ELSEVIER

Hyperspectral Mixture Modeling for Quantifying Sparse Vegetation Cover in Arid Environments

Kenneth McGwire, Timothy Minor,* and Lynn Fenstermaker**

A linear mixture model based on calibrated, atmospherically corrected Probe-1 hyperspectral imagery was compared with three vegetation indices to test its relative ability to measure small differences in percent green vegetative cover for areas of sparse vegetation in arid environments. The goal of this research was to compare multispectral and hyperspectral remote sensing approaches for detecting human disturbance of arid environments. The normalized difference vegetation index (NDVI) was tested using both narrow and broad bandwidths. Broadband NDVI provided results (r^2*=0.63) similar to NDVI derived from individual hyperspectral channels* (r^2*=0.60). While the soil-adjusted vegetation index (SAVI) was designed as an improvement to NDVI for sparse vegetation, in this study SAVI performed significantly worse than NDVI* (r^2*=0.51). The modified soil-adjusted vegetation index (MSAVI) provided an insignificant improvement over NDVI* (r^2*=0.64). Linear mixture modeling provided significantly better results,* r^2 *of 0.74. Cross-validation was used to test the significance of differences between the various methods and to determine the standard error associated with each method. Results suggest that any improvements provided by adjusted vegetation indices over NDVI may be strongly dependent on those adjustments being derived from local conditions. The use of a linear mixture model with multiple soil endmembers appears to provide the best method for quantifying sparse vegetative cover. Though present in small amounts, a single plant species,* Krameria erecta*, was strongly correlated with residuals of the mixture model. Inclusion of a spectral endmember for this species increased the* r^2 *of the fit with percent green cover to 0.86. However, it is not clear if the explained variation was actually due to* K. erecta *or a correlated phenomena. Problems were also identified with the use of multiple vegetation endmembers.* ©*Elsevier Science Inc., 2000*

INTRODUCTION

Arid and semiarid environments comprise approximately 35% of the earth's land surface. However, quantifying the health and abundance of vegetation in these environments with vegetation indices derived from multispectral imagery has been problematic because vegetation cover is sparse, surface spectral reflectance may be dominated by soil backgrounds, and at-sensor radiance may be strongly affected by atmospheric scattering (Huete and Tucker, 1991). This study tests the potential for hyperspectral remote sensing to improve the quantification of small amounts of vegetation cover in arid environments.

The difference between infrared and red reflectance has been used to study vegetation and phenology since the early 1970s (e.g., Colwell, 1973; Colwell, 1974; Carneggie et al., 1974; Rouse et al., 1974). Such studies often use the normalized difference vegetation index (NDVI) because it minimizes the effects of topography (Holben and Justice, 1981), requires no prior knowledge of ground conditions, is sensitive to the amount of photosynthetically active vegetation (Myneni et al., 1992; Tucker, 1979), and is conveniently scaled between −1.0 and 1.0. Sellers (1985, 1987) has demonstrated a meaningful physiological relationship between NDVI and photosynthetic activity. However, NDVI transformations of multispectral image data have proven problematic in regions of sparse vegetation cover due to strong soil background effects (Huete and Tucker, 1991). To address these problems, improved variants of the NDVI transformation, such as the soil-adjusted vegetation index (SAVI) and modified SAVI (MSAVI), have been developed that

* Desert Research Institute, Biological Sciences Center

Address correspondence to Kenneth McGwire, Desert Research Institute, Biological Sciences Center, 2215 Raggio Parkway, Reno, NV 89512 USA. E-mail: kenm@dri.edu

Received 25 March 1999; revised 22 November 1999.

REMOTE SENS. ENVIRON. 72:360–374 (2000)

655 Avenue of the Americas, New York, NY 10010

0034-4257/00/$–see front matter
PII S0034-4257(99)00112-1

incorporate a soil adjustment factor (Qi et al., 1994; Huete et al., 1992; Huete, 1988).

More recently, Elvidge et al. (1993) have shown that the improved spectral resolution of hyperspectral remote sensing data in the region of the "red edge" may improve the detection of changes in sparse amounts of vegetative cover. The "red edge" refers to the position of the edge of a strong chlorophyll absorption feature in the red wavelengths. Hurcom and Harrison (1998) describe the use of hyperspectral field measurements for estimating the abundance of semiarid plant species in Spain and Portugal, though the density of the plant canopies they tested was much higher (>30% cover) than those studied here (<11% cover). Hurcom and Harrison (1998) found that one of the components of a factor analysis corresponded to the "red edge" and thus had a strong correlation with NDVI.

Linear mixture modeling has been promoted as a method that can make effective use of the information content of multispectral and hyperspectral imagery (Roberts et al., 1993; Smith et al., 1990; Gillespie et al., 1990; Mustard and Pieters, 1989). This approach decomposes each image pixel into a linear combination of reference spectra, referred to as "endmembers," that characterize the various spectrally unique components of the landscape. The reference endmembers might be developed from laboratory or field spectra (Tompkins et al., 1997), or they might be derived directly from image data (Adams et al., 1993). Smith et al. (1990) used the linear mixture model technique with Landsat TM imagery to derive vegetation abundance estimates for a large region of relatively sparse vegetation in the Owens Valley, California. Fractions of vegetation, soils, and shadow were calculated for TM pixels by applying a mixture model based on laboratory and field reference spectra. The vegetation endmember derived from the image data was compared to vegetation cover estimates collected from ground transects. Results indicated that this method provided estimates of vegetation abundance that were consistent with the projected vegetation cover. In the results section of this paper we will compare the results of Smith et al. (1990) to those developed in this study.

METHODS

Study Site Description

Six study sites were selected on the Providence Mountain bajadas within the Mojave National Preserve in southeastern California (Fig. 1). The six sites are located on mixed plutonic material of Quaternary age (McDonald and McFadden, 1994). Five of the six sites are composed of fluvial materials, and the northernmost site is composed of eolian sand. Surfaces range in age from the Holocene, ca. 4 ka, to Pleistocene, ca. 8 to 12 ka (McDonald, 1994), with older surfaces at higher elevations. Surface designations used to identify the geomorphic settings (e.g., Qf5, Qf3, and Qe3) are abbreviations for Quaternary fluvial (f) or eolian (e) surfaces, with larger numbers indicating younger surfaces. All of the surfaces have some degree of desert pavement with dark rock varnish except the Qe3 surface, which is composed of loose sand from old dunes. The Qf5 distal site has the most highly developed pavement with homogenous, darkly varnished coarse fragments approximately 1 cm to 2.5 cm in diameter. The Qf6 surface has large coarse fragments with a small amount of varnish. The Qf5 and Qf4 surfaces have a coarser pavement with boulders present on the surface. The Qf3 surface is devoid of boulders and has areas of well-developed pavement; however, there are small areas where rodent activity has disturbed the pavement (e.g., 30 cm to 50 cm long by 10 cm to 30 cm wide).

Two field plots were identified in each of the six geomorphic settings, using sites originally established by McAuliffe and McDonald (1995). The plot dimensions were 20×20 m for all of the plots except the two plots located on the Qf5 distal study site. The Qf5 distal site is a narrow lobe of material, so plot size was limited to 16×20 m. Two opposing corners of each field plot were marked with reflective material several meters away (to minimize pixel contamination) at the time of overflights to allow accurate identification of the plots in the hyperspectral imagery. The six geomorphic settings and two plots per setting provided a fairly complete assessment of the plutonic surfaces on the Providence Mountain bajadas, though there were some surfaces of mixed age, and limestone or mixed limestone/plutonic surfaces north of the study area. Other omitted geomorphic settings were: (1) the washes that were too narrow to establish a 3-by-3 pixel study area (i.e., 15 m by 15 m); and (2) Qf1, Qf2, and Qf7 surfaces that had similar vegetation characteristics to other surfaces of close age (e.g., Qf7 is very similar to Qf6, and Qf-1 and -2 are very similar to Qf3) (McAuliffe, personal communication).

A 1994 survey by J. R. McAuliffe recorded 37 perennial species; however, two shrubs, creosote bush (*Larrea tridentata*) and white bursage (*Ambrosia dumosa*) accounted for 75% to over 90% of the total perennial plant canopy volume. Table 1 lists all of the plant species that had green-leaf area during the early June data collection period for this study. McAuliffe and McDonald (1995) showed that the differences in *L. tridentata* and *A. dumosa* cover at these study sites varied with site-specific soil characteristics related to infiltration and water-holding capacity, and that vegetation cover decreased with increasing age of the surface.

Field Data Acquisition

Vegetation Cover

Detailed vegetation cover data were acquired for each of the field plots, starting the week before the aircraft acquisition and extending through the following week. Each plot was subdivided into equal-size quadrants (10

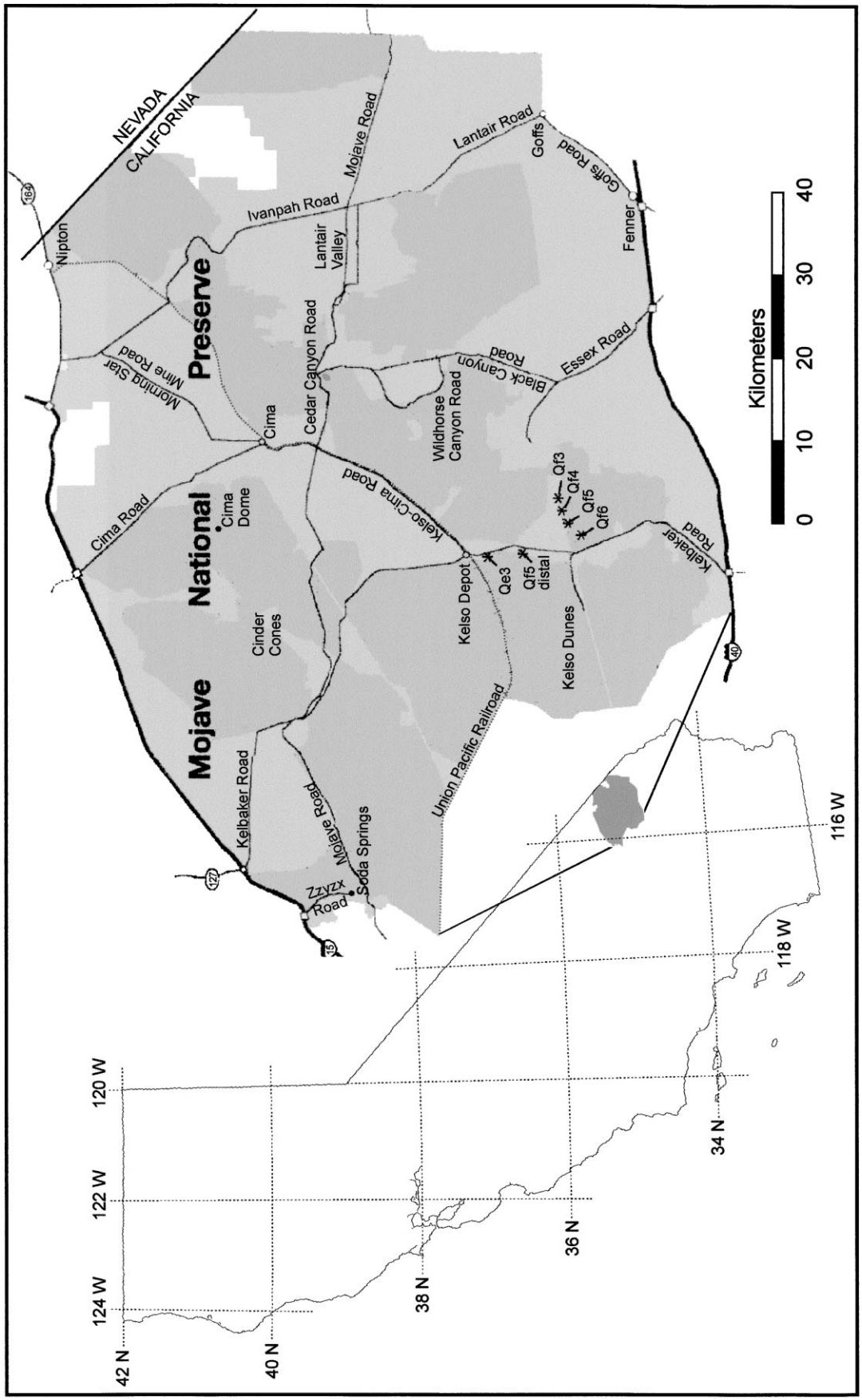


Figure 1. Map of study area (two field sites at each *).

Table 1. Species with green-leaf area at time of overpass

Creosotebush, *Larrea tridentata*	Mojave yucca, *Yucca schidigera*
White bursage, *Ambrosia dumosa*	Pencil cholla, *Opuntia ramosissima*
Range ratany, *Krameria erecta*	Beavertail cactus, *Opuntia basilaris*
Desert thorn, *Lycium andersonii*	Barrel cactus, *Ferocactus acanthodes*
Turpentine broom, *Thamnosma montana*	Galleta grass, *Pleuraphis rigida*
Desert senna, *Cassia armata*	Dune primrose, *Oenothera deltoides*
Sandpaper plant, *Petalonyx thurveri*	Locoweed, *Astragalus layneae*

m by 10 m for all plots, except the Qf5 distal plots that were 8 m by 10 m). Species, height, diameter in two dimensions, and estimated percent green cover were recorded for every green plant in each quadrant. Percent green cover refers to the projected area of green canopy on the ground (ignoring overlap) and was measured at 15% increments. Hurcom and Harrison (1998) showed that this type of field-of-view measurement corresponds better with spectral indices than leaf area index (LAI). Percent cover was selected because the final application was for simple change detection due to human disturbance, such as off-road vehicle use, rather than ecological analysis. Because of this, the relationships described here may not be as strong for more ecologically relevant parameters such as above ground biomass or LAI. Any herbaceous plants present at the time of image acquisition were totally senescent.

Visual estimates of percent green cover were made for each plant at 15% increments. To ensure reasonably accurate visual estimates, a 20-by-20 cm grid was used to objectively measure percent green cover for portions of several *L. tridentata* shrubs. These tests were performed at several plots on shrubs with high, medium, and low levels of percent green cover. Because the other predominant shrubs did not have growth forms that permitted the placement of a grid beneath their canopy, only *L. tridentata* estimates of green cover were evaluated in this manner. *L. tridentata* canopies with high green cover that were recorded as 45% green had grid estimates of green cover ranging from 46% to 54%. Visual estimates of medium green cover (30%) had grid measurements ranging from 25% to 31%. Canopies recorded as having low green cover (15%) had grid estimates ranging from 12% to 16%. Plant-level measures were converted to quadrant-level percent green cover estimates by summing the percent green cover times areal extent of each plant and dividing by quadrant area. The areal extent of plants was based on approximate circular area using the mean canopy diameter. Quadrat data for each plot were combined in the final analysis, based on an assessment of the image pixel size and geometric registration.

Field Spectrometry

Field spectra were acquired between 10:00 and 14:00 hours on the day of the Probe-1 data acquisition and the day after the aircraft data collection. It was not possible to acquire all field spectra during the same day due to the rapid development of cloud cover immediately after the aircraft collection. Field spectra were acquired with an Analytical Spectral Devices, Inc. (ASD) FieldSpec Full Range that acquires continuous spectra from 0.4 μm to 2.5 μm. Dark-current (closed aperture) and white reference (reference-grade spectralon panel) data were acquired for the FieldSpec approximately every 5 minutes. Each spectra used in this analysis was an average of fifty individual measurements that were automatically acquired by the FieldSpec. Spectra were acquired for the following features: canopy-level spectra for each major plant species at each plot; leaf-level spectra for *L. tridentata* and *A. dumosa*; the soil matrix at each plot; dead plant material (primarily grasses); and various shadows at each plot. Measurements were made using a fiber optic tip that was 2 mm in diameter with a 25 degree field of view. All leaf-level measurements were made with consistent sun/fiber tip/spectrometer-operator geometry and the tip held approximately 3 mm to 4 mm above the leaf, resulting in a footprint of 2 mm to 3 mm in diameter that fit entirely within the smallest leaf measured. Canopy-level measurements were nadir-view, acquired at a height of approximately 30 cm above the canopy, and had a 15 cm diameter footprint on the ground.

Field-measured canopy spectra were examined to identify cases for which data acquisition conditions were poor. Plots of individual spectral response curves were visually assessed for each species and strong outliers were removed from subsequent use. Fieldspec data were resampled to Probe-1 wavelengths using a gaussian model with a full width, half-maximum (FWHM) equal to the band spacings. FWHM is the width of an instrument response (band pass) at half of the band depth.

Aircraft Hyperspectral Data

Earth Search Sciences Inc. (ESSI) of McCall, Idaho acquired Probe-1 hyperspectral data over the 12 field plots on June 3 1998 between 10:45 and 11:15 A.M. Probe-1 is a whiskbroom instrument manufactured by Integrated Spectronics in Australia. Probe-1 has 128 spectral channels in the visible through short-wave infrared wavelengths (0.4 μm to 2.5 μm), with a signal-to-noise ratio around 1,500/1 for the visible and NIR wavelengths, and 800/1 for the SWIR (Vance et al., 1999; ESSI, 1999). The data were acquired from an average altitude of

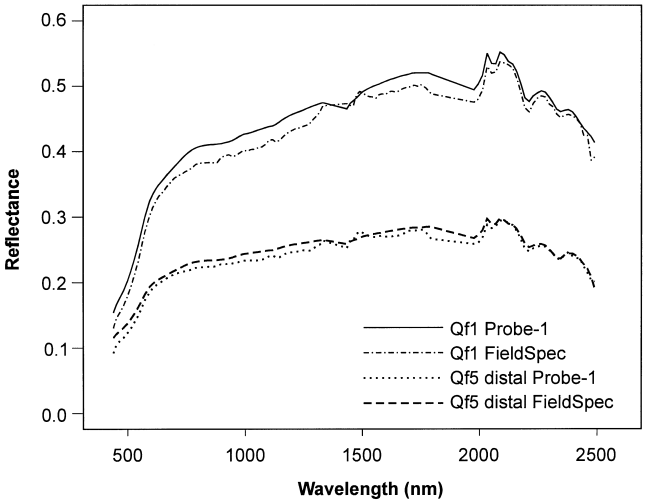


Figure 2. Correspondence of corrected aircraft data with averaged field spectra.

2,850 m ASL over the six study sites, resulting in a nominal ground resolution of 5 m.

A two-step atmospheric correction method was used that employed the ATREM radiative transfer-based atmospheric correction code (CSES, 1997) and the modified flat field (MFF) technique (Farrand and Harsanyi, 1997). The ATREM correction improved the correspondence between aircraft and field spectra. However, due to misregistration between the wavelength positions of the Probe-1 spectral bands and the positions calculated by ATREM (at the time ATREM had only been designed to work with AVIRIS and HYDICE data), the correction was not satisfactory, as NIR and SWIR band passes were poorly corrected. Therefore, a ground-based correction method was superimposed on the ATREM corrected data to mitigate these problems. In the MFF correction, the average ATREM-derived surface reflectance of a spatially and spectrally homogenous area on the ground was divided into each pixel of the image data set. In this case we used a transmission line service road for the correction. Each pixel of the resulting data set was then multiplied by the field-measured road spectra. This second stage of the MFF process returns the data to units of reflectance. This normalization effectively rescaled the aircraft-measured reflectance to field measurements. No correction was made for solar illumination, as it was assumed that a shadow endmember would compensate for such effects. The ATREM/MFF correction procedure was used on all three flight lines of Probe-1 data and resulted in reflectance images that appeared to correspond well with averaged field spectra. Comparisons of selected field and aircraft spectra are shown in Figure 2.

A reflectance image was created by concatenating subsets from the three flight lines that contained the six study sites. After correction, the reflectance values in the imagery were consistent with those observed in the field plots. Examination of the resultant 128-band data set revealed problems that required the removal of eight bands. Bands 31 and 32 overlapped at a crossover point between the first two Probe-1 detectors and were removed because they caused the spectral reflectance curve to double back on itself. Bands 63 through 67 of the Probe-1 image data were removed due to water vapor absorption features. In addition, band 128 was removed due to its very low signal output.

Baseline Vegetation Indices

NDVI and two of its variants, SAVI and MSAVI, were calculated from the Probe-1 reflectance data to provide a baseline for evaluating the utility of the hyperspectral linear mixture modeling approach. NDVI is a ratio of shortwave infrared and red reflectance in the form shown in Eq. (1):

$$\mathrm{NDVI}=\frac{\mathrm{NIR}-\mathrm{red}}{\mathrm{NIR}+\mathrm{red}} \tag{1}$$

that provides a convenient, rapid estimate of the amount and/or productivity of vegetation in an image pixel (Tucker, 1979). NDVI was calculated using both narrowband and broadband approaches. The narrowband method used individual channels of the Probe-1 data that were subjectively selected to provide the most consistent representation of red absorption and the NIR plateau for plant species in the study area (667 nm and 1086 nm, respectively). The broadband method used channels that were averaged to approximate spectral channels 3 and 4 of Landsat Thematic Mapper.

The second set of baseline calculations used the SAVI and MSAVI indices. SAVI, which is formulated as shown in Eq. (2)

$$\mathrm{SAVI}=\left(\frac{\mathrm{NIR}-\mathrm{red}}{\mathrm{NIR}+\mathrm{red}+\mathrm{L}}\right)\cdot(1+\mathrm{L}) \tag{2}$$

may provide a more accurate estimate of vegetation productivity or biomass than the NDVI by accounting for first-order soil-vegetation spectral interactions (Huete, 1988). The development of an empirical adjustment factor, L, minimized the backscatter effect of soil background reflectance being transmitted through the canopy. Huete (1988) reported $L=1.0$ to be the appropriate adjustment factor in reducing effects of background soil for sparse canopy covers, so that value was used here.

The modified SAVI (MSAVI) (Qi et al., 1994) is formulated as seen in Eq. (3):

$$\mathrm{MSAVI}=(1+\mathrm{L})\cdot\frac{\mathrm{NIR}-\mathrm{red}}{\mathrm{NIR}+\mathrm{red}+\mathrm{L}} \tag{3}$$

where:

$$L=1-2a\cdot\mathrm{NDVI}\cdot\mathrm{WDVI}$$

$$a=\text{slope of the soil line}$$

$$\mathrm{WDVI}=\mathrm{NIR}-(a\cdot\mathrm{red})$$

MSAVI attempts to account for soil background vari-

ation by utilizing a variable *L* function based on the weighted difference vegetation index (WDVI: Clevers and Verhoef, 1993) and the slope of the soil line. Note that this *L* function is different from the empirical *L* factor of SAVI. In a study that compared the sensitivity of NDVI, SAVI, MSAVI, and several other vegetation indices in arid environments with less than 25% vegetation cover, Rondeaux et al. (1996) determined that MSAVI may perform better than the other indices. For this study, SAVI and MSAVI were tested using the broadband approach in which Probe-1 bands were averaged to approximate Landsat TM channels 3 and 4.

Hyperspectral Data Analysis

The purpose of this effort was to test the effectiveness of linear mixture modeling with hyperspectral data relative to standard multispectral vegetation indices. Given that the goal was to perform this comparison in sparsely vegetated ecosystems, it was not possible to derive a pure vegetation endmember directly from the image data. The highest level of green vegetative cover measured at the field sites was 10.74% for a 20-by-20 m field plot. As a result, the green vegetation endmember for the linear mixture model was generated by averaging leaf-level spectra collected in the field from *L. tridentata* and *A. dumosa* plants. Figure 3 displays the field-sampled, canopy spectral reflectance curves for several of the 14 plant species that had green foliage at the time of the overflight, along with the average green vegetation endmember. While each species did tend to deviate from the average in a systematic way, there was substantial variability in field-measured canopy reflectance among plants of a given species. The coherence of the overall pattern in spectral curves at the canopy level, combined with a relatively high ratio of within- to between-class variance, suggested that it may be reasonable to use a "universal" green vegetation endmember. *L. tridentata* and *A. dumosa* were selected for the average vegetation endmember because they had the most dominant cover in the region, and the variability of field-measured reflectance for these two species encompassed much of the variability demonstrated by other plants in the region. Despite this, reflectance curves for two of the plant species, *Yucca schidigera* and *Oenothera deltoides*, differed noticeably from this pattern (Fig. 3). However, the contribution of these two species to total canopy cover was quite small.

Given the goal of detecting sparse vegetation in arid environments, it was believed that it would be possible to develop spectral endmembers for the various soil types in the area directly from the image data by identifying unvegetated pixels with relatively homogenous soil composition. This would have the advantages of reducing the dependence on field data acquisition and reducing problems with undersampling soil types in the region. To generate the soil endmembers, the reflectance image was used as input into the minimum noise fraction (MNF) noise-whitened transform function (Green et al., 1988), to yield 120 principle component eigen-images of the data. Eigenvalues were very high for the first three eigen-images, but quickly tailed off for the remaining bands. Visual inspection determined that after the fourteenth component the MNF transforms became very noisy. A pixel purity index (PPI) function, using a convex geometry technique, was applied to the first 14 eigen-images, and a high number of iterations (5,000) was used to rank relatively pure soil background pixels in the MNF eigen-images (Boardman, 1993).

To ensure that the PPI algorithm was only evaluating soil-dominated pixels, an NDVI image mask based on a threshold value of 0.09 was applied during the computation of the PPI to mask out those pixels exhibiting some vegetation response. In addition, a relatively high threshold value of 2.5 was specified for "pixel purity." The NDVI-masked, PPI threshold procedure substantially reduced the number of pixels within the image data to be considered for generating soil endmembers. Only 1,148 pixels remained, less than 1% of the entire image. This small set of "pure" pixels were evaluated using multidimensional visualization software, in which the pixels were interactively clustered, examined, and compared in the dimensional space of the 14 MNF bands. Selection of endmember spectra was based on the cohesion of data clusters as they were rotated about multiple MNF band axes. Eight distinct spectral soil endmembers were identified; three related to the eolian sand site, three related to the bajada site containing fluvial materials, and two endmembers related to the distal site containing desert pavement with dark rock varnish. Although Okin et al. (1998) suggest that three to four endmembers might typically be enough to characterize soils in arid environments, these eight soil endmembers did represent distinct extremes in the MNF space.

The final endmember that was used in the mixture model was a shadow endmember. The shadow endmember was generated using a simple flat response of 0.0 reflectance across all bands. This approach was used as a simplification to the ambiguity of shadow reflectance over various substrates that, in turn, may or may not be affected by vegetation depending on whether it resulted from canopy or topographic shading. It was expected that shadows cast by vegetation were not likely to be greatly depleted in visible wavelengths or enriched in the infrared given the open canopies and small or microphyllous leaves of the arid-adapted species.

The averaged green-leaf endmember, the eight image-derived soil endmembers, and the shadow endmember were then used in a partially constrained (unit-sum constraint) linear spectral unmixing algorithm (ENVI version 2.6: RSI, 1997) on the reflectance image. The unit-sum constraint was added to the system of simultaneous equations in the inversion process to ensure that the weights calculated for the various endmembers

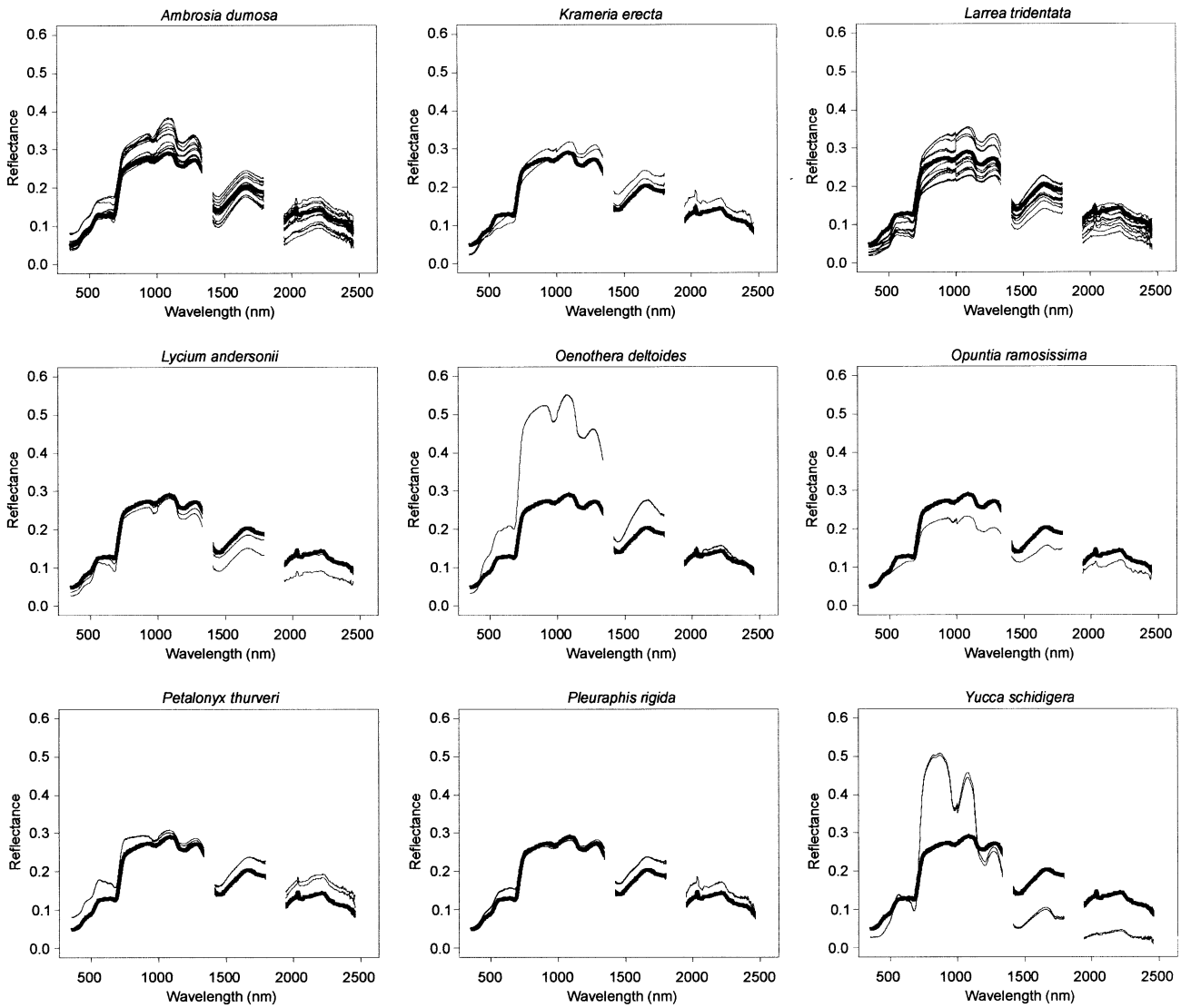


Figure 3. Field measured canopy spectra (light lines) and the average green-leaf endmember (heavy line).

summed to 1.0. This approach is commonly used to relate output from the mixing model to the fraction of cover in each pixel by different endmembers.

The relationship between the various methods and the field measurement of percent green cover was examined using the r^2 values of a linear regression. To test the significance of differences between methods, a cross-validation method was used in which each sample was sequentially omitted from the regression analysis and then predicted from the remaining samples. This cross-validation allowed a test of the significance of differences in r^2 between the various methods.

RESULTS

The results for narrowband and broadband NDVI using all samples are presented in Figures 4 and 5. Table 2 displays the mean and standard deviation for the regression parameters as determined from the leave-one-out cross-validation method. The broadband approach performed slightly better than the narrowband approach, with r^2 of 0.63 and 0.60, respectively. This difference was probably due to a reduction in noise associated with the averaging of the Probe-1 bands and a possible reduction in species-specific reflectance characteristics that may be expressed at finer spectral resolutions. However, the results of a t-test on the iterations of r^2 in the cross-validation (Table 3) show that this difference has a large p value and cannot be considered significant with this number of samples.

Results using all samples for the adjusted vegetation indices are presented in Figures 6 and 7. SAVI displays a lower r^2 (0.51) than either NDVI method, while MSAVI has a better r^2 than the NDVI methods (0.64; Table 2). As seen in Table 3, the poor performance of SAVI was quite significant relative to all other methods,

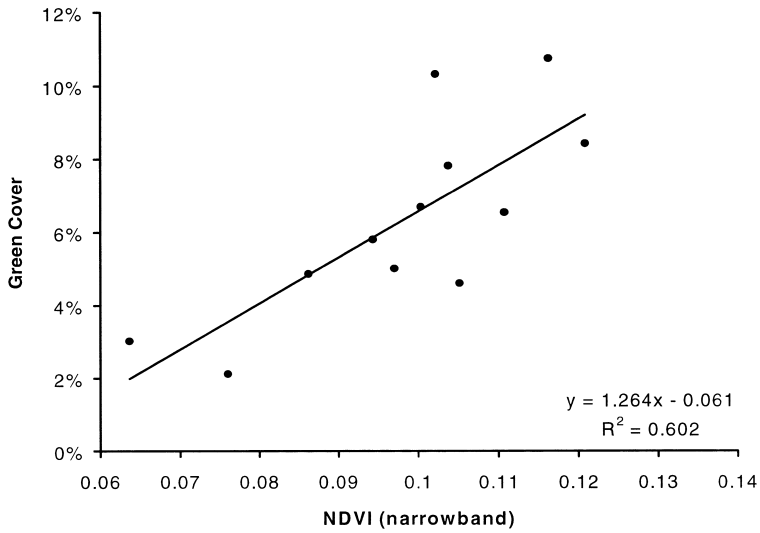


Figure 4. Regression of narrowband NDVI vs. percent canopy cover.

while differences between MSAVI and NDVI were not very significant. Thus, although it is not certain that MSAVI is significantly better than the broadband NDVI, it appears important that adjustments to NDVI be based on locally derived soil effects (MSAVI) rather than a "universal" *L* factor (SAVI). By extension, we might expect that the more diverse the IR/red ratios of soils in a region, the less effective a single, locally developed correction like MSAVI will be.

With an r^2 of 0.74, the linear mixture modeling approach corresponds better with field measurements than any of the vegetation indices (Fig. 8). The improvement of the mixture model over the other methods is quite significant (mixture 1 in Tables 2 and 3). This improvement is likely due to the manner in which the linear mixture model explains underlying variability in background soils on a per-pixel basis. It is also possible that the hyperspectral vegetation reflectance curve provides a better signal than can be obtained from just two spectral bands. Given eight soil endmembers, it is also possible that some variance that was not explained by the ratios (e.g., stems, litter) was unintentionally compensated for by endmembers of soils not actually present in a given pixel. The linear mixture model provided a negative value for one of the sample points. This physically impossible result was not constrained in the mixture model, and it is likely that it arose from the sample lying outside the convex hull that was defined in spectral space by the 10 endmembers.

In examining Figures 4 through 8 it can be seen that for those cases with low r^2, the regressed relationship is driven to a large extent by two isolated samples with very low green cover. In these cases, samples with higher green cover on the right side of the figures are not well explained. This suggests that background soil influences allowed only the coarsest changes in green cover to drive the regressions. Because the spread of variance is lopsided (nonstationary), these indices may be very suscepti-

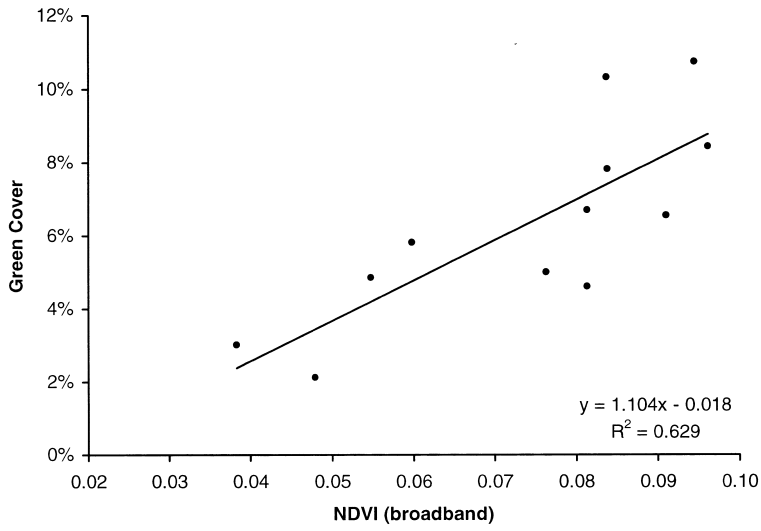


Figure 5. Regression of broadband NDVI vs. percent canopy cover.

Table 2. Cross-Validated Regression Parameters

	$NDVI_1$	$NDVI_2$	*SAVI*	*MSAVI*	*Mixture #1*	*Mixture #2*
R^2						
Mean	0.604	0.629	0.513	0.644	0.737	0.855
Standard Deviation	0.053	0.046	0.060	0.044	0.042	0.021
Intercept						
Mean	−0.061	−0.018	−0.038	−0.051	0.027	0.031
Standard Deviation	0.010	0.005	0.011	0.009	0.002	0.002
Slope						
Mean	1.271	1.105	1.408	1.078	1.075	0.530
Standard Deviation	0.105	0.074	0.152	0.085	0.068	0.022
Standard Error	1.9%	1.8%	2.1%	1.8%	1.6%	1.2%

$NDVI_1$=narrowband; $NDVI_2$=broadband; Mixture #1=average; Mixture #2=with *K. erecta*.

ble to statistical outliers. There is some improvement in the distribution of variance along the regression line with broadband NDVI and MSAVI, but the linear mixture model displays the best spread of variance throughout the regressed relationship.

The r^2 reported here for the linear mixture model is less than the r^2 of 0.83 reported by Smith et al. (1990) when using Landsat TM data in the Owens Valley of California. This shortfall might be somewhat surprising given the contrast of high spatial resolution, hyperspectral data vs. the much coarser spatial and spectral resolution of TM imagery. It is possible that the higher resolution of the Probe-1 aircraft data increased the relative importance of error in registering the imagery to the field sites (McGwire et al., 1993). In other words, the chance inclusion or exclusion of shrubs along the boundaries of pixels that covered the field plots would cause a spurious increase in variance regardless of the quality of the image transformation. However, the regression plot in Smith et al. (1990; p. 18) shows that their highest r^2 is likely to be strongly influenced by two highly leveraged samples, resulting in some degree of uncertainty in their reported r^2 value. The range of vegetation density found in Smith et al. (1990) was also two to three times greater than that reported here. It is quite possible that the very limited range of vegetation density in our study area did not allow the proportion of variance explained by the regression line to be fully expressed relative to the unexplained variance. If we were to extend the pattern of points observed in Figures 4 through 8 to a larger range of percent cover values, we would expect the resulting r^2 values to be noticeably higher.

It is interesting to note that the output of the linear mixture model must be rescaled to match the field measurements of green-leaf cover (Fig. 8). The Y intercept for the regression may imply that sparse cover that is less than 3% is not detectable, possibly due to a limitation in the signal-to-noise ratio. The slope for the mixture model regression was 1.075, whereas in a perfect case the slope of the line would be 1.0 (dashed line in Fig. 8), or perhaps less to compensate for the Y intercept being greater than 0.0. This expectation is based on the numerical constraint that weights were to be scaled from 0.0 to 1.0. While this small discrepancy might be explained by a bias in field measures, if the regression parameters are accurate they imply that the overall spectral response function is slightly nonlinear since the projected percent cover based on an output weight of 1.0 is greater than 100%. Ray and Murray (1996) document the existence of nonlinear reflectance for creosote bush (*Larrea tridentata*), one of the dominant species in this study area. However, the relationship does appear linear in this range of percent cover values. Smith et al. (1990) report a much greater inflation in the slope of the regression between linear model output and percent cover, with slopes being 1.5 in summer and 1.82 in winter for xeric vegetation on desert bajadas. In their assessment, this discrepancy arose primarily from assuming continuous, opaque canopies in the field sampling technique, though part of this difference is likely due to their rescaling of endmember fractions to remove the effect of shadows. If it is true that field data are responsible for the discrepancy between regressed slope and the expected value of 1.0 in Smith et al. (1990), then it appears that the field

Table 3. Significance of Differences in r^2 (p Value)

	$NDVI_2$	*SAVI*	*MSAVI*	*Mixture #1*	*Mixture #2*
$NDVI_1$	0.2258	<0.0001	0.0551	<0.0001	<0.0001
$NDVI_2$		<0.0001	0.4224	<0.0001	<0.0001
SAVI			<0.0001	<0.0001	<0.0001
MSAVI				<0.0001	<0.0001
Mixture #1					<0.0001

See Table 2 for abbreviations.

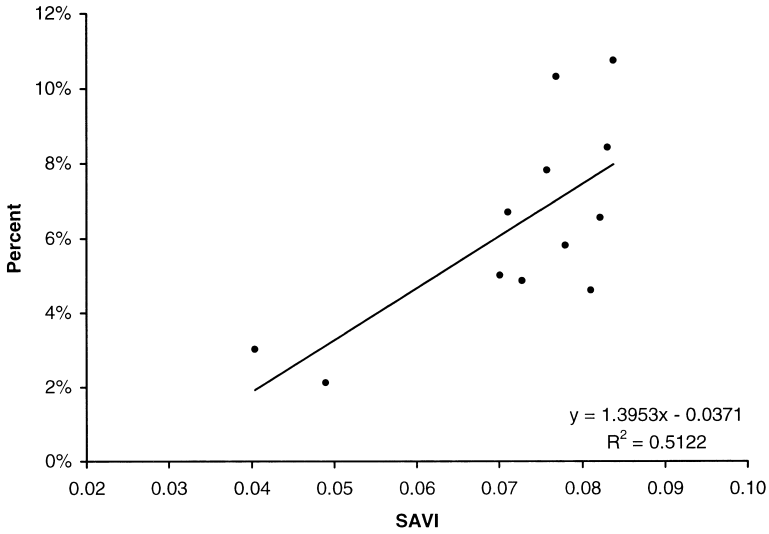


Figure 6. Regression of SAVI vs. percent canopy cover.

sampling method used here that involved visual estimation of canopy density for individual plants provided some improvement.

Figure 9 displays the residuals from the mixture model regression vs. the field-measured percent green cover of various species in each plot. Only four of the species were present in enough of the plots to allow for a meaningful interpretation. The two species used in generating the green-leaf endmember (*L. tridentata* and *A. dumosa*) both appeared to have little relationship to the residuals. The percentage of *Krameria erecta* (previously *K. parvifolia*), by contrast, appeared to have a strong relationship to the residuals, although the greatest level of percent green cover for this species at the plot level was only 1.5%. Given this small percent cover and the general correspondence between *K. erecta* and the average vegetation endmember (Fig. 3), it is quite possible that the residuals were actually due to some other unmeasured spectral variable that happened to be strongly correlated with the presence of *K. erecta.*

To test the possible improvement in the mixture model, the mixture model was run again using an additional spectral endmember for *K. erecta.* The spectral data obtained for *K. erecta* was only for canopy-level reflectance, so for this test, canopy-level reflectance was also used for the average vegetation endmember. The canopy-level, average spectra by itself provided results that were not significantly different from the leaf-level results presented in Figure 8 (r^2=0.72; difference of regressed estimates from leaf-level results: p=0.958). The only noticeable difference between the methods was that the slope of the regression line for the canopy-level results was 0.42, which corresponds approximately to the maximum percent green cover of the measured canopies. The Y intercept for the average canopy-level endmember was 3.1% (cross-validated), again suggesting a detection

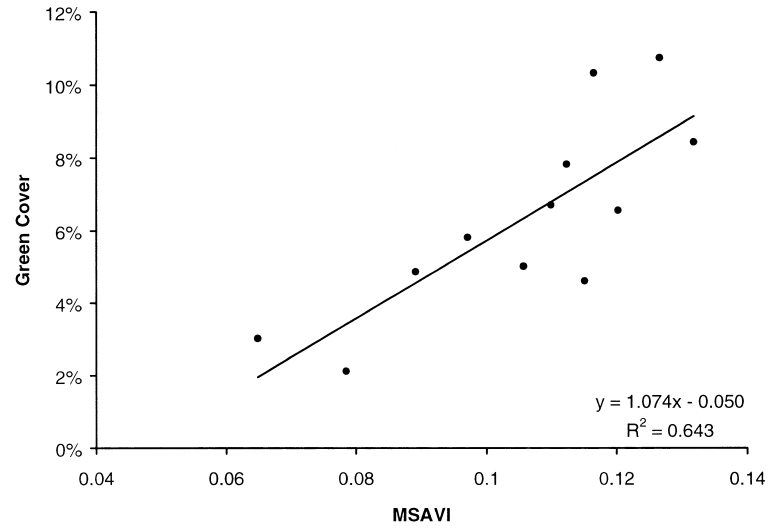


Figure 7. Regression of MSAVI vs. percent canopy cover.

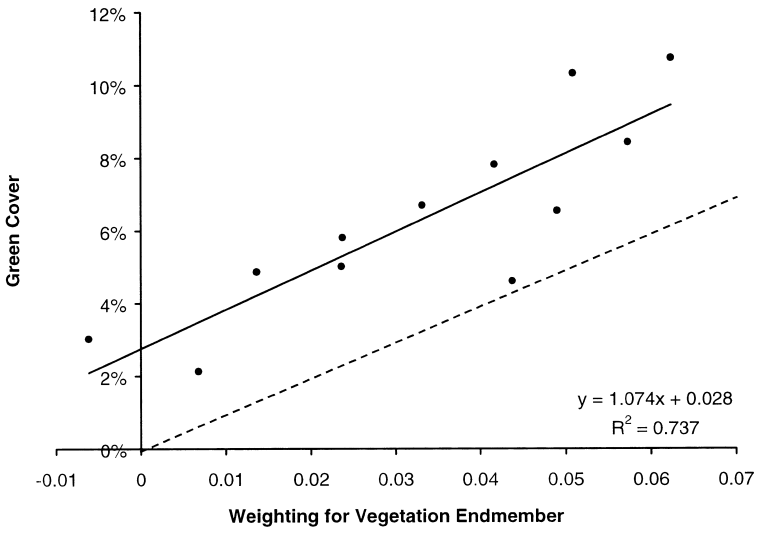
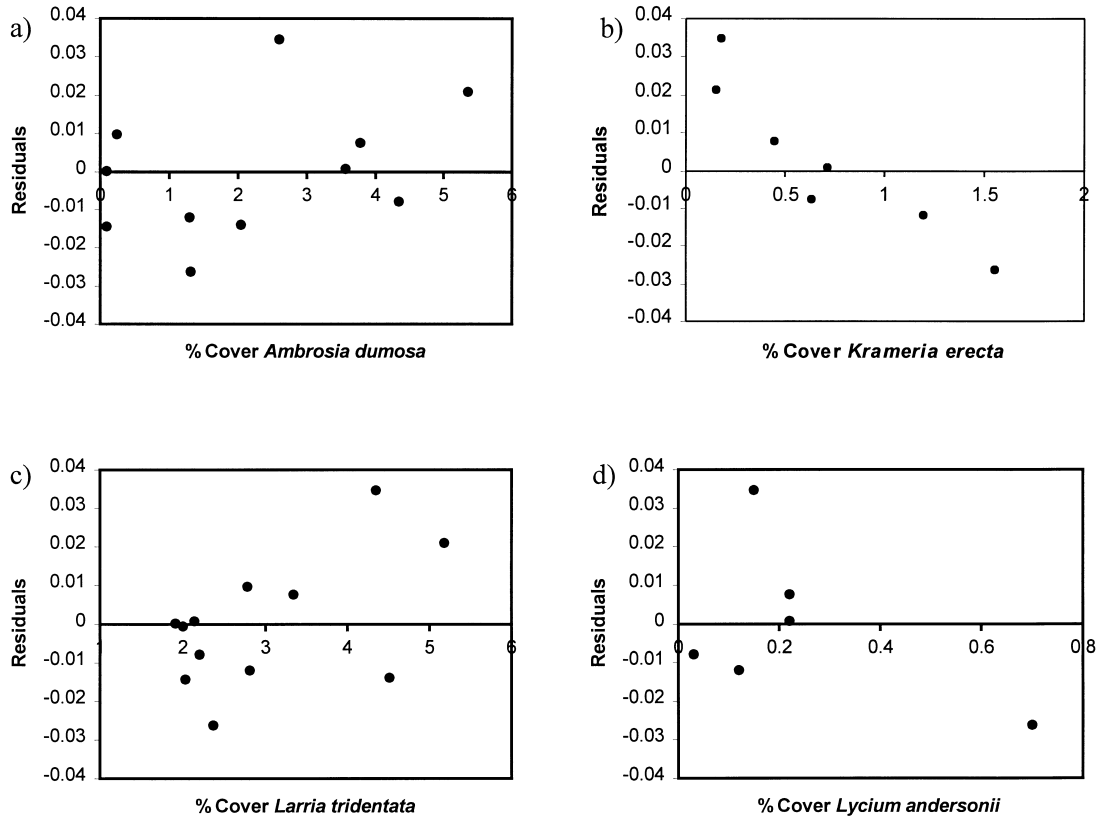


Figure 8. Regression of mixture model output vs. percent canopy cover (1:1 relationship shown as dashed line).

threshold. It appears that the contribution of background soils, stems, litter, and shadows did not have a strong effect on the results because the average canopy-level endmember covered so many different plants and substrates.

The linear mixture model was run using the average canopy-level vegetation endmember, a canopy-level endmember for *K. erecta*, the eight soil endmembers, and a shadow endmember. Output weights for the average vegetation and the *K. erecta* endmembers were summed and regressed against total green cover, and the results are presented in Figure 10. The cross-validated r^2 jumped to 0.86 (mixture 2 in Table 2), significantly better than all other methods tested (Table 3). Despite the improved performance for predicting total green cover, it appears that a problem arises at the level of the individual endmembers. As shown in Figures 11 and 12, the

Figure 9. Species effects on mixture model residuals.



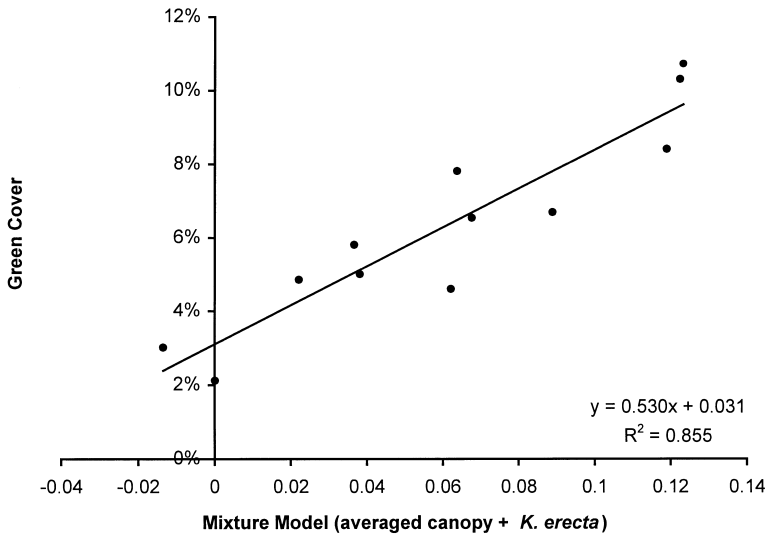


Figure 10. Regression for mixture model with *K. erecta* added.

r^2 for each individual vegetation endmember is very poor. The average canopy-level endmember does not correlate well with the total green cover minus the cover of *K. erecta* (Fig. 11; r^2=0.06), nor does the *K. erecta* endmember correlate with its own abundance (Fig. 12; r^2=0.02). This may occur because the two spectral endmembers are so similar, creating nonunique solutions to the mixture equations where noise plays a large factor in pushing the weights to one endmember or the other. This would be analogous to the statistical problem of multicollinearity.

Taking the apparent success and shortcomings of the multiple vegetation endmembers to a further level, a test of the linear mixture modeling method was performed using separate endmembers for the canopy-level reflectance of seven different plant species. The summed output weights for all vegetation endmembers had little relationship to the total green vegetative cover (r^2=0.07). This breakdown in relationship is likely due to a lack of uncorrelated dimensionality in the spectral curves, inherent variability of plant reflectance, insufficient signal-to-noise ratio, and numerical instability in unmixing technique. Problems associated with too many spectral endmembers were also documented by Roberts et al. (1993).

Table 2 contains the cross-validated standard error in percent green cover for each of the methods. The standard error was generated by predicting back on the samples that were omitted from each regression of the cross-validation. The linear mixture model with a single, leaf-level vegetation endmember provides a 24% improvement over SAVI and an 11% improvement over broadband NDVI and MSAVI. The standard error for the mixture model that incorporated an endmember for *K. erecta* presents a further 33% improvement over the single vegetation endmember. However, until further work clarifies the appropriateness and limitations of multiple vegetation endmembers, it may not be appropriate to assume the *K. erecta* results are repeatable.

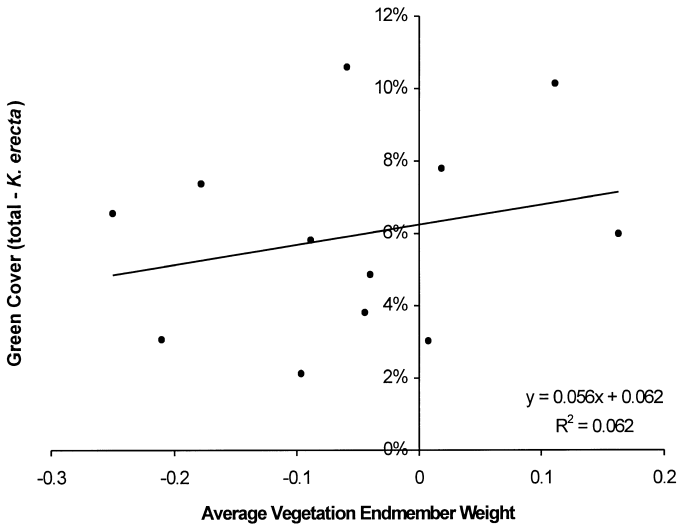


Figure 11. Average vegetation endmember vs. (total percent cover−percent *K. erecta*).

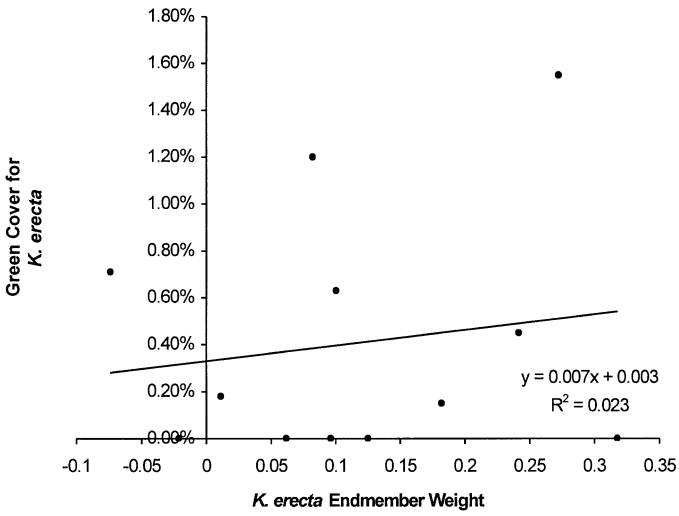


Figure 12. K. erecta endmember vs. percent cover of *K. erecta.*

While the difference in standard error between the various methods appears to be small, small improvements are important in the context of an environment where the percent green cover is generally less than 10%. Multiplying the standard error by 2.2 to develop a 95% confidence interval (*t* distribution, $df=11$) on percent cover, the linear mixture model with a single vegetation endmember would be expected to be accurate to within ±3.6%. However, this estimate assumes that error is randomly distributed, an assumption that is typically untrue in image data (McGwire et al., 1993; Congalton, 1988; Campbell, 1981). Thus, while we would not be certain of absolute green cover to within ±3.6%, we would very likely be able to detect smaller relative changes over distances that were less than the autocorrelation length of the two-dimensional error field.

Due to the field measurement technique, some small uncertainty is associated with the precision of the regression results and detection thresholds that are reported here. While it was found that the visual estimates of percent cover for *L. tridentata* compared well with an independent grid count, visual assessments were somewhat coarse (15% increments), were inherently subjective, and their accuracy may have varied by species.

CONCLUSION

The use of linear mixture modeling with hyperspectral data provided significantly better results than the standard vegetation indices that were tested. Given an average green-leaf spectra that is representative of plants in a region, the mixture modeling method described here could be run in a data-driven manner like NDVI. However, these results are derived from a small study area, and further work should be done to document geographical and temporal variability for such relationships. Unfortunately, since the output from the mixture model needs to be rescaled and some species may have strong effects on the results, field work will still be required to match the model output to actual percent cover. There appears to be a detection threshold of 3% green cover for linear mixture modeling with the Probe-1 imagery.

It is interesting that the SAVI technique performed so poorly, and that the MSAVI technique did not provide a significant improvement over a simple broadband NDVI. Apparently, some degree of caution is required in using these methods. By directly incorporating multiple soil endmembers into the solution on a pixel-by-pixel basis, mixture modeling with hyperspectral data is well suited to the challenges of quantifying sparse vegetation coverage. In essence, the interaction of these soil endmembers with the shadow endmember goes beyond the single soil adjustment of MSAVI to compensate for multiple soil lines in the image.

The *K. erecta* results suggest that interspecies variability may play a significant role in the performance of the linear mixture model. However, given that this effect likely arose from some other phenomena that was correlated with *K. erecta*, the lesson may have more to do with omitted endmembers. Because of certain strong differences in spectral response, future work should include the development of a spectral taxonomy of xeric species that provides information on the potential for aggregating sets of species. However, variability in spectral response for a given species may provide a fundamental limit to the implementation of such a taxonomy (Price, 1994), as is apparent in Fig. 3. In addition, there are likely to be problems with the generation of unique solutions to mixture model equations when spectral endmembers are very close in spectral space. An understanding of how dispersion between endmembers affects the reliability of results is required. Problems with numerous endmembers might be alleviated by using multiple endmember spectral mixture analysis (Okin et al., 1998) to provide an a priori selection of expected endmembers.

This work was funded by the Department of Energy through the Hyperspectral-Multispectral Algorithm Research Center at Bechtel Nevada (Contract No. DE-AC099-96NV11718) and by the Desert Research Institute of the University and Community College System of Nevada. The authors are indebted to J. McAuliffe for assistance with the field site and Bill Farrand for his assistance with the Probe-1 data calibration.

REFERENCES

Adams, J., Smith, M., and A. Gillespie (1993), Imaging spectroscopy: Interpretation based on spectral mixture analysis. In *Remote Geochemical Analysis: Elemental and Minerological Composition* (C. M. Pieters, P. A. J. Englert, Eds.), Cambridge University Press, New York, pp. 145–166.

Boardman, J. W. (1993), Automated spectral unmixing of AVIRIS data using convex geometry concepts. In *Summaries, Fourth JPL Airborne Geoscience Workshop*, JPL Publication 93-26, NASA Jet Propulsion Laboratory, Pasadena, CA, pp. 11–14.

Campbell, J. (1981), Spatial correlation effects upon accuracy of supervised classification of land cover. *Photogramm. Eng. Remote Sens.* 47(3):355–364.

Carneggie, D. M., Degloria, S. D., and Colwell, R. N. (1974), Usefulness of ERTS-1 and supporting aircraft data for monitoring plant development and range conditions in California's annual grassland, BLM Final Report 53500-CT3-266 (N), U.S. Bureau of Land Management, Washington, D.C.

Clevers, J. G. P. W., and Verhoef, W. (1993), LAI estimation by means of the WDVI: A sensitivity analysis with a combined PROSPECT-SAIL model. *Remote Sens. Environ.* 7:43–64.

Colwell, J. E. (1973), Bidirectional spectral reflectance of grass canopies for determination of above ground standing biomass, Ph.D. thesis, University of Michigan, University Microfilm 75-15, 693.

Colwell, J. E. (1974), Vegetation canopy reflectance. *Remote Sens. Environ.* 3:175–183.

Congalton, R. (1988), Using spatial autocorrelation analysis to explore the error matrices for assessing the accuracy of maps generated from remotely sensed data. *Photogramm. Eng. Remote Sens.* 54(5):587–592.

CSES (Center for the Study of Earth from Space) (1997), Atmosphere REMoval Program (ATREM) User's Guide, Version 3.0, Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado, USA.

Elvidge, C. D., Chen, Z., and Groenveld, D. P. (1993), Detection of trace quantities of green vegetation in 1990 AVIRIS data. *Remote Sens. Environ.* 44:271–279.

ESSI (Earth Search Sciences Inc.) (1999), The Probe-1, http://www.earthsearch.com/probe1.html.

Farrand, W. H., and Harsanyi, J. (1997), Mapping the distribution of mine tailings in the Coeur d'Alene River Valley, Idaho, through the use of a constrained energy minimization technique. *Remote Sens. Environ.* 59:64–76.

Gillespie, A. R., Smith, M. O., Adams, J. B., Willis, S. C., Fischer III, A. F., and Sabol, D. E. (1990), Interpretation of residual images: Spectral mixture analysis of AVIRIS images, Owens Valley, California. In *Proceedings of the Airborne Science Workshop*, JPL Publication 90-54, Jet Propulsion Laboratory, Pasadena, CA, pp. 243–270.

Green, A. A., Berman, M., Switzer, P., and Craig, M. D. (1988), A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Transact. Geosci. Remote Sens.* 26(1):65–74.

Holben, B. N., and Justice, C. O. (1981), An examination of spectral band ratioing to reduce the topographic effect on remotely sensed data. *Int. J. Remote Sens.* 2:115–121.

Huete, A. R. (1988), A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* 25:295–309.

Huete, A. R., Hua, G., Qi, J., Chehbouni, A., and van Leeuwen, W. J. D. (1992), Normalization of multidirectional red and NIR reflectances with the SAVI. *Remote Sens. Environ.* 41:143–154.

Huete, A. R., and Tucker, C. J. (1991), Investigation of soil influences in AVHRR red and near-infrared vegetation index imagery. *Int. J. Remote Sens.* 12:1223–1242.

Hurcom, S., and Harrison, A. (1998), The NDVI and spectral decomposition for semi-arid vegetation abundance estimates. *Int. J. Remote Sens.* 19(16):3109–3125.

McAuliffe, J. R., and McDonald, E. V. (1995), A piedmont landscape in the Eastern Mojave Desert: Examples of linkages between biotic and physical components. *San Bernardino County Museum Assoc. Quarterly* 42:53–63.

McDonald, E. V. (1994), The relative influences of climatic change, desert dust, and lithologic control on soil-geomorphic processes and hydrology of calcic soils formed on Quaternary alluvial-fan deposits in the Mojave Desert, California, PhD Diss., Univ. New Mexico, Albuquerque.

McDonald, E. V., and McFadden, L. D. (1994), Quaternary stratigraphy of the Providence Mountains piedmont and preliminary age estimates and regional stratigraphic correlation of Quaternary deposits in the eastern Mojave Desert, California. In *Quaternary Stratigraphy and Dating Methods: Understanding Geologic Processes and Landscape Evolution within Southern California* (S. G. Wells, J. Tinsley, L. D. McFadden, and N. Lancaster, Eds.), GSA Cordilleran Section Guidebook: Geological Investigations of an Active Margin, pp. 205–210.

McGwire, K., Friedl, M., and Estes, J. (1993), Spatial structure, sampling design, and scale in remotely sensed imagery of a California savanna woodland. *Int. J. Remote Sens.* 14: 2137–2164.

Mustard, J. F., and Pieters, C. M. (1989), Photometric phase functions of common geologic minerals and applications to quantitative analysis of mineral mixture reflectance spectra. *J. Geophys. Res.* 94(13):619–634.

Myneni, R. B., Ganapol, B. D., and Asrar, G. (1992), Remote sensing of vegetation canopy photosynthetic and stomatal conductance efficiencies. *Remote Sens. Environ.* 42:217–238.

Okin, G., Okin, W., Roberts, D., and Murray, B. (1998), Multiple endmember spectral mixture analysis: Application to an arid/semi-arid landscape. In *Proc. 7th AVIRIS Earth Science Workshop*, JPL Publication 98-21, Pasadena, CA, pp. 291–299.

Price, J. (1994), How unique are spectral signatures? *Remote Sens. Environ.* 49:181–186.

Qi, J., Chehbouni, A., Huete, A., Kerr, Y., and Sorooshian, S.

(1994), A modified soil-adjusted vegetation index (MSAVI). *Remote Sens. Environ.* 48:119–126.

Ray, T. W., and Murray, B. C. (1996), Nonlinear spectral mixing in desert vegetation. *Remote Sens. Environ.* 55:59–64.

RSI (Research Systems, Inc.) (1997), *ENVI*, The Environment of Visualizing Images, Version 2.6, Research Systems, Inc., Boulder, Colorado.

Roberts, D., Smith, M., and Adams, J. (1993), Green vegetation, vegetation, and soils in AVIRIS data. *Remote Sens. Environ.* 44:255–269.

Rondeaux, G., Steven, M., and Baret, F. (1996), Optimization of soil-adjusted vegetation indices. *Remote Sens. Environ.* 55:95–107.

Rouse, J. W., Haas, R. H., Schell, J. A., Deering, D. W., and Harlan, J. C. (1974), Monitoring the vernal advancement and retrogradation (greenwave effect) of natural vegetation, NASA/GSFC Type III Final Report, Greenbelt.

Sellers, P. (1985), Canopy reflectance, photosynthesis and transpiration. *Remote Sens. Environ.* 21:143–183.

Sellers, P. (1987), Canopy reflectance, photosynthesis and transpiration. II. The role of biophysics in the linearity of the interdependence. *Int. J. Remote Sens.* 6(8):1335–1372.

Smith, M. O., Ustin, S. L., Adams, J. B., and Gillespie, A. R. (1990), Vegetation in deserts: I. A regional measure of abundance from multispectral images. *Remote Sens. Environ.* 31:1–26.

Tompkins, S., Mustard, J. F., Pieters, C. M., and Forsyth, D. W. (1997), Optimization of endmembers for spectral mixture analysis. *Remote Sens. Environ.* 59:472–489.

Tucker, C. J. (1979), Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* 8:127–150.

Vance, L., Stewart, R., and Farrand, W. (1999), Hyperspectral imagery applications using Probe-1. In *Thirteenth International Conference on Applied Geologic Remote Sensing*, Vancouver, March 1–3, 1999, pp. I-553 to I-559, ERIM, Ann Arbor, Michigan.