Techniques Developed for Geologic Analysis of Hyperspectral Data
Applied to Near-Shore Hyperspectral Ocean Data**

Fred A. Kruse
Analytical Imaging and Geophysics LLC
4450 Arapahoe Ave, Suite 100, Boulder, Colorado 80303 USA

Laurie L. Richardson
Department of Biological Sciences and Drinking Water Research Center
Florida International University, Miami, Florida, 33199 USA

and

Vince G. Ambrosia
JCWS, Inc., MS 242-4, NASA Ames Research Center
Moffett Field, California, 94035-1000 USA

Abstract

Imaging spectrometers or “Hyperspectral Sensors” simultaneously collect spectral data as both images and as individual spectra. A broad range of techniques have been examined, refined, and put into operational practice for analysis of geologic problems. This paper describes a successful geologic case history using an end-to-end approach on Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) data, including data calibration to reflectance, use of a linear transformation to minimize noise and determine data dimensionality, location of the most spectrally pure pixels, extraction of endmember spectra, and spatial mapping of specific endmembers. Several supporting case studies using AVIRIS data of near-shore marine environments demonstrate the viability of these methods for studying the coastal zone. The methods described provide a starting point for image segmentation, material identification, and mapping of marine processes in the near-shore environment.

1.0 Introductions

Hyperspectral data (Imaging Spectrometer data) consisting of hundreds of spectral bands can be analyzed using multispectral analysis techniques, however, classical methods do not take full advantage of the spectral dimensionality of these datasets. Geologists have been analyzing hyperspectral images since 1983, and consequently, a broad range of techniques have been examined, refined, and put into operational practice. The methods described here utilize the unique capabilities of hyperspectral data to locate, map, and identify the materials present on the Earth’s surface. This paper describes a successful geologic case history using this end-to-end approach on Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) data, including data

calibration to reflectance, use of a linear transformation to minimize noise and determine data
dimensionality, location of the most spectrally pure pixels, extraction of endmember spectra, and
spatial mapping of specific endmembers. While ocean data are generally more complex, many of
the concepts developed for geologic analysis are applicable to analysis of near-shore
hyperspectral data.

The Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) is an imaging spectrometer that
simultaneously acquires 224 spectral images at 20 meter spatial resolution and 10 nanometer
spectral resolution (Porter and Enmark, 1987). Each “scene” covers an approximately 10 x 12
kilometer area. Imaging spectrometers collect images, however, in addition, a complete
spectrum can also be extracted from each picture element (pixel) of the image, thus allowing
detailed mapping based on the spectroscopic characteristics of materials (Goetz et al., 1985,
Boardman et al., 1995).

2.0 Geologic Example - Goldfield/Cuprite Area, Nevada
The Goldfield/Cuprite area, Nevada, is a well-known geologic site used for testing of new
sensors and processing strategies. Considerable previous remote sensing information as well as
ground truth exists for the area. AVIRIS data were acquired during July 1995 as part of a group
effort between NASA, AIG, and private industry (Kruse et al., 1996). Analysis results from this
site serve to act as a guide to state-of-the-art processing of AVIRIS data for geologic
applications. Processing conducted on the Goldfield/Cuprite AVIRIS data included calibration
to apparent reflectance using “ATREM”, spectral compression and noise suppression using the
Maximum Noise Fraction (MNF) transformation, determination of endmembers using geometric
methods, and production of mineral maps using a linear spectral unmixing procedure. A
summary of the “standardized” procedures is given in Table 1.

2.1 ATREM Calibration
The Goldfield/Cuprite AVIRIS data were calibrated to apparent reflectance using the “ATREM”
software available from the Center for the Study of Earth from Space (CSES) at the University of
Colorado, Boulder. This software can be obtained via anonymous FTP from “cses.colorado.edu”
in the directory “pub/atrem. Get the readme file for download instructions.

ATREM is an atmospheric model-based calibration routine, and requires input of data
parameters such as the acquisition date and time, the latitude and longitude of the scene, and the
average elevation, along with atmospheric model parameters (CSES, 1992). AIG used ATREM
version 1.31. The output of the ATREM procedure is apparent reflectance calibrated data and a
water vapor image for each scene. Typically, the water vapor image mimics topographic
expression. Higher water vapor concentrations occur in the valleys, and lower water vapor
concentrations over the higher elevations. In the Goldfield/Cuprite case, however, it appears as if
there may have been some modulation by clouds. No clouds were visible in the images,
however, and the ATREM apparent reflectance data appear to have adequately removed water
vapor contributions from the spectra.
Table 1. AIG “Standardized” AVIRIS Processing Methodology

1. Download AVIRIS Quicklook Data
2. Review Spatial Coverage
3. Preliminary Assessment of Spatial Data Quality
4. Define Areas for Further Processing
5. Order AVIRIS Radiance Data for Selected Scenes
6. Download Radiance Data from Tape
7. Data Quality Assessment
   - Spatial Browsing
   - Spectral Browsing
   - SNR Calculations
   - ATREM
   - Spectral Browsing
8. Calibration to Apparent Reflectance
9. Report Data Problems to JPL
10. MNF Transform (spectral compression)
11. Pixel Purity Index (limited iterations for identification of bad pixels)
12. Masking of bad pixels
13. Pixel Purity Index (maximum iterations for endmember determination)
14. n-Dimensional Visualization (Endmember Definition)
15. Comparison of endmembers to spectral library for identification
16. Spectral Angle Mapper (SAM)
17. Spectral Unmixing and/or Match Filtering
18. Annotation, Output, and Report

2.2 MNF Transformation
A “Minimum Noise Fraction” (MNF) Transform was used to reduce the number of spectral dimensions to be analyzed. The MNF transformation is a linear transformation related to principal components that orders the data according to signal-to-noise-ratio (Green et al., 1988). It can be used to determine the inherent dimensionality of the data, to segregate noise in the data, and to reduce the computational requirements for subsequent processing (Green et al., 1988; Boardman and Kruse, 1994). The MNF transformation can be used to partition the data space into two parts: one associated with large eigenvalues and coherent eigenimages, and a second with near-unity eigenvalues and noise-dominated images. By using only the coherent portions in subsequent processing, the noise is separated from the data, thus improving spectral processing results. For the Goldfield/Cuprite AVIRIS data, the eigenvalue plots fall sharply for the first 10 eigenvalues and flatten out. Examination of the eigenimages shows that while the first 10 images contain most of the information, images 11 through 20 still contain coherent spatial detail. The higher numbered MNF bands contain progressively lower signal-to-noise.

2.3 Pixel Purity Index (PPI)
Based on the above MNF results, the lower order MNF bands were discarded and the first 20 MNF bands were selected for further processing. These were used in the “Pixel Purity Index” (PPI), processing designed to locate the most spectrally extreme (unique or different or “pure”) pixels (Boardman et al., 1995). The most spectrally pure pixels typically correspond to mixing endmembers. The PPI is computed by repeatedly projecting n-dimensional scatterplots onto a random unit vector. The extreme pixels in each projection are recorded and the total number of
times each pixel is marked as extreme is noted. A PPI image is created in which the digital
number of each pixel corresponds to the number of times that pixel was recorded as extreme. A
histogram of these images shows the distribution of “hits” by the PPI. A threshold was
interactively selected using the histogram and used to select only the purest pixels in order to
keep the number of pixels to be analyzed to a minimum. These pixels were used as input to an
interactive visualization procedure for separation of specific endmembers.

2.4 n-Dimensional Visualization
Spectra can be thought of as points in an n-dimensional scatterplot, where n is the number of
bands (Boardman, 1993; Boardman et al., 1995). The coordinates of the points in n-space
consist of “n” values that are simply the spectral reflectance values in each band for a given
pixel. The distribution of these points in n-space can be used to estimate the number of spectral
endmembers and their pure spectral signatures, and provides an intuitive means to understand the
spectral characteristics of materials. In two dimensions, if only two endmembers mix, then the
mixed pixels will fall in a line in the histogram. The pure endmembers will fall at the two ends
of the mixing line. If three endmembers mix, then the mixed pixels will fall inside a triangle,
four inside a tetrahedron, and so on. Mixtures of endmembers "fill in" between the endmembers.
All mixed spectra are "interior" to the pure endmembers, inside the simplex formed by the
endmember vertices, because all the abundances are positive and sum to unity. This "convex set"
of mixed pixels can be used to determine how many endmembers are present and to estimate
their spectra.

The Goldfield/Cuprite AVIRIS data were analyzed using these geometric techniques. The
thresholded pixels from the MNF images above were loaded into an n-dimensional scatterplot
and rotated in real time on the computer screen until “points” or extremities on the scatterplot
were exposed. These projections were “painted” using Region-of-Interest (ROI) definition
procedures and then rotated again in 3 or more dimensions (3 or more bands) to determine if their
signatures were unique in the AVIRIS MNF data. Once a set of unique pixels were defined, then
each separate projection on the scatterplot (corresponding to a pure endmember) was exported to
a ROI in the image. Mean spectra were then extracted for each ROI to act as endmembers for
spectral unmixing. Using the IR data only from 2.0 to 2.4 µm, 9 endmembers were defined for
the Goldfield/Cuprite AVIRIS data. These include the minerals calcite, buddingtonite, kaolinite,
muscovite, alunite, and zeolite-group minerals. Another mineral, with an “unknown 2.2 µm
absorption feature” was also located. Based on the spatial distribution of spectra matching this
endmember and known information about the sites, this endmember was identified as
representing opaline silica. Unfortunately, similar spectra also occur on alluvial fans away from
the altered areas, probably because of weathering and/or spectral mixing. Finally, both “light”
and “dark”, relatively aspectral endmembers were defined. These endmembers or a subset of
these endmembers were used for subsequent classification and other processing.
2.5 Spectral Angle Mapper (SAM) Classification
The Spectral Angle Mapper (SAM) is an automated method for comparing image spectra to individual spectra (Boardman, Unpublished data; Kruse et al., 1993). The algorithm determines the similarity between two spectra by calculating the “spectral angle” between them, treating them as vectors in a space with dimensionality equal to the number of bands. Because this method uses only the vector “direction” of the spectra and not their vector “length”, the method is insensitive to illumination. The result of the SAM classification (not shown) is an image showing the best SAM match at each pixel. Additionally, rule images are calculated that show the actual angular distance (in radians) between each spectrum in the image and each reference or endmember spectrum. Darker pixels in the rule images represent smaller spectral angles and thus spectra that are more similar to the endmember spectra. For the purposes of display, the dark pixels are inverted, so that the best matches appear bright. These images present a good first cut of the mineralogy at the sites.

2.6 Spectral Unmixing
While the SAM algorithm does provide a means of identifying and spatially mapping minerals, it only picks the best match to a spectrum. Geologic surfaces are rarely composed of a single uniform material, thus it is necessary to use mixture modeling to determine what materials cause a particular spectral “signature” in imaging spectrometer data. Spectral mixing is a consequence of the mixing of materials having different spectral properties within a single image pixel. If the scale of the mixing is large (macroscopic), then the mixing occurs in a linear fashion. A simple additive linear model can be used to estimate the abundances of the materials measured by the imaging spectrometer. Each mixed spectrum is a linear combination of the "pure" spectra, each weighted by their fractional abundance within the pixel, a simple averaging (Boardman, 1991).

In order to determine the abundances, we must first determine what materials are mixing together to give us the spectral signature measured by the instrument. Selection of “endmembers” is the most difficult part of linear spectral unmixing. The ideal spectral library used for unmixing consists of endmembers that when linearly combined can form all other observed spectra. This can be presented as a simple mathematical model in which the observed spectrum (a vector) is the result of a multiplication of the mixing library of pure endmember spectra (a vector) by the endmember abundances (a vector). An inverse of the original spectral library matrix is formed by multiplying together the transposes of the orthogonal matrices and the reciprocal values of the diagonal matrix (Boardman, 1989). A simple vector-matrix multiplication between the inverse library matrix and an observed mixed spectrum gives an estimate of the abundance of the library endmembers for the unknown spectrum.

Linear Spectral Unmixing was used to produce mineral maps for the Goldfield/Cuprite AVIRIS data. The endmember library defined using the n-dimensional visualization procedure was used in the unmixing process and abundance estimates were made for each mineral. These results can be presented in two ways. First, a set of gray-scale images stretched from 0 to 50% (black to white) provides a means of estimating relative mineral abundances. Selected results are shown in Figure 1. Secondly, color composite images can be used to highlight specific minerals and
mineral assemblages. Pure colors in these images represent areas where the mineralogy is relatively pure. Mixed colors indicate spectral mixing, with the resultant colors indicating how much mixing is taking place and the relative contributions of each endmember. For example, in a color composite for Goldfield (not shown), the minerals Kaolinite, Alunite, and Muscovite when assigned to Red, Green, and Blue in the color output result in distinctive image colors. Areas that are pure red in such an image correspond to areas where kaolinite is the spectrally dominant (~most abundant) mineral. Areas that are green are dominated by Alunite. Areas that are blue contain primarily muscovite. The yellow pixels are an example of mixed pixels, where the contribution of red from kaolinite and of green from alunite results in the mixed yellow color.

Figure 1. Selected 1995 Goldfield AVIRIS Unmixing Results. Bright pixels represent higher abundances stretched from 0 - 50% (black to white).

2.7 Matched Filtering
Matched filtering is based on well-known signal processing methodologies, maximizing the response of a known endmember and suppressing the response of the composite unknown background (Chen and Reed, 1987; Stocker et al., 1990; Yu et al., 1993; Harsanyi and Chang, 1994). It provides a rapid means of detecting specific minerals based on matches to specific library or image endmember spectra. This technique produces images similar to the unmixing as described above, but with significantly less computation. Matched filter results are presented as gray-scale images with values from 0 to 1.0, which provide a means of estimating relative degree of match to the reference spectrum (where 1.0 is a perfect match). For the Goldfield/Cuprite AVIRIS data, this approach produced detailed image-maps of the occurrence of the best matches to the endmember spectra.

3.0 Application to Marine Environments
3.1 South San Francisco Bay

AVIRIS data of salt evaporation ponds near South San Francisco Bay (Moffett Field), California, were acquired in April 1994. Data were calibrated to reflectance using the ATREM model and analyzed using the standardized geologic approach detailed above. A description and previous analysis of these data were published by Richardson et al. (1994).

The reflectance properties of salt evaporation ponds are typically determined by concentrations of algae and photosynthetic bacteria containing a variety of photosynthetic and photoprotective pigments (Richardson et al., 1991; Richardson, 1996). The characteristic reflectance spectrum of a specific pond consists of overlapping spectral features caused by the presence of individual pigments related to single species and mixtures of species. Previous work using derivative spectroscopy has shown that the diagnostic peaks within the individual pigments can be deconvolved from AVIRIS reflectance spectra, thus allowing mapping of the spatial distribution of specific pigments (Richardson et al., 1994).

For this example, the AVIRIS data themselves were used to extract representative spectral endmembers. The MNF procedure described above was used to reduce the spatial dimensionality of the data. The PPI procedure further reduced the dimensionality by selecting those spectra most likely to be pure endmembers. Figure 2 shows selected endmember spectra extracted from the data.

Figure 2. Selected AVIRIS endmember spectra for the Moffett AVIRIS scene.
These spectra represent the relatively “pure” occurrences of specific materials. For example the bottom spectrum in the plot (Mean:n_D Class 1) corresponds to areas of high red reflectance, previously described by Richardson et al (1994) for AVIRIS data of this same site acquired in 1989. The dominant spectral features observed in this spectrum are caused by the red pigment bacteriohodopsin present in photosynthetic Halobacteria (Richardson et al., 1994). Other spectra, show characteristic chlorophyll absorption features near 0.68 mm, both in the water (for example Mean:n_D Class 11) and for near-shore (Mean:n_D Class 14) and on-shore vegetation (Mean: n_D Class 1). Figure 3 shows the results of linear spectral unmixing analysis using these endmembers on the 1994 AVIRIS data. Brighter pixels represent higher abundances. While mixing of materials in water is generally not a simple linear process, examination of n_Dimensional scatterplots for these data indicate that, to the first order, the mixing is linear in this case (as exhibited by straight-line mixing lines between endmembers).

A. Red Pigment Abundance  
B. Green Pigment Abundance  
C. Green Vegetation 1 Abundance  
D. Green Vegetation 2 (Veg. in Water?)

Figure 3. AVIRIS linear spectral unmixing results. Brighter pixels represent higher abundances.
3.2 White Point, California AVIRIS
AVIRIS data of the White Point/San Pedro Channel area, California, were acquired during March 1991. This image covers the outfall area for approximately 374 million gallons of treated wastewater per day from Los Angeles County (Davis et al., 1993). AVIRIS data calibrated to reflectance using a combination of the LOWTRAN 7 radiative transfer model and surface measurements were provided to the authors by Davis. The calibrated AVIRIS data were analyzed using the standardized geologic approach detailed above.

Again, following the geologic procedures, the AVIRIS data themselves were used to extract representative spectral endmembers. The MNF procedure was used to reduce the spatial dimensionality of the data and the PPI procedure was used to determine pure endmember spectra. Figure 4 shows selected endmember spectra extracted from the data. Figure 5 shows matched filter results. Brighter pixels on the images represent higher abundances of the endmember materials.

Figure 4. San Pedro Endmember Spectra

Figure 5a. On-Shore Vegetation

Figure 5b. Kelp Beds
The White Point AVIRIS data allow detailed mapping of the character of the near-shore environment based on spectral properties. Both on-shore and marine vegetation (kelp) are easily mapped as are the active sediment plumes. This methodology also was useful in mapping the distribution of near-shore sediments. Note in Figure 5d above, the scattered occurrence of these sediments away from the immediate shore-line area. This was also observed in previous analysis of these data by Davis et al. (1993) and attributed to resuspended bottom sediment or effluent from the outfall. It should also be noted, that these data are clearly non-linear as demonstrated by curvilinear mixing lines in the n-Dimensional scatterplots. In order to extract quantitative information from these data using the geologic methods, it will first be necessary to linearize the data.

3.3 Florida Bay AVIRIS

AVIRIS data of a portion of northern Florida Bay near Rankin Key were acquired during March 1996. These data are preview or “Beta” radiance data provided by JPL for evaluation of the 1996 JPL calibration procedures, and do not represent final JPL calibration. The Beta radiance data were calibrated to reflectance using the ATREM model and analyzed using the standardized geologic approach detailed above. Previous analysis of Landsat TM data for this general area were published by Richardson et al. (1995) and initial results of AVIRIS analysis are in this proceedings (Ambrosia and Richardson, 1997; Richardson and Ambrosia, 1997).

The 1996 AVIRIS data were used to extract representative spectral endmembers. The MNF procedure was used to reduce the spatial dimensionality of the data. The PPI and n-Dimensional Visualization procedures further reduced the dimensionality by selecting pure endmember spectra. Figure 6 shows selected endmember spectra extracted from the data. Richardson (1996) describes the reflectance characteristics of algal populations, photosynthetic bacteria, and the effect of overlying water on the spectral signatures of mixed populations. Many of the features and characteristics described are present in the AVIRIS endmember spectra from Rankin Key. Diagnostic indications are of a blue-green algal bloom with a diatom component.
Table 2: Key to Reflectance Spectra

<table>
<thead>
<tr>
<th>ID</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>Blue Green Algae 1</td>
</tr>
<tr>
<td>11</td>
<td>Blue-Green Algae 2</td>
</tr>
<tr>
<td>16</td>
<td>Blue-Green Algae 3</td>
</tr>
<tr>
<td>7</td>
<td>Dense near-surface Phytoplankton</td>
</tr>
<tr>
<td>33</td>
<td>Combination dilute phytoplankton/sea grass</td>
</tr>
<tr>
<td>36</td>
<td>Sea Grass</td>
</tr>
<tr>
<td>10</td>
<td>Clear water over sediment 1</td>
</tr>
<tr>
<td>40</td>
<td>Clear water over sediment 2</td>
</tr>
<tr>
<td>31</td>
<td>On-shore Vegetation</td>
</tr>
<tr>
<td>23</td>
<td>Sand/Exposed Sediment</td>
</tr>
</tbody>
</table>

Figure 6. Selected AVIRIS endmember spectra for the Florida Bay AVIRIS scene.

Figure 7. Selected SAM rule images. Brighter pixels represent best spectral match.
4.0 Discussion

Case studies using AVIRIS data of the coastal marine environment demonstrate the viability of methods originally developed for analysis of geologic targets for analysis of ocean hyperspectral data. This analysis allows separation of distinct near-shore characteristics such as bottom type, pigment concentrations and suspended solids, as well as mapping of coastal marshlands, marine vegetation and urban encroachment. In some cases, the spectral properties of materials in the near-shore environment are explained by near-linear mixing, however, clearly, non-linearity is an issue in analysis of these data, and adaptations of this methodology will require that this non-linearity be dealt with by linearizing the data using appropriate transformations prior to analysis.
5.0 References


