Curve Shape Matching, End-Member Selection and Mixture Modelling of AVIRIS and GER Data for Mapping Surface Mineralogy and Vegetation Communities

Steve Mackin
Adolf Ellissen Weg 16,
3400 Göttingen,
Germany.

Nick Drake
Dept. of Geography,
University of Reading,
United Kingdom.

Jeff Settle
NUTIS,
University of Reading,
United Kingdom.

Steve Briggs
BNSC/RSADU,
Monks Wood Station,
Abbots Ripton, Cambridgeshire,
United Kingdom.

Abstract. The availability of high spectral resolution data and laboratory collected spectra of surface materials allows the automatic identification of spectra collected by airborne instruments without the need of ground data. Using relatively simple matching techniques, the identity of the major components in each surface pixel can be obtained by comparison of the pixel spectra to each library spectra. A further step using an expert system can determine whether to accept or reject the basic shape match using several parameters such as flatness and spectral contrast (or separation) of materials. The output consists of a map of the identified materials with an indication of their purity. These end-members are input into a linear mixture model to determine the proportions of each identified material in each image pixel. An application of the method using both GER and AVIRIS data is given.

I. Introduction

Identification and mapping of surface mineralogy and vegetation community distribution are now possible by applying a shape analysis of high spectral resolution data using a minimum deviation algorithm. Using comparisons in shape of both overall curve and primary absorption features of each pixel to a library of reflectance spectra, the identity of each pixel can be determined.

Although an identity is established, the data is processed further using an expert system, to estimate the doubt associated with the matched identity and whether to accept or reject the identity based on the calculated doubt. Several parameters extracted from the data such as curve flatness and spectral contrast (or separation) are used at this stage. The output consists of relatively pure pixels of each identified component.

The purest pixels of each component identified by the expert system are input into a linear mixture model to estimate proportions of each component in each image pixel. Mixture modelling complements the expert system approach, as mixed pixels are not readily identified by matching to a library of pure materials.

II. Area of study

The methods were tested using both AVIRIS and GER data sets, collected over a site near Golconda in north central Nevada. The site covers the boundary between the Cambrian Preble Formation and the Ordovician Valmy Formation along the Getchell fault trend.

The geology of the Preble formation is primarily of interbedded folded quartzites and micaceous phyllites, with limited outcrops of blue grey, partly dolomitised limestone. Mineralisation is confined to N-S trending high angle vein systems, close to the Getchell fault. Some disseminated gold mineralisation is presently being mined in the area and seems to be related to silicification and the presence of abundant illite.

Vegetation cover is highly variable in this semi-arid terrain and consists primarily of scattered sage brush and grass, with limited development of annual shrubs in the stream beds.

The area is shown in Slide 15. The slide consists of four separate images. The image, top left, is a brightness image of the AVIRIS data covering the 2.0 mm to 2.5 mm wavelength region. The area of interest is shown outlined in red. There are two dark circular regions in the area of interest that correspond to a mineralised area currently being mined. North is to the right.

III. Processing and analysis of the data

The data processing and analysis consist of three stages. The first stage is shape matching of the reflectance spectrum of each data pixel to the library reflectance spectra. The data is converted to a set of percentage variations about the mean value, which allows matching of the data in a manner independent of the reflectance scale, which side steps the problems of shadowing and brightness variations due to grain size. This step is performed on both the overall curve shape and on smaller wavelength regions that define areas of characteristic absorption for each library material. This eliminates false matches, where the overall curve shape is well matched, but the error is concentrated on a small but characteristic absorption feature.

The second stage is a simple expert system, which evaluates parameters extracted from the data and spectral library to determine the doubt associated with the simple shape match. For example the accuracy of the match will be of the same order for the whole curve shape and for the characteristic absorptions; variations in accuracy between these two measures increase doubt. Another parameter is spectral contrast, or separation, for example, if a match score is moderate for the match of two materials in the library to the pixel spectrum and the separation of the two materials is quite low in the pure (library) state (e.g. Calcite and Siderite), the match may be accepted, with the proviso that the actual identity could be one or the other. While with the alternative, that given the moderate match score for the match of two materials in the library to the pixel spectrum and the separation of the two materials is quite high in the pure state (e.g. Calcite and Muscovite), the match may be rejected on the basis that it is most likely to be a mixture of dissimilar materials. The expert system basically ties together various parameters from the data and makes a decision to accept or reject the assigned pixel identity.

The third stage uses the output from the expert system stage to determine the end-members for the linear mixture model. The pixels with the highest match scores are the purest in the image; these are input as end-members into the mixture algorithm which estimates the proportion of each identified component in each image pixel. Estimation of the proportions **f** is based upon minimisation of the quadratic function:

$$(\mathbf{x} - \mathbf{Mf})^{T}\mathbf{N}^{-1}(\mathbf{x}-\mathbf{Mf})$$

where **x** = a pixel spectra
M = a matrix whose columns are end-member spectra
N = a variance-covariance matrix independent of **f**.

The above equation is subject to the following constraints,

$$0 < f_i < 1 \qquad i = 1,...c \quad \text{and}$$

$$f_1 + f_2 .... f_c = 1$$

Using this modified classical estimator error estimates can be calculated, for a two component mixture with the sum to one constraint imposed and with **N** given by **sI**, this quantity is given by :

$$\mathbf{E} = s/(m_1-m_2)^{T}(m_1-m_2)$$

and in this case the same for each component. This shows that the error on the estimate depends both on the level of noise in the scene and on the spectral separation of the components. When more than two components are present, **E** depends on the distance between the component and the closest mixture to it of the remaining end-member spectra.

IV. Geological application

Slide 15 contains four images. The top left image is the previously described AVIRIS brightness image. The image bottom left is a corresponding material identification map, again with north to the right. The map is based on shape analysis alone, prior to the expert system stage. The colour coding is as follows. Undifferentiated clays (muscovite, illite and kaolinite) are shown in red, dry vegetation in blue, chlorite in yellow and green vegetation in dark green. The clays are seen to dominate the area of interest (outlined in red in the brightness image) and are confined primarily to the phyllites of the Preble formation to the east of the Getchell fault (bottom quarter of the image). The only distinct feature to the west of the fault is an area rich in chlorite (yellow).

The image top right is another material identification map of the GER data over the outlined area of interest of the AVIRIS data. In this image, the clays are separated, red is illite, green is kaolinite and blue is muscovite. Note the rotation of the image relative to the AVIRIS data, here north is to the top. The bright red area (centre left of the top right image) is an area of illite, that corresponds to an active mine area, with disseminated gold mineralisation. While the blue-green area consists of muscovite and kaolinite with no economic mineralisation. The image below (bottom right) is the corresponding map **after** the expert system stage, although the number of accepted pixels is reduced, the general distribution of clay materials remains unchanged, except for an area of illite (red) in the shape image (top right image, centre right), that is rejected in the expert system output (bottom right image). This essentially confines the illite distribution to the area of economic mineralisation currently being exploited (red, bottom right image).

The pixels with the highest matches from the material identification maps are the purest in the image and were used for input as end-members in the mixture algorithm. Seven end-members were identified by the expert system in the GER data. These were input into the mixture model and their **E** values are shown in column A, below.

End-members	A (%)	B(%)
Muscovite	16.5	9.0
Illite	8.4	
Dry vegetation	26.0	18.1
Calcite	14.2	11.9
Kaolinite	36.1	
Gypsum	32.2	
Green vegetation	11.9	8.4

The high errors using seven end-members provide proportion maps dominated by noise. To reduce this problem, those with the highest **E** values are dropped. Apart from the muscovite, these are the materials with 2.2 mm absorption features. As a result, muscovite acts as a composite "clay" end-member, but the errors are now reasonable (column B). The clay mixture map produces high estimates in the regions with clearly identified pixels from the shape match and expert system stages.

V. Vegetation application

Three vegetation communities are found in this region. In the dry periods the densities and the amount of dry and green material in their canopies differ. The graminaceous community consists of dense dry vegetation, while the chaemophytic community has an open canopy, with lower proportions of these materials and a lower density. As the amount of dry and green vegetation in each pixel is known it can be used to classify their distribution. If the geological mixture maps are added together a map is derived that depicts the amount of bare ground. These vegetation communities can now be thought of in terms of their distribution in the ternary mixture space of bare ground, green vegetation and dry vegetation. Where each community is dominant, it occupies a specific region of this mixture space.

A simple decision rule classifier can then be implemented to classify each community. If an area has >90% green vegetation for example, the annual shrub community will be dominant. A similar threshold on the dry vegetation will outline where the graminaceous community is dominant. A rather arbitrary threshold of >90% has been used to outline areas of dominantly bare ground. The chaemophytic shrub community is highly variable in its density and also exhibits variations in the amount of dry and green vegetation in its canopy, it lies in the large region of this ternary mixture space not defined by the other decision rules.

Slide 16 shows the classified map resulting from this mixture space interpretation. The classes that are relatively pure such as the annual and graminaceous communities and the bare ground have solid colouring of green, yellow and red respectively. The region, where the chaemophytic shrub community dominates, exhibits large variations in density to show that a vegetation density map (derived by adding the dry and green vegetation maps together) is used as a backdrop. The community is displayed in various shades of cyan, where grey indicates a low density and bright cyan a high density.

VI. Conclusions

1. With current high spectral resolution data and a library of reflectance spectra, it is relatively easy to carry out a shape analysis of each pixel spectra and thus determine the identity of the spectrally dominant surface component.

2. An enhancement using an expert system stage that cross checks various data and library parameters enables a much clearer identification of materials present. With improvements it should allow the separation of components that would be misidentified using only the simple shape match.

3. Extraction of the purest material pixels provides an alternative method for determining end-members for a mixture model.

4. The mixture modelling extends the shape matching of pure materials to give a distribution of the identified components in each image pixel. However, very similar materials cannot be separated using mixture modelling due to their low contrast and high noise level of the imagery. They can be separated by the expert system as it uses a more robust matching technique, although the expert system cannot measure abundance or classify mixed pixels.

5. Linear mixture modelling, combined with field work to determine the vegetation phenology at the time of image acquisition and the various components in the plant canopy, is an effective method of deriving ecological parameters from imaging spectrometry data.