Assessment and Comparison of Classification Techniques for Forest Inventory Estimation: A Case Study using IRS-ID Imagery

Mukherjee, S.,[1] and Mukherjee, P.,[2]
Remote Sensing and GIS Application Lab, School of Environmental Sciences, Jawaharlal Nehru University
New Delhi – 100067, India, E-mail: dr.saumitramukherjee@usa.net[1], prabir@yahoo.com[2]

Abstract
Traditional satellite image classifications are mostly confined to supervised, unsupervised and hybrid methods. An alternative to these approaches, subpixel classification is gradually changing the concept of image classification. This technique is advantageous particularly for medium to low resolution satellite image, by removing the influence of associated features within a pixel. The Spectral Mixture Analysis (SMA), is a typical subpixel classifier, and is applied here in forest classification of Gurdaspur, Hosiarpur and Rupnagar districts of Punjab, India to explore its authenticity and accuracy. A special effort was made to accurately calculate the forest inventory in Punjab using the SMA, and the results were compared with the data obtained through traditional classification methods. This technique enables estimation of proportional forest type in a single pixel and may be used to estimate various aspects of forest vegetation important for different forest modeling (forest growth, forest yield etc.), carbon budgeting and decision making.

1. Introduction

Extraction of accurate and detailed land use/ land cover (LULC) information from remotely sensed images still continues to be a challenge. A wide range of alternative approaches such as image slicing, segmentation and multispectral image classification including supervised and unsupervised techniques have been explored for land cover mapping. Multispectral image classification is considered as one of the most efficient methods for identification of land use/cover classes (Ketting et al., 1976). In image classification, pixels of an image are assigned to different classes based on their spectral properties. By comparing unclassified pixels to those of known identity, it is possible to assemble groups of similar pixels into classes that match the predefined categories. Basically this classification is based on information derived from "training areas" selected through ground truthing. This has certain disadvantages in terms of cost and time requirements for training area selection. Particularly in case of forest classification, selection of training sites in remote areas is a major challenge. In this regard, desktop-based advanced technique of image classification with increased accuracy is advantageous. Since forest mapping is generally carried out at regional scale, medium (Landsat, IRS-IC/ID, SPOT, Aster etc.) to low resolution (MODIS, NOAA, SPOT VEGETATION Sensor, WiFS etc.) satellite images are more economical than high-resolution images. However, images with coarser resolution lead to greater inaccuracy. This is more so when a pixel is occupied by more than one land cover classes. Hence the reflectance measured by the sensor can be treated as a sum of interactions among various classes present within a pixel as weighted by their relative proportions (Strahler et al., 1986). Mixed-pixel problem affects the effective use of remote sensing data for LULC classification (Fisher, 1997 and Jaisawal et al., 2002). Almost in all the hard classification techniques, an entire pixel is assigned to a specific land-cover class on the basis of the classification scheme. Unsupervised classification procedure makes groups or clusters of multispectral values of the image into distinct classes (e.g. sparse forest, dense forest, bare soil) based purely on the image statistics, and generates a new raster map displaying the class designations within the image (Hall, 1994). This process does not require the user to feed any information about the features contained within the image. The objective is to group multiband spectral response patterns into clusters that are statistically separable. The two most frequently used grouping algorithms are K-means and the ISODATA clustering algorithms. The present study applies ISODATA algorithm while using unsupervised classification technique. Supervised classifications involve the use of ground truth data and train the computer to group pixels into clusters and to put them into certain land-cover classes (Hall, 1994). Various algorithms (e.g.

maximum likelihood, nearest neighbor etc) can be employed to classify pixels not falling within training areas (Cambell, 1996). The technique of supervised classification considers the statistics of every pixel in all the bands, and calculates the probability that a given pixel belongs to a specific class (Settle and Drake, 1993). In absence of any specific probability threshold, all pixels are classified or assigned to a specific class. Therefore, each pixel is assigned to the class to which it has the highest probability (i.e. "maximum likelihood") of belonging. Supervised and unsupervised classifiers cannot effectively solve the mixed-pixel problem of complex landscape. On the other hand, most vegetation indices are derived through band ratioing. Chlorophyll pigment of vegetation is known to absorb visible light, whereas leaf's mesophyll tissue strongly reflects near infrared light (Cambell, 1996). As the amount of vegetation increases within a pixel, surface reflectance in the visible-red decreases and in the near infrared increases (Hall, 1994). Vegetation indices such as the Normalized Difference Vegetation Index (NDVI), takes the advantage of these properties. The NDVI can provide a measure of the vegetation content within each pixel. By employing the characteristic spectral response of vegetation, this index allows the analyst to develop a measure of one component of each pixel. An alternative approach to solve the mixed-pixel problem is the Spectral Mixture Analysis (SMA) technique as sub-pixel classifier. This method recognizes that a single pixel is typically made up of a number of varied spectral types (i.e. soil, water, vegetation etc.) (Atkinson et al., 1997), and determines the class of the pixel based on the relative percentage of occupancy by different land-cover classes. The SMA technique reduces the cost of ground truthing and improves classification accuracy. In the present study, the SMA has been used on multispectral IRS-ID data to differentiate mixed-pixels irrespective of the percentage of their occupancy by different land-cover classes. The present method was tested over forest cover mapping. Ideally the classification of image will give the best result if one pixel represents one type of land cover. In reality however, land-cover surfaces are often composed of a mixture of materials and the separation between them is not easy particularly in remotely sensed images. When a sensor scans land cover units with dimensions smaller than the spatial resolution of the sensor, the pixel will bear the mixed signatures of the land covers and the reflectance spectra produced will not match any of the pure spectra of the materials present within a pixel.

The objective of SMA is to identify primary spectral contributions within each pixel (Adams et al., 1995). The SMA allows decomposing each pixel into the percentage of the pixel that is represented by the major land cover classes that can be derived from the image. A spectral mixture model is a physically based model in which a mixed spectrum is modeled a combination of 'pure' spectra, called endmembers (Adams et al., 1995). Linear SMA is the process of solving for endmember fractions, assuming that the spectrum in each pixel on the image represents a linear combination of endmember spectra that corresponds to the physical mixture of some components on the surface, weighted by surface abundance. In this way, a profile of constituent parts of each pixel is created and by aggregating those values, the percentage of earth's surface covered by a particular class is determined more accurately. Nevertheless, the SMA is a useful method of image classification; particularly for defining proportions of land cover types in pixels for coarser resolution satellite imagery. While conventional image classification matches pixels to broad classes of features, the SMA attempts to identify surfaces from their spectral data much more precisely than it was done previously, particularly for medium resolution satellite sensors. This is probably the first attempt to classify the forest inventory with medium resolution IRS-ID satellite data.

2. Study Area

The present study has been carried out in the north, northeast and eastern borders of Punjab state of India, consisting of Gurdaspur, Hosiarpur and Rupnagar districts (Figure 1). The area is confined within 30° 33 54″ to 32° 30 16″ N latitudes and 74° 52 53″ to 76° 50 41″ E longitudes, covering 9050 km^2 areas. Gurdaspur, the northern most district of Punjab is confined between rivers Ravi and Beas with latitude 31° 35 18″ to 32° 30 16″ N and longitude 74° 52 53″ to 75° 56 13″ E. Hosiarpur district is located in the north-eastern part of the state. The district is sub-mountainous and stretched along the river Beas to the north-west with latitude 30° 58 00″ to 32° 04 16″ N and longitude 75° 28 52″ to 76° 31 02″ E. Rupnagar district located within 30° 33 54″ and 31° 25 40″ N latitudes and 76° 17 50″ and 76° 50 41″ E longitudes, occupies the eastern part of the state, bordering the Satluj river (2 to 5 km). Topographically hills and valleys characterize the region with forest dominating in hills and agriculture in the plains. The Himalaya in the north and the Thar Desert in south and southwest influence rainfall and temperature in this region.

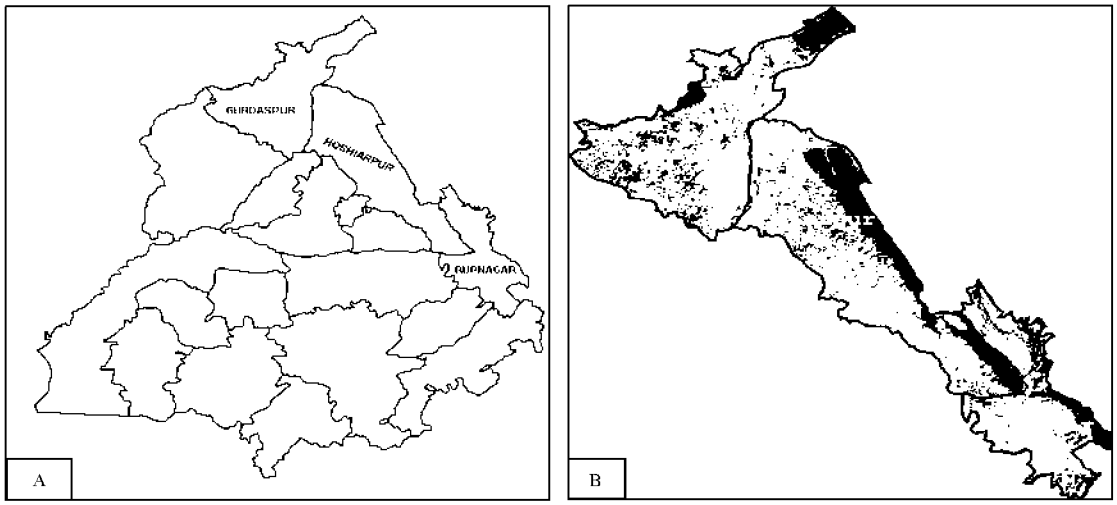

Figure 1: A) District maps of Punjab highlighting the study area B) False Color Composite (FCC) changed to grey scale of IRS-1D LISS III Satellite Image.

The mean annual rainfall varies from 300 mm to 1400 mm. Rainfall gradually increases from the southeast towards the northwest. More than 90 % of annual rainfall occurs during the monsoon season (early July to early October). While the rainfall regime in this region is unimodal, the annual cropping cycle is mostly bimodal. The rainy-season crop (Khariff crop) is directly supported by rainwater while the winter crop (Rabi crop) is cultivated with the aid of irrigation water. Depending upon water availability, summer crop is cultivated in some places within the region. Cotton and sugarcane are the main Khariff crops, while wheat is the main Rabi crop. The summer crops include vegetables and cereals. Natural vegetation is dominated by open forest and shrubs, and is denser in the hills.

3. Data

In this investigation Indian Remote Sensing satellite image of IRS-ID LISS III, of October-November, 2003 have been used. The IRS satellite images consist of three bands in the visible and near infrared (VNIR) range, including B2: green (0.520-0.590 μm), B3: red (0.620-0.690 μm), and B4: near infrared (0.770-0.860 μm) with 23.5-meter spatial resolution, and one band (B5) in short wave infrared (SWIR – 1.50-1.70 μm) with 70.5 meter spatial resolution. In the present study, only first three bands (B2, B3, and B4) have been used and the SWIR band is ignored due to its low resolution. Digital topographic maps of 1:50,000 scales were used for selection of Ground Control Points (GCP) and also to evaluate precision of geometric correction. The IRS-ID LISS III images were corrected by 40 GCP using 2nd degree polynomials (Root Mean Square Error = 0.25 pixel). The pixels were resampled by the nearest neighbor method to maintain their originality.

3. Methods

The image was digitally processed using PC based ERDAS Imagine and ENVI software. In this investigation both traditional (viz. supervised, unsupervised and vegetation index based classification) and the advanced SMA techniques were performed using IRS-ID LISS III satellite image to assess the potential advantages of the SMA over the standard methods. For a comparative analysis of different methods, a common input was used for all the studies. Since in this investigation only 3 bands of IRS-ID LISS III multispectral image were selected, the maximum number of endmembers for running the SMA model could be three. However, the study area was found to be occupied by more than three endmembers. To remove the complexity of mixing among these members, two distinct land cover classes (agricultural standing crops and water bodies) were removed from the study area based on their distinct spectral properties. Water shows lowest reflectance in the NIR band compared to green and red bands and agricultural crops show a typical pinkish-red shade in False Color Composite (FCC) image. Due to its similarity, the built-up class was merged with open bare soil. After merging and masking out the above mentioned classes, the study area was left with mostly three land-cover classes: (1) dense forest (2) sparse forest and (3) open bare soil, which were selected as the endmembers. The same resultant image was used for traditional approaches (like supervised, unsupervised and vegetation index

based). However, for regions with agricultural land or any other land-cover classes bordering forest lands, endmembers should be selected accordingly. An attempt was made to map the forest areas and its two types namely dense forest (crown cover is 40% or more) and sparse forest (crown cover is within 10-40%) apart from the open bare soil. The accuracy assessment of the output data was done in two ways: (1) comparing the forest areas with the published statistical data by Forest Survey of India, State of Forest Report -2003, and (2) using the standard error matrix using the random sampling data (total 30) taken from high resolution satellite image and limited field survey. Ground truthing was carried out by assessment of relative NDVI values and limited ground truthing by using ISRO make 8 band ground truth radiometer. Overall accuracy, user's accuracy and producer's accuracy (Congalton and Mead, 1983) were calculated based on the error matrix for all four-classification results.

3.1 Supervised Classification

The classification of IRS-ID LISS III multispectral satellite image was carried out using maximum likelihood classifier. The following basic sequences of operations were used to perform supervised classification.

1. Defining the training sites
2. Extraction of signature
3. Final image classification

In the first step of supervised classification the training sites were selected from the areas within the forest (dense and sparse forest) and from surrounding open/bare areas.

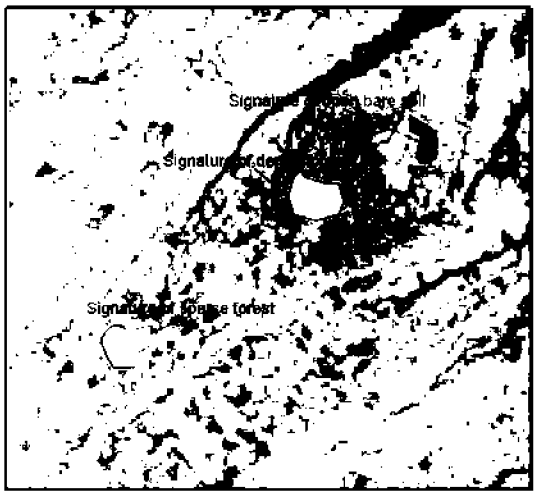

Figure 2: Sample training areas for supervised classification of the study area

This is generally done by on screen digitization of multispectral images. In this investigation, a total of 30 training sites were selected, with 10 sites in each of dense, sparse and open areas. The training sites were gathered from reference data sources including existing topographic maps, and high-resolution satellite images. The training sites were assumed to represent pixels of known identity and covered rather homogenous regions of land cover (Figure 2). Training sites were linearly stretched into their respective classes. Selected training sites went through statistical characterization of information. These were basically signatures for individual class. The signature file contains a variety of information about the land cover classes. The final step was image classification with the maximum likelihood classifier. Supervised classification started with computing statistics for the selected training sites of land-cover classes and the results of the statistical summary were used to classify the image. The final image was classified into the required three classes a) dense forest b) sparse forest and c) open bare soil.

3.2 Unsupervised Classification

Unsupervised classification (ISODATA) works on the principal of clustering of pixels of similar spectral properties. This method requires determination of the input spectral bands, desired number of output clusters and the cluster threshold radius. Firstly, the number of spectral classes for the image was decided. Following the general rule of 20 spectral classes for each land cover class in an image, the image was classified into 60 spectral classes. The advantage of obtaining a large number of spectral classes was to improve the ability to distinguish differences in the spectral appearance of single land cover class. An initial unsupervised classification was performed using the ERDAS Imagine package with intent of separating the forest areas from surrounding land cover classes. Image containing band 2, 3, and 4 are processed using ISODATA unsupervised classification algorithm and forced into 60 initial classes. Classes were manually labeled into three categories (dense forest, sparse forest and open bare soil) same as was derived using supervised classification techniques. The labeling process was accomplished primarily by analyzing the locations of pixels in each spectral class and determining which land-cover class is most likely represented.

3.3 Normalized Difference Vegetation Index (NDVI)

The NDVI map was created by the ratio of (Band4 – Band3) and (Band4 + Band3).

$$NDVI = \frac{Band4 - Band3}{Band4 + Band3}$$

Equation 1

The NDVI model was run on the IRS-ID satellite image. The output was sliced based on the NDVI values (Figure 3). Thresholds for dense and sparse forest have been decided based on the NDVI values, and the final forest map was generated.

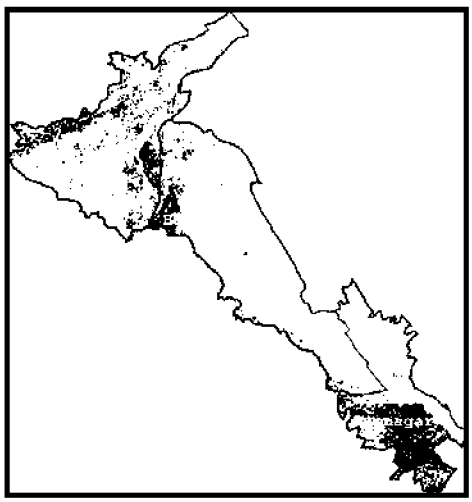

Figure 3: Normalized Difference Vegetation Index (NDVI) map of the study area

3.4 Spectral Mixture Analysis (SMA)

The procedure used in this study was based on a linear mixture model to derive continuous fields of (1) dense forest, (2) sparse forest, and (3) open bare soil. The SMA was performed on the three bands (Band 2, band 3 and Band 4 i.e. VNIR bands) of the IRS-ID multispectral image. The fourth band (SWIR), which is of 70.5 m spatial resolution, was not considered for this study. Commercially available ENVI software package was used for this SMA functionality with the name of linear unmixing model. Since the visible bands are highly correlated between the adjacent spectral wavebands (Barnsley, 1999), Principal Component Analysis (PCA) was run to transform the data from highly correlated bands to an orthogonal subset. The steps involved in the SMA process have been described in the flow chart (Figure 4). The training data selected for the SMA were same as used for training sites in case of supervised classification to maintain the same base information for forest mapping. The idealized pure signature for a class from these training sites is called an endmember. A variety of methods has been developed to determine the endmembers. Endmembers can be obtained for manual selection through (1) a spectral library (2) the image itself or high-order PCA eigenvectors (Boardman, 1993); (3) spectrally pure pixel identification using Pixel Purity Index (PPI) (Boardman et al., 1995) etc.

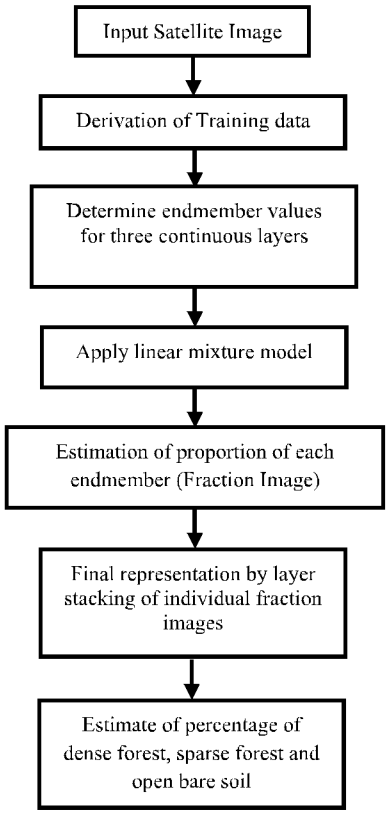

Figure 4: Flow diagram of the SMA process

In this study the endmembers were selected from the image without using any spectral library. Scatter plots of the 3 bands helped in locating the three purest endmembers by taking the extreme corner pixels. In two dimensions, if only two endmembers mix, then the mixed pixels fall in a line in the histogram. The pure endmembers fall at the two ends of the mixing line. If three endmembers mix, then the mixed pixels fall inside a triangle and pure pixels are concentrated at the three vertices (Figure 5). Because of sensor-noise and within-class signature variability, endmembers only exist as a conceptual convenience and as idealizations in real images. The linear mixture model in the SMA approach assumes that the spectrum measured by a sensor is a linear combination of the spectra of all components within a pixel (Adams et al., 1995). Linear unmixing was performed using the endmembers obtained, and keeping the unit sum constraint as 1. The linear mixture model can be mathematically described in a linear vector-matrix equation:

$$DN_i = \sum_{k=1}^{n} (R_{ik} x F_k) + E$$

Equation 2

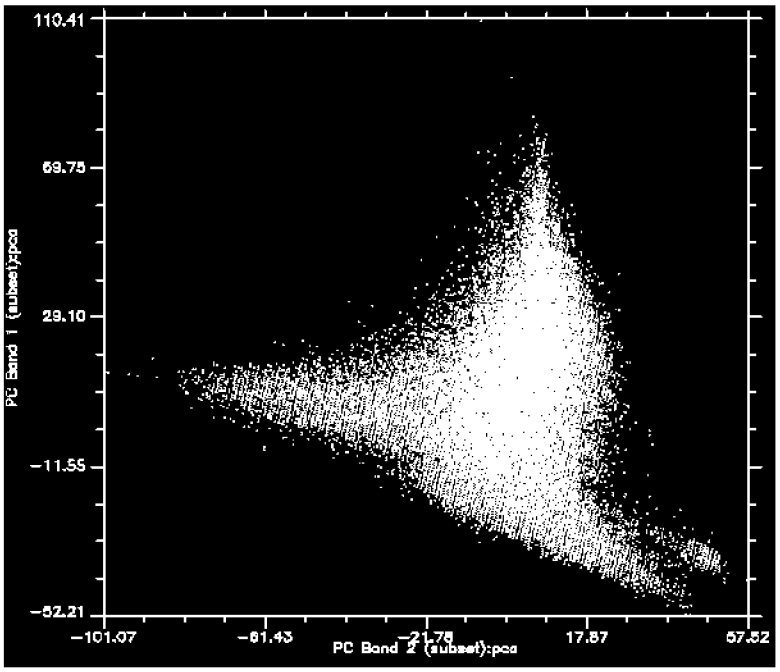

Figure 5: Scatter plots of PC Band 1 and PC Band 2

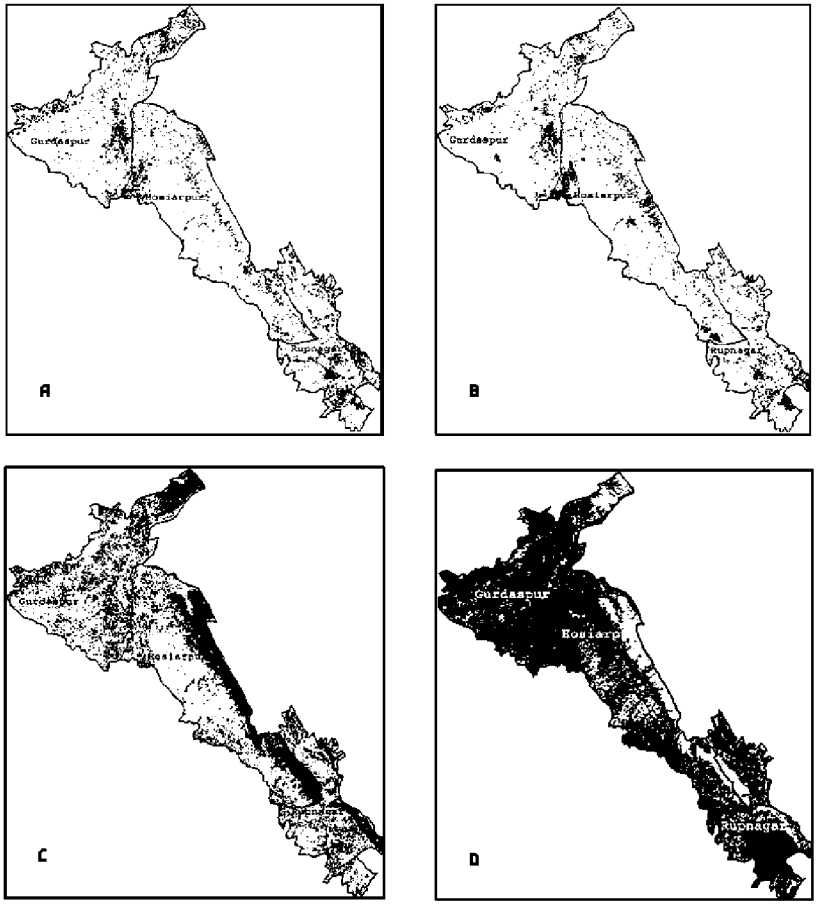

Figure 6: Fraction images of A) dense forest, B) sparse forest, C) open bare soil, D) Color composite changed to grey scale of the fraction images for the three endmembers. Open bare soil as very dark tone, sparse forest as less dark tone and dense forest as very light tone.

Where, i= 1,……, m (number of bands)
k= 1,……, n (number of endmembers)
DN_i = Spectral reflectance of the i^{th} spectral band of a pixel
R_{ik} = known spectral reflectance of k^{th} component
F_k = the fraction coefficient of the k^{th} component within the pixel
E_i = error for the i^{th} spectral band

The satellite image used here is of three bands (excluded the 4^{th} band due to its low resolution). When the linear mixture model was applied to this 3-band image to estimate the combination of three endmembers (x = dense forest, y = sparse forest and z = open bare soil), the mixture model becomes:

$$DN = [(R_{x1} \times F_x) + (R_{y1} \times F_y) + (R_{z1} \times F_z)] + [(R_{x2} \times F_x) + (R_{y2} \times F_y) + (R_{z2} \times F_z)] + [(R_{x3} \times F_x) + (R_{y3} \times F_y) + (R_{z3} \times F_z)] \quad 3 \text{ bands}$$
$$x \qquad y \qquad z$$

Where, DN is the spectral reflectance of a pixel in IRS-ID 3-band composite image, R_{ik} is the known spectral reflectance or endmember values for dense forest, sparse forest and open bare soil. The three bands of IRS-ID scene are represented by the parameter *i* and each of the three endmember is represented by factor *k*. F_k is the fraction coefficient of the k^{th} component within the pixel or the fractional cover for dense forest, sparse forest and open bare soil. The fraction image corresponding to each endmember was generated (Figure 6) and color composite was prepared of the fraction endmembers. The fractions represent the areal proportions of the endmembers within a pixel. The bright pixels of the image corresponds the higher fraction of that particular endmember within the pixel. The proportion of the endmembers in the fraction image ranges from 0 to 1. Where 0 indicates absence of the endmember and increasing value shows higher abundance with 1 representing 100% presence of the endmember in a particular pixel. The fraction composition of individual endmembers can be studied in Figure 7. The open bare soil pixel has a higher fraction of open bare soil class and a lower fraction of sparse forest and least fraction of dense forest. The sparse forest pixel has very low fraction of open bare soil and comparatively lower fraction of dense forest. Similarly, dense forest pixel has very higher fraction of dense forest and a lower fraction of sparse forest and lowest fraction of open bare soil.The SMA allows decomposing each pixel into the percentage of the pixel represented by the major land-cover classes that can be derived from the image. In order to compare the output of the SMA with published data (by Forest Survey of India, 2003) for dense and sparse forest, firstly the SMA fractional image was converted into the percent covered by the three endmembers for individual pixel. The fractional files contain the estimates of the fractional coverage associated with each endmember and the values within 0 to 100 could be interpreted as percentages of that particular endmember. In this way the profile of each pixel was created of its constituents parts and by aggregating those values, the aerial extent of individual land cover class within the study area was calculated.

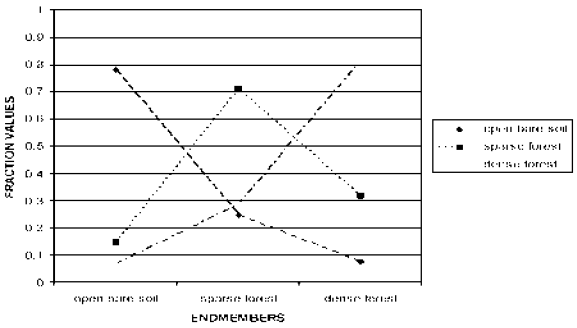

Figure 7: Comparative study of fraction features among three endmember (open bare soil, sparse forest and dense forest) within a pixel.

4. Results

The outputs obtained from the four methods (Figure 8) as explained above, were compared with the State of Forest Report, 2003 published by the Forest Survey of India. The area statistics for individual forest types in three districts of Punjab are available in Forest Survey of India site (http://www.fsiorg.net /fsi2003/states/index.asp?state_code=22).The utputs were also compared with the limited field sampling data collected across the three districts during the study. The overall statistical data (Tables 1-4 and Figure 9) reveals that the output derived through the SMA technique has the highest correlation with the data published in the State of Forest Report 2003 by the Forest Survey of India. The next best correlation has been found with the output of supervised approach followed by NDVI and the unsupervised approach. The error matrix computation (Table 5) using four different classification approaches detect the superior performance of the SMA over other three approaches. The overall accuracy of the classified map determined to be 90 % using the SMA approach over the supervised classification (76.67 % accuracy), the NDVI approach (63.33 % accuracy) and the unsupervised approach (53.33 % accuracy).The result complements the SMA technique in mapping the forest cover along with its

composition and areal estimates due to its ability to produce fractions representative of subpixel components directly related to forest type and relative area. Although the analysis used a spectral unmixing technique based on the assumptions of the linear unmixing model, the mixture proportion and areal estimates that were derived, matches well with the published state forest report. The results were also validated with limited field survey across three districts. Although the field survey was supported by the use of ground truth radiometer data also.

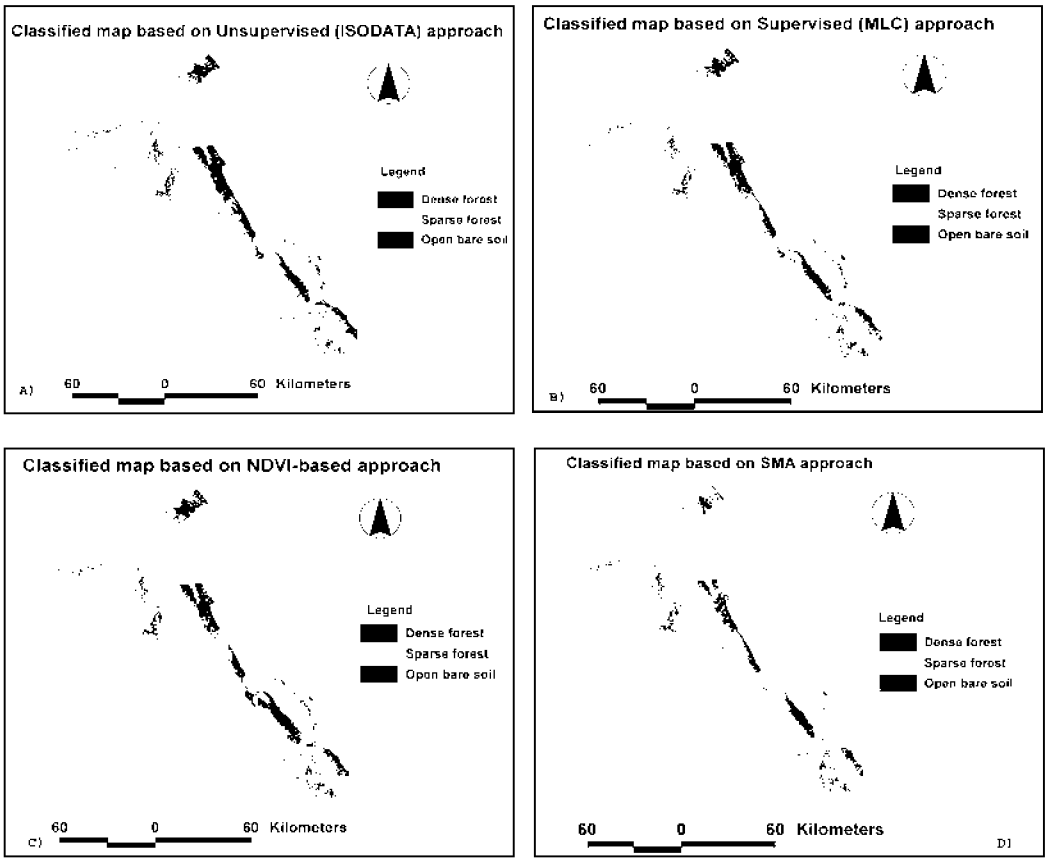

Figure 8: Classified output of four approaches A) Unsupervised approach B) Supervised approach C) NDVI approach D) SMA approach

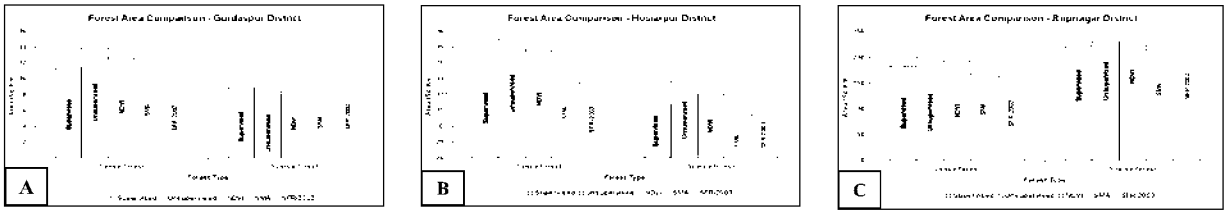

Figure 9: Graphical representation of outputs obtained through different models for three districts of Punjab (A: Gurdaspur District B: Hosiarpur District C: Rupnagar District)

Table 1: Areal estimation of forest types from supervised classification

District Name	Forest Cover Type	No. of Pixels	Area (Km^2) using Supervised classification	Area (Km^2) in State of Forest Report - 2003
Gurdaspur	Dense forest	180579	112.86	96
	Sparse forest	142520	89.07	96
Hosiarpur	Dense forest	545185	340.74	327
	Sparse forest	502584	314.11	307
Rupnagar	Dense forest	292322	182.70	162
	Sparse forest	353593	220.99	212

Table 2: Areal estimation of forest types from unsupervised classification

District Name	Forest Cover Type	No. of Pixels	Area (Km^2) using Unsupervised classification	Area (Km^2) in State of Forest Report - 2003
Gurdaspur	Dense forest	223547	139.71	96
	Sparse forest	127554	79.72	96
Hosiarpur	Dense forest	568844	355.53	327
	Sparse forest	525784	328.61	307
Rupnagar	Dense forest	321249	200.78	162
	Sparse forest	369025	230.64	212

Table 3: Areal estimation of forest types from Normalized Difference Vegetation Index (NDVI)

District Name	Forest Cover Type	No. of Pixels	Area (Km^2) using NDVI Model	Area (Km^2) in State of Forest Report - 2003
Gurdaspur	Dense forest	202159	126.35	96
	Sparse forest	135167	84.48	96
Hosiarpur	Dense forest	556574	347.86	327
	Sparse forest	512245	320.15	307
Rupnagar	Dense forest	308068	192.54	162
	Sparse forest	361581	225.99	212

Table 4: Areal estimation of forest types from Spectral Mixture Analysis (SMA)

District Name	Forest Cover Type	No. of Pixels	Area (Km^2) using SMA Model	Area (Km^2) in State of Forest Report - 2003
Gurdaspur	Dense forest	167463	104.66	96
	Sparse forest	150580	94.11	96
Hosiarpur	Dense forest	528479	330.30	327
	Sparse forest	494634	309.15	307
Rupnagar	Dense forest	268603	167.88	162
	Sparse forest	342470	214.04	212

Table 5: Comparison of results and assessment of Classification Accuracy amongst the SMA, Supervised, NDVI and Unsupervised approaches

Method	Classified Data	Reference Data			Ref. Totals	Class Totals	Number Correct	Producer's Accuracy	User's Accuracy
		Dense forest	Sparse forest	Open bare soil					
SMA	Dense forest	10	0	0	12	10	10	83.33%	100%
	Sparse forest	2	8	0	9	10	8	88.88%	80%
	Open bare soil	0	1	9	9	10	9	100.00%	90%
	Overall Classification Accuracy = 90% (i.e. 27/30)								
Supervised Classification	Dense forest	8	2	0	10	10	8	80.00%	80.00%
	Sparse forest	2	7	1	11	10	7	63.64%	70.00%
	Open bare soil	0	2	8	9	10	8	88.89%	80.00%
	Overall Classification Accuracy = 76.67% (i.e. 23/30)								
NDVI	Dense forest	6	2	2	10	10	6	60.00%	60.00%
	Sparse forest	3	6	1	10	10	6	60.00%	60.00%
	Open bare soil	1	2	7	10	10	7	70.00%	70.00%
	Overall Classification Accuracy = 63.33% (i.e. 19/30)								
Unsupervised Classification	Dense forest	5	3	2	9	10	5	55.56%	50.00%
	Sparse forest	3	5	2	11	10	5	45.45%	50.00%
	Open bare soil	1	3	6	10	10	6	60.00%	60.00%
	Overall Classification Accuracy = 53.33% (i.e. 16/30)								

5. Discussion

Forest is one of the most important and precious components in natural resources. Because of the many benefits that can be gained from forests, public and political interest is directed towards sustainable management and development of it. Planning for sustainable management of forests requires accurate information about forest resources such as area and nature of forests and types of plants. Mapping of forest in large areas is not easy through field survey or by means of aerial photo interpretation. However, satellite data with their advantages such as synoptic view, revisit frequency, constant spatial resolution, enhanced spectral resolution and automatic analysis capability have created a high potential in forest classification and

forest-resources estimation. The present study explored the usefulness and relative efficiency of different standard classification techniques in forest classification using IRS satellite image. The study also assessed the relative advantage of the Spectral Mixture Analysis (SMA) model over the standard approaches. From the previous studies it was understood that the SMA model runs better on satellite images with higher number of bands and a few studies have already been executed with Landsat ETM+ satellite images. But present research also aimed to know the applicability of the SMA with limited band satellite image like IRS LISS III. As described earlier, data of only three comparable bands – green (0.52-0.59 μm), red (0.62-0.68 μm), and near infrared (0.77-0.86 μm) were used for digital analysis. The short-wave infrared band (1.50-1.70μm) was not included in the SMA model due to mis-registration with the other three spectral bands of LISS III data. Further more, to make the image more simplified and to reduce the influence of greater number of features we removed the spectrally distinguishable features like agricultural lands and water bodies from the image. The resultant masked image was used for further digital processing. The SMA is a material-based promising image analysis process that allows extraction of subpixel level accurate quantitative information. Classification of image using pixel-level classification leads to inaccuracy due to the presence of a number of land-cover classes within a single pixel. Using the SMA approach, the spectral variability in a multispectral image can be modeled by mixtures of a small number of surface materials with distinct reflectance spectra (endmember). Unlike supervised, unsupervised and NDVI based image classification, the SMA did not follow the pixel-wise clustering of similar spectral signature. Rather, it was able to consider each pixel individually and assess the presence of proportion of selected endmembers. Selection of endmember is a vital component in running the SMA model. Pure endmember selection is often difficult and the process passes through a number of iterations till the independent endmembers are selected. Within the present study the endmembers were selected manually based on Principal Component Analysis (Bateson and Curtiss, 1996). Since we selected three endmembers (dense forest, sparse forest and open bare soil), the scatter plots of the bands helped in locating the purest endmembers by taking the extreme corner pixels of the triangle. These scatter plots were generated in iterative manner to get the purest pixels within the vertex zone of the triangle. This ultimately helped in generating the high-quality fraction images. The SMA produced fraction images that were pixel-by-pixel measures of the percent composition for each endmember in the spectral mixing model. The results described above showed that the SMA technique was able to generate more accurate areal estimates of the endmember classes, matching the field verified data with higher accuracy. For field verification the pixel values of satellite data were matched with the values of ground truth radiometer data of same co-ordinate. Since supervised, unsupervised and NDVI based methods were based on classification of entire pixels through predefined classification schemes, they caused a "rounding-off error", often producing too high or low estimates of land cover classes due to the inability to distinguish at subpixel level. The SMA technique proved to generate higher accuracy in forest classification and provided a more realistic areal estimate of forest type rather than a patchy output of traditional image classification methods. There is a potential to improve the output by selecting more accurate training samples for endmembers. The results could be validated using higher number of field verified information. Although in this research a liner model has been considered for the spectral mixture, it may not follow the linear relation always, and in that case a non-linear model has to be framed.

6. Conclusions

Conventional remote sensing classification methodology, as generally applied in forestry is based on qualitative analysis of information derived from "training areas" (ground-truth) or through clustering of pixels with similar spectral signature. This has certain disadvantages in terms of time and cost required for training area establishment, as well as to ensure higher accuracy. Unlike the conventional qualitative approach, the SMA follows the physical based classification at subpixel level and this approach is readily applied and replicated by others working in distant comparable environmental regions. Furthermore, by reporting endmember characteristics and locations and applying the SMA to generate endmember fraction images, the researcher can define classifications grounded on physical measure of the Earth's surface (i.e. reflectance) and based on the types and amount of materials present. Results indicate that the SMA approach in mapping forest composition and the corresponding areal estimates is one of the best classification methods due to its ability to produce fractions representative of subpixel components directly related to forest tree type and relative area. Although the analysis used a spectral unmixing technique based on the assumptions of the linear mixture model, the mixture proportions and area

estimates that were collected matched well with the available field information and the published result of the Forest Survey of India. This suggests that the SMA technique based on the linear mixture model is efficient for vegetation mapping. Furthermore, areal estimates from the SMA had higher accuracy when compared with the results from traditional supervised, unsupervised and NDVI based classification approaches that were used to discrete classification schemes. Accuracy could be further increased through selection of pure training sites and the result can be more accurately validated with sufficient number of field verification by increasing the frequency of ground truthing by using ground truth radiometer. This study gave emphasis on evaluating the potentiality of the SMA model in case of IRS-ID LISS III image with fewer spectral bands. Various studies have been carried out on application of the SMA technique on Landsat ETM+ (higher number of bands) images. However, very limited studies have been performed regarding usability of the SMA model on images with fewer bands such as the IRS LISS III images. The results suggest that the SMA technique based on the linear mixture model is an adequate means of land-cover mapping not only for Landsat ETM+ images of large number of bands but also for multispectral images like IRS with limited bands. The outcome of this analysis can be used effectively for several vegetation and environmental analyses like forest growth model, environmental monitoring models, landscape models and different policy-making decisions. In recent years, all over the world there is a major concern over environmental degradation including global warming (Townshend et al., 1994). Forests are an important component of the global carbon cycle as carbon dioxide is one of the key ingredients in photosynthesis. In view of the above importance of accurate forest mapping, the high precision technique of the SMA will provide higher accuracy also in carbon cycle modeling.

Acknowledgements

The authors thank to the anonymous reviewer of *International Journal of Geoinformatics* for reviewing the article.

References

Adams, J. B., Sabol, D. E., Kapos, V., Filho, R. A., Roberts, D. A., Smith, M. O., and Gillespie, A. R., 1995, Classification of Multispectral Images Based on Fraction Endmembers: Application to Land-Cover Change in the Brazilian Amazon. *Remote Sensing of Environment*, 52, 137–154.

Atkinson, P. M., Cutler, M. E. J., and Lewis, H., 1997, Mapping Sub-Pixel Proportional Cover with AVHRR Imagery. *International Journal of Remote Sensing*, 18, 917– 935.

Bateson, A. and Curtiss, B., 1996, A Method for Manual Endmember Selection and Spectral Unmixing. *Remote Sensing of Environment*, 55, 229–243.

Campbell, J. B., 1996, Introduction to remote sensing (2nd ed.). New York: Guilford.

Congalton, R. G., and Mead, R. A., 1983, A quantitative method to test for consistency and correctness in photo interpretation, *Photogrammetric Engineering & Remote Sensing*, 49, 69–74.

Dengsheng, Q., 2004, Spectral Mixture Analysis of the Urban Landscape in Indianapolis with Landsat ETM+ Imagery. *Photogrammetric Engineering & Remote Sensing*, 70, 1053–1062.

Drury, S. A., 1990, *A Guide to Remote Sensing: Interpreting Images of the Earth.* (Oxford, UK: Oxford University Press).

Fisher, P., 1997, The pixel: A snare and a delusion. *International Journal of Remote Sensing*, 18, 679–685.

Forest Survey of India, Dehradun 2005: *State of Forest Report 2003*. (http://www.fsiorg.net/fsi-2003/states/ index.asp? state_code=22).

Hall, F. G., 1994, Adaptation of NASA remote sensing technology for regional-level analysis of forest ecosystems. In *Remote sensing and GIS in ecosystem management*. V. A. Sample (ed), (Washington, DC: Island Press).

Jaisawal, R. K., Mukherjee, S., Raju, K. D., and Saxena, R., 2002, Forest fire risk zone mapping from satellite imagery and GIS. International Journal of Applied Earth Observation and Geoinformation, 4 (2002) 1-10

Settle, J., and Drake, N. A., 1993, Linear Mixing and the Estimation of Ground Cover Proportions. *International Journal of Remote Sensing*, 14, 1159–1177.

Strahler, A. H., Woodcock, C. E., and Smith, C. E., 1986, On the nature of models in remote sensing. *Remote Sensing of Environment*, 70,121–139.

Townshend, J. R. G., Justice, C. O., Skole, D., Malingreau, J. P., Cihlar, J., Teillet, P. P., Sadowski, F., and Ruttenberg, S., 1994, The 1 km Resolution Global Data Set: Needs of the International Geosphere Biosphere Programme. *International Journal of Remote Sensing*, 15, 3417–3441.