CHAPTER VII

LANDUSE / LAND COVER ANALYSIS BY DIGITAL TECHNIQUES

7.1 Introduction

7.2 Landuse/ Land Cover Using Digital Techniques
CHAPTER – VII

LANDUSE / LAND COVER ANALYSIS BY DIGITAL TECHNIQUES

7.1 Introduction:

In the previous chapter discussed points with the introduction, resource mapping, landuse mapping and landuse/land cover of the study region.

This chapter deals with the introduction and totally devoted with the landuse / land cover using by digital techniques.

7.2 Landuse/ Land cover Using Digital Techniques:

The methodology adopted for classifying landuse/ land cover using digital techniques was already described in chapter V. In this chapter an attempt has been made to analyze the results of landuse / land cover classification done by different algorithms such as maximum likelihood, parallelepiped and minimum distance for distinctly different landuse/ land cover categories in the different parts of the study region. To classify landuse/ land cover two image processing systems viz., multispectral data analysis system (MDAS) and interactive multispectral data analysis system (IMDAS) were utilized. Only one algorithm i.e. maximum likelihood was used using MDAS for classification covering areas such as Manjara River and its environs, Manjara dam of the study region. In these areas more than 90 percent accuracy was achieved. The main advantages of digital techniques for landuse/ land cover classification are 1) fast processing, 2) classification of inaccessible areas by using similar known areas and 3) retrievable whenever it is required. At the same time there are certain limitations such as resolution, scale, cloud.
cover, shadows etc., the limitations can be avoided by selecting the correct season data and by limiting the use of maps upto certain level. For example maps generated on 1:50000 district level planning and so on.

The Plate No. 7.1 shows landuse classification of different areas as cited above. The colour coding is as follows:

<table>
<thead>
<tr>
<th>Colour</th>
<th>Corresponding landuse/ land cover Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td>Standing Crops</td>
</tr>
<tr>
<td>Blue</td>
<td>Water</td>
</tr>
<tr>
<td>Grey</td>
<td>Fallow Black Soil</td>
</tr>
<tr>
<td>Yellow</td>
<td>Fallow Red Soil</td>
</tr>
<tr>
<td>Cyan</td>
<td>Scrub Land</td>
</tr>
<tr>
<td>Green</td>
<td>Forest</td>
</tr>
<tr>
<td>White</td>
<td>Sand</td>
</tr>
<tr>
<td>Purple</td>
<td>Grass lands/ grazing lands</td>
</tr>
</tbody>
</table>

This is assumed to be correct as field checks and accuracy estimation analysis were done for the entire area following the methodology as described in chapter V. Covering the same area, analysis was done on IMDAS using the three algorithms as mentioned earlier. Plate No. 7.2 shows the false colour composite (FCC) of Landsat TM of the part of the study area for comparison purpose. The FCC can be visually interpreted. But here the corresponding CCT
(covering the same area) was used for classification of landuse / land cover classification using the different algorithms through IMDAS.
Plate No. 7.1, 7.3 and 7.4 show the landuse / land cover classification using maximum likelihood, parallelepiped and minimum distance respectively. Plate No. 7.1 which is an output of maximum likelihood classifier is comparable with the visual interpretation of the raw data. This indicates that the maximum likelihood classifier gave a good result when compared with other algorithms. Plate No. 7.3, which is an output of parallelepiped algorithm shows maximum / grazing lands (Purple). The colour coding and training sets are same in all the cases. Sandy areas (white) are shown more when compared to plate 7.1. The cropland area in plate 7.3 (parallelepiped) under Manjra dam was classified as forest. It indicates that parallelepiped algorithm could not distinguish cropland from forest. The outputs of small areas of parallelepiped algorithms are shown in plate’s No. 7.5, 7.6, 7.7 and 7.8.

The landuse / land cover classification shown in plate no. 7.4 is an output of minimum distance classifier. Here, forest is (green) classified considerably on par with the maximum likelihood algorithm output (plate no. 7.1). But the scrub land (cyan) was classified as if it is distributed in maximum area. This means that the minimum distance classifier could not distinguish crops from forest areas. But under the Manjra dam the crop land was properly classified. This means wherever the crops and forests have distinctly different spectral signatures, which could be separated. Thus the maximum likelihood classifier was found to be far superior compared to parallelepiped and minimum distance classifiers.
Plate No. 7.1: Landuse Classification using Maximum Likelihood Algorithm

Plate No. 7.2: Landsat TM FCC of a Part of the Study region
Plate No. 7.3: Landuse Classification of a Part of the Study Region using Parallelopiped Algorithm

Plate No. 7.4: Landuse Classification using Minimum Distance to Mean's Algorithm
Plate No. 7.5: Landuse Classification using Parallelopiped Algorithm for Ahmedpur Sandy Area

Plate No. 7.6: Landuse Classification using Parallelopiped Algorithm for Tawarja Dam
Plate No. 7.7: Landuse Classification using Parallelopiped Algorithm for Tawarja dam, Shiur, Uti and Surrounding Area

Plate No. 7.8: Landuse Classification using Parallelopiped Algorithm for Tawarja dam, Shiur, Uti and part of Nilanga Hill Area
With the help of the analysis by digital techniques it can be concluded that the maximum likelihood classifier should be adopted for proper landuse / land cover classification for any given area.

7.2.1 Application, Automated 3D Surface Model:

3D geological information systems provide a means to capture, model, manipulate, retrieve, analyze, and present geological situations. Traditional geological maps which illustrate the distribution and orientation of geological materials and structures on a 2D ground surfaces provide vast amounts of raw data. It is thus vital to develop a set of intelligent maps that shows features of geological formations and their relationships.

A. Digital Elevation Model of the Study Region:

DEM is a representation of the terrain surface by coordinates and numerical descriptions of altitude. DEM is easy to store and manipulate, and it gives a smoother, more natural appearance of derived terrain features. Therefore, the created DEM is the foundation of 3D geological maps when the co-ordinates of the vertices of geological formations can be interpolated. The data consists of 4 topographical map sheets, with 3D coordinates of terrain, contour lines, and other information. The maps are in GEOTIFF format at a scale of 1:150000 (Fig. No. 7.1). These DEMs were then integrated into a whole DEM of the study region using a DEM Global Mapper. The final gridded DEM data with 5-metre intervals for the study region was obtained (Fig. No. 7.1).
Figure No. 7.1: Tiled DEM of Latur District (courtesy USGS)
Figure No. 7.2: Cropped DEM using Latur district base map
B. Cropping DEMs using the Study Region:

After integrating DEMs tiles, the next process is to extract (crop) the required region of the study region from integrated DEMs using the area base map of shape file. For this process, we use the software GLOBAL Mapper 11v to crop the DEMs with only required region's terrain data. The remaining area is considered as null data as shown in Fig. No. 7.2.

C. Accessing and Concatenating DEMs in MATLAB:

After the successful cropping of all the DEM data sheets (tiles), we import them in MATLAB for further processes. The DEMs can be converted in to DTED (Digital Terrain Elevation Data) version 0,1,2 any format, and import them in MATLAB. The DTED0 files have 120-by-120 points. DTED1 files have 1201-by-1201. The edges of adjacent tiles have redundant records.

Acquiring all the data sheets with their specified location (projection) and sequence of data sheets are very important here. Concatenation of the DEM tiles with respect to their locations needs horizontal and vertical concatenation.

1) Horizontal Concatenation:

First, we concatenate the matrices of top-left and top-right tiles (Fig.2), i.e. Horizontal concatenation.

\[ H_1 = TL \text{ (horzcat)} TR \]  

Where, \( H_1 \) is a concatenated matrix of top-left

(TL) and top right (TR) matrices.
Next, we concatenate the matrices of Bottom-Left and Bottom-Right tiles, i.e. again Horizontal concatenation.

\[ H_2 = \text{BL (horzcat) BR} \]  

Where, \( H_2 \) is a concatenated matrix of Bottom-left (BL) and Bottom-right (BR) matrices.

2) Vertical Concatenation:

Next, we need to concatenate \( H_1 \) and \( H_2 \) matrices vertically, i.e.

\[ H = H_1 \text{ (vertcat) } H_2 \]  

Where, \( H \) is a complete concatenated matrix of \( H_1 \) and \( H_2 \).

D. Visualizing 3D Geographical Surface Model:

A workflow was chosen, on the one hand, by applying GIS methods using ESRI shape files and global mapper software for data acquisition, maintenance, and presentation and on the other hand, by applying three-dimensional spatial modeling with an interactive 3D modeling in MATLAB. Based on Non-Uniform Rational data, any geometric shape can be modeled. Besides surfaces of the different engineering geological units, solids using boundary representation techniques were modeled. In MATLAB it is one of the easiest ways to visualize the well-defined projected data sets in 3D view using mathematical functions surf and mesh. To visualize the acquired projected data set over a rectangular region, we need to create colored parametric surfaces specified by \( X, Y, \) and \( Z \), with color specified by \( Z \).
A parametric surface is parameterized by two independent variables, i and j, which vary continuously over a rectangle; for example, 1<=i<=m and 1<=j<=n. The three functions \( x(i,j) \), \( y(i,j) \), and \( z(i,j) \) specify the surface. When i and j are integer values, they define a rectangular grid with integer grid points. The functions \( x(i,j) \), \( y(i,j) \), and \( z(i,j) \) become three m-by-n matrices, \( X \), \( Y \), and \( Z \). Surface color is a fourth function, \( c(i,j) \), denoted by matrix \( C \). Each point in the rectangular grid can be thought of as connected to its four nearest neighbours.

\[
\begin{array}{c}
  i-1,j \\
  i,j - i,j - i,j + 1 \\
  i+1,j \\
\end{array}
\]

Surface color can be specified in two different ways: at the vertices or at the centers of each patch. In this general setting, the surface need not be a single-valued function of \( x \) and \( y \). Moreover, the four-sided surface patches need not be planar. For example, one can have surfaces defined in polar, cylindrical, and spherical coordinate systems.

The shading function sets the shading. If the shading is interpolates, \( C \) must be of the same size as \( X \), \( Y \), and \( Z \); it specifies the colors at the vertices. The color within a surface patch is a bilinear function of the local coordinates. If the shading is faceted (the default) or flat, \( C(i,j) \) specifies the constant color in the surface patch:

\[
\begin{array}{c}
  (i,j) \\
  C(i,j) \\
  (i+1,j) \\
\end{array}
\]

\[
\begin{array}{c}
  (i,j+1) \\
  (i+1,j+1) \\
\end{array}
\]
In this case, C can be the same size as X, Y, and Z and its last row and column are ignored. Alternatively, its row and column dimensions can be one less than those of X, Y, and Z.

E. Assigning Axes to 3D Model:

MATLAB automatically creates axes, if one does not already exist, when you issue a command that creates a graph, but the default axes assigned by MATLAB don’t match with real coordinate systems of this projected area. This existing model is built with 3 axes data x, y and z respectively. The X and Y axis represents the latitude and longitude values for this model i.e.

UPPER LEFT X=76.2076079218
UPPER LEFT Y=18.8385493143
LOWER RIGHT X=77.2934412815
LOWER RIGHT Y=17.8677159574
WEST LONGITUDE=76° 12' 27.3885" E
NORTH LATITUDE=18° 50' 18.7775" N
EAST LONGITUDE=77° 17' 36.3886" E
SOUTH LATITUDE=17° 52' 3.7774" N

The above shown values are associated with all four tiles of DTED files. The Z axis itself represents the terrain (height) values of ground surface objects. Here in this model the elevation data is assigned in feet scale format i.e. 0 to 3000 feets.
Figure No. 7.3: 70° camera view point of surface model with gray color scheme
Figure No. 7.4: 45° camera view point of surface model with HSV color scheme
Figure No.7.5. A true color composite scheme (Atlas shader) 3D model
With reference to the processes discussed above, the 3D visualization experimental results are shown in the Fig. No. 7.3, 7.4 and 7.5 for 3D model of Latur district geological surface.

Some key processes for automated 3D geological surface modeling such as data acquisition, concatenation, 3D surface modeling and axes data managing have been presented. The visualization experiments are done using data for the study region. In the future work, we attempt to overlay real time map layers on this 3D surface model.

7.2.2 Application, Digital Elevation Model:

Continuous overexploitations of natural resources like land, after and forest have caused serious threat to the local population of the semiarid region. Thus, problems like little scope for, declining ground water level and shortage of drinking water prevail. In this study an attempt has been made to estimate the availability of water resources and water scarcity regions using advanced topographic mapping techniques.

Topography is basic to many earth surface processes and thus finds applications in ecology, hydrology, security, agriculture, climatology, geology, pedology, geomorphology and a host of other domains and constitutes the basis for explaining processes and predicting them through the process of modelling. The tremendous role of topographic mapping in national development continues to receive recognition by national, state and local governments the world over. The importance of topographic mapping as a national project is therefore growing and accurate topographic maps as its
major products are considered as indispensable components of national geospatial data infrastructure. In countries with developed and stable economies, a clearly articulated road map is usually made for regular, fresh topographic mapping using current data and the state-of-the-art mapping technologies. In less-developed countries however, the problem of poverty and inadequate technical capacities in the area of geo-information production and management culminate in the inability of such countries to carry out fresh topographical mapping at regular time intervals using expensive but accurate topographical mapping techniques. For such countries, an alternative and cost-effective strategy over the years has been topographical map revision using existing topographical maps, satellite imageries and other readily available spatial data sources. Unfortunately, the process of extracting topographical information (contours) from existing topographical maps and integrating same in a new digital topographical map, often at a larger scale, is usually a lengthy and time-consuming process due to differences in map units and contour intervals between the existing base maps and the new map. Moreover, most of the topographical map sheets to be used as base maps for a revision exercise are either missing or, where they exist, are very old and suffer from severe distortion. A recent development representing a tremendous forward leap in remote sensing technology that will significantly eliminate some of the lacunae associated with topographic map revision from existing topographical maps is the launching of the Shuttle Radar Topography Mission (SRTM) in February, 2000.

A digital elevation model (DEM) is a digital representation of a terrain's surface-commonly for a planet (including Earth), moon, or
asteroid-created from terrain elevation data. Using the Synthetic Aperture Radar (SAR) interferometry to produce the first near-global high resolution digital elevation model (DEM) of the Earth, SRTM has created an unparalleled set of global elevations that is freely available for modelling, aping and environmental applications (Gorokhovich and Voustianiouk, 2006). The global availability (about 80% of the Earth surface, covering land masses between 60°N and 56°S) makes it the most widely used set of baseline elevation information for a wide range of applications and this development has been identified by professionals in the geo-information arena as a significant landmark that will tremendously revolutionize medium-scale topographic mapping.

The near-global SRTM digital elevation model (DEM) product was processed and compiled at a resolution of 90m by the Consultative Group for International Agriculture Research Consortium for Spatial Information (CGIAR-CSI) and hosted on a Web portal for free public access and download. Although this product presents attractive promise for terrain analysis for an impressively wide range of applications, several researchers have proposed a thorough evaluation of its vertical accuracy.

A. Source Data:

The studies employed in this research are with three major sources of spatial data (90-m resolution CGIAR-CSI SRTM digital elevation data, 1:50,000 topographical map of the study region and contour lines randomly distributed over the study region). The SRTM Digital Elevation Model was processed and maintained by the Consultative Group for International Agriculture Research
Consortium for Spatial Information (CGIAR-CSI). In the form compiled and maintained by CGIAR-CSI, the SRTM elevation data have a spatial resolution of 90m. This data set is seamless with all voids filled using a methodology based on spatial filtering (Gorokhovich and Voustianiouk, 2006). The CGIAR-CSI SRTM 90m digital elevation data sets are provided to the general public in 5º X 5º tiles in computer-compatible raster formats (Geo-Tiff and ARC/INFO ASCII Grid). The data set is in Lat-Lon coordinate system projected on the WGS 84 Ellipsoid.

One topographic map sheet at the scale of 1:50,000 covering the chosen study site was selected for use as a reference for analyzing the SRTM elevation data. The map sheet was digitized into different layers as part of the input data sets within the framework. The maps are based on the UTM projection (Zone 43) on WGS 1984 Ellipsoid and had a contour interval of 25 feet.

B. Materials:

Five major software packages were employed for the processing of the data and the visualization and analysis of the results. These included the MATLAB, open-source Integrated Land and Water Information System (ILWIS 3.4), Quantum GIS (QGIS) 1.6, Natural Resource Data Base (NRDB 2.7) and Microsoft Excel. In addition, we developed a number of in-house programs in Visual Basic 6.0 for performing some specialized functions such as coordinate transformation, terrain profiling and elevation data extraction from ASCII raster data sets.
C. Methodology:

- Quantitative statistical and geo-statistical tests were performed on the two spatial data sources for different terrain configurations and contexts to determine their suitability for topographical mapping in different scenarios. In particulars of this study consisted of: measuring the vertical accuracy of the DEM derived from the 1:50,000 topographic map and that of the 90-m resolution CGIAR-CSI SRTM digital elevation data against higher precision GPS measurements within the same region.
- Interpolating digital elevation models from an existing topographic map covering the same area and comparing measurements from the two sources.
- Implementing a processing strategy to minimize errors emanating from contour interpolation using SRTM data as a base.
- Contour lines generation with 25 meters interval.
- Extraction of water bodies from DEM.
- And finding which villages fell under critical water scarcity regions of the study region.

The data sets employed in this study emanated from disparate sources based on different formats, coordinate systems and projections. The first step in the exploitation of the data sets was therefore the transformation of all the data sets into a common system. Since the CGIAR-CSI SRTM 90m digital elevation data sets were in Lat-Lon WGS84 system, the topographic map layers (contours and rivers) and the GPS elevation data in UTM Clarke 1880 system were transformed into the Lat-Lon WGS84 system using tools available in ILWIS 3.4 software. The QGIS 1.6 is used to process DEMs
and polygene them to find the water-bodies/lakes of the study region for finding water scarcity zones. In MATLAB we processed the raster map (JPG) of DEM to extract the river basins.

And finally all the processed data imported into our local spatial data base to find out the villages which fell under the critical water scarcity zones.

D. Land Elevation Contours:

Contouring is the most common method for terrain mapping. Contour line connects points of equal elevation, the contour interval represents the vertical distance between contour lines, and the base contour is the contour from which contouring starts. Contour lines are lines drawn on a map connecting points of equal elevation. The contour line represented by the shoreline separates areas that have elevations above sea level from those that have elevations below sea level. We refer to contour lines in terms of their elevation above or below sea level.

Contour lines are useful because they allow us to show the shape of the land surface (topography) on a map. Suppose a DEM has elevation readings from 362 to 750 meters. If the base contour is set to 400 and the contour interval at 100, then contouring would create the contour lines of 400, 500, 600 and so on.

Contour lines can be drawn for any elevation, but to simplify things only lines for certain elevations are drawn on a topographic map. The contour interval is constant for each map. It will be noted on the margin of the map. One can also determine the contour
interval by looking at how many contour lines are between labeled contours.

**Table No. 7.1: Particulars of Covered Elevation Area**

<table>
<thead>
<tr>
<th>ID</th>
<th>Elevation (Mtr)</th>
<th>Area (Ha)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>415</td>
<td>315</td>
</tr>
<tr>
<td>2</td>
<td>425</td>
<td>2127</td>
</tr>
<tr>
<td>3</td>
<td>450</td>
<td>8068</td>
</tr>
<tr>
<td>4</td>
<td>475</td>
<td>26511</td>
</tr>
<tr>
<td>5</td>
<td>500</td>
<td>49710</td>
</tr>
<tr>
<td>6</td>
<td>525</td>
<td>47635</td>
</tr>
<tr>
<td>7</td>
<td>550</td>
<td>179404</td>
</tr>
<tr>
<td>8</td>
<td>575</td>
<td>5962</td>
</tr>
<tr>
<td>9</td>
<td>600</td>
<td>182758</td>
</tr>
<tr>
<td>10</td>
<td>625</td>
<td>170194</td>
</tr>
<tr>
<td>11</td>
<td>650</td>
<td>46997</td>
</tr>
<tr>
<td>12</td>
<td>675</td>
<td>5333</td>
</tr>
<tr>
<td>13</td>
<td>700</td>
<td>83</td>
</tr>
</tbody>
</table>

*Source: Compiled by Researcher.*

The study region is situated on Balaghat ranges and has minimum 415 and maximum 700 meters of heights. The study region has 551.15 meters (1808.23 feet) of average elevation. The above said heights are generated in contour lines/areas and stored in Latur district spatial database. The areas of specific heights, mean elevation, standard deviation of area and standard deviation of elevation are shown in Table No. 7.1. The process of generating contour lines/area using digital elevation models are performed in Quantum GIS 1.6. The vector map of elevation contour line / area of the study region are generated with 25 meters of intervals and are shown in below Map No. 7.1.
Map No.7.1: Elevation Contours of Latur District
E. Extraction of Water Bodies from DEM:

The slope map was derived using the DEM. The Quantum GIS package is used to extract the reservoirs or lakes of the study area using polygonization. The polygonization utility creates vector polygons for all connected regions of pixels in the raster sharing a common pixel value. Each polygon is created with an attribute indicating the pixel value of that polygon. A raster mask may also be provided to determine which pixels are eligible for processing. The utility will create the output vector data source if it does not already exist, defaulting to GML format. The extracted water bodies are compared with hydrological map toposheets for accuracy assessment. The slope map is shown in Map No. 7.2, and the extracted water body map for reservoirs/lakes of the study region is shown in following Map No. 7.3.

The entire region contains 08 major rivers namely Manjra, Manyad, Terna, Tavarja, Tiru, Lendi, Gharni and Deoni. The length of each individual river in the study region is Manjra 278 km, Manyad 95 km, Terna 122 km, Tavarja 47 km, Tiru 71 km, Lendi 63 km, Gharni 31 km, and Deoni 55 km. The reservoirs map stored into spatial database.

The total area covered by lakes/reservoirs in the region is 6262 hectares. The total average district area occupied by the reservoirs in the district is 0.86%. The location and flow of each individual river in the region is shown in Map No. 7.4.
Map No. 7.2: **Slope map of Latur district**
Map No. 7.3: Hydrological map (reservoirs/lakes) of Latur District
Map No.7.4: River map of Latur district
Map No. 7.5: Water Scarcity map of Latur district
F. Water Scarcity Locations:

Taking all the above processes and results into account, the elevation maps, slope map, reservoir map and river map are imported into spatial data base of the study region. Through examining the villages locations and processed results in the spatial database, the following Map No.7.5 shows the critical locations of water scarcity regions.

The Map No. 7.5 shows the critical regions of water scarcity. The yellow colored polygon used to showcase normal water scarcity regions, the orange color polygon visualized for critical water scarcity regions and the red color polygon shows the very critical water scarcity region in the study region.

The techniques with different methods for analysis of elevation contour lines, water body (reservoirs and lakes) extraction, and finding critical zones of water scarcity in the region. Because the continuous overexploitations of natural resources like land, water and forest have caused serious threat to the local population of the semiarid region. Thus, problems like little scope for, declining ground water level and shortage of drinking water prevail. So some step should be taken for planning and developments of those areas which are fell in the water scarcity zones. It's for e-governance.

7.2.3 Image Classification based on Satellite Imagery:

Image classification based on satellite imagery is a widely used technique for extracting thematic information on land cover. This image processing step is the translation from spectral reflectance or digital numbers (DN) to thematic information. We classify objects by
reducing a multiplicity of phenomena to a relatively small number of general classes (Tso and Mather, 2001). Classification is often performed to generalize a complex image into a relatively simple set of classes. A classified map is then used as input into a geographic information system (GIS) for further processing or analysis. Such inference is most often less than perfect and there is always an element of uncertainty in a classification result. As it can affect further processing steps and even decision making, it is important to understand, quantify and visualize the classification process.

Visual Data Mining (VDM) is a powerful tool which is often overlooked in favour of traditional purely non-visual data mining, defined as the process of (semi-)automatically discovering meaningful patterns in data (Witten, 2005). VDM uses visual interaction to allow a human user to visually extract and explore patterns in data.

When conducting a non-visual data mining, no matter how unbiased it may seem, the fact is that by simply choosing to carry out an automated analysis a priori assumptions have been made about what form the important results will take before analysis has actually begun (Simoff, 2002). By visually mining the data this prior bias can be removed. Whilst the bias is removed, subjectivity of the analysis is increased as it is based on a user’s perception, a point highlighted by many machine learning purists. However, this increased subjectivity is compensated for by a vastly increased degree of confidence in the analysis (Keim, 2002). VDM not only seeks to allow a human user to visually mine data but also to augment the non-visual data mining process.
This augmentation usually takes the form of making the automated process more transparent to the user, hence providing increased confidence. VDM is not commonly applied in remote sensing applications. A traditional supervised remote sensing classification starts with a selection of training pixels or areas that represent specific land cover classes. The spectral and statistical properties of these pixels are then used to classify all unlabelled pixels in the image with a classification algorithm such as the widely used maximum likelihood classifier (commonly implemented in commercial remote sensing software). The accuracy of the classified map is tested with reference pixels that are not used in the training stage. Accuracy assessment usually takes the form of an error matrix with derived accuracy values such as the overall accuracy and the Kappa statistic. Although the error matrix provides an overall assessment of classification accuracy, it does not provide an indication of the spectral dissimilarity of class clusters, uncertainty related to the attribution of class labels to individual pixels, or the spatial distribution of classification uncertainty. In this study, we argue that VDM is an important tool for visual exploration of the data to improve insight into the classification algorithm and identify sources of spatial and thematic uncertainty.

Recent studies showed that exploratory visualization tools can help to improve the image analyst's understanding of uncertainty in a classified image scene. They proposed a combination of static, dynamic and interactive visualizations for exploration of classification uncertainty in the classification result. Lucieer (2004) and Lucieer and Kraak (2004) developed a visualization tool that allowed for visual interaction with the parameters of a fuzzy
Figure No. 7.6: **Subset of IKONOS image of Latur district area showing the locations of ground reference samples as colored circles**
classification algorithm. The study showed that visualization of a fuzzy classification algorithm in a 3D feature space plot dynamically linked to a satellite image improves a user's understanding of the sources and locations of uncertainty. In this study, we develop and present a new VDM prototype to visualize irregular shapes of class clusters and their spectral overlap in a 3D feature space plot. The tool helps to identify the location and shape of class clusters (showing spectral variance) and the overlap of these class clusters in 3D feature space to highlight sources of uncertainty in the training data for a spectral image classifier. To showcase the visualization prototype we present a classification study based on high-resolution IKONOS imagery of Latur district to assess the value of VDM in semi-automated image classification. This study is limited to a pixel-based classification approach; however, the visualization tool can be used for object-oriented classification as well.

**Operating Environment:**

The prototype tool has been built in IDL (the Interactive Data Language) as an extension to ENVI (the Environment for Visualising Images) using the ITools (Intelligent Tools) framework (ITTVIS, 2006). ENVI provides various image analysis and manipulation features, such as Regions of Interest (ROI), which are used by the prototype to extract sample pixels from the image. The ITools framework provides the 3D engine for the feature space plots and the user interface for interaction with the 3D objects.
3D space feature plot:

When considering a classification problem it is useful to visualize the image data in a 2D or 3D feature space with selected image bands on each of the axes (similar to a scatter plot). This visualization provides important insight into both the patterns in the image data and the operation of classification algorithms. Most commercial remote sensing software offer the tools to visualize a 2D scatter plot. In this study, we extend these common plots to a 3rd dimension to increase the amount of information (image bands) in the visualization. To generalize the large amount of image pixels, 3D feature space can be internally represented by a volume allowing for visualization of the density of class clusters. This volume based representation divides feature space into cubes or voxels. Each voxel is represented by a density value: a count of how many pixels fall in the region of feature space generalized by this voxel. In this way the volume is a 3D frequency histogram with each voxel recording the frequency at which ranges of pixel values occur. The size of the volume, as specified by the user, determines the degree of generalization and the storage and processing requirements for operations on the volume.

Visualization of class clusters:

Creating an isosurface for each ROI and simultaneously visualizing these in a feature space plot offers two insights. Firstly, the user can examine each training cluster at varying levels of density. This is useful for traditional exploratory data analysis (EDA). Traditionally, visual EDA was used in data mining only as a means of
checking that data conformed to assumptions prior to analysis (Wegman, 2001).

For the maximum likelihood classification algorithm this means checking the training data for a normal distribution. Thematic classes in satellite imagery often do not conform to assumptions made by classification algorithms.

Secondly, the user may explore overlap between training clusters. This is another use of traditional EDA to check underlying assumptions. Many classifiers struggle to deal appropriately with overlapping training data introducing uncertainty in the classification result. It is important to visualize both of these phenomena prior to supervise classification in order to interpret the results such analyses.

Results

To showcase the visualization prototype we present a classification study based on high-resolution satellite imagery of Latur district area to assess the value of visualization in semi-automated image classification. The study is a simple 4 class problem with training regions as shown in Fig. No. 7.7 A random sample of 200 pixels was extracted from each training area for classification and a further 200 independently, randomly sampled pixels extracted for accuracy assessment. Visualization of the training regions is performed using all pixels in the regions.

Firstly, bands 4, 2 and 1 are selected to be used for classification, and hence visualization. The tool is configured to display each region as an isosurface. The result is shown in Fig. 7.7.
The shapes are colored according to the legend in Fig. No. 7.6 and Fig. No. 7.7 shows some possible overlap between the Rock (red) and Water (blue) classes.

The tool is used to compute this overlap and display it as a new isosurface. The tool is also used to identify the pixels causing this overlap and highlight their location in image space. The new volume produced by the intersection operation is shown as a yellow isosurface. The pixels from the Rock class causing this intersection are highlighted with a purple overlay, and those from the Water class with a yellow overlay. This visualization tool also has the ability to visualize decision boundaries and parameters.

This feature of the prototype is used to visualize the decision boundaries and parameters for the 3 class problem for the minimum distance and maximum likelihood classifiers. The mean is the only property of the data used by the minimum distance classifier. Pixels are classified according to the closest mean point in feature space. The visualization tool shows that some misclassification between the Water and Rock classes may be expected.

VDM is a useful technology for image classification. This study has showcased the added value that visualization can provide to analysis of satellite imagery. A novel volume based representation was used as a basis for visualization using isosurfaces. Isosurfaces and ellipsoids where used to construct 3D feature space plots showing the relationships between training regions and decision boundaries used during classification. Linkage of feature space visualizations and geographic space image views allowed a thorough investigation of patterns in the image data.
Fig. No. 7.7: **Feature space plot showing isosurface for 3 training regions** (X-axis = band4, Y-axis = band 2, Z-axis = band 1)
7.2.4 Visual Data Mining:

Visual Data Mining (VDM) is a powerful tool which is often overlooked in favour of traditional purely non-visual data mining, defined as the process of (semi-)automatically discovering meaningful patterns in data (Tso B. and Mather P.M, 2002). VDM uses visual interaction to allow a human user to visually extract and explore patterns in data. When conducting a non-visual data mining, no matter how unbiased it may seem, the fact is that by simply choosing to carry out an automated analysis a priori assumptions have been made about what form the important results will take before analysis has actually begun. By visually mining the data this prior bias can be removed. Whilst the bias is removed, subjectivity of the analysis is increased as it is based on a user’s perception, a point highlighted by many machine learning purists. However, this increased subjectivity is compensated for by a vastly increased degree of confidence in the analysis. VDM not only seeks to allow a human user to visually mine data but also to augment the non-visual data mining process. This augmentation usually takes the form of making the automated process more transparent to the user, hence providing increased confidence (GeoEye, 2006).

VDM is not commonly applied in remote sensing applications. A traditional supervised remote sensing classification starts with a selection of training pixels or areas that represent specific land cover classes. The spectral and statistical properties of these pixels are then used to classify all unlabelled pixels in the image with a classification algorithm such as the widely used maximum likelihood classifier (commonly implemented in commercial remote sensing software).
The accuracy of the classified map is tested with reference pixels that are not used in the training stage. Accuracy assessment usually takes the form of an error matrix with derived accuracy values such as the overall accuracy and the Kappa statistic. Although the error matrix provides an overall assessment of classification accuracy, it does not provide an indication of the spectral dissimilarity of class clusters, uncertainty related to the attribution of class labels to individual pixels, or the spatial distribution of classification uncertainty. In this study, we argue that VDM is an important tool for visual exploration of the data to improve insight into the classification algorithm and identify sources of spatial and thematic uncertainty. Recent studies showed that exploratory visualization tools can help to improve the image analyst's understanding of uncertainty in a classified image scene. They proposed a combination of static, dynamic and interactive visualizations for exploration of classification uncertainty in the classification result. Keim D. A. 2002 developed a visualization tool that allowed for visual interaction with the parameters of a fuzzy classification algorithm. The study showed that visualization of a fuzzy classification algorithm in a 3D feature space plot dynamically linked to a satellite image improves a user's understanding of the sources and locations of uncertainty. In this study, we develop and present a new VDM prototype to visualize irregular shapes of class clusters and their spectral overlap in a 3D feature space plot. The tool helps to identify the location and shape of class clusters (showing spectral variance) and the overlap of these class clusters in 3D feature space to highlight sources of uncertainty in the training data for a spectral image classifier. To showcase the visualization prototype we present a classification study based on high-resolution
LULC (Land Use/Land Cover) imagery of Latur district to assess the value of VDM in semi-automated image classification as shown in Map No. 7.6. This study is limited to a pixel-based classification approach; however, the visualization tool can be used for object-oriented classification as well.

**Visualization of Class Clusters**

Creating an isosurface for each ROI and simultaneously visualizing these in a feature space plot offers two insights. Firstly, the user can examine each training cluster at varying levels of density. This is useful for traditional exploratory data analysis (EDA). Traditionally, visual EDA was used in data mining only as a means of checking that data conformed to assumptions prior to analysis (Simoff S. J., 2001). For the maximum likelihood classification algorithm this means checking the training data for a normal distribution. Thematic classes in satellite imagery often do not conform to assumptions made by classification algorithms. Secondly, the user may explore overlap between training clusters. This is another use of traditional EDA to check underlying assumptions. Many classifiers struggle to deal appropriately with overlapping training data introducing uncertainty in the classification result. It is important to visualize both of these phenomena prior to supervise classification in order to interpret the results such analyses.

To showcase the visualization prototype we present a classification study based on high-resolution LULC imagery of Latur district area to assess the value of visualization in semi-automated image classification. The study is a simple 5 class problem with training regions as shown in Map No.7.6 A random sample of 200
pixels was extracted from each training area for classification and a further 200 independently, randomly sampled pixels extracted for accuracy assessment. Visualization of the training regions is performed using all pixels in the regions. Firstly, bands 3, 2 and 1 are selected to be used for classification, and hence visualization. The tool is configured to display each region as an isosurface. The result is shown in Fig. No. 7.10. The shapes are overlap between the Rock (red) and Water (blue) classes. The tool is used to compute this overlap and display it as a new isosurface. The tool is also used to identify the pixels causing this overlap and highlight their location in image space. The new volume produced by the intersection operation is shown as a yellow isosurface. The pixels from the Rock class causing this intersection are highlighted with a purple overlay, and those from the Water class with a yellow overlay.

Table No. 7.2: LULC classified values of 5 area classes

<table>
<thead>
<tr>
<th>Classes</th>
<th>Area (%)</th>
<th>Red</th>
<th>Green</th>
<th>Blue</th>
<th>Frequency (Pixels)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constructed</td>
<td>0.379638</td>
<td>239</td>
<td>71</td>
<td>62</td>
<td>7932</td>
</tr>
<tr>
<td>Water</td>
<td>1.313144</td>
<td>115</td>
<td>146</td>
<td>206</td>
<td>27489</td>
</tr>
<tr>
<td>Rocky / Barren</td>
<td>12.63076</td>
<td>245</td>
<td>240</td>
<td>172</td>
<td>264383</td>
</tr>
<tr>
<td>Vegetation</td>
<td>28.1574</td>
<td>170</td>
<td>195</td>
<td>58</td>
<td>589333</td>
</tr>
<tr>
<td>Agriculture</td>
<td>57.55275</td>
<td>248</td>
<td>250</td>
<td>129</td>
<td>1204596</td>
</tr>
</tbody>
</table>

Source: Compiled by Researcher.
Map No. 7.6: LULC image of Latur district area

(Courtesy NRSC (ISRO), Hyderabad)
Fig. No. 7.8: **Feature space plot showing is surface for 5 training regions**
Fig. No. 7.9: Color histogram of LULC

Count: 3838632
rMean: 122.43 rSD: NaN rMode: 0
gMean: 126.64 gSD: NaN gMode: 0
bMean: 63.16 bSD: 65.55 bMode: 0
Fig. No. 7.10: Isosurface of classified regions of Latur LULC (5 classes)
This visualization tool also has the ability to visualize decision boundaries and parameters. This feature of the prototype is used to visualize the decision boundaries and parameters for the 5 class problem for the minimum distance and maximum likelihood classifiers. The mean is the only property of the data used by the minimum distance classifier.

Pixels are classified according to the closest mean point in feature space. The visualization tool shows that some classification between the Water, constructed area, and vegetation, agriculture and rocky/barren area classes in the form of isosurface. The table no. 7.2 showing the classified values of 5 different classes.

VDM is a useful technology for image classification. This study has showcased the added value that visualization can provide to analysis of satellite (LULC) imagery. A novel volume based representation was used as a basis for visualization using isosurfaces. Isosurfaces and ellipsoids where used to construct 3D feature space plots showing the relationships between training regions and decision boundaries used during classification. Linkage of feature space visualizations and geographic space image views allowed a thorough investigation of patterns in the image data. This is a very easiest and fastest method to classify LULC imageries for land assessment, analysis, change detection of area, planning and development, etc.