Soft Classification based Sub-pixel Allocation model

Uttam Kumar[1] [uttam@ces.iisc.ernet.in]
Mukhopadhyay C[2]. [cm@mgmt.iisc.ernet.in]
Kumar Raja S[3] [s_kumar_raja@yahoo.com]
&
Ramachandra T. V[4],* [cestvr@ces.iisc.ernet.in]

[1] Department of Management Studies, Centre for Sustainable Technologies
[2] Department of Management Studies, Indian Institute of Science
[3] Department of Electrical Engineering, Indian Institute of Science
[4] Centre for Ecological Sciences, Centre for Sustainable Technologies,
Indian Institute of Science

_____

* Address for Correspondence:
Dr. T.V. Ramachandra
Energy & Wetlands Research Group
Centre for Ecological Sciences
Indian Institute of Science,
Bangalore – 560 012, INDIA

Tel : 91-80-23600985/22932506/22933099
Fax : 91-80-23601428/23600085/23600683[CES-TVR]
E-mail: cestvr@ces.iisc.ernet.in
         energy@ces.iisc.ernet.in
http://wgbis.ces.iisc.ernet.in/energy

Soft Classification based Sub-pixel Allocation model

Abstract

Land cover (LC) refers to what is actually present on the ground and provide insights into the underlying solution for improving the conditions of many issues, from water pollution to sustainable economic development. One of the greatest challenges of modeling LC changes using remotely sensed (RS) data is of scale-resolution mismatch: that the spatial resolution of detail is less than what is required, and that this sub-pixel level heterogeneity is important but not readily knowable. However, many pixels consist of a mixture of multiple classes. The solution to mixed pixel problem typically centers on soft classification techniques that are used to estimate the proportion of a certain class within each pixel. However, the spatial distribution of these class components within the pixel remains unknown. This study investigates Orthogonal Subspace Projection - an unmixing technique and uses pixel-swapping algorithm for predicting the spatial distribution of LC at sub-pixel resolution. Both the algorithms are applied on many simulated and actual satellite images for validation. The accuracy on the simulated images is ~100%, while IRS LISS-III and MODIS data show accuracy of 76.6% and 73.02% respectively. This demonstrates the relevance of these techniques for applications such as urban-nonurban, forest-nonforest classification studies etc.

Key words: Soft classification, sub-pixel, orthogonal subspace projection, pixel-swapping, resolution

1.   Introduction

Land cover (LC) describes the physical state of the earth's surface in terms of the natural environment (such as vegetation, soil and water, etc.) and the man-made structures (e.g. buildings). These LC features have been classified using the remotely sensed (RS) satellite imagery of different spatial (panchromatic), spectral (multispectral, superspectral and hyperspectral) and temporal resolutions. Superspectral (SS) imagery provides additional benefits over multispectral (MS) imagery in many applications such as detection, discrimination, classification and identification of objects. Earlier SS imagery were processed and analysed by MS image processing algorithms via preprocessing such as feature extraction, dimensionality reduction and band selection. However, such MS to SS approaches are not straightforward extension of MS image processing. When the spectral resolution is low (3-4 bands) as in MS images, the image processing techniques are generally developed to explore spatial information for spatial domain analysis. When spectral resolution is increased as in SS imagery, such spatial domain-based MS imaging techniques may be found to be less effective in certain applications. For example, in some applications, where the object size is smaller than the pixel resolution, such as houses/buildings in a 250 m spatial resolution data, the processing relies at sub-pixel

level. In order to address this problem, unmixing techniques have been developed to exploit sub-pixel level information for image analysis [1]. Sub-pixel class composition is estimated through the use of techniques, such as linear mixture modelling [2], supervised fuzzy-c means classification [3] and artificial neural networks [4]. Linear mixture model is briefly stated here:

Let $h(x, y)$ be the signature collected by the sensor at the pixel with spatial coordinates (x, y). This signature can be considered an N-dimensional (N-D) vector, where N is the number of spectral bands; it can also be modeled as a linear combination of endmembers (pure pixels or pixels representing only one category in the image) vectors e_i, $i = 1, \ldots, E$ using equation (1).

$$h(x, y) = \sum_{i=1}^{E} \varphi_i(x, y).e_i \qquad (1)$$

where $\varphi_i(x, y)$ is a scalar value representing the functional coverage of endmember vector e_i at pixel $h(x, y)$. Two constraints are usually imposed in equation (1). These are the abundance nonnegativity constraint and abundance sum-to-one constraint as defined by equation (2) and (3):

$$\varphi_i(x, y) \geq 0, \ \forall \, i : \ 1 \leq i \leq E \qquad (2)$$

and

$$\sum_{i=1}^{E} \varphi_i(x, y) = 1 \qquad (3)$$

This allows proportions of each pixel to be partitioned between classes. The output of this technique produce abundance maps that display the portion of a certain class within each pixel and there are as many abundance maps as the number of classes. The abundance ranges from 0 to 1 in any given pixel, where 0 indicates absence of a particular class and 1 indicates full presence of that class in that particular pixel. Intermediate values between 0 and 1 may represent a fraction of that class. For example, 0.4 may represent 40% presence of a class in an abundance map and the remaining 60% could be some other class.

Orthogonal Subspace Projection (OSP) proposed by Chang [5] has been implemented here using SS data based on 1) best utilization of *a priori* target knowledge and 2) effective use of information in numerous spectral bands. The prior target knowledge is characterized in accordance with target signature of interest, referred to as desired target signature **d** and undesired target signature matrix **U** formed by those target signatures that are known but not wanted in image analysis. OSP separates **d** from **U** in a signal detection model, then eliminate the undesired target signatures in **U** prior to detection of **d** so as to improve signal detectability. This approach is effective when the number of spectral bands is higher than the target signatures of interest. The abundance maps obtained from this technique is less error prone representation of LC than obtained using hard classification techniques [6]. However the spatial distribution of these class components within the pixel remains unknown. Atkinson [7] examined a pixel-swapping optimisation algorithm within a geostatistical framework for allocating classes of sub-pixels. The proportion of LC within each pixel derived using various soft classification methods described above is used to map the location of class components within the

pixels. Soft pixel proportions are randomly allocated to hard sub-pixel binary classes initially and iterated by swapping pixel values in the hard classified map. The objectives of this paper are: (i) Generation of abundance maps from SS RS data of 250 m spatial resolution using OSP; (ii) Predict LC's spatial distribution at sub-pixel resolutions by maximizing the spatial autocorrelation between neighbouring sub-pixels.

2. Methods

2.1 **Orthogonal Subspace Projection (OSP):** SS data (such as MODIS with 250 to 1000 m spatial resolution), more often than not, encompass several different LC classes forming a mixed pixel. OSP eliminates all undesired spectral signatures within a pixel and then uses a matched filter to extract the desired spectral signature (endmember). The proportion of a class within a pixel is estimated as follows:

Let λ be the number of spectral bands and $\mathbf{r}$ be a λ x 1 dimensional column vector of digital numbers (reflectance). Assume that there are p targets, t_1, t_2, …, t_p present in an image scene. Therefore, there are p spectrally distinct endmembers. Let m_1, m_2, …, m_p denote their corresponding target signatures. A linear mixture model $\mathbf{r}$, models the spectral signatures of $\mathbf{r}$ as a linear combination of m_1, m_2, …, m_p with appropriate abundance fractions specified by α_1, α_2, …, α_p. Therefore M is a λ x p matrix representing target spectral signature (made up of linearly independent columns), denoted by [m_1, m_2, …, m_p], where $\lambda > p$ (overdetermined system) and m_j is an λ x 1 column vector represented by the spectral signature of the jth target t_j resident in the pixel vector $\mathbf{r}$. Let $\alpha = (\alpha_1, \alpha_2, …, \alpha_p)^T$ be a p x 1 abundance column vector associated with $\mathbf{r}$ where α_p denotes the fraction of the pth target signature m_p present in the pixel vector $\mathbf{r}$. A mixed pixel problem assumes that the spectral signature of the pixel vector $\mathbf{r}$ is linearly mixed by m_1, m_2, …, m_p, the spectral signature of the p targets, t_1, t_2, …, t_p and equation 1 can also be written as

$$\mathbf{r} = \mathbf{M}\alpha + \mathbf{n} \tag{4}$$

where $\mathbf{n}$ is the noise or can be interpreted as a measurement or model error. Equation 4 is a standard signal detection model where $\mathbf{M}\alpha$ is a desired signal vector to be detected and $\mathbf{n}$ is a corrupted noise. Since we are interested in detecting one target at a time, we can divide the set of p targets, t_1, t_2, …, t_p into a desired target, say t_p and a class of undesired targets, t_1, t_2, …, t_{p-1}. We need to eliminate the effects caused by the undesired targets t_1, t_2, …, t_{p-1} that are considered as interferers to t_p before the detection of t_p takes place. With annihilation of the undesired target signatures the detectability of t_p can be enhanced. First m_p is separated from m_1, m_2, ..., m_p in $\mathbf{M}$ and equation 4 is rewritten as

$$\mathbf{r} = \mathbf{d}\alpha_p + \mathbf{U}\gamma + \mathbf{n} \tag{5}$$

where $\mathbf{d} = m_p$ is the desired target signature of t_p and $\mathbf{U}$ is λ x (p-1) matrix = [m_1, m_2, …, m_{p-1}] is the undesired target spectral signature matrix made up of m_1, m_2, …, m_{p-1} which are the spectral signatures of the remaining $p − 1$ undesired targets, t_1, t_2, …, t_{p-1}. Equation 5 is called a ($\mathbf{d}$, $\mathbf{U}$) model; $\mathbf{d}$ is a λ x 1 column vector [d_1, d_2, …, d_λ]^T, γ is a (p-

1) x 1 column vector containing (p-1) component fractions of $\alpha = [\alpha_1, \alpha_2, …, \alpha_{p-1}]^T$ and $\mathbf{n}$ is a λ x 1 column vector representing additive and zero-mean white noise with cov=σ^2I where I is a λ x λ identity matrix. Using the ($\mathbf{d}$, $\mathbf{U}$) model, OS projector can annihilate $\mathbf{U}$ from the pixel vector $\mathbf{r}$ prior to detection of t_p similar to [8].

$$P = \mathbf{I} - \mathbf{U}\mathbf{U}^{\#}$$ (6)

where $\mathbf{U}^{\#}=(\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T$ is the pseudo-inverse of $\mathbf{U}$. The projector P is a λ x λ matrix operator that maps the observed pixel vector $\mathbf{r}$ into the orthogonal complement of $\mathbf{U}$. $\mathbf{U}$ has same structure as the orthogonal complement projector from the theory of least squares. Applying P to the ($\mathbf{d}$, $\mathbf{U}$) model results in a new signal detection model

$$P\mathbf{r} = P\mathbf{d}\alpha_p + P\mathbf{U}\gamma + P\mathbf{n}$$ (7)

where the undesired signal in $\mathbf{U}$ have been annihilated and the original noise $\mathbf{n}$ has been also suppressed to ñ = P$\mathbf{n}$. Equation 7 is called a OSP model. The operator minimizes energy associated with the signatures not of interest as opposed to minimizing the total least square error. A general signal detection-in-noise model given in (4) separates a signal source from noise. The ($\mathbf{d}$, $\mathbf{U}$) model in (5) is derived from the general signal-detection-in-noise model by breaking up the considered signal source into two types of signal sources $\mathbf{d}$ and $\mathbf{U}$ that can be processed separately on the basis of prior knowledge. The OSP model is a one signal source ($\mathbf{d}$) model derived from the ($\mathbf{d}$, $\mathbf{U}$) model with the U in the ($\mathbf{d}$, $\mathbf{U}$) model annihilated by P. Thus OSP model is a custom-designed signal detection-in-noise model from (4) where the signal source has been preprocessed by P for signal enhancement. It should be noticed that P operating on $\mathbf{U}\gamma$ reduces the contribution of $\mathbf{U}$ to about zero. So, $P\mathbf{r} = P\mathbf{d}\alpha_p + P\mathbf{n}$ (8)

On using a linear filter specified by a weight vector $\mathbf{x^T}$ on the OSP model, the filter output is given by

$$\mathbf{x}^T P\mathbf{r} = \mathbf{x}^T P\mathbf{d}\alpha_p + \mathbf{x}^T P\mathbf{n}.$$ (9)

An optimal criteria here is to maximize signal to noise ratio (SNR) of the filter output

$$SNR(\mathbf{x}) = \frac{\mathbf{x}^T P\mathbf{d}\alpha_p^2 \mathbf{d}^T P^T \mathbf{x}}{\mathbf{x}^T P E\{\mathbf{n}\mathbf{n}^T\} P^T \mathbf{x}} = \frac{\alpha_p^2}{\sigma^2} \frac{\mathbf{x}^T P\mathbf{d}\mathbf{d}^T P^T \mathbf{x}}{\mathbf{x}^T P P^T \mathbf{x}}$$ (10)

where E{} denotes the expected value. Maximisation of this is a generalized eigenvalue-eigenvector problem

$$P\mathbf{d}\mathbf{d}^T P^T \mathbf{x} = \tilde{\lambda} P P^T \mathbf{x}$$ (11)

where $\tilde{\lambda}=\lambda(\sigma^2/\alpha_p)$. The eigenvector which has the maximum λ is the solution of the problem and it turns out to be $\mathbf{d}$. The idempotent ($P^2 = P$) and symmetric ($P^T = P$) properties of the interference rejection operator are used. One of the eigenvalues is $\mathbf{d}^T P\mathbf{d}$ and it turns out that the value of $\mathbf{x^T}$ (filter) which maximizes the SNR is

$$\mathbf{x}^T = k\mathbf{d}^T$$ (12)

where k is an arbitray scalar. This leads to an overall classification operator for a desired target in the presence of multiple undesired target and white noise given by the 1 x λ vector

$$q^T = \mathbf{d}^T P \tag{13}$$

This result first nulls the interfering signatures, and then uses a matched filter for the desired signature to maximize the SNR. When the operator is applied to all of the pixels in a scene, each λ x 1 pixel is reduced to a scalar which is a measure of the presence of the signature of interest. The final result reduces the λ images comprising the SS images into a single image where the high intensity indicates the presence of the desired signal. Applying $\mathbf{d}^T P$ on (8)

$$\mathbf{d}^T PP\mathbf{r} = \mathbf{d}^T PP\mathbf{d}\alpha_p + \mathbf{d}^T PP\mathbf{n} .$$

Therefore,
$$\boxed{\alpha_p = \frac{\mathbf{d}^T P\mathbf{r}}{\mathbf{d}^T P\mathbf{d}}} \tag{14}$$

is the abundance estimate of the pth target material. In the absence of noise, the estimate matches with the exact value in (5).

2.2   **Pixel-swapping algorithm:** The pixel-swapping algorithm is designed to receive, as input, an image of LC proportions in $K=2$ classes (probably obtained by application of a soft classifier). Assumption here is that LC pixels have intra and inter spatial dependence, forcing the sub-pixel allocation model as an optimisation problem to maximize the autocorrelation between the pixels. First, the pixel-level LC proportions are transformed into sub-pixel hard LC classes. Thus, if 10 by 10 sub-pixels are to be mapped within each pixel, a LC proportion of 57 percent would mean that 57 sub-pixels were allocated to that class. Second, these sub-pixels are allocated randomly within each pixel. Once allocated, only the spatial arrangement of the sub-pixels can vary, not the actual attribute values. Furthermore, the number of sub-pixels allocated within each pixel remains fixed. Once the random initialization is done, the objective is to vary the spatial arrangement of the sub-pixels in such a way that the spatial correlation between neighboring sub-pixels (both *within* and *between* pixels) is maximized given that the pixel-level proportions cannot vary. The three basic steps are:

1.   For every sub-pixel the atractiveness A_i of a pixel i is predicted as a distance-weighted function of its $j = 1, 2, ..., j$ neighbours:

$$A = \sum_{j=1}^{j} \lambda_{ij} z(\mathbf{x}_j) \tag{15}$$

where $z(\mathbf{x}_j)$ is the (binary) class of the jth pixel at location $\mathbf{x}_j$ and λ_{ij} is a distance-dependent weight predicted as:

$$\lambda_{ij} = \exp\left(\frac{-h_{ij}}{\alpha}\right) \tag{16}$$

where h_{ij} is the distance between the location $\mathbf{x_i}$ of pixel i for which the attractiveness is desired, the location x$_j$ of a neighbouring pixel j, and α is the non-linear parameter of the

exponential model. The exponential weighting function chosen here is arbitrary, and other alternatives such as simple inverse distance weighting function of the Gaussian model could also be used. The choice of non-linear parameter and the number of nearest neighbours are also important considerations.

2.   Once the attractiveness of each sub-pixel location has been predicted based on the current arrangement of sub-pixel classes, the algorithm ranks the scores on a pixel by-pixel basis. For each pixel, the least attractive location at which the sub-pixel is currently allocated to a "1" (i.e., a "1" surrounded mainly by "0"s) is stored:

$$\text{candidate A} = (\mathbf{x}_j : A_i = \min(A) \,|\, z(\mathbf{x}_j) = 1) \tag{17}$$

Similarly, the most attractive location at which the pixel is currently allocated to a "0" (i.e., a "0" surrounded mainly by "1"s) is also stored:

$$\text{candidate B} = (\mathbf{x}_j : A_i = \min(A) \,|\, z(\mathbf{x}_j) = 0). \tag{18}$$

3.   Sub-pixel classes are swapped on the following basis. If the attractiveness of the least attractive location is less than that of the most attractive location, then the classes are swapped for the sub-pixels in question:

$$\left. \begin{array}{l} z(\mathbf{x}_j) = 0 \\ z(\mathbf{x}_j) = 1 \end{array} \right\} \text{if } A_i < A_j. \tag{19}$$

If it is more attractive, no change is made. The above 3 steps are repeated such that a solution is approached iteratively. The process is stopped either at a fixed number of iterations or when the algorithm converges on a solution. The algorithm is summarized below in simple terms:
1. Allocate sub-pixels to classes based on the pixel level proportions.
2. For each iteration:
   (a) For each pixel
      (i) For each sub-pixel within the pixel:
         1. For each neighbouring sub-pixel within a window calculate A per sub-pixel,
         2. Find minimum attractiveness A_j, for all sub-pixels currently allocated to 1 (i.e., $(A_i = \min(A)|z(x_i) = 1)$)
         3. Find maximum attractiveness A_j for all sub-pixels currently allocated to 0 (i.e., $(A_j = \max(A)|z(x_j) = 0)$)

      (ii) If $A_i < A_j$ swap the single pair of sub-pixel allocations.

3.   **Data:**   To test the performance of the algorithms several datasets of various spatial resolutions were simulated. Algorithm implementation and artificial dataset generation were done using C language and were validated on RS data of different spatial and spectral resolutions.

3.1   Simple and Polygonal targets: To test the performance of algorithms two class images of simple geometric shapes (line of size 1000 rows x 1000 columns and circle of size 700 x 700) were generated. These class maps were then used to generate 100 sub-pixels by 100 sub-pixels target images for line and 70 by 70 sub-pixels for circle respectively; which is one-tenth of the original class map resolution. The purpose of these simulated images was to use them in lieu of real soft classified images of LC proportions. Similarly, a complex shape class map (1360 x 1400) was generated to test the performance of the algorithm on irregular shapes that might correspond to any LC feature. The image was degraded to 136 x 140 to see the performance of the pixel-swapping algorithm.

3.2   High Resolution IRS LISS-III data: IRS LISS-III (Indian Remote Sensing Satellite Linear Imaging Self Scanner-III) data acquired on 28th March, 2006 (procured from NRSA, Hyderabad, India) having 3 spectral bands (Green, Red and NIR) with 23.5 m spatial resolution for Bangalore (12°39'00'' to 131°3'00''N and 77°22'00'' to 77°52'00''E) were used. The data were georegistered with known ground control points (GCPs) and geometrically corrected and resampled to 25 m (1360 x 1400 pixels) which helped in pixel level comparison of abundance maps obtained by unmixing MODIS data with LISS-III classified map.

3.3   SS MODIS data: MODIS (Moderate Resolution Imaging Spectroradiometer) data were downloaded from GLCF (Global land Cover Facility), NASA, USA. These data have 7 bands with band 1 and 2 at 250 m and band 3 to 7 at 500 m spatial resolution. The data were georegistered with GCPs, geometrically corrected and bands 3 to 7 were resampled to 250 m (136 x 140 pixels) so that all 7 bands are at same spatial resolution.

4.   Results and discussion

4.1   Classification of LISS-III data: The class spectral characteristics for four LC categories (buildings/urban, vegetation, water bodies and others) using LISS-III MSS bands 2, 3 and 4 were obtained from the training pixels spectra to assess their inter-class separability and the images were classified using Gaussian Maximum Likelihood classifier with training data uniformly distributed over the study area collected with pre calibrated GPS (figure 1). This was validated with the representative field data (training sets collected covering the entire city and validation covering ~ 10% of the study area) and also using Google Earth image (http://www.earth.google.com). LC statistics, producer's accuracy, user's accuracy and overall accuracy computed are listed in table 1.

4.2   Spectral unmixing of MODIS data: Unmixing of MODIS data were done to obtain abundance maps using OSP. Visual inspection as well as accuracy assessment of abundance images corresponding to each endmember showed that OSP algorithm maps LC categories spatially and proportionally similar to supervised classified image of LISS-III. The proportions of the endmembers in these images (figure 2) range from 0 to 1, with 0 indicating absence of the endmember and increasing value showing higher abundance. Bright pixels represent higher abundance of 50% or more stretched from black to white.
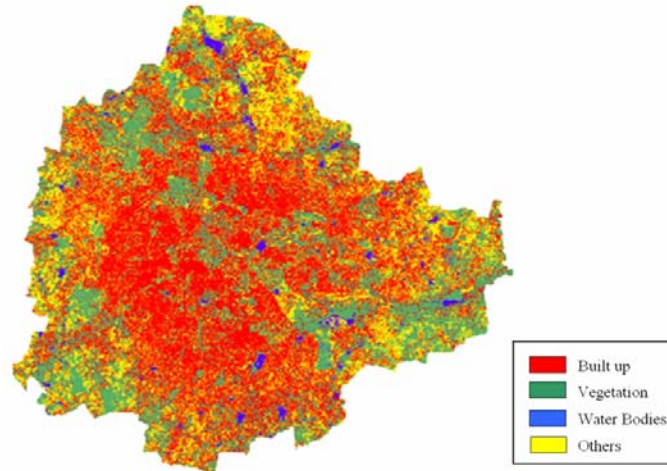
Figure 1: LC classification using LISS-III MSS.

Table 1: LC details using LISS -III MSS

Class	Area (Ha)	Area (%)	Producer's Accuracy (%)	User's Accuracy (%)	Overall Accuracy
Builtup	29535	43.23	85.21	84.54	
Vegetation	19696	28.83	83.47	88.11	84.63%
Water bodies	1073	1.57	90.73	80.20	
Others	18017	26.37	78.27	77.73	

Accuracy assessment at pixel level: A pixel of MODIS (250 m) corresponds to a kernel of 10 x 10 pixels of LISS-III spatially. The builtup abundance map was validated pixel by pixel with the LISS-III builtup map. The proportion of builtup percentage obtained in these 136 x 140 cells were compared using three measures of performance – Root Mean Square (RMS) errors (0.160), Bivariate Distribution Functions (BDFs) between LISS-III classified (real proportion) and MODIS based abundance map (predicted proportions) and Wilcoxon signed rank test. The bivariate plot (figure 3) was helpful to visualise the accuracy of prediction by mixture models. While most of the points fall between the two 20% difference lines, some points are outside the lines indicating that there could be higher or lower prediction than the actual values. The Pearson's product-moment correlation was 0.84 (p-value $< 2.2e^{-16}$). 10% of the difference pixels between the LISS-III proportion and the MODIS abundance were randomly selected at the same spatial locations in the two images by random number generation. These difference data values failed when subjected to normality test (table 2).

Table 2: Normality tests for the difference in LISS-III and MODIS proportions

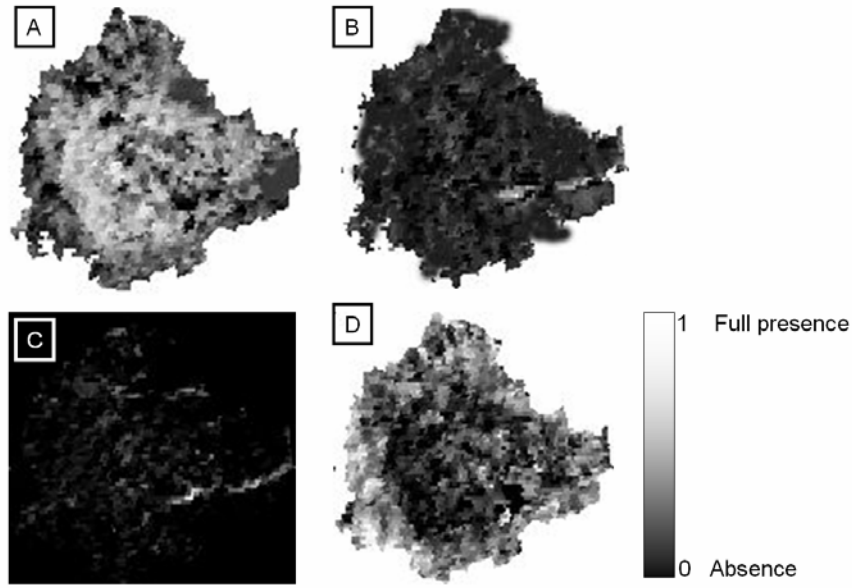| Normality test	p-value
Shapiro-Wilk	$2.2e^{-16}$
Anderson-Darling	$2.2e^{-16}$
Kolmogorov-Smirnov	$2.2e^{-16}$
Cramer-von Mises	Infinity
Lilliefors	$2.2e^{-16}$

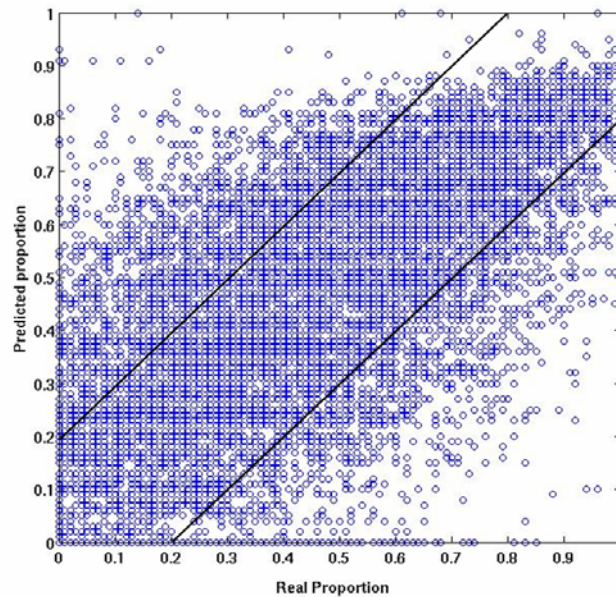**Figure 2: LC abundance maps of builtup, vegetation, water bodies and others.**

Figure 3: BDF of LISS-III classified map and MODIS based abundance map.

The data points corresponding to builtup proportion in a MODIS pixel and corresponding LISS-III pixels were then subjected to box-plot and histogram analysis. The distribution was symmetrical and hence Wilcoxon signed rank test with continuity correction was performed (p-value = 0.6633), which showed less evidence against null hypothesis – true location of the median of the difference between two data sets is equal to 0.

4.3 Pixel swapping algorithm on simple and irregular shape images: The simulated class maps of line and circle are shown in figure 4(A) and 4(E). Pixel swapping algorithm was implemented on the simulated abundance maps of line [figure 4(B)], circle

[figure 4(F)] and polygon shape. The algorithm initially allocated each sub-pixel to a binary hard class randomly, such as to maintain the original pixel proportions (figure 4(C) and figure 4(G)). Thereafter, the algorithm made a maximum of one swap per-pixel for each iteration. The number of nearest neighbour used was 3 and the non-linear parameter of the exponential model α was also set to 3. The line image took 34 iterations, circle - 18 iterations and complex shape took 43 iterations to converge to class maps [figure 4(D) and 4(H)] from their abundance maps. For all the three features the results are acceptable visually. The overall accuracy of the final converged map for line is 99.97, circle is 99.94 and polygon is 99.84%.
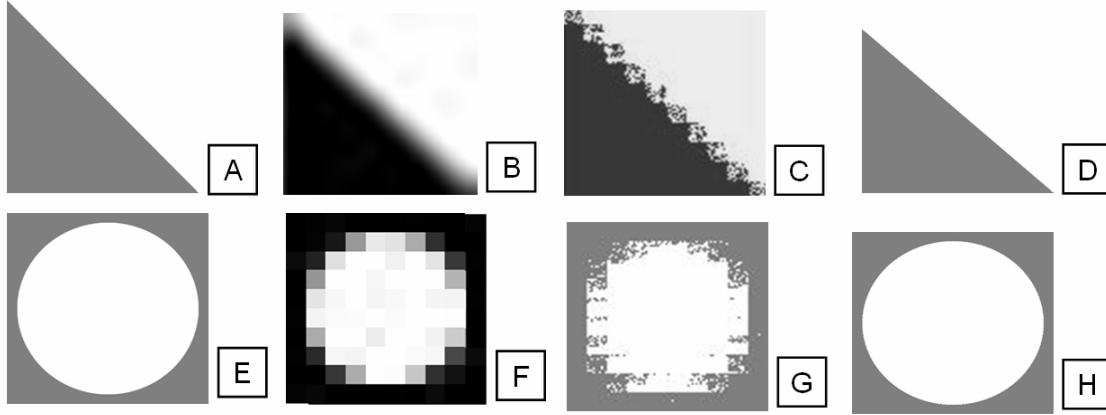

Figure 4: Sub-pixel mapping of a linear feature and a circle: (A) Test image of line, (B) Abundance image, (C) Random initial allocation to sub-pixels, (D) Solution after convergence of line, (E) Test image of circle, (F) Abundance image of circle, (G) random initial allocation, (H) Converged map of the circle after pixel swapping iterations.

4.4   Pixel swapping on LISS-III image: From the supervised classification of LISS-III image (figure 1) of 25 m spatial resolution (1360 x 1400 pixels), builtup class was assigned 1 while other pixels were assigned 0 as the algorithm can handle only two classes as shown in figure 5(A). This was simulated to 250 m class proportions (136 x 140 sub-pixels) as shown in figure 5(B). Initially, the sub-pixels within each pixel were allocated to one of the two classes (builtup/nonbuiltup). The nonlinear parameter of the exponential model and the bandwidth were changed and adjusted at each execution and finally set to 3 which gave the best output. In a two-class prediction problem (binary classification as in this case: builtup/nonbuiltup), where the outcomes are labeled either as positive (*p*) or negative (*n*) class, there are four possible outcomes - *true positive* (TP); *false positive* (FP), *true negative* (TN) and *false negative* (FN). The sensitivity, specificity, positive predicted value (PPV), and negative predicted value (NPV) were computes as given by equations 20– 23.

p - *Sensitivity*= TP / (TP + FN)   (20)
n - *Specificity*= TN / (FP + TN)   (21)
p' - *PPV* = TP / (TP + FP)   (22)
n' - *NPV* = TN / (TN + FN)   (23)
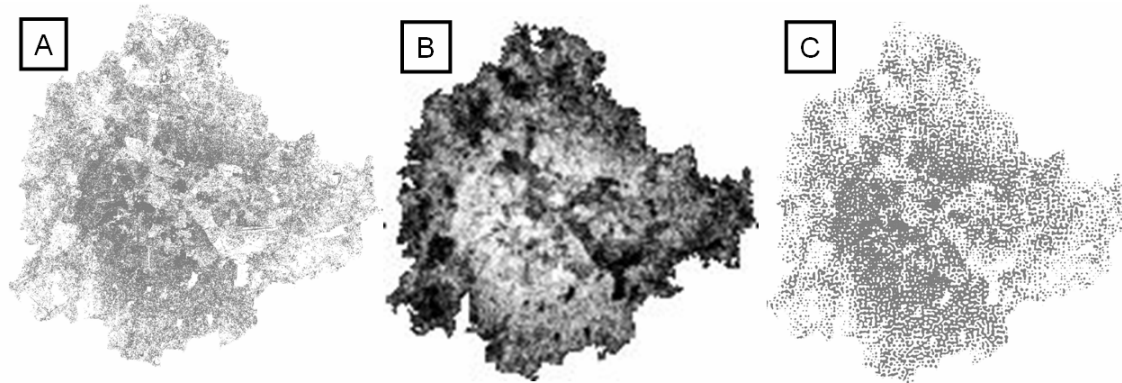
Figure 5: (A) Builtup pixels shown in black and non-built shown in white from LISS-III classified image, (B) Sub-pixel map of builtup, (C) Converged map of the builtup after applying pixel swapping algorithm.

The sensitivity measures the proportion of actual positives which are correctly identified. The specificity measures the proportion of negatives which are correctly identified. PPV is the precision of positives that were correctly identified. NPV is the precision of negatives correctly identified. Table 3 is the contingency table for LISS-III class map of builtup and converged map. The analysis showed that sensitivity was 0.6, specificity was 0.69, PPV was 0.6 and NPV was 0.69. Accuracy assessment was also performed using error matrix. The producer's and user's accuracy was 60% and the overall accuracy was 65% with respect to the LISS-III original class image.

Table 3: Contingency table for LISS-III MSS class image and converged image

		Actual	
		p	*n*
Predicted	*p'*	290741	192612
	n'	193359	442448

However, when the converged map was compared with the ground truth, the accuracy was much higher (76.6%). Next, the proportion of the number of builtup pixels before and after applying the pixel swapping algorithm were verified. The original as well as the converged map were resampled to 1400 x 1400 and divided into small-images of 100 by 100 pixels (14 x 14 blocks). Proportion of builtup per block was computed for each block. The BDF of actual and converged builtup pixels is given in figure 6, which shows class map obtained from simulated abundance LISS-III map using pixel-swapping algorithm is as good as the original image (r =0.9998275, p < 2.2e-16).

Pixel swapping on MODIS abundance map: Similar to LISS-III, pixel swapping algorithm was implemented on MODIS abundance map. Initially the data were grouped into two categories (builtup/nonbuiltup) as shown in figure 7(A). Then sub-pixels within each pixel were randomly allocated to one of the two classes (builtup/nonbuiltup) as shown in figure 7(B). The nonlinear parameter of the exponential model and the bandwidth were adjusted at each run to get the best converged map at 3. A 2 by 2 contingency table was generated to assess the accuracy of the prediction (table 4).
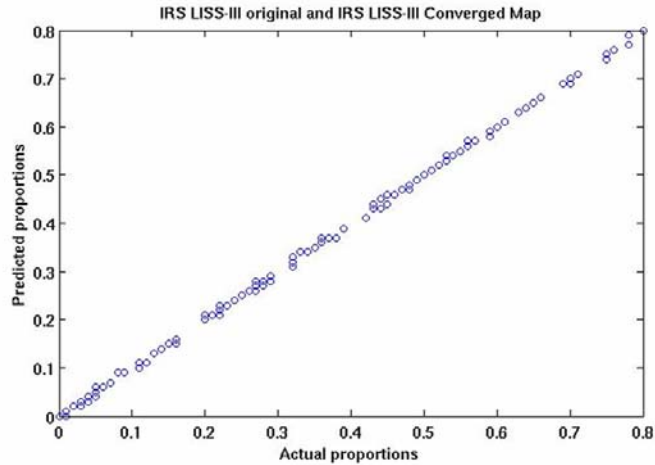
Figure 6: BDF of LISS-III class map and LISS-III converged map.

Table 4: Contingency table for LISS-III MSS class and MODIS converged image

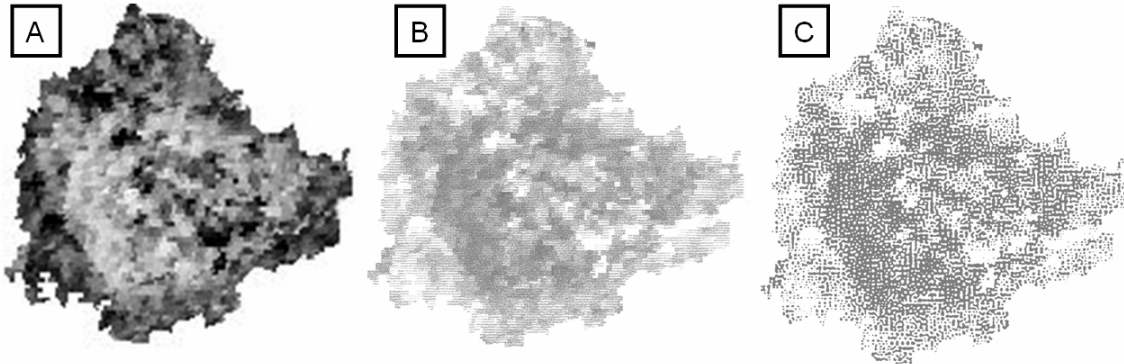|		Actual	
		p	*n*
Predicted	*p'*	255995	241150
	n'	228105	393910

Figure 7: (A) Sub-pixel map of builtup from MODIS, (B) Random allocation of builtup sub-pixels, (C) Converged map after applying pixel swapping algorithm.

Accuracy assessment showed that the producer's and user's accuracy was 52% and the overall accuracy was 58% with respect to LISS-III original classified image. Sensitivity was 0.52, specificity was 0.62, PPV was 0.51 and NPV was 0.63, indicating 50% match of the MODIS converged map with LISS-III class map. However, when the converged map was compared with the ground truth, the accuracy was 73.1%. Images of 100 by 100 pixels (14 x 14 blocks) were generated as earlier, and the BDF of actual and predicted builtup was plotted (figure 8), which indicates significant match (r = 0.9736454, p < 2.2e^{-16}) in terms of the pixel proportions. The proportion of the class is almost retained in the converged map compared to the class map even though the spatial location may shift due to the inherent correlation maximisation property of the pixel swapping algorithm.
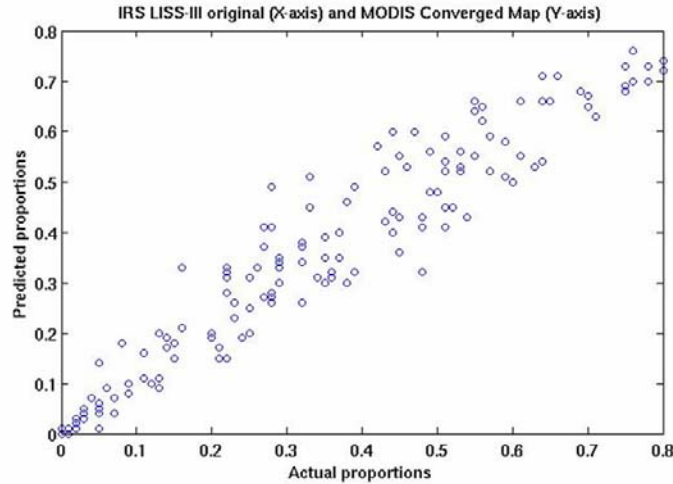
Figure 8: BDF of LISS-III class map and MODIS converged map.

There are certain conditions under which the pixel-swapping does not perform accurately. The algorithm fails if a fuzzy set represents the classes in the scene, such that the mixed pixels arise through vagueness rather than ambiguity. The formation of compact convex shapes is LISS-III and MODIS converged maps [figure 5(C) and 7(C)] is observed as the pixel swapping algorithm attempts to maximize the spatial correlation between neighbouring sub-pixels. Presence of multiple small objects in a pixel, also leads to coalescing of smaller objects, leading to incorrect prediction of a single large object. There are some issues related to choice of parameters – the distance-weighted function, the number of neighbours and the non-linear parameter of the exponential function. With the smaller number of neighbours, the speed of the algorithm is very high, however it produces unsatisfactory results. An important consideration is that the number of neighbours should not be so large that a given sub-pixel is attracted to other sub-pixel that it cannot neighbour such as belonging to a feature that exist in a pixel that its own pixel does not neighbour [7]. Thus the maximum non-linear parameter should be one sub-pixel less than the number of sub-pixels along a pixel side. Image resample factor (image of 136 x 140 is resampled to 1360 x 1400 which is 10 times) is another important parameter which is a relative increase is spatial resolution from the sub-pixel level image of proportions to the pixel level image of hard LC classes.

5.   Conclusion

Statistical analysis and validataion of the LISS-III proportion of 10 by 10 builtup pixels and corresponding MODIS pixel confirm the utility of OSP (Orthogonal Subspace Projection) for generation of abundance maps. The validation showed higher accuracy in LISS-III converged image (76.6%) as well as MODIS converged image (73.1%) using pixel swapping algorithm. These techniques would be useful for detecting objects of interest (such as urban sprawl detection, forest/nonforest classification etc.) from coarse resolution data in the absence of high spatial resolution image.

6. Acknowledgement

7. References

[1] Shimabukuro, Y. E. and Smith, A. J., 1991, The least-squares mixing models to generate fraction images derived from remote sensing multispectral data. IEEE Transactions on Geoscience and Remote Sensing, 29(1), pp. 16–20.

[2] Kumar, U., Kerle, N., and Ramachandra T. V., 2008, Contrained linear spectral unmixing technique for regional land cover mapping using MODIS data. In: Innovations and advanced techniques in systems, computing sciences and software engineering / ed by Elleithy, K., Berlin: Springer, 2008. ISBN: 978-1-4020-8734-9. paper 87. 9 p.

[3] Foody, G.M., and Cox, D. P., 1994. Subpixel land cover composition estimation using a linear mixture model and fuzzy membership functions. International Journal of Remote Sensing, 15(3), pp. 619–631.

[4] Kanellopoulos, I., Varfis, A., Wilkinson, G. G., and Megier, J., 1992, Land-cover discrimination in SPOT imagery by artificial neural network-a twenty class experiment, International Journal of Remote Sensing, 13(5), pp. 917–924.

[5] Chang, C-I., 2005, Orthogonal Subspace Projection (OSP) Revisited: A Comprehensive Study and Analysis. IEEE Transactions on Geoscience and Remote Sensing, 43(3), pp. 502–518.

[6] McKelvey, K.S., and Noon, B. R. 2001, Incorporating uncertainties in animal location and map classification into habitat relationships modeling, Spatial Uncertainty in Ecology: Implications for Remote Sensing and GIS Applications, (C.T. Hunsaker, M.F. Goodchild, M.A. Friedl, and T.J.Case, editors), Springer, New York, pp. 72–90.

[7] Atkinson, P. M., 2005, Sub-pixel target mapping from soft-classified, remotely sensed imagery. Photogrammetric Engineering & Remote Sensing, 71(7), pp. 839–846.

[8] Harsanyi, J. C. and Chang, C.-I, 1994, Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection. IEEE Transaction on Geoscience and Remote Sensing, 32(4), pp. 779 – 785.