# Hyperspectral Data Analysis Procedures with Reduced Sensitivity to Noise[1]

David A. Landgrebe

School of Electrical Engineering
Purdue University
West Lafayette, IN 47907-1285, USA
Internet: landgreb@ecn.purdue.edu
Tel: (317) 494-3486; Fax: (317) 494-6440

*Abstract*

Multispectral sensor systems have become steadily improved over the years in their ability to deliver increased spectral detail. With the advent of hyperspectral sensors, including imaging spectrometers, this technology is in the process of taking a large leap forward, thus providing the possibility of enabling delivery of much more detailed information. However, this direction of development has drawn even more attention to the matter of noise and other deleterious effects in the data, because reducing the fundamental limitations of spectral detail on information collection raises the limitations presented by noise to even greater importance.

Much current effort in remote sensing research is thus being devoted to adjusting the data to mitigate the effects of noise and other deleterious effects. A parallel approach to the problem is to look for analysis approaches and procedures which have reduced sensitivity to such effects.

In this presentation we shall discuss some of the fundamental principles which define analysis algorithm characteristics providing such reduced sensitivity. One such analysis procedure including an example analysis of a data set will be described illustrating this effect.



[1] Proceedings of the Workshop on Atmospheric Correction of Landsat Imagery, pp 172-176, Torrance California, June 29 - July 1, 1993. This work was supported in part by NASA under Grant NAGW-925.

# HYPERSPECTRAL DATA ANALYSIS PROCEDURES WITH REDUCED SENSITIVITY TO NOISE

David Landgrebe

School of Electrical Engineering
Purdue University
West Lafayette IN 47907-1285
Internet: Landgreb@ECN.Purdue.EDU

Noise occurs in multispectral data from many sources, a significant one being that due to the effects of the atmosphere. The problem of adjusting the data to minimize the effect of the atmosphere has proven to be a daunting one. This leads one to consider if one can construct analysis procedures which have a reduced sensitivity to such noise. In this presentation an approach is given to the analysis of hyperspectral data which is based upon some fundamental principles of signal processing and data analysis. The focus is placed upon hyperspectral data because it presents some new opportunities to deal more effectively with such noise sources as atmospheric effects. The presentation begins pointing to the difference between hyperspectral data and more conventional multispectral data, then outlines some basic characteristics of the analysis process . It concludes with an example analysis of an AVIRIS data set illustrating some of these principles.

Although TM would appear to be a logical extension of MSS, hyperspectral data such as that of AVIRIS is a very large step beyond TM. Not only has the spectral detail increased (4 bands to 6 bands to 210 bands), but the signal-to-noise ratio has as well (6 bit to 8 bit to 10 bit data). It is thus reasonable to suggest that large paralleling advances in data analysis methods are needed if the full value of hyperspectral data is to be realized. Because of the complexity of the new data, such data analysis research should proceed in a very rigorous and fundamentally based fashion.

For example, a key question to be addressed for the new environment of large numbers of spectral bands is that of finding or constructing optimal sets of features to be used in a given analysis problem. Methods useful in the past tended to be dependent upon there being small numbers of bands. Thus, in addition to algorithm complexity and computation time being less important, useful features tended to involve a large percentage (e.g., 3 out of seven) of the total number available. In the new environment, useful features may involve either broad regions or be confined to very narrow spectral regions, or some of both. Tools such as the principal components transformation would tend to suppress important features which are narrow. Optimal selection of individual feature subsets would not be feasible due to the large amount of computation required.

Engineering research over the years has resulted in much fundamental knowledge about the process of analysis of complex data. Results drawn from the fields of the communication sciences, pattern analysis, and signal processing are particularly relevant to noise in remote sensing problems. Basic principles which have emerged are useful as a point of departure.

Next, we will outline some relevant ideas in the areas of,

- Means for the quantitative representation of signals, and
- Analysis algorithm characteristics.

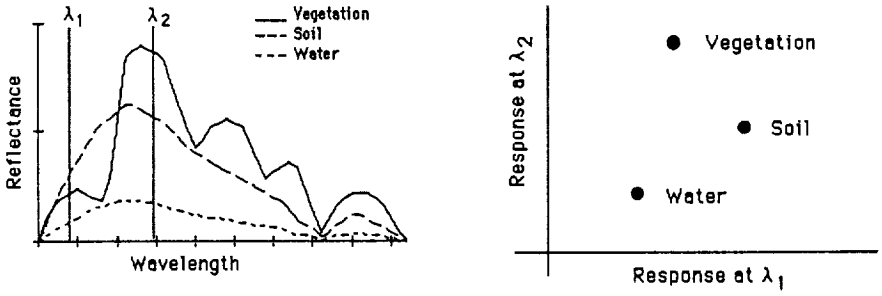
In order that one not unknowingly overlook any information that might be present in such data, one must carefully select the means of data representation which forms the basis for the analysis approach. The data representation should have the following characteristics. It should be

- Broadly Applicable
- Mathematically Rigorous
- Must Not Ignore Any Information-Bearing Attributes

The method proposed is defined by the following pair of equations.

$$x_n = \int_a^b X(\lambda)\,\phi_n(\lambda)\,d\lambda \qquad\qquad X(\lambda) = \sum_{n=1}^{N} x_n\,\phi_n(\lambda)$$

These equations give a general and very powerful method. Basically the process required is a transformation from a continuous function to a discrete, finite dimensional multivariate space. We refer to this as a transformation from spectral space to feature space. A simplified way of thinking of this process is one of sampling, however, the sampling function may take many forms other than that of simple impulse sampling as used in this illustration.




Spectral Space Feature Space

One of the serious concerns in working in higher dimensional feature spaces is that many of the usual properties of two or three dimensional space do not necessarily remain valid. The following are two illustrations of this.

- Borsuk's Conjecture: If you break a stick in two, both pieces are shorter than the original.
- Keller's Conjecture: It is possible to use cubes (hypercubes) of equal size to fill an n-dimensional space, leaving no overlaps nor underlaps.

Counter-examples to both have been found for higher dimensional spaces.[1] Thus one must be very careful about using one's intuition in projecting what may be true with high dimensional data analysis.

The most straightforward way of thinking of a pixel is that a "pure pixel" contains a single material which has a specific spectral response. Given the greater discriminant power of higher dimensional data, this may be an over-simplified point of departure. In reality, any pixel viewed at any resolution is (a) a mixture of a large number of things which (b) involve a variety of observational parameter values. For example, a vegetative canopy pixel would contain a conglomerate of reflectance from the leaf surface, the stems, the background or understory, etc. , and these would be observed under various conditions of illumination and view, at various levels of the canopy, etc.

---

1 Science, Vol. 259, 1 Jan 1993, pp 26-27

Thus different pixels of the same material, having slightly different mixes of these parameters, would have slightly different spectral responses. These mixes of parameter values are characteristic of the material. This means that a material is defined not by a single spectral response, but by a family of (characteristically related) spectral responses.

With regard to analysis methods, there are a number of general characteristics which are axiomatic to obtaining optimal results. Examples are that relative measurements can be made more accurately than absolute ones, situation-specific methods will out-perform general ones (i.e., the jack of all trades is master of none), and one must make full use of all relevant ancillary data. In addition we require that there be no requirement for concomitant observations from the ground.

In the context of hyperspectral data analysis this implies that one wants to discriminate between a set of classes rather than attempt to identify a single class outright. Further, rather than making the algorithm automatic, the algorithm should utilize reference data which is situation specific.

It is useful at this point to step back from the problem and take a broad overview of the entire remote sensing system. The sensor basically acts as a transducer, converting the radiation from the Earth surface to electrical signals. These signals are next transmitted to the processing center, where the data may be preprocessed in some way, e.g., calibrated. Next follows the application of one or more analysis algorithms. A key step in this process is the merging with additional ancillary or reference data, and with the expertise of the analyst.

So far as the extraction of information from the data is concerned, the merging of the new data with the reference ancillary data is a very key step. Indeed, this pretty much defines the broad outlines of the analysis process, i.e., it is a process of merging the new data to be analyzed with reference data or information so as to make the analysis algorithm to be used effective.

However, to make this merging successful there is another key operation required, the reconciliation of the circumstances of the collection of the new data with that of the reference data. This "reconciliation of conditions" step may be carried out in any of three possible ways:

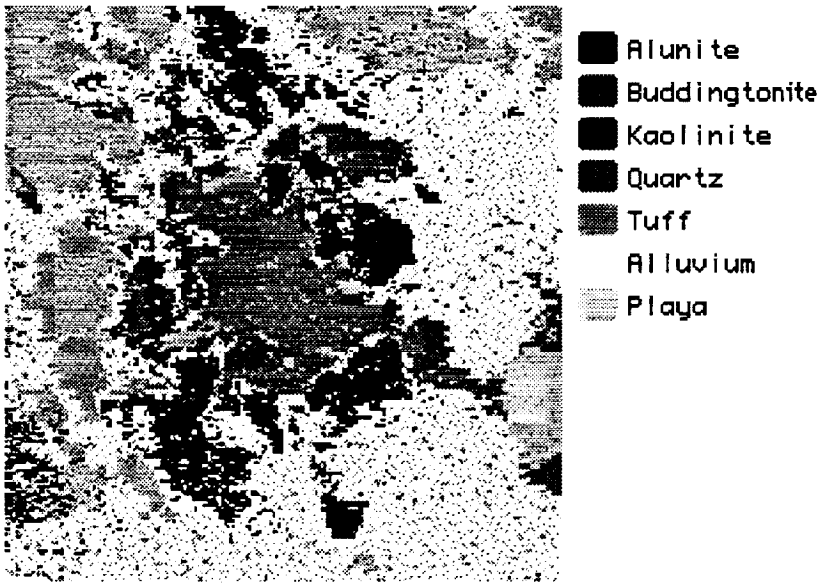
1. The new data may be adjusted to the conditions present in the reference data. This is referred to as the "Stored Signature Approach."
2. Both the new data and the reference data may be adjusted to a third set of conditions. This is the case when both data sets are "Calibrated" to an absolute set of geophysical units, for example.
3. The reference data may be adjusted or referenced to those conditions existing when the new data was collected.

The first two necessarily require a very substantial amount of processing, since they are applied to all of the new data as a minimum. Further, if the level of precision is to be maintained, the precision of the data used in these adjustments must be very high, and as sensor technology continues to advance, so must the level of this precision. A simple but powerful way of accomplishing this reconciliation is to use the third possibility by manually identifying examples of each material in the current data. This observation focuses one's attention upon the reference data, and just what is meant by that term.

By the term, Reference Data, is meant all the relevant data and information which can be merged with the data stream in a favorable way during the analysis process , whether it be quantitative or subjective, e.g., whether it be calibration data or the expertise of the analyst. It was previously argued that one must make maximum use of all information-bearing attributes of the data, and further we argue that the analysis performance is in direct proportion to the effective use of the reference data. Care must be

exercised that both the quantitative and the subjective reference data are used in such a way as to not bias the results inappropriately.

An example analysis of an AVIRIS data set will illustrate how some of these principles come into play. For this example, a data set was chosen purposely as one with a high noise level, and no correction for the atmosphere was used. Four minerals were of particular interest, alunite, buddingtonite, kaolinite, and quartz. The result of the analysis is shown in the following figure. This result compares favorably with more conventional published results obtained from higher quality data.




The quantitative reference data used in this case were known spectroscopic absorption features for each of the four minerals, represented as reflectance (as compared to a radiance) vs. wavelength. Conventional techniques might have one attempt to calibrate all of the data set to absolute units of reflectance so that each pixel could be compared to the four curves to see if it was adequately similar. This approach, though quite functional in many circumstances, requires a quite high signal-to-noise ratio in order to avoid errors, and it is quite dependent upon having accurate calibration information. Further it utilizes only the (obvious) information about the above classes which is apparent from the above known spectral features. It ignores less obvious information which may be available elsewhere in the spectra.

The classification of the data set was done with ECHO, a spectral spatial maximum likelihood algorithm. The processing steps used in obtaining the result have some fundamental differences from conventional ones. Instead of using the four reference spectra directly, they were only used to locate and label training samples in the data set itself. In this way, not only was it possible to avoid all the processing involved in correcting or calibrating all of the data, but the procedure automatically normalizes out many of the observational variations not related directly to the classes of interest. This process also allows in an objective and effective way for the analyst to use his/her knowledge about the positional relationship between different materials and how they might be expected to occur in the scene. The feature selection algorithm used, an algorithm that calculates optimal features in terms of a linear combination of the bands in the range, allows for making use of any characteristic in the wavelength range which will assist in discriminating between the materials, and not simply the four absorption features above.


The above result is an improved one in which several of the limitations arbitrarily imposed in the previous case have been lifted, allowing it to be more typical of what would normally be expected. It was obtained with basically the same procedure, but with the following improvements.

1. A data set which has a higher signal-to-noise ratio was used.
2. Though no diagnostic spectral features beyond the spectral region 2.02-2.35 μm are apparent by manual examination, are there characteristics elsewhere in the 0.4-2.4 μm region which might be useful? A feature extraction technique referred to as discriminate analysis was used to construct a linear combination of the 210 spectral bands which would be optimal for discriminating between the desired classes. The optimal eight dimensional subspace of this new space was then used to classify the data.
3. Significantly greater separation of classes was observed at this point and an additional class was added to the list.
4. The class separations were great enough that it was not necessary to use the ECHO spectral/spatial classifier, although doing so results in a substantially faster classification. The ECHO result is essentially identical to that of standard maximum likelihood, but takes only 60% as much computer time.

All of the processing was done with MultiSpec, a software system implemented on the Apple Macintosh, and working in conjunction with Matlab, thus demonstrating the fact that though the data are complex, a low-cost analysis system could be used quite effectively, and no programming skills in a compiler language are required.

In summary, rather than focusing upon calibrating the data or correcting for atmospheric effects, we have devised a set of algorithms and procedures for their use which significantly reduces the need for such data correction. In doing so, we do not suggest that atmospheric corrections should not be made, for indeed their use when accomplished with adequate precision should provide even further potential for information extraction. However, we do suggest that such correction procedures should be done being fully aware that they may be helpful, hurtful, or have little or no effect at all.