0.7828855 for R^2. It can be concluded that this model is acceptable.
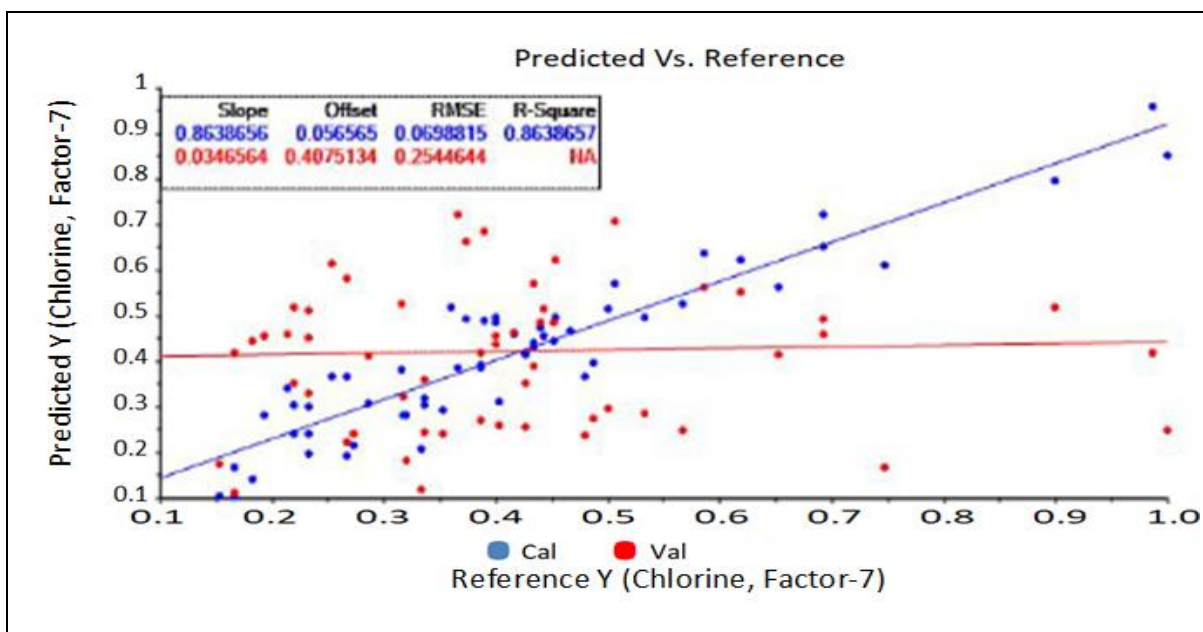

Figure 6.27 Scattered plot between the ref and predicted pH value with PLS regression between 700 to 2500 nm range (MSC-1D)

In the Figure 6.27 plot is shown for reference and predicted values with multiscattering correction first derivative for pH values, it is clear that the root mean square error for calculated value is 0.0181479 and coefficient of determination is 0.9257077, which shows it is acceptable and for validation the value for RMSE is 0.0877675 and it is not acceptable for R^2 so it can be concluded that this model is not acceptable.

Figure 6.28 Scattered plot between the ref and predicted chlorine value with PLS regression between 700 to 2500 nm range (MSC)

In the Figure 6.28 plot is shown for reference and predicted values with multiscattering correction for chlorine, it is clear that the root mean square error for calculated value is 0.1362157 and coefficient of determination is 0.4827533, which shows it is not acceptable and for validation the value for RMSE is high and it is not acceptable for R^2 so it can be concluded that this model is not acceptable.
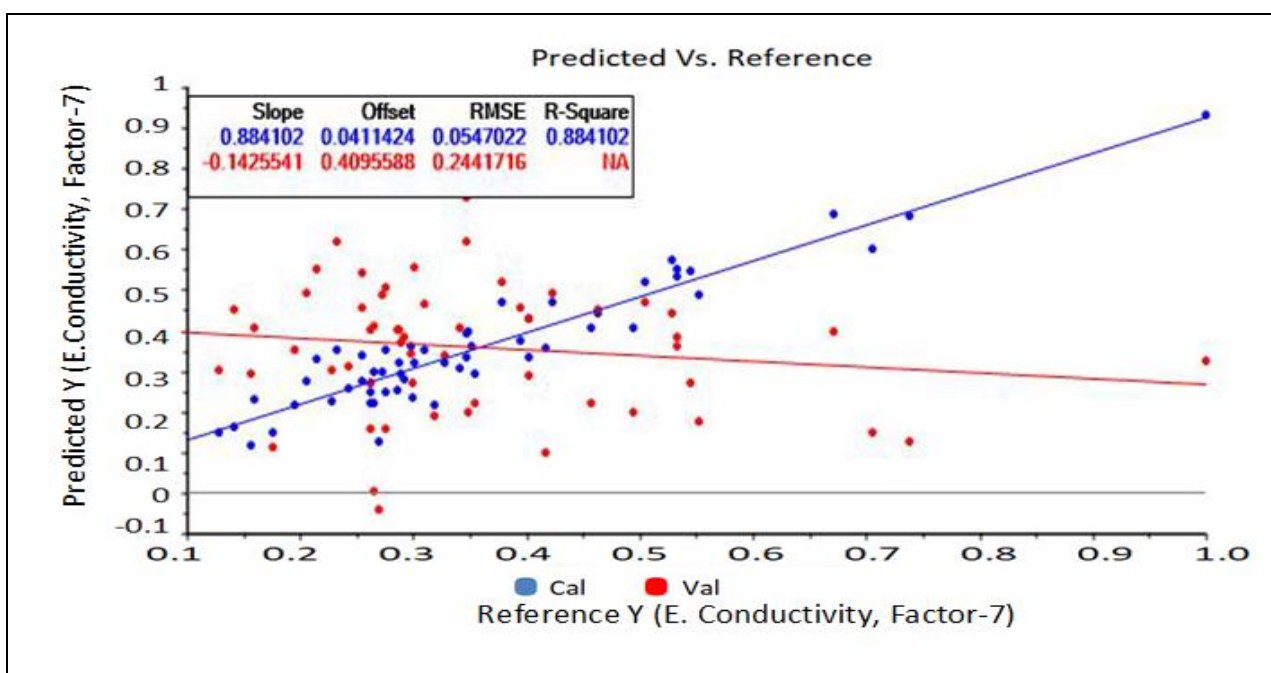

Figure 6.29 Scattered plot between the ref and predicted electrical conductivity value with PLS regression between 700 to 2500 nm range (MSC)

In the Figure 6.29 plot is shown for reference and predicted values with multiscattering correction for electrical conductivity, it is clear that the root mean square error for calculated value is 0.1124864 and coefficient of determination is 0.5099214, which shows it is not acceptable and for validation the value for RMSE is high and it is not acceptable for R^2 so it can be concluded that this model is not acceptable.
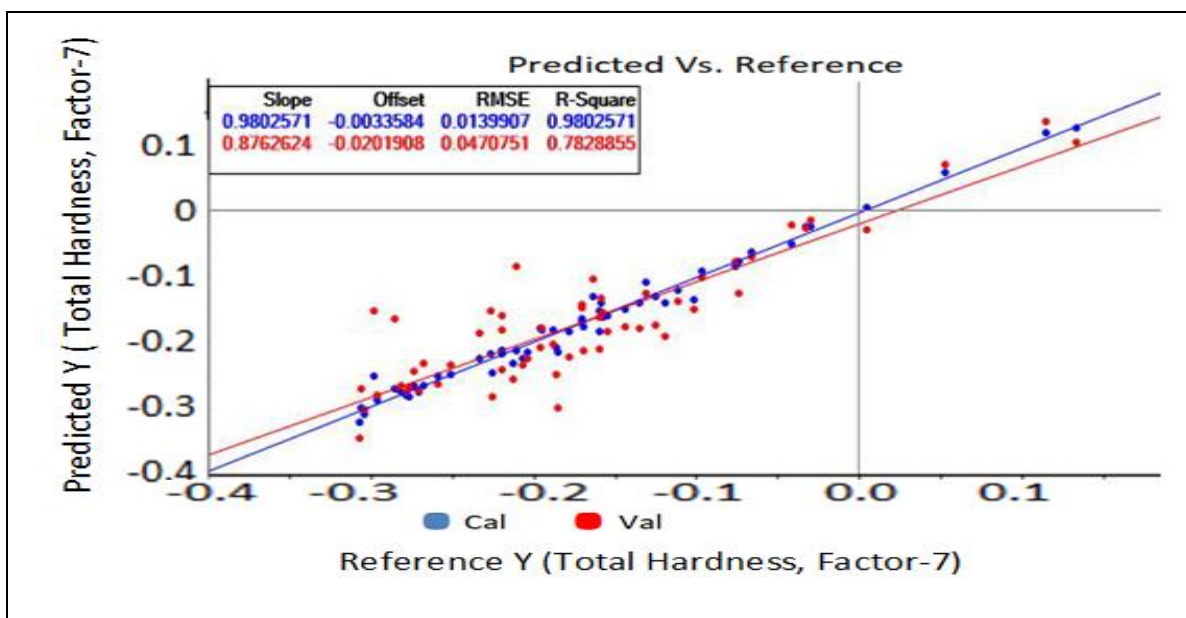

Figure 6.30 Scattered plot between the ref and predicted pH value with PLS regression between 700 to 2500 nm range (MSC)

In the Figure 6.30 plot is shown for reference and predicted values with multiscattering correction for pH, it is clear that the root mean square error for calculated value is 0.0453001 and coefficient of determination is 0.5371005, which shows it is not acceptable and for validation the value for RMSE is high and it is not acceptable for R^2 so it can be concluded that this model is not acceptable.
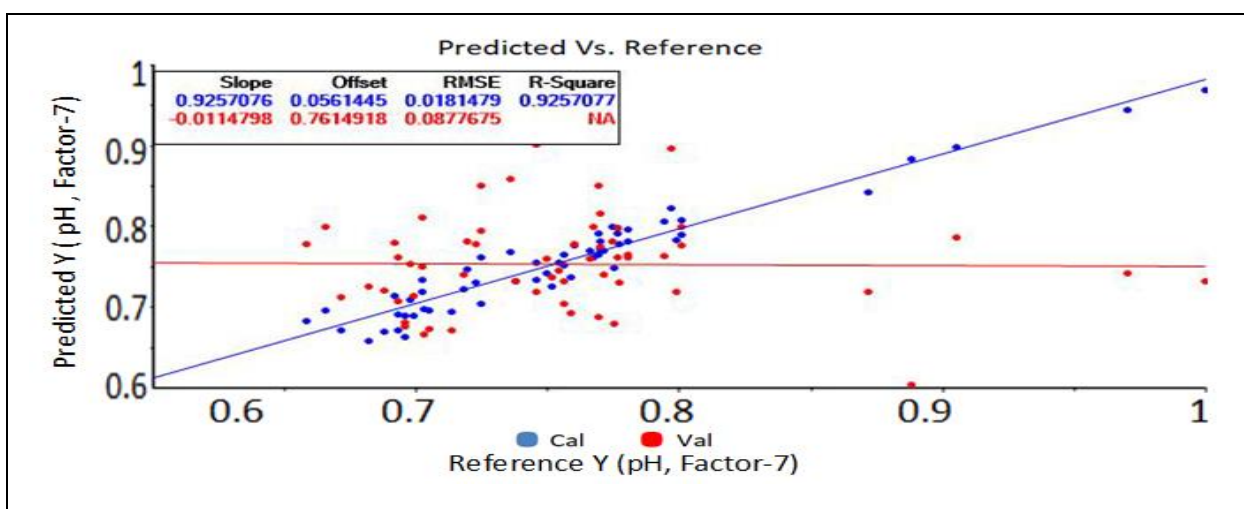

Figure 6.31 Scattered plot between the ref and predicted TDS value with PLS regression between 700 to 2500 nm range (MSC)

In the Figure 6.31 plot is shown for reference and predicted values with multiscattering correction for TDS, it is clear that the root mean square error for calculated value is 0.1125655 and coefficient of determination is 0.5084694, which shows it is not acceptable and for validation the value for RMSE is high and it is not acceptable for R^2 so it can be concluded that this model is not acceptable.
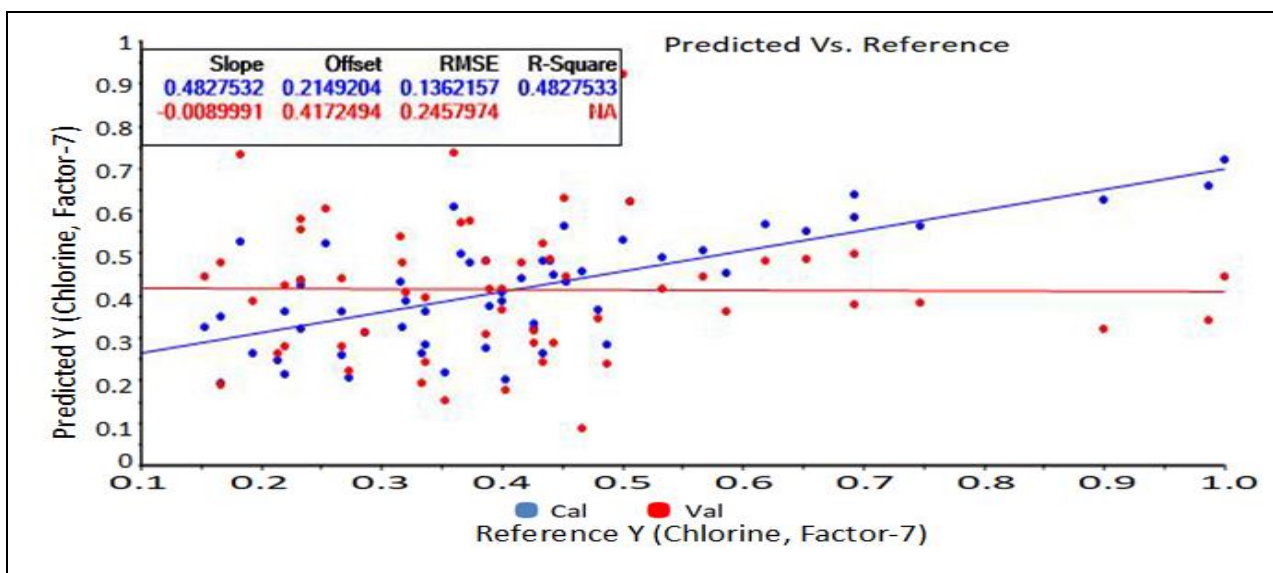

Figure 6.32 Scattered plot between the ref and predicted total hardness value with PLS regression between 700 to 2500 nm range (MSC)

In the Figure 6.32 plot is shown for reference and predicted values with multiscattering correction for total hardness, it is clear that the root mean square error for calculated value is 0.1304051 and coefficient of determination is 0.4843409, which shows it is not acceptable and and for validation the value for RMSE is high and it is not acceptable for R^2 so it can be concluded that this model is not acceptable.
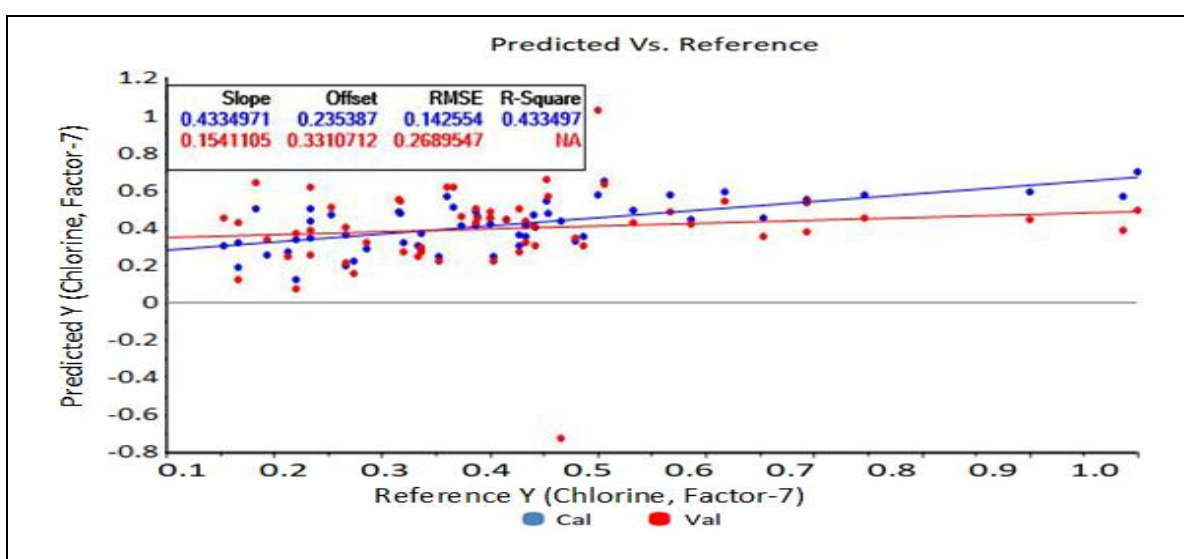
Figure 6.33 Scattered plot between the ref and predicted chlorine value with PLS regression between 700 to 2500 nm range (Raw)

In the Figure 6.33 plot is shown for reference and predicted values with raw data for chlorine, it is clear that the root mean square error for calculated value is 0.142554 and coefficient of determination is 0.433497, which shows it is not acceptable and for validation the value for RMSE is high and it is not acceptable for R^2 so it can be concluded that this model is not acceptable.
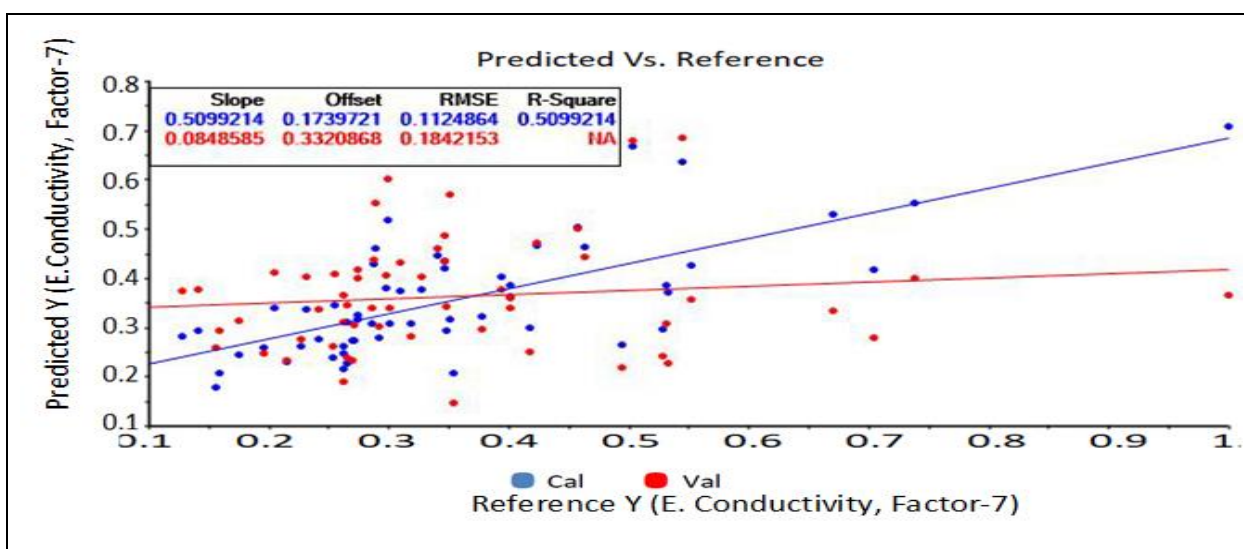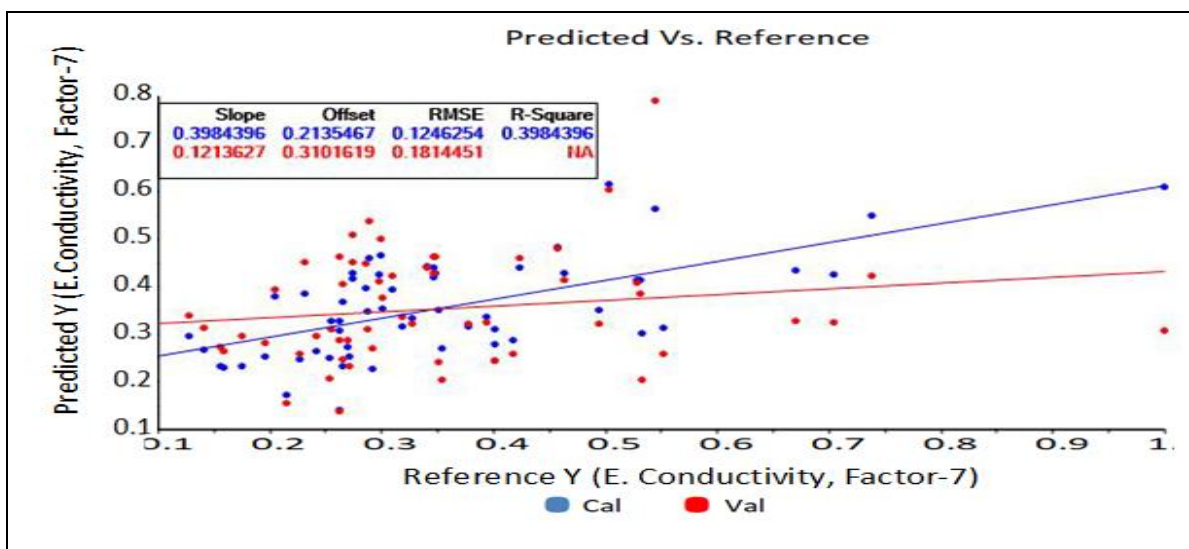

Figure 6.34 Scattered plot between the ref and predicted conductivity value with PLS regression between 700 to 2500 nm range (Raw)

In the Figure 6.34 plot is shown for reference and predicted values with raw data for conductivity, it is clear that the root mean square error for calculated value is 0.1246254 and coefficient of determination is 0.3984396, which shows it is not acceptable and for validation the value for RMSE is high and it is not acceptable for R^2 so it can be concluded that this model is not acceptable.
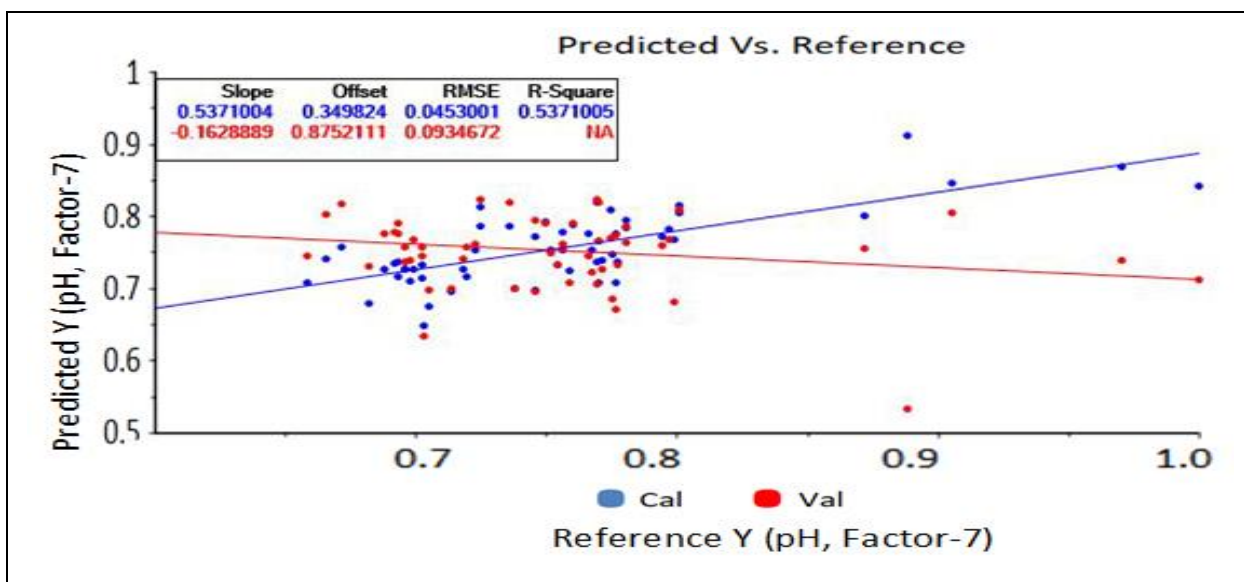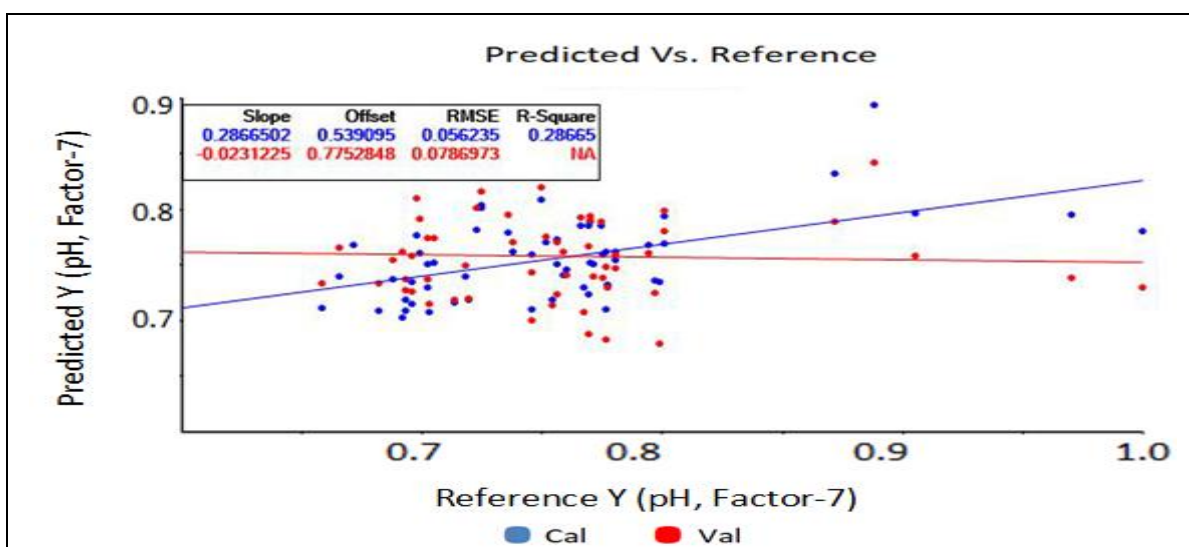

Figure 6.35 Scattered plot between the ref and predicted pH value with PLS regression between 700 to 2500 nm range (Raw)

In the Figure 6.35 plot is shown for reference and predicted values with raw data for pH value, it is clear that the root mean square error for calculated value is 0.056235 and coefficient of determination is 0.28665, which shows it is not acceptable and for validation the value for RMSE is high and it is not acceptable for R^2 so it can be concluded that this model is not acceptable.
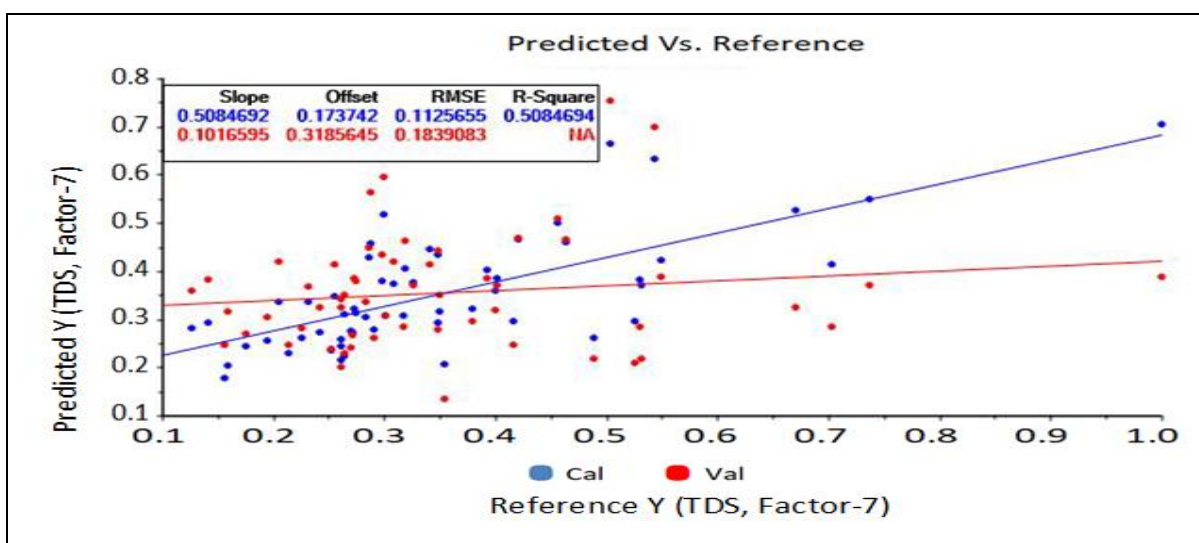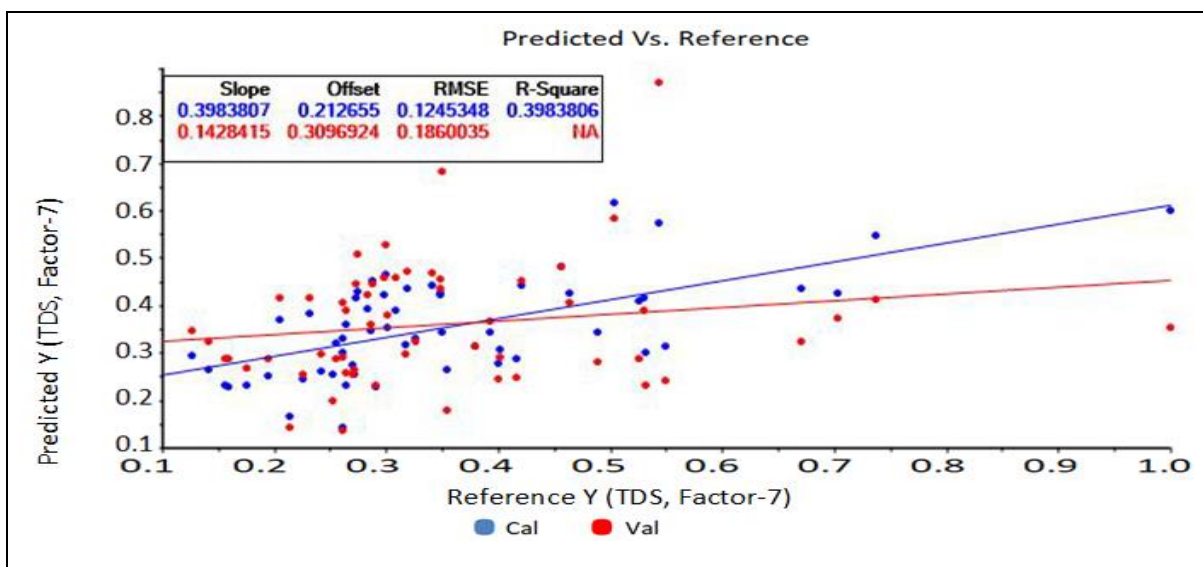

Figure 6.36 Scattered plot between the ref and predicted TDS value with PLS regression between 700 to 2500 nm range (Raw)

In the Figure 6.36 plot is shown for reference and predicted values with raw data for TDS value in the range 700 to 2500 nm, it is clear that the root mean square error for calculated value is 0.1245348 and coefficient of determination is 0.3983806, which shows it is not acceptable and for validation the value for RMSE is high and it is not acceptable for R^2 so it can be concluded that this model is not acceptable.
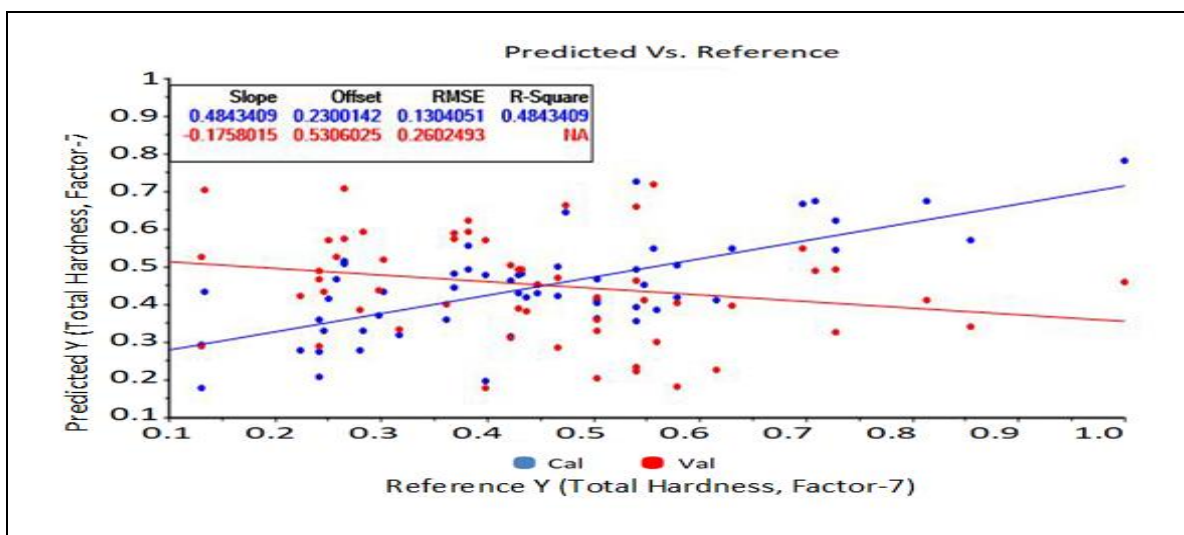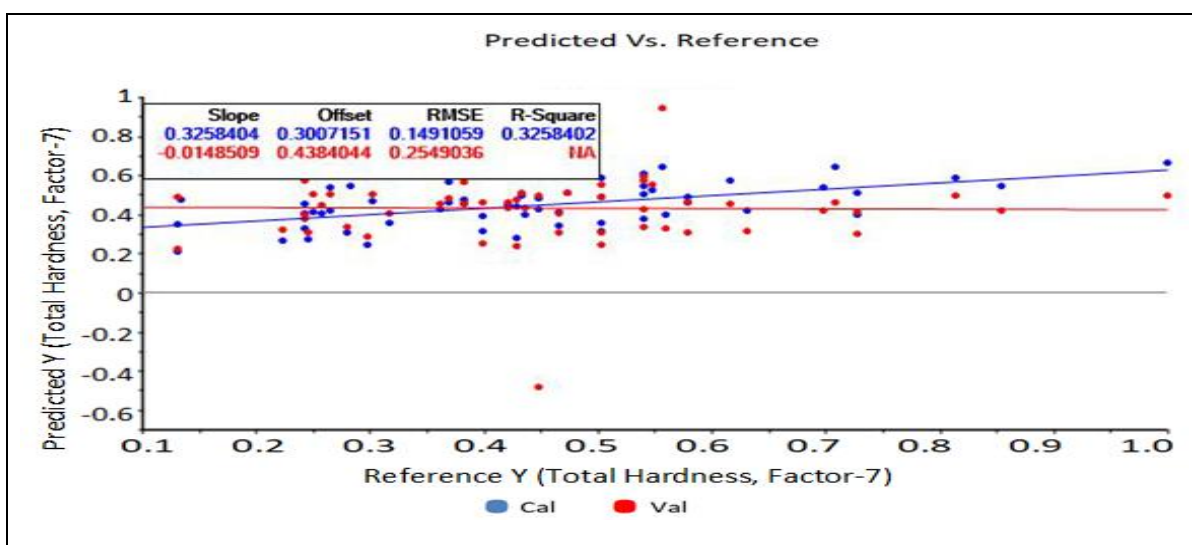

Figure 6.37 Scattered plot between the ref and predicted total hardness value with PLS regression between 700 to 2500 nm range (Raw)

In the Figure 6.37 plot is shown for reference and predicted values with raw data for total hardness, in the range 700 to 2500 nm, it is clear that the root mean square error for calculated value is 0.1491059 and coefficient of determination is 0.3258402, which shows it is not acceptable and for validation the value for RMSE is high and it is not acceptable for R^2 so it can be concluded that this model is not acceptable.


Figure 6.38 Chlorine raw data preprocessing and smoothing in the range 2200-2500 nm

In the Figure 6.38 plot is shown for reference and predicted values for chlorine content, the graph is obtained after preprocessing and smoothing, in the range from 2200-2500 nm, it is clear that the root mean square error for calculated value is 0.1040735 and coefficient of determination is 0.6980579, which shows it is acceptable to some extent and for validation the value for RMSE is 0.1207094 and it is 0.5938134 for R^2 so it can be concluded that this model is not acceptable.
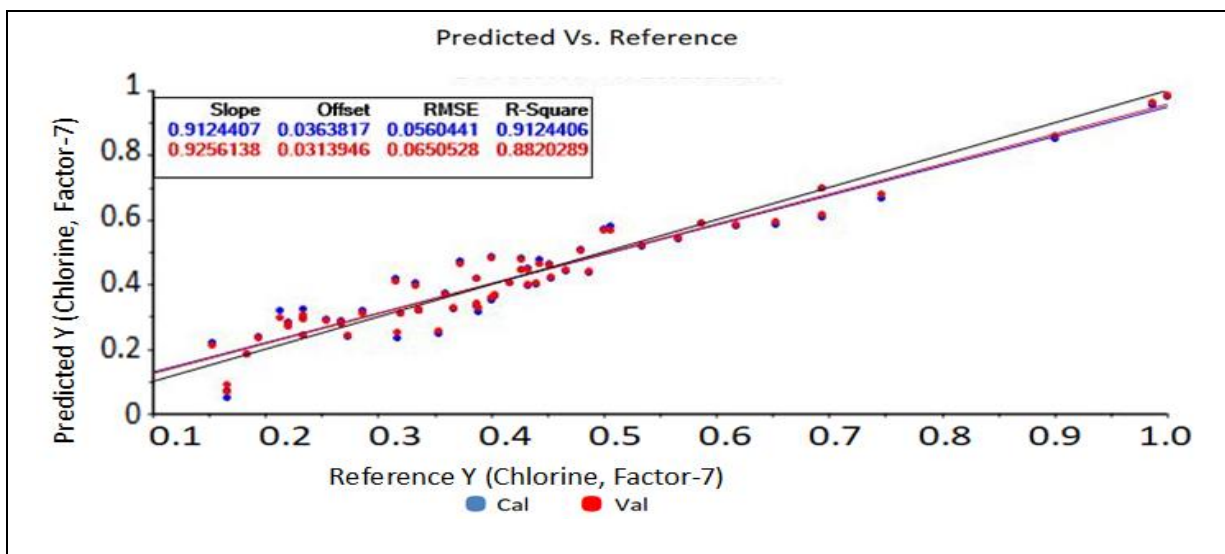
Figure 6.39 Chlorine preprocessing and smoothing plus derivative in the range 1900-2200 nm

In the Figure 6.39 plot is shown for reference and predicted values for chlorine content, the graph is obtained after preprocessing and smoothing plus derivative , in the range from 1900-2200 nm, it is clear that the root mean square error for calculated value is 0.0560441 and coefficient of determination is 0.9124406, which shows it is acceptable and for validated the values for RMSE is 0.0650528 and it is 0.8820289 for R^2 so it can be concluded that this model is acceptable.

In the Figure 6.40 plot is shown for reference and predicted values for chlorine content, the graph is obtained after preprocessing and smoothing plus derivative , in the range from 2200-2500 nm, it is clear that the root mean square error for calculated value is 0.067093 and coefficient of determination is 0.8745131, which shows it is acceptable and for validated the values for RMSE is 0.0797169 and it is 0.8228488 for R^2 so it can be concluded that this model is acceptable.

Figure 6.40 Chlorine preprocessing and smoothing plus derivative in the range 2200-2500 nm

In the Figure 6.41 plot is shown for reference and predicted values with raw data for electrical conductivity, the graph is obtained from raw data, in the range from 2200-2500 nm, it is clear that the root mean square error for calculated value is 0.0917773 and coefficient of determination is 0.6725728, which shows it is not acceptable and for validated the values for RMSE is 0.1081051 and it is 0.5457066 (on lower side) for R^2 so it can be concluded that this model is not acceptable.

Figure 6.41 Electrical conductivity raw data in the range 2200-2500 nm

In the Figure 6.42 plot is shown for reference and predicted values for electrical conductivity, the graph is obtained after preprocessing and smoothing, in the range from 2200-2500 nm, it is clear that the root mean square error for calculated value is 0.0921975 and coefficient of determination is 0.6695678, which shows it is not acceptable and for validation the value for RMSE is is high and it is small for R^2 so it can be concluded that this model is not acceptable.
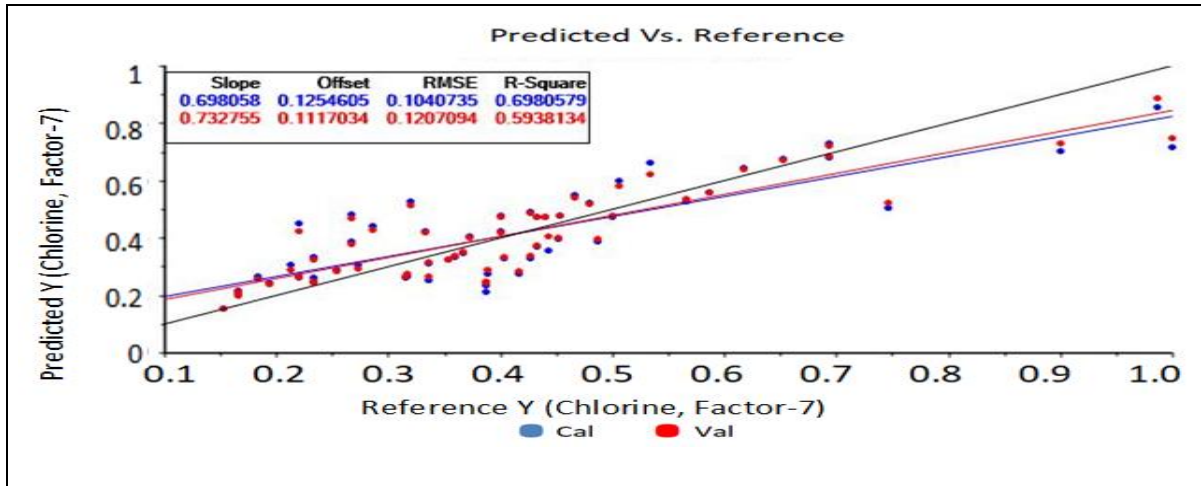

Figure 6.42 Electrical conductivity preprocessing and smoothing in the range 2200-2500 nm

Figure 6.43 Electrical conductivity preprocessing and smoothing plus derivative in the range 1900-2200 nm

In the Figure 6.43 plot is shown for reference and predicted values for electrical conductivity, the graph is obtained after preprocessing and smoothing plus derivative, in the range from 1900-2200 nm, it is clear that the root mean square error for calculated value is 0.0521099 and coefficient of determination is 0.8948265, which shows it is acceptable and for validation the value for RMSE is 0.0605751 and it is 0.8578801 for R^2 so it can be concluded that this model is acceptable.

In the Figure 6.44 plot is shown for reference and predicted values for electrical conductivity, the graph is obtained after preprocessing and smoothing plus derivative, in the range from 2200-2500 nm, it is clear that the root mean square error for calculated value is 0.0423345 and coefficient of determination is 0.9042509, which shows it is not acceptable and for validation the value for RMSE is 0.0502971 and it is 0.8648448 for R^2 so it can be concluded that this model is acceptable.
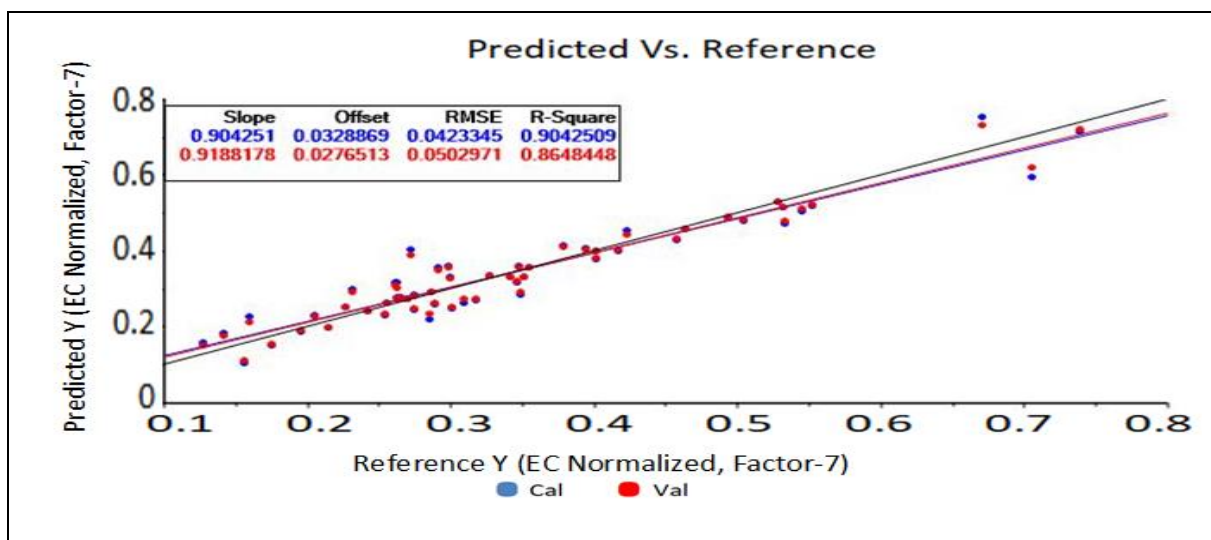
Figure 6.44 Electrical conductivity preprocessing and smoothing plus derivative in the range 2200-2500 nm
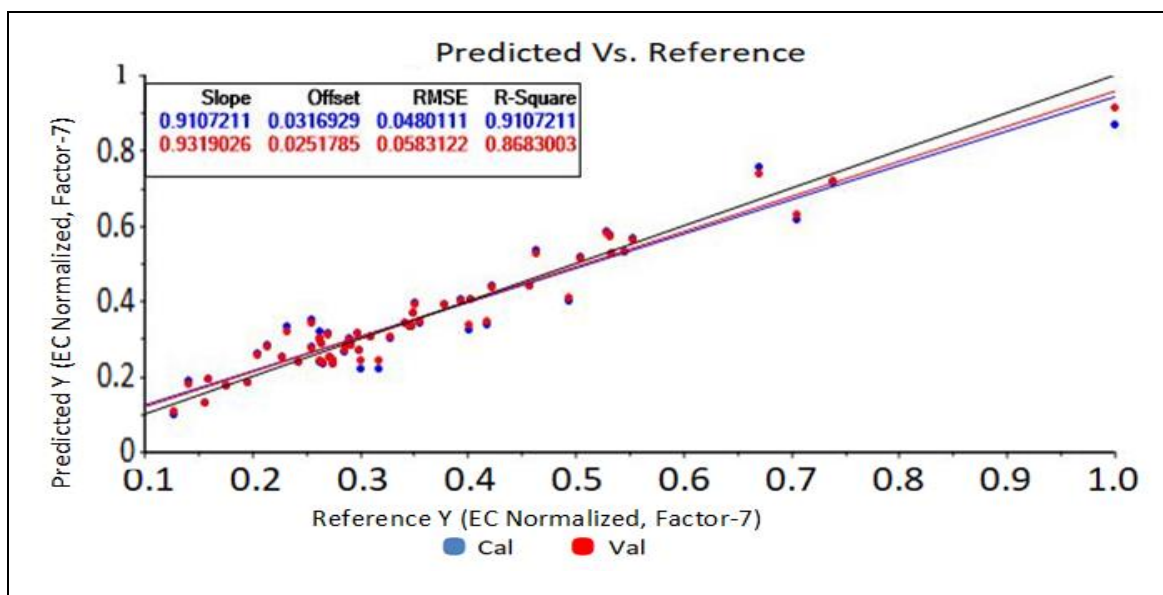
Figure 6.45 Electrical conductivity preprocessing and smoothing plus derivative in the complete range of NIR

In the Figure 6.45 plot is shown for reference and predicted values for electrical conductivity, the graph is obtained after preprocessing and smoothing plus derivative, in the complete range of NIR, it is clear that the root mean square error for calculated value is 0.0480111 and coefficient of determination is 0.9107211, which shows it is acceptable and for validation the value for RMSE is 0.0583122 and it is 0.8683003 for R^2 so it can be concluded that this model is acceptable.
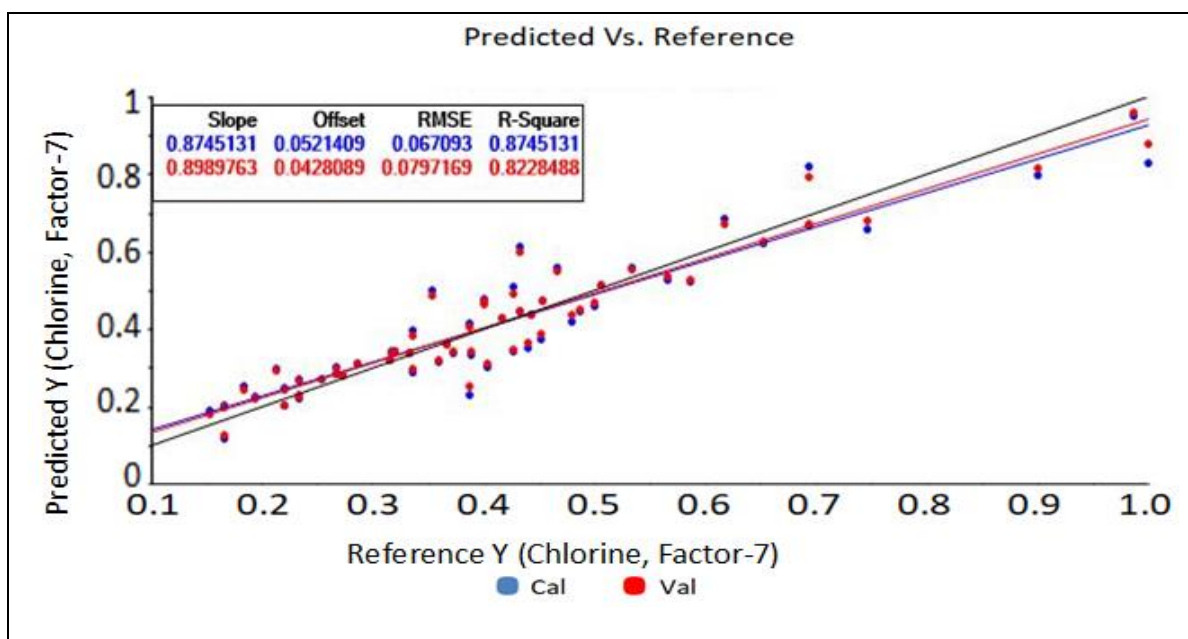
In the Figure 6.46 plot is shown for reference and predicted values for TDS, the graph is obtained form raw data, in the range from 1900-2200 nm, it is clear that the root mean square error for calculated value is 0.0986702 and coefficient of determination is 0.6210923, which shows it is not acceptable and for validation the value for RMSE is high and it is small for R^2 so it can be concluded that this model is not acceptable.

Figure 6.46 Total Dissolved Solids raw data   in the range 1900-2200 nm
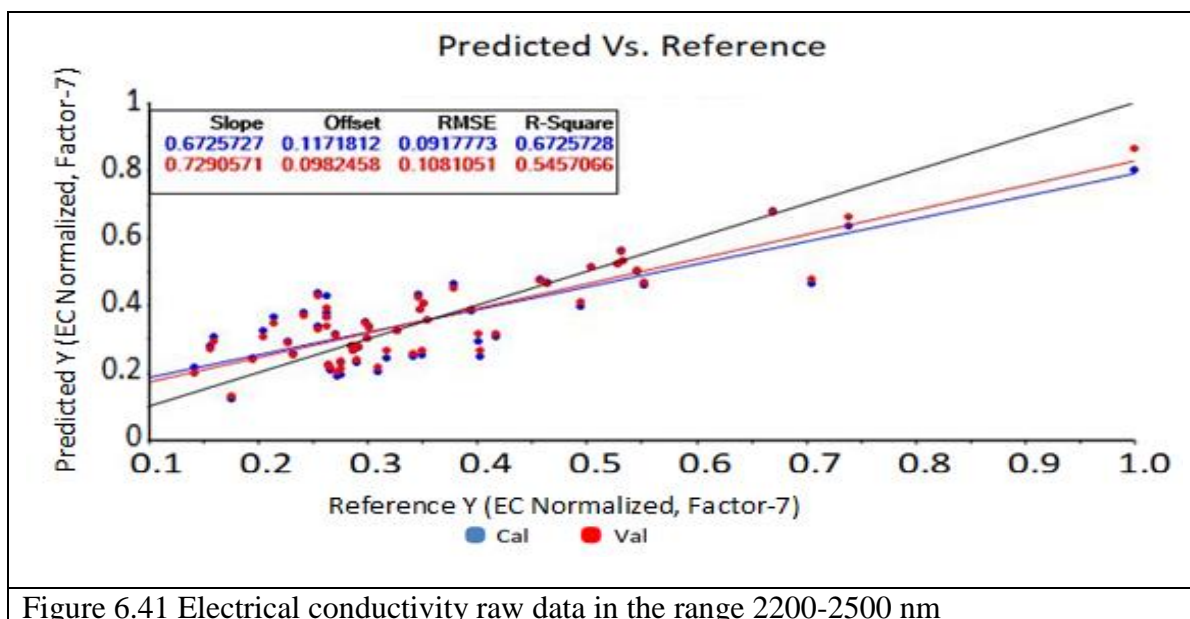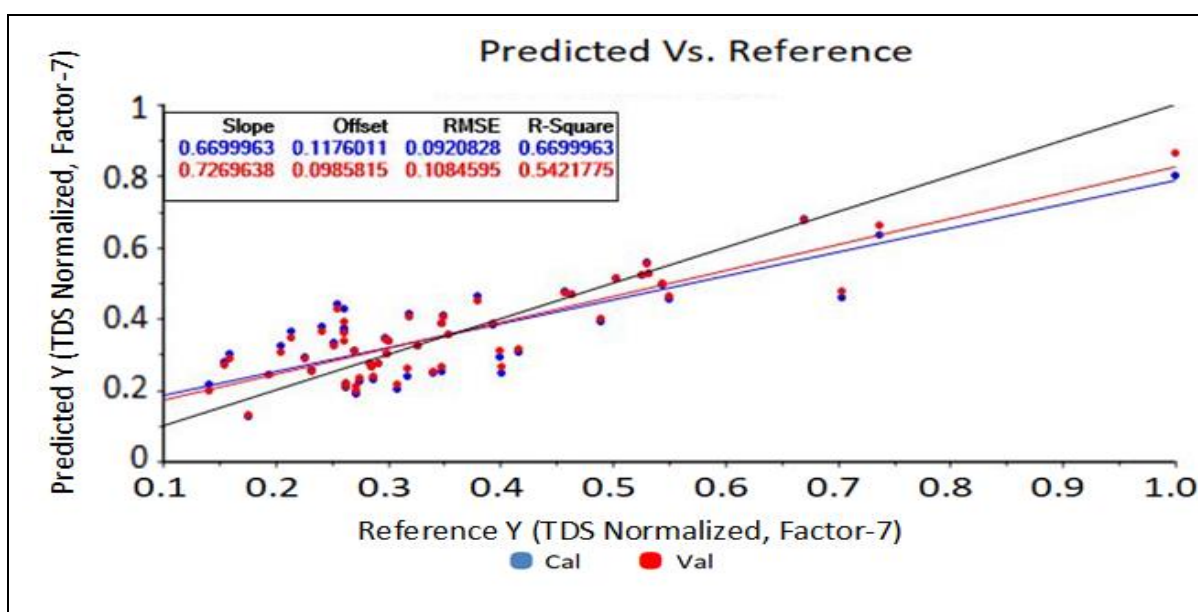

Figure 6.47 Total Dissolved Solids raw data   in the range 2200-2500 nm

In the Figure 6.47 plot is shown for reference and predicted values for TDS, the graph is obtained from the raw data, in the range from 2200-2500 nm, it is clear that the root mean square error for calculated value is 0.0920828 and coefficient of determination is 0.6699963, which shows it is not acceptable and for validation the value for RMSE is high and is small for R^2 so it can be concluded that this model is not acceptable.
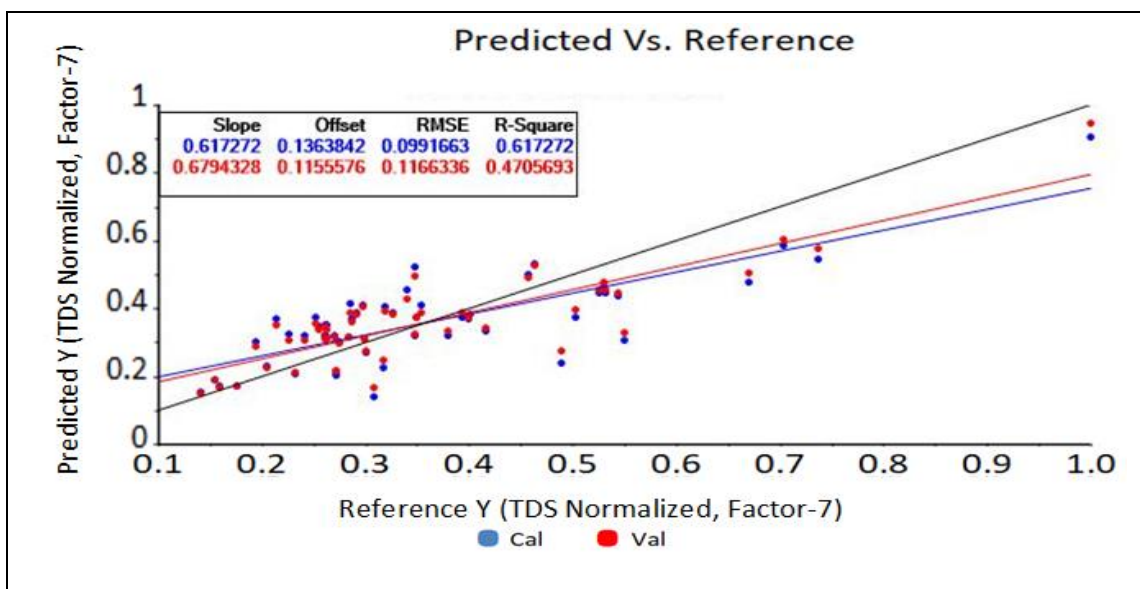
Figure 6.48 Total Dissolved Solids preprocessing and smoothing in the range 1900-2200 nm
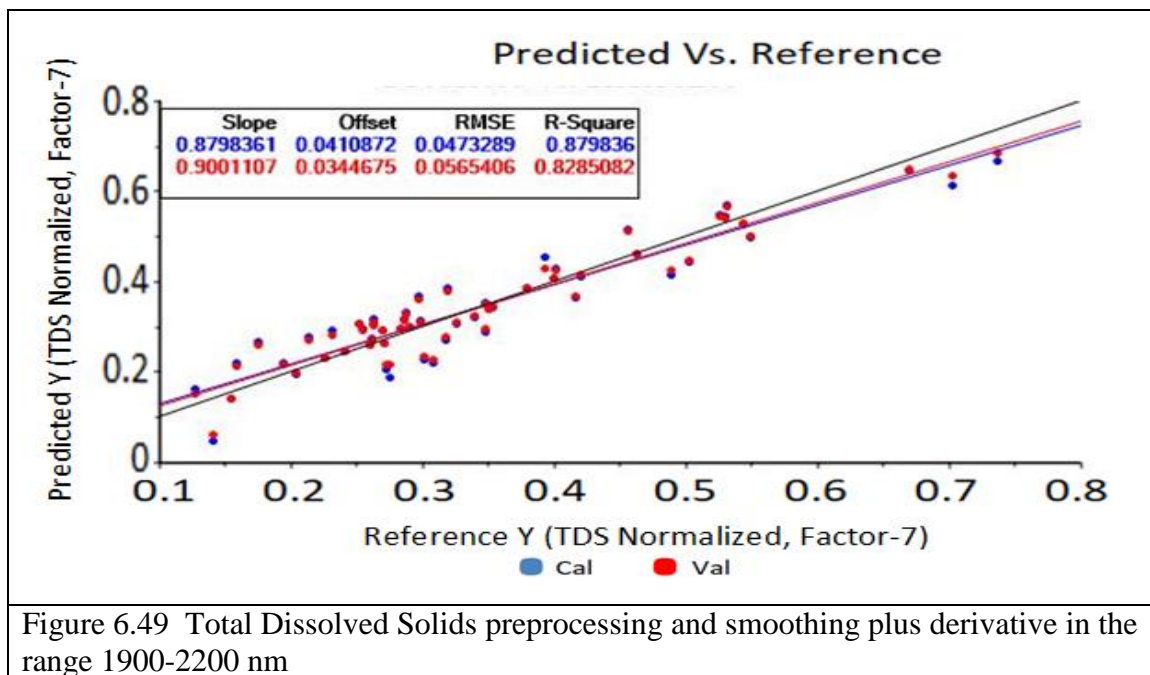
Figure 6.49 Total Dissolved Solids preprocessing and smoothing plus derivative in the range 1900-2200 nm

In the Figure 6.48 plot is shown for reference and predicted values for TDS, the graph is obtained preprocessing and smoothing of the data, in the range from 1900-2200 nm, it is clear that the root mean square error for calculated value is 0.0991663 and coefficient of determination is 0.617272, which shows it is not acceptable and for validation the value for RMSE is high and is small for R^2 so it can be concluded that this model is not acceptable.
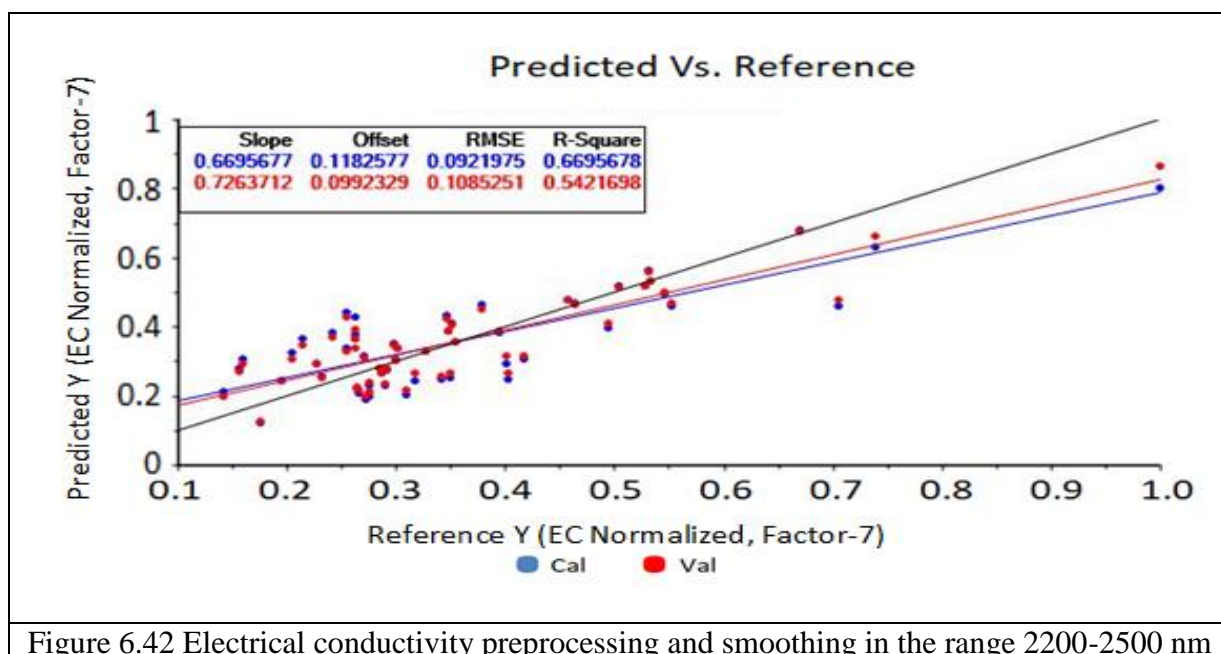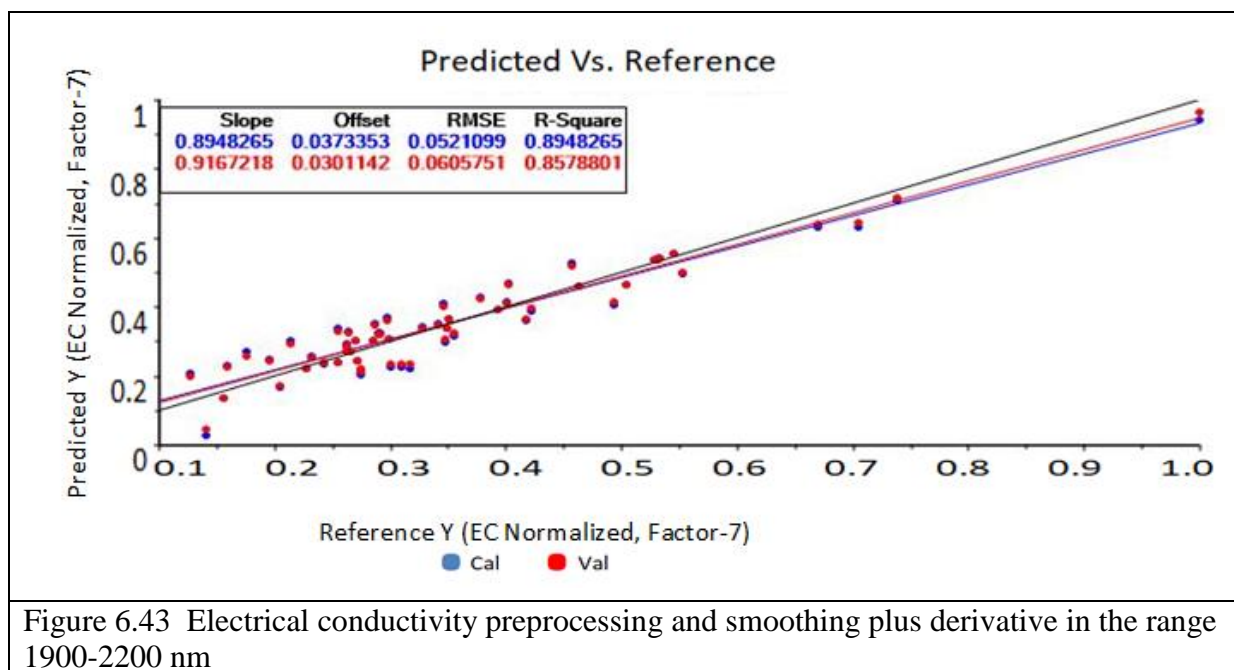
In the Figure 6.49 plot is shown for reference and predicted values for TDS, the graph is obtained preprocessing and smoothing plus derivative of the data, in the range from 1900-2200 nm, it is clear that the root mean square error for calculated value is 0.0473289 and coefficient of determination is 0.879836, which shows it is acceptable and for validation the value for RMSE is 0.0565406 and it is 0.8285082 for R^2 so it can be concluded that this model is acceptable.

Figure 6.50 Total Dissolved Solids preprocessing and smoothing plus derivative in the range 2200-2500 nm

In the Figure 6.50 plot is shown for reference and predicted values for TDS, the graph is obtained preprocessing and smoothing plus derivative of the data, in the range from 2200-2500nm, it is clear that the root mean square error for calculated value is 0.0421802 and coefficient of determination is 0.904558, which shows it is acceptable and for validation the value for RMSE is 0.0501537 and it is 0.8650639 for R^2 so it can be concluded that this model is acceptable.

Figure 6.51 plot is shown for reference and predicted values for TDS, the graph is obtained preprocessing and smoothing plus derivative of the data, in the complete range of NIR, it is clear that the root mean square error for calculated value is 0.0478548 and coefficient of determination is 0.9111636, which shows it is acceptable and for validation the value for RMSE is 0.0582449 and it is 0.8684 for R^2 so it can be concluded that this model is acceptable.

Figure 6.51 Total Dissolved Solids preprocessing and smoothing plus derivative in the range NIR complete

Figure 6.52 Total hardness raw data   in the range 1900-2200 nm
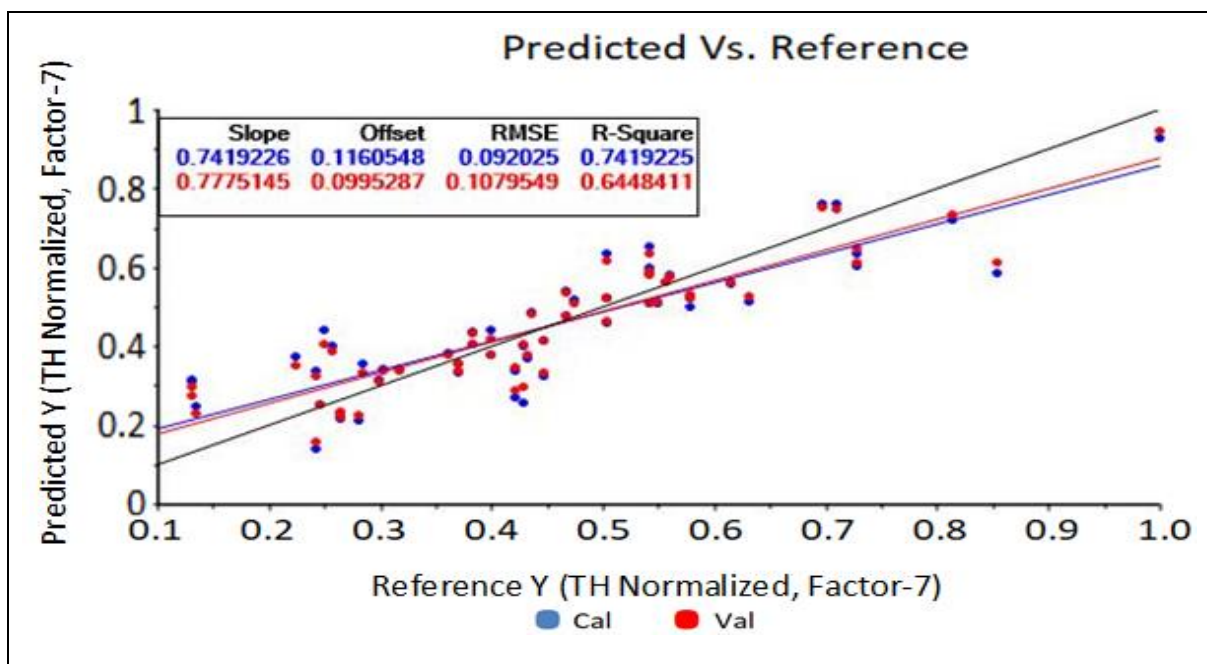
Figure 6.53 Total hardness preprocessing and smoothing in the range 1900-2200 nm

In the Figure 6.52 plot is shown for reference and predicted values for total hardness, the graph is obtained from the raw data, in the range from 1900-2200 nm, it is clear that the root mean square error for calculated value is 0.0914342 and coefficient of determination is 0.7452261, which shows it is acceptable and for validation the value for RMSE is low and it is near 0.7 for R^2 so it can be concluded that this model is acceptable to some extent.

In the Figure 6.53 plot is shown for reference and predicted values for total hardness, the graph is obtained after preprocessing and smoothing of the data, in the range from 1900-2200 nm, it is clear that the root mean square error for calculated value is 0.092025 and coefficient of determination is 0.7419225, which shows it is acceptable and for validation the value for RMSE is high and it is near 0.7 for R^2 so it can be concluded that this model is acceptable.
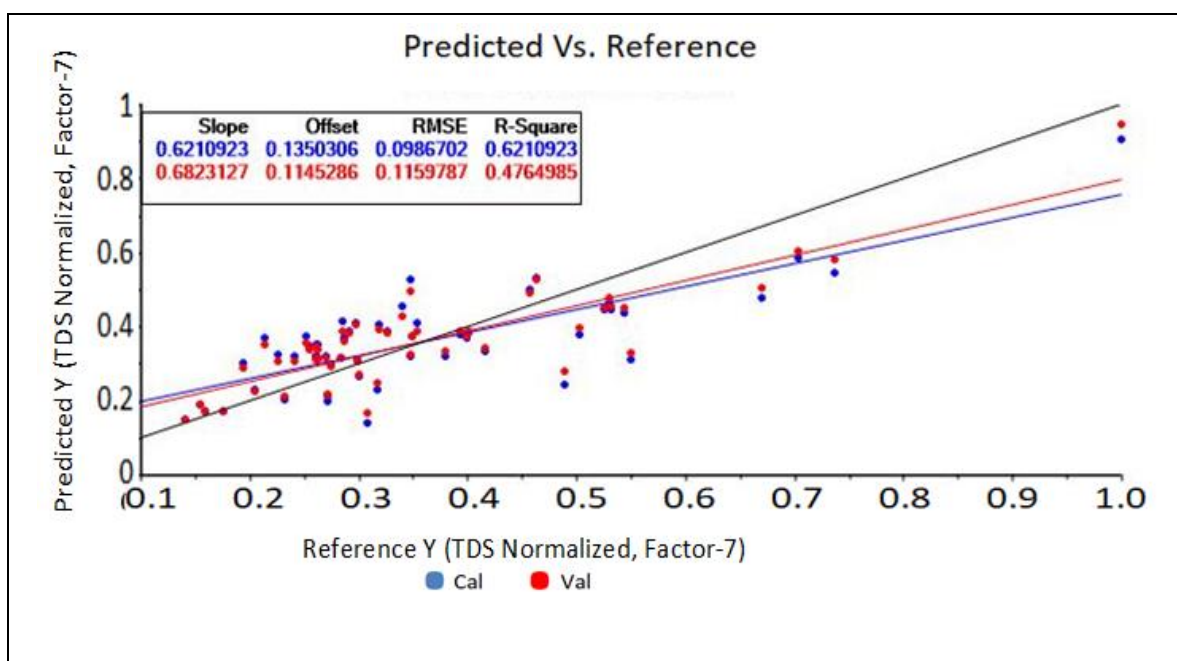
In the Figure 6.54 plot is shown for reference and predicted values for total hardness, the graph is obtained after preprocessing and smoothing plus derivative in the range 1300-1600nm, it is clear that the root mean square error for calculated value is 0.0847258 and coefficient of determination is 0.7858918, which shows it is acceptable and for validation the value for RMSE is 0.1008808 and it is near 0.7 for R^2 so it can be concluded that this model is acceptable to some extent.
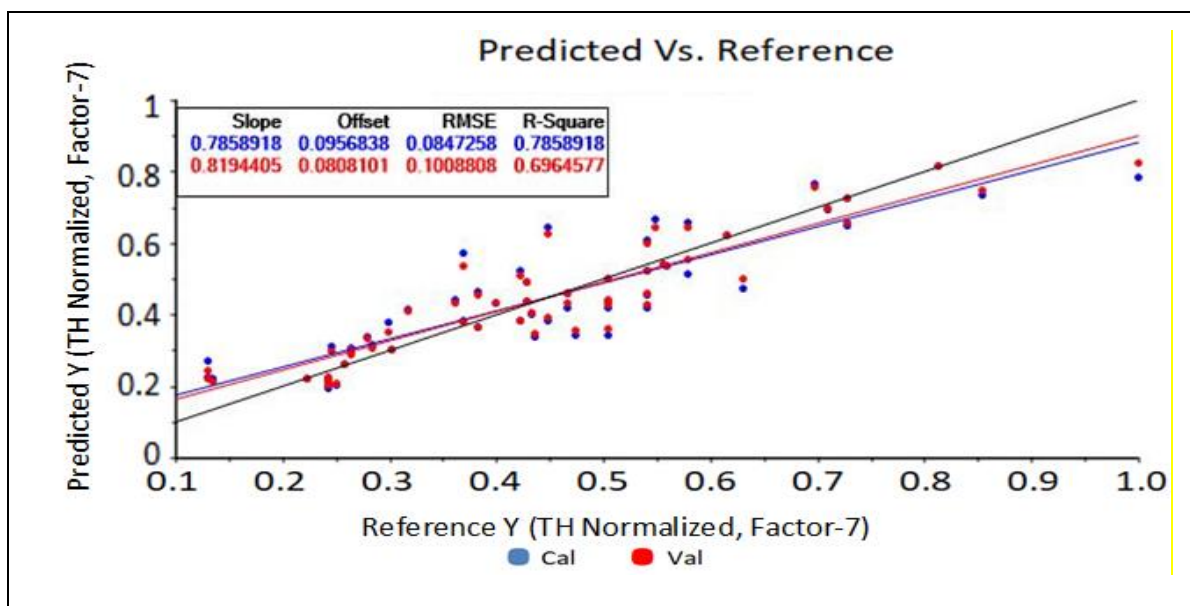
Figure 6.54 Total hardness preprocessing and smoothing plus derivative in the range 1300-1600 nm
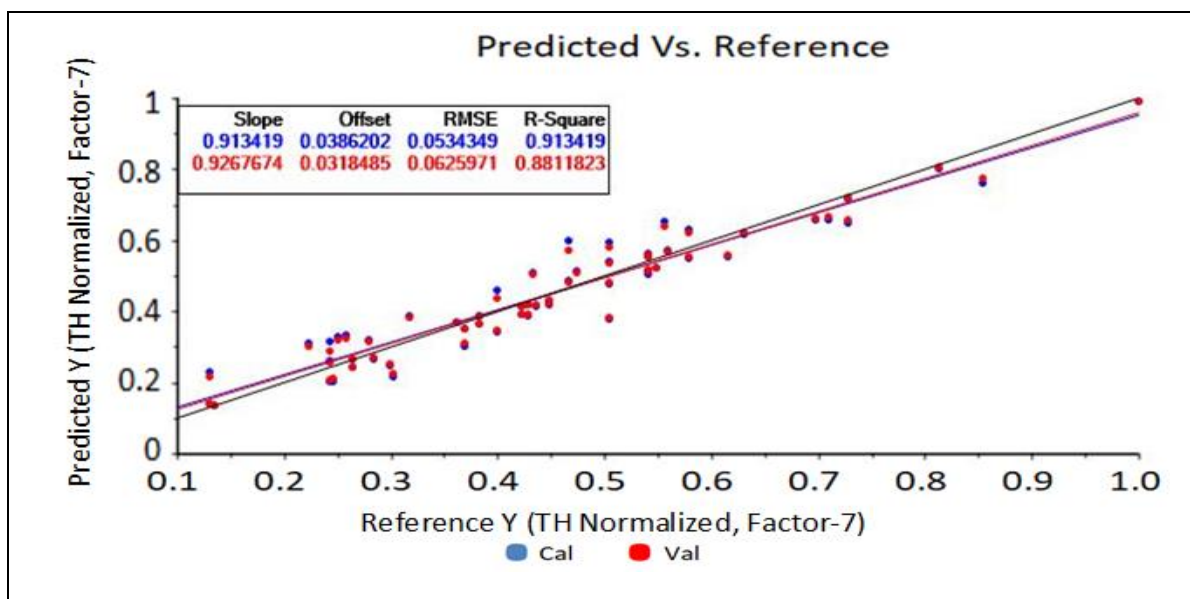
Figure 6.55 Total hardness preprocessing and smoothing plus derivative in the range 1900-2200 nm

In the Figure 6.55 plot is shown for reference and predicted values for total hardness, the graph is obtained after preprocessing and smoothing plus derivative in the range 1900-2200 nm, it is clear that the root mean square error for calculated value is 0.0534349 and coefficient of determination is 0.913419, which shows it is acceptable and for validation the value for RMSE is 0.06225971 and it is 0.8811823 for R^2 so it can be concluded that this model is acceptable.
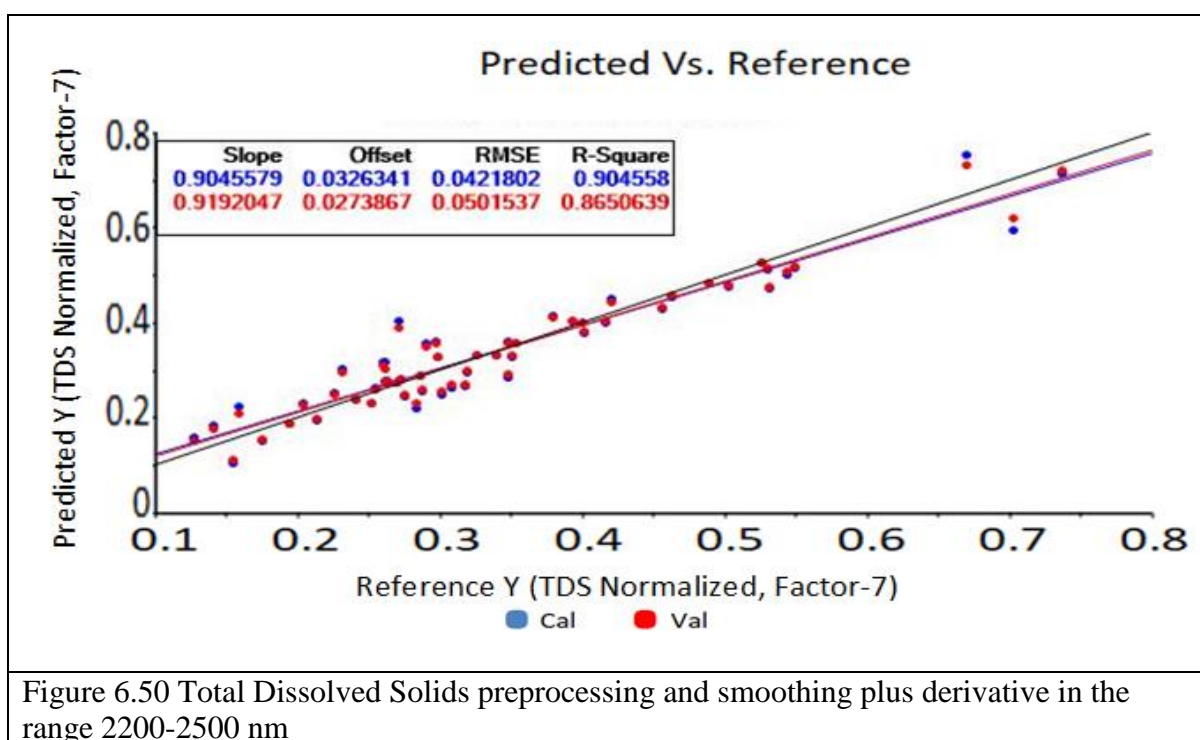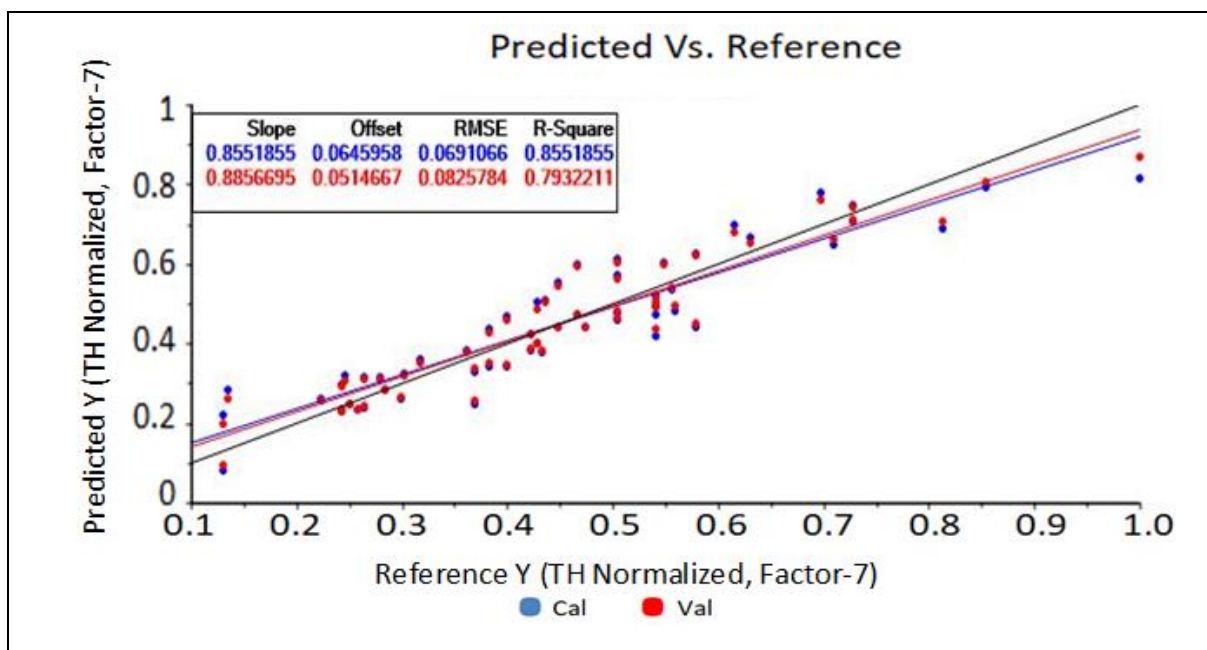
Figure 6.56 Total hardness preprocessing and smoothing plus derivative in the range 2200-2500 nm

In the Figure 6.56 plot is shown for reference and predicted values for total hardness, the graph is obtained after preprocessing and smoothing plus derivative in the range 2200-2500 nm, it is clear that the root mean square error for calculated value is 0.0691066 and coefficient of determination is 0.8551855, which shows it is acceptable and for validation the value for RMSE is 0.0825784 and it is 0.7932211 for R^2 so it can be concluded that this model is acceptable to some extent.
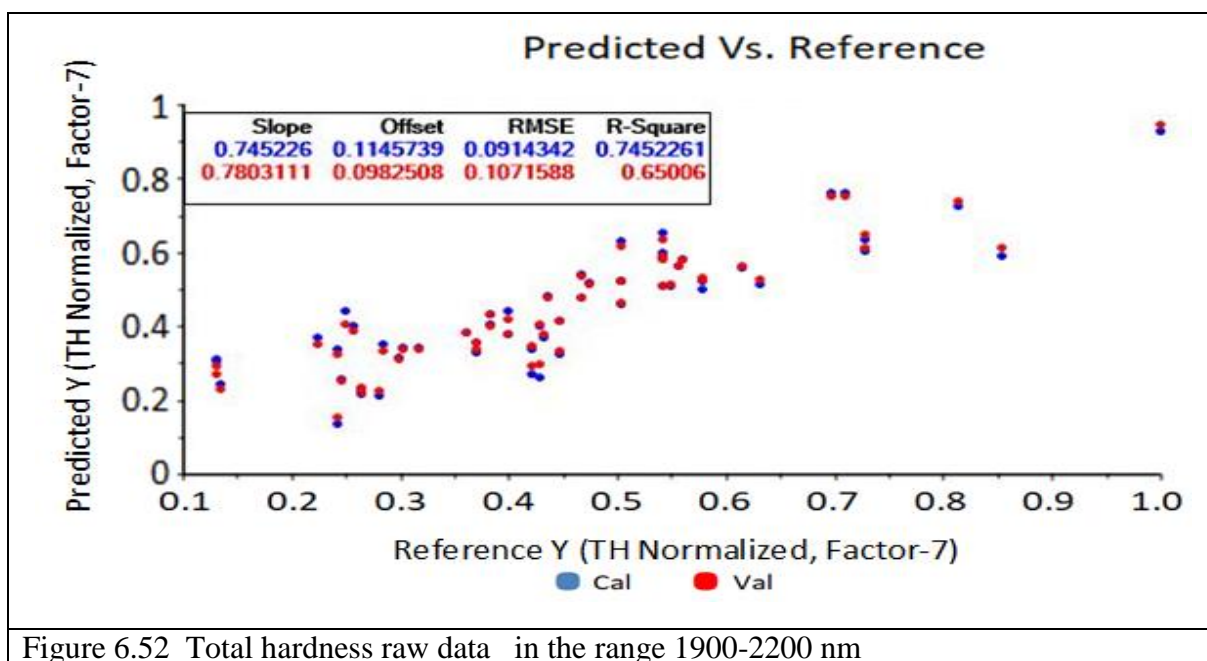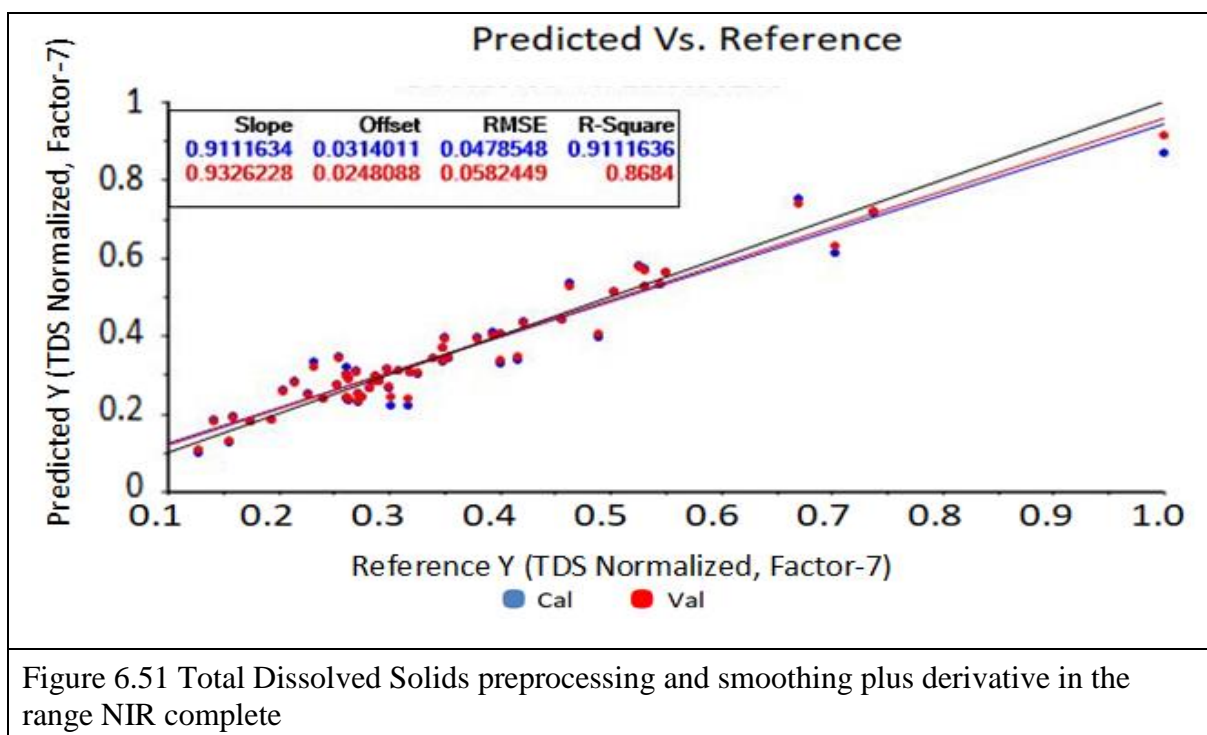
In the Figure 6.57 plot is shown for reference and predicted values for total hardness, the graph is obtained after preprocessing and smoothing plus derivative in the complete range of NIR, it is clear that the root mean square error for calculated value is 0.0518287 and coefficient of determination is 0.9195174, which shows it is acceptable and for validation the value for RMSE is 0.061771 and it is 0.8856776 for R^2 so it can be concluded that this model is acceptable.
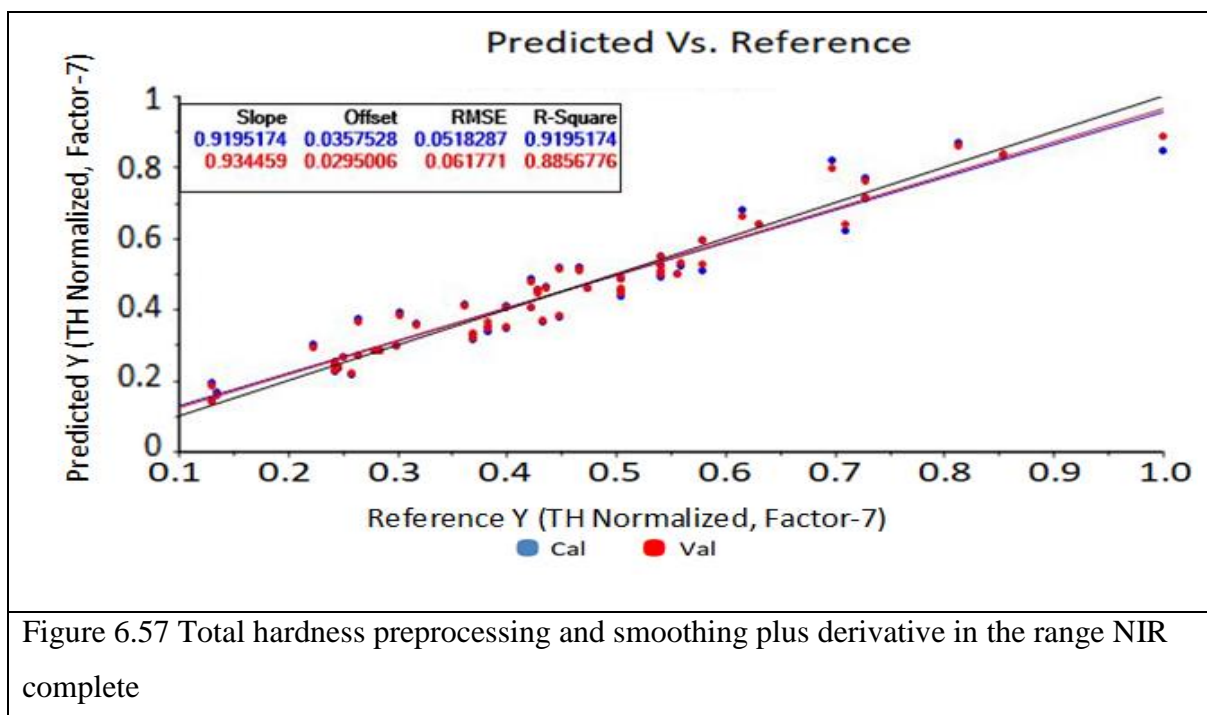
Figure 6.57 Total hardness preprocessing and smoothing plus derivative in the range NIR complete

NIR data analysis had been performed using unscrambler $\times$ 10.5 software. Preprocessing such as normalization, double derivative, baseline and smoothening has been applied on the NIR data. Multivariate analysis (PLS) was used to develop the rapid and non-destructive statistical regression model. Results show that for wavelength range 700-2500 nm, R^2 is near or greater than 0.7 for the parameters (pH, TDS, EC, turbidity, total hardness, chlorine). Here R^2_C is coefficient of determination for calibration, $RMSE_C$ is root mean square error for calibration and R^2_V is coefficient of determination for validation, $RMSE_V$ is root mean square error for validation.

From some of the plots shown in figures 6.24, 6.25, 6.27 to 6.38, 6.41 to 6.42 and 6.46 to 6.48, it is clear that the values of R^2 are very low and the values of RMSE are high so such models are not suitable and are rejected.

6.5 INTERVAL PARTIAL LEAST SQURE REGRESSION

Table 6.1 Electrical conductivity Smoothing Derivative

WAVELENGTH RANGE	$RMSE_C$	R^2_C	$RMSE_V$	R^2_V
400 - 2500	0.1590264	0.0268089	0.1708374	NA
700 - 1000	0.1603976	0.0099539	0.1689193	NA
1000 - 1300	0.1610709	0.0016241	0.1670146	NA
1300 - 1600	0.159863	0.016543	0.1671623	NA
1600 - 1900	0.1589516	0.0277241	0.168249	NA
1900 - 2200	0.1550671	0.0746652	0.1765049	NA
2200 - 2500	0.1360368	0.2878481	0.1959717	NA

Table 6.2 PLS regression Model parameters for Electrical conductivity by applying cross validation with preprocessing (smoothing + Derivative)

WAVELENGTH RANGE	$RMSE_C$	R^2_C	$RMSE_V$	R^2_V
400 - 2500	0.1589954	0.02718885	0.170882	NA
700 - 1000	0.1603971	0.0099597	0.1719436	NA
1000 - 1300	0.1610887	0.0014035	0.1650219	NA
1300 - 1600	0.1598626	0.0165473	0.16646	NA
1600 - 1900	0.1589465	0.0277862	0.1680432	NA
1900 - 2200	0.1549896	0.0755892	0.1763851	NA
2200 - 2500	0.1356196	0.2922094	0.1992115	NA

Table 6.3 PLS regression Model parameters for Electrical conductivity by applying cross validation with preprocessing (smoothing)

WAVELENGTH RANGE	RMSE$_C$	R^2_C	RMSE$_V$	R^2_V
400 - 2500	0.1606392	0.0069697	0.1746471	NA
700 - 1000	0.1611129	0.001104	0.1662181	NA
1000 - 1300	0.1610324	0.0021017	0.1682357	NA
1300 - 1600	0.1607686	0.005368	0.1755944	NA
1600 - 1900	0.160731	0.0058336	0.1699645	NA
1900 - 2200	0.1600304	0.0144813	0.1854234	NA
2200 - 2500	0.1602868	0.011321	0.185764	NA

Table 6.4 PLS regression Model parameters for Electrical conductivity by applying cross validation

WAVELENGTH RANGE	RMSE$_C$	R^2_C	RMSE$_V$	R^2_V
400 - 2500	0.1606391	0.0069702	0.1762505	NA
700 - 1000	0.1611129	0.0011041	0.1687277	NA
1000 - 1300	0.1610322	0.0021043	0.1666885	NA
1300 - 1600	0.1607686	0.0053684	0.1743991	NA
1600 - 1900	0.160731	0.0058339	0.1690509	NA
1900 - 2200	0.1600303	0.0144825	0.1860168	NA
2200 - 2500	0.1602863	0.0113272	0.185204	NA

Table 6.5 PLS regression Model parameters for total dissolved solids by applying cross validation with preprocessing (smoothing)

WAVELENGTH RANGE	$RMSE_C$	R^2_C	$RMSE_V$	R^2_V
400 - 2500	0.1604994	0.0068961	0.1723414	NA
700 - 1000	0.160957	0.0012244	0.1659579	NA
1000 - 1300	0.1608774	0.0022121	0.1661279	NA
1300 - 1600	0.1606305	0.0052725	0.1740312	NA
1600 - 1900	0.1606157	0.0054564	0.1691956	NA
1900 - 2200	0.374624	0.0141319	0.1857771	NA
2200 - 2500	0.1602126	0.0104415	0.1856937	NA

Table 6.6 PLS regression Model parameters for Turbidity by applying cross validation with preprocessing (smoothing + Derivative)

WAVELENGTH RANGE	$RMSE_C$	R^2_C	$RMSE_V$	R^2_V
400 - 2500	0.2201467	0.0287655	0.2247566	0.0117006
700 - 1000	0.2203293	0.0271533	0.2268407	0.0163869
1000 - 1300	0.2220728	0.0116966	0.2277792	NA
1300 - 1600	0.2207628	0.023322	0.2259308	0.0170445
1600 - 1900	0.2211453	0.0199343	0.2268212	0.0165561
1900 - 2200	0.2215921	0.0159701	0.2279843	NA
2200 - 2500	0.2100488	0.1158215	0.2319486	NA

Table 6.7 PLS regression Model parameters for Turbidity by applying cross validation with preprocessing (smoothing)

WAVELENGTH RANGE	RMSE$_C$	R^2$_C$	RMSE$_V$	R^2$_V$
400 - 2500	0.2210183	0.0210593	0.2268964	0.0040196
700 - 1000	0.2224962	0.0079243	0.2337049	NA
1000 - 1300	0.2211371	0.0200073	0.2290306	NA
1300 - 1600	0.2210885	0.0204375	0.2258975	0.0090879
1600 - 1900	0.2208183	0.0228303	0.2258763	0.0120372
1900 - 2200	0.2222291	0.0103047	0.2286738	NA
2200 - 2500	0.222221	0.0103762	0.2294818	NA

Table 6.8 PLS regression Model parameters for Turbidity by applying cross validation

WAVELENGTH RANGE	RMSE$_C$	R^2$_C$	RMSE$_V$	R^2$_V$
400 - 2500	0.2210183	0.0210596	0.2266933	0.0058024
700 - 1000	0.2224962	0.0079243	0.2300743	NA
1000 - 1300	0.2211365	0.0200127	0.2276037	NA
1300 - 1600	0.2210885	0.0204375	0.2243803	0.0112924
1600 - 1900	0.2208183	0.0228303	0.2248712	0.0171557
1900 - 2200	0.2222291	0.0103047	0.2285692	NA
2200 - 2500	0.222221	0.0103762	0.2288961	NA

Table 6.9 Total dissolved solids raw data

WAVELENGTH RANGE	RMSE$_C$	R^2_C	RMSE$_V$	R^2_V
400 - 2500	0.1604993	0.0068967	0.1736819	NA
700 - 1000	0.160957	0.0012244	0.1653857	NA
1000 - 1300	0.1608772	0.0022146	0.1678511	NA
1300 - 1600	0.1606305	0.0052732	0.1749782	NA
1600 - 1900	0.1605941	0.0057242	0.1714033	NA
1900 - 2200	0.1599135	0.0141328	0.18274	NA
2200 - 2500	0.1602122	0.0104473	0.182195	NA

Table 6.10 Total dissolved solids smoothing first derivative

WAVELENGTH RANGE	RMSE$_C$	R^2_C	RMSE$_V$	R^2_V
400 - 2500	0.1589042	0.0265391	0.1715839	NA
700 - 1000	0.1602717	0.0097111	0.1682069	NA
1000 - 1300	0.160927	0.001597	0.1658035	NA
1300 - 1600	0.1597445	0.0162157	0.1669793	NA
1600 - 1900	0.1588224	0.0275412	0.1679805	NA
1900 - 2200	0.1550404	0.073303	0.1757682	NA
2200 - 2500	0.136156	0.2853038	0.196552	NA

Table 6.11 Electrical conductivity raw data leverage

WAVELENGTH RANGE	RMSE$_C$	R^2_C	RMSE$_V$	R^2_V
400 - 2500	0.1606391	0.0069702	0.1694901	NA
700 - 1000	0.1611129	0.0011041	0.1664603	NA
1000 - 1300	0.1610322	0.0021044	0.1678368	NA
1300 - 1600	0.1607686	0.0053685	0.1704683	NA
1600 - 1900	0.160731	0.0058337	0.1700266	NA
1900 - 2200	0.1600304	0.0144823	0.164005	NA
2200 - 2500	0.118563	0.4590491	0.1377639	0.2696502
3OR	0.0956252	0.6458306	0.1136608	0.4996348
2OR	0.0917773	0.6725728	0.1081051	0.5457066

Table 6.12 Electrical conductivity smoothing polynomial

WAVELENGTH RANGE	RMSE$_C$	R^2_C	RMSE$_V$	R^2_V
400 - 2500	0.1606392	0.0069695	0.1694904	NA
700 - 1000	0.1611129	0.001104	0.1664603	NA
1000 - 1300	0.1610324	0.0021016	0.1678358	NA
1300 - 1600	0.1607687	0.0053678	0.170469	NA
1600 - 1900	0.160731	0.0058334	0.1700269	NA
1900 - 2200	0.1600304	0.0144812	0.1640054	NA
2200 - 2500	0.1187993	0.4568902	0.1382666	0.2643106
3OR	0.0961339	0.6420525	0.1142486	0.4944457
2OR	0.0921975	0.6695678	0.1085251	0.5421698
1OR	0.0886035	0.6776691	0.1069985	0.5299379
3OR	0.0739637	0.6844711	0.0877574	0.558098

Table 6.13 Electrical conductivity smoothing derivative

WAVELENGTH RANGE	RMSE$_C$	R^2$_C$	RMSE$_V$	R^2$_V$
400 - 2500	0.1074405	0.5557827	0.1266894	0.3823529
3OR	0.0480111	0.9107211	0.0583122	0.8683003
4OR	0.0413472	0.938186	0.0503793	0.9082307
3OR	0.0286138	0.9715534	0.0345334	0.9585658
700 - 1000	0.1603976	0.0099538	0.1695196	NA
1000 - 1300	0.1010709	0.001624	0.1656494	NA
1300 - 1600	0.159863	0.0165429	0.1649716	NA
1600 - 1900	0.1589516	0.0277241	0.1627983	NA
1900 - 2200	0.069259	0.815408	0.0903449	0.6859006
3OR	0.0521099	0.8948265	0.0605751	0.8578801
1OR	0.0472438	0.8807563	0.056485	0.8295441
2200 - 2500	0.068458	0.8196534	0.806135	0.7499219
3OR	0.0577867	0.8706631	0.07006	0.8098895
1OR	0.0423345	0.9042509	0.0502971	0.8648448
2OR	0.0377487	0.9174125	0.0442422	0.8805556
4OR	0.0330497	0.9231973	0.0393743	0.89099

Table 6.14 Total hardness raw data

WAVELENGTH RANGE			$RMSE_C$	R^2_C	$RMSE_V$	R^2_V
400	-	2500	0.1701019	0.1450456	0.175385	0.091113
700	-	1000	0.1614965	0.229361	0.1782856	0.060801
1000	-	1300	0.1714788	0.1311482	0.1783276	0.060359
1300	-	1600	0.1710856	0.1351277	0.1763505	0.081079
1600	-	1900	0.1715113	0.1308184	0.1782426	0.061254
1900	-	2200	0.1522303	0.3152581	0.1692476	0.153611
		2OR	0.1005632	0.7000948	0.1170983	0.593363
		2OR	0.0914342	0.7452261	0.1071588	0.65006
2200	-	2500	0.1823174	0.0178422	0.2010119	NA

Table 6.15 Total hardness smoothing

WAVELENGTH RANGE			$RMSE_C$	R^2_C	$RMSE_V$	R^2_V
400	-	2500	0.1701045	0.1450187	0.1753877	0.091086
700	-	1000	0.1615106	0.229226	0.1783001	0.060649
1000	-	1300	0.1714871	0.1310643	0.1783399	0.06023
1300	-	1600	0.1710854	0.1351305	0.1763501	0.081083
1600	-	1900	0.1715114	0.1308177	0.1782443	0.061237
1900	-	2200	0.152439	0.3133787	0.1690531	0.155556
		2OR	0.1010965	0.6969054	0.117527	0.59033
		2OR	0.092025	0.7419225	0.1079549	0.644841
		2OR	0.0925438	0.7428085	0.108349	0.647458
2200	-	2500	0.1823176	0.0178408	0.2010117	NA

Table 6.16 Total hardness smoothing derivative

WAVELENGTH RANGE			RMSE$_C$	R^2_C	RMSE$_V$	R^2_V
400	-	2500	0.1147122	0.611184	0.1299643	0.500916
		4OR	0.0518287	0.9195174	0.061771	0.885078
		2OR	0.0509891	0.9230821	0.0624115	0.88476
		3OR	0.0465573	0.9379413	0.0567039	0.907944
700	-	1000	0.1031368	0.2136272	0.1719958	0.125901
		3OR	0.166379	0.160598	0.1777455	0.041989
1000	-	1300	0.1345985	0.46469	0.1563689	0.27752
		4OR	0.1407946	0.4060722	0.1638585	0.195549
1300	-	1600	0.104103	0.6797781	0.1349961	0.461523
		4OR	0.0847258	0.7858918	0.1008808	0.696458
1600	-	1900	0.174241	0.1029318	0.1795709	0.047211
1900	-	2200	0.704485	0.8533546	0.0823761	0.799494
		3OR	0.0534349	0.913419	0.0625971	0.881182
2200	-	2500	0.0865299	0.7787631	0.0976713	0.718124
		3OR	0.0691066	0.8551855	0.0825784	0.793221
		2OR	0.06077	0.8840063	0.0733216	0.8311

Table 6.17 Chlorine raw data leverage

WAVELENGTH RANGE			RMSE$_C$	R^2_C	RMSE$_V$	R^2_V
400	-	2500	0.1742943	0.1555636	0.1832883	0.066165
700	-	1000	0.1892831	0.0040799	0.1948724	NA
1000	-	1300	0.1832711	0.0663407	0.1925339	NA
1300	-	1600	0.177751	0.1217368	0.1865631	0.032497
1600	-	1900	0.1735861	0.1624116	0.1830455	0.068637
1900	-	2200	0.1451869	0.4140566	0.17844	0.150228
		2OR	0.1758908	0.1404465	0.1861341	0.037416
2200	-	2500	0.1890336	0.0067045	0.1953679	NA

Table 6.18 Chlorine smoothing leverage

WAVELENGTH RANGE			$RMSE_C$	R^2_C	$RMSE_V$	R^2_V
400	-	2500	0.1742949	0.1555578	0.1832896	0.066151
700	-	1000	0.1892831	0.0040799	0.1948724	NA
1000	-	1300	0.1832742	0.0663084	0.1925413	NA
1300	-	1600	0.1777512	0.121735	0.1865636	0.032492
1600	-	1900	0.1741202	0.1572492	0.18348087	0.064938
1900	-	2200	0.145435	0.4120523	0.1755006	0.143835
2200	-	2500	0.1081879	0.6746449	0.1799127	0.100245
		3OR	0.1040735	0.6980579	0.1207094	0.593813
		3OR	0.099119	0.729377	0.1161746	0.62804
		2OR	0.0999929	0.7230027	0.1172198	0.619338

Table 6.19 Chlorine smoothing derivative

WAVELENGTH RANGE			$RMSE_C$	R^2_C	$RMSE_V$	R^2_V
400	-	2500	0.1103453	0.6615392	0.1597072	0.290995
700	-	1000	0.187005	0.0279081	0.1999059	NA
1000	-	1300	0.1469673	0.3995984	0.1753478	0.145324
1300	-	1600	0.1557345	0.3258285	0.1683546	0.212137
1600	-	1900	0.1779848	0.1194242	0.1864946	0.033208
1900	-	2200	0.0723062	0.8546712	0.0807327	0.818824
		3OR	0.0560441	0.9124406	0.0650528	0.882029
		2OR	0.0504362	0.9174837	0.059398	0.885544
2200	-	2500	0.074052	0.8475689	0.0865788	0.791635
		3OR	0.067093	0.8745131	0.0797169	0.822849
		4OR	0.0488636	0.8979786	0.0568963	0.861679

Table 6.20 Total dissolved solids raw data leverage correction

WAVELENGTH RANGE			RMSE$_C$	R^2$_C$	RMSE$_V$	R^2$_V$
400	-	2500	0.1604993	0.0068967	0.1692966	NA
700	-	1000	0.160957	0.0012244	0.1662923	NA
1000	-	1300	0.1608772	0.0022146	0.1676793	NA
1300	-	1600	0.1606305	0.0052734	0.17028	NA
1600	-	1900	0.160594	0.0057243	0.1698506	NA
1900	-	2200	0.13905	0.2545988	0.1502864	0.1292627
		3OR	0.1040146	0.5731649	0.1249493	0.3840595
		2OR	0.0986702	0.6210923	0.1159787	0.4764985
2200	-	2500	0.1185203	0.458456	0.1391125	0.2539284
		3OR	0.0959628	0.6427715	0.1146192	0.4903696
		2OR	0.0920828	0.6699963	0.1084595	0.542177

Table 6.21 Total dissolved solids

WAVELENGTH RANGE			RMSE$_C$	R^2$_C$	RMSE$_V$	R^2$_V$
700	-	2500	0.1604994	0.006896	0.169297	NA
700	-	1000	0.1609571	0.0012243	0.1662923	NA
1000	-	1300	0.1608775	0.0022119	0.1676783	NA
1300	-	1600	0.1606305	0.0052734	0.17028	NA
1600	-	1900	0.1605941	0.0057239	0.1698509	NA
1900	-	2200	0.1392296	0.2526719	0.1522694	0.1061319
		3OR	0.1044049	0.5699558	0.1257415	0.3762239
		2OR	0.0991663	0.617272	0.1166336	0.4705693
2200	-	2500	0.1185203	0.458456	0.1391125	0.2539284
		3OR	0.0959628	0.6427715	0.1146192	0.4903696
		2OR	0.0920828	0.6699963	0.1084595	0.5421775

Table 6.22 Total dissolved solids leverage smoothing derivative

| WAVELENGTH RANGE	RMSE$_C$	R^2$_C$	RMSE$_V$	R^2$_V$
400 - 2500	0.1073917	0.5553797	0.1290011	0.3584441
3OR	0.0478548	0.9111636	0.0582449	0.8684
4OR	0.0412049	0.9385229	0.0502144	0.9086997
700 - 1000	0.1602717	0.0097111	0.169339	NA
1000 - 1300	0.160927	0.0015969	0.1654956	NA
1300 - 1600	0.1597445	0.0162154	0.1648413	NA
1600 - 1900	0.1588224	0.0275412	0.1626483	NA
1900 - 2200	0.0694549	0.8140253	0.0909013	0.6814422
3OR	0.0538092	0.8876809	0.0632203	0.8449564
1OR	0.0473289	0.879836	0.0565406	0.8285082
2OR	0.0412105	0.9075364	0.0482358	0.8733239
2200 - 2500	0.0685291	0.8189503	0.0807988	0.7483147
3OR	0.057881	0.870039	0.0703487	0.8080215
1OR	0.0421802	0.904558	0.0501537	0.865063
3OR	0.0379732	0.9171062	0.0442593	0.8873901

6.6 RESULTS AND DISCUSSIONS

Graphs were plotted for the total dissolved solids, chlorine, electrical conductivity, and total hardness. The graphs were plotted for raw data, after preprocessing, smoothing and with first derivative, multi scatter correction first derivative etc. It is observed from the Figure 6.26 that the value for RMSE is very less and of R^2 is greater than 0.7 the model for total hardness from 700-2500 nm with multi scattering correction first derivative is acceptable model. From Figure 6.39 and Figure 6.40 for Chlorine in the range 1900-2200 nm and 2200-2500 nm after preprocessing and smoothing plus derivative the values of for RMSE is very less and of R^2 is greater than 0.7 the model for Chlorine content is acceptable. From the Figure 6.43, Figure 6.44 and Figure 6.45 it is observed that in the range 1900-2200 nm, 2200-2500 nm and complete range of NIR after preprocessing and smoothing plus derivative the values of RMSE is very less and of R^2 is greater than 0.7 the model for electrical conductivity is acceptable. From the Figure 6.49, Figure 6.50 and Figure 6.51 it is observed that in the range 1900-2200 nm, 2200-2500 nm and complete range of NIR after preprocessing and smoothing plus derivative the values of for RMSE is very less and of R^2 is greater than 0.7 the model for total dissolved solids is acceptable. From the Figure 6.52, it is observed that in the range 1900-2200 nm for raw data of total hardness and from Figure 6.53 and in the range of 1900-2200 nm with preprocessing and smoothing and from the Figure 6.54 in the range 1300-1600 nm with preprocessing and smoothing plus derivative the values of RMSE is less and of R^2 is nearly equal to 0.7 the model for total hardness is acceptable. Model for total hardness is acceptable in the range 1900-2200 nm and 2200-2500 nm with preprocessing and smoothing plus derivative also as can be observed from the Figure 6.55 and Figure 6.56. From Figure 6.57 model is also acceptable for complete NIR range. From the data available NIR spectroscopy and the physicochemical properties of water samples prediction and validation of the results were done NIR data analysis had been performed using Unscrambler $\times$ 10.5 software. Preprocessing such as normalization, derivative, baseline and smoothening has been applied on the NIR data. Multivariate analysis (PLS) were used to develop the rapid and non-destructive statistical regression model.

6.7 FOURIER TRANSFER INFRARED SPECTROSCOPY ANALYSIS:

Figure 6.58 Turbidity correlation loading based on FTIRS
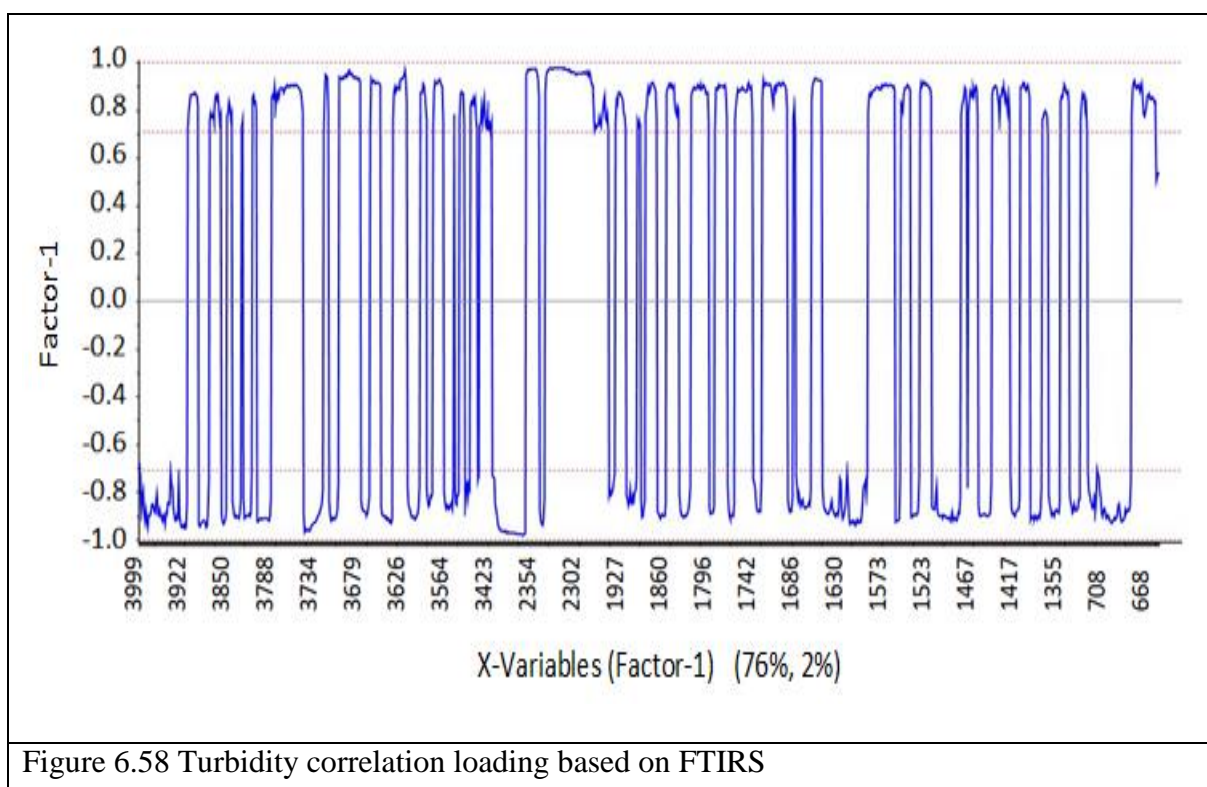
From the Figure 6.58 it is observed that there are various wavelengths where correlation is high in case of turbidity and the wavelengths range from 670-4000 nm.

Figure 6.59 Total hardness correlation loading based on FTIRS
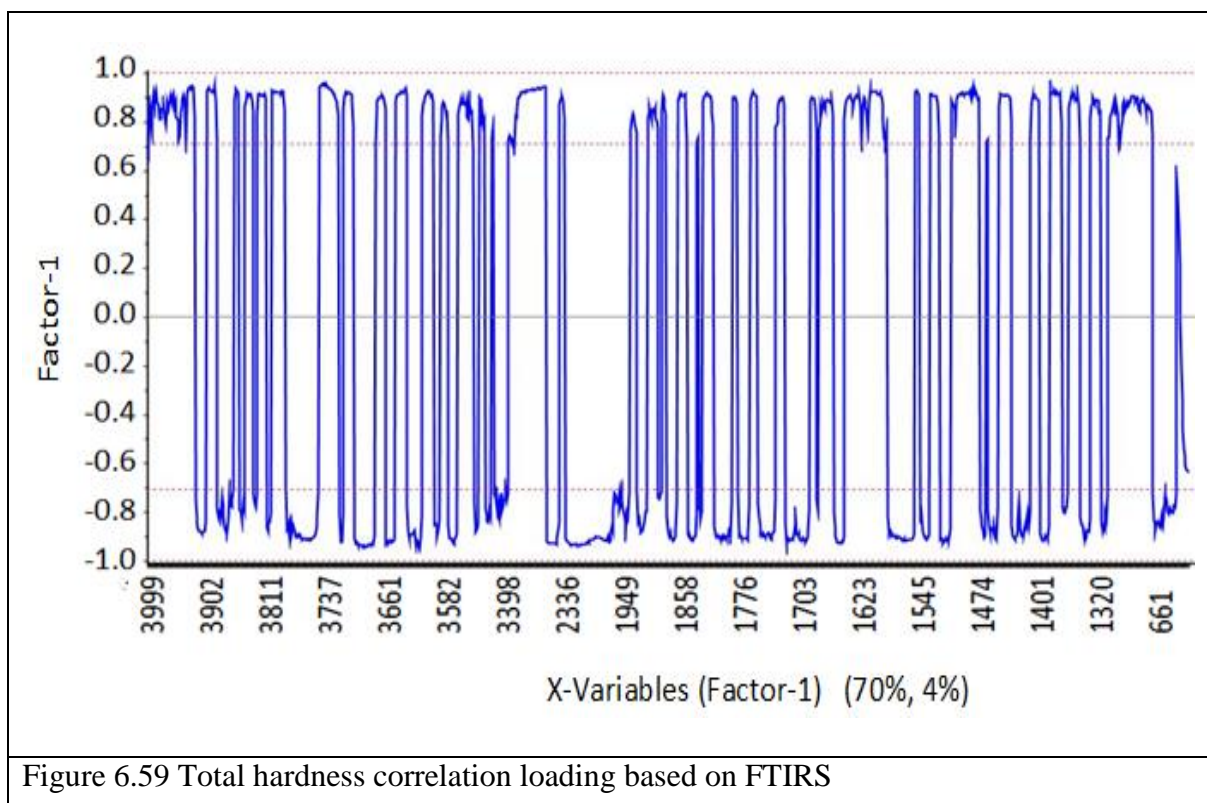
From the Figure 6.59 it is observed that there are various wavelengths where correlation is high in case of total hardness for wavelengths range from 660-4000 nm.

Figure 6.60 TDS correlation loading based on based on FTIRS
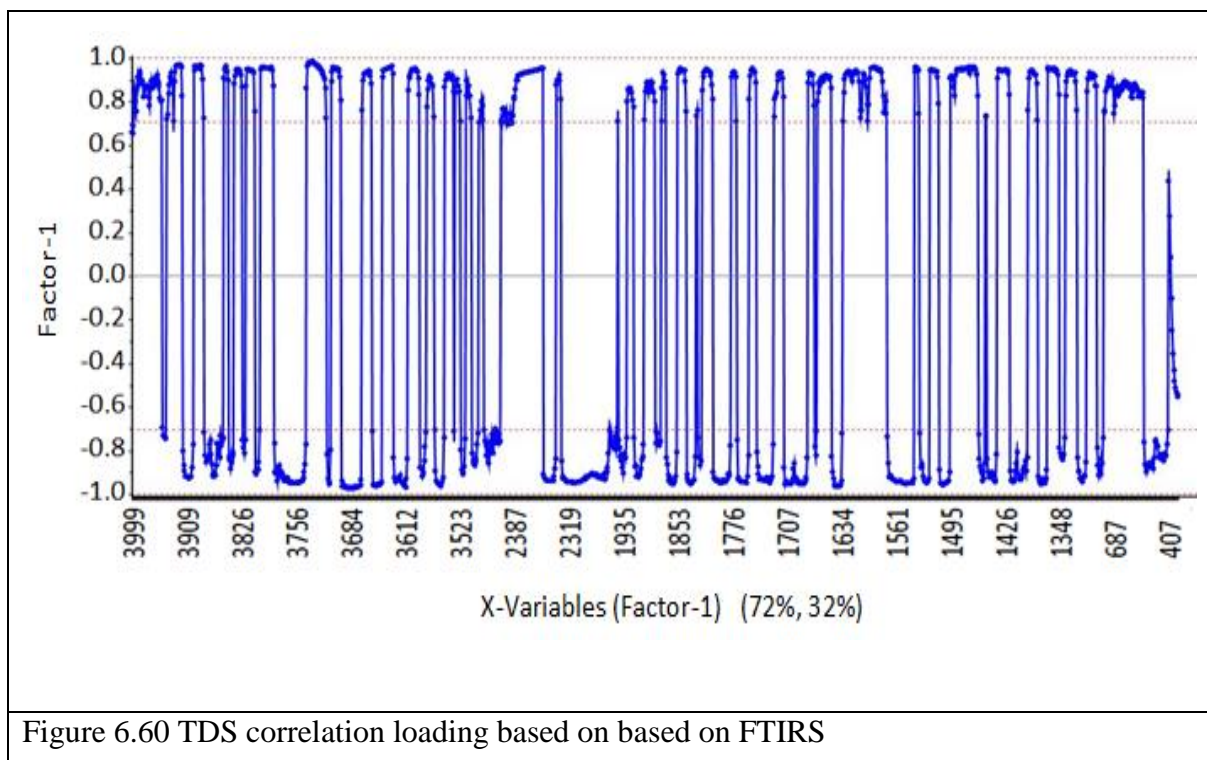
From the Figure 6.60 it is observed that there are various wavelengths where correlation is high in case of TDS for wavelengths range from 400-4000 nm.
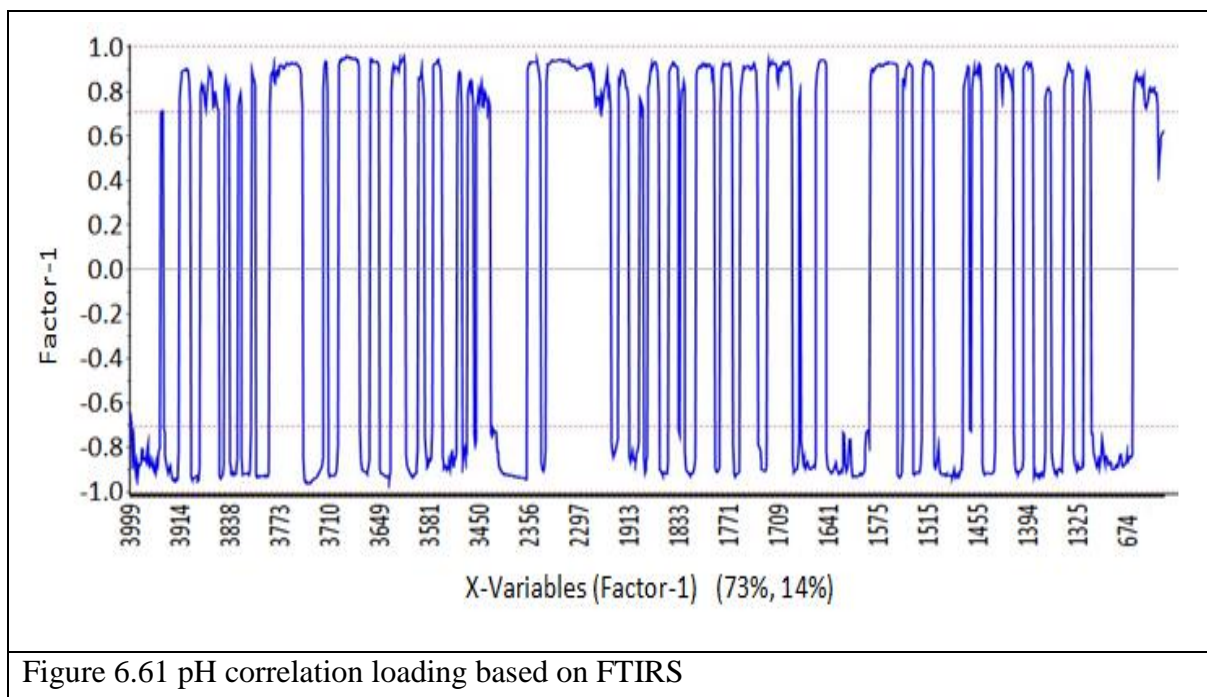

Figure 6.61 pH correlation loading based on FTIRS

From the Figure 6.61 it is observed that there are various wavelengths where correlation is high in case of pH for wavelengths range from 670-4000 nm.

Figure 6.62 EC correlation loading based on FTIRS

From the Figure 6.62 it is observed that there are various wavelengths where correlation is high in case of EC for wavelengths range from 660 - 4000 nm.

Figure 6.63 Cl correlation loading based on FTIRS
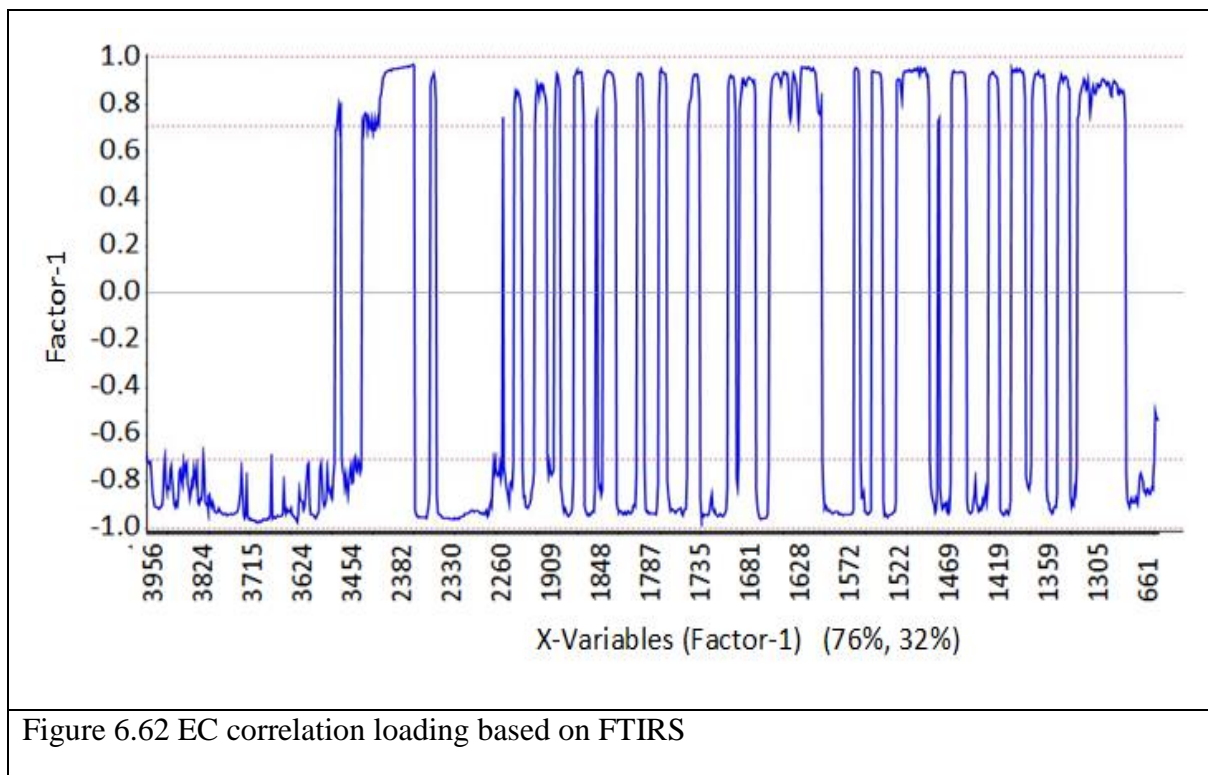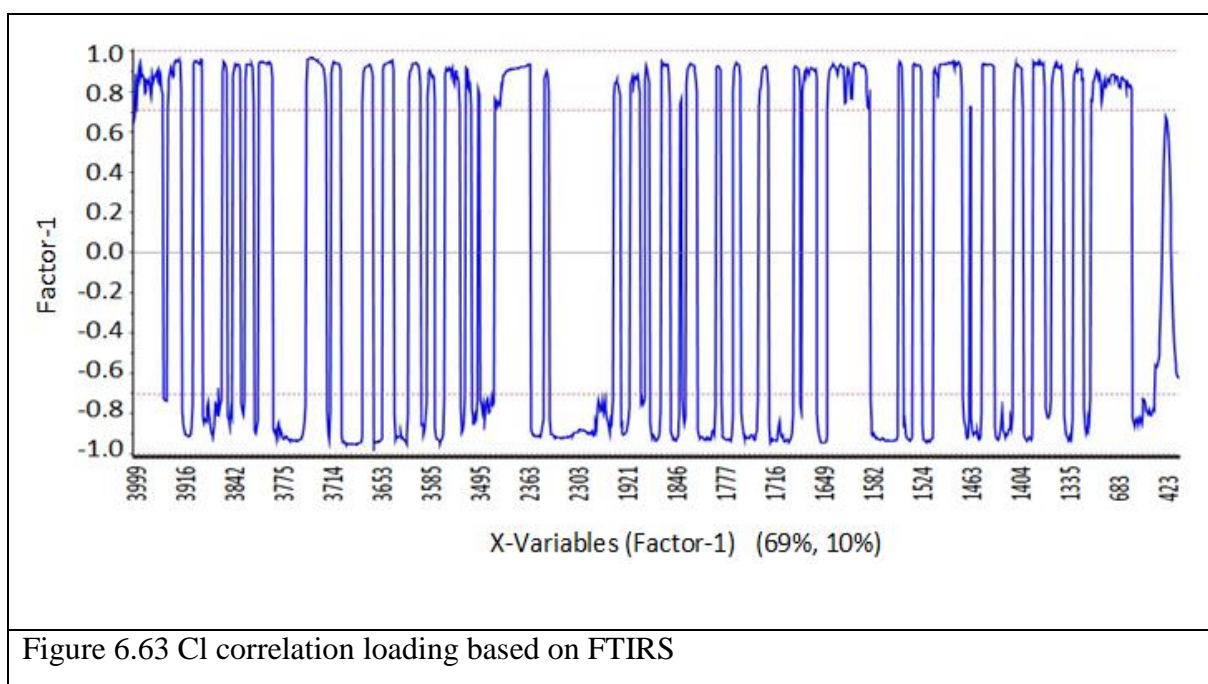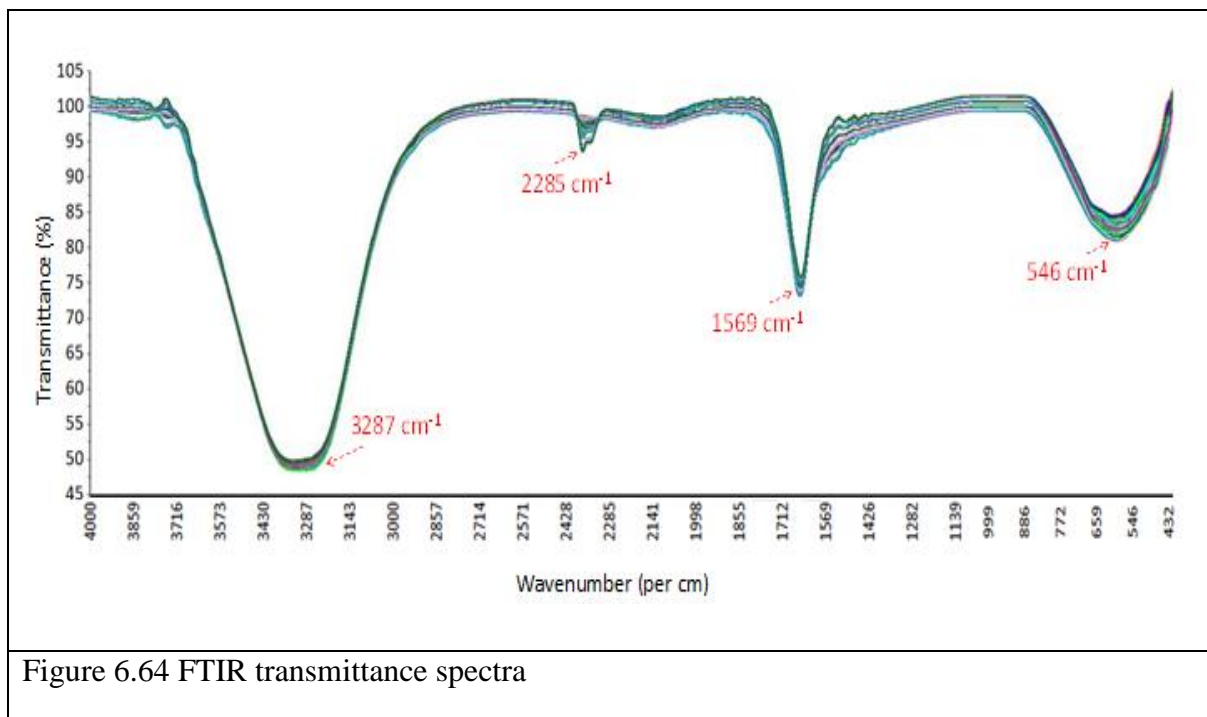
From the Figure 6.63 it is observed that there are various wavelengths where correlation is high in case of chlorine for wavelengths range from 420 - 4000 nm.

Figure 6.64 FTIR transmittance spectra

The Figure 6.64 gives the FTIR spectra for all the samples and there are peaks near 3280 cm^{-1} (50% transmittance), near 2300 cm^{-1} (95 % transmittance), near 1600 cm^{-1} (75% transmittance), 550 cm^{-1} (80% transmittance). There is possibility of bonds at these peaks.

Figure 6.65 FTIR spectra absorbance
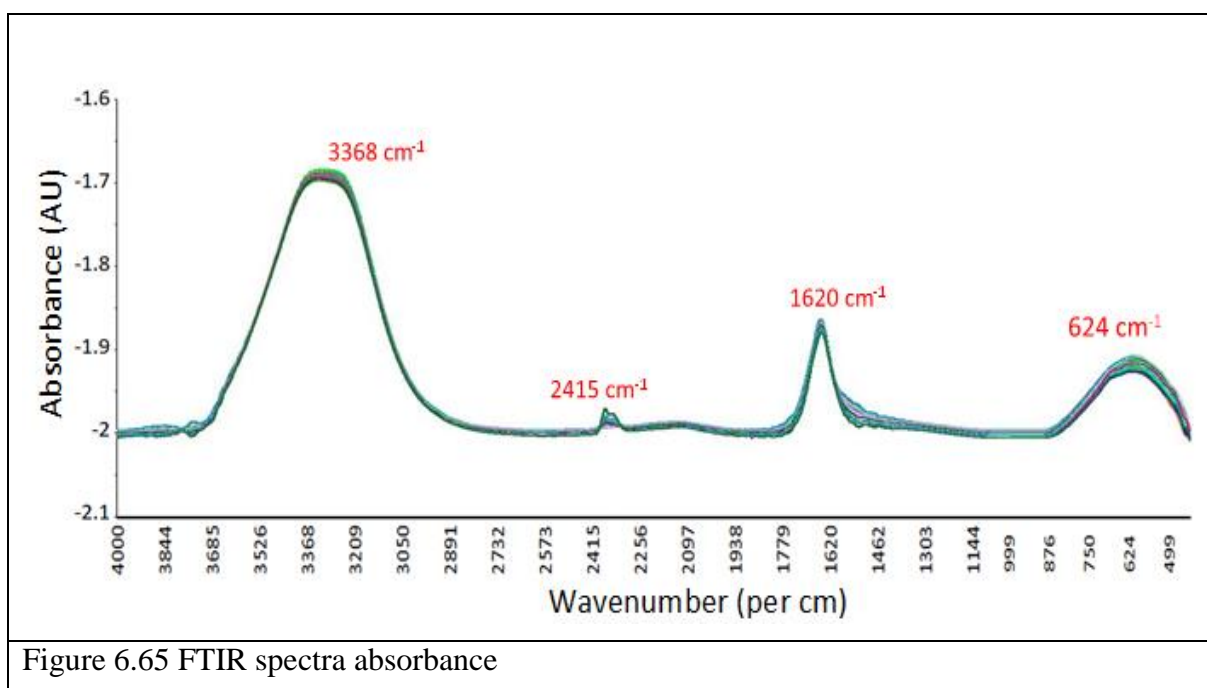
In the Figure 6.65 FTIR spectra is shown in absorbance mode. This also shows peaks at the values peaks near 3330 cm^{-1}, near 2300 cm^{-1}, near 1600 cm^{-1}, 550 cm^{-1}. There is possibility of bonds at these peaks.

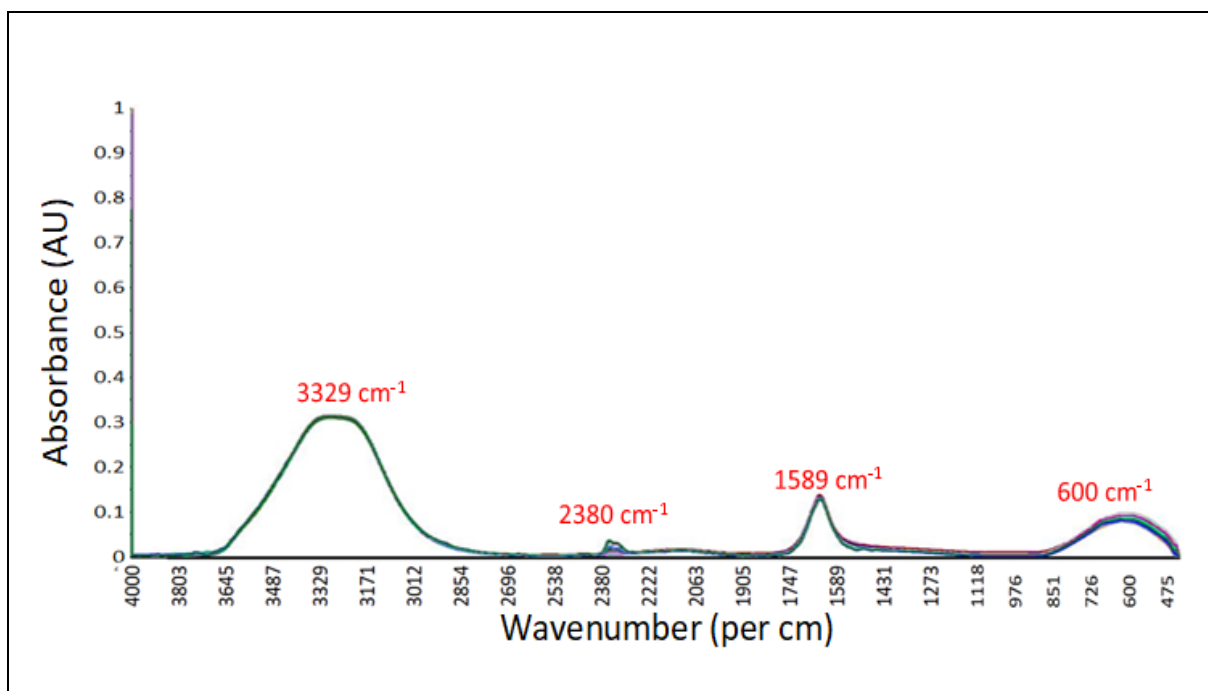Figure 6.66 FTIR spectra absorbance base line

In the Figure 6.66 the FTIR spectra in absorbance is shown with baseline. Showing peaks near 3300 cm^{-1} (absorbance near to 0.3), near 2380 cm^{-1} (absorbance near to 0.05), near 1600 cm^{-1} (absorbance near to 0.15), 600 cm^{-1} (absorbance near to 0.1).

Figure 6.67 FTIR spectra absorbance base line normalize

In the Figure 6.67 the FTIR spectra in absorbance is shown with baseline normalization. Showing peaks near 3300 cm^{-1} (absorbance near to 1.0), near 2365 cm^{-1} (absorbance near to 0.1), near 1600 cm^{-1} (absorbance near to 0.4), 600 cm^{-1} (absorbance near to 0.3).
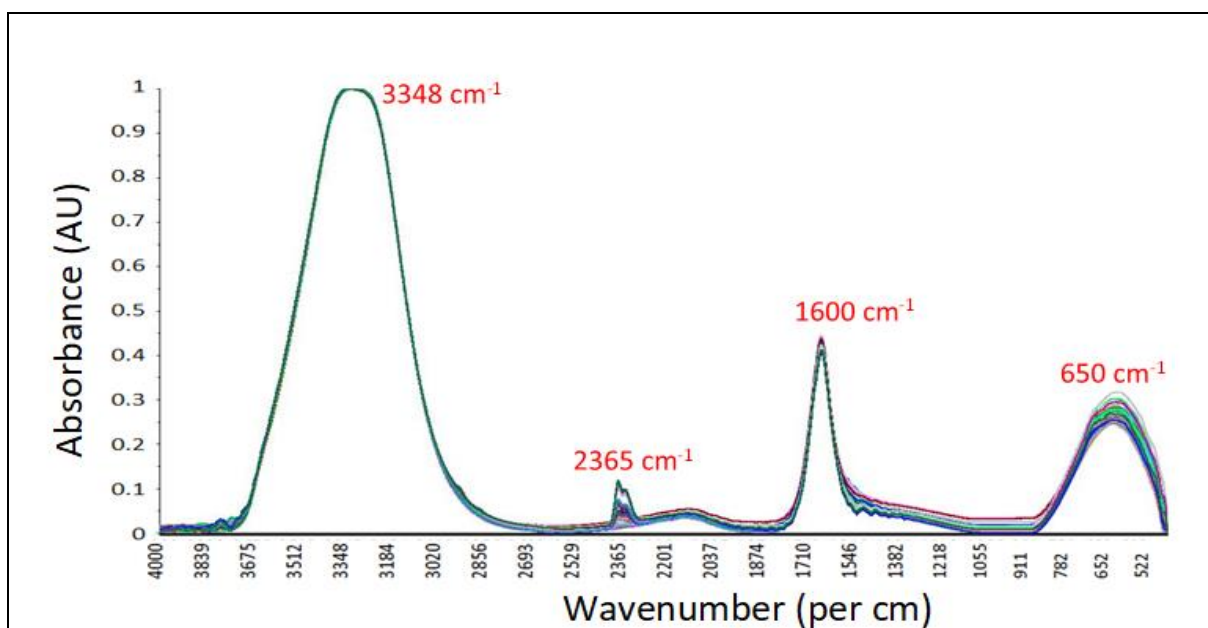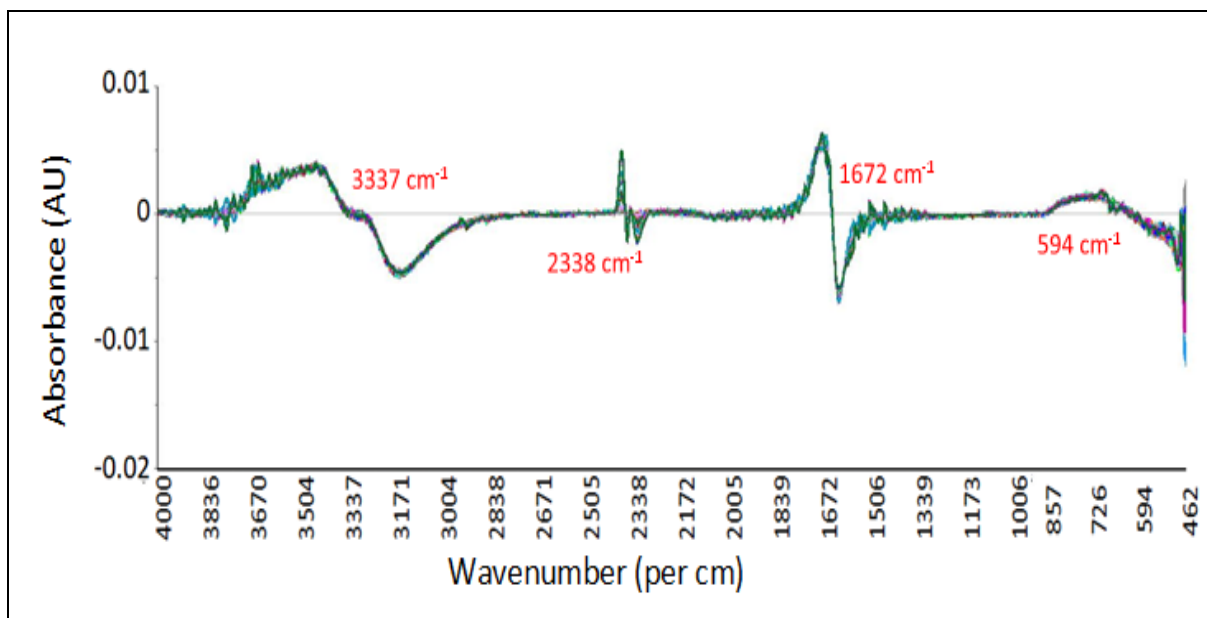

Figure 6.68 FTIR spectra absorbance base line normalize derivative

In the Figure 6.68 the FTIR spectra in absorbance is shown with baseline normalization with derivative, showing peaks near 3300 cm^{-1}, near 2380 cm^{-1}, near 1600 cm^{-1}, 600 cm^{-1}.
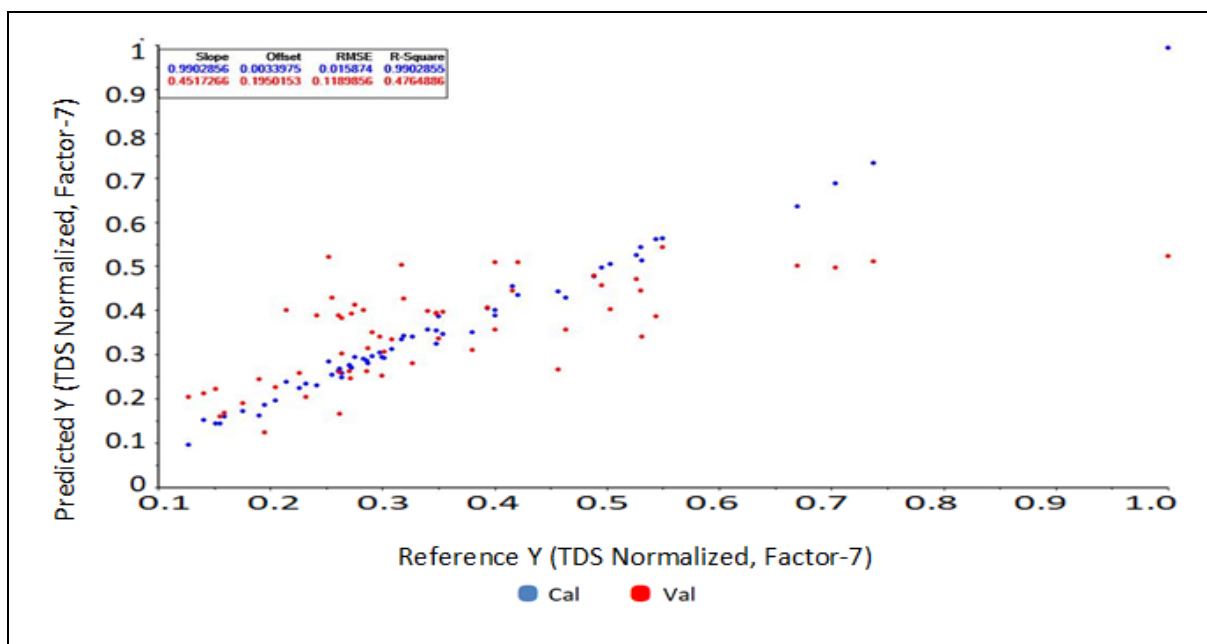
Figure 6.69 TDS prediction FTIR

In the Figure 6.69 it is observed that in case of total dissolved solids (TDS) the R^2 value in case of calibration is 0.9902855 and value of RMSE is 0.015874 and in case of validation the R^2 value is 0.4764886 and value of RMSE is 0.1189856.

Figure 6.70 pH prediction FTIR
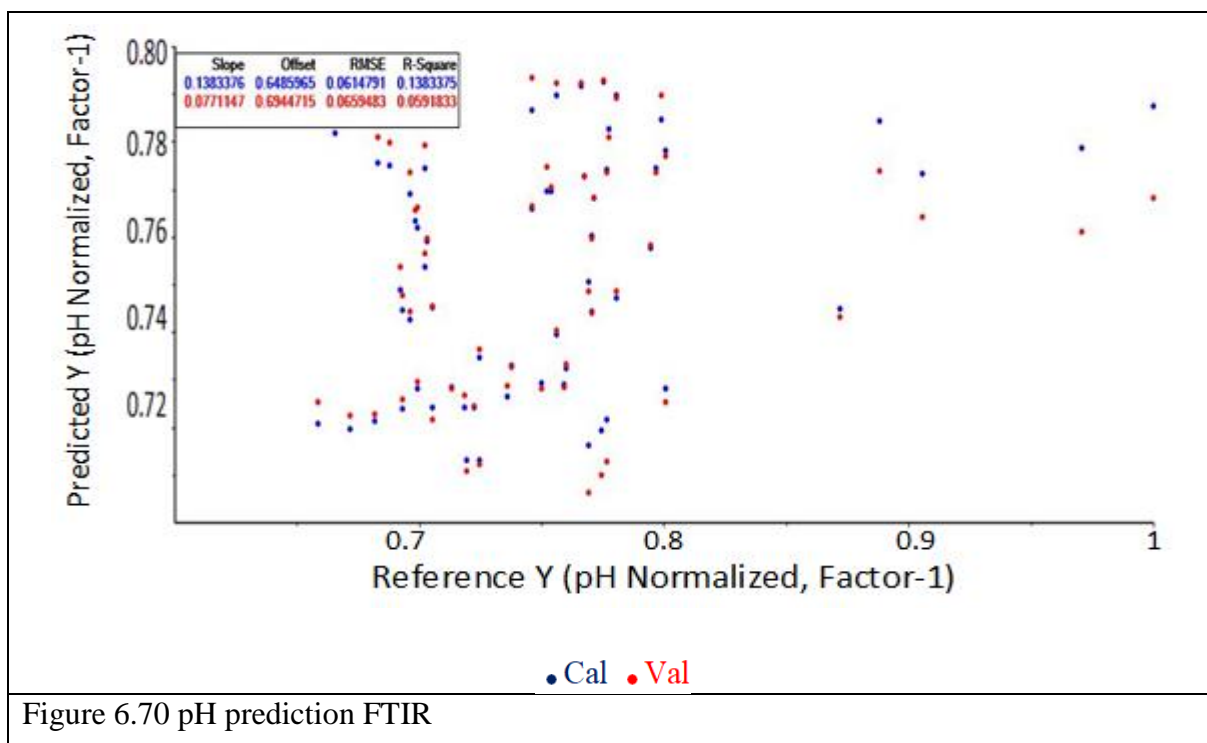
In the Figure 6.70 it is observed that in case of pH the R^2 value in case of calibration is 0.1383375 and value of RMSE is 0.0614791 and in case of validation the R^2 value is 0.0591833 and value of RMSE is 0.0659483.

Figure 6.71 EC prediction FTIR
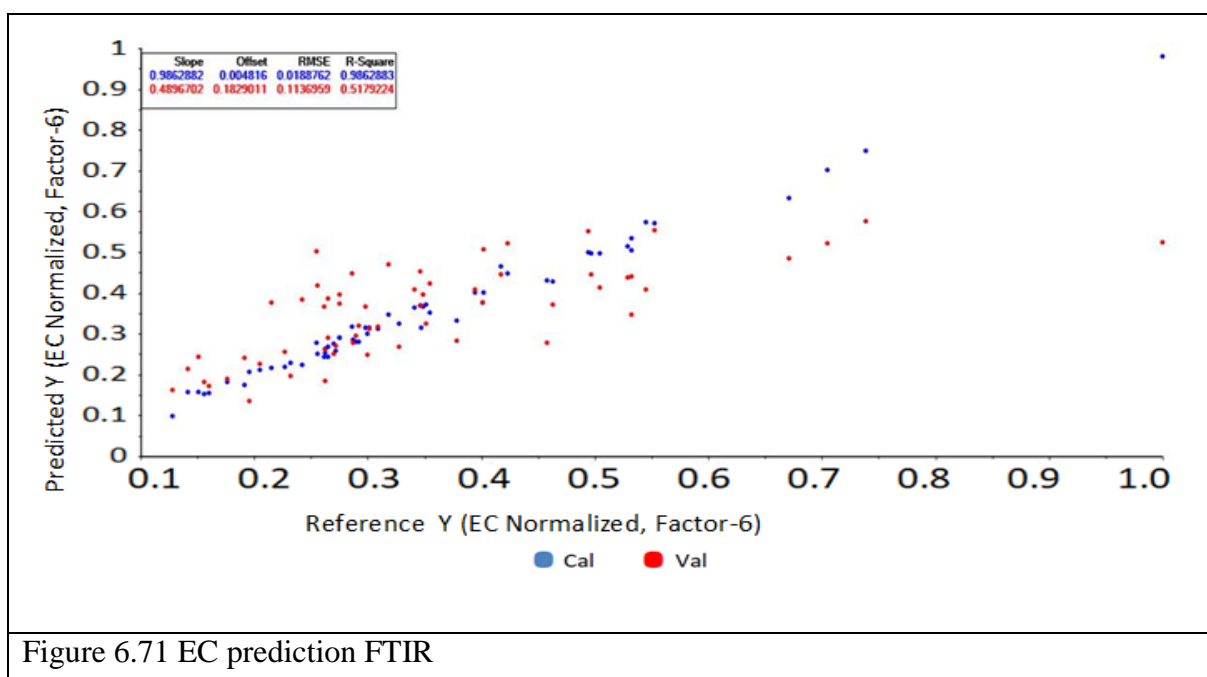
In the Figure 6.71 it is observed that in case of electrical conductivity the R^2 value in case of calibration is 0.9862883 and value of RMSE is 0.0188762 and in case of validation the R^2 value is 0.5179224 and value of RMSE is 0.1136959.

Figure 6.72 Cl prediction FTIR

In the Figure 6.72 it is observed that in case of chlorine content the R^2 value in case of calibration is 0.0963258 and value of RMSE is 0.1803041 and in case of validation the R^2 value is 0.0499089 and value of RMSE is 0.1871124.
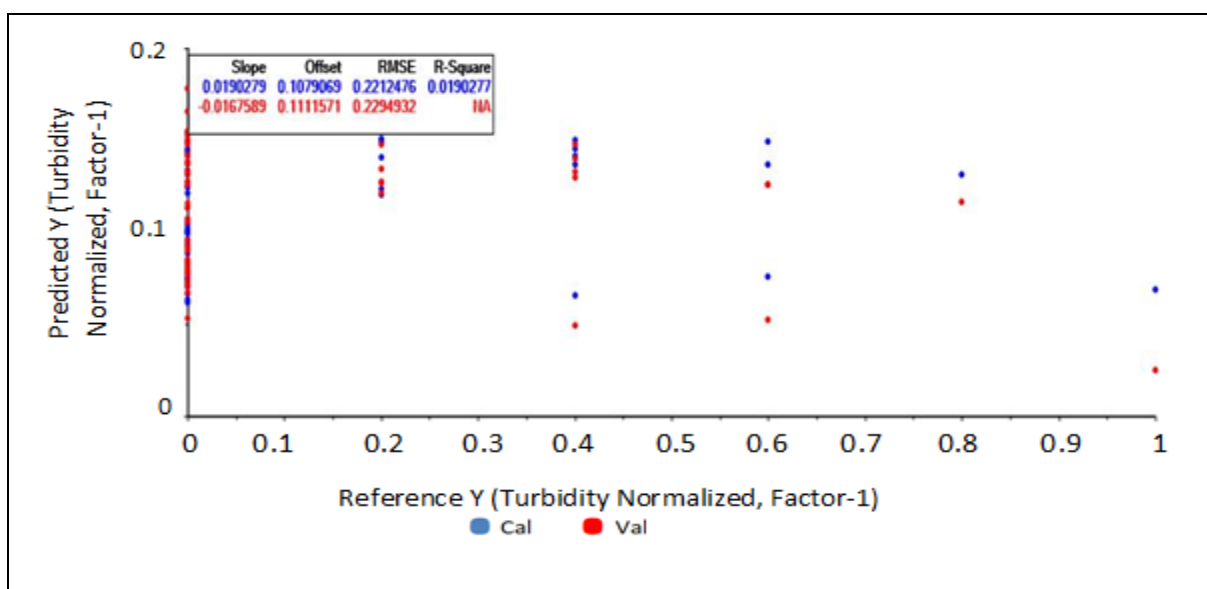
Figure 6.73 Turbidity prediction FTIR

From the Figure 6.73 it is observed that in case of turbidity the R^2 value in case of calibration is 0.0190277 and value of RMSE is 0.2212476 and in case of validation the R^2 value is NA and value of RMSE is 0.2294932.
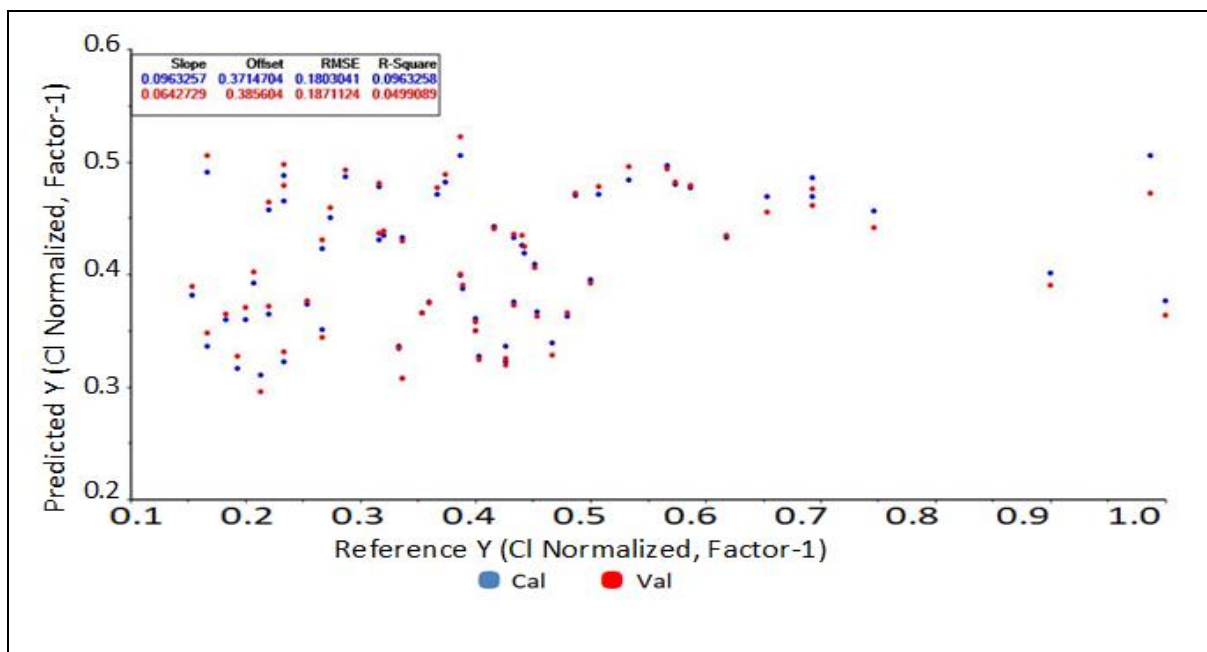

Figure 6.74 Total hardness prediction FTIR
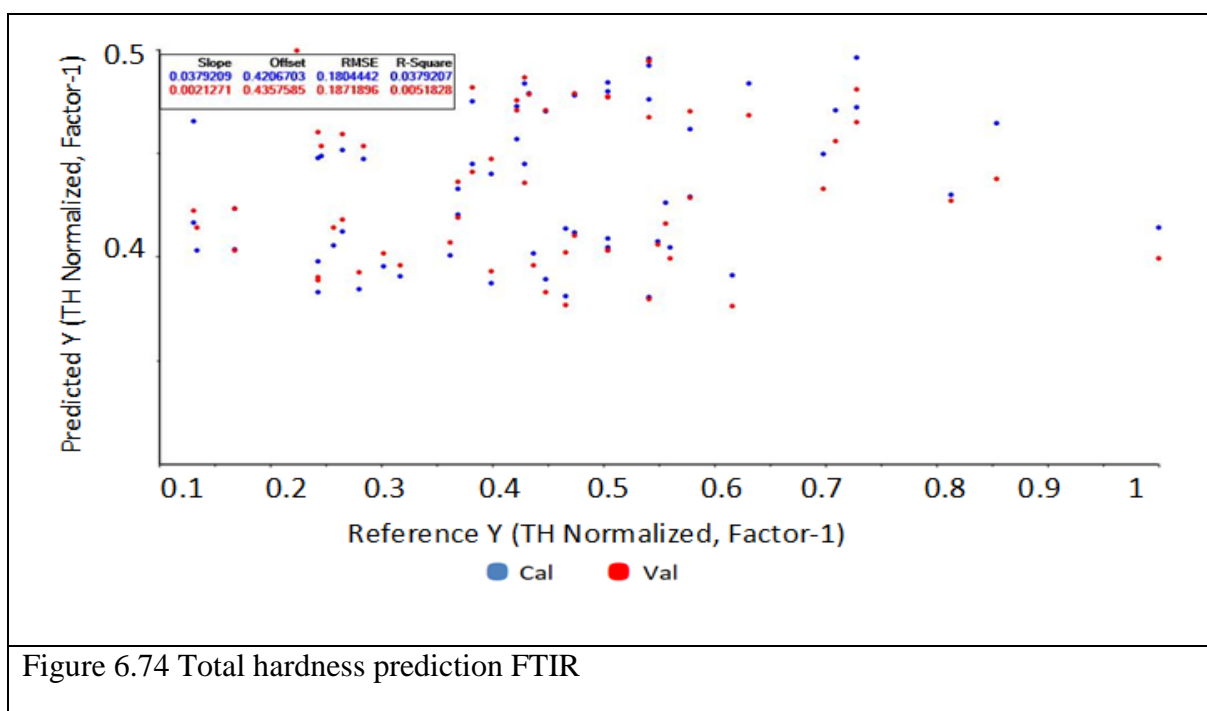
From the Figure 6.74 it is observed that in case of total hardness the R^2 value in case of calibration is 0.0379207 and value of RMSE is 0.1804042 and in case of validation the R^2 value is 0.0051828 and value of RMSE is 0.1871896.

6.8 CONCLUSION

Graphs were plotted for the total dissolved solids, chlorine, electrical conductivity, and total hardness for fluorescence, near infrared and FTIR spectra to find the correlations. The graphs were plotted for raw data, after preprocessing, smoothing and with first derivative, multi scatter correction first derivative etc. and the values of coefficient of determination for calibration and validation were calculated. Multivariate analysis (PLS) was used to develop the rapid and non-destructive statistical regression model.

It is observed from the graphs of fluorescence that at 415 nm and in the range 335 to 355 there is a weak correlation with turbidity, a weak correlation in case of total hardness near 330 nm and in the range 335 nm to 365 nm, for TDS there is positive correlation at 330 nm and negative correlation in the range 330 nm to 360 nm, in case of pH value, there is negative correlation from 325-330 nm and from 410-430 nm and there is positive correlation

from 370-380 nm, in case of electrical conductivity there is positive correlation near 330 nm, and negative correlation in the range 335-365 nm, in case of chlorine content there is positive correlation at 330 nm and near 400 nm and negative correlation in the range 335-360 nm. From the fluorescent spectra of the samples it is observed that all the samples are behaving somewhat similarly. There are emission peak at 325 nm for all the samples and then the emission goes on decreasing for all the samples and then somewhat linearly goes on increasing from 330 to 400 nm and the emission are ranging from all the samples goes from 8000 to 28000.

Those cases where the value for RMSE is very less and of R^2 is greater than or near to 0.7 then the model is acceptable e.g. total hardness from 700-2500 nm with multi scattering correction first derivative is acceptable model. For Chlorine in the range 1900-2200 nm and 2200-2500 nm after preprocessing and smoothing plus derivative the values of for RMSE is very less and of R^2 is greater than 0.7 the model for Chlorine content is acceptable. It is observed that in the range 1900-2200 nm, 2200-2500 nm and complete range of NIR after preprocessing and smoothing plus derivative the values of RMSE is very less and of R^2 is greater than 0.7 the model for electrical conductivity is acceptable. It is observed that in the range 1900-2200 nm, 2200-2500 nm and complete range of NIR after preprocessing and smoothing plus derivative the values of for RMSE is very less and of R^2 is greater than 0.7 the model for total dissolved solids is acceptable. It is observed that in the range 1900-2200 nm for raw data of total hardness, in the range of 1900-2200 nm with preprocessing and smoothing and in the range 1300-1600 nm with preprocessing and smoothing plus derivative the values of RMSE is less and of R^2 is nearly equal to 0.7 the model for total hardness is acceptable. Model for total hardness is acceptable in the range 1900-2200 nm and 2200-2500 nm with preprocessing and smoothing plus derivative.

The model obtained is based on the physicochemical properties and spectroscopic data of the water and can be used for on-site quality check of the water samples. The model can be used to predict the quality of the water, high pollution risk areas can be easily identified and actions can be taken in short time. Since the process of using models is economical and time saving, the results can be obtained at higher frequencies, it helps the policy makers to take relevant decisions on the basis of the results for high risk areas and otherwise. PCA is a statistical technique which helps in simplifying the complex data. This technique finds patterns in the complex data and reduces the dimensions of the complicated data. As an example, if there are thousands of data of many variables then PCA helps to

combine similar variables, thus making things easy for us and we have to deal with less number of variables. The PLSR is another statistical technique which is used when we have to predict something using many different measured parameters. PLSR finds that part of the data which is more useful in making predictions. The regression analysis helps in finding the relationship between the variables. These advanced statistical techniques are used to shrink the multidimensional data into information which we can understand easily.

CHAPTER 7

CONCLUSION AND FUTURE SCOPE

CHAPTER 7- CONCLUSION AND FUTURE SCOPE

7.1 CONCLUSION

The present Ph.D. work aimed at the quantitative and qualitative analysis of groundwater samples collected from the four districts of Majha region of Punjab, India. The objectives of the work were to identify the pollutants in the water using the Fourier transform infrared spectroscopy, near infrared spectroscopy and fluorescent spectroscopy. The water samples were analysed for six physicochemical parameters viz. total dissolved solids, pH value, electrical conductivity, turbidity, total hardness and chlorine content.

Statistical parameters like mean, median, standard deviation, variance, correlation etc. The Turbidity values ranges from 0 to 5 NTU and its mean values comes out to be 0.5500 NTU and standard deviation is 0.1454 NTU, the value of TDS ranges from 198 to 1557 mg/L, its mean value is 544.5367 mg/L, its standard deviation is 32.6467 mg/L. Electrical conductivity ranges from 305 to 2390 μS/cm, its mean value is 839.4333 μS/cm and standard deviation is 50.1582 μS/cm, pH value ranges from 6.48 to 9.84, and mean value comes out to be 7.4068 and standard deviation is 0.0848, total hardness ranges from 70 to 536 mg/L, and mean value is 234.3667 mg/L and its standard deviation is 12.8374 mg/L, Chloride content range from 23 to 150 mg/L, the mean is 61.6600 mg/L and standard deviations is 3.7039 mg/L.

The correlation between various pairs is also analysed and the pairs of physicochemical parameters which showed some correlation are given here – total dissolved solids - total hardness (0.5997), total dissolved solid - chloride (0.7001), electrical conductivity - total dissolved salts (0.9998), electrical conductivity - total hardness (0.5967), electrical conductivity - chloride (0.6987), chlorine-total hardness (0.8064); the number in the bracket is the value of coefficient of correlation between the pairs. Regression analysis was done for these six pairs and in chapter 5 the linear regression equations are given. These equations help us to predict one parameter from the values of other parameters.

The study aimed at two pollutants i.e. urea and Arsenic trioxide. The Fourier transfer infrared spectroscopy was done for all the groundwater samples and also for the pollutants in question. The comparison revealed that both the pollutants were not present in the samples.

Principal component analysis (PCA) was also done and from the score plot, it is observed that chlorine content, Turbidity and pH value are all on the same side and seems to be closely related to each other as compared to electrical conductivity, total dissolved solids

and total hardness as these parameters are far away from the closely related parameters. Classification of the samples is also done on the basis of source of water, depth of water and region, but the samples could not be classified on these basis. The NIR spectral data was also obtained for all the samples and the partial least square analysis was done to develop regression models and to determine the wavelength intervals for which good results could be obtained. This is the last objective of the Ph.D. work to develop a model using the NIR, FTIR and fluorescence spectral data to predict the chemical parameters in the ground water samples. The model will be beneficial in the determination of parameter like pH, chlorine content, total dissolved solids, electrical conductivity, total hardness and turbidity in the Majha region of Punjab in non-destructive, time saving and economical way. Partial least square regression technique was used to develop models for the near infrared range 700-2500 nm, models may predict the parameters for this range of wavelengths. The complete range of NIR (700-2500 nm) was split in to small intervals of 300 nm gap to further develop the models to minimize the span of time required in the analysis and this will also become cost effective for the future instruments to be developed. The selected wavelength range interval is important for a particular parameter, as it can predict the parameter in this range of wavelength. The regression model is developed all the small ranges of wavelengths and tested on the basis of R^2 (the coefficient of determination) and RMSE (root mean square error). The wavelength ranges for which the correlation of prediction is high are considered. It was observed that the wavelength range 1900-2200 nm and 2200-2500 nm with preprocessing, smoothing and derivative give high value of R^2 for both calibration and validation for chlorine content. In the case of electrical conductivity the suitable range is 1900 - 2200 nm and 2200 - 2500 nm with preprocessing, smoothing and derivative having high value of R^2 for calibration and validation. In the case of total dissolved solids the suitable range is 1900 - 2200 nm and 2200 - 2500 nm with preprocessing, smoothing and derivative having high value of R^2 for calibration and validation. The range for total hardness where R^2 is high comes out be 1900 - 2200 nm, 2200 - 2500 nm and 1300 - 1600 nm. It is observed that the ranges 1900 - 2200 nm and 2200 - 2500 nm of near infrared radiations seem to be suitable for chlorine, electrical conductivity, total dissolved solids and total hardness for the groundwater samples of four districts of Majha region Punjab, India.

From the fluorescent data graphs are plotted and correlation is observed for physicochemical parameters at particular wavelengths or for a range of wavelengths. For turbidity at the wavelength 415 nm, for total hardness a range is there from 335 nm to 365 nm, and at 330 nm, for total dissolved solids the correlation is found for the wavelength range

from 330 nm to 360 nm, for pH value the correlation is found between 325 to 330 nm and from 410 to 430 nm and 370 to 380 nm, for electrical conductivity there is correlation from 335 to 365 nm and near 330 nm. For chlorine content the correlation is observed for 330 nm and 400 nm and in the range 335 to 365 nm. It can be concluded that fluorescent spectroscopy shows correlation for many of the parameters.

The model developed using interval partial least regression technique may be successfully used for the determination of the parameters of ground water samples quantitatively in a rapid, non-destructive and ecofriendly manner. The developed models based on partial least square regression technique can be incorporated in softwares dealing with multivariate analysis. These softwares are used in the industries and the models can be used for the prediction of groundwater quality parameters from the region of Majha region, Punjab, India.

7.2 FUTURE SCOPE

- To develop a handy, robust and rapid spectroscopic instrument having in built programing software for multivariate analysis, that can analyze ground water for the water quality parameters for whole of the Majha region and then to whole Punjab of India.

- To identify various types of contaminants or pollutants quantitatively using spectroscopic techniques linked with the regression model developed using selected wavelength interval ranges.

- To explore other spectroscopic techniques like ultraviolet-visible spectroscopy and to cover more water quality parameters.

- A longitudinal study could be done by collecting samples over different seasons and years and analyzing the data to understand the contamination and pollution dynamics.

- In future studies microbial, microsplastics, heavy metals, pesticides water quality assessment can be combined with chemical and spectroscopic study

- Future studies can include other spectroscopic methods such as Raman spectroscopy, mass spectroscopy and X-ray spectroscopy, these can be used to develop the model which will be more accurate and will broaden the range.

- The research can be expanded to make self-learning models with the inclusion of new data and then to enhance water quality monitoring. This will need the inclusion of interdisciplinary collaborations from the fields like environmental, chemical sciences,

computer technologies, spectroscopic data advanced statistical techniques, and then policy making.

PUBLICATION OF PAPERS

1. Y. Kumar, N. Munjal, U. Kamboj (2024): Classification on the Basis of Physicochemical Properties of Ground Water of Majha Region of Punjab, India, African Journal of Biological Sciences, Issue 5, VOL. 6, ISSN 2663-2187/2406-2418.

2. Yogesh Kumar, Neha Munjal, Uma Kamboj (2024): Infrared Spectroscopy Analysis of Ground Water- A Pilot Study, African Journal of Biological Sciences, Issue 5, VOL. 6, ISSN 2663-2187/8439-8445.

CONFERENCES ATTENDED

1. Participated in the "Recent Advances in Fundamental and Applied Sciences" (RAFAS)-2021 at Lovely Professional University, Phagwara .

2. Presented a paper in the "International Conference on Functional Materials Manufacturing and Performances" (ICFMMP)-2022 held on Jul 29-30th, at Lovely Professional University, Phagwara -2022 having the title "Spectroscopy Analysis of Ground Water-A Pilot Study"

3. Attended a short term course on research methodology and Data analysis organized by Lovey Professional University, Phagwara from October 31, 2022 to November 05, 2022.

4.. Presented   paper in the "Recent Advances in Fundamental and Applied Sciences" (RAFAS)- 2023 having the title "Classification on the basis of Physicochemical properties of Ground water of Majha Region of Punjab, India".

5. Participated in **Science, Space and Energy, Aerospace Innovation and Sustainability Workshop** held on 21 Feb, 2025 by COLOMBIAN AEROSPACE FORCE, ECOPETROL (ICPET)-CODALTEC.

REFERENCES

[1]   Gerhard Herzberg,1945, Atomic spectra and atomic structure, New York, Dover publication,.

[2]   Cathal Wilson, Gerard McGranaghan, 2014, Infrared heating comes of age, Reinforced Plastics, Volume 58, Issue 2, Pages 43-47, https://doi.org/10.1016/S0034-3617(14)70109-2.

[3]   Rita kakkar, 2015, Atomic and molecular spectroscopy, Basic concepts and applications, India,Cambridge university press,

[4]   David W. Ball, 1962, The basics of spectroscopy, Washington USA, SPIE press, Bellingham,.

[5]   James Withrow, 2016, Infrared Spectroscopy, Revised edition, New York, Published by Research world,.

[6]   Ferraro JR, Nakamoto K, 2003, Introductory Raman spectroscopy. 2nd ed. Amsterdam: Elsevier;

[7]   Hollas JM., 2002, Basic atomic and molecular spectroscopy. Cambridge (UK): The Royal Society of Chemistry;

[8]   Banwell CN, McCash EM. 2003, Fundamentals of molecular spectroscopy. 4th ed. New Delhi (India): Tata McGraw Hill Publishing Company Limited;.

[9]   Siesler HW, Ozaki Y, Kawata S, Heise HM, 2002, Near-infrared spectroscopy: principles, instruments, applications. Weinheim (Germany): WILEY-VCH Verlag GmbH;.

[10]   Hollas JM. 2004, Modern spectroscopy. 4th ed. Chichester (UK): John Wiley & Sons Ltd.

[11]   Burns DA, Ciurczak EW, editors. 2008, Handbook of near-infrared analysis. 3rd ed. Boca Raton (FL): CRC Press, Taylor & Francis Group.

[12]   Smith BC. 2011, Fundamentals of Fourier transform infrared spectroscopy. 2nd ed. Boca Raton (FL): CRC Press, Taylor & Francis Group;.

[13]   Gauglitz G, Vo-Dinh T, editors. 2003, Handbook of spectroscopy. Weinheim (Germany): WILEY-VCH Verlag GmbH & Co. KGaA;.

[14]   Ozaki Y, Huck CW, Bec KBB. Near-IR spectroscopy and its applications. In: Molecular and laser spectroscopy. Amsterdam: Elsevier Inc.; 2018. doi:10.1016/3978-0-12-849883-5.00002-4.

[15]   Skoog DA, West DM, Holler FJ, Crouch SR. 2014 Fundamentals of analytical chemistry. 9th ed. Belmont (CA): Brooks/Cole, Cengage Learning;.

[16]   https://chem.libretexts.org/Courses/University_of_California_Davis/CHE_115:_Instrumental_Analysis_-_Lab_Manual/Lab_4:_Molecular_Fluorescence

[17]   Radley JA, Grant J. 1933. Fluorescence analysis in ultra-violet light. New York: D. Van Nostrand Company, Inc.;.

[18]   Kılıç Z. 2020. The importance of water and conscious use of water. Int J Hydrol. 4:239–241.

[19]   Akter R. 2020. Importance of safe drinking water for human life. Res Mol Cell Clin Biochem. 1:10–14.

[20]   Kaur P. 2021. Drinking water contaminants and health risks: a systematic review. [Zeichen Journal]. 7:366–379.

[21]   World Health Organization. 2003. Total dissolved solids in drinking water: background document for development of WHO guidelines for drinking water quality. WHO/SDE/WSH/03.04/16.

[22]   https://cdn.who.int/media/docs/default-source/wash-documents/wash-chemicals/tds.pdf? sfvrsn= 3e6d651e_4

[23]   World Health Organization. 2007. pH in drinking-water revised background document for development of WHO guidelines for drinking-water quality. WHO/SDE/WSH/07.01/1

[24]   World Health Organization. 2003. Chlorine in drinking-water: background document for development of WHO guidelines for drinking-water quality. WHO/SDE/WSH/03.04/45.

[25]   Bureau of Indian Standards. 2012. Drinking water — specification (second revision). IS 10500. ICS 13.060.20.

[26]   https://www.punjabdata.com/Majha-Malwa-Doaba.aspx

[27]   Gupta SP. Statistical methods. 34th ed. New Delhi: Sultan Chand & Sons, 2005.

[28]   Montgomery DC, Peck EA, Vining GG. 2012. Introduction to regression analysis. 5th ed. Hoboken (NJ): John Wiley & Sons, Inc. Wiley series in probability and statistics.

[29]   Tobias RD. 1997. An introduction to partial least squares regression. Cary (NC): SAS Institute.

[30]   Seasholtz MB, Kowalski BR. 1992. The effect of mean centering on prediction in multivariate calibration. J Chemom. 6(2):103–111. doi:10.1002/cem.1180060208.

[31]   https://www.cgwb.gov.in/old_website/WQ/GROUND%20WATER%20QUALITY%20SCENAR IO_%20IN%20INDIA.pdf

[32]   Singh A, Guite LT. 2022. Assessing the prevalence of water borne diseases in Muktsar district of Punjab, India. Trans Inst Indian Geogr. 44:1–12.

[33]   Büning-Pfaue H. 2003. Analysis of water in food by near infrared spectroscopy. Food Chem. 82(1):107–115. doi:10.1016/s0308-8146(02)00583-6.

[34]   Xu P. 2018. Research and application of near-infrared spectroscopy in rapid detection of water pollution. Desalination Water Treat. 122:1–4. doi:10.5004/dwt.2018.22559.

[35]   Munawar AA, Zulfahrizal, Meilina H, Pawelzik E. 2022. Near infrared spectroscopy as a fast and non-destructive technique for total acidity prediction of intact mango: comparison among regression approaches. Comput Electron Agric. 193:106657. doi:10.1016/j.compag.2021.106657.

[36]   Shawky E, Selim DA. 2019. NIR spectroscopy-multivariate analysis for discrimination and bioactive compounds prediction of different Citrus species peels. Spectrochim Acta A Mol Biomol Spectrosc. 219:1–7. doi:10.1016/j.saa.2019.04.026

[37] Liu Z, Pan L, Hu F, Hu Y. 2020. Advanced landfill leachate biochemical effluent treatment using Fe-Mn/AC activates O3/Na2S2O8 process: process optimization, wastewater quality analysis, and activator characterization. Environ Sci Pollut Res. 27:15337–15349. doi:10.1007/s11356-020-08046-2.

[38] Muñoz MA, Carmona C, Balón M. 2007. FTIR study of water clusters in water–triethylamine solutions. Chem Phys. 335:37–42. doi:10.1016/j.chemphys.2007.03.015.

[39] Galapate RP, Baes AU, Ito K, Mukai T, Shoto E, Okada M. 1998. Detection of domestic wastes in Kurose River using synchronous fluorescence spectroscopy. Water Res. 32:2232–2239.

[40] Thomas A, Sharma PK, Sharma MK, Sood A. 1999. Hydrogeomorphological mapping in assessing ground water by using remote sensing data - a case study in Lehra Gaga block, Sangrur district, Punjab. J Indian Soc Remote Sens. 27:31–42.

[41] Virk HS, Walia V, Bajwa BS. 2001. Radon monitoring in underground water of Gurdaspur and Bathinda districts of Punjab, India. Indian J Pure Appl Phys. 39:746–749.

[42] Howe KJ, Ishida KP, Clark MM. 2002. Use of ATR FTIR to study fouling in natural waters. Desalination. 147:251–255. doi:10.1016/j.desal.2002.03.015.

[43] Farooqi A, Masuda H, Siddiqui R, Naseem M. 2009. Sources of arsenic and fluoride in highly contaminated soils causing groundwater contamination in Punjab, Pakistan. Arch Environ Contam Toxicol. 56:693–706.

[44] Sohn KJ, Himmelsbach DS, Barton FE 2nd, Fedorka-Cray PJ. 2009. Fluorescence spectroscopy for rapid detection and classification of bacterial pathogens. Appl Spectrosc. 63(11):1251–1255. doi:10.1366/000370209789806993.

[45] Singh P, Saharan JP. 2010. Elemental analysis of Satluj River water using EDXRF. Nat Sci. 24–28.

[46] Li J, Zhang Y, Cai W, Shao X. 2011. Simultaneous determination of mercury, lead and cadmium ions in water using near-infrared spectroscopy with preconcentration by thiol-functionalized magnesium phyllosilicate clay. Talanta. 84:679–683.

[47] Ajay Kumar, Rout S, Narayanan U, Mishra MK, Tripathi RM, Singh J, Kumar S, Kushwaha HS. 2011. Geochemical modeling of uranium speciation in the subsurface aquatic environment of Punjab State in India. J Geochem Model Res. 3(5):137–146. Article Number - D2EA49C6003. https://doi.org/10.5897/JGMR.9000028

[48] Srivastava S, Tandon P, Singh R, Uttam KN. 2013. Elemental investigation of river Ganga water by LIBS. Natl Acad Sci Lett. 36:57–60.

[49] Sidhu M, Mahajan P, Bhatt SM. 2014. Highly sensitive & low cost colorimetric method for quantifying arsenic metal in drinking water of Malwa Punjab and comparison with ICAP-AES. J Environ Chem. 5:105–109.

[50] Chen Y, Chen ZP, Jin JW, Yu RQ. 2015. Quantitative determination of ametryn in river water using surface-enhanced Raman spectroscopy coupled with an advanced chemometric model. Chemom Intell Lab Syst. 142:166–171.

[51] Saikia BJ, Parthasarathy G, Borah RR, Borthakur R. 2016. Raman and FTIR spectroscopic evaluation of clay minerals and estimation of metal contaminations in natural deposition of surface sediments from Brahmaputra River. Int J Geosci. 7:873–883.

[52] Virk HS. 2017. A crisis situation due to uranium and heavy metal contamination of ground waters in Punjab State, India: A preliminary report. RRJoT Res Rev A J Toxicol. 7:6–11

[53] Wen Z, Song K, Liu G, Lyu L, Shang Y, Fang C, Du J. 2019. Characterizing DOC sources in China's Haihe River basin using spectroscopy and stable carbon isotopes. Environ Pollut. 258:113684.

[54] Tanu Sharma, Arvesh Sharma, Kaur I, Mahajan RK, Litoria PK, Sahoo SK, Bajwa BS. 2019. Uranium distribution in groundwater and assessment of age dependent radiation dose in Amritsar, Gurdaspur and Pathankot districts of Punjab, India. Chemosphere. 219:607–616. https://doi.org/10.1016/j.chemosphere.2018.12.039.

[55] Quintelas C, Melo A, Costa M, Mesquita DP, Ferreira EC, Amaral AL. 2020. Environmentally-friendly technology for rapid identification and quantification of emerging pollutants from wastewater using infrared spectroscopy. Environ Toxicol Pharmacol. 80:103458

[56] Mamera M, van Tol JJ, Aghoghovwia MP, Kotze E. 2020. Sensitivity and calibration of the FT-IR spectroscopy on concentration of heavy metal ions in river and borehole water sources. Appl Sci. 10:1–16. https://doi.org/10.3390/app10217785.

[57] Christian Scherer, Annkatrin Weber, Friederike Stock, Sebastijan Vurusic, Harun Egerci, Christian Kochleus, Niklas Arendt, Corinna Foeldi, Georg Dierkes, Martin Wagner, Nicole Brennholt, Georg Reifferscheid. 2020. Comparative assessment of microplastics in water and sediment of a large European river. Sci Total Environ. 738:139866. https://doi.org/10.1016/j.scitotenv.2020.139866.

[58] Kumar R, Mittal S, Peechat S, Sahoo PK, Sahoo SK. 2020. Quantification of groundwater–agricultural soil quality and associated health risks in the agri-intensive Sutlej River Basin of Punjab, India. Environ Geochem Health. 42:4245–4268. https://doi.org/10.1007/s10653-020-00636-w.

[59] Kumar A, Singh CK. 2020. Arsenic enrichment in groundwater and associated health risk in Bari doab region of Indus basin, Punjab, India. Environ Pollut. 256:113324. https://doi.org/10.1016/j.envpol.2019.113324.

[60] Kumar R, Mittal S, Sahoo PK, Sahoo SK. 2021. Source apportionment, chemometric pattern recognition and health risk assessment of groundwater from southwestern Punjab, India. Environ Geochem Health. 43:733–755. https://doi.org/10.1007/s10653-020-00736-7.

[61] Rani R, Kumar S, Munjal N, Kamboj U. 2022. Characterization of groundwater using spectroscopic techniques. J Phys Conf Ser. 2267:01202. https://doi.org/10.1088/1742-6596/2267/1/012022.

[62] Gautam D, Vaid U, Khaki QZ. 2023. Suitability evaluation of groundwater for drinking and agriculture purpose in Una District, Himachal Pradesh (India). IOP Conf Ser: Earth Environ Sci. 1110.

[63] Singh K, Singh R, Pandey G. 2022. Hydrogeochemistry, solute source identification, and health risk assessment of groundwater of cancer-prone region in India. Water Supply. 23(1):317–342. https://doi.org/10.2166/ws.2022.435.

[64] Iqbal J, Su C, Wang M, Abbas H, Baloch MYJ, Ghani J, Ullah Z, Huq ME. 2023. Groundwater fluoride and nitrate contamination and associated human health risk assessment in South Punjab, Pakistan. Environ Sci Pollut Res. 30:61606–61625. https://doi.org/10.1007/s11356-023-25958-x.

[65] Gautam VK, Kothari M, Al-Ramadan B, Singh PK, Upadhyay H, Pande CB, Fahad Alshehri, Zaher Mundher Yaseen. 2024. Groundwater quality characterization using an integrated water quality index and multivariate statistical techniques. PLOS ONE. 19(2):e0294533. https://doi.org/10.1371/journal.pone.0294533.

[66] Tamanna, Gupta NC. 2024. A study and evaluation of groundwater quality and heavy metal contamination in Bathinda district of Punjab, India. Environ Earth Sci. https://doi.org/10.53550/EEC.2023.v29i04.065.

[67] IS 10500:2012. 2012. Drinking Water Specifications, Second Revision. Bureau of Indian Standards, New Delhi

[68] World Health Organization. 2017. Guidelines for drinking-water quality, fourth edition incorporating the first addendum. ISBN 978-92-4-154995-0. World Health Organization

[69] Kaur P. 2021. Drinking Water Contaminants. Zeichen J. 7(6):366–379. ISSN No: 0932-4747.

[70] Virk HS. 2017. A crisis situation due to uranium and heavy metal contamination of ground waters in Punjab State, India: A preliminary report. Res Rev A J Toxicol. 7(2):6–11.

[71] Wang Y, Li J, Ma T, Xie X, Deng Y, Gan Y. 2020. Genesis of geogenic contaminated groundwater: As, F, and I. Crit Rev Environ Sci Technol. 50(1):1–24.

[72] Virk HS, Walia V, Bajwa BS. 2001. Radon monitoring in underground water of Gurdaspur and Bathinda districts of Punjab, India. Indian J Pure Appl Phys. 39:746–749.

[73] Singh S, Rani A, Mahajan RK, Walia TPS. 2003. Analysis of uranium and its correlation with some physico-chemical properties of drinking water samples from Amritsar, Punjab. J Environ Monit. 5:917–921.

[74] Sinha Ray SP, Elango L. 2019. Deterioration of groundwater quality: implications and management. In: Singh A, Saha D, Tyagi AC, editors. Water Governance: Challenges and Prospects. Singapore: Springer. p. 87–101. https://doi.org/10.1007/978-981-13-2700-1_5.

[75] Sharma T, Bajwa BS, Kaur I. 2022. Hydro-geochemical characteristics and quality appraisal of aquifers using multivariate statistics and associated risk assessment in Tarn-taran district, Punjab, India. Environ Sci Pollut Res. 29:54916–54938. https://doi.org/10.1007/s11356-021-16327-7.

[76] Ahada CPS, Suthar S. 2018. Groundwater nitrate contamination and associated human health risk assessment in southern districts of Punjab, India. Environ Sci Pollut Res. 25: 18360–18371. https://doi.org/10.1007/s11356-018-2581-2.

[77] Adimalla N, Wu J. 2019. Groundwater quality and associated health risks in a semi-arid region of south India: Implication to sustainable groundwater management. Hum Ecol Risk Assess. https://doi.org/10.1080/10807039.2018.1546550.

[78] Khatri P, Gupta KK, Gupta RK. 2021. A review of partial least squares modeling (PLSM) for water quality analysis. Model Earth Syst Environ. 7:703–714. https://doi.org/10.1007/s40808-020-00995-4.

[79] Sharifi A. 2016. Partial least squares-regression (PLS-regression) in chemometrics. In: Proceedings of the 1st National Conference on Achievement in Chemistry and Chemical Engineering (ACCE), Shiraz, Iran.

[80] Wold S, Sjöström M, Eriksson L. 2001. PLS-regression: a basic tool of chemometrics. Chemom Intell Lab Syst. 58(2):109–130. https://doi.org/10.1016/S0169-7439(01)00155-1..

[81] Nørgaard L, Saudland A, Wagner J, Nielsen JP, Munck L, Engelsen SB. 2000. Interval Partial Least-Squares Regression (iPLS): A comparative chemometric study with an example from near-infrared spectroscopy. Appl Spectrosc. 54(3):413–419. https://doi.org/10.1366/0003702001949500.

[82] da Silva DJ, Wiebeck H. 2017. Using PLS, iPLS and siPLS linear regressions to determine the composition of LDPE/HDPE blends: A comparison between confocal Raman and ATR-FTIR spectroscopies. Vibrational Spectroscopy. 92:259–266. https://doi.org/10.1016/j.vibspec.2017.08.009.

[83] Li P., Karunanidhi D., Subramani T., Srinivasamoorthy K. (2021). Sources and consequences of groundwater contamination. *Archives of environmental contamination and toxicology*, *80*(1), 1-10. https://doi.org/10.1007/s00244-020-00805-z

[84] Central Ground Water Board, Ministry of Water Resources, Govt. of India, Faridabad (2014): Concept note on geogenic contamination of ground water in India (with a special note on nitrate)

[85] https://punjab.gov.in/know-punjab/

[86] Gosal, G. 2004. Physical geography of the Punjab. JPS. 11(1):20.

[87] https://upload.wikimedia.org/wikipedia/commons/thumb/f/fa/Punjab_district_map.png/800px-Punjab_district_map.png

[88] Kothari C.R., Garg G. 2019. Research Methodology: Methods and Techniques. 4th ed. New Age International Publishers.

[89] Mardia K.V., Kent J., Bibby M. 1995. Multivariate Analysis. 10th ed. Academic Press Harcourt Brace & Company Publishers. ISBN 0-12-471250-9.

[90] Takeuchi K, Yanai H, Mukherjee B.N. 1981. The Foundations of Multivariate Analysis: A Unified Approach by Means of Projection onto Linear Subspaces.

[91] Camo. 2015. The Unscrambler X v10.3 User Manual. Oslo, Norway: Camo Software AS

[92] Jolliffe, I.T. 2002. Principal Component Analysis. Springer. ISBN 0-387-95442-2.

[93] Nørgaard L, Saudland A, Wagner J, Nielsen JP, Munck L, Engelsen SB. 2000. Interval Partial Least-Squares Regression (iPLS): A comparative chemometric study with an example from near-infrared spectroscopy. Appl Spectrosc. 54(3):413–419. https://doi.org/10.1366/0003702001949500.

[94] Mohammed Abdul Khaja Shameem, Sudheera Sammanthi Jayasinghe, Ediriweera P.S. Chandana, Channa Jayasumana, P. Mangala C.S. De Silva. 2015. Arsenic and human health effects: A review. Environ Toxicol Pharmacol. 40(3):828–846. https://doi.org/10.1016/j.etap.2015.09.016.

[95] Nguyen K.M., Nguyen B.Q., Nguyen H.T., Nguyen H.T.H. 2019. Adsorption of arsenic and heavy metals from solutions by unmodified iron-ore sludge. Appl. Sci. 9:619. https://doi.org/10.3390/app9040619.

[96] Virk, Hardev Singh. 2020. Groundwater contamination in Punjab due to arsenic, selenium and uranium heavy metals. Res Rev J Toxicol. 10(1):1–6..

[97] He Y, Huang M, García A, Hernández A, Song H. 2007. Prediction of soil macronutrients content using near-infrared spectroscopy. Comput Electron Agric. 58(2):144–153. https://doi.org/10.1016/j.compag.2007.03.011..

[98] Singh, S. K., Ghosh, A. K., Kumar, A., Kislay, K., Kumar, C., Tiwari, R. R., Parwez, R., Kumar, N., and Imam, M. D. 2014. Groundwater arsenic contamination and associated health risks in Bihar, India. Int. J. Environ. Res. 8(1):49–60.

[99] Nie P, Dong T, He Y, Qu F. 2017. Detection of soil nitrogen using near infrared sensors based on soil pretreatment and algorithms. Sensors. 17(5):1102. https://doi.org/10.3390/s17051102.

[100] Edward Raja, C. 2020. Assessment of physicochemical characteristics of groundwater collected from different taluks, Dindigul District, Tamilnadu, India. Environmental Research and Technology. 3(1):1–7. https://doi.org/10.35208/ert.658910

[101] Brown, William H., Iverson, Brent L., Anslyn, Eric, Foote, Christopher S. 2013. Organic Chemistry. Cengage Learning. ISBN 1133952844, 9781133952848.

[102] The Groundwater Project. 2020. The Importance of Groundwater. Available at: https://gw-project.org/the-importance-of-groundwater/.

[103] Sarkar, L. 2024. Groundwater contamination in perspective of different elements, water quality index and health implications. Curr World Environ. 19(1). DOI: http://dx.doi.org/10.12944/CWE.19.1.9.

[104] Kumar, Love, Deitch, Matthew J., Tunio, Imran Aziz, Kumar, Avinash, Memon, Sheraz Ahmed, Williams, Lauren, Tagar, Uroosa, Kumari, Ramna, Basheer, Sarosh. 2022. Assessment of physicochemical parameters in groundwater quality of desert area (Tharparkar) of Pakistan. Case Studies in Chemical and Environmental Engineering. 6:100232. https://doi.org/10.1016/j.cscee.2022.100232..

[105] IS 3025 (Part 10): Method of Sampling and Test (Physical and Chemical) for Water and Wastewater, Part 10: Turbidity (First Revision). Bureau of Indian Standards (BIS).

[106] IS 3025 (Part 11): Method of Sampling and Test (Physical and Chemical) for Water and Wastewater, Part 11: pH Value (First Revision). Bureau of Indian Standards (BIS). 1983. Available at: https://law.resource.org/pub/in/bis/S02/is.3025.11.1983.pdf.

[107] IS 3025 (Part 14): Method of Sampling and Test (Physical and Chemical) for Water and Wastewater, Part 14: Specific Conductance (Wheatstone Bridge, Conductance Cell) (First Revision). Bureau of Indian Standards (BIS). Available at: https://archive.org/details/gov.law.is.3025.14.1984.

[108] IS 3025 (Part 16): Methods of Sampling and Test (Physical and Chemical) for Water and Wastewater, Part 16: Filterable Residue (Total Dissolved Solids) (First Revision). Bureau of Indian Standards (BIS). Available at: https://archive.org/details/gov.in.is.3025.16.1984.

[109] IS 3025 (Part 26): Method of Sampling and Test (Physical and Chemical) for Water and Wastewater, Part 26: Chlorine, Residual (First Revision). Bureau of Indian Standards (BIS). 1986. Available at: https://law.resource.org/pub/in/bis/S02/is.3025.26.1986.pdf.

[110] IS 3025 (Part 21): Method of Sampling and Test (Physical and Chemical) for Water and Wastewater, Part 21: Hardness (Second Revision). Bureau of Indian Standards (BIS). 2009. Available at: IS 3025-21 (2009) PDF.

[111] World Health Organization. 2003. Total Dissolved Solids in Drinking Water, Background Document for Development of WHO Guidelines for Drinking Water Quality. WHO/SDE/WSH/03.04/16. Available at: https://iris.who.int/bitstream/handle/10665/44584/9789241548151_eng.pdf.

[112] R. S. Aswal, Mukesh Prasad, Narendra K. Patel, A. L. Srivastav, Johnbosco C. Egbueri, G. Anil Kumar, & R. C. Ramola. 2023. Occurrences, sources and health hazard estimation of potentially

toxic elements in the groundwater of Garhwal Himalaya, India. *Scientific Reports*, 13, 13069. https://doi.org/10.1038/s41598-023-40266-7.

[113] World Health Organization. 2003. Chlorine in Drinking Water, Background Document for Development of WHO Guidelines for Drinking Water Quality. WHO/SDE/WSH/03.04/45. Available at: https://www.who.int/water_sanitation_health/dwq/chlorine.pdf.

[114] World Health Organization. 2017. Water Quality and Health - Review of Turbidity: Information for Regulators and Water Suppliers. Technical Brief, WHO/FWC/WSH/17.01. Available at: https://www.who.int/publications/i/item/water-quality-and-health-review-of-turbidity-information-for-regulators-and-water-suppliers.

[115] World Health Organization. 2011. Hardness in Drinking Water, Background Document for Development of WHO Guidelines for Drinking Water Quality. WHO/HSE/WSH/10.01/10/Rev/1. Available at: WHO Document.

[116] Bureau of Indian Standards (BIS). 2012. IS 10500: Drinking Water — Specification (Second Revision). ICS 13.060.20.

[117] World Health Organization. 2021. Guidelines for Drinking-water Quality: Fourth Edition Incorporating the First and Second Addenda. ISBN 978-92-4-004506-4. Available at: https://www.who.int/publications/i/item/9789240045064.

[118] Karki, D., & Thapa, Y. N. 2022. Assessing Physico-Chemical Parameters of Drinking Water in Majkhola, Tansen, Palpa. International Journal of Applied Sciences and Biotechnology, 10(1), 60-70. DOI: 10.3126/ijasbt.v10i1.44161.

[119] International Association of Hydrogeologists. "Groundwater - More About the Hidden Resource." Available at: https://iah.org/education/general-public/groundwater-hidden-resource.

[120] Seametrics. "Facts About the World's Shrinking Groundwater Resources." Available at: https://www.seametrics.com/blog/groundwater/.

[121] Li, P., Karunanidhi, D., Subramani, T., & Srinivasamoorthy, K. 2021. Sources and Consequences of Groundwater Contamination. *Archives of Environmental Contamination and Toxicology*, 80, 1–10. https://doi.org/10.1007/s00244-020-00805-z.

[122] Wu, J., Zhang, Y., & Zhou, H. 2020. Groundwater chemistry and groundwater quality index incorporating health risk weighting in Dingbian County, Ordos basin of northwest China. *Geochemistry*, 80(4), 125607. https://doi.org/10.1016/j.chemer.2020.125607.

[123] Gibson, J. M., & Kavanaugh, M. C. 1994. Restoring Contaminated Groundwater: An Achievable Goal? *Environmental Science & Technology*, 28(8), 349A-355A. https://doi.org/10.1021/es00057a001. .

[124] Subba Rao, N., Ravindra, B., & Wu, J. 2020. Geochemical and health risk evaluation of fluoride rich groundwater in Sattenapalle Region, Guntur district, Andhra Pradesh,

India. *Human and Ecological Risk Assessment: An International Journal*, 26(9), 2316–2348. https://doi.org/10.1080/10807039.2020.1741338.

[125] Bokobza, L. 1998. Near Infrared Spectroscopy. *Journal of Near Infrared Spectroscopy*, 6, 3-17. https://doi.org/10.1255/jnirs.100.

[126] Ring, E. F. J. 2000. The discovery of infrared radiation in 1800. *The Imaging Science Journal*, 48(1), 1–8. https://doi.org/10.1080/13682199.2000.11784339.

[127] Kaye, W. 1954. Near-infrared spectroscopy: I. Spectral identification and analytical applications. *Spectrochimica Acta*, 6(4), 257-E2. https://doi.org/10.1016/0371-1951(54)80011-7.

[128] LibreTexts. "Infrared Spectra of Some Common Functional Groups." In Organic Chemistry (Wade). Available at:
https://chem.libretexts.org/Bookshelves/Organic_Chemistry/Map%3A_Organic_Chemistry_(Wade)/11%3A_Infrared_Spectroscopy_and_Mass_Spectrometry/11.05%3A_Infrared_Spectra_of_Some_Common_Functional_Groups.

[129] Workman, J. J. 1996. Interpretive Spectroscopy for Near Infrared. *Applied Spectroscopy Reviews*, 31(3), 251–320. https://doi.org/10.1080/05704929608000571.

[130] Christy, A. A., Kasemsumran, S., Du, Y., & Ozaki, Y. 2004. The detection and quantification of adulteration in olive oil by near-infrared spectroscopy and chemometrics. *Analytical Sciences*, 20(6), 935–940. https://doi.org/10.2116/Analsci.20.935.

[131] Luypaert, J., Massart, D. L., & Vander Heyden, Y. 2007. Near-infrared spectroscopy applications in pharmaceutical analysis. *Talanta*, 72(3), 865-883. https://doi.org/10.1016/j.talanta.2006.12.023.

[132] Cleve, E., Bach, E., & Schollmeyer, E. 2000. Using chemometric methods and NIR spectrophotometry in the textile industry. Analytica Chimica Acta, 420(2), 163-167. https://doi.org/10.1016/S0003-2670(00)00888-6.

[133] Ozaki, Y. 2012. Near-Infrared Spectroscopy—Its Versatility in Analytical Chemistry. *Analytical Sciences*, 28, 545–563. https://doi.org/10.2116/analsci.28.545.

[134] Inagaki, T., Yonenobu, H., Asanuma, Y., & Tsuchikawa, S. 2018. Determination of physical and chemical properties and degradation of archaeological Japanese cypress wood from the Tohyamago area using near-infrared spectroscopy. Journal of Wood Science, 64, 347-355.