School of Applied Science

Reducing the Dimensionality of Hyperspectral Remotely Sensed Data with Applications for Maximum Likelihood Image Classification

Norman Ty Santich

Presented as Part of the Award of the Degree of Doctor of Philosophy of Curtin University of Technology

December 2007

Abstract

As well as the many benefits associated with the evolution of multispectral sensors into hyperspectral sensors there is also a considerable increase in storage space and the computational load to process the data. Consequently the remote sensing community is investigating and developing statistical methods to alleviate these problems.

The research presented here investigates several approaches to reducing the dimensionality of hyperspectral remotely sensed data while maintaining the levels of accuracy achieved using the full dimensionality of the data. It was conducted with an emphasis on applications in maximum likelihood classification (MLC) of hyperspectral image data. An inherent characteristic of hyperspectral data is that adjacent bands are typically highly correlated and this results in a high level of redundancy in the data. The high correlations between adjacent bands can be exploited to realise significant reductions in the dimensionality of the data, for a negligible reduction in classification accuracy.

The high correlations between neighbouring bands is related to their response functions overlapping with each other by a large amount. The spectral band filter functions were modelled for the HyMap instrument that acquires hyperspectral data used in this study. The results were compared with measured filter function data from a similar, more recent HyMap instrument. The results indicated that on average HyMap spectral band filter functions exhibit overlaps with their neighbouring bands of approximately 60%. This is considerable and partly accounts for the high correlation between neighbouring spectral bands on hyperspectral instruments.

A hyperspectral HyMap image acquired over an agricultural region in the south west of Western Australia has been used for this research. The image is composed of 512 × 512 pixels, with each pixel having a spatial resolution of 3.5 m. The data was initially reduced from 128 spectral bands to 82 spectral bands by removing the highly overlapping spectral bands, those which exhibit high levels of noise and those bands located at strong atmospheric absorption wavelengths. The image was examined and

found to contain 15 distinct spectral classes. Training data was selected for each of these classes and class spectral mean and covariance matrices were generated.

The discriminant function for MLC makes use of not only the measured pixel spectra but also the sample class covariance matrices. This thesis first examines reducing the parameterization of these covariance matrices for use by the MLC algorithm. The full dimensional spectra are still used for the classification but the number of parameters needed to describe the covariance information is significantly reduced. When a threshold of 0.04 was used in conjunction with the partial correlation matrices to identify low values in the inverse covariance matrices, the resulting classification accuracy was 96.42%. This was achieved using only 68% of the elements in the original covariance matrices.

Both wavelet techniques and cubic splines were investigated as a means of representing the measured pixel spectra with considerably fewer bands. Of the different mother wavelets used, it was found that the Daubechies-4 wavelet performed slightly better than the Haar and Daubechies-6 wavelets at generating accurate spectra with the least number of parameters. The wavelet techniques investigated produced more accurately modelled spectra compared with cubic splines with various knot selection approaches. A backward stepwise knot selection technique was identified to be more effective at approximating the spectra than using regularly spaced knots. A forward stepwise selection technique was investigated but was determined to be unsuited to this process.

All approaches were adapted to process an entire hyperspectral image and the subsequent images were classified using MLC. Wavelet approximation coefficients gave slightly better classification results than wavelet detail coefficients and the Haar wavelet proved to be a more superior wavelet for classification purposes. With 6 approximation coefficients, the Haar wavelet could be used to classify the data with an accuracy of 95.6%. For 11 approximation coefficients this figure increased to 96.1%.

First and second derivative spectra were also used in the classification of the image. The first and second derivatives were determined for each of the class spectral means

and for each band the standard deviations were calculated of both the first and second derivatives. Bands were then ranked in order of decreasing standard deviation. Bands showing the highest standard deviations were identified and the derivatives were generated for the entire image at these wavelengths. The resulting first and second derivative images were then classified using MLC. Using 25 spectral bands classification accuracies of approximately 96% and 95% were achieved using the first and second derivative images respectively. These results are comparable with those from using wavelets although wavelets produced higher classification accuracies when fewer coefficients were used.

Acknowledgments

Thanks go to my supervisors, Brendan McGann and Merv Lynch from Curtin and Norm Campbell and Harri Kiiveri from CMIS. Also special thanks to Peter Caccetta, who as group leader of RSII provided a lot of the direction and many helpful suggestions for this work.

Many thanks also to Carmel McManus, Tiffanie Harrington and Candace Culyer for all their administrative support and for just being such wonderful people!

Special thanks also to CMIS for providing the financial support, IT facilities and resources which I would never have received if I was based at Curtin. CMIS has been a great place to work and the people have been fantastic.

I also need to thank Dr Tom Cudahy of CSIRO Exploration and Mining. I had tried several times to obtain the HyMap spectral filter function data with no success. It wasn't until after I had given up trying and had resorted to modelling the responses mathematically that he came through with the data. As a result, the thesis is much more complete for it. Thank you Tom!

I would also like to thank the staff at Satellite Remote Sensing Services. Their support through the final stages of the thesis was very much appreciated.

And finally, to all my close friends and wonderful, wonderful family who have supported me so much over the last few years. I'm sure I would have gone insane without it.

Table of Contents

List of Figures

List of Tables

1 Introduction

As satellite remote sensing technology advanced, it was inevitable that the number of bands would increase, the bands would become narrower and that multispectral sensors would evolve into hyperspectral sensors. The first airborne spectroradiometer was developed in 1978 (Chiu & Collins 1978; Collins et al. 1983) and was referred to as the Collins Airborne Spectrometer by Bodechtel (2001). This was followed by many other airborne hyperspectral sensors such as the Airborne Imaging Spectrometer (AIS) (Vane & Goetz 1988), the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) (Vane et al. 1993; Green et al. 1998) and the Hyperspectral Mapper (HyMap) (Cocks et al. 1998). On November 21, 2000, Hyperion was launched on board the Earth Observing One (EO-1) satellite from Vandenburg Air Force Base (Ungar et al. 2003) and was the first satellite-borne imaging spectrometer to routinely collect data from an earth orbit (Pearlman et al. 2003). Kramer (1994) gives a comprehensive list and description of many spaceborne and airborne sensors and the more significant hyperspectral sensors are summarised in Appendix D on page 158.

The number of spectral bands in hyperspectral data is frequently referred to as the *dimensionality* of the data. Many of the spectral bands are highly correlated and the effective number of uncorrelated bands is sometimes referred to as the *intrinsic dimensionality* of the data. Most hyperspectral data sets have an intrinsic dimensionality of approximately 5% (or 5 – 10 bands), which is significantly less than the original dimensionality.

The "hyper" prefix in hyperspectral implies an excess of spectral data and this can have several different implications. Historically, hyperspectral sensors have had at least 48 spectral bands with a spectral resolution of 20 nm or better (Aspinall, Marcus & Boardman 2002). This compares with multispectral sensors which traditionally have had less than 10 bands. Typically, hyperspectral sensors have had more than 100 spectral bands. These spectral bands are usually located in the visible/near infrared region of the electromagnetic spectrum between wavelengths of 400 – 2500 nm. The data sets associated with these high numbers of bands can be

extremely large. Data sets of this magnitude have a significant influence on the data transmission times, amount of storage space required and the computational power and times needed for processing.

Despite these problems, hyperspectral imagery is a valuable tool in many different applications of remote sensing. In atmospheric science, hyperspectral imagery has been used for atmospheric correction (Sanders, Schott & Raqueño 2001) and trace gas detection (Gallagher, Wise & Sheen 2003). Over oceans and other marine features; water quality has been sensed (Hakvoort et al. 2002), concentrations of chlorophyll and suspended sediments have also been estimated (Herutet et al. 1999) and coastal bathymetry has been mapped (Sandidge & Holyer 1998).

For terrestrial environments, hyperspectral imagery has applications in geology, agriculture and vegetation monitoring. Geological applications have included soil and mineral mapping (Vaughan, Calvin & Taranik 2003), detecting rock weathering (Riaza et al. 2001) and also mine site waste (Mars & Crowley 2003). Farmers and agricultural scientists have been assisted by hyperspectral imagery for precision agriculture (Yang et al. 2002), weed detection (Goel et al. 2003), mapping of saline affected areas (Metternicht & Zinck 2003) and monitoring crop conditions such as nutrient status (Strachan, Pattey & Boisvert 2002), vegetation water content (Serrano et al. 2000) and chlorophyll concentrations (Haboudane et al. 2002).

Remotely sensed data frequently contains pixels having more than one type of land cover class within each pixel. Each class within the pixel is known as an endmember. Endmember analysis (also known as spectral unmixing) involves determining what endmembers occupy each pixel and what fraction of the pixel they occupy. When used for land cover classification and spectral unmixing, hyperspectral data has the advantage in that the narrower bands are able to detect finer spectral details and characteristics. As a result the data is able to detect less predominant endmembers within image pixels. This is significant, as an analyst may be looking to detect the presence of a particular endmember within a pixel. If the fraction of the endmember is relatively small compared to the pixel size, data from broad band sensors will lack the spectral resolution required to be able to detect it.

The fine spectral resolution of hyperspectral data however makes it more suitable for detecting the presence of low prevalence signals within pixel spectra.

The accuracy of thematic maps can also be improved by using hyperspectral data. The capacity to detect finer spectral details gives the classification algorithms more power when discriminating between spectral classes. There are many different spectral classification algorithms and Mallet, Coomans and de Vel (1996) reviews several of these. Established traditional discriminant analysis techniques such as Fisher's linear discriminant analysis and Bayesian classification are reviewed. New emerging techniques such as flexible discriminant analysis, regularized discriminant analysis and penalized discriminant analysis are also introduced. The latter two are both able to handle data with large dimensionalities and where the ratio of training samples to spectral bands is low.

The most widely used classification technique is maximum likelihood classification (MLC). When deciding on which class label to allocate to a pixel, MLC makes use of both the sample class means and the class covariance statistics. One of the disadvantages of using MLC for hyperspectral data is that if the number of training samples for a class is low there can be problems in reliably estimating the class covariance statistics. When the number of training pixels is less than the dimensionality, N, the resulting class covariance matrix will be singular, and it is recommended that at least $10N$ or even as high as $100N$ training samples be used in the estimation of the class covariance matrices (Richards & Jia 1999). The measured pixel spectra will be represented as $N \times 1$ column vectors and the class covariance information will be contained in $N \times N$ matrices. When N is large, as with hyperspectral data, there is an enormous number of parameters needed by the MLC algorithm and the computational load is massive.

It is for these reasons that this work focuses on the problems of reducing the dimensionality and redundancy inherent in hyperspectral remotely sensed data, with an emphasis on applications in MLC of hyperspectral imagery. Also, the MLC algorithm is widely available and supported on several image processing software platforms, such as ENVI, ER Mapper and MultiSpec. The spectral bands on hyperspectral sensors are closely spaced giving near-continuous spectral

measurements. Consequently, neighbouring spectral bands exhibit very high correlations which can be exploited as a large source of redundancy within the data. It may be possible to remove a significant number of these highly correlated bands yet still retain the levels of classification accuracy possible when using the full set of available bands.

This research begins with an overview of established dimensionality reduction techniques in Chapter 2. The MLC algorithm is introduced in Chapter 3 as well as a review of previous approaches to significantly reduce the parameterization of the covariance matrix. Many of these make use of the typically sparse nature of the inverse covariance matrices associated with hyperspectral data. The nature of the spectral band filter functions is examined in Chapter 4 and the degree to which these filter functions overlap is assessed. In Chapter 5, methods are applied to reduce the parameterization of the class covariance matrices and these are subsequently used with the MLC algorithm to classify the data. The hyperspectral image data used was acquired at Toolibin in the south west corner of Western Australia by the HyMap sensor. Training data was selected for 15 distinct spectral classes and from this the original class covariance matrices were generated. The benefit of having the full dimensionality of the measured pixel spectra is retained yet the number of parameters used in the computation is significantly reduced. Chapter 5 extracts the spectra from a randomly selected strip of 20 pixels and trials different methods for estimating the spectra using significantly fewer spectral bands/parameters. The methods used include wavelet techniques and cubic splines using different knot selection approaches. Cubic splines are used for interpolating the data by taking advantage of the fact that the digitized spectra approach smooth spectral curves. In Chapter 7 the entire image is reduced dimensionally using similar approaches to those employed in Chapter 6. By using pixel spectra of a much lower dimensionality, the associated class covariance matrices are also significantly smaller automatically and no further processing is required to reduce their size. The problem of insufficient training samples to reliably estimate the class covariance matrices is also eased. The resulting dimensionally reduced images are then classified using MLC. Chapter 7 summarises and compares the performance of the approaches taken and suggests some possible extensions to the work performed.

2 Previous Approaches to Dimensionality Reduction of Hyperspectral Data

As remote sensing instruments advance from multispectral to hyperspectral the demands on processing time, computational load and storage space increase. Even though hyperspectral data typically has greater than 100 bands, it has been suggested that the intrinsic spectral dimensionality is significantly less, perhaps even lower than 10 (Harsanyi & Chang 1994). By definition, the term *hyperspectral* implies that there is an excess of spectral information. In this chapter, previous approaches to reducing the dimensionality are reviewed. These approaches can be grouped into two main categories: approaches where the data are spectrally transformed, and those approaches in which individual bands are selected to approximate the entire suite of bands.

2.1 Dimensionality Reduction via Spectral Transforms

A common approach to the problem of dimensionality reduction involves spectrally transforming the original set of bands into a new set of component images. The new set of component images will typically be ordered according to variance or information content and the effective number of bands can be reduced by discarding those new image components exhibiting low variances. Even though the dimensionality will be significantly reduced, ideally the loss in information content will be negligible. In this section, the Principal Components Transform (PCT), Maximum Noise Fraction (MNF) and Noise Adjusted Principal Components (NAPC) transforms are presented. These transforms decorrelate the original data and reorder the resulting features according to variance or noise content. Finally, canonical variate analysis (CVA) is reviewed. CVA transforms the data so that the classes have maximum spectral separability.

2.1.1 Principal Components Transform

The PCT (also known as the Karhunen-Loève or Hotelling Transform) is a method for decorrelating multispectral data and arranging the resulting components in order of decreasing variance. If $\mathbf{x}^T = (x_1, x_2, \ldots, x_N)$ is a pixel vector in an N-band scene containing L pixels, the mean pixel vector is given by

$$\overline{\mathbf{x}} = \frac{\sum_{l=1}^{L} \mathbf{x}_l}{L} \tag{2.1}$$

and the $(N \times N)$ covariance matrix, $\mathbf{C}_x$, is given by

$$\mathbf{C}_x = \frac{\sum_{l=1}^{L} (\mathbf{x}_l - \overline{\mathbf{x}})(\mathbf{x}_l - \overline{\mathbf{x}})^T}{L-1} \tag{2.2}$$

The PCT can then be defined as

$$\mathbf{z} = \mathbf{A}_{PCT}^T \mathbf{x} \tag{2.3}$$

where $\mathbf{A}_{PCT}^T$ is the $(N \times N)$ matrix of normalized eigenvectors of $\mathbf{C}_x$ and $\mathbf{z}$ is the newly transformed pixel vector.

The transformed data will be uncorrelated and the components will be in order of decreasing variance. Because the majority of the data variance will be contained in the lower order principal components, the higher order principal components can be rejected, resulting in a negligible loss of information content and a significant reduction in the data dimensionality. As the components are ordered in terms of global variance, and not local detail, the components need to be examined before they are rejected (Richards & Jia 1999). Higher order components may contain local detail that is not present in the lower order components. As a result of this, the PCT may not be optimal for class discrimination or separability (Chang et al. 1998).

As the transformed data is uncorrelated, its covariance matrix will be diagonal, with the diagonal elements being the eigenvalues of $\mathbf{C}_x$ (Roger 1996a). The diagonal elements, e are arranged such that $e_1 > e_2 > \ldots > e_N$ and also represent the variances of

the principal components. The proportion of the variance of the entire scene contained in the j^{th} component is given by:

$$\frac{e_j}{\sum_{i=1}^{N} e_i}. \tag{2.4}$$

However, the PCT has several disadvantages. Green et al. (1988) have noted that even though the PCT orders components by decreasing variance, this does not necessarily result in the components being ordered by image quality, or signal-to-noise ratio (SNR). Noise may contribute significantly to the variance of a component and as a result, the component may contain less information than a component with lower variance (Roger 1994a). The PCT is not invariant to changes in radiometric scale and is biased to high variance bands. An example of this is the effect of the shape of the solar irradiance spectrum where the shorter wavelength bands will have much higher variances and consequently bands in the visible/near infrared will tend to dominate the PCT.

Implementing the PCT on hyperspectral data has a huge computational load, which can be made more efficient if the segmented PCT devised by Jia and Richards (1999) is used. Jia and Richards (1999) have observed that neighbouring bands are highly correlated and that these high correlations appear in blocks within the correlation matrix. If the data set is partitioned into H subgroups, the PCT can be performed separately on each individual subgroup. Ideally, each subgroup should be composed of highly correlated bands and if the number of bands in subgroups 1, 2, …, H is N_1, N_2, …N_H respectively then

$$\sum_{h=1}^{H} N_h = N. \tag{2.5}$$

Whereas the conventional PCT requires N^2 multiplications to transform each pixel, the segmented PCT only requires N_h^2 multiplications for each subgroup. The total number of multiplications needed to transform a pixel using the segmented PCT is therefore $\sum_{h=1}^{H} N_h^2$. As well as reducing computational time, the segmented PCT will also ease the problem of solar spectrum weighting (Jia & Richards 1999).

2.1.2 Maximum Noise Fraction

As a result of the PCT being unable to order components in terms of image quality, Green et al. (1988) developed the MNF. The data measured by the sensor, $\mathbf{x}$, can be thought of as having a signal component, $\mathbf{x}_s$, and a noise component, $\mathbf{x}_n$,

$$\mathbf{x} = \mathbf{x}_s + \mathbf{x}_n. \tag{2.6}$$

The covariance matrix of $\mathbf{x}$ can then be expressed as:

$$\mathbf{C}_x = \mathbf{C}_s + \mathbf{C}_n \tag{2.7}$$

where $\mathbf{C}_s$ and $\mathbf{C}_n$ are the signal and noise covariance matrices respectively. The MNF transform is defined as

$$\mathbf{z} = \mathbf{A}_{MNF}^{T}\mathbf{x} \tag{2.8}$$

where $\mathbf{z}$ is the transformed data and $\mathbf{A}_{MNF}^{T}$ is the eigenvector matrix of $\mathbf{C}_n\,\mathbf{C}_x^{-1}$. $\mathbf{C}_x^{-1}$ is the inverse covariance matrix of $\mathbf{x}$.

If dark current measurements are available, these can be used to estimate $\mathbf{C}_n$. When the dark current measurements are not available, Green et al. (1988) describe a method to estimate $\mathbf{C}_n$ using near neighbour differences. The technique assumes that the signal at any point in an image will be strongly correlated with the signal at neighbouring pixels and that noise will only exhibit weak spatial correlations. The technique is applicable to salt and pepper noise and also image striping.

The components generated by the MNF transform will be ordered by increasing image quality, or decreasing noise fraction. Noise fraction is defined as

$$\text{Var}\,\{\mathbf{x}_n\} \,/\, \text{Var}\,\{\mathbf{x}\}, \tag{2.9}$$

That is, the ratio of the noise variance in a band to the total variance of that band.

Using the MNF transform, noise can be effectively removed from multispectral data. This is accomplished by either smoothing or rejecting the noisiest MNF components and then retransforming back to the original space. Unlike the PCT, the MNF transform is invariant to scale changes in any bands as it depends on the noise fraction which is a *ratio* of the noise variance to the total band variance. MNF is equivalent to the PCT when the noise is uncorrelated and has equal variance across all bands. This could explain the success of the PCT at ordering bands in terms of image quality for many remotely sensed data sets: the noise may have had equal variance and been uncorrelated across all bands (Green et al. 1988).

2.1.3 Noise Adjusted Principal Components

Lee, Woodyatt and Berman (1990) have modified the MNF transform to produce the NAPC Transform. The NAPC is a variation of the PCT and is mathematically equivalent to the MNF transform. The NAPC transform is derived in Roger (1994a) using the following steps:

1) Compute the orthonormalized eigenvector matrix, $\mathbf{V}_{\text{ort}}$, and the diagonal eigenvalue matrix, $\mathbf{E}_n$, of $\mathbf{C}_n$.

2) Using the renormalization matrix, $\mathbf{F} = \mathbf{V}_{\text{ort}}\mathbf{E}_n^{-1/2}$, generate the noise adjusted data covariance matrix, $\mathbf{C}_{\text{adj}}$

$$\mathbf{C}_{\text{adj}} = \mathbf{F}^T\mathbf{C}_x\mathbf{F}. \tag{2.10}$$

3) Compute the eigenvector matrix, $\mathbf{V}_{\text{adj}}$, of $\mathbf{C}_{\text{adj}}$

4) The NAPC transformed data, $\mathbf{z}$, is then given by

$$\mathbf{z} = \mathbf{A}_{NAPC}^T\mathbf{x} \tag{2.11}$$

where $\mathbf{A}_{NAPC}^T = \mathbf{F}\mathbf{V}_{\text{adj}} = \mathbf{V}_{\text{ort}}\mathbf{E}_n^{-1/2}\mathbf{V}_{\text{adj}}$ and is the NAPC transform matrix.

From the above steps, the NAPC transform can be summarised as a two-stage transformation. First, the data covariance matrix, $\mathbf{C}_x$, is weighted so the noise covariance matrix becomes the identity matrix, that is, a matrix with a value of 1 for every element along the main diagonal and a value of 0 for every element off the main diagonal. This results in the noise being decorrelated and obtaining unit variance across all bands. Secondly, a PCT is implemented on the resulting transformed data set (Lee, Woodyatt & Berman 1990). The transformed data set will have components ordered by decreasing SNR, with band 1 having the highest SNR and band N having the lowest SNR.

2.1.4 Canonical Variate Analysis

With CVA (Campbell 1984; Kiiveri 1992; Richards & Jia 1999), uncorrelated, linear combinations of the original set of variables/bands are chosen to maximize the ratio of the between class to within class variances. CVA is different to the PCT in that CVA utilizes the within and between class covariance matrices, whereas the PCT examines the global covariance matrix of the data. CVA is therefore more sensitive to class structure and more suited for class separation.

Given N dimensional data consisting of G classes, the within class covariance matrix, $\mathbf{C}_W$, is defined as

$$\mathbf{C}_W = \left\{ \sum_{k=1}^{G} (L_k - 1)\mathbf{C}_k \right\} / L_G \tag{2.12}$$

where L_k is the number of pixels in the k^{th} class, $\mathbf{C}_k$ is the covariance matrix for the k^{th} class and $L_G = \sum_{k=1}^{G} L_k$, i.e., the total number of training pixels used. The between class covariance matrix $\mathbf{C}_B$ is given by

$$\mathbf{C}_B = \frac{1}{L} \sum_{k=1}^{G} L_k (\bar{\mathbf{x}}_k - \bar{\mathbf{x}}_0)(\bar{\mathbf{x}}_k - \bar{\mathbf{x}}_0)^T \tag{2.13}$$

where L is the total number of pixels, $\bar{\mathbf{x}}_k$ is the mean of the k^{th} class and $\bar{\mathbf{x}}_0 = (\Sigma L_k \bar{\mathbf{x}}_k)/L$. CVA involves finding canonical vectors, $\mathbf{t}$, such that the ratio $o_k = \mathbf{t}^T\mathbf{C}_B\mathbf{t}/\mathbf{t}^T\mathbf{C}_W\mathbf{t}$ is maximized. o_k is the ratio of between class to within class variance and gives an indication of the separability of the classes. This leads to the generalized eigenvalue equation

$$(\mathbf{C}_B - \mathbf{O}\mathbf{C}_W)\mathbf{T} = 0 \tag{2.14}$$

where $\mathbf{O}$ is the diagonal matrix of the canonical roots o_k and $\mathbf{T}$ is the matrix of the canonical vectors. $\mathbf{O}$ and $\mathbf{T}$ need to be solved for in (2.14), subject to the constraints

$$\mathbf{T}^T\mathbf{C}_W\mathbf{T} = \mathbf{I}, \quad \mathbf{T}^T\mathbf{C}_B\mathbf{T} = n_w\mathbf{O} \tag{2.15}$$

where n_w is the within class degrees of freedom. The diagonal elements of $\mathbf{O}$ will be in order of decreasing magnitude and if $N > G - 1$, there will only be $G - 1$ non-zero roots for (2.14). The dimensionality of the transformed data will therefore be reduced.

2.2 Feature Selection

Feature selection involves selecting a subset of the original bands to reduce the classification cost and minimize the classification error. By reducing the number of features, the number of training samples required for classification is also reduced and classification will become more computationally efficient. In this section, previous approaches to the problem of feature selection/reduction are described.

2.2.1 Spectral Basis Functions

Price (1975) developed a method to significantly reduce the number of spectral bands/independent variables that are required to explain observed spectra to within the noise level of the instrument. The method also determines the spectral location and width of these spectral bands. Initially, the method was developed for use with Infrared Interferometer Spectrometer (IRIS) data having dimensionality $N = 862$ in the spectral range 6.6 – 25 μm (400 – 1600 cm^{-1}). However, the methodology was

awkward and inefficient when the number of independent variables was > 12 – 15. It has since been improved (Price 1994; 1997) to allow for the simultaneous selection of multiple spectral bands. The evolved methodology is described in Price (1994; 1997) and is summarised here.

Spectral basis functions can be used to approximate observed *N*-dimensional reflectance spectra as

$$\mathbf{x} \approx \sum_{h=1}^{H} W_h \boldsymbol{\varphi}_h \tag{2.16}$$

where $\mathbf{x}$ are the observed spectra, W_h are spectral integrals relating to the original $\mathbf{x}$ and $\boldsymbol{\varphi}_h$ are the *N*-dimensional spectral basis functions. Each $\boldsymbol{\varphi}_h$ has an associated spectral band, $[\lambda_h(\min) - \lambda_h(\max)]$ that corresponds to the interval over which W_h is integrated. Within this interval, the mean value of $\boldsymbol{\varphi}_h$ is 1, and this value decreases outside the spectral interval according to the spectral correlation inherent in the reflectance data. Each term in the expansion of (2.16) represents a successive approximation to $\mathbf{x}$. For the approximation of $\mathbf{x}$ to be practical, the number of basis functions, *H*, needed to estimate $\mathbf{x}$ to within a very small residual must be < *N*.

If $\delta\mathbf{x}_h$ is the difference between the measured reflectance and its approximation to order *h*, then W_h is defined as the average value of $\delta\mathbf{x}_h$ over the selected spectral interval, $[\lambda_h(\min) - \lambda_h(\max)]$, i.e.

$$W_h = \frac{1}{\lambda_h(\max) - \lambda_h(\min)} \int_{\lambda_h(\min)}^{\lambda_h(\max)} \delta\mathbf{x}_h d\lambda \tag{2.17}$$

Initially, $\delta\mathbf{x}_1 = \mathbf{x}$. The basis function $\boldsymbol{\varphi}_h$ is defined as

$$\boldsymbol{\varphi}_h = \frac{\langle \delta\mathbf{x}_h W_h \rangle}{\langle (W_h)^2 \rangle} \tag{2.18}$$

where the angle brackets represent averaging over the *L* measured spectra. By definition, $\delta\mathbf{x}_h$ will have a value of zero somewhere in the interval $[\lambda_h(\min) -$

$\lambda_h(\max)]$. As h increases, $\delta\mathbf{x}_h$ passes through zero at more values of λ and the magnitude of the residuals $\int(\delta\mathbf{x}_h)^2 d\lambda$ decreases.

The basis function at level h is computed by subtracting the contribution of lower order basis functions and integrating the residual over a spectral interval. Hence, we need

$$W_h = \frac{1}{\lambda_h(\max) - \lambda_h(\min)} \int_{\lambda_h(\min)}^{\lambda_h(\max)} \left(\mathbf{x} - \sum_{\gamma=1}^{h-1} W_\gamma \boldsymbol{\varphi}_\gamma \right) d\lambda \tag{2.19}$$

However, for $\gamma < h$, $\boldsymbol{\varphi}_\gamma$ are already known so we can set

$$b_{h\gamma} = \frac{1}{\lambda_h(\max) - \lambda_h(\min)} \int_{\lambda_h(\min)}^{\lambda_h(\max)} \boldsymbol{\varphi}_\gamma d\lambda \tag{2.20}$$

i.e., $b_{h\gamma}$ is the integral over the h^{th} spectral interval of the γ^{th} basis function. Also, we define

$$w_h = \frac{1}{\lambda_h(\max) - \lambda_h(\min)} \int_{\lambda_h(\min)}^{\lambda_h(\max)} \mathbf{x} d\lambda \tag{2.21}$$

where $W_1 = w_1$ and for $h > 1$

$$W_h = w_h - \sum_{\gamma=1}^{h-1} b_{h\gamma} W_\gamma \tag{2.22}$$

Rewriting (2.22) in terms of the original spectra w_h gives

$$W_h = \sum_{\gamma=1}^{h} a_{h\gamma} w_\gamma \tag{2.23}$$

where $a_{h\gamma} = 1$ and for $a_{h > \gamma}$

$$a_{h\gamma} = -\sum_{\alpha=\gamma}^{h-1} b_{h\alpha} a_{\alpha\gamma} \,. \tag{2.24}$$

Finally, the basis functions $\boldsymbol{\varphi}_h$ are given by

$$\boldsymbol{\varphi}_h = \frac{\sum_{\gamma=1}^{h} a_{h\gamma} \langle \mathbf{x} w_\gamma \rangle - \sum_{\gamma=1}^{h-1} J_{h\gamma} \boldsymbol{\varphi}_\gamma}{J_{hh}} \tag{2.25}$$

where

$$J_{h\gamma} = \sum_{\alpha=1}^{h} \sum_{\beta=1}^{\gamma} a_{h\alpha} a_{\gamma\beta} \langle w_\alpha w_\beta \rangle \,. \tag{2.26}$$

The computational sequence for generating the basis functions is:

For h and γ from 1 to H

1) Calculate $\langle \mathbf{x} w_h \rangle$ and $\langle w_h w_\gamma \rangle$ and generate the first basis function $\boldsymbol{\varphi}_1 = \langle \mathbf{x} w_h \rangle \,/\, \langle w_h^2 \rangle$

For $h = 2, 3, \ldots, H$

2) Calculate $b_{h\gamma}$ using (2.20)

3) Calculate $a_{h\gamma}$ using (2.24)

4) Generate the $\boldsymbol{\varphi}_h$ using (2.25)

The above discussion shows how to approximate spectral reflectance data in terms of basis functions without mentioning how to determine the spectral intervals associated with the basis functions. The procedure for determining these intervals is discussed here. The dimensionality of the data is reduced by first dividing the range of the data into equal broad-band intervals and averaging each of the spectral data points within

these broad-band intervals. ie, $\bar{x}_1 = \sum_{i=1}^{N_h} x_i$, $\bar{x}_2 = \sum_{i=N_h+1}^{2N_h} x_i$, etc. where i is the original spectral band number and N_h is the number of spectral data points in each broad-band interval. Then a cosignal matrix is calculated

$$\mathbf{Y} = [y_{h\gamma}] = (1/L)\sum_l \bar{x}_h^l \bar{x}_\gamma^l = \langle \bar{x}_h \bar{x}_\gamma \rangle \tag{2.27}$$

where $\mathbf{Y}$ is a symmetric matrix and the trace, given by $\sum_h \langle \bar{x}_h^2 \rangle$, gives an indication of the total variance of the data. If a spectral interval $\Delta\lambda$ is selected for the description of the basis functions, then the other spectral intervals can be reduced by their correlation with $\Delta\lambda$, i.e., $\bar{x}_h' = \bar{x}_h - r_{h\Delta\lambda}\sigma_h / \sigma_{\Delta\lambda}\bar{x}_{\Delta\lambda}$, where $r_{h\Delta\lambda} = \langle \bar{x}_h \bar{x}_{\Delta\lambda} \rangle / (\sigma_h \sigma_{\Delta\lambda})$ and $\sigma_h^2 = \langle \bar{x}_h^2 \rangle$. After selecting $\Delta\lambda$ the new cosignal matrix becomes $\mathbf{Y}' = [y'_{h\gamma}] = y_{h\gamma} - (y_{h\Delta\lambda} y_{\gamma\Delta\lambda})/\sigma_{\Delta\lambda}^2$ and with row and column $\Delta\lambda$ deleted the trace of $\mathbf{Y}'$ becomes $\sum_{h \neq \Delta\lambda} \sigma_h^2 (1 - r_{h\Delta\lambda}^2)$. The spectral interval $\Delta\lambda$ should be chosen so that the residual trace is minimized. Once this is chosen then $\mathbf{Y}'$ is used to select the next spectral interval. The procedure is continued until H spectral intervals are chosen.

The method just described only selects preliminary spectral intervals and does not yield an optimum set of basis functions. To find the optimum spectral intervals, the preliminary basis functions need to be examined. The range of the spectral intervals needs to be adjusted so that the value of φ within the range exceeds 0.85. If adjacent basis functions overlap where the value of both basis functions is above 0.9 then a wavelength should be selected that separates the two intervals where the preliminary basis functions are equal. With the improved spectral intervals, the spectral basis functions can be recalculated.

2.2.1.1 Applications of the Price (1994, 1997) Method to Hyperspectral Data

The concept for band selection as described above has been applied to several different applications. Price (1975) applied an earlier version of the method to Infrared Interferometer Spectrometer (IRIS) data. The IRIS, aboard the Nimbus 4 satellite, measures spectral data over 6.25 – 25 μm with a spectral dimensionality of

N = 862. It was found that 9 spectral bands were able to describe the majority of the spectra to within the noise levels of the IRIS data.

The method was then modified in Price (1990) for use with visible/near infrared data and applied to a collection of 564 soil spectra. The spectral data were acquired with an Exotech Model 20 radiometer in the spectral range 0.55 – 2.32 μm with a spectral resolution of 0.01 μm. The dimensionality of the data is N = 178. 4 spectral bands were found that could describe 99.6% of the variability of the data. The same method was then applied to two different crop data sets (Price 1992). The first was field data measured with an Exotech model 20C spectroradiometer. The data was measured in the spectral range 0.5 – 2.31 μm at a resolution varying between 0.0007 – 0.0024 μm and contained 1387 crop/soil spectra. The different crops included corn, soybeans, winter wheat, sunflowers and alfalfa. The spectral resolution of the data was reduced to 0.01 μm resulting in a dimensionality of N = 147. 5 spectral bands were found that could describe 99.8% of the spectral variability of the data.

The second crop data set was acquired by a helicopter-borne Field Spectrometer System (FSS) in the spectral range 0.4 – 2.4 μm. The spectral resolution of the data was 0.02 μm and 0.05 μm between the ranges 0.4 – 1.1 μm and 1.1 – 2.4 μm respectively. The data set showed more noise than the field data, and as a result, 5 spectral bands could describe only 98.7% of the spectral variability of the data.

In Price (1997) 45 AVIRIS scenes, which are representative of most surface types on Earth, are analysed using the band selection method described above. 5 spectral bands described 98.45% of the spectral variability of the data and 20 spectral bands described 99.90%. 39 spectral bands were found that were able to describe 99.95% of the spectral variability of the data.

A comparison of the different applications of Price's (1994; 1997) band selection method for VIS/NIR reflectance data is shown in table 2.1. The first five spectral intervals selected by the method in each application have been shown as well as the residuals resulting from the approximations. In the spectral regions 1.35 – 1.47 μm and 1.81 – 2.02 μm, the atmospheric transmittance is strongly affected by water vapour. For this reason, these spectral regions were omitted while spectral intervals were selected to approximate the spectral reflectance data.

2.2.2 Spatial Autocorrelation

Warner and Shank (1997) have devised methods for feature selection that utilizes the spatial autocorrelation information contained within an image. Three broad classes of feature selection are considered:

1) narrow-band feature selection

2) broad-band feature selection

3) non-adjacent, multiple-band feature selection

Normalized difference ratio images are formed for each combination of spectral bands and then the autocorrelation of these ratio images is calculated. Band ratios are used, as they tend to suppress illumination effects. Only neighbouring pixels are considered when calculating the spatial autocorrelation.

Table 2.1: A comparison of the applications of Price's (1994; 1997) band selection method.

Measuring Instrument / Surface type	Price (1990) Exotech Model 20 Radiometer soils		Price (1992) Exotech Model 20C Spectroradiometer crops/soils		Price (1992) Field Spectrometer System crops/soils		Price (1997) AVIRIS all	
Spectral Band	Spectral Interval (μm)	Residual (%)	Spectral Interval (μm)	Residual (%)	Spectral Interval (μm)	Residual (%)	Spectral Interval (μm)	Residual (%)
1	0.93 – 1.13	25.8	0.99 – 1.13	37.5	1.05 – 1.38	41.1	0.99 – 1.08	26.4
2	2.03 – 2.31	5.8	1.48 – 1.80	2.5	2.03 – 2.38	6.8	2.08 – 2.20	11.7
3	0.63 – 0.74	2.0	0.57 – 0.71	1.3	0.55 – 0.69	3.9	0.44 – 0.55	3.88
4	1.61 – 1.80	0.4	2.03 – 2.31	0.5	0.77 – 0.85	1.8	1.12 – 1.16	1.92
5			0.73 – 0.83	0.2	1.53 – 1.78	1.3	0.66 – 0.69	1.55

With the narrow-band feature selection approach the spatial autocorrelation is calculated for each normalized difference ratio image. The pairs of ratioed bands are then sorted in order of decreasing spatial autocorrelation. The first two narrow-band features to be selected are the two bands that result in the normalized difference ratio image with the highest spatial autocorrelation. The next feature to be selected is the one that when ratioed with each of the features already selected, has the highest spatial autocorrelation for both. Additional narrow-band features are selected in a manner similar to that used to select the third feature.

For the class of broad-band feature selection there are two different approaches. The first selects the pair of adjacent bands that when ratioed have the lowest spatial autocorrelation (i.e., the pair of adjacent bands that are the most similar). This pair of adjacent bands is then grouped together to form the first feature. This is repeated until the final number of desired features has been selected. The second approach for broad-band feature selection begins with the same best bands selected by the narrow-band feature selection method. The number of narrow-band features to begin with is the same as the final number of desired features. Adjacent bands are added to each of the selected narrow-band features. A spectral band should only be added to a group if it improves the spatial autocorrelation.

The non-adjacent, multiple-band feature selection technique is similar to the broad-band feature selection approach except it is not constrained by the necessity of only selecting adjacent spectral bands. As with the broad-band case, the procedure begins by selecting the best narrow-band features or the two most similar bands (although not necessarily the two most similar *adjacent* bands). Bands are added to these features by selecting other similar bands, ensuring that the spatial autocorrelation is improved at each step.

After testing with synthetic data, Warner and Shank (1997) applied the above methods to AVIRIS data of part of the Arid Lands Ecology (ALE) preserve near Richland, Washington. Warner and Shank (1997) note that the AVIRIS SNR is poor at wavelength regions below 400 nm and above 2400 nm. The results from the narrow-band feature selection method are shown in table 2.2, with the best 10 selected bands shown. Only 2 bands were selected at wavelengths greater than 1400

Table 2.2: First ten narrow-band features selected by Warner and Shank (1997) using spatial autocorrelation information.

Rank	AVIRIS Band Number	Central Wavelength (nm)
1	19	557
1	64	962
3	124	1521
4	31	676
5	39	723
6	47	799
7	25	616
8	186	2123
9	37	704
10	93	1240

nm and the majority of bands were from the VIS/NIR. The broad-band feature selection method selected 20 broad-band features from the AVIRIS data with an average width of ~ 50 nm. It is noted that the growth of these features was limited by the growth of adjacent groups.

Twenty features were selected from the AVIRIS data using the non-adjacent, multiple-band feature selection approach. All groups selected between 440 nm and 718 nm were constituted of contiguous bands. Because there are so many spectral bands on AVIRIS, the non-adjacent, multiple-band feature selection algorithm takes much longer to execute. The groups are generally larger at the infrared wavelengths, particularly between 2000 – 2500 nm.

3 Maximum Likelihood Classification of Remotely Sensed Data and Methods for Simplifying the Covariance Matrices

This chapter begins with an overview of the maximum likelihood classification (MLC) process and examines the structure of the covariance information used by MLC. It will be shown that this structure can easily be simplified or modelled and existing methods to achieve this are reviewed. The first of these is the minimum distance classifier, which forgoes the covariance information completely. This simplification can be improved by incorporating the standard deviation information. Other methods are introduced which assume that either the inverse covariance matrix has a band-diagonal form or the correlation matrix has a block-diagonal form. These assumptions are then relaxed and two algorithms are presented, which set individual terms in the inverse covariance matrix to zero, and thus reduce the number of parameters required to reliably estimate the covariance information.

In this chapter the focus shifts from dimensionality reduction towards a major application for remotely sensed data — image classification. Each remotely sensed scene will contain several different elements, such as roads, forests, houses and lakes, etc. These different elements (referred to as classes) can be identified using either supervised or unsupervised classification techniques. The most commonly used supervised classification technique for multispectral remotely sensed data is maximum likelihood classification. When applied to hyperspectral data, MLC is time consuming and computationally expensive. It is therefore highly beneficial to be able to simplify the process and hence reduce the amount of computation and time required for the task of classification.

3.1 Maximum Likelihood Classification

The process of MLC assumes that the data has a multivariate normal distribution and is distinguished by making use of the second order class statistics. The discriminant function for MLC is

$$g_k(\mathbf{x}) = -\ln|\mathbf{C}_k| - (\mathbf{x} - \overline{\mathbf{x}}_k)^T \mathbf{C}_k^{-1}(\mathbf{x} - \overline{\mathbf{x}}_k). \quad (3.1)$$

A pixel will be classified as being a member of class k if $g_k(\mathbf{x})$ is a maximum over all classes. If $\mathbf{x}$ is normally distributed, the quadratic term on the right of (3.1) will have a χ^2 distribution with N degrees of freedom. Unlike simpler classifiers, the decision surfaces are not restricted to being linear, but are instead quadratic. Theoretically, the minimum number of training samples per class required for MLC is $N + 1$. Any less than this and the covariance matrix will be singular and as a result its inverse will be impossible to calculate. In practise, however, the number of training samples should be at least $10N$ and desirably $100N$ (Swain & Davis 1978). When a class only has a small number of pixels present in a scene, problems can arise as there may be insufficient training samples to estimate the covariance information.

3.2 Minimum Distance Classification

One of the biggest simplifications that can be made to MLC is to completely disregard the second order statistics and replace the inverse covariance information in (3.1) with the identity matrix. The resulting discriminant function becomes

$$g_k(\mathbf{x}) = (\mathbf{x} - \overline{\mathbf{x}}_k)^T (\mathbf{x} - \overline{\mathbf{x}}_k), \quad (3.2)$$

where a pixel $\mathbf{x}$ will be classified as belonging to class k if g_k is a minimum over all classes. This new classifier is known as the minimum distance classifier and is essentially a measure of the Euclidean distance between the pixel and the class mean.

The minimum distance classifier is faster than MLC, although without the covariance information the minimum distance classifier lacks the flexibility when allocating classes to pixels. Unlike MLC, where the decision surfaces are second order multidimensional curves, the decision surfaces for the minimum distance classifier are linear. However for small classes where the number of available training samples is limited, the minimum distance classifier proves to be much more effective than MLC (Richards & Jia 1999). With hyperspectral data, MLC will perform poorly, as a large number of training samples are required to reliably estimate the covariance information.

The minimum distance approximation to MLC can be improved by incorporating the class standard deviation information into the discriminant function. The class covariance matrix, $\mathbf{C}_k$, can be rewritten in terms of the correlation matrix, i.e.,

$$\mathbf{C}_k = \mathbf{S}_k \mathbf{R}_k \mathbf{S}_k, \tag{3.3}$$

where $\mathbf{R}_k$ is the correlation matrix and $\mathbf{S}_k$ is a diagonal matrix whose elements contain the class standard deviations. The inverse covariance matrix can therefore be written as

$$\mathbf{C}_k^{-1} = \mathbf{S}_k^{-1} \mathbf{R}_k^{-1} \mathbf{S}_k^{-1}. \tag{3.4}$$

If this new expression for $\mathbf{C}_k^{-1}$ is substituted into (3.1) and the inverse correlation matrix replaced with the identity matrix, the resulting quadratic term in (3.1) becomes

$$Q = (\mathbf{S}_k^{-1}\mathbf{d})^T (\mathbf{S}_k^{-1}\mathbf{d}), \tag{3.5}$$

where $\mathbf{d}$ denotes the difference vector, $\mathbf{x} - \overline{\mathbf{x}}_k$. This is known as the standardized distance to the mean (Roger 1996b). Because $\mathbf{S}_k^{-1}$ is diagonal, a significant saving in processing time can be realized.

A further improvement to the minimum distance classifier is possible if the $\mathbf{C}_k^{-1}$ is expressed in terms of the partial correlation matrix. The elements of the partial correlation matrix are the negative values of the partial correlation coefficients. Partial correlation coefficients indicate how strongly correlated two variables are when the effect of other variables has been removed or held constant (Minium, King, & Bear 1993). For three bands labelled A, B and C, the partial correlation coefficient of bands A and B with band C held constant is given by (Kendall & Stuart 1961)

$$\rho_{AB \cdot C} = \frac{r_{AB} - r_{AC} r_{BC}}{\sqrt{\left(1 - r_{AC}^2\right)\left(1 - r_{BC}^2\right)}} \tag{3.6}$$

where r_{ij} is the correlation coefficient between bands i and j. When there are four bands A, B, C and D, the partial correlation coefficient is given by

$$\rho_{AB\cdot CD} = \frac{\rho_{AB\cdot D} - \rho_{AC\cdot D}\rho_{BC\cdot D}}{\sqrt{\left(1-\rho^2_{AC\cdot D}\right)\left(1-\rho^2_{BC\cdot D}\right)}} \tag{3.7}$$

and for N bands this generalizes to (Roger 1994b)

$$\rho_{12\cdot 34\ldots N} = \frac{\rho_{12\cdot 4\ldots N} - \rho_{13\cdot 4\ldots N}\rho_{23\cdot 4\ldots N}}{\sqrt{\left(1-\rho^2_{13\cdot 4\ldots N}\right)\left(1-\rho^2_{23\cdot 4\ldots N}\right)}}. \tag{3.8}$$

So the inverse covariance matrix can be written as

$$\mathbf{C}^{-1} = \mathbf{UPU} \tag{3.9}$$

where **P** is the partial correlation matrix and **U** is a diagonal matrix whose elements are the square roots of the elements of the main diagonal of the inverse covariance matrix. The elements in **U** can be thought of as representing the reciprocal of standard deviations of the residuals or 'unexplained standard deviations' (Roger 1996b).

This allows the quadratic term in (3.1) to be written as

$$Q = (\mathbf{U}_k\mathbf{d})^T\mathbf{P}_k(\mathbf{U}_k\mathbf{d}). \tag{3.10}$$

With hyperspectral data, **P** typically takes on a sparse band-diagonal form, and can be approximated by the identity matrix. This gives

$$Q = (\mathbf{U}_k\mathbf{d})^T(\mathbf{U}_k\mathbf{d}), \tag{3.11}$$

which has been termed as the 'residual-scaled distance to the class mean' (Roger 1996b). Because $\mathbf{U}_k$ is diagonal, the processing time is significantly reduced, as was the case with the standardized distance to the mean classifier.

3.3 Maximum Likelihood Classification Assuming Block-Diagonal Correlation Matrices

The correlation matrices for hyperspectral remotely sensed data typically exhibit highly correlated blocks along the main diagonals and blocks containing lower correlations at a distance from the main diagonal. Jia and Richards (1994) have recognised this and have modified the MLC methodology to exploit this observation and achieve faster classification times. Their method assumes that hyperspectral correlation matrices have a band-diagonal form. The global correlation matrix is examined for boundaries between highly correlated blocks and the complete set of bands is divided into subgroups of these highly correlated blocks. It is possible for highly correlated blocks to appear off the main diagonals. If this should occur, the bands need to be reordered so that these highly correlated regions are moved closer to the main diagonal of the correlation matrix. When classifying classes containing a small number of pixels, this new approach is more effective than the standard MLC as the number of training pixels required only depends on the number of bands used in the *largest subgroup*.

If the $\mathbf{x} - \overline{\mathbf{x}}_k$ terms in (3.1) are substituted as $\mathbf{d}$ and $\mathbf{d}$ is then partitioned into H subvectors corresponding to the H chosen subgroups, $\mathbf{d}$ can be represented as

$$\mathbf{d} = [\mathbf{d}_1^T, \cdots, \mathbf{d}_H^T]^T . \tag{3.12}$$

A significant reduction in the computational load can be achieved by realising that the determinant of the associated block-diagonal covariance matrix will be the product of the determinants of the individual blocks, i.e.

$$|\mathbf{C}_k| = \prod_{h=1}^{H} |\mathbf{C}_{kh}| \tag{3.13}$$

which leads to

$$\ln |\mathbf{C}_k| = \sum_{h=1}^{H} \ln |\mathbf{C}_{kh}| . \tag{3.14}$$

Also, the inverse of a block-diagonal matrix can be calculated by inverting each constituent matrix. The quadratic term in (3.1) can then be written as

$$\mathbf{d}^T \mathbf{C}_k^{-1} \mathbf{d} = \sum_{h=1}^{H} \mathbf{d}_h^T \mathbf{C}_{kh}^{-1} \mathbf{d}_h \ . \tag{3.15}$$

The new MLC discriminant function that exploits the typical block-diagonal nature of correlation matrices to save on the computation time is given by

$$g_k(x) = -\sum_{h=1}^{H} \left\{ \ln \left| \mathbf{C}_{kh} \right| + \left(\mathbf{x} - \overline{\mathbf{x}}_k \right)^T \mathbf{C}_{kh}^{-1} \left(\mathbf{x} - \overline{\mathbf{x}}_k \right) \right\}, \ k = 1 \cdots G, \ h = 1 \cdots H \tag{3.16}$$

where G is the number of classes. An alternative to subdividing the original set of bands into groups of highly correlated blocks is to divide the original set of bands into regular sized blocks of, say, 2 × 2 or 3 × 3 bands.

Jia and Richards (1994) have tested the block-diagonal technique on two poorly separated classes and compared the results with those from the conventional MLC approach. The data originated in an AVIRIS scene measured over the Jasper Ridge Biological Preserve in California. The conventional method was 85% accurate whereas the new method achieved an accuracy of 89% and only required 20% of the processing time.

In a second experiment both variable and uniform sized blocks were used as well as the minimum distance classifier. Each was used for the classification of two AVIRIS scenes — one from Jasper Ridge, California and the other from Moffett Field, California. Seven classes were identified in the Jasper Ridge scene and 9 classes were identified in the Moffett Field scene. Using variable sized highly correlated blocks, the overall classification accuracy for the Jasper Ridge and Moffett Field scenes were 96% and 100% respectively. For uniform block sizes of 3 × 3, 2 × 2, and 1 × 1 the overall accuracy for the Jasper Ridge and Moffett Field scenes were 94% and 99% respectively. The minimum distance classifier performed the poorest of the classifiers, only achieving accuracies of 91% and 82% for the Jasper Ridge and Moffett Field scenes, respectively. In terms of the classification time needed

relative to the conventional MLC approach, the variable sized block classification required only 23 – 30% of the time and the other methods (i.e., minimum distance classification and uniform block size classification) only required 3 – 4% of the time (Jia & Richards 1994).

3.4 Maximum Likelihood Classification Assuming Band-diagonal Partial Correlation Matrices

The inverse covariance matrix can be written in terms of the partial correlation matrix, as shown in section 3.2 (3.9). The partial correlation matrix has a similar form to the correlation matrix, in that it is symmetrical, has values of 1 along the main diagonal and all other elements have values between ± 1. If the correlation matrix is thought of as a scaled version of the covariance matrix, the partial correlation matrix can be regarded as a scaled version of the inverse covariance matrix. Dividing the elements in the covariance and inverse covariance matrices by the square roots of the corresponding diagonal elements will result in the correlation and partial correlation matrices respectively.

Roger (1996b) notes that with hyperspectral data, both the inverse covariance and partial correlation matrices tend to take on sparse, band-diagonal forms and demonstrates how this can be used to reduce the number of parameters required for MLC. If the approximate bandwidth of the partial correlation matrix is determined, all off-band elements can be set to zero. So in the case of a partial correlation matrix having a bandwidth of 3, the partial correlation matrix can be approximated by

$$\mathbf{P} \approx \begin{bmatrix} 1 & p_{12} & p_{13} & 0 & & & 0 \\ p_{12} & 1 & p_{23} & p_{24} & 0 & & \\ p_{13} & p_{23} & 1 & p_{34} & p_{35} & 0 & \\ & & & \ddots & & & \vdots \\ & 0 & p_{N-4,N-2} & p_{N-3,N-2} & 1 & p_{N-2,N-1} & p_{N-2,N} \\ & & 0 & p_{N-3,N-1} & p_{N-2,N-1} & 1 & p_{N-1,N} \\ 0 & & & 0 & p_{N-2,N} & p_{N-1,N} & 1 \end{bmatrix} \tag{3.17}$$

This approximation may not result in positive definite matrices, in which case Q would take on negative values. One way to avoid this problem is to find the Cholesky factor (Watkins 1991) of the matrix **P** and then set the off band terms to zero. The approximation for **P** is then obtained by multiplying the band-diagonal Cholesky factor by its transpose.

The Cholesky decomposition theorem (Watkins 1991) states that any positive definite matrix can be expressed as the product of a lower triangular matrix with its (upper triangular) transpose. This lower triangular matrix is known as the Cholesky factor. Alternatively, the product of any lower triangular matrix with its transpose will result in a positive definite matrix. The Cholesky factor will also preserve the determinant of the original matrix. The determinant of any triangular matrix will always be the product of the diagonal elements (Fraleigh & Beauregard 1990, Schwarz, Rutishauser & Stiefel 1973). Therefore, if the off-diagonal terms of the lower triangular matrix are set to zero, the determinant will not be affected.

3.5 Algorithm for Fitting Covariance Selection Models Based on Simple Graphs

The covariance matrix (and its inverse) can be thought of as being a set of $\frac{1}{2}N(N+1)$ parameters. By setting elements in the inverse to zero, the number of parameters can be reduced and the covariance structure simplified. This is convenient, as the inverse covariance matrices associated with hyperspectral data typically take on a sparse form. The idea of covariance selection was introduced by Dempster (1972) and has since been studied by Wermuth (1976a, b), Wermuth and Scheidt (1977) and Speed and Kiiveri (1986).

Speed and Kiiveri (1986) have developed two similar algorithms that allow elements in the inverse covariance matrix to be set to zero, yet, unlike the method suggested by Roger (1996b), are not limited by the assumption that the inverse must take on a band-diagonal form. The methods link the elements of the inverse covariance matrix with a simple, undirected graph, where non-zero elements are denoted as pairs of adjacent nodes. Once the elements to be set to zero have been determined, maximal sets of vertices where every pair is adjacent (known as cliques) are identified. These

cliques are then sorted such that each successive clique contains a node not contained in any of the previous cliques. The algorithms then proceed to cycle through each of the cliques, constructing the maximum likelihood estimate of the covariance matrix from the sample covariance matrix. This new estimate of the covariance matrix will have the property that elements in the positions of the non-zero elements in the inverse will agree with those in the sample covariance matrix and all other elements in the estimated inverse are set to zero.

The first algorithm in Speed and Kiiveri (1986) is a backward selection procedure and begins with the sample inverse covariance matrix. As it cycles through the cliques of the associated complementary graph, off-diagonal terms are set to zero. Starting with $\hat{\mathbf{\Sigma}}_0 = \mathbf{C}$, if f is any non-negative integer, H is the number of cliques and h is the clique number, a sequence of estimates $\hat{\mathbf{\Sigma}}_{fH+h} = \mathbf{Z}_h(\hat{\mathbf{\Sigma}}_{fH+h-1})$ is generated using the equation

$$\mathbf{W}^{-1} = \mathbf{X}^{-1} + \begin{bmatrix} \mathbf{B}_a{}^{-1} - \mathbf{X}_a{}^{-1} & 0 \\ 0 & 0 \end{bmatrix} \tag{3.18}$$

where $\mathbf{Z}_h(\hat{\mathbf{\Sigma}}) = \mathbf{W}^{-1}$, $\mathbf{X}^{-1} = \hat{\mathbf{\Sigma}}$, a denotes the set of bands in the h^{th} clique and $\mathbf{B}_a{}^{-1}$ is the diagonal matrix formed by the diagonal elements of $(\hat{\mathbf{\Sigma}}^{-1})_a{}^{-1}$.

The second algorithm is a forward selection procedure and starts off with the identity matrix, ie, with all off-diagonal terms in the inverse covariance matrix set to zero. As it cycles through the cliques, the covariance matrix is constructed by adding in the non-zero terms. Starting with $\hat{\mathbf{\Sigma}}_0$ as the identity matrix, a sequence of estimates $\hat{\mathbf{\Sigma}}_{fH+h} = \mathbf{Y}_h(\hat{\mathbf{\Sigma}}_{fH+h-1})$ is generated using the equation

$$\mathbf{W} = \begin{bmatrix} \mathbf{B}_a & \mathbf{B}_a\mathbf{X}_a{}^{-1}\mathbf{X}_{a,a'} \\ \mathbf{X}_{a',a}\mathbf{X}_a{}^{-1}\mathbf{B}_a & \mathbf{X}_{a'} - \mathbf{X}_{a',a}\mathbf{X}_a{}^{-1}(\mathbf{I} - \mathbf{B}_a\mathbf{X}_a{}^{-1})\mathbf{X}_{a,a'} \end{bmatrix} \tag{3.19}$$

where $\mathbf{Y}_h(\hat{\mathbf{\Sigma}}) = \mathbf{W}$, $\mathbf{X} = \hat{\mathbf{\Sigma}}$, a' denotes the complete set of bands not contained in the h^{th} clique and $\mathbf{B} = \mathbf{C}$. If the cliques are ordered properly, the second algorithm will

converge after one cycle. In order for the cliques to be ordered properly, the graph must be triangulated. This means that all cycles in the graph must contain a chord. A cycle is a sequence of (≥ 4) nodes where each successive node is connected to the previous one and the first and last nodes are also connected.

The choice of which algorithm to use is influenced by the number and size of cliques in both the graph and the complementary graph. However, because the first algorithm initially requires **C** to be inverted and the second algorithm converges after one cycle, the second algorithm is more likely to be chosen.

4 Overlapping Spectral Bands — A Source for the Correlation Between Hyperspectral Bands

One of the reasons why measured spectra approach smooth curves is due to the close spacing between the large number of spectral bands. Targets measured by remote sensing instruments tend to have broad spectral features and when measured by hyperspectral sensors, these features span across many adjacent spectral channels. These spectral channels are capable of detecting small variations in the level of the reflected/emitted signal, and hence the resulting spectra can resemble smooth curves.

Another contributing factor to the relative smoothness of the measured spectral curves is due to the fact that the spectral bands have filter functions that are much wider than the spectral band separations. As a result the spectral bands will overlap with neighbouring spectral bands. The amount of overlap becomes significant with adjacent spectral bands. The affect of this on spectral measurements is to smooth the spectra and increase the level of correlation between neighbouring bands. Clearly, the level of overlap between spectral bands will be a function of the width and separation of the bands.

If the central wavelengths and Full Width at Half Maximum values are known for each spectral band, the spectral band filter functions can be modelled as rectangular, triangular or Gaussian functions. For HyMap spectral filter functions, Jupp (2001) reports that they are said to be well approximated by triangular functions. When modelling spectral filter functions, it is important to ensure that each filter function has a constant area of 1. The peak value for each spectral band filter function will be dependant on the FWHM value.

In this chapter, the HyMap spectral band filter functions will be modelled as triangular functions and the amount of overlap between adjacent bands will be estimated. In section two, the band filter functions are then modelled as Gaussian functions and again the overlap between adjacent bands will be assessed. The filter functions were not available for the 128 band HyMap instrument that acquired the data used in this research, only the band central wavelengths and the FWHM values

were known. For comparison with actual measured filter function data, the spectral bands on a more recent 126 band HyMap instrument will be modelled as both triangular and Gaussian functions in section three and the overlaps between neighbouring spectral bands will be assessed. These calculations will then be compared with the measured filter functions for the more recent HyMap instrument. These were measured during November 2000.

4.1 Estimating the Amount of Band Overlap Using Triangular Filter Functions

In this section, the HyMap spectral band filter functions will be modelled as triangular functions, as suggested by Jupp (2001). For spectral band *i*, given the band central wavelength (λ_{ic}) and the FWHM ($\lambda_{i½}$) values, the peak value of the filter function, (ϕ_{ip}), is calculated to give the area under the curve is equal to one. This will be ensured if ϕ_{ip} is given by:

$$\phi_{ip} = \frac{1}{\lambda_{i½}} \tag{4.1}$$

The left hand edge of the triangular filter function can be modelled as a linear equation with respect to the wavelength, λ, and is given by:

$$\phi_{il}(\lambda) = \frac{\lambda + \lambda_{i½} - \lambda_{ic}}{\lambda_{i½}^{\,2}} \tag{4.2}$$

Similarly, the right hand edge is given by the equation:

$$\phi_{ir}(\lambda) = \frac{-\lambda + \lambda_{i½} + \lambda_{ic}}{\lambda_{i½}^{\,2}} \tag{4.3}$$

For two neighbouring bands *i* and *j*, with central wavelengths of λ_{ic} and λ_{jc} respectively, and FWHM values of $\lambda_{i½}$ and $\lambda_{j½}$ respectively, their spectral filter functions will intersect at a wavelength given by:

$$\lambda_x = \frac{\lambda_{i½}\lambda_{j½}^{\,2} + \lambda_{ic}\lambda_{j½}^{\,2} + \lambda_{jc}\lambda_{i½}^{\,2} - \lambda_{j½}\lambda_{i½}^{\,2}}{\lambda_{i½}^{\,2} + \lambda_{j½}^{\,2}} \tag{4.4}$$

At this wavelength the filter functions will have a value of:

$$\phi_x = \frac{\lambda_{i½}\lambda_{j½}^{\,2} + \lambda_{ic}\lambda_{j½}^{\,2} + \lambda_{jc}\lambda_{i½}^{\,2} - \lambda_{j½}\lambda_{i½}^{\,2}}{\left(\lambda_{i½}^{\,2} + \lambda_{j½}^{\,2}\right)\lambda_{j½}^{\,2}} + \frac{\lambda_{j½} - \lambda_{jc}}{\lambda_{j½}^{\,2}} \tag{4.5}$$

With the coordinates of where the two neighbouring bands intersect being known, it then becomes possible to calculate the amount of overlap between the two bands. The overlapping region will form a triangle with a height found in (4.5) and a base width given by $(\lambda_{ic} + \lambda_{i½}) - (\lambda_{jc} - \lambda_{j½})$. The area of the overlapping region will therefore be given by:

$$Area = ½\left[\left(\lambda_{ic} + \lambda_{i½}\right) - \left(\lambda_{jc} - \lambda_{j½}\right)\right]\left[\frac{\lambda_{i½}\lambda_{j½}^{\,2} + \lambda_{ic}\lambda_{j½}^{\,2} + \lambda_{jc}\lambda_{i½}^{\,2} - \lambda_{j½}\lambda_{i½}^{\,2}}{\left(\lambda_{i½}^{\,2} + \lambda_{j½}^{\,2}\right)\lambda_{j½}^{\,2}} + \frac{\lambda_{j½} - \lambda_{jc}}{\lambda_{j½}^{\,2}}\right] \tag{4.6}$$

Because the filter functions all have a total area of one, the area calculated in (4.6) will also be the proportion of the total area.

Most spectral bands will overlap with the neighbouring spectral bands on both sides. For example, band i will overlap on the left with band $i - 1$ and on the right with band $i + 1$. It is not uncommon for bands $i - 1$ and $i + 1$ to also overlap with each other. When this occurs, it should be noted that the total overlap of a band i with its neighbours is not the sum of the overlap with the left hand neighbour $i - 1$ and the overlap with the right hand neighbour $i + 1$. In this case, the overlapping region between bands $i - 1$ and $i + 1$ needs to be subtracted.

The spectral bands in the HyMap instrument used to acquire the data for this research were modelled as triangular functions from their central wavelengths and their FWHM values. The results are summarised in the plot of the fractional spectral filter function overlaps versus the spectral band central wavelengths in figure 4.1. The

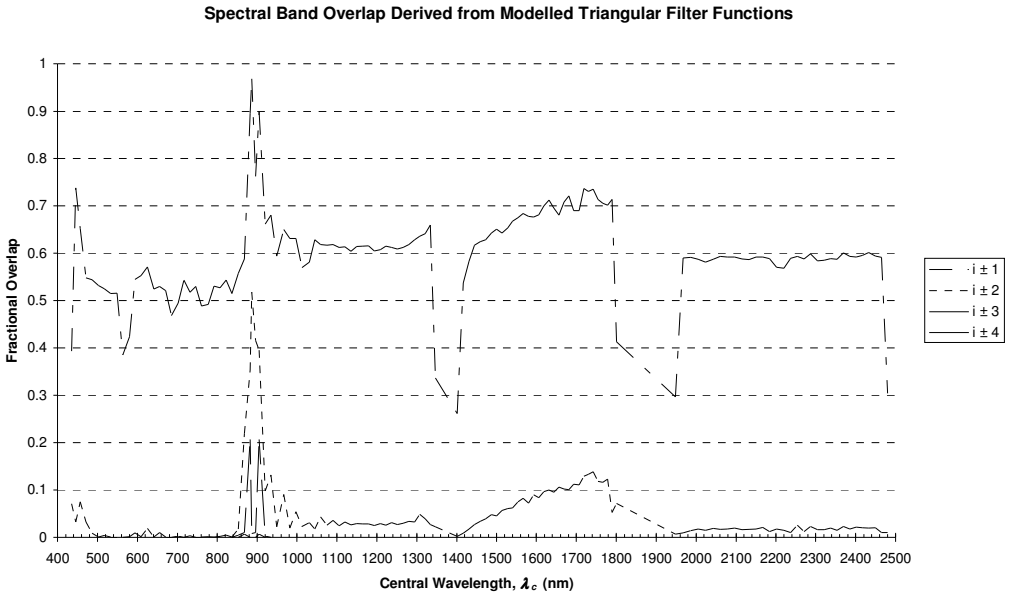


Figure 4.1: Fractional overlap of spectral band filter functions with neighbouring spectral bands when the filter functions are modelled as triangular functions. The plot shows the total overlap between a spectral band i and its neighbours (i ± 1), the overlap with bands that are ± 2 bands away (i ± 2), the overlap with those that are ± 3 bands away (i ± 3) and the overlap between a spectral band and those that are ± 4 bands away (i ± 4)

plot shows that the average value of the total fractional filter function overlap is 0.6, or 60%.

The spectral bands on HyMap are arranged in banks of four modules, with each module having 32 spectral bands. The first module spans the spectral range between 450 nm – 890 nm, with the bands having an average spacing of 15 nm. The second module spans 890 nm – 1350 nm, also with an average spacing of 15 nm between bands. The third module spans 1400 nm – 1800 nm and the fourth module spans 1950 nm – 2480 nm. The average band spacing of these modules is 13 nm and 17 nm respectively. The location of the spectral modules is evident in figure 4.1, where the region between modules 3 and 4 at 1800 nm – 1950 nm can be seen as a reduction in the total fractional overlap between neighbouring spectral bands. A similar effect can also be seen between 1350 nm and 1400 nm, corresponding to the region between the second and third modules. The opposite effect is evident at 890 nm and this is due to the fact that there is no separation between modules one and two. As a result, the spectral bands in this region are much closer and hence the amount of spectral overlap is significantly increased.

In modules 2 and 4, figure 4.1 shows that the average total spectral overlap is approximately 0.6. For module 3, the total spectral overlap has a value ranging from 0.6 – 0.74. In the first module the total overlap between spectral bands tends to be between 0.38 – 0.74. These are large fractional overlap values and an argument could be made that the number of spectral bands could be reduced by increasing the average spacing between them for only a small reduction in information content.

When looking at a spectral band and those that are separated by ± 2 bands, the amount of spectral overlap between them decreases significantly. This can be seen in figure 4.1 (i ± 2). The average overlap between a band i and its neighbours $i \pm 2$ reduces to 0.05. Excluding the region where module 1 and 2 overlap, module 3 exhibits the highest level of overlap, with the values peaking at 0.14. Spectral bands in module 1 show the lowest levels of overlap with their ($i \pm 2$) neighbours, with the values typically being approximately 0.003, but getting as high as 0.075. When modelled as triangular functions most spectral bands have an overlap of zero with their ($i \pm 3$) and ($i \pm 4$) neighbours, except for a few bands where the first and second modules overlap.

4.2 Estimating the Amount of Band Overlap Using Gaussian Filter Functions

Spectral filter functions can also be modelled as functions having a normal or Gaussian distribution. If a Gaussian probability distribution function is used, the spectral filter function for band i can be modelled by

$$\phi_i(\lambda) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-(\lambda-\lambda_{ic})^2/2\sigma_i^2} \tag{4.7}$$

where σ_i is the standard deviation about the central wavelength of the filter function and can be related to the FWHM. At wavelengths of ($\lambda_{ic} \pm \lambda_{i½}/2$), (4.7) should have a value of $\frac{1}{2\sqrt{2\pi\sigma_i^2}}$. If $\lambda = \lambda_{ic} \pm \lambda_{i½}/2$ is substituted into λ in (4.7), it can be shown that σ_i^2 is related to the FWHM by

$$\sigma_i^2 = \frac{\lambda_{i½}^2}{8\ln(2)}. \tag{4.8}$$

To calculate the amount of overlap between two spectral bands, the wavelength at which the filter functions intersect needs to be determined. This can be found by equating the filter functions for two bands *i* and *j* and then by solving for λ in (4.7). When $\lambda_{i½}$ is not equal to $\lambda_{j½}$ this yields:

$$\lambda_x = \frac{\sigma_i^2\lambda_{jc} - \sigma_j^2\lambda_{ic} - \sqrt{\sigma_i^2\sigma_j^2\left\{(\lambda_{ic}-\lambda_{jc})^2 + 2(\ln\sigma_i - \ln\sigma_j)(\sigma_i^2 - \sigma_j^2)\right\}}}{\sigma_i^2 - \sigma_j^2} \tag{4.9}$$

If $\lambda_{i½}$ and $\lambda_{j½}$ are equal then the denominator in (4.9) will be equal to zero. When this occurs the wavelength at which the two filter functions intersect will be equal to the mean of the two central wavelengths of the two filter functions and is given by:

$$\lambda_x = \frac{\lambda_{ic} + \lambda_{jc}}{2}. \tag{4.10}$$

Once the wavelength at which two spectral bands intersect at has been determined, the overlapping area of the two filter functions can then be determined by referring to normal distribution cumulative probability lookup tables. These are included with most scientific calculators, spreadsheet programs and mathematical software packages such as Maple or Matlab.

Given two bands *i* and *j* ($\lambda_{ic} < \lambda_{jc}$) that have been modelled as Gaussian functions and intersect at a wavelength λ, the left hand area of the intersecting region is found by using λ_{jc} and σ_j as the mean and standard deviation values to look up the corresponding cumulative probability at a value of λ. This cumulative probability is the area of the left hand region. The area of the right hand region is found by using λ_{ic} and σ_i as the mean and standard deviation values to look up the corresponding cumulative probability at a value of λ. This cumulative probability then needs to be subtracted from unity to give the area of the right hand region. The total area of the

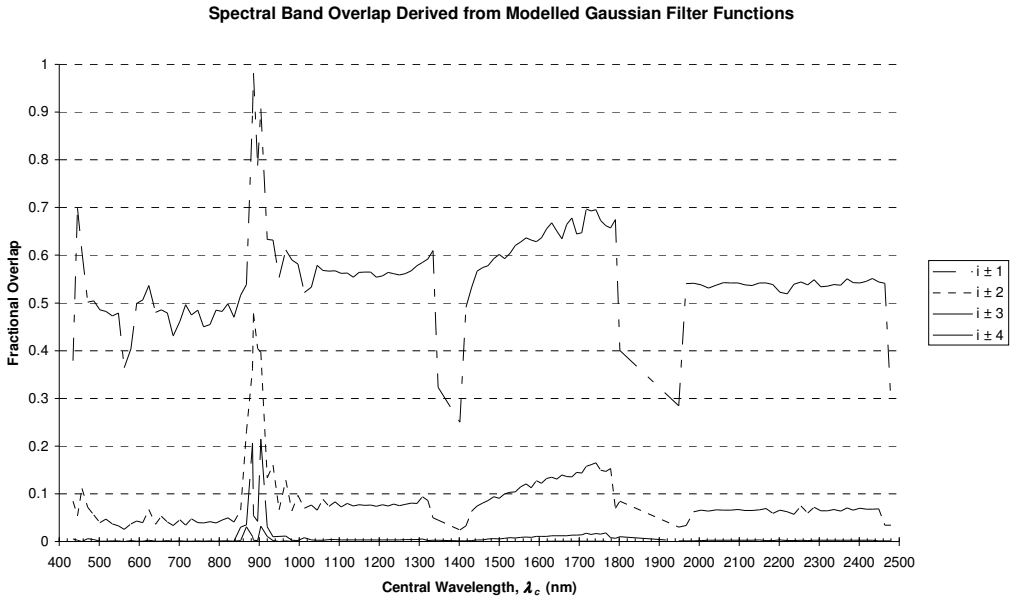


Figure 4.2: Fractional overlap with neighbouring spectral bands when modelled as Gaussian functions. The plot shows the total overlap between a spectral band and its neighbours (i ± 1), the overlap with bands that are ± 2 bands away (i ± 2), the overlap with those that are ± 3 bands away (i ± 3) and the overlap between a spectral band and those that are ± 4 bands away (i ± 4)

overlapping region is then simply the sum of the areas of the left and right hand regions.

The spectral bands on the HyMap instrument used to acquire the data for this research have been modelled as Gaussian functions and for each band *i*, the total spectral overlap has been calculated. The overlap with bands $i \pm 2$, $i \pm 3$ and $i \pm 4$ has also been calculated. The results from this are shown in figure 4.2, and appear to be very similar to those in figure 4.1. When modelled as Gaussian functions, the spectral filter functions have an average total overlap of 0.56. This is slightly less than the value of 0.60 calculated when the filter functions were modelled as triangular functions.

When examining the results for the amount of overlap between bands *i* and $i \pm 2$, the Gaussian modelled filter functions show an average overlap of 0.09. This is greater than the value of 0.05 calculated for the overlap when the filter functions were modelled as triangular functions. This is to be expected as triangular functions are limited to a width that is twice the FWHM value.

4.3 Comparing Modelled Filter Functions with Measured Filter Function Data

The filter functions for a more recent HyMap instrument (having only 126 spectral bands) have been measured and this data was recently made available by Dr Tom Cudahy from the Division of Exploration and Mining, CSIRO. It is now possible to do a direct comparison between the modelled filter functions and the actual measured filter function responses. From the measured filter function data the central wavelengths and FWHM values can be determined for each spectral band and hence the spectral band responses can be modelled as triangular or Gaussian functions as outlined in sections 4.1 and 4.2, respectively. It is also now possible to assess how well the modelled filter functions can predict the amount of overlap between the spectral bands.

Figure 4.3 shows, as an example, a plot of the measured filter function responses for bands 75 – 82 of the more recent HyMap instrument. Also shown are plots of the same spectral band filter functions when modelled as Gaussian and triangular functions. It can be seen that there is fairly good agreement between the measured and modelled data. One noticeable difference between the measured and modelled

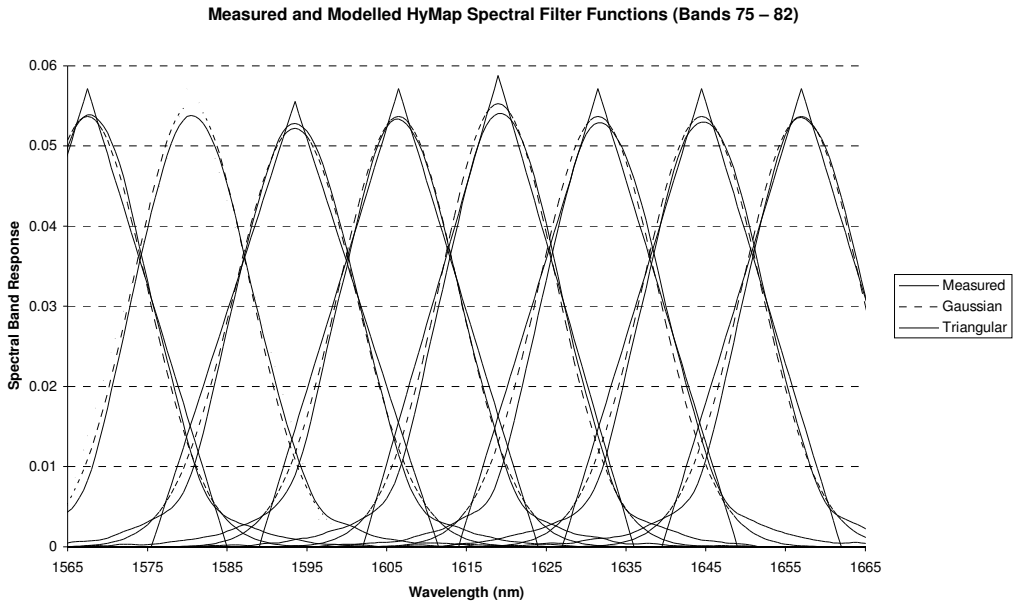


Figure 4.3: Measured spectral band responses for bands 75 – 82 of the more recent 126 band HyMap instrument (solid lines). Also shown are the responses modelled as Gaussian functions (dotted lines) and triangular functions (dashed lines)

data is that the modelled data tend to underestimate the tails and hence the overall width of the filter functions. Another difference is that the modelled triangular functions overestimate the peak response value of the actual filter functions. The higher estimated peak response values can be explained, as the modelled triangular functions only have a finite width of 2 × FWHM and underestimate the tails of the measured filter function responses.

Root Mean Square (RMS) errors have been calculated for each HyMap spectral band filter function when they were modelled as triangular and Gaussian functions. The RMS errors are plotted in figure 4.4. The Gaussian modelled functions perform consistently better than the triangular modelled functions. The mean RMS error for the Gaussian modelled functions was 0.00169 and the triangular modelled functions had a mean RMS error of 0.00238. This is a difference in the RMS error of 0.00069 between the Gaussian and triangular modelled spectral filter functions.

With the measured filter function data being available in digital form, the amount of overlap between neighbouring bands can be calculated as follows. Given two overlapping spectral bands i and j (where it is assumed $\lambda_{ic} < \lambda_{jc}$) having response functions ϕ_i and ϕ_j, respectively, the wavelength, λ_x, at which their two response

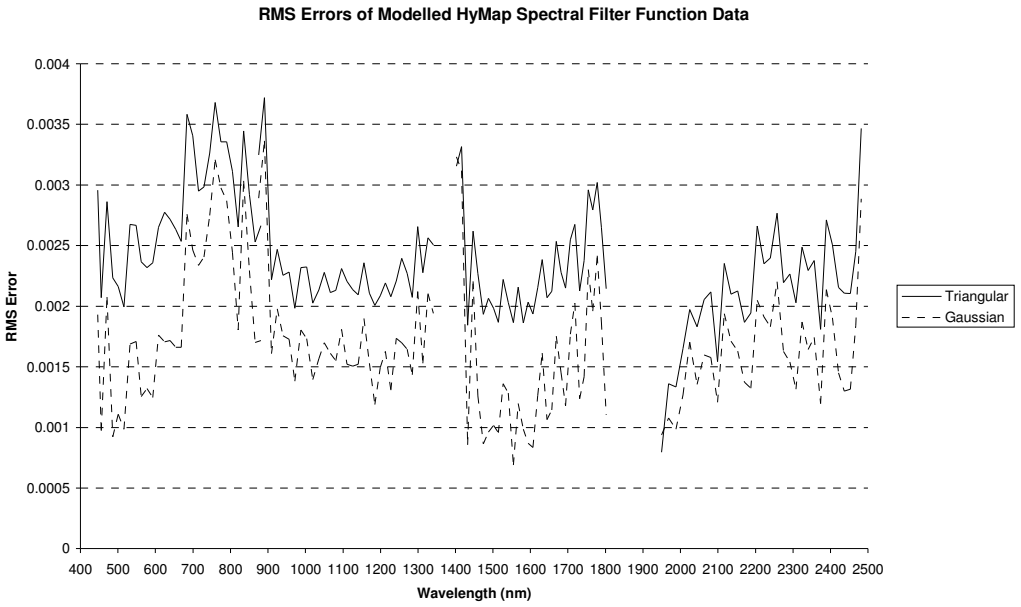


Figure 4.4: RMS errors of HyMap spectral band filter functions when modelled as triangular and Gaussian functions.

functions intersect needs to be identified. For values of $\lambda \leq \lambda_x$, the sum of values for ϕ_j needs to be found and this is added to the sum of values of ϕ_i for $\lambda > \lambda_x$. This total then needs to be multiplied by the wavelength sampling interval of the measured filter function data to obtain the total overlapping area.

The areas of the overlapping regions between neighbouring spectral filter functions have been calculated for the measured data. Figure 4.5 shows a plot of the total spectral overlap for each band as calculated from the measured filter function data. Excluding those bands that are near the edges of the spectral modules, the total fractional overlap between the bands ranges between 0.44 and 0.71. Module 3 exhibits the highest amount of overlap between bands, with an average value of 0.68, and module 1 shows the lowest, with an average value of 0.47. When comparing with the plots of total overlap values predicted by the modelled functions (see figure 4.1 and figure 4.2) for the 128 band HyMap instrument that was used to acquire the data for this research, the overlaps calculated from the measured data appear very similar.

Figure 4.5 also shows the total overlap for each band of the 126 band HyMap

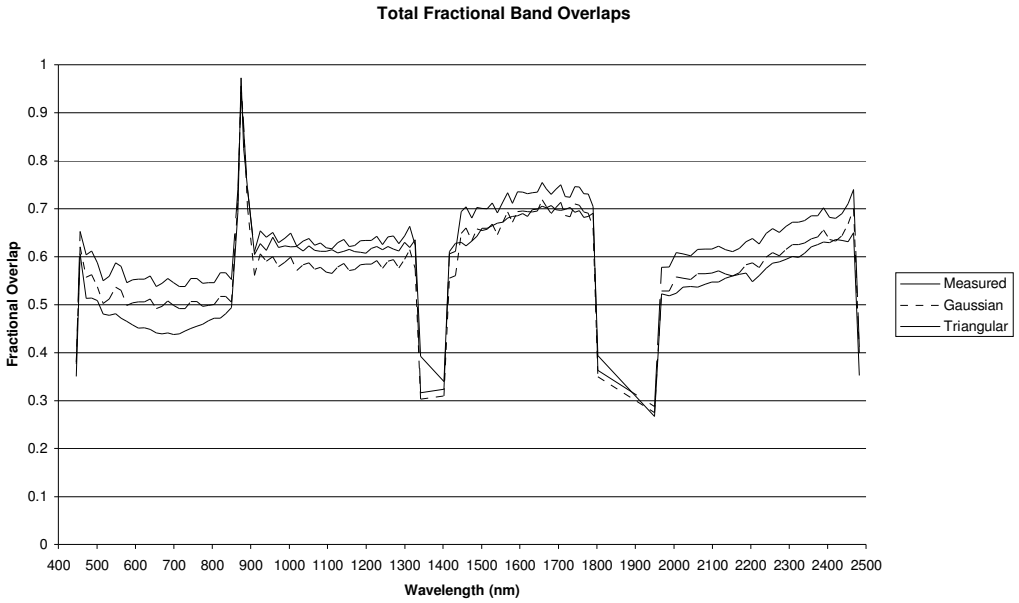


Figure 4.5: Total fractional overlaps for the spectral bands on the more recent 126 band HyMap instrument. *Note*: Shown are the values calculated from the measured response data (solid line), the values predicted from Gaussian-modelled filter function data (dotted line) and the values predicted from the triangular-modelled filter function data (dashed line)

instrument as predicted by the modelled filter function data. The plot in figure 4.5 shows that the total overlaps predicted by the modelled filter function data agree well with the measured data, particularly when the filter function data is modelled by Gaussian functions. Using Gaussian modelled filter functions the total overlap between each band is overestimated in the first and fourth module and underestimated in the second module. Very good agreement is demonstrated in the third module between the total overlaps predicted by the Gaussian modelled functions and the actual total overlaps as calculated from the measured data. When modelled as triangular functions, the calculated total overlap between each band is overestimated in all four modules.

Figure 4.6 shows the fractional overlaps between bands i and $i \pm 2$ as calculated from the measured filter functions and also for those modelled from Gaussian and triangular functions. The modelled data captures the trends of the overlaps well when compared with the overlaps that were calculated from the measured data. In the second module though, the measured data does give values higher than those predicted by the modelled filter functions. The overlap values calculated from the measured data in the second module average approximately 0.19 compared with an average value of 0.09 for the Gaussian modelled filter functions and 0.04 for the triangular modelled filter functions. As the distance between neighbouring spectral bands increases the triangular modelled filter functions will consistently underestimate the fractional overlap between the two bands. This is because of the finite width of the triangular functions when compared with the overall width of the measured filter responses and the Gaussian modelled functions.

When the fractional overlaps between bands i and $i \pm 3$ were calculated from the measured spectral band responses, the results showed negligible values in modules 1 and 4 and in modules 2 and 3 the average fractional overlaps were found to be 0.08 and 0.06 respectively. The Gaussian modelled responses showed negligible overlaps in modules 1, 2 and 4, and in module 3 the average fractional overlap was found to be less than 0.02. The overlaps calculated from the triangular modelled responses were all found to be zero except for the region where module 1 overlaps with module 2. Here, the maximum overlap between the triangular functions was 0.09.

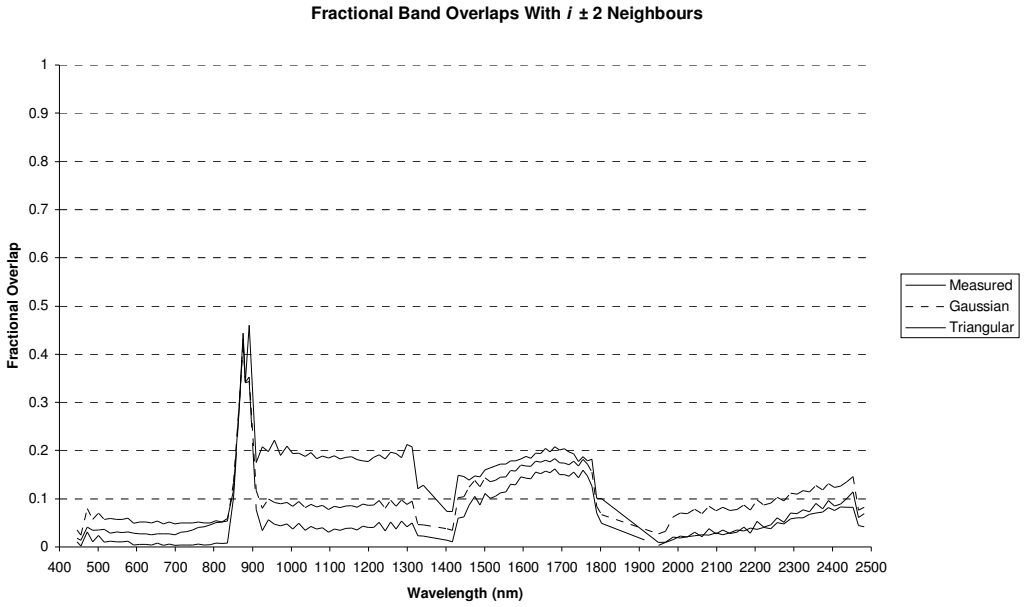


Figure 4.6: Fractional overlaps between bands *i* and *i* ± 2 for the spectral bands on the more recent 126 band HyMap instrument. Shown are the values calculated from the measured response data (solid line), the values predicted from Gaussian-modelled filter function data (dotted line) and the values predicted from the triangular-modelled filter function data (dashed line)

It has been shown in this chapter that the amount of overlap between neighbouring bands on hyperspectral sensors can be considerable. In the example shown here, the measured filter functions from a 126 band HyMap instrument exhibited total fractional overlaps for each band averaging approximately 0.6 and being as high as 0.7. This high level of overlap between the spectral band response functions will have a significant contribution to the correlations between neighbouring bands, as the neighbouring bands are not completely unique and are measuring part of the same data. As the separation between bands increases, the amount of overlapping reduces. When examining the overlap between a band *i* and bands *i* ± 2, the average amount of overlap reduced to approximately 0.1.

With regard to being able to model the spectral band filter functions, the results presented here are very encouraging. The results indicate that the filter functions for spectral bands on hyperspectral instruments such as HyMap can be well approximated by using either triangular or Gaussian functions. Modelling of the filter functions may be required if the actual filter function data for a hyperspectral remote sensing instrument are not available, as is the case with this research. Gaussian functions were found to perform better than triangular functions as triangular functions are restricted because their overall width is limited. The total

width of a triangular function is $4\lambda_{½}$, whereas the measured response functions typically had overall widths that were much wider than this. Gaussian functions are much more suited to capturing this characteristic than are triangular functions. In the example presented here, Gaussian functions achieved an average RMS error of 0.0017. The average RMS error when triangular functions were used was 0.0024.

As a result of being able to accurately model the spectral band responses, it also proved possible to approximate the total amount of overlap for each spectral band, particularly when Gaussian functions were used to model the band responses and when the spectral bands were closer together. The results from calculating the total overlapping areas between neighbouring filter functions from the measured and modelled data were very similar. The results also implied that a different HyMap instrument with a similar configuration of spectral bands would have similar levels of overlap between the spectral bands.

5 Maximum Likelihood Classification Using Parameter-Reduced Covariance Matrices

In Chapter 3, the idea was introduced that the (inverse) covariance matrix could be thought of as a set of ½$N(N + 1)$ parameters, where N is the dimensionality of the data. Methods were presented which could approximate the covariance matrix by setting a high percentage of the parameters in the inverse covariance matrix to zero and hence simplify the covariance structure. In this chapter some of the methods that were presented in Chapter 3 are tested on remotely sensed, hyperspectral data in an attempt to significantly reduce the number of parameters needed to describe the covariance information while still achieving comparable accuracies in the application of MLC. The covariance matrix is an important and significant component of the MLC discriminant function and has a considerable influence on the computational load of MLC particularly with hyperspectral data.

The covariance matrices are initially modelled by approximating the partial correlation matrices as band-diagonal matrices. The assumption that the partial correlation matrix takes on a band-diagonal form is then relaxed and a method that identifies low valued terms in the partial correlation matrix using thresholds is used instead. The remaining terms having absolute values greater than the threshold are used to model the covariance matrix using an algorithm in Speed and Kiiveri (1986). The modelled (inverse) covariance matrices will then be used for MLC of the data by replacing the full covariance matrix in the MLC discriminant function (3.1) with the modelled covariance matrices. The data can then be classified and the accuracy of the resulting classified images will be assessed and compared.

5.1 The Data

The scene used in this work is a 512 × 512 pixel subset of a 128-band HyMap image acquired over Toolibin, Western Australia, and is shown as a false colour RGB composite in figure 5.1. Toolibin, located at 32°57’ S, 117°37’ E, is in the south-west corner of Western Australia approximately 200 km south of Perth and

Figure 5.1: A false colour RGB (bands 111, 31 and 9, respectively) composite of the HyMap scene used for this work. The data was acquired over Toolibin in Western Australia on 20/11/1998.

approximately 40 km east of Narrogin. The Toolibin scene (TB3) was acquired during clear conditions on the afternoon of November 20, 1998 (Newnham, Renzullo & Lynch 1999) and has been atmospherically corrected for atmospheric absorption and scattering using ATREM (Dunne 1999). For this mission, the HyMap instrument was flown on a twin engine Cessna Titan at an altitude to produce data having pixel resolutions of 3.5 m at nadir. This resolution was due to a special request by CSIRO Mathematical and Information Sciences (CMIS). HyMap is generally flown at an altitude to produce data with a spatial resolution of 5 m at nadir. The data was collected as part of the CSIRO Earth Observation Centre's (EOC) Hyperspectral Task campaign. The data have been recorded as 16-bit, signed integers.

HyMap has four spectrometers, with each spectrometer taking measurements in 32 spectral bands. The 128 spectral bands cover the 0.43 – 2.5 μm spectral region and have spectral resolutions varying between 13 – 23 nm. A list of all 128 spectral bands can be found in Appendix C, showing the band central wavelengths and the full width at half maximum (FWHM) values. At a solar zenith angle of 30° and a

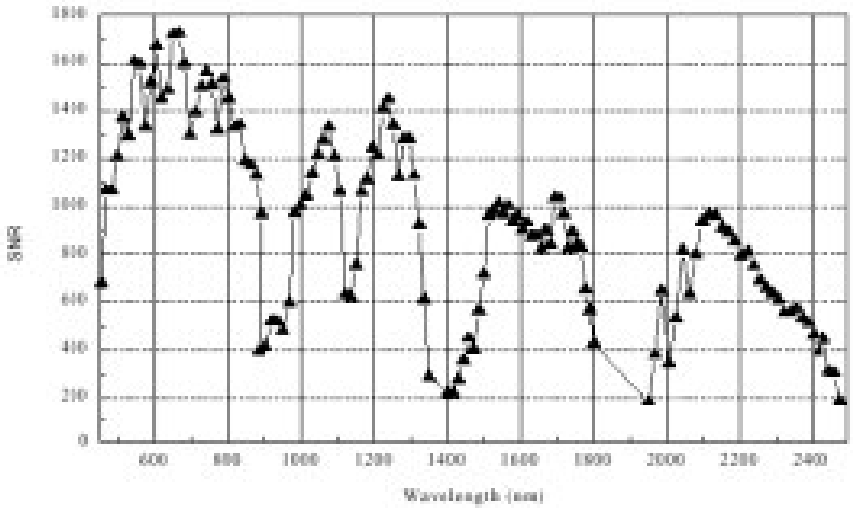


Figure 5.2: Signal-to-noise ratio for the HyMap sensor for a target having a reflectance of 0.5 when the solar zenith angle is 30° (Cocks et al. 1998).

reflectance of 0.5, the signal-to-noise ratio is greater than 500 (Cocks et al. 1998). A plot of the SNR values for each HyMap band is shown in figure 5.2.

The data were examined and of the original 128 spectral bands, 82 were retained for analysis. There were a number of factors which contributed to the other 46 bands being omitted. Several of the spectral bands are located in regions of significant atmospheric absorption and hence exhibit high noise levels and very low signal content. Other bands contained pixels with negative values. It is possible that this is an artefact from part of the atmospheric correction process, specifically, the dark current correction process. A high level of spectral overlap between bands was another reason for bands being omitted. The first two bands of the second spectrometer (bands 33 and 34) overlap with bands 31 and 32 and were removed. All other bands were examined for unusually high noise levels due to reasons other than those just stated. Those bands that were found to exhibit unusually high noise levels were also omitted.

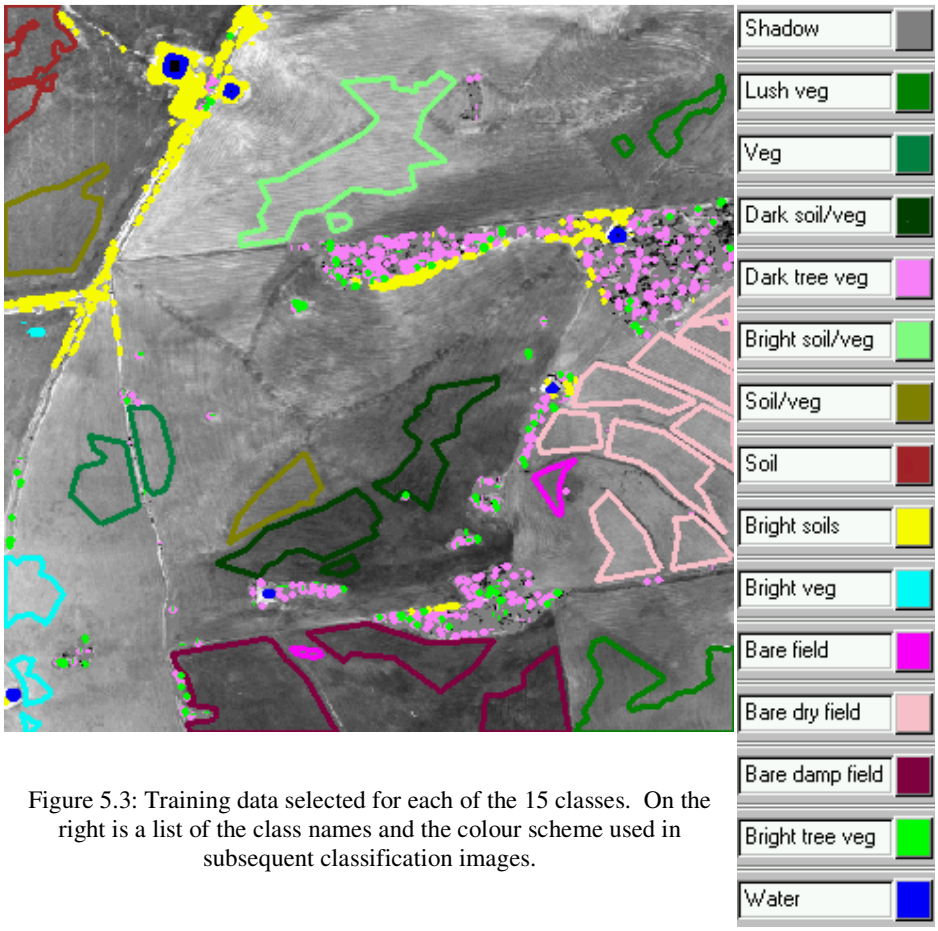


Figure 5.3: Training data selected for each of the 15 classes. On the right is a list of the class names and the colour scheme used in subsequent classification images.

The TB3 scene was examined and 15 distinct spectral classes were identified. Training data was selected for each of these classes and these were used to generate the class first and second order statistics such as the spectral means and the covariance matrices for each class. Table 5.1 contains a list of the 15 classes identified and the number of training samples used for each class. The covariance matrices were then used to generate the correlation, inverse covariance, and partial correlation matrices. Both the global and class correlation matrices are shown in Appendix E and show regions of highly correlated blocks along the main diagonals as is typical for hyperspectral data. With these matrices available, the methods for covariance matrix parameter reduction can be tested on the class covariance matrices.

Table 5.1: Number of training samples selected for each spectral class and the determinant of the covariance matrix (|C|).

Class	Number of Training Samples	|C|
Bare damp field	10028	1.3068E+148
Bare dry field	11059	5.1608E+153
Bare field	475	2.1823E+145
Bright soil/veg	6319	1.5716E+155
Bright soils	1742	3.4548E+190
Bright tree veg	307	1.1233E+166
Bright veg	1889	1.2265E+156
Dark soil/veg	5516	1.6046E+148
Dark tree veg	1282	2.2526E+161
Lush veg	5367	7.3142E+152
Shadow	1809	5.1606E+154
Soil	2384	2.5919E+150
Soil/veg	5010	7.7316E+152
Veg	3036	1.9423E+152
Water	346	7.2247E+168

Table 5.1 shows that the determinants of the class covariance matrices are enormous. This is due to the large size of the covariance matrices (82 × 82 elements) and also because the image data are 12-bit integer digital counts and not reflectance values ranging between 0 – 1. Several classes such as "Bright soils", "Bright tree veg", "Dark tree veg" and "Water" occupy relatively small regions within the image, unlike other classes which cover large areas within the image. Consequently, there were significantly fewer training pixels available to determine the class statistics for these smaller classes. These classes tend to exhibit even larger determinants for their covariance matrices. This is likely to be due to the possibility that some of the training pixels selected may not have been 100% "pure" pixels and actually contained a fraction of other classes within the pixels. As a result a higher degree of variability could have been introduced to the class statistics.

5.1.1 Significance of the Toolibin Region

The scene shown in figure 5.1 is located just south of Toolibin Lake. Toolibin is a long term research site and was selected as a primary EOC hyperspectral task field site. A priority for the area is the assessment of vegetation health, including both agricultural crops and native vegetation species. Toolibin Lake is situated at an altitude of 310 m above sea level and is one of the last wooded freshwater lakes in the southern wheatbelt and is the only wetland in the wheatbelt to have swamp sheoaks (*Casuarina obesa*) growing on the lakebed (Hooper & Wallace 1994). Other species growing on the lakebed include paperbarks (*Melaleuca strobophylla*)

and flooded gums (*Eucalyptus rudis*). The surrounding catchment has an area of 483 km^2 and farming in the catchment area is primarily for cereal crops (i.e. wheat, barley) and livestock grazing. The area surrounding the lake is vegetated with Sighing Sheoak (*Allocasuarina huegeliana*), Acorn Banksia (*Banksia prionotes*), York Gum (*Eucalyptus laxophleba*) and Jam Wattle (*Acacia acuminata*).

The vegetation on and around Toolibin Lake provides shelter for many species of waterbirds. More species of rare and protected waterbirds use the Toolibin reserve for breeding than any other wetland. Twenty four species of waterbird have been recorded breeding at Toolibin Lake and 41 species have been observed using the lake (Smith 1999). Without Toolibin Lake, many species of waterbirds may not be able to breed in the south west of Australia. Toolibin Lake is recognized as a lake of international significance under the Ramsar Convention and has been listed on the Register of the National Estate (*Toolibin Lake Recovery Plan* 1994).

Freshwater wetlands were common in the south west of Western Australia until clearing of the land began in the 1890s for agricultural purposes. This caused waterlogging and salinity, which killed many of the trees. By the 1950s and 1960s, freshwater wetlands downstream from Toolibin Lake had become highly saline and there was concern that Toolibin Lake would follow. Toolibin Lake is being threatened by salinity from rising groundwater levels and saline surface water inflows. In the 1970s, trees along the western edge of Toolibin Lake started dying and the surrounding farmland became increasingly salty and waterlogged.

In the 1980s, an effort to revegetate the surrounding land of Toolibin Lake commenced. A recovery plan was prepared in 1994, which outlined both emergency short-term actions and longer-term solutions to the salinity problem. The short-term actions are expensive to maintain, yet they are required until the longer-term solutions start to have an effect. These actions will hopefully both lower the groundwater level and significantly reduce the amount of saline surface water flowing into the lake.

In order to lower the groundwater levels, bores have been installed to pump the saline groundwater. This is then pumped to Taarblin Lake, southwest of Toolibin

Lake, which was also a freshwater lake but is now badly affected by salinity. A long-term revegetation program has also started, with perennial woody vegetation being planted as well as deep-rooted grasses. In April 1995 a 5.5 km diversion drain and gate was constructed to prevent the highly saline surface water from flowing into Toolibin Lake. The gates are opened to allow the highly saline water to be diverted to Lake Taarblin. Once the saline water has been removed, the gates are closed and the relatively fresh surface water rises up over the levee and flows into Lake Toolibin.

If these actions are not successful, Toolibin Lake will inevitably become saline. This would result in the trees on the lakebed and in the surrounding catchment areas dying. The waterbirds that use the lake for breeding and nesting would not be able to survive with these conditions and would disappear. Early results however have been positive. The groundwater level has been lowered by up to 15 m and trees in the vicinity of the pumps appear visually healthier.

5.2 Approximating the Inverse Covariance Matrix with Band-Diagonal Partial Correlation Matrices

In this section the inverse covariance matrix for each of the classes are approximated as being band-diagonal before being used in an MLC of the data. An overview of the theory behind this is given in section 3.4, which was developed in Roger (1996b). It will be shown that even though setting off-diagonal terms to zero in the inverse covariance matrix may make a good approximation, the resulting covariance matrix shows little resemblance to the original covariance. This is not necessarily a problem for MLC however, as the full covariance matrix does not appear in the MLC discriminant function (3.1).

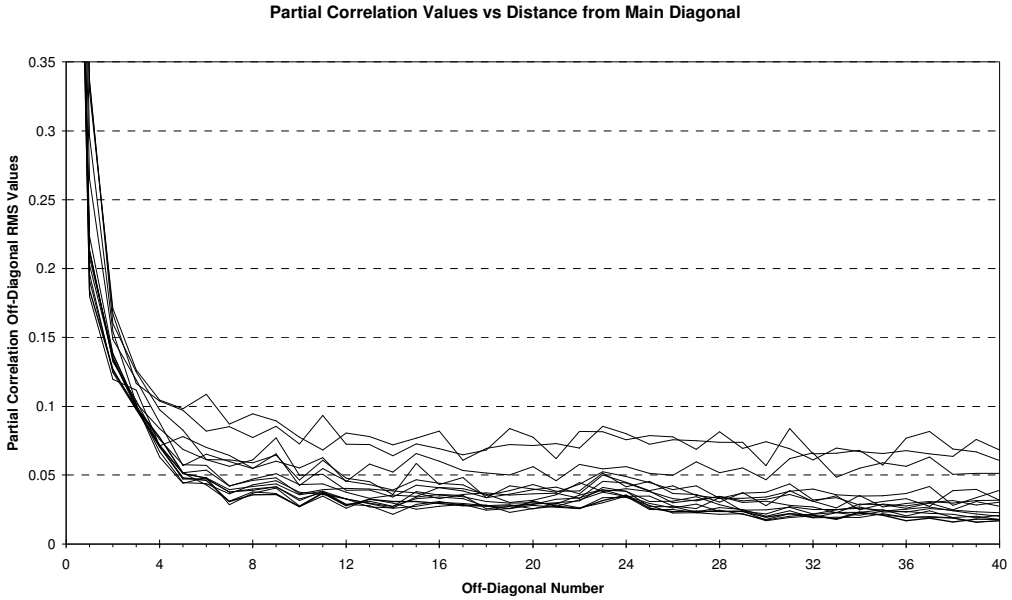


Figure 5.4: Partial correlation RMS values as a function of distance from the main diagonal. Each curve represents the values for a different class.

The first step in approximating the inverse covariance matrix as a band-diagonal matrix is to determine the effective bandwidth of the inverse covariance matrix. To do this the RMS (root-mean-squared) value of each off-diagonal in the partial correlation matrices was calculated and plotted against the off-diagonal number for each class. This is shown in figure 5.4 for the first 40 off-diagonals, where each curve represents one of the 15 classes identified in the data. The elements of the partial correlation matrices have a value of one along the main diagonal and as was expected, figure 5.4 shows that the RMS value sharply decreases from the main diagonal. This means that the partial correlation between neighbouring bands is, in general, higher than that between bands with large separations. Eventually the curves start to level off from about the fifth off-diagonal. Beyond the fifth off-diagonal, the RMS values are very low and suggest that the off-diagonal elements in the partial correlation matrices could be set to zero. Therefore, the inverse covariance matrices will be approximated as band-diagonal using bandwidths of 5, 6 and 7.

By just setting the off-diagonal elements to zero, there is a possibility that the resulting matrix will no longer be positive definite. A non positive definite matrix can be prevented by first finding the Cholesky factor for each of the partial

correlation matrices and then making the resulting lower triangular matrices band-diagonal by setting the off-diagonal terms to zero. Calculating the Cholesky factor for each matrix ensures that the resulting matrices remain positive definite. It also ensures that the determinants of the resulting matrices stay the same as those of the original matrices. The approximated inverse covariance matrices are then reconstructed using (3.9), replacing $\mathbf{P}$ with $\mathbf{L}_c\mathbf{L}_c^T$, where $\mathbf{L}_c$ is the band-approximated Cholesky factor of $\mathbf{P}$.

A symmetric 82 × 82 inverse covariance matrix will contain 6724 elements, or 3403 parameters. When approximated as a band-diagonal matrix, the number of parameters decreases significantly. For example, when bandwidths of 5, 6 and 7 are used the number of parameters required reduces to 477, 553 and 628, respectively. These represent only 14% – 18.5% of the original number of parameters.

As a final check, the band-approximated class inverse covariance matrices were inverted, plotted and compared visually with the original class covariance matrices. Due to the initial Cholesky factorization the determinants will be preserved but this does not guarantee the resulting matrices will look similar. In fact, it was found that the resulting matrices showed little resemblance to the original covariance matrices. At first, this would seem to nullify any value of using this method for MLC. However, referring to (3.1) it is realised that the full covariance matrix is not actually used in the MLC discriminant function. The determinant of the covariance matrix appears which does not change as a result of the initial Cholesky factorisation of the partial correlation matrix. Also, the inverse covariance matrix appears in (3.1), which is what has actually been approximated.

To illustrate this finding, figure 5.5 shows both the original covariance matrices and those resulting from approximating the inverse covariance using a bandwidth of 6 for the classes “Soil/veg” and “Veg”. These classes have been chosen purely because they contain the number of training samples (5010 and 3036 respectively) closest to the average number of training samples for all classes (3771 training samples). The plots have not been generated with the same scale in order to enhance the structure in each matrix and produce the maximum contrast. Despite this, it is still obvious that the matrices resulting from a band-approximated inverse covariance matrix show

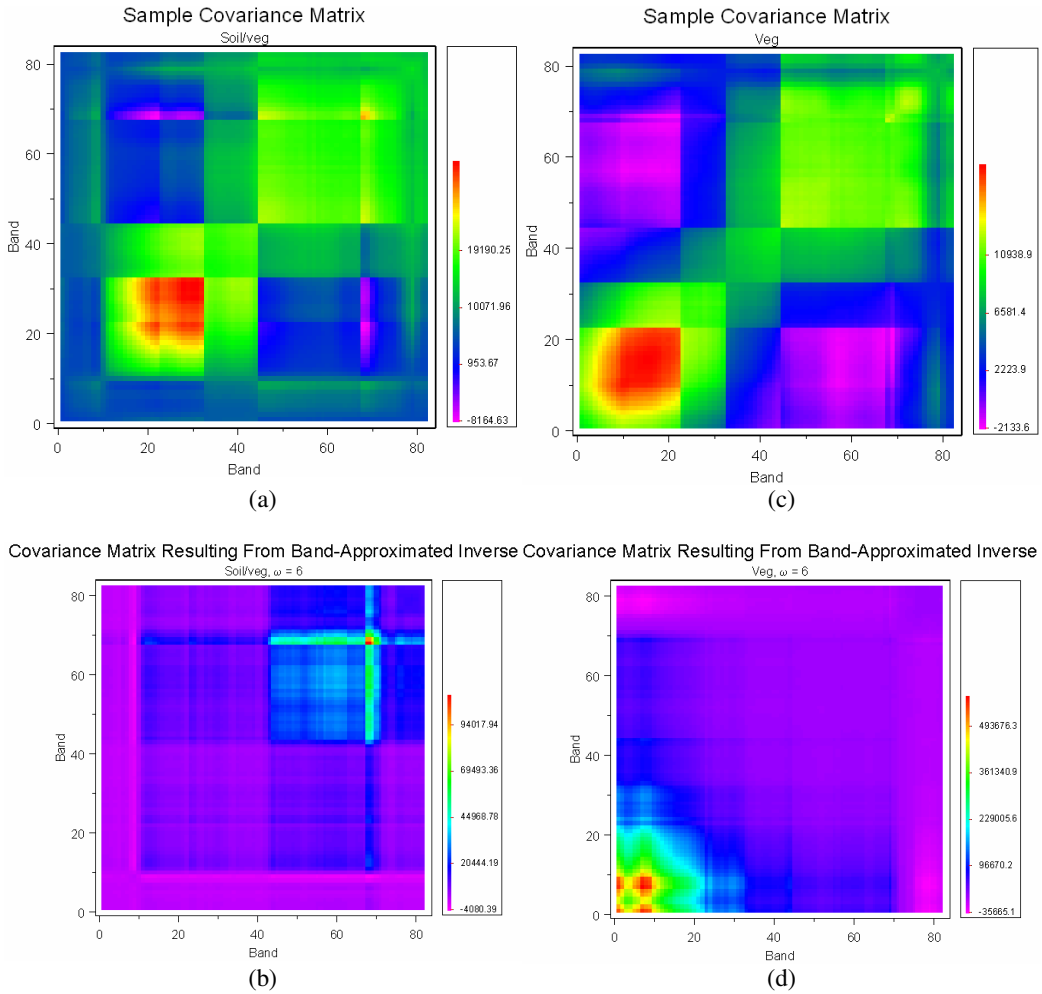


Figure 5.5: Comparison of selected class covariance matrices with those generated from band-approximated inverse covariance matrices: (a) Soil/veg covariance matrix, (b) Matrix resulting from inverse covariance matrix using bandwidth = 6, (c) Veg covariance matrix, (d) Matrix resulting from inverse covariance matrix using bandwidth = 6.

little resemblance to the original covariance matrix in terms of actual values or structure.

5.3 Setting Low Valued Terms in the Partial Correlation Matrix to Zero in the Inverse Covariance Matrix

In section 5.2 the inverse class covariance matrices were approximated by forming the generalization that values away from the main diagonals are approximately zero. The inverse covariance matrices were assumed to be band-diagonal and all off-diagonal values were set to zero. In this section, the band-diagonal assumption is relaxed and low values in the inverse covariance matrix are identified by

thresholding the partial correlation matrix and setting the corresponding low values to zero in the inverse covariance matrix.

The theory for setting individual terms in the inverse covariance matrix was outlined in section 3.5 and was developed in Speed and Kiiveri (1986). To select a threshold to use for identifying which terms to set to zero, figure 5.4, which shows the partial correlation RMS values as a function of distance from the main diagonal, was examined. For most classes the partial correlation RMS values level off below a value of 0.04. It was decided therefore to test the method developed in Speed and Kiiveri (1986) on the class covariance matrices using thresholds of 0.030, 0.035, 0.040, 0.045 and 0.050.

With the thresholds selected, the first step involved generating the adjacency matrices to show those elements which have an absolute value greater than the selected threshold in the partial correlation matrices. Elements in the adjacency matrix will have a value of one if their corresponding partial correlation value has an absolute value of greater than the threshold. If the absolute value in the partial correlation matrix was less than the threshold then the element in the adjacency matrix was set to zero. The adjacency matrices can be thought of as simple undirected graphs, where each vertex represents a spectral band. A value of one indicates a connection between two vertices/bands. The maximal sets of vertices where every pair is adjacent are known as cliques. Once the adjacency matrices were generated for each threshold and for each class, the graphs were triangulated and the cliques extracted. Before the cliques were used as input into the algorithm outlined in section 3.5, they were ordered in such a way that each successive clique contained a band not present in the preceding cliques. This was done so that the modelled covariance matrices would converge after one cycle.

For each class, the original class covariance matrix was input into the algorithm and for each of the selected thresholds the corresponding list of ordered cliques was also input into the algorithm (3.19) to generate the modelled covariance matrices. Tables 5.2, 5.3 and 5.4 summarise the results showing for each class and threshold used: the determinant of the modelled covariance matrix, the number of zero terms in the adjacency matrix, the number of cliques and the number of terms in the longest

Table 5.2: Summary of Results from Modelling Class Covariance Matrices on the First Group of 5 Classes Using the Method Developed in Speed and Kiiveri (1986).

Class	Threshold	Determinant	# of Zero Terms	# of Cliques	# of Terms in Longest Clique
Bare damp field	0	1.3068E+148	0	1	82
	0.03	1.7138E+148	4846	24	54
	0.035	2.2423E+148	5172	30	47
	0.04	4.3137E+148	5460	30	42
	0.045	1.1838E+149	5636	35	36
	0.05	2.8948E+149	5778	38	32
Bare dry field	0	5.1608E+153	0	1	82
	0.03	6.9744E+153	4898	24	55
	0.035	9.9588E+153	5206	27	47
	0.04	3.3475E+154	5460	35	35
	0.045	1.712E+155	5664	41	29
	0.05	4.315E+155	5778	42	25
Bare field	0	2.1823E+145	0	1	82
	0.03	2.2219E+145	2604	6	77
	0.035	2.2397E+145	2990	6	77
	0.04	2.2856E+145	3366	8	74
	0.045	2.363E+145	3736	8	74
	0.05	2.6216E+145	4028	11	71
Bright soil/veg	0	1.5716E+155	0	1	82
	0.03	2.1746E+155	4824	24	58
	0.035	3.0905E+155	5168	31	47
	0.04	5.9475E+155	5436	39	37
	0.045	1.006E+156	5586	34	32
	0.05	2.0412E+156	5728	35	30
Bright soils	0	3.4548E+190	0	1	82
	0.03	-9.3842E+195	3616	9	74
	0.035	7.3348E+195	4066	14	69
	0.04	-6.3443E+209	4458	17	65
	0.045	6.7584E+219	4770	19	62
	0.05	1.2285E+213	5038	21	60

clique. In the tables, a threshold of zero corresponds to the original class covariance matrix which can be thought of as being modelled from a single clique containing all the bands and having no zero terms in its adjacency matrix.

The tables show that for thresholds of 0.030, 0.035, 0.040, 0.045 and 0.050, a significant number of terms in the partial correlation matrix have absolute values less than the threshold. This ranges from 29% (i.e., 1972 out of 6724 elements) for class "Bright tree veg" when using a threshold of 0.03 to almost 86% (i.e., 5778 out of 6724 elements) for classes "Bared damp field" and "Bare dry field" when using a threshold of 0.05. An average of 57.8% of the elements in the class partial correlation matrices had values less than a threshold of 0.03. For a threshold of 0.05, this figure increases to 75.5% of the elements. When thresholds of 0.03 and 0.04 are used on the class "Bright soils" the resulting modelled covariance matrices are not

Table 5.3: Summary of Results from Modelling Class Covariance Matrices on the Second Group of 5 Classes Using the Method Developed in Speed and Kiiveri (1986)

Class	Threshold	Determinant	# of Zero Terms	# of Cliques	# of Terms in Longest Clique
Bright tree veg	0	1.1233E+166	0	1	82
	0.03	9.1584E+165	1972	5	78
	0.035	1.4549E+166	2276	5	78
	0.04	1.1842E+166	2596	6	77
	0.045	1.2271E+166	2870	6	77
	0.05	1.0938E+166	3186	7	76
Bright veg	0	1.2265E+156	0	1	82
	0.03	1.2896E+156	4008	11	72
	0.035	1.3697E+156	4462	11	71
	0.04	1.6515E+156	4834	15	66
	0.045	2.0395E+156	5112	18	61
	0.05	3.1781E+156	5346	24	53
Dark soil/veg	0	1.6046E+148	0	1	82
	0.03	2.1007E+148	4730	19	63
	0.035	3.2183E+148	5146	28	50
	0.04	4.3429E+148	5404	29	44
	0.045	6.4809E+148	5580	27	40
	0.05	1.0842E+149	5704	30	35
Dark tree veg	0	2.2526E+161	0	1	82
	0.03	2.3516E+161	3568	9	74
	0.035	2.6021E+161	3990	12	71
	0.04	2.7086E+161	4388	12	70
	0.045	3.5329E+161	4700	15	67
	0.05	4.3673E+161	4974	18	62
Lush veg	0	7.3142E+152	0	1	82
	0.03	9.4561E+152	4566	21	58
	0.035	1.1132E+153	4970	26	53
	0.04	1.5541E+153	5232	33	46
	0.045	2.6919E+153	5438	33	41
	0.05	6.6284E+153	5614	32	37

positive definite. This presents a problem, as MLC requires that the covariance matrices are positive definite.

It is to be expected that as the selected threshold is increased then: the number of parameters set to zero increases, the number of cliques increase and the length of the shortest clique decreases. This is demonstrated in tables 5.2, 5.3 and 5.4. What is not apparent from the tables is that there is also a relationship between the number of training pixels used to calculate the statistics for each class and the number of parameters set to zero. It was found that as the number of training pixels used for a class increased, so to would the number of parameters set to zero for a given threshold. For example, Bright tree veg was the smallest class, having 307 training pixels. At thresholds of 0.03 and 0.05 the number of parameters set to zero was 1972 and 3186 respectively. This compares with the largest class Bare dry field which had

11059 training pixels. At thresholds of 0.03 and 0.05 the number of parameters set to zero was 4898 and 5778.

Figure 5.6 shows a few examples of the modelled covariance matrices for the classes "Soil/veg" and "Veg". The original class covariance matrices are compared to the modelled results using a threshold of 0.05. Visually, the modelled and the original covariance matrices look extremely similar and it is difficult to spot any differences between them. In terms of their determinants, the modelled covariance matrices also show good agreement with the originals. For Soil/veg, the determinant of the sample covariance matrix was 7.7316×10^{152} compared with 5.9237×10^{153} for the modelled covariance matrix using a threshold of 0.05. For Veg, the determinants for the original covariance matrix and that using a threshold of 0.05 were 1.9423×10^{152} and 5.1014×10^{152}, respectively.

Table 5.4: Summary of Results from Modelling Class Covariance Matrices on the Last Group of 5 Classes Using the Method Developed in Speed and Kiiveri (1986)

Class	Threshold	Determinant	# of Zero Terms	# of Cliques	# of Terms in Longest Clique
Shadow	0	5.1606E+154	0	1	82
	0.03	5.9557E+154	3780	11	71
	0.035	6.3954E+154	4238	13	70
	0.04	6.2731E+154	4606	20	63
	0.045	8.8182E+154	4910	23	58
	0.05	1.0851E+155	5202	25	53
Soil	0	2.5919E+150	0	1	82
	0.03	2.8512E+150	4044	15	68
	0.035	2.8882E+150	4498	15	67
	0.04	3.6388E+150	4872	19	56
	0.045	4.435E+150	5142	21	53
	0.05	6.8095E+150	5368	24	49
Soil/veg	0	7.7316E+152	0	1	82
	0.03	9.5241E+152	4510	18	63
	0.035	1.2127E+153	4946	25	53
	0.04	1.6592E+153	5224	30	46
	0.045	2.5867E+153	5452	32	41
	0.05	5.9237E+153	5632	34	35
Veg	0	1.9423E+152	0	1	82
	0.03	2.166E+152	4392	17	66
	0.035	2.3611E+152	4752	20	63
	0.04	2.8855E+152	5100	24	57
	0.045	3.8233E+152	5358	29	51
	0.05	5.1014E+152	5538	33	45
Water	0	7.2247E+168	0	1	82
	0.03	3.2898E+172	1966	6	77
	0.035	1.7289E+171	2278	7	76
	0.04	1.3326E+173	2626	8	75
	0.045	3.3692E+172	2932	8	75
	0.05	9.5759E+172	3232	8	75

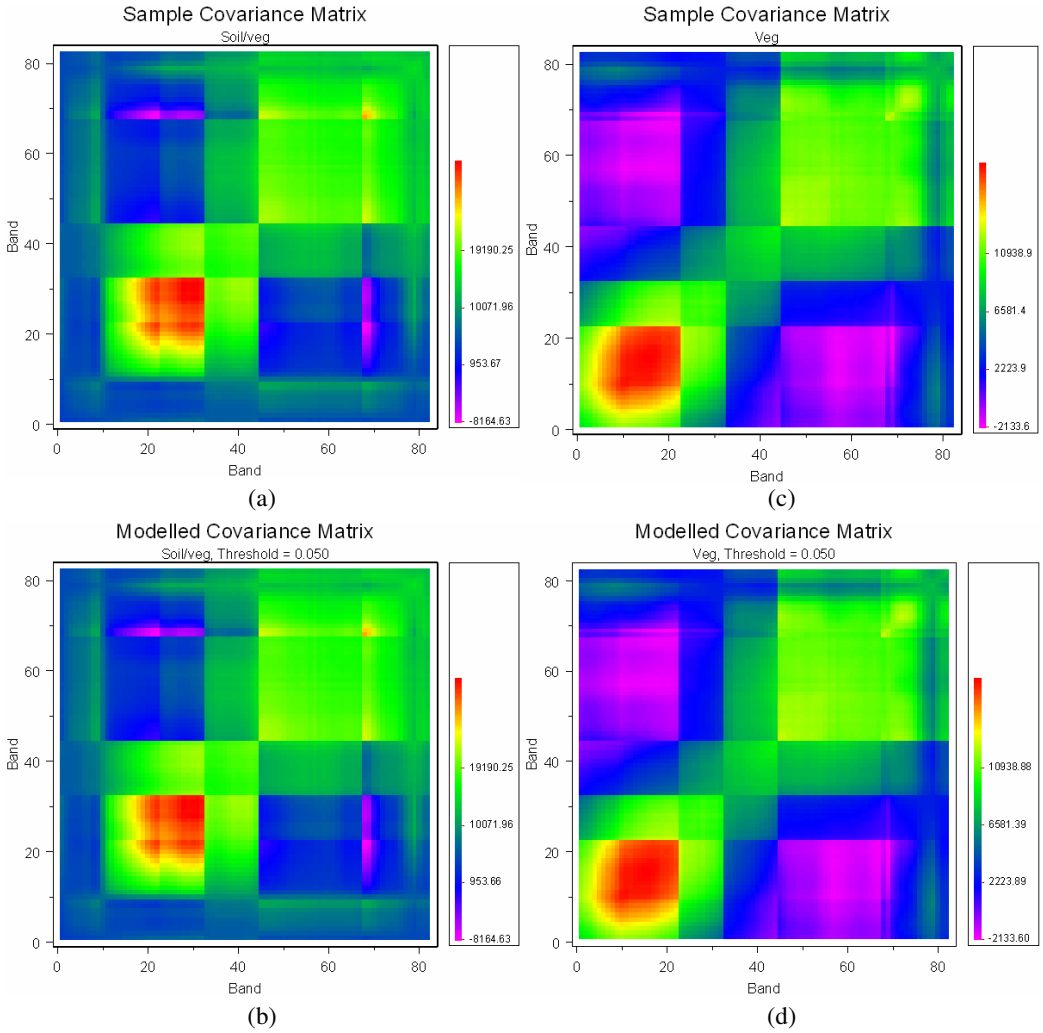


Figure 5.6: Comparison of selected class covariance matrices with those generated from Speed and Kiiveri (1986) algorithm: (a) Soil/veg covariance matrix, (b) Modelled Soil/veg covariance matrix using a threshold of 0.05, (c) Veg covariance matrix, (d) Modelled Veg covariance matrix using a threshold of 0.05.

5.4 Classification Results Using Parameter-Reduced Class Covariance Matrices

The sparse nature of the hyperspectral can be exploited to set a large number of terms to zero and the approximated inverse covariance matrices can then be used with MLC to classify the data. The commercially available software ER Mapper 6.2 was used to perform MLC with the data. The class covariance matrices needed for MLC can be found in the ER Mapper header file associated with the image data, and before a classification was performed, each class covariance matrix was replaced with the newly modelled covariance matrix. Once this had been done the standard

MLC was executed using ER Mapper. As well as a standard MLC, ER Mapper also includes an option to use the spatial correlations and use neighbouring pixels in the classification. It was decided to use the Maximum Likelihood Standard option and not use neighbouring pixels. This is because in this work the spectral correlations are being examined, and not the spatial correlations.

Table 5.2 shows negative determinants for thresholds of 0.03 and 0.04, indicating that the resulting covariance matrices for "Bright soils" are not positive definite, and therefore are not suitable for MLC. To overcome this problem, the original sample covariance matrix for the class Bright soils was used in place of the covariance matrix resulting from a threshold of 0.03. The covariance matrix resulting from a threshold of 0.04 was replaced with the covariance matrix for Bright soils resulting from a threshold of 0.035.

Some of the resulting classified images are shown in figure 5.7. The result from a standard MLC is shown along with classifications using a bandwidth of 6 in the inverse covariance matrix and thresholds of 0.035 and 0.045 applied to the partial correlation matrix. The colour scheme used for the classification images is the same as that used to depict the training data shown in figure 5.3. The result from using a standard MLC is very similar to the result from using an MLC with a threshold of 0.035. When a threshold of 0.045 was used, the Bright soils class was clearly over-classified. In fact, both the standard MLC and the MLC using a threshold of 0.035 have also over-classified Bright soils, but not to the extent as that when using a threshold of 0.045. A possible explanation is that some of the pixels erroneously classified as Bright soils could in fact be mixed pixels that contained a proportion of Bright soils and other classes. Another possible explanation relates to the original training data. Unlike many of the classes, Bright soils did not occur in large contiguous regions within the image. As a result, when training the data, in order to get enough samples to reliably estimate the covariance information for Bright soils, individual pixels had to be selected. It is probable that some of the pixels selected were not 100% pure but were a mixture of Bright soils and other classes. This would have an effect on the estimation of the mean spectra and the covariance information for the class, and this would propagate into the classification results.

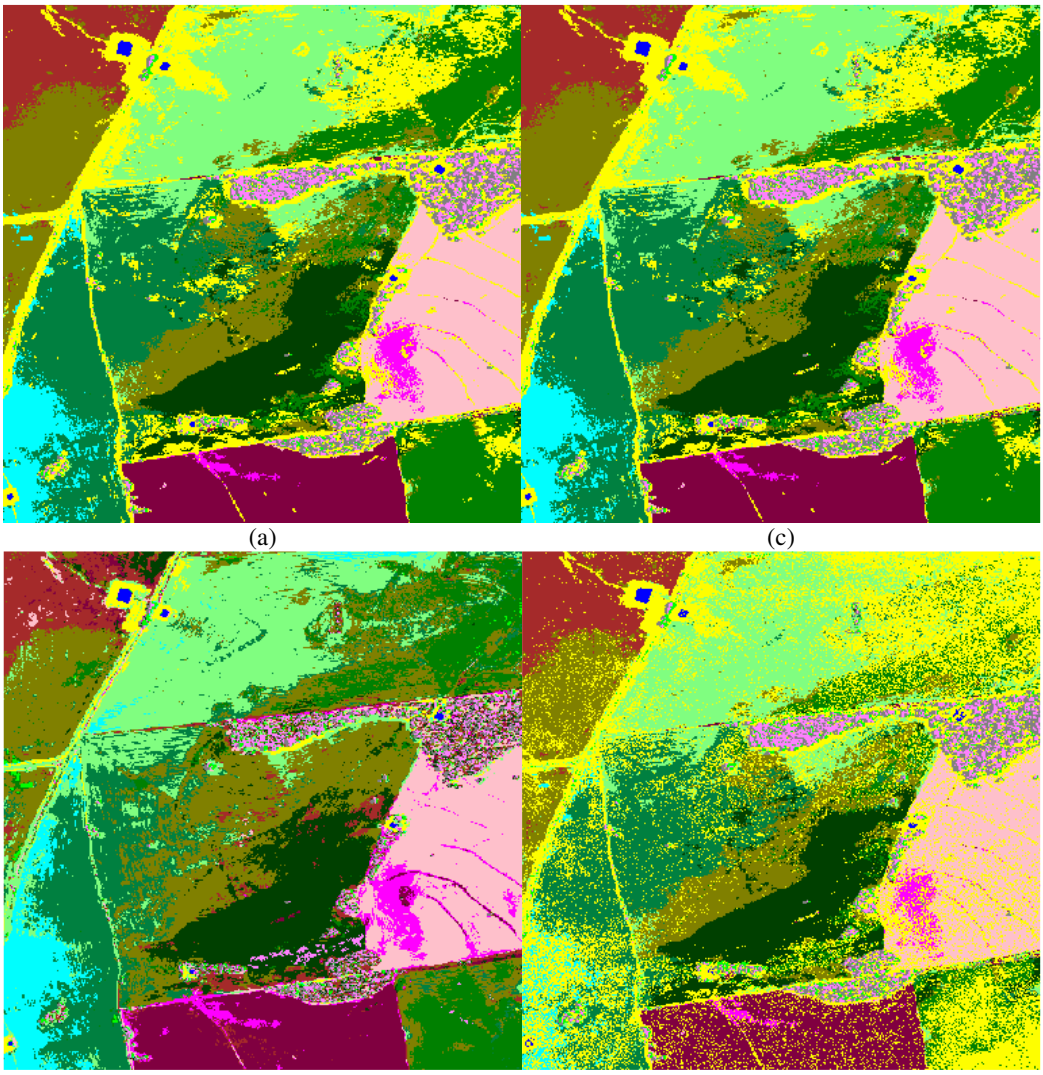


Figure 5.7: Comparison of selected classified images: (a) Standard MLC image, (b) Classification image resulting from a bandwidth of 6, (c) Classification image resulting from a threshold of 0.035, (d) Classification image resulting from a threshold of 0.045.

At the time when the image was acquired, no adequate ground truth data had been collected. In order to assess the accuracy of the classifications then, the original training data was used for comparison. Table 5.5 shows the accuracies of the different schemes implemented when assessed against the original training data. When using the Speed and Kiiveri (1986) algorithm with thresholds of 0.03, 0.035 and 0.04, the classification accuracies are very similar and comparable to those from the standard MLC (96.46%). As the threshold is increased to 0.045 and 0.05, the classification accuracies decrease to 77% and 61% respectively. When using band-diagonal inverse covariance matrices for classification, the classification accuracies

Table 5.5: Accuracies from the Different MLC Implementations

Speed and Kiiveri (1986)		Band-diagonal Approximated Inverse	
Threshold	Classification Accuracy (%)	Bandwidth, ω	Classification Accuracy (%)
0.03	96.465	5	93.070
0.035	96.458	6	92.985
0.04	96.417	7	92.362
0.045	76.929		
0.05	60.949	Standard MLC	96.460

are slightly lower than that from a standard MLC, yet still range between 92% – 93% for bandwidths of 5, 6 and 7.

The classification accuracies for each class using each of the different MLC implementations are shown in table 5.6. The classes that were classified with the lowest accuracy (Bright tree veg, Shadow, Dark tree veg, bright soils and Water) were those whose spatial distributions did not occur in large contiguous groups within the image, but were scattered throughout the image in small groups containing only a few pixels of the same class. This further reinforces the idea that the training data selected for these classes may not have been 100% pure. If this was the case, it was unavoidable as at least $(N + 1)$ training samples are required to avoid the covariance matrices being singular and to get reliable estimates of the covariance information, at least $10N$ training samples are needed (Swain & Davis 1978). The poorly classified classes were also the classes which had the fewest number of training pixels selected, with the exception of the class Bare field, which only had 475 pixels selected as training data, yet was accurately classified.

When classified using a threshold of 0.045 both Lush veg and Bright veg unexpectedly had significantly lower classification accuracies (15% and 33% respectively) than when they were classified using a threshold of 0.05. For the other classes, lower thresholds resulted in higher classification accuracies. Looking at the classification images in figure 5.7, it can be seen that these classes have actually been classified predominantly as Bright soils. When modelling the covariance matrix for Bright soils using a threshold of 0.045, the determinant is 6.7584×10^{219}. This is considerably higher than the determinant when using a threshold of 0.05, i.e. 1.2285×10^{213}. In the majority of cases, as the threshold is increased, the determinant of the modelled covariance matrix also increases.

Table 5.6 demonstrates the success of the methods tested, particularly when using the method from Speed and Kiiveri (1986) with thresholds ≤ 0.04. When approximating the inverse covariance matrices as band-diagonal with bandwidths of 5, 6 and 7, 10 of the 15 classes were classified with accuracies > 90%, giving an overall classification accuracy ranging from 92% – 93%. Using thresholds of ≤ 0.04 with the method from Speed and Kiiveri (1986) resulted in 10 out of the 15 classes being classified with accuracies > 95%. This compared very well with the results from a standard MLC. Overall, the classification accuracies from both the standard MLC and thresholds ≤ 0.04 gave accuracies of 96%.

These results demonstrate that there is a large degree of redundancy inherent in the covariance structure of hyperspectral, remotely sensed data. Many of the terms in the inverse covariance matrices can be set to zero and yet still achieve comparable classification accuracies when compared with the results from a standard MLC. This indicates that a significant saving in the number of parameters required can be realised, by identifying parameters having values close to zero and systematically setting them to zero. The structure of the covariance matrices is considerably simplified and the amount of disk space needed for storage is also significantly reduced.

This chapter has examined ways of reducing the number of parameters required to describe the class covariance information of hyperspectral remotely sensed data, with

Table 5.6: Classification Accuracies (%) for Each Class for the Different MLC Implementations

Class	Standard	Thresholds					Bandwidths			Average
	MLC	0.03	0.035	0.04	0.045	0.05	5	6	7	
Shadow	53	53	54	53	50	36	34	35	35	45
Lush veg	100	100	100	100	15	93	97	97	95	89
Veg	100	100	100	100	90	86	99	99	97	97
Dark soil/veg	100	100	100	100	100	77	99	98	98	97
Dark tree veg	57	58	57	57	57	45	46	48	46	52
Bright soil/veg	100	100	100	100	90	71	99	99	99	95
Soil/veg	100	100	100	100	87	78	97	96	96	95
Soil	100	100	100	100	99	51	96	97	97	93
Bright soils	86	86	86	85	83	86	52	54	53	75
Bright veg	99	99	99	99	33	95	99	99	99	91
Bare field	98	98	98	98	79	57	99	99	99	92
Bare dry field	100	100	100	100	84	55	99	99	99	93
Bare damp field	100	100	100	100	83	18	97	97	96	88
Bright tree veg	35	35	35	31	39	25	31	31	32	33
Water	88	87	86	88	71	46	77	77	77	77
Overall	96	96	96	96	77	61	93	93	92	

applications in MLC. The class covariance matrices play an important role in MLC and when the data is of a hyperspectral nature, these matrices will be large. Consequently, there will be a large number of elements in each class covariance matrix, which requires larger amounts of storage space and increases the amount of processing time considerably. Two approaches were used, one that approximates the inverse covariance matrices as band-diagonal and one that identifies low valued elements within the partial correlation matrices using thresholds and subsequently sets the low valued elements to zero.

When a bandwidth of 7 was used, 5550 out of 6724 elements were set to zero in each inverse covariance matrix. Using these band-diagonal approximated inverse covariance matrices, the resulting MLC achieved an accuracy of 92.4%. When a threshold of 0.04 was used to identify low valued elements in the class partial correlation matrices an average of 4604 elements were set to zero. The overall classification accuracy from using the resulting modelled class inverse covariance matrices was 96.4%. In this work the selected threshold was applied to all class partial correlation matrices. It was noticed later that the number of terms in the class partial correlation matrix having absolute values less than the threshold was also dependant on the number of training pixels selected for each particular class. In future work, class inverse covariance matrices having a small number of training pixels could be modelled by using a smaller threshold than for larger classes.

The individual pixel spectra also play a major role in MLC and are used in the MLC discriminant function. The pixel spectra are a possible source of redundancy in hyperspectral data and if the effective dimensionality of the pixel spectra can be reduced, then this will reduce the amount of processing time needed for MLC of the data. Not only will the size of the spectra be reduced, the size of the resulting class covariance matrices will also decrease considerably. This issue is examined in the next chapter.

6 Dimensionality Reduction for the Characterization of Remotely Sensed Spectral Curves

The covariance matrix is a potential source of redundancy in hyperspectral remote sensing. This is significant in applications such as MLC, where the covariance matrix is required for the discriminant function. As well as the class covariance matrix, the individual pixel spectra and the mean class spectra also appear in the discriminant function for MLC (3.1). When working with hyperspectral data, these spectra are also a potential source of high redundancy. The spectral bands on hyperspectral sensors are spaced close together (typically < 20 nm separation) and neighbouring bands located in atmospheric windows are usually very highly correlated. The spectra measured by hyperspectral sensors can resemble smooth curves, and this combined with the high inter-band correlations can be exploited to approximate the spectral curves using a significantly smaller number of parameters. This will ease the problems of storage and transmission of the data. A reduction in the dimensionality of the data will also lead to a reduction in the size of the covariance matrices used for the classification of the data.

In this chapter, methods are investigated to characterize the measured spectral curves using fewer bands or parameters, with an emphasis on minimizing the RMS error. The spectra from 20 selected pixels have been extracted and this is presented in section 1. The 20 sample spectra are then approximated using the discrete wavelet transform (DWT). Unlike the PCT, the DWT is a relatively new technique when used for the analysis of remotely sensed hyperspectral data (Li et al. 2001; Koger et al. 2003; Bruce, Koger & Li 2002). As the measured spectra resemble smooth curves, piecewise cubic smoothing splines will be used for their approximation.

6.1 The Data

A row of 20 pixels has been selected from the image in figure 5.1 in order for the methods to be tested. The pixels are from line 196 in the image and stretch between

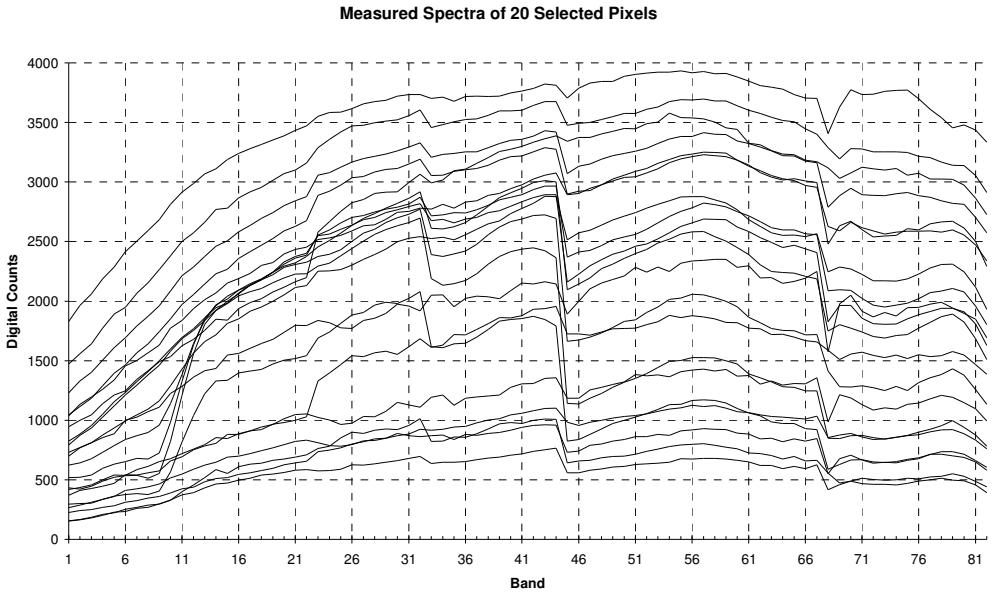


Figure 6.1: Spectra of the 20 pixels used in this study. The pixels are from row 196 between columns 406 – 425 in the HyMap image used for this research.

columns 406 – 425. These pixels were chosen as they cover several different classes and will therefore exhibit diverse spectra having different characteristics. The 82-band spectra have been extracted from each of the pixels and are shown in figure 6.1.

6.2 Approximating Spectra Using Wavelets

The field of wavelet analysis is a relatively recent development in applied mathematics (Daubechies 1992) and particularly in hyperspectral remote sensing (Li et al. 2001). Wavelets are oscillating functions with compact support. The compact support of a wavelet function $\psi(t)$ refers to the maximal interval outside of which the wavelet function has a value of zero. When compared with traditional Fourier analysis techniques wavelets are a significant advancement. Wavelets have advantages for representing discontinuous functions or functions with sharp peaks. Wavelets are also better for deconstructing and reconstructing finite or non-periodic functions. The wavelet functions are formed from dilated and translated versions of a single basic wavelet, known as the mother wavelet. Unlike Fourier analysis which is a global analysis, wavelets operate on the signal simultaneously at different scales and different locations.

A wavelet analysis simultaneously separates a signal into a coarse approximation and coefficients representing the finer details. These finer detail coefficients (the wavelet coefficients) are saved and the coarse approximation is submitted to the next level of the wavelet analysis. The coarse approximation is then separated into a coarser approximation and finer detail wavelet coefficients. This process continues recursively until a predetermined level has been reached. At each level the wavelet analysis can be thought of as a simultaneous high-pass and low-pass filtering of the data.

In order for an oscillating function to be considered as a wavelet function it must meet certain criteria. The first is that it must have an average value of zero. The second criteria is that the wavelet function ψ must have a magnitude of one, i.e. $\|\psi\| = 1$. The final criterion is that the function must satisfy the admissibility condition, i.e.

$$\int_{-\infty}^{\infty} \frac{|\hat{\psi}(\omega)|^2}{\omega} d\omega < \infty, \tag{6.1}$$

where $\hat{\psi}$ denotes the Fourier transform of the wavelet function.

The continuous wavelet transform (also known as the integral wavelet transform) of a signal f is given by

$$W_\psi f(\beta,\alpha) = \langle f, \psi_{\beta,\alpha} \rangle = \frac{1}{\sqrt{\alpha}} \int_{-\infty}^{\infty} f(t) \overline{\psi\left(\frac{t-\beta}{\alpha}\right)} dt, \tag{6.2}$$

where α and β are the dilation and translation parameters respectively, and the over bar notation represents the complex conjugate. Equation (6.2) is a function of two *real* variables. To arrive at the discrete wavelet transform, the dilation and translation parameters first need to be discretized. The dilation parameter, α, is recast as α_0^j and the translation parameter, β, is recast as $n\alpha_0^j\beta_0$. The j and n terms are both integers. A typical value used for α_0 is 2 and for β_0 a good choice is a value

of 1. If the original signal, *f*, has been discretized then (6.2) can be rewritten to become the discrete wavelet transform

$$W_\psi f(\alpha_0^j, n\alpha_0^j\beta_0) \approx \sum_n f(t)\alpha_0^{-j/2}\psi\left(\frac{t - n\alpha_0^j\beta_0}{\alpha_0^j}\right) = 2^{-j/2}\sum_n f(t)\psi(2^{-j}t - n) \tag{6.3}$$

The inversion formula for the continuous wavelet transform is given by

$$f(t) = \frac{1}{C_\psi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{\alpha^2}\left[W_\psi f(\beta,\alpha)\right]\psi_{\beta,\alpha}(t)\,d\alpha\,d\beta \tag{6.4}$$

where C_ψ is given by

$$C_\psi = \int_{-\infty}^{\infty}\frac{|\hat{\psi}(\omega)|^2}{\omega}d\omega < \infty \tag{6.5}$$

This should be recognized as the admissibility condition from (6.1).

A wavelet analysis of a signal is essentially a multiresolution analysis consisting of a sequence of successive approximation subspaces $\ldots V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} \subset \ldots \subset L^2$ (Daubechies 1992; Blatter 1998; Kaarna et al. 2000). These V_j have the properties that $\bigcap_j V_j = \{0\}$ and $\overline{\bigcup_j V_j} = L^2$. If the orthogonal compliment of V_j is denoted as W_j then $V_{j-1} = V_j \oplus W_j$. But $V_j = V_{j+1} \oplus W_{j+1}$, and $V_{j+1} = V_{j+2} \oplus W_{j+2}$, etc., and so it can be written $V_j = W_{j+1} \oplus W_{j+2} \oplus W_{j+3} \oplus \ldots \oplus W_{j+J} \oplus V_{j+J}$ for some level J. The V subspaces constitute the low-pass filters (l_k) and the W subspaces constitute the high-pass filters (h_k). A function $\phi \in V_0$ exists such that its translates $\{\phi(t - n) | n \in Z\}$ form an orthonormal basis for V_0. The function ϕ is known as the scaling function and is sometimes referred to as the father wavelet (Aboufadel & Schlicker 1999). $\phi(t)$ can be expressed as

$$\phi(t) = \sqrt{2}\sum_{k=-\infty}^{\infty} l_k \phi(2t - k)\,, \tag{6.6}$$

where l_k are the low-pass filter coefficients and can be calculated from $l_k = \langle \phi, \phi_{-1,k} \rangle$. The value of k ranges from $0 \le k \le 2N - 1$. The $\phi_{j,k}$ are defined as

$$\phi_{j,k}(t) = 2^{-j/2} \phi\left(\frac{t}{2^j} - k \right). \tag{6.7}$$

In (6.6), $\phi(t)$ was defined in terms of the filter coefficients l_k. In much the same way, $\psi(t)$ can be defined as

$$\psi(t) = \sqrt{2} \sum_k h_k \phi(2t - k), \tag{6.8}$$

where the h_k are the high-pass filter coefficients and are related to l_k by

$$h_k = (-1)^k \overline{l_{2N-1-k}}. \tag{6.9}$$

The first stage in analysing a signal f is to generate the approximation coefficients $a_{j,n}$. Initially, $a_{0,n}$ is set to $f(n)$ and as j increases the approximation of f becomes coarser (by a factor of 2). Both $\phi_{j,n}$ and $\psi_{j,n}$ can be written recursively as

$$\phi_{j,n} = \sum_k l_k \phi_{j-1,2n+k} \quad \forall j, \forall n \tag{6.10}$$

and

$$\psi_{j,n} = \sum_k h_k \phi_{j-1,2n+k} \quad \forall j, \forall n. \tag{6.11}$$

The $a_{j,n}$ are given by

$$a_{j,n} = \langle f, \phi_{j,n} \rangle = \sum_k \overline{l_k} \langle f, \phi_{j-1,2n+k} \rangle. \tag{6.12}$$

Recursively, this can be written as

$$a_{j,n} = \sum_k \overline{l_k} a_{j-1,2n+k} \ . \tag{6.13}$$

The next stage involves generating the actual wavelet coefficients, $d_{j,n}$ (i.e the detail coefficients) and these are given by

$$d_{j,n} = \langle f, \psi_{j,n} \rangle = \sum_k \overline{h_k} \langle f, \phi_{j-1,2n+k} \rangle \, . \tag{6.14}$$

As before, this can be written recursively as

$$d_{j,n} = \sum_k \overline{h_k} a_{j-1,2n+k} \ . \tag{6.15}$$

The wavelet analysis algorithm is shown diagrammatically in (6.16). The maximum number of steps, J, in the algorithm is determined by the length of the digitized signal, $f(n)$. If f has a length of 2^J then the wavelet analysis should be terminated after $\leq J$ steps.

$$\begin{array}{ccccccccc} a_{0,n} & \xrightarrow{\overline{l_k}} & a_{1,n} & \xrightarrow{\overline{l_k}} & a_{2,n} & \xrightarrow{\overline{l_k}} & a_{3,n} & \xrightarrow{\overline{l_k}} \cdots & \xrightarrow{\overline{l_k}} a_{J,n} \\ \downarrow \overline{h_k} & & \downarrow \overline{h_k} & & \downarrow \overline{h_k} & & \downarrow \overline{h_k} & & \downarrow \overline{h_k} \\ d_{1,n} & & d_{2,n} & & d_{3,n} & & d_{4,n} & \cdots & d_{J,n} \end{array} \tag{6.16}$$

To reconstruct the signal with a wavelet synthesis, the process in (6.16) needs to be reversed. An equation for the $a_{j-1,n}$ is

$$a_{j-1,n} = \sum_k a_{j,k} \langle \phi_{j,k}, \phi_{j-1,n} \rangle + \sum_k d_{j,k} \langle \psi_{j,k}, \phi_{j-1,n} \rangle \tag{6.17}$$

which becomes

$$a_{j-1,n} = \sum_k l_{n-2k} a_{j,k} + \sum_k h_{n-2k} d_{j,k} \tag{6.18}$$

The wavelet synthesis algorithm is shown schematically in (6.19)

$$
\begin{array}{ccccccccc}
a_{J,n} & \xrightarrow{l_k} & a_{J-1,n} & \xrightarrow{l_k} & a_{J-2,n} & \xrightarrow{l_k} & \cdots & \xrightarrow{l_k} & a_{1,n} & \xrightarrow{l_k} & a_{0,n} \\
 & & \uparrow h_k & & \uparrow h_k & & \uparrow h_k & & \uparrow h_k & & \uparrow h_k \\
 & & d_{J,n} & & d_{J-1,n} & & d_{J-2,n} & & d_{2,n} & & d_{1,n}
\end{array}
\qquad (6.19)
$$

It should be noted that apart from the $a_{J,n}$ coefficients, the $a_{j,n}$ resulting from the analysis phase are only temporary coefficients and it is not necessary for them to be stored in order to reconstruct the original signal at a later date. It is only the $a_{J,n}$ and the wavelet coefficients $d_{j,n}$ that need to be retained for a perfect reconstruction of the signal. By storing all the $d_{j,n}$ and $a_{J,n}$, no reduction in the amount of storage space is achieved. A significant reduction in the amount of storage space can be realised however by recognizing that a significant number of the $d_{j,n}$ are relatively close to zero. If a threshold is set and all $d_{j,n}$ with an absolute value less than this threshold are set to zero, the number of remaining non-zero $d_{j,n}$ terms is usually quite small, depending on the form of the original signal. The reconstruction of the signal will be reasonably accurate, although the accuracy will depend on the size of the threshold used.

Interestingly, neither the wavelet function ψ nor the scaling function ϕ appears in the wavelet analysis and synthesis algorithms, nor do they need to be known. All that is needed is the l_k and h_k values, and these can be found in many texts covering wavelets (e.g. Daubechies (1992)) for many existing wavelets, such as the Haar wavelet or the Daubechies family of wavelets. There is also commercially available software such as S-Plus that can be used for generating the l_k and h_k.

The Haar wavelet and some of the Daubechies wavelets are presented in the next two sections. These will be used to analyse the spectra presented in section 6.1. Different thresholds will be used and many of the wavelet coefficients will be set to zero. The original spectra will then be approximated from a wavelet synthesis using the remaining non-zero wavelet coefficients.

6.2.1 The Haar Wavelet

The Haar wavelet was the first of the orthonormal wavelets and is one of the more popular wavelets. It is not continuous and resembles a simple step function. The wavelet function has compact support of [0, 1] and the function is given by:

$$\psi_{\text{Haar}}(t) = \begin{cases} 1 & (0 \le t < ½) \\ -1 & (½ \le t < 1) \\ 0 & (\text{otherwise}) \end{cases} . \tag{6.20}$$

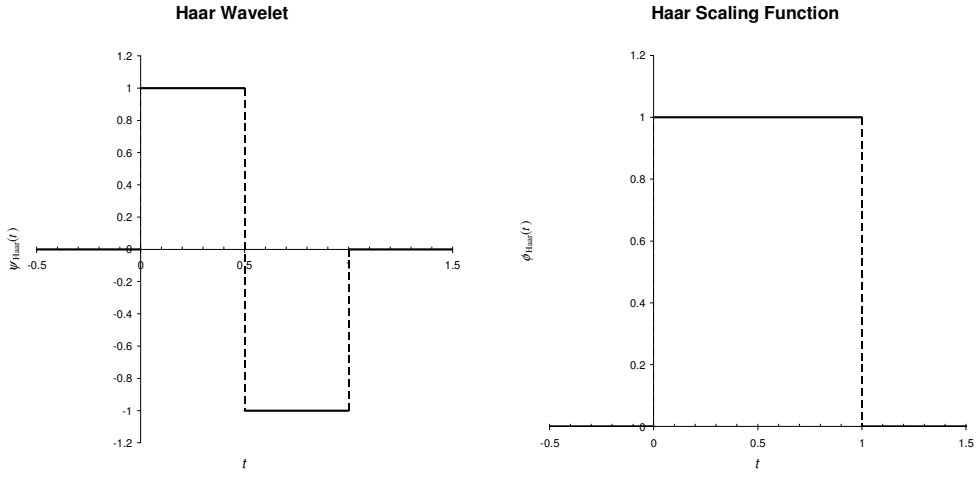
The corresponding scaling function $\phi_{\text{Haar}}(t)$ is given by:

$$\phi_{\text{Haar}}(t) = \begin{cases} 1 & 0 \le t < 1 \\ 0 & (\text{otherwise}) \end{cases} . \tag{6.21}$$

Both the Haar wavelet and the associated scaling function are shown graphically in figure 6.2. The low and high-pass filter coefficients l_k and h_k are given in table 6.1. It should be noted that with the l_k

$$\sum_k l_k = \sqrt{2} \quad \text{and} \quad \sum_k |l_k|^2 = 1$$

and with the h_k



(a) (b)

Figure 6.2: (a) The Haar Wavelet and (b) The Haar scaling function

Table 6.1: Filter Coefficients for the Haar Wavelet.

k	l_k	h_k
0	0.7071068	0.7071068
1	0.7071068	–0.7071068

$$\sum_k h_k = 0 \quad \text{and} \quad \sum_k |h_k|^2 = 1.$$

The Haar wavelet was used to analyse each of the spectra shown in figure 6.1 and using different thresholds the low valued wavelet coefficients were set to zero. The spectra were then approximated by synthesizing them from the remaining non-zero wavelet coefficients. For each approximated spectra, the number of non-zero wavelet coefficients used was determined and the RMS error was calculated. The RMS values for each of the spectra approximated with the Haar wavelet are given in table 6.2. As expected, lower thresholds resulted in more accurately approximated spectra. With a threshold of 50, the RMS errors for the 20 spectra ranged from 15.3 – 20.41, at an average of 17.46. When the threshold was increased to 150, the range of RMS errors increased to 35.39 – 56.61 with an average of 46.12.

The number of wavelet coefficients used to approximate each of the spectra can be found in table 6.3. The analysis of the spectra was terminated after $J = 5$ levels and

Table 6.2: RMS Values of Approximated Spectra When Using the Haar Wavelet with Various Thresholds

Pixel	Threshold				
	50	75	100	125	150
1	17.27	30.58	37.60	44.93	47.57
2	18.41	29.81	35.55	43.22	48.48
3	17.54	27.16	34.86	36.97	43.55
4	16.49	29.17	35.78	41.48	51.25
5	15.30	26.37	41.62	43.18	56.61
6	15.40	32.29	33.33	39.75	47.37
7	18.69	29.52	34.57	43.82	51.16
8	20.41	25.65	39.11	42.43	53.13
9	16.63	22.58	32.74	37.25	40.02
10	16.44	26.35	29.74	31.48	35.39
11	17.44	23.61	34.10	37.57	41.45
12	16.21	22.83	24.80	37.77	40.43
13	18.96	24.31	29.31	35.11	38.37
14	17.24	25.02	34.37	40.99	43.81
15	20.20	24.06	36.46	39.96	52.11
16	17.88	26.53	36.05	45.38	47.61
17	18.50	26.97	34.19	42.29	47.42
18	15.66	23.53	32.89	34.99	42.80
19	17.91	28.58	30.52	39.75	42.38
20	16.55	29.06	38.42	42.59	51.41
Average	17.456	26.700	34.301	40.046	46.116

Table 6.3: Number of Wavelet Coefficients Used to Approximate Spectra With the Haar Wavelet at Different Thresholds.

Pixel	Threshold				
	50	75	100	125	150
1	50	37	32	28	27
2	47	36	32	28	26
3	42	33	28	27	25
4	47	34	29	26	22
5	47	37	26	25	19
6	44	27	26	23	20
7	45	34	30	25	22
8	41	36	27	25	21
9	28	24	17	15	14
10	26	17	15	13	12
11	33	27	20	17	15
12	41	35	34	29	28
13	41	36	33	31	30
14	48	40	34	31	30
15	47	43	35	33	28
16	45	37	31	26	25
17	37	28	23	19	17
18	31	24	18	17	14
19	38	27	26	22	21
20	49	35	28	26	22
Average	41.35	32.35	27.20	24.30	21.90

at this stage the number of temporary coefficients had been reduced to 3. These have not been included in the values in table 6.3. For a threshold of 50, an average of 41 wavelet coefficients had not been set to zero and were used in the reconstruction of the spectra. This decreased to an average of 22 wavelet coefficients when using a threshold of 150. For pixel 10, as few as 12 wavelet coefficients were used when the threshold was 150, but for pixel 1 and a threshold of 50, as many as 50 wavelet coefficients were used. This is still a significant reduction compared to the 82 bands required to characterize the original spectra.

As a graphical example of the results, the spectra from pixels 6, 10 and 14 are shown in figures 6.3, 6.4 and 6.5, respectively, as well as the resulting approximations from using the Haar wavelet with thresholds of 50, 100 and 150. These particular pixels were chosen to be used as examples as their spectra exhibit different forms. For comparability, the spectra are plotted using the same scale. One thing that is evident with the approximations produced by the Haar wavelet is that they resemble a series of step functions. This effect is more pronounced at slow changing sections of the spectral curves.

Figure 6.3 shows that the measured spectrum for pixel 6 is well approximated when using the Haar wavelet with thresholds of 50, 100 and 150. When a threshold of 50 was used, the Haar wavelet was able to approximate pixel 6 using 44 wavelet coefficients to an RMS error of 15.40. The RMS error increased to 33.33 and the number of wavelet coefficients decreased to 26 when the threshold was increased to 100. When the threshold was raised to 150 the RMS error increased to 47.37 while the number of wavelet coefficients used decreased to 20.

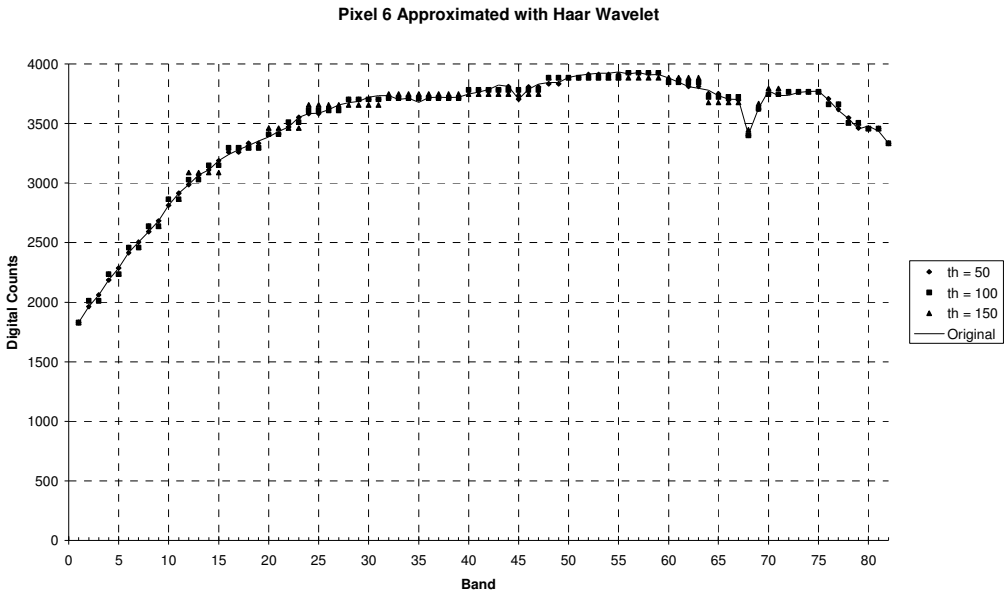


Figure 6.3: Spectra for pixel 6 when approximated with the Haar wavelet using thresholds of 50, 100 and 150.

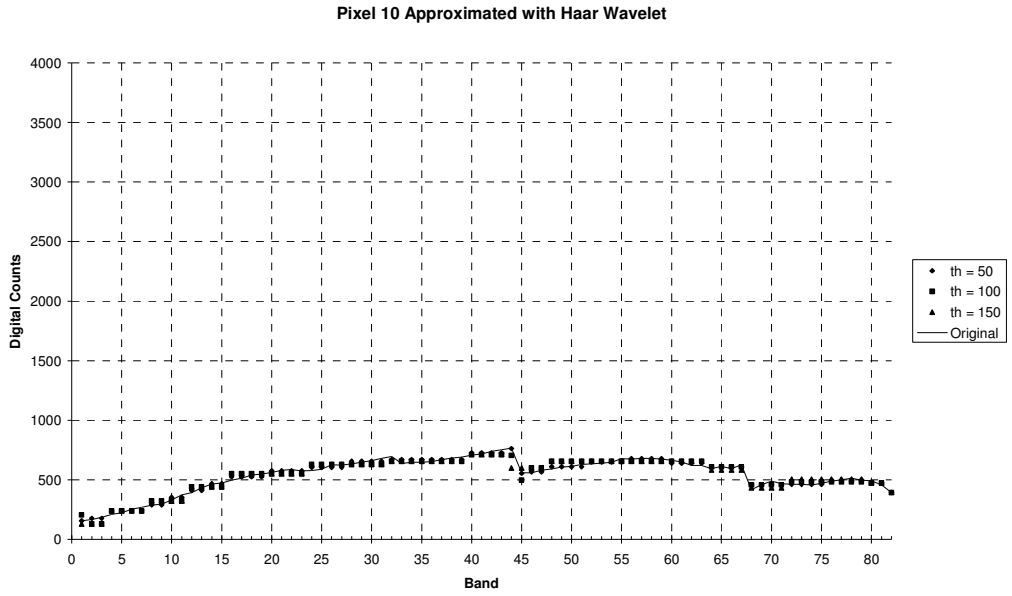


Figure 6.4: Spectra for pixel 10 when approximated with the Haar wavelet using thresholds of 50, 100 and 150.

For all 5 thresholds used, the approximations for pixel 10 (figure 6.4) consistently used the least number of wavelet coefficients and for thresholds of 125 and 150 they also had the lowest RMS error. Of the 20 sample spectra, the spectrum for pixel 10 is relatively flat and void of absorption features. This would explain the fact that the approximations for pixel 10 used the fewest wavelet coefficients: the spectrum does not have many fine details that need to be characterized. The RMS errors of the approximations ranged between 16.44 and 35.39 and the number of wavelet coefficients used ranged from 12 – 26.

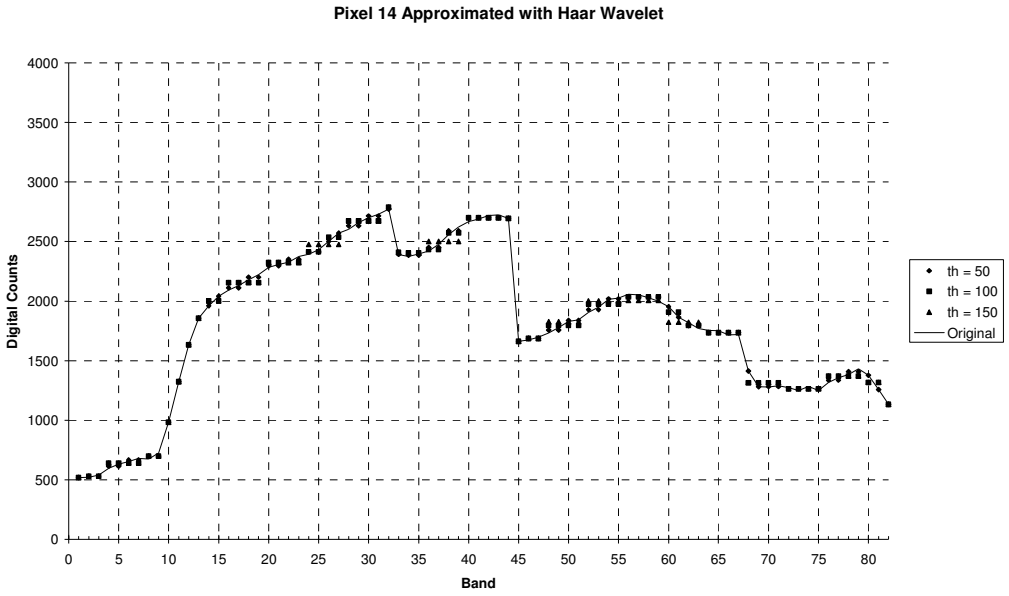


Figure 6.5: Spectra for pixel 14 when approximated with the Haar wavelet using thresholds of 50, 100 and 150.

The spectrum for pixel 14 shows a lot more variability than is present in the spectra for pixels 6 and 10, including a "red edge" feature between bands 10 and 12. As with the other two spectra, the Haar wavelet was able to well approximate the spectrum using as few as 30 and as many as 48 wavelet coefficients for thresholds ranging from 50 to 150. The corresponding RMS errors ranged from 43.81 down to 17.24.

6.2.2 The Daubechies Family of Wavelets

As well as the Haar wavelet, there are many other famous existing orthonormal wavelets such as the Battle-Lemarié wavelets, the Meyer wavelets and the Shannon wavelet. Another popular group of wavelets is the Daubechies family of wavelets. The Daubechies wavelets were developed by Ingrid Daubechies to be orthogonal and have compact support of $0 \leq t < 2N - 1$. Unlike the Haar wavelet, the Daubechies wavelets are continuous and as N increases, the interval of support gets wider and the wavelet functions become smoother (Mallet et al. 1997; Walnut 2002). The number of filter coefficients for each of the Daubechies wavelets is given by $2N$. In this section some examples of the Daubechies wavelets are given and these will then be used to approximate the 20 spectra shown in figure 6.1.

Software packages such as Maple and S-Plus are available which have the capabilities to calculate the scaling functions, wavelet functions and filter coefficients for many of the existing popular wavelets, including those from the Daubechies family. The wavelets module in S-Plus can calculate these for the first N = 10 wavelet functions from the Daubechies family. As an example, the scaling and wavelet functions for N = 2, 3 and 5 are shown in figures 6.6 and 6.7. In this study, the Daubechies scaling functions for N = 2, 3 and 5 will be denoted as ϕ_{D4}, ϕ_{D6} and

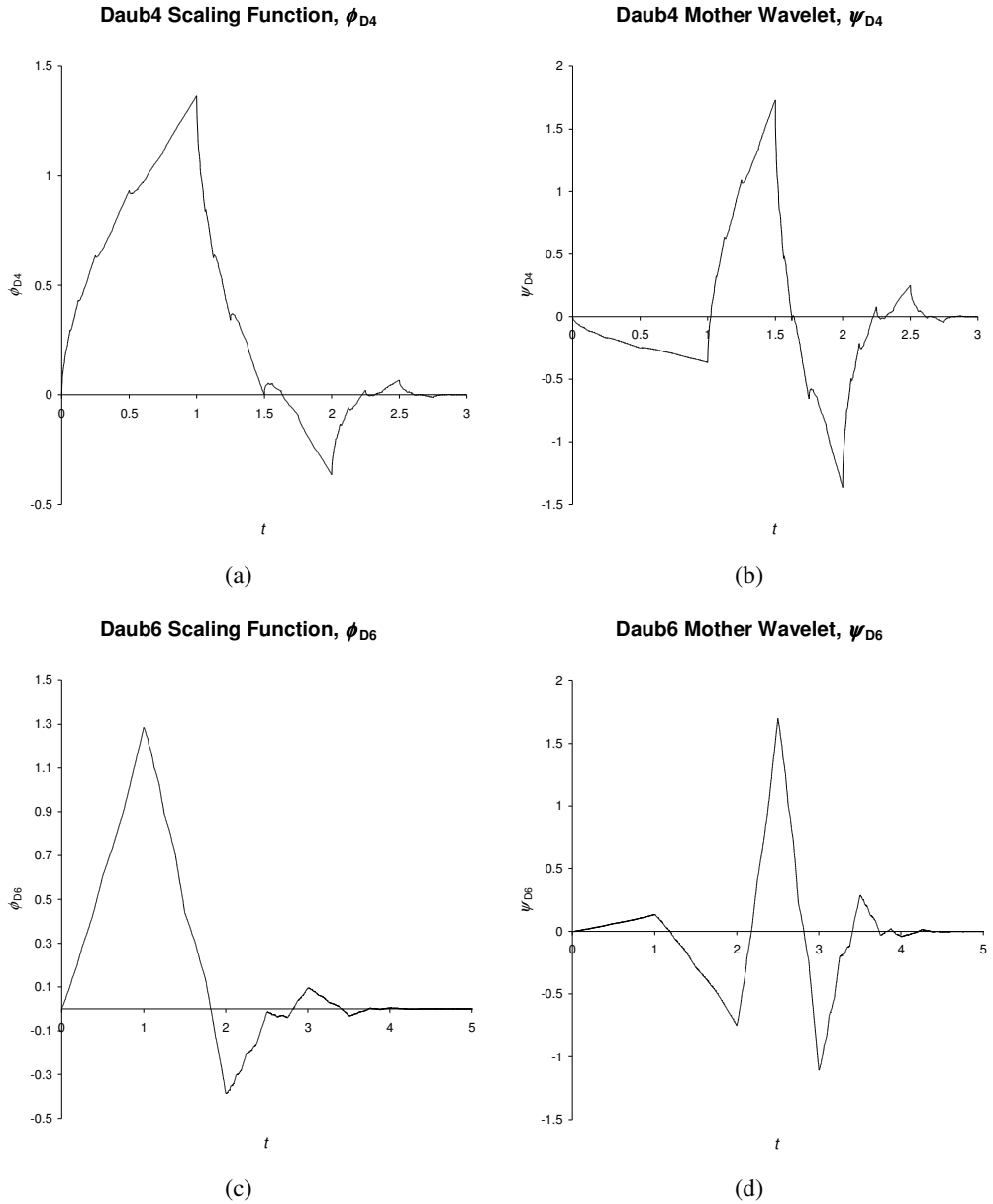
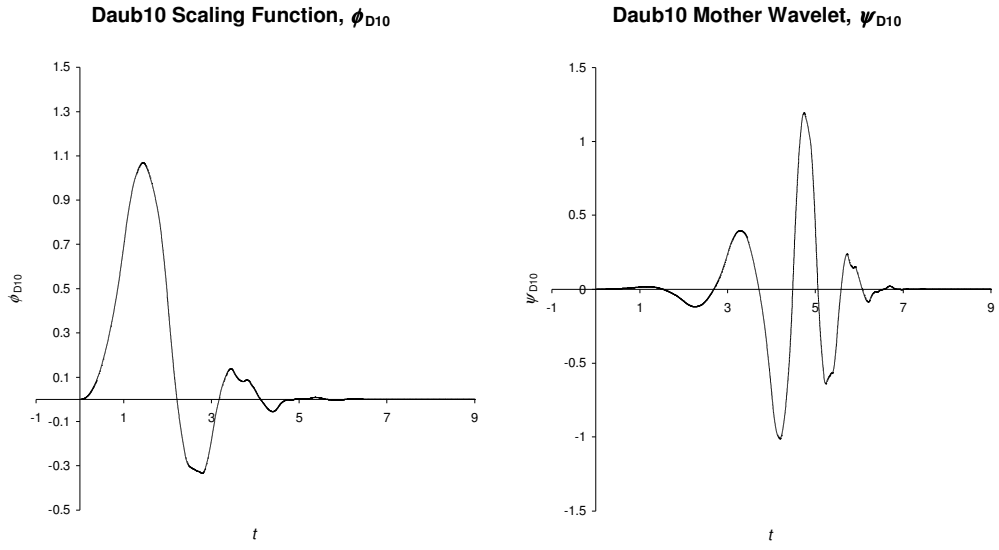


Figure 6.6: Examples of some of the Daubechies Scaling functions and Wavelets. (a) Scaling function for N = 2, ϕ_{D4}; (b) Wavelet function for N = 2, ψ_{D4}; (c) Scaling function for N = 3, ϕ_{D6}; (d) Wavelet function for N = 3, ψ_{D3}.



(a) (b)

Figure 6.7: Examples of some of the Daubechies Scaling functions and Wavelets cont. (a) Scaling function for $N = 5$, ϕ_{D10}; (b) Wavelet function for $N = 5$, ψ_{D10}.

ϕ_{D10} and the corresponding wavelet functions will be denoted as ψ_{D4}, ψ_{D6} and ψ_{D10}. The corresponding filter coefficients for ψ_{D4}, ψ_{D6} and ψ_{D10} are given in table 6.4, and were generated with the wavelets module in S-Plus.

Each of the spectra in figure 6.1 were analysed to $J = 5$ levels using the ψ_{D4}, ψ_{D6} and ψ_{D10} wavelets and were reconstructed by thresholding the wavelet coefficients. The thresholds used were values of 50, 75, 100, 125, and 150. Table 6.5 shows the RMS error resulting from the spectra being approximated by each of the three wavelets for the five different thresholds used. In terms of the RMS errors, the Daubechies wavelets performed significantly better than the Haar wavelet at approximating the

Table 6.4: Low-pass and High-pass Filter Coefficients for the ψ_{D4}, ψ_{D6} and ψ_{D10} Wavelets.

k	ψ_{D4}		ψ_{D6}		ψ_{D10}	
	l_k	h_k	l_k	h_k	l_k	h_k
0	0.482962900	-0.129409500	0.332670600	0.035226290	0.160102400	0.003335725
1	0.836516300	-0.224143900	0.806891500	0.085441270	0.603829300	0.012580750
2	0.224143900	0.836516300	0.459877500	-0.135011000	0.724308500	-0.006241490
3	-0.129409500	-0.482962900	-0.135011000	-0.459877500	0.138428100	-0.077571490
4			-0.085441270	0.806891500	-0.242294900	-0.032244870
5			0.035226290	-0.332670600	-0.032244870	0.242294900
6					0.077571490	0.138428100
7					-0.006241490	-0.724308500
8					-0.012580750	0.603829300
9					0.003335725	-0.160102400

spectra for a given threshold. For a threshold of 50, the Haar wavelet could approximate the spectra with an average RMS error of 17.5. The three Daubechies wavelets approximated the spectra using the same threshold with an average RMS error of 15.1. With a threshold of 150, the improvement was even greater. The Haar wavelet approximated the spectra to an average RMS error of 46.1, compared with the three Daubechies wavelets used which gave an average RMS error of 38.9.

Of the three Daubechies wavelets used, ψ_{D4} was consistently the better performer. For each of the five different thresholds used the spectra approximated using ψ_{D4} had a lower average RMS error than the spectra approximated using either ψ_{D6} or ψ_{D10}. The range in average RMS errors at each threshold was not large, with the largest range 37.98 – 40.13, occurring at a threshold of 150. For the larger valued thresholds used (125 and 150), ψ_{D6} produced approximated spectra having lower RMS errors than those produced by ψ_{D10}. For thresholds of 50, 75 and 100 however, ψ_{D10} was able to approximate the spectra more accurately than ψ_{D6} could.

Table 6.5: RMS Values of Approximated Spectra Using the ψ_{D4} ψ_{D6} and ψ_{D10} Wavelets with Different Thresholds

Pixel	Threshold														
	50			75			100			125			150		
	ψ_{D4}	ψ_{D6}	ψ_{D10}	ψ_{D4}	ψ_{D6}	ψ_{D10}	ψ_{D4}	ψ_{D6}	ψ_{D10}	ψ_{D4}	ψ_{D6}	ψ_{D10}	ψ_{D4}	ψ_{D6}	ψ_{D10}
1	16.02	16.22	14.66	23.45	21.71	20.86	28.47	28.30	26.46	33.52	28.71	34.02	39.86	32.66	41.95
2	13.12	13.89	14.69	20.84	22.75	22.51	29.76	25.91	24.69	29.86	31.69	32.79	29.87	38.54	38.59
3	17.74	11.04	15.31	21.94	23.80	21.18	27.85	29.07	27.15	30.48	29.07	33.03	33.97	32.46	47.51
4	15.37	15.43	16.19	24.73	21.23	19.98	24.75	22.97	27.58	35.06	25.95	34.61	38.67	30.27	41.20
5	14.90	14.37	13.45	19.39	19.81	24.47	24.57	30.54	28.13	34.94	35.68	28.13	40.38	38.31	42.18
6	15.26	13.41	15.27	19.30	19.96	20.25	28.55	23.97	26.46	30.67	34.81	26.46	30.67	34.84	30.67
7	14.53	14.23	14.27	21.94	18.94	18.59	28.18	28.25	23.57	30.58	30.38	32.30	34.28	36.73	40.97
8	19.49	18.52	18.99	27.26	26.21	26.24	31.17	33.30	29.04	33.71	40.23	31.07	40.25	45.41	47.43
9	14.71	14.19	15.37	19.29	21.83	24.60	28.89	27.06	26.86	37.46	37.17	31.40	39.93	44.56	35.14
10	12.31	14.33	15.95	18.84	23.23	22.25	24.54	26.45	28.15	33.11	34.79	36.67	39.02	38.40	36.67
11	14.37	17.03	14.21	22.74	25.97	18.27	28.90	27.18	29.45	28.90	33.81	40.29	36.51	39.99	45.44
12	13.88	15.94	15.89	20.87	20.30	23.86	24.06	24.94	30.64	36.86	31.92	34.83	50.58	38.24	41.33
13	11.05	12.71	11.04	18.93	23.77	21.78	22.77	23.65	29.26	31.54	29.23	34.49	44.85	35.87	38.25
14	13.87	14.56	11.79	21.57	19.57	17.30	25.72	29.53	30.43	33.18	37.70	33.42	42.07	46.28	34.18
15	14.65	14.69	14.71	20.99	19.19	20.43	29.32	29.20	24.43	29.35	36.87	39.94	35.99	41.78	48.64
16	14.83	15.92	13.23	18.19	23.68	18.67	27.53	28.07	24.54	32.90	28.43	36.93	36.79	41.48	39.70
17	18.05	20.06	18.78	21.04	28.68	24.15	32.39	31.39	29.38	34.91	40.15	36.56	47.78	43.46	37.76
18	13.48	13.38	14.40	19.03	23.63	19.10	24.15	29.08	27.11	30.08	32.53	29.74	33.38	38.75	29.74
19	15.77	16.99	16.82	19.96	24.94	25.01	22.64	27.60	34.18	28.54	30.84	36.13	28.70	37.37	45.39
20	16.65	17.05	16.26	24.20	24.40	20.74	25.61	28.28	26.71	28.32	33.30	29.85	36.07	36.96	39.92
Average	15.00	15.20	15.06	21.23	22.68	21.51	26.99	27.74	27.71	32.20	33.16	33.63	37.98	38.62	40.13

In terms of reducing the number of terms needed to accurately characterize the spectra, ψ_{D4} performed better than ψ_{D6} and ψ_{D10}. Table 6.6 shows the number of wavelet coefficients needed to approximate the spectra for a given threshold. It shows that on average, ψ_{D4} used fewer wavelet coefficients to approximate the spectra than did ψ_{D6} and ψ_{D10}, particularly with the lower valued thresholds. When compared with the Haar wavelet, the results are mixed. For the larger thresholds (100, 125 and 150), the Haar wavelet outperformed the Daubechies wavelets, although for thresholds of 50 and 75, the Daubechies wavelets performed better. It should be noted that table 6.6 does not include the number of approximation coefficients that need to be stored to initialize the synthesis phase to approximate the spectra. For ψ_{D4}, the number of approximation coefficients that need to be stored after $J = 5$ levels of analysis is 5. For ψ_{D6} this increases to 7 and for ψ_{D10} the number of stored a coefficients needed is 11. When taking this into account, ψ_{D4} is clearly superior over ψ_{D6} and ψ_{D10} at approximating the spectra, both in terms of accuracy and the number of parameters needed.

To demonstrate some of the results graphically, figures 6.8, 6.9 and 6.10 show the

Table 6.6: Number of Wavelet Coefficients Used to Approximate Spectra Using the ψ_{D4} ψ_{D6} and ψ_{D10} Wavelets with Different Thresholds. These Values do not Include the Approximation Coefficients Needed to Initialize the Synthesis Process.

Pixel	Threshold														
	50			75			100			125			150		
	ψ_{D4}	ψ_{D6}	ψ_{D10}	ψ_{D4}	ψ_{D6}	ψ_{D10}	ψ_{D4}	ψ_{D6}	ψ_{D10}	ψ_{D4}	ψ_{D6}	ψ_{D10}	ψ_{D4}	ψ_{D6}	ψ_{D10}
1	42	42	52	36	36	47	33	32	40	29	31	37	27	28	32
2	43	44	53	36	35	45	31	33	40	30	31	37	29	27	34
3	38	46	53	34	36	47	31	31	44	30	31	39	28	29	32
4	39	41	51	33	35	48	32	33	43	28	30	39	27	29	34
5	37	43	53	34	36	43	30	29	40	26	26	40	24	24	34
6	36	39	45	33	33	42	28	30	39	26	26	38	26	25	36
7	38	42	47	31	38	43	28	30	41	27	29	35	25	27	30
8	42	48	49	34	38	41	30	32	35	28	27	33	25	25	28
9	31	32	35	26	24	23	20	21	19	15	15	16	14	10	15
10	26	25	24	20	16	17	15	11	14	12	8	10	10	6	10
11	29	30	30	20	20	25	16	16	19	16	13	14	14	10	12
12	43	41	47	37	34	39	35	32	31	28	28	28	22	26	26
13	46	48	53	38	37	43	36	36	38	33	34	35	27	31	31
14	45	48	54	38	43	45	34	37	37	31	31	35	28	28	32
15	45	48	51	40	43	44	34	37	40	32	33	32	30	29	26
16	38	41	50	35	34	44	29	31	39	27	29	31	24	25	30
17	35	38	41	31	26	32	24	22	27	23	18	21	17	16	19
18	27	26	31	22	17	24	19	14	16	16	10	14	14	7	14
19	35	37	47	30	26	37	27	23	29	25	22	25	24	19	22
20	38	39	49	30	32	44	29	28	40	27	25	37	25	23	33
Average	38	40	46	32	32	39	28	28	34	25	25	30	23	22	27

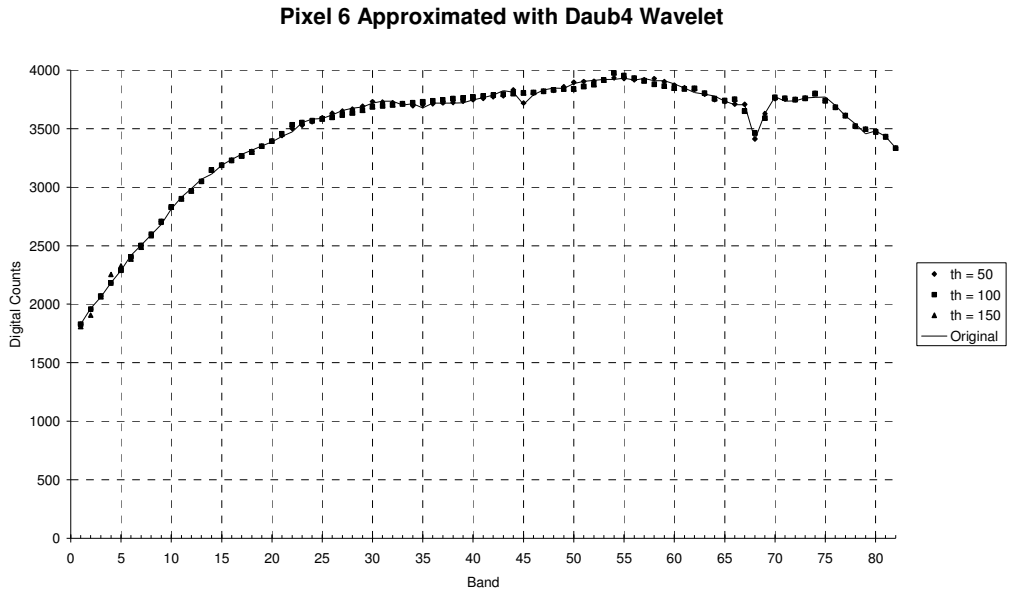


Figure 6.8: Pixel 6 spectra: original and approximated using ψ_{D4} wavelet with thresholds of 50, 100 and 150.

spectra for pixels 6, 10 and 14, along with the spectra approximated using ψ_{D4} with thresholds of 50, 100 and 150. The spectra have been plotted using the same scale for comparability. When the ψ_{D4} wavelet was used to approximate the measured spectrum for pixel 6, the number of wavelet coefficients used was 36, 28 and 26 for thresholds of 50, 100 and 150 respectively. At these thresholds the corresponding RMS errors were 15.26, 28.55 and 30.67. These approximations were more accurate than those produced by the Haar wavelet although more parameters were required with thresholds of 75 – 150. The improvement in accuracy resulting from the use of ψ_{D4} can be seen in figure 6.8. The shape of the spectrum has been better captured and it is no longer being approximated as a series of step functions, as compared to figure 6.3.

When the ψ_{D4} wavelet was used to approximate the measured spectrum for pixel 10, the results produced (figure 6.9) were more accurate than those for the Haar wavelet at lower valued thresholds, but not for thresholds of 125 and 150. It can be seen in figure 6.9 that for a threshold of 150, the approximation does not adequately capture the sharp features in the spectrum. This would indicate that 150 is too high a threshold to represent these types of features. In terms of parameterisation, fewer

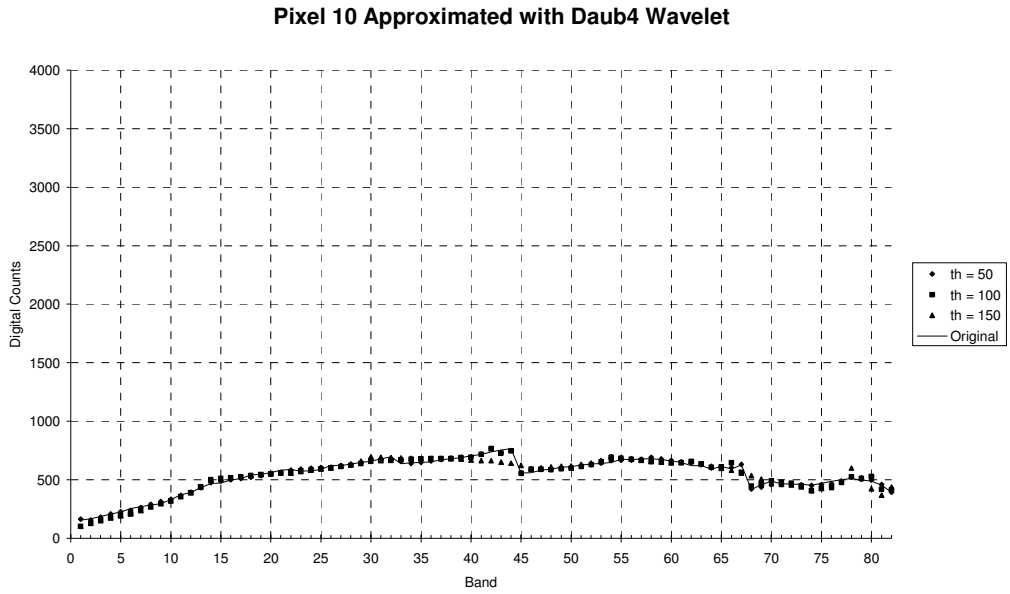


Figure 6.9: Pixel 10 spectra: original and approximated using ψ_{D4} wavelet with thresholds of 50, 100 and 150.

wavelet coefficients are used with the ψ_{D4} wavelet for thresholds of 125 and 150 than when the Haar wavelet was applied.

When used to approximate pixel 14, the ψ_{D4} wavelet performs better than the Haar wavelet for all thresholds trialled, both in terms of accuracy and reducing the number of wavelet coefficients needed. The results from using ψ_{D4} to approximate pixel 14 with thresholds of 50, 100 and 150 are shown in figure 6.10. The spectral curve exhibits many different features and these are all well characterized by the ψ_{D4} wavelet.

It has been shown that of the three members from the Daubechies family of wavelets used, the ψ_{D4} wavelet performs better at accurately approximating remotely sensed spectra. The ψ_{D4} wavelet also uses fewer parameters to characterize the spectra than both ψ_{D6} and ψ_{D10}, particularly when the number of approximation coefficients used is considered.

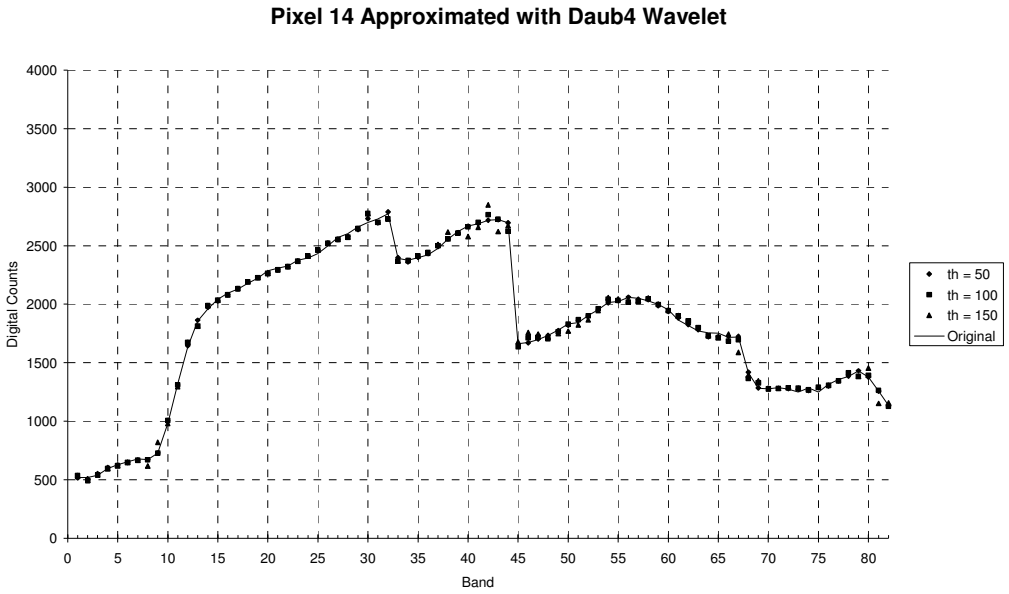


Figure 6.10: Pixel 14 spectra: original and approximated using ψ_{D4} wavelet with thresholds of 50, 100 and 150.

6.3 Approximating Spectra Using Cubic Splines

Spectra can also be characterised using fewer parameters by approximating the spectra with piecewise polynomials. The spectra are divided into subintervals and subject to continuity conditions at the join points of the subintervals, or knots, the spectra are fitted with a different polynomial at each subinterval. A subset of the bands is chosen for the knots and the rest can be interpolated from the fitted polynomials. Typically, more knots are selected in rapidly changing regions of the curve.

When linear functions are used to approximate the spectral curves, the resulting functions will not have continuous first derivatives at the knots and hence are not smooth curves. Quadratic functions are the simplest polynomials that will result in smooth curves and can be defined with only three constants. However, cubic polynomials are most commonly used as they are the lowest order polynomials that, unlike quadratic functions, give points of inflection. For these reasons, cubic splines will be used in this work. Given a function f defined on the interval $[a, b]$ and a set of $n + 1$ knots, $a = x_0 < x_1 < \ldots < x_n = b$, cubic splines, S, are fitted to the n subintervals so that the following continuity conditions are satisfied:

1) S_j is a cubic polynomial defined on the subinterval $[x_j, x_{j+1}]$ for each $j = 0, 1, \ldots, n-1$;
2) $S(x_j) = f(x_j)$ for each $j = 0, 1, \ldots, n$;
3) $S_{j+1}(x_{j+1}) = S_j(x_{j+1})$ for each $j = 0, 1, \ldots, n-2$;
4) $S'_{j+1}(x_{j+1}) = S'_j(x_{j+1})$ for each $j = 0, 1, \ldots, n-2$;
5) $S''_{j+1}(x_{j+1}) = S''_j(x_{j+1})$ for each $j = 0, 1, \ldots, n-2$;
6) Either $S'(x_0) = f'(x_0)$ and $S'(x_n) = f'(x_n)$ for clamped boundary conditions or $S''(x_0) = S''(x_n) = 0$ for free or natural boundary conditions.

For higher order polynomials, there are additional continuity conditions for each additional derivative. Discontinuities will exist at the highest order derivatives for each knot, regardless of the degree of the polynomials used at each subinterval. If the highest order derivatives were continuous at a knot, then this would imply the polynomials fitted to the subintervals adjoining the knot are identical and the knot would then be superfluous.

With selected knots, cubic splines are constructed by applying the continuity conditions to the cubic polynomials. For each subinterval, $j = 0, 1, \ldots, n-1$

$$S_j(x) = a_j + b_j(x - x_j) + c_j(x - x_j)^2 + d_j(x - x_j)^3, \tag{6.22}$$

where clearly at x_j,

$$S_j(x_j) = f(x_j) = a_j. \tag{6.23}$$

Applying condition 3) for each $j = 0, 1, \ldots, n-2$ gives

$$a_{j+1} = S_{j+1}(x_{j+1}) = S_j(x_{j+1}) = a_j + b_j(\Delta_j) + c_j(\Delta_j)^2 + d_j(\Delta_j)^3 \tag{6.24}$$

where $\Delta_j = x_{j+1} - x_j$. Similarly as with a_j, $b_j = S'_j(x_j)$, and applying condition 4) for each $j = 0, 1, \ldots, n-1$ gives

$$b_{j+1} = b_j + 2c_j(\Delta_j) + 3d_j(\Delta_j)^2. \tag{6.25}$$

Applying condition 5) for each $j = 0, 1, \ldots, n-1$ gives

$$c_{j+1} = c_j + 3d_j\Delta_j. \tag{6.26}$$

Equation (6.26) can be rearranged to solve for d_j, and this is then substituted into (6.24) and (6.25), to yield,

$$a_{j+1} = a_j + b_j\Delta_j + \frac{\Delta_j^2}{3}(2c_j + c_{j+1}) \tag{6.27}$$

and

$$b_{j+1} = b_j + \Delta_j(c_j + c_{j+1}). \tag{6.28}$$

Equation (6.27) can be rearranged to solve for b_j and also for b_{j-1} by reducing the index. If the resulting equations are substituted into (6.28) after reducing the indices by 1, the resulting equation is

$$\Delta_{j-1}c_{j-1} + 2(\Delta_{j-1} + \Delta_j)c_j + \Delta_j c_{j+1} = \frac{3}{\Delta_j}(a_{j+1} - a_j) - \frac{3}{\Delta_{j-1}}(a_j - a_{j-1}). \tag{6.29}$$

If one of the boundary conditions from condition 6) is applied and it is recognized that $c_n = S''(x_n)/2$, the c_j can be calculated by solving a linear system of the form $\mathbf{Ax} = \mathbf{b}$, where in the case of $\mathbf{A}$ is the $(n + 1) \times (n + 1)$ matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & \cdots & \cdots & 0 \\ \Delta_0 & 2(\Delta_0 + \Delta_1) & \Delta_1 & & & \vdots \\ 0 & \Delta_1 & 2(\Delta_1 + \Delta_2) & \Delta_2 & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & 0 \\ \vdots & & & \Delta_{n-2} & 2(\Delta_{n-2} + \Delta_{n-1}) & \Delta_{n-1} \\ 0 & \cdots & \cdots & 0 & 0 & 1 \end{bmatrix}$$

for natural cubic splines, $\mathbf{x}$ and $\mathbf{b}$ are the $(n + 1) \times 1$ vectors

$$\mathbf{x} = \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_n \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 0 \\ \frac{3}{\Delta_1}(a_2 - a_1) - \frac{3}{\Delta_0}(a_1 - a_0) \\ \vdots \\ \frac{3}{\Delta_{n-1}}(a_n - a_{n-1}) - \frac{3}{\Delta_{n-2}}(a_{n-1} - a_{n-2}) \\ 0 \end{bmatrix}.$$

Once the c_j coefficients have been obtained, it is then simply a matter of substituting them into (6.27) and (6.26) to get the b_j and d_j coefficients respectively. Press et al. (1986) provide standard subroutines that will calculate the coefficients a_j, b_j, c_j and d_j and fit polynomial splines to the data. These subroutines are available in most of the scientific languages (e.g. FORTRAN, Pascal and C) and there are commercial packages available that have there own spline routines included. One example is Maple, which will fit natural splines to the data given a set of knots and offers a choice as to the degree of polynomial to use.

Thus, given a set of knots, it is a relatively straightforward process to fit cubic splines to spectral data. It remains to find a suitable knot selection strategy that will approximate spectra with a minimised RMS error. With hyperspectral data it is impractical to trial every possible combination of knots from the available list of bands in order to find the best combination. For example, if 40 knots were to be chosen from a list of 80 bands, the number of possible combinations would be more than 1.075×10^{23}. If only 20 knots were to be selected, the number of combinations would reduce to 3.535×10^{18}. This is too many combinations to have to trial and assess. The most convenient method is to select the knots at evenly spaced intervals, although this is unlikely to select the most optimal set of knots.

A review of the different knot selection schemes that have been used in the statistical literature is given in Wand (2000). More recently, He, Shen & Shen (2001) presented a two stage method that recognises the fact that more knots are required in rapidly changing regions of the curve. He, Shen & Shen (2001) propose that a wavelet decomposition scheme be used to find an initial subset of knots and then reduce this subset using a stepwise knot deletion scheme.

Cubic splines have found applications in hyperspectral remote sensing. Furby, Kiiveri and Campbell (1990) used them for analysis of a hyperspectral data set prior to executing a CVA. After the data had been approximated by cubic splines, it was found that important spectral information was retained. Gao, Liu and Davis (1998) used cubic splines to smooth hyperspectral AVIRIS data for geological applications. After atmospheric correction of the data, small spikes are systematically present in the spectral curves, which can mimic important spectral absorption or reflectance features and can be a distraction when analysing the data. The authors successfully used cubic smoothing splines to smooth the data and retain the significant mineral features.

The results from fitting cubic splines to the data shown in figure 6.1 are presented next. At first, evenly spaced knots are used with cubic splines to approximate the spectra. These results are then compared with those from a forward stepwise knot selection technique. Starting with the first and last band as end points, each additional knot is selected as the band which approximates the spectrum with the lowest RMS error. This is continued until the required number of knots has been selected. A backward stepwise knot deletion scheme is then trialled, where initially all available bands are used as knots and one by one a knot is subsequently removed. The knot to be removed is that for which all other remaining knots produce an approximated spectrum having the lowest RMS error. This is continued until the full set of knots has been reduced to the required number. Either the band numbers or the wavelengths could be supplied as knots to the spline fitting procedure. In order for the results to be comparable with those from approximating the spectra using wavelets, the band numbers will be used as knots and not the band wavelengths.

6.3.1 Cubic Splines Fitted with Regularly Spaced Knots

When the number of knots to be used for a cubic spline fit of the data is known, the most straightforward knot selection technique is to select the knots at evenly spaced intervals. If n +1 knots are to be selected from N bands, the band number x_j to be selected as knot j (j = 0, 1, …, n) is given by

$$x_j = \frac{j(N-1)}{n} + 1$$

and x_j is rounded off to the nearest integer to correspond with the nearest band number.

A cubic polynomial is characterized by four parameters. If the number of parameters needed to approximate the spectra is to be reduced then each spectral curve needs to be approximated using a maximum of 20 knots. The spectra shown in figure 6.1 were approximated using cubic splines with 16 – 20 regularly spaced knots. The results are summarised in table 6.7, which shows that of the five combinations of knots used, the best result was achieved on average using 19 regularly spaced knots. When 19 knots were used to approximate the spectra the average RMS error was 58.4. The largest average RMS error of 85.0 occurred when 18 regularly spaced knots were used. Of the 20 spectra approximated, pixel 18 was approximated with the most accuracy. The RMS errors for this pixel ranged from 21.8 when 19 regularly spaced knots were used to 32.6 when the number of regularly spaced knots used was 18. Pixel 13, conversely, was approximated by cubic splines the least accurately; with the RMS errors ranging from 124.5 – 206.2 when the number of

Table 6.7: RMS Errors of Spectra Approximated Using Cubic Splines with Knots Spaced at Regular Intervals

Pixel	Number of Regularly Spaced Knots				
	20	19	18	17	16
1	96.55	78.00	129.04	85.59	119.98
2	79.61	66.05	101.79	71.33	99.27
3	59.16	52.06	70.04	54.40	71.21
4	54.74	47.08	73.38	53.81	65.97
5	40.14	39.41	40.36	33.21	45.39
6	36.40	35.23	58.14	42.45	44.19
7	50.59	50.57	86.80	61.10	63.81
8	72.41	68.40	120.96	83.56	78.79
9	38.13	33.79	56.88	39.20	45.53
10	30.85	28.04	41.70	31.75	37.83
11	47.65	42.84	53.34	37.57	58.84
12	116.40	100.48	140.77	95.84	159.77
13	145.11	126.77	176.87	124.49	206.19
14	124.62	109.09	148.63	105.92	174.93
15	112.46	93.53	139.95	95.68	151.69
16	53.83	45.69	71.94	53.31	72.84
17	45.61	42.97	70.75	48.40	51.25
18	23.22	21.78	32.61	24.91	29.61
19	50.83	41.01	46.07	37.09	56.80
20	52.53	44.75	40.24	39.18	54.85
Mean	66.54	58.38	85.01	60.94	84.44

regularly spaced knots used was 17 and 16 respectively.

Given the set of knots, the process of fitting a cubic spline approximation is quite fast. Hence, instead of storing the cubic spline coefficients, it may be desirable to save only the knots and the associated data values. The cubic splines could then be generated for each spectral curve at a later date. In this case, it is now possible to reduce the number of stored parameters by fitting cubic splines with up to 40 knots. With more knots, it would be expected that the spectra could be approximated more accurately. Table 6.8 summarises the results when the measured pixel spectra are approximated using cubic splines with 36 – 40 regularly spaced knots. The accuracy of the approximations has improved significantly, with the largest average RMS error decreasing from 85.0 to 48.8. The lowest average RMS error was 42.3, and this was achieved using 37 knots spaced at regular intervals. Pixel 18 was approximated the most accurately, having RMS errors ranging from 13.8 – 20.6 when the number of knots used was 37 and 40 respectively. Pixel 13 was again approximated with the least accuracy for all combinations of knots used. In this instance the RMS errors ranged from 78.8 – 87.7 when the number of knots used was 39 and 40 respectively.

Table 6.8: RMS Errors of Spectra Approximated Using Cubic Splines with Knots Spaced at Regular Intervals

Pixel	Number of Regularly Spaced Knots				
	40	39	38	37	36
1	72.79	63.04	64.85	60.02	65.29
2	58.75	48.27	51.24	46.48	51.82
3	43.68	33.05	35.56	32.94	36.31
4	47.44	38.85	40.07	36.51	40.49
5	25.26	21.33	22.76	25.29	24.54
6	40.86	33.79	32.91	30.57	32.20
7	53.27	42.30	42.32	36.84	41.56
8	80.12	70.71	69.84	66.25	70.35
9	34.50	29.30	29.72	26.88	29.61
10	26.23	20.91	22.05	19.53	22.39
11	30.47	27.90	30.44	32.12	30.81
12	73.29	66.71	71.43	72.62	71.16
13	87.66	78.81	84.52	83.97	85.60
14	73.46	65.73	71.42	70.53	72.63
15	71.36	63.77	67.84	65.48	67.72
16	39.50	33.67	34.70	30.84	35.17
17	43.75	41.39	41.93	39.84	41.54
18	20.57	15.63	16.73	13.78	16.80
19	26.87	25.56	27.33	27.73	27.23
20	25.68	26.51	27.10	28.23	27.58
Mean	48.78	42.36	44.24	42.32	44.54

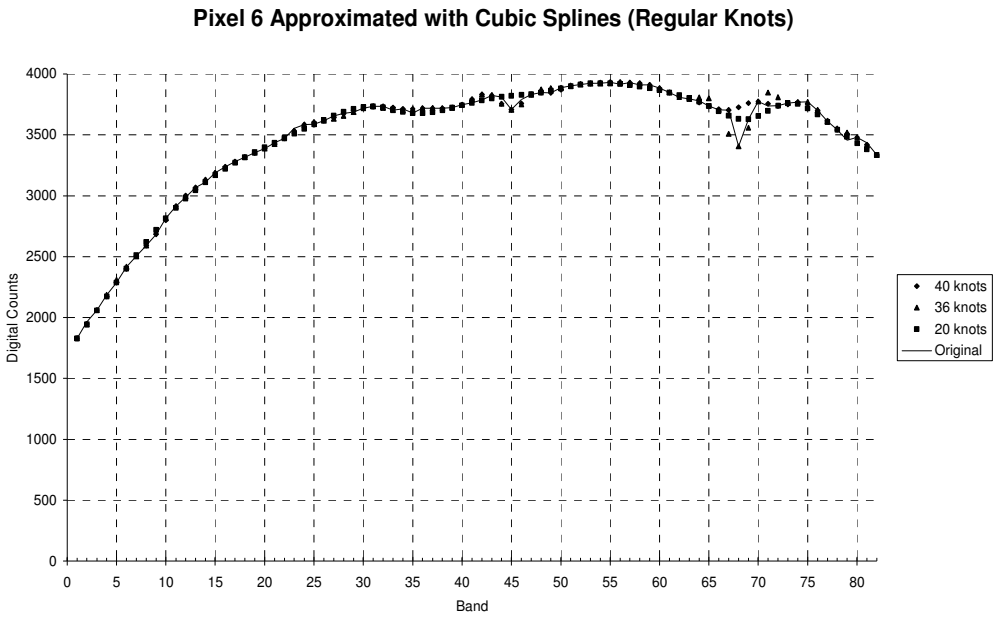


Figure 6.11: Pixel 6 spectra: original and approximated using cubic splines with 20, 36 and 40 knots spaced at regular intervals.

To illustrate the results graphically, figures 6.11, 6.12 and 6.13 show the spectra from approximating pixels 6, 10 and 14 respectively using cubic splines with 20, 36 and 40 regularly spaced knots. The plots generally show good agreement between the original and approximated spectra. The main anomalies between the measured

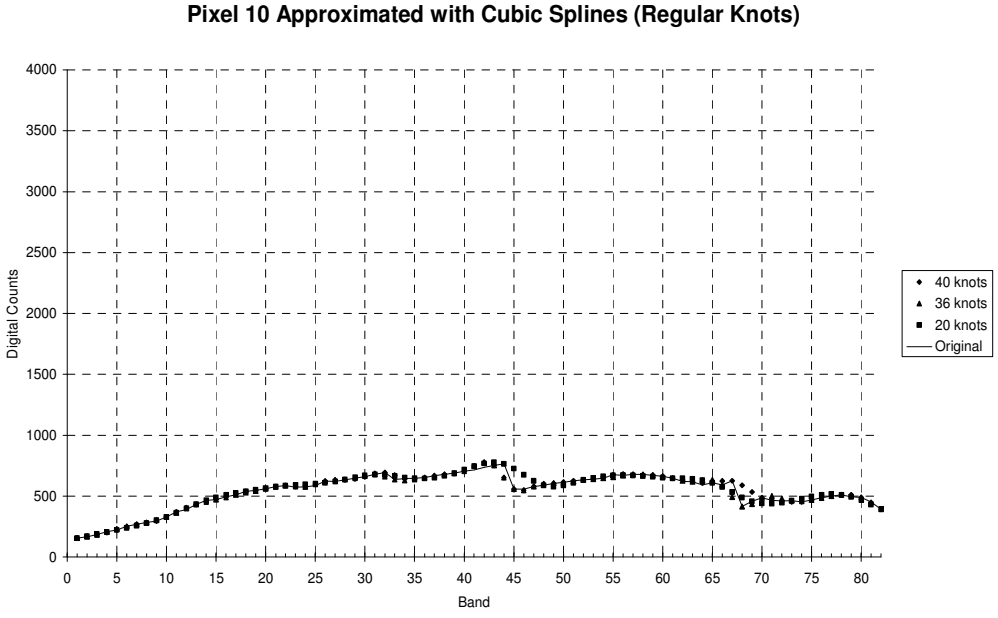


Figure 6.12: Pixel 10 spectra: original and approximated using cubic splines with 20, 36 and 40 knots spaced at regular intervals.

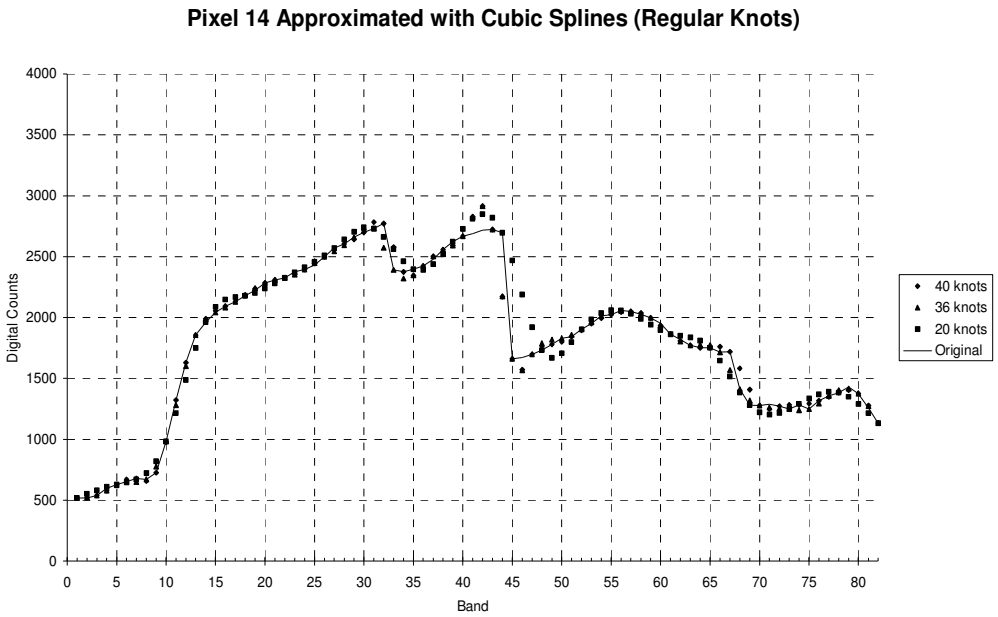


Figure 6.13: Pixel 14 spectra: original and approximated using cubic splines with 20, 36 and 40 knots spaced at regular intervals.

and the approximated spectra occur where there are rapid changes in the measured spectra, particularly near bands 45 and 68. This is the main disadvantage to using knots at regularly spaced intervals. Ideally, more knots should be located in regions where the spectral curve undergoes rapid changes and fewer knots in regions where the curve changes slowly.

When compared with the results from the measured spectra approximated with wavelets, the results from splines are significantly more inaccurate. Even when a threshold of 150 was used with the Haar wavelet to set the wavelet coefficients to zero, the average RMS error was as low as 46.1. Cubic splines with 40 evenly spaced knots could only achieve approximations with an average RMS error of 48.8.

The pixels that were approximated with the least accuracy (pixels 12 – 15) are those whose spectra possess the largest range between their maximum and minimum values. Another characteristic of the spectra that were not as accurately approximated by cubic splines is that they exhibit many undulating features and rapidly varying regions within the curves. These accuracies could theoretically be improved by locating the knots where the spectral curves are changing faster than in other regions. In the next section, a knot selection scheme is trialled that commences

with the end points and sequentially selects the next knot that gives the lowest RMS error.

6.3.2 Cubic Splines Using Forward Stepwise Knot Selection

In the previous section the spectra shown in figure 6.1 were approximated using cubic splines with regularly spaced knots. The resulting approximations gave generally good fits to the spectral curves, although in rapidly varying regions of the curves, the approximations did not perform as well. In this section, the spectra are again approximated using cubic splines, this time with knots selected from a forward stepwise selection technique. Stepwise selection techniques are reviewed in Wand (2000) and have been used with hyperspectral data in Furby, Kiiveri and Campbell (1990) and He, Shen and Shen (2001). The method begins with bands 1 and N as the endpoints and selects the next knot as the band which gives the approximation having the smallest RMS error. This process continues until a predetermined number of knots have been selected.

The results from fitting the spectra using cubic splines with knots selected from the

Table 6.9: RMS Errors of Spectra Approximated Using Cubic Splines with Knots Selected with a Forward Stepwise Procedure

Pixel	Number of Knots				
	40	39	38	37	36
1	47.95	48.01	48.09	48.46	48.52
2	43.40	43.49	43.59	43.81	43.92
3	38.91	38.91	38.91	38.92	38.93
4	45.16	47.27	52.15	53.97	65.92
5	20.29	21.04	21.25	22.31	22.37
6	36.92	37.88	38.35	38.88	50.52
7	30.55	30.60	30.69	30.80	30.85
8	60.04	60.08	60.13	60.37	60.63
9	28.19	28.26	28.30	28.32	28.35
10	22.31	22.43	22.43	22.43	22.44
11	53.07	53.12	53.11	53.11	53.11
12	53.74	54.99	56.07	56.51	56.84
13	182.53	182.53	182.53	182.53	182.53
14	64.70	64.82	64.99	65.88	67.82
15	87.51	87.51	87.52	87.54	87.55
16	34.04	34.06	34.08	34.11	34.14
17	41.60	41.60	41.60	41.61	41.65
18	14.53	14.88	15.09	15.71	15.75
19	70.95	70.95	70.95	71.06	70.96
20	29.41	29.45	29.49	29.52	29.57
Mean	50.29	50.59	50.97	51.29	52.62

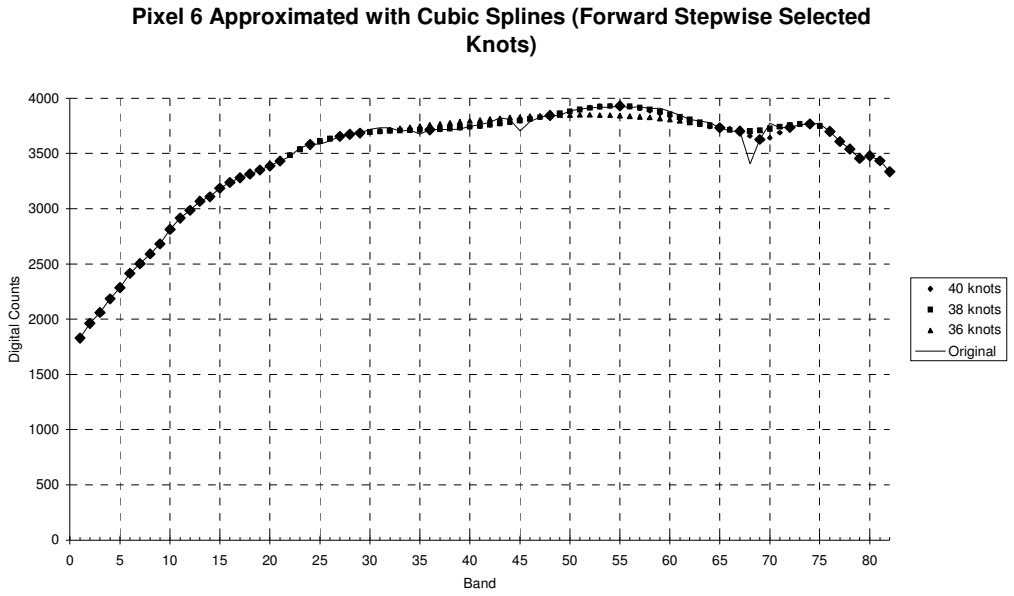


Figure 6.14: Pixel 6 spectra: original and approximated using cubic splines with 40 knots selected using a forward stepwise selection technique. The positions of the 40 knots selected are shown as larger diamonds.

forward stepwise selection scheme are summarised in table 6.9. Unexpectedly, the results show that for the same number of knots there has been a slight decrease in the overall accuracy of the approximations when compared to those fitted with splines with regularly spaced knots. For some of the pixels the accuracy has improved, but in the majority of cases the spectra have actually been approximated with less accuracy.

The reason for the decrease in accuracy resulting from the use of the forward stepwise selection scheme is likely to be due to the fact that the technique only selects one knot at a time. A fluctuating region within a spectral curve will typically need to be fitted using several knots. By starting to fit the fluctuating section of the curve one knot at a time, the RMS error of the approximation to that section of the curve will possibly increase until a sufficient number of knots have been selected to fit the curve adequately. The forward stepwise selection technique chooses the next knot according to which band will give the lowest RMS error. Given the choice of choosing a knot in a relatively linear section of the curve over starting to fit a fluctuating section of the curve which will initially result in a higher RMS error, naturally the knot selection technique will opt for the knot within a linear section of the curve. A possible solution, therefore, is to use all available bands as knots and

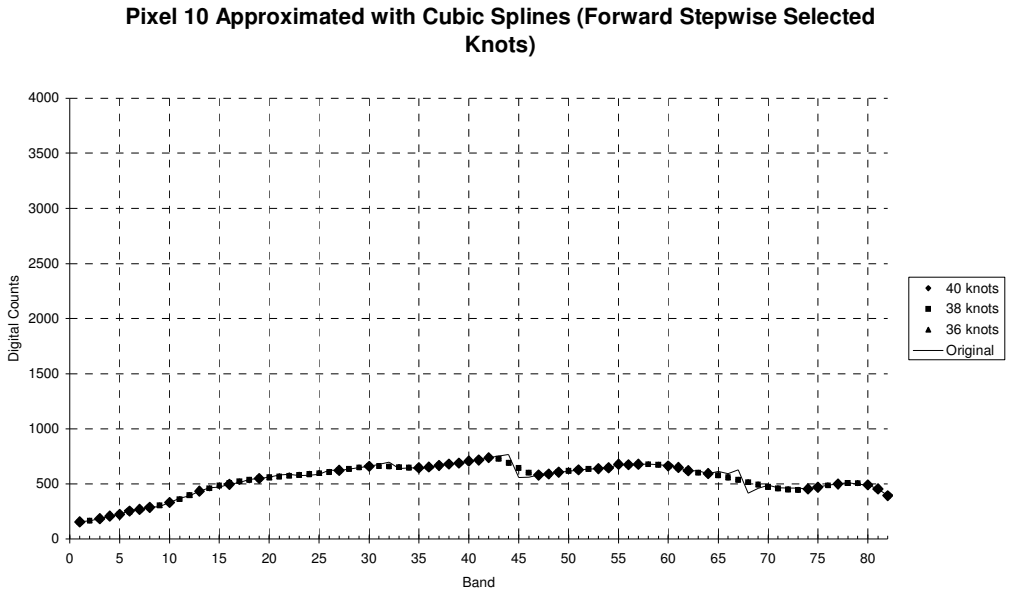


Figure 6.15: Pixel 10 spectra: original and approximated using cubic splines with 40 knots selected using a forward stepwise selection technique. The positions of the 40 knots selected are shown as larger diamonds.

one by one remove those knots which are redundant, i.e., a backwards knot deletion scheme.

The results for pixels 6, 10 and 14 are shown graphically in figures 6.14, 6.15 and

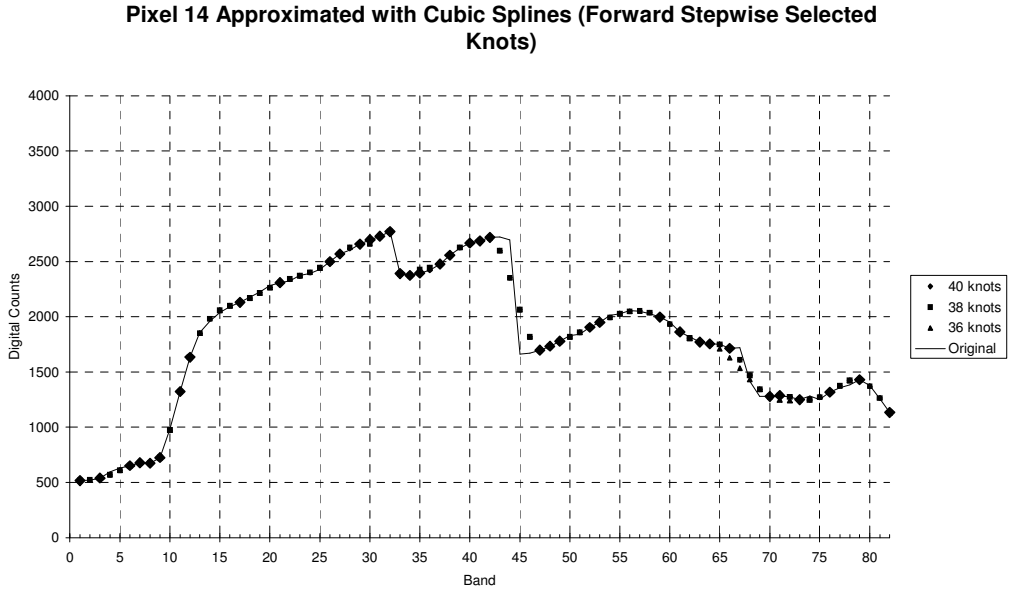


Figure 6.16: Pixel 14 spectra: original and approximated using cubic splines with 40, 38 and 36 knots selected using a forward stepwise selection technique. The positions of the 40 knots selected are shown as larger diamonds.

6.16 respectively. The locations of the first 40 knots selected for each pixel are shown as the larger diamonds along the original pixel spectra. The plots appear similar to those generated from using regularly spaced knots. One obvious difference occurs when pixel 6 is approximated using 36 knots selected using the forward stepwise technique. There is an absence of selected knots between bands 48 and 65 and consequently the convex nature between these bands has not been sufficiently captured. Band 55 was selected as the 37th knot and it can be seen that the fit to the convex feature has been improved in the plots for 38 and 40 knots.

6.3.3 Cubic Splines Using Backward Stepwise Knot Selection

Fitting cubic splines by selecting the knots with a forward stepwise approach actually gave less accurate results than from using regularly spaced knots. Given a set of knots that had already been chosen, the subsequent knot to be selected would be the one which resulted in the minimum RMS error in the approximation to the spectral curve being fitted. This resulted in knots being selected in regions of the curve that were already well fitted, and contradictory to the objective, would actually increase the redundancy. In this section the reverse approach is taken. The spectral curves are fitted using the full set of available bands as knots, and one by one knots are subsequently removed from the fitted curve to minimize the RMS error in the resulting approximation to the curve. All the "redundant" knots are removed from the well approximated regions of the spectral curves first, leaving the more essential knots to fit the spectra.

When the backwards knot deletion method was used to select a subset of knots to approximate the spectra shown in figure 6.1, the results were significantly improved upon when compared with those from the forward stepwise selection procedure. In most cases, even using only 20 knots resulted in an improvement in the RMS errors from the approximated spectral curves when fitted with 40 regularly spaced or forward stepwise selected knots. A subset of the results from using the backwards knot deletion method is summarised in table 6.10. The table shows the RMS errors for the 20 pixel spectra when approximated using 40, 35, 30, 25 and 20 knots. When using 40 knots, the average RMS error for the 20 approximated spectra was 9.3.

Table 6.10: RMS Errors of Spectra Approximated Using Cubic Splines with Knots Selected with a Backward Stepwise Procedure

Pixel	Number of Knots				
	40	35	30	25	20
1	8.8	14.1	17.2	26.3	35.6
2	6.4	10.7	16.6	23.6	30.4
3	7.2	11.3	15.5	23.7	38.2
4	9.7	12.0	18.3	21.3	30.5
5	7.7	10.9	14.2	17.8	29.0
6	7.5	10.4	12.6	18.5	24.5
7	10.4	12.0	13.5	17.1	22.2
8	19.7	22.7	23.9	32.1	39.9
9	7.2	8.4	10.1	12.4	19.7
10	5.4	7.2	7.5	9.0	16.2
11	8.1	10.3	13.1	17.6	28.8
12	11.2	16.0	25.6	38.1	54.6
13	9.0	12.6	17.0	39.8	74.1
14	12.1	14.8	24.9	37.8	53.6
15	7.7	12.1	22.3	37.7	76.7
16	6.5	8.9	12.9	20.7	34.1
17	12.9	18.1	18.5	25.3	29.0
18	4.6	5.3	6.8	9.2	13.7
19	9.4	11.8	15.3	28.2	32.3
20	13.6	14.6	16.8	18.8	23.9
Average	9.3	12.2	16.1	23.7	35.3

This increases to 16.1 when the number of knots selected is reduced to 30 and to 35.3 when the number of selected knots is further reduced to 20.

The results for pixels 6, 10 and 14 are shown graphically in figures 6.17, 6.18 and 6.19 respectively. For each pixel, the graphs show the resulting approximation using 40, 30 and 20 knots, and for comparison, these are shown along with the original spectral curve. The approximations are in good agreement with the original spectra and fit the curves well. The main factor affecting the goodness of fit is clearly the number of knots used for the fit and their location along the curve. The other significant factor is the number of spectral features present in the curve being modelled. As an example, the spectral curve for pixel 10 (figure 6.18) shows a relatively flat spectral curve with few spectral features. The resulting approximations show a high level of accuracy, with RMS errors ranging from 5.4 – 16.2 for 40 – 20 knots used respectively. This contrasts with the spectral curve for pixel 14 (figure 6.19) which exhibits several absorption/reflectance features and consequently the accuracy of the approximations is reduced. When using 40 knots selected with the backward deletion scheme, the RMS error in the approximation for pixel 14 is 12.1. This increases to 24.9 with 30 knots and 53.6 when only 20 knots

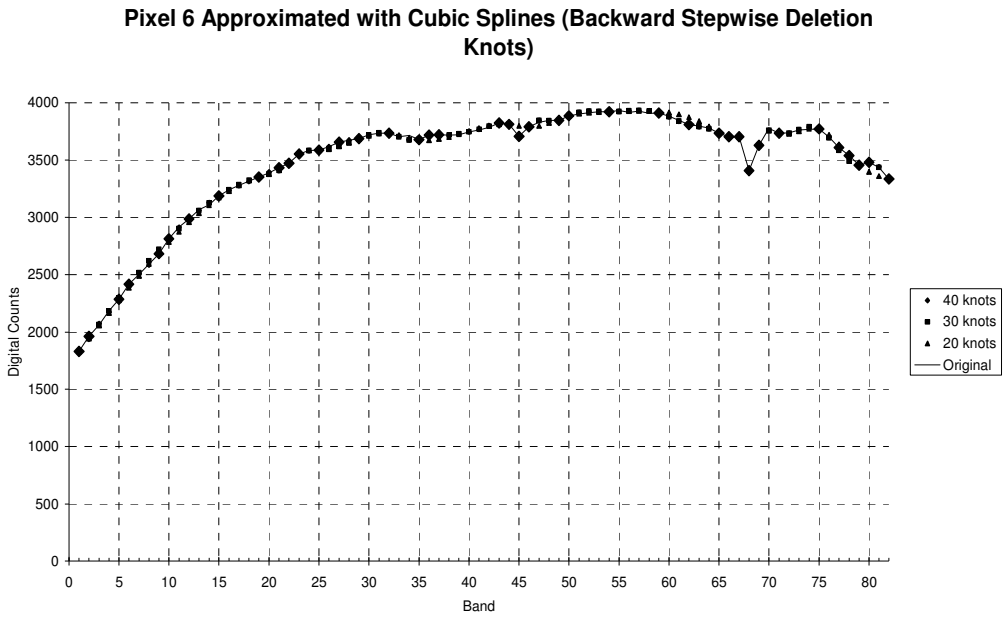


Figure 6.17: Pixel 6 spectra: original and approximated using cubic splines with 40, 30 and 20 knots selected using a backward stepwise selection technique. The positions of the 40 knots selected are shown as larger diamonds.

are used. The decrease in accuracy is visibly noticeable when only 20 knots are used, particularly near bands 36, 67, 72 and 80, where the spectral features have not been fully emphasized.

It has been shown in this chapter that to fit spectral curves with cubic splines, the location of the knots is critical to the accuracy of the approximations. Regularly spaced knots gave adequate approximations to the spectral curves although there is no guarantee that important spectral features will be properly modelled. Given a predetermined number of knots to use, the optimal solution would be found by testing each combination for the highest accuracy. When dealing with hyperspectral data however, this solution is completely impractical as the number of combinations to assess is enormous. In the examples presented here, the data set was composed of 82 spectral bands. In order to choose 20 knots, the number of combinations to assess in order to find the optimal solution would be $> 6.4 \times 10^{15}$. This is if the first two knots are automatically chosen as bands 1 and 82 for the end points. If 40 knots are required, the number of combinations increases to $> 9.7 \times 10^{22}$. A more strategic method is needed that does not require so many combinations to be tested.

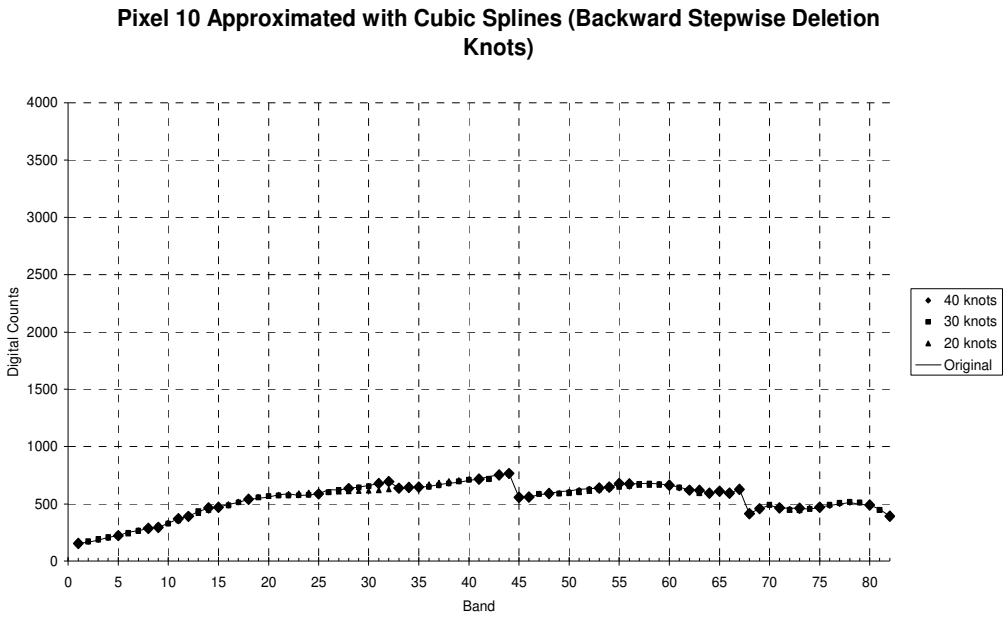


Figure 6.18: Pixel 10 spectra: original and approximated using cubic splines with 40, 30 and 20 knots selected using a backward stepwise selection technique. The positions of the 40 knots selected are shown as larger diamonds.

Two different stepwise knot selection techniques were investigated which sequentially select or remove a knot. The first was a forward stepwise procedure which commences with the first two knots as the end points and selects each subsequent knot as the one giving the smallest RMS error. This process continues until the total number of required knots has been selected. The other stepwise technique was a backward stepwise procedure. Initially all knots are used to fit the curve and a knot is removed at each step. The knot chosen to be removed is the one which once deleted results in the approximated curve having the lowest possible RMS error.

When trialled on 20 sample pixel spectra the forward stepwise procedure chose knots which gave unexpectedly poor results. In 48% of the cases tested the cubic splines using knots from the forward stepwise scheme gave less accurate approximations than those from when the same number of regularly spaced knots was used. Upon closer examination it was found that knots were sometimes being selected which were actually redundant. Typically, a spectral feature within a curve would need several knots to be well approximated. By selecting the first of these knots could actually increase the RMS error in the fit significantly, so the procedure would opt for a knot in a region of the curve that was already well characterized. Choosing a

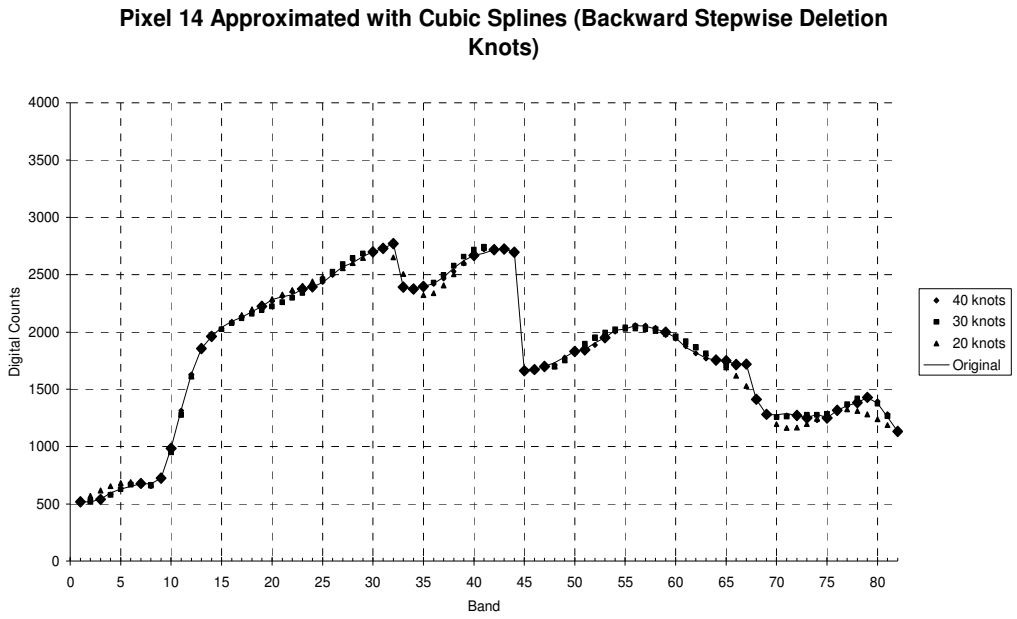


Figure 6.19: Pixel 14 spectra: original and approximated using cubic splines with 40, 30 and 20 knots selected using a backward stepwise selection technique. The positions of the 40 knots selected are shown as larger diamonds.

knot in a region of the curve which is already accurately described is superfluous, and contradicts the objectives of this research.

The backward stepwise technique was shown to be successful at selecting knots to fit cubic splines to the data. Instead of choosing the superfluous knots, it would actually remove them one by one. This resulted in cubic spline fits to the data that needed fewer knots to produce spectra with higher levels of accuracy. It should not be assumed though that the set of knots chosen by this method is necessarily the best possible combination of knots. The backward stepwise knot selection scheme is however a much more successful method than the forward stepwise knot selection scheme or by simply choosing knots at regularly spaced intervals. The forward stepwise knot selection scheme could perhaps be improved by selecting the next knot at the point at which the difference between the actual and approximated spectra is a maximum. Choosing knots in this manner is unlikely though to outperform the backward stepwise selection technique.

To compare the performance of wavelets and cubic splines at approximating the spectra the average RMS errors were plotted against the number of parameters. For each of the different wavelets used, the RMS errors and the number of parameters were averaged over the 20 pixels at different thresholds and the RMS error was plotted against the equivalent number of parameters used. For comparison, the same was done in the case of cubic splines where the knots had been selected using the backward stepwise procedure. The average RMS error and number of parameters used were averaged over the 20 pixels for the different numbers of knots used. The number of knots trialled was restricted to a maximum of 40. The results are shown in figure 6.20 and show the results from using the Haar, ψ_{D4}, ψ_{D6} and ψ_{D10} wavelets with thresholds of 50, 75, 100, 125 and 150. Also shown are the results from approximating the spectra with cubic splines with knots selected with the backward stepwise procedure. To make the comparison clearer, lines of best fit have been fitted for equations of the form $y = ax^{-b}$, where b is a positive integer. In all five cases, the lines of best fit had Pearson product-moment correlation coefficients > 0.99. Ideally, these curves would occupy the lower left corner of the graph, where both the RMS errors and the number of parameters needed have low values. Figure 6.20 shows the ψ_{D4} wavelet is the best method of those trialled for approximating

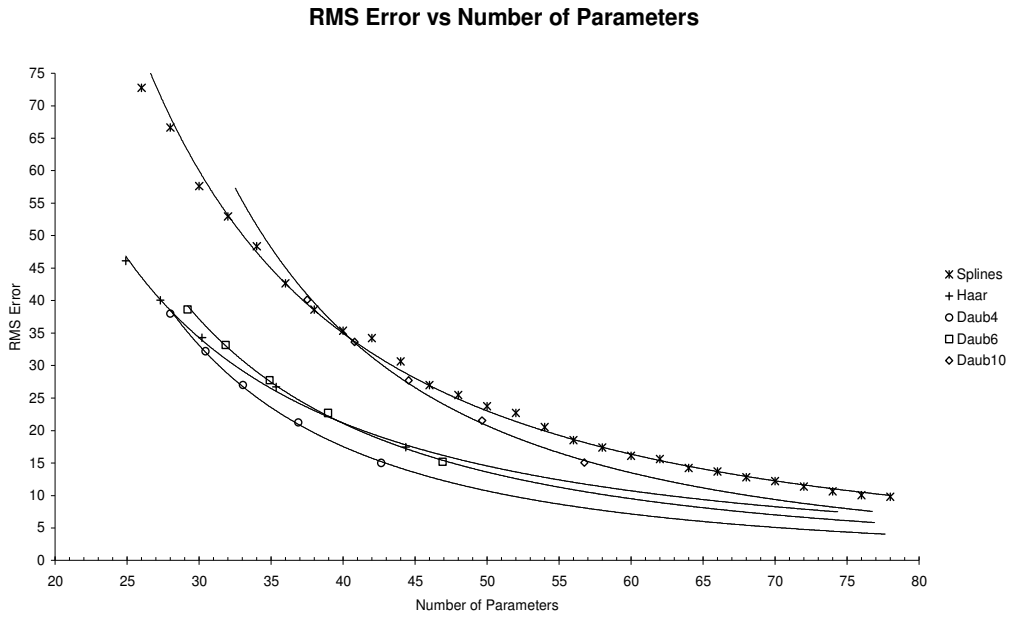


Figure 6.20: RMS errors plotted as a function of the number of parameters used to approximate the spectra shown in figure 6.1. Lines of best fit are for equations of the form $y = ax^{-b}$, where b is a positive integer.

spectra, both in terms of accuracy and efficiency. This is closely followed by the Haar and ψ_{D6} wavelets. Figure 6.20 also shows that the least successful methods used are the ψ_{D10} wavelet and cubic splines with knots selected by the backward stepwise technique.

The approaches presented in this chapter have been tested with the emphasis of minimizing the RMS error in the approximations to the spectra. There is no indication that these methods can be used to accurately classify remotely sensed hyperspectral data. This issue will be investigated in Chapter 7. In this instance, the focus will shift from an emphasis on the accuracy of the approximated spectra to the maximization of the class separability and classification accuracy.

7 Maximum Likelihood Classification Using Dimensionally-Reduced Spectra

In Chapter 6 various techniques such as wavelets and cubic splines were used to investigate how well the measured spectra could be approximated using significantly fewer coefficients. Some of those techniques are adapted slightly and the focus moves away from minimizing the RMS error in the approximation towards an emphasis on the accuracy of the classification of the entire scene. Instead of reconstructing the measured spectra, the output coefficients from the analysis are used as direct inputs to the classification algorithm.

After each successive level of a wavelet analysis the numbers of approximation and detail coefficients are both halved. Given that both these sets of coefficients are characteristic of the original spectra at a particular scale, they can be used to calculate the corresponding sample mean and covariance statistics for each class and then used as direct inputs to a maximum likelihood classifier. Dimensional reduction will be achieved if the coefficients from any one particular level of analysis are used. This will avoid the need to further use either a statistical feature selection method or threshold out the detail coefficients as was the focus of the previous chapter.

Koger et al. (2003) note that when analysing hyperspectral signatures with the Haar mother wavelet, the resulting detail coefficients are equivalent to first-order approximations of the first derivatives of the original signatures. Derivative analysis is an established technique in analytical chemistry and has been used for many years in spectroscopy (Demetriades-Shah, Steven & Clark 1990; Bruce & Li 2001; Koger et al. 2003). Derivative spectra show great promise for dealing with hyperspectral remote sensing problems although this area has not been researched widely by the remote sensing community (Demetriades-Shah, Steven & Clark 1990; Tsai & Philpot 1998).

An object may be more easily identifiable based on the shape of its spectral signature rather than on the magnitude of the spectral values. It was noted by Furby, Kiiveri and Campbell (1990) that the first and second derivatives of a spectral curve give

information on the slope and curvature of the curve respectively. If the first and second derivative spectra are used for classification purposes, this offers another possibility for dimensional reduction of the spectra. Regions of the spectral curves that are fairly linear will be reduced to zero in the second derivative spectra. Spectral bands in which this occurs frequently could be removed from the data and not used for the classification.

Both the approximation and detail wavelet coefficients are investigated for their ability to classify the scene shown in figure 5.1. The research then examines using dimensionally-reduced first and second derivative spectra as a means for classifying the same scene. The bands are selected by finding the first and second derivatives of the class means and then selecting those bands exhibiting the highest standard deviations amongst the mean derivative spectra. This typically occurs where the spectral bands are located near reflection/absorption features or where there are steep slopes/edges in the spectral signatures.

7.1 MLC Using Wavelets

Wavelets have found applications in hyperspectral remote sensing, in particular, discriminating between classes. It has not been common however for wavelets to have been used for entire scene based classifications. Mallet et al. (1997) used optimized adaptive wavelets with Bayesian linear discriminant analysis to discriminate between three different classes using simulated spectral data. The authors also compared their adaptive wavelets with the Daubechies and Coiflet families of wavelets when discriminating between spectral data of five different mineralogical groups (amphilolites, calsilicates, granite, mica and soil). In each case the Coiflet wavelets gave the lowest classification accuracies with the adaptive wavelets tending to achieve slightly higher classification accuracies than those from the Daubechies wavelets.

Koger et al. (2003) used the detail coefficients from 36 different mother wavelets to detect Pitted Morning Glory (*Ipomoea lacunosa*) amongst soybean (*Glycine max*). A portable field spectroradiometer was used to collect spectral measurements of soybean and soybean intermixed with Morning Glory at different growth stages. The

measurements were taken at a height of 1.22 m to achieve a 0.25 m spatial resolution and to ensure that the measurements contained a background soil reflectance component. Measurements were also taken from close range (6 cm) of each of the three components (soybean, Morning Glory and bare soil) so that synthetically mixed reflectance spectra could be generated. Of the 36 mother wavelets used, 10 of them achieved classification accuracies of 90% or better and included mother wavelets from the Daubechies, Symlet, and Biorthogonal families of wavelets plus the Haar wavelet. Fisher's linear discriminant analysis was used to discriminate between the two classes. The ability of the Haar wavelet detail coefficients to discriminate between the two classes was compared with a PCA. In every case, the Haar wavelet was able to discriminate between the two classes at different growth stages with a higher accuracy than that of the PCA.

7.1.1 Approximation Coefficients

The same mother wavelets used in section 6.2 are now investigated for their ability to classify an entire scene, specifically, the one shown in figure 5.1. First, the approximation coefficients from different levels of analysis are used and then in the next section the detail coefficients will be tested. Dimensional reduction of the data is achieved due to the fact that after each successive level of analysis the number of coefficients is halved. To ensure the size of the resulting image files are smaller than the originals, the coefficients are rounded off to the nearest whole number and saved as 16 bit signed integers and not as 32 or 64 bit reals. This would have negligible effect on the data as the approximation coefficients tend to range from $10^2 - 10^4$ in magnitude.

Table 7.1: Classification Accuracies Resulting from Using Wavelet Approximation Coefficients After Different Levels of Analysis

Mother Wavelet	Classification Accuracies (%)					
	$J = 1$	$J = 2$	$J = 3$	$J = 4$	$J = 5$	Mean
Haar	96.33	96.30	96.10	95.61	93.20	95.51
Daub4	96.37	96.29	96.18	95.42	95.32	95.92
Daub6	96.41	96.31	96.20	95.86	95.51	96.06
Daub10	96.38	96.29	96.16	95.98	95.63	96.09
Mean	96.37	96.30	96.16	95.72	94.92	

Table 7.1 shows the resulting classification accuracies when the approximation coefficients are used as inputs to a standard MLC algorithm. The wavelets used were the Haar, Daub4, Daub6 and Daub10 wavelets. Approximation coefficients were generated after $J = 1 - 5$ levels of analysis and the table shows the classification accuracies from each level of analysis. As the level of analysis increases, the classification accuracies decrease. This is to be expected though, because each time the level of analysis increases the number of coefficients is reduced by a factor of a half. Upon examining the average results the Daub10 wavelet gives the higher classification accuracies and the Haar wavelet gives the lower accuracies when using the approximation coefficients.

When the number of coefficients used for the classification is taken into account however, the situation changes. Figure 7.1 shows the classification accuracies plotted as a function of the number of approximation coefficients used. When the number of coefficients is small (< 25), the Haar wavelet gives the best classification results and between 11 and 25 coefficients the Daub10 wavelet yields the lowest classification accuracies. In terms of classification accuracy the differences between the four wavelets used is negligible and the four curves in figure 7.1 are relatively

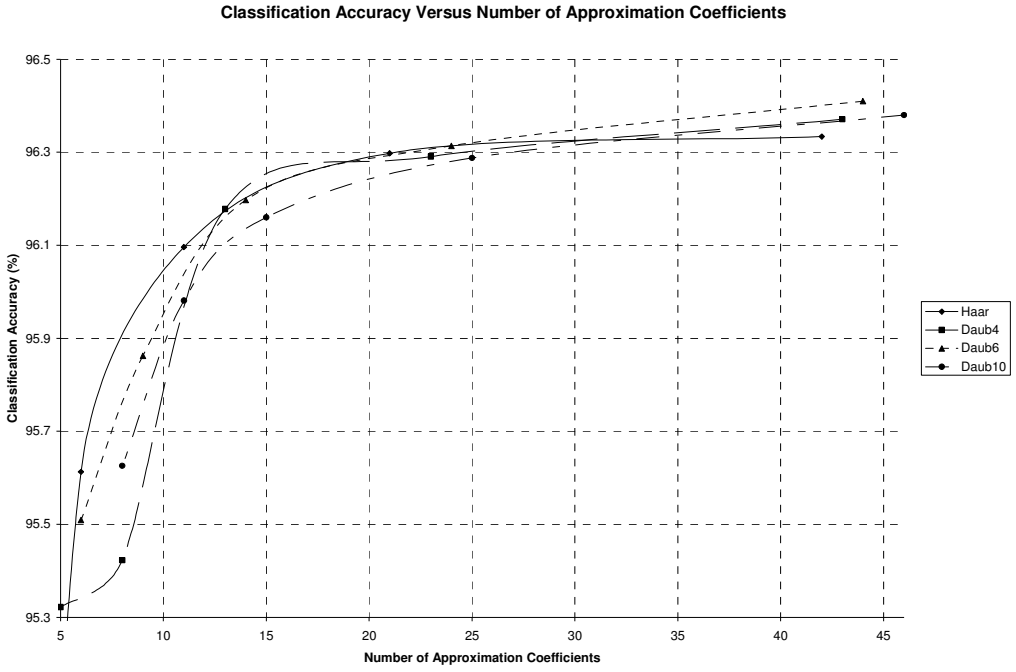


Figure 7.1: Classification accuracies as a function of the number of approximation coefficients used for different wavelets.

close together.

The number of approximation coefficients used in the classifications for a particular level of analysis is shown in table 7.2. The numbers do not reduce by exactly 0.5 after each level of analysis due to the width of the wavelet filter used, edge effects at the first and last few spectral bands and N initially not being equal to 2^n. Also, when attempting to classify the scene with the Daub6 and Daub10 wavelets for $J = 3 - 5$ and $J = 2 - 5$ respectively, the determinants of some of the class covariance matrices were equal to zero. This was due to the fact that for each pixel the last few approximation coefficients had values close to zero. To overcome this problem the low valued coefficients had to be removed and as a result the total number of approximation coefficients input to the classifier was reduced. When J increases from 4 to 5, the number of approximation coefficients for the Haar wavelet reduces to 3 and the corresponding classification accuracy reduces sharply from 95.6% to 93.2%.

7.1.2 Detail Coefficients

As with the approximation coefficients, when the wavelet detail coefficients were used for classifying the data at specific levels of analysis, the Daub10 wavelet performed the best of the four wavelets used, with the Haar wavelet resulting in lower classification accuracies. The results are summarised in table 7.3 and show that the best results overall were achieved at $J = 2$. When compared with the results from the approximation coefficients, the detail coefficients give more accurate results at higher levels of analysis ($J = 4, 5$). This suggests that for a small number of coefficients, wavelet detail coefficients will result in a more accurate classification than will the approximation coefficients. When the number of coefficients used to characterize the pixel spectra is higher, i.e. at $J = 1, 2, 3$, the approximation

Table 7.2: Number of Approximation Coefficients Generated by Wavelets at Different Levels of Analysis

Mother Wavelet	$J = 1$	$J = 2$	$J = 3$	$J = 4$	$J = 5$
Haar	42	21	11	6	3
Daub4	43	23	13	8	5
Daub6	44	24	14	9	6
Daub10	46	25	15	11	8

Table 7.3: Classification Accuracies Resulting from Using Wavelet Detail Coefficients After Different Levels of Analysis

Mother Wavelet	Classification Accuracies (%)					
	$J = 1$	$J = 2$	$J = 3$	$J = 4$	$J = 5$	Mean
Haar	96.27	96.19	95.90	95.56	94.19	95.62
Daub4	96.11	96.08	96.10	95.73	95.40	95.88
Daub6	96.06	96.12	96.08	95.98	95.23	95.89
Daub10	95.77	96.09	95.95	96.04	95.77	95.92
Mean	96.05	96.12	96.01	95.82	95.15	

coefficients proved to be better for classifying the data.

The number of detail coefficients used at each particular level of analysis for each wavelet used is shown in table 7.4. When the classification accuracies are plotted as a function of the number of detail coefficients used (figure 7.2), it is apparent that the Daub10 wavelet gives the lowest classification accuracies with an exception at 11 coefficients (corresponding to $J = 4$). The Haar wavelet produced the highest classification accuracies except for when 11 detail coefficients were used (corresponding to $J = 3$).

In both figures 7.1 and 7.2, it is ideal to have the data points located in the top left portions of the plots. This is where the classification accuracy is the highest and the number of coefficients used is the lowest. It is not ideal for points to be located in the bottom right section of the plots. There is not a significant difference between the approximation and detail coefficients in the top left sections of the plots however for higher numbers of coefficients used the detail coefficient points are situated closer to the bottom right corner than the approximation coefficient points. Based on these results, it is recommended that the approximation coefficients be used for a maximum likelihood classification of remotely sensed hyperspectral image data given the choice of using wavelet detail or approximation coefficients.

Table 7.4: Number of Detail Coefficients Generated by Wavelets at Different Levels of Analysis

Mother Wavelet	$J = 1$	$J = 2$	$J = 3$	$J = 4$	$J = 5$
Haar	42	21	11	6	3
Daub4	43	23	13	8	5
Daub6	44	24	15	9	7
Daub10	46	27	17	11	9

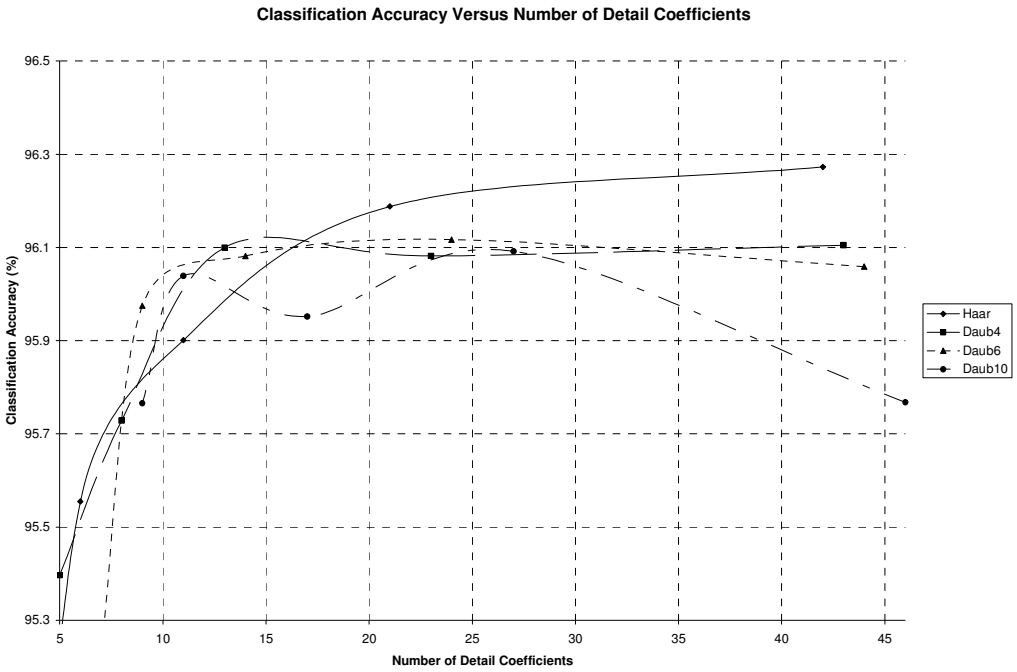


Figure 7.2: Classification accuracies as a function of the number of detail coefficients used for different wavelets.

7.2 MLC Using First and Second Derivative Spectra

It was noted in Koger et al. (2003) that when the Haar wavelet is used to analyse a hyperspectral signal the resulting detail coefficients approximate the first derivative. As with extracting wavelet detail coefficients from a signal, the process of differentiation is effectively a high-pass filtering operation (O'Haver 1982). The high-pass filtering nature of differentiation has seen it find applications in reducing low frequency background noise in signals (Demetriades-Shah, Steven & Clark 1990).

One advantage of finding derivatives is that for second or higher orders there should be little variation with changes in illumination intensity (Tsai & Philpot 1998). The solar angle, topography and cloud cover will all have an influence on illumination intensity and this will affect the magnitude of the measured spectra for an object/pixel. The location of spectral features and the spectral shape, however, will only be slightly affected by changes in illumination. Information on the shape of spectral curves can be derived by taking derivatives of the spectra. For example, the first derivative gives information on the slope of the spectra and the curvature of the

measured spectra is given via the second derivative. If information on the shape of spectra is characteristic of the object then the derivative spectra could be useful information to submit to a classification algorithm.

The idea of derivative spectra was first proposed in the early 1920s by Rutherford (Fell & Smith 1982) and has since become a well established spectroscopic technique. Initially, researchers were not enthusiastic in adapting derivative spectra for remote sensing applications, but the concept has recently started to gain in popularity. Demetriades-Shah, Steven and Clark (1990) used spectral derivatives to detect chlorosis in vegetation and showed how derivative spectra can be used to remove background soil signatures from the spectra. Philpot (1991) used derivatives to develop the derivative ratio algorithm to avoid atmospheric effects in hyperspectral data.

Chen, Curran and Hansom (1992) investigated the use of spectral derivatives for estimating the suspended sediment concentrations (SSC) in water. It was found that the first derivative produced higher maximum correlation coefficients between SSC and wavelength than did the measured spectral reflectance. Their results were based on both laboratory and *in situ* measured reflectance spectra. Spectral derivatives have been used by Gong, Pu and Yu (1997) for vegetation species discrimination. Both neural networks and linear discriminant analysis were used with *in situ* first derivative spectra as the input and training data to discriminate between six different conifer species. The effect of smoothing and band separation on derivative spectra has been investigated by Tsai and Philpot (1998) and as an application they applied derivative spectroscopy to band decomposition and locating absorption features within remotely sensed spectra. Blackburn (1998) used spectral derivative approaches to quantify vegetation pigment concentrations at both leaf and canopy scales. First and second derivatives of pseudo absorbance (log[1/reflectance]) were found to be strongly related to pigment concentration per unit area but were better suited to estimating the concentration per unit mass. The pigments for which concentrations were being quantified in the study were chlorophyll *a*, chlorophyll *b* and the carotenoids.

Most recently Tsai and Philpot (2002) have used derivative spectra for classifying vegetation in an AVIRIS image. A principal components analysis was initially performed on the 97-band image data and the first 10 principal components were used to construct a new base image. An MLC was performed on the new base image and to improve the classification accuracies, derivative spectral features were iteratively appended to the 10-band base image. After each derivative feature was added the new image would be classified with the ML classifier. The added derivative features were selected such that larger Jeffries-Matusita (JM) distances resulted between classes. The JM distance between normally distributed classes k and l is given by

$$J_{kl} = 2(1 - e^{-\alpha}) \tag{7.1}$$

where

$$\alpha = \frac{1}{8}(\overline{\mathbf{x}}_k - \overline{\mathbf{x}}_l)^T \left\{ \frac{\mathbf{C}_k + \mathbf{C}_l}{2} \right\}^{-1} (\overline{\mathbf{x}}_k - \overline{\mathbf{x}}_l) + \frac{1}{2} \ln \left\{ \frac{\left| \frac{\mathbf{C}_k + \mathbf{C}_l}{2} \right|}{\left| \mathbf{C}_k \right|^{1/2} \left| \mathbf{C}_l \right|^{1/2}} \right\} \tag{7.2}$$

Differential calculus defines the derivative of a function as

$$f'(x) = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} \tag{7.3}$$

(Thomas & Finney 1988). The previous chapter demonstrated that hyperspectral data curves could be accurately represented by piecewise polynomial cubic spline functions. Cubic spline functions have both continuous first and second derivatives and these are straightforward to compute given the parameters for the cubic spline functions. If each cubic spline piece is given by

$$S_j(x) = a_j + b_j x + c_j (x)^2 + d_j (x)^3$$

then the first derivative for the piece is given by

$$S_j'(x) = b_j + 2c_j x + 3d_j (x)^2$$

and the second derivative will be given by

$$S_j''(x) = 2c_j + 6d_j x$$

The task of fitting cubic splines to an entire hyperspectral scene for the purpose of differentiating the spectra would be computationally enormous. However, the continuous nature of hyperspectral data lends itself to numerical approximations for differentiation and it is possible to avoid having to initially fit cubic splines to the data. Given pixel spectra **x**, the first derivative for each of the spectra can be approximated using

$$\frac{d\mathbf{x}}{di} \approx \frac{\mathbf{x}_{i+\Delta i} - \mathbf{x}_i}{\Delta i}, \quad \frac{\mathbf{x}_{i+\Delta i} - \mathbf{x}_{i-\Delta i}}{2\Delta i} \quad \text{or} \quad \frac{\mathbf{x}_i - \mathbf{x}_{i-\Delta i}}{\Delta i}, \tag{7.4}$$

where i is the band number and Δi is the separation between bands. The differentiation in (7.4) is with respect to band number and not wavelength in order to simplify any associated algorithms and further reduce the computational load. If the derivative is required with respect to wavelength it is simply a matter of exchanging i with λ and Δi with $\Delta\lambda$ in (7.4). An approximation for the second derivative is similarly obtained from that for the first derivative as

$$\frac{d^2\mathbf{x}}{di^2} \approx \frac{\mathbf{x}_{i+\Delta i} - 2\mathbf{x}_i + \mathbf{x}_{i-\Delta i}}{(\Delta i)^2}. \tag{7.5}$$

In regions where the spectra are fairly flat, the first and second derivatives at those points will have values close to zero. Alternatively, steep edges or strong absorption/reflection features in the original spectra will have correspondingly large values in the first and second derivative spectra. This presents a potential opportunity for effective dimensionality reduction. Significant features within the original spectra can be identified with the first and second derivative spectra and the corresponding bands can be retained and used for subsequent classification. Conversely, the insignificant bands can be removed from the data and excluded from any further classification.

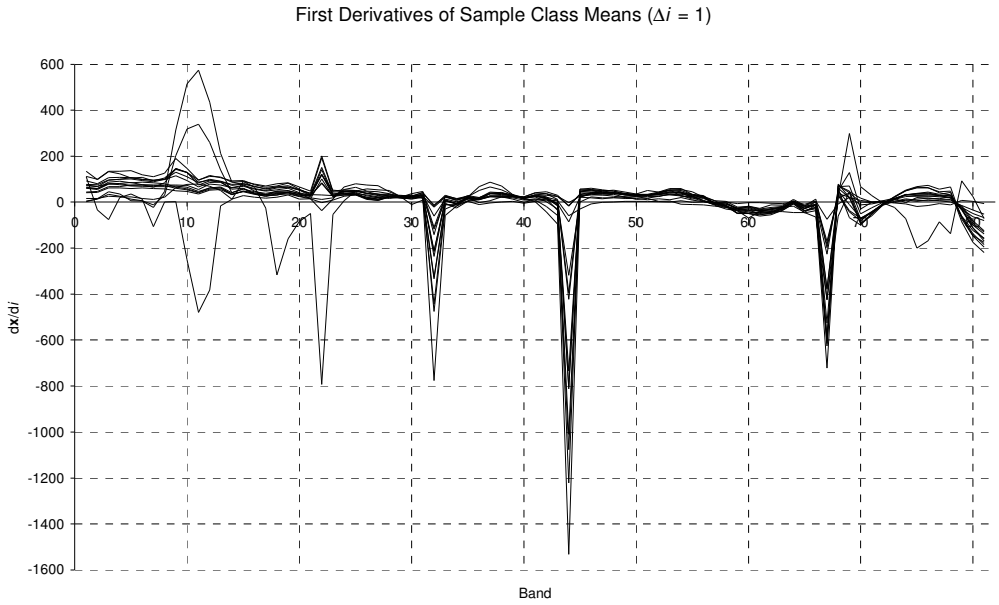


Figure 7.3: First derivatives of the sample class mean spectra with a band separation of $\Delta i = 1$. The values are very similar between classes except for where significant spectral features occur.

Equation (7.4) was used to calculate the first derivative of the sample mean spectra for each of the classes identified in figure 5.3. Figure 7.3 shows the first derivatives of each of the class means for a band separation of $\Delta i = 1$ and shows that most of the classes have similar values of the first derivative at many of the bands. At several bands however, the value of the first derivatives are somewhat varied. These bands correspond to the location of significant features in the original spectra. The standard deviation in the first derivatives was calculated for all 15 classes at each band and these were sorted in order from highest to lowest. The first N bands exhibiting the highest standard deviations were selected and the first derivatives were calculated at these bands for every pixel in the scene. The resulting image was then classified with a maximum likelihood classifier. This was done for band separation values of $\Delta i = 1$, 2 and 3 and reduced dimensionalities of $N = 5$, 10, 15, 20, 25 and 30.

The classification results when using the first derivative spectra are summarised in table 7.5. It can be seen that for different band separations there is little effect on the classification accuracies. For a given dimensionality N, the variation in classification accuracy as the band separation changes is small. When $N = 10$ and the difference in classification accuracy changes by ~ 3% from 88.7% to 91.8% when the band

Table 7.5: Classification Accuracies when Using First Derivative Spectra with Different Band Separations, Δi Using only N Selected Bands

Number of Bands,	Classification Accuracies (%)		
N	$\Delta i = 1$	$\Delta i = 2$	$\Delta i = 3$
5	77.79	78.62	78.15
10	91.84	90.11	88.71
15	94.89	94.31	93.68
20	95.43	95.48	95.41
25	95.71	95.68	95.81
30	95.90	95.90	95.97

separation changes from $\Delta i = 3$ to $\Delta i = 1$, but generally the change in classification accuracy with change in Δi is much less. Also apparent in table 7.5 is that both wavelet approximation and detail coefficients perform better than the first derivative spectra when used for classification purposes. Even when as few as 11 coefficients were used, the wavelet approximation and detail coefficients were achieving consistently better than 95% classification accuracies. For a similar number of bands, the first derivative spectra could only achieve classification accuracies of ~ 90%. These are still promising results, although when compared to the results from using wavelet coefficients they appear inferior.

The use of second derivative spectra is now investigated. Using (7.5) the second derivatives of the sample class mean spectra were calculated for different band separations. The result for $\Delta i = 1$ is shown in figure 7.4 and shows that as with the first derivatives, the values are very close together with the exception at those bands where there are significant spectral features. The standard deviations of the second derivatives were calculated between the different classes for each band and were ranked from highest to lowest. The bands corresponding to the N highest standard deviation values were selected and the second derivatives were calculated at these bands for all pixels in the scene. The resulting N-band image was classified using a maximum likelihood classifier and this was done for band separations of $\Delta i = 1$, 2 and 3.

The resulting classification accuracies when using the second derivative spectra as input are summarised in table 7.6. With the exception when $N = 5$, as the band separation is increased, the classification accuracy also increases. This effect is more pronounced for lower dimensionalities ($N = 10$, 15) than at higher ones ($N = 25$, 30). Comparing the classification results with those from using first derivative spectra,

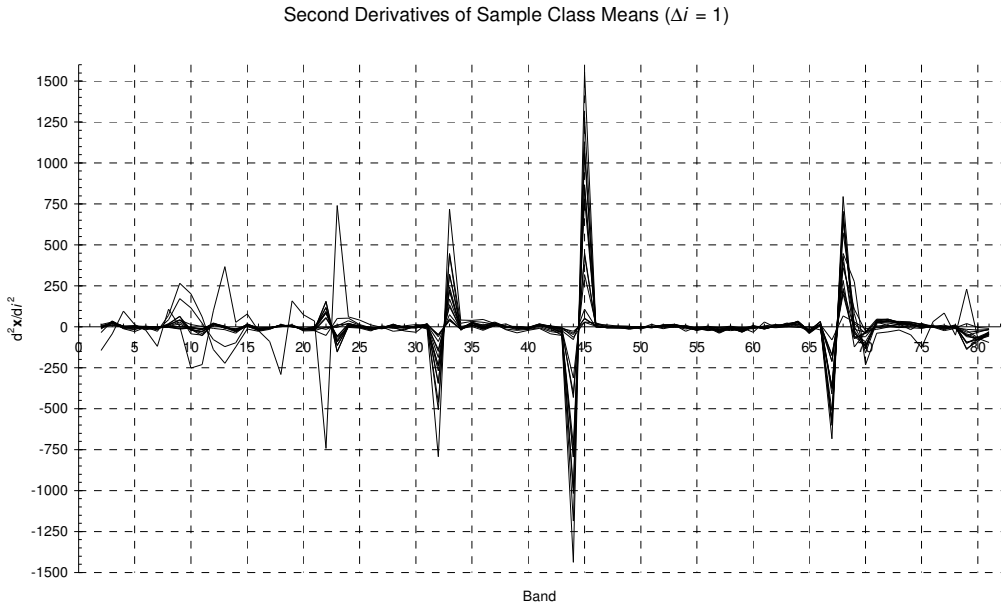


Figure 7.4: Second derivatives of the sample class mean spectra with a band separation of $\Delta i = 1$. As with the first derivatives, the values are very similar between classes except for where significant spectral features occur.

the second derivative spectra do not give as accurate classifications. One possible explanation for this is the high-pass filtering nature of the differentiation process.

When a signal is differentiated, the low frequency signal components are suppressed and the high frequency components are amplified. A side effect of this is that any high frequency noise present within the signal will also be amplified as higher order derivatives are found for the signal. Despite the fact that noise increases with derivative order, higher order derivatives still have their advantages in spectroscopy. An example is that it is easier to identify a peak or a trough in a signal by the corresponding minima/maxima in the second derivative than it is by the shape of the first derivative (Chadburn 1982).

Table 7.6: Classification Accuracies when Using Second Derivative Spectra with Different Band Separations, Δi Using only N Selected Bands

Number of Bands,	Classification Accuracies (%)		
N	$\Delta i = 1$	$\Delta i = 2$	$\Delta i = 3$
5	64.43	62.48	60.38
10	80.10	87.15	87.83
15	86.80	89.60	93.54
20	90.74	92.34	94.28
25	93.30	93.92	95.02
30	93.71	94.67	95.26

The results could be improved by first smoothing the pixel spectra prior to differentiation, an issue which has been examined by several researches (Savitzky & Golay 1964; O'Haver 1982; Tsai & Philpot 1998). One of the more popular smoothing techniques was developed by Savitzky and Golay (1964). Savitzky and Golay (1964) proposed a least squares convoluting method that makes use of quadratic, cubic and other higher order polynomial functions. Simple mean filter smoothers distort spectral peaks by attenuating the amplitude and increasing the width (O'Haver 1982). With polynomial convolution functions this effect is reduced. As the derivative order is increased, the level of noise also increases and hence the SNR is reduced. The amount of smoothing needed therefore, will also need to be increased. Instead of increasing the width of the smoothing filter to increase the amount of smoothing, O'Haver (1982) demonstrated that the data should be filtered several times to achieve this. O'Haver (1982) recommended that the data be filtered with $n + 1$ passes for the nth order derivative.

When the Savitzky and Golay (1964) method was compared with the mean filter smoother, Tsai and Philpot (1998) found there was little difference in performance when the spectra contained broad spectral features and relatively high frequency noise. Tsai and Philpot (1998) opted for a mean filter smoothing algorithm based on simplicity and the fact that computation time was less. Tsai and Philpot (1998) also found that the band separation Δi could be chosen to have a quasi-smoothing effect.

The choice of Δi is a compromise between noise reduction and loss of spectral detail. As Δi is increased, the level of noise will tend to be reduced but spectral details smaller than Δi will also be eliminated (Tsai & Philpot 1998). Also, as Δi is increased the values of the resulting spectral derivatives will be attenuated due to its presence in the denominator (particularly with higher order derivatives). Another disadvantage of using larger Δi values is that values within Δi of the ends of the spectra are unable to be calculated.

For comparison of these results, figure 7.5 shows the classification accuracies plotted against the number of bands used for the classification for some of the different dimensionality reduction techniques used. Also shown in the graph is the accuracy attained when all 82 bands were used to classify the data (96.46%). The compromise

between classification accuracy and the number of bands used is apparent and ideally the results would be located in the top left hand corner of the plot. This is where the classification accuracy is maximized and the number of bands used is minimized. Figure 7.5 reveals that the wavelet techniques used performed the best, followed by the dimensionally reduced first derivative spectra. The second derivative spectra are revealed to have performed the poorest.

The most significant feature of figure 7.5 is the level of accuracy attained by the wavelet methods when using only a fraction of the original bands. With as few as five bands, both the Haar wavelet approximation and detail coefficients achieve classification accuracies above 95.5%. This is very close to 96.46% which was attained using all 82 bands. For a 94% reduction in the number of bands used, the classification accuracy has slipped by less than 1%.

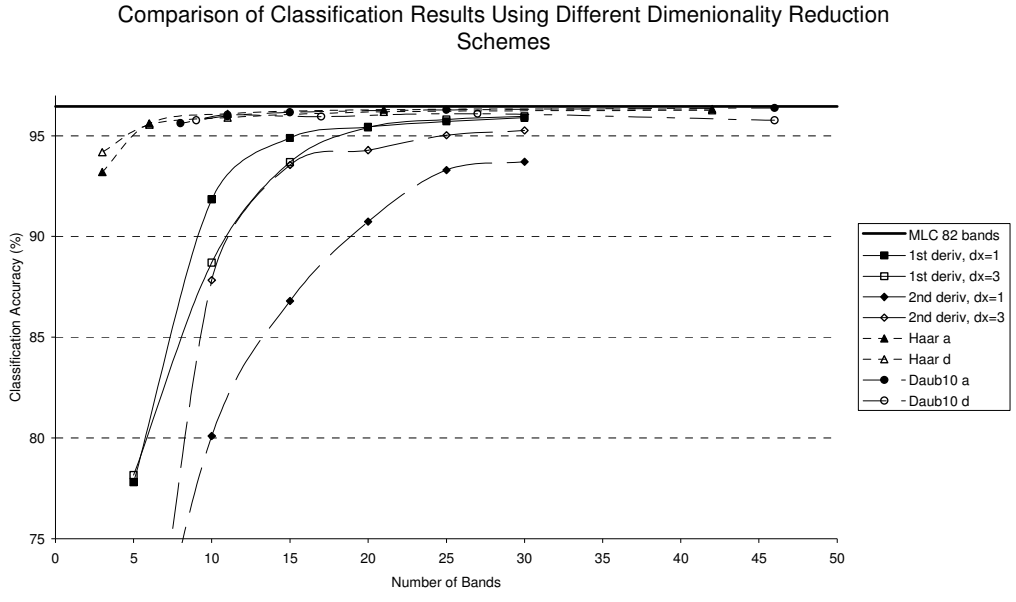


Figure 7.5: A comparison of the classification accuracies resulting from some of the different dimensionality reduction techniques examined by this research. The horizontal black line represents the classification accuracy when all 82 bands are used to classify the data (96.46%).

8 Conclusion

The evolution of hyperspectral sensors has brought both advantages and disadvantages to the field of remote sensing. The finer spectral resolution and increase in the number of spectral bands has given geologists, agriculturalists, horticulturalists, marine biologists and atmospheric scientists the capacity to measure characteristics of the land, oceans and atmosphere with much finer detail. The disadvantages of hyperspectral data relate to the immense increase in the amount of data. Hyperspectral data contains typically more than 100 spectral bands and this is a huge increase when compared with multispectral sensors such as Landsat or AVHRR which have typically less than 10 bands. As a result there is a large increase in the time required to transmit the data and in the amount of disk space needed to store the data. The analysis of hyperspectral data also increases the demand on computational power and processing time significantly.

This research has examined the problem of reducing the dimensionality and redundancy inherent in hyperspectral data. The emphasis focussed on applications in MLC of hyperspectral image data. As the covariance matrix is a major component of the MLC discriminant function the work also examined reducing the number of parameters required to adequately describe the class covariance information. The spectra associated with the pixels of hyperspectral images are also used by the MLC discriminant function and methods were investigated to model the spectra from significantly fewer parameters.

The research commenced by examining the nature of the spectral bands on HyMap instruments. Under examination was by what amount do the neighbouring spectral band response functions overlap with each other? High levels of overlap between spectral bands will have a significant influence on the correlation between neighbouring bands. In the absence of measured spectral filter function data, the filter functions can be modelled from triangular functions and especially Gaussian functions. When modelling the spectral filter functions with triangular functions the mean RMS error for each band was 0.0024. When Gaussian functions were used the mean RMS error improved to 0.0017. The main differences between the triangular

modelled functions and the Gaussian modelled were that the triangular functions tended to underestimate the tails of the band spectral response functions and overestimate the central peak. From these results it was found that the amount of overlap between neighbouring HyMap bands is considerable. The mean total overlap for each spectral band with its neighbouring bands was found to be approximately 60%. It would be expected that this high level of overlap between the bands would have a large affect on the level of correlation between the bands.

Two different approaches to reducing the parameterization of the inverse covariance matrices were investigated. Both methods make use of the sparse nature of the inverse covariance matrices and the partial correlation matrices to set off-diagonal elements in these matrices to zero. The first approach, developed in Roger (1996b) approximates the inverse class covariance matrices as band diagonal and experiments with different bandwidths. The second approach, based on theory developed in Speed and Kiiveri (1986) does not make the band diagonal assumption and uses a threshold to identify low values in the partial correlation matrices and set them to zero. The resulting approximated class covariance matrices were then used to classify a hyperspectral image acquired by the HyMap sensor over Toolibin, WA. The number of elements in an 82×82 matrix is 6724, but by setting a threshold of 0.04, an average of 4604 elements were set to zero per classification type and the resulting classification accuracy was > 96.4%. When the threshold was increased to 0.045 the number of elements with an absolute value less than this increased to an average of 4859 and the overall classification accuracy dropped sharply to 77%.

When assuming a band diagonal form for the inverse class covariance matrices, the results were not as successful as using thresholds with the partial correlation matrices. The resulting class covariance matrices showed little resemblance to the original matrices, although this need not necessarily result in classification accuracy. A review of the MLC discriminant function demonstrated that it is actually the inverse covariance matrices that are used by the algorithm. If these are well approximated, it should still be possible to achieve accurate classification results. Calculating the Cholesky factor for each partial correlation matrix ensured that the determinants of the modelled inverse covariance matrices were the same as the original matrices. While this approach produced classification accuracies greater

than 90%, it proved to yield less accurate results than when the values in the partial correlation matrix were thresholded. Bandwidths of 5, 6 and 7 were trialled and resulted in classification accuracies of 93.1%, 93.0% and 92.4% respectively. Hence, by using these methods it was demonstrated that a significant number of parameters in the inverse class covariance matrices can be set to zero with the full dimensionality of the original measured pixel spectra being retained. In this instance, using the method based on Speed and Kiiveri (1986) with a threshold ≤ 0.04 yielded excellent results.

Several methods including wavelets and splines were tested for the ability to accurately approximate the measured pixel spectra. A variety of commonly supported mother wavelets were used and various knot selection techniques were trialled to fit cubic splines to interpolate the spectral curves. The mother wavelets used were the Haar wavelet and also the Daub4, Daub6 and Daub10 from the Daubechies family of wavelets. Spectra were analysed and for a number of different thresholds, low valued detail coefficients were identified and set to zero. The reduced set of detail coefficients were then used to resynthesize the original spectra. When comparing the number of parameters used and the resulting RMS errors in the approximations, the Haar, Daub4 and Daub6 wavelets all achieved similar RMS errors for the equivalent number of parameters used, with the Daub4 wavelet achieving slightly higher accuracies. Of the four wavelets trialled, the Daub10 wavelet yielded larger RMS errors and required a larger number of parameters to estimate the spectra.

The different knot selection techniques used were a forward and backward stepwise knot selection procedure as well as knots positioned at regularly spaced intervals. The forward stepwise selection technique performed the most poorly and resulted in spectral approximations having the highest RMS errors. This is because the points that were being selected as knots were already closely fit to the spectral curve. If a point was selected in a region of the curve that was not well defined, the selection may increase the RMS error of the fit because the surrounding points were not closely fit. With 40 selected knots, the forward stepwise knot selection technique achieved a mean RMS error of 50.3 when fitting the spectral curves. The backward stepwise selection method performed the best of the three techniques trialled as it

would actively reject the most redundant knots first. With 40 selected knots the backward stepwise selection technique achieved a mean RMS error of 9.3. With 40 knot points selected at regular intervals the mean RMS error was 48.8. When compared with the different wavelets, the backward stepwise knot selection technique was similar in performance to the Daub10 wavelet but was not as accurate as the other three wavelets trialled. In terms of the number of parameters used, the backward stepwise knot selection technique was slightly more accurate than the Daub10 wavelet when less than 40 knots were used. For more than 40 knots, the backward stepwise knot selection technique was slightly less accurate than the Daub10 wavelet when using the equivalent number of parameters.

The methods used to approximate the measured pixel spectra from dimensionally-reduced spectra were adapted and used to process the entire hyperspectral image. The resulting dimensionally-reduced images were then classified. Both wavelet approximation and detail coefficients were generated at different levels of analysis and the coefficients from individual levels of analysis were used to estimate the class mean and covariance information and then submitted to the MLC algorithm. There was very little difference between using approximation coefficients or wavelet detail coefficients when comparing classification accuracies. After one level of analysis, the mean classification accuracies achieved from the approximation coefficients and wavelet detail coefficients were 96.4% and 96.1% respectively. The number of coefficients used at one level of analysis ranged from 42 coefficients for the Haar wavelet and 46 for the Daub10 wavelet. These numbers of coefficients were for both the approximation coefficients and wavelet detail coefficients. After four levels of analysis, the mean classification accuracies achieved from the approximation coefficients and wavelet detail coefficients decreased slightly to 95.7% and 95.8% respectively. The number of coefficients used by both the approximation and wavelet detail coefficients for four levels of analysis ranged from 6 coefficients for the Haar wavelet and 11 for the Daub10 wavelet. These results are very close to the original 96.5% classification accuracy that was achieved with a standard MLC using all 82 bands.

Interpolating spectral data with cubic splines results in smooth curves with continuous first and second derivatives. The spline coefficients make calculating

these first and second derivatives a straightforward task but to fit cubic splines to an entire image and then determine the first and second derivatives at particular bands would be a considerable computational task. To overcome this problem a numerical approximation was used to generate the first and second derivatives. A small subset of the bands was first selected by finding the first and second derivatives of the class means. Regions of the spectra where the curves were relatively flat or linear would have low valued first and second derivative values respectively. This proved to be the case for a large number of bands. Bands that are located at significant absorption or reflectance features exhibit significant first and second derivatives. The first and second derivatives were approximated at these bands and the resulting values were used to generate other images of much lower dimensionality. The resulting images were then classified using MLC.

The results indicated that the first derivative spectra were able to classify the data more effectively than the second derivative spectra. Using 30 bands with different band separations, the first derivative spectra gave classification accuracies of approximately 96%. The results from using first derivative spectra compared with the second derivative spectra which resulted in classification accuracies between 93.7% and 95.3%. When ten bands were used with different band separations, the range of classification accuracies using the first derivative spectra was between 88.7% and 91.8%, while the second derivative spectra classification accuracies range between 80.1% and 87.8%.

One problem that can arise when differentiating signals is that noise present within the signal can be amplified. The results from using derivative spectra for classifying the hyperspectral image could be improved if the spectra were smoothed prior to differentiation. By using larger band separations a slight smoothing effect can be realised and this was apparent in the results from the second derivative spectra. When 10 bands were used for the classification the accuracy increased from 80.1% to 87.8% when the band separation was increased from $\Delta i = 1$ to $\Delta i = 3$. If a proper smoothing approach is to be employed it needs to be noted that the amplitude of absorption/reflection features will be attenuated and the widths of the features will be broadened.

One of the main factors affecting the success of MLC is the number of training samples selected to represent each spectral class. For a data dimensionality of N, it is recommended that at least $10N$ training samples be selected to reliably estimate the class covariance matrices (Swain & Davis 1978; Richards & Jia 1999). Many classes in the image appear in large contiguous blocks and selecting ample training data for these classes is a straightforward matter. Some classes such as shadow, dark tree veg, bright tree veg and bright soils however typically appear in clusters comprised of only a few pixels and are not very prevalent within the image. In order to select enough training samples and avoid generating singular class covariance matrices, it was necessary to select pixels on edges of these small clusters. Inevitably, these pixels would not have been pure pixels, but would have been contaminated to some degree by the spectral signatures from neighbouring classes. The presence of these foreign endmember signals within the pixel spectra could have caused them to be misclassified and this would reduce the classification accuracies. In circumstances such as these, spectral unmixing/endmember analysis is desirable.

Recommendations for future work would include revising the knot selection techniques when approximating spectral curves with cubic splines. The results from fitting cubic splines using a forward stepwise selection procedure were inadequate and this was due to the knot selection approach used. Given an existing set of knots, each successive knot was selected such that it resulted in the lowest RMS error to the approximation to the spectral curve. This was later observed to be selecting knots in regions of the spectral curves that were already well approximated. It may be worthwhile trialling an alternate forward knot selection approach which, given an existing set of knots, selects the next knot as the point on the spectral curve which has been approximated the worst. Specifically, the spectral curve needs to be compared to the approximation and the point on the curve having the largest difference between the original and approximated curve would be selected as the next knot.

This research has principally focussed on the compromise between the accuracy of MLC and the number of coefficients used. The issue of computer processing time has not been examined. Another recommendation for future work would be to compare the computer processing times for the different methods tested and compare

with the results achieved. This was not practical as each method was implemented with different software on differing computer hardware.

In summary, it has been demonstrated that despite the wealth of data provided by hyperspectral sensors, it is possible to use only a small fraction of this data yet still obtain comparable classification accuracies to those when the full suite of data is used. The data typically exhibits very high interband correlations, rendering much of the data largely redundant. This can be exploited to bring about a significant reduction in the dimensionality of the data, considerably reducing the amount of storage space and processing needed.

9 References

Aboufadel, E. & Schlicker, S. 1999, *Discovering Wavelets*, John Wiley & Sons, Inc., New York.

Aspinall, R.J., Marcus, W.A. & Boardman, J.W. 2002, 'Considerations in Collecting, Processing, and Analysing High Spatial Resolution Hyperspectral Data for Environmental Investigations', *Journal of Geographical Systems*, vol. 4, pp. 15 – 29.

Blackburn, G.A. 1998, 'Quantifying Chlorophylls and Carotenoids at Leaf and Canopy Scales: An Evaluation of Some Hyperspectral Approaches', *Remote Sensing of Environment*, vol. 66, pp. 273 – 285.

Blatter, C. 1998, *Wavelets: A Primer*, A K Peters, Ltd., Natick, Massachusetts.

Bodechtel, J. 2001 'Requirements on Optical Sensors for Quantitative Definition of Surface Parameters Multispectral-Hyperspectral', *Advances in Space Research*, vol. 28, no. 1, pp. 241 – 250.

Bruce, L.M. & Li, J. 2001, 'Wavelets for Computationally Efficient Hyperspectral Derivative Analysis', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 7, pp. 1540 – 1546.

Bruce, L.M., Koger, C.H. & Li, J. 2002, 'Dimensionality Reduction of Hyperspectral Data Using Discrete Wavelet Transform Feature Extraction', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 10, pp. 2331 – 2338.

Campbell, N.A. 1984, 'Canonical Variate Analysis — A General Model Formulation', *Australian Journal of Statistics*, vol. 26, no. 1, pp. 86 – 96.

Chadburn, B.P. 1982, 'Derivative Spectroscopy in the Laboratory: Advantages and Trading Rules', *Analytical Proceedings*, vol. 19, pp. 42 – 43.

Chang, C.-I, Zhao, X.-L., Althouse, M.L.G. & Pan, J.J. 1998, 'Least Squares Subspace Projection Approach to Mixed Pixel Classification for Hyperspectral Images', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 3, pp. 898 – 912.

Chen, Z., Curran, P.J. & Hansom, J.D. 1992, 'Derivative Reflectance Spectroscopy to Estimate Suspended Sediment Concentration', *Remote Sensing of Environment*, vol. 40, no. 1, pp. 67 – 77.

Chiu, H.-Y. & Collins, W. 1978, 'A Spectroradiometer for Airborne Remote Sensing', *Photogrammetric Engineering and Remote Sensing*, vol. 44, no. 4, pp. 507 – 517.

Cocks, T., Jenssen, R., Stewart, A., Wilson, I. & Shields, T. 1998, 'The HyMap™ Airborne Hyperspectral Sensor: The System, Calibration and Performance', presented at *1st EARSel Workshop on Imaging Spectroscopy*, Zurich, October 1998.

Collins, W., Chang, S.-H., Raines, G., Canney, F. & Ashley, R. 1983, 'Airborne Biogeophysical Mapping of Hidden Mineral Deposits', *Economic Geology*, vol. 78, no. 4, pp. 737-749.

Daubechies, I. 1992, *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics, Philadelphia.

Demetriades-Shah, T.H., Steven, M.D. & Clark, J.A. 1990, 'High Resolution Derivative Spectra in Remote Sensing', *Remote Sensing of Environment*, vol. 33, pp. 55 – 64.

Dempster, A.P. 1972, 'Covariance Selection', *Biometrics*, vol. 28, pp. 157 – 175.

Dunne, R.A. 1999, 'Analysis of Hyperspectral Data: HyMap Data for the Toolibin Region', Internal Report, CSIRO Mathematical and Information Sciences, Floreat, Western Australia.

Fell, A.F. & Smith, G. 1982, 'Higher Derivative Methods in Ultraviolet-Visible and Infrared Spectrophotometry', *Analytical Proceedings*, vol. 19, pp. 28 – 33.

Fraleigh, J.B. & Beauregard, R.A. 1990, *Linear Algebra*, 2nd edn, Addison-Wesley Publishing Company, Reading, Massachusetts.

Furby, S., Kiiveri, H., & Campbell, N. 1990, 'The Analysis of High-Dimensional Spectral Curves', *Proceedings of the 5th Australasian Remote Sensing Conference*, Perth, pp. 175 – 184.

Gallagher, N.B., Wise, B.M. & Sheen, D.M. 2003, 'Estimation of Trace Vapor Concentration-Pathlength in Plumes for Remote Sensing Applications from Hyperspectral Images', *Analytica Chimica Acta*, vol. 490, pp. 139 – 152.

Gao, B-C., Liu, M. & Davis, C.O. 1998, 'A New and Fast Method for Smoothing Spectral Imaging Data', *Summaries of the Seventh JPL Airborne Earth Science Workshop*, Pasadena, pp. 131 – 140.

Goel. P.K., Prasher, S.O., Landry, J.A., Patel, R.M., Bonnell, R.B., Viau, A.A. & Miller, J.R. 2003, 'Potential of Airborne Hyperspectral Remote Sensing to Detect Nitrogen Deficiency and Weed Infestation in Corn', *Computers and Electronics in Agriculture*, vol. 38, no. 2, pp. 99 – 124.

Gong, P., Pu, R. & Yu, B. 1997, 'Conifer Species Recognition: An Exploratory Analysis of *In Situ* Hyperspectral Data', *Remote Sensing of Environment*, vol. 62, pp. 189 – 200.

Green, A.A., Berman, M., Switzer, P. & Craig, M.D. 1988, 'A Transformation for Ordering Multispectral Data in Terms of Image Quality with Implications for Noise Removal', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 26, no. 1, pp. 65 – 74.

Green, R.O., Eastwood, M.L., Sarture, C.M., Chrien, T.G., Aronsson, M., Chippendale, B.J., Faust, J.A., Pavri, B.E., Chovit, C.J., Solis, M., Olah, M.R. & Williams, O. 1998, 'Imaging Spectroscopy and the Airborne

Visible/Infrared Imaging Spectrometer (AVIRIS)', *Remote Sensing of Environment*, vol. 65, no. 3, pp. 227 – 248.

Haboudane, D., Miller, J.R., Tremblay, N., Zarco-Tejada, P.J. & Dextraze, L. 2002, 'Integrated Narrow-band Vegetation Indices for Prediction of Crop Chlorophyll Content for Application to Precision Agriculture', *Remote Sensing of Environment*, vol. 81, pp. 416 – 426.

Hakvoort, H., de Haan, J., Jordans, R., Vos, R., Peters, S. & Rijkeboer, M. 2002, 'Towards Airborne Remote Sensing of Water Quality in The Netherlands—Validation and Error Analysis', *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 57, no. 3, pp. 171 – 183.

Harsanyi, J. C. & Chang, C.-I 1994, 'Hyperspectral Image Classification and Dimensionality Reduction: An Orthogonal Subspace Projection Approach', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, no. 4, pp. 779 – 785.

He, X., Shen, L. & Shen, Z. 2001, 'A Data-Adaptive Knot Selection Scheme for Fitting Splines', *IEEE Signal Processing Letters*, vol. 8, no. 5, pp. 137 – 139.

Herut, B., Tibor, G., Yacobi, Y.Z. & Kress, N. 1999, 'Synoptic Measurements of Chlorophyll-a and Suspended Particulate Matter in a Transitional Zone from Polluted to Clean Seawater Utilizing Airborne Remote Sensing and Ground Measurements, Haifa Bay (SE Mediterranean)', *Marine Pollution Bulletin*, vol. 38, no. 9, pp. 762 – 772.

Hooper, K. & Wallace, K. 1994, 'Recovering Lake Toolibin', *Landscope*, vol. 10, no. 1, pp. 41 – 44.

Jia, X., & Richards, J.A. 1994, 'Efficient Maximum Likelihood Classification for Imaging Spectrometer Data Sets', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, no. 2, pp. 274 – 281.

Jia, X., & Richards, J.A. 1999, 'Segmented Principal Components Transformation for Efficient Hyperspectral Remote-Sensing Image Display and Classification', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 1, pp. 538 – 542.

Jupp, D. L. B. 2001, *FWHM Notes*, Earth Observation Centre, CSIRO, Canberra, [Online], Available: http://www.eoc.csiro.au/hswww/oz_pi/docs/fwhm_notes.pdf [2001, February].

Kaarna, A., Zemcik, P., Kälviäinen, H. & Parkkinen, J. 2000, 'Compression of Multispectral Remote Sensing Images Using Clustering and Spectral Reduction', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 2, pp. 1073 – 1082.

Kendall, M.G. & Stuart, A. 1961, *The Advanced Theory of Statistics, Volume 2, Inference and Relationship*, 3rd edn, Charles Griffin & Company Limited, London.

Kiiveri, H.T. 1992, 'Canonical Variate Analysis of High-Dimensional Spectral Data', *Technometrics*, vol. 34, no. 3, pp. 321 – 331.

Koger, C.H., Bruce, L.M., Shaw, D.R. & Reddy, K.N. 2003, 'Wavelet Analysis of Hyperspectral Reflectance Data for Detecting Pitted Morningglory (*Ipomoea lacunosa*) in Soybean (*Glycine max*)', *Remote Sensing of Environment*, vol. 86, pp. 108 – 119.

Kramer, H.J. 1994, *Observation of the Earth and its Environment: Survey of Missions and Sensors*, 2nd edn, Springer-Verlag, Berlin.

Lee, J.B., Woodyatt, A.S. & Berman, M. 1990, 'Enhancement of High Spectral Resolution Remote-Sensing Data by a Noise-Adjusted Principal Components Transform', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 28, no. 3, pp. 295 – 304.

Li, J., Bruce, L.M., Byrd, J. & Barnett, J. 2001, 'Automated Detection of Pueraria Montana (Kudzu) Through Haar Analysis of Hyperspectral Reflectance Data', *IEEE 2001 International Geoscience and Remote Sensing Symposium*, Sydney, 9 – 13 July 2001, pp. 2247 – 2249.

Mallet, Y., Coomans, D., & De Vel, O. 1996, 'Recent Developments in Discriminant Analysis on High Dimensional Spectral Data', *Chemometrics and Intelligent Laboratory Systems*, vol. 35, pp. 157 – 173.

Mallet, Y., Coomans, D., Kautsky, J. & De Vel, O. 1997, 'Classification Using Adaptive Wavelets for Feature Extraction', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, No. 10, pp. 1058 – 1066.

Mars, J.C. & Crowley, J.K. 2003, 'Mapping Mine Wastes and Analyzing Areas Affected by Selenium-rich Water Runoff in Southeast Idaho Using AVIRIS Imagery and Digital Elevation Data', *Remote Sensing of Environment*, vol. 84, no. 3, pp. 422 – 436.

Metternicht, G.I. & Zinck, J.A. 2003 'Remote Sensing of Soil Salinity: Potentials and Constraints', *Remote Sensing of Environment*, vol. 85, no. 1, pp. 1 – 20.

Minium, E.W., King, B.M. & Bear, G. 1993, *Statistical Reasoning in Psychology and Education*, 3rd edn, John Wiley & Sons, Inc, New York.

Newnham, G.J., Renzullo, L.J. & Lynch, M.J. 1999, 'Airborne Hyperspectral Scanner (HyMap) Validation Program at Toolibin, Western Australia: Field Trip Report', Internal Report SPS 702/1999/AP79, Department of Applied Physics, Curtin University of Technology, Western Australia.

O'Haver, T.C. 1982, 'Derivative Spectroscopy: Theoretical Aspects', *Analytical Proceedings*, vol. 19, pp. 22 – 28.

Pearlman, J.S., Barry, P.S., Segal, C.C., Shepanski, J., Beiso, D. & Carman, S.L. 2003, 'Hyperion, a Space-Based Imaging Spectrometer', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 6, pp. 1160 – 1173.

Philpot, W.D. 1991, 'The Derivative Ratio Algorithm: Avoiding Atmospheric Effects in Remote Sensing', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 29, no. 3, pp. 350 – 357.

Press, W.H., Flannery, B.P., Teukolsky, S.A. & Vetterling, W.T. 1986, *Numerical Recipes: the Art of Scientific Computing*, Cambridge University Press, Cambridge.

Price, J.C. 1975, 'Information Content of IRIS Spectra', *Journal of Geophysical Research*, vol. 80, no. 15, pp. 1930 – 1936.

Price, J.C. 1990, 'On the Information Content of Soil Reflectance Spectra', *Remote Sensing of Environment*, vol. 33, pp. 113 – 121.

Price, J.C. 1992, 'Variability of High-Resolution Crop Reflectance Spectra', *International Journal of Remote Sensing*, vol. 13, no. 14, pp. 2593 – 2610.

Price, J.C. 1994, 'Band Selection Procedure for Multispectral Scanners', *Applied Optics*, vol. 33, no. 15, pp. 3281 – 3288.

Price, J.C. 1997, 'Spectral Band Selection for Visible-Near Infrared Remote Sensing: Spectral-Spatial Resolution Tradeoffs', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 5, pp. 1277 – 1285.

Riaza, A., Strobl, P., Beisl, U., Hausold, A. & Müller, A. 2001, 'Spectral Mapping of Rock Weathering Degrees on Granite Using Hyperspectral DAIS 7915 Spectrometer Data', *International Journal of Applied Earth Observation and Geoinformation*, vol. 3, no. 4, pp. 345 – 354.

Richards, J.A., & Jia, X. 1999, *Remote Sensing Digital Image Analysis: An Introduction*, 3rd edn, Springer-Verlag, Berlin.

Roger, R.E. 1994a, 'A Faster Way to Compute the Noise-Adjusted Principal Components Transform Matrix', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, no. 6, pp. 1194 – 1196.

Roger, R.E. 1994b, 'Sparse Inverse Covariance Matrices and Efficient Maximum Likelihood Classification of Hyperspectral Data', Technical Report, Department of Electrical Engineering, Australian Defence Force Academy, Canberra 2600, Australia.

Roger, R.E. 1996a, 'Principal Components Transform with Simple, Automatic Noise Adjustment', *International Journal of Remote Sensing*, vol. 17, no. 14, pp. 2719 – 2727.

Roger, R.E. 1996b, 'Sparse Inverse Covariance Matrices and Efficient Maximum Likelihood Classification of Hyperspectral Data', *International Journal of Remote Sensing*, vol. 17, no. 3, pp. 589 – 613.

Sanders, L.C., Schott, J.R. & Raqueño, R. 2001, 'A VNIR/SWIR Atmospheric Correction Algorithm for Hyperspectral Imagery with Adjacency Effect', *Remote Sensing of Environment*, vol. 78, no. 3, pp. 252 – 263.

Sandidge, J.C. & Holyer, R.J. 1998, 'Coastal Bathymetry from Hyperspectral Observations of Water Radiance', *Remote Sensing of Environment*, vol. 65, no. 3, pp. 341 – 352.

Savitzky, A. & Golay, M.J.E. 1964, 'Smoothing and Differentiation of Data by Simplified Least Squares Procedures', *Analytical Chemistry*, vol. 36, no. 8, pp. 1627 – 1639.

Schwarz, H.R., Rutishauser, H. & Stiefel, E., translated by Hertelendy, P. 1973, *Numerical Analysis of Symmetric Matrices*, Prentice-Hall, Inc, Englewood Cliffs, New Jersey.

Serrano, L., Ustin, S.L., Roberts, D.A, Gamon, J.A. & Peñuelas, J. 2000, 'Deriving Water Content of Chaparral Vegetation from AVIRIS Data', *Remote Sensing of Environment*, vol. 74, no. 3, pp. 570 – 581.

Smith, A. 1999, *Decade of Landcare — Celebrating our Successes. The Toolibin Lake Recovery Plan*, presented at WA BankWest Landcare Conference,

Esperence Civic Centre, Esperence, 8 – 10 September 1999, [Online], Available: www.wn.com.au/landcareconf/Asmith.doc

Speed, T.P. & Kiiveri, H.T. 1986, 'Gaussian Markov Distributions Over Finite Graphs', *The Annals of Statistics*, vol. 14, no. 1, pp. 138 – 150.

Strachan, I.B., Pattey, E. & Boisvert, J.B. 2002, 'Impact of Nitrogen and Environmental Conditions on Corn as Detected by Hyperspectral Reflectance', *Remote Sensing of Environment*, vol. 80, no. 2, pp. 213 – 224.

Swain, P.H. & Davis, S.M. 1978, *Remote Sensing: The Quantitative Approach*, McGraw-Hill, New York.

Thomas, G.B., & Finney, R.L. 1988, *Calculus and Analytic Geometry*, 7th edn, Addison-Wesley Publishing Company, Reading, Massachusetts.

Toolibin Lake Recovery Plan, [Online], 1994, Available: http://www.dec.wa.gov.au/pdf/projects/salinity/toolibin/toolibin_lake_recovery_plan.pdf [1994, September]

Tsai, F. & Philpot, W. 1998, 'Derivative Analysis of Hyperspectral Data', *Remote Sensing of Environment*, vol. 66, pp. 41 – 51.

Tsai, F. & Philpot, W.D. 2002, 'A Derivative-Aided Hyperspectral Image Analysis System for Land-Cover Classification', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 2, pp. 416 – 425.

Ungar, S.G., Pearlman, J.S., Mendenhall, J.A. & Reuter, D. 2003, 'Overview of the Earth Observing One (EO-1) Mission', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 6, pp. 1149 – 1159.

Vane, G. & Goetz, A.F.H. 1988, 'Terrestrial Imaging Spectroscopy', *Remote Sensing of Environment*, vol. 24, no. 1, pp. 1 – 29.

Vane, G., Green, R.O., Chrien, T.G., Enmark, H.T., Hansen, E.G. & Wallace M. Porter, W.M. 1993, 'The Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)', *Remote Sensing of Environment*, vol. 44, pp. 127 – 143.

Vaughan, R.G., Calvin, W.M. & Taranik, J.V. 2003, 'SEBASS Hyperspectral Thermal Infrared Data: Surface Emissivity Measurement and Mineral Mapping', *Remote Sensing of Environment*, vol. 85, no. 1, pp. 48 – 63.

Walnut, D.F. 2002, *An Introduction to Wavelet Analysis*, Birkhäuser, Boston.

Wand, M.P. 2000, 'A Comparison of Regression Spline Smoothing Procedures', *Computational Statistics*, vol. 15, pp. 443 – 462.

Warner, T.A. & Shank, M.C. 1997, 'Spatial Autocorrelation Analysis of Hyperspectral Imagery for Feature Selection', *Remote Sensing of Environment*, vol. 60, pp. 58 – 70.

Watkins, D.S. 1991, *Fundamentals of Matrix Computations*, John Wiley & Sons, New York.

Wermuth, N. 1976a, 'Analogies Between Multiplicative Models in Contingency Tables and Covariance Selection', *Biometrics*, vol. 32, pp. 95 – 108.

Wermuth, N. 1976b, 'Model Search Among Multiplicative Models', *Biometrics*, vol. 32, pp. 253 – 263.

Wermuth, N. & Scheidt, E. 1977, 'Fitting a Covariance Selection Model to a Matrix', Algorithm AS 105, *Applied Statistics*, vol. 26, no. 1, pp. 88 – 92.

Yang, C-C., Prasher, S.O., Whalen, J. & Goel, P.K. 2002, 'PA—Precision Agriculture: Use of Hyperspectral Imagery for Identification of Different Fertilisation Methods with Decision-tree Technology', *Biosystems Engineering*, vol. 83, no. 3, pp. 291 – 298.

Appendix A: List of Symbols Used

In this appendix, a list of symbols used in this thesis is provided for the convenience of the reader.

α	Wavelet dilation parameter
β	Wavelet translation parameter
Δ_j	Separation between knots j and $j+1$
ϕ	Wavelet scaling function
ϕ_x	Filter function value at intersection with another filter function
ϕ_{ip}	Peak value of filter function for spectral band i
$\boldsymbol{\varphi}$	Spectral basis function
λ	Wavelength
λ_{ic}	Central wavelength of spectral band i
$\lambda_{i½}$	Full width at half maximum value of spectral band i
λ_x	Wavelength at which two spectral filter functions intersect
π	Pi, 3.14159265
θ	Angle
ρ	Partial correlation coefficient

σ_i Standard deviation about the central wavelength for filter function of spectral band *i*

$\hat{\boldsymbol{\Sigma}}$ Covariance matrix estimate

τ Atmospheric transmittance

ω Bandwidth; angular frequency

ψ Wavelet function

$\hat{\psi}$ Fourier transform of wavelet function

a Set of bands in clique; x^0 cubic spline parameter

$a_{j,n}$ Wavelet approximation coefficient

A First knot

A Transformation matrix

b x^1 cubic spline parameter

B Last knot

B Matrix for selecting non zero elements and setting terms to zero in inverse covariance matrix

c Covariance; x^2 cubic spline parameter

$C\psi$ Admissibility condition

C Sample covariance matrix

d x^3 cubic spline parameter

$d_{j,n}$	Wavelet (detail) coefficient
$\mathbf{d}$	$\mathbf{x} - \bar{\mathbf{x}}$
e	Eigenvalue
$\mathbf{E}$	Eigenvalue matrix
f	A non-negative integer; a signal
$\mathbf{F}$	Renormalization matrix
g	Discriminant function
G	Number of classes
h	Block index; clique number
h_k	High-pass filter coefficient
H	Number of blocks; number of cliques
i	Band index
$\mathbf{I}$	Identity matrix
j	Band index; level of wavelet analysis
J	Highest level of wavelet analysis
k	Class index; filter coefficient index
$\mathbf{K}$	Lower Cholesky matrix
l	Pixel index

l_k Low-pass filter coefficient

L Number of pixels

L Lower-triangular Cholesky factor

m Multiple correlation coefficient

M Multiple correlation matrix

n Wavelet coefficient index

N Dimensionality, number of bands; order of Daubechies wavelet function

o Canonical root

O Canonical root matrix

p Partial correlation element

P Partial Correlation matrix

Q Quadratic term in MLC discriminant function

R Correlation matrix

r Correlation coefficient

s Standard deviation

S Cubic spline polynomial

S Standard deviation matrix

t time

t	Canonical vector
T	Canonical vector matrix
u	"Unexplained" variance
U	Unexplained variance matrix
v	Eigenvector
V_j	Wavelet approximation subspaces
V	Eigenvector matrix
W	New modelled estimate of covariance matrix
W	Wavelet transform
W_j	Wavelet detail subspaces
x	Pixel vector
X	Previous iterative estimate of $\hat{\boldsymbol{\Sigma}}$, $\hat{\boldsymbol{\Sigma}}^{-1}$
Y	Cosignal matrix; sequence of modelled estimates of covariance matrices using forward selection procedure
z	Transformed pixel
Z	Sequence of modelled estimates of covariance matrices using forward selection procedure

Appendix B: Acronyms

ALE	Arid Lands Ecology
AVIRIS	Airborne Visible Infrared Imaging Spectrometer
CMIS	CSIRO Mathematical and Information Sciences
CSIRO	Commonwealth Scientific and Industrial Research Organisation
CVA	Canonical Variate Analysis
DWT	Discrete Wavelet Transform
EO-1	Earth Observing One
EOC	Earth Observation Centre
FSS	Field Spectrometer System
FWHM	Full Width at Half Maximum
HyMap	Hyperspectral Mapper
IRIS	Infrared Interferometer Spectrometer
JM	Jeffries-Matusita
MLC	Maximum Likelihood Classification
MNF	Maximum Noise Fraction
NAPC	Noise Adjusted Principal Components
PCT	Principal Components Transform

RGB Red, Green, Blue

RMS Root-Mean-Squared

SNR Signal to Noise Ratio

SSC Suspended Sediment Concentration

VIS/NIR Visible/Near Infrared

Appendix C: Spectral Characteristics of the HyMap Bands

Band	Band # in this research	Wavelength (nm)	FWHM (nm)	Band	Band # in this research	Wavelength (nm)	FWHM (nm)
1		435	15	33		886.2	17.9
2		446.2	15	34		904.6	23.7
3		456.7	14.6	35		919.8	17.6
4		471.2	15.5	36		934.4	19
5		486.4	17.1	37		949.3	16.9
6		502.3	15.4	38	23	966.8	21.7
7		517.1	15.8	39	24	982.3	17.2
8		533	16	40	25	997.3	17.6
9		548.4	15.7	41	26	1012.4	17.6
10	1	562.6	12.9	42	27	1030.1	17.7
11	2	579.9	14.4	43	28	1044.6	16.9
12	3	593.6	15.3	44	29	1059.6	17.4
13	4	609	15.9	45	30	1074.8	17.3
14	5	624.5	18.9	46	31	1090	17.6
15	6	640.8	15	47	32	1105.2	16.8
16	7	655.5	15.5	48		1119.9	17
17	8	670.6	15.4	49		1134.9	16.6
18	9	685.2	14.4	50		1149.6	16.7
19	10	701.7	16	51	33	1164.2	16.7
20	11	716.3	15.4	52	34	1178.9	16.7
21	12	731.4	15.4	53	35	1193.4	16.4
22	13	746.5	15.3	54	36	1208.2	16.5
23	14	761	15	55	37	1222.5	16.2
24	15	777.3	15.5	56	38	1236.7	16.1
25	16	792	15.4	57	39	1250.9	16.1
26	17	807.3	15.6	58	40	1265.3	16.5
27	18	822.4	15.9	59	41	1279.8	16.4
28	19	837.2	15.1	60	42	1293.9	16.1
29	20	853.5	17.3	61	43	1307.7	16.3
30	21	868.2	16	62	44	1321.5	15.7
31	22	883.1	16.3	63		1334.5	15.8
32		896.2	15.1	64		1347.2	14.5

Band	Band # in this research	Wavelength (nm)	FWHM (nm)
65		1401.4	15.9
66		1416.7	15.4
67		1431.2	15.3
68		1444.8	15.6
69		1458.5	15.9
70		1472.4	16.3
71	45	1486.5	16.4
72	46	1500	16
73	47	1513.4	16.2
74	48	1527.4	16.4
75	49	1540.6	16.2
76	50	1553.9	16.4
77	51	1567	16.4
78	52	1579.9	15.9
79	53	1592.7	16.3
80	54	1606	16.1
81	55	1618.5	16.4
82	56	1631.4	16.6
83	57	1643.6	16.1
84	58	1656.6	15.9
85	59	1669.1	16.1
86	60	1681.2	16.3
87	61	1693.4	16
88	62	1706.4	15.7
89	63	1718.3	16.4
90	64	1730.1	16.2
91	65	1742.1	16.4
92	66	1753.9	15.9
93	67	1766.2	15.8
94		1778.1	15.4
95		1790	15.1
96		1801.1	16
97		1948.6	20.2
98	68	1967.4	21.1
99	69	1986.4	20.9
100		2005.2	20.7
101		2023.9	20.2
102	70	2042.4	20.3
103	71	2060.7	20.1
104	72	2078.7	20
105	73	2096.9	20.1
106	74	2114.8	19.6
107	75	2132.6	19.4
108	76	2150.2	19.5
109	77	2167.6	19
110	78	2184.5	18.4
111	79	2201.5	19.3
112	80	2219.8	19
113	81	2236.8	18.6
114	82	2253.9	19.3
115		2270.9	18.1
116		2287.4	18.7
117		2304	18.1
118		2320.6	18
119		2336.9	18.2
120		2353.4	17.8
121		2369.5	18.1
122		2385.4	17.4
123		2401.2	17.6
124		2417.2	17.8
125		2433.1	17.7
126		2448.7	17.3
127		2464.3	17
128		2479.5	16.9

Appendix D: Some Previous and Future Hyperspectral Sensors

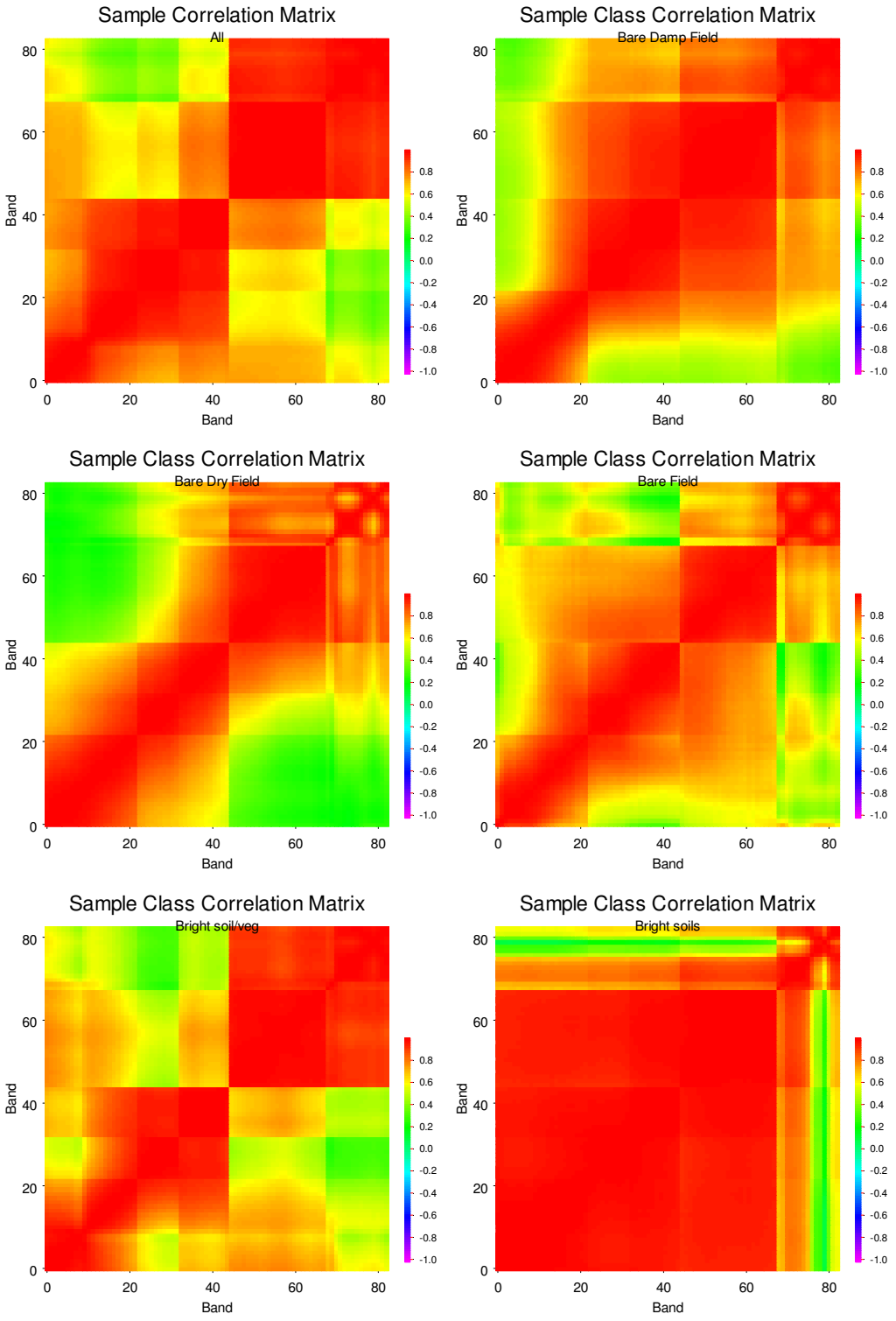
Airborne Sensors

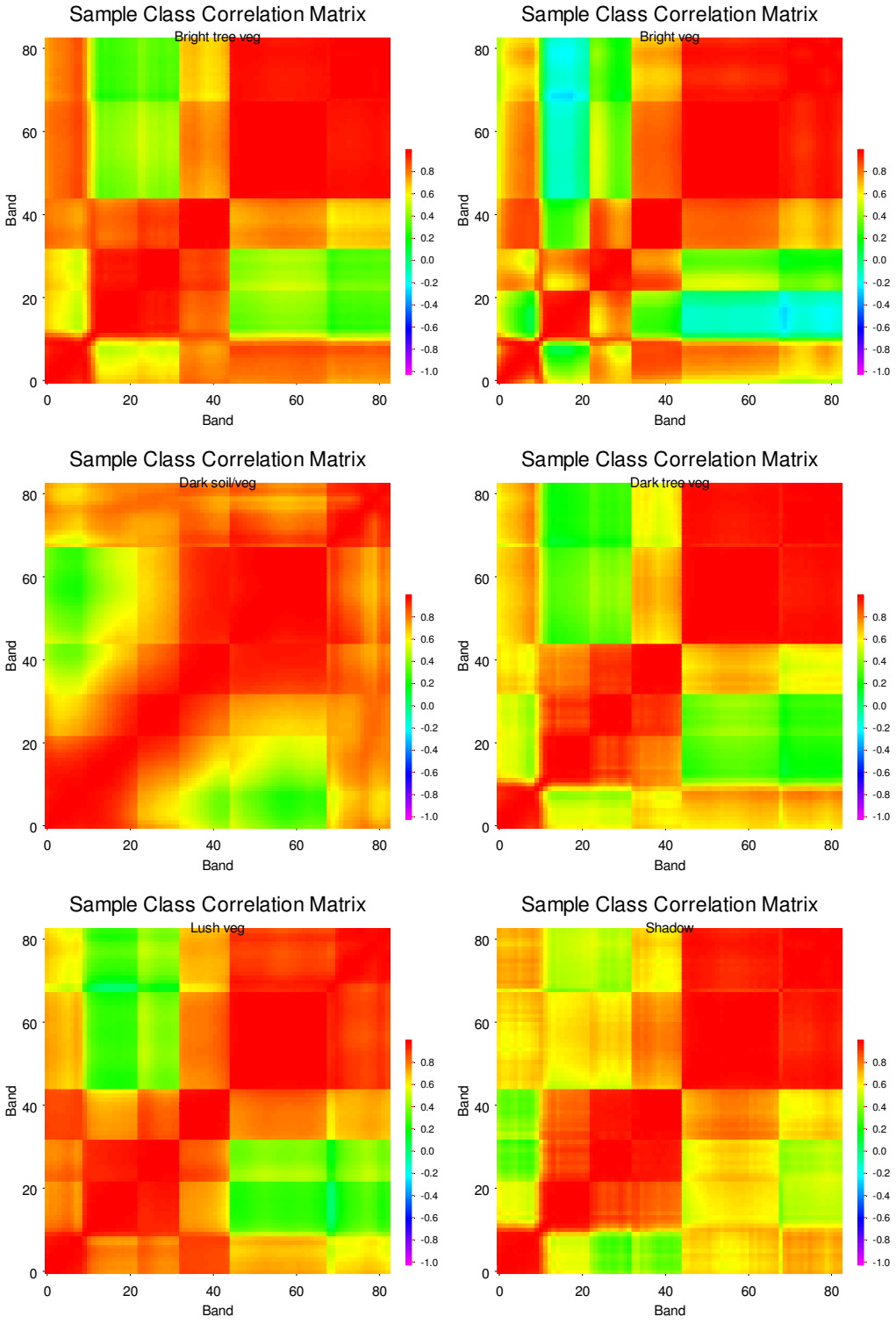
Sensor	# of Bands	Spectral range (nm)	FWHM (nm)
Airborne Imaging Spectrometer-1 AIS-1	128	900 – 2100/1200 – 2400	9.3
Airborne Imaging Spectrometer-1 AIS-2	128	800 – 1600/1200 – 2400	10.6
Geophysical and Environmental Research Imaging Spectrometer GER-63	24, 4, 29, 6	400 – 1000, 1500 – 2000, 2000 – 2500, 8000 – 12500	25 125 17.2 750
Airborne Visible/Infrared Imaging Spectrometer AVIRIS	32, 64, 64, 64	380 – 690, 670 – 1270, 1260 – 1880, 1880 – 2500	9.7 9.5 10.0 12.0
Hyperspectral Digital Imagery Collection Experiment HYDICE	210	400 – 2500	7.6 – 14.9
Digital Airborne Imaging Spectrometer DAIS-7915	32, 8, 32, 1, 6	400 – 1010, 1500 – 1788, 1970 – 2450, 3000 – 5000, 8700 – 12700	10 – 16, 36, 36, 2000, 600
Digital Airborne Imaging Spectrometer DAIS-16115	76, 32, 32, 6, 12	400 – 1000, 1000 – 1800, 2000 – 2500, 3000 – 5000, 8000 – 12000	8, 25, 16, 333, 333
Hyperspectral Mapper HyMap	128	435 – 2500	13 – 24
Operational Airborne Research Spectrometer OARS	190	500 – 2500	12

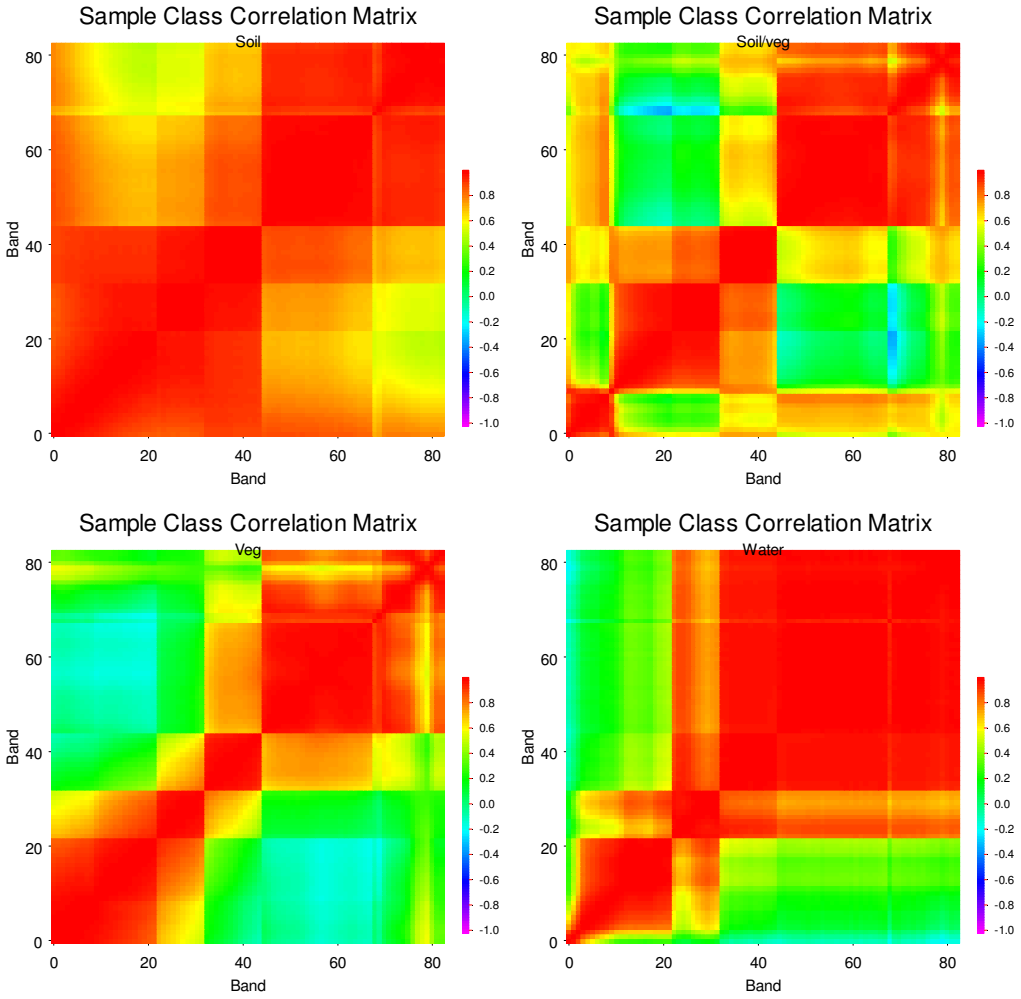
Spaceborne Sensors

Sensor	# of Bands	Spectral range (nm)	FWHM (nm)
Hyperion	220	400 – 2500	10
Linear etalon imaging spectrometer array Atmospheric Corrector LAC	256	900 – 1600	2 – 6
Compact High Resolution Imaging Spectrometer CHRIS	63	400 – 1050	2 – 12
Atmospheric Infrared Sounder AIRS	514, 602, 1262	3740 – 4610, 6200 – 8220, 8800 – 15400	1.6 – 1.9, 2.6 – 3.4, 3.7 – 6.4

Appendix E: Sample Global and Class Correlation Matrices for Toolibin Data




Sample Class Correlation Matrix
Bright tree veg
Sample Class Correlation Matrix
Bright veg
Sample Class Correlation Matrix
Dark soil/veg
Sample Class Correlation Matrix
Dark tree veg
Sample Class Correlation Matrix
Lush veg
Sample Class Correlation Matrix
Shadow
Band
0
20
40
60
80
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0


Sample Class Correlation Matrix
Soil
Band
Band
Sample Class Correlation Matrix
Soil/veg
Band
Band
Sample Class Correlation Matrix
Veg
Band
Band
Sample Class Correlation Matrix
Water
Band
Band