QUALITY ASSESSMENT OF GROUND WATER OF WESTERN MALWA REGION, PUNJAB USING NEAR INFRARED SPECTROSCOPY AND CHEMOMETRICS

Thesis Submitted For the Award of the Degree of

DOCTOR OF PHILOSOPHY

in

Physics

By

Reena Rani

41900387

Supervised By   **Co-Supervised by**

Dr. Uma Kamboj (21636)   **Dr. Neha Munjal(18869)**

Assistant Professor   **Associate Professor**

Department of Physics   **Department of Physics**

Lovely Professional University   **Lovely Professional University**

LOVELY PROFESSIONAL UNIVERSITY

PUNJAB

2023

DECLARATION

I, hereby declared that the presented work in the thesis entitled "**QUALITY ASSESSMENT OF GROUND WATER OF WESTERN MALWA REGION, PUNJAB USING NEAR INFRARED SPECTROSCOPY AND CHEMOMETRICS"** in fulfilment of degree of **Doctor of Philosophy (Ph. D.)** is outcome of research work carried out by me under the supervision. **Dr. Uma Kamboj**, working as Assistant 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 Supervisor)

Name of the scholar: Reena Rani

Registration No.: 41900387

Department/school: 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 "**QUALITY ASSESSMENT OF GROUND WATER OF WESTERN MALWA REGION, PUNJAB USING NEAR INFRARED SPECTROSCOPY AND CHEMOMETRICS"** submitted in fulfillment of the requirement for the reward 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 **Reena Rani**, **41900387**, 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]   [Signature of Co-Supervisor]

Name of supervisor: Dr. Uma Kamboj   Name of Co-Supervisor: Dr. Neha Munjal

Designation: Assistant Professor   Designation: Associate Professor

Department/school: School of Chemical   Department/school: School of Chemical

Engineering and Physical Sciences   Engineering and Physical Sciences

University: Lovely Professional   University: Lovely Professional

University, Punjab, India   University, Punjab, India

Abstract

Near infrared spectroscopy (NIRS) is successfully used in field of agriculture, food products and beverage industry. Once the model is developed and validated using NIRS, it is used for qualitative and quantitative analysis of the products. It is important to maintain the robustness of the developed chemometrics model towards instrumental and environmental effects. Spectral data along with reference data can be successfully used to build a regression model using multivariate techniques to analyze the sample qualitatively and quantitatively.

In the present PhD work, main aim was to analyze the quality of the ground water samples from the two districts of Malwa region of Punjab for presence of urea and arsenic content in them. Further the objective was to analyze the characteristics of the water sample quantitatively by measuring their pH, total dissolved solids (TDS), electrical conductivity (EC) and chloride content. All the measurements for the physicochemical parameters were done from a soil testing lab at Jind, Haryana. The physicochemical data of the ground water samples was then used to classify the samples collected on the basis of their region – urban (U), rural (R) and industrial (I), on basis of the source of collected water samples – submersible (SS), waterbox (WB) and handpump (HP), on basis of their depth – less than 100ft and more than 100ft using the principal component analysis (PCA) from clustvis online open software and CAMO Unscrambler X10.5 software at Council of Scientific and Industrial Research (CSIR) – CSIO, Chandigarh.

To check the dependency of the four parameters of ground water samples and to develop a general equation for their measurement, correlation analysis was run on the physical and chemical parameters. Analysis of Variance (ANOVA) was also performed on the reference (physicochemical) dataset to check if there was any significant difference between the samples on the basis of the four properties. The correlation analysis showed relation between the three parameters – TDS, EC and chloride content with correlation coefficient greater than 0.8. As there was good correlation observed between these parameters, bivariate linear regression equations were developed to build a relation which could help the determination of other two parameters if any one parameter if known from EC, TDS and chloride content.

The use of reference data to develop a relation between TDS, pH and EC of the collected ground water samples, helped in establishing an equation where only using any one parameter all the parameters can be determined. This was useful as it resulted in time saving and environmental friendly. As now there was need to only measure one parameter from the chemical analysis and rest two parameters can be calculated using the equation.

The literature review also revealed that presence of Uranium, urea and other pollutants can be estimated using the values of TDS, EC and depth of the ground water sample. Due to this reason analysis was done to see the dependency of depth on the values of TDS, EC and chloride. It was observed that the ground samples which were collected from depth greater than 100ft had their parametric values in the permissible range according to international standards of world health organization (WHO) and national standards of Bureau of Indian Standards (BIS).

The spectral data was acquired from the near infrared (NIR) spectrometer and Fourier transform infrared spectrometer (FTIR) from the central instrument laboratories at Council of Scientific and Industrial Research (CSIR) – CSIO, Chandigarh and Lovely Professional University, respectively. PCA was used to classify the ground water samples using the NIR spectral data on the basis of district – Ferozepur, Fazilka; region –I,U,R; source of sample collected – SS, WB, HP and depth – less than 100ft and more than 100ft.

Next objective of the Ph.D. work was to identify the pollutants – urea and arsenic tri oxide in the samples. FTIR spectra of pure urea and arsenic trioxide was obtained and compared with the FTIR spectra of the ground water samples for the identification of the characteristic bond. The FTIR peak data for urea and arsenic trioxide was also compared with the NIR spectra of the samples to identify the 1st overtone or higher harmonics if any in the samples. It was observed that most samples were found to have few bands / peaks, overtones corresponding to the signature wavelengths of Urea.

Last objective focused on developing a regression model using partial least square (PLS) regression to determine the quality parameters of the ground water samples quantitatively. Regression model was applied after pre-processing of the spectral data to reduce noise and scattering of the samples. PLS model was first developed from

wavelength range 700-2500nm. Score and loading plot of the PLS model were used to check the relation between the samples and dependency of wavelengths with the parameter, respectively. Further it was observed that some wavelengths were not important for the prediction of these parameters and had very less value of coefficients.

To overcome this difficulty the spectral data set from 700-2500 nm was divided into intervals of equal width by keeping intervals of 100 nm, 200nm, 300nm and 600nm. Regression models were developed for all the parameters for the four set of data set and the models were compared on the basis of coefficient of determination (R^2) and root mean square error (RMSE) for calibration and prediction. In all 16 iPLS regression models were developed. The wavelengths sets with low values of RMSE for calibration and prediction were noted for further analysis.

To further enhance the robustness of the model, the interval sets with values of r^2 higher than 0.8 were combined together irrespective of their interval and width for all the parameters. This method is known as synergy interval PLS. The regression model with this selective wavelength combination was compared and the wavelength set with highest R^2 values was finally selected and the model was saved for future industrial use.

The complete work focused on identification of bonds relative of pollutants, classification of the data set on the basis of region, source and depth and developing a regression model with selected important wavelengths to predict the physicochemical parameters of ground water samples. The research work was able to successfully select a set of wavelengths from the complete spectral data set, to predict the parameters. This model developed will help to reduce the analysis time, experimental time and cost of the instrument. The cost of the instrument is decreased as now there will be requirement of only selected filters of important wavelengths rather than the complete monochromator from 700-2500nm. The developed regression model can be successfully used to predict the values of pH, TDS, EC and chloride content of the ground water samples of Ferozepur and Fazilka district with good accuracy.

Finally, I would again like to thank God, for letting me through all the difficulties. I have experienced your guidance day by day. You are the one who let me finish my degree. I will keep on trusting you for my future.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS

WHO	World health organization
BIS	Bureau of Indian standards
TDS	Total dissolves solids
EC	Electrical conductivity
EM	Electromagnetic
NIRS	Near Infrared Spectroscopy
FTIR	Fourier Transform Infrared spectroscopy
LED	Light emitting diodes
InGaAs	Indium Gallium Arsenide
PbS	Lead Sulphide
PLS	Partial least Square
PLSR	Partial least Square Regression
PCA	Principal Component Analysis
PCR	Principal Component Regression
MLR	Multiple Linear Regression
SVM	Support Vector Machine
PCs	Principal Components
$RMSE_V$	Root mean square error of validation
$RMSE_C$	Root mean square error of calibration
$RMSE_P$	Root mean square error of prediction
R	Coefficient of correlation
R^2	Coefficient of determination
iPLS	Interval Partial least Square
siPLS	Synergy interval Partial least Square
Cl	chloride content
HP	Hand Pump
SS	Submersible
WB	Water Box
ANOVA	Analysis of Variance
SPSS	Statistical Package for Social Sciences

Chapter 1 – Introduction

1. Chapter 1 - Introduction

Spectroscopy is the field of physics that uses the absorption and emission spectra of the element – a molecule or an atom, when they are incident by electromagnetic spectrum and throw light on the material. It provides information about the arrangement and motion of the outer electrons, their angular momentum, magnetic momentum, distribution of charge and magnetism of nucleus.

Sir Isaac Newton originated spectroscopy by showing that a prism refracts low wavelength light more than the higher wavelength radiations, thus forming a band of colors known as 'spectrum'. The complete spectrum ranges from the Gamma Rays to Radio waves. The wavelength λ and the frequency ν of a monochromatic wave is related as,

Equation 1.1 Relation between frequency and wavelength of electromagnetic radiation

$$\lambda \nu = v \qquad \text{------------------------- [1.1]}$$

Where, v is the phase velocity in the medium.

The Electromagnetic Spectrum:

The electromagnetic [EM] spectrum starts from the smallest wavelength, gamma rays to the longest wavelength, radio waves, including ultraviolet, visible, infrared and microwave radiation. Radiation is the energy which transvers when it passes through an element / compound / molecule. When this radiation penetrates the element, transitions from electronic to vibrational to rotational depending upon their energy bands.

Figure 1.1   shows the details of the every part of the electronic radiation spectrum with their wavelength, frequency and their specific regions.
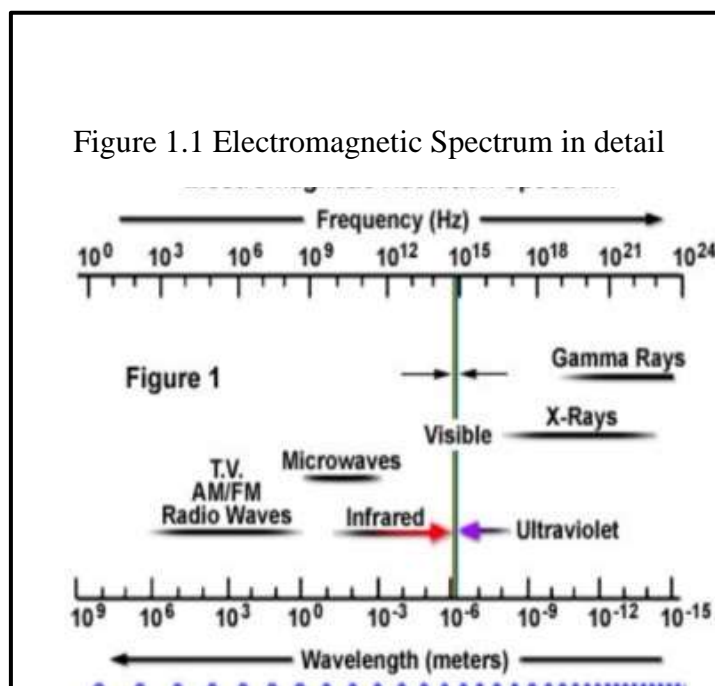
Figure 1.1 Electromagnetic spectrum showing all the radiations with their wavelengths (nm)

The electromagnetic spectrum varies from longest wavelength to shortest. [Fig.1]. [Credit: NASA's Imagine the Universe]

- Radio: Radio waves are the radiations which are emitted by radio station and many astronomical objects. There frequency is less than 3×10^{11} Hz.

- Microwave: Microwave radiation ranges from 3×10^{11} to 3×10^{13} Hz and are mostly used in microwaves for cooking and heating the food materials. Microwaves are also used for studying astronomical objects.

- Infrared: Infrared light ranges from 10^{13} to 10^{14} Hz. These radiations are used for heating purposes, night vision measurements by observing the infrared light emitted by our body, studying astronomical objects.

- Visible: The region which has biggest astronomical windows and our eyes are capable of observing this radiation.

- Ultraviolet: Ultraviolet radiation is blocked by the atmospheric window. Thus for its study and observation, instruments are placed

- X-ray: The radiation with frequency far higher than radiowaves are used for ortho measurements and many other medical uses.

- Gamma ray: Gamma rays are having highest frequency and these rays are used for medical purposes.

Spectroscopy can be divided into two fields - Atomic Spectroscopy and Molecular

Spectroscopy

Atomic spectroscopy: Atomic spectroscopy is the spectroscopy, where the emission and absorption of the radiation study the atom's structure, chemical composition and bonding when electromagnetic radiation is incident on them. In atomic study, when atom absorbs the energy the electrons get excited and jump to the higher excited state and due to instability of electrons in excited state these return in ground state and emission of energy occurs as a result considered in spectrum. This helps in the study of the structure of atoms [Fig.2].
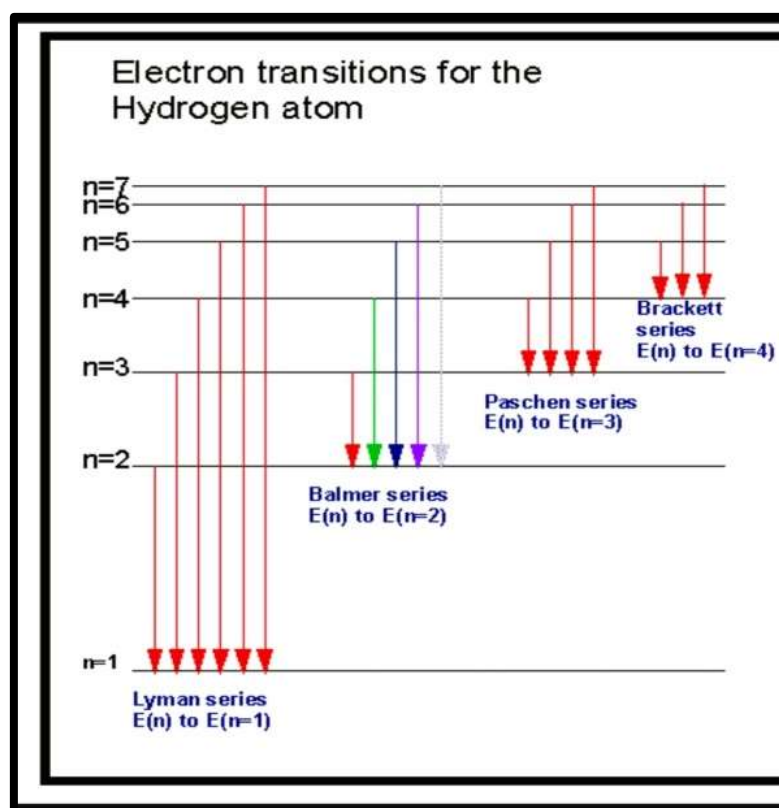
Figure 1.2 Electronic Transitions for the Hydrogen atom

Molecular spectroscopy

In molecular spectroscopy the radiations of the electromagnetic spectrum are incident on the molecule and the absorption spectrum is observed which helps in identifying the structural and compositional information. The emissions and absorption spectra of molecules arise from the transition between the allowed states of energy of the molecule. In the study of molecules, the energy is absorbed by molecules as result in

electronic, vibrational and rotational transitions which are depending upon the amount of energy absorbed. Electronic, vibrational and rotational energy levels of molecule are known as molecular energy levels. The transitions of energies taking place between these levels give the result in the molecular spectrum.

Types of Molecular Spectra

1. Electronic Spectra: These spectra observed both in emissions and absorption in the visible and ultraviolet. Each spectrum consists quite a large number of bands.

2. Vibrational Spectra: Vibrational molecular spectra arise from transition between vibrational energy state associated with same electronic state of the molecule and observed in near infrared region of electromagnetic spectrum. Vibrational spectra obtained only for molecule having permanent dipole moment such as hetro nuclear diatomic molecules [Fig. 3].

3. Rotational Spectra: Rotational molecular spectra arise from transition between rotational energy states and observed in far infrared or microwave region of the electromagnetic spectrum. Only molecule that has permanent electric dipole moments can observe or emit electromagnetic radiation in such transitions

Raman Spectra: Raman spectra are observed when a beam of radiation is incident on a material in any state, and the light radiation is then scattered in different directions. The spectrum of the light scattered consists of lines of frequency of similar frequency to the incident beam [Rayleigh lines] and also consists of some weak lines of different frequencies. These lines of new frequencies other than incident are known as Raman Lines [Fig4.].
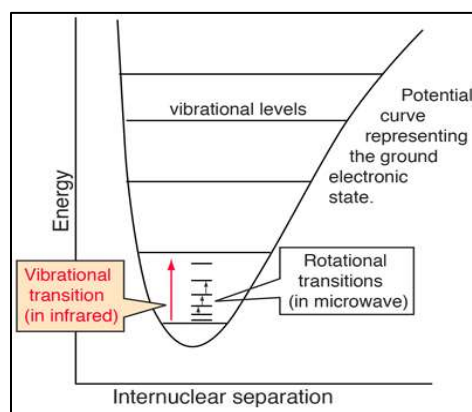
Figure 1.3 Vibrational and Rotational bands of a molecule
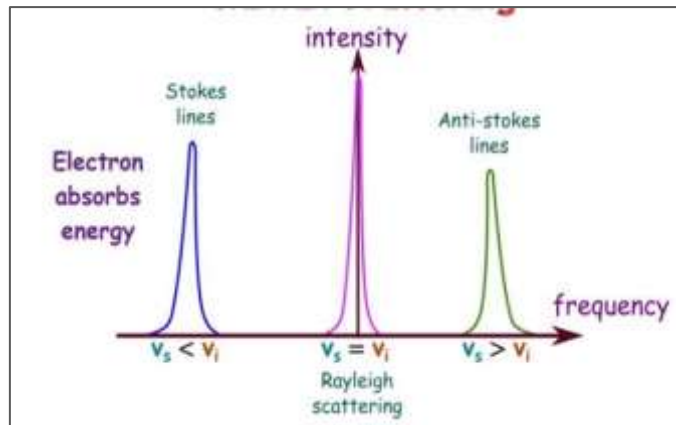
Figure 1.4 Rayleigh Scattering

Near Infrared Spectroscopy – Introduction and Principle

Near-infrared spectroscopy [NIRS] is a technique in spectroscopic measurements where near infrared radiation [from 780 nm to 2500 nm] is used to analyze the food products and beverage qualitatively and quantitatively. NIR spectroscopy occurs when there is interaction between a sample [e.g. medicinal, millets, beverages, cereals, oils]. Here each wavelength which is dispersed using a prism, monochromator or diffraction grating is incident on the sample and respective absorption / transmittance spectra is observed to understand and study the structure of the element (Fig .)
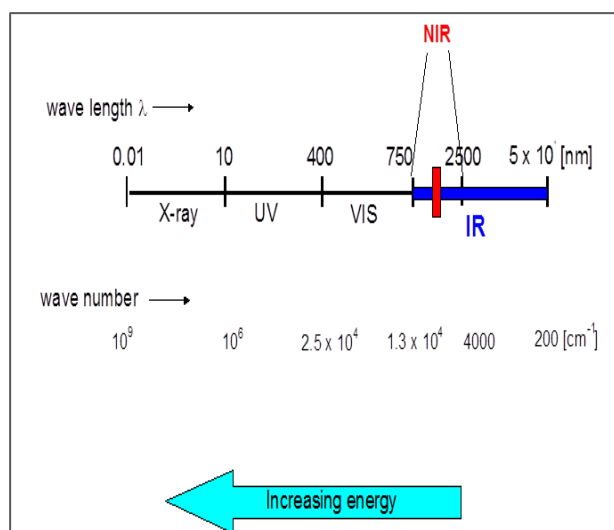
Figure.1.5 Near Infrared Spectrum in the Electromagnetic Radiation

Near Infrared spectroscopy is capable of measuring the composition of organic compounds with the use of transmittance or reflectance measurements of the sample at different wavelengths [1]. Depending on the energy absorption by functional group in the sample, intensity and number of absorption bands or peaks varies. NIR is capable of measuring the overtones and combinational bands for the study of hydrogen bonds- namely N-H, S-H, C-H, O-H. Further NIR is also capable of measuring the concentration of different structures – pH, water, protein, Electrical conductivity, fiber, carbohydrates etc. [2]. NIR light can pass readily through organic materials which can be measured using reflectivity. As the NIR light is not capable of electronic transitions, thus the technique is non-destructive, nonhazardous and rapid in nature. The absorption of radiation by an organic compound in the NIR region causes molecular vibrations. The basic principle of NIRS is based the harmonic vibrations and their combination produced by the molecules. Fundamental and overtone vibrations are evaluated by anharmonicity of the diatomic molecule. The first overtone exists in the NIR region because the X-H functional group absorbs at wavenumbers must be greater than $2000cm^{-1}$ [3]. Bands are becoming weaker in intensity towards shorter wavelengths. The lower NIR intensities imply that solid samples do not need dilution and the impacts of non-linearity are less probable owing to powerful absorption. The energy [per photon] of infrared light is very less as compared to visible, ultraviolet and x-rays. Infrared light is therefore relatively safe since its photons cannot break down molecules.

NIRS has been widely used in the chemical, pharmaceutical, food sectors, forestry and agricultural sectors, water quality assessment, food products, cereals, grains and beverages.

Instrumentation

Near-IR [NIR] spectroscopy instrumentation is nearly comparable to those for the ultra violet and mid-Infra-Red ranges. A near infrared spectrometer consists of a light source (which emits radiation in near infrared region), a dispersive element (monochromator or diffracting grating), sample holder and a detector (mainly silicon, lead sulphide and indium gallium arsenide).
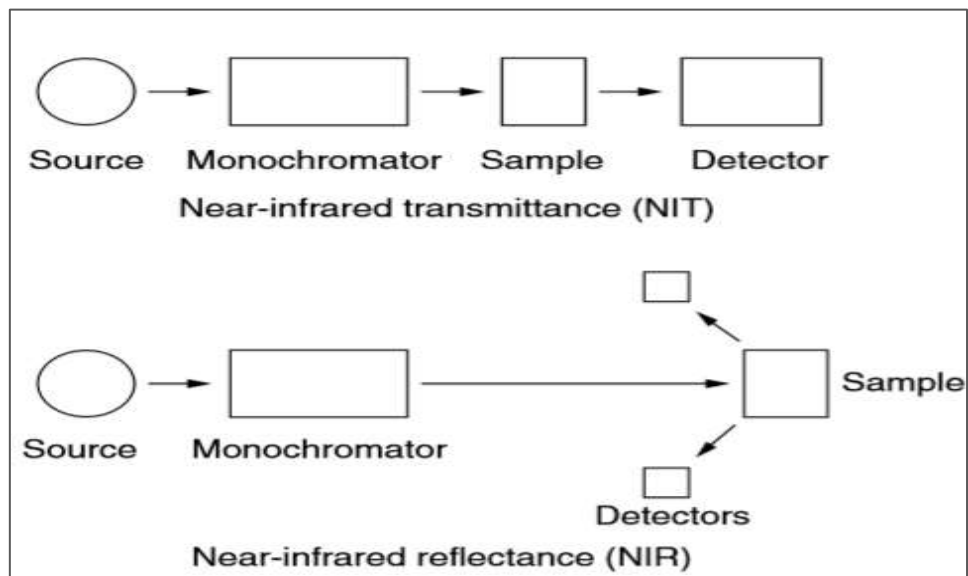
Figure 1.6 Basic design of Near Infrared Spectrometer

Source: [4]

Fig 1.6 shows that basic diagram of Near Infrared Spectroscopy which consist of transmittance and reflectance mode of collecting the spectra. The difference between both designs can appear when assembling the spectroscopic measurements. In the reflectance measurements, the light penetrates only $1\text{-}4\times10^{-3}$m from the surface of the sample. When the energy of small penetration power passes into the sample it will bring about greater variation then transmittance technique [when measuring non-homogeneous sample]. The path length of the sample in the transmittance measurement is investigated into spectrum measurements. Due to transmittance technique measurement can be done on large particles [Fig.7].
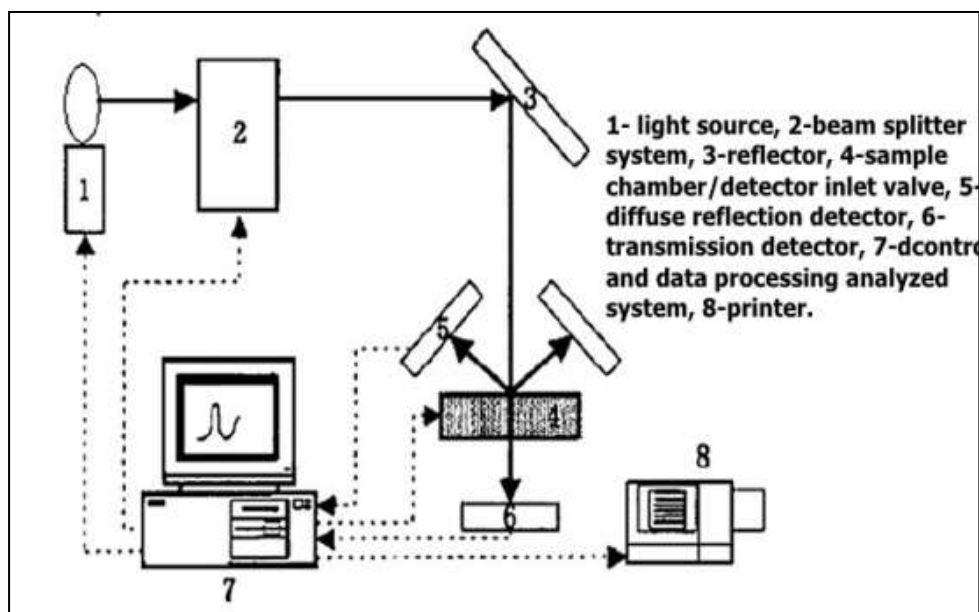
Figure 1.7 Block diagram of NIR spectrometer

For transmittance measurements, particle size can be small but it should not be too small, that signals do not get detected by the detectors. For analytical measurements, the common sources used are incandescent or quartz halogen light bulbs as they emit in the near infrared region with high transmittance. Light-emitting diodes [LEDs] in the required range have a larger lifetime and spectral stability. The light from the source is incident on the sample either directly or after getting dispersed from dispersive element. Depending upon the placement of the dispersive element, spectrometers may be pre dispersive or post dispersive. The dispersive components consist of gratings, monochromator or prism.

The light incident on the sample is either transmitted or absorbed by the sample itself, and sample undergo different vibrational transitions depending upon the bonds present in the sample. The sample may undergo different transitions from ground to excited state [1st overtone], ground state to third excited state [2nd overtone] transition for different dipole active bonds present in the sample.

Reflected / transmitted light from the sample is detected by the detector. Detectors are used to check the absorbance of the sample at a particular wavelength. Silicon-based CCDs are suitable for detection in the range of 700-1100 nm. InGaAs and PbS devices are used for the detection in the higher range in NIR. In some NIRS instruments, both Si based and Indium Gallium Arsenic detectors are employed in the

9

single instrument so that detection in the range of UV, Visible and NIR can be detected simultaneously. Absorbance is calculated from the transmitted or the reflected spectra obtained using Lambert's Beer Law. The range of Near Infrared instrument depends upon the source and the detector used for the instrumentation process. The absorption spectra are used along with the reference data to analyse and interpret the characteristics qualitatively and quantitatively with the help of chemometrics.

Applications of NIR Spectroscopy in Qualitative and Quantitative Analysis

NIR spectra along with chemometrics is a rapid, non destructive technique used for the analysis of food products, pharmaceuticals, beverages, combustion products and a major branch of astronomical spectroscopy. It has various applications in the field of astronomy, medical industry, food and beverage industry, agriculture, remote sensing, material science.

Chemometrics

Near Infrared Spectra of the samples yield information about the hydrogen bonds in the sample for their characterization. As the NIR spectra consists of overtones and combinational bands, the spectra is quite complex and is not easy to analyze. For analysis of the spectra, it is considered as a secondary data and the physical, chemical, biological and quality parameters of the sample act as primary data for further analysis. To analysis the bands in NIR spectra for qualitative and quantitative analysis, there is need of multivariate analysis. Multivariate analysis helps to identify, quantify and predict the quality of the sample with the use of partial least square [PLS] regression, multiple linear regression [MLR], principal component regression [PCR] and support vector machine [SVM] learning techniques. Chemometrics is used to analyse large amount of chemical data to extract the qualitative and quantitative information of the sample with use of statistical and mathematical methods. Chemometrics helps in bridging the chemical / physical information of a sample with its spectral measurements. Due to its wide applications it is an important part of modern research. Chemeometric software packages have been widely used by industry, researchers, engineers, academicians of all fields for qualitative and

quantitative analysis and development of regression models to predict ensure and classify the raw material in a nondestructive way in real time analysis [5]. The presence of the bands in the sample can be estimated by checking the absorbance at different wavelengths and comparing the same with the standard band chart for NIR absorbance (Fig 1.8). The chart shows that if the absorbance at a particular wavelength or band is observed, then the bond and its harmonics can be estimated in element.
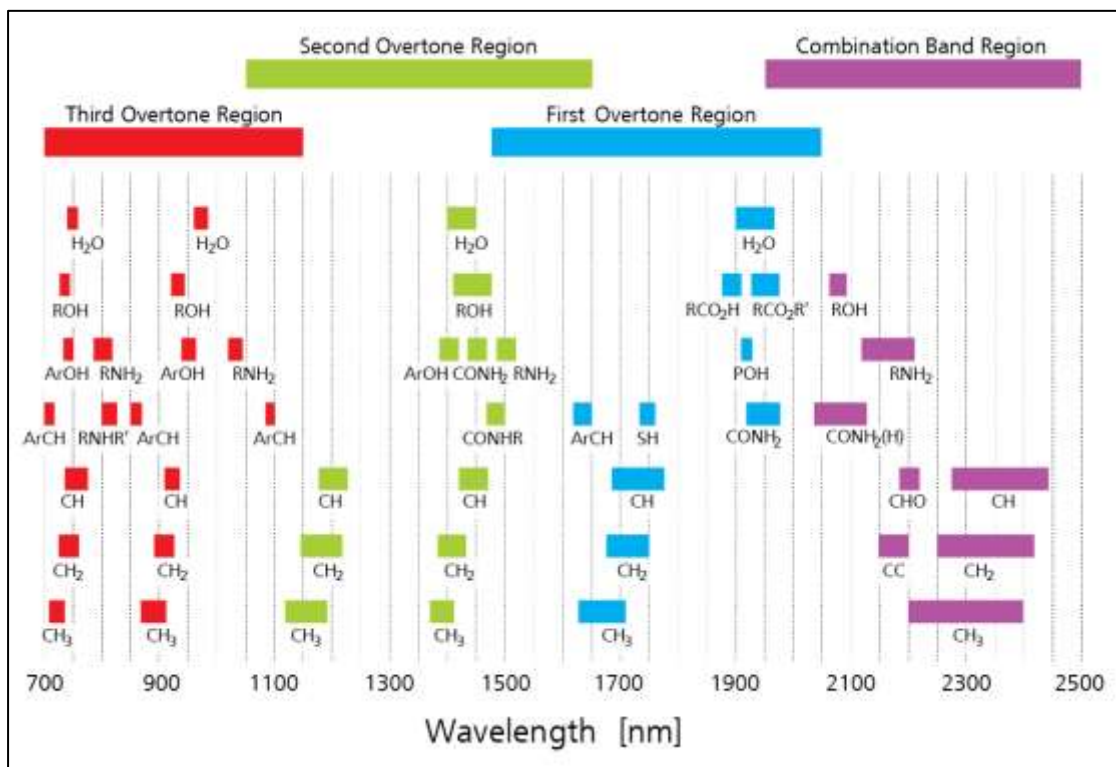
Figure 1.8 Near Infrared absorption band chart showing the first, second, third overtone regions and combinational bands.

Source [1]

Regression Methods

Regression methods are used to find the relationship between variables and samples. It is used when there is need to find the cause and effect relationship between different variables. For example to find the relation of effect of intake of calories on the body weight of the person, or check the effect of concentration of medicine on relaxation of a disease. To understand such relation, researcher performs regression

analysis to find the relation between the samples and variables and estimate the effect on the variables in term of quantity of a chemical. Regression analysis helps in determining the degree and strength of correlation between the variables.

Principal Component Analysis

Principal component analysis [PCA] is used to build linear models of large data set. PCA models use orthogonal vectors [loadings] to develop the model. Aim of PCA is to avoid noise associate with unwanted or outlier samples / variables, reduce the dimensions of complex data set and minimize the random error effect. With the use of new set of variables known as principal components [PCs] PCA is able to explain the covariance of data. The PCs are linear combinations of original data which help us in developing the relations between the samples [scores] and variables [loadings]. To analyse the big data set, PCA is the first step, followed by clustering / classification and other chemometrics techniques – PLS, PCR, MLR, SVM. The first PC refers to the direction where the projected data set have maximum variance. The second PC is orthogonal to the first and explains the maximum variance among the data points which are projected on it. This way the whole projected data set is covered using different PCs. Data reduction based on PCA helps in identifying the relationships between samples and variable and also helps in determining the outlier in the sample data set [6].

Partial Least Square Regression

Partial Least Square Regression is one step further to PCA, to find a coordinate system which is advantageous over the system used by PCA and PCR. In PLS factors are determined for both the spectral and reference / chemical data. It first uses the capability of PCA to remove the noise from the reference and the spectral data. PLS rotates the vectors until they are aligned so that a noise free spectral is obtained. In addition to this new coordinate axes for spectral data [x- variables] and reference data [y variable] is obtained.

PLS have three different steps in its procedure – calibration, validation and prediction. Calibration and validation involve development of a regression model using the spectral and the reference data. The model's accuracy is determined by the factors namely – root means square error of calibration [RMSEC], root mean square of

validation [RMSEV], coefficient of determination [R^2]. The model is further used to predict the parameters for the new samples using their spectral information.

Spectral Preprocessing

The application of the chemometric in NIR spectra is affected negatively due to the presence of unwanted variation in the data set. These effects may occur due to small variations in the properties of the sample- texture, sample size, temperature etc. or some parameters of the NIR spectrometer.

To remove the unwanted data form the spectral data, pre-processing techniques are applied prior to the regression analysis on the data set.

Different pre-processing techniques are

Mean Centering: - Mean centering helps in translating the data to the origin of chemometric space where analysis will be performed. This process removes the need of intercept from the regression model [7].

Smoothening: - This preprocessing technique helps in removing signal to noise ratio from the spectral data. Running average for a small window of points is calculated to apply the smoothing. Smoothened spectra are obtained by taking the average of the window.

Derivatives: - Derivate preprocessing helps in removal of baseline offsets, because the derivate of the constant is zero. Derivative of a digitized curve is easy to be approximated by statistical methods to get rid of baseline offsets.

Normalization: - In this every spectrum is normalized to fixed area, which helps in removal of the effect of fluctuating signals. Normalization can be of different types – maximum normalization, mean normalization, area normalization, the spectrum values are set from 0 to 1 in each case. The normalization applied depends upon the need of the application.

Multiplicative Scatter Correction and Standard Normal Variate:- Multiplicative Scatter Correction [MSC] is used to correct the error due to baseline offsets and particle size variance. In Standard Normal Variate [SNV] mean of each spectrum is subtracted from the data and then normalized to 1, the results produced in both cases are almost same and many times difficult to analyze.

Water

Water is a transparent, odorless, colorless liquid that build the seas, rivers, lakes.it is very basic needs for all living organism. Chemical formula of water is H_2O, It is essential for most plant and animal life and the most widely used for all solvents. Freezing point of water is 0°C [32°F], boiling point is 100°C [212°F].

Ground Water

Ground water is main component of complete cycle of water. Infiltration of water into ground occurs by rainfall water from stream flow, reservoirs and lakes, which are natural resources and by surfaces irrigation which is artificial resource. Under the effect of gravity water move into the ground [Fig 1.9].
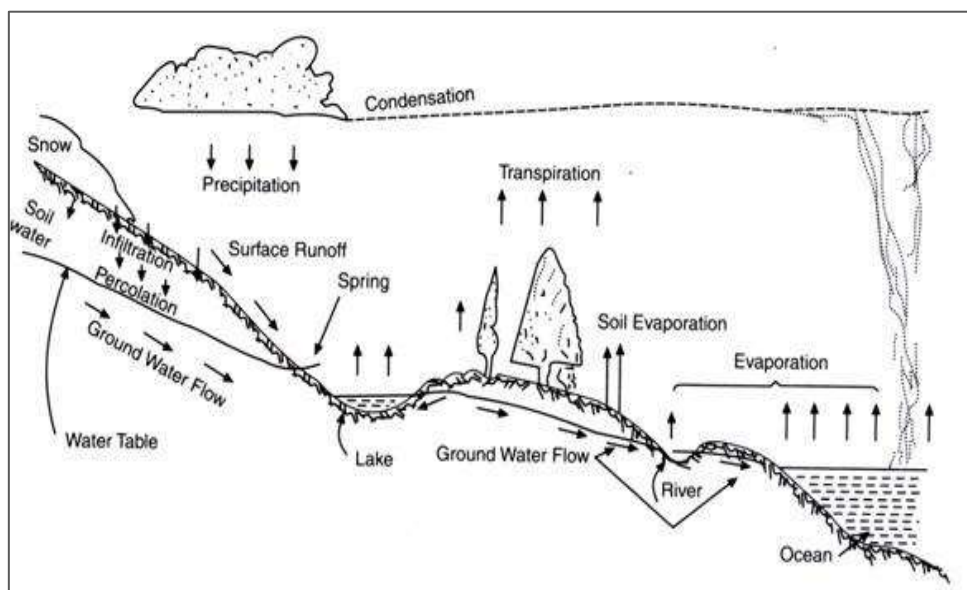
Figure 1.9 Process of Ground Water generation

Source: [4]

Difference between surface water and ground water:

Water which is available on surface of earth and is available readily for example-river, lake, pond is called surface water, whereas the one which is under the earth's surface is called ground water [Fig 1.10]. Ground water is used in drinking, cooking purposes but the surface water is mainly used for washing, agriculture and industrial purposes. Both the ground and surface water is also used for irrigation. Due to the

availability of surface water on the exposed area of earth, there is variation in its temperature but the ground water, which is at a depth, normally maintains same temperature throughout the year.

Ground water contamination:

Groundwater is the main source for drinking and cooking purposes. It's also used for irrigation purpose. Ground water is used by installing hand pumps, submersible motors. Though at a depth, ground water gets polluted as lot of industrial waste and other waste elements are dumped in the earth. The contaminants and harsh chemicals from these waste is absorbed by the soil and affects the quality of groundwater. The effect can be observed my measuring the characteristics parameters of the drinking water and comparing them with the national and international standards which are given by WHO and BIS. The dumped waste mainly contains oil, harmful chemicals, petrochemicals which are not suitable for human and animal health. The fertilizers and pesticides used for irrigation are also absorbed by the soil and lead to polluting the ground water. Materials from the land's surface can move through the soil and end up in the groundwater.

Types of Pollutant

Contaminants which pollute the ground water are the one which are absorbed by the soil because of industrial, human and agricultural waste. The contaminants can be organic, chemical, radioactive and bacteriological in nature.

Main pollutants which are harmful for human health and are found mostly in ground water are arsenic, uranium, fluoride, pathogens carried by human waste, nitrate components and some organic elements. Ground water is also polluted by presence of metals because of dumping of industrial waste by the industries. The metals whose composition is harmful are lead, cadmium, chromium. The medical waste dumped also lead to contaminating the ground water.
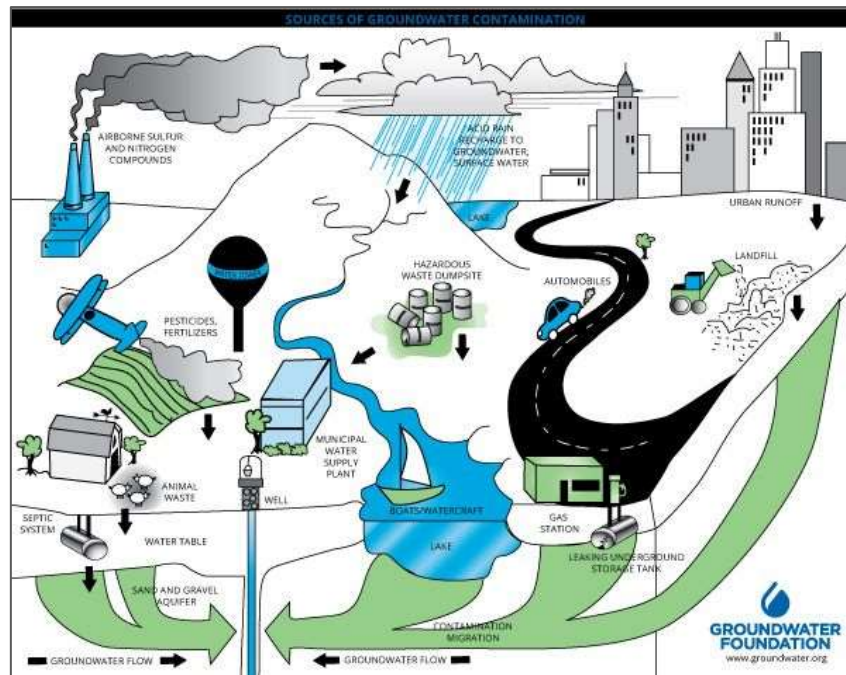
Figure 1.10 Ground water Foundation

Characterization of Ground Water

Groundwater is a valuable natural resource that plays a vital part in the economy. Ground water is majorly used for agriculture, industrial and domestic purpose. About 2 billion people in world do not have safe drinking water and more than twice of 2 billion do not have safe sanitation according to UNESCO [2018] [3]. Around 15% of the industry requirement for water is met by groundwater, for domestic purpose 65% requirement is fulfilled by groundwater and for irrigation 20% demand of water is met by groundwater [8]. It is critical to achieve a harmonious interaction between human activities and water resources in developing countries with high economic progress. However, in developing countries, where groundwater has long been an essential source of water, are experiencing major groundwater pollution concerns, owing primarily to human activity. Groundwater is affected by human activities such as urbanisation, industrialization, and agriculture.

Now days, the availability of safe drinking water availability is a major challenge for the living beings. Due to rapid industrialization, urbanization and unsustainable agricultural practices, the quality of ground water is below the

acceptable levels as mentioned by the concerned regulatory bodies [9]. According to a previous study, the groundwater of Malwa region of Punjab, India is highly contaminated with NO_3. The 93% of samples collected from this highly populated region of the India presented unacceptable and high level of NO_3 content. In this region, the groundwater is a major source of water for the rural population. Thus, contamination of water may cause serious health issue in of the rural population of this region [10]. The main anthropogenic source in this region was industrial wastage, combusted coal, use of excessive fertilizers and pesticides in agricultural lands. Due to dissolve of fluorine from fertilizer, water becomes highly alkaline. Due to infiltration of this alkaline water, ground water becomes contaminated. Presence of arsenic is due to air pollutants from coal combusted in brick factories [11]. It was mentioned that the groundwater of around 107 countries is contaminated with arsenic. Among these countries, the groundwater of Asian countries such as India, Pakistan, China Bangladesh, Nepal, Burma, Cambodia and Thailand are the most affected due to arsenic. Whereas, In India, the groundwater of four union territories and 20 states are highly affected by arsenic contamination [12].

In rural area, people are using groundwater for drinking and domestic purpose without knowing the suitability and quality of water in their region.

Thus, millions of people are suffering with life-threatening diseases such as cholera, diarrhoea, dysentery, hepatitis A and polio due to the consumption of unsafe water. Thus, due to the consumer's awareness towards adverse impact of consuming contaminated water the demand for water quality monitoring is increasing [13]. The groundwater suitability for domestic, drinking, irrigation and industrial purpose is determined by a fundamental factor known as groundwater hydrochemistry [14].

Thus groundwater contamination is big challenge to human survival and growth in a particular region. Hence, since that two to three decades, researchers, governments and renowned agencies such as national groundwater association [16-17] and World Bank [18] have been actively working on to control the water contamination. Table 1.1 shows the values for the quality parameters of

ground water as given by National [BIS] standards as well as International [WHO] standards.

The main parameters required for measurement of quality of ground water are:-

pH

It is negative logarithm of the hydrogen ion such as pH = −log [H+] [19]. The pH values of water samples were measured to know the variation in the acidity / basicity of samples collected from different locations. In neutral solution [H^+] is 10^{-7} / pH is 7. pH mainly reveals the free hydrogen or hydroxyl ion present in sample. It also measures the corrosive nature of water. The water samples with high value of pH are of high corrosive nature.

Total Dissolved Solids [TDS]

Total Dissolved Salts [TDS] measure the salt dissolved in a water samples after removing the suspended solids. When the evaporation of water is completed, the residue is TDS of the water sample. TDS of samples determine the various solids such as salts, minerals and organic matter present in water [Krishna et.al, 2017]. TDS is calculated using the equation 1.2:

Equation 1.2 Calculation of TDS of the sample   [1.2]

$$\text{Total Solids } [\frac{\text{mg}}{\text{L}}] = \frac{(W_1 - W_2)[1000]}{\text{Sample Volume [mL]}}$$

Where W_1 = Weight of dried residue + dish and W_2 = Weight of empty dish

Table 1.1 National and International standard values for the parameters of ground water

Water quality parameters	WHO* International Standards [Maximum desired limit]	WHO International Standards [Maximum allowed limit]	BIS* Standards [Acceptable limit]	BIS Standards [maximum desired limit]
pH	6.5	8.5	6.5-8.5	-
TDS [mg/L]	500	1500	500	2000
EC [μS/cm]	1400	1400	1400	1400
Chloride [mg/L]	200	600	250	1000

*WHO [World Health Organization]; *BIS [Bureau of Indian Standards]; TDS [Total Dissolved solid]; EC [Electrical Conductivity]

Source: [19]

Electrical Conductivity [EC]

The amount of total dissolved solids [TDS] is directly related to electrical conductivity [EC; mhos/cm] or specific conductance. Electrical conductivity measures the capacity by which water can transmit electrical current. TDS and conductivity affect the water sample and the solubility of slightly soluble compounds and gases in water [e.g. $CaCO_3$, and O_2]. In general, the corrosiveness of the water increases as TDS and EC increase, assuming other variables are kept constant.

EC value measures the concentration of ions presents in water [20]. There can be three type of EC such as type I, type II, type III on basis of conductivity. Type I class have conductivity up to 1500 μs/cm, type II have 1500 and 3000 μs/cm whereas, and Type III has EC 3000μS/cm [21].

Chloride content

Chloride is measured by titration method. Silver nitrate is added with potassium chromate into solution. Color of solution noted, when it became red, then it indicate chloride content in solution. Chloride has been found in all types of water. Chloride mostly comes from agricultural activities, industries.

Outline of Thesis

Present PhD work focuses on the measurement of the characteristics of ground water samples and then classifying using PCA and deploying a regression model using PLS and MLR techniques.

Chapter1 includes the introduction of the basic topics and techniques which are used in present research work. Starting from electromagnetic spectrum to interaction of light with matter, atomic and molecule vibrations, thesis then covers the basic introduction of Near Infrared Spectroscopy, its instrumentation and application in different field. Introduction also covers different statistical methods used for analysis, classification and development of statistical model. Further an introduction to groundwater, its importance, reason of contamination and preventive measures are covered. An introduction is given to different physicochemical parameters of ground water in brief.

Chapter 2 covers the literature review and highlights the work which is already carried to determine the characterization and analysis of groundwater quality in terms of quality and quantity measurements.

Chapter 3 covers the hypothesis defined for the research work, research gap estimated from the review of literature. The chapter also focuses on the materials and methods used throughout the research work. Starting from sample collection, chemicals used, parameters used to characterize the ground water to the spectra data acquisition technique, statistical methods employed in the research have been listed in this chapter.

Chapter 4 to chapter 7 of the thesis covers the results obtained during the research and their interpretation and discussion.

Chapter 4 covers the results obtained while identifying the presence of pollutants in the collected ground water samples using NIR and FTIR analysis.

Chapter 5 covers the results for the reference analysis of the samples, their descriptive data analysis and discussion of the PCA analysis of all the collected ground water samples on the basis of region, source, depth and source of collection of the ground water samples.

Chapter 6 covers the PCA analysis interpretation and discussion of the spectral data for all the ground water samples.

Chapter 7 covers the final objective, where the regression models are developed using the PLS, iPLS and siPLS regression techniques.

References

[1].   Workman Jr, Jerry, and John Shenk. "Understanding and using the near-infrared spectrum as an analytical method." *Near-infrared spectroscopy in agriculture* 44 (2004): 1-10.

[2].   P.C. Williams, K. Norris [Eds.], Chemical principles of near-infrared technology. Pages 17–34 *in*: Near-Infrared Technology in the Agricultural and Food Industries, American Association of Cereal Chemists, Inc., St. Paul, MN [1987]

[3].   Moreno, Leidy Jhoana Agudelo, et al. "Evaluation of aquifer contamination risk in urban expansion areas as a tool for the integrated management of groundwater resources. Case: Coffee Growing Region, Colombia." *Groundwater for Sustainable Development* 10 (2020): 100298.

[4].   Donald and A. Burns [2007], Handbook of Near Infrared Analysis, 3rd Ed. Boca Raton, CRC Press, https://doi.org/10.1201/9781420007374

[5].   Vandeginste, B. G. M., et al. "Chemometrics: a textbook." [1988]: 20-21.

[6].   Jolliffe, I. Principal component analysis. John Wiley & Sons, Ltd, 2005

[7].   Seasholtz, Mary Beth, and Bruce R. Kowalski. "The effect of mean centering on prediction

[8].   Adimalla, Narsimha, Sathish Kumar Vasa, and Peiyue Li. "Evaluation of groundwater quality, Peddavagu in Central Telangana (PCT), South India: an insight of controlling factors of fluoride enrichment." *Modeling Earth Systems and Environment* 4.2 (2018): 841-852.

[9].   Singh, Abhay Kumar. "Quality assessment of surface and subsurface water of Damodar river basin." *Indian Journal of Environmental Health* 44.1 (2002): 41-49.

[10].   Ahada, Chetan PS, and Surindra Suthar. "Assessing groundwater hydrochemistry of Malwa Punjab, India." Arabian journal of geosciences 11.2 (2018): 1-15.

[11].   Farooqi, Abida, et al. "Sources of arsenic and fluoride in highly contaminated soils causing groundwater contamination in Punjab, Pakistan." Archives *of environmental contamination and toxicology* 56.4 (2009): 693-706.

[12].   Shaji, E., et al. "Arsenic contamination of groundwater: A global synopsis with focus on the Indian Peninsula." *Geoscience frontiers* 12.3 (2021): 101079.

[13].   Li, P., and J. Wu. "Drinking water quality and public health. Exposure and Health, 11 (2), 73–79." (2019)..

[14].   Li, Peiyue, and Jianhua Wu. "Sustainable living with risks: meeting the challenges." *Human and Ecological Risk Assessment: An International Journal* 25.1-2 (2019): 1-10.

[15].   Shah, Tushaar. *Groundwater governance and irrigated agriculture*. Stockholm: Global Water Partnership (GWP), 2014.

[16].   Shah, Tushaar, et al. "Sustaining Asia's groundwater boom: an overview of issues and evidence." *Natural Resources Forum*. Vol. 27. No. 2. Oxford, UK: Blackwell Publishing Ltd, 2003

[17].   Foster, Stephen D., Ricardo Hirata, and Ken WF Howard. "Groundwater use in developing cities: policy issues arising from current trends." *Hydrogeology Journal* 19.2 (2011): 271-274.

[18].   WHO Guidelines

[19].   Krishna Kumar, S., et al. "Evaluation of water quality and hydrogeochemistry of surface and groundwater, Tiruvallur District, Tamil Nadu, India." *Applied Water Science* 7.5 (2017): 2533-2544.

[20].   Prasanth, SV Sarath, et al. "Evaluation of groundwater quality and its suitability for drinking and agricultural use in the coastal stretch of Alappuzha District, Kerala, India." *Applied Water Science* 2.3 [2012]: 165-175.

[21].   Sunitha, V., and Y. Sudharshan Reddy. "Hydrogeochemical evaluation of groundwater in and around Lakkireddipalli and Ramapuram, YSR District, Andhra Pradesh, India." *HydroResearch* 2 (2019): 85-96.

Chapter 2- Review of Literature

2. Chapter 2- Review of Literature

Groundwater is one of the most important natural resources available on earth. According to a report of UNESCO, "Groundwater represent 98% of the unfrozen fresh water" [1].

All living organism need water for their survival. Water is most abundant element natural resource on earth. In developing countries, human activities such as Human urbanization, agricultural and industrial activities lead to contamination of ground water. At present ground water contamination is universal problem and it has direct impact on human health. Ground water quality varies with region to region and with depth also. So ground water study is important for world. The Malwa region of Punjab state is facing major issues regarding water contamination due to uncontrolled human activity. Water quality parameters should be checked regularly, because due to contaminated water, there can be many water borne diseases [2].

This chapter covers the research work carried out by researchers showing the feasibility and wide application of near infrared spectroscopy, methods for quality analysis of ground water, characterization of ground water of Malwa region, Punjab, application of spectroscopy mainly near infrared and Fourier transform infrared for analysis of ground water.

Application of infrared spectroscopy

Near Infrared [NIR] Spectroscopy is a tool to study different kinds of samples because of understanding intermolecular interaction, phase transitions, molecular structure. NIR helps in providing information about the sample. To extract information from the complex spectra multivariate analysis methods have to be applied, as NIR spectra is more complex as compared to MIR and FIR region [3].

These characteristics and ability of NIR to check the quality of organic compounds leads to its wide applications in the field of medical, food products, polymer industry, beverage industry and many more. NIR spectroscopy is used for the qualitative and quantitative analysis of wheat, millet, milk, oats and many more food products. Spectroscopy as a secondary data and chemical reference data as primary along with multivariate analysis was successfully used to determine the β-Glucan content in hulled and naked oats, for detection of adulterant and its quantity in milk [4-7].

Due to its feasibility to extract information from organic bonds, namely – N-H, S-H, O-H, C-H for overtones and combinational bands, its used in all fields for quality and quantity estimation. Chemical, physical, biological parameters under study are measured using the chemical and standard methods and act as a primary data set, NIR spectra collected for the samples is used as a secondary data set. Both reference and spectral data is then used to develop statistical regression models using various multivariate techniques [8-9].

Analysis of groundwater using chemical properties

Groundwater being an important part as a source of livelihood and irrigation. Various research is carried out to find its parameters, check its quality using different chemical, physical and biological methods.

Research work carried at International level

Set of 28 ground samples were collected from area of Taiwan and relationship between 13 different hydrochemical parameters were tested. Factor analysis was used to explain the Blackfoot disease in the reason. The work clearly shows the quality index of the collected samples in terms of the concentration of EC, TDS, Cl^-, $SO_4^{2-,}$ Na^+, K^+ and $Mg^{2+,}$ Alk, TOC and arsenic [10].

A research was carried out on ground water collected from 11 districts of Bangladesh and various water indices were measures such as sodium adsorption ratio [SAR], soluble sodium percentage [SSP], magnesium adsorption ratio [MAR], permeability index [PI] and irrigation water quality [IWQ] index. The results were compared with the IWQ index and it was observed that 41% samples showed medium sustainability for irrigation and 25% in highly sustainability water type. Such findings can be useful in managing the groundwater for irrigation purposes [11].

Work has also been carried to monitor the quality of groundwater by extending the use of soil and water assessment tool to simulate the groundwater quality to measure the total dissolved solids and nitrate level. A model was developed between the calculated and the observed groundwater quality. Satisfactory results were obtained with coefficient of determination between 0.6-0.83. The work showed that extended SWAT models are useful for quality analysis of groundwater [12].

Research works also shows that Bayesian network can be used to measure and predict the nitrate and TDS concentrations for groundwater quality analysis. In simulation correlation coefficient [R^2] of 0.86 for sodium chloride, chloride, drawdown, return flow, and precipitation for samples collected from Iran. Value of nitrate was calculated using concentrations of sulfate, calcium, hardness, electric conductivity, drawdown, aquifer recharge. The research work showed that combination of the different prediction results with proper zoning will be useful in land development and improvement [13].

Frequency domain electromagnetic data for the groundwater samples around the dumpsites from the districts of Nigeria were measured and multiple linear regression [MLR] was used to develop a model for measurement of TDS and other parameters. The results showed that variation in the values of TDS help on understanding the integrity and quality of groundwater around a given place [14].

Research work have been reported from Lahore, Pakistan to check the quality of groundwater contaminated because of toxic elements present in soil and water due to waste from industries and agricultural pesticides. The work aimed in estimating the concentration of TDS in water collected from different parts of area under investigation. Analysis was done on groundwater, drain water and river water. [15].

Presence of the Calcium and sodium in groundwater samples mainly used for irrigation, procured from western part of Wadi Nisah was carried out by the researchers Saturation indices were calculated and it was observed that samples were under saturated due to the concentration of gypsum, halite, aragonite etc. The facies analysis showed that the samples were having high concentration of Ca, Na, SO_4 type water, indicating rock–water interaction. Principal component analysis was used to study the relation between the concentration of NO_3 and Pb. Grouping of the samples were done using R-mode cluster analysis. The values of TDS help in determining the water as brackish or freshwater. Reason for the high TDS was listed because of soil–water interaction, agricultural and household wastewater. Results showed that the ground water was highly polluted in the region of study [16].

In a research work carried out at Nullah Lai Basin in Pakistan, all physical and chemical parameters measured with meters and chemical methods were found with in safe range but microbial pollutants were high in the tested samples [17].

Another research was carried on the samples from China to check the temporal and spatial variation of the quality of the groundwater. Principal component analysis [PCA] along with geographic information system [GIS] showed that the hydrogeochemical process was major factor to determine the quality of water samples, in addition agricultural activities, industrial and domestic sewage discharges are responsible for the same. Nitrogen concentration was also high due to anthropogenic activities in groundwater [18].

Research work carried at National level

Literature also shows that work has been done in India to identity the quality of the samples. Research was carried out on the samples collected from districts of Maharashtra. Reference methods were used to find the major chemical and physical parameters and compare with the national and international standards. Correlation coefficients were also calculated to know quality of groundwater samples [19].

Seventy three groundwater samples collected from district of Tamilnadu were tested for the pH, EC, Turbidity, Total Hardness, TDS, Iron, Chlorides and metals. Piper's trilinear diagram was used to study the quality of groundwater samples. Statistical tools namely descriptive statistics, cluster analysis and correlation were applied on the data set. Results showed that the variation in TDS had a high influence in classifying the ground water samples [20].

Ground water samples from kheda district was checked in pre monsoon season for quality assessment by measuring the parameters - electrical conductivity [EC], pH, total dissolved solids [TDS], sulphate, chloride and total cation and anion. Minerals were also measured for the collected samples using reference methods. Correlation coefficients and regression equations were determined for the parameters to know the relation between them. Research showed that developing the regression equation helps in quick determination of the parameters [21-22].

Research was carried out to check the quality parameters of waste water in drains of regions of Coimbatore, Tamil Nadu, India. Quality of the groundwater samples near these drains were tested by measuring the chemical and physical parameters - pH, EC, TDS, chlorides, sulphates, and hardness. Principal component analysis was used to

check the relation between different factors responsible for contamination of groundwater samples [23].

Research work carried at State level

Punjab is divided into three regions, namely – Majha, Malwa and Doaba. The water quality in all the three regions is affected by the industrial, agricultural and human waste dumped in ground.

Quality of ground water procured from south west region of Punjab was tested for uranium concentrations. Hydro geochemical data was used for uranium measurements. It was observed that samples with high TDS, pH and ORP values had higher chance of uranium concentration in high range. It was observed that climatical and agriculture based factors are responsible for high value of Uranium in ground water [24].

Authors discussed about the ground water quality of two district Patiala and Muktsar and compared its suitability for drinking and irrigation purpose. The research work carried out by measuring the concentrations of pH, EC, dissolved oxygen, ORP and temperature. Results showed that ground water in Patiala district is better for drinking and irrigation as compared to that of Muktsar region [25].

In another research uranium concentration in ground water samples of six districts of Punjab were examined. Uranium fluorimeter was used to measure the concentrations of uranium. It was observed that uranium concentration was higher in south west district of Punjab than the limit given by WHO, USEPA and AERB and the western and northern region had the concentration of uranium in limit given by AERB [26].

In another research hydrochemistry of ground water samples from 76 sites of Malwa region of Punjab were tested for the parameters - pH, EC, TDS, TH, Na^{2+}, K^+ , Mg^{2+}, Ca^{2+}, F^- and Cl^- SO_4^{2-} , PO_4^{3-} , NO_3^-. Ground water samples procured from Malwa region of Punjab was not found suitable for agriculture and drinking. Region for the same was found to be chemical and agriculture regions [27].

Literature also showed that research was carried to test the ground water samples for 18 sited of Ropar Punjab. Work was carried out to measure the heavy metal contents of groundwater samples were measured by Flame Atomic Absorption Spectrometer. Results showed that all groundwater samples were non-saline slightly alkaline, and

very hard in nature. It was found that water is contaminated with heavy metals and can cause serious health problem [28].

Authors did research to measure the fluoride and nitrate contamination in drinking water sources from Bathinda, Ferozepur and Faridkot districts of Punjab. Measurements for testing contaminations were pH, TDS, fluoride and nitrate content. The measurements were done by TDS meter, pH meter and fluoride and nitrate field test kits of Merck. Colorimetric determination of fluoride was also measured using the SPADNS method, nitrate by cadmium reduction method and phosphate by ascorbic acid method, respectively. Results showed that rural parts of south-western districts of Punjab are suffering from both nitrate and fluoride and nitrate in water [29].

Another research was carried out in eight districts of south-western part of the Punjab, India. The groundwater quality in about 6% region of study area was found fit, 18% was marginal and 76% was unfit for irrigation purpose which is a serious problem [30].

Another research work done on the samples collected from Dera Basi of Punjab, India. Physiochemical parameters were measured and compared with WHO and Indian standards. All parameters except measured alkalinity were found in the safe range according to the national and international standards [31].

Analysis of groundwater using spectroscopic techniques

Literature work shows that lot of work is reported for measuring the ground water quality using chemical and standard methods. Very less literature is available where rapid techniques using spectroscopic data are used for measuring the quality parameters estimation and prediction of ground water samples.

Research has been reported where spectral variation due to change in temperature have been studied in NIR region of the electromagnetic spectrum. Work was carried out by comparing the reference data with the spectroscopic data using chemometrics. Spectra was obtained using FT-NIR spectra in the **range of** 9000–5500 cm^{-1} over a temperature range of 5–85°C. Multilinear regression [MLR], principal component regression [PCR] and partial least squares [PLS] regression, and various preprocessing techniques were applied for the analysis. Results showed and explained the presence of peaks at different wavenumbers. It was explained that band around

6900 cm^{-1} may be due to the combination of antisymmetric and symmetric stretching modes of water. Temperature of the samples could be predicted by the multivariate analysis MLR, PCR and PLS regression. Results show that prediction can be done for spectrum of water between 5 and 85°C. Thus the results showed that spectroscopy and chemometrics can be used for predicting the temperature variations of the water samples [32].

Near Infrared spectroscopy had been useful for studying the water spectrum bonds. Bands near to 1440 and 1930 nm were responsible for complex analysis of the water in NIR spectrum. Results showed that the hydrogen bonding in water affects the reliability of this rapid and nondestructive method. [33].

Literature reported that Soil water content [WC] can be estimated using visible [VIS] and near infrared [NIR] spectroscopic and can also be used to measure other properties of soil like C, N, and other nutrients. WS contribution can be estimated by subtracting its contribution to spectra data. Classification of the spectral data was done and is useful in determining water content. 275 samples procured from a single field and the 360 samples from multiple fields in Belgium and northern France were used for the study. Partial least squares [PLS] regression and factorial discriminant analysis [FDA] were used for prediction of WC in the water samples. Results showed that PLS and FDA can successfully be used to predict the WC Samples of single field samples were higher as compared to the multifield samples. Thus FDA and PLS can be successfully used to classify soil VIS-NIR spectra into different WC levels [34].

On the basis of the OH covalent and other hydrogen bonds the spectral data is highly affected. These bonds are affected by the environment factors. Thus researchers worked on the field of aquaphotomics to analyze the water using near infrared [NIR] spectra, as a function of internal and external factors. Aquaphotomics analysis were done to analyse different groundwater samples using the NIR spectral data. The unique patterns helps in identifying the spectral variations due to water absorption bands. Two groundwaters obtained at different depths and different parameters were analysed for subsequent years. The obtained NIR spectra was preprocessed using the smoothing and multiplicative scatter correction [MSC] to remove noise and baseline effects. Analysis showed that NIR aquaphotmics spectral patterns of groundwater samples were affected by temperature, humidity, time and others. Short wavelength of

NIR region is highly influenced by humidity and temperature. Results confirmed that vis/NIR spectral pattern can be used as an specific marker of water status, providing the basis for a cost-effective and efficient method for monitoring of water functionality [35].

Research work was reported where geostatistical techniques were applied to check the quality of groundwater quality in a region of Ghana. Analysis were carried upon on 41 samples for physico-chemical and trace metals. Water quality index was predicted using partial least square regression, suggesting that PO_4^{3-}, As, Cd and Fe are important parameters for the prediction quality [36].

Importance and power of regression have been also listed in a review by the authors, where they have discussed the use of partial least squares modeling [PLSM] in the field of water analysis for different kinds of water samples – ground water, river and coastal water [37].

According to the literature, chemometrics has been used along with fluorescent spectra. Bacterial contamination has been accessed using fluorescence spectroscopy along with partial least squares [PLS] algorithms. In all 90 groundwater samples from 6 different wells of Israel were used for the study. It was observed that satisfactorily results were obtained for measuring microbial quality of water using PLS [38].

Wavelength selection methods along with PLS, namely – iPLS, siPLS have been implement by researchers for composition of samples using the FTIT / NIR data [39].

Reported literature work summarized that the quality of the ground water samples are affected by temperature, humidity and other environmental factors. The quality is also affected by the dumping of agricultural, industrial and human waste. Physical, chemical and reference methods have been used to test the quality parameters and compared with the national and international standards given by BIS and WHO respectively. Research has been carried at international, national and state level. Maximum work is reported using the chemical and reference methods. Very little research has been reported where spectroscopic techniques has been used to quantify the water samples using the regression methods PLS, MLR and FDA. Classification

of the samples was done on the basis of temperature and other environmental factors. Regression models were used to predict the quality of the water samples.

References

[1].   https://en.unesco.org/themes/water-security/hydrology/groundwater.

[2].   Simpi, Basavaraja, et al. "Analysis of water quality using physico-chemical parameters Hosahalli Tank in Shimoga District, Karnataka, India." *Global Journal of Science Frontier Research* 11.3 (2011): 31-34.

[3].   Czarnecki, Mirosław A., et al. "Overview of application of NIR spectroscopy to physical chemistry." *Near-Infrared Spectroscopy*. Springer, Singapore, 2021. 297-330.

[4].   Meenu, Maninder, et al. "Rapid Determination of β-Glucan Content of Hulled and Naked Oats Using near Infrared Spectroscopy Combined with Chemometrics." *Foods* 11.1 (2021): 43

[5].   Kamboj, Uma, et al. "Application of selective near infrared spectroscopy for qualitative and quantitative prediction of water adulteration in milk." *Materials Today: Proceedings* 24 (2020): 2449-2456.

[6].   Kamboj, Uma, Paramita Guha, and Sunita Mishra. "Characterization of chickpea flour by near infrared spectroscopy and chemometrics." *Analytical Letters* 50.11 (2017): 1754-1766.

[7].   Meenu, Maninder, et al. "Green method for determination of phenolic compounds in mung bean (Vigna radiata L.) based on near-infrared spectroscopy and chemometrics." *International Journal of Food Science & Technology* 51.12 (2016): 2520-2527. https://doi.org/10.1111/ijfs.13232

[8].   Zareef, Muhammad, et al. "An overview on the applications of typical non-linear algorithms coupled with NIR spectroscopy in food analysis." *Food Engineering Reviews* 12.2 (2020): 173-190.. https://doi.org/10.1007/s12393-020-09210-7

[9].   Ozaki, Yukihiro, et al., eds. *Near-infrared spectroscopy: theory, spectral analysis, instrumentation, and applications*. Berlin/Heidelberg, Germany: Springer, 2021.

[10].   Liu, Chen-Wuing, Kao-Hung Lin, and Yi-Ming Kuo. "Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan." *Science of the total environment* 313.1-3 (2003): 77-89.

[11].   Islam, Md Shajedul, and M. G. Mostafa. "Groundwater suitability for irrigated agriculture in alluvial Bengal delta plain: a review." *Int J Adv Appl Sci ISSN* 2252.8814 (2021): 8814.

[12].   Meshesha, Tesfa Worku, et al. "Modelling groundwater quality of the Athabasca River Basin in the subarctic region using a modified SWAT model." *Scientific Reports* 11.1 (2021): 1-12.

[13].   Hamid Kardan Moghaddam, Amir Rajaei, Zahra Rahimzadeh kivi, Hossein Kardan Moghaddam, Prediction of qualitative parameters concentration in the groundwater resources using the Bayesian approach, Groundwater for Sustainable Development,   Volume   17, May   2022,   100758, https://doi.org/10.1016/j.gsd.2022.100758

[14].   Aduojo, Ameloko Anthony, Ayolabi Elijah Adebowole, and Okezie Uchegbulam. "Modeling groundwater total dissolved solid from derived electromagnetic data using multiple linear regression analysis: a case study of groundwater contamination." *Modeling Earth Systems and Environment* 6.3 (2020): 1863-1875.

[15].   Akhtar, M. M., and Zhonghua Tang. "Identification of contamination sources and TDS concentration in groundwater of second biggest city of Pakistan." *International Journal of Environmental Science and Development* 4.3 (2013): 341.

[16].   Mohammed Gubran, Habes Ghrefat, Faisal Zaidi & Mohamed Shehata [2019]. NIR modelling in ground water analysis, Environmental Earth Sciences, Integration of hydrochemical, GIS, and remote-sensing data for assessment of shallow groundwater aquifers in Wadi Nisah, Central Saudi Arabia, volume 78, Article number: 161

[17].   Ali, Arshad, Mazhar Iqbal, and Abdul Waheed. "Groundwater quality assessment near Nullah Lai stream of Pakistan." *Central Asian Journal of Environmental Science and Technology Innovation* 2.2 (2021): 45-51.

[18].   Li, Qiling, et al. "Groundwater pollution source apportionment using principal component analysis in a multiple land-use area in southwestern China." *Environmental Science and Pollution Research* 27.9 (2020): 9000-9011. https://doi.org/10.1007/s11356-019-06126-6

[19].   Patil, V. T., and P. R. Patil. "Physicochemical Analysis of Selected Groundwater Samples of Amalner Town inJalgaon District, Maharashtra, India." *E-Journal of Chemistry* 7.1 (2010): 111-116.. https://doi.org/10.1155/2010/820796

[20].   Samson, S., and K. Elangovan. "Multivariate statistical analysis to assess groundwater quality in Namakkal district, Tamil Nadu, India." (2017).

[21].   Ashiyani, Nikunj, Falguni Parekh, and T. M. V. Suryanarayana. "Analysis of Physico-chemical properties of groundwater." *International Journal of Innovative Research in Science, Engineering and Technology* 4.3 (2015): 1094-1098

[22].   Chakraborti, Dipankar, Bhaskar Das, and Matthew T. Murrill. "Examining India's groundwater quality management." (2011): 27-33.

[23].   Mohammed Siraj Ansari, A., and R. Saraswathi. "use of Principal Component Analysis for Parameter Selection to Compute Water Quality Index and Assessment of Groundwater Quality Status: A Case Study of Sanganur Stream, Coimbatore City." *Journal of the Geological Society of India* 98.3 (2022): 417-422.

[24].   Sahoo, Prafulla Kumar, et al. "Meta-analysis of uranium contamination in groundwater of the alluvial plains of Punjab, northwest India: Status, health risk, and hydrogeochemical processes." *Science of The Total Environment* 807 (2022): 151753. [https://doi.org/10.1016/j.scitotenv.2021.151753

[25].   Kumar, Manish, et al. "A comparative evaluation of groundwater suitability for irrigation and drinking purposes in two intensively cultivated districts of Punjab, India." *Environmental geology* 53.3 (2007): 553-574. https://doi.org/10.1007/s00254-007-0672-3

[26].   Saini, Komal, Parminder Singh, and Bikramjit Singh Bajwa. "Comparative statistical analysis of carcinogenic and non-carcinogenic effects of uranium in groundwater samples from different regions of Punjab, India." *Applied Radiation and Isotopes* 118 (2016): 196-202. https://doi.org/10.1016/j.apradiso.2016.09.014

[27].   Ahada, Chetan PS, and Surindra Suthar. "Assessing groundwater hydrochemistry of Malwa Punjab, India." *Arabian journal of geosciences* 11.2 (2018): 1-15. https://doi.org/10.1007/s12517-017-3355-8

[28].   Sharma, Sakshi, Avinash Kaur Nagpal, and Inderpreet Kaur. "Appraisal of heavy metal contents in groundwater and associated health hazards posed to human

population of Ropar wetland, Punjab, India and its environs." *Chemosphere* 227 (2019): 179-190. https://doi.org/10.1016/j.chemosphere.2019.04.009

[29]. Sharma, Chetna, Amita Mahajan, and Umesh Kumar Garg. "Fluoride and nitrate in groundwater of south-western Punjab, India—occurrence, distribution and statistical analysis." *Desalination and Water Treatment* 57.9 (2016): 3928-3939. https://www.researchgate.net/publication/284427753

[30]. Chopra, Rajinder Pal Singh, and Gopal Krishan. "Analysis of aquifer characteristics and groundwater quality in southwest Punjab, India." *Journal of earth science and engineering* 4.10 (2014): 597-604. doi: 10.17265/2159-581X/2014.10.002

[31]. Sharma, Nitish, Upma Vaid, and Sanjay Kumar Sharma. "Assessment of groundwater quality for drinking and irrigation purpose using hydrochemical studies in Dera Bassi town and its surrounding agricultural area of Dera Bassi Tehsil of Punjab, India." *SN Applied Sciences* 3.2 (2021): 1-13. https://doi.org/10.1007/s42452-021-04199-y

[32]. Maeda, Hisashi, et al. "Near infrared spectroscopy and chemometrics studies of temperature-dependent spectral variations of water: relationship between spectral changes and hydrogen bonds." *Journal of Near Infrared Spectroscopy* 3.4 (1995): 191-201.

[33]. Büning-Pfaue, Hans. "Analysis of water in food by near infrared spectroscopy." *Food Chemistry* 82.1 (2003): 107-115.

[34]. Mouazen, Abdul Mounem, et al. "Characterization of soil water content using measured visible and near infrared spectra." *Soil Science Society of America Journal* 70.4 (2006): 1295-1302. https://doi.org/10.2136/sssaj2005.0297

[35]. Kovacs, Zoltan, et al. "Aquaphotomics for monitoring of groundwater using short-wavelength near-infrared spectroscopy." *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* (2022): 121378.

[36]. Abdul-Wahab, Dickson, et al. "Application of geostatistical techniques to assess groundwater quality in the Lower Anayari catchment in Ghana." *HydroResearch* 5 (2022): 35-47

[37]. Khatri, Punit, Karunesh Kumar Gupta, and Raj Kumar Gupta. "A review of partial least squares modeling (PLSM) for water quality analysis." *Modeling Earth*

Systems and Environment 7.2 (2021): 703-714 https://doi.org/10.1007/s40808-020-00995-4

[38]. Nakar, Amir, et al. "Quantification of bacteria in water using PLS analysis of emission spectra of fluorescence and excitation-emission matrices." *Water Research* 169 (2020): 115197, https://doi.org/10.1016/j.watres.2019.115197.

[39]. da Silva, Daniel José, and Hélio Wiebeck. "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 (2017): 259-266.

Chapter 3 – Hypothesis, Objectives and Methodology

3. Chapter3 – Hypothesis, Objectives and Methodology

Research Gap

Literature review reveals that ground water is very important for our livelihood. It is required for drinking, washing and irrigation purposes but the ground water is contaminated with different harmful chemicals from human, industrial and agricultural waste dumped in land. Organic contaminants, chemicals, minerals are the main adulterant found in ground water. To analyse and test the quality of the samples various physical and chemical methods have been reported and the results have been compared with the international and national standards given by WHO and BIS respectively. Different reference methods such as Gas Chromatography, X-ray Diffraction, Scanning Electron Microscope and Tunneling Electron Microscope have been used for the quality analysis. Review shows that the work is done at international, national and state level for quality analysis of ground water. Few research is carried out where infrared spectroscopy along with regression methods such as PLS, MLR have been used to investigate the soil and water quality on data collected using geospatial images.

It has been observed that the very few or minimal research has been conducted in India for the identification of pollutants in ground water and estimating the characteristics of ground water by developing a regression model using the FTIR and NIR spectra with the reference data obtained from the chemical methods.

Punjab is divided into three different regions – Malwa, Majha and Doaba. It has been observed from different techniques that the Malwa region belt have contaminated ground water which is responsible for the many diseases such as cancer.

Present research focuses on quality estimation, classification of the procured samples on the basis of their collection parameters and developing a regression model using complete near infrared and FTIR spectra to characterize the water samples qualitatively and quantitatively.

Research Hypothesis

Null Hypothesis – Near Infrared spectra have no significant relation with the properties – pH, Total Dissolved Solids, Electrical Conductivity and chloride content of ground water samples.

Alternative Hypothesis – Near Infrared spectra have significant relation with the properties – pH, Total Dissolved Solids, Electrical Conductivity and chloride content of ground water samples.

Research Objective

1. Identification of pollutants in ground water by using chemical methods for the western Malwa belt of Punjab - Ferozepur and Fazilka districts.

2. Absorbance measurements of ground water using Visible-Near Infrared spectroscopy.

3. Development of a regression model using spectral, chemical data with chemometrics to predict the pollutants.

Sample Collection

Samples for the research work were collected from Malwa Region, Punjab. The districts which were selected for study were – Ferozepur and Fazilka districts of Punjab. These districts lie in the Malwa region. Fazilka is located 10 km east to the India-Pakistan border near the tri-junction of Punjab, Haryana and Rajasthan states of India between north latitude 30.403˚ and east longitude 74.025˚ with an average elevation of 177 m [581 feet]. Fazilka is major rice and cotton growing region and exporter in India. The soil in this region is mostly sandy loam. The Ferozepur district is situated on the bank of the Sutlej River, between north latitude 30.9331˚ and east longitude 74.6225˚. Soil type of this district is sandy loam to silt [Figure 3.1]. This type of soil has water holding capacity for long time, so it is suitable for rice growth. ˙
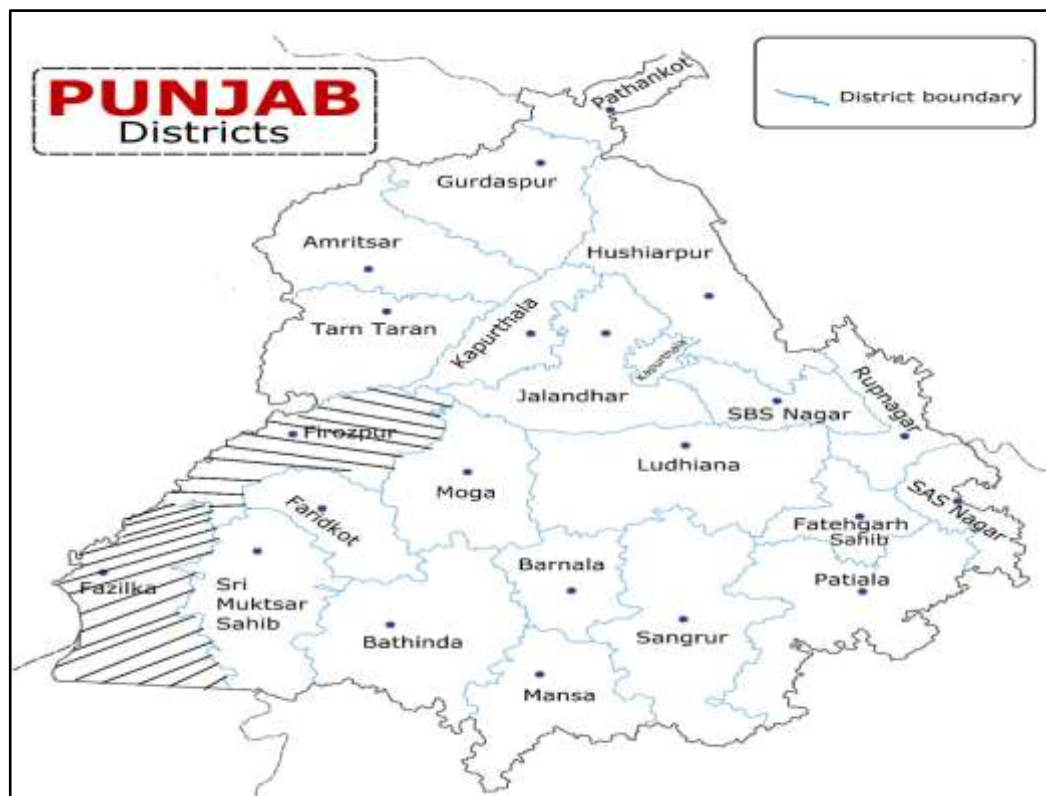
Figure 3.1 The highlighted region of Punjab, India map are shown from where the ground water samples were collected.

Sample coding was done for the samples with A as collected from Fazilka and B from Ferozepur. The water samples were collected from rural, industrial and

urban region on the basis of depth, from three different sources– water box [WB], submersible [SS], hand pump [HP]. Depth range for hand pump, submersible and water box were 28m to 50m, 60m to 500m and 340 to 540m respectively. Samples were stored in screw capped polythene bottles at room temperature till further analysis. Analysis for chemical and spectral data collection were done within a week of collection of the samples.

In all 50 samples were collected from the two districts. Details of the collected samples from the two districts with differentiation of region, source of collection and depth are listed in Table 3.1, 3.2 respectively. Complete details of the collected samples are given in Table 3.3.

Table 3.1 Distribution of the collected samples on the basis of district and region – Urban, Industrial and rural basis.

Row Labels	Count of Place	Count of IRU[Region]
A	**27**	**27**
I	9	9
R	9	9
U	9	9
B	**23**	**23**
I	7	7
R	8	8
U	8	8
Grand Total	**50**	**50**

Table 3.2 Distribution of the collected samples on the basis of district and source collected – Handpump-HP, Submersibal- SS, Waterbox-WB.

Row Labels	Count of Place	Count of Source
A	**27**	**27**
HP	4	4
SS	17	17
WB	6	6
B	**23**	**23**
SS	14	14
WB	9	9
Grand Total	**50**	**50**

Table 3.3 Description of the collected samples. Sample coding for district: Fazilka –A, Ferozepur – B; region: Rural – R, Industrial- I, Urban – U; water sources: Handpump – HP, Submersible – SS, Waterbox - WB

Sample No.	Place	IRU[Region]	Sample code	Source	Depth [feet]	Sample No.	Place	IRU[Region]	Sample code	Source	Depth [feet]
1	A	R	1AR	HP	50	26	A	U	26AU	WB	450
2	A	R	2AR	HP	28	27	A	U	27AU	WB	550
3	A	R	3AR	HP	40	28	B	R	28BR	WB	550
4	A	I	4AI	HP	40	29	B	I	29BI	SS	100
5	A	R	5AR	SS	60	30	B	I	30BI	SS	260
6	A	U	6AU	SS	160	31	B	R	31BR	WB	450
7	A	U	7AU	SS	60	32	B	R	32BR	SS	100
8	A	R	8AR	SS	95	33	B	R	33BR	SS	120
9	A	R	9AR	WB	450	34	B	R	34BR	SS	90
10	A	R	10AR	SS	108	35	B	R	35BR	SS	90
11	A	R	11AR	WB	410	36	B	I	36BI	SS	220
12	A	I	12AI	SS	200	37	B	I	37BI	SS	140
13	A	I	13AI	SS	220	38	B	R	38BR	WB	350
14	A	R	14AR	WB	400	39	B	I	39BI	SS	120
15	A	I	15AI	SS	200	40	B	R	40BR	SS	85

16	A	I	16AI	SS	205	41	B	U	41BU	WB	450
17	A	I	17AI	SS	220	42	B	U	42BU	WB	340
18	A	I	18AI	SS	230	43	B	I	43BI	SS	180
19	A	I	19AI	SS	250	44	B	I	44BI	SS	150
20	A	I	20AI	SS	250	45	B	U	45BU	WB	550
21	A	U	21AU	SS	220	46	B	U	46BU	WB	500
22	A	U	22AU	SS	120	47	B	U	47BU	WB	550
23	A	U	23AU	WB	500	48	B	U	48BU	SS	450
24	A	U	24AU	SS	385	49	B	U	49BU	SS	250
25	A	U	25AU	SS	500	50	B	U	50BU	WB	495

Reference Analysis

All the ground water samples were tested for pH, TDS, EC and chloride content. The pH of water samples was measured using digital pH meter. EC was determined by microprocessor conductivity meter. The TDS value of samples were measured by using TDS meter [TDS –salinity Temp meter G-3020 of HPG system], whereas, the chloride content of water samples was determined by employing chemical method. In this method two drops of potassium chromate solution was added in ground water samples, then water color was observed. If the color turned into yellow, then silver nitrate was added in the solution until it turned into red and reading indicate amount of chloride in ground water. Details and need of the parameters are explained in Chpater1. The entire reference analysis tests were done from the National Laboratory. Reviewing the literature showed that these are the main characteristics tested or quality analysis of ground water. It was observed that TDS and pH were the major parameters which helped in estimating the presence of Uranium content in ground water It was observed in the literature that the studies done on the soil of Malwa region of Punjab showed that TDS is highly correlated with the Uranium concentration. Further it was also observed that the ground water samples collected from depth < 60 m had high presence of Uranium [1]. Thus the literature corresponding to the quality of ground water and soil showed that there is need to understand and study the physicochemical parameters for groundwater.

Spectral Analysis

All the ground water samples were analysed on the basis of their spectral data. Vis-NIR and FTIR data was collected using NIR DS 2500 spectrometer at Council of Scientific and Industrial Research –Central Scientific Instrumentals Organization, Chandigarh, India [Fig 3.2] and FTIR spectra was obtained from Central Instrumentation Facility, Lovely Professional University, Jalandhar, Punjab, India.

Vis - NIR spectra was obtained in the range of 400-2500 nm with a gap of 0.5 nm. The spectrometer used glass beakers with gold reflectors [Fig 3.3]. Two

detectors were used in the spectrometer – 400-1100 nm was analysed using Si detectors and 1100-2500 nm was obtained from PbS detectors. Data was obtained in reflectance mode in triplicate and was converted to absorbance by the inbuilt VISION software of the instrument. Instrument is resistance to dust, temperature, vibrations and moisture, thus gives accurate results. It's a pre-dispersive spectrometer, where polychromatic light is dispersed into different wavelengths before illuminating the sample.

Figure 3.2 NIR DS2500 Spectrometer by Metrohm used for the vis-NIR spectra collection

Figure 3.3 The liquid sample holder for the NIR DS2500 instrument with gold reflector

FTIR data was obtained using the Perkin Elmer Diamond ATR Fourier Transform Infrared Spectrometer with resolution 0.5 cm^{-1} [Fig 3.4]. Infrared Spectroscopy can be used to study the chemical composition of the sample due to their vibrational transitions. Molecules with dipole moment, when interacted with the IR radiations fluctuates the EM. These are the footprints of the sample. Thus to understand the chemical composition of the sample these fluctuation of electromagnetic radiations are recorded by this spectrophotometer. Spectra data was obtained from 350- 8300 cm^{-1}. The instrument was equipped with atmospheric vapour compensation [AVC] facility to subtract the H_2O and CO_2 absorptions.

Figure 3.4 FT-IR Spectrophotometer by Perkin Elmer for collecting the FTIR spectra

Fluorescence Spectra was obtained using the Perkin Elmer Fluorescence spectrophotometer at CIF, LPU, and Punjab. Fluorescence spectra helps to analyse the state of the sample by studying their interactions with fluorescent molecules. Instrument measures the changes in fluorescent optical properties of the sample. Xenon Pulse lamp was used to obtain the fluorescence light. Excitation and emission range for the instrument lies from 200-900nm. Instrument was equipped with slit of variable width from 1 to 20 nm.
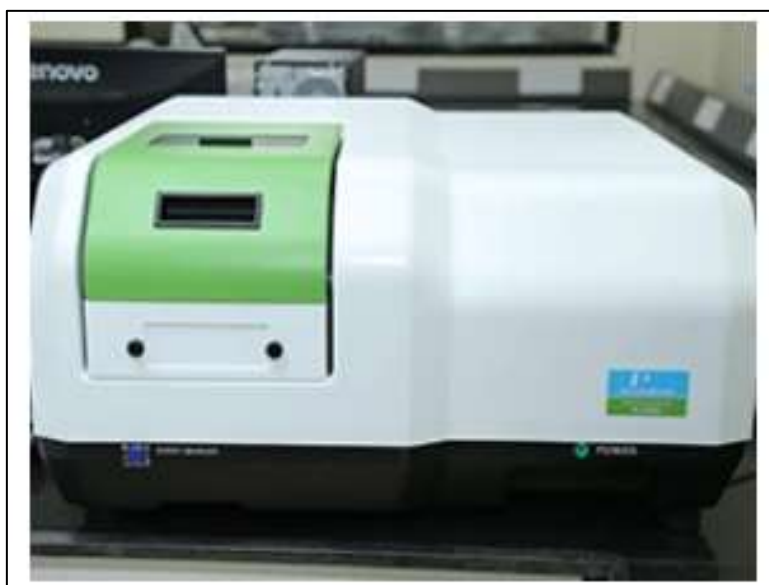
Figure 3.5 Fluorescence Spectrophotometer by Perkin Elmer to analyse the samples

Statistical Analysis

Descriptive Statistics are used to know the data better. In a research study one may have large data sets, thus there is need to understand the data by knowing the distributions, diversity of the data. This statistics is applied primarily to the data set. There are two main categories covered in this research work – measures of central tendency and measures of dispersion.

Measures of Central tendency

These are also known as complete averages or a value around which all other values have a tendency to cluster. The three main averages are – mean median and mode. In the present research work emphasis is given on mean and median.

Mean is the simplest and most common measure among the measures of central tendency. It is obtained by dividing sum of all values with total number of values. Mean values is affected if one or more data has extreme values. Data set obtained was individual data and not a frequency distribution data. Thus mean was calculated by the equation 3.2.

Equation 3.3 Calculation of Mean for the individual data set

$$Mean\ [\bar{X}] = \frac{\sum X_i}{n} \qquad [2]$$

where, $\bar{X}$ symbol used for mean, X_i values of the ith element and n is the total number of elements.

Median is the measure of central tendency which is used when mean becomes flowed as representative of data set. Median is not used in the data where every element is to be assigned relative weightage. Median divided the complete data in two halves, where one half contained elements less than median and other half are larger than median [equation 3.3].

Equation 3.4 Calculation of Median for the individual data set

$$Median = \left(\frac{n+1}{2}\right) when\ number\ of\ observation\ is\ odd \qquad [3]$$

$$Median = \frac{1}{2}\left[\left(\frac{n}{2}\right) th\ observation + \left(\frac{n}{2} + 1\right) th\ observation\right], n\ is\ even$$

Measure of Dispersion

Average can represent the data set as a one value, but cannot explain the variation in the complete dataset. For example – For the data set 1 [5,5,5,5, 5] mean and median are 5 and for data set 2 [3,4,5,6,7] also mean and median are 5, but these two data sets are different. One data set has no variation whereas other data set varies from value 3

to 7. Thus to understand the variability in the data set there is need to measure the dispersion parameters – range, standard deviation and variance.

Range is the simplest way to calculate the dispersion of the data set. It is calculated by finding the difference between the maximum and the minimum value of the data set.

Standard deviation is the most widely used statistics to measure the dispersion is represented by σ. It is the square root of the averages of square of deviations of every element from the mean. Standard deviation is calculated using equation 3.4.

Equation 3.5 Calculation of Standard Deviation for the individual data set

$$Standard\ deviation\ (\sigma) = \sqrt{\frac{[X_i - \bar{X}]^2}{n}}$$ [4]

Variance is the measure of variability which is calculated as square of the standard deviation. Though both standard deviation and variance of measures of variability, they both differ in their units. Standard deviation is expressed in same units as the values whereas the variance is expressed in larger units.

Measures of Relation -Correlation and Bivariate Regression Analysis

To know the relation between different parameters, correlation coefficient was measured using the Data Analysis – Correlation add in feature of MS Excel. It is important to know the relation between different parameters, so that one parameter can be derived from the other if they are highly correlated.

Correlation is used to find how the given two variables behave with respect to each other. To find the same either scatter plot is useful and helps in estimating the relation direction between two variables. If in scatter plot with increase in one variable, second variable is increasing, it refers to positive correlation and

if second variable decreases with increase in one, it is called negative correlation. Karl Pearson's Coefficient of Correlation is used to find the strength of the correlation [equation 3.5]

Equation 3.6 Determination of Coefficient of correlation between two variables using covariance and standard deviation

$$Coeffcient\ of\ Correlation\ (r) = \frac{Cov\ [X.Y]}{STDEV\ (X).STDEV\ [Y]} \qquad [3.5]$$

Coefficient of correlation is not affected by the change in location of the scale of the values. It is a unit free relationship between the two variables. Value of r lies between [-1 to +1]. If the value of r is close to +1 [-1], it shows a strong positive [negative] correlation between the two variables [2].

Bivariate Linear regression analysis was run for the parameters with strong correlation to obtain a linear regression equation between the parameters.

Multivariate Analysis

The spectral data obtained is complex because of the presence of first overtone, second overtone and combinational bands and it is not possible to analyse and directly find the wavelength set responsible for the quality and quantity analysis of the ground water samples. There are more than two variables for the analysis of the complete data set thus there is need of multivariate analysis where more than two variables can be used to develop regression equations. Camo Unscrambler X 10.5 version was used to do the multivariate analysis for pre-processing of the data, perform principal component analysis [PCA], regression analysis – partial least square [PLS] regression analysis, interval partial least square [iPLS] and synergy interval partial least square [si-PLS] regression analysis.

ANOVA

Analysis of Variance [ANOVA] is a test to check the statistical significance between values of different variables. ANOVA checks the means of different groups to find that variances are similar or not. ANOVA is an inferential test, so hypothesis is tested using ANOVA for a given significance level α. The data set to be used for ANOVA have to from a normal distribution data, selection of the samples should be random, and there should be equal standard deviations among the samples. If ANOVA is to apply on k groups with n as total number of samples, then degree of freedom for between groups is k-1 and within groups is n-k. To test the similarity among the groups using ANOVA, F-distribution / F value is calculated. To calculate the F value, variance between the samples and within the samples is calculated and mean sum of squares for between and within groups is estimated. F value is then calculated using the equation 3.6.

Equation 3.7 Calculation of F statistics value from mean sum of squares for between and within the groups

$$F\ value = \frac{Mean\ sum\ of\ squares\ between\ the\ groups}{Mean\ Sum\ of\ squares\ within\ the\ groups} \qquad [3.6]$$

The calculated F value is then compared with tabular value from F table for given alpha α for degree of freedom [k-1, n-k]. If the calculated value is less than the tabular value then it is considered that there is no significant difference between the values of the groups or vice versa [3].

Principal Component Analysis [PCA] is a multivariate data analysis which is performed for quantitative data analysis. Main of PCA is to reduce the dimensions and noise of the data set and retrieve the important information from a large data set. PC is the projection of data on 2D plane. By taking the average of the horizontal and vertical data average of the sets is done and then a new origin is fixed and the first axis – first principal component [PC] which

maximizes the variance between maximum data set. Subsequent PCs are set orthogonal to the previous PC and maximizes the remaining variance.

After projecting the data in all PCs, relation between the data is obtained among the variables [columns / loadings] and samples [rows / scores]. By accessing the difference between variation data points on the PCA, relationship among different variables or samples can be analysed.

Score plot is the sample PCA plot where the relationships can be found by comparing the distance from the origin and other points.

Loading plot the variables PCA plot, used to interpret relation between variables by accessing the angles between different variables [4]. Drawback of PCA is that it is only used to access the linear relationship between the data sets. PCA may not represent the relationships which are non linear in nature.

Partial least Square Regression

Partial Least Square Regression [PLSR] is a method for soft modeling for constructive predictive models for a large data set with multiple variables and collinear in nature. PLS does not focuses on understanding the relation between the variables and samples, rather its aim is to predict the responses on the basis of predictor variables. PLS developed in the 1950s by Herman Wold as an economic technique for regression analysis is now widely used in predictive modelling in the field of bioinformatics, food and beverage industry, medicinal industry and agricultural field. If the number of factors gets too large i.e. larger than the observation, there are latent factors which can be estimated. General idea of PLS method is to extract these latent variables. Principal of PCA is like robust form of redundancy analysis for the prediction of the variables.

PLS builds a linear predictive model based on the spectrum, where each spectrum contains more than 1000 wavelengths / wavenumbers. PLS factors are first calculated using linear combinations of spectral amplitudes, parameters are then predicted based on these spectral amplitudes. PLS is prediction on the basis of the spectral values in an linear form.

PLS has become an important tool in chemometrics modeling as it helps in deriving hard modeling on the information from the soft models [5]. PLS is an evolving technique. Different software are used for the multivariate analysis using PLS namely – SPSS, R software, MATLAB and CAMO Unscrambler.

In present work, SPSS and CAMO Unscrambler version X 10.5 were used for the PLS modeling of the samples. The Unscrambler software facility was accessed at CSIR-CSIO, Chandigarh. SPSS was employed @LPU.

First Linear regression was employed using SPSS, it was observed that SPSS used wavelengths with high loadings to generate the regression model.

Thus regression analysis was also employed using the CAMO Unscrambler software. The PLS modeling uses all the spectrum for soft modeling to predict the parameters. Several models for PLS, such as ipLS, siPLS, backward PLS, forward PLS are used by researchers to reject the wavelengths which are not important for a particular parameter and are only responsible for noise in the data set.

Interval PLS [i-PLS]

Interval PLS [i-PLS] is a wavelength selection method mainly employed for spectroscopic data for qualitative and quantitative analysis. Main idea of this method is to apply regression models on different parts of the data and at specific intervals to find the most important wavelength required for prediction of a particular parameter. iPLS is employed in steps – [a] the complete spectral data set is split in N intervals of equal width [the gap is same between the two wavelengths / wavenumbers of the data], [b] Regression models are applied in forward direction on the intervals starting from lowest to highest wavelengths. The last step is to compare the regression models of all the intervals and neglecting the models with less value of correlation coefficient [in general less than 0.4 or 0.5], or high root mean square error [RMSE] [6].

Synergy Interval PLS [siPLS]

Synergy interval PLS [siPLS] regression method uses the combination of different equally placed intervals together to develop a regression model. siPLS

is employed to obtain robust models with less error as least important wavelengths are rejected and important wavelengths are combined together to develop the regression models. The regression model with lowest RMSE is selected for prediction of that particular parameter [7].

Use of iPLS and siPLS methods for regression models helps in development of fast and robust model. These techniques have been used by many authors to create a regression method with selected wavelengths resulting in low cost, rapid and robust models [8].

In present PhD research work these methods have been used for sample procurement, data collection, analysis for classification, correlation an regression analysis.

References

[1]. Sahoo, Prafulla Kumar, et al. "Meta-analysis of uranium contamination in groundwater of the alluvial plains of Punjab, northwest India: Status, health risk, and hydrogeochemical processes." *Science of The Total Environment* 807 (2022): 151753. doi: 10.1016/j.scitotenv.2021.151753. Epub 2021 Nov 23.

[2]. C R Kothari, Gaurav Garg, research Methodology Methods and Techniques, Fourth Edition, New Age International Publishers, 2019.

[3]. Barbara Illowsky, Susan Dean, Introductory Statistics, 13.2 The F Distribution and the F-Ratio, ISBN-13: 978-1-50669-823-6 by OpenStax is licensed under Creative Commons Attribution License v4.0

[4]. https://bookdown.org/brian_nguyen0305/Multivariate_Statistical _Analysis_with_R/what-is-pca.html

[5]. Tobias, Randall D. "An introduction to partial least squares regression." *Proceedings of the twentieth annual SAS users group international conference*. Vol. 20. Cary, NC, USA: SAS Institute Inc., 1995.

[6]. Kucheryavskiy, Sergey. "mdatools–R package for chemometrics." *Chemometrics and Intelligent Laboratory Systems* 198 (2020): 103937.

[7]. Zuo, Xiao-Bo, Sheng Fang, and Xian-Li Liang. "Synergy interval partial least square (siPLS) with potentiometric titration multivariate calibration for the simultaneous determination of amino acids in mixtures." *Advance Journal of Food Science and Technology* 6.11 (2014): 1209-1218.

[8]. da Silva, Daniel José, and Hélio Wiebeck. "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 (2017): 259-266.

Chapter 4 – Identification of pollutants - heavy metals in ground water by using chemical methods for the Malwa belt.

4. Chapter 4 – Identification of pollutants - heavy metals in ground water by using chemical methods for the Malwa belt.

Introduction

Contamination of ground water by arsenic, urea and other pollutants has been reported in many countries. The concentration of these pollutants in ground water is due to the industrial waste, fertilizers and pesticides used in agriculture. Interaction of matter and electromagnetic radiations in the infrared range helps the sample undergo electronic, vibrational and rotational transitions, which helps in understanding the structure, bonds present in the sample. Fourier transform infrared (FTIR) spectroscopy is a rapid, quantitative powerful chemical analytical technique. Near Infra-red (NIR) spectroscopy is also a strong technique to find physicochemical propertied of the samples in a rapid and non-destructive way.

Urea, arsenic and other contaminants are harmful for human body and there is need to check the presence of their concentration in the ground water samples. Along with the ground water soil is also contaminated and a lot of research is carried out to determine the presence of these contaminants in soil and water. A research was carried where 165 samples were collected from a particular region, analysis of nutrient content was done using PCA and PLS [1-4]. Measurement of nitrogen content in the samples also help in determining the presecnce of contaminants in samples [5-6]. Arsenic content in ground water was studied for the ground water samples of Bihar, India. Arsenic content was measured by Silverdiethyledithiocarbamate method at 520 nm in UV spectroscopy. It was reported that As content in some samples were more than the permissible limit of WHO (10µg/L) [7]. Study of water samples of Sunderban Delta, India were also reported in literature. Review of literature was presented related to contamination of ground water samples in the region [8]. Further it was reported that the contamination of the ground water samples leads to severe changes in the physicochemical properties – pH, TDS, EC, chloride content, chemical oxygen demand (COD) and dissolved oxygen (DO) in the water samples [9].

The chapter focuses on the work carried out to identify the bonds of the contaminants present in groundwater using the FTIR and NIR spectroscopy.

Results and Discussion

Fourier transform infrared spectroscopy of pure Urea and arsenic trioxide was obtained using the ATR- FTIR spectrometer at CIF lab, Lovely Professional University. The FTIR spectra of the pollutants showed the signature peaks in the transmittance spectra at characteristics wavenumber. These spectral peaks were then compared with the spectral peaks and bands of FTIR and NIR spectra of the ground water samples, to identify bonds of these pollutants in the ground water samples procured from the two districts of Malwa region of Punjab. Figure 4.1 shows the FTIR spectra of urea.

FTIR analysis

Urea consists of C-N bond, C-O bond, N-H and C=O bond. According to literature and the FTIR spectra of urea it can be observed that the N-H stretching modes appear around $3300 – 3400$ cm^{-1}. Two more bonds appear due to in phase stretching vibration around $3000-3300$ cm^{-1}. Further a peak around 1670 cm^{-1} was observed corresponding to the C=O and NH_2 motions. Peaks observed between $1500-2000$ cm-1 are observed due to C=O, C-N and N-H stretching vibrations. These peaks were used to check the presence of urea in ground water samples. The signature wavenumbers observed from the FTIR spectra were also compared with the 1st and 2nd harmonics wavelengths from the NIR spectra of ground water sample. Signature / fundamental wavelength vibration when divided by 2 gives the 1st overtone wavelength, divided by 3 given 2nd overtone and so on calculation was done for higher harmonics.
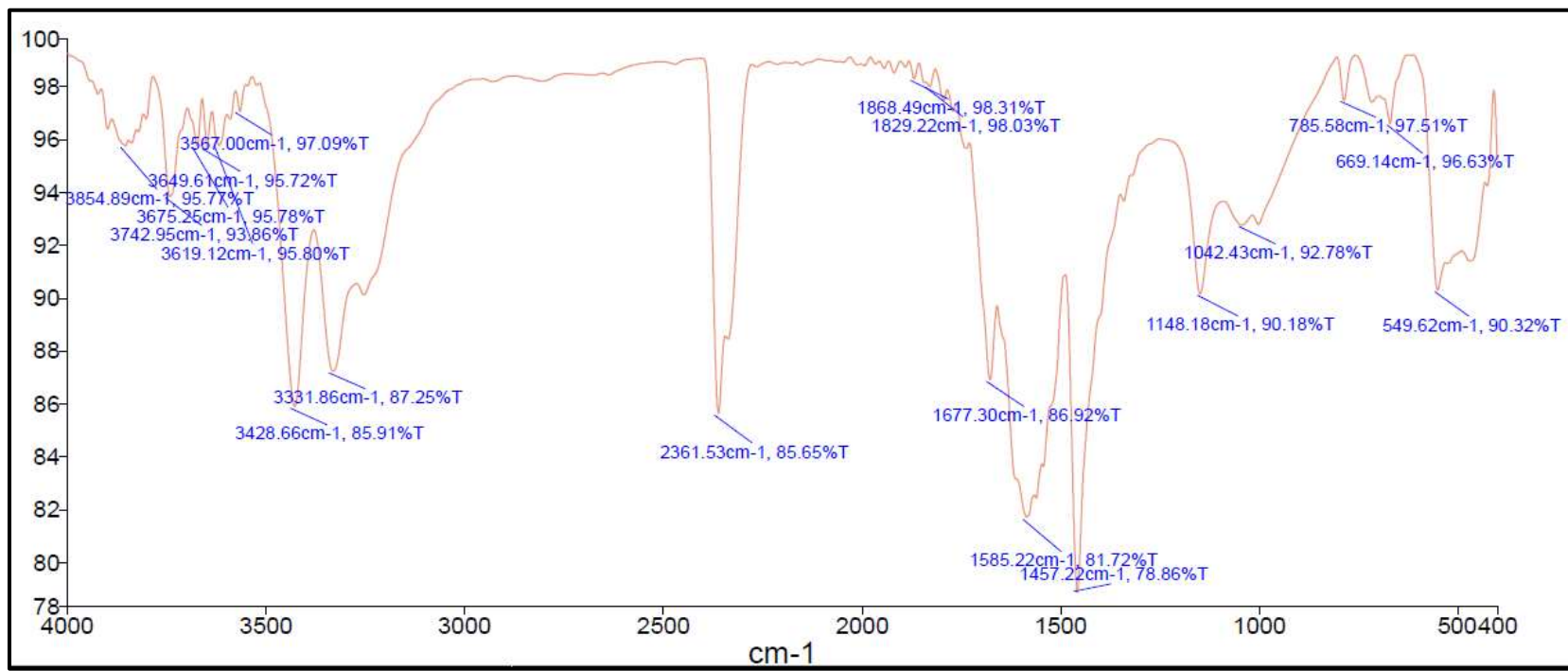
Figure 4.1 Fourier Transform spectra of Urea using the ATR-FTIR spectrometer at CIF, LPU
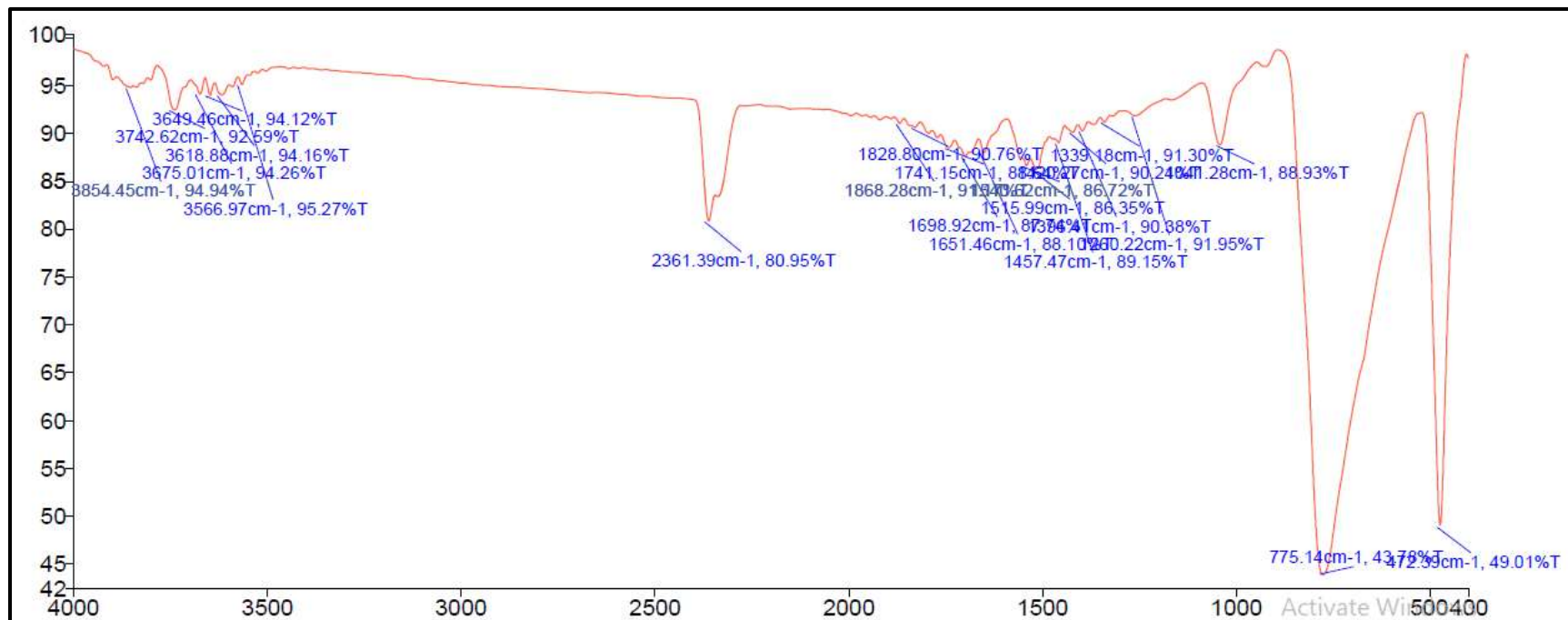
Figure 4.2 Fourier Transform spectra of As_2O_3 using the ATR-FTIR spectrometer at CIF, LPU

Identification of Urea using Fourier Transform Spectra

Arsenic trioxide structure consists of As=O, and As-O bonds present in it. Signature peaks observed from FTIR spectra for As_2O_3 were around 3500-4000, 2300-3400 and 770cm^{-1} in the transmittance spectra (Fig 4.2). Further peaks of very high transmittance were observed in between 1000-2000cm^{-1}. These signature wavenumbers were then compared with the peaks in the FTIR of the ground water samples and their overtone wavelengths were compared with the NIR spectra of the samples.

The FTIR spectra of 3 random samples of ground water collected from the two districts of Malwa region of Punjab are shown in fig 4.3 - 4.5.

From the FTIR spectra of ground water samples it was observed that there was one peak observed witth minimum transmittance (maximum absorbance) from 400-500cm^{-1}, Two peaks around 3000-3500 cm^{-1} and 1500-1750 cm^{-1} with transmittance % between 50-70%. Continuous peaks were observed in the region from 1750-2500cm^{-1} with high transmittance % of value above 90%.

A comparison was done for all the ground water sample peaks with Urea and Arsenic trioxide for their FTIR spectra. Table 4.1 shows the complete comparison details for all the samples.
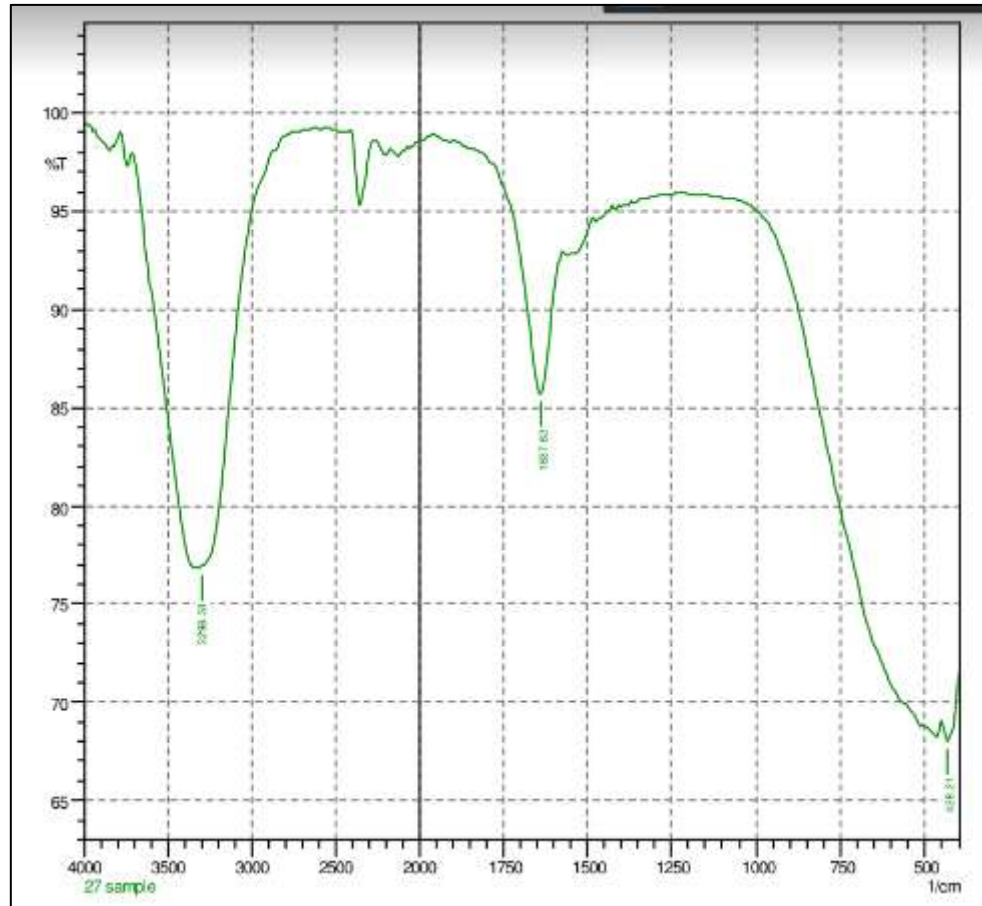
Figure 4.3 FTIR spectra of sample no-27 of the ground water samples collected in transmittance mode
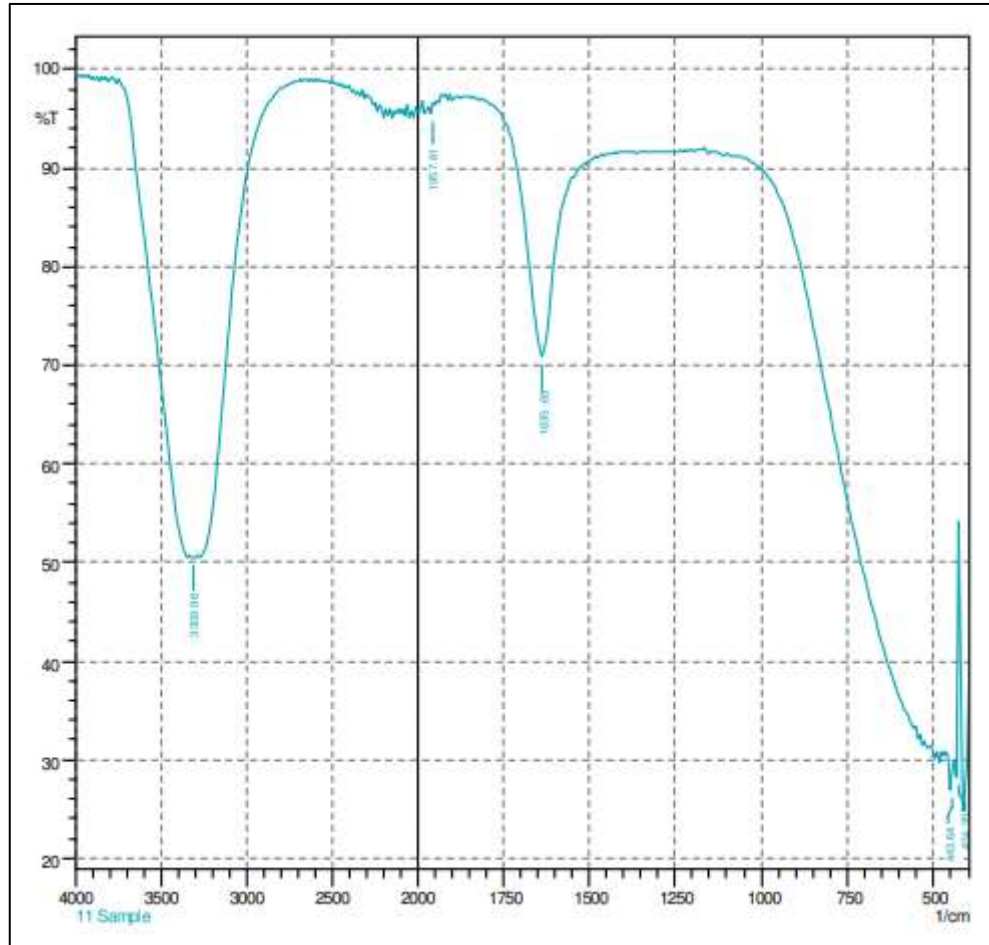
Figure 4.4 FTIR spectra of sample no-11 of the ground water samples collected in transmittance mode
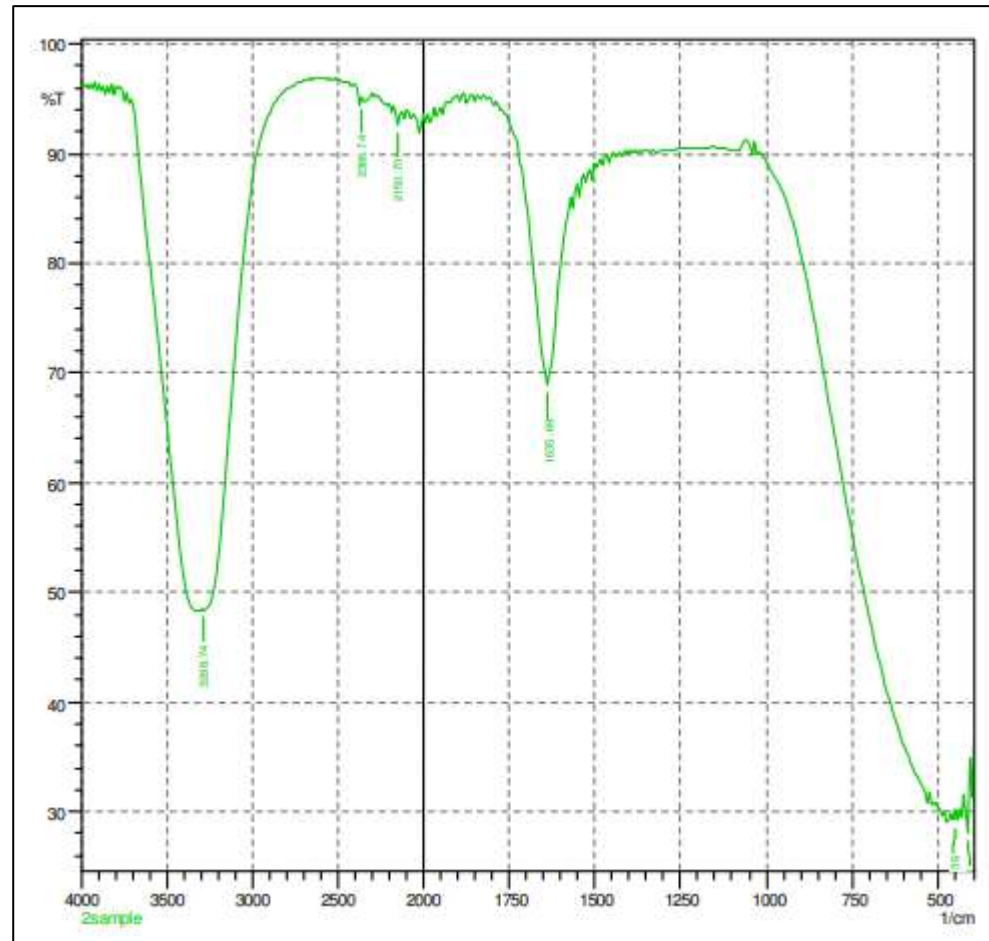
Figure 4.5 FTIR spectra of sample no-2 of the ground water samples collected in transmittance mode

Table 4.1 Identification for presence of urea and arsenic tri oxide in all the sample by comparing their FTIR spectra

Sample id	Depth(feet)	Source	IRU(Region)	Place	Sample No.	Urea	As$_2$O$_3$
1AR	50	HP	R	A	1	Present	Present
2AR	28	HP	R	A	2	Present	Present
3AR	40	HP	R	A	3	Present	Present
4AI	40	HP	I	A	4	Present	Present
5AR	60	SS	R	A	5	Not Present	Not Present
6AU	160	SS	U	A	6	Present	Not Present
7AU	60	SS	U	A	8	Not Present	Not Present
8AR	95	SS	R	A	9	Not Present	Not Present
9AR	450	WB	R	A	10	Present	Present
10AR	108	SS	R	A	11	Present	Not Present
11AR	410	WB	R	A	12	Present	Not Present
12AI	200	SS	I	A	13	Present	Not Present
13AI	220	SS	I	A	14	Present	Present
14AR	400	WB	R	A	15	Present	Not Present
15AI	200	SS	I	A	16	Not Present	Not Present
16AI	205	SS	I	A	17	Present	Not Present
17AI	220	SS	I	A	18	Present	Not Present
18AI	230	SS	I	A	19	Present	Not Present
19AI	250	SS	I	A	20	Not Present	Not Present
20AI	250	SS	I	A	21	Present	Not Present
21AU	220	SS	U	A	22	Present	Not Present
22AU	120	SS	U	A	23	Present	Not Present

23AU	500	WB	U	A	24	Present	Not Present
24AU	385	SS	U	A	25	Not Present	Present
25AU	500	SS	U	A	27	Present	Not Present
26AU	450	WB	U	A	28	Present	Present
27AU	550	WB	U	A	30	Present	Present
28BR	550	WB	R	B	31	Present	Not Present
29BI	100	SS	I	B	32	Present	Present
30BI	260	SS	I	B	33	Not Present	Not Present
31BR	450	WB	R	B	34	Not Present	Not Present
32BR	100	SS	R	B	35	Not Present	Not Present
33BR	120	SS	R	B	36	Not Present	Not Present
34BR	90	SS	R	B	37	Present	Not Present
35BR	90	SS	R	B	38	Not Present	Present
36BI	220	SS	I	B	39	Not Present	Not Present
37BI	140	SS	I	B	40	Not Present	Not Present
38BR	350	WB	R	B	41	Not Present	Not Present
39BI	120	SS	I	B	42	Present	Present
40BR	85	SS	R	B	43	Not Present	Not Present
41BU	450	WB	U	B	44	Present	Present

						Not	
42BU	340	WB	U	B	46	Present	Not Present
						Not	
43BI	180	SS	I	B	47	Present	Not Present
						Not	
44BI	150	SS	I	B	48	Present	Not Present
45BU	550	WB	U	B	49	Present	Not Present
						Not	
46BU	500	WB	U	B	50	Present	Not Present
						Not	
47BU	550	WB	U	B	52	Present	Not Present
48BU	450	SS	U	B	54	Present	Not Present
49BU	250	SS	U	B	55	Present	Not Present
						Not	
50BU	495	WB	U	B	56	Present	Not Present

If all the peaks with high values of transmittance were to be considered (wavenumbers with least absorbance) then it may be considered that, there is presence of arsenic tri oxide in all samples. The small absorbance peaks shows that the presence of As_2O_3 is very small or minimal amount.

Identification of bonds using FTIR and NIR Spectra

On the basis of the presence of peaks, bond identification was carried out from the literature at each wavenumber. Results for the same are listed in Table 4.2.

Table 4.2 Bond identification on the basis of peaks observed for FTIR spectra of all the samples

Sample id	Depth(feet)	Source	IRU(Region)	Place	Sample No.	Bond Identification
1AR	50	HP	R	A	1	1. C-N ring formation, N-O deformation, CH_2 wagging 2. Saturated amide $O = C <$ 3. C≡N Alkynes 4. R-OH alchols
2AR	28	HP	R	A	2	1. C-N ring formation, N-O deformation, CH_2 wagging 2. Saturated amide $O = C <$ 3. C≡C 4. -C≡N 5. -C≡C-H
3AR	40	HP	R	A	3	1. C-N ring formation, N-O deformation, CH_2 wagging 2. Saturated amide $O = C <$ 3. -C≡N 4. -C≡C-H
4AI	40	HP	I	A	4	
5AR	60	SS	R	A	5	

6AU	160	SS	U	A	6		
7AU	60	SS	U	A	8		
8AR	95	SS	R	A	9		
9AR	450	WB	R	A	10		
10AR	108	SS	R	A	11		
11AR	410	WB	R	A	12		
12AI	200	SS	I	A	13		
13AI	220	SS	I	A	14		
14AR	400	WB	R	A	15		
15AI	200	SS	I	A	16		
16AI	205	SS	I	A	17		
						1. Saturated amide $O = C <$ 2. N-O deformation	
17AI	220	SS	I	A	18		
18AI	230	SS	I	A	19		
19AI	250	SS	I	A	20		
20AI	250	SS	I	A	21		
21AU	220	SS	U	A	22		
22AU	120	SS	U	A	23		
						1. Saturated amide $O = C <$ 2. C-N ring formation, N-O deformation, CH_2 wagging	
23AU	500	WB	U	A	24		
24AU	385	SS	U	A	25		
25AU	500	SS	U	A	27		
26AU	450	WB	U	A	28		
27AU	550	WB	U	A	30		
28BR	550	WB	R	B	31	1. Saturated amide	

						$O = C <$ 2. -C≡C-
29BI	100	SS	I	B	32	
						1. Saturated amide $O = C <$ 2. C-N ring formation, N-O deformation, CH_2 wagging 3. Alkynes
30BI	260	SS	I	B	33	
31BR	450	WB	R	B	34	
32BR	100	SS	R	B	35	
33BR	120	SS	R	B	36	
34BR	90	SS	R	B	37	
						1. Saturated amide $O = C <$ 2. -C≡C- 3. -C≡N
35BR	90	SS	R	B	38	
36BI	220	SS	I	B	39	
37BI	140	SS	I	B	40	
38BR	350	WB	R	B	41	
39BI	120	SS	I	B	42	
40BR	85	SS	R	B	43	
						1. Saturated amide $O = C <$ 2. No triple bond exists
41BU	450	WB	U	B	44	
42BU	340	WB	U	B	46	1. Saturated amide $O = C <$

						1. Saturated amide $O = C <$
						2. C-N ring formation, N-O deformation, CH$_2$ wagging
43BI	180	SS	I	B	47	
44BI	150	SS	I	B	48	
						1. Saturated amide $O = C <$
						2. -C≡C-
						3. Amines two bonds occur , 1 bond for 1[0] amain and one for 2[0] amine, one is weaker than other
45BU	550	WB	U	B	49	
						1. Saturated amide $O = C <$
						2. C-N ring formation, N-O deformation, CH$_2$ wagging
46BU	500	WB	U	B	50	
47BU	550	WB	U	B	52	
48BU	450	SS	U	B	54	
49BU	250	SS	U	B	55	
50BU	495	WB	U	B	56	

Summary

Present work was carried out to identify pollutants namely – urea and arsenic trioxide in the ground water samples. FTIR spectra of the pollutants and the ground water samples were compared to check the presence of the bonds of Urea and arsenic trioxide in ground water samples. It was observed that peaks were observed around $3300cm^{-1}$, which were similar to present in Urea. Bands in 1700-2500 cm-1, were showing the presence of arsenic trioxide in the sample. It was observed that for all the samples irrespective of their place of collection, there was presence of peaks which showed the presence of Urea and arsenic trioxide. Further the first and second harmonic bands / peaks were checked from the NIR spectra of the ground water samples. It was observed that there was presence of urea and arsenic trioxide present in all the samples. It was also observed that samples with high TDS and EC values had the presence of Urea and arsenic from the literature.

References

[1].   Y. He, M. Huang, A. García, et al. "Prediction of soil macronutrients content using near infrared spectroscopy," Computers and Electronics in Agriculture, vol. 58, no. 2, pp. 144–153, 2007.

[2].   V. Bellon-Maurel, E. Fernandez-Ahumada, et al, "Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy," TrAC Trends in Analytical Chemistry, vol. 29, no. 9, pp. 1073–1081, 2010.

[3].   P. Nie, T. Dong, Y. He and F. Qu, "Detection of soil nitrogen using near infrared sensors based on soil pretreatment and algorithms," Sensors, vol. 17, no. 5, pp. 1–12, 2017.

[4].   D. Cozzolino and A. Moron, "Influence of soil particle size on the measurement of sodium by near-infrared reflectance spectroscopy," Communication in Soil Science and Plant Analysis, vol. 41, no. 19, pp. 2330–2339, 2010.

[5].   E. Ben-Dor, S. Chabrillat, J. A. M. Demattê, G. R. Taylor, et al., "Using imaging spectroscopy to study soil properties," Remote Sensing of Environment, vol. 113, no. 1, pp. S38–S55, 2009.

[6].   R. F. Muñoz-Huerta, R. G. Guevara-Gonzalez, L. M. Contreras-Medina, I. Torres-Pacheco, J. Prado-Olivarez et al., "A review of methods for sensing the nitrogen status in plants: Advantages, disadvantages and recent advances," Sensors, vol. 13, no. 8, pp. 10823–10843, 2013

[7].   Singh, S. K., et al. "Groundwater arsenic contamination and associated health risks in Bihar, India." *International Journal of Environmental Research* 8.1 (2014): 49-60.

[8].   Nath, Anirban, et al. "Threat of arsenic contamination, salinity and water pollution in agricultural practices of Sundarban Delta, India, and mitigation strategies." *SN Applied Sciences* 3.5 (2021): 1-15. https://doi.org/10.1007/s42452-021-04544-1

[9].   Ranjana, Agrawal. "Physico-chemical analysis of some groundwater samples of Kotputli town Jaipur, Rajasthan." *International Journal of Chemical Environmental and Pharmaceutical Research* 1.2 (2010): 111-113.

Chapter 5 – Measurement of Physicochemical Properties of ground water samples and classification on the basis of their region, source and depth using Principal Component Analysis

5. **Chapter 5– Measurement of Physicochemical Properties of ground water samples and classification on the basis of their region, source and depth using Principal Component Analysis**

Groundwater is a valuable natural resource that plays a vital part in the economy. Ground water is majorly used for agriculture, industrial and domestic purpose. However, in developing countries, where groundwater has long been an essential source of water, are experiencing major groundwater pollution concerns, owing primarily to human activity. Groundwater is affected by human activities such as urbanisation, industrialization, and agriculture. In rural area, people are using groundwater for drinking and domestic purpose without knowing the suitability and quality of water in their region. Research work was carried out to estimate the salinity of the ground water using the TDS and EC concentration of the ground water samples collected from Haryana, India. It was observed that hand pump water was more saline than bore well and dug well water. Work was useful to study depth and spatio-temporal variation of ground water samples [1]. In another research GIS data was used to estimate the ground water quality using the TDS values of the samples. Research focused on testing the vulnerability of ground water of south Darb El Arbaein for irrigation. [2].

The present part of the research work focuses on measurement of the physical and chemical parameters of the ground water collected from two districts of Malwa region of Punjab, India.

Physicochemical measurements

Physical and chemical properties of ground water samples measured for the study were – pH, TDS, EC and chloride content from Soil and Water testing Lab, Jind, Haryana.

The physical and chemical parameters of all the samples are listed in Table 5.1.

Table 5.1 Physicochemical parameters of all the 50 ground water samples

Sample Id	Place	IRU [Region]	Depth [feet]	Source	pH	TDS[mg/L]	EC [µS/cm]	Chloride [mg/L]
1AR	A	R	50	HP	7.2	1398	4520	1109.56
2AR	A	R	28	HP	7.1	454	834	184.93
3AR	A	R	40	HP	7	1280	3760	839.66
4AI	A	I	40	HP	6.9	151	250	59.98
5AR	A	R	60	SS	7.3	548	1034	234.91
6AU	A	U	160	SS	7	676	1345	294.88
7AU	A	U	60	SS	7.2	1134	3070	709.72
8AR	A	R	95	SS	7.5	1236	3410	834.67
9AR	A	R	450	WB	7.5	683	1375	284.89
10AR	A	R	108	SS	7.3	1124	2940	709.72
11AR	A	R	410	WB	7.4	411	730	159.94
12AI	A	I	200	SS	7.4	782	1674	394.84
13AI	A	I	220	SS	7.6	667	1350	324.87
14AR	A	R	400	WB	7.6	464	860	194.93
15AI	A	I	200	SS	7.3	407	730	174.93
16AI	A	I	205	SS	7	341	606	149.94

17AI	A	I	220	SS	7.3	280	482	109.96
18AI	A	I	230	SS	7.3	556	1050	279.89
19AI	A	I	250	SS	7.5	501	930	224.91
20AI	A	I	250	SS	7.4	753	1550	344.86
21AU	A	U	220	SS	7.8	798	1690	379.85
22AU	A	U	120	SS	7.6	743	1552	359.86
23AU	A	U	500	WB	7.3	666	1340	304.88
24AU	A	U	385	SS	7.1	656	1307	299.88
25AU	A	U	500	SS	7.4	952	2400	589.76
26AU	A	U	450	WB	7.2	628	1220	304.88
27AU	A	U	550	WB	7.1	706	1420	349.86
28BR	B	R	550	WB	7.4	377	653	159.94
29BI	B	I	100	SS	7.4	366	630	154.94
30BI	B	I	260	SS	7.4	1017	2410	569.77
31BR	B	R	450	WB	7.6	519	1027	249.9
32BR	B	R	100	SS	7.2	1088	2890	694.72
33BR	B	R	120	SS	7.1	649	1370	324.87
34BR	B	R	90	SS	7.1	547	1130	274.89
35BR	B	R	90	SS	7.2	367	700	159.94

36BI	B	I	220	SS	7.4	309	546	119.95
37BI	B	I	140	SS	7.4	172	301	69.97
38BR	B	R	350	WB	7.4	184	305	74.97
39BI	B	I	120	SS	7.3	508	950	209.92
40BR	B	R	85	SS	7.4	165	290	64.97
41BU	B	U	450	WB	7.6	313	540	124.95
42BU	B	U	340	WB	7.4	281	490	114.95
43BI	B	I	180	SS	7.2	439	800	184.93
44BI	B	I	150	SS	7.2	659	1362	319.88
45BU	B	U	550	WB	7	351	640	149.94
46BU	B	U	500	WB	7.4	317	550	109.96
47BU	B	U	550	WB	6.9	430	780	184.93
48BU	B	U	450	SS	7.1	708	1424	354.86
49BU	B	U	250	SS	7.5	370	664	154.94
50BU	B	U	495	WB	7.6	361	650	149.94

Table 5.2 Descriptive analysis of water quality parameters

Statistical parameters/Water quality parameters	pH	TDS[#] [mg/L]	EC[#] [µS/cm]	Chloride content [mg/L]
Mean	7.31	589.84	1290.62	303.78
Standard Error	0.03	43.15	134.51	32.27
Standard Deviation	0.21	305.08	951.14	228.16
Sample Variance	0.04	93076.06	904660.16	52057.56
Range	0.90	1247	4270	1049.58
Minimum	6.90	151	250	59.98
Maximum	7.80	1398	4520	1109.56

[#]TDS [Total Dissolved Solids]; [#]EC [Electrical Conductivity]

All the water samples were tested for pH, TDS, EC and Chloride content. The pH of water samples was measured using digital pH meter. As discussed in literature review it is important to study pH and TDS to know the quality of the water. EC and chloride content are further measured to check the characteristics of all the 50 collected samples. The parameters were compared with the standard measurement as given by WHO [3-4]. Descriptive statistics – measures of central tendency and measures of variation were calculated using the descriptive analysis add in tab for MS excel and results for the same are shown in Table 5.2. It was observed that the TDS values of water samples varied from 151mg/L to 1398 mg/L. Mean and standard deviation of TDS was found 589.84mg/L and 305.08 mg/L respectively. Large value of standard deviation showed that as the samples were collected from different depths thus the TDS values are having higher variation. Such large variations have been observed for the research done on ground water collected from Mewat district of Haryana [5]. Research carried on samples of Punjab, also showed high variation in TDS samples and it was found that the water was not suitable for irrigation or drinking [6].

Similar variation was observed in EC. EC for the collected samples lies between 250µS/cm and 4520µS/cm. Mean and standard deviation of EC was found 1290.62 µS/cm and 951.14 µS/cm respectively. Chloride content also showed large variations lies between 1109.56mg/L and 59.98mg/L. Mean and standard deviation of chloride was found 303.78mg/L and 228.16mg/L respectively. pH values for the ground water samples had least or negligible variation. pH values lies between 6.9 and 7.8. Mean and standard deviation of pH was found 7.31 and .21 respectively.

Impact of source of ground water on parameters of ground water

The literature revealed that there is a relation between the source of the ground water sample and TDS values. Thus Analysis of Variance was applied on the three different sources – hp [hand pump], ss [submersibal] and wb [waterbox]. Null Hypothesis for ANOVA was that there is no difference between the TDS values of the samples with respect to the sources of collection. It was observed

from ANOVA results that calculated F value was more than the critical value for degree of freedom [4, 31] and also p value for analysis was less than 0.05 [Table 5.3]. Thus results revealed that the TDS values of the water collected from different sources have significant difference between their values and null hypothesis was rejected. Similar variation in TDS values were also observed by a research on the ground water samples collected from Great Britain. Results showed that distribution of ground water depends upon the hydrogeological setting and TDS provide a useful explanation for the source of collected samples [7]. ANOVA was also employed on the EC, chloride content and pH values of the samples to check their variation on the basis of source of the collected samples. Results shown in Table 5.4, 5.5 and 5.6 for variation of EC, chloride content and pH respectively and the results showed that though there is large variance for the EC, TDS and chloride content values as compared to the pH values for the three sources but all the three sources have significant difference between their values.

Table 5.3 Analysis of variance results for the TDS variation for different sources of procured samples

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
hp	4	3283	820.75	375732.9
ss	31	19518	629.6129	83250.45
wb	15	6691	446.0667	25974.07

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	572377.7	2	286188.8	3.372542	0.042789681	3.195056
Within Groups	3988349	47	84858.49			
Total	4560727	49				

Table 5.4 Analysis of variance results for the EC variation for different sources of procured samples

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
hp	4	9364	2341	4468311
ss	31	42587	1373.774	715666
wb	15	12580	838.6667	125854.4

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	7691473	2	3845737	4.933544	0.011352	3.195056
Within Groups	36636875	47	779508			
Total	44328348	49				

Table 5.5 Analysis of variance results for the chloride content variation for different sources of procured samples

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
hp	4	2194.13	548.5325	256799.1
234.91	30	9841.09	328.0363	42785.66
wb	15	2918.86	194.5907	7056.799

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	436004.1	2	218002.1	4.752704	0.013296	3.199581706
Within Groups	2109977	46	45869.06			
Total	2545981	48				

Table 5.6 Analysis of variance results for the chloride content variation for different sources of procured samples

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
hp	4	28.2	7.05	0.016667
Ss	31	226.9	7.319355	0.033613
wb	15	110.4	7.36	0.049714

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	0.310613	2	0.155306	4.160657	0.021693	3.195056
Within Groups	1.754387	47	0.037327			
Total	2.065	49				

Variation of the physicochemical parameters with depth

To check the effect of depth of the collected samples on values of EC, TDS and chloride content, graph was plotted for these variables with respect to the depth of the samples. As shown in Fig. 5.1, the quality parameters of water are significantly dependent on the depth of the sample collection site. It was observed that as the depth of ground water sample increases, the range of quality parameters under investigation decreases. It was also observed that the quality parameters of water samples collected from depth greater than 100 feet were in the permissible range. This trend observed was observed in case of EC, TDS and chloride content. The water samples with EC values out of permissible limit were further analyzed and it was observed that the ground water sample collected from depth more than 100 feet exhibit values in the standard range.
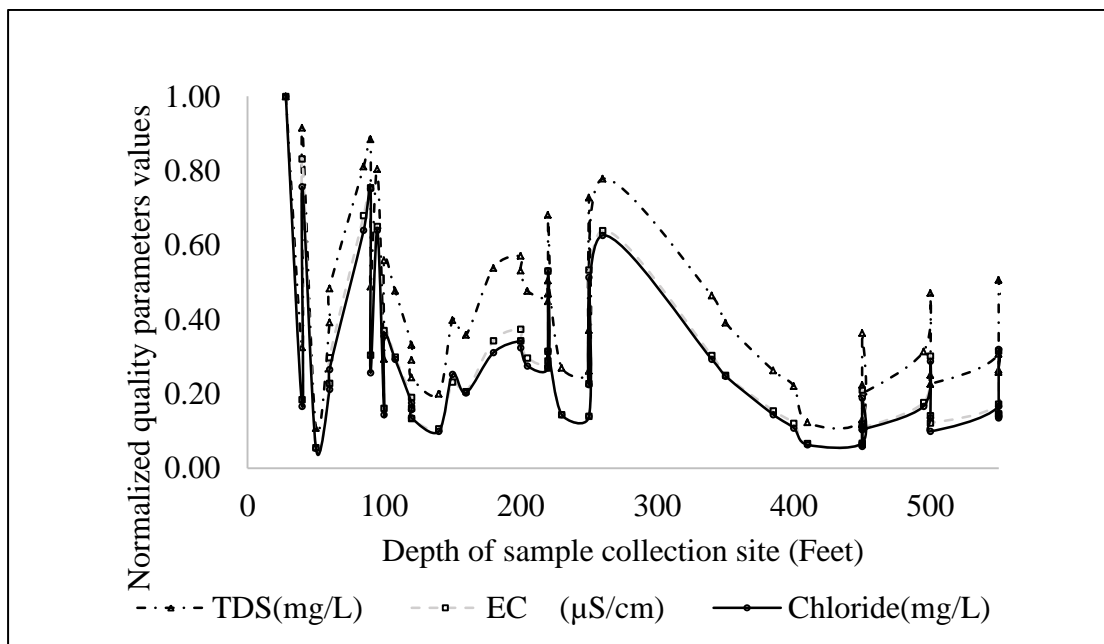
Figure 5.1 Variation among the normalized TDS, EC and chloride content values with depth of the sample collection site

Correlation and regression Analysis

As the variations showed that the TDS, EC and chloride content had similar trends, thus decision was made to check the correlation among the samples and then develop a regression equation for the samples having coefficient of correlation values higher than |0.6|. However, no or minimal correlation was observed among pH and other parameters of water samples [Table 5.7]. Whereas, a strong correlation [r > 0.9] was observed between the TDS, EC and chloride content values. Thus, further regression analysis was performed to explore the relation between these three parameters. Table 5.7 shows the correlation values as obtained from MS Excel data analysis add in tab.

Table 5.7 Correlational analysis for all the parameters

Quality parameters	pH	TDS [mg/L]	EC [µS/cm]	Chloride [mg/L]
pH	1			
TDS[ppm]	-0.0493	1		
EC [µS/cm]	-0.0782	0.9793	1	
Chloride	-0.0795	0.97482	0.9976	1

Regression Analysis

The regression analysis of quality parameters resulted in following regression equations between TDS and EC, TDS and chloride content, EC and chloride content. These equations provide the information about mutual relation between these parameters. Line plots are shown in Fig 5.2. Thus, it can be concluded that by measuring one parameter using chemical methods, the other two quality parameters of water samples can be successfully predicted by employing equations 1, 2 and 3 in a non-destructive way with good accuracy. Equation 1 revealed the relation between EC and TDS values with slope of 3.0532 and intercept of 510.27. Whereas, equation 2 shows the relationship between chloride content and TDS values of water samples with slope of 0.729 and intercept of 126.23 and the equation 3 presents the relation between chloride content and EC values with slope of 0.2393 and intercept of 5.0752.
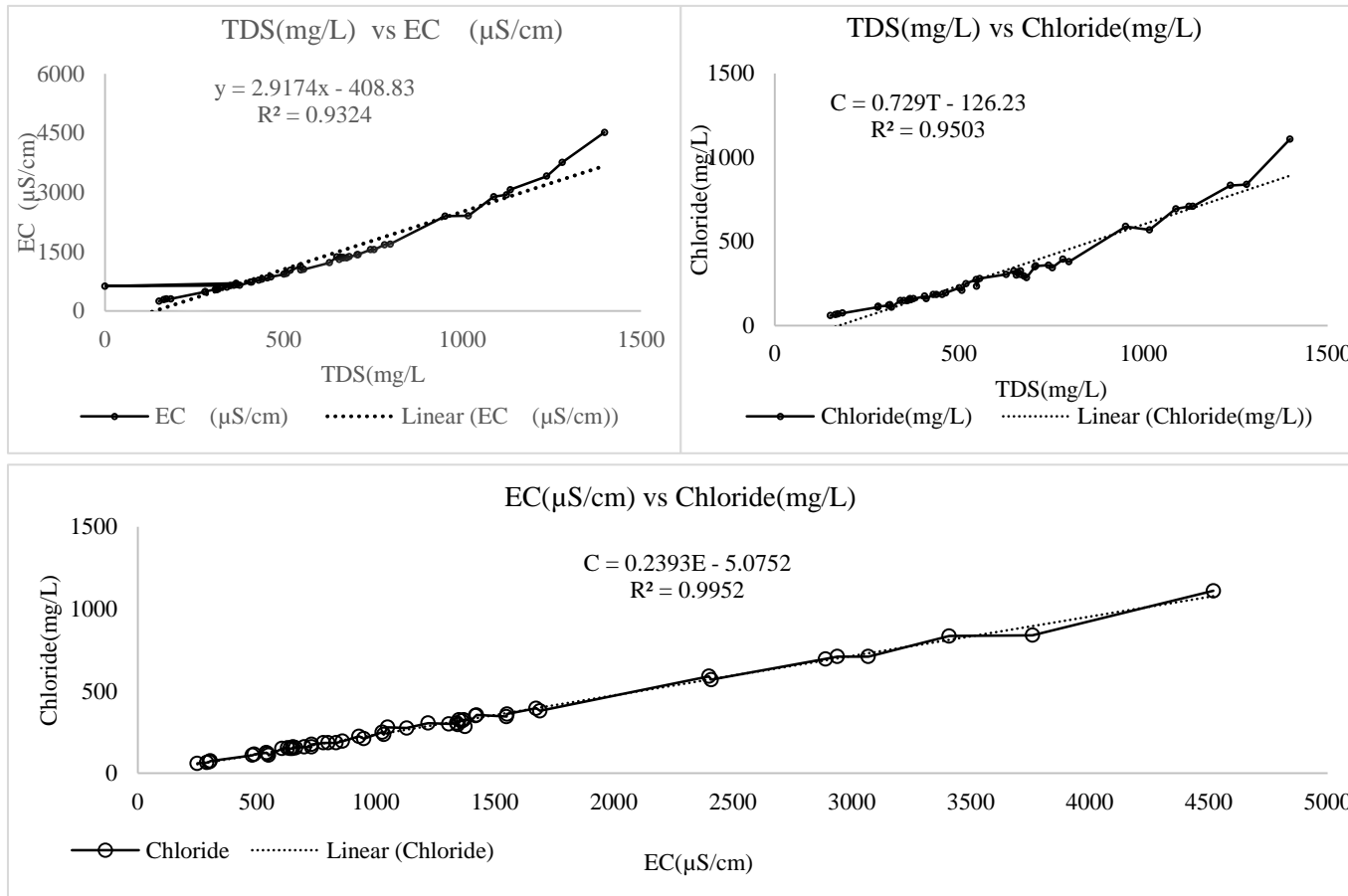
Figure 5.2 Line plots for the regression for TDS vs EC, TDS vs Chloride content and EC vs chloride content

Equation 5.8 Regression equation between Electrical Conductivity E and TDS values

$$E = 3.0532T - 510.27 \qquad [5.1]$$

Equation 5.9 Regression equation between Chloride content [C] and TDS values

$$C = 0.729T - 126.23 \qquad [5.2]$$

Equation 5.10 Regression equation between Chloride content [C] and Electrical Conductivity EC

$$C = 0.2393E - 5.0752 \qquad [5.3]$$

Where, E represents EC, T is used for TDS and C presents chloride content of water samples.

Furthermore, Analysis of Variance [ANOVA] was performed on the three quality parameters namely: EC, TDS and chloride content of water as shown in Table 4.8. It was observed that F value for EC and chloride content was less than the F observed, which shows that there is no difference between these two parameters. Henceforth, the two parameters EC and chloride content of water samples under investigation exhibit similar values.

Table 5.8 ANOVA results for the EC and chloride content of the samples

SS [sum of squares]; df [degree of freedom]; MS [Mean Squares]; F [F value].

Source of Variation	SS	df	MS	F	p-value	F crit
Between Groups	0.003453	1	0.003453	0.079767	0.778209	3.938111
Within Groups	4.241687	98	0.043283			
Total	4.24514	99				

Classification of samples using Principal Component Analysis

The data was maximum normalized for the further analysis. Principal Component Analysis [PCA] was performed on the physio-chemical data of water samples using ClustVis online platform [https://biit.cs.ut.ee/clustvis/]. The collected data for physico-chemical parameters of water samples was divided in three regions namely, urban, rural and industrial for the two districts. Furthermore, the PCA analysis was performed to classify the water samples based on their collection regions [urban / rural / industrial], district [Ferozepur / Fazilka] wise.

District wise classification using the physicochemical data

For PCA analysis, unit variance scaling was applied to rows containing different samples; SVD [Singular value decomposition] with imputation was used to calculate principal components. X and Y axis presented principal component 1 and principal component 2 that explain 96.9% and 3% of the total variance, respectively. N = 4 data points. As shown in Fig. 5.3, the PCA results are in support with the results of correlation and regression analysis. It was observed that the EC and chloride content of all the water samples under investigation have similar trend, whereas the pH values presented an opposite trend compared to the trend shown by EC, TDS and chloride content [Fig 5.3].
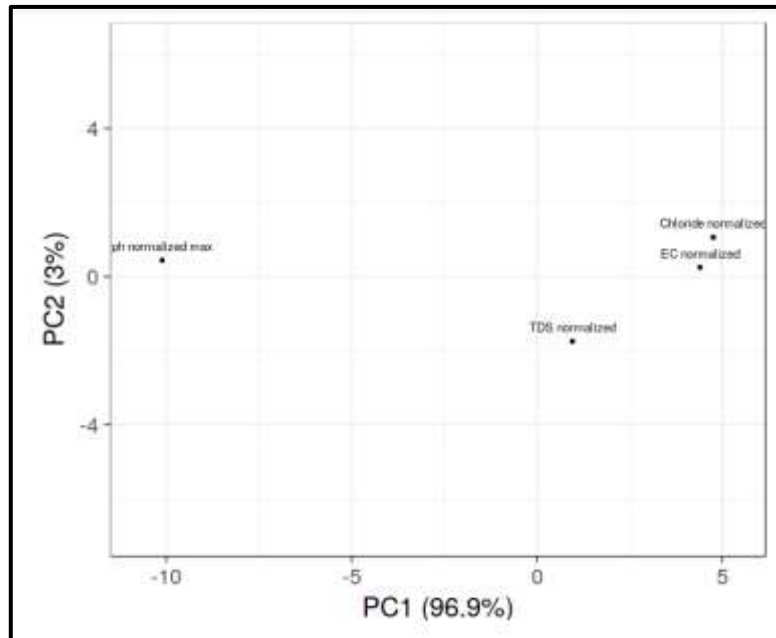
Figure 5.3 Principal Component Analysis loading plot showing the

relation between the physico-chemical parameters of water samples

The PCA results showed that the parameters EC, TDS and chloride content have similar properties and are related with each other whereas pH was on the other side of the PCA. Reason for the same, was that the pH values of all the samples had no significant difference. Furthermore, it was observed that the water samples can also be classified on the basis of district using physico-chemical parameters (Fig 5.3).

The data was further divided based on the urban, industrial and rural regions of two districts A-Fazilka, B-Ferozepur, Punjab. PCA was performed on the data and a unit variance scaling was applied to the rows; SVD with imputation was employed to calculate principal components. X and Y axis presented principal component 1 and principal component 2 that explain 74.5% and 24.9% of the total variance, respectively. N = 17 data points [Fig 5.4]. It was observed that

the water samples collected from two districts can be successfully classified with district A on right side and B on the left side of the graph.
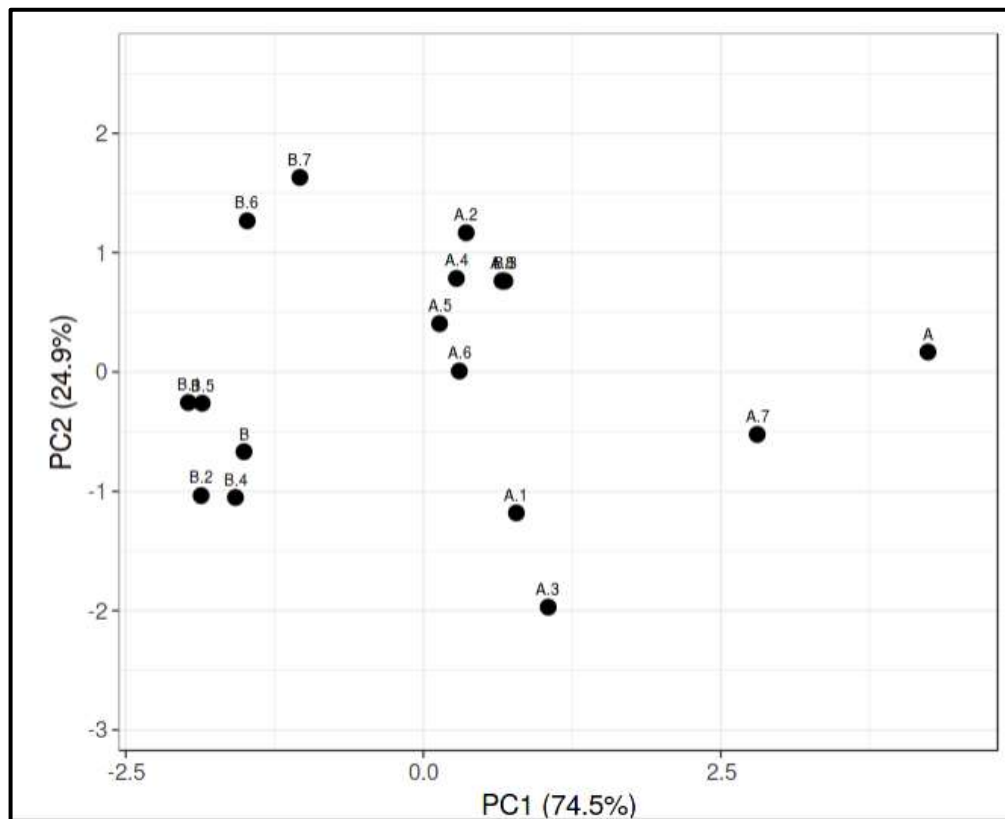
Figure 5.4 Classification for the urban ground water samples based on district A-Fazilka and B-Ferozepur

Furthermore, to understand the relationship and correlation between the water samples collected from two different districts, PCA was applied separately on the samples based on the Urban, Industrial and Rural regions. As shown in Fig. 5.5, with increase in the PC1 value from negative to zero, the separation of rural, urban to industrial ground water samples can be observed.
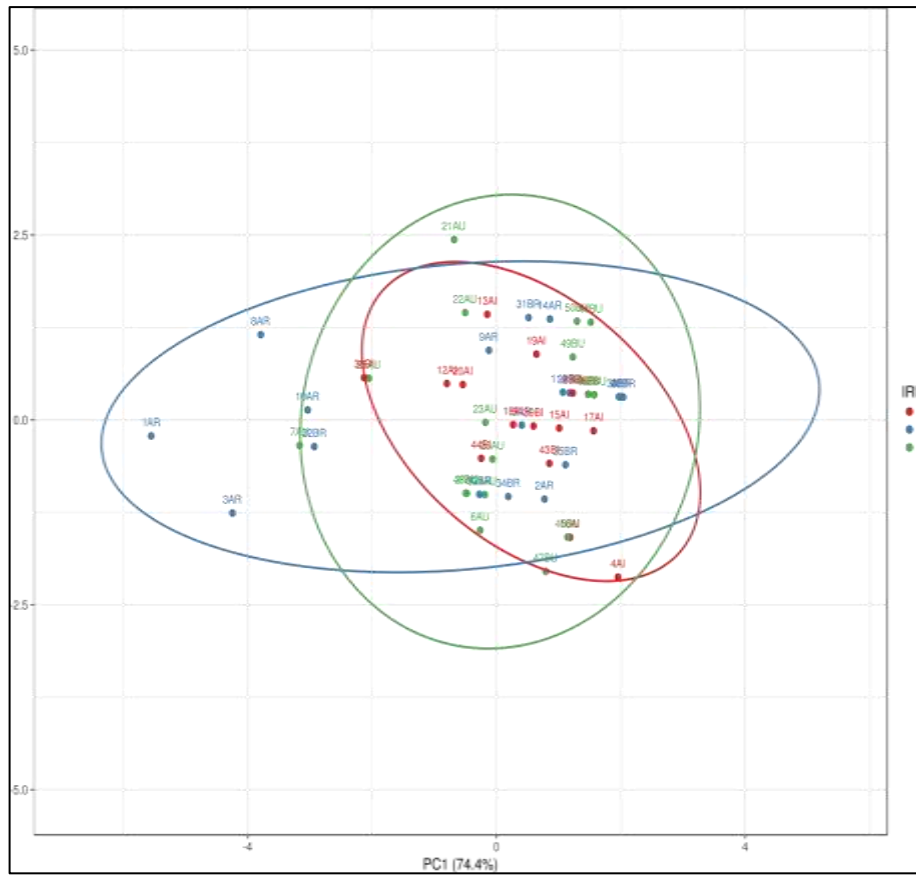
Figure 5.5 Classification for the water samples based on Industrial, Rural and Urban region of district A-Fazilka and B-Ferozepur

In addition, ground water samples of two districts were further classified based on the source of ground water, namely – HP [hand pump], SS [submersibal pump] and WB [water pump]. PCA was applied with the following conditions- unit variance scaling was applied to rows containing samples and NIPLAS PCA was used to calculate principal components. X and Y axis presented principal component 1 and principal component 2 that explain 74.4% and 24.8% of the total variance, respectively [Fig 5.6]. In prediction ellipses, with probability 0.95, a new observation from the same group was fall inside the ellipse. N = 50 data points. When the PC1 and PC2 values move from negative to positive values, samples were grouped from HP, SS to WB. The classification of samples based on different source of water was efficient as the depth of the different sources varied significantly.

Thus the research analysis using ANOVA, descriptive statistics and PCA was
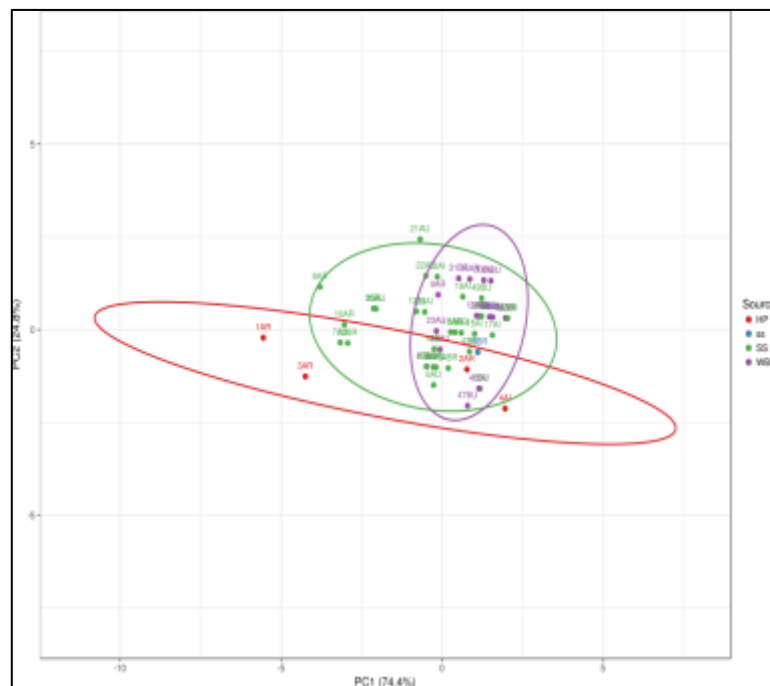
Figure 5.6 Classification for the water samples based on different source of water of district A-Fazilka and B-Ferozepur (HP-Hand Pump, SS-Submersible, WB- Water Box)

Conclusion

Present work focuses on the quality analysis of groundwater from two districts of Malwa Region, Punjab based on their physicochemical parameters such as pH, TDS, EC and Chloride. The pH and TDS values of all samples were in safe limit according to WHO and BIS standards. The EC values of water samples collected from depth greater than 100 ft. were in the permissible range, but as the depth decreases the values of parameters does not lie in the range and are having higher values than standard. It was also observed that all the water samples from Ferozepur districts lies in the acceptable range because all were collected from the sources having the depth greater than 100 ft during random sampling. The ground water samples collected from water box as a source in rural area were also in the permissible range. The samples from urban and industrial region had their values greater than the desired limit irrespective of more depth and the source of water for both the districts. There was no correlation of pH with EC, Chloride and TDS. The correlation values between any of two parameters - TDS, EC and chloride were greater than 0.90. Furthermore, by using developed regression equation for EC and TDS, TDS and chloride and EC and chloride content, by measuring one parameter, the other two parameters can be calculated. This will reduce the cost and time for chemical analysis. PCA results shows that TDS, EC and Chloride have same trend and pH shows different trend from these, further PCA results shows that sample can be classified on the basis of districts [Ferozepur and Fazilka], regions [urban, rural and industrial] and sources[hand pump, submersible and water box]. Thus the research analysis using ANOVA, descriptive statistics and PCA was capable of classifying the samples from two districts on the basis of physicochemical parameters, districts Ferozepur and Fazilka] region [urban, rural or industrial] and sources of ground water sample.

References

[1]. Kishor, Kamal, et al. "Sources, spatio-temporal distribution and depth variations in groundwater salinity of semi-arid Rohtak district, Haryana, India." *Groundwater for Sustainable Development* (2022): 100790.

[2]. ElKashouty, Mohamed. "Groundwater vulnerability risk assessment in south Darb El Arbaein, south Western Desert." *Applied Water Science* 11.5 (2021): 1-21.

[3]. World Health Organization. [2017]. Guidelines for drinking-water quality: first addendum to the fourth edition.

[4]. World Health Organization. "pH in Drinking-water: Revised background document for development of WHO Guidelines for Drinking-water Quality." World Health Organization: Geneva, Switzerland [2007].

[5]. Krishan, Gopal, et al. "Analysis of water level fluctuations and TDS variations in the groundwater at Mewat (Nuh) district, Haryana (India)." *Current World Environment* 11.2 (2016): 388.

[6]. Manish Kumar, Kalpana Kumari, AL. Ramanathan & Rajinder Saxena. "A comparative evaluation of groundwater suitability for irrigation and drinking purposes in two intensively cultivated districts of Punjab, India." Environmental Geology volume 53, pages553–574 [2007]

[7]. Bloomfield, J. P., et al. "Characterising variations in the salinity of deep groundwater systems: A case study from Great Britain (GB)." *Journal of Hydrology: Regional Studies* 28 (2020): 100684.

Chapter 6 - Near Infrared Spectroscopy as a tool to Classify groundwater on the basis of depth, region and source

6. *Chapter 6 - Near Infrared Spectroscopy as a tool to Classify groundwater on the basis of depth, region and source*

Punjab is an agricultural land, due to this use of fertilizers have increased, so water analysis becomes necessary in this region. Present research work focuses on classification of ground water sample of two districts of Punjab, India using near infrared [NIR] spectral data. NIR spectra have been collected for all the samples in the range of 700-2500 nm with a gap of 0.5 nm. Principal Component Classification was applied on NIR spectra to classify samples on the basis of districts [Ferozepur and Fazilka], region [urban, rural and industrial], sources [hand pump, water box and submersible] and Depth. Results reveal that the NIR spectra can classify the samples on the basis of region, district, sources and depth with good accuracy.

Introduction

Groundwater is most precious and major natural resource on earth. It is used for agricultural, domestic and industrial purpose. About one third population of world depends on ground water for drinking purpose [1]. For development of a country there should be supply of safe and clean drinking water. Main threats to groundwater quality are urbanization, industrial and agricultural activity and climate changes. Water is most important component for life on earth. It had been contaminated by toxic element and inorganic industrial pollutant which result into it become unsafe for human and irrigation purpose. Chemical pollution include agriculture activities cities mining area and industries. Water pollutant mainly comes from physical biological and chemical reaction. Rainfall also responsible for groundwater contamination. Contaminants from agriculture field such as pesticide her besides and fertilizers, oil urban road salt animal wastage and acid toxic from minds and all these transfer to groundwater. Most of the contaminants are due to dissolution of natural mineral from geogenic origin. Contaminants such as toxic metals, hydrocarbons, micro plastics, nanoparticles, organic contaminants are not good for human health and socioeconomic development [2, 3]. Now chemical contamination is modern threat in ground water. Ground water being the basic human need and so it a great challenge to have safe and clean water, so it provide great opportunity to researchers to provide main source of contaminants and provide method for safe and clean drinking water

[4]. For sustainable development of society it is necessary that natural resources should be used in judicial manual not only for present population but also for future generation. This contamination can be due to road salt, viruses, chemicals, bacteria, fertilizers, medication fuel and fertilizers. Ground water contamination is different from surface water contamination. [9-10]. Effect of contaminated ground water on human health is very chronic and difficult to detect. [11].

Due to pressure of population growth, there is need of more food [12] Due to this green revolution technologies has grown and uses of fertilizers has increased. Punjab is the most agricultural state in India with the highest consumption of fertilizers. So ground water monitoring becomes necessary in this state. Because water is basic need of all human being , so before using groundwater for drinking all parameters [pH, TDS, EC and Chloride content] should be in acceptable range.

Infrared region is between electromagnetic spectrum, it has three main region such as near-infrared [2,500 to 750 nm], mid-infrared [10 to 2.5 μm] and far infrared [1 mm to 10 μm]. There occurs molecular vibration by absorption of radiation of any wavelength. Due to this there occur different types of vibration. Near infrared [NIR] is most energetic region and it is close to visible region. NIR discovered by Herschel in 1800 [13-14]. Near infrared spectra is made up of overtones and combination band. NIR spectra mainly consist of four overtones, but forth overtone is weak. These overtones are from absorption of aromatic C-H, methyl C-H, methoxy C-H, methylene, N-H from primary and secondary amides, N-H from amine salts , O-H, S-H, C=O , N-H from amides and carbonyl associated C-H. All these group found in organic molecule and water [15]. In 1900-2500nm of NIR combination band region is found [16]. Again starting of NIR done by Karl Norris from US department of agriculture in 60s. They found moisture content from seed using NIR by using multivariate calibration approach [17, 18]. Later in 1980 protein in wheat has been determined [19]. Now NIR proved useful not in grain sector but in non-agricultural [material science, pharmaceuticals, textiles etc.], also proved to be time and money saving [20]. Due to presence of overtones and combination band of C-H, N-H and O-H in food give rise to absorption NIR region. NIR is nondestructive method. But extraction of functional group from huge data is not easy task, but it is easy with the

help of Chemo metrics. [21]. NIR spectrometers main components are Sample compartment, Light source,

Results and discussion

Samples have been collected from different region, sources and depth from Ferozepur and Fazilka districts of Punjab, India and then PCA is applied on the NIR spectral data to check the feasibility of NIR data to classify the groundwater samples on the basis of different factors.

Depth range for hand pump, submersible and water box is 28m to 50m, 60m to 500m and 340 to 540m respectively. Samples were stored in screw capped polythene bottles at room temperature till further analysis.

Principal component analysis of the data was performed using Clustvis web tool. Principal component analysis was run on the data category wise – Region, depth, district and source of the ground water samples using Single value decomposition [SVD] method.

Clustvis web tool application has been performed for this statistical analysis. NIR data has been imported into clustvis in form of square symmetric matrix. A numeric data extracted as an input variable and performed principle components analysis [PCA] to classify the data on the basis of sources, regions [hand pump, water box and submersible] and districts [Ferozepur, Fazilka].

Classification of whole data on basis of source, region and depth

Firstly the PCA was applied to classify the samples on the basis of region [Rural R, Urban U and Industrial I] [Fig 6.1 a] and further PCA was done to check the classification of the samples on the basis of depth [Fig 6.1 b], third PCA was applied to classify on the basis of district [Fig 6.1 c] and lastly PCA was performed to do classification on the basis of sources of water samples [Fig 6.1 d].

For calculating principal components, SVD methodology was used with imputation. X axis shows principal component 1 with 78.2% of total variance. Y axis shows component 2 with 11% of total variance [Fig 6.1]. It was observed that the PCA for complete data set on basis of region, depth or source of ground water samples. Best classification of the water samples can be observed on the basis of district wise when

PCA is applied on the NIR spectral data [Fig 6.1c]. Fazilka district-A samples lie on the right side of PCA with negative values of Principal component (PC) 1, whereas the Ferozepur district – B samples lie on left side with positive values of PC2. Thus it can be inferred that NIR spectral data is capable of classification of the samples of two different places, thought they are located at a gap of 70-80 km. The results showed that spectral data collected using a nondestructive, rapid and non-hazardous technique is capable of classifying the ground water samples on the basis of district.
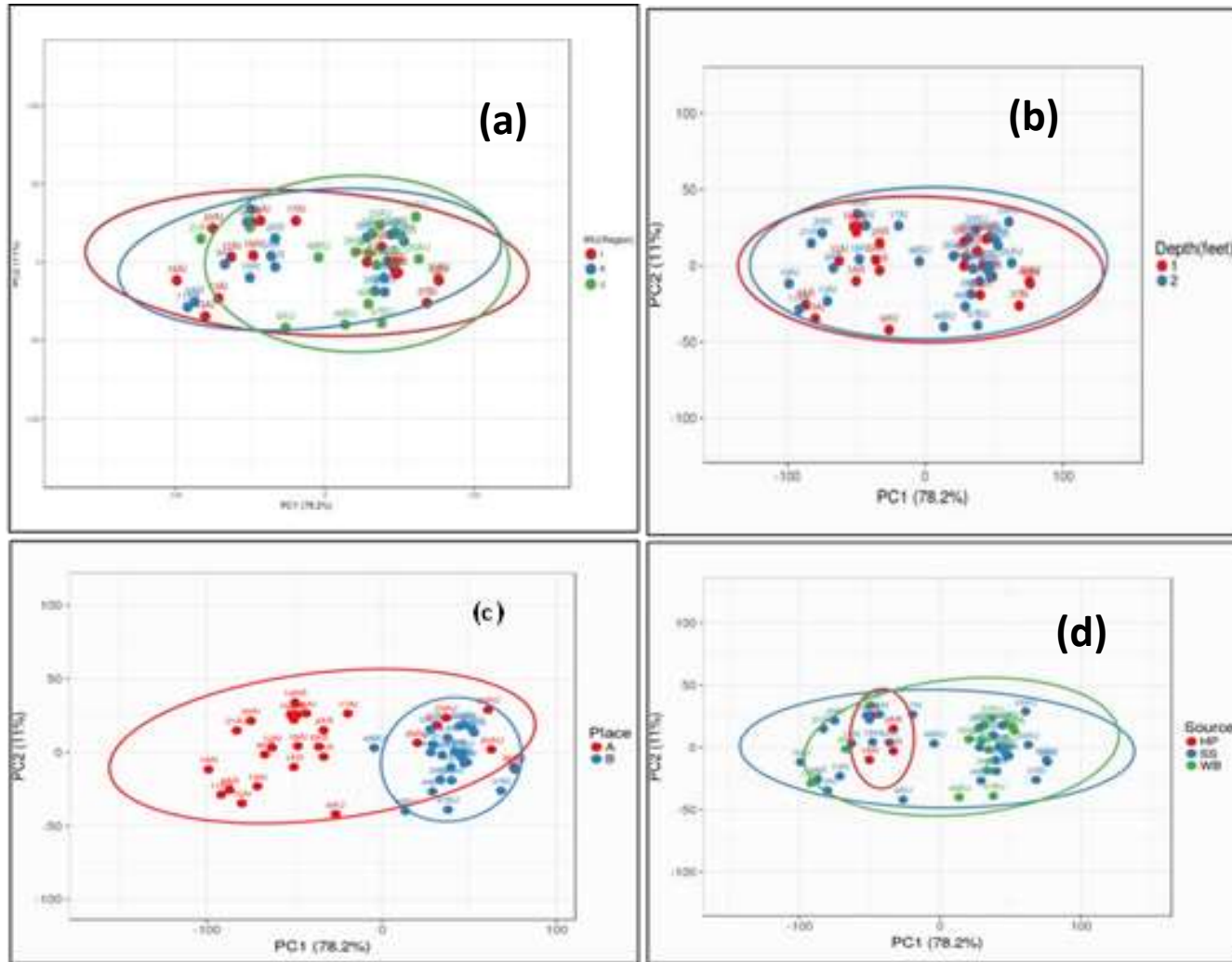
Figure 6.1 Classification of the samples on the basis of region (a), depth (b), district (c) and source (d) using Principal Component Analysis of the spectral data

Classification of each district on basis of source, region and depth

To further classify the samples with more accuracy, PCA was employed on the samples of two districts separately [Figure 6.2]. For District A, It can be observed that the best classification of the water samples can be observed on the basis of district wise when PCA is applied on the NIR spectral data. For calculating principal components SVD with imputation were used. X axis shows principal component 1 with 74.3% of total variance. Y axis shows component 2 with 12.1 % of total variance. From graph new observation [from same group] can be predicted with probability 0.95. New observation will fall inside the ellipse. N = 27 data points. For district B, for calculating principal components SVD with imputation were used. X axis shows principal component 1 with 36.8% of total variance. Y axis shows component 2 with 32.4 % of total variance. It was observed that when classification was done one district at a time. Data collected from urban regions could be separated from the one collected from Industrial and rural region successfully. Further for district A the classification of R and I region was difficult as they were in the same PCs values. The classification of region for district B were pretty clear. Thus the spectral data is capable of classifying the urban and rural samples with good accuracy,
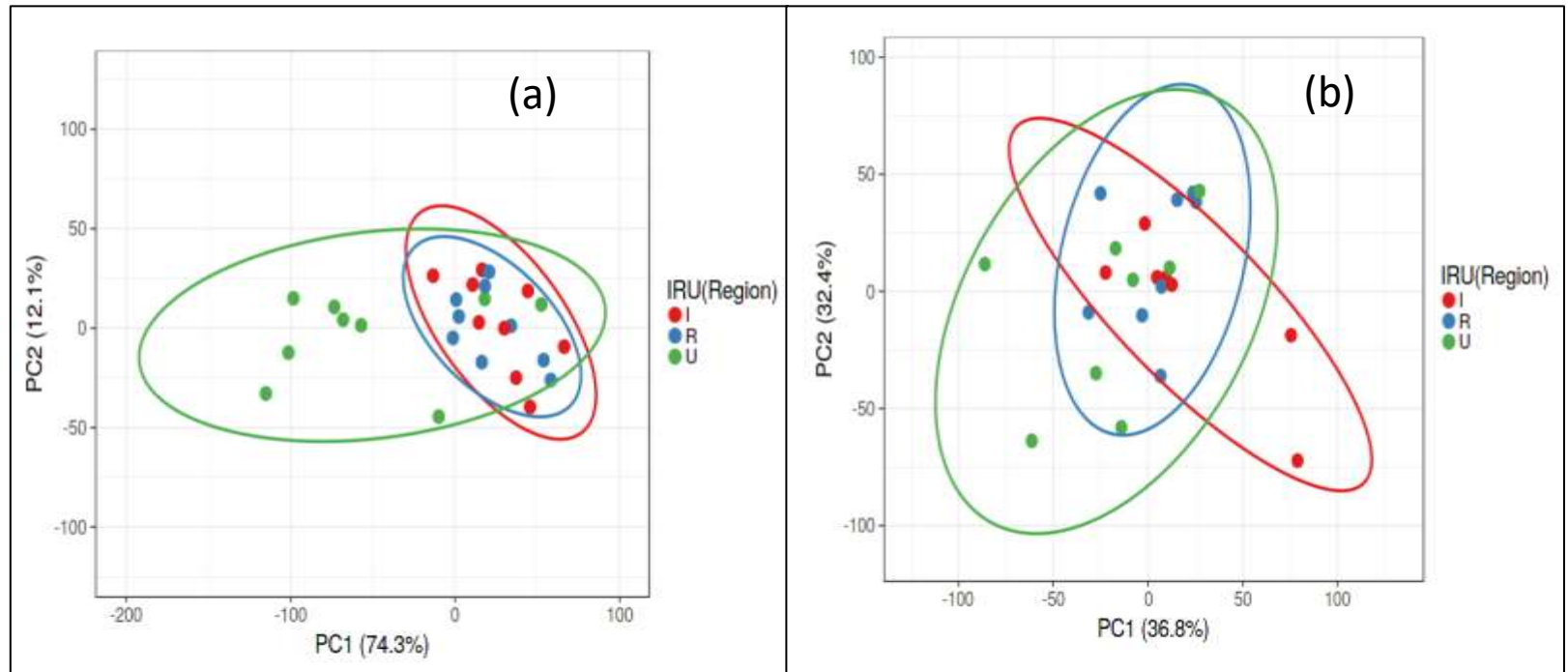
Figure 6.2 Principal Component Analysis of data using Near Infrared Spectroscopy on the basis of region I-Industrial, U-

Urban and R-rural (a) district A and (b) district B
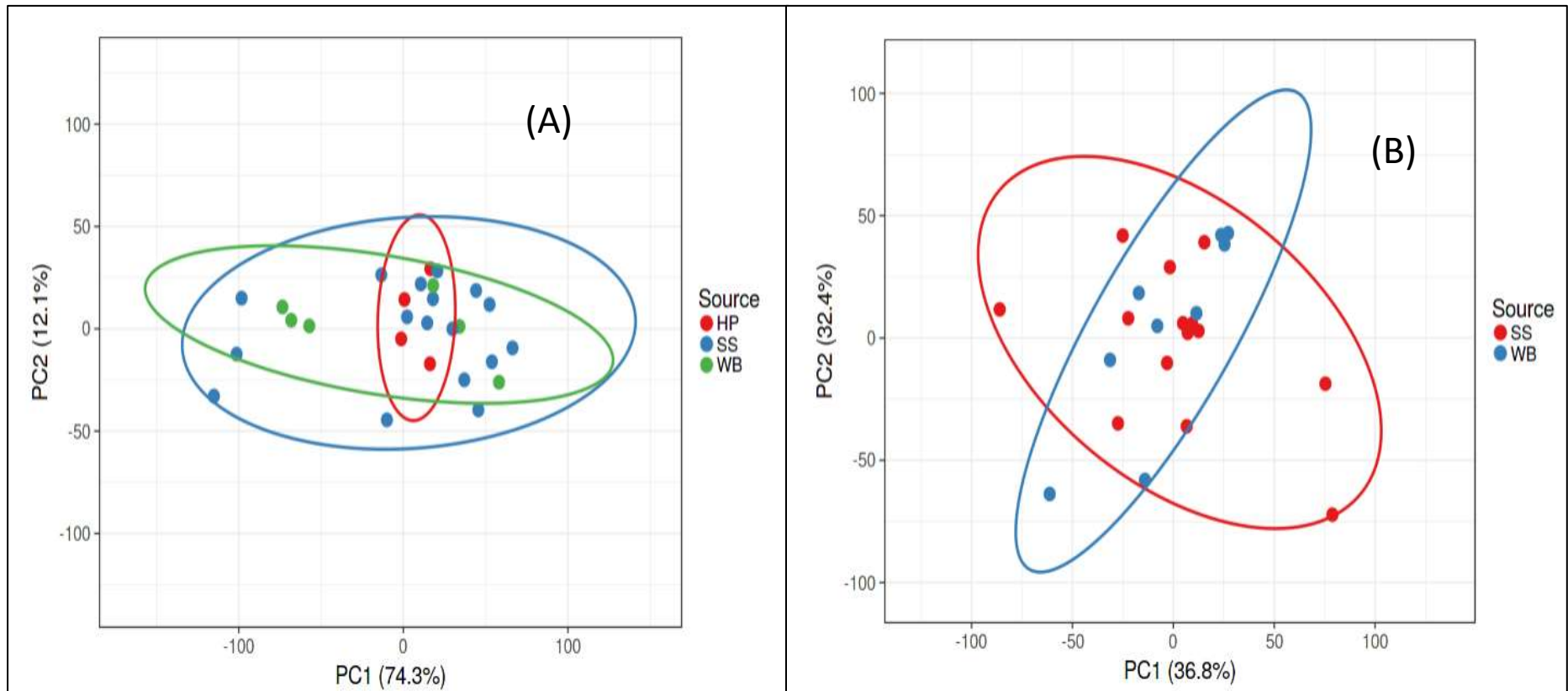
Figure 6.3 Principal Component Analysis of data using Near Infrared Spectroscopy on the basis of sources f collected ground water samples (a) district A and (b) district B
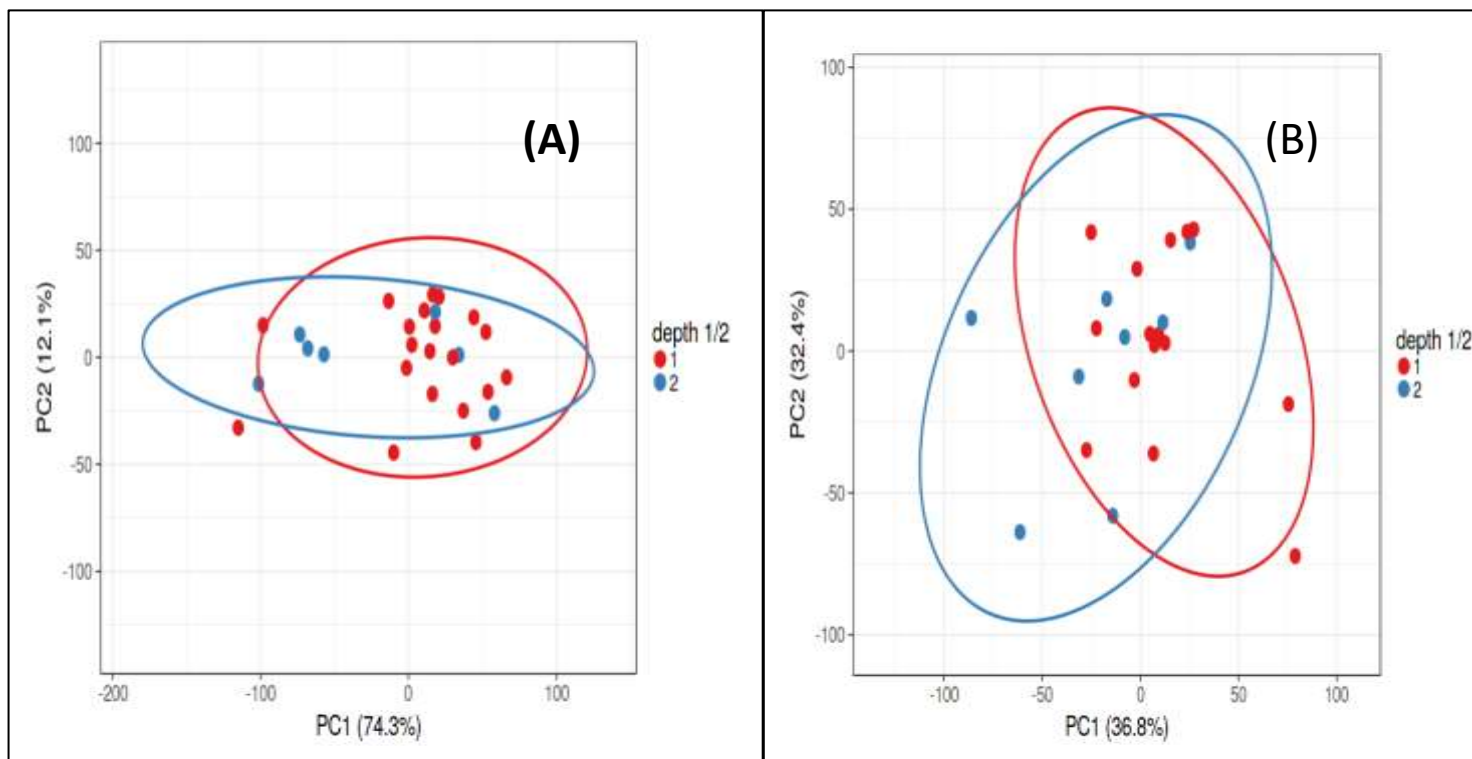
Figure 6.4 Principal Component Analysis of data using Near Infrared Spectroscopy on the basis of depth less than

400 ft =1, more than 400 ft =2 (a) district A and (b) district B

PCA was also carried out on the basis of source [Figure 6.3] and depth [Figure 6.4] of ground water samples. Here the depth coding was change to less than and equal to 400ft as 1 and more than 400 ft as 2 for the two districts separately. For district A, for calculating principal components SVD with imputation were used. X axis shows principal component 1 with 74.3% of total variance. Y axis shows component 2 with 12.1 % of total variance. From graph new observation [from same group] can be predicted with probability 0.95. For district B, for calculating principal components SVD with imputation were used. X axis shows principal component 1 with 36.8% of total variance. Y axis shows component 2 with 32.4 % of total variance. From graph new observation [from same group] can be predicted with probability 0.95. New observation will fall inside the ellipse. N = 23 data points.

Classification of samples on the basis of depth is shown in Fig 6.4 and it was observed that samples can be grouped in two groups using spectral data when data was analyzed district wise.

Conclusion:

Present research work was carried out to check the feasibility of NIR spectral data to classify the ground water samples on the basis of different categorized. Application of PCA on the spectral data showed that the NIR spectral data can successfully classify the samples on the basis of different district of same state. The reason for the same may be soil of different districts resulting in spectral data which can classify the samples. Further PCA was applied separately on each district for classification on basis of source, depth and region. Results concluded that NIR spectral data is capable of classifying the ground water samples on the basis of region, source, depth and district wise.

References

[1].   International Association of Hydrogeologists [2020] Groundwater— more about the hidden resource. https://iah.org/education/gener al-public/groundwater-hidden-resource. Accessed 13 Nov 2020

[2].   Lin, Hangsheng. "Earth's Critical Zone and hydropedology: concepts, characteristics, and advances." *Hydrology and Earth System Sciences* 14.1 [2010]: 25-45.

[3].   Basu, Ankita, et al. "A review on sources, toxicity and remediation technologies for removing arsenic from drinking water." *Research on Chemical Intermediates* 40.2 [2014]: 447-485.

[4].   Pandey, H. K., S. K. Duggal, and Ajit Jamatia. "Fluoride contamination of groundwater and it's hydrogeological evolution in District Sonbhadra [UP] India." *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences* 86.1 [2016]: 81-93.

[5].   Subba Rao, N., B. Ravindra, and Jianhua Wu. "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 [2020]: 2316-2348.

[6].   He, Xiaodong, et al. "Poor groundwater quality and high potential health risks in the Datong Basin, northern China: research from published data." *Environmental Geochemistry and Health* 43.2 [2021]: 791-812.

[7].   MacDonald, Jacqueline A., and Michael C. Kavanaugh. "Restoring contaminated groundwater: an achievable goal?." *Environmental Science & Technology* 28.8 [1994]: 362A-368A.

[8].   Chakraborti, Dipankar, et al. "Groundwater arsenic contamination in Bangladesh—21 Years of research." *Journal of Trace elements in Medicine and Biology* 31 [2015]: 237-248.

[9].   Smil, Vaclav. "Long-range perspectives on inorganic fertilizers in global agriculture, 1999 Travis P Hignett lecture." *Alabama, USA* [1999].

[10].   Bokobza, L. "Near infrared spectroscopy." *Journal of Near Infrared Spectroscopy* 6.1 [1998]: 3-17.

[11]. Agelet, Lidia Esteve, and Charles R. Hurburgh Jr. "A tutorial on near infrared spectroscopy and its calibration." *Critical Reviews in Analytical Chemistry* 40.4 [2010]: 246-260. Bokobza, L. "Near infrared spectroscopy." *Journal of Near Infrared Spectroscopy* 6.1 [1998]: 3-17.

[12]. Hart, Joe R., Karl H. Norris, and Calvin Golumbic. "Determination of the moisture content of seeds by near-infrared spectrophotometry of their methanol extracts." *Cereal Chem* 39.2 [1962]: 94-99.

[13]. NORRIS, KARL H. "Direct spectrophotometric determination of moisture content of grain and seeds." *Proc. of the 1963 International Symposium on Humidity and Moisture*. Vol. 4. 1965.

[14]. American Association of Cereal Chemists, and American Association of Cereal Chemists. Approved Methods Committee. "Approved Methods of the American Association of Cereal Chemists." AACC, 1983.

[15]. Weyer, Lois. "History of Near-Infrared [NIR] Applications." *Practical Guide to Interpretive Near-Infrared Spectroscopy*. CRC Press, 2007. 119-124.

[16]. Christy, Alfred A., et al. "The detection and quantification of adulteration in olive oil by near-infrared spectroscopy and chemometrics." *Analytical Sciences* 20.6 [2004]: 935-940.

[17]. Spellman, Frank R. *The drinking water handbook*. CRC Press, 2017.

[18]. Edzwald, James, and American Water Works Association. *Water quality & treatment: a handbook on drinking water*. McGraw-Hill Education, 2011.

[19]. Hammer MJ. Water and Wastewater Technology. 7th ed. Upper Saddle River: Pearson education; 2011

[20]. Alley, E. Roberts. *Water quality control handbook*. Vol. 2. New York: McGraw-Hill, 2007.

[21]. Edition, Fourth. "Guidelines for drinking-water quality." *WHO chronicle* 38.4 [2011]: 104-108.

Chapter 7 – Prediction of parameters of ground water using Near Infrared Spectroscopy and Chemometrics

7. Chapter 7 - Prediction of parameters of ground water using Near Infrared Spectroscopy and Chemometrics

Introduction:

Groundwater is most precious and major natural resource on earth. Groundwater is used for agricultural, domestic and industrial purpose. About one third population of world depends on ground water for drinking purpose [1]. For development of a country there should be supply of safe and clean drinking water. Main threats to groundwater quality are urbanization, industrial and agricultural activity and climate changes. Contaminants such as hydrocarbons, toxic metals, nanoparticles, organic contaminants, micro plastics are not good for human health and socioeconomic development [2, 3]. Now chemical contamination is modern threat in ground water. Because ground water is basic human need and so it a great challenge to have safe and clean water, so it provide great opportunity to researcher to provide main source of contaminants and provide method for safe and clean drinking water [4]. For sustainable development of society it is necessary that natural resources should be used in judicial manual not only for present population but also for future generation. Groundwater is one such natural resource. Water is most important component for life on earth. It had been contaminated by toxic element and inorganic industrial pollutant which result into it become unsafe for human and irrigation purpose. Chemical pollution include agriculture activities cities mining area and industries. Water pollutant mainly come from physical biological and chemical reaction. Rainfall also responsible for groundwater contamination. Contaminants from loans and agriculture field such as pesticide her besides and fertilizers, oil urban road salt animal wastage and acid toxic from minds and all these transfer to groundwater. Most of the contaminants are from geogenic origin due to dissolution of natural mineral.

Addition of unwanted substances in ground water is known as groundwater contamination [9]. This contamination can be due to road salt, viruses, chemicals, bacteria, fertilizers, medication fuel and fertilizers. Ground water contamination is different from surface water contamination [10]. Effect of contaminated ground water on human health is very chronic and difficult to detect [11].

Due to pressure of population growth, there is need of more food [12] Due to this green revolution technologies has grown and uses of fertilizers has increased. Punjab is the most agricultural state in India with the highest consumption of fertilizers. So ground water monitoring become necessary in this state. Because water is basic need of all human being , so before using groundwater for drinking all parameters [pH, TDS, EC and Chloride content] should be in acceptable range. In present paper we have focused on two Ferozepur and Fazilka district of Punjab, India.

Wavelength, frequency and energy are main properties of electromagnetic energy and light is electromagnetic energy. Energy is directly proportional to frequency. To form electromagnetic spectrum light can be arranged according to wavelength or frequency [13]. Infrared region is between electromagnetic spectrum, it has three main region such as far infrared [1 mm to 10 μm], mid-infrared [10 to 2.5 μm], and near-infrared [2,500 to 750 nm]. There occur molecular vibrartion by aborption of radiation of any wavelength. Due to different energy of three region provide different types of vibration. Near infrared [NIR] is most energetic region and it is close to visible region. NIR discovered by Herschel in 1800 [14]. Near infrared spectra is made up of overtones and combination band. NIR spectra mainly consist of four overtones, but forth overtone is weak. These overtones are from absorption of aromatic C-H, methyl C-H, methoxy C-H, methylene, N-H from primary and secondary amides, N-H from amine salts , O-H, S-H, C=O , N-H from amides and carbonyl associated C-H. All these group found in organic molecule and water [15]. At 1900-2500nm f NIR combination band region is found. [16]. Again starting of NIR done by Karl Norris from US department of agriculture in 60s. They found moisture content from seed using NIR by using multivariate calibration approach. [17, 18] . Later in 1980 protein in wheat has been determined [19]. Now NIR proved useful not in grain sector but in non-agricultural [material science, pharmaceuticals, textiles etc.] also proved to be time and money saving [20] . Due to presence of overtones and combination band of C-H, N-H and O-H in food give rise to absorption NIR region .NIR is nondestructive method. But extraction of functional group from huge data is not easy task, but it is easy with the help of Chemo metrics [21].

Literature review study of different paper revealed that all physiochemical parameters have been found by different chemical methods. But all these methods are laborious,

time consuming and costly. By using regression model we can determine these parameters by NIR. NIR is fast and nondestructive method.

NIR spectral analysis

Spectral data for all the 50 ground water samples were collected from NIRDS 2500 spectrometer. NIR spectra for all the samples is shown in Fig 7.1.
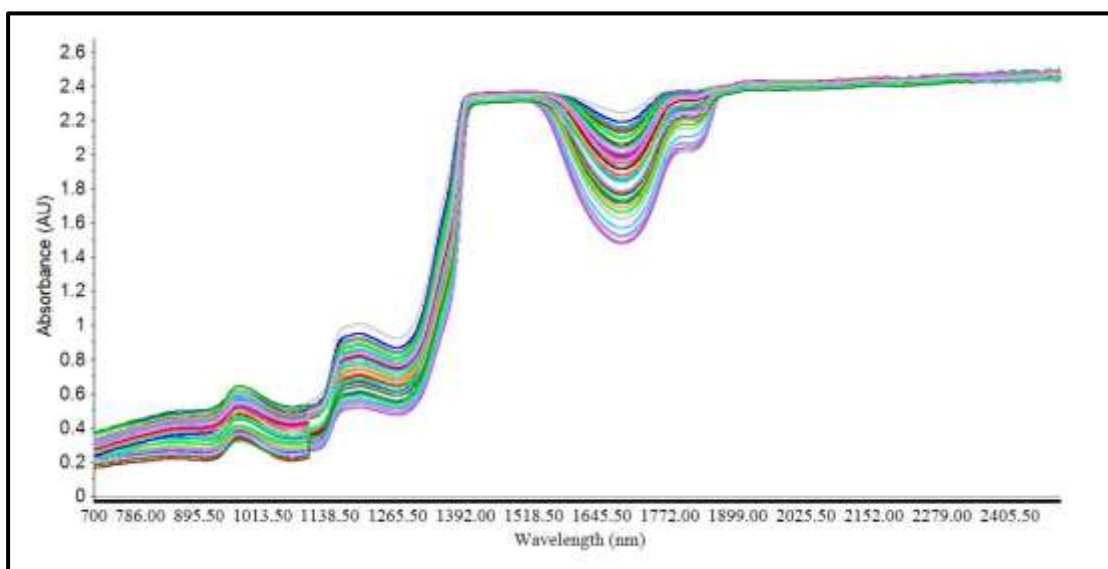
Figure 7.1 Near Infrared Spectra of all the 50 samples from 700-2500 nm

Spectral data shows that there is change in detector around 1100 nm as the spectra shows a very small shift at 1100 nm. Further peaks / bands were observed around 1000 nm, 1200 nm, 1650nm showing the presence of the bonds at these wavelengths.

Statistical Regression Model using SPSS and CAMO Unscrambler X10.5

As there were bands in the spectrum of the samples thus to reveal the information about the parameters – bonds relative to pH, TDS, EC and chloride content, NIR and reference data analysis had been performed using SPSS (@LPU, Punjab) and CAMO Unscrambler X 10.5   software (CSIR – CSIO, Chandigarh). Multivariate analysis namely - PLS, iPLS, siPLS. (Details about these regression methods discussed in Chapter 1) were used to develop the rapid and non-destructive statistical regression model. Preprocessing such as

normalization, double derivative, baseline and smoothening has been applied on the NIR data.

Partial Regression Analysis

Regression model was first developed using the range 700-2500 nm of spectral data for all the characteristics parameters of ground water samples using CAMO Unscramabler. Results showed that for wavelength range 700-2500nm coefficient of determination (R^2) for calibration and validation is greater than 0.9 for all parameters (pH, TDS, EC and chloride content) and the root mean square error (RMSE) for calibration and validation is 0.05 for all parameters at factor 7 (Table 7.1).

Table 7.1 Partial Least Square Regression model results - RMSE- Root Mean Square Error, R^2- Coefficient of Determination for all the parameters for ground water samples.

700-2500nm	$RMSE_C$	R^2_c	$RMSE_V$	R^2_v
pH	0.00575	0.95122	0.00695	0.92877
TDS	0.04856	0.94947	0.05722	0.92986
EC	0.0414	0.96051	0.04867	0.94542
Cl	0.04072	0.956	0.04793	0.94455

Regression graphs showing the measured values of the parameters and as predicted by the PLS regression models are shown in fig 7.2.
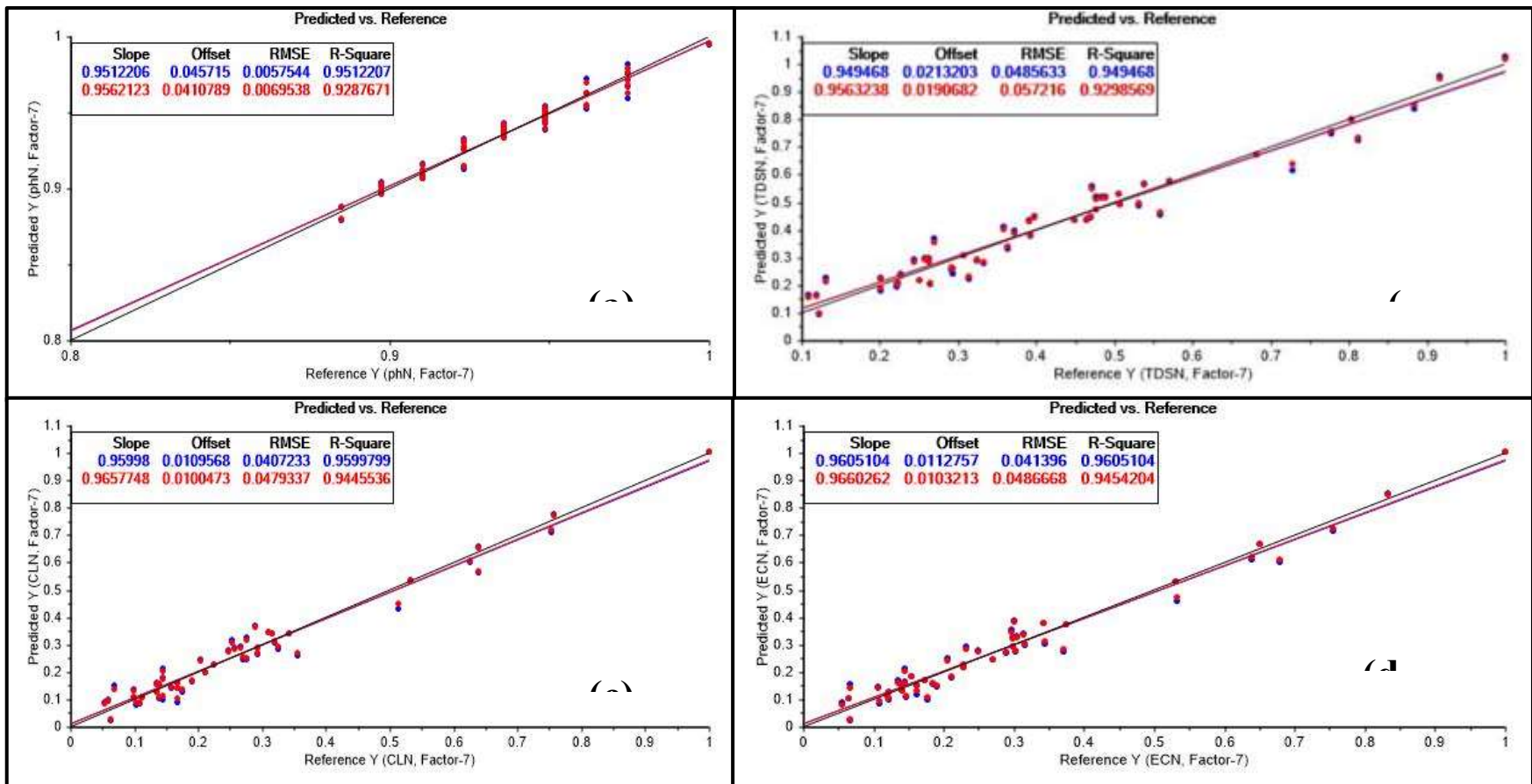
Figure 7.2 Predicted vs measured graphs for PLS regression model 700-2500 nm

Though the results from 700-2500nm range of PLS models were good, effort was made to check the important wavelengths by dividing the complete spectra in different parts.

Wavelength selection methods

As discussed in Chapter 1 about the iPLS and siPLS methods. First iPLS was employed to the data set.

Interval Partial Least Square Regression methods

Interval PLS (iPLS) has been applied to the complete data set in different groups. Complete data set was divided into equal intervals of equal width. iPLS regression models were developed by dividing the complete wavelength range with an equal interval of 100nm, 200nm, 300nm and 600nm.

Regression model developed using SPSS

In SPSS, separate regression models were developed for prediction of pH, TDS, EC and chloride content for all the intervals. Thus in all 16 regression models were developed. Four regression models for each parameter for every interval – gap of 100nm, 200nm, 300nm and 600 nm.

To select the appropriate wavelength range, R^2 and RMSE values were compared for all the intervals.

The details of all the 16 models developed using SPSS in detail are shown in Table 7.2, 7.3, 7.4 and 7.5 for the interval of 100nm, 200nm, 300nm and 600nm respectively for the parameters pH, EC, TDS and chloride content.

In general it was observed that the value of R^2 was less than 0.5 for the wavelength range from 700-1200nm for all the parameters. It shows that these wavelengths are least important for prediction of any of these parameters. Further it was observed that R^2 value was greater than 0.8 for the wavelengths in the range 1900-2500 nm for all the parameters.

Table 7.2 Coefficient of determination for the prediction of pH for gap of 100nm,
200nm, 300nm and 600 nm

Wavelength Range (gap of 100 nm)	r^2	r	Adjusted r^2	Standard Error
700-800	0.092	0.303	0.032	0.025899
800-900	0.092	0.304	0.054	0.025603
900-1000	0.125	0.354	0.068	0.0254092
1000-1100	0.119	0.345	0.041	0.0257744
1100-1200	0.039	0.197	-0.024	0.026629
1200-1300	0.031	0.177	-0.032	0.0267325
1300-1400	0.119	0.346	0.019	0.0260619
1400-1500	0.324	0.569	-0.104	0.0276521
1500-1600	0.131	0.361	-0.151	0.0282416
1600-1700	0.054	0.232	0.013	0.0261409
1700-1800	0.06	0.246	-0.01	0.0263308
1800-1900	0.26	0.51	-0.066	0.0271747
1900-2000	0.427	0.654	-0.17	0.0284649
2000-2100	0.536	0.732	-0.033	0.0267489
2100-2200	0.461	0.679	-0.2	0.0288359
2200-2300	0.55	0.741	-0.051	0.0269754
2300-2400	0.66	0.813	0.126	0.0246051
2400-2500	0.651	0.807	-0.005	0.0263904
Wavelength Range (gap of 300 nm)	r^2	r	Adjusted r^2	Standard Error
700-1000	0.108	0.328	0.029	0.0259401
1000-1300	0.094	0.306	0.013	0.0261462
1300-1600	0.588	0.767	-0.01	0.0264445
1600-1900	0.324	0.569	-0.004	0.0263766

Wavelength Range (gap of 600 nm)	r^2	r	Adjusted r^2	Standard Error
1900-2200	1	1	-	-
2200-2500	1	1	-	-
700-1300	0.119	0.345	-0.004	263727
1300-1900	0.685	0.828	-0.186	0.0286593
1900-2500	1	1	-	

Wavelength Range (gap of 200 nm)	r^2	r	Adjusted r^2	Standard Error
700-900	0.092	0.303	0.011	0.026176
900-1100	0.125	0.353	0.047	0.0256926
1100-1300	0.035	0.188	-0.028	0.0266802
1300-1500	0.412	0.642	-0.108	0.0276976
1500-1700	0.16	0.4	-0.176	0.0285399
1700-1900	0.317	0.563	0.016	0.0261104
1900-2100	0.953	0.976	0.671	0.0151056
2100-2300	0.922	0.96	0.039	0.0257979
2300-2500	0.994	0.997	0.847	0.01028281

Table 7.3 Coefficient of determination for the prediction of EC for gap of 100nm, 200nm, 300nm and 600 nm

Wavelength Range (gap of 100 nm)	r	r^2	Adjusted r^2	Standard Error
700-800	0.316	0.1	0.041	0.2060577
800-900	0.296	0.087	0.049	0.2052463
900-1000	0.383	0.147	0.091	0.2005946
1000-1100	0.381	0.145	0.069	0.2030537
1100-1200	0.504	0.254	0.206	0.1875428
1200-1300	0.4	0.16	0.105	0.1990521
1300-1400	0.433	0.187	0.095	0.2001849
1400-1500	0.634	0.402	0.203	0.2080295
1500-1600	0.506	0.256	0.014	0.2089045
1600-1700	0.202	0.041	0	0.210443
1700-1800	0.194	0.038	0.025	0.213049
1800-1900	0.532	0.284	-0.033	0.2138256
1900-2000	0.696	0.484	-0.054	0.2159877
2000-2100	0.672	0.452	-0.221	0.2325929
2100-2200	0.837	0.701	0.333	0.1718616
2200-2300	0.735	0.54	-0.073	0.2179503
2300-2400	0.766	0.586	-0.067	0.2173818
2400-2500	0.696	0.476	-0.511	0.2587043
Wavelength Range (gap of 300 nm)	r	r^2	Adjusted r^2	Standard Error
700-1000	0.433	0.187	0.095	0.2001849
1000-1300	0.428	0.184	0.111	0.1984128
1300-1600	0.84	0.706	0.28	0.1785358
1600-1900	0.664	0.441	0.17	0.1917532

Wavelength Range (gap of 600 nm)	r	r^2	Adjusted r^2	Standard Error
1900-2200	1	1	-	-
2200-2500	1	1	-	-
700-1300	0.39	0.152	0.034	0.2068067
1300-1900	0.862	0.743	0.03	0.2072194
1900-2500	1	1	-	-

Wavelength Range (gap of 200 nm)	r	r^2	Adjusted r^2	Standard Error
700-900	0.313	0.098	0.018	0.2085601
900-1100	0.382	0.146	0.07	0.202945
1100-1300	0.428	0.183	0.13	0.1962752
1300-1500	0.658	0.433	-0.069	0.217612
1500-1700	0.509	0.259	-0.037	0.2143289
1700-1900	0.54	0.292	-0.021	0.212603
1900-2100	0.945	0.893	0.253	0.1818955
2100-2300	0.987	0.974	0.684	0.1182114
2300-2500	0.996	0.991	0.783	0.0979399

Wavelength Range	r	r^2	Adjusted r^2	Standard Error
800-1000	0.309	0.096	0.037	0.2065316
1000-1200	0.368	0.135	-0.079	0.2019606
1200-1400	0.429	0.184	0.092	0.2005647
1400-1600	0.809	0.655	0.231	0.1845493
1600-1800	0.475	0.226	0.157	0.1931972
1800-2000	0.807	0.652	-0.219	0.2323334
2000-2200	0.954	0.911	0.124	0.1969511
2200-2400	0.996	0.992	0.813	0.0909251

Table 7.4 Coefficient of determination for the prediction of chloride content for gap of 100nm, 200nm, 300nm and 600 nm

Wavelength range (gap of 100nm)	r	r^2	Adjusted r^2	Standard Error
700-800	0.307	0.094	0.035	0.2019577
800-900	0.283	0.08	0.041	0.201396
900-1000	0.375	0.141	0.085	0.1967433
1000-1100	0.372	0.138	0.062	0.1991997
1100-1200	0.495	0.245	0.196	0.1843533
1200-1300	0.389	0.151	0.096	0.1955045
1300-1400	0.416	0.173	0.079	0.1973021
1400-1500	0.611	0.374	-0.23	0.208001
1500-1600	0.501	0.251	0.008	0.204788
1600-1700	0.197	0.039	-0.002	0.2058281
1700-1800	0.191	0.036	-0.026	0.2083341
1800-1900	0.527	0.278	-0.041	0.2097631
1900-2000	0.691	0.477	-0.067	0.2123942
2000-2100	0.671	0.45	-0.225	0.227622
2100-2200	0.83	0.689	0.306	0.1712742
2200-2300	0.726	0.528	-0.102	0.2158584
2300-2400	0.758	0.575	-0.097	0.2153388
2400-2500	0.69	0.476	-0.511	0.2587043
Wavelength range (gap of 300nm)	r	r^2	Adjusted r^2	Standard Error
700-1000	0.309	0.095	0.015	0.2040743
1000-1300	0.419	0.176	0.102	0.194832
1300-1600	0.836	0.699	0.262	0.1766497
1600-1900	0.655	0.429	0.152	0.189343

Wavelength range (gap of 200nm)	r	r²	Adjusted r²	Standard Error
1900-2200	1	1	-	-
2200-2500	1	1	-	-
700-1300	0.389	0.151	0.033	0.2022458
1300-1900	0.855	0.73	-0.016	0.2072934
1900-2500	1	1	-	-
Wavelength range (gap of 200nm)	**r**	**r²**	**Adjusted r²**	**Standard Error**
700-900	0.306	0.093	0.013	0.2042995
900-1100	0.371	0.138	0.061	0.1992544
1100-1300	0.418	0.175	0.121	0.1927582
1300-1500	0.638	0.407	-0.117	0.2172952
1500-1700	0.502	0.252	-0.047	0.2103974
1700-1900	0.529	0.28	0.037	0.209415
1900-2100	0.943	0.889	0.223	0.1812835
2100-2300	0.987	0.974	0.681	0.1160872
2300-2500	0.994	0.989	0.719	0.109042
Wavelength range (gap of 200nm)	**r**	**r²**	**Adjusted r²**	**Standard Error**
800-1000	0.295	0.087	0.027	0.2028139
1000-1200	0.368	0.135	0.079	0.2019606
1200-1400	0.413	0.17	0.076	0.1976563
1400-1600	0.802	0.644	0.206	0.1832243
1600-1800	0.471	0.222	0.153	0.1892813
1800-2000	0.809	0.654	-0.212	0.2263711
2000-2200	0.954	0.91	0.116	0.1933045
2200-2400	0.997	0.994	0.864	0.0757698

Table 7.5 Coefficient of determination for the prediction of TDS content for gap of 100nm, 300nm and 600 nm

Wavelength range (gap of 100nm)	r	R²	Standard error
700-800	0.094	0.307	0.2143795
800-900	0.095	0.295	0.2129042
900-1000	0.13	0.361	0.2100571
1000-1100	0.13	0.361	0.2123776
1100-1200	0.218	0.467	0.1991278
1200-1300	0.379	0.144	0.2084368
1300-1400	0.175	0.418	0.2092363
1400-1500	0.406	0.637	0.2149454
1500-1600	0.287	0.536	0.2120416
1600-1700	0.049	0.221	0.2172899
1700-1800	0.042	0.205	0.2204714
1800-1900	0.315	0.562	0.2167778
1900-2000	0.485	0.697	0.2237337
2000-2100	0.47	0.686	0.2370249
2100-2200	0.677	0.823	0.1850207
2200-2300	0.594	0.77	0.2125169
2300-2400	0.582	0.763	0.2266896
2400-2500	473	0.688	0.2688823
Wavelength range (gap of 600nm)	**r**	**r²**	**Standard error**
700-1300	0.146	0.382	0.2037539
1300-1900	0.732	0.856	2191532
1900-2500	1	1	0
Wavelength	**r**	**r²**	**Standard**

range (gap of 300nm)			error
700-1000	0.099	0.314	0.2162091
1000-1300	0.161	0.402	0.2085451
1300-1600	0.667	0.817	0.1969711
1600-1900	0.467	0.683	0.1942035
1900-2200	1	1	0
2200-2500	1	1	0

Regression model developed using CAMO Unscrambler X 10.5

Interval PLS was employed using Unscrambler software. The process to apply the iPLS algorithm was same as discussed in chapter 1. Complete data set was divided into equal intervals of equal width with an interval of 100, 200, 300 and 600 nm.

Results for calibration are shown in Fig 7.3 to Fig 7.6 for the iPLS regression models for different wavelength intervals. It was observed from Fig 7.3 that 1200-1300nm, 1700-1800nm and higher wavelengths showed less error for prediction of all the four parameters of ground water samples. Again it was observed that none of the interval alone was developing a model with similar robustness as compared with the model developed using 700-2500nm. Further iPLS was employed by taking an interval of 200nm (fig 7.4). It was observed that wavelength groups higher than 1500 nm onwards had less error and there R^2 was near to the R^2 of the model developed using 700-2500nm. The intervals of 300nm (Fig 7.5) showed that regression model developed using 1600-1900, 1900-2100 an 2100-2500nm showed similar R^2 as the model developed using complete wavelength. Further the iPLS for the gap of 600 nm made clear that the wavelength range 700-1300nm was least important and 1900-2500 nm was most important for the prediction of parameters of ground water sample (Fig 7.6). For more precise selection of wavelengths for prediction of pH, TDS, EC and chloride content of sample, siPLS was employed on the combination of the wavelength intervals with minimum RMSE and high values of R^2.

Validation results of the iPLS regression model are shown in Fig 7.6 -7.9. Calibration and validation results showed that the model robustness decreases from calibration to validation.
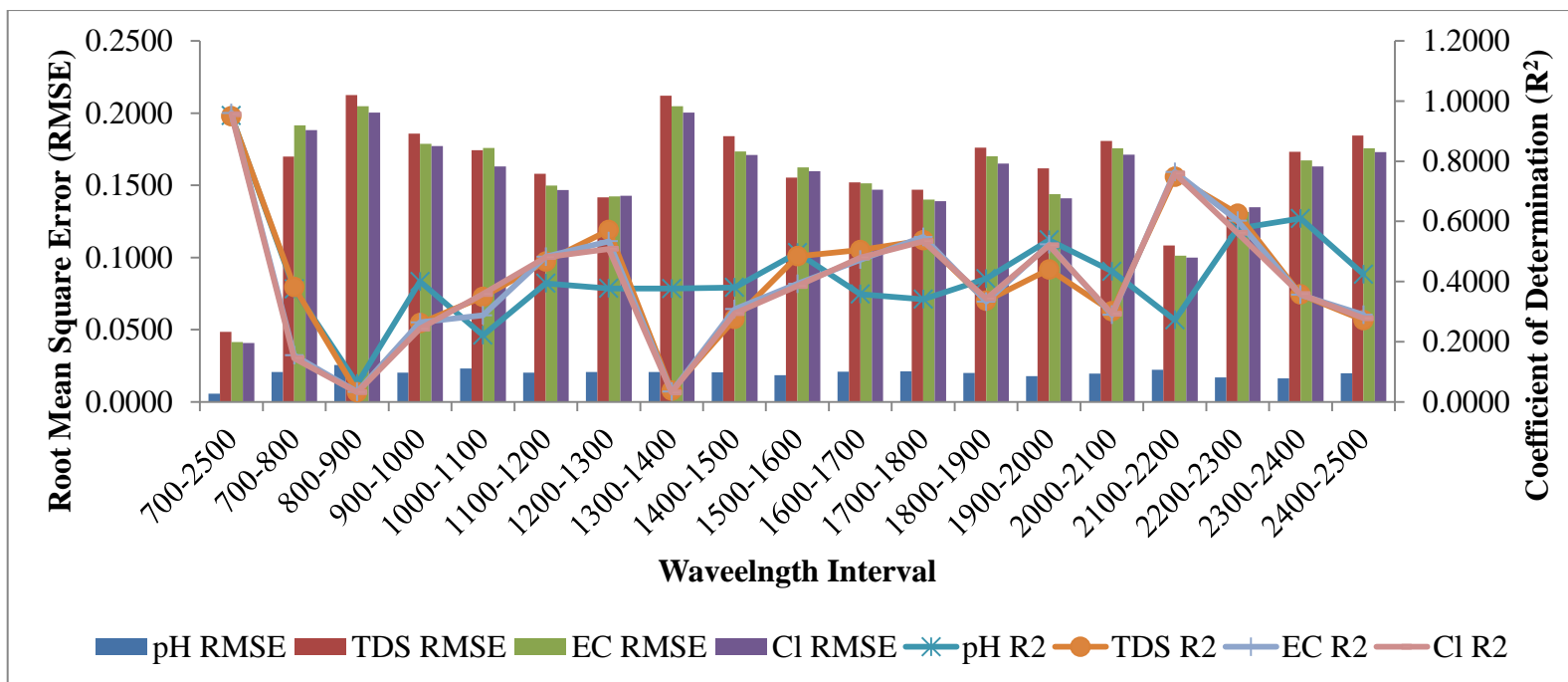
Figure 7.3 Root Mean Square Error and R2 for the iPLS (gap of 100 nm) for calibration and complete wavelength
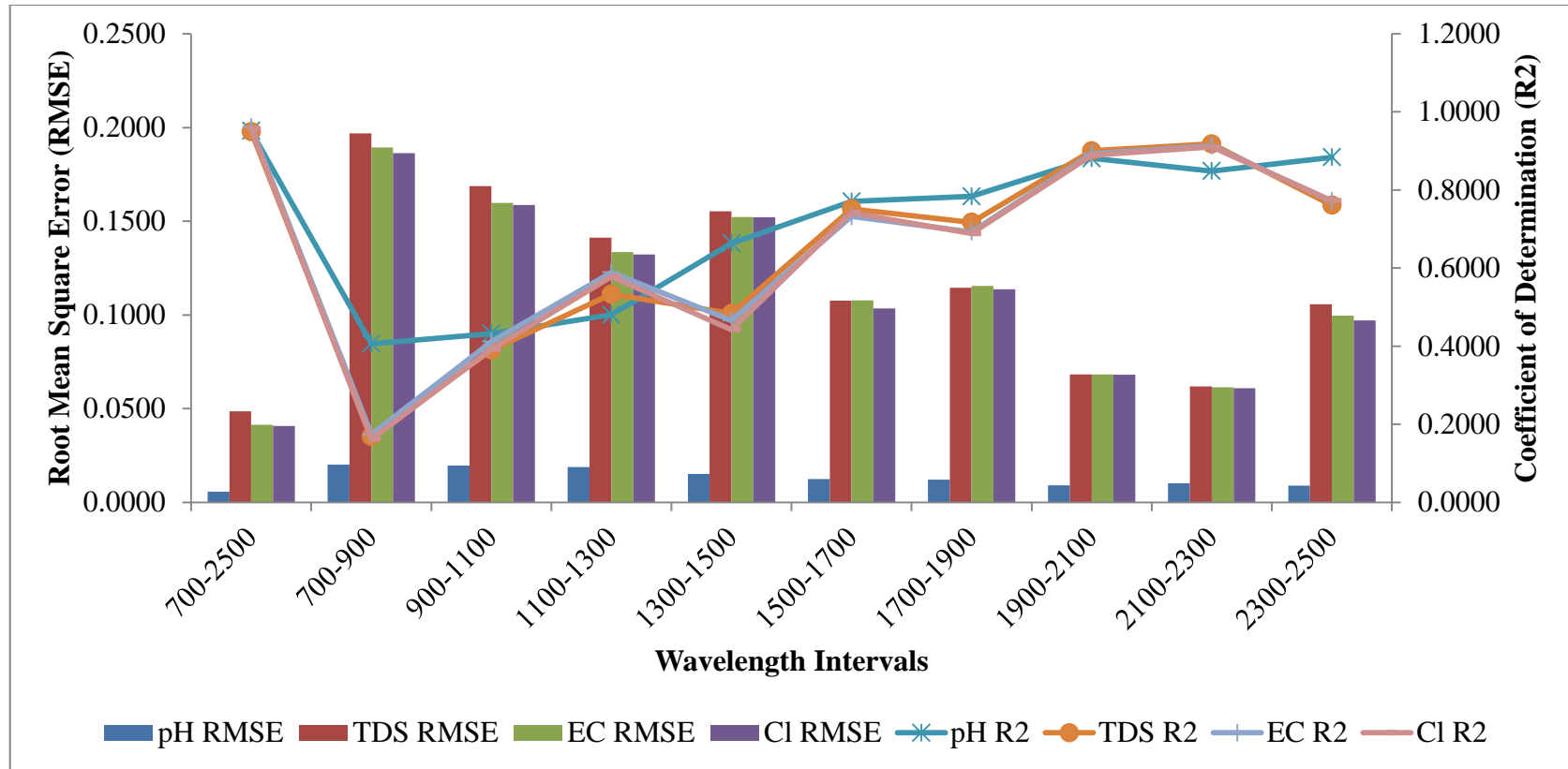
Figure 7.4 Root Mean Square Error and R2 for the iPLS (gap of 200 nm) for calibration and complete wavelength

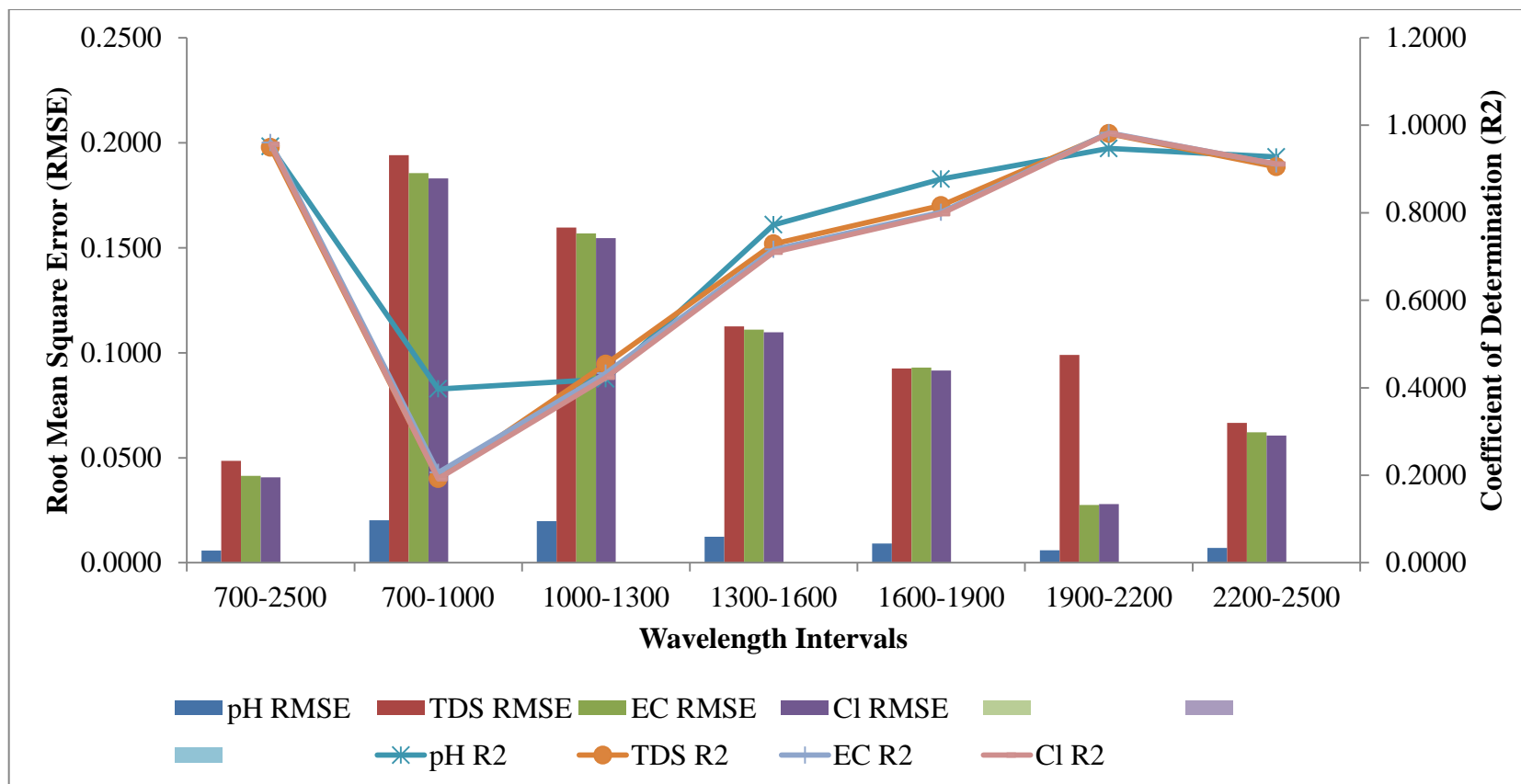**Figure 7.5** Root Mean Square Error and R2 for the iPLS (gap of 300 nm) for calibration and complete wavelength

Figure 7.6 Root Mean Square Error and R2 for the iPLS (gap of 600 nm) for calibration and complete wavelength

Figure 7.7 Root Mean Square Error and R2 for the iPLS (gap of 100 nm) for validation and complete wavelength

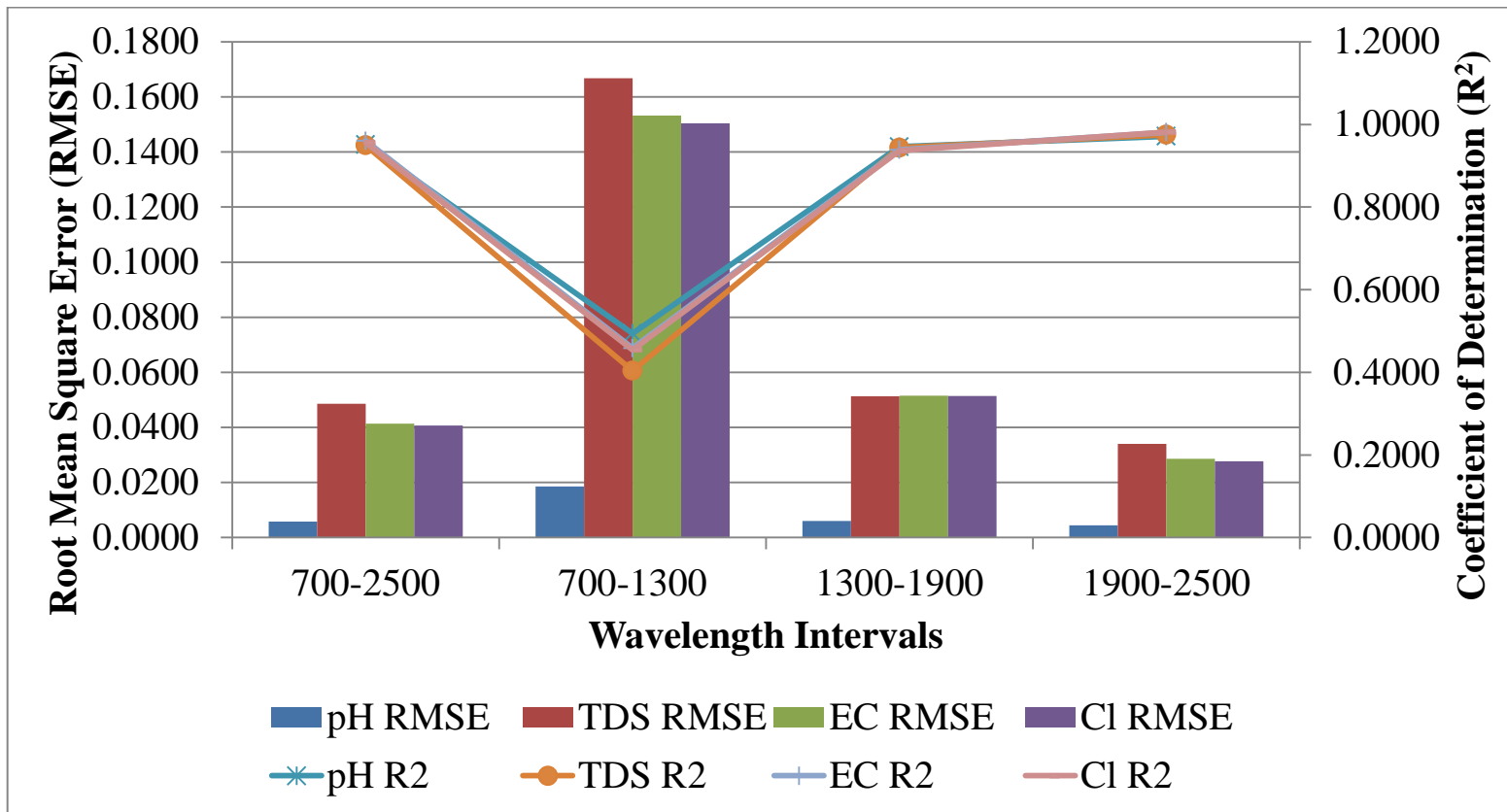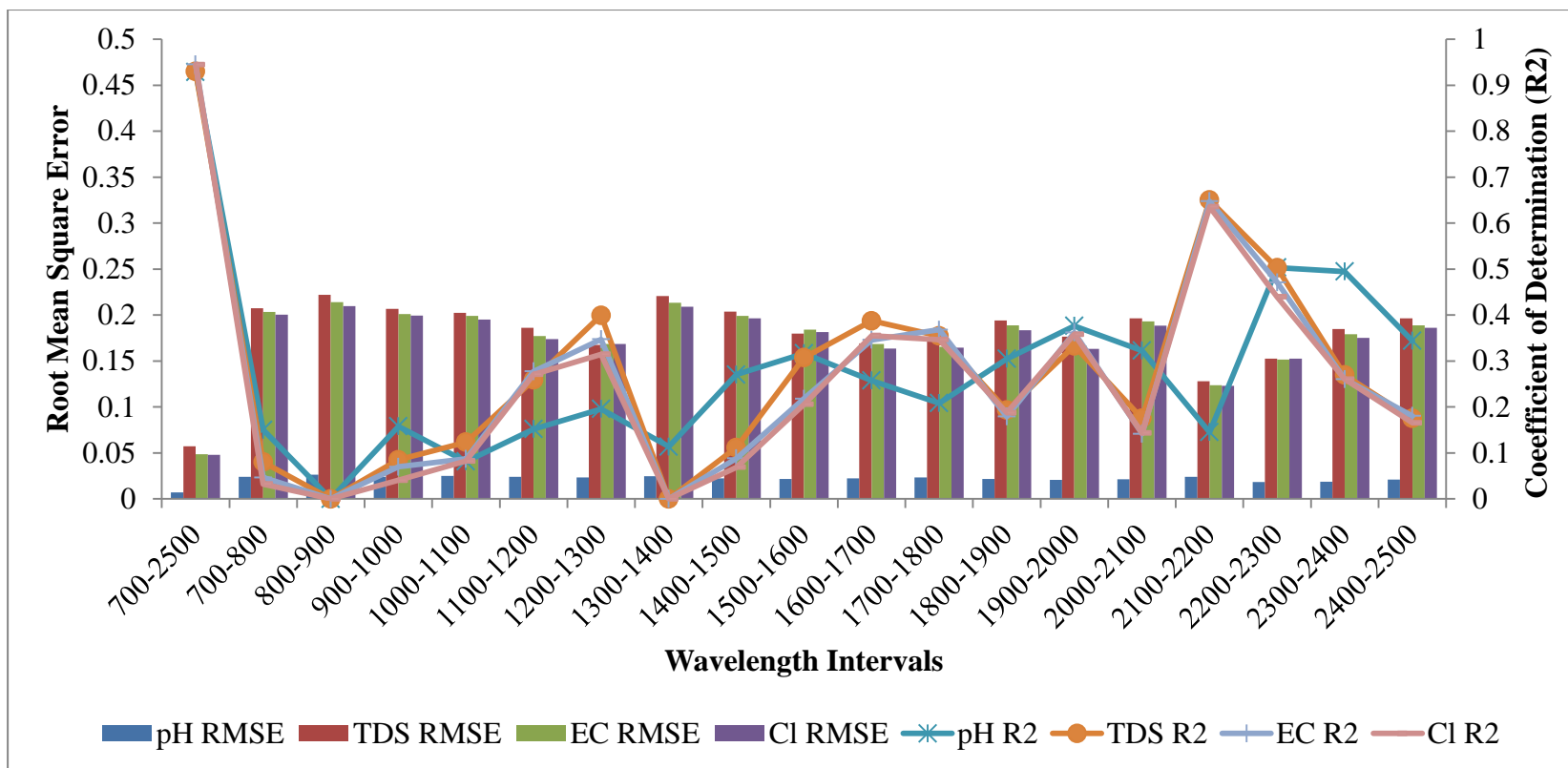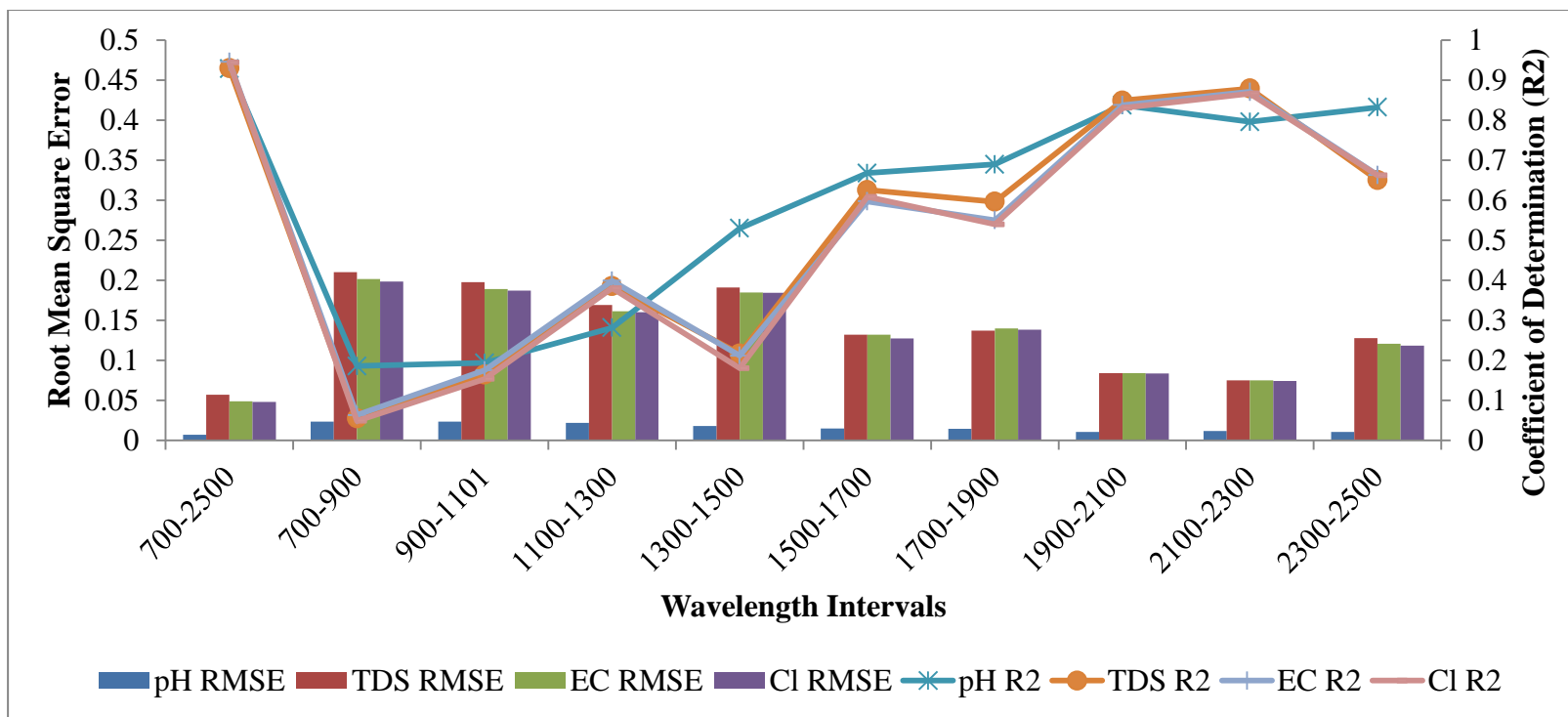Figure 7.8 Root Mean Square Error and R2 for the iPLS (gap of 200 nm) for validation and complete wavelength
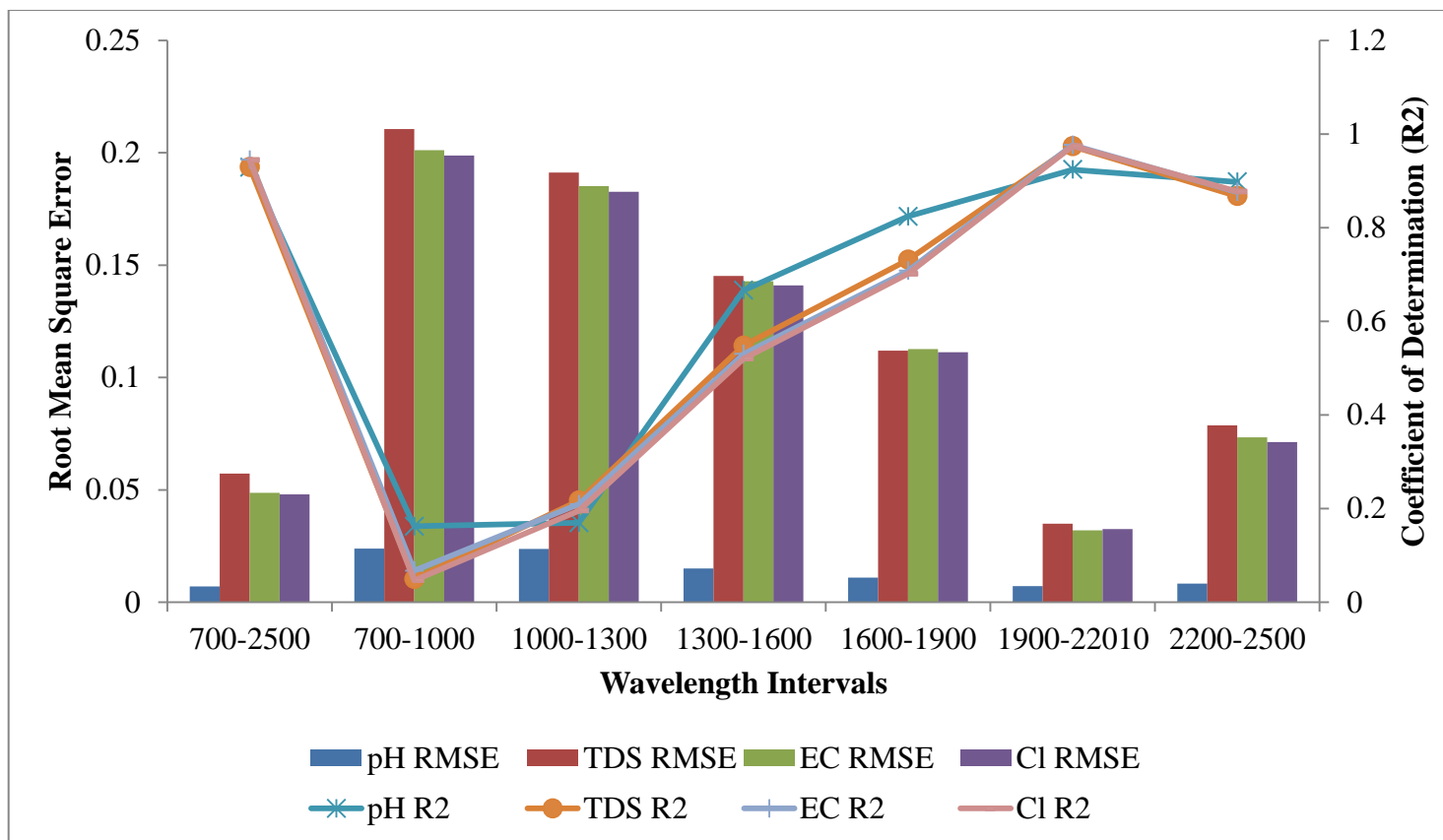
Figure 7.9 Root Mean Square Error and R2 for the iPLS (gap of 300 nm) for validation and complete wavelength
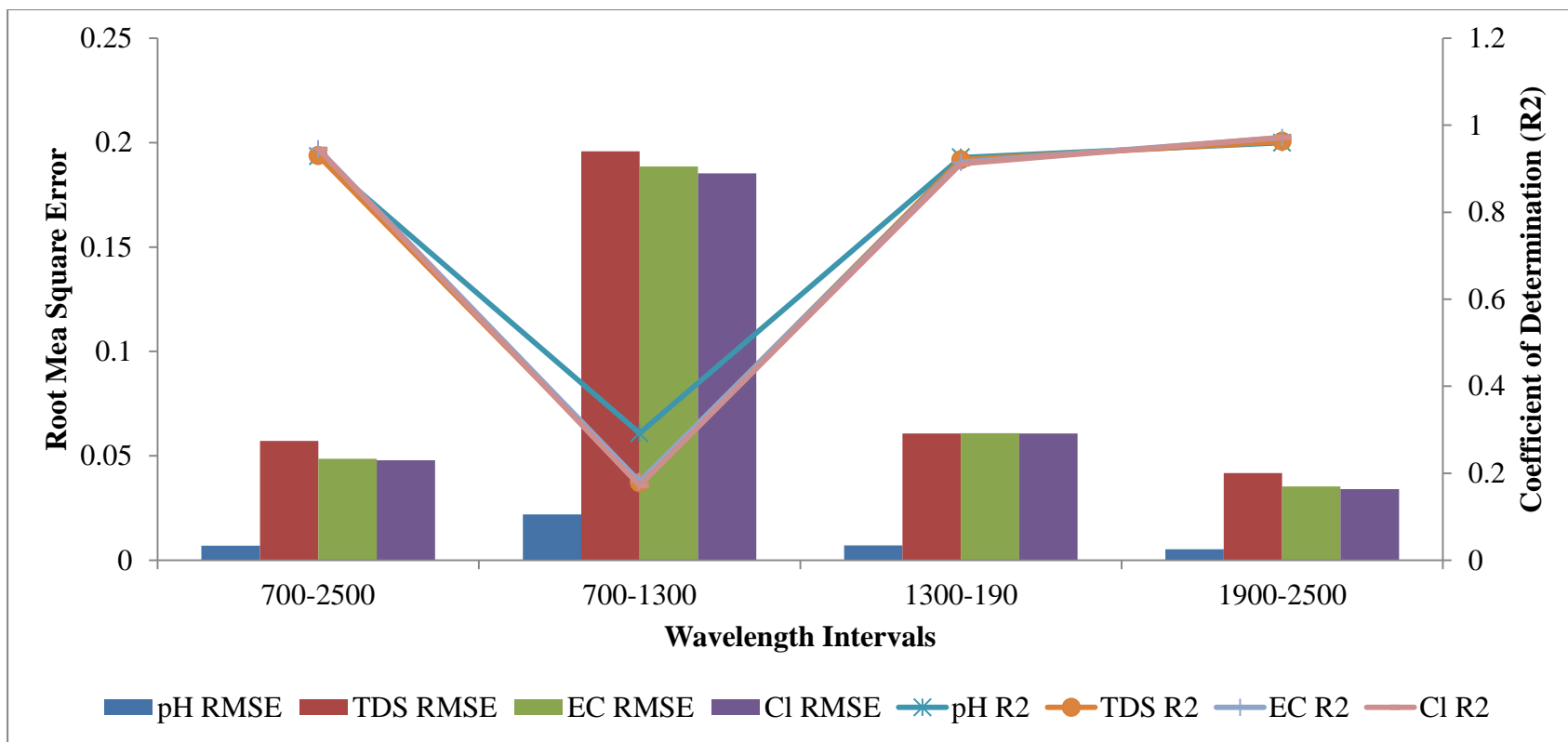
Figure 7.10 Root Mean Square Error and R2 for the iPLS (gap of 600 nm) for validation and complete wavelength

Wavelength selection method using siPLS

Further siPLS was performed using CAMO Unscrambler X 10.5 software, where the intervals with high value of correlation coefficient r^2 were combined and then regression model was developed for each parameter. Different iPLS models were developed for every parameter by making different combination of unequal widths,, which had low root mean square error. Results shows that siPLS for wavelength range of 1200-1300, 1900-2100 ; 12001300, 2100-2300; 1700-1800,1900-2100; 1700-1800,2100-2300 r^2 is greater than 0.8 (Table 7.6).

Table 7.6 Partial Least Square Regression model for the siPLS for all the parameters on the gap of 100nm iPLS results

	Wavelength Range	$RMSE_c$	R^2_c	$RMSE_v$	R^2_v
pH		0.00899	0.88096	0.01054	0.83625
TDS	1200-1300, 1900-	0.07119	0.89142	0.088	0.83409
EC	2100	0.07065	0.88498	0.08688	0.82605
Cl		0.70328	0.88064	0.08654	0.81929
pH		0.01014	0.84867	0.01177	0.79577
TDS	1200-1300, 2100-	0.06171	0.9184	0.07436	0.88152
EC	2300	0.06095	0.91439	0.07404	0.87368
Cl		0.06051	0.91163	0.07361	0.86926
pH		0.00582	0.95011	0.00686	0.93068
TDS	1700-1800, 1900-	0.06751	0.90234	0.08161	0.85729
EC	2100	0.06678	0.89722	0.08087	0.84927
Cl		0.06658	0.89303	0.08094	0.84192
pH		0.00773	0.91204	0.00913	0.87708
TDS	1700-1800, 2100-	0.05595	0.93292	0.06653	0.90517
EC	2300	0.0523	0.93696	0.06242	0.9102
Cl		0.05248	0.93354	0.06263	0.90535

The tables shows the siPLS results to check the accuracy of the regression models for the combination of the selected wavelengths. It was observed from the regression models that the range of wavelengths from 1200-1300nm were important for the prediction of all the parameters of ground water with their R^2 above 0.85. With comparison of the wavelength band with the standard NIR absorption band chart given in Fig 1.8 it can be observed that the in this region absorbance is due to the presence of second overtones of C-H, CH_2 and CH_3 bonds. Wavelength range 1900-2500 nm was also important for the determination of the parameters with more accurate regression model with R^2 greater than 0.9. Absorbance in this region can be observed for the determination of all the parameters as it corresponds to absorbance due to water molecules from 1900-2000 nm and combination bands in the region from 2000-2500nm. Further it was observed that wavelength range from 1300-1500 nm was only important to predict the pH of the ground water samples. It is because of the absorbance of H_2O and ROH bands in this region. Wavelength range 1700-1800nm was also considered important for prediction of all the parameters as these wavelengths correspond to absorbance due to the presence of first overtones of C-H, CH_2 and CH_3 bonds.

The results of the prediction showed that only the wavelengths in the region of 1200-1300nm, 1700-1800nm and 1900-2300nm were sufficient to predict the quantity of the pH, EC, TDS and chloride content of the ground water samples from the two districts of Malwa region of Punjab.

The developed model will be useful to determine these parameters and further estimate the presence of the pollutants on the basis of the values of the parameters using only the NIR spectra of the samples.

Conclusion

The chapter focuses on the use of NIR spectral data and parameter data for ground water samples to develop the regression models using partial least square regression. In PLS model it was observed that the analysis was done on 700-2500nm for all the parameters. It was observed that all the models had a high value of coefficient of determination (greater than 0.9) and very small root mean square error (less than 0.5). The literature revealed that all the 700-2500 nm are not important for determination of parameters, thus wavelength selection methods – interval PLS and synergy interval PLS were employed to find the important wavelengths for prediction of the parameters. 700-2500nm data was divided in different sets. Data was divided in equal intervals of equal width. Data was divided in intervals of 100nm, 200nm, 300 nm, and 600nm of equal width. PLS regression models were developed using the Camo Unscrambler software for all the intervals. The capability of the models was evaluated using the coefficient of determination and root mean square error. Results showed that wavelengths from 700-1300nm are least important and wavelengths from 1900-2500nm are more important for prediction of all parameters.

To further clarify with the important wavelengths. Combination of different bands was made. Wavelengths bands with high correlation values were combined in different combinations. PLS regression models were developed for all the combination and the best models were chosen for every parameter.

Research work showed that only few wavelengths were sufficient to determine the characteristics of ground water. The PLS regression model with the least wavelength sets and maximum coefficient of determination can be used by the public / societies to check the quality of ground water in a very quick, non hazardous and robust method.

References

[1].   International Association of Hydrogeologists [2020] Groundwater— more about the hidden resource. https://iah.org/education/gener_al-public/groundwater-hidden-resource. Accessed 13 Nov 2020

[2].   Li, Peiyue. "To make the water safer." *Exposure and health* 12.3 [2020]: 337-342.

[3].   Li, Peiyue, and Jianhua Wu. "Sustainable living with risks: meeting the challenges." *Human and Ecological Risk Assessment: An International Journal* 25.1-2 [2019]: 1-10.

[4].   Lin, Hangsheng. "Earth's Critical Zone and hydropedology: concepts, characteristics, and advances." *Hydrology and Earth System Sciences* 14.1 [2010]: 25-45.

[5].   Basu, Ankita, et al. "A review on sources, toxicity and remediation technologies for removing arsenic from drinking water." *Research on Chemical Intermediates* 40.2 [2014]: 447-485.

[6].   Pandey, H. K., S. K. Duggal, and Ajit Jamatia. "Fluoride contamination of groundwater and it's hydrogeological evolution in District Sonbhadra [UP] India." *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences* 86.1 [2016]: 81-93.

[7].   Subba Rao, N., B. Ravindra, and Jianhua Wu. "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 [2020]: 2316-2348.

[8].   He, Xiaodong, et al. "Poor groundwater quality and high potential health risks in the Datong Basin, northern China: research from published data." *Environmental Geochemistry and Health* 43.2 [2021]: 791-812.

[9].   Li, Peiyue, et al. "Sources and consequences of groundwater contamination." *Archives of environmental contamination and toxicology* 80.1 [2021]: 1-10.

[10].   MacDonald, Jacqueline A., and Michael C. Kavanaugh. "Restoring contaminated groundwater: an achievable goal?." *Environmental Science & Technology* 28.8 [1994]: 362A-368A.

[11]. Chakraborti, Dipankar, et al. "Groundwater arsenic contamination in Bangladesh—21 Years of research." *Journal of Trace elements in Medicine and Biology* 31 [2015]: 237-248.

[12]. Smil, Vaclav. "Long-range perspectives on inorganic fertilizers in global agriculture, 1999 Travis P Hignett lecture." *Alabama, USA* [1999].

[13]. Bokobza, L. "Near infrared spectroscopy." *Journal of Near Infrared Spectroscopy* 6.1 [1998]: 3-17.

[14]. Agelet, Lidia Esteve, and Charles R. Hurburgh Jr. "A tutorial on near infrared spectroscopy and its calibration." *Critical Reviews in Analytical Chemistry* 40.4 [2010]: 246-260.

[15]. Davies, A. M. C. "An introduction to near infrared spectroscopy." *NIR news* 16.7(2005):9-11.

[16]. Bokobza, L. "Near infrared spectroscopy." *Journal of Near Infrared Spectroscopy* 6.1 [1998]: 3-17.

[17]. Hart, Joe R., Karl H. Norris, and Calvin Golumbic. "Determination of the moisture content of seeds by near-infrared spectrophotometry of their methanol extracts." *Cereal Chem* 39.2 [1962]: 94-99.

[18]. NORRIS, KARL H. "Direct spectrophotometric determination of moisture content of grain and seeds." *Proc. of the 1963 International Symposium on Humidity and Moisture*. Vol. 4. 1965.

[19]. American Association of Cereal Chemists, and American Association of Cereal Chemists. Approved Methods Committee. "Approved Methods of the American Association of Cereal Chemists." AACC, 1983.

[20]. Weyer, Lois. "History of Near-Infrared [NIR] Applications." *Practical Guide to Interpretive Near-Infrared Spectroscopy*. CRC Press, 2007. 119-124.

[21]. Christy, Alfred A., et al. "The detection and quantification of adulteration in olive oil by near-infrared spectroscopy and chemometrics." *Analytical Sciences* 20.6 [2004]: 935-940.

[22]. Spellman, Frank R. *The drinking water handbook*. CRC Press, 2017.

[23]. Edzwald, James, and American Water Works Association. *Water quality & treatment: a handbook on drinking water*. McGraw-Hill Education, 2011.

[24]. Hammer MJ. Water and Wastewater Technology. 7th ed. Upper Saddle River: Pearson education; 2011

[25]. Alley, E. Roberts. *Water quality control handbook*. Vol. 2. New York: McGraw-Hill, 2007.

[26]. Edition, Fourth. "Guidelines for drinking-water quality." *WHO chronicle* 38.4 [2011]: 104-108

Conclusion

The Ph.D research work focuses on quantitative and qualitative analysis of ground water samples procured from the two districts of Malwa region.

Primary task was to identify the pollutants on the basis of the FTIR spectra of the samples and pollutants. Two pollutants were considered for observation – Urea and Arsenic trioxide. Comparison of spectra of FTIR of the samples and pollutants revealed that Urea was present in maximum samples due to presence of absorbance around 3300 cm^{-1}. Further the absorbance around 1700-2500cm^{-1} also shows that there may be arsenic trioxide present in the samples (Chapter4).

Next task was to analyse the samples on the basis of their physicochemical (reference data set) parameters. Descriptive statistics of the reference data set revealed that the mean value of the pH of the sample was 7.31 and there was least variation between the pH of all the samples with standard deviation of 0.21. Average value of TDS, EC and chloride content were 589.84 mg/L, 1290.62 µS/cm and 303.78 mg/L respectively. Large variation was observed in this value of the ground water samples with standard deviation of 305.08 mg/L, 951.14 µS/cm, 228.16 mg/L respectively. The results moved the research to be more focused on the variation and further analysis of the TDS, EC and chloride content of the samples. The correlation analysis was performed between these parameters and it was observed that there is very strong correlation between TDS-EC, EC-chloride content and TDS-chloride content with their coefficient of correlation greater than 0.95. Thus regression analysis were performed and relations were developed to calculate one parameter from the other (Chapter 5).

The relation between the samples was also understood by developing the score plot of principal component analysis (PCA) which showed that the chloride content and EC are more closely related as compared to their relation with TDS and properties of TDS, EC and chloride content are opposite to the pH parameters (Chapter 5).

Classification of the samples using the reference data could clearly classify the samples of the two districts. It was observed that the samples from Fazilka were present on the right side of the PCA with positive values of PC 1 and that of Ferozepur district was present on the left of the PCA with negative values of PC1.

Reference data was used to classify the samples on the basis of region – I, U and R, source of samples – WB, SS and HP. It was observed that samples could be classified on the basis of above parameters (Chapter 5).

Effect of depth was also studied on the collected samples. It was observed that the samples with width more than 100ft had their values in the permissible levels given by WHO and BIS. To check the relation between the different properties, correlation and regression analysis were employed on the data. It was observed that there was strong correlation observed between EC, TDS and chloride content of all the sample, irrespective of their origin, source and depth. To predict one parameter from the other, regression equations were built for the parameters.

Classification of the samples were also done using only the NIR spectral data (Chapter 6). The complete data set was classified using the PCA analysis for spectral data on the basis of region – I, U, R; district, depth and source of the collected ground water samples. It was observed that the samples of different districts could be classified into two groups on the basis of spectral data, but results were not satisfactory for other PCAs, thus next PCA was applied on the two districts separately. PCA results on two districts shows that the samples can be classified on the basis of region. Samples from urban region can be distinguished from the samples of rural and industrial region. Source of the samples can also be identified using the spectral data. The results showed the samples from the different region, sources and depth can be classified in different groups with the use of only the NIR spectral data, which is a non-destructive, rapid and non-hazardous method.

It has been observed that there is presence of Urea and Arsenic trioxide in the sample and also that their presence can be confirmed with the values of TDS, EC and chloride content. Last objective of the Ph.D. work was to develop a model to determine the parameters-pH, TDS, EC and chloride content of the samples using the NIR spectra. Development of the model will be useful in determining the values of the parameters of the ground water samples of the two districts of Malwa region in a non destructive, rapid and non hazardous manner. PLS models for the prediction of parameters were developed using the complete range from 700-2500nm. To reduce the analysis time and cost of the instrument to be developed in future, effort was made to select the wavelengths important for the prediction of specific properties. The complete spectral

data was divided into equal intervals of 100nm, 200nm, 300nm and 600nm of equal width and regression model for the same was developed and their R^2 and RMSE values were compared to pick the wavelengths with more correlation of prediction (iPLS technique). It was observed that wavelength range from 700-1200 nm had their value of R^2 less than 0.5. Thus for the applying synergy interval PLS where the wavelength ranges with values of R^2 are higher than 0.6 are considered. Different wavelength range combinations were made on the basis of the values of iPLS and siPLS was employed for all the parameter. After developing different siPLS regression models it was found that the wavelength range 1200-1300 nm due to the absorbance of C-H, CH_2 and CH_3 bonds, 1900-2300nm due to combinational bands and 1700-1800nm due to absorbance of O-H and H_2O were important for prediction of all parameters with R^2 values greater than 0.85.

Thus it was finally confirmed that only the wavelength ranges from 1200-1300nm, 1700-2300nm in all were sufficient for the prediction of pH, EC, TDS and chloride content of ground water samples of ground water samples of two districts of Malwa region of Punjab, India.

The developed siPLS models can be successfully used to determine the quantity of the parameters of ground water samples in a fast, non-destructive, non-hazardous way, resulting in green and friendly methodology.

Regression Model developed can be incorporated in the multivariate software used by the industries and the model can be used to predict the water parameters which may be used to know the presence of the contaminants for the sample collected from Fazilka and Ferozepur Districts.

FUTURE SCOPE

- Developing of an instrument which is handheld and fast for ground water quality analysis using spectroscopy and chemometrics for the complete Malwa region and then whole Punjab region.

- Identification of more contaminants quantitatively using spectroscopic techniques and also regression model can be developed with selected wavelengths for estimation of these contaminants with this method.

- There are some other spectroscopic techniques like Fourier Transform infrared spectroscopy, UV-Vis spectroscopy which can be explored for the same.

CERTIFICATE OF PUBLICATION OF PAPERS FOR PH.D.

This is to certify that Mr./Ms. **Reena Rani** pursuing Ph.D. (**Full Time/ Part Time**) programme in Department of **Physics** with Registration Number **41900387** under the Guidance of **Dr. Uma Kamboj** has the following Publications / Letter of Acceptance in the Referred Journals / Conferences mentioned thereby fulfilling the minimum programme requirements as per the UGC.

Sno.	Title of paper with author names	Name of journal / conference	Published date	ISSN no/ vol no, issue no	Indexing in Scopus/ Web of Science/UGC-CARE list (*please mention*)
1.	Classification of Groundwater Samples from Two Major Districts of the Malwa Region of Punjab, India Based on Their Physicochemical Properties.	Water and Energy International	Accepted on 16 August 2022.	0974-4207	Indexed in Scopus SJR 0.13 https://www.scopus.com/sourceid/21100920050

	Reena Rani, Neha Munjal and Uma Kamboj*				
2.	A Review on Different Techniques for Analysis for Uranium as a Ground Water Contaminant at AFMD. Reena Rani, Neha Munjal and Uma Kamboj*	Advanced Functional Materials and Devices, Springer Springer Proceedings in Materials book series	02 Dec 2021	vol 14. 978-981-16-5970-6	Scopus Springer https://doi.org/10.1007/978-981-16-5971-3_22
3.	Characterization of ground water using spectroscopic techniques. Reena Rani, Sunil Kumar, Neha Munjal and Uma Kamboj*	Journal of Physics: Conference Series 3rd International Conference on Recent Advances In Fundamental And Applied Sciences 25-26 June 2021	26 June 2022	2267 (2022) 012022 Print ISSN: 1742-6588	Scopus and WOS doi:10.1088/1742-6596/2267/1/012022
4.	Spectroscopic technique	AIP Conference	Accepted		

	for pollutants identification in water a review Reena Rani, Neha Munjal and Uma Kamboj*	Proceedings International Conference on Materials for Emerging Technologies (ICMET-21) from 23-24 December 2021	Manuscript ID: MS-262 20Aug 2022		

Signature of Candidate with Date, Registration No, Email ID

Signature of Guide with Date & UID   **Signature of Co-Guide with Date & UID**