

# Gaussian Mixture Discriminant Analysis and Sub-pixel Land Cover Classification in Remote Sensing<sup>1</sup>

*Junchang Ju<sup>†</sup>, Eric D. Kolaczyk<sup>‡</sup>, and Sucharita Gopal<sup>†2</sup>*

<sup>†</sup> Department of Geography and <sup>‡</sup> Department of Mathematics and Statistics  
Boston University, Boston, MA 02215.

## Abstract

Mixture analysis is a necessary component for capturing sub-pixel heterogeneity in the classification of land cover from remotely sensed images. Mixture analysis approaches in remote sensing vary from simple, conventional linear mixture models to nonlinear neural network mixture models. Linear mixture models are fairly simple and generally result in poor classification accuracy. Neural network models can achieve much higher accuracy, but typically lack interpretability. In this paper we present a mixture discriminant analysis (MDA) model for inferring fractional land covers within forest stands from Landsat Thematic Mapper images. Specifically, individual classes are modeled as mixtures of subclasses of Gaussian distributions, and classification is performed based on the appropriate posterior distribution. Compared to a benchmark study on mixture models with Plumas National Forest data, this MDA model easily outperforms traditional linear mixture models and parallels the performance of the ARTMAP neural network mixture model. In other words, the MDA model is observed to successfully combine the performance characteristics of more complex neural network models (due to the nonlinear nature of its classification rules), with the ease of interpretation associated with linear mixture models (due to its relatively simple structure). MDA models therefore offer a viable alternative for addressing the mixture modeling problem in remote sensing.

## 1 Introduction

The extraction of land cover information from remote sensing images traditionally is viewed as a classification problem. Satellite sensors integrate the spectral radiation coming from an area on the land surface known as the Instantaneous Field of View (IFOV) and discretize it to integers known as digital numbers (DN). These numbers can then be displayed as an image in a variety of ways. IFOV loosely refers to spatial resolution, or relative ground area per pixel. For Landsat Thematic Mapper(TM), the spatial resolution is 30m × 30m

---

<sup>1</sup>This research was supported by NSF Grant BCS 0079077 and ONR Award N00014-99-1-0219.

<sup>2</sup>Correspondence author, E-mail: suchi@bu.edu

per pixel except for the thermal band. TM has 6 solar reflective spectral bands located in the visible, near infrared, and shortwave infrared spectral region and 1 thermal band in the thermal infrared.

Most classification algorithms assume that a pixel represents a pure class and each pixel is labeled as one of only a few possible classes. In reality, however, a pixel may contain a mixture of two or more land cover types due to the limited spatial resolution of the image and the intrinsic mixed nature of most land covers. Even if a classification system does not assume homogeneity of land cover categories, the usually fairly limited number of classes cannot fully capture the actual complexity of land cover information. For example, a classification system may define a pixel of 80% or more fractional coverage of *conifer* as *conifer* land cover. Then a labeled *conifer* pixel in the classification map realistically may contain *conifer* coverage anywhere between 80% and 100% and 20% or less of *brush* and/or other classes. Hence, there are a variety of ways in which discretization of land cover into categories contributes to a loss of information.

Mixture modeling in remote sensing aims to capture the respective proportions of land cover classes within pixels and characterize land cover more accurately. A mixture map represents the proportions of each pure land cover within pixels. This information is very important for land resource management and global climate modeling [7]. One nontrivial use of the mixture information is that discrete classification maps of any type can be produced out of the continuous land cover information if desired. Mixture-based classifiers typically require training pixels, often called “image endmembers,” which can be either pure pixels [6] or mixed pixels with known class proportions [12, 21].

Attempts to model mixtures in remote sensing include linear mixture models [1, 18, 19, 21], neural networks [6, 12, 9, 2], fuzzy classifier [8], maximum likelihood classifiers [10, 20], and decision trees [16]. Prior research has shown that linear mixture models generate poor or moderate results. However, as demonstrated by [4], linear mixture models may not be suitable in cases when multiple scattering results in nonlinear mixture. In this context, a nonlinear model can produce better results. Atkinson *et al.* [2] applied Multilayer Perceptron neural network based mixture models to decompose AVHRR imagery. The mixture information from their model was better compared to that generated through linear mixture models and fuzzy c-means classifiers. Foody [9] used a simple regression and contouring based approach to produce a sub-pixel land cover map based on fuzzy classification. Carpenter *et al.* [6] present a nonlinear algorithm for mixture estimation based on an ARTMAP neural network, and use that for identifying life form components of the vegetation mixture. Landsat TM imagery is used to estimate the sub-pixel information for the life-form components. The ARTMAP-based mixture model was able to capture nonlinear effects and thus performed better than the conventional linear mixture models. ARTMAP is based on Adaptive Resonance Theory (ART), as introduced by Grossberg [14]. The main feature of ART systems is a pattern matching process that compares the current input with a selected learned category representation. ART systems develop stable clusters in response to arbitrary sequences of input patterns by self-organization. Details on ART models are provided in [5].

For purposes of comparison, we will take linear mixture methods and ARTMAP methods as representative extremes from this literature. The commonly used linear mixture method is effectively a constrained linear model, wherein the spectra of a mixed pixel is modeled as a linear combination of endmember spectra (with weights constrained to be positive and sum to one). Independent sampling and a common, multivariate Gaussian error model are assumed, and least squares is used to estimate class proportions from observed data. The end result is a linear classifier that, while capturing the effect of land cover mixing at the level of the mean, is unable to truly match the multi-modal nature of the underlying data. Correspondingly, these methods tend to exhibit a relatively poor level of classification accuracy in most situations. On the other hand, neural network methods like those based on the ARTMAP architecture of Carpenter *et al.* [6], have been reported to estimate fractional coverage with much higher accuracy, due to their ability to represent highly complex nonlinear classification rules. However, this same strength has also led to criticism that neural networks can be difficult to use and yield little in the way of explanation or interpretation because of their "black box" nature [11].

In this paper we develop a mixture discriminant analysis (MDA) model for estimating *conifer* and *hardwood* covers within stands from Landsat TM images. The use of finite mixture models, in which each class is modeled with one simple density component of an overall mixture density, has a long history in remote sensing and image analysis (e.g., [17]). However, while MDA models (in which each class is modeled as a mixture itself) have been around informally for some years now in areas like statistics and pattern recognition, they seem to have been explored formally only recently by Hastie and Tibshirani [15] within statistics and, to the best of our knowledge, have found no application in the remote sensing community to date. Employing this framework, we model each land cover class as a mixture of subclasses of multivariate Gaussian distributions. We discuss the training of this model and propose a classifier based on the posterior distribution of classes, given data. In the spirit of Carpenter *et al.* [6], and using the same data, we conduct a numerical study in which we compare our MDA-based classifier with those resulting from the traditional linear mixture model and from the ARTMAP neural network architecture. We find that with little loss in simplicity and interpretability over the linear mixing approach, our MDA approach is able to nearly match the performance of the neural network.

## 2 Mixture Discriminant Analysis (MDA) model description

We briefly describe here the mixture discriminant analysis (MDA) modeling framework, as outlined in Hastie and Tibshirani [15]. These authors consider MDA in some generality, focusing in particular on a number of variations on the basic modeling and fitting strategy, and consider its application to tasks such as the recognition of hand-written digits. Here our focus is on the adaptation

of this framework to land cover classification in remote sensing. MDA can be viewed as an extension of linear discriminant analysis (LDA), in which classes are modeled as a mixtures of subclasses, with each subclass represented by a Gaussian distribution. For remote sensing this distinction is crucial, as the assumption of a single distribution for each class (as in a conventional maximum likelihood classifier (MLC)) typically does not match the nature of the variation in the data. Even pure forest pixels can show variation as a result of species, age, etc, let alone mixed pixels.

Suppose we have  $J$  classes in our classification system and the number of subclasses in each class is  $R_j$ ,  $j = 1, 2, \dots, J$ . Then we write the mixture density for class  $j$  as

$$m_j(x) = P(X = x|G = j) = |2\pi\Sigma_j|^{-1/2} \sum_{r=1}^{R_j} \pi_{jr} \exp\{-D(x, \mu_{jr})/2\}, \quad (1)$$

where  $X$  is a vector of measurements (e.g., DN's in 6 reflective spectral bands in our examples),  $G$  refers to the class of a given object (e.g., a pixel),  $\Sigma_j$  is a covariance matrix assumed common to the mixture sub-classes of class  $j$ , and  $\pi_{jr}$  and  $\mu_{jr}$  are the mixing probability and mean of the  $r$ -th subclass of the  $j$  class. The quantity  $D(x, \mu_{jr}) = (x - \mu_{jr})^T \Sigma_j^{-1} (x - \mu_{jr})$  here refers to the Mahalanobis distance between  $x$  and  $\mu_{jr}$ .

The purpose of the classification in remote sensing often will decide the number  $J$  of classes. On the other hand, the number of subclasses  $R_j$  for each class  $j$  is decided from the variation of the training data, and we will elaborate on this in next section. Classification rules can be built using the posterior class probabilities which, via Bayes theorem, are of the form

$$P(G = j|X = x) \sim \Pi_j \text{Prob}(x|j) \sim \Pi_j |\Sigma_j|^{-1/2} \sum_{r=1}^{R_j} \pi_{jr} \exp\{-D(x, \mu_{jr})/2\}, \quad (2)$$

where  $\Pi_j$  is the prior probability for class  $j$ .

The parameters  $\mu_{jr}$ ,  $\pi_{jr}$  and  $\Sigma_j$  used in Equation (2) can be estimated from the training data using the EM algorithm. The EM algorithm maximizes the conditional log-likelihood,  $l^{mix}(\theta)$ , for the training data.

$$l^{mix}(\theta) = l^{mix}(\mu_{jr}, \Sigma_j, \pi_{jr}) = \sum_{i=1}^N \log m_{gi}(x_i) \quad (3)$$

Proceeding in the usual manner, it may be found that the E- and M-steps reduce to iteratively calculating the values

$$\begin{aligned} \hat{p}(c_{jr}|x, j) &= \text{Prob}(x \in r\text{th subclass of class } j|x, j) \\ &= \frac{\pi_{jr} \exp\{-D(x, \mu_{jr})/2\}}{\sum_{k=1}^{R_j} \pi_{jk} \exp\{-D(x, \mu_{jk})/2\}}, \end{aligned} \quad (4)$$

$$\hat{\pi}_{jr} \propto \sum_{g_i=j} p(c_{jr}|x_i, j), \quad \sum_{r=1}^{R_j} \hat{\pi}_{jr} = 1, \quad (5)$$

$$\hat{\mu}_{jr} = \frac{\sum_{g_i=j} x_i p(c_{jr}|x_i, j)}{\sum_{g_i=j} p(c_{jr}|x_i, j)} \quad (6)$$

$$\hat{\Sigma}_j = \frac{1}{N_j} \sum_{g_i=j} \sum_{r=1}^{R_j} p(c_{jr}|x_i, j) (x_i - \mu_{jr})(x_i - \mu_{jr})^T. \quad (7)$$

Here  $N$  refers to the total number of training data, of the form  $(x_i, g_i)$ ,  $N_j$  is the number of training observations in class  $j$ , and the notation  $\sum_{g_i=j}$  means summing over all observations of class  $j$ .

Note we make a small change to the original model in [15] through our use of the class-dependent covariance matrices in Equations (2) and (7). This change is natural for two reasons. First, as can be seen in the scatter plot of Figure 1, showing data from TM Bands 3 and 4 of the training data plotted against each other, the clusters in the four land cover classes (i.e., *conifer*, *hardwood*, *brush*, and *barren*) clearly have different shapes and sizes. Secondly, by way of similar reasoning, covariance matrices traditionally are estimated separately for each class in fitting conventional maximum likelihood classifiers in remote sensing.

### 3 Data and Training

#### 3.1 Satellite Image and Field Measurement

The data consist of two components: satellite image data and field data. These had been collected and processed ready for analysis previously, and were made available by Carpenter et al [6] as a byproduct of their mixture study. The study area, Plumas National Forest of California, is characterized by temperate *conifer* forests mixed with chaparral *brush* field and *hardwood* forests. For the purpose of forest management, the quantification of conifers and *hardwood* within stands is useful [6].

The satellite data are from a Landsat TM image taken on June 20, 1990. The image was registered to a map projection and resampled using a nearest neighbor algorithm. Landsat TM Bands 1-5 and 7 are used for analysis (Band 6 is thermal band). The DN range for these bands is  $0 \sim 255$ . Field data was collected in August 1995 at 388 widely distributed field sites, which were delineated as polygons on the aerial photographs and satellite image. A site could be a pure *conifer* or *hardwood* stand, a *brush* patch, simply *barren*, or a mixture of all of them. The change in land cover between 1990 and 1995, almost negligible, was accounted for [6]. Sites range in size from 11 pixels to

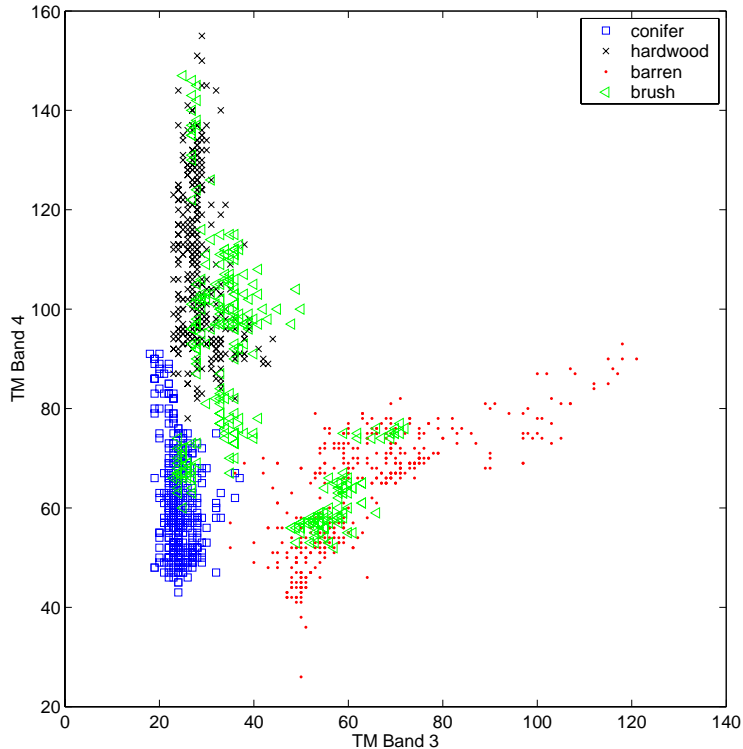


Figure 1: Scatter plot of Bands 3 and 4 of the training data. There are four main classes. Clusters in different classes have different size and shape. This warrants the class-dependent estimation of subclass covariance matrices.

224 pixels, with an average of 52 pixels per site. Fractional covers were estimated by means of visual interpretation of the aerial photographs and field traversing. The estimation was the consensus of a group of experts who conducted field data validation and the error bound was estimated to be 10%. For each site, fractional covers were estimated for *conifers* and *hardwoods*. For 263 of the 388 sites, the non-forest land cover is *barren* only. For the remaining 125 sites, the presence of *brush* as a non-forest land cover was recorded but its cover was not estimated; *brush/barren* fraction was estimated as a whole.

### 3.2 Training MDA

Training the MDA model is straight-forward. We used the same training data that Carpenter et al [6] used to train a MLC, consisting of 10 'pure' sites chosen randomly for the four land cover classes - *conifer*, *hardwood*, *barren* and *brush*. Training data consisted of 413 *conifer*, 399 *hardwood*, 303 *barren* and 291 *brush* pixels. The remaining 348 sites (19128 pixels) were used for testing.

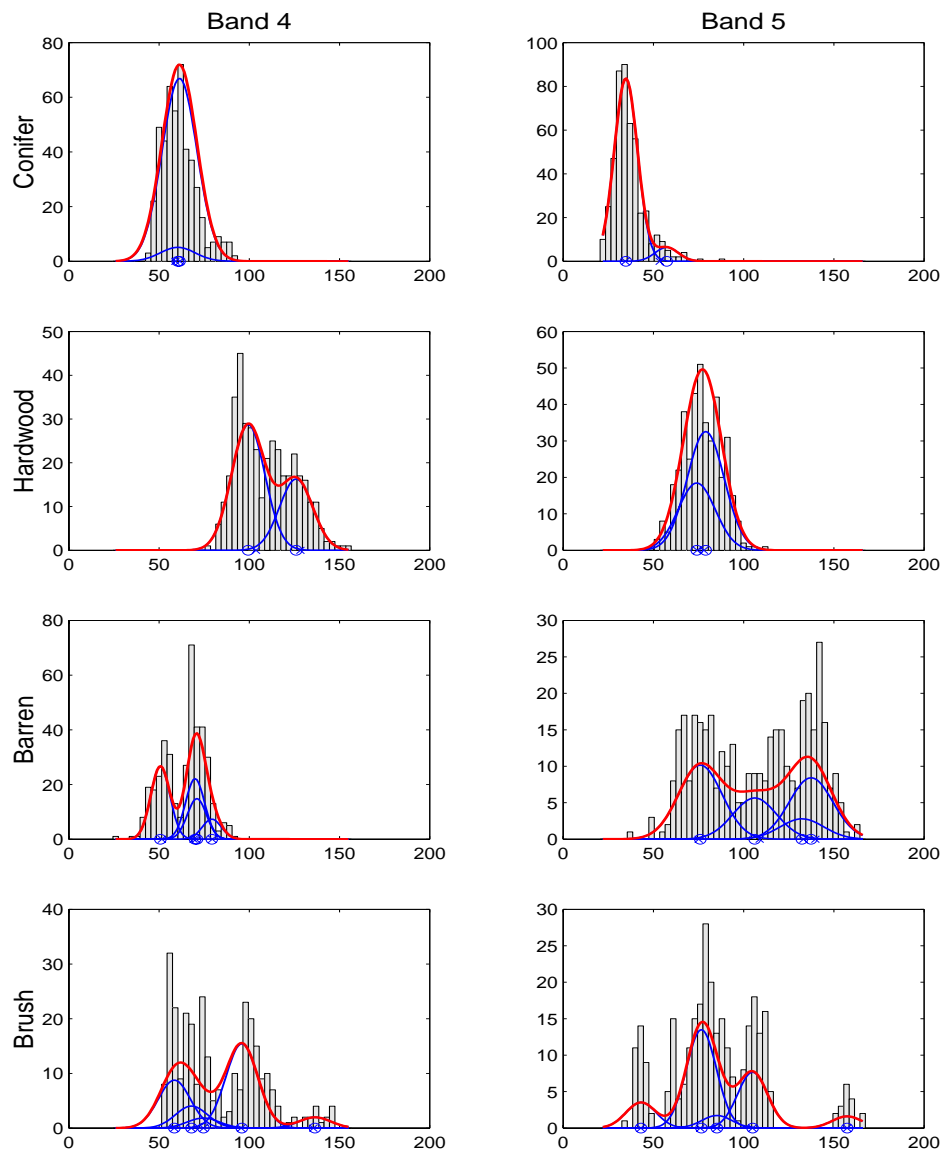


Figure 2: Histograms of TM Bands 4 and 5, for *conifer*, *hardwood*, *barren* and *brush* classes. Multi-modal distribution for *hardwood*, *barren* and *brush* classes are very clear. Fitted mixture components (blue lines) and mixture distributions (red lines) are plotted over the histograms and can be seen to capture this multi-modal character. The subclass means from *ustats*, serving as starting values for the EM algorithm, are marked with a cross. The final mean values produced by iterating EM to convergence are marked with a circle.

The number of subclasses was chosen based on examination of the histograms of the training data in all 6 bands. A clustering program called *ustats* in the image processing package IPW [13] was used to produce starting values for  $\mu_{jr}$ ,  $\pi_{jr}$  and  $\Sigma_j$  for executing EM. A histogram of Band 4 and Band 5 for the four land cover classes is shown in Figure 2. The clustering in the endmember spectrum is obvious. *Hardwood*, *barren* and *brush* classes are characterized by multi-modal distributions which may result from contamination by the presence of conifers of different species, age and other factors. Even the distribution of the relatively pure *conifer* class shows some evidence of multi-modality, in the form of a slight skewness. Clearly a single Gaussian distribution at the class level would not represent the data well. Hence, it is more realistic to treat a class as a mixture of subclasses. Note that the number of visible modes in the histogram of a given band does not necessarily seem to match the number of subclasses used in some of these plots, since the histogram only shows the marginal distribution in a single dimension (i.e., a spectral band) but the mixtures were chosen based on consideration of the multivariate data (i.e., all six spectral bands). One clear example is the *hardwood* class histogram which shows two very similar subclasses in Band 5 but two distinct subclasses in Band 4. The choice of the number of subclasses for MDA model was based on the combined information in histograms of all 6 bands. Then the clustering program *ustats* is used to calculate the mean ( $\mu_{jr}$ ), covariance matrix ( $\Sigma_j$ ) for 2 *conifer*, 2 *hardwood*, 4 *barren* and 5 *brush* subclasses. *ustats* also provides the size of the clusters whose proportions are used as the mixing probability ( $\pi_{jr}$ ) for subclasses.

The EM algorithm uses the  $\mu_{jr}$ ,  $\pi_{jr}$  and  $\Sigma_j$  output from *ustat* as starting values and re-estimates these quantities in an iterative fashion until convergence, as described in Equations (4) – (7). For example, in Figure 2 the initial  $\mu_{jr}$  estimates from *ustats* are marked with a cross and the final values resulting from the EM algorithm are marked with a circle. Many of these values are little changed although some, such as in Band 4 for *hardwood*, illustrate movement resulting from the EM algorithm towards a better fit.

Prior probabilities are important parameters for MDA but may be difficult to obtain in many contexts. In this research, we used the proportions of the four pure land cover classes in 388 sites as prior probabilities, 0.26, 0.17, 0.24, 0.33 for *conifer*, *hardwood*, *barren* and *brush* respectively. With all the parameters available, the posteriors for each testing pixel were computed with Equation (2). The posteriors represent the fractional covers in a pixel.

### 3.3 Performance Measures

The classifier performance measures used by Carpenter et al [6] are used here. Performance was evaluated in terms of root means squared error (RMSE) and percentage of predictions within a certain error bound. Since *barren/brush* cover is estimated as a unit for many sites, the predictions for *barren* and *brush* covers are merged as a single class called *brush+barren* for comparison over the test sites. The predictions for all pixels in a site are averaged to get the site-level fractional covers. For a land-cover class (*conifer*, *hardwood*, *barren+brush*), the

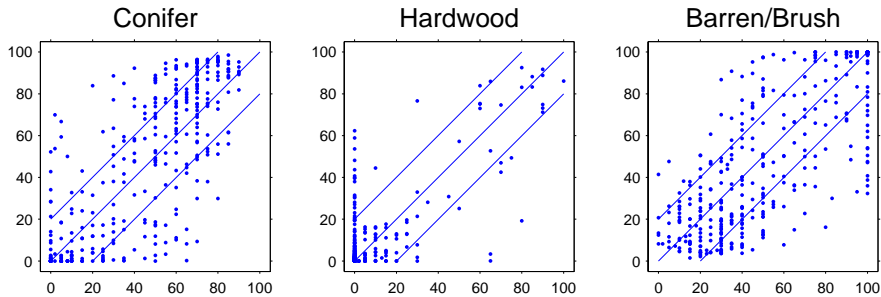


Figure 3: Fractional covers for the MDA classifier: measured( $x$  axis) against predicted( $y$  axis). The lines parallel to the diagonal are the 20% error bound. 73% of the predictions are within the 20% error bound.

RMSE is

$$\text{RMSE}_j = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{x}_{ji} - x_{ji})^2}$$

where  $\hat{x}_{ji}$  and  $x_{ji}$  are the predicted proportion and measured proportion for a land cover class  $j$  at site  $i$  respectively, and  $n = 344$  is the total number of testing sites.

Another measure is the percentage of the predictions for *conifer*, *hardwood* and *other* that falls within a certain absolute error. For example, at 10% level, the measure is

$$\frac{1}{3n} \sum_{j=1}^3 \sum_{i=1}^n n_{ji}$$

where  $n = 344$  and  $n_{ji}$  is the number of fractional cover predictions  $\hat{x}_{ji}$  such that

$$|\hat{x}_{ji} - x_{ji}| \leq 0.1$$

## 4 Results

Figure 3 shows the scatter plot of measured ( $x$ -axis) and MDA-predicted fractional cover ( $y$ -axis) of *conifer*, *hardwood*, and *brush+barren* classes for our test data set. Most of *conifer* and *hardwood* predictions fall well within the 20% error bounds. These results compare favorably to similar plots shown in Carpenter *et al* [6] using mixture models.

Table 1 compares the performance of our MDA classifier to that of other mixture models examined by Carpenter *et al* [6]. In their study, ARTMAP neural networks outperform linear mixture models consistently. The comparison here shows that the RMS errors for MDA are much smaller than the errors for linear mixture models for predictions of all land cover classes. For *conifer* sites, MDA results in a higher error than ARTMAP classification, but lower error for

	RMS Error			%Total	%Total
	C	H	Barren+Brush	Predictions within 10%	Predictions within 20%
ARTMAP classification	0.19	0.16	0.23	41%	78%
Exterior mixture	0.27	0.20	0.37	20%	56%
Interior mixture	0.28	0.17	0.34	34%	63%
ARTMAP mixture	0.18	0.13	0.20	50%	84%
MDA mixture	0.22	0.12	0.22	53%	73%

Table 1: Predictive accuracy of MDA, as compared to that of ARTMAP and other mixture methods studied by Carpenter *et al* [6].

*hardwood* and *brush/barren* sites. The RMS errors for MDA are slightly greater than those of ARTMAP mixture for *conifer* and *barren/brush* predictions, but smaller for *hardwood* predictions. Furthermore, the fraction of predictions by MDA that fall within 10% of the measurement is 53%, higher than than any other mixture methods including ARTMAP mixture models. For the 20% error bound, the fraction for MDA is 73%, higher than the linear mixture models, and closer to ARTMAP methods. All in all, the results of Table 1 show that the performance of MDA comes very close to that of ARTMAP mixture model.

## 5 Discussion

Linear mixture models and artificial neural network models can be viewed as two ends of a spectrum of mixture models in remote sensing. Linear mixture models are simple, but generally result in poor performance. Neural network model are highly nonlinear, and hence are better able to capture complex structure in the data. Their usage is gaining popularity, but their intrinsic “black box” character presents a major obstacle, due to the difficulty it lends to explaining how results were derived and extracting interpretations on the dynamics of the underlying physical process [11].

The MDA methodology is readily understandable in that it offers a relatively simple statistical framework that also captures the apparent variability in the data in a direct manner (e.g., see Figure 2). It also is easy to use, as the clustering software used here is part of a standard package and the EM steps involve iteration of closed-form expressions. The user need only specify the number of subclasses to the clustering program. In some sense MDA is even simpler to use than the linear mixture models, in that linear mixture models used by Carpenter *et al.* [6] require a proper identification of endmembers. The selection of endmembers is subjective and hence each choice of the endmembers produces different results. For example, in Carpenter *et al.* [6] there are 2 linear mixture models, exterior and interior linear mixture models, that resulted from the choice of 2 different endmember selections (see Table 1).

Modeling a land cover class in our MDA method as a mixture of subclasses, each modeled in turn as a Gaussian distribution, represents the data more real-

istically, in that it captures the multi-modal variation evident within the data. There have been other methods proposed that attempt similarly to accommodate the variation in endmembers, for example, such as the endmember bundle approach [3] and the multiple endmember approach [19]. However, endmember bundle, which uses certain principal components as the abstract endmembers, has not been tested on real data. And similarly, the multiple endmember models, which choose the best models from a pool of hundreds of candidate models, are very hard to implement unless very detailed information is available for the study area.

The current implementation of MDA requires that the training data represent pure land cover classes. In reality, “pure” classes are not always pure. For some “pure” *hardwood* sites, the canopy cover is only 80% *hardwood* with the remaining 20% a mixture of other classes. This reflects the inherent nature of land cover and the mapping process. Noisy and poor training data introduces error. For coarser resolution image of sparse vegetation, pure training data may be more difficult to obtain. This is one of a number of issues we plan to address in future research. Classification of land cover is one of the primary objectives in the use and analysis of data gathered by remote sensing. Increasingly, global climate models and terrestrial ecosystem models require specification of mixtures of land cover. The MDA technique introduced in this paper extends the application of mixture modeling in remote sensing by providing an easy and convenient tool to model mixtures in a non-linear fashion and provide interpretability in analysis.

## References

- [1] J. Adams, D. Sabol, V. Kapos, R.A. Filho, D.A. Roberts, M.O. Smith, and A.R. Gillespie. Classification of multispectral images based on fraction endmembers, application to land cover change in the Brazilian Amazon. *Remote Sensing of Environment*, 52:137–154, 1995.
- [2] P. Atkinson, M. Cutler, and H. Lewis. Mapping sub-pixel proportional land cover with AVHRR imagery. *International Journal of Remote Sensing*, 18:917–935, 1997.
- [3] C.A. Bateson, G.P. Asner, and C.A. Wessman. Endmember bundles: A new approach to incorporating endmember variability into spectral mixture analysis. *IEEE Trans. Geosci. Remote Sens.*, 38:1083–1094, 2000.
- [4] C. Borel and S. Gerstl. Nonlinear spectral mixing models for vegetative and soil surfaces. *Remote Sensing of Environment*, 47:403–416, 1994.
- [5] G. Carpenter. Distributed learning, recognition, and prediction by ART and ARTMAP neural networks. *Neural networks*, 10:1473–1494, 1997.

- [6] G.A. Carpenter, S. Gopal, S. Macomber, S. Martens, and C.E. Woodcock. A neural network method for mixture estimation for vegetation mapping. *Remote Sensing of Environment*, 70:135–152, 1999.
- [7] R.S. DeFries, J.R.G. Townshend, and M.C. Hansen. Continuous fields of vegetation characteristics at the global scale at 1-km resolution. *Journal of geophysical research*, 104:16911–16923, 1999.
- [8] G. Foody. Approaches for the production and evaluation of fuzzy land cover classifications from remotely-sensed data. *International Journal of Remote Sensing*, 17:1317–1340, 1996.
- [9] G. Foody. Sharpening fuzzy classification output to refine the representation of sub-pixel land cover distribution. *International Journal of Remote Sensing*, 19:2593–2599, 1998.
- [10] G. M. Foody, N.A. Campbell, N.M. Trod, and T.F. Wood. Derivation and applications of probabilistic measures of class membership from the maximum-likelihood classification. *Photogrammetric Engineering & Remote Sensing*, 58:1335–1341, 1992.
- [11] G.M. Foody. *Image classification with a neural network: from completely-crisp to fully-fuzzy situations*, in Atkinson, P.M and Tate N.J., Eds, *Advances in remote sensing and GIS analysis*, pages 17–37. John Wiley & Sons, 1999.
- [12] G.M Foody, R.M. Lucas, and P.J. Curran. Non-linear mixture modelling without endmembers using an artificial neural network. *International Journal of Remote Sensing*, 18:937–953, 1997.
- [13] J.E. Frew. *The image processing workbench*. Ph.D. dissertation, University of Santa Barbara, 1990.
- [14] S. Grossberg. Adaptive pattern classification and universal coding, II: Feedback, expectation, olfaction, and illusions. *Biol. Cybernet*, 23:187–202, 1976.
- [15] T. Hastie and R. Tibshirani. Discriminant analysis by Gaussian mixtures. *Journal of the Royal Statistical Society*, B 58:155–176, 1996.
- [16] D.K. McIver. *Adapting machine learning approaches for coarser resolution land cover classification*. Ph.D. dissertation, Boston University, 2001.
- [17] P. L. Odell and Basu J.P. Concerning several methods for estimating crop acreages using remote sensing data. *Commun. Statist.-Theory. Meth.*, A5:1091–1114, 1976.
- [18] D. Roberts, M. Smith, and J. Adams. Green vegetation, nonphotosynthetic vegetation, and soils in AVIRIS data. *Remote Sensing of Environment*, 44:255–269, 1993.

- [19] D.A. Roberts, M. Gardner, and R. Church. Mapping chaparral in the Santa Monica mountains using multiple endmember spectral mixture models. *Remote Sensing of Environment*, 65:267–279, 1998.
- [20] R.A. Schowengerdt. On the estimation of spatial-spectral mixing with classifier likelihood functions. *Pattern recognition letters*, 17:1379–1387, 1996.
- [21] M.O. Smith, S.L. Ustin, J.B. Adams, and A.R. Gillespie. Vegetation in deserts: I. a regional measure of abundance from multispectral images. *Remote Sensing of Environment*, 31:1–26, 1990.