CHAPTER - II
DATA SOURCES AND METHODOLOGY

2.1 Data sources

The present study has utilized different types of data sets, both primary and secondary, which were generated, collected and collated for analysis using GIS. The data sets used for analysis were first made compatible to GIS and further analysis has been carried out in GIS. The following sets of data have been used for the present study.

(a) Survey of India topographic map No. 63L/12 on 1:50,000 scale (surveyed in 1969 – 1970) was obtained from Survey of India, Dehradun. Topographic map was utilized to get acquainted with location of different features present in the study area.

(b) The following data of Indian Remote Sensing Satellite (IRS) was procured from National Remote Sensing Centre (NRSC), Hyderabad.

(i) Standard Geocoded False colour composite (FCC) of Indian Remote Sensing Satellite (IRS – 1B) of Linear Imaging Self scanning (LISS-II) falling in sub-scene A1 (Path 23-Row 51) band combination 234 of 4th May, 1993 and 8th May, 2001 having a spatial resolution of 36.25 meter (Figure 2.1& 2.2).

(ii) Standard Geocoded False colour composite (FCC) and digital data of Indian Remote Sensing Satellite (IRS – P6 Resourcesat) of Linear Image Self scanning (LISS-III) falling in sub-scene A2 (Path 102 - Row 55) band combination 234 of 4th May, 2010 having a spatial resolution of 23.5 meter (Figure 2.3).

(c) Meteorological data (rainfall and temperature) for Singrauli meteorological station for the period 1978 – 2010 were obtained from India Meteorological Department (IMD), Pune. The temperature data corresponding to 1978 – 2003 only as 2004 – 2010 was not available at India Meteorological Department (IMD), however rainfall data is from 1978 – 2010.
Figure 2.1: IRS-IB LISS II FCC (1993) of the study area
Figure 2.2: IRS-IB LISS II FCC (2001) of the study area
Figure 2.3: IRS-P6 LISS III FCC (2010) of the study area
(d) Advance Spaceborne Thermal Emission and Reflection Radiometer (ASTER) data downloaded from the website (http://www.gdem.aster.ersdac.or.jp/search.jsp) having 30 meter resolution within 82°15’ to 82°45’ N latitude and 24°00’ to 24°15’ E longitude which was used for generation of Digital Elevation Model and consequently used for slope map preparation.

(e) Secondary data/information were collected and utilized to supplement the present study. Relevant information was extracted from the following sources:

(i) District statistical hand book, 2007 obtained from the Collectrate Office, Sidhi district, M.P.


(iii) Coal production, overburden and plantation data from 1986 to 2010 and annual reports for the years 2005, 2007, 2008 and 2009 were obtained from Northern Coalfield Ltd (NCL), Singrauli and other relevant data on its web site (www.ncl.nic.in), (www.coalindia.nic.in).

(iv) District map of Singrauli with tehsil boundaries, from District Collectrate Office, Singrauli M.P.

(v) Published research papers, technical reports, special volumes and memoirs of Geological Society of India, and information from other government and non-government sources were consulted for the present study. Research papers in journals namely, International Journal of Remote Sensing, Journal of Indian Society of Remote Sensing, Journal of Geoinformatics, Mausam, Journal of Geological Society of India, Environmental Geology, Journal of Earth science system, Journal of Environmental science and Engineering, International Journal of Applied Earth Observation and Asian Journal of Geoinformatics were also consulted as part of review of literature. Papers and articles relevant to the topic of research were also downloaded from various websites using the Google search engine.
(f) Data collected from the field in the form of ground truth verification and field photographs of various land use/land cover types were also incorporated in analysis and writing the thesis.

(g) The physico-chemical analysis of water samples (surface and underground) collected from the study area was carried out in the Geo-chemistry lab of the Department of Geology. A.M.U, to determine major and minor element concentration in the samples to ascertain water quality.

(h) Various software were utilized for different data input, data generation, editing, analysis, Geo-coding, and output. Softwares which were used in the present study are Arc View GIS 3.2, Arc GIS 10, ERDAS IMAGINE 9.2 for digitizing, analysis, data input, editing, generation of thematic maps and digital image processing. The open source softwares utilized in the present study are SAGA 2.0 version (http://www.saga-gis.uni-goettingen.de/html.index.hph) for generation of elevation map and slope map from ASTER data, Georeferencing software (http://www.mapwindow.org/download.hph?show_details=17) used to rectify the secondary data with respect to toposheet of the study area, Geotrans 2.3 version (http://earth-info.nga.mil/gandg/geotrans/) used to convert the latitude and longitude values into UTM coordinates. R software has also been used for analyzing temperature data. The infrastructure facilities available at Remote Sensing and GIS lab of the Department of Geology, Aligarh Muslim University, were availed during the course of the study.

2.2 Methodology:

A systematic approach involving multiple steps was followed to carry out the present work. It includes preparation of base map, generation of thematic maps using satellite data, ground truthing, laboratory analysis of water, land use/land cover analysis, land use/land cover change detection etc. Figure 2.4 shows major steps followed for the present study.
Figure 2.4: Methodology followed for the present study
2.2.1 Preparation of Base Map:

Base map was prepared on a tracing sheet (Mylar) using Survey of India (SOI) topographic maps on 1:50,000 scale. The basic information such as latitudes, longitudes, major roads, railway line, important towns/city, major rivers, streams and water bodies were plotted on the base map. The base map was then superimposed on the IRS, geocoded FCC so that the important features on the base map overlay the features on the satellite data. Base map was used as a key map for subsequent thematic mapping based on remote sensing data.

2.2.2 Preparation of Drainage Map:

Drainage network was initially derived from SOI topographic maps on a tracing sheet (Mylar) and subsequently updated using IRS-P6 LISS III FCC data of 4th May 2010. The drainage map was scanned and geo-referenced with respect to the topographic maps and was subsequently imported in Arc-View 3.2 GIS software for digitization and creation of drainage layer. Drainage was created using line coverage, assigning different unique ids for various stream orders during digitization. The drainage map was edited and digitization errors were removed using edit module in the Arc-View 3.2.

2.2.3 Georeferencing:

The thematic maps generated from visual interpretation of satellite data were geo-referenced using ground control points (GCPs) obtained from Survey of India (SOI) topographic maps. The maps were projected in Universal Transverse Mercator (UTM) projection, taking World Geodetic System (WGS84) as the datum using Georeferencing and Geotrans 2 software. The study area lies in UTM zone 44 North. For georeferencing, the latitudes and longitudes of reference map were converted into X, Y co-ordinates in the software Geotrans 2, and the corresponding X-Y co-ordinates were transferred on the thematic map.
2.2.4 Land use/land cover mapping and change detection:

Land use/land cover mapping was carried out through standard visual image interpretation method based on photographic recognition elements such as tone, texture, size, shape, pattern, association and field knowledge. Land use/land cover mapping was carried out through multi-temporal time series data of IRS 1B LISS II of 4th May 1993, IRS 1B LISS II of 8th May 2001 and IRS P6 LISS III of 4th May 2010. The information about road, railways, main towns and water bodies from the base map was transferred onto the tracing film (Mylar) which was superimposed on IRS FCCs of 1993, 2001 and 2010 to delineate various land use/land cover categories. Interpretation of satellite data led to the identification and delineation of fifteen land use land cover categories namely dense forest, open forest, open scrub, plantation, cultivated land, uncultivated land, mine pit, overburden dumps, waste land, rocky/barren area, settlement, ash pond, water bodies, thermal power plant and dry river. Ground truth verification was carried out three times in the month of May 2009, 2010, and 2011, to check the veracity of the remote sensing data and spectral signature of land cover features. Interpretation key were developed from field inputs and were incorporated wherever required in the final mapping. Errors in the pre field interpretation were corrected and accordingly the land use/land cover maps were finalized after necessary changes in spatial spread of land use/land cover categories.

Land use/land cover details on the tracing sheet (Mylar) were scanned, georeferenced, digitized and imported to Arc view 3.2 GIS software for digitization and generation of area statistics by using different modules. Every land use land cover category was digitized as polygon coverage and a unique id was assigned to each one of them. Area under each land use/land cover category was computed in square kilometers as well as in percentage by using calculate area module in X tools in Arc view 3.2 version. Change in area under each land use/land cover category from 1993 to 2001, 2001 to 2010, and 1993 to 2010 was analyzed in order to find out the changes in land cover in a time series domain, as a result of expansion of coal mining operations.
2.2.5 Digital Image Processing:

Digital image processing involves manipulation and interpretation of digital images with the aid of a computer (Lillisand et al., 2004). The basic purpose of image processing is to rectify the data of its errors due to technical or systematic errors or external errors and improve the visual quality of imagery.

While the pre-processing is generally done by the agency which supplies satellite data i.e. NRSC, before sending the data to the user. Image enhancement and information extraction are normally done by the user depending upon his requirement. For the present study a limited analysis to support the results obtained by visual interpretation of the data has been attempted using ERDAS Imagine 8.6.

Several image processing techniques were applied to IRS-P6 LISS III digital data in ERDAS IMAGINE 8.6 which include supervised classification, unsupervised classification, spatial filtering, band rationing and Normalized difference vegetation index (NDVI).

a) **Supervised Classification:**

The initial step involves defining image properties as pixel DN values that represent a group of information of training classes. The module classifier, signature editor and supervised classification were used to generate supervised classified map (Lillesand et al., 2004; Campbell and Wynne, 2011). In case of supervised classification 42 known signatures were selected from the known land cover types which fall under 15 land use/land cover categories in the image and were specified to the computer as training areas, which were saved as .sig format. Statistical measures were generated for the training areas and input to the classifier, which then determines other areas in the image that have similar spectral characteristics. Finally in supervised classification both image and signature editor (.sig format) was taken as an input for generation of land use/land cover map.
b) *Unsupervised Classification:*

Unsupervised classification uses an automatic clustering algorithm that analyzes the “unknown” pixels in the database and divides them into a number of spectrally distinct classes based upon their natural grouping (clusters) in n-spectral dimensions (Lillesand et al., 2004; Campbell and Wynne, 2011). Unsupervised classification was generated using the module classifier, unsupervised classification. After the classification process is complete the classes with same spectral reflectance characteristics were merged and classified into 9 land use/land cover classes like forest, cultivated land, Uncultivated land, mining pits, overburden dumps, wasteland, settlement, ash pond, water body.

c) *Spatial Filtering:*

Filtering was applied to the IRS-P6 data using the module raster, filtering, convolution filtering. Edge enhancement filters were used to emphasize the visual transition between regions of contrast brightness. A typical edge enhancement 5/5 algorithm moved through the image and new digital value is calculated using original value on the local average five adjacent pixels.

d) *Band Ratioing:*

Band rationing was carried out using the module interpreter, utilities and operators for generating different combination of bands. First combination was made of i.e. green and red, second red and green, third between red and near infra red and fourth between green and near infra red.

e) *Normalized difference vegetation index (NDVI):*

Normalized difference vegetation index was developed using the module interpreter, utilities and operators for NDVI for 4/3 bands. Normalized difference vegetation index was the successful vegetation index based on band ratioing. The NDVI is a numerical indicator that uses the visible and near infrared bands to analyze
whether the target being observed contain live green vegetation or not. NDVI image has been generated to delineate land use categories that have higher concentration of green vegetation like dense forest, open forest, open scrub and plantation.

2.2.6 Digital Elevation Model and Slope:

The Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) is an advanced multispectral Japanese sensor which is one of the five remote sensing devices on board the Terra satellite launched into earth orbit by NASA in December, 1999. ASTER provides high-resolution images of the planet Earth in 15 different bands of the electromagnetic spectrum, ranging from visible – near infrared (VNIR), Short wave- near infrared (SWIR) to thermal infrared (TIR) with global resolution of 30 meter. The digital elevation data for the study area was downloaded from ERSDAC by visiting the link http://www.gdem.aster.ersdac.or.jp/search.jsp. The ASTER GDEM is in 1° X 1° tiles in GeoTIFF format with geographic lat/long coordinates and a 1 arc-second (30 m) grid of elevation postings and is referenced to the WGS84/EGM96 geoid. The horizontal and vertical accuracy has been reported to be less than 30 and 20 meters with 95% accuracy respectively. The 1° X 1° tile in GeoTIFF format was imported in SAGA software and cropped by shape file to the limits of the study area. Since the tiles are available in geodetic coordinate system, the cropped tile was transformed into UTM projection (zone 44N). The study area boundary was superimposed on the imported ASTER DEM and a color coded Digital Elevation Model was generated.

Taking the Digital Elevation Model (DEM) as input for creation of slope map of the area, a command was given in SAGA software through module, terrain analysis and standard terrain analysis, an automatic slope map was created which shows slope ranges from 0° to 29°. The slope map was then imported into the Arc view 3.2 GIS software where it was classified into four categories of equal intervals viz; gentle (0° to 7°), moderate (7° to 15°), steep (15° to 22°) and very steep (22° to 29°).
2.3 Climate data analysis:

The climate data for the period 1978 - 2010 procured from India metrological department, Pune, were statistically analyzed in respect of temperature and annual rainfall. For analyzing temperature variations the data has been divided into three seasons viz., summer (March, April, May, October), winter (November, December, January, February), and monsoon (June, July, August, September). The variation in annual average maximum and minimum temperature has been calculated in the R software by using the average growth rate formula,

\[
\text{Growth Rate} = \frac{(V_{\text{present}} - V_{\text{past}})}{V_{\text{past}}} \times 100
\]

Where,

\[V_{\text{present}} = \text{present or future value}\]
\[V_{\text{past}} = \text{past}\]

The average growth rate for all the years were calculated in R software by using average growth function i.e. [avg. growth< -function (X)], the mean of all those average growth values for the whole period gives the change in annual average maximum and minimum temperature.

The annual rainfall data was computed and analyzed for 1978 to 2010 period to know the variation during the period of 32 years. The variation in average annual rainfall were calculated in Microsoft Excel by plotting the annual rainfall against the years, a trend line was drawn and difference in the y value give the variation in the rainfall during the whole period of study. The standard deviation and co-efficient of variation was also calculated to know the deviation and variation from the mean value by applying formulas, as

\[
\text{Standard deviation} = \sqrt{\frac{\Sigma}{N}}
\]
\[
\text{Coefficient of variation} = \frac{\text{Standard deviation}}{\text{Average}} \times 100
\]
2.4 Water Sampling and Analysis:

Twenty seven water samples were collected for analysis based on the preliminary field survey carried out to understand the overall impact of mining and industrialization on the surface and ground water resources. Out of twenty seven samples, ten samples are taken from surface water bodies and seventeen from ground water (tube wells) sources. The tube wells which are selected are used for both domestic and agricultural purposes in the study area. The samples were collected in sterilized screw capped polythene bottles of one litre. The samples from tube wells were collected after 10 min of pumping and stored with necessary precautions. Immediately after the sampling, pH, electric conductivity (EC) and total dissolved solids (TDS) were measured in the field by using multi parameter analysis kit (INOVA-WTFK). There after bottles were labeled and were bought to laboratory for determining chemical constituents such as total hardness, calcium, magnesium, sodium, potassium, sulfate, chloride, bicarbonates, copper, nickel, iron, cobalt, manganese, zinc, chromium. The analysis of the sample was done on the basis of standard methods suggested by the American Public Health Association (APHA, 1995; Trivedy and Goel, 1986).

Titration and Photometric methods were used to determine the chemical constituents of water samples.

2.4.1 Titration method procedure:

Titration is a common laboratory method of quantitative analysis that is used to determine the unknown concentration of a known reactant. Because volume measurements play a key role in titration, it is also known as volumetric analysis. Titrant is a reagent of known concentration and volume is used to react with the solution of the analyte whose concentration is not known. A typical titration begins with a beaker or Erlenmeyer flask containing a precise volume of the reactant and a small amount of indicator, placed underneath a burette or buretting syringe containing the reagent. By controlling the amount of reagent added to the reactant, it is possible to detect the point at which the indicator changes color. As long as the indicator has been chosen correctly, this should also be the point where the reactant and reagent
neutralize each other, and by reading the scale on the burette, the volume of reagent can be measured. In simple acid-base titrations a pH indicator may be used, such as phenolphthalein, which becomes pink when a certain pH (about 8.2) is reached or exceeded. Another example is methyl orange, which gives red color in acids and yellow color in alkali solutions. In a titration, both titrant and analyte are required to be in a liquid (solution) form. The majority of titrations are carried out in aqueous solution, other solvents such as glacial acetic acid or ethanol are used for special purposes. The mathematical result of the titration can be calculated directly with the measured amount. Sometimes the sample is dissolved or diluted beforehand, and a measured amount of the solution is used for titration. In this case the dissolving or diluting must be done accurately with a known coefficient because the mathematical result of the titration must be multiplied with this factor. As the concentration of the reagent is known, the number of moles of reagent can be calculated [since Molarity = number of moles / volume (L)]. Then, from the chemical equation involving the two substances, the number of moles present in the reactant can be found. Finally, by dividing the number of moles of reactant by its volume the concentration is calculated.

Ca$^{++}$, Mg$^{++}$, HCO$_3$ and Cl$^-$ were analyzed by using volumetric titration method. Concentration of Ca$^{++}$ and Mg$^{++}$, were determined by using standard EDTA and those of HCO$_3$, and Cl$^-$ by H$_2$SO$_4$ and AgNO$_3$ respectively.

2.4.2 Photometric Method:

The Concentration of Na$^+$ and K$^+$ were measured using a flame photometer (Model EEL PAT NO. 712700). Trace metals like Cu$^{++}$, Ni, Fe$^{++}$, Co$^{++}$, Mn$^{++}$, Zn$^{++}$, Cr were determined by Atomic Absorption spectrum (Perkin Elmer AAnalyst 800) using multi element Perkin-Elmer standard solution. The procedures followed for the analysis are discussed in the next section.
a) **Flame photometer:**

This method was used to determine the concentration of the Sodium (Na\(^+\)) and Potassium (K\(^+\)) in collected samples. A characteristic light is produced due to excitation of electrons when the sample containing sodium and potassium is sprayed into a flame. The intensity of this radiation is proportional to the concentration of sodium and potassium can be read at 589 nm and 768 nm respectively by using flame photometer. The filter of the flame photometer is set to 589 nm (marked for Sodium, Na) and 768 nm (marked for potassium, K). By feeding distilled water the scale is set to zero and maximum using the standard of highest value. A standard curve between concentration and emission is prepared by feeding the standard solutions. The sample is filtered through filter paper and fed into the flame photometer and the concentration is found from graph or by direct readings. Calibration curve is prepared in the ranges for the various standards of 10, 20, 40, 60 ppm and blank for Na and K. If the sample is having higher concentrations, it can be diluted to come in the range of determination and the dilution factor is taken into account during the estimation. The standard curve is a linear one at lower concentrations of sodium, however at higher concentrations it has got a tendency to level off. This curve is used to estimate the concentration of sodium in the sample (Ramachandra and Solanki, 2007).

b) **Atomic Absorption Spectrometer:**

Atomic Absorption Spectroscopy in analytical chemistry is a technique for determining the concentration of a particular metal element within a sample. Atomic absorption spectroscopy can be used to analyze the concentration of over 62 different metals in a solution. Typically, the technique makes use of a flame to atomize the sample, other atomizers such as a graphite furnace are also used. Three steps are involved in turning a liquid sample into an atomic gas:

1. Desolvation – the liquid solvent is evaporated, and the dry sample remains
2. Vaporizations – the solid sample vaporizes to a gas
3. Volatilization – the compounds making up the sample are broken into free atoms.
In flame atomic absorption spectroscopy, a liquid sample is aspirated and mixed as an aerosol with combustible gasses (acetylene and air or acetylene and nitrous oxide). The mixture is ignited in a flame of temperature ranging from 2100 to 2800 degrees C (depending on the fuel gas used). During combustion, atoms of the element of interest in the sample are reduced to the atomic state. The flame is arranged such that it is laterally long (usually 10cm) and not deep. The height of the flame must also be controlled by controlling the flow of the fuel mixture. A beam of light is focused through this flame at its longest axis (the lateral axis) onto a detector past the flame. A light beam from a lamp whose cathode is made of the element being determined is passed through the flame into a monochronometer and detector. Free, unexcited ground state atoms of the element absorb light at characteristic wavelengths; this reduction of the light energy at the analytical wavelength is a measure of the amount of the element in the sample (http://www.galbraith.com/spectroscopy.htm).

The concentrations of major and minor elements were known after they were analyzed by different methods. Results obtained were interpreted in terms of coal mining affecting the quality of water.

2.5 Ground and Surface water quality mapping:

Various parameters like, pH, TDS, Hardness and Conductivity were extrapolated to know the spatial distribution in the study area. Arc GIS spatial analyst tool has been used for the generation of spatial distribution maps which show variation in the values at different locations in the study area. Inverse Distance Weighing (IDW) method has been applied to generate the raster graphical output from the known values measured at different location by different laboratory techniques. These raster output were also used for overlay analysis with land use/land cover details to correlate various parameter values with different land use categories in the study area.