ENDMEMBER SELECTION FOR MULTIPLE ENDMEMBER SPECTRAL MIXTURE ANALYSIS

Philip E. Dennison[1] and Dar A. Roberts[1]

1. Introduction

Spectral mixture analysis (SMA) models image spectra as the linear combination of endmembers (Adams et al., 1993). By utilizing an invariable set of endmembers, SMA does not account for the absence of one of the endmembers or spectral variation within "pure" materials. Multiple endmember spectral mixture analysis (MESMA) addresses these issues by allowing endmembers to vary on a per pixel basis (Roberts et al., 1998). MESMA has been applied in a variety of environments for vegetation and snow mapping. Roberts et al. (1997; 1998; 2003) and Dennison et al. (2000) used MESMA to map vegetation species and land cover type in Southern California chaparral. Painter et al. (1998; 2003) mapped snow grain size in the Sierra Nevada of California using MESMA. MESMA has also been used to map vegetation in semi-arid environments (Okin et al., 2001).

2. Background

Since the number of possible materials in an image can be very large, and since MESMA permits multiple endmembers for each material, an appropriate spectral library can contain hundreds of spectra. A large number of potential endmembers decreases computation efficiency and increases the complexity of the model output, so a parsimonious spectral library is desirable. Several methods of endmember selection for MESMA have been proposed. Painter et al. (1998) and Okin et al. (2001) used a limited number of reference spectra or *a priori* knowledge to select endmembers for their analyses. Roberts et al. (1997) devised a hierarchical endmember selection rule that used specialist endmembers to unmix a scene and then used generalist endmembers to model the remaining unmodeled or poorly modeled spectra. Roberts et al. (1998) selected endmembers to maximize the area mapped and minimize the overlap between models using a solution to the maximal covering problem (Church and Revelle, 1974). This paper presents the application of a new technique for selecting endmembers for MESMA using the endmembers that best model the spectra within their own class. The endmember with the minimum average root mean square error (RMSE) within a class is selected as the most representative endmember for the class.

Each spectrum in a spectral library can be modeled by any other spectrum within the library and shade using a two endmember model. Each of these models has a goodness of fit as measured by the RMSE. Endmember average RMSE (EAR) is the average RMSE for an endmember modeling the library spectra within its own material class. EAR is calculated as:

$$EAR_{A_i,B} = \frac{\sum_{j=1}^{n} RMSE_{A_i,B_j}}{n} \quad (1)$$

where A is the endmember class, A_i is the endmember, B is the modeled spectra class, and n is the number of modeled spectra in class B. For example, a "soil" class within a spectral library could contain 8 spectra. EAR can be calculated for the spectrum "soil3" as the average RMSE of "soil3" and a shade endmember modeling all the spectra within the soil class. EAR measures the actual model performance of an endmember for modeling spectra within its class. The spectrum with the lowest EAR best models the class, and is thus most representative of the class.

3. Methods

High altitude AVIRIS data were acquired on 5 dates over the Santa Barbara front range, including the city of Santa Barbara, California and the south-facing slope of the Santa Ynez Mountains. Six land cover classes were identified as dominant at a scale of 20 meters within the study area, including 5 vegetation classes (*Adenostoma fasciculatum*, *Arctostaphylos spp.*, *Ceanothus megacarpus*, *Quercus agrifolia*, and mixed introduced grasses) and an urban class. 65 reference polygons for the vegetation classes were identified using field inspection and hard copy 1 meter resolution United States Geological Survey digital orthophotos in June, 2002. Ten urban reference polygons were identified from the digital orthophotos in January, 2003. Polygons were required to be at least 50% dominated by one of the six land cover classes and be at least 40 meters by 40 meters in size, so that at least one pixel was contained entirely within the polygon.

AVIRIS data were acquired between 1998 and 2001, in the months of May, June and September (Table 1). All dates were processed to apparent surface reflectance using a modified version of the MODTRAN radiative transfer model (Green et al., 1993) and calibrated using the field-measured reflectance of a sand target. The data were

registered to an orthorectified SPOT mosaic resampled to 20 meters. Since the date is not a reliable indicator of vegetation water stress due to the variable nature of precipitation in Southern California, a simple soil water balance model was used to rank the relative moisture status of the 5 AVIRIS dates. Precipitation measured at the El Estero Water Treatment Plant in Santa Barbara was compared to reference evapotranspiration (ET_0) measured at a California Irrigation Management Information System (CIMIS) station approximately 4 kilometers to the northwest of the treatment plant. Soil water balance was set to zero for the full dry season preceding each date. Soil water balance was determined by cumulatively summing the daily ET_0 subtracted from the daily precipitation. Runoff and soil infiltration were not included in the model. The beginning of the dry season was determined to be the date on which the water balance reached zero after the last significant precipitation (> 3mm). Positive and negative soil water balance are referenced from this date (Table 1).

Image spectra from pixels entirely inside the reference polygons were extracted from the 5 registered AVIRIS reflectance images. 988 spectra from 59 polygons at least 75% dominated by a single land cover class were included in a separate spectral library for each date (Table 2). For each date, the library of 988 image spectra was unmixed by each of its component spectra and photogrammetric shade using MESMA. The non-shade endmember fraction was constrained to less than 106%, based on optimal constraints from Halligan (2002). For best-fit models with non-shade endmember fractions in excess of 106%, RMSE was calculated using the maximum non-shade endmember fraction of 106%. Permitting higher non-shade endmember fractions allows dark endmembers to have low EAR values that are not representative of their ability to model the spectra within their class. EAR was calculated for each endmember by averaging the RMSE for modeled spectra within the same land cover class. Endmembers with the minimum EAR within their class for each date were selected for mapping the AVIRIS images. Each AVIRIS image was modeled using 6 two endmember models corresponding to the 6 minimum EAR endmembers for each date. Non-shade endmember fractions were constrained to between -6% and 106%. Residuals were not allowed to exceed 2.5% reflectance for more than 7 contiguous bands and RMSE was constrained to below 2.5% reflectance (Roberts et al., 1998).

4. Results

The selected minimum EAR endmembers displayed significant spectral changes through the AVIRIS time series (Figure 1). Solar zenith was smaller for the 1998 and 2001 images (Table 1), and brightness effects due to lighting geometry were evident in the selected spectra of all of the land cover classes. All of the vegetation endmembers selected from the positive water balance images possessed a distinct red edge and chlorophyll absorption. Grassland endmembers exhibited the greatest changes in spectral shape due to the complete senescence of the grasslands. The red edge and shortwave infrared absorption features of the grassland endmember selected from the 1998 image were greatly reduced in the 2001 and 2002 endmembers, and were largely absent from the 1999 and 2000 endmembers (Figure 1d). A red edge was apparent in all of the selected urban spectra (Figure 1e), indicating subpixel scale vegetation was present in the urban environment. The presence of the red edge in both positive and negative soil water balance images indicates this vegetation was irrigated. The most interesting spectral changes occurred in *A. fasciculatum* and *Arctostaphylos* (Figure 1a,b). As soil water balance decreased, the presence of non-photosynthetic vegetation became more pronounced in the spectra of these land cover classes. Positive water balance spectra in these two land cover classes showed pronounced chlorophyll absorption and little ligno-cellulose absorption, while negative water balance spectra showed increased ligno-cellulose absorption and decreased chlorophyll absorption. This trend was less distinct in the selected *C. megacarpus* and *Q. agrifolia* spectra (Figure 1c,e).

Large areas corresponding to mixed residential and riparian areas were unmodeled by the 2 endmember models in all 5 AVIRIS images (Figure 2). Neither class was spectrally similar to the selected endmembers from the 6 land cover classes, and the high heterogeneity of residential neighborhoods made them difficult to model with only two endmembers. Both images with positive soil water balance (Figure 2a,b) were well mapped. *C. megacarpus* dominates the south-facing slope of the Santa Ynez Mountains, with bands of *A. fasciculatum* on rockier soils and *Q. agrifolia* on more mesic slopes and valley bottoms. *Arctostaphylos spp.* was properly limited to higher altitude rocky soils. Grassland was poorly modeled in the 1998 image, most likely due to varying degrees of grassland senescence. Three of the vegetation classes in the images with negative soil water balance were poorly modeled: *A. fasciculatum, Arctostaphylos spp.,* and *C. megacarpus* (Figure 2c,d,e). *Arctostaphylos spp.* and *A. fasciculatum* were overmodeled in all three images, while *C. megacarpus* was undermodeled. *Q. agrifolia* was overmodeled in the 2002 image, but was adequately modeled in the 1999 and 2000 images. Urban and grassland classes were well modeled in all the images with negative soil water balance.

5. Discussion

The accuracy of the modeled images was assessed using the entire set of 75 reference polygons. Land cover class accuracy was assessed by grouping all of the modeled image spectra within a reference polygon and selecting the most frequently modeled land cover class as the dominant class for the polygon. Unmodeled image spectra and polygons with equally dominant land cover classes were excluded from the accuracy assessment. Overall accuracy, kappa coefficient, and kappa variance were calculated for each date (Table 3) (Cohen, 1960; Congalton, 1991). The 2001 modeled image had the highest accuracy (0.90) and kappa coefficient (0.87) of the five AVIRIS dates. The 1998 modeled image also possessed overall accuracy and kappa coefficients over 0.80. Kappa and kappa variance were used to calculate a Z-statistic for each pair of dates to determine whether the kappa coefficients for each date were significantly different (Congalton, 1991). The kappa values of the two positive water balance images were found to be significantly better than the kappa values for two of the three negative water balance images, at the 95% confidence level. All three negative soil water balance images suffered from low accuracy. This is directly the result of *C. megacarpus* and *A. fasciculatum* polygons being modeled by *A. fasciculatum* and *Arctostaphylos* endmembers, as is apparent in Figure 2. The confusion matrix for the 2000 image highlights the confusion between these three vegetation classes (Table 4).

As the soil water balance decreases, the amount of senesced and dead material in a stand of vegetation increases. Even if the dominant species is not prone to senescence or dieback, subdominant components (grasses, *Artemisia californica, Salvia spp.*) of the stand may be. The vegetation classes modeled in the AVIRIS images become less distinct under drought conditions due to varying amounts of non-photosynthetic vegetation (NPV). The selected *C. megacarpus* endmembers model fewer *C. megacarpus* polygons in the negative water balance images than in the positive water balance images. *C. megacarpus* polygons with a higher fraction of NPV are modeled by the *A. fasciculatum* and *Arctostaphylos* endmembers, which displayed spectral features characteristic of increased NPV. Similarly, greener *A. fasciculatum* spectra were better modeled by *C. megacarpus* endmembers than by *A. fasciculatum* endmembers.

6. Conclusions

Endmember average RMSE was used to select the most representative image endmembers of six land cover classes from five AVIRIS images with varying soil moisture availability. Confusion between endmembers increased as soil water balance changed from positive to negative, reducing the accuracy of the modeled negative water balance images. Considering that many areas of all five AVIRIS images were unmodeled, a single two endmember model for each land cover class is not adequate for comprehensive mapping. Using multiple 2 endmember models for each class or adding 3 endmember models will reduce the number of unmodeled spectra. The use of additional models may also diminish confusion between *A. fasciculatum, Arctostaphylos,* and *C. megacarpus*. Endmembers with varying amounts of NPV could be selected for each species, allowing more accurate mapping in the negative water balance images. Even with an expanded set of endmembers, it is likely that wet season images with positive water balances will still be modeled with higher accuracies. This has implications for mapping vegetation using broadband sensor data, which will also be sensitive to seasonal spectral variability.

7. References

Adams, J.B., Smith, M.O., and Gillespie, A.R., 1993. Imaging Spectroscopy: Interpretation Based on Spectral Mixture Analysis. In: C.M. Pieters, and P.A.J. Englert (Eds.), *Remote geochemical analysis : elemental and mineralogical composition*. Press Syndicate of University of Cambridge, Cambridge, England, pp. 145-166.

Church, R. and Revelle, C., 1974. The maximal covering location problem. *Papers of the Regional Science Association,* 32:101-118.

Cohen, J., 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20:37-46.

Congalton, R., 1991. A review of assessment the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37:35-46.

Dennison, P., Roberts, D., and Regelbrugge, J., 2000. Characterizing chaparral fuels using combined hyperspectral and synthetic aperture radar. *Proceedings of the Ninth JPL Airborne Earth Science Workshop*. Jet Propulsion Laboratory, Pasadena, California, pp. 119-124.

Green, R., Conel, J., and Roberts, D., 1993. Estimation of aerosol optical depth and additional atmospheric parameters for the calculation of apparent surface reflectance from radiance as measured by the Airborne Visible-Infrared Imaging Spectrometer (AVIRIS). *Summaries of the Fourth Annual JPL Airborne Geosciences Workshop*, Jet Propulsion Laboratory, Pasadena, California, pp. 73-76.

Halligan, K.Q., 2002. Multiple endmember spectral mixture analysis of vegetation in the northeast corner of Yellowstone national park. Master's Thesis, University of California Santa Barbara.

Okin, G.S., Roberts, D.A., Murray, B., and Okin, W.J., 2001. Practical limits on hyperspectral vegetation discrimination in arid and semiarid environments. *Remote Sensing of Environment,* 77:212-225.

Painter, T.H., Roberts, D.A., Green, R.O., and Dozier, J., 1998. The effect of grain size on spectral mixture analysis of snow-covered area from AVIRIS data. *Remote Sensing of Environment,* 65:320-332.

Painter, T. H., J. Dozier, D. A. Roberts, R. E. Davis, and R. O. Green, 2003. Retrieval of subpixel snow-covered area and grain size from imaging spectrometer data. In press, *Remote Sensing of Environment*.

Roberts, D.A., Dennison P.E., Gardner, M., Hetzel, Y.L., Ustin, S.L., and Lee, C., 2003. Evaluation of the potential of Hyperion for fire danger assessment by comparison to the Airborne Visible Infrared Imaging Spectrometer. In press, *IEEE Transactions on Geoscience and Remote Sensing*.

Roberts, D.A., Gardner, M., Church, R., Ustin, S.L., and Green, R.O., 1997. Optimum strategies for mapping vegetation using multiple endmember spectral mixture models. *Imaging Spectrometry III*, SPIE Optical Engineering Press, Bellingham, Washington, pp. 108-119.

Roberts, D.A., Gardner, M., Church, R., Ustin, S., Scheer, G., and Green, R.O., 1998. Mapping chaparral in the Santa Monica Mountains using multiple endmember spectral mixture models. *Remote Sensing of Environment,* 65:267-279.

Table 1. Dates of AVIRIS data, ordered by soil water balance.

AVIRIS Dates	Solar Zenith	Soil Water Bal. (cm)
May 30, 1998	12.5°	+66.2
June 14, 2001	11.2°	+12.5
May 5, 2002	19.7°	-18.4
Sept. 16, 2000	37.0°	-37.9
Sept. 11, 1999	32.9°	-64.0

Table 2. Spectral library constituents from each land cover class.

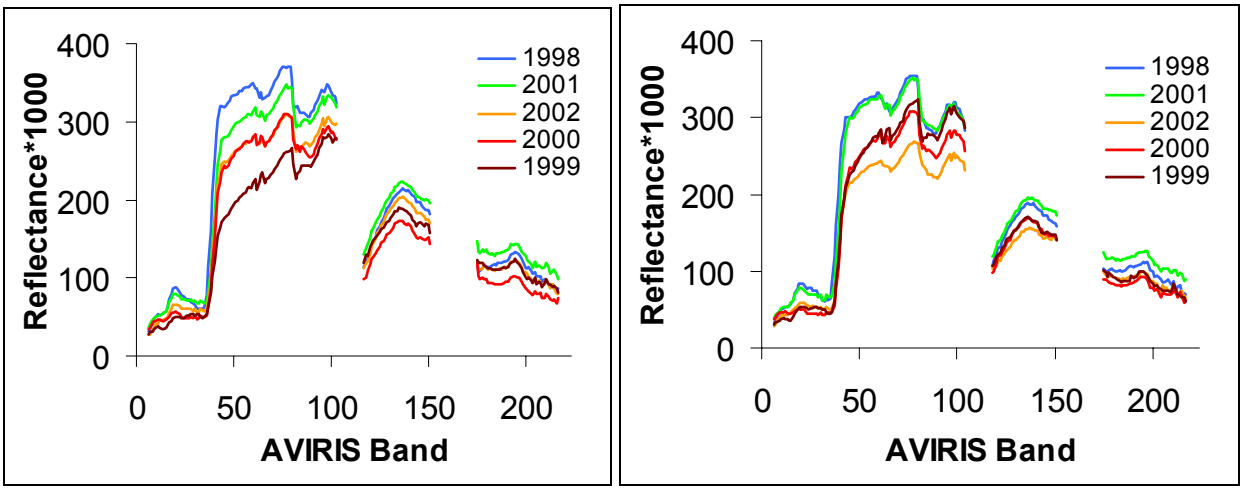
Class	# of spectra
A. fasciculatum	76
Arctostaphylos spp.	111
C. megacarpus	398
grassland	117
Q. agrifolia	107
urban	179
total	988

Table 3. Accuracy, kappa, and kappa variance for each AVIRIS date.

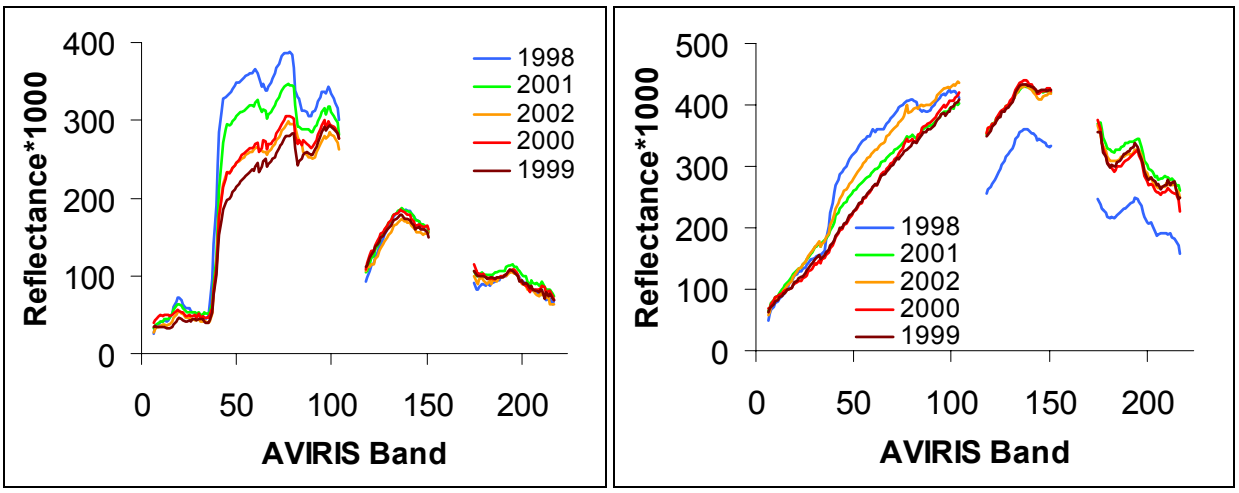
Year	Accuracy	Kappa	Kappa Var.
1998	0.85	0.81	0.0034
2001	0.90	0.87	0.0021
2002	0.63	0.56	0.0054
2000	0.68	0.62	0.0046
1999	0.79	0.74	0.0037

Table 4. Polygon dominant land cover class confusion matrix for the modeled 2000 AVIRIS image, including user's and producer's accuracies.

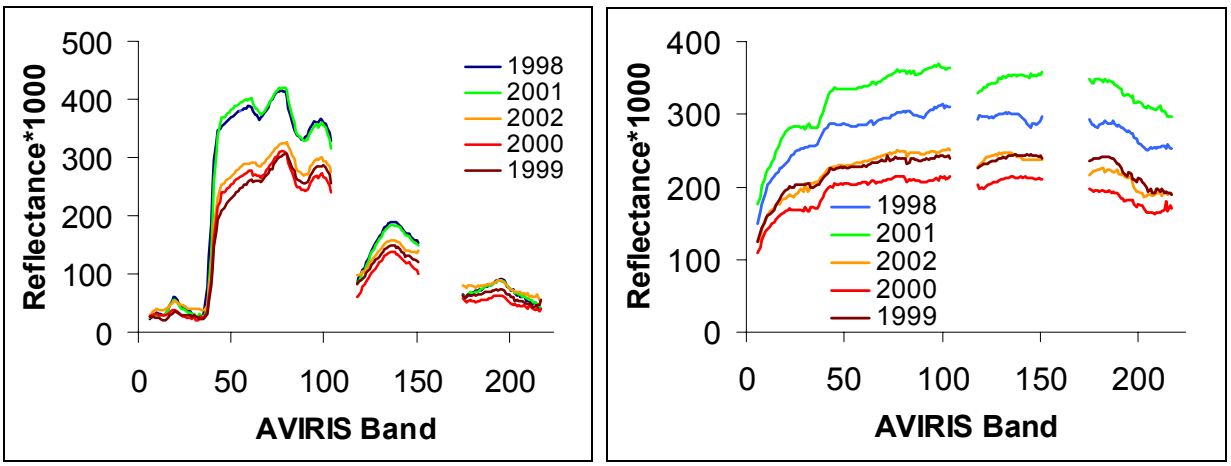
		Reference Dominant						
		A. fasc.	*Arcto.*	*C. mega*	grass	*Q. agri.*	urban	user's
Image Dominant	*A. fasc.*	5	1	5	0	0	0	0.45
	Arcto.	4	6	9	0	0	0	0.32
	C. mega.	2	0	9	0	0	0	0.82
	grass	0	0	0	10	0	0	1.00
	Q. agri.	0	0	1	0	8	0	0.89
	urban	0	0	0	0	0	9	1.00
	producer's	0.45	0.86	0.38	1.00	1.00	1.00	



a) *Adenostoma fasciculatum* b) *Arctostaphylos spp.*



c) *Ceanothus megacarpus* d) grassland



e) *Quercus agrifolia* f) urban

Figure 1. Selected minimum EAR endmembers for each land cover class, by year.

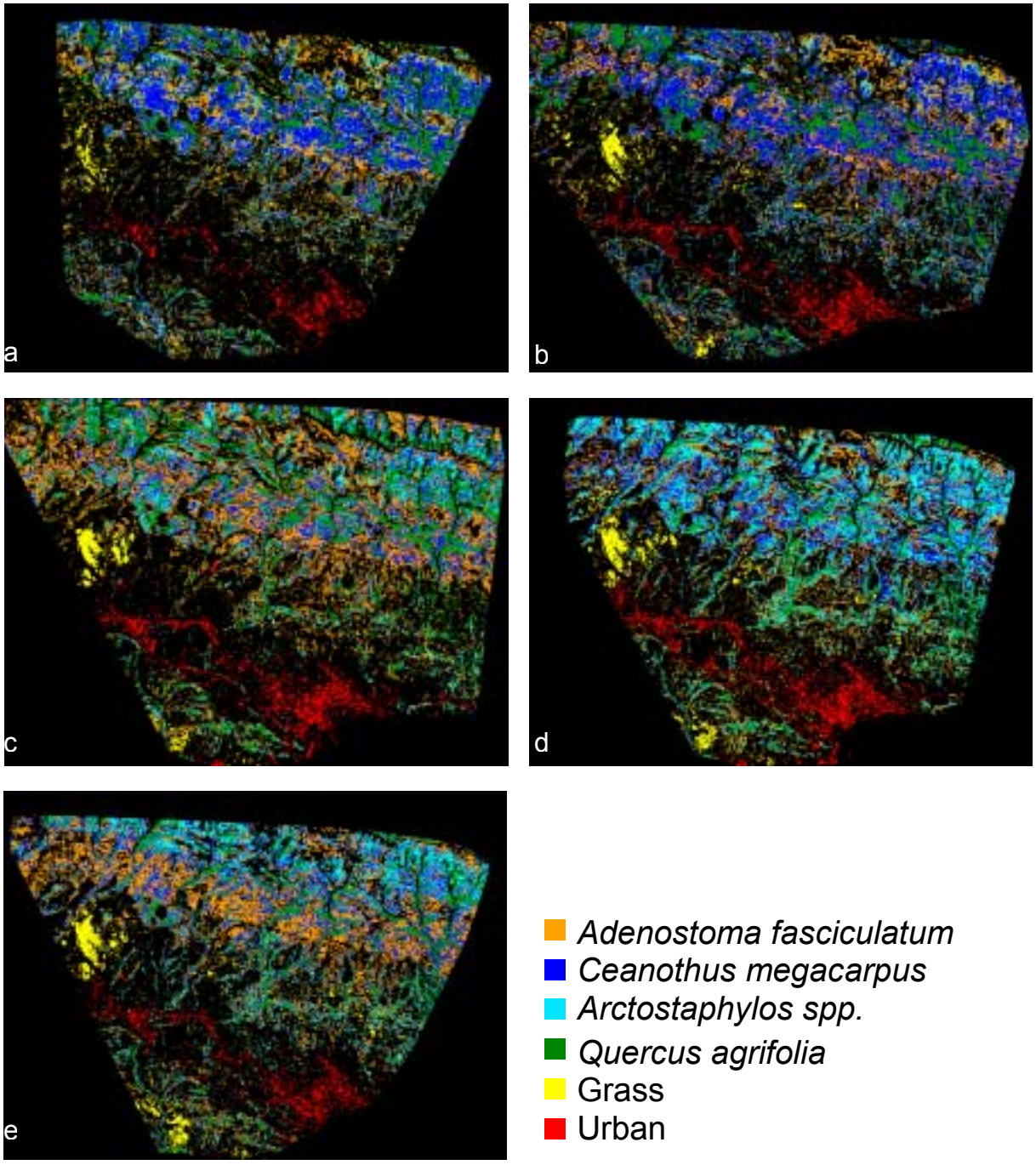


Figure 2. AVIRIS images of the Santa Barbara front range modeled using the minimum EAR endmembers for each date. Image letters correspond to the following dates: a) May 30, 1998; b) June 14, 2001; c) May 5, 2002; d) Sept. 16, 2000; and e) Sept. 11, 1999. Black areas are unmodeled.