REMOTE SENSING FOR ENVIRONMENTAL BASELINING AND MONITORING

David Reister, Robert Washington-Allen, and Art Stewart

INTRODUCTION

Hyperspectral remote sensing is an emerging technology with the potential to identify plant species, map vegetation, characterize soil properties, identify contamination, classify ecological units and habitat characteristics, and differentiate causes of vegetation stress. The main benefit of this project, both to the DOE Fossil Energy Program and to our petroleum industry partners, will be an improved ability to provide detailed environmental data for a region, rapidly, at low cost. Increasingly, environmental data are needed to assess present conditions of lands owned, leased or managed by petroleum companies, and to characterize and quantify changes in the environmental conditions of these lands through time. Present methods of assessing large areas depend extensively on field surveys, which can take weeks or months to complete. Such methods are inconvenient and can be expensive. Further, some areas that are inaccessible or very large are difficult or virtually impossible to monitor accurately using field-survey techniques only.

Investigators from ChevronTexaco, the University of California at Davis (UCD), the Carnegie Institution of Washington at Stanford University (CIW), the U.S. Geological Survey (USGS), the U.S. Department of Agriculture's Agricultural Research Service (ARS), and the Oak Ridge National Laboratory (ORNL) are collaborating to develop remote (airborne or satellite) hyperspectral sensor techniques specifically to characterize conditions at exploration and production (E&P) sites, where vegetation and soil may be impacted by oil or gas production. Project objectives include: (1) identifying hyperspectral signatures of plant responses to hydrocarbons, toxic metals and other stressors; (2) developing advanced analytical methods for evaluating hyperspectral data, so as to more clearly reveal environmental impacts from spectrally "noisy backgrounds;" and (3) demonstrating techniques for revealing environmental impacts by analysis of hyperspectral data. These techniques are not currently available. This report will discuss three topics: (1) Analysis of remote hyperspectral data at the Jornada Experimental Range near Las Cruces, NM (Jornada), (2) Analysis of field hyperspectral data from Jornada, and (3) Plans to collect airborne high spatial resolution hyperspectral imagery in Osage County , OK, analyze the data set to identify distinct groups of similar spectra, and perform field validation.

REMOTE HYPERSPECTRAL ANALYSIS

An ideal (but costly) experiment for this project would be to apply petroleum hydrocarbons on a plot, and obtain hyperspectral images of the plants on the plot before and after the spill, to characterize changes in vegetative condition through time. In June 2000, a near-ideal "natural" experiment occurred when a road grader accidentally cut an oil pipeline, allowing oil to spray over a two-hectare site at Jornada (see

Fig.1). Jornada has been operated as a research park by the Agricultural Research Service (ARS) since 1912 (see http://usda-ars.nmsu.edu/). Thus, the oil-spill accident at this park was a particularly fortuitous event for our project. In 1981, the National Science Foundation selected Jornada as a Long-Term Ecological Research (LTER) site. The Jornada Experiment (JORNEX) was begun in 1995 to collect remotely sensed data from ground, airborne, and satellite platforms. In May 1997, an 11-day field sampling campaign occurred at Jornada. This sampling campaign was designed to collect ground and airborne hyperspectral data to validate satellite images. During the campaign, the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor was used to collect data at Jornada. Subsequently, this sensor has been used to collect data at Jornada at least twice a year. AVIRIS is an imaging hyperspectral sensor with 224 bands; it is carried aboard a NASA ER-2 airplane (a modified U2 plane) at an elevation of ~20 km. An AVIRIS flight over the Jornada site occurred on June 10, 2000, just ten days after the oil-spill. Unfortunately, the AVIRIS image missed the oil spill site. Subsequent AVIRIS flights occurred in September 2000 and June 2001. The Jornada site also was imaged with the AVIRIS sensor before the spill (September 1999), and hyperspectral data on vegetation and soils at the site have been obtained during site-characterization projects by ARS investigators. Thus, Jornada has a rich set of before-and-after hyperspectral data, both from ground and airborne platforms.

Fig. 1. Oil spill at Jornada.

The Jornada oil spill site has a plant canopy dominated by creosote bush (Larrea tridentata (Sesse & Moc. Ex DC.) Coville) shrub land that is one of the five major plant community types in the Jornada basin. Qualitative field inspection in early February 2001 indicated that upper plant canopy contact with the diesel fuel was manifest as: (1) etiolation (lack of photosynthetic activity) and probable chlorosis (loss of chlorophyll) that resulted in a grey to white color of the upper canopy and a white to slight reddening of the lower canopy graminoids and litter, (2) partial and complete defoliation of shrubs, (3) apparent high mortality of much of the above ground phytomass, including grasses, cactoids, and biological crusts, and (4) darkening of the orange-red alluvial soil (see Fig. 1). It was also evident that the spill boundary could be delineated on the basis of smell, as diesel was still volatizing from the soil. These features were still evident one year after the release. It must be noted that the canopy dominant, creosote bush, is expected to recover from the diesel spill because it resprouts after injury. This aspect of plant physiology is significant for studies of resilience in desert ecosystems.

The AVIRIS sensor was used to collect data at Jornada at least twice annually, including prior to and after the diesel spill. Consequently, three scenes from September 1999 and 2000, and June 2001 are available for the diesel spill site. Time series analysis of hyperspectral data is rare, only one other study prior to this has been conducted and that was by Dr. Susan Ustin and one her students at UCD. Consequently, this is an opportunity for us to study the response and recovery of a desert plant community to a diesel spill using indicators derived from hyperspectral imagery. Past studies have shown that plant stress is usually detected between 400 to 735 nm and hydrocarbons near 1700 and 2200 nm. The three AVIRIS scenes were acquired from Dr. Greg Asner at CIW. Analysis has been completed for the two September scenes and indicated separation between the before and during spectral responses, with the “during” response showing plant response in the 400 to 735 nm and evidence of absorption in the 1700 nm region. The June 2001 “after” scene is currently being processed by both ORNL and UCD for comparison to the 1999 and 2000 scenes. A manuscript of these results is in preparation and will be submitted to the Journal: Ecological Applications.

FIELD HYPERSPECTRAL ANALYSIS

In April 2002, we received data for the 665 field measurements that were made at the Jornada Experimental Range by ARS investigators in September 2000. Using the field data, our goal was to explore the resolution limits for pixel unmixing and plant species identification. We have completed an analysis of 211 measurements that were made at the grass site. Each measurement consists of 2151 numbers that are the reflectance values at wavelengths ranging from 350 nm to 2500 nm. Each measurement is described with one or two of the following 12 labels: “Bare,” “Litter,” “Aristada,” “Capa,” “Datu,” “Forb,” “Grass,” “Mesquite,” “Senna,” “Snakeweed,” “Tila,” and “Yucca.” Here, “Bare”

represents bare soil, “Litter” represents plant litter, and the other 10 labels designate plant species. This data set can be used to test whether or not each of the 12 labels has a distinct hyperspectral signature. We have completed a paper: Endmembers, Pixel Unmixing, and Clustering. and submitted it to the IEEE Transactions on Geoscience and Remote Sensing. The three authors are: David B. Reister (ORNL), Jerry C. Ritchie (ARS), and Albert Rango (ARS). This section will provide a short summary of the paper.

Endmembers have been used to analyze multispectral and hyperspectral images since the paper by Adams, et al. [1] that analyzed an image of Mars using four image endmembers, that included a shade endmember. The three nonshade image endmembers were identified by comparison with a library of laboratory reflectance spectra. The measured spectrum (ρ_i) of a pixel in the image is the weighted average of the spectra of the endmembers (a_{ij}) and the weights (p_j) are the fraction of the total area of the image that is occupied by each subcomponent:

$$\rho_i = \sum_j a_{ij} p_j \quad , \tag{1}$$

where the weights are nonnegative numbers that sum to one.

Gillespie, et al. [2] analyzed hyperspectral AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) data using image endmembers. The number of endmembers is one more than the numerical rank (K) of the measurements, which can be determined by using Principal Component Analysis (PCA) to find the number of dimensions required to explain the data variance. However, PCA does not identify the endmembers, which are extreme pixels in data clusters. Boardman and Goetz [3] estimated an illumination factor at each pixel and constrained the weights to be nonnegative and to sum to one or less. Boardman [4] uses the framework of the geometry of convex sets to identify the K+1 end members as the vertices of the smallest simplex that bounds the measured data. Roberts, et al. [5] have developed multiple endmember spectral mixture analysis, in which they evaluated 889 two-endmember models and used the evaluation results to generate 276 three-endmember models. The best 24 two-endmember and 12 three-endmember models were chosen to generate vegetation maps.

Boardman and Roberts are using endmembers in quite different ways. Boardman generates a single set of endmembers and calculates the weights for each pixel to determine the fractional abundance patterns in the region. Roberts has a large library of laboratory and field measurements and evaluates many two-endmember models. Since a two-member model has one free parameter, he finds which model provides the best fit for each pixel. However, the results are not unique and one or more of the models may provide a good fit for a pixel.

Our objective is the same as Boardman and Roberts: to group hyperspectral measurements into as many distinct classes as possible. Richards and Jia [6] draw the distinction between information classes

and spectral classes, where information classes are human classifications and spectral classes are groups or clusters of measurements made by a computer.

We will consider (1) to be a general linear mixing model and allow the spectra and weighs to be any real number. The spectra are columns of the matrix $\boldsymbol{A}$. All linear combinations of the spectra form a subspace (the range space of $\boldsymbol{A}$) that has a numerical rank (K) and can have many sets of basis vectors. The null space of $\boldsymbol{A}^{\mathrm{T}}$ is all vectors that are orthogonal to $\boldsymbol{A}^{\mathrm{T}}$.

We distinguish between two problems: forward and inverse. For the forward problem, we are given the spectra of the subcomponents and the weights, and we calculate the measured spectra. For the inverse problem, we are given the spectra of the subcomponents and the measured spectra, and we calculate the weights. While the forward problem always has a unique solution, the inverse problem usually does not have a unique solution. Given a solution to the inverse problem, another solution can be found by adding any vector from the null space of $\boldsymbol{A}^{\mathrm{T}}$.

Since pixel unmixing is an inverse problem, there should be many solutions to the pixel unmixing problem. The basic problem is that the measured spectra are not linearly independent, but are highly correlated. Singular value decomposition (SVD) [7] can be used to create an uncorrelated orthonormal basis for the spectra, that will provide a solution to the inverse problem. The solution for the inverse problem could also be expressed as a linear combination of any K linearly independent members of the original spectral library. However, this form of pixel unmixing is neither unique nor informative.

The second section of the paper describes a method for partitioning hyperspectral measurements into distinct classes. When we began our research, we expected that the distinct classes would be a basis for the measurements. As the research has progressed, we realized that the basis vectors are not classes. Instead, the classes are groups (clusters) of distinct measurements. There can be many more classes than there are basis vectors. A familiar example is the population distribution on the Earth's surface. We know that there are many distinct towns and cities on the 2D surface of the Earth.

The third section discusses the methods used to collect a hyperspectral field data set at the Jornada Experimental Range near Las Cruces, NM. In the fourth section, we use our clustering method to partition the field measurements into distinct groups and compare the clusters (spectral classes) with the labels (information classes) that were given to each of the field measurements at the time of data collection.

Our objective is to group hyperspectral measurements into as many distinct classes as possible. A new clustering algorithm for unsupervised classification has been developed. The algorithm has been applied to a hyperspectral field data set (with information classes). Using SVD, we calculated an orthonormal basis for the 211 measurements. When we define the numerical rank (K) as the number of basis vectors required to explain 99% of the variance in the data, we find that K = 4 for this data set.

In our clustering algorithm, we vary the cluster radius to find the maximum number of clusters. For this data set, the maximum number of clusters is 20. Each cluster consists of the neighbors of a root measurement vector that has a label. In all 20 cases, the label of the root measurement in the cluster is consistent with the cluster membership. All of the 12 labels are at the root of at least one of the clusters.

Using the first five vectors in the SVD basis for the measurements, we have plotted the location of the 20 root vectors on 2D "slices" through a 5D space. While axis 1 is not useful for distinguishing distinct groups in the central clusters (the 12 root vectors that have more than a 0.95 correlation with axis 1), it is quite useful for separating the central clusters from the outlier clusters (the 8 root vectors that have lower correlations with axis 1). The slices reveal that each of the 12 root vectors for the central clusters has a distinct location in the 5D space.

ENVI has two clustering algorithms for unsupervised classification: Isodata and K-means. For our data set, the clusters found by both Isodata and K-means were significantly different than our clusters. We found that the most of the cluster radii were much larger than our radius and the largest radius was more than a factor of 10 larger. When a cluster has a large radius, many members of other clusters can be inside the ball defined by the radius. Further analysis found that both Isodata and K-means placed more than half of the measurements in the wrong cluster. Since our algorithm does not require iterations, it requires less computation for large data sets than Isodata and K-means. Thus, our new algorithm appears to advance the state-of-the-art.

REMOTE DATA for OSAGE COUNTY

Osage County, OK is the Osage Indian Reservation and has been a major oil producing area (38,500 oil wells) since 1896. The county is large (2,260 square miles) and 1,480 square miles are within a quarter mile of an oil well. The mineral rights are owned by the Osage Nation and there are 500 independent operators in the county. Many areas in the county have brine scars or weathered oil pits (The USGS has sample photographs: http://ok.water.usgs.gov/skiatook/Skiatook.Photo.html). Figure 2 is a wintertime photo of an Osage County salt scar area. The maximum depth of erosion is about 2 meters. The original soil profile is preserved in the pedestal in the upper right. Saline water in the middle drops below the surface toward the foreground.

Hyperspectral remote data will be collected in selected regions in Osage County and analyzed to detect brine scars, oil pits, and plant stress associated with brine and oil. The imagery will be collected by the HyVista Corporation (http://www.hyvista.com/). The hyperspectral imagery has 126 spectral bands and 3.2 m spatial resolution. Our new data analysis techniques will identify clusters (distinct groups of similar spectra). Limited fieldwork will be required to identify the geological and botanical characteristics of the members of each cluster.

Fig. 2. Osage County salt scar.

REFERENCES

1. J. B. Adams, M. O. Smith, and P. E. Johnson, "Spectral mixture modeling: A new analysis of rock and soil types at the Viking Lander 1 site," *J. Geophys. Res*., vol. 91, no. B8, pp. 8098–8112, 1986.
2. A. R. Gillespie, M. O. Smith, J. B. Adams, S. C. Willis, A. F. Fischer III, and D. E. Sabol, "Interpretation of residual images: Spectral mixture analysis of AVIRIS images, Owens Valley, California," *AVIRIS Workshop Proc.*, JPL Publication 90-54, Jet Propulsion Laboratory, Pasadena, CA, pp. 243–270, 1990.
3. J. W. Boardman and A. F. H. Goetz, "Sedimentary facies analysis using AVIRIS data: A geophysical inverse problem," *AVIRIS Workshop Proc.*, JPL Publication 91-28, Jet Propulsion Laboratory, Pasadena, CA, pp. 4–13, 1991.
4. J. W. Boardman, "Automated spectral unmixing of AVIRIS data using convex geometry concepts," *AVIRIS Workshop Proc.*, JPL Publication 93-26, Jet Propulsion Laboratory, Pasadena, CA, pp.11–14 (1993).

5. D. A. Roberts, M. Gardner, R. Church, S. Ustin, G. Scheer, and R. O. Green, “Mapping chaparral in the Santa Monica mountains Using multiple endmember spectral mixture models,” *Remote Sens. Environ.*, vol. 65, no. 3, pp. 267–279, 1998.
6. J. A. Richards and X. Jia, “Remote Sensing Digital Image Analysis,” Third Edition, Springer-Verlag, Berlin (1999).
7. G. H. Golub and C. F. Van Loan, “Matrix Computations,”3rd ed., Baltimore, MD, The Johns Hopkins University Press, 1996.