Chapter 05: Materials and Methods

5. MATERIALS & METHODS

During the complete study periods, time interval survey was carried out targeted to every community development blocks in Paschim Medinipur district for 36 months. During survey monthly data of various hydrological parameters of the available freshwater bodies also collected using water analyzer kit, record actual geolocation of the sampling stations by the handheld GPS device, fish species from the culture ponds and wild from the lowland areas and paddy fields, from the fishermen and local fish markets. The water samples from the unreachable aquatic bodies are collected in a dark, sterile, airtight bottle and brought to the laboratory for further analysis. These significant steps followed to execute the research work are described below:

5.1 Remote sensing

The present study is mainly based on real time satellite data, which have been used for the accurate mapping of the water bodies and conservation priority area. Multispectral satellite data acquired by IRS P6, 1D (LISS-III of resolution 23.5 m) and PAN 5.8 m (Black/White) are the best available option in India to study and classify water bodies. Different imageries generated using remote sensing has been used in this current study were procured for and are explained as following:

5.1.1 High-resolution images from Google Earth®

Google Earth® is a web based utility which has augmented GIS and remote sensing knowledge in a comprehensive way. The utility has primarily offered free updated data of remote sensing and the GIS directed to public domain which enhances aptitude to understand complex science through very interactive and participatory way. Now days, provider has updated the whole region consist with high resolution images by TeleAtlas, Europa Technologies, TerraMetrics, GeoEye, DigitalGlobe, Cnes/Spot, Data SIO , MRC/JHA, NOAA, US and Mapabc at 0.8 m-2m resolution.

The updated utility version of Google Earth Pro® was accessed and premium resolution (4800X3883 dpi) image of the study site was downloaded at 2000-6000m of eye altitude and scenes were sensitively captured and stored into JPEG file format and gone for geo-referenced. Approximately 200 images per district were captured, processed and merged in 24 bit BGR (true color) format for accuracy during the study periods. The main features like river systems, ponds and reservoirs were manually extracted from the images.

Most remote sensing job is sophisticated and needs the expertise to work with this. Currently it has been established that, Google Earth has the eminence potentiality to do for remote sensing. Google Earth generally aggregating and organizing the satellite imagery, collecting aerial photography from various ranges of sources and presenting in a very lucid way to take up and easily acceptable by the general end user. Indeed, its hassle free use and way of presentation is mainly drawing public interest in geospatial technology, its applications, attaching images and valuable other information to the geospatial data. Functionality stands Google Earth a key tool for natural resource conservation. While Google web system is playing an active role in this development through its Google Earth Outreach program. Spatially as a visual tool for communication researchers have established that Google Earth can publish results through effective and proper way in accessible format.

5.1.2 Indian remote sensing data

The primary data source for the study was satellite imagery of both pre monsoon and post monsoon data of NRSA (National Remote Sensing Agency), ISRO. IRS image were analyzed and a cloud free area of (column 8346 by 4975 rows) 23007714 pixels (DN value>=1) was subset from the original 4 full scene (Table-1) image using TNTmips 7.3 of Micro Image software®. This sub-scene image was geometrically rectified to the projection parameter system- Lat/Long, datum- India-India, Everest-1956. The rectification has been achieved at the accuracy level of 0.5 m in Google image, 2.5 m in fused image and 12.5 mt in LISS-III images. All the geo-referencing (registration) has been done on the basis of online Google Earth instead of toposheet map.

Pre monsoon			Post monsoon		
Sat	Path/Row	Acquin date	Sat	Path/Row	Acquin date
IRS-P6	106/55	8th Jan,12	IRS-1D	106/55	23rd Mar,12
IRS-P6	106/56	8th Jan,12	IRS-1D	106/56	23rd Mar,12
IRS-P6	107/55	26th Nov,13	IRS-1D	107/55	3rd Apr, 08
IRS-1D	107/56	19th Jan,09	IRS-1D	107/56	3rd Apr,09
(Sat = Satellite, Acquin = Date of acquisition)					

Table 01: Used satellite imageries in the study area.

5.1.3 Software

TNTmips version 7.1 and 9 of Microimage® has been utilized in satellite image processing, classification and data transformation, whereas TNTLite has been used during the onscreen display. For word processing, graphs and databases MS-ACCESS 2017, MS-WORD 2017, MSEXCEL 2017 and SPSS 22.0 Windows version have been used.

Maps (Made available by ICAR-CIFRI, Barrackpore)

1. DPMS (Survey of India, 1972)
2. Mouza boundary (Survey of India, 2005).

Other accessories

1. GPS (Germin- *etrex* Vista® H)
2. Census data (2011)

5.1.4 Image preparation

The images were pre-treated with under mentioned digital process before analysis to extract, classify water bodies and other real objects.

5.1.5 Import of satellite images

In TNTmips a variety of different formats are allowed to import/export in its own format rvc along with subset. This file is further rectified for different process eg. images from IRS (IRS super structure) and Google (JPEG) were imported to rvc format.

5.1.6 Geometric Correction and referencing

Satellite images received from National Remote Sensing Agency, Hyderabad were geographically corrected to an accuracy level for the present study. TNTmips with Ground Control Points (GCP), online Google Earth correction and rectification had been done by standard georeferencing techniques (Ref. module); subsequently images from DPMS, mouza maps were corrected and referenced to ensure that all images were made uniform with the same geographic coordinates system.

This sub-scene image was geometrically rectified to the projection parameter system-Latitude/Longitude, datum- India-India, Everest-1956. Basically, online Google Earth was used as base map. To minimize error in geo-referencing 8 control points were taken to the object that was geo-referenced in order to embed the geographic location of cells in raster objects.

5.1.7 Histogram

TNTmips works based on raster histograms for getting a suitable contrast that improves the image quality. Full histogram data analysis of image was done for the retrieval of actual value of the pixel and pyramiding for averaging size of every pixel. Remote sensing images often appear dark, dull, and in visible detail for that *trim and stretch* enhancement operation was done by based on information in the raster histogram.

5.1.8 Band Reduction

Multi-spectral data were composed of 4 contiguous bands which were highly correlated and therefore one band can represent many correlated bands, so reduction is extremely advantageous (TNTmips Tutorial, 2007). In order to reduce this redundancy in the multi-spectral data, two methods of band reduction were used in the study (i) Principal Component Analysis (PCA) and (ii) Noise reduction (NR). PCA performed on the entire region in TNTmips to reduce data dimensionality. By utilizing the PCA reduction of 4 bands principle component bands will reduce processing time and avoid useless information contained in the entire data set. PCA on only selected bands determined to be more important for water quality.

Another method of band reduction is the noise reduction (NR). This process reduces or removes noise which is present in the sensor itself such as striping or banding. It was done on all of the images of this region and compared with the PCA.

5.1.9 Fusion and band combination

Satellite data (IRS-P6 and PAN) was fused to extract more information through an equation S = PAN x3 xR/ (R+G+B+1) where S is numerical value of fused image. Resulted fused image was more valuable for remote sensing information.

5.1.10 Mosaicking

A mosaic corresponds to an assembly of satellite image prepared for obtaining a view of the entire area. Four subset images were mosaic into one by contrasting ‘none’ and set value ‘0’ instead of 255.

5.1.11 Feature mapping

After geometric correction and fusion Image-interpretation was performed to convert remotely sensed imagery to digital object by feature mapping. Feature Mapping was performed to develop specifically to assist in classification of these images. The aims and objects of Feature mapping was to identify and measure the features in a set of processed raster data and by combining the knowledge of study site with TNTmips' processing power (TNTmips tutorial, 2007). However feature mapping is a technique of unsupervised classification, primarily it was used for marking water area as a whole.

5.1.12 Classification

Classification of the information contained in the imagery was one of the major applications of image interpretation. TNTmips offers automatic classification (supervised and unsupervised) respectively. Automatic classification was applied where uniform properties remain throughout, such as satellite imagery. Several classes of features extracted from LISS-III/fused images eg. WATER class was being extracted for marking various signature of water related pixel, SAND class for marking sand feature of the riverine system and FOREST for forest coverage area.

Classification methods that were followed in the analysis of images especially, on IRS data included k- mean (un-supervised), Maximum Likelihood (supervised), through TNTmips by entering the classes and threshold number or creating area of interest at the specific region. This was done by considering average of PCA (Principal Component Analysis) values for the pixels surrounding the location where water sample were taken or by visual interpretation. This had been accomplished by an algorithm which transforms raw image data into a new image based upon the formula:

$$NDWI = (Band1 - Band3)/(Band1 + Band3)$$

$$NDVI = (Band3 - Band2)/(Band3 + Band2)$$

The images produced in this manner provided information on vegetation and water. NDVI minimizes the effect of shadow and used in separation of water bodies. The wetland areas come in the zone of lower NDVI than the terrestrial vegetation (Menon and Bawa 1997).

To extract the maximum amount of data possible, a step-wise classification was performed on a subset of fused image or simple subset image. All classified information after categorization was made into raster object.

5.1.13 Vector making

All classified raster objects were transformed into vector format through auto transformation (raster to vector) in TNTmips and used for further analysis.

The GIS is commonly described as vector-based data when its data structures contain x, y coordinates for basic data elements like points, lines etc. A vector object includes nodes, polygons, lines, points and labels. The vector elements and the attributes assigned to them are stored together in a TNT Project File. Each element is assigned to a class with attached database for display. TNT supports more than one type of vector topology eg. 2D, 3D vector objects, a vector object represent particular theme like water area, roads etc. and the theme will establish necessary topology level. The topology type can be changed for any vector object in the spatial data editor. Nodes automatically positioned at the intersection of lines that cross and lines are created automatically when polygons are created, and polygons created as per need.

5.1.14 Database creation

The databases, considerate as the store house of all kind of data, can be managed, retrieved, and manipulated for specific reason. TNTmips allows building an internal database when a raster is converted to vector format. User based database can be attached or created within the object through a number of gateways.

5.1.15 Spatial database

GIS has a baseline data created internally on every object during vector making. They are spatial data eg. Y-coordinate (Latitude)/ X-coordinate (Longitude), polystate, bound-length etc.

5.1.16 Attribute database

Ground truthing data were collected during entire field survey. These ground truthing data of water bodies were the attribute data which were created for GIS as shown in (Table-2).

Class	Attribute	Description
Water body	District, Block, Vill.	Name of the attribute
	Area(ha)	By enquiry/estimate
	Depth(m)	Depth of the water body (min/max)
	Weed cover (%)	By visual observation
	Water transparency	High/medium/low
	Type	Pond, tank, beel, reservoir etc.
	Status	Perennial and seasonal
Fish and fishery	Fish species	Number of species recorded/reported
	Culture type	Scientific or traditional
	Fishing right	Private, free, lease, license, community
	Production	Fish production Kg/ha/yr

Table 02: Various attribute data

5.1.17 Combination of attribute and spatial data

All spatial and the attribute data were combined to analyze various themes for generation of species richness area, market place, conservation network area etc. in GIS.

5.1.18 Structure of data for GIS layers

Metadata was broadly classified into 4 categories (Table-3). These data were used to display various themes through GIS viz. species richness map, Transport map etc.

Type	Vector data	Layer
Region	Polygon/region	State, district, Block
Water body	Polygon and line	Closed water body, river and stream
Transport	Line	Rail, road
Spot mark	Point	Ground thruthing/survey spot

Table 03: Structure of different layer of vector data

5.1.19 Thematic mapping

Thematic mapping is one of the ways to develop mapping applications. Map components were logically organized into sets of layers called *layout*. A theme, a group of similar geographic features like water body (polygon), road network (lines), or place of survey (point) was manipulated using classification and legend (symbology) to create different thematic maps and classify attribute data into group data. These groups were displayed using different legends reflecting different colors, line thicknesses, and symbols. Combination of thematic layer produced thematic map at 1:50000 scale.

5.2 Preliminary water body mapping for field survey

After map creation, the next important step was to analyze spatial data to find certain areas where field based investigations were required and to achieve this, supervised classification of the satellite data was performed and confusion classes were identified and ground truthing was carried out with hard copy of layout.

5.2.1 Use of GPS

The surveyed water body location and positioning data were collected using GPS (model used Garmin eTrex Vista HCx) device. The data collected from entire Blocks inPaschim Medinipur during survey period of 2014-2017. The coordinate system considered in geo-referencing was also used in taking reading in GPS. All the GCPs had been fed in to GIS layer and checked for accuracy.

5.2.2 Field survey

Field survey was performed for the period from year 2013 to 2016. Data collected randomly since the entire study area. A total of 212 numbers of ground truthing data for reservoirs were considered especially large closed water bodies. The physico-chemical properties investigated using water analysis kit and in support of laboratory. The data were properly documented.

5.3 Collection of fish samples

Inhabitant freshwater ichthyofauna were collected using fishing gears viz., cast and drag net etc. Fishes also collected secondarily from the fishermen and from local fish market. The photograph taken immediately after collection of the fish samples. Then they were subjected to prepare a fish data.

5.3.1 Preservation and identification of fishes

The fish samples having collected was subjected to preservation in sterilized container using 4% formalin. After that they were examined and identified by following the methodologies after FishBase (2017) and the available literature of Talwar & Jhingran, 1981. Photographs were taken immediately after sampling of the fish specimens to get the actual body colour and contour patterns. Fishes were identified taxonomically to species level except for juveniles form.

5.4 Diversity measurement

Mathematical expression related to diversity of the species within a community is called diversity index. It provides information about the composition of community. Diversity indices consider species relative abundance. It gives also the information like community commonness and rarity. Quantification of diversity to the biologists is a key tool to study about community structure. The diversity indices used in this research study are Shannon’s diversity index, Simpson’s diversity index, Margalef’s richness index, Evenness index, Pielou’s equitability index, Chao-1.

The Shannon diversity index (H) (Kasturirangan, Aravamudan et al. 1996):

Shannon's index calculates the diversity of species within community. It considers the abundance and evenness both. The pi is calculated at first which means proportion of species within total species. This value then multiplied by natural logarithm of pi and total of the results is taken and finally multiplying with -1.

Shannon's diversity index $H = -\sum_{i=1}^{S} p_i \ln p_i$

Where H= Shannon's index of diversity

S= total taxa number within a community

pi= proportion of the species

ln= natural logarithm

Shannon's equitability (E_H) is generally obtained after division of H by H_{max} Equitability value ranges from 0 to 1 where 1 mean complete evenness.

Shannon's equitability index $E_H = \frac{H}{H\max} = \frac{H}{\ln S}$

Where H_{max}= lnS

Disturbance in any ecological environment changes results to the diversity of any community. In order to preserve biodiversity in a particular area, one needs to understand the conservation and management strategies affecting the diversity. Biodiversity indices act as effective tools in getting the diversity status to the researchers and help the conservationist to protect the desired species.

Simpson's Index (D) (Prasad, Ramachandra et al. 2002):

Simpson's index determines probability among two individuals sampled randomly from a large community will be of same species.

Simpson's index $D = \frac{\sum n(n-1)}{N(N-1)}$

Where, n = species wise total individuals

N = total individuals of all the species.

D value ranges from 0 to 1.

In this index, 0 = infinite diversity where 1 = no diversity. The higher value of D means lower diversity.

Simpson's Diversity Index 1-D:

This index represents probability of two individuals selected randomly from a community are from different species. This value also ranges 0 to 1. Here, higher value=higher diversity.

Margalef's richness index (Shannon 1948): $\frac{(S-1)}{\ln(n)}$

Here S= number of taxa

Chao-1(Simpson 1949):

This index determines the total species richness.

$$Chao1 = S + \frac{F1(F1-1)}{2(F2+1)}$$

F1 = Singleton species number

F2=Doubleton species number

S= Taxa number

Dominance Indices:

Dominance indices are weighted toward the abundance of the commonest species. Commonly used dominance index is Simpson's diversity index. It considers richness and evenness both.

Species evenness:

Species evenness means the available species numbers are close to each other in an environment. It quantifies the numerical equality of species in the specified community. Evenness of community is represented following the Pielou's evenness index (Margalef 1958):

$$J' = \frac{H'}{H'max}$$

Here H′= Shannon diversity index

and H′max = maximum value of species diversity (in case every species are likely equal)

$$H'max = -\sum_{i=1}^{s} \frac{1}{S} \ln \frac{1}{S} = \ln S$$

S = total species number

J' value ranges 0 to 1. Lower the value of J' means less evenness among the species in communities. Conversely it shows the dominancy of fewer species in that community.

SHE analysis:

SHE analysis deduces relationship among the S (the species richness), H (the Shannon index) and the E (the evenness) of the sampled data. Therefore, it contributes species number and their equitability to the changes in diversity. SHE analysis describes the changes in parameters with increasing of sampling. The output is a plot on spreadsheet of calculated S, H and E for all the selected samples. SHE analysis is useful for identifying ecotones (Chao 1984). It is most effective method in testing 'goodness-of-fit' to these models.

Individual rarefaction

This model is used to compare diversity of different sampling sizes. Samples have to be taxonomically similar and obtained by standardized sampling from similar habitat.

This model estimates the expected species number from a very small sample size. The algorithm is taken from Pielou equation (Pielou 1966).

Rarefaction normally assesses species richness from sampling data. Rarefaction generates rarefaction curves which are obtained by plotting species number. These curves generally grow rapidly initially so long the most common species are found and it forms plateau after that getting rarest species sampled.

The formula used to this model is:

$$E\left(S_n\right)=\sum_{i=1}^{S}\left[1-\frac{\left(\frac{N-N_i}{n}\right)}{\frac{N}{n}}\right]$$

Here = expected species number

N = total individuals of the rare taxa,

Ni =total individuals in each of the original taxa and n = sub-sample

The curve highlights on sampling efforts based on the number of species. The greater the sampling size the chance of getting the less common and rare species. It generates abundances on equivalency based on various sampling size. The total abundance imbalances are due to sampling differences and not due to actual abundances (rarity) differences.

Diversity profile:

This profile compares diversities of several samples. The validity of the diversities using arbitrarily different diversity indices is not out of question. One sample may contain large taxa number while the other has higher Shannon index value. It can be minimized by defining a family of various diversity indices based on single continuous parameter (Hayek and Buzas 2010).

PAST uses exponential of Renyi index, that depends on α value. α=0 means total species number. α =1 is proportional to Shannon index, while α =2 is proportional to Simpson index.

$$\exp(H_\alpha) = \exp\left(\frac{1}{1-\alpha}\ln\sum_{i=1}^{s} p_i^\alpha\right)$$

This program plot several diversity profiles together. The obtained profile of the diversities cannot be compared if they cross each other.