CHAPTER-4

MATERIAL AND METHODOLOGY

4.1 General Information

The extraction of minerals from the Earth's surface generally results in an imbalance and adverse effects on the surrounding environment. However, the magnitude of these impacts differs among different mines, depending on factors such as the extent of mining activities, hydrological conditions, geological characteristics, and other relevant aspects specific to each location. Mining has been associated with several adverse impacts, such as deforestation, disruption of water resources, alteration of local climate conditions, changes in rainfall patterns, degradation of habitats for wildlife and fisheries, and ecological modifications. This study focuses on the environmental impact of the KCF mine, which is the largest mine in Asia and the second largest globally. To achieve the objectives, a systematic approach was employed. The following paragraph provides a brief overview of the materials and methods.

4.2 The required data collection

There is the few secondary information, including publications, study-related maps, and site features, has been obtained from various government entities. Furthermore, an extensive examination of the relevant literature within the study topic was conducted by accessing a number of publications, exploring online resources, and carefully reviewing significant papers. The mentioned government organisations were very helpful in achieving the study's goal.

i. South Eastern Coalfield Limited (SECL)

ii. Korba Coalfield Chhattisgarh

iii. CMPDIL Rachi

iv. Bhuvan-National Remote Sensing Centre (ISRO)

v. Bhukosh-Geological Survey of India (GSI), Kolkata

vi. USGS Earth Explorer

vii. Ministry of Mines

viii. Indian Meteorological Department, Kolkata

ix. Global Energy Monitor Wikipedia

x. Census India

xi. Central Ground Water Board (CGWB)

xii. Power Data Assess Viewer (NASA)

4.3 Important software and Instrument

The GPS was utilised to obtain the accurate geographic coordinates of the selected sites in the study region during surface and groundwater sampling. The following enumeration comprises several software applications, such as ArcGIS 10.8, ERDAS Imagine, QGIS, and Origin Pro 2017, which are extensively utilized for the purpose of preparing thematic maps from significant satellite imagery. Furthermore, the study employed many other software packages, including Past 4.03, Grapher, Diagrammes, Surfer 16, and Minitab 18, to analyse a variety of maps concerning the presence of anthropogenic sources of elemental chemistry in water resources. likewise, a variety of instruments, including multiparameter analysers, ion chromatography systems, and inductively coupled plasma mass spectrometry (ICP-MS) instruments, were employed in this study to evaluate both surface water and groundwater.

4.4 Water sampling, processing, and laboratory testing

4.4.1 Sampling of water

A preliminary survey of the study region was conducted prior to collecting the water sample. Multiple locations throughout the KCF region had been selected for the surface and groundwater samples collection. The criteria for selection of water sampling locations depends on available villages around the mining area where people are living and using water for their livelihood from bore well, hand pump and dug well. The collection, preservation, and analysis of water samples followed the standard guideline forward by the American Public Health Association (APHA, 2005). Some of the uncertainty in data collection, methods of analyses, and findings can be happened due to sampling procedures, spatial and temporal coverage, analytical precision and accuracy, human error and statistical variability. Employing standardized protocols, advanced analytical techniques, robust statistical methods, and comprehensive uncertainty analysis can help mitigate these uncertainties and lead to more accurate and reliable assessments of water quality in the study area. The sterilized polypropylene bottles were washed twice with distilled water having 400 ml capacity and used to collect the water samples. In the case of surface water, bottles were used to collect samples from a depth of 0.5 m to avoid the suspended particles. There are few key precautions to take while collecting random water samples to avoid contamination such as use clean and sterile equipment, proper handling techniques, sample collection method, equipment decontamination, sample preservation and storage, sample transportation, site-specific considerations. In order to assess the chemical composition and overall quality of surface and groundwater in the KCF region during pre and post monsoon season of 2022, a comprehensive collection of 120 water samples were conducted by following the standard guideline of APHA (2005). These samples consisted of 34 surface water samples and 86 groundwater samples, all of which

were obtained from 60 distinct locations effectively selected that covered the entirety of the study region. The study area map and sampling location map is shown in Fig.4.1 and Fig.4.2. The few photographs have captured during water sampling in both seasons as shown in Fig.4.3.

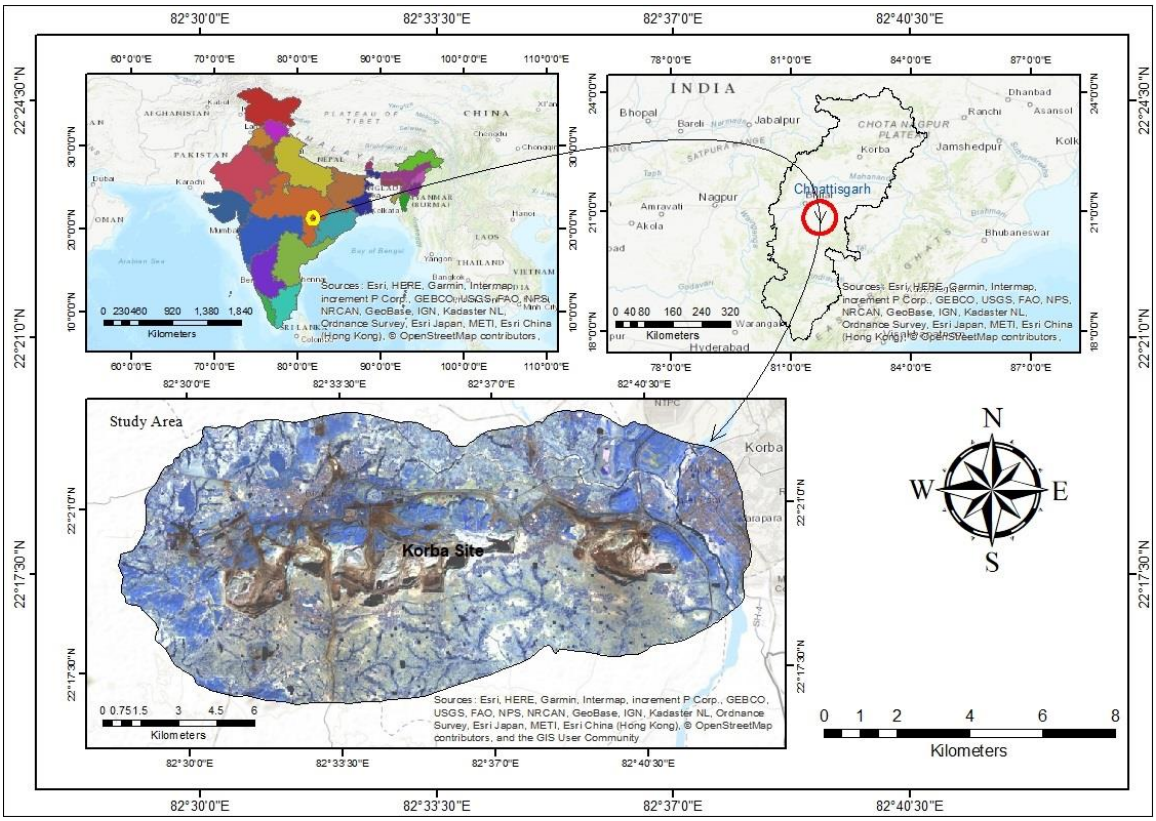

Fig.4.1 Study area map of KCF region

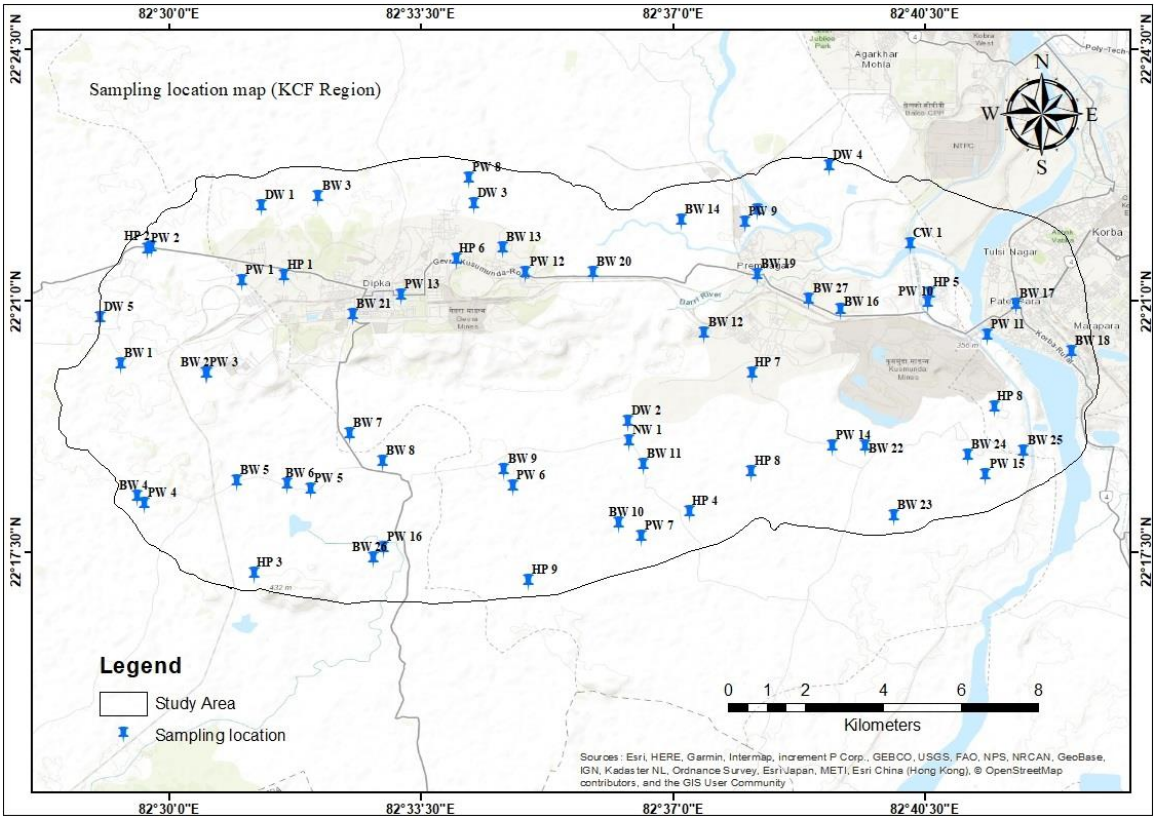

Fig.4.2 Sampling locations map of study region

Fig.4.3 Photograph showing water sampling from different sources in study area

4.4.2 Bottles labelling and Samples processing

Every container was properly tagged with all of the necessary printed information before the sampling began. This information included the sample code, the name of the sampling

location, the time and date of the sampling, and the source of the water as depicted in Fig.4.4. The surface and groundwater samples were obtained from selected locations and stored in sterilised polypropylene bottles. All the collected samples must be treated with the necessary care. It is preferable to promptly analyse the water samples that were collected. However, due to logistical constraints, immediate analysis was not feasible. Consequently, the water samples were sent to the laboratory and subsequently stored at a temperature of 4°C to mitigate any potential alterations. Each site yielded two sample sets. One set was filtered through a 0.22 µm Nylon syringe filter and acidified with nitric acid to achieve a pH below 2. This set was intended for metal analysis. The other set of samples were collected for ion analysis (APHA 2005).

Fig.4.4 Photograph showing the collected water samples with its proper labelling

4.4.3 Method of analysis

The analytical procedures for groundwater samples in detail are shown in Fig.4.5, while for surface water samples are shown in Fig.4.6. The following section offers a comprehensive explanation of each significant step described within the analytical methodologies.

4.4.3.1 In-situ Analysis

A few parameters were examined immediately after sampling. These parameters included pH, TDS, and EC, were measured using a calibrated multiparameter device named as ISO-TECH System as given in Table 4.1. The use of a multiparameter instrument was considered for in-situ monitoring during sampling period.

Table.4.1 In-situ analysis of physicochemical parameters

S No.	Parameters	Instruments	Instrument Facilitated by
1.	pH	Multiparameter analyser	Mining Engineering Lab
2.	TDS	Multiparameter analyser	Mining Engineering Lab
3.	EC	Multiparameter analyser	Mining Engineering Lab

4.4.4.1 Laboratory Testing

Prior to analysis for parameters, all water samples underwent filtration using a 0.22 µm Nylon syringe filter to remove particulates matters, protect the instrument, and ensure accurate and reliable results. Ion-Chromatography (IC) and Induced Coupled Plasma Technique (ICP-MS) instruments were employed to measure the concentrations of cations (calcium, magnesium, sodium, and potassium), anions (fluoride, chloride, nitrate, sulphate, and bicarbonate), along with heavy metals (aluminium, barium, cadmium, chromium, copper, iron, manganese, lead, nickel, and zinc) in water samples in Central Instrument Facility (CIF) IIT (BHU) Varanasi by following the standard analytical methods proposed by (APHA 2005). Some photographs were presented here of laboratory analysis as shown in Fig.4.7.

Table.4.2 Analytical methods for different parameters and heavy metals used for water quality assessment

S No.	Parameters	Instruments	Lab Facility
1.	Calcium (Ca^{2+})	Ion-Chromatography	CIF, IIT (BHU)
2.	Magnesium (Mg^{2+})	Ion-Chromatography	CIF, IIT (BHU)
3.	Sodium (Na^{+})	Ion-Chromatography	CIF, IIT (BHU)
4.	Potassium (K^{+})	Ion-Chromatography	CIF, IIT (BHU)
5.	Fluoride (F^{-})	Ion-Chromatography	CIF, IIT (BHU)
6.	Chloride (Cl^{-})	Ion-Chromatography	CIF, IIT (BHU)
7.	Nitrate (NO_3^{-})	Ion-Chromatography	CIF, IIT (BHU)
8.	Sulphate (SO_4^{2-})	Ion-Chromatography	CIF, IIT (BHU)
9.	Bicarbonate (HNO_3^{-})	Ion-Chromatography	CIF, IIT (BHU)
10.	Total Hardness	Titration method	Environmental Lab
11.	Aluminium (Al)	ICP-MS	CIF, IIT (BHU)
12.	Barium (Ba)	ICP-MS	CIF, IIT (BHU)
13.	Cadmium (Cr)	ICP-MS	CIF, IIT (BHU)
14.	Chromium (Cd)	ICP-MS	CIF, IIT (BHU)
15.	Copper (Cu)	ICP-MS	CIF, IIT (BHU)
16.	Iron (Fe)	ICP-MS	CIF, IIT (BHU)
17.	Manganese (Mn)	ICP-MS	CIF, IIT (BHU)
18.	Lead (Pb)	ICP-MS	CIF, IIT (BHU)
19.	Nickel (Ni)	ICP-MS	CIF, IIT (BHU)
20.	Zinc (Zn)	ICP-MS	CIF, IIT (BHU)

ICP-MS- Inductively Coupled Plasma-Mass Spectrometry, CIF-Central Instrument Facility

The titrating standard solution of ethylene diamine tetra acid (EDTA) was used to determine the total hardness using complexometric titration methods. The titration method was used to determine the concentration of bicarbonate in the sample, complying with the standard guideline established by (APHA 2005).

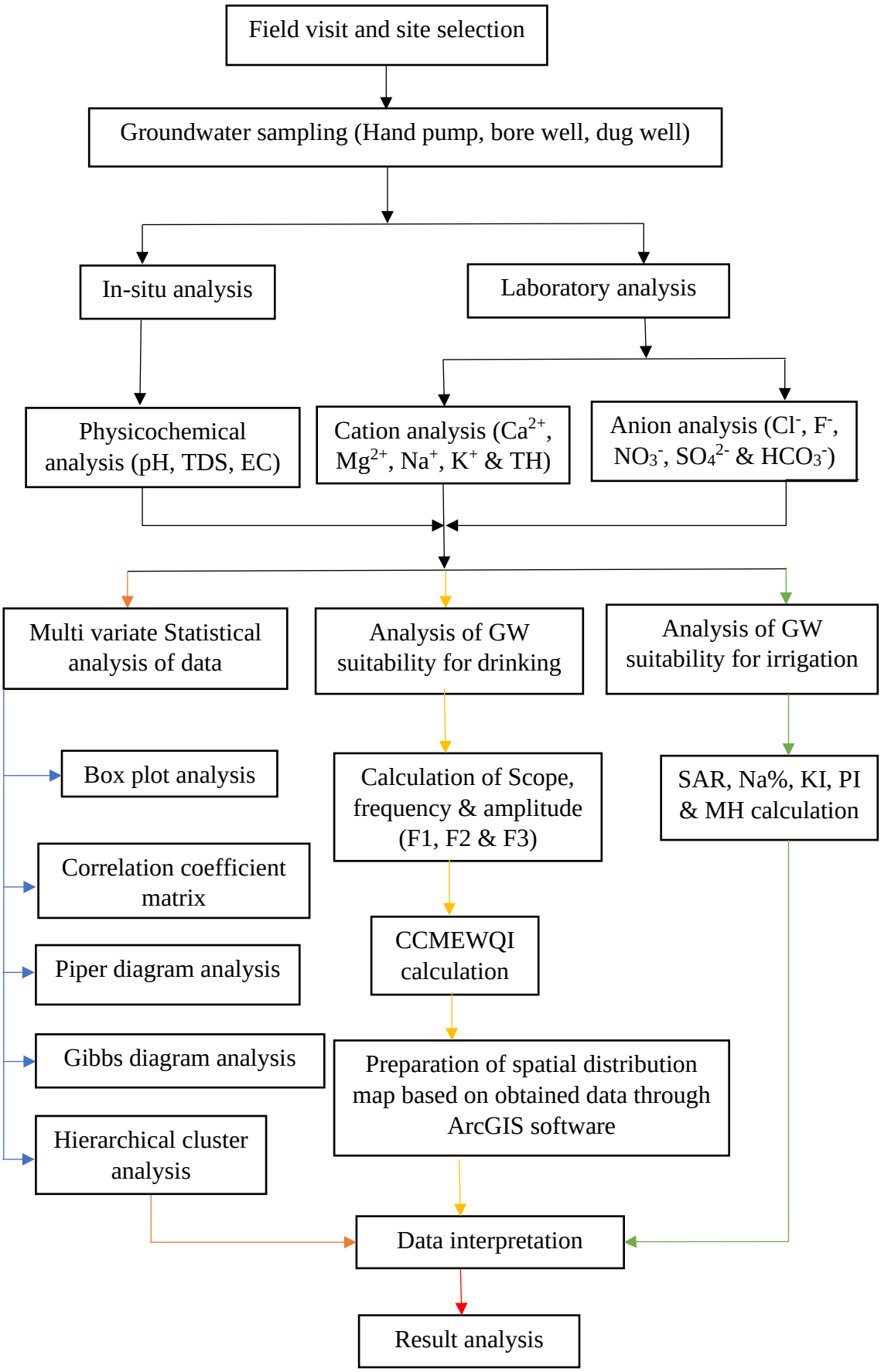

Fig.4.5 Flow chart showing analytical method for groundwater

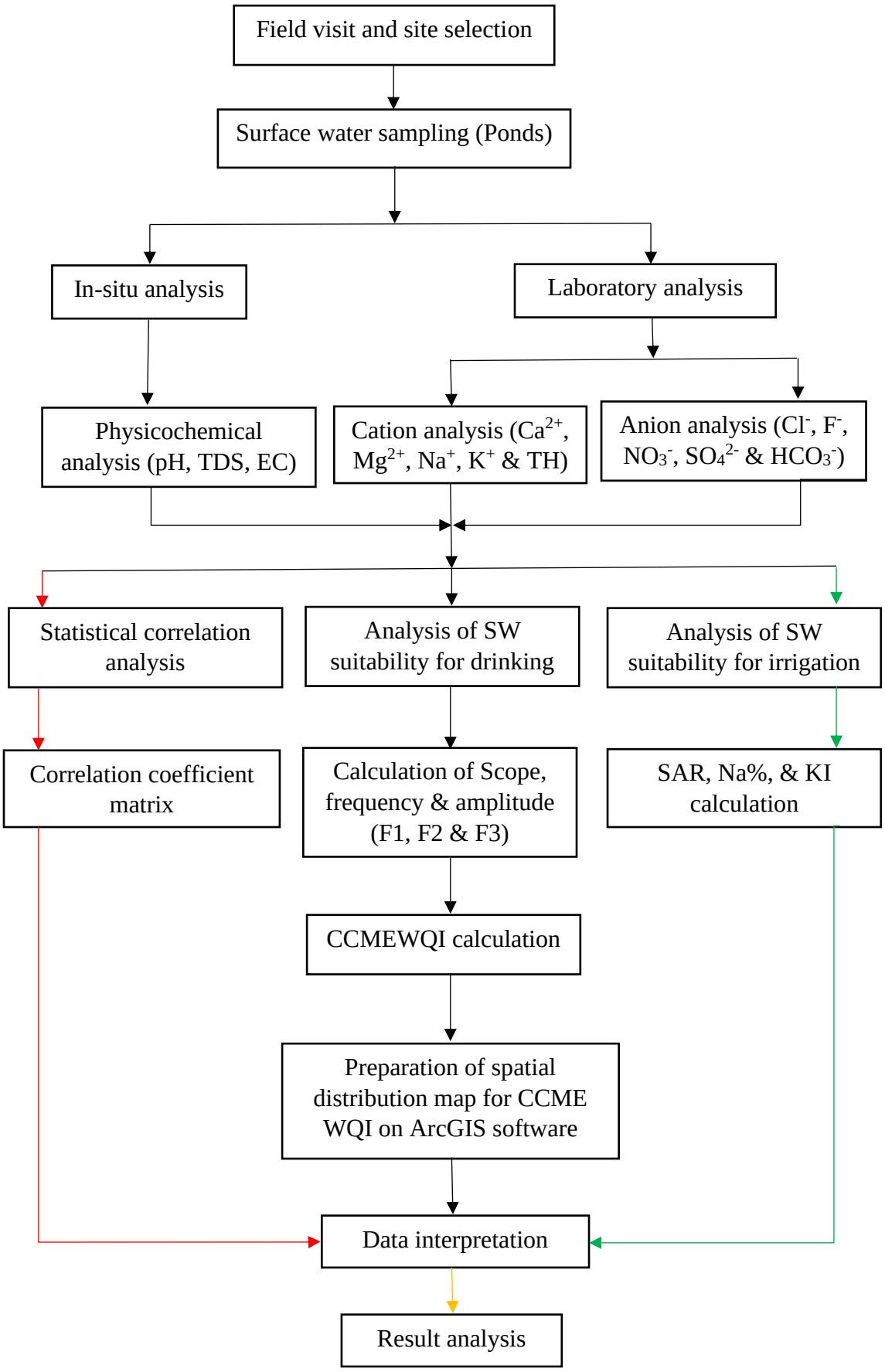

Fig.4.6 Flow chart showing analytical method for surface water

Fig.4.7 Photograph showing Instruments along with parameters analysis

4.5 Water quality assessment technique for drinking purposes

The WHO and BIS recommended limits were employed to compare the measured concentrations of various parameters in order to assess the suitability of water for drinking and domestic purposes.

4.5.1 CCME WQI calculation Procedure

The water quality index was determined by using the Canadian Council of Ministers of the Environment Water Quality Index (CCME WQI). The British Columbia Ministry of Environment, Lands, and Parks initially formulated this index, which has since been updated with revisions from the province of Alberta. The CCME WQI offers a practical method for condensing complex data on water quality and making it easier for readers to understand.

The CCME WQI proves valuable in monitoring and evaluating alterations in water quality at a specific site across time, while also enabling comparisons with other locations. Based on the approach proposed by the CCME WQI assessment, the index values are computed within a range of 0 to 100. A higher index value is indicative of an increase in water quality, while a lower value suggests a deterioration in water quality (CCME 2001). The CCME WQI calculation relies on three key factors: the scope (F1), frequency (F2), and amplitude of excursions (F3). Water quality is segmented into five categories determined by the CCME WQI value as listed in Table 5.6 in next chapter.

Scope (F1) and Frequency (F2) are relatively simple to calculate, while Amplitude (F3) takes some additional procedures. The Technical Report of CCME WQI 1.0, has given the related formula of CCME WQI calculation (CCME 2001).

Step-1: F1 (Scope) reflects the percentage of parameters that do not fulfil their standards at least once during the time period under consideration ("failed parameters"), as a percentage of all parameters measured:

$$F1 = \frac{Number\ of\ failed\ variables}{Total\ number\ of\ variables} \times 100 \quad \ldots\ldots (4.1)$$

Step-2: F2 (Frequency) denotes the proportion of individual tests that fail to fulfil criteria ("failed tests"):

$$F2 = \frac{Number\ of\ failed\ tests}{Total\ number\ of\ tests} \times 100 \quad \ldots\ldots (4.2)$$

Step-3: F3 (Amplitude) represents the amount by which failed test values do not reach their standards and F3 calculated in 3 stages.

Stage-I: The number of times by which an individual concentration is greater than (or less than, when the guideline is a minimum) the guideline is termed an "excursion" and is expressed as follows. When the test value must not exceed the guideline:

$$Excursion_i = \frac{Failed\ test\ value\ (i)}{Objective\ (j)} - 1 \quad \ldots\ldots (4.3)$$

For the cases in which the test value must not fall below the guideline:

$$Excursion_i = \frac{Objective\ (j)}{Failed\ test\ value\ (i)} - 1 \quad \ldots\ldots (4.4)$$

Stage-II: The total amount by which individual tests are out of compliance is derived by adding the excursions of each test from their standards and dividing by the total number of tests. This measure, known as the normalized sum of excursions, or nse, is calculated as

$$nse = \frac{\sum_{i=1}^{n} excursion\ (i)}{Number\ of\ tests} \quad \ldots\ldots (4.5)$$

Stage-III: F3 is then computed using an asymptotic function that scales the normalized sum of the excursions from guidelines (nse) to yield a range between 0 and 100.

$$F3 = \frac{nse}{0.01nse+0.01} \quad \ldots\ldots (4.6)$$

And Finally, CCME WQI is calculated with following equation (5):

$$\text{CCME WQI} = 100 - \frac{\sqrt{F1^2+F2^2+F3^2}}{1.732} \quad \ldots\ldots (4.7)$$

The divisor 1.732 adjusts the resulting values to a range of 0 to 100, with 0 reflecting "worst" water quality while 100 reflecting "best" water quality as categorized by CCME WQI ranking.

4.6 Water quality assessment technique for irrigational purposes

The irrigation water used should be sufficient to meet the soil and plant needs for healthy plant growth. Electrical conductivity and Na^+ play an important role in plant growth. The osmotic pressure in the soil solution increases due to the application of irrigation water containing a high concentration of salts. In addition to having a direct influence on plant growth, salts can also exert an indirect influence on plant growth by modifying soil permeability, aeration, and structure. There are many indices that can be used to assess the suitability of irrigation water, including SAR, %Na, KI, PI and MH as mentioned below.

4.6.1 Sodium Absorption Ratio (SAR)

SAR refers to the concentration of sodium ions (Na^+) in the water extract derived from a saturated soil paste, in comparison with the concentrations of calcium ions (Ca^{2+}) and magnesium ions (Mg^{2+}). It is calculated by dividing the Na^+ content by the sum of squares root of one-half of the (Ca^{2+} + Mg^{2+}) concentrations. Elevated levels of Na+ can modify soil properties, diminishing soil permeability and leading to the development of alkaline soil conditions. The absolute and relative concentrations of cations are utilized to compute the Na+ or alkali hazard, often represented by SAR. SAR can be used to express the salt hazard for irrigation water (Richards, 1954).

$$SAR = \frac{Na}{\sqrt{(Ca+Mg)/2}} \quad \text{…. (4.8)}$$

There are four alkali categories for irrigation waters (Rechards 1954). SAR-based classifications consist of four intensity levels: low (0–10), medium (11–18), high (18–26), and very high (>26).

The SAR is depicted in a US salinity diagram, with EC representing the salinity hazard and SAR indicating the alkalinity hazard. In the US salinity diagram, the four classes for sodium hazards (S1-Low, S2-Medium, S3-High, and S4-Very High) are positioned on the y-axis, while the four classes for salinity hazards (C1-Low, C2-Medium, C3-High, and C4-Very High) are situated on the x-axis.

4.6.2 Percent Sodium (%Na)

The percentage of Sodium is widely utilized to evaluate the suitability of water quality for irrigation purposes (Wilcox 1955). The excessive Na^+ in irrigation water decreases permeability, which ultimately leads to soil with poor internal drainage by causing Na^+ in water to exchange with Ca^{2+} and Mg^{2+} in the soil. As a result, air and water movement are constrained when it is wet, and these soils typically become hard once dried (Collins and Jenkins 1996; Saleh et al. 1999). The percent of sodium (%Na) in irrigation water is determined, with a maximum allowable sodium level specified at 60% (BIS 1991). The %Na is employed as an index for delineating water suitability for irrigation purposes. This is calculated by applying the equation given by (Wilcox, 1955).

$$\%Na = \frac{Na+K}{(Ca+Mg+Na+K)} \times 100 \quad \text{…. (4.9)}$$

The Wilcox plot is used to discuss the % Na evidence. The Wilcox plot is typically drawn between sodium percentage on y-axis and electrical conductivity on x-axis Wilcox (1955).

4.6.3 Kelly Index (KI)

The ratio of $Na^{+}/(Ca^{2+}+Mg^{2+})$, commonly known as KI, is used to define water for irrigation. Water with a KI exceeding 1 indicates an excessive sodium content, rendering it unsuitable for irrigation purposes. According to the findings of Kelley (1946) and Paliwal (1967), it is established that water possessing KI below 1.0 is considered to be sufficient primarily for the purpose of irrigation. The Kelly index is calculated with following formula.

$$KI = \frac{Na}{(Ca+Mg)} \quad \text{…. (4.10)}$$

4.6.4 Permeability Index (PI)

The permeability of soil varies over time due to the application of irrigated water, leading to consequential effects on agricultural productivity. Therefore, assessing the suitability of water for irrigation purposes based on its Permeability Index (PI) becomes essential. Permeability index is an important measure for assessing bicarbonate and carbonate risks in groundwater. Doneen (1964) classification states that groundwater with a PI value > 75% is in Class I, which means it can be used for irrigation. Water with a PI value between 25% and 75% falls into Class II, indicating suitability for irrigation, while water with a PI value less than 25% is classified as Class III, indicating unsuitability for irrigation.

4.6.5 Magnesium Ratio (MR)

The excess concentration of Magnesium changes the quality of soil and makes it alkaline. The crops yield gets decreased as a result, poor agricultural returns. Magnesium hazard has been calculated by formula.

$$MR = \frac{Mg*100}{(Ca+Mg)} \quad \text{…. (4.11)}$$

4.7 Basic Descriptive statistical Techniques

The degree of linear relationship and the proximity between two independent variables are measured via correlation matrix analysis. Through carrying out an analysis of correlation matrices, it is possible to ascertain the degree of proximity between two variables (Singh et al. 2012; Dheeraj et al. 2020). Multivariate statistics is widely regarded as a highly effective approach for comprehensively evaluating large datasets. This method not only facilitates the identification of significant patterns but also serves as a powerful instrument for interpreting complex data matrices within a specific field. The different variables in the surface and groundwater assessments show trends that are either similar to or different from one another. Therefore, as result, multivariate analysis is an essential technique for identifying the similar as well as dissimilar features to be used in the evaluation of the water samples during the study. Hence, this study has utilised many multivariate approaches like hierarchical cluster analysis, Box plot analysis, Piper plot, and correlation analysis.

4.7.1 Box Plot Diagrams

The box plot serves as a graphic representation of the results obtained from the various parameter concentrations analysed in this study, including cations, anions, physicochemical variables, and heavy metals. The graph often displays the data concerning the minimum, maximum, and mean values of the parameter. The utilisation of statistical techniques proves to be very helpful for performing a comparative analysis of several parameter's concentration. The box plot for the present examination was generated using software programmes, namely Past 4.03 and Origin 8 Pro.

4.7.2 Correlation Analysis

The correlation is a statistical measure of the linear correlation of two variables. The correlation coefficient (r) has been determined between the water samples for both

seasons using a mathematical method in Microsoft Excel 2019 software. This analysis aimed to investigate the correlation between each of the variables of the water samples that were taken into account.

The correlation coefficient is typically calculated using the following formula.

$$r = \frac{n\sum(xy) - \sum(x)(y)}{\sqrt{n\sum(x)2 - (\sum x)2}\sqrt{n\sum(y)2 - (\sum y)2}} \quad \text{..... (4.12)}$$

where n indicates the number of data, r is the correlation coefficient, and x & y denote the variables considered for analysis. Absolutely, the correlation coefficient indeed ranges from -1 to +1. A value of -1 indicates a strong negative correlation, suggesting that as one variable increases, the other tends to decrease. Conversely, a value of +1 indicates a strong positive correlation, implying that as one variable increases, the other also tends to increase. A correlation coefficient of 0 suggests no linear correlation between the variables.

4.7.3 Piper Diagram

Piper plots, also known as trilinear diagrams, are widely recognized as effective methods for visually representing the relative abundance and spatial distribution of significant ions in samples collected in various areas of hydrogeology analysis. The ability to plot multiple data on a single plot makes this particular kind of plot very helpful. The trilinear diagram was introduced by Piper in 1944 as a means of explaining the chemical characteristics of a given water sample (Back, 1961). A piper plot has three parts: first part lies in the lower left known as cations (such as Ca^{2+}, Mg^{2+}, and Na^{+} + K^{+}) while second part in the lower right represents anions (such as Cl^{-}, SO_4^{2-}, CO_3^{2-} + HNO_3^{-}), and third one in the middle called as diamond plot that is a matrix transformation of the two ternary diagrams. Each sample has been standardised so that the relative concentrations are expressed as in percentages (sum of cations and anions equal to 100).

Additionally, it facilitates in the identification of samples collected based on the presence of cations or anions. In this study, the software tool Diagramme was utilised to do the computation of the Piper trilinear diagram.

4.7.4 Gibbs diagram

The Gibbs plot is utilised to ascertain the correlation between lithology, hydrologic processes, and water chemistry within an aquifer. In order to understand how the water and environment interact to show their specific chemistry, such relation must be examined (Egbueri et al. 2019). In the Gibbs plot, the weight proportions for cations and anions, such as $Na^+/ (Na^+ + Ca^{2+})$ and $Cl^-/ (Cl^- + HCO_3^-)$, are shown on the abscissa, while the variation in TDS is shown on the ordinate. A well-organized concept that could be serve as a general approach for a discussion of the various mechanisms that influence global water chemistry (Gibbs, 1970). The evaporation crystallisation, rock dominance and Atmospheric precipitation are the three major processes that govern global groundwater chemistry.

4.7.5 Hierarchical Cluster Analysis

The hierarchical clustering algorithm, also known as hierarchical cluster analysis, groups objects based on their degree of similarity. The endpoint consists of multiple clusters, each of which is unique from the others, and the elements within each cluster tend to share similarities with one another. The dendrogram shows the clustering and gives important information about the connections between the clusters (Shrestha 2007). Ward's method employing square Euclidean distance was utilized for hierarchical cluster analysis (HCA). This method determines the distance between two stations by calculating the square root of the sum of the squares of the differences between their respective values.

$$Dij = \sum_{k=1}^{p} [(Z_{i,k} - Z_{j,k})^2]^{0.5} \quad \text{…….. (4.13)}$$

Where p represents number of variables, $Z_{i,k}$ represents standardized values for variables k at site i and j. Cluster analysis serves the purpose of identifying similarities or dissimilarities among samples, as well as determining the most prominent physicochemical variable responsible for the temporal and spatial variations in water quality. The dendrogram employed in this study was generated using data that was performed using SPSS 22.0 software.

4.8 Mapping of Spatial Distribution Contour map on ArcGIS

4.8.1 Spatial Distribution Map

Geographical Information System (GIS) seems to be an excellent tool to collect, storing, managing, and retrieving a wide range of spatial and non-spatial information. Additionally, it's widely used for spatial analysis and data integration purposes (Gupta and Srivastava et al. 2010). In this study, the spatial analyst component of ArcGIS Software was used to build a spatial distribution map for several water quality parameters. The algorithms, namely IDW (Inverse Distance Weighting), has been used to interpolate a spatial distribution map using ArcGIS software. Both physicochemical and heavy metal concentrations have been studied using this algorithm. The value for each grid node has been determined using the IDW algorithm by looking at adjacent data points that are contained within a user-specified search radius (Burrough and McDonnell 1998). The interpolation method employs each and every data point, as well as the cluster value is determined by averaging the weighted total of the points.

4.9 Heavy Metal Pollution Indices and its Calculation Procedure

The water is incredibly important for all living creatures to survive on earth; hence it is important to periodically assess both its quantity and quality. Heavy metal is defined as any metallic chemical element that possesses a relatively high density and exhibits potential danger or hazard even at low concentrations. The phenomenon of rising demand for portable water can be attributed to two main factors: groundwater scarcity and substantial population growth in different parts of the world. The remote sensing & GIS were used to analyse the distribution of metals in groundwater in this study. The suitability of groundwater for drinking purposes has been assessed using metal pollution indices, namely the HPI, HEI, and MI. These indices have been employed to compare the concentrations of pollutants with the guidelines established by the WHO (2006) and BIS (2012). The HPI and HEI offer an evaluation of overall water quality by factoring in the presence of heavy metals. Conversely, the Metal Index (MI) quantifies how heavy metals affect human health and the overall quality of water used for drinking and domestic purposes. These pollution evaluation indicators are very helpful for finding and analysing trends in water quality in terms of heavy metals present in the water. Fig.4.8 illustrates the flow chart that presents the methods incorporated in this particular objective.

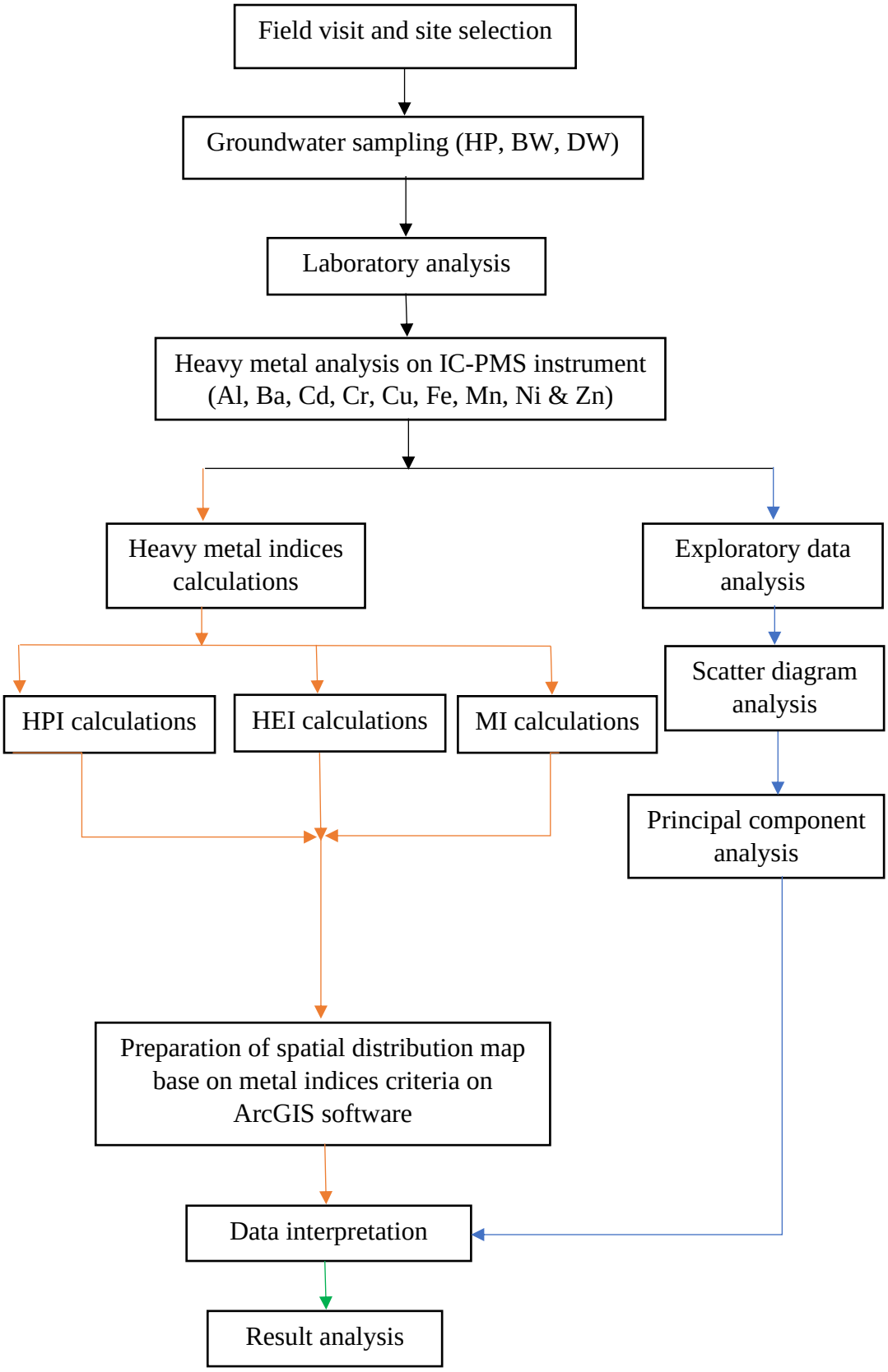

Fig.4.8 Flow chart of methodology adopted for heavy metals indices analysis

4.9.1 Heavy metal pollution index (HPI)

The HPI, an empirical evaluation system, is utilised to evaluate the overall effect of individual metals on water quality. The grading system assigns a numerical score between 0 and 1, which reflects the relative importance of each parameter in determining the overall grade. The critical value for the HPI concerning drinking water is set at 100. Additionally, the unit weightage (W_i) assigned to the relative variables (Al, Ba, Cd, Cr, Cu, Fe, Mn, Pb, Ni, and Zn) used in the computation of the HPI is inversely proportional to their corresponding standard values (S_i) (Prasad and Bose 2001).

The value of HPI is estimated with following steps (Prasad and Bose 2001) given by equation (4.14).

$$HPI = \frac{\sum_{i=1}^{n} WiQi}{\sum_{i=1}^{n} Wi} \qquad \ldots\ldots(4.14)$$

Where W_i – unit weight, n – Number of variables and Q_i – Sub-index of i^{th} parameter calculated by Equation (4.15).

$$\mathrm{Qi} = \sum_{i=1}^{n} \frac{(Mi - Ii)}{(Si - Ii)} \times 100 \qquad \ldots\ldots(4.15)$$

Where M_i – Analytical data, I_i – Ideal value and S_i – Standard value

(-) – represents a numerical difference between value, not used for algebraic importance.

4.9.2 Heavy metal evaluation index (HEI)

According to (Edet and Offiong 2002), it is suggested that the HEI index serves as a substitute for the HPI index in determining the heavy metals content present in groundwater samples. The heavy metal pollution classes found in water samples are analysed using this HEI (Prasanna et al. 2012). HEI is calculated as given by Equation (4.16).

$$HEI = \sum_{i=1}^{n} \frac{Mi}{Si} \qquad \ldots\ldots\ldots(4.16)$$

Where M_i – Monitored value, S_i – Standard permissible values and n – Number of parameters

4.9. 3 Metal Index (MI)

The MI quantifies the extent to which heavy metals negatively impact human health, along with the overall quality of water utilised for drinking, domestic activities, and other applications (Balakrishnan and Ramu, 2016). A higher concentration of metals in comparison to the MAC value indicates poor water quality. A value of MI greater than 1 serves as a threshold indicating warnings (Ameh and Akpah, 2011). This index is calculated by the equation (4.17) below:

$$MI = \sum_{I=1}^{N} \frac{Ci}{(MAC)i} \qquad \ldots\ldots\ldots\ldots\ldots\ldots(4.17)$$

where C seems to be the monitored value, MAC is the maximum permissible values, and MI stands for the metal index. A higher MI score implies that the water quality is poorer, posing a greater risk to human health as well as other activities. The six categories that comprise the MI index are presented in Table 5.7, which can be found in the next chapter.

4.10 Exploratory data analysis

4.10.1 Principal component analysis

In this study, PCA analysis was performed for groundwater using the Origin Pro 2017 software programme which is designed for quantitatively analysing complex data sets. The PCs were extracted using PCA analysis with an eigenvalue cut-off of 1 (Wilk-Wo´zniak et al. 2014). The loading values for each variable within the primary principal components were computed using the varimax normalized rotation technique (Ali et al. 2014; Yasin and Usman, 2017). The basic processes involved in analysing and assessing

the quality of the water are interpreted using a set of associated variables for each PC with positive and negative loading values. The variance and cumulative variance (%) were both attained for all PCs. In statistical analysis, variance is used to determine dispersion, which reveals the distances of data values from their mean (Wu et al. 2016). The PC loading values were categorised into three distinct groups: strong, moderate, and weak. The strong category included values > 0.75, the moderate category included values between 0.75 and 0.5, and the weak category included values between 0.5 and 0.4. The loading values of variables below 0.40 were ignored due to their less importance. These principal components were used to interpret both natural and anthropogenic processes involving in water systems. The values of all variables from the laboratory analysis were normalized by dividing them by their respective maximum values (Vongdala et al. 2019). PCA analysis was performed using these normalized value sets.

4.10.2 Scatter Plot

A scatter plot is a type of data visualization that displays individual data points as dots on a two-dimensional coordinate system. It is used to show the relationship or correlation between two continuous variables. Each data point in the scatter plot represents a unique observation with values for both variables. The scatter plots are often used as a method of identifying and examining correlational relationships. The dots on a scatter plot indicate not only the values of individual data points, but also reveal patterns when the data is seen as a whole. Overall, it is a valuable tool for exploratory data analysis, providing a visual representation of the relationships and patterns between two variables in groundwater samples.

4.11 Land use and Land cover (LULC) Change Detection

4.11.1 LULC Classification

The LULC are two terms, with land use referring to anthropogenic activities on the land that are typically not visible in imagery and land cover referring to vegetation and man-made constructions on the surface of the land (Lo 1986; Burley 1961). The land is a natural resource that mankind employs to generate material wealth, and exploiting these resources implies changes in land usage and land cover (Prakash and Gupta, 1998). The mining operation is done with large machinery and advanced technology. Consequently, these operations can change the land cover pattern economically or hydrologically in that area (Singh 2007; Tiwari et al. 2015; Turner and Ruscher, 1988). Furthermore, mining operations give rise to limited alterations, such as changes in topography (Manna and Maiti, 2014), drainage pattern (Akiwumi and Butler 2008; Khan and Javed 2012), landscape change and degradation (Bajocco et al. 2012), soil erosion and degradation (Singh et al. 1997), and overall environmental changes (Dhar et al. 1991). It is critical to investigate the impact that mining has on the evolution of LULC in order to minimize the negative consequences that mining has on the environment, land management, and decision-making (Bocco and colleagues, 2001).

4.11.2 LULC Technique

The current study has been applied the unsupervised classification for observing the LULC changes with the help of the ERDAS 15 software. In this approach, the classification process does not rely on predefined training samples or labelled data. Instead, the algorithm groups pixels with similar spectral characteristics into clusters or classes based on the similarity of their spectral values. The unsupervised classification technique was used to create six distinct classes of LULC in accordance with the broad spatial specifications of the area. The categorised images have undergone a cleaning

process to enhance the accuracy of the results and minimize instances of misclassification (Harris and Ventura, 1995). The land use and land cover (LULC) data for the years 2001, 2011, and 2021 have been compiled, with a time interval of 10 years. The conventional false colour composite (FCC) technique was used on each image before it was subjected to pre-processing in preparation for LULC analysis. Fig.4.9 illustrates the flow chart that presents the methods incorporated in this particular objective.

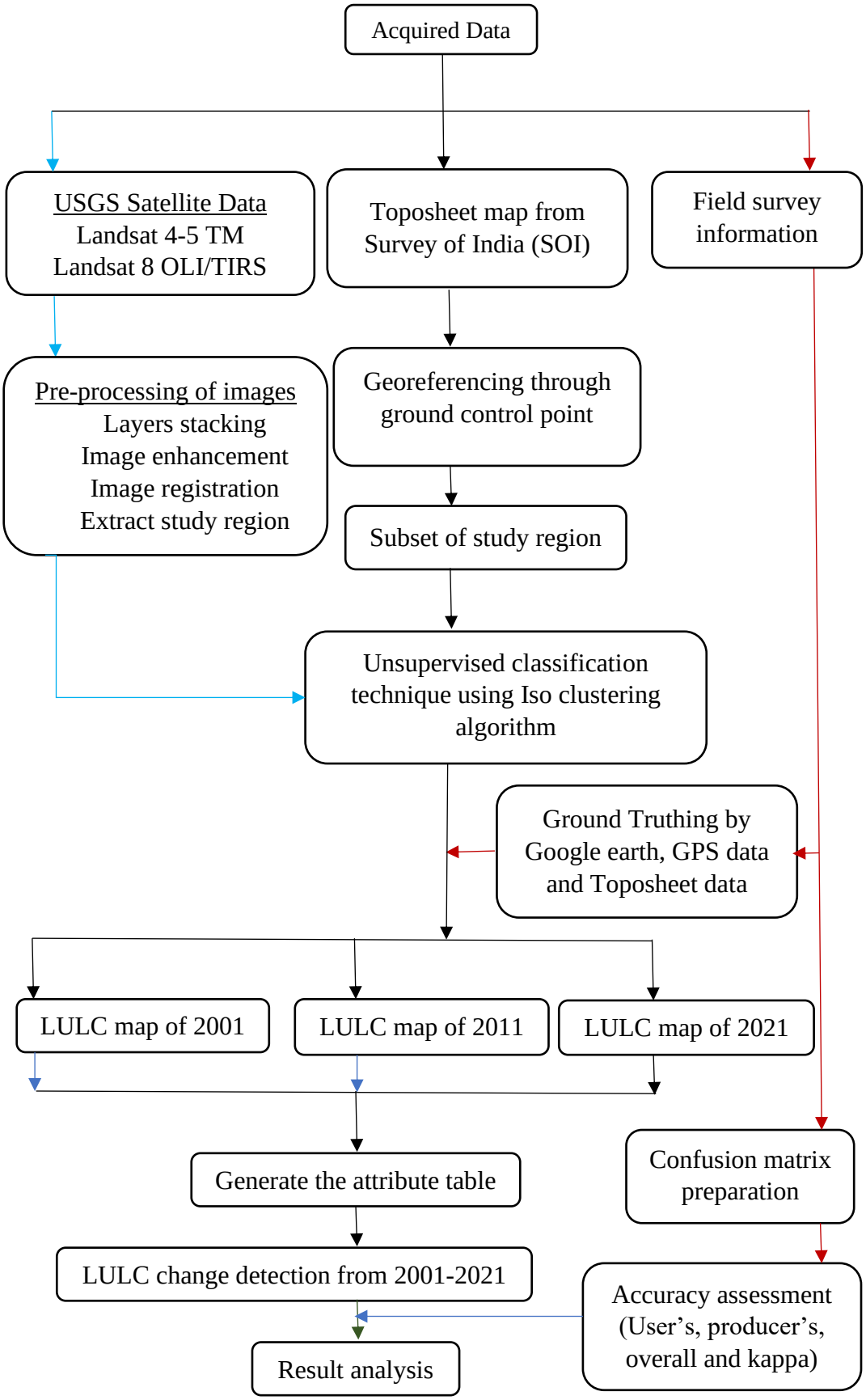

Fig.4.9 Flow chart showing the methodology of LULC change analysis

4.11.3 Acquisition of Data Types

To ascertain LULC change, a minimum of two time-period satellite images are necessary. Additionally, multi-dated satellite imagery is essential to accurately quantify changes in LULC over time. The satellite datasets were obtained from the USGS Earth Explorer website (https://earthexplorer.usgs.gov/). The projection reference system employed for all downloaded imagery from the USGS website was WG 1984 UTM Zone 44N. The present study utilised three satellite data sets, namely Landsat 4-5 data for the years 2001 and 2011, and Landsat 8-9 OLI/TIRS data for the year 2021 (Table.4.3), in order to identify and analyse LULC changes within the selected study region. The resolution of the data that was retrieved from Landsat 4-5 and Landsat 8-9 was respectively 30 m.

Table.4.3 USGS and BHUVAN Satellites data along with their specifications

Specifications	**Landsat 4-5 Data**	**Landsat 4-5 Data**	**Landsat 8-9 Data**
Acquisition Date	09 February 2001	05 February 2011	31 January 2021
Data types	Landsat 5 TM	Landsat 5 TM	Landsat 8 OLI/TIRS
Spatial resolutions	30 m	30 m	30 m
Number of Band	7	7	8
Land cloud cover	0	0	0
Path & Row/Orbit	142 & 045	142 & 45	142 & 45
Source	Earth Explorer	Earth Explorer	Earth Explorer

4.11.4 Shapefile of study area

At least two time period satellite imagery are required to determine LULC change, and multi-dated satellite images are required to quantify land use/land cover change. The satellite data sets were obtained from the USGS Earth Explorer website (https://earthexplorer.usgs.gov/). The projection reference system employed for all downloaded imagery from the USGS website was WG 1984 UTM Zone 44N. The present study utilised three satellite data sets, namely Landsat 4-5 data for the years 2001 and 2011, and Landsat 8-9 OLI/TIRS data for the year 2021 (Table.4.4), in order to identify

and analyse LULC changes within the selected study region. The resolution of the data that was retrieved from Landsat 4-5 and Landsat 8-9 was respectively 30 m.

4.11.5 Image Processing

In this study, a radiometric correction procedure was implemented prior to conducting change detection analysis. This correction aimed to mitigate any potential environmental attenuation and radiometric inaccuracies that could arise from the utilisation of various sensor systems. The satellite imagery was georeferenced using an image-to-image overlapping technique (Demirel et al. 2011). The USGS Sentinel-2A satellite imagery was used as reference data for the radiometric and geometric processing. For the georeferencing process, around 8 ground control points were chosen randomly from clearly visible areas such as road crossings and river junctions. The quality of the imagery can be increased using the ArcGIS software.

4.11.6 Change Detections analysis

The post-classification strategy has been identified as the most commonly utilised method in change detection analysis, as employed in this study (Jensen et al. 1993). The significant changes were observed in the prepared LULC map from 2001 to 2021, which includes net changes, change patterns, percent changes, and rate of change of LULC from 2001 to 2011, and 2011 to 2021.

4.11.6.1 Percent change calculation

In order to calculate the percentage change in LULC, both the initial and final LULC coverage have been utilised, and an appropriate formula is provided by equation 4.18.

$$\text{Change percentage } (\%) = \left[\frac{(\text{the area of present LULC} - \text{the area of previous LULC})}{\text{The area of previous LULC}} \times 100\right] \ldots (4.18)$$

4.11.6.2 Rate of LULC change calculation

The rate of LULC change for the period of 2001-2011 and 2011-2021 has been calculated using prepared LULC map. The following formula (4.19) has been used to calculate the rate of LULC change.

$$R\Delta\left(\frac{Km^{2}}{year}\right) = \left[\frac{\text{area of present LULC in Km2 (X)} - \text{area of previous LULCin Km2(Y)}}{\text{time interval between X and Y (in year)}}\right] \quad \ldots\ldots (4.19)$$

Where RΔ= Rate of LULC change

4.12. Post LULC Classification Change Estimation

A post classification of change detection map has been created for the period of time from 2001 to 2021. This method is essential for determining the rate, nature, and location of changes based on the independent categorization, environmental factors, and acquired satellite imagery (Almutairi and Warner, 2010). In addition, this detection method typically depicts the quantity and dispersion of every changing land cover class across the different multi-temporal imagery (Ojaghi et al. 2017). However, this approach neglects the temporal correction between the equivalent pixels across two distinct time periods (Wu et al. 2017). In this study, the LULC image classification for three different time periods was evaluated initially, and then overlaying prepared images was done in ArcGIS for showing spatial changes of LULC from 2001 to 2021. Later, using this methodology, it was possible to characterise changes in the specified area based on "From-to" statistics from six different classes for three distinct year's maps.

4.13 Accuracy assessment

The classification of LULC may be susceptible to a limited number of errors during preparation, thus the prepared LULC map requires some accuracy evaluation testing to assess the error status using an accepted statistical technique by the user. An accuracy evaluation index, such as user, producer, overall, and Kappa coefficient accuracy, is

frequently necessary for clear classification and the measurement of descriptive precision when working with LULC maps (Olofsson et al. 2013; Lu and Weng, 2005). According to a researcher, the term "accuracy" refers to the evaluation of the "correctness" of a prepared LULC classification, which can be assessed by using of an error matrix (Foody 2002). The Kappa coefficient technique and the confusion matrix are both suitable methods for carrying out an accuracy evaluation (Story and Congalton, 1986; Lea and Curtis, 2010). It was helpful to use stratified random sample procedures for the accuracy testing. The overall, user, producer, and the Kappa coefficient accuracy have each been computed using the formula that is described in the following set of equations: (4.20), (4.21), (4.22), and (4.23) respectively.

$$\text{Overall Accuracy} = \frac{\text{Total No.of Correctly classified pixel (Diagonal)}}{\text{Total No.of Reference Pixel}} * 100 \quad .. (4.20)$$

$$\text{User accuracy} = \frac{\text{Number of correctely classified pixel in each category} \times 100}{\text{Total No.of classified pixel in that category (the row total)}} \quad .. (4.21)$$

$$\text{Producer accuracy} = \frac{\text{Number of correctely classified pixel in each category} \times 100}{\text{Total No.of reference pixel in that category (the column total)}} \quad .. (4.22)$$

$$\text{Kappa Coefficient } K = \frac{(\text{TS} \times \text{TCS}) - \sum(colomn\ total \times row\ total)}{TS^2 - \sum(colomn\ total - row\ total)} * 100 \quad .. (4.23)$$

Where TS-Total sample, TCS-Total corrected samples

4.14 Retrieving LST from Landsat data

The LST refers to the radiative skin temperature of the land, which is derived from solar radiation. It is measured in the direction of the distant sensor. The mono-window technique was employed in this study to extract LST from multi-temporal Landsat satellite data (Qin et al. 2001; Guha et al. 2017, 2018, 2019). In order to determine an estimation of LST using the mono-window approach, it is essential to consider three key parameters such as ground emissivity, atmospheric transmittance, and effective mean atmospheric temperature. The method used for determining LST involves the use of mathematically derived equations. The LST procedures were carried out in the Raster Calculator of ArcGIS 10.8 software. The flow chart illustrating the process that is incorporated in achieving this objective can be found in Fig.4.10. The linear regression model is employed to evaluate the regression coefficient between the dependent variable LST and the independent variable NDVI values.

4.14.1 Scan Line Error for Landsat 7 ETM+ data

The Landsat 7 Enhanced Thematic Mapper Plus (ETM+) sensor experienced an error in its scan-line corrector (SLC) in 2003, leading to an improper scanning of roughly 22% of the pixels in each scene. The SLC failure has severely constrained the scientific applications of ETM+ data. These scan line errors are easily corrected using the Landsat Toolbox function in ArcGIS software prior to the utilisation of the tiles.

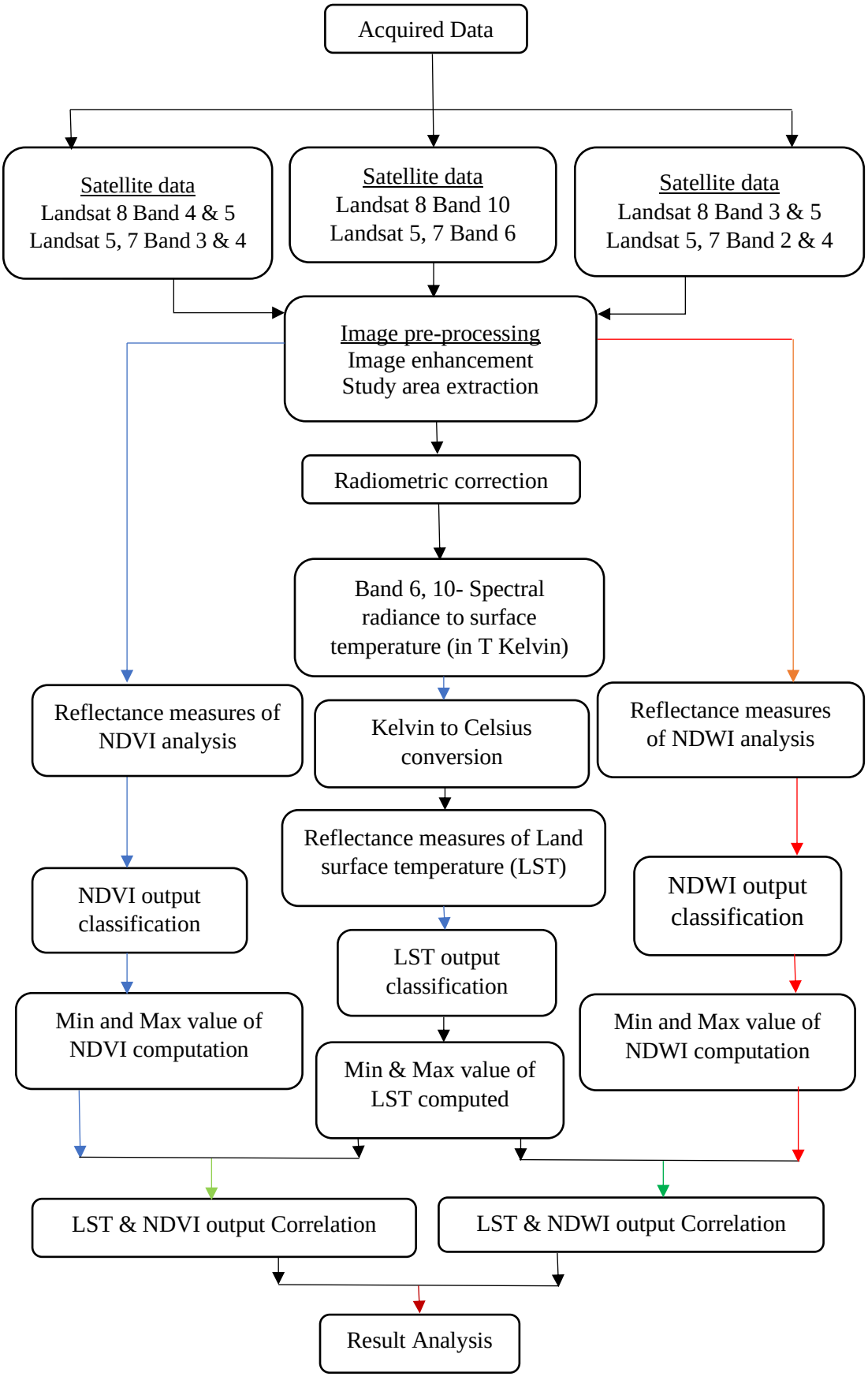

Fig.4.10 Flow chart showing correlational methodology of LST with NDVI and NDWI

4.14.2 Data Type

The estimation of LST, NDVI, and NDWI in the study region for the month of April was conducted using a time series of Landsat satellite data covering the years 2001, 2006, 2011, 2016, and 2021. This study exclusively utilised three series of Landsat data, namely Landsat 4-5 TM, Landsat 7 ETM+, and Landsat 8-9 OLI /TIRS. The selected Landsat images were employed for image geo-referencing via image-to-image co-registration. The selected Landsat satellite data were cloud-free except 2006. The different satellite data were being used for the LST, NDVI and NDWI calculation. Additionally, a description of the entire data set, along with its characteristics, is tabulated in Table 4.4.

Table.4.4 Landsat satellite data along with their characteristics

S No.	Date	Data Type	Spatial Resolution	Path & Row	Land cloud cover
1.	22 April 2001	Landsat 7 TM ETM+ C2 L2	30×30	142 & 045	00
2.	28 April 2006	Landsat 4-5 TM C2 L2	30×30	142 & 044	02
3.	26 April 2011	Landsat 4-5 TM C2 L2	30×30	142 & 045	00
4.	23 April 2016	Landsat 8-9 OLI /TIRS C2 L2	30×30	142 & 045	00
5.	05 April 2021	Landsat 8-9 OLI /TIRS C2 L2	30×30	142 & 045	00

4.14.3 Calculation Steps for LST for Landsat 5 & 7 data

The surface temperature is a general word that refers to the average temperature of all accessible objects, including the actual ground surface. The LST is maintained by outgoing terrestrial infrared radiation and long wave radiation from the sun and other sources (Li et al. 2013). The measurement serves as a reliable indicator of the energy balance occurring at the surface of the Earth. It represents the temperature of the ground, soil, rocks, vegetation, and other natural surfaces.

Following are a few procedures for calculating LST using a mathematical formula for a specific domain pixel. All of the LST stages were carried out using the raster calculator in the Arc toolbox of the ArcGIS 10.8 software.

Calculation of spectral Radiance-

$$L_\lambda = \left(\frac{\text{LMAX}\,\lambda - \text{LMIN}\,\lambda}{\text{QCALMAX} - \text{QCALMIN}}\right) \times (QCAL - QCALMIN) + LMIN\lambda \quad \ldots\ldots\ldots (4.22)$$

Where L_λ – Spectral radiance

QCAL - quantized calibrated pixel value in DN,

$LMAX_\lambda$ - Spectral radiance scaled to $QCAL_{MAX}$,

LMINλ - Spectral radiance scaled to $QCAL_{MIN}$,

$QCAL_{MIN}$ – Minimum quantized calibrated pixel value in DN -1 and

$QCAL_{MAX}$ – Minimum quantized calibrated pixel value in DN -255

$$T = \left(\frac{K2}{\ln\left(\frac{K2}{L\lambda}+1\right)} - 273.15\right) \quad \ldots\ldots\ldots\ldots (4.23)$$

Where T – Satellite temperature (°C),

K1 & K2= Calibration constant as listed in Table: 5 and

L_λ – Spectral radiance

4.14.3 Calculation procedure for LST for Landsat 8 OLI/TIRS data

Top of atmospheric (L_λ) spectral radiance

$$L_\lambda = M_L \times Q_{Cal} + A_L \quad \ldots. (4.24)$$

Where M_L - Band-specific multiplicative rescaling factor,

Q_{Cal} – Correspond to Band 10 for Landsat 8 data and

A_L - Band-specific additive rescaling factor.

Brightness temperature conversion from top of atmospheric

$$BT = \frac{K2}{\ln\left(\frac{K1}{L\lambda}+1\right)} - 273.15 \quad \text{……. (4.25)}$$

Where K1 & K2= Calibration constant **(**$Wm^{-2}\,sr^{-1}\,mm^{-1}$) as listed in Table.4.5.

Lλ – Top of atmospheric and to get the results in Celsius, the radiant temperature is adjusted by adding the absolute zero (approx. -273.15°C) ($Wm^{-2}\,sr^{-1}\,mm^{-1}$)

NDVI calculation-

$$NDVI = \frac{(Band\ 5 - Band\ 4)}{(Band\ 5 + Band\ 4)} \quad \text{……. (4.26)}$$

(P_v) Proportion of vegetation calculation-

$$P_v = \left(\frac{((NDVI - NDVImin)}{(NDVImax - NDVImin)}\right)^2 \quad \text{……. (4.27)}$$

Emissivity (ε) Calculation

$$(\varepsilon) = 0.0004 \times P_v + 0.986 \text{ (correction value)} \quad \text{…… (4.28)}$$

Land surface temperature calculation-

$$LST = \frac{BT}{\left(1+\left(0.00115*\frac{BT}{1.4388}*\ln(\varepsilon)\right)\right)} \quad \text{…… (4.29)}$$

Where BT - Brightness temperature and (ε) – Emissivity

Table.4.5 Calibration constant for thermal bands of Landsat data

Sensors	K1	K2
Landsat 5	607.76	1260.56
Landsat 7	666.09	1282.71
Landsat 8	774.8853	1321.0789

4.15 NDVI Estimation

The Normalized Difference Vegetation Index (NDVI) has commonly been utilised as an accurate representation of vegetation coverage on the Earth's surface (Purevdorj et al.

1998; Ke et al. 2015). It measures the difference between near-infrared and red light to quantify vegetation. It is a measure of vegetation proportion calculated using variations in the NIR region of electromagnetic spectrum (EMS). These are reflected by green vegetation in addition to the red portion of the spectrum that is absorbed by vegetation. The calculation of the NDVI for Landsat 5 and Landsat 7 data involves the use of band 3 and band 4, as indicated in equation (4.30). Conversely, for Landsat 8 data, the calculation of NDVI involves the utilisation of band 4 and band 5, as shown in equation (4.31).

Calculation Formula for NDVI of Landsat 5 & 7 data

$$NDVI = \frac{(Band\ 4 - Band\ 3)}{(Band\ 4 + Band\ 3)} \quad .. (4.30)$$

Calculation Formula for NDVI of Landsat 8 data

$$NDVI = \frac{(Band\ 5 - Band\ 4)}{(Band\ 5 + Band\ 4)} \quad .. (4.31)$$

Vegetation strongly reflects near-infrared light and absorbs red light due to the presence of chlorophyll. NDVI takes advantage of these characteristics to distinguish between healthy and non-healthy vegetation and quantify vegetation density. unhealthy or sparse vegetation reflects more visible light and less near infrared, while bare sands reflect a small amount of both the red and infrared parts of the EMS and finally, there will be zero outcomes (Lillsand et al. 2009).

4.16 Correlation between LST and NDVI

The surface radiation temperature as well as NDVI have the opposite tendency. The alterations in NDVI have the potential to induce modifications in LST because to the high sensitivity of NDVI towards variations. Although not clearly consistently, the correlation between NDVI and LST changed from season to season (Malik et al. 2019).

A statistical study was conducted on the ArcGIS platform version 10.8 to ascertain the correlation between the LST and NDVI parameters. Logistic regression is utilised to estimate the numerical comparison of input factors that lead to final outcomes. To calculate the regression coefficient for this study, the LST was used as the dependent variable and the NDVI as the independent variable.

4.17 NDWI Estimation

The Normalized Difference Water Index (NDWI) is a remote sensing index used to monitor and quantify the presence of water in vegetation and soil. It is particularly useful for analysing water bodies and detecting changes in water content over time. NDWI is derived from the reflectance measurements obtained from satellite or aerial imagery, typically in the visible and near-infrared (NIR) spectral bands.

The green and near-infrared (NIR) bands are used to calculate NDWI. The bands 2 and band 4 are used as the green band and the NIR band, for TM and ETM+ data as shown in equation (4.32). whereas, the band 3 and band 5 are employed as a green and NIR bands for OLI/TIRS data as shown in equation (4.33) as listed in Table.4.5. NDWI values typically range from -1 to 1, with higher positive values indicating a higher presence of water, while lower or negative values indicate the absence of water. A negative NDWI score is commonly associated with built-up regions and unvegetated areas without water bodies, while a positive NDWI number typically signifies places with both vegetation and water (McFeeters 1996, 2013).

Calculation Formula for NDWI of Landsat 5 & 7 data

$$\text{NDWI} = \frac{(\text{Band 2} - \text{Band 4})}{(\text{Band 2} + \text{Band 4})} \quad \text{.. (4.32)}$$

Calculation Formula for NDWI of Landsat 8 data

$$NDWI = \frac{(Band\ 3 - Band\ 5)}{(Band\ 3 + Band\ 5)} \quad .. (4.33)$$

4.18 Correlation between LST and NDWI

It's important to note that while LST and NDWI are related in the context of water content and temperature, they are not directly dependent on each other. Other factors, such as land cover, vegetation type, and atmospheric conditions, can also influence both indices. The LST and NDWI have the opposite tendency. The variable of concern, LST, was considered the dependent variable, while the NDWI parameter was treated as the independent variable (Guha et al. 2020). The measurement of the LST and NDWI parameters was conducted through the utilisation of correlation coefficients (R^2) obtained from linear regression analysis.