

Article

A Neural Network Based Hybrid Mixture Model to Extract Information from Non-linear Mixed Pixels

Uttam Kumar ^{1,2,3}, Kumar S. Raja ^{4,5}, Chiranjit Mukhopadhyay ³ and T.V. Ramachandra ^{1,2,*}

¹ Energy & Wetlands Research Group, Centre for Ecological Sciences, Indian Institute of Science, Bangalore 560012, India; E-Mail: uttam@ces.iisc.ernet.in

² Centre for Sustainable Technologies, Indian Institute of Science, Bangalore 560012, India

³ Department of Management Studies, Indian Institute of Science, Bangalore 560012, India; E-Mail: cm@mgmt.iisc.ernet.in

⁴ Institut de Recherche en Informatique et Systèmes Aléatoires, 35042 Rennes cedex, France; E-Mail: sk.kumar.raja@gmail.com

⁵ Technicolor Research & Innovation, 35576 Cesson Sévigné, France

* Author to whom correspondence should be addressed; E-Mail: cestvr@ces.iisc.ernet.in; Tel.: +91-80-22933099; Fax: +91-80-23601428/23600085/23600683 [CES-TVR].

Received: 1 July 2012; in revised form: 20 August 2012 / Accepted: 22 August 2012 /

Published: 14 September 2012

Abstract: Signals acquired by sensors in the real world are non-linear combinations, requiring non-linear mixture models to describe the resultant mixture spectra for the endmember's (pure pixel's) distribution. This communication discusses inferring class fraction through a novel *hybrid mixture model* (HMM). HMM is a three-step process, where the endmembers are first derived from the images themselves using the N-FINDR algorithm. These endmembers are used by the linear mixture model (LMM) in the second step that provides an abundance estimation in a linear fashion. Finally, the abundance values along with the training samples representing the actual ground proportions are fed into neural network based multi-layer perceptron (MLP) architecture as input to train the neurons. The neural output further refines the abundance estimates to account for the non-linear nature of the mixing classes of interest. HMM is first implemented and validated on simulated hyper spectral data of 200 bands and subsequently on real time MODIS data with a spatial resolution of 250 m. The results on computer simulated data show that the method gives acceptable results for unmixing pixels with an overall RMSE of 0.0089 ± 0.0022 with LMM and 0.0030 ± 0.0001 with the HMM when compared to actual class proportions. The unmixed MODIS images showed overall RMSE with HMM as

0.0191 ± 0.022 as compared to the LMM output considered alone that had an overall RMSE of 0.2005 ± 0.41 , indicating that individual class abundances obtained from HMM are very close to the real observations.

Keywords: mixture model; sub-pixel classification; non-linear unmixing; MODIS

1. Introduction

Hyper spectral imaging spectrometers collect data in the form of an image cube that represents reflected energy from the Earth's surface materials, where each pixel has the resultant mixed spectrum of the reflected source radiation [1]. The mixed spectrum phenomenon causes a mixed pixel problem because the intrinsic scale of spatial variation in land cover (LC) due to the heterogeneous and fragmented landscapes [2] is usually finer than the scale of sampling imposed by the image pixels (for example, MODIS data at 250 m to 1 km spatial resolution) resulting in mixed pixels. Mixed pixels thus are a mixture of more than one distinct object and exist for one of two reasons. Firstly, if the spatial resolution of the sensor is not high enough to separate different LC types, these can jointly occupy a single pixel, and the resulting spectral measurement will be a composite of the individual spectra that reside within a pixel. Secondly, mixed pixels can also result when distinct LC types are combined into a homogeneous mixture. This happens independently of the spatial resolution of the sensor [3].

Commonly used approaches to mixed pixel classification have been linear spectral unmixing [4], supervised fuzzy-c means classification [5], ANN (artificial neural networks) [6,7] and Gaussian mixture discriminant analysis [8], *etc.* which use a linear mixture model (LMM) to estimate the abundance fractions of spectral signatures lying within a pixel. LMM assumes that the reflectance spectrum of a mixture is a systematic combination of the component's reflectance spectra in the mixture (called endmembers). The combination of these endmembers is linear if the component of interest regarding a pixel appears in spatially segregated patterns. If, however, the components are in intimate association, the electromagnetic spectrum typically interacts with more than one component as it is multiply scattered, and the mixing systematics between the different components are highly non-linear. In other words, non-linear mixing occurs when radiance is modified by one material before interacting with another one under the assumption that incident solar radiation is scattered within the scene itself and that these interaction events may involve several types of ground cover materials [9] and require non-linear mixture model for unmixing the components of interest. In such cases, LMM have mostly failed in modeling a mixed pixel [10–12] and non-linear models have been found to be appropriate as evident from various studies [2], including vegetation and canopy discrimination [13] water quality assessment [12,14], *etc.*

If there are M spectral bands and N classes, then associated with each pixel is a M -dimensional vector \mathbf{y} whose components are the gray values corresponding to the M bands. Let $\mathbf{E} = [\mathbf{e}_1, \dots, \mathbf{e}_{n-1}, \mathbf{e}_n, \mathbf{e}_{n+1}, \dots, \mathbf{e}_N]$ be a $M \times N$ matrix, where $\{\mathbf{e}_n\}$ is a column vector representing the spectral signature (endmember) of the n th target material. For a given pixel, the abundance or fraction of the n th target material present in a pixel is denoted by α_n , and these values are the components of the N -dimensional abundance vector $\boldsymbol{\alpha}$. Assuming LMM [15], the observation vector \mathbf{y} is related to \mathbf{E} by

$$\mathbf{y} = \mathbf{E}\mathbf{a} + \boldsymbol{\eta} \quad (1)$$

where $\boldsymbol{\eta}$ accounts for the measurement noise. We further assume that the components of the noise vector $\boldsymbol{\eta}$ are zero-mean random variables that are i.i.d. (independent and identically distributed). Therefore, the covariance matrix of the noise vector is $\sigma^2 \mathbf{I}$, where σ^2 is the variance, and \mathbf{I} is $M \times M$ identity matrix. Two constraints imposed on the abundances in equation (1) are the non-negativity and sum-to-one given as

$$\alpha_n \geq 0, \forall n: 1 \leq n \leq N \quad (2)$$

and

$$\sum_{n=1}^N \alpha_n = 1 \quad (3)$$

This allows proportions of each pixel to be partitioned between classes. A non-linear mixture model (NLMM) is expressed as:

$$\mathbf{y} = f(\mathbf{E}, \mathbf{a}) + \boldsymbol{\eta} \quad (4)$$

where, f is an unknown non-linear function that defines the interaction between \mathbf{E} and \mathbf{a} . Theory and experiments demonstrate that we will get the fractions of endmembers wrong by using a linear model when spectral mixing actually is non-linear [10,11]. Non-linear effects are an area of active research in particular applications where LMM generally results in poor accuracy [12].

In this context, ANN based NLMMs outperform the traditional linear unmixing models. ANNs have been widely studied as a promising alternative to accomplish the difficult task of estimating fractional abundances of endmembers. Atkinson *et al.* [2] applied a MLP (multi-layer perceptron) model to decompose AVHRR imagery, and it was superior to the linear unmixing model and a fuzzy c-means classifier. Another popular ANN model—ARTMAP—introduced to identify the life form components of the vegetation mixture [13] using Landsat data could capture non-linear effects, performing better than LMM [16]. ART MMAP, an extension of ARTMAP was designed specifically for mixture analysis with enhanced interpolation function and it provides better prediction of mixture information than ARTMAP [17]. A regression tree has also been used as a non-linear unmixing model [7]. All of these methods stand alone and work on the data directly when endmembers are known *a priori*. The objective of this paper is to develop an automated procedure to unmix hyperspectral imagery for obtaining a fraction that accounts for the non-linear mixture of the class types. We call this model the Hybrid Mixture Model (HMM). HMM is carried out in three stages: (i) Endmembers are extracted from the image itself using an iterative N-FINDR algorithm; (ii) the endmembers are used in the linear unmixing model for abundance estimation; (iii) the abundance values along with the actual ground proportions are used to refine the abundance estimates using MLP for the individual classes to account for the non-linear nature of the mixing classes of interest.

This paper is structured in six sections. Methods for automatic endmember extraction, linear unmixing and MLP are discussed in Section 2 followed by the description of HMM in Section 3. Data preparation is dealt with in Section 4 with the experimental results and discussion in Section 5. Section 6 concludes with model limitations.

2. Methodology

2.1. Automatic Endmember Extraction—The N-FINDR Algorithm

The N-FINDR algorithm [18] is a fully automatic technique for endmember extraction from the image, which is briefly described here:

- (i) Let N denote the number of classes or endmembers to be identified.
- (ii) Perform a PCA-decomposition of the data and reduce the data to $N-1$ dimension space.
- (iii) Pick N pixels from the set and compute the simplex volume generated by the spectra of the N pixels. The volume of the simplex is proportional to

$$V = \det \begin{vmatrix} 1 & 1 & \dots & 1 \\ e_{11} & e_{12} & \dots & e_{1N} \\ e_{21} & e_{22} & \dots & e_{2N} \\ \dots & \dots & \dots & \dots \\ e_{N-11} & e_{N-12} & \dots & e_{N-1N} \end{vmatrix} \quad (5)$$

- (iv) Replace each endmember with the spectrum of each pixel in the data set and recompute the simplex volume. If the volume increases, the spectrum of the new pixel is retained as a potential endmember.

- (v) The above steps are executed iteratively considering all pixels, and the final set of retained spectra is taken as the endmembers.

2.2. Orthogonal Subspace Projection (OSP) to Solve Linear Mixture Model

OSP proposed by Chang [19] involves (a) finding an operator which eliminates undesired spectral signatures, and then (b) choosing a vector operator which maximizes the signal to noise ratio (SNR) of the residual spectral signature.

If we assume that there are N targets, $t_1, \dots, t_{n-1}, t_n, t_{n+1}, \dots, t_N$ present in an image scene, then there are N spectrally distinct endmembers with corresponding target signatures as $e_1, \dots, e_{n-1}, e_n, e_{n+1}, \dots, e_N$, where $M > N$ (over a determined system), where Equation (1) is a standard signal detection model. Since we are interested in detecting one target at a time, we can divide the set of N targets into a desired target, say t_n , and a class of undesired targets, $t_1, \dots, t_{n-1}, t_{n+1}, \dots, t_N$. We need to eliminate the effects caused by the undesired targets that are considered as interferers to t_n before the detection of t_n takes place. With annihilation of the undesired target signatures, the detectability of t_n can be enhanced. In order to find the abundance of the n th target material (α_n), first e_n is separated from $e_1, \dots, e_{n-1}, e_n, e_{n+1}, \dots, e_N$ in \mathbf{E} . Let the corresponding spectral signature of the desired target material be denoted as \mathbf{d} . The term $\mathbf{E}\boldsymbol{\alpha}$ can be rewritten to separate the desired spectral signature \mathbf{d} from the rest as:

$$\mathbf{E}\boldsymbol{\alpha} = \mathbf{d}\alpha_n + \mathbf{U}\boldsymbol{\gamma} \quad (6)$$

and Equation (1) is rewritten as

$$\mathbf{y} = \mathbf{d}\alpha_n + \mathbf{U}\boldsymbol{\gamma} + \boldsymbol{\eta} \quad (7)$$

where $\mathbf{d} = e_n$ is the desired target signature of t_n and \mathbf{U} is $M \times (N-1)$ matrix $= [e_1, \dots, e_{n-1}, e_{n+1}, \dots, e_N]$, is the undesired target spectral signature, which are the spectral signatures of the remaining $N-1$

undesired targets, $t_1, \dots, t_{n-1}, t_{n+1}, \dots, t_N$. Equation (7) is called a (\mathbf{d}, \mathbf{U}) model; \mathbf{d} is a $M \times 1$ column vector $[d_1, d_2, \dots, d_M]^T$, γ is a $(N-1) \times 1$ column vector containing $(N-1)$ component fractions of $\alpha = [\alpha_1, \dots, \alpha_{n-1}, \alpha_{n+1}, \dots, \alpha_N]^T$. Using the (\mathbf{d}, \mathbf{U}) model, OSP can annihilate \mathbf{U} from the pixel vector \mathbf{y} prior to detection of t_n similar to [20] by the operator

$$P = \mathbf{I} - \mathbf{U}\mathbf{U}^\# \quad (8)$$

where $\mathbf{U}^\# = (\mathbf{U}^T \mathbf{U})^{-1} \mathbf{U}^T$ is the pseudo-inverse of \mathbf{U} . The projector P is a $M \times M$ matrix operator that maps the observed pixel vector \mathbf{y} 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 (OSP model) given by

$$P\mathbf{y} = P\mathbf{d}\alpha_n + P\mathbf{U}\gamma + P\boldsymbol{\eta} \quad (9)$$

where the undesired signal in \mathbf{U} has been annihilated and the original noise has also been suppressed to $P\boldsymbol{\eta}$. The operator minimizes energy associated with the signatures not of interest as opposed to minimizing the total least squares error. It should be noted that P operating on $\mathbf{U}\gamma$ reduces the contribution of \mathbf{U} to about zero. So,

$$P\mathbf{y} = P\mathbf{d}\alpha_n + P\boldsymbol{\eta} \quad (10)$$

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

$$\mathbf{x}^T P\mathbf{y} = \mathbf{x}^T P\mathbf{d}\alpha_n + \mathbf{x}^T P\boldsymbol{\eta} \quad (11)$$

an optimal criterion here is to maximize SNR of the filter output

$$\begin{aligned} \text{SNR}(\mathbf{x}) &= \frac{\mathbf{x}^T P\mathbf{d}\alpha_n^2 \mathbf{d}^T P^T \mathbf{x}}{\mathbf{x}^T P E\{\boldsymbol{\eta}\boldsymbol{\eta}^T\} P^T \mathbf{x}} \\ &= \frac{\alpha_n^2}{\sigma^2} \frac{\mathbf{x}^T P\mathbf{d}\mathbf{d}^T P^T \mathbf{x}}{\mathbf{x}^T P P^T \mathbf{x}} \end{aligned} \quad (12)$$

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

$$P\mathbf{d}\mathbf{d}^T P^T \mathbf{x} = \lambda P P^T \mathbf{x} \quad (13)$$

where $\lambda = \lambda(\sigma^2/\alpha_n)$. 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 the value of \mathbf{x}^T (filter), which maximizes the SNR is

$$\mathbf{x}^T = k\mathbf{d}^T \quad (14)$$

where k is an arbitrary scalar. It leads to an overall classification operator for a desired target in the presence of multiple undesired targets and white noise given by the $1 \times M$ vector

$$\mathbf{q}^T = \mathbf{d}^T P \quad (15)$$

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 the pixels in a scene, each $M \times 1$ pixel is

reduced to a scalar which is a measure of the presence of the signature of interest. The final result reduces the M images into a single image where the high intensity indicates the presence of the desired signal. Applying $\mathbf{d}^T P$ on (10) gives

$$\mathbf{d}^T P P \mathbf{y} = \mathbf{d}^T P P \mathbf{d} \alpha_n + \mathbf{d}^T P P \boldsymbol{\eta} \quad (16)$$

therefore,

$$\alpha_n = \frac{\mathbf{d}^T P \mathbf{y}}{\mathbf{d}^T P \mathbf{d}} \quad (17)$$

is the abundance estimate of the n th target material. In the absence of noise, the estimate matches with the exact value in Equation (7). Another way of removing the undesired signal based on band ratios is hinted by [21]. For a noise subspace projection method, see [22]. Settle [23] showed that the full linear unmixing and OSP used here, and as described by Harsanyi and Chang [20] are identical. Full linear unmixing can be performed when the spectra for all the endmembers present in the image are known *a priori*. Often, knowledge of all the endmembers spectra is not available. Therefore partial unmixing methods for estimating the presence of one or a few desired, known spectra only are desirable [24]. In general, these approaches are effective when the number of spectral bands is higher than the target signatures of interest.

The value of α_n is the abundance of the n th class (in an abundance map) and ranges from 0 to 1 in any given pixel and there are as many abundance maps as the number of classes. Zero indicates absence of a particular class and 1 indicates presence of only that class in a particular pixel. Intermediate values between 0 and 1 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.

2.3. Artificial Neural Network (ANN) based Multi-layer Perceptron (MLP)

The advent of ANN approaches is mainly due to their power in pattern recognition, interpolation, prediction, forecasting, classification and process modeling [24]. A MLP network comprises a number of identical units organized in layers, with those on one layer connected to those on the next layer so that the output of one layer is used as input to the next layer. A detailed introduction on MLP can be found in the literature [24–28]. The main aspects here are: (i) The order of presentation of training samples should be randomized from epoch to epoch; and (ii) the momentum and learning rate parameters are typically adjusted (and usually decreased) as the number of training iterations increases. Individual algorithms were implemented in C programming language. GRASS (Geographic Resources Analysis Support System)—a free and open source package—was used for visualization of results, and statistical analysis was carried in R in a Linux system running on a 3 GHz Pentium-IV processor with 3.5 GB RAM.

3. Hybrid Mixture Model (HMM)

Despite many attempts of using ANN for unmixing models, ANN-based non-linear unmixing techniques remain largely unexplored for general-purpose applications [12]. Only [1,7,9,29] have produced some of the pioneering work in NLMM to be considered as a general model for ANN-based

non-linear unmixing independent of physical properties of the observed classes. Some of these applications are, however, difficult and complex in their implementation. LMM is easy to implement, generalize and reconstruct. Therefore, in our approach, we make use of the LMM output as the input to NLMM to refine the fraction estimates. The MLP architecture can be extended to produce a continuous-values output for sub-pixel classification problems. The entries to the MLP model is the abundance a^i (see Figure 1) output obtained from LMM, which is denoted by a^i_{LMM} where $i = 1, \dots, E$, and the neuron count at the input layer equals the number of endmember classes (estimated by a fully constrained LMM) as shown in Figure 2 [30].

Figure 1. Architecture of the multi-layer perceptron (MLP) model.

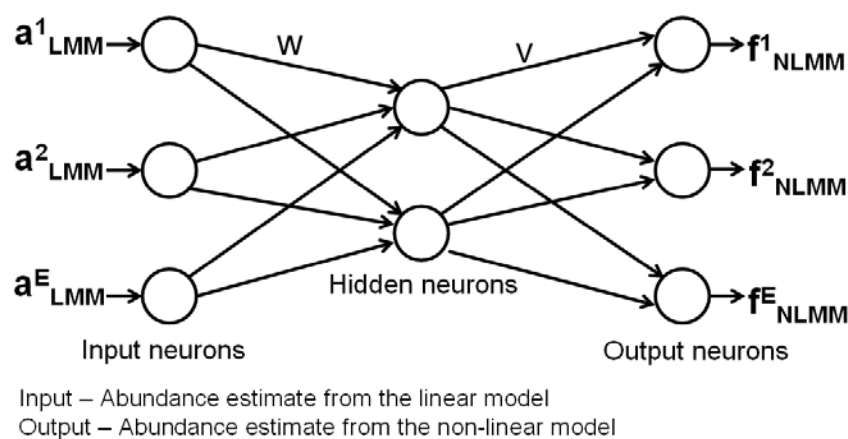
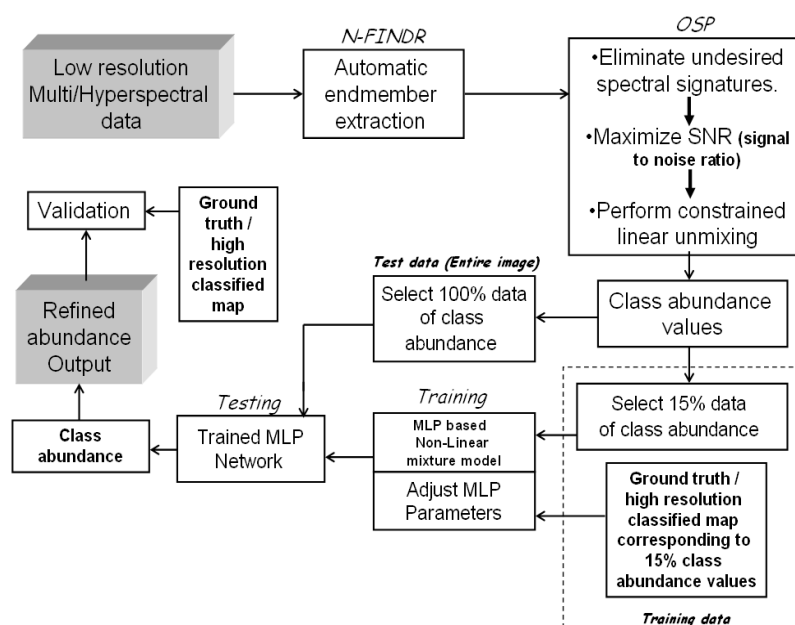


Figure 2. Hybrid mixture model (HMM) method for spectral mixture analysis.

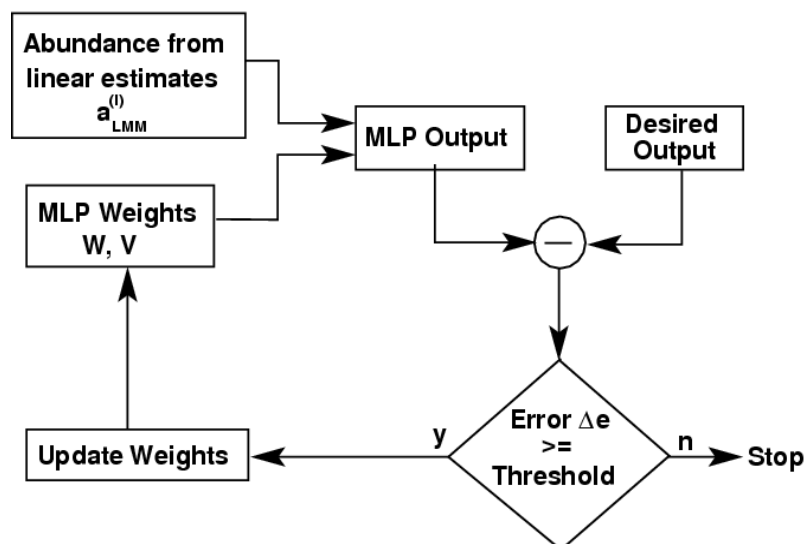


The training process is based on an error back-propagation algorithm [29], where the respective weights in the output and hidden nodes (W and V in Figure 3) are modified depending on the error (δe), the input data and the learning parameter α (α). The activation rule used here for the hidden and output layer nodes is defined by the logistic function

$$f(x) = \frac{1}{1 + e^{-x}} \quad (18)$$

δe of the output layer is calculated as the difference between the fraction (f) estimation outputs f_{NLMM}^i , $i = 1, \dots, E$, provided by the network architecture and a set of desired outputs given by actual fractional abundances available for the training samples. The resulting error is back-propagated until the convergence is reached. One of the earlier works by Plaza *et al.* [12] attempted a similar NLMM methodology, which made use of a modified MLP neural network (NN), whose entries were determined by a linear activation function provided by a Hopfield NN (HNN). The combined HNN/MLP method used the LMM to provide an initial abundance estimation and then refined the estimation using a non-linear model. As per Plaza *et al.* [12], this was the first and only approach in the literature that integrated linear and NLMM.

Figure 3. MLP structural diagram.



4. Data

4.1. Computer Simulations

One of the major problems involved in analyzing the quality of fractional estimation methods is the fact that ground truth information about the real abundances of materials at sub-pixel levels is difficult to obtain in real scenarios [29]. In order to avoid this shortcoming, a simulation of hyperspectral imagery was carried out to examine the algorithm's performance in a controlled manner. Spectral libraries of four minerals—alunite, buddingtonite, kaolinite and calcite [31] were used to generate synthetic data. Plaza *et al.* [12] used the signatures of soil (e_1) and vegetation (e_2) to create a simulated image with non-linear mixtures using a simple logarithmic function. The abundance of e_1 and e_2 were assigned according to Equation (19)

$$y(x,y) = \sum_{p=1}^2 e_p \cdot s_p(x,y) \quad (19)$$

where, \mathbf{y} denotes a vector containing the simulated discrete spectrum of the pixel at spatial coordinates (x,y) of the simulated image, $s_p(x,y) = \log \alpha_p(x,y)$ is the contribution of endmember e_p and $\alpha_p(x,y)$ is the fractional abundance of e_p at (x,y) . A limitation here is that even though all the pixels are mixed in different proportions, there are no instances of pure pixels. If α is 1, we expect the observed hyperspectral signature to be solely from one material, and therefore, ideally it should be identical to the endmember itself. Here, as the abundance increases towards 1, $\log(\alpha)$ approaches 0, thereby suppressing the contribution of that particular endmember. On the other hand, if α is 0, $\log(\alpha)$ approaches $-\infty$, and therefore, starts dominating in the observed spectral signature. Of course, it will appear as negative numbers. This is against our physical understanding as to how a material, which is almost not present in the pixel, contributes to the observation in a dominant way. That is, the model is not able to highlight the endmember of the correct material when its contribution is 1 and gives a wrong endmember when its contribution is 0. To overcome this limitation, we modify the model in Equation (19) by Equation (20):

$$\mathbf{y}(x,y) = \sum_{p=1}^4 \mathbf{sig}_p \cdot s_p(x,y) \quad (20)$$

where, \mathbf{sig}_p is the signature corresponding to p th mineral, $s_p(x,y) = \log(1 + \alpha_p(x,y))$ is the contribution of endmember e_p and $\alpha_p(x,y)$ is the fractional abundance of e_p in the pixel at (x,y) .

Simulated synthetic non-linear mixture hyperspectral data of 200 bands (250×250) using four minerals were classified using Maximum Likelihood Classifier (MLC) with signatures from the spectral libraries. This constitutes high-resolution (HR) images. These images were used to generate synthetic mixed pixels of 25×25 (referred to as low-resolution (LR) images). Four endmembers were extracted from LR images, and subsequently, abundance images were estimated corresponding to each endmember. Percentage abundance for a group of 10×10 pixels was computed for this entire HR classified image (250×250) obtained from MLC. This new image of a size of 25×25 was used as reference for validating the LR abundance output. However, the HR MLC based classified output (250×250) was not validated as the same spectral library which was used for generating the individual class signatures for classification of the HR image and was also used to create the synthetic images. Abundance values from 15% of the pixels obtained from linear unmixing along with the corresponding proportions obtained from the 250×250 classified image obtained by MLC were used to train the neurons in MLP. For example, each input sample to the MLP has the abundance values obtained from OSP for each of the four classes (0.2, 0.3, 0.1, 0.4 = 1 or 100% of a pixel) and the proportion of each class as derived from HR MLC based classified map (0.18, 0.27, 0.2, 0.35 = 1 or 100% of a pixel) by considering 25×25 classified pixels and finding the percentage of each class separately which is equivalent to 1×1 LR pixel spatially. Testing was done on the entire output abundance images (100% pixels).

4.2. MODIS Data

The training and testing data (pertaining to Kolar district, Karnataka State, India) used to study the model consisted of (i) IRS LISS-3 Multispectral with three spectral bands of 23.5×23.5 m spatial resolution acquired on December 25, 2002 and (ii) MODIS eight-day composite (of 19 December to 26 December 2002) data with seven bands at 250 and 500 m. The fractional LC for each MODIS pixel

was computed in four steps: (i) LISS-III data of 1000×1000 pixels were geo-corrected, resampled to 25 m and classified into six LC classes (agriculture, built-up/settlement, forest, plantation/orchard, wasteland/barren land and water bodies) using MLC; (ii) MODIS images (100×100 pixels; 10 times smaller than the size of LISS-III) were co-registered to LISS-III data and resampled to 250 m; (iii) Six endmembers were extracted using N-FINDR from the MODIS bands and the data were unmixed to estimate abundances for each pixel at the MODIS image scale; (iv) Finally, 15% MODIS abundance pixels obtained from LMM were randomly selected to be associated with the corresponding LISS-III classified pixels (as ground truth) at the same spatial locations to train the neurons in MLP based HMM. The weights were adjusted until fractions of LC obtained from HMM were nearly the same as that of LISS-III (desired output). The learned network was applied on the test data set that included all the abundance values for all the classes in the entire image obtained from LMM. The HMM outputs were six abundance maps, one for each class.

5. Experimental Results and Discussion

5.1. Simulated Data

Three images from the 200-bands are shown in Figure 4 and the classified output of the 250×250 hyper spectral 200 bands data is shown in Figure 5. The proportions of each of the four minerals were computed based on 10×10 groups of pixels for 625 groups [(250×250) divided by (10×10)]. N-FINDR was used to extract the endmembers from the synthetic mixed pixels, which are shown in Figure 6. The endmembers identified by the algorithm (drawn in red) have a good match with the actual ones (green in color). Abundances of each of the minerals from the artificial mixed pixels obtained from LMM are as shown in Figure 7 b–e. Figure 7 a is the 10 times down-sampled image of the original mineral classified image (250×250) shown in Figure 5 to compare hard classification with the abundance map visually. A three-layer MLP architecture was made with four input, one hidden and four output layers.

Figure 4. A 200 band hyperspectral image generated from spectral libraries of four different minerals: (a) band 1; (b) band 100; (c) band 200.

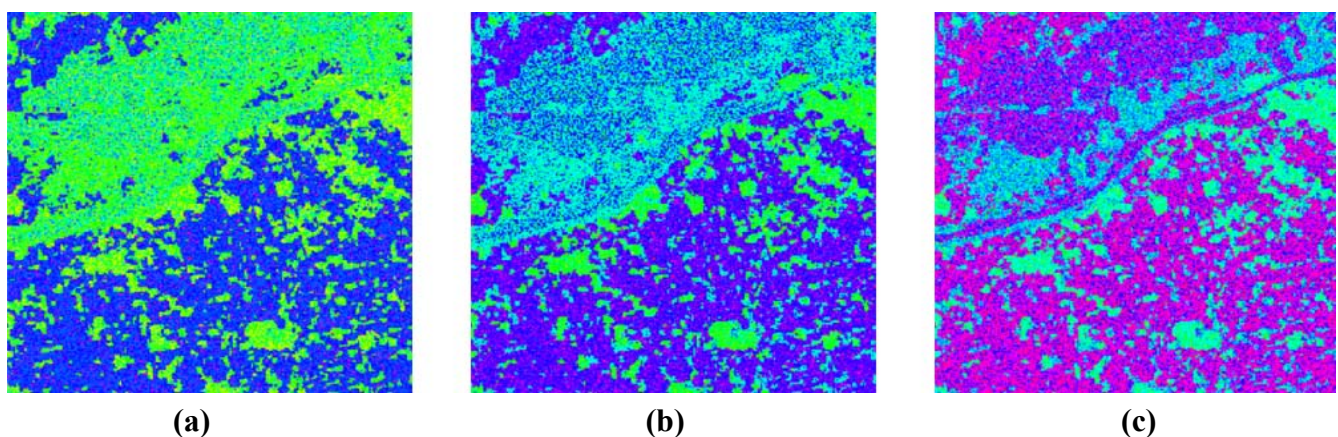


Figure 5. Mineral classified map.

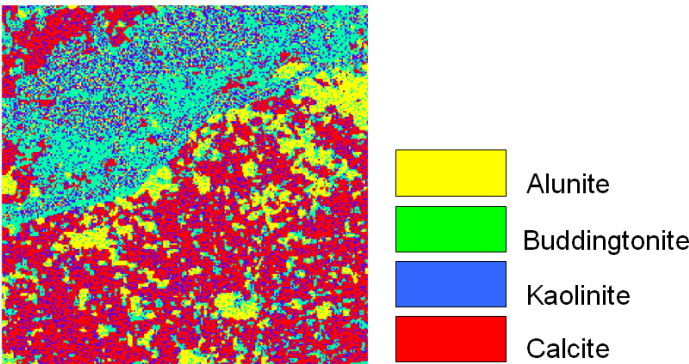


Figure 6. Comparison between the true endmembers and endmembers computed from the N-FINDR algorithm. (X-axis: band number, Y-axis: reflectance value.)

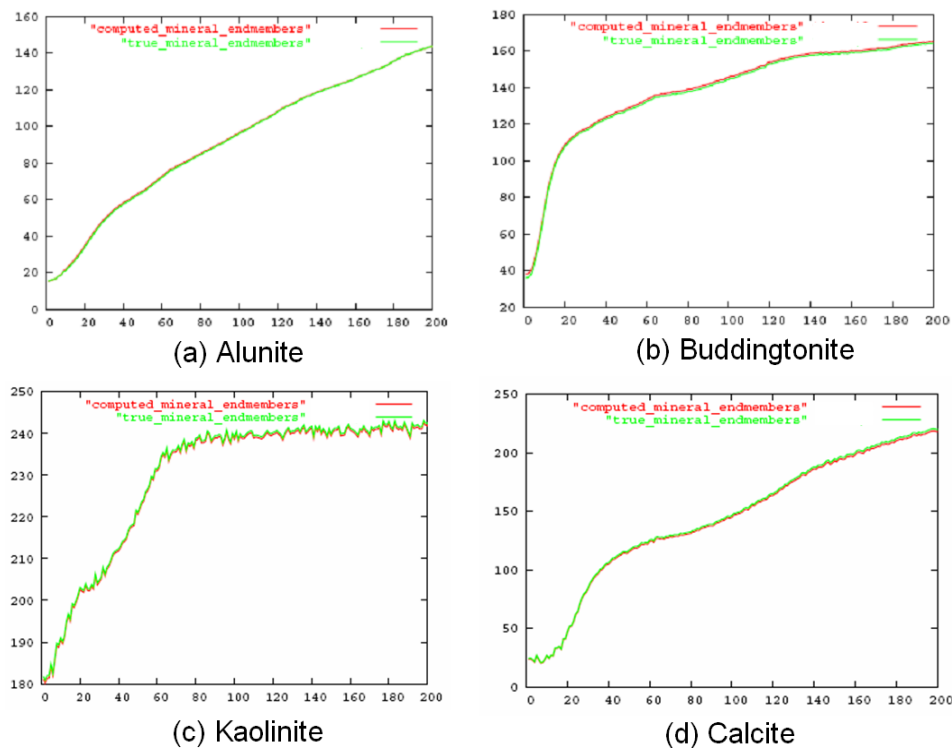
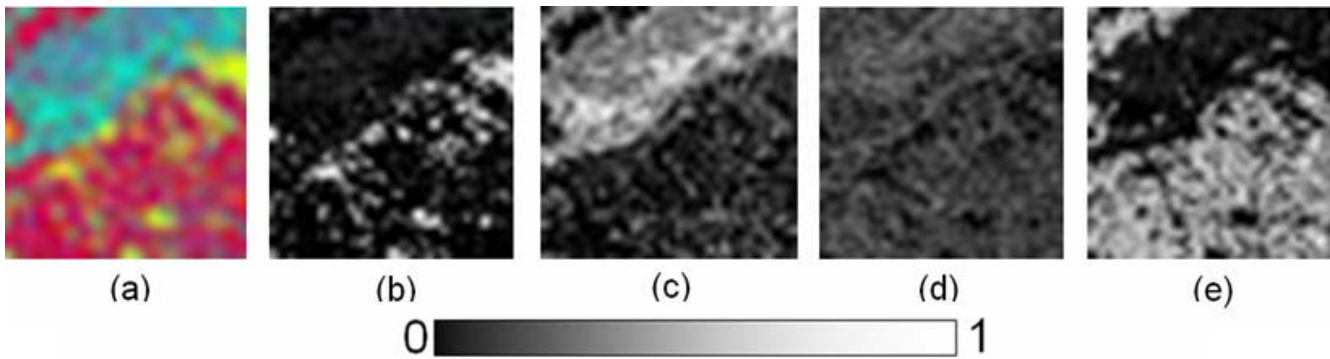


Figure 7. (a) Ten times down-sampled image of the original mineral classified image; (b)–(e) abundances maps of the four minerals obtained from linear mixture model (LMM).



The number of hidden nodes in the hidden layer, learning rate, momentum and epoch were varied in steps to estimate the best abundance values that could account for the non-linearity in the mineral mixtures (as shown in Figure 8), until the performance saturated. Table 1 lists the values of the training parameters along with the training time and the overall RMSE of the MLP network for every 500 epochs. Three measures of performance were used to evaluate the output from artificial dataset—RMSE, correlation, Bivariate Distribution Functions (BDFs). BDF is helpful to visualize the accuracy of prediction by mixture models. BDFs were plotted against the real proportions as shown in Figure 9. Pearson's product-moment correlation at 95% confidence interval and RMSE between the actual and estimated proportion from LMM and HMM are given in Table 2. The average RMSE of the LMM was 0.0089 ± 0.0022 while the average RMSE of the HMM was 0.0030 ± 0.0001 demonstrating the superiority of the HMM over the LMM. The MLP network can successfully approximate virtually any function when trained correctly.

Figure 8. Abundances maps of the four minerals obtained from HMM.

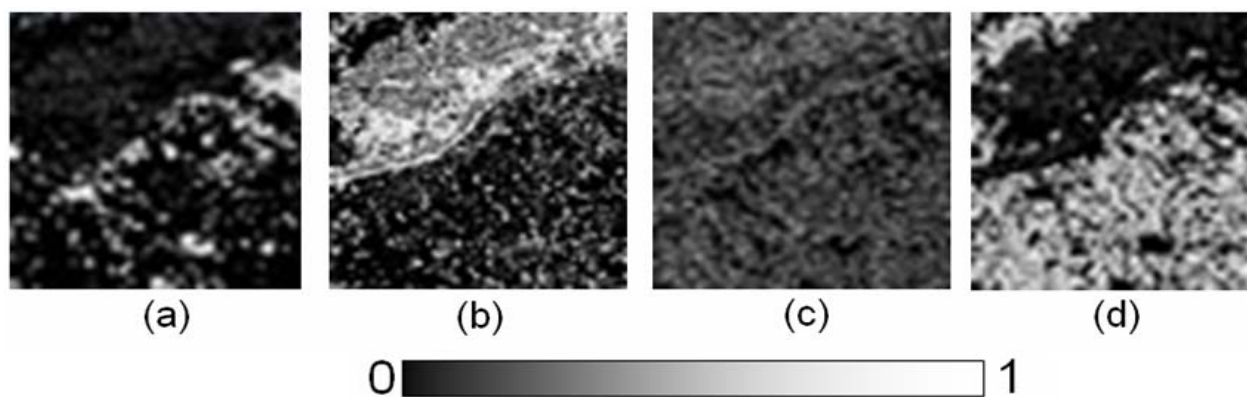


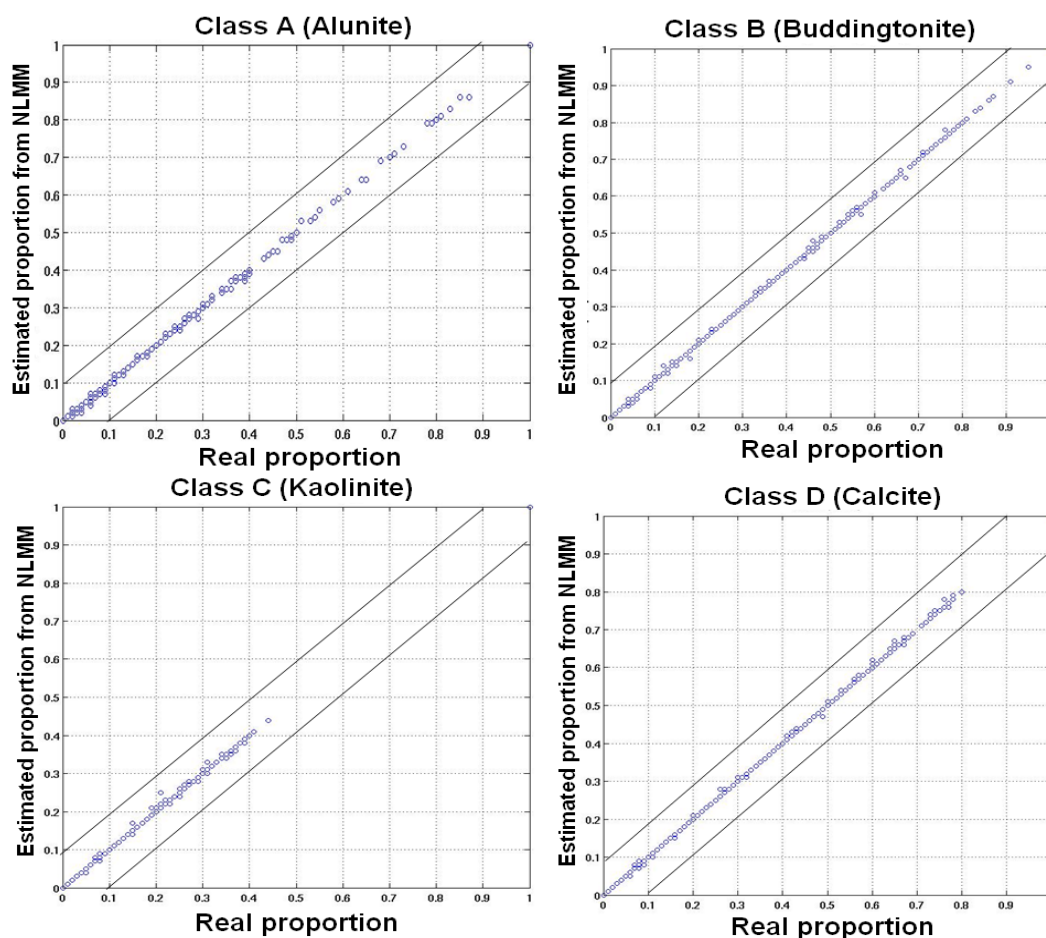
Table 1. Details of training for unmixing of simulated dataset.

No. of epochs	Learning rate	Momentum term	Training time (sec)	Unmixing time (sec)	Overall RMSE
500	0.90	0.5	4	8	0.0160
1000	0.85	0.4	5	8	0.0117
1500	0.80	0.3	7	7	0.0030
2000	0.70	0.2	7	6	0.0071
2500	0.60	0.1	8	5	0.0115

Table 2. Correlation and RMSE between actual and predicted proportions for simulated data.

Classes	Correlation (r) ($p < 2.2e^{-16}$)		RMSE	
	LMM	HMM	LMM	HMM
Alunite	0.67	0.97	0.0120	0.0032
Buddingtonite	0.71	0.98	0.0073	0.0029
Kaolinite	0.73	0.98	0.0088	0.0031
Calcite	0.75	0.99	0.0076	0.0029

Figure 9. Bivariate Distribution Functions (BDFs) of simulated test data for the four minerals obtained from HMM.



5.2. MODIS Data

In order to validate the MODIS unmixed image, a LISS-III classified image with an overall accuracy of 95.63% and individual class producers accuracy ranging from 92% to 97% and users accuracy ranging from 88% to 98% was used. Linear unmixing was applied on MODIS data to obtain the abundance maps. 15% MODIS abundance pixels obtained from LMM were randomly selected to relate with the corresponding LISS-III classified pixels (as ground truth) at the same geographical locations to train the neurons in HMM. MLP architecture with seven inputs (since seven bands of MODIS data were used), one hidden and six output layers (as six different LC classes) was constructed. The MLP based HMM was executed with varied learning rates, momentum and epochs. The momentum term and the learning rate were altered after every 500 epochs. Table 3 lists the values of the training parameters along with the training time and the overall RMSE of the MLP network on the MODIS images after every 500 epochs. The fraction maps obtained from LMM and HMM are shown in Figures 10 b–g and 11b–g.

BDFs against the real and estimated proportions from MODIS data for LMM and HMM were plotted as shown in Figure 12 and Figure 13 respectively. The Pearson's product-moment correlation at 95% confidence interval and RMSE between the actual and estimated proportion from LMM and HMM are given in Table 4.

Table 3. Details of training for unmixing of MODIS images.

No. of epochs	Learning rate	Momentum term	Training time (sec)	Unmixing time (sec)	Overall RMSE
500	0.90	0.05	25	11	0.0220
1000	0.85	0.05	22	11	0.0197
1500	0.80	0.03	22	10	0.0195
2000	0.70	0.02	18	9	0.0191
2500	0.60	0.01	18	8	0.0195

Figure 10. (a) LISS-III classified map resampled to 100×100 pixels. Abundance maps for (b) agriculture; (c) built-up/settlement; (d) forest; (e) plantation/orchard; (f) wasteland/orchard; (g) water bodies obtained from LMM.

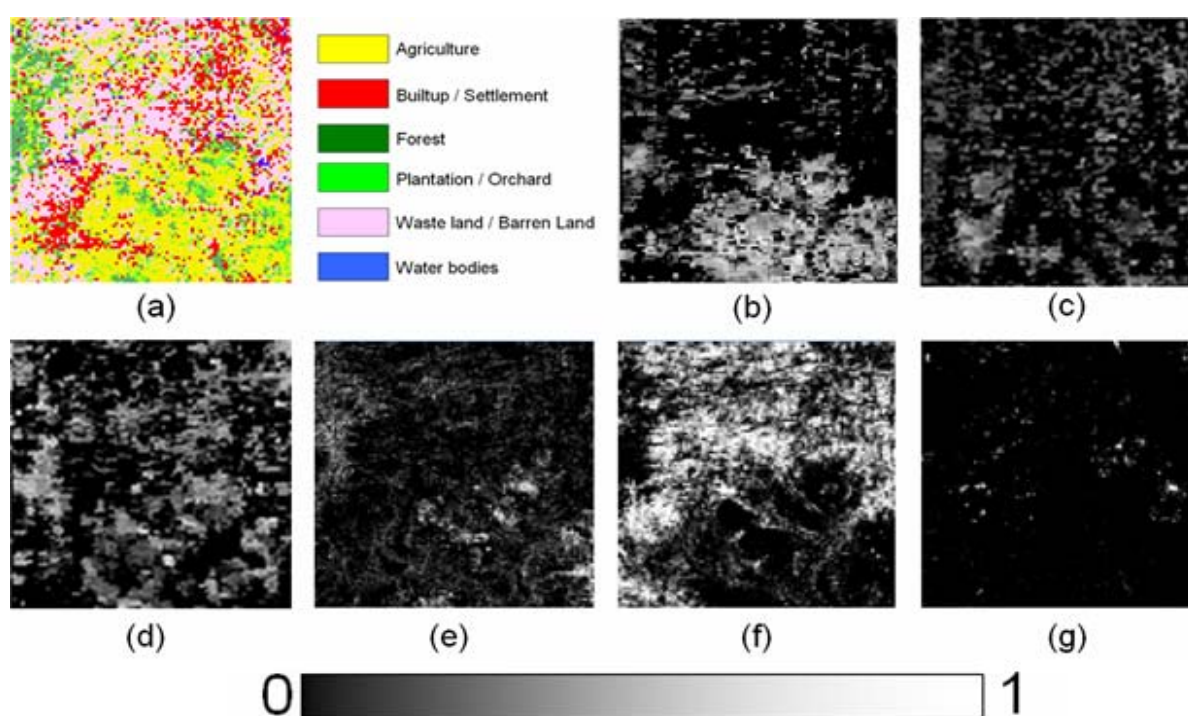


Figure 11. (a) LISS-III classified map resampled to 100×100 pixels. Abundance maps for (b) agriculture; (c) built-up/settlement; (d) forest; (e) plantation/orchard; (f) wasteland/orchard, (g) water bodies obtained from HMM.

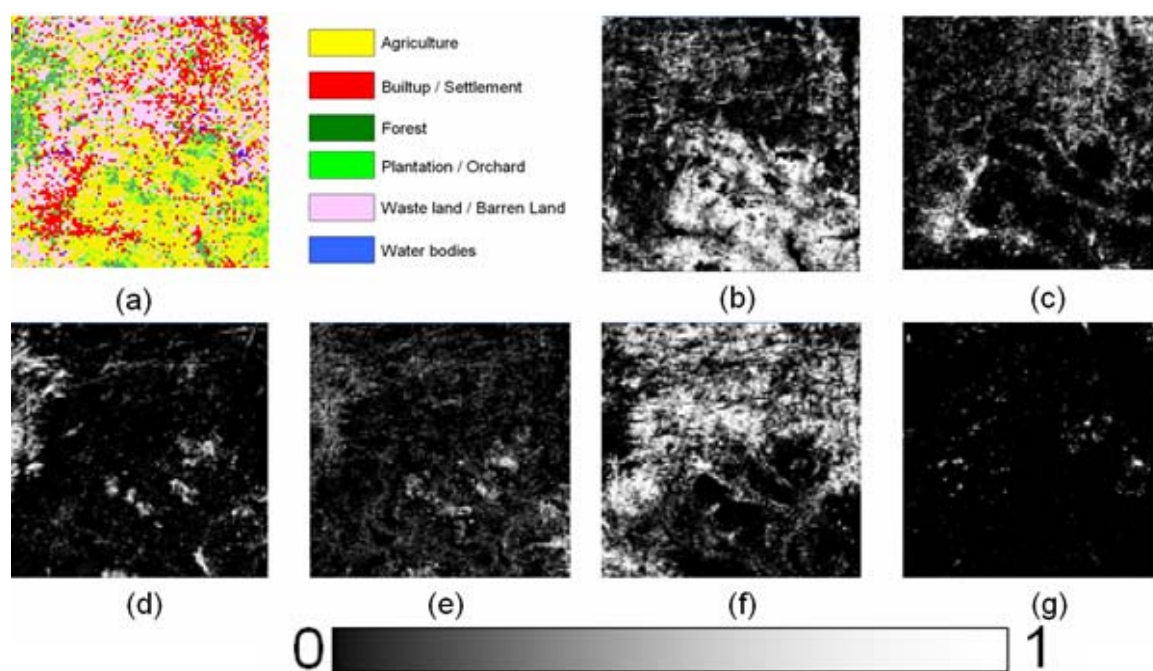


Figure 12. BDFs of MODIS test data from LMM. (a) agriculture; (b) built-up/settlement; (c) forest; (d) plantation/orchard; (e) wasteland/orchard; (f) water bodies.

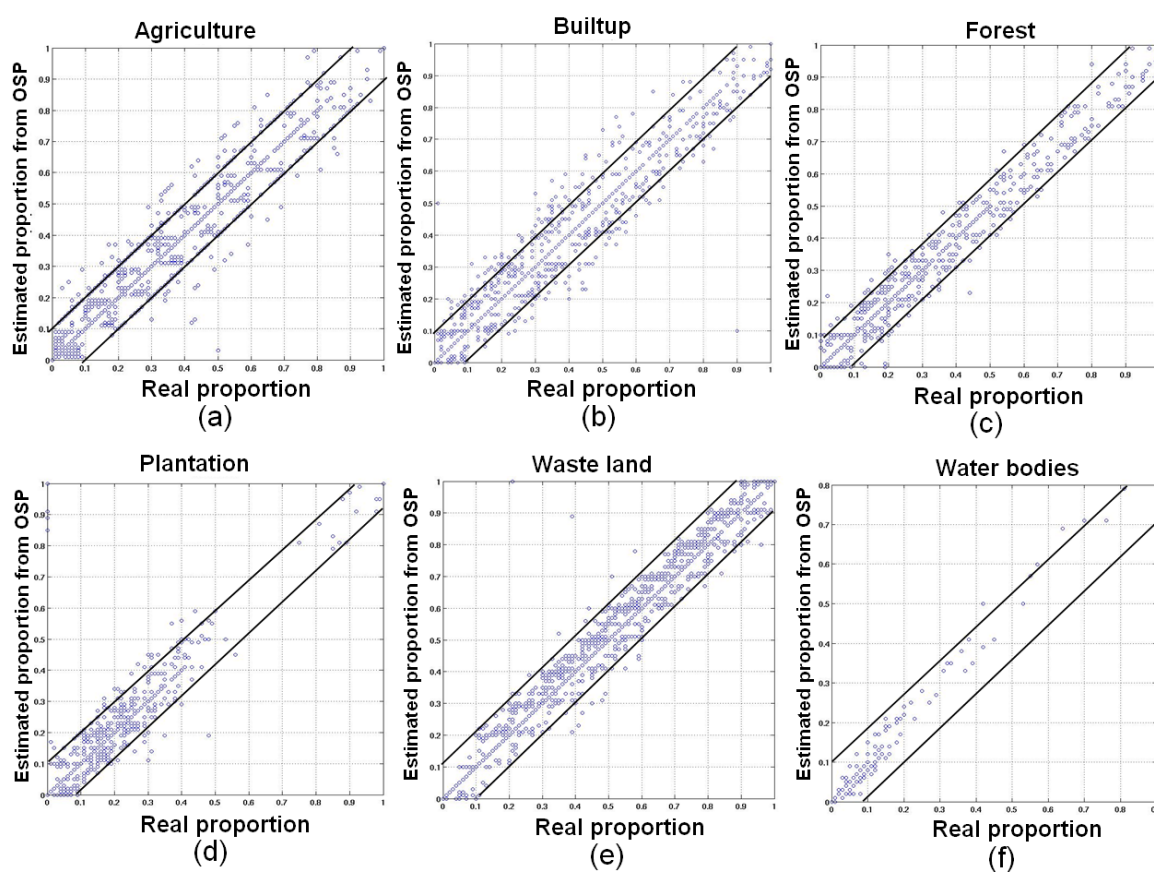


Figure 13. BDFs of MODIS test data from HMM. (a) agriculture; (b) built-up/settlement; (c) forest; (d) plantation/orchard; (e) wasteland/orchard; (f) water bodies.

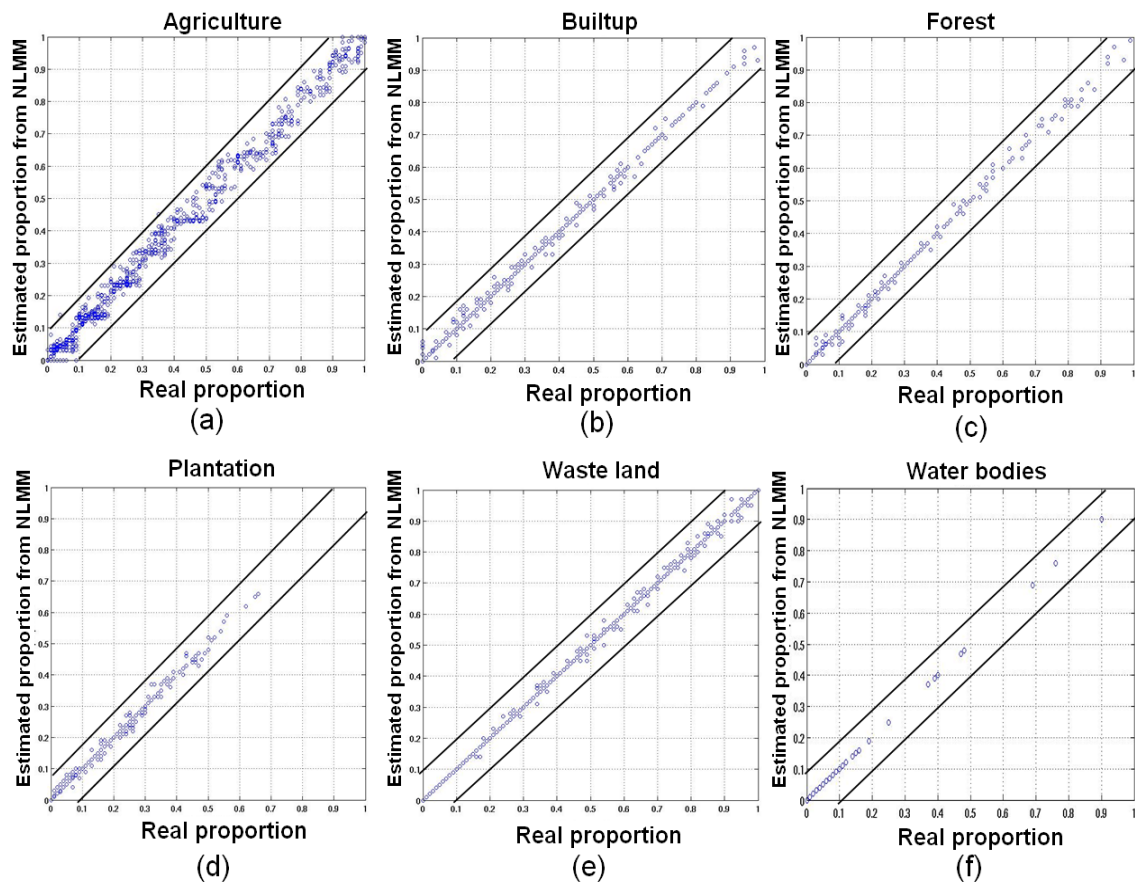


Table 4. Correlation and RMSE between real/reference and predicted proportions for MODIS data.

Classes	Correlation (r) ($p < 2.2e^{-16}$)		RMSE	
	LMM	HMM	LMM	HMM
Agriculture	0.6730	0.9110	0.0518	0.0271
Builtup/Settlement	0.6390	0.9345	1.0519	0.0083
Forest	0.7310	0.9411	0.0257	0.0062
Plantation/Orchard	0.6990	0.9447	0.0280	0.0061
Waste/Barren land	0.6599	0.9342	0.0431	0.0073
Water bodies	0.7799	0.9855	0.0061	0.0016

The 100×100 pixels from both LMM (Figure 14) and HMM (Figure 15) abundance maps were subjected to error distribution graphs to get a general idea of the range and magnitude of the errors (absolute values of the prediction error *i.e.*, difference between real and estimated class proportions) and their location. Agriculture class had the highest error among the other classes and water had the lowest error. Agriculture had more prediction error in the area marked by circles mainly existing in the mixing region. The HR classified image showed a mixture of agricultural and wasteland in that area. Here, the similarity of the two samples in feature space has caused a learning difficulty so that the model generates prediction errors. Field verification revealed that some of the agricultural parcels were

fallow (non-cultivation period), which is very similar to barren/open ground and this resulted in a mismatch. No similar errors were found in other spatial locations since fallow lands were absent. The error distribution patterns are also seen in Figure 16. For each class, the graph indicates how many predictions fall within a given percentage of field measurement. It is observed that the percentage of samples with zero prediction error is high and almost equal for built-up, forest, plantation and wasteland. Water class has the least samples with zero error followed by agriculture. But, with the increase of error bound, water class outperforms agriculture class after the error bound is over 10%.

The non-linear model uses abundance vectors and associated true mixture vectors from a training set to estimate mixtures from spectra by interpolation. MLP has several experimental parameters that affect the model's performance at a certain level such as the number of hidden layers, learning rate, momentum, epoch, *etc.* We started with a simple architecture with one hidden layer and varied the other parameters.

The algorithm took 7 seconds to train and 7 seconds to generate the fraction images (refer to Table 1). For MODIS data, the performance of MLP model was stable and acceptable with lowest RMSE in 2000 epochs, 0.7 learning rate, 0.02 momentum and took 18 seconds to train and 9 seconds to generate fraction images (refer Table 3). The correlation is high and significant for both artificial and MODIS data for the HMM compared to the LMM. This implies that training constitutes a vital stage in MLP-based classification. Fifteen percent of the LMM based abundances along with the actual abundances were used for training and the MLP model could interpolate to produce many more combinations of class proportions to match the testing samples.

Figure 14. Error distribution of MODIS abundance obtained from LMM (X and Y axes are the two dimensions in feature space and Z axis is the absolute difference between real/reference and estimated class proportion) for the six classes.

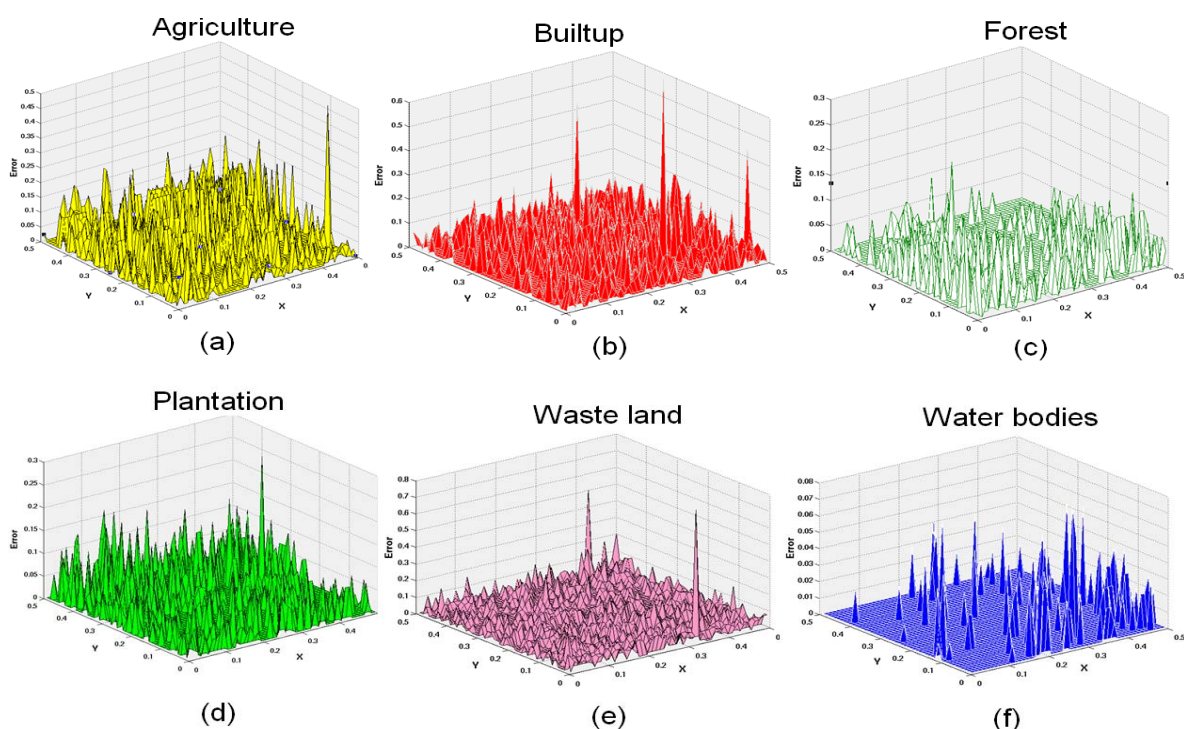


Figure 15. Error distribution of MODIS abundance obtained from HMM (X and Y axes are the two dimensions in feature space and Z axis is the absolute difference between real/reference and estimated class proportion) for the six classes.

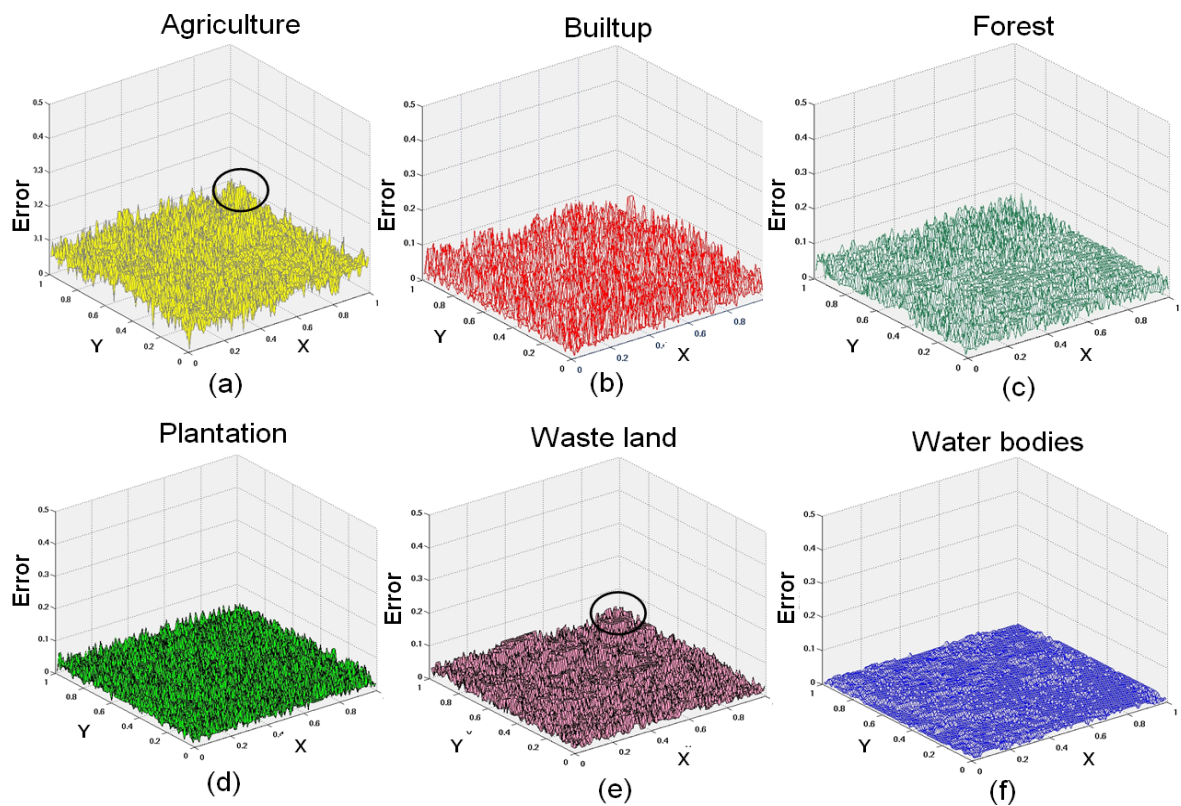
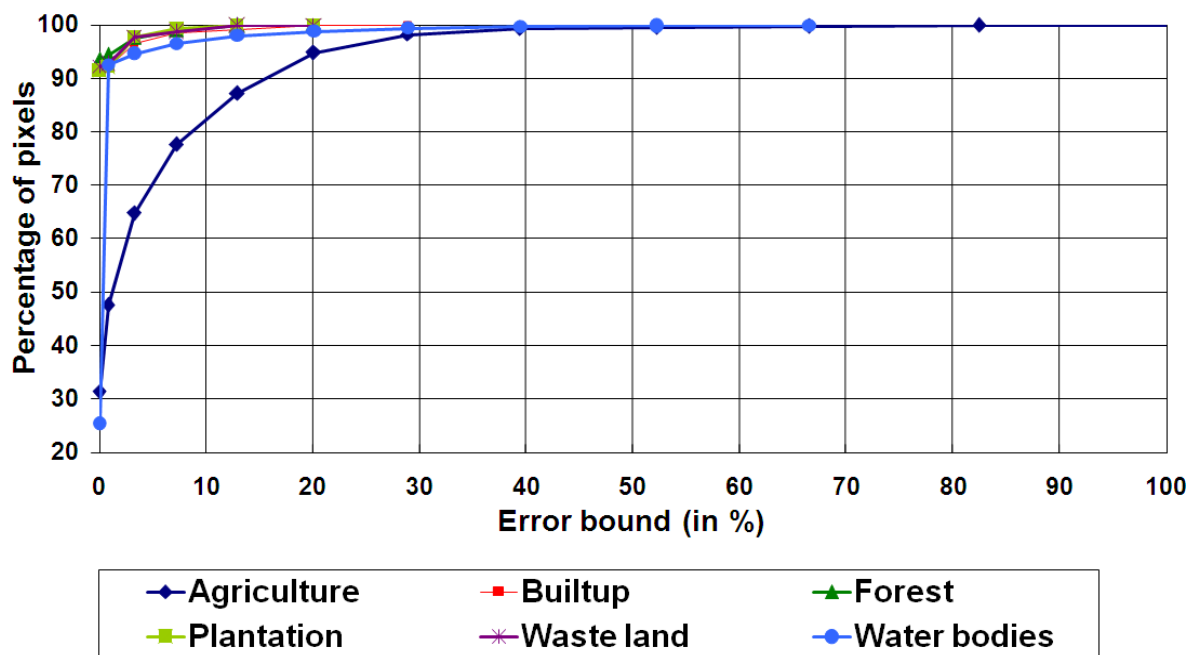
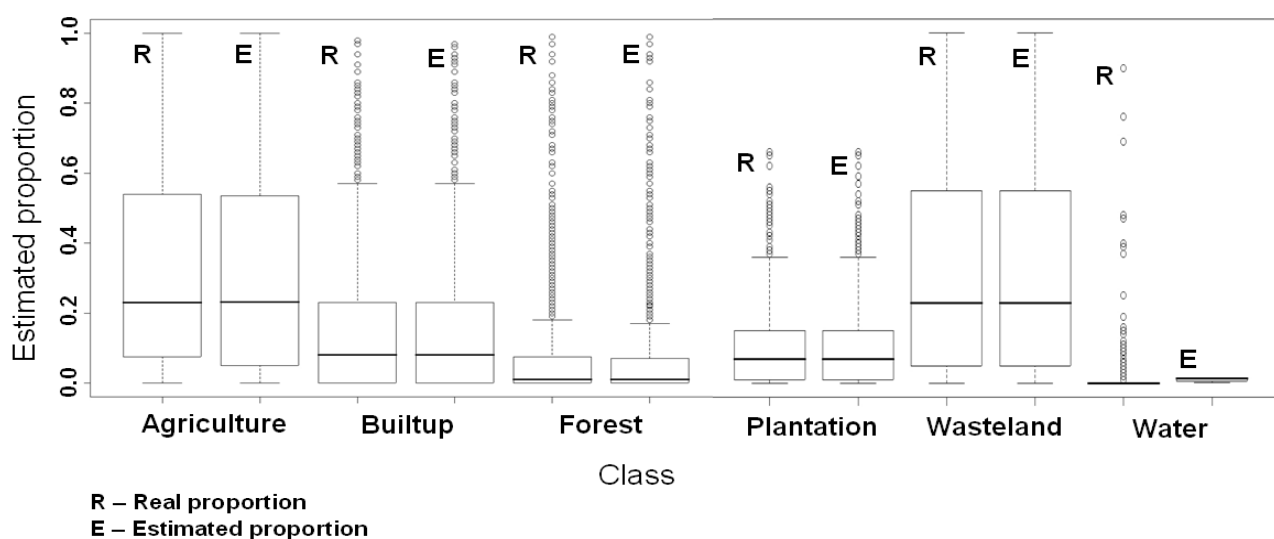


Figure 16. For the HMM, the graph shows the percent of pixels that lie within a given percent of the actual class proportion.



Box plot analysis of the real and estimated proportions for each class indicated that all the classes except water had medians at the same location for both real and estimated proportions. Many pixels both have absence (0%) and full presence (100%) of agriculture. Since the median lies towards the 1st quartile, most of the pixels have less abundance (Figure 17). Many pixels have zero or only a small proportion of built-up and forest with some outliers. Wasteland presents a similar picture as that of agriculture with both absence and full presence, indicating that agriculture and wasteland are the two dominant categories in the area. There is a much smaller number of water bodies and most of them are small in area ($\leq 1000 \text{ m}^2$), unevenly distributed as captured by classified image. However, these have not been properly estimated by the algorithm as can also be seen in the box plot. This could be due to the size of water bodies that are smaller than the intrinsic scale of the MODIS pixels, and hence the absence of pure pixel for endmembers, which was not detected by N-FINDR algorithm. The experiments performed in this analysis also indicated areas where non-linear detection methods for targets in hyper spectral imagery have proven to be more effective than linear methods. These scenarios may be targets of small abundance situations, where small target areas are dispersed throughout the image or situations where accuracy is required. Since the approach in this work is a hybrid of linear and non-linear estimators, it can be easily acclimatized to either the linear or non-linear mixture model.

Figure 17. Box plot of the actual and estimated proportions.



Initialization and training are the two issues in a MLP network that have to be dealt with carefully. However, an effective learning algorithm should not depend on initial conditions, which can only affect the convergence rate but should not alter the final results [32]. This is not the case of learning algorithms used for NNs. In order for a mixture model to be effective, initial values must be representative and cannot be arbitrary. In a recent study, Plaza *et al.* [33] used ANN based models to select higher informative samples in order to effectively train the neural architecture using (1) a border-training algorithm which selects training samples located in the vicinity of the hyperplanes that can optimally separate the classes; (2) a mixed-signature algorithm to select the most spectrally mixed pixels; and (3) a morphological-erosion algorithm which incorporates spatial information through mathematical morphology to select spectrally mixed training samples located in spatially

homogeneous regions. The experimental results were demonstrated using non-linear mixed spectra with absolute ground truth.

The non-linear mixture output also depends on selection of endmembers, which is an important issue for successful application of this approach. One of the potential drawbacks of the N-FINDR algorithm is that it requires at least one pixel in the image to be pure. This may not always be the case. In such cases, alternative methods of endmember extraction have to be studied and integrated in the algorithm. Also, the simulated dataset in our work is noise-free (has no noise component) so, further research is required to analyze the impact of the interference of noise.

6. Conclusions

The proposed HMM algorithm integrates the concept of both linear and non-linear mixing. Endmembers are first extracted from the image, eliminating the undesired spectral signatures followed by unmixing and then interpolating the fractions for the whole image. Unmixed MODIS output was compared with a HR classified output (IRS LISS-III MS) to estimate the accuracy. It showed that the overall RMSE of HMM was 0.0191 ± 0.022 as compared to the LMM output considered alone that had an overall RMSE of 0.2005 ± 0.41 indicating that individual class abundances obtained from HMM are very close to what is present on the ground and observed in the HR classified output. This emphasizes that the influence due to multiple reflections among ground cover targets has to be considered for the abundance estimation. While a linear detection method might work adequately for many scenarios, a non-linear model would perform better. The only disadvantage of this method is that the endmember selection through N-FINDR requires at least one pure pixel. Future work will involve developing methods to obtain pure pixels when there are no endmembers in the scene.

Acknowledgments

We thank the NRDMS Division, the Ministry of Science and Technology (DST), Government of India and Indian Institute of Science for their financial and infrastructure support. Land Processes Distributed Active Archive Center (LP DAAC) and National Remote Sensing Centre (NRSC), Hyderabad, India are acknowledged for providing the MODIS and IRS LISS-III data, respectively.

References

1. Guilfoyle, K.J.; Althouse, M.L.; Chang, C.-I. A quantitative and comparative analysis of linear and nonlinear spectral mixture models using radial basis neural networks. *IEEE Trans. Geosci. Remote Sens.* **2001**, *39*, 2314–2318.
2. Atkinson, P.M.; Cutler, M.E.J.; Lewis, H. Mapping sub-pixel proportional land cover with AVHRR imagery. *Int. J. Remote Sens.* **1997**, *18*, 917–935.
3. Keshava, N.; Mustard, J.F. Spectral unmixing. *IEEE Signal Process. Mag.* **2002**, *19*, 44–57.
4. Kumar, U.; Kerle, N.; Ramachandra, T.V. Constrained linear spectral unmixing technique for regional land cover mapping using MODIS data. In *Innovations and Advanced Techniques in Systems, Computing Sciences and Software Engineering*; Elleithy, K., Ed.; Springer: Berlin, Germany, 2008; pp. 87–95.

5. Foody, G.M.; Cox, D.P. Subpixel land cover composition estimation using a linear mixture model and fuzzy membership functions. *Int. J. Remote Sens.* **1994**, *15*, 619–631.
6. Kanellopoulos, I.; Varfis, A.; Wilkinson, G.G.; Megier, J. Land-cover discrimination in SPOT imagery by artificial neural network—a twenty class experiment. *Int. J. Remote Sens.* **1992**, *13*, 917–924.
7. Liu, W.; Wu, E.Y. Comparison of non-linear mixture models: Sub-pixel classification. *Remote Sens. Environ.* **2005**, *94*, 145–154.
8. Ju, J.; Kolaczyk, E.D.; Gopal, S. Gaussian mixture discriminant analysis and sub-pixel and sub-pixel land cover characterization in remote sensing. *Remote Sens. Environ.* **2003**, *84*, 550–560.
9. Arai, K. Non-linear mixture model of mixed pixels in remote sensing satellite images based in Monte Carlo simulation. *Adv. Space Res.* **2008**, *42*, 1715–1723.
10. Mustard, J.F.; Sunshine, J.M. Spectral analysis for Earth science: investigations using remote sensing data. In *Remote Sensing for the Earth Sciences: Manual of Remote Sensing*, 3rd ed.; Rencz, A.N., Ed.; John Wiley & Sons: New York, NY, USA, 1999; Volume 3, pp. 251–306.
11. Adams, J.B.; Gillespie, A.R. *Remote Sensing of Landscapes with Spectral Images: A Physical Modeling Approach*; Cambridge University Press: New York, NY, USA, 2006.
12. Plaza, J.; Martinez, P.; Pérez, R.; Plaza, A. Nonlinear neural network mixture models for fractional abundance estimation in AVIRIS Hyperspectral Images. In *Proceedings of the NASA Jet Propulsion Laboratory AVIRIS Airborne Earth Science Workshop*, Pasadena, CA, USA, 31 March–2 April 2004; pp. 1–12.
13. Carpenter, G.A.; Gopal, S.; Macomber, S.; Martens, S.; Woodcock, C.E. A neural network method for mixture estimation for vegetation mapping. *Remote Sens. Environ.* **1999**, *70*, 138–152.
14. Cantero, M.C.; Pérez, R.; Martínez, P.; Aguilar, P.L.; Plaza, J.; Plaza, A. Analysis of the behaviour of a neural network model in the identification and quantification of hyperspectral signatures applied to the determination of water quality. In *Chemical and Biological Standoff Detection II*, Proceedings of SPIE Optics East Conference, Philadelphia, PA, USA, 25–28 October 2004; Jensen, J.O., Thériault, J.-M., Eds.; SPIE: Bellingham, WA, USA, 2004; pp. 174–185.
15. Shimabukuro, Y.E.; Smith, A.J. The least-squares mixing models to generate fraction images derived from remote sensing multispectral data. *IEEE Trans. Geosci. Remote Sens.* **1991**, *29*, 16–20.
16. Liu, W.; Gopal, S.; Woodcock, C. ARTMAP Multisensor-resolution framework for land cover characterization. In *Proceedings 4th Annual Conference information fusion*, Montreal, Canada, 7–10 August, 2001; pp. 11–16.
17. Liu, W.; Seto, K.; Wu, E.; Gopal, S.; Woodcock, C. ART-MMAP: A neural network approach to subpixel classification. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 1976–1983.
18. Winter, M.E. N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data. In *Proceedings of the SPIE: Imaging Spectrometry V*; Descour, M.R., Shen, S.S., Eds.; SPIE: Bellingham, WA, USA, 1999; Volume 3753, pp. 266–275.
19. Chang, C.-I. Orthogonal Subspace Projection (OSP) Revisited: A comprehensive study and analysis. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 502–518.
20. Harsanyi, J.C.; Chang, C.-I. Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach. *IEEE Trans. Geosci. Remote Sens.* **1994**, *32*, 779–785.

21. Crippen, R.E.; Bloom, R.G. Unveiling the lithology of vegetated terrains in remotely sensed imagery via forced decorrelation. In *Proceedings of the Thirteenth International Conference on Applied Geologic Remote Sensing*, Vancouver, Canada, 1–3 March 1999; p. 150.
22. Tu, T.-M.; Chen, C.-H.; Chang, C.-I. A noise subspace projection approach to target signature detection and extraction in an unknown background for hyperspectral images. *IEEE Trans. Geosci. Remote Sens.* **1998**, *36*, 171–181.
23. Settle, J.J. On the relationship between spectral unmixing and subspace projection. *IEEE Trans. Geosci. Remote Sens.* **1996**, *34*, 1045–1046.
24. Nielsen, A.A. Spectral mixture analysis: Linear and semi-parametric full and iterated partial unmixing in multi-and hyperspectral image data. *Int. J. Comput. Vis.* **2001**, *42*, 17–37.
25. Haykin, S. *Neural Networks: A Comprehensive Foundation*. Macmillan: New York, NY, USA, 1998.
26. Duda, R.O.; Hart, P.E.; Stork, D.G. *Pattern Classification*. Wiley: Indianapolis, IN, USA, 2000; pp. 517–598.
27. Kavzoglu, T.; Mather, P.M., The use of backpropagating artificial neural networks in land cover classification. *Int. J. Remote Sens.* **2003**, *24*, 4907–4938.
28. Mas, J.F. Mapping land use/cover in a tropical coastal area using satellite sensor data, GIS and artificial neural networks. *Estuar. Coast. Shelf Sci.* **2003**, *59*, 219–230.
29. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–535.
30. Kumar, U.; Raja, K.S.; Mukhopadhyay, C.; Ramachandra, T.V. A Multi-layer Perceptron based Non-linear Mixture Model to estimate class abundance from mixed pixels. In *Proceedings of the 2011 IEEE Students' Technology Symposium*, Indian Institute of Technology, Kharagpur, India, 14–16 January 2011; pp. 148–153.
31. Jet Propulsion Laboratory (JPL) Homepage. Available online: <http://speclib.jpl.nasa.gov/> (accessed on 27 July 2009).
32. Plaza, J.; Plaza, A.; P'erez, R.; Martinez, P. Joint linear/nonlinear spectral unmixing of hyperspectral image data. In *Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS'07)*, Barcelona, Spain, 23–27 July 2007; pp. 4037–4040.
33. Plaza, J.; Plaza, A.; Perez, R.; Martinez, P. On the use of small training sets for neural network-based characterization of mixed pixels in remotely sensed hyperspectral images. *Pattern Recognition* **2009**, *42*, 3032–3045.