REMOTE SENSING

MODULE OF

REMOTE SENSING DATA ANALYSIS

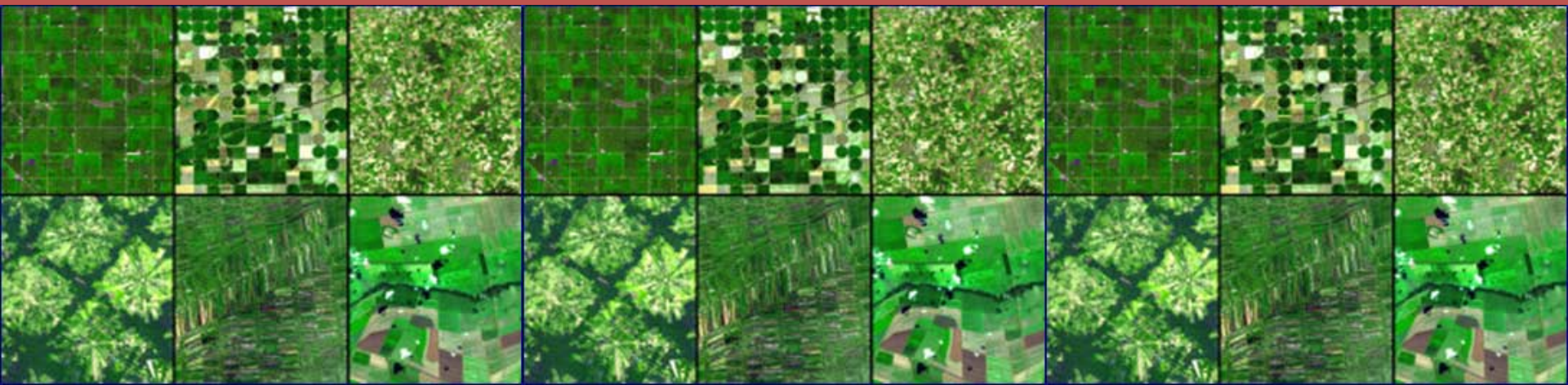
(6 CFU)

A.Y. 2013/14

MASTER OF SCIENCE IN COMMUNICATION TECHNOLOGIES AND MULTIMEDIA

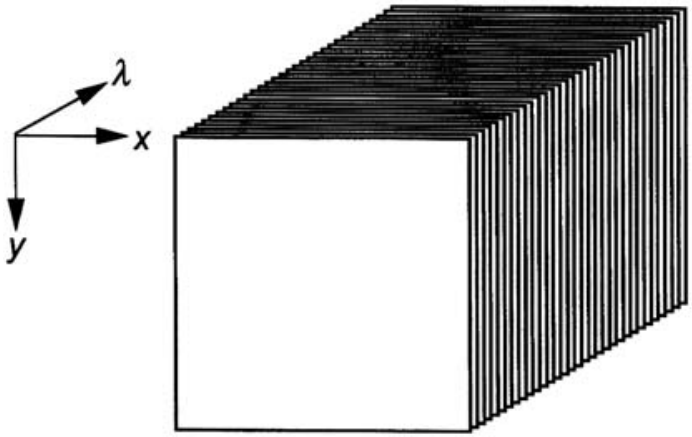
PROF. ALBERTO SIGNORONI

INTERPRETATION OF HYPERSPECTRAL IMAGE DATA



Data Characteristics

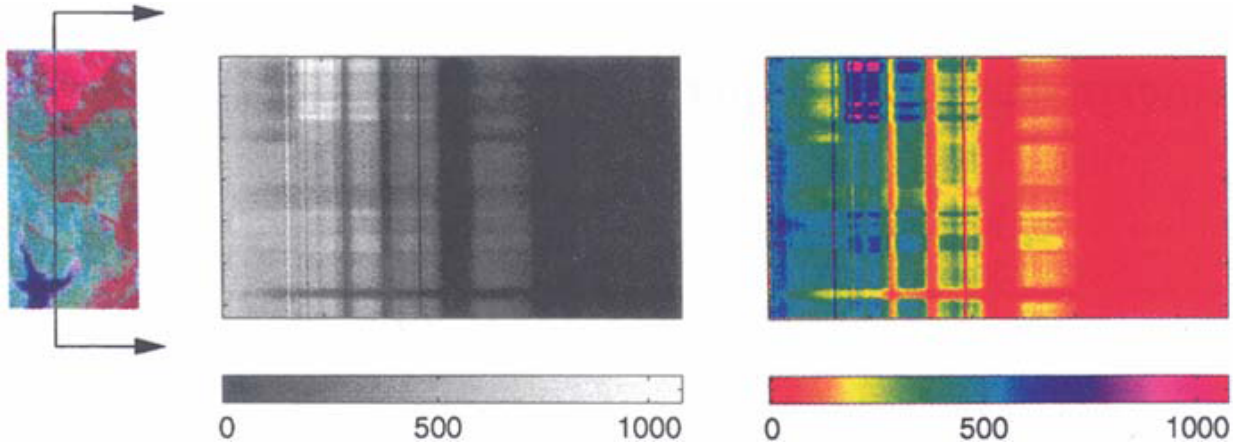
- The data produced by the **imaging spectrometers** is different from that of *multispectral* instruments owing to the enormous number of wavebands recorded – leading to the term ***hyperspectral***.
  - For a given geographical area imaged, the data produced can be viewed as a *cube*, as shown in Figure, having two dimension that represent spatial position and one that represents wavelength.



- When displaying multispectral data, such as that from Landsat, both spatial dimensions are generally used, with three of the spectral bands written to the red, green and blue colour elements of the display device.
  - Sometimes, careful band selection is required in this process to ensure the most informative display, while on other occasions multispectral transformations, such as principal components, are used to enhance the richness of the displayed data.

Data Characteristics

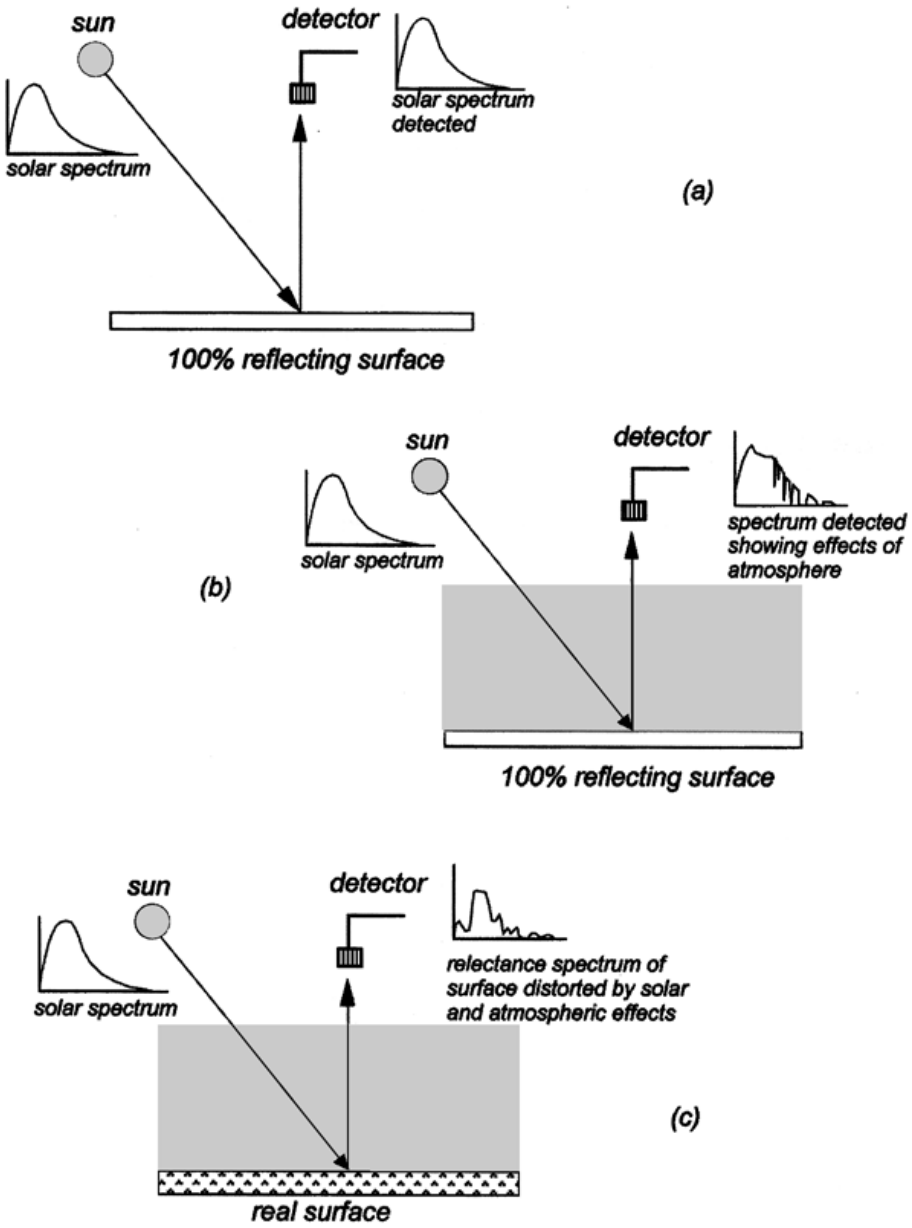
- With **hyperspectral data** there are both challenges and opportunities presented in creating data displays.
  - First, choosing the most appropriate three channels to use is not straightforward and, in any case, would invariably lead to substantial *loss* of the spectral benefits offered by this form of data gathering.
    - Nevertheless, unless spectral transformations are employed, a set of three bands comparable to those used with multispectral imagery are often adopted (near IR, red, green) for simple display of the data.
  - Secondly, because of the large number of bands available, a two dimensional display using one geographical dimension and the spectral dimension can be created as shown in Figure.
    - Such a representation allows changes in spectral profiles with position (either along track or across track) to be observed. Usually the greyscale is mapped to colour to enhance the interpretability of the displayed data.



Data Characteristics

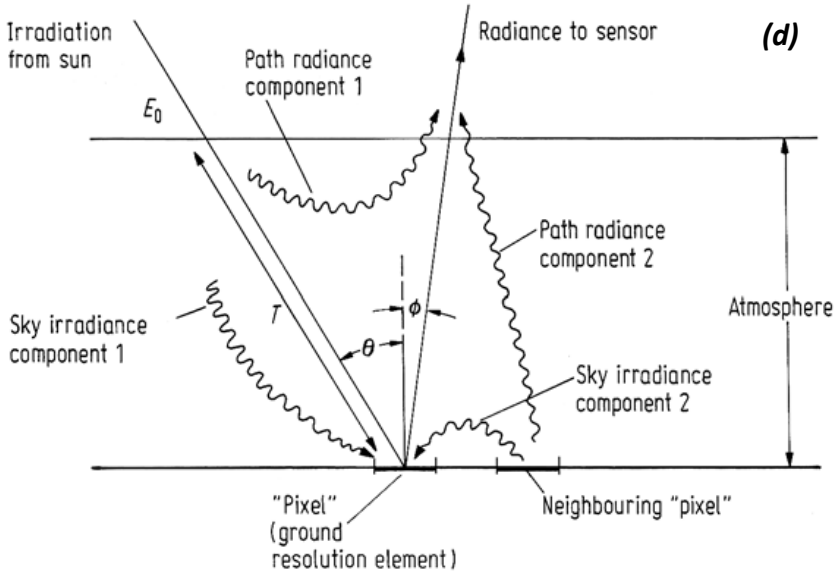
- To understand much of what is to follow it is useful to envisage how recorded hyperspectral data is affected by the presence of the **atmosphere** and the nature of the **solar spectrum**.

1. Imagine the region being imaged has a *uniform 100% spectral response* – in other words it will reflect all of the incident sunlight over all wavelengths, as depicted in **Figure (a)**; also assume that there is *no atmosphere* above the surface.
   - A detector capable of taking many spectral samples (say 200 or so) will then essentially record the solar spectrum as shown.
   - If the spectral resolution of the detector were sufficiently fine then the recorded solar spectrum would include the Fraunhofer absorption lines, resulting from the gases in the solar atmosphere (Slater, 1980).



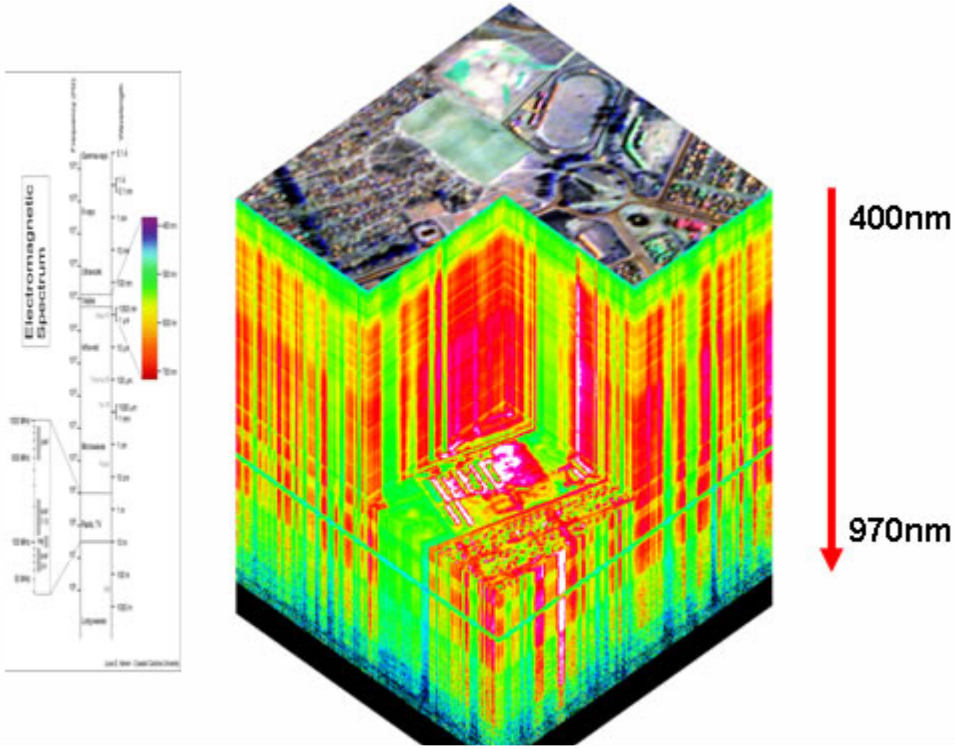
Data Characteristics

2. Now suppose there is a *normal terrestrial atmosphere* in the path between the sun, the surface and the detector.
   - The spectrum recorded will be modified by the extent to which the atmosphere selectively absorbs the radiation.
   - There are well known absorption features caused by the presence of oxygen and water vapour in the atmosphere and these appear in the recorded data as depicted in **Figure (b)**.
   - Also, the atmosphere scatters the solar radiation leading to the sky irradiance and path radiance terms of **Figure (d)**.



- So for a start, if we wished to determine the (uniform) spectrum of the ideally reflecting surface, the atmospheric absorption features need to be removed, as does the shape of the solar spectrum and the effect of atmospheric scattering.

3. **Figure (c)** suggests how the reflectance spectrum of a *real* surface might appear before compensation for solar and atmospheric effects. The spectrum recorded is a combination of the actual spectrum of the real surface, modulated by the effects of the solar curve and the atmosphere. We will addresses a range of techniques used for removing those effects.



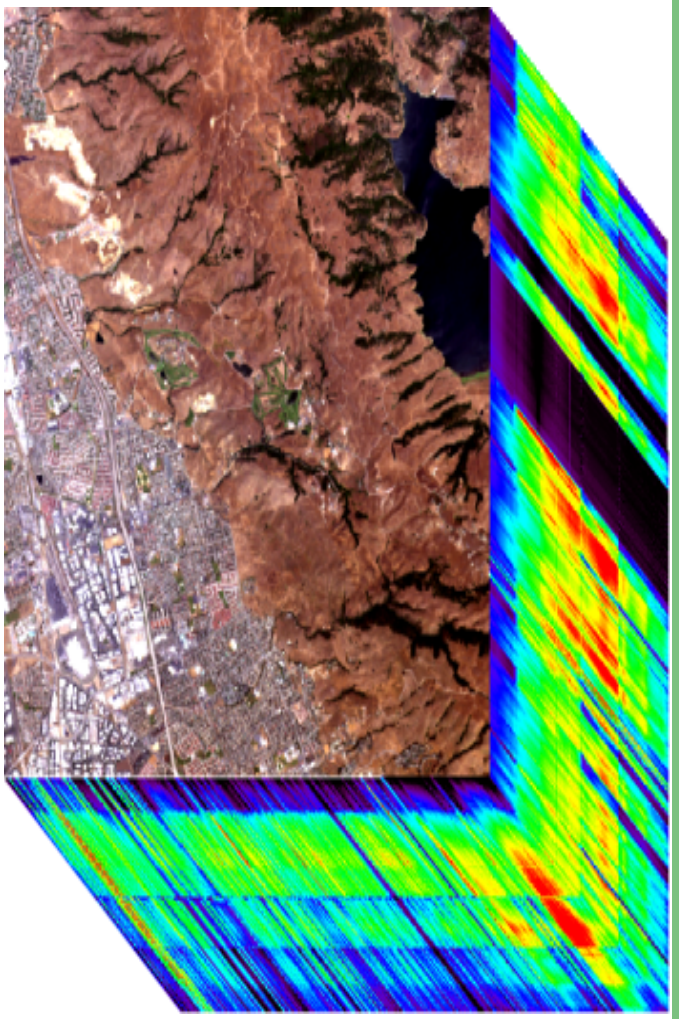
THE CHALLENGE TO INTERPRETATION

The Challenge to Interpretation

- We learned that there are essentially two classes of analytical technique used with multispectral data – *photointerpretation* and *machine analysis* (classification).
  - The former depends upon the use of image enhancement procedures for improving the visual interpretability of image data
  - whereas the latter is based usually on statistical or other forms of numerical algorithms for labelling individual pixels.
- When the data has hundreds of spectral bands, traditional image processing and data handling techniques face difficulties.
- On the other hand, enough information is readily available in the data to allow analysis based on a knowledge of **spectroscopic principles**, as we will see.
- It is important to understand the limitations placed on the more traditional analytical approaches since those methods still find application with hyperspectral data, not the least reason for which is the substantial investment in image processing software.
  - In the following ,the features which distinguish hyperspectral from multispectral data are highlighted as a precursor to a discussion on the methods of analysis that can be used with hyperspectral data, either modified or in original form.
  - These differences include **data volume**, **redundancy** and **dimensionality**.

Data Volume

- Although data volume strictly does not pose any major data processing challenges with *contemporary computing systems* it is nevertheless useful to examine the relative magnitudes of data for say Landsat Thematic Mapper multispectral imagery and *AVIRIS hyperspectral data*.
  - Clearly, the major differences to note between the two is the number of wavebands (7 versus 224) and the radiometric quantizations used (8 versus 10 bits per pixel per band).
  - Ignoring differenences in spatial resolution, the relative data volumes, per pixel, are 7 × 8 : 224 × 10 – i.e. 56 : 2240.



- Per pixel there are **40 times as many bits** therefore for AVIRIS as for TM data.
- Consequently, storage and transmission of hyperspectral data are issues for consideration,
  - suitable data compression techniques will be discussed.

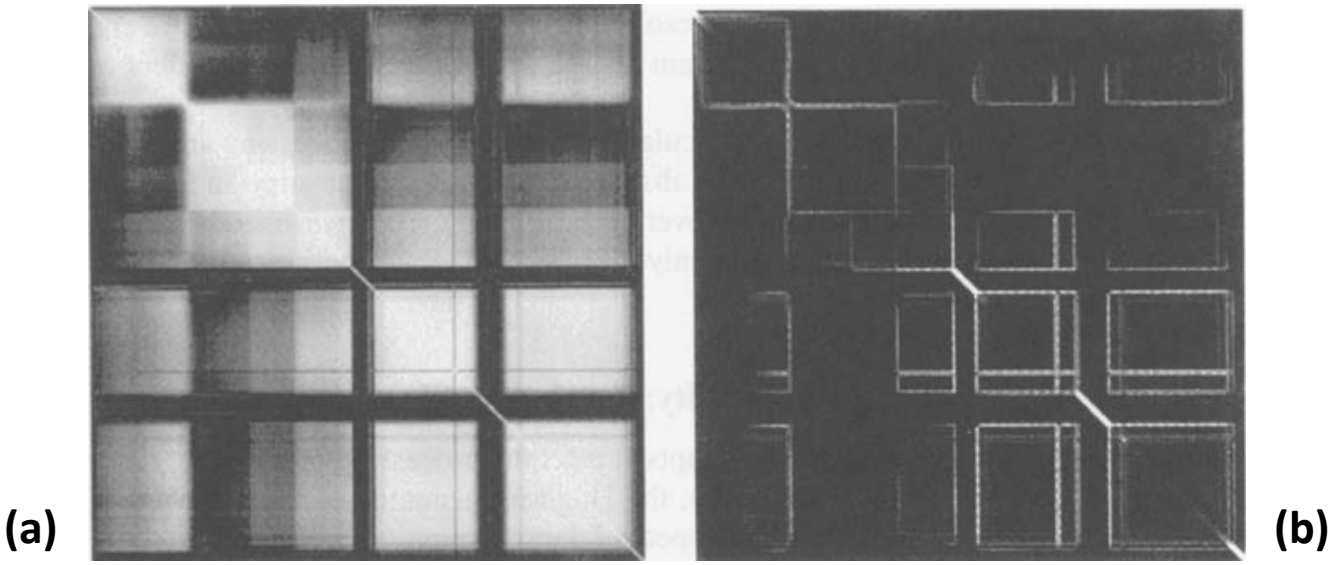
Redundancy

- With 40 times as much data per pixel one is led to question *whether 40 times as much information can be obtained* about the ground cover types being imaged.
  - Generally, of course, that is not the case – **much of the additional data does not add to the inherent information content** for a particular application **even though it often helps in discovering that information**.
  - In other words it contains **redundancies**.
- Much of the data we deal with in everyday life is highly redundant.
  - Take the English language as an example. If we remove certain letters from a word we can often still understand what word is intended. For example ***rmte sesng*** would be recognised by most people who read this book as *remote sensing* because there are sufficient redundant letters that losing some is not critical to understanding.
  - The same is true with remote sensing data, especially that recorded by hyperspectral sensors – there is often **substantial overlap of information content over the bands of data recorded for a given pixel**.
  - In such cases not all of the data is needed to characterise a pixel properly, although redundant data may be different for different applications.
- In remote sensing data redundancy can take two forms: **spatial** and **spectral**.
  - Exploiting spatial redundancy is behind the spatial context methods seen in statistical supervised classification.
  - Spectral redundancy means that the information content of one band can be fully or partly predicted from the other bands in the data (as we saw when talking about (multi)spectral transforms such as PCA).

Redundancy

- An interesting way to view spectral redundancy is to form the correlation matrix for an image (or portion of an image) of interest; the correlation matrix $|v_{ij}|$ can be derived from the covariance matrix using the already seen $\varrho_{ij} = v_{ij}/\sqrt{v_{ii}v_{jj}}$.
  - High correlations between band pairs indicate high degrees of redundancy.
- Because there are so many bands with hyperspectral data it is not practical to list all the correlations numerically, such as is done for multispectral data.
- Instead, it is better to **display** the inherent correlations (redundancies) pictorally as shown in Figure (a), next slide, where a grey scale is used to represent levels of correlation.
  - This representation is often used with hyperspectral data and is a *useful tool for identifying correlations among bands when applying traditional processing tools* as will be seen later.
- An interesting by-product of representing the correlation (or covariance) matrix in this form is that **image processing procedures can be applied** to it.
  - For example its block structure can be emphasised by using a simple edge detection filter to give the result shown in Figure (b).
- Means for removing inherent redundancy are often not readily apparent, although techniques such as the principal components transformation assist in the task since decorrelation followed by discarding low variance components amounts to redundancy-reduction.

Redundancy



- Figure: (a) The correlation matrix for 196 wavebands covering 400 nm to 2400 nm for the AVIRIS Jasper Ridge image (white represents correlations of 1 or – 1, while black indicates a correlation of 0). (b) The result of edge detecting the correlation matrix.
  - Overlapping bands result from the use of four individual spectrometers in the AVIRIS instrument; these and the significant water absorption bands and bands which have very small means (*< 2*) have been deleted from the original 224 bands, leaving 196 bands for image processing.

The Need for Calibration

- The high spectral resolution of hyperspectral datasets means that fine **atmospheric absorption features** will be detected and displayed as discussed before.
  - In order that they not be confused with absorption features of the ground cover type being imaged it is important to account for them and "remove" them from the data.
- Moreover, because the high spectral resolution suggests that recorded spectra can be interpreted scientifically it is important also to remove the **modulating effect of the solar spectrum**.

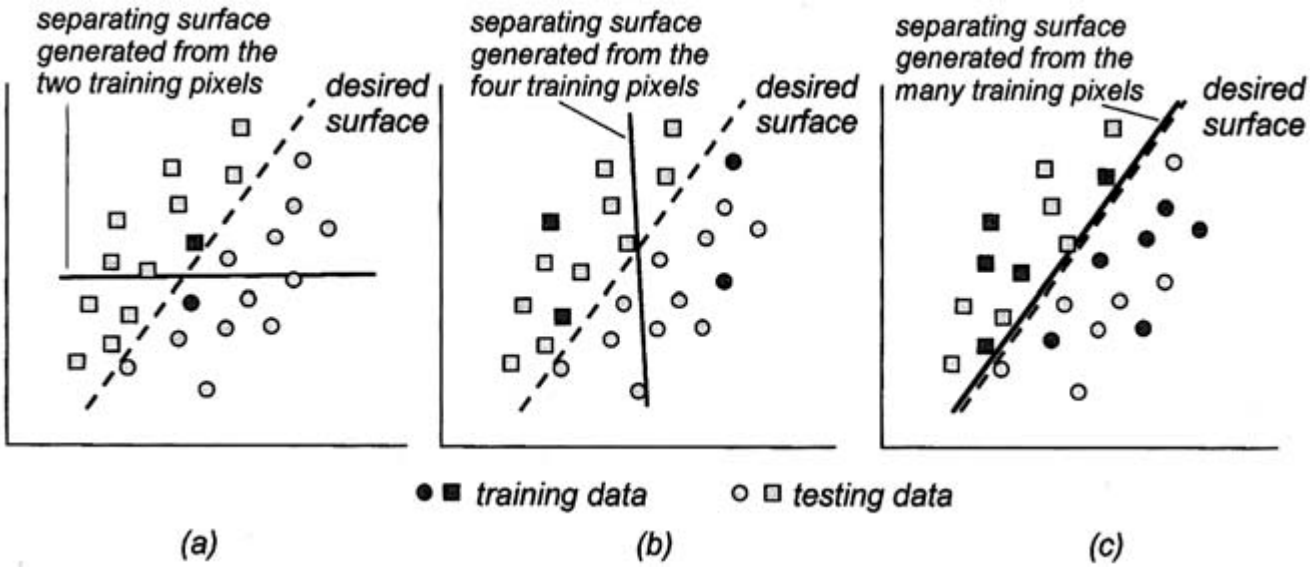
- Neither of those effects has been particularly important in the processing and analysis of multispectral data because of the absence of well defined absorption features and the use of average solar irradiance over each of the recorded wavebands.
  - With multispectral data only the effects of atmospheric scattering and transmittance are corrected.

The Problem of Dimensionality: The Hughes Phenomenon

- While *recognised since the earliest attempts at machine processing of remotely sensed image* data (Swain and Davis, 1978), the **Hughes phenomenon** *had not been of major concern until the advent of hyperspectral data*.
- Briefly, **a minimum ratio of the number of training pixels to number of spectral bands is needed to ensure reliable estimates of class statistics are obtained when training supervised classifiers**; as the dimensionality of the data set increases the minimum number of training pixels per class must be increased to preserve the accuracy of the statistical estimates.
  - Thus, adding more spectral bands, as in the case of AVIRIS, MODIS and Hyperion, is not helpful unless more training pixels per class are available.
- This turns out to be one of the major limitations in attempting to apply traditional image classification procedures to hyperspectral data.
- A simple example, based on determining a reliable linear separating surface, can be used to illustrate the problem.

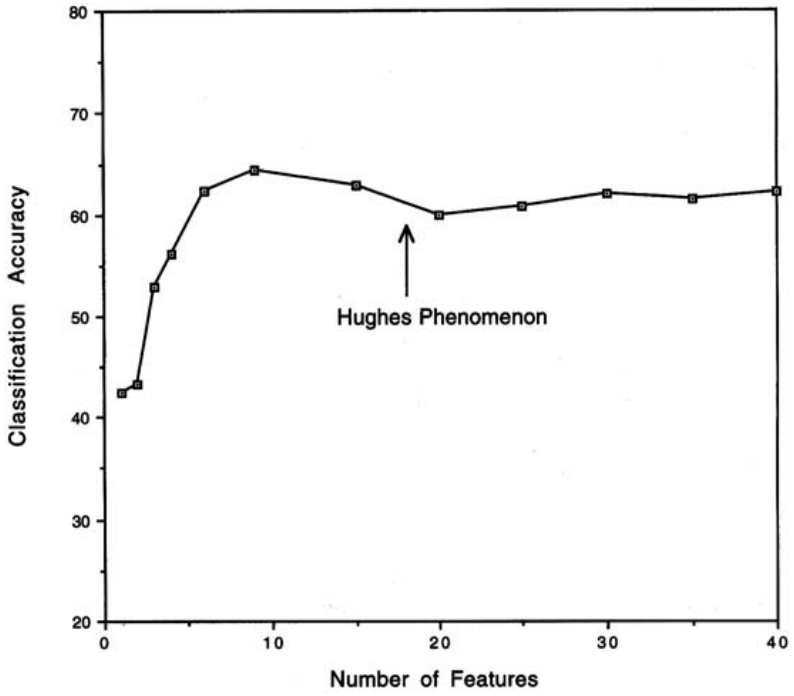
The Problem of Dimensionality: The Hughes Phenomenon

- The Figure *shows three different training sets of data for the same two dimensional (band) data set*.
  - The first (Figure (a)) has only one pixel per class. As seen, while a separating surface can be found it may not be accurate.
  - Having two training pixels per class as in the case of Figure (b) provides a better estimate of the separating surfaces, but it is not until we have many pixels per class, when compared to the number of channels in the data, that we will obtain good estimates of the parameters of the supervised classifier (Figure (c)).



The Problem of Dimensionality: The Hughes Phenomenon

- This is simply another way of looking at what we already said about the number of training pixels required for each class (it was suggested to be at least 10N).
- However, rather than increase the number of training pixels for a given number of bands, consider now the case of increasing the number of bands for a set number of training pixels; the same problem is observed as illustrated in Figure.
  - We note that the performance of the classifier is compromised by the poor estimates of the training statistics beyond about ten features.



DATA CALIBRATION TECHNIQUES

Detailed Radiometric Correction

- As we know, the upwelling radiance measured by a sensor results from incident solar energy scattered and reflected from the atmosphere and earth surface.
  - Detailed radiometric correction to obtain surface reflectance for hyperspectral data follows similar procedures as for the examples given when we dealt the subject.

- However, since hyperspectral data covers the whole spectral range from 0.4 to 2.4µm, including water absorption features, and has high spectral resolution, a **more systematic process is generally required**, consisting of **three possible steps**:

1. Compensation for the shape of the solar spectrum. The measured radiances are divided by solar irradiances above the atmosphere to obtain the *apparent reflectances* of the surface.
2. Compensation for atmospheric gaseous transmittances and molecular and aerosol scattering. Simulating these atmospheric effects allows the *apparent reflectances* to be converted to *scaled surface reflectances.*
3. Scaled surface reflectances are converted to *real surface reflectances* after consideration of any topographic effects.

  - If topographic data is not available, real reflectance is assumed to be identical to scaled reflectance under the assumption that the surfaces of interest are Lambertian (isotropic).
  - Procedures for solar curve and atmospheric modelling are *incorporated in a number of models* (Gao et al., 1993), including Lowtran 7 (Low Resolution Atmospheric Radiance and Transmittance), 5S Code (Simulation of the Satellite Signal in the Solar Spectrum) and Modtran 3 (The Moderate Resolution Atmospheric Radiance and Transmittance Model – see Anderson et al., 1995).

Data Normalisation

- When detailed radiometric correction is not feasible (for example, because the necessary ancillary information is unavailable) normalisation is an alternative which makes the corrected data independent of multiplicative noise, such as topographic and solar spectrum effects. This can be performed using *Log Residuals* (Green and Craig, 1985), based on the relationship between radiance (raw data) and reflectance:

$$x_{i,n} = T_i R_{i,n} I_n, \quad i = 1, \ldots K; n = 1, \ldots N$$

where $x_{i,n}$ is radiance for pixel i in waveband n. T_i is the topographic effect, which is assumed constant for all wavelengths. $R_{i,n}$ is the real reflectance for pixel i in waveband n. I_n is the (unknown) illumination factor, which is assumed independent of pixel. K and N are the total number of the pixels in the image and the total number of bands, respectively.

- There are two steps which remove the topographic and illumination effects respectively. $x_{i,n}$ can be made independent of T_i and I_n by dividing $x_{i,n}$ by its geometric mean over all bands and then its geometric mean over all pixels. The result is not identical to reflectance but is independent of the multiplicative illumination and topographic effects present in the raw data. The procedure is carried out logarithmically so that the geometric means are replaced by arithmetic means and the final result obtained for the normalised data is

$$\begin{aligned} \log z_{i,n} &= \log x_{i,n} - \log m_n - \log m_i \\ &= \log x_{i,n} - \frac{1}{N}\sum_{n=1}^{N} \log x_{i,n} - \frac{1}{K}\sum_{i=1}^{K} \log x_{i,n} \end{aligned}$$

Approximate Radiometric Correction

- As with multispectral data, approximate correction is acceptable for some applications.
- One approach is the *Empirical Line procedure* (Roberts et al., 1985).
  1. Two spectrally uniform targets in the site of interest, one dark and one bright, are selected;
  2. their actual reflectances are then determined by field or laboratory measurements.
  3. The radiance spectra for each target are extracted from the image and then mapped to the actual reflectances using linear regression techniques.
  4. The gain and offset so derived for each band are then applied to all pixels in the image to calculate their reflectances.
- While the computational load is manageable with this method, field or laboratory data may not be available.
- The *Flat Field method* (Roberts et al., 1986), an approximate correction technique that relies purely on the image data itself, is then an alternative.
  - This depends on locating a large, spectrally uniform area in an image (such as sand or clouds) and finding its average radiance spectrum.
  - It is assumed that the shape and the absorption features presented in this spectrum are caused by solar and atmospheric effects.
  - The reflectance of each pixel is then obtained by dividing the average radiance spectrum into the image spectrum of the pixel.

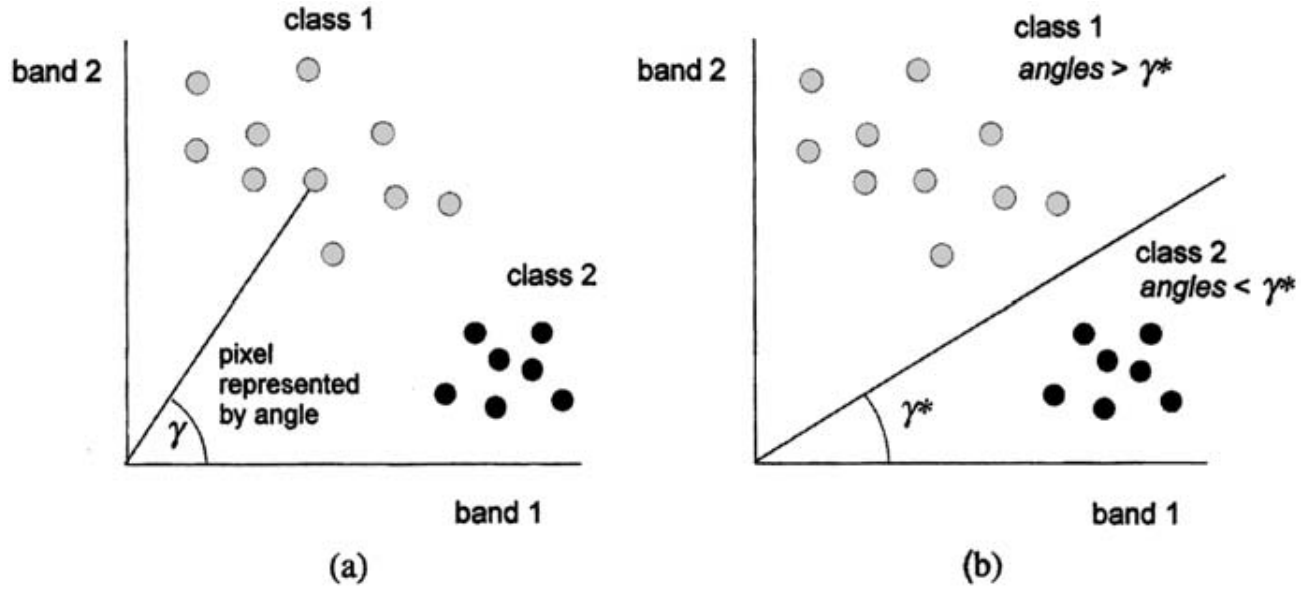
INTERPRETATION USING SPECTRAL INFORMATION

Spectral Angle Mapping

- As will be seen later on, pixel labelling techniques in hyperspectral data analysis, based on standard classification procedures, can often fail because of the difficulty in obtaining reliable class definitions with data of such high dimensionality.
- One means for coping with the problem is to reduce the dimensionality by some means.
- A candidate approach is to ignore the magnitudes of the pixel vectors in hyperspectral space and attempt classification instead **using just their angular orientations** as their sole describing characteristic.
  - In *N* dimensional multi-(hyper-)spectral space a pixel vector ***x*** has both magnitude (length) and an angle measured with respect to the axes that define the coordinate system of the space.
  - In the **spectral angle mapper (SAM)** technique for identifying pixel spectra only the angular information is used.
- Clearly the SAM technique will fail if the vector magnitude is important in providing discriminating information, which it will in many instances.
  - However, if the pixel spectra from the different classes are well distributed in the space there is a high likelihood that angular information alone will provide good separation.
  - The technique functions well in the face of scaling noise. Details for implementing SAM will be found in Kruse et al. (1993).

Spectral Angle Mapping

- Figure (a) shows a two dimensional (i.e. two band) example where spectra are characterised entirely by their angles from the horizontal axis.
  - The spectra can be distinguished from each other provided the angles are sufficiently different.
- Using this concept, **angular decision boundaries** can be set up (from library information or training data) that segment the space as shown in Figure (b).
  - Spectra are then labelled according to the sector within which they fall.



Using Expert Spectral Knowledge and Library Searching

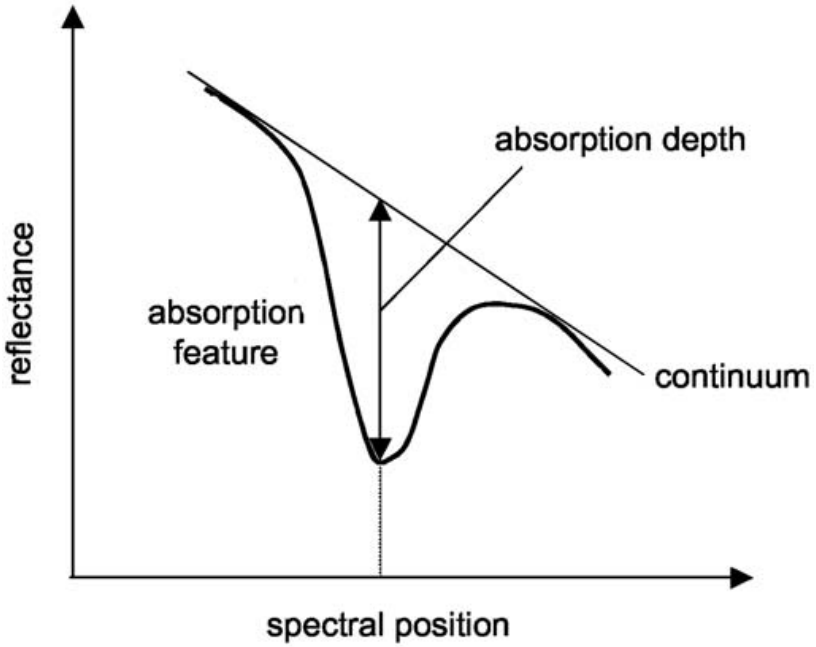
- Having a *well-defined spectrum* means that a ***scientific approach*** *to interpretation* can in principle be carried out, much as a sample is identified using *spectroscopy in the laboratory through a knowledge of spectral features*.

- Absorption features (seen as localised dips) are often observed in the reflectance spectra of specific minerals provided sufficient spectral resolution is available.
  - It is those **absorption features** that provide the information needed for identification.
  - They are sometimes referred to therefore as "**diagnostically significant features**".

- Characterisation and thus automatic detection of such absorption features, when they occur, is of particular interest in hyperspectral image recognition.
  - Absorption features can be characterised by their **locations** (bands), **relative depths and widths** (full width at half the maximum depth), and used in pixel identification.
  - To make that possible it is important to *separate the absorption features from the background continuum of the spectrum that results from light transmission and scattering*, as against the absorption features themselves that are due to photon interaction with the atomic structure of the chemicals present in the material being observed.

Using Expert Spectral Knowledge and Library Searching

- The importance of continuum removal can be seen in Figure.
- Often the background will not be “horizontal”, so definition of the depth of the feature can then be ambiguous.
  - If the continuum in the vicinity of the feature is defined by a line of best fit between those portions of the spectrum either side of the feature then a reasonably consistent measure of band depth can be established.



Using Expert Spectral Knowledge and Library Searching

- Usually, a complete spectrum is *divided into several spectral regions* (often under the guidance of an expert) and *absorption features are detected in each of the regions*.
- An unknown pixel is then labelled as belonging to a given class if the properties of its diagnostically significant absorption features match those of the spectrum for that class held in a **spectral feature library**.
- A complication that can arise with library searching in general, and with seeking to match absorption features in particular, is that **mixtures are often encountered**, and **some materials have very similar spectral features**.
- An excellent treatment of the complexities that arise, and how they can be handled, is given in Clark et al (2003).

Library Searching by Spectral Coding

- Because the pixel spectrum is so well specified and can be corrected for atmospheric and solar distortions, spectral comparison is possible – either with previously recorded data or with laboratory spectra – for pixel identification.
- The reference spectra are usually stored in a **spectral library**.
- It is clear that the searching and matching processes must be efficient in such a procedure. Full spectral matching using original radiometric data is not practical.
- However, *given the degree of redundancy* spectrally and radiometrically that one would anticipate with the data recorded by an imaging spectrometer, *coding techniques can be employed to represent a pixel spectrum in a simple and effective manner* so that *fast library searching* and matching can be achieved.

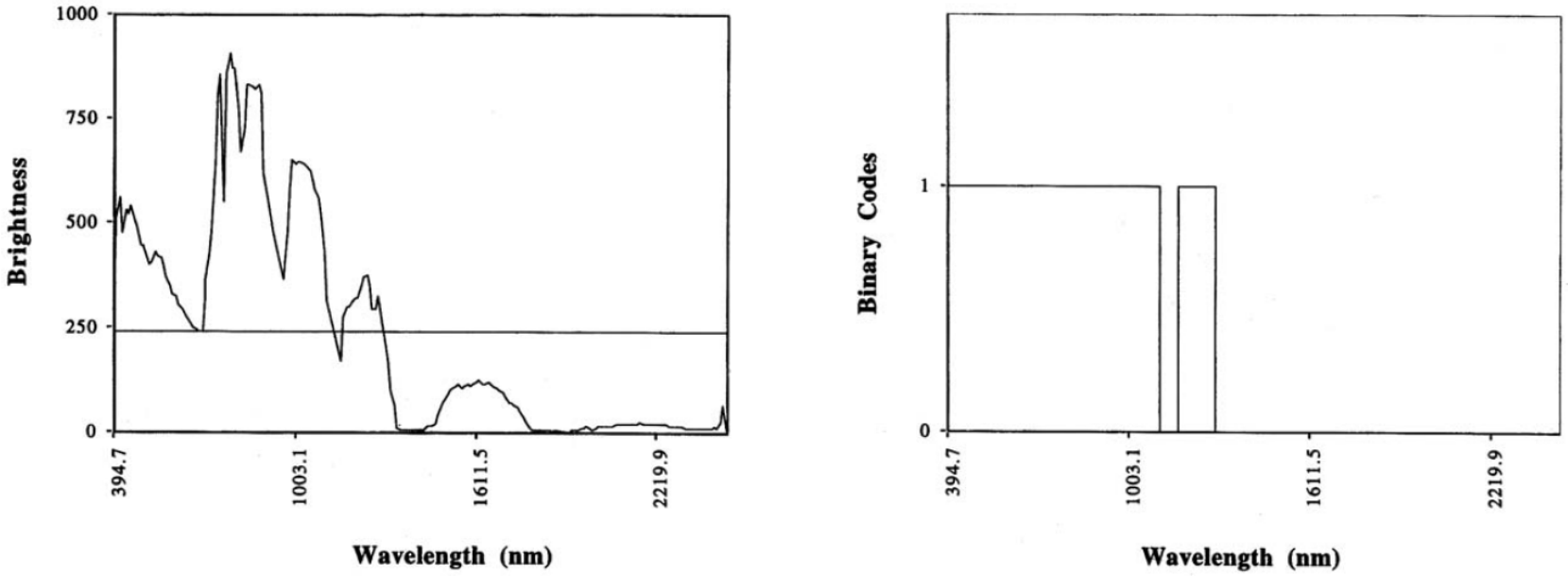
Library Searching by Spectral Coding: *Binary Spectral Codes*

- A simple **binary code** for a reflectance spectrum can be formed according to

$$h(n) = 0 \quad \text{if } x(n) \le T$$
$$1 \quad \text{otherwise} \qquad n = 1, \ldots N$$

  where $x(n)$ is the brightness value of a pixel in the n^{th} band, T is a user specified threshold for forming the binary code, and $h(n)$ is the resulting binary code symbol for the pixel in the n^{th} spectral band.
  - Usually T is chosen as the average brightness value of the spectrum.
  - Figure (next slide) demonstrates a typical spectrum encoded in this manner.
- Instead of using the average brightness of the complete spectrum as a threshold, the *local average* over the adjacent channels could be employed.
- *Such a simple binary code will not always provide reasonable separability* between the spectra in a library, *nor will it guarantee that a measured spectrum will match* with either only one or a small number of library spectra.
- Consequently, more sophisticated codes may need to be adopted. For example, more than one threshold could be used.

Library Searching by Spectral Coding: *Binary Spectral Codes*



- With *three thresholds* a two binary digit code for the brightness of a pixel will be created:

  $h(n) =$ 00 if $x(n) \leq T\,1$

  01 if $T\,1 < x(n) \leq T\,2$

  11 if $T\,2 < x(n) \leq T\,3$

  10 if $T\,3 < x(n)$.

- The mean brightness over the spectrum can be one threshold; the other two are chosen above and below this value.
- Note the use of a Gray code so that there is only one binary digit difference between levels.

Library Searching by Spectral Coding: *Binary Spectral Codes*

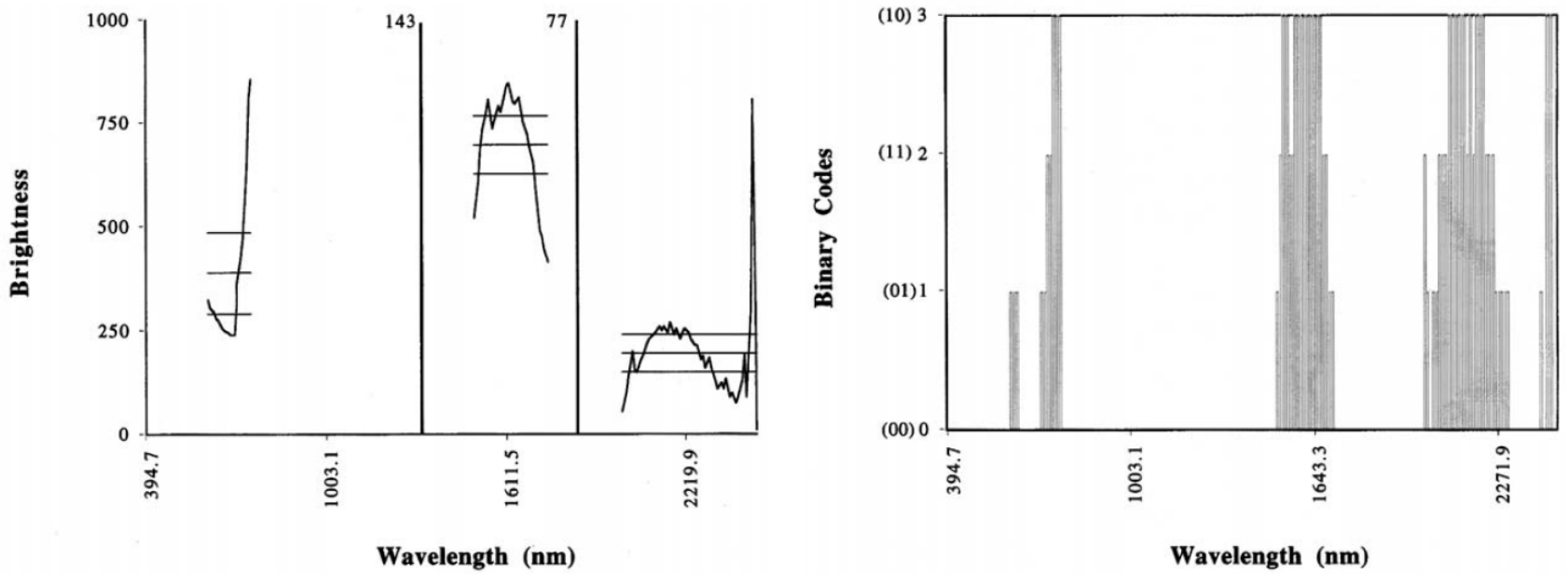
- Spectral slope can also be used as part of a code.
  - One binary representation of the local slope, *s(n),* at each waveband is:

$$s(n) = 0 \quad \text{if } (x(n+1) - x(n-1)) \leq 0$$
$$1 \quad \text{otherwise} \qquad n = 1, \ldots N$$

- Another variation is to develop separate codewords for different regions of the spectrum.
  - The resulting codes accommodate spectral coarse structure better.
  - Uniformly spaced regions could be used or perhaps those regions of the spectrum suspected as being most significant in differentiating cover types could be adopted.
- The latter is based upon the knowledge that in different wavelength ranges the reflectance spectrum is dominated by different physical characteristics of the surface being imaged.

Library Searching by Spectral Coding: *Binary Spectral Codes*

- Figure shows an example of coding on selected bands with 3 thresholds.



Library Searching by Spectral Coding: *Matching Algorithms*

- Comparison of binary coded spectra can be made by measuring the Hamming distance between them, defined as

$$D_H(h_i, h_j) = \sum_{l=1}^{L} (h_i(l) \oplus h_j(l))$$

where h_i and h_j are two spectral codewords of length (i.e. number of bits) L.

  - For simple thresholding, $L = N$, the number of bands. $L = 2N$ if slope coding is also used or three thresholds are employed.
  - $\oplus$ denotes the exclusive OR operator.
  - It is applied on a bit-by-bit basis for a pair of binary codewords and records a difference as '1'and no difference as '0'.
  - For example, the exclusive OR of two spectral codewords 01110011 and 00101011 becomes 01011000.
- Hamming distance is then calculated by summing the number of times the binary digits are different.
  - In this example, the Hamming distance is 3.
  - If the distance is within a user-specified threshold, the two pixels are identified as belonging to the same class.
  - When one, say $h_i(n)$, is a class signature code, the comparison leads to labelling for pixel j .

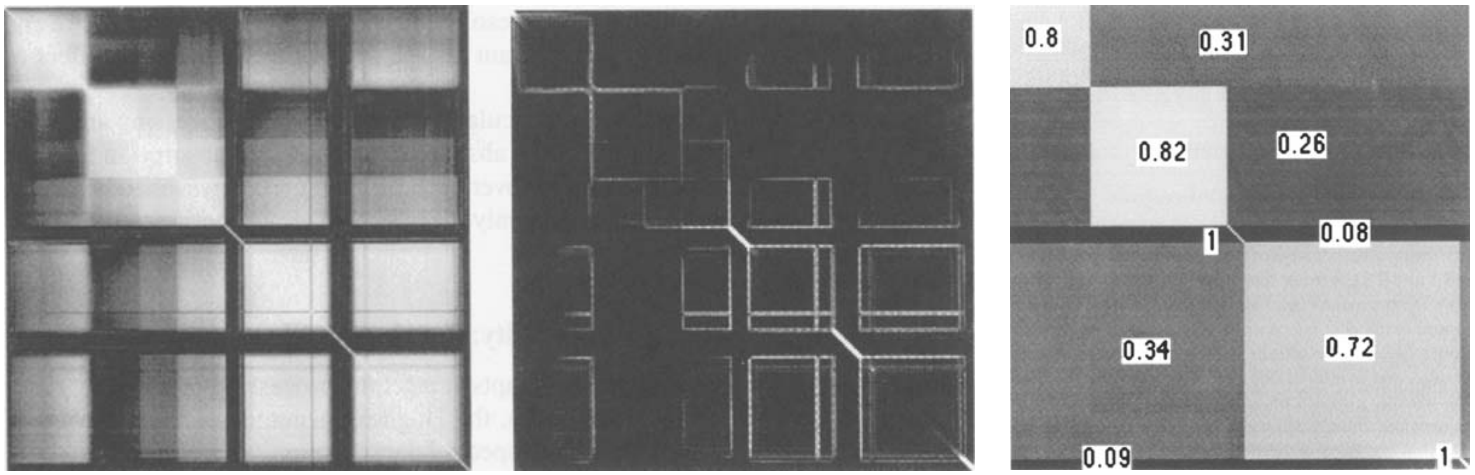
HYPERSPECTRAL INTERPRETATION BY STATISTICAL METHODS

Limitations of Traditional Thematic Mapping Procedures

- **Traditional supervised and unsupervised classification techniques will require very long processing times for hyperspectral** data because of the dependence on the number of wavebands.
- A more serious problem, however, is the need to estimate class signatures – i.e. the mean vector and covariance matrix – when using algorithms, such as maximum likelihood, based on second order statistics.
  - The difficulty lies in the **small number of available training pixels** per class compared with the number of wavebands used, and is related directly to the **Hughes phenomenon**.
  - If too few training samples are used then the class model may be very accurate for the training data and *classification accuracy on **training data** can be very **high***.
  - However, *classification accuracy on **testing data** will be **poor***.
- In this case, the classifier is **overtrained** (or **overfitted**) and the statistics estimated are unreliable.
  - To avoid the problem of unreliable class statistics and thus poor classifier performance the number of training pixels per class should be at least ten times the dimensionality of the data, with desirably 100 times (as already discussed for supervised classification).
- In the following we treat a technique developed for dealing with the small training set problem.

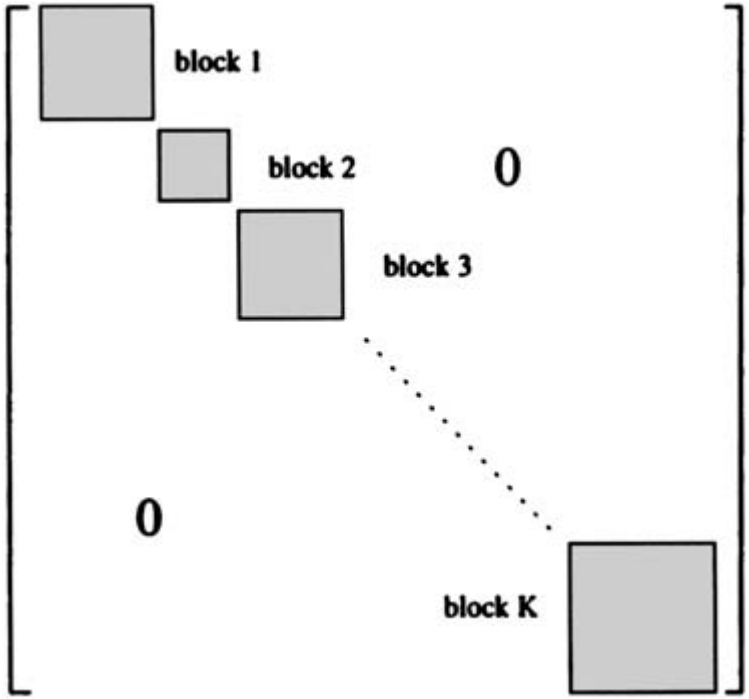
Block-based Maximum Likelihood Classification

- In general, correlations between neighbouring bands in hyperspectral data sets are higher than for bands further apart and highly correlated bands appear in groups.
  - As a result, the correlation matrix is roughly block diagonal in form as already shown (see the leftmost figure), in which a greyscale is used to represent the degree of correlation.
- The rightmost Figure shows the data of the leftmost one but, for purposes of illustration, with the *correlations averaged within identifiable blocks* demonstrating the strongly block diagonal form of the correlation and thus the covariance matrix.
  - Those blocks can be identified visually or with the assistance of edge detection on the correlation matrix as shown in the central Figure.



Block-based Maximum Likelihood Classification

- Now assume that the low off-diagonal correlations are zero.
- The matrix is then fully block diagonal as depicted in general terms in Figure.
- By assuming that the subgroups of bands within each block are independent of those in other subgroups, maximum likelihood classification can then be applied to each subgroup **independently**.



Block-based Maximum Likelihood Classification

- Noting that the block diagonal form of the correlation matrix leads to a covariance matrix of the same structure, the discriminant function becomes the sum of the logarithmic discriminant values of the individual groups of wavebands (blocks):

$$g_i(\boldsymbol{x}) = -\sum_{k=1}^{K}\{\ln|\Sigma_{ik}| + (\boldsymbol{x}_k - \boldsymbol{m}_{ik})^t \Sigma_{ik}^{-1}(\boldsymbol{x}_k - \boldsymbol{m}_{ik})\}$$

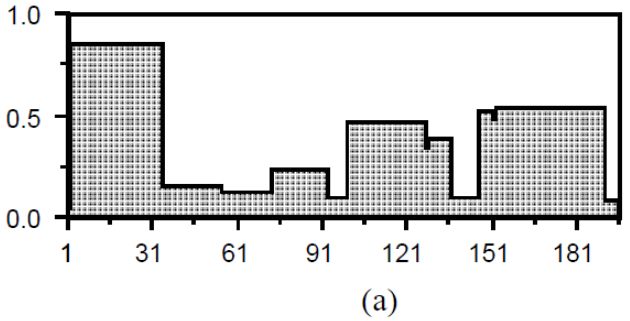
$$i = 1, \ldots M; \quad k = 1, \ldots K \qquad (13.1)$$

In (13.1) the dimensions of $\boldsymbol{x}$, $\boldsymbol{m}_i$, and Σ_i are reduced to $n_k (n_k < N)$, the size of the k^{th} subgroup of bands, so that advantage can be taken of the corresponding quadratic reduction in classification time (see Sect. 8.5). Also, the number of training pixels required per class for reliable statistics, determined by the size of *the biggest* subgroup, is much smaller than when all bands are used.

The sizes of subgroups to use are generally guided by observation of the boundaries of the high correlation blocks along the principal diagonal of a correlation matrix, which will be different for different images.

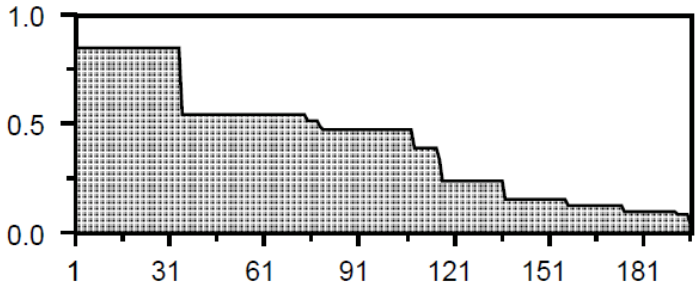
Block-based Maximum Likelihood Classification

- If training data is limited, some relatively high correlations may have to be ignored.
  - However, this approach will still be better than, say, minimum distance classification (often used when training pixels are limited) since at least some correlations are taken into account.
- With some data sets, highly correlated blocks of bands will occur away from the diagonal.
  - They can be moved onto the diagonal by reordering the bands before the correlation matrix is computed.
  - Such an operation makes no difference to the information contained in the matrix or to subsequent image analysis operations.
  - However, it does mean that a reconstructed pixel spectrum will have some bands out of order in the sequence of wavelengths.
- A simple and effective means for re-ordering the bands is to consider the first set of rows in the image of the already seen correlation matrix corresponding to the first highly correlated (diagonal) block of bands.
  - That block covers bands 2–35 in this example.
  - Moving across those 34 rows as a single group, blocks of similar correlation are identifiable (they are correlations of the respective bands with bands 2–35).
  - If we average the correlations in those blocks, the graph of Figure (a) is produced.

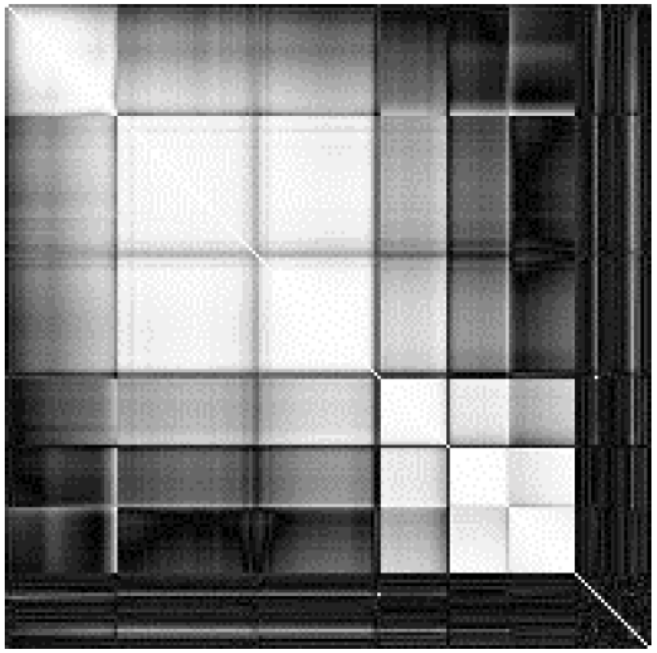


Block-based Maximum Likelihood Classification

- If we then re-arrange the bands as shown in Figure (b), by moving the more highly correlated blocks of bands to the left and the less correlated blocks to the right
  - then that has the effect of re-arranging the blocks of bands in the correlation matrix
  - such that the lower correlated blocks are shifted towards the off-diagonal corners
  - and the more highly correlated blocks are moved to the diagonal as shown in Figure (c).



(b)


(c)

Block-based Maximum Likelihood Classification

- For interest, the Table shows how the band blocks for this example have been re-ordered.

Group Number	Reordered Band Number	Original Band Number
1	1–34	2–35
2	35–38	148–151
3	39–77	153–191
4	78–105	101–128
5	106–113	130–137
6	114	129
7	115	152
8	116–135	74–93
9	136–153	56–73
10	154–173	36–55
11	174	1
12	175–181	94–100
13	182–191	138–147
14	192–196	192–196

COMPRESSION OF HYPERSPECTRAL DATA

Compression of Hyperspectral Data

- Owing to the large data volumes involved, storage and transmission of data from imaging spectrometers benefit from the application of procedures that will reduce data volume without substantially affecting the information content.
- Those procedures are generally in the form of codes that represent the spectra in reduced form.
- The binary codes seen in this lesson are typical of codes that could be used, although with such reductions in the spectra significant information loss (allowing the spectra to be used over a large number of applications) could be expected.
- More sophisticated codes minimise information loss while compressing the data.
- The principal components transformation is an example.
  - The higher order components with low variance can be discarded without significant information loss and yet with a reduction in storage requirement in proportion to the number of bands discarded.
  - Also, the original spectral or image data can be reconstructed from the reduced representation (using an inverse principal components transform) although with loss of information.
- Sometimes the information loss is referred to as **distortion** since the reconstructed data will differ, depending on the level of loss of detail, from the original.

Compression of Hyperspectral Data

- An alternative transformation widely used in the television and video industry is the Discrete Cosine Transform (Rao andYip, 1990).
  - The DCT is similar in principle to the Discrete Fourier Transform but with cosine expansion functions instead of complex exponentials.
- If the user can tolerate substantial amounts of distortion then significant compression of remote sensing imagery is possible; figures as high as 100 times reduction in volume have been reported, but one is then led to question the integrity of the compressed data.
- Generally, those compression schemes that allow the original image to be reconstructed without error (so-called lossless compression algorithms) will give compression ratios of about 2 to 3.
- A compression scheme well matched to the needs of remote sensing is referred to as **vector quantisation**, based upon the use of a so-called code book.
  - That book contains a number of representative pixel vectors (for example class means) that could be obtained from training data, or possibly could even be prototypical reference spectra.
  - Each code book vector is given a label (such as a number or even a class symbol).

Compression of Hyperspectral Data by Vector Quantization

- Now imagine an image has to be transmitted over a telecommunications channel.
  - If the spectrum matches exactly one of the stored spectra then only the label need be transmitted.
  - The receiver also has a copy of the code book and can retrieve the spectrum in question through matching the label.
  - If the spectrum does not match a code book entry exactly then transmitting the label of the nearest match will incur an error.
- Whether that error is acceptable, or whether a correction needs to be transmitted with the label of closest match, will depend on the application.
- The efficacy of the scheme depends upon how well the code book represents the range of pixel vectors in the image.
  - A good code book will give rise to small differences (errors) between code book entries and pixel vectors to be transmitted.
  - Such small differences can be encoded using a small number of bits (substantially smaller than the number of bits in the original pixel vector), so that good data compression is achieved.
- A simple illustration is given in the following Table in which 10 SPOT multispectral vectors are to be sent over a channel.

Compression of Hyperspectral Data

Cluster 1 (e.g. vegetation) Cluster 2 (e.g. soil)

Original pixel vectors:

1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
50	55	60	58	48	48	49	55	53	51
10	11	12	9	9	70	69	73	71	68
150	152	148	154	160	171	163	165	167	160

Code book entries (cluster means):

54
10
153

–

51
70
165

Differences between pixel vectors and nearest code book entry:

1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
−4	1	6	4	−6	−3	−2	4	−2	0
0	1	2	−1	−1	0	−1	3	−1	−2
−3	−1	−5	1	7	6	−2	0	2	−5

Compression of Hyperspectral Data

- Ordinarily, with each band represented by 8 bits, the ten pixels require 10 × 3 × 8 = 240 bits to be transmitted.
- However, **recognising** there are two clusters in the data and using the cluster means as code book vectors, it is possible to represent each of the pixels to be transmitted by their difference (error) from the nearest mean.
  - There are 8 distinct differences (between 0 and 7); they can be distinguished from each other (including sign) by allowing a 4 bit word for coding them.
  - Thus the number of bits then to be transmitted is 10×3×4 = 120 bits, plus one bit per pixel to indicate the code book vector label (one bit is enough to represent just two labels – i.e. 0 or 1) and 2×3×8 = 48 bits to transmit the code book beforehand.
  - Thus the vector quantised scheme requires 120 + 10 + 48 = 178 bits for the 10 pixels.
- The "compression ratio" is 240/178 = 1.35 with the ability to reconstruct the original pixel vectors without loss (distortion).
- Further compression of the data is possible by using a more efficient coding process on the errors.
  - Rather than simply allocating (in this example) 3 bits per difference (based on the observation that there are 8 different errors to transmit) shorter code words (in terms of numbers of bits) can be ascribed to the most commonly encountered errors (in this example 1 and 2).
- Details of this refinement, vector quantisation in general and the overall issue of compression in remotely sensed data can be found in Ryan and Arnold (1997a,b).

SPECTRAL UNMIXING: END MEMBER ANALYSIS

Spectral Unmixing: End Member Analysis

- A challenge that has faced interpreters throughout the history of remote sensing has been the need to handle mixed pixels – i.e. those pixels that represent a mixture of cover types or information classes. Several early studies attempted to resolve the proportions of pure cover types within mixed pixels by assuming that the measured radiance is a linear combination of the radiances of the "pure" constituents in each of the imaging wavebands used.
- With low resolution (multispectral) data the approach generally did not meet with a great deal of success because most cover types are not well differentiated in the small number of wavebands used. However, with hyperspectral data, the prospect of uniquely characterising a vast number of earth cover types, and thus differentiating them from each other spectroscopically, suggests that the mixing approach should be re-visited as a means for establishing mixture proportions of pure cover types in pixels. This has particular relevance in minerals mapping where abundance maps for minerals of interest can then be produced based upon the proportions determined for all pixels in a given image.
- The process can be developed mathematically in the following manner. Assume there are M pure cover types in the image of interest. In the nomenclature of mixing models these are referred to as *endmembers*. Let the proportions of the various endmembers in a pixel be represented by $f_m, m = 1, \ldots M$. These are the unknowns in the process which we wish to find, based on observation of the hyperspectral reflectance of the pixel.

Spectral Unmixing: End Member Analysis

- Let $R_n, n = 1, \ldots N$ be the observed reflectance of the pixel in the n^{th} spectral band of the sensor and $a_{n,m}$ be the spectral reflectance in the n^{th} band of the m^{th} endmember. Then we assume

$$R_n = \sum_{m=1}^{M} f_m a_{n,m} + \xi_n \quad n = 1, \ldots N$$

where ξ_n is an error in band n. The equation says that the observed reflectance in each band is the linear sum of the reflectances of the endmembers; the extent to which that does not work exactly in a given situation is encapsulated in the error term.

- An assumption that allows us to use linear mixing in this form is that the incident energy is scattered only once to the sensor from the landscape and does not undergo multiple scatterings among, for example, foliage components.

The above mixing equation can be expressed in matrix form as

$$\boldsymbol{R} = A\boldsymbol{f} + \boldsymbol{\xi}$$

where $\boldsymbol{f}$ is a column vector of size M, $\boldsymbol{R}$ and $\boldsymbol{\xi}$ are column vectors of size N and A is an $N \times M$ matrix of endmember spectral signatures (by column).

- Spectral unmixing, as the process is called, involves finding a set of endmember proportions that will minimise the error vector $\boldsymbol{\xi}$. On the assumption that the correct set of endmembers has been chosen the problem then becomes one of solving the simpler equation

Spectral Unmixing: End Member Analysis

- $$\boldsymbol{R} = A\boldsymbol{f}$$

  Normally there are more equations than unknowns so that simple inversion of the last equation to find the vector of mixing proportions is not possible. Instead, a least squares solution is found by using the pseudo inverse

  $$\boldsymbol{f} = (A^t A)^{-1} A^t \boldsymbol{R} \ .$$

- It should be mentioned that there are two constraints that the mixing proportions of the endmembers are expected to satisfy. The first is that the proportions should sum to unity and the second is that they should all be non-negative:

  $$\sum_{m=1}^{M} f_m = 1$$

  $$0 \leq f_m \leq 1 \text{ for all } m$$

  As discussed by Gross and Schott (1998) these constraints are sometimes violated if the endmembers are derived from average cover type spectra or the endmember selection is poor.