ERROR CHARACTERIZATION OF SPECTRAL PRODUCTS USING A FACTORIAL DESIGNED EXPERIMENT

John W. Klatt

B.Eng. Engineering and Management
Royal Military College of Canada
(1994)

A thesis submitted in partial fulfilment of the

requirements for the degree of Masters of Science

in the Chester F. Carlson Center for Imaging Science

of the College of Science

Rochester Institute of Technology

April, 2001

Signature of Author:_____

Accepted by:_____
        Co-ordinator, M. S. Degree Program

CHESTER F. CARLSON
CENTER FOR IMAGING SCIENCE
COLLEGE OF SCIENCE
ROCHESTER INSTITUTE OF TECHNOLOGY
ROCHESTER, NEW YORK

CERTIFICATE OF APPROVAL

M.S. DEGREE THESIS

The M.S. Degree Thesis of John W. Klatt has been
examined and approved by the
thesis committee as satisfactory for the
thesis requirement for the Master of Science Degree.

_____

Dr. John R. Schott, Thesis Advisor

_____

Dr. Anthony Vodacek

_____

Mr. Scott Brown

_____

Major Wayne Farrell

_____

Date

THESIS RELEASE PERMISSION

ROCHESTER INSTITUTE OF TECHNOLOGY

COLLEGE OF SCIENCE

CHESTER F. CARLSON CENTER FOR IMAGING SCIENCE

Title of Thesis:

ERROR CHARACTERIZATION OF SPECTRAL PRODUCTS USING A

FACTORIAL DESIGNED EXPERIMENT

I, John W. Klatt, hereby grant permission to the Wallace Memorial Library of the Rochester Institute of Technology to reproduce my thesis in whole or part. Any reproduction will not be for commercial use or profit.

Signature: _____

Date: _____

ERROR CHARACTERIZATION OF SPECTRAL PRODUCTS USING A FACTORIAL DESIGNED EXPERIMENT

John W. Klatt

Abstract

The main objective of any imaging system is to collect information. Information is conveyed in remotely sensed imagery by the spatial and spectral distribution of the energy reflected/emitted from the earth. This energy is subsequently captured by an overhead imaging system. Post-processing algorithms, which rely on this spectral and spatial energy distribution, allow us to extract useful information from the collected data.   Typically, spectral processing algorithms include such procedures as target detection, thematic mapping and spectral pixel unmixing. The final spectral products from these algorithms include detection maps, classification maps and endmember fraction maps. The spatial resolution, spectral sampling and signal-to-noise characteristics of a spectral imaging system share a strong relationship with one another based on the law of conservation of energy. If any one of these initial image collection parameters were changed then we would expect the accuracy of the information derived from the spectral processing algorithms to also change.

The goal of this thesis study was to investigate the accuracy and effectiveness of spectral processing algorithms under different image levels of spectral resolution, spatial resolution and noise. In order to fulfill this goal a tool was developed that degrades hyperspectral images spatially, spectrally and by adding spectrally correlated noise. These degraded images were then subjected to several spectral processing algorithms. The information utility and error characterization of these "degraded" spectral products is assessed using algorithm-specific metrics.   By adopting a factorial designed experimental approach, the joint effects of spatial resolution, spectral sampling and signal-to-noise with respect to algorithm performance was also studied.   Finally, a quantitative performance comparison of the tested spectral processing algorithms was made.

Acknowledgements

Although only one name appears on the title page of this thesis, it is only through the combined efforts of several people that this work materialized. Without their assistance, support and guidance, this thesis study would not have been possible. I would like to extend my sincere appreciation to Dr. John Schott for his suggestions and guidance throughout the duration of this project. It is through him that I have discovered an area of science that will continue to fascinate and challenge me throughout my career – military or otherwise. To Scott Brown, for his friendship, endless help (especially when results went astray), suggestions and patience when acting as my sounding board for ideas. To Erich Hernandez-Baquero for his lessons and insights into programming ENVI routines – which proved to be a critical component to this study. To Bob Alexander for sponsoring this research project and his continued support to the DIRS program. To Nick Schad, for the long hours of work in performing image degradations and data collection – my best wishes for your future studies. Most importantly, I would like to extend my most sincere appreciation to Joe Kloiber. His assistance and support was truly instrumental in the completion of this thesis. Joe, you are a great worker, and above all else - a great friend (even if you are a Sabres fan).

I would like to extend my appreciation to the Canadian Forces for providing me the opportunity to pursue my education as an integral part of my career. Along these lines, I would also like to thank my fellow officers, both American and Canadian, attending RIT – your comradeship will not be forgotten. In today's world it is important that our military teams remain both technologically advanced and in close partnership.

Finally, I would like to thank all of the friends I have made at the Rochester Institute of Technology and in the Rochester area. The friendship and fun we have shared during my two years of study has made the experience that much more enjoyable. Thank you for the memories. Cheers!

Dedication

In loving memory of my deceased grandmother, Hilda Elizabeth Feeney.

Her last words to me regarding faith continue to both challenge and inspire me.

To my Mum, Dad and sister, Ruth, for your love and support.

The light can carry all the visions of the sea
The light can carry all the images to me….

…All four winds together
Can't bring the world to me
Shadows hide the play of light
So much I want to see
Chasing the light around the world
I want to look at life – In this available light.

Neil Peart "Available Light", 1989

Table of Contents

Appendices

List of Figures

List of Tables

Chapter 1

Introduction

Ever since man was provided with a "bird's eye" view of the earth from the first manned balloon flight in 1783 earth observation technology has evolved in a manner that allows us to gain more information about the world we live in. The evolution of remote sensing started with simple photographs taken from a balloon by Nadar in 1858 (Schott, 1997) and has grown into what presently includes hi-tech aerial and satellite based electro-optical sensors. Every remote sensing system, no matter how simple or complex, has the primary goal of gaining information from an object or the world below.

In essence, information is conveyed in remote sensing data by the spatial and spectral distribution of energy that is either reflected or emitted from the earth (Landgrebe, 1978a) and is subsequently captured by an imaging system. Spatial characteristics of an image relate directly to the size, shape, pattern, site and geometry of objects within the image. Spectral information refers to the electromagnetic (EM) distribution of light and/or thermal energy. A very simple explanation of spectral information is to say what "colour" an object is. Beyond the human visual system's range more information may be extracted. For instance, vegetation appears bright at 1.0 µm and dark at 0.65 µm, whereas soil appears bright at both these wavelengths and water appears dark at 1.0 µm and dull at 0.65 µm. It is these types of differences in the spectral signature of materials that allow precise identification and discrimination of materials (Richards, 1995; Wolfe, 1997).

The study of spectral signatures of materials, known as spectroscopy, is important in the field of remote sensing. The advent and use of hyperspectral remote sensing systems exploit the spectral domain of image acquisition by sampling the spectrum at intervals of 10 to 20 nm

(Wolfe, 1997). The motivation behind spectral sampling at these minute intervals is that "detailed spectral profiling of absorption features in liquid, solids and some gaseous materials" can be accomplished (Stoner and Resmimi, 1996), thereby resulting in easier material identification and discrimination. The main idea behind hyperspectral imaging is that each image pixel contains the spectral signature of all materials located within that pixel.

Remote sensing can be simply defined as the "study associated with extracting information about an object without coming into physical contact with it" (Landgrebe, 1978a). Our eyes and ears gather data from the world around us, yet it is our brain and cognitive processes that form this data into information by which we make decisions. Analogous to this is the manner in which information is extracted from hyperspectral imagery. The electro-optical sensor gathers data from earth reflectance/thermal emittance, yet it is a series of post-processing algorithms that actually extract information from this raw data. Generally speaking, hyperspectral processing algorithms include such procedures as target detection, thematic mapping/classification routines and unmixing algorithms. The final products from these algorithms include such items as material abundance/ fraction maps, classification maps and target detection/prediction maps.

At this point it should be noted that spectral and spatial resolution share a strong relationship with one another. Furthermore, this spectral-spatial relationship is also shared with the signal-to-noise ratio (SNR) of an electro-optical imaging system. The inter-relationship of the three parameters of spatial resolution, spectral resolution and noise create a "trade-off space" that is based on the law of conservation of energy (Landgrebe, 1978b). In colloquial terms, the photons reaching a sensor can only be divided up so that they (photons) contribute to either improving one or two parameters (SNR, producing finer spatial resolution or spectral resolution) but not improving all three simultaneously. The selection of two parameters will automatically fix the third parameter given a certain sensor design.

Continuing with the previous human analogy, if our eyes and ears are not perfect then it is quite possible that we will miss valuable information. Just the same, as we vary the spectral resolution, spatial resolution and noise characteristics in a remote sensing system, we would expect the accuracy of information derived from post-processing algorithms to change. This is the main thrust behind this thesis study – an examination of how differing spectral resolution, spatial resolution and noise values effect the performance of hyperspectral algorithms and the utility of the information derived from them. We wish to characterise the error in spectrally based

information products. In essence, the utility of processed hyperspectral images will be examined. This will allow us to better predict and understand the effectiveness of these algorithms under different sensor parameters. This type of assessment also allows us to make performance comparisons between different types of processing algorithms.

In brief, the basic approach that will be followed in this research is to begin with a few differing scenes (two real images and one synthetic image) acquired from a hyperspectral imaging sensor and subject these to several hyperspectral algorithms in the pursuit of information extraction. These processed images and their final products will act as "references" for later comparisons. Subsequently, the original images will be degraded by conducting a change in the spatial and spectral resolution of the image. Different amounts of spectrally correlated noise will also be added to the images in order to degrade the SNR. These degraded images will be subjected to the same hyperspectral processing algorithms. The final product information extracted from these degraded images will be quantitatively compared, using several metrics, to the reference information. This will provide a measure of how the fidelity or utility of the image is effected by varying sensor parameters. This approach will also provide a performance comparison of different algorithms. Since a factorial designed experimental approach will be adopted for altering the levels of the sensor parameters we will also be able to see how the joint effects of spatial and spectral resolution and noise effect algorithm performance.

Chapter 2 contains a wealth of information providing a background to this research. In this chapter the hyperspectral algorithms that will be tested and the metrics employed will be discussed. Additionally, a more detailed look at the sensor parameters of spectral resolution, spatial resolution and spectrally correlated noise will be conducted. Chapter 3 outlines the approach and experimental design of this thesis. In Chapter 4 the results of the experiment are presented with an accompanying discussion of their analysis. Chapter 5 provides a summary of the work completed in this thesis study and makes recommendations towards future work that is needed in the field of algorithm performance comparison and testing.

Chapter 2

Background – Literature Review

2.1   The Very Basics

Prior to diving into the detailed background material surrounding this thesis, it is essential that some basic definitions and concepts be understood. Most importantly, a hyperspectral imaging system produces hundreds of copies of the same image each at a different wavelength. As seen, in Figure 2-1, each image pixel is essentially the spectrum of all the materials at that spatial location on the ground.
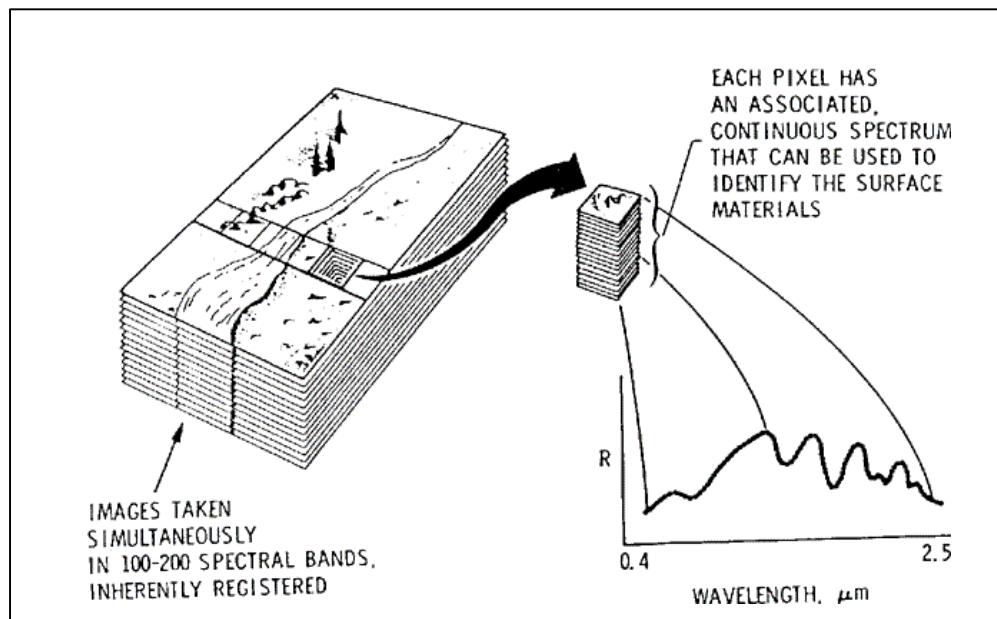
Figure 2-1: Hyperspectral Imaging Concept (RSI, 1998)

4

The image pixel is the projection of the imaging detector onto the ground from above and is called the ground instantaneous field of view (GIFOV) or ground spot size. The GIFOV represents the smallest spatial feature that can be resolved by an imaging system and is therefore used interchangeably with the term spatial resolution. Obviously the larger the GIFOV, the more ground coverage on a per pixel basis. This inevitably means more materials are included within a pixel and therefore more combining of material spectral signatures. Similar to spatial resolution, spectral resolution refers to the smallest spectral feature or rate of spectral sampling conducted by the imaging system. In Figure 2-1, if the stack of images to the left is an image taken every 10 nm, then the spectral resolution, also known as bandwidth, is 10 nm. These brief definitions serve only as starting point and will be discussed in more detail later on.

2.2   The Imaging Chain

In order to study the acquisition of remotely sensed data and the transformation of this data into useful information, a systems approach is advantageous. In this sense, the steps or chain of events of a remote sensing system, from initial data acquisition to a final useful product, can be termed as a system called an "imaging chain" (Schott, 1997).

Basically, a remote sensing system can be divided into three major subsystems – the scene, the sensor and processing (Kerekes, 1987, 1996). Although seemingly crude, this type of simplified breakdown was used in the development of the General Image Quality Equation (GIQE) which predicts image interpretability/utility based on the target, sensor and processing characteristics of a panchromatic electro-optical system (Leachtenauer et al., 1997). The GIQE is a successful demonstration that the imaging chain can be modelled quantitatively. Figure 2-2 is adopted from the GIQE work and work by Kerekes (1987, 1996) to show all of the components of the remote sensing imaging chain. An excellent and brief description of the entire system is provided by Kerekes (1987):

> "The scene contains the spectral, spatial and temporal variations of the surface reflectance and in the transmitting medium (atmosphere) which are then present at the input of the sensor. These variations include both the information bearing and information degrading types. The sensor includes all electro-mechanical effects of transforming the incident electromagnetic wave signal that represents the scene and is suitable for processing. The processing sub-system includes all effects of obtaining the desired output information from the data obtained by the sensor…in each subsystem many factors contribute to the data."

SENSOR
- detector size and optics
- System MTF (spatial resolution)
- spectral resolution
- radiometric and calibration errors
- noise
-communication/storage errors

INFORMATION

PROCESSING
- measurement conversion errors
- algorithm errors/assumptions
- computer processing errors
- human error

SCENE
- sun position
- surface characteristics
- target geometry
- atmosphere

Figure 2-2: Imaging Chain Model of a Remote Sensing System

It is clearly evident from the Figure 2-2 that there are numerous factors contributing to the final product at the end of the imaging chain. It should also be noted that many of these factors are related. It has already been established in Chapter 1 that the three sensor parameters being studied in this thesis - spatial resolution, spectral resolution and noise are interrelated based on the law of conservation of energy (Landgrebe, 1978). Another example of factor interaction is that certain atmospheric inversion techniques (discussed in more detail later) work more effectively depending on the scene characteristics. For instance, the Internal Average Relative Reflectance (IARR) method works best for arid regions with little slope and vegetation. (RSI, 1998).

Schott (1997) states that by adopting an imaging chain approach in analysing the output product or image, "we can better understand what the product means, the limitations of the product and where those limitations were introduced by weak links" in the system. Furthermore, he purposefully emphasises the point that image utility is only as strong as the weakest link in the imaging chain. The main goal of this thesis study is to quantitatively assess the accuracy and effectiveness of hyperspectral processing algorithms under differing values of spectral resolution,

spatial resolution and noise. A systematic approach to this study is paramount since we are examining how factors from one component in the imaging chain effect the results from another component in the imaging chain.

2.2.1 Processing Levels

Another way of looking at the imaging chain is to look at the evolution of different products along the imaging chain. We can define a "product" as the output from one link of the image chain. Using the chain analogy from the previous section, we can easily see that the output, or product, from each step along the image chain becomes the input of the next step. In this sense, products are the "interface points" along the processing chain (Alexander and Cheatham, 1998). Each product has defining characteristics that relate to how it was produced and how it is used. These products can more easily be discussed by dividing the processing chain into six unique levels (Alexander and Cheatham, 1998). These levels can be then grouped into three separate domains, as seen in Figure 2-3.

Processing Chain Levels and Domain Groupings

SYSTEM DOMAIN	Level 0.	Raw Image Formation
	Level 1.	Calibration
PRODUCT DOMAIN	Level 2.	Data Resampling
	Level 3.	Spectral Products
INFORMATION DOMAIN	Level 4.	Data Exploitation
	Level 5.	Reporting and Decisons

Figure 2-3: Processing Levels and Domain groupings

The System Domain consists of the first two levels of processing. Level 0 is the initial hyperspectral image data cube formed by the electro-optical sensor and either stored for later use

or transmitted via a communication link. Level 1 Calibration involves the radiometric calibration of the image data cube to generally produce a geolocated radiometrically calibrated image cube. This calibrated cube becomes the input to Level 2 processing.

Data Resampling (Level 2) is the first processing level of the Product Domain. Data resampling involves two types of transforms to the image cube. The first transform is a geospatial rectification transform of the image and the second transform is an atmospheric inversion transform that converts the data cube from units of radiance to reflectance/emittance. These two transforms will be discussed in more detail in the next section. In general, the products resulting from this processing level are directly related to these two transforms – atmospheric inversion products or geo-rectification products. A detailed list of all the products from this level can be found in Appendix A.

Spectral Data Extraction (i.e. Spectral Products) is the third processing level and the second portion of the Product Domain. This is a "transition" level in the sense that from Level 3 and onward the products are aimed more towards application specialists who may not be versed in spectral analysis (Alexander and Cheatham, 1998). The Spectral Products, or output from Level 3, are essentially information formed by the exploitation of the image's spectral content. This directly relates back to the field of spectroscopy and the ability to identify/discriminate between materials based on their spectral signature. These products fall into one of five main categories. The categories are temperature maps, classification/thematic maps, endmember or fraction maps, anomaly maps and spectral matching. It is these types of products, the algorithms that form these products and the testing of their accuracy that form the main crux of this thesis study.

The Information Domain is the last grouping of processing levels. It consists of the Level 4 – Data Exploitation in which Spectral Products from the previous level are grouped and used in direct application to solving or studying a certain problem. These products, which are typically very problem specific, become the input to the final level (5) of processing where all pertinent information that has been extracted is compiled into a report so that well informed decisions can be made in order to solve a problem or further a study.

The main goal of any imaging system is to collect information. The description of a remote sensing system as a level of interrelated products helps in describing the imaging chain. It does this by providing a qualitative description of products as they relate to the evolution of image data into image information used for decisions. The type of system definition (Alexander

and Cheatham, 1998) will provide a focus point of products levels that this thesis will investigate. This thesis is primarily looking at the accuracy and effectiveness of algorithms that result in Level 3 products based on the changes of Level 0 or sensor parameters of spectral resolution, spatial resolution and noise.   The specific types of algorithms being studied will be discussed in more detail in a subsequent section of this chapter.

2.3   Geo-rectification and Atmospheric Inversion (Level 2 Products)

Level 2 processing involves two different types of transforms to the initial image data. The first transform is a geospatial rectification transform of the image and the second transform is an atmospheric inversion transform that converts the data cube from units of radiance to reflectance/emittance. It is **<u>not</u>** within the scope of this thesis study to examine hyperspectral algorithm performance with respect to errors introduced by Level 2 processing. However, a very brief discussion of the processing involved with these two transforms and accompanying errors may prove beneficial for a better understanding of the entire imaging chain.

2.3.1   Geo-rectification/Registration

When an image is captured at two different time periods it is possible that the two images do not share the same spatial location due to varying sensor position, view angle and resolution. The process to transform the geo-metric co-ordinate system of one image to another, so that a common spatial co-ordinate system is shared, is known as registration or rectification. A hyperspectral imaging sensor captures the same image simultaneously across multiple spectral bands. Depending on the design of a hyperspectral sensor, band to band registration may be necessary (Wrigley et al., 1984; Wolfe, 1997).

The goal of rectification/resampling is to transfer the sample image/band so that it has the same geometric co-ordinates as a given reference image/band. In order to do this a relationship between the reference and sample image/band must be established by using a least-squares-fit to a polynomial equation (Schott, 1997). This type of transform equation will account for rotating, scaling, skewness, shifted pixels and perspective changes and is solved for by the selection of Ground Control Points from the two images. Another method of registration is to use block correlation techniques. This type of registration was done for the different bands of LANDSAT TM (Wrigley et al, 1984). Once these transformations are complete for either image-to-image or band-to-band registration, a method to resample the data is required. Resampling is done by an

interpolation scheme. Examples of interpolation techniques include nearest neighbour resampling, bilinear interpolation or cubic spline (Schott, 1997; Easton, 1998). Accompanying each one of these interpolation techniques are trade-offs with respect to introduced edge artifacts, image blur and preservation of spectral information (Schott, 1997).

Errors due to misregistration will inevitably occur. Misregistration errors occur in the form of shifted pixels. Conversely, spectral integrity may be sacrificed depending on the type of interpolation method that was used. These types of errors and their associated consequences must be kept to a minimum since classification accuracy may be effected by any more than a 0.3 pixel registration error (Wrigley et al., 1984).

2.3.2 Atmospheric Inversion

Information in hyperspectral imagery is based upon the ability to discriminate between and identify materials based on their spectral signatures and characteristic absorption features. This information is typically extracted through the use of various algorithms based on the retrieved surface reflectance of the scene. In this way, each pixel is a reflectance spectrum. The surface reflectance is calculated by following two conversions of the data.

The first conversion is a calibration of the sensor output, be it in digital counts or voltage, to radiance reaching the sensor (Schott, 1997). More information regarding radiance calibration can be read in Schott (1997). The second conversion of the data is known by many names, all of which mean the same thing - atmospheric correction, calibration, compensation or inversion. The atmosphere scatters and absorbs information-carrying light; thus robbing light from reaching the sensor. This results in the atmosphere acting as a low-pass filter by attenuating higher spatial frequencies in an image. The amount of blur that results in an image is dependent upon many factors including type of atmosphere (e.g. hazy or clear), constituents within the atmosphere and height of the sensor. Generally speaking, the atmosphere also has strong absorption bands at certain wavelengths thereby diminishing the spectral signature information of the objects of interest at those wavelengths. An example of this type of atmospheric absorption bands is those imposed by water at 940nm and 1140nm. The justification for methods to "erase" the atmosphere from hyperspectral imagery is obvious.

There are numerous algorithms available that perform atmospheric inversion and transform the image data into surface reflectance. Although it is beyond the scope of this study to mention each algorithm with all of its respective advantages and disadvantages, we can group

these algorithms into one of three categories. These three categories are in-scene techniques, ground-truth techniques and radiative propagation/transfer models (Schott, 1997; Kerekes, 1998). The use of a ground-truth method will be discussed in more detail in the following paragraphs since it was used in this thesis study. A full discussion regarding other approaches is available in many sources including Schott (1997).

Ground-truth methods are based on a linear regression that solve for atmospheric variables based on the observed radiance at the sensor produced by a target of known emissivity or reflectance (Schott, 1997). These targets are known either by measuring the data at the same time as the remotely sensed data or from library spectra. Therefore, with an observed radiance at the sensor and known reflectance values we are able to calibrate the image from radiance to reflectance via the following equations:

$$L_{obs} = (E_s\boldsymbol{p}^{-1} + L_d)\boldsymbol{tr} + L_u$$
$$L_{obs} = mr + b$$

(2-1)

Where L_{obs} is the observed radiance at the sensor, L_u is the upwelled atmosphere radiance, L_d is the downwelled atmosphere radiance, τ is the atmospheric transmission, r is the object reflectance and E_s is the solar spectral irradiance. Slope ($m = E_s\pi^{-1}\tau + L_d\tau$ or $\tau(E_s\pi^{-1} + L_d)$) and intercept (L_u) values found via the regression are applied to the radiance spectra for each pixel to produce an apparent reflectance spectrum at each pixel. This method, known as Empirical Line calibration method (ELM), is a very popular ground-truth approach that forces the image data to match selected field reflectance spectra (Roberts et al, 1985). This approach assumes a constant atmosphere over the image and that the known targets are Lambertian. The accuracy of these methods are highly dependent upon the accuracy of the ground-truth collection, the sensor calibration and "the uniformity of reference objects over the spatial scale of the remotely sensed measurement" (Kerekes, 1998).

An investigation of the impact of atmospheric correction techniques on the effectiveness of hyperspectral algorithm performance is **not** included as part of this thesis. The effectiveness of hyperspectral algorithm performance with respect to only spatial, spectral resolution and noise is done here. However, this would be an interesting study using this thesis as a starting point.

11

2.4 Spectral Products (Level 3 Products)

As discussed previously, various spectral algorithms are available that use the spectral content of the acquired image to form spectral products (Level 3 products). The final spectral products include such items as temperature maps, classification/thematic maps, endmember or fraction maps, anomaly maps and spectral matching. It is these types of products, the algorithms that form these products and the testing of their accuracy that form the aim of this thesis study. These algorithms are based on the field of spectroscopy and mathematically identifying/discriminating between materials based on their spectral signature. The five algorithms tested in this thesis may be categorized as either classification algorithms, spectral unmixing algorithms or target detection algorithms. In the following paragraphs each of these algorithm categories will be discussed in more detail, as will the specifics of the five algorithms being tested in this thesis study. It is important to note that the spectral algorithms tested are the algorithms as they are implemented in ENVI 3.2 (RSI, 1998).

2.4.1 Classification Algorithms

Classification or thematic mapping algorithms segment an image into its class components or materials - such as vegetation, concrete and types of minerals, for example. The algorithms assign each pixel to a class based on that pixel's spectral signature in comparison to reference spectral signatures. The reference spectra form part of a larger spectral library or are extracted as regions of interest from the image itself. The final product of a classification algorithm is a single "class map" in which each pixel is assigned to a class or material type and is indicated by a colour code, as seen in Figure 2-4. Classification maps are used in a variety of fields that include cartography, urban planning, agriculture, mining and defence.
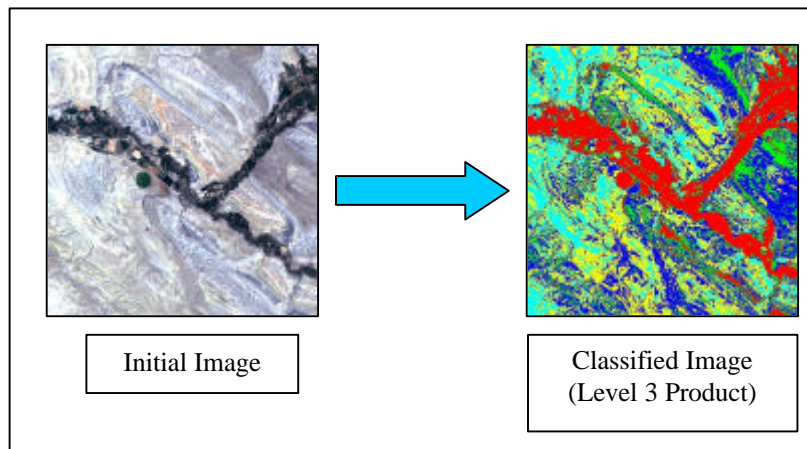
Figure 2-4: Schematic of the Classification Process

There are several methods of class assignment, which form the basis of algorithm operation. The three classification algorithms tested as part of this thesis are Binary Encoding, Spectral Angle Mapper and Gaussian Maximum Likelihood. Each classification algorithm tested in this thesis is a supervised classification method meaning that it requires information from the user regarding what endmembers to include - either from a spectral library, training data or thresholds.

Binary Encoding (BE)

The binary encoding classification technique first calculates the mean of both image and reference/endmember spectra across all bands. Subsequently, both image data and endmember spectra are encoded into 0s and 1s based on whether the spectrum's value in each band falls below or above the spectral mean, across all bands. This results in each spectrum being stored as an integer with the number of bits equal to the number of image bands and each bit representing a point in the spectrum (Kruse et al, 1993a). In ENVI, an exclusive OR function (XOR) is used to compare each encoded reference spectrum with the encoded image spectra and a classification image is produced (RSI, 1998). Pixels are classified to the endmember with the greatest number of bands that match.   The implementation of this algorithm in ENVI allows the user to set a minimum match threshold expressed as a percentage. A schematic of this algorithm can be seen in Figure 2-5.

The binary encoding algorithm is simple and fast. It has been found that this algorithm is quite accurate for identifying materials with distinct absorption bands, is insensitive to albedo variations and is not susceptible to high frequency noise (Kruse et al., 1993a).

13

Hypothetical Spectral Signature - IMAGE

Reflectance

0.9 0.8 0.7 0.6 0.5 0.4 0.3 0.2 0.1 0

400 600 800 1000 1200 1400 1600 1800 2000 2200 2400
Wavelength (nm)

Material A — Average(Material A)

Binary Encode Image Spectra

1 0 1 0 0 0 1 1 1 1 1 0 0 0 0 0 0 0 1 1 0 0

XOR = 90.9% match

1 0 1 0 0 0 1 1 1 1 1 0 0 0 0 0 1 0 1 1 0 1

Binary Encode Reference Spectra

Hypothetical Spectral Signature - REFERENCE

Reflectance

0.9 0.8 0.7 0.6 0.5 0.4 0.3 0.2 0.1 0

400 600 800 1000 1200 1400 1600 1800 2000 2200 2400
Wavelength (nm)

Ref. Material 1 — Average(Ref. Material 1)

Figure 2-5: Schematic of Binary Encoding Algorithm

<u>Spectral Angle Mapper (SAM)</u>

The SAM algorithm operates under the premise that both the image spectra and reference/endmember spectra can be represented as vectors in N-dimensional space, where N is equal to the number of spectral bands. The algorithm determines the similarity between image and reference spectra by computing the "spectral angle" between these two vectors (Kruse et al., 1993b). Following along with the explanation provided by Kruse et al. (1993b), consider two-band reference spectrum and image spectrum. The two spectra may be represented as plotted points, as seen in Figure 2-6. The line connecting each point to the origin contains all possible illuminations of that same material. It should be noted that this simple approach does not account for shading by transmissive object. That is, each line represents the material and a point on that line represents how much that material was illuminated in the image. In this regard, the SAM algorithm is insensitive to illumination factors and unknown gain factors (RSI, 1998).

14
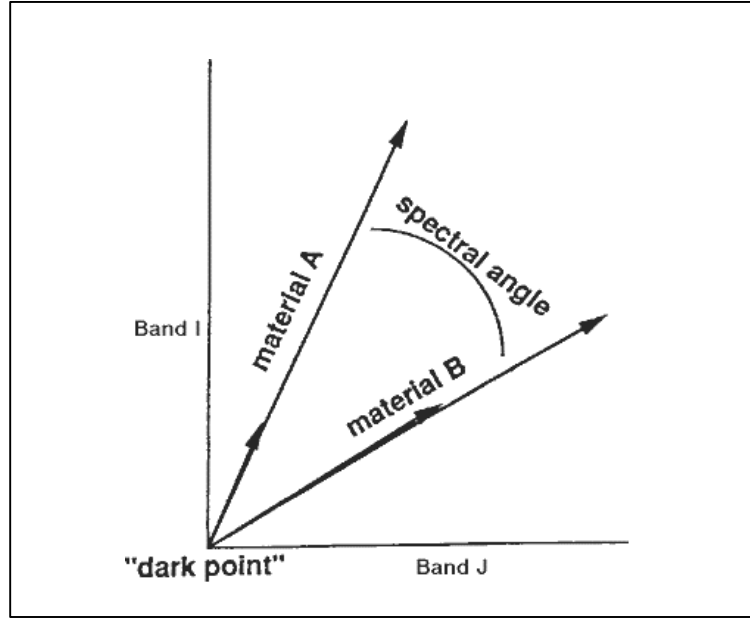
Figure 2-6: Schematic of SAM algorithm (RSI, 1998)

Geometrically, the angle between these vectors remains constant despite their length (i.e. "illumination"). SAM determines the similarity of an image spectrum t to a reference/endmember spectrum r by the following equation:

$$a = \cos^{-1}\left(\frac{\vec{t} \bullet \vec{r}}{\left\| \vec{t} \right\| \bullet \left\| \vec{r} \right\|} \right) \tag{2-2}$$

The spectral angle α (in radians) is calculated for every spectrum image in the image with respect to each reference spectrum used. A small angle between the two vectors indicates more similarity in the vectors, and the materials they represent, than a larger angle. Image spectra t are assigned to the endmember r which yields the smallest spectral angle α between them. Pixels further away than the user specified maximum angle threshold, in radians, are not classified to the class represented by **r**. More information regarding the SAM algorithm can be found in Kruse et al., 1993b.

Gaussian Maximum Likelihood (GML)

Gaussian Maximum Likelihood classification uses Bayesian probability theory and assumes that the statistics for each class in each band are normally distributed. GML calculates the probability that a given pixel belongs to a specific class and assigns each image pixel (i.e. spectrum) to the class that the pixel has the highest probability of belonging to. The following brief explanation and derivation of the GML algorithm is similar to that found in Schott (1997) and Richards (1993). These two references may be consulted for a more in-depth coverage of this algorithm.

In determining which class, j, that a pixel/spectrum X, belongs to we are most interested in the conditional probability $p(j \mid X)$. Classification is performed according to: $X \, \varepsilon \, j$ if $p(j_a \mid X) > p(j_b \mid X)$ for all $a \neq b$, where a and b represent different classes. The conditional probability that pixel X belongs to class j can be expressed as:

$$p(j \mid X) = \frac{p(j)p(X \mid j)}{p(X)} \tag{2-3}$$

where p(j) is the *a priori* probability that any class j will be observed or more simply put it is the proportion of classes. The probability, p(X), represents the chance that X occurs in the image and is the normalized multivariate histogram of the image. The term p(X) may be dropped from equation 2-3 without effecting the final results since it merely scales the calculation of $p(j \mid X)$ (Schott, 1997). The term $p(X \mid j)$ is the probability that a pixel value or spectrum is observed given a certain class selection. This probability is calculated for all pixel/spectrum and class values based on the training data provided by the user through his/her selection of endmembers and choosing image regions of interest. The term $p(X \mid j)$ is calculated, assuming that the statistics for each class in each band are normally distributed, by the following equation:

$$p(X \mid j) = \frac{1}{(2p)^{nb/2}\left|\Sigma_j\right|^{1/2}} \exp\left[-\frac{1}{2}(X - \overline{X}_j)^T \Sigma_j^{-1}(X - \overline{X}_j)\right] \tag{2-4}$$

where nb is the number of spectral bands, Σ_j is the covariance matrix for class j, $|\Sigma_j|$ is the determinant of the covariance matrix, Σ_j^{-1} is the inverse of the covariance matrix and $(X - \overline{X}_j)^T$

is the transpose of (X - $\overline{X}_j$) where $\overline{X}_j$ is the spectral mean of class j. It should be noted that Σ_j is a square matrix with dimension equivalent to the number of spectral bands. It should also be noted that "the location of the multivariate normal distribution for a class is fully characterised by the mean vector $\overline{X}_j$ and the shape of the distribution provided by the covariance matrix Σ_j" (Schott, 1997). A method to calculate Σ_j can be found in Schott (1997). Equation 2-4 is simply substituted into equation 2-3 to find the conditional probability p(j | X) of pixel/spectrum X belonging to class j given that the data is normally distributed. This final formula can be further simplified into several other discriminant-based forms (Richards, 1993). Despite these other forms the basis of assigning pixel X to class j based on a maximum conditional probability p(j | X) never changes.

Although this algorithm is quite complex it is one of the more popular classification routines since it minimises classification errors by taking into account the spectral shape of each individual class (Schott, 1997). In order that this shape is accurately predicted, large training sets of image pixels/spectra are required. The statistics of these data training sets must also reveal that the data is approximately Gaussian distributed. As we will see in another section the results of the GML classification algorithm can be further enhanced by reducing the spectral dimensionality of the image data via transforms like the Maximum Noise Fraction transform (see section 2.6.3) prior to classification. This is a common practice and will be used in this thesis study when using the GML algorithm.

2.4.2   Unmixing Algorithms

Before considering the use of unmixing algorithms and the results derived from them it is useful to understand the phenomenology behind "mixed pixels." As stated earlier, an image pixel is the projection of the imaging detector onto the ground from above and is called the ground instantaneous field of view (GIFOV) or ground spot size. Obviously the larger the GIFOV, the more ground coverage on a per pixel basis. This inevitably means more materials are included within one pixel and therefore more "mixing" of material spectral signatures. In this sense, an imaging spectrometer is an integration device in which the photons reflected from a finite GIFOV, and all the materials within that GIFOV, are integrated onto a single detector (Boardman, 1994).

Spectral mixing also occurs naturally by the spatial mixing of materials represented within a single pixel. This type of mixing falls into three different categories. **Aggregate** mixing

is the combination of materials on the macroscopic scale where the radiance in a pixel is an average of the individual materials that the sensor could not spatially separate. **Areal** mixing is occurs due to the limited GIFOV of the sensor and the pixel constituents could be separated with a higher spatial resolution imager. **Intimate** mixtures are defined by materials being combined at the microscopic level and will involve multiple interactions between materials and the incident photons (Konno, 1999). Areal and aggregate mixing models may be modelled by the linear addition of material reflectances within the finite sized GIFOV. However, intimate mixing exhibits non-linear behaviour. It is apparent that "the degree of linearity of the mixing depends on the spatial resolution of the sensor, the physical distribution of endmember materials within that GIFOV and the definition of the endmembers" (Boardman, 1994). Although there are many unmixing models available, linear unmixing techniques are an excellent approximation and work well in many circumstances (Boardman and Kruse, 1994). The linear spectral unmixing algorithm, as implemented in ENVI 3.2, is the algorithm being tested as part of this thesis study.

The required inputs into any unmixing algorithm are the hyperspectral image and the endmembers that are to be "unmixed." These are chosen from a spectral library or user-defined regions of interest. The output from this algorithm is a series of images, one for each selected endmember, that contain the relative fraction of each specific endmember at each pixel location. These pixels of these "fraction maps" ideally range from 0 to 1 and their brightness indicates relative abundance. A cartoon depicting the unmixing process can be seen in Figure 2-7.

Linear Spectral Unmixing

As mentioned in the above section, a linear spectral unmixing model is an excellent approximation for calculating the abundance or fraction of an endmember in an image pixel. The linear model can be expressed by the following equation:

$$R_b = \sum_{em=1}^{NE} F_{em} R_{em,b} + Err_b$$

(2-5)

where R_b is either the reflectance or radiance in band b, F_{em} is the weighting fraction of each endmember R_{em} in band b and $Error_b$ is obviously the error term for any unaccounted signal in band b (Pinzon et al., 1998). This says that the observed signal (reflectance/radiance) is the
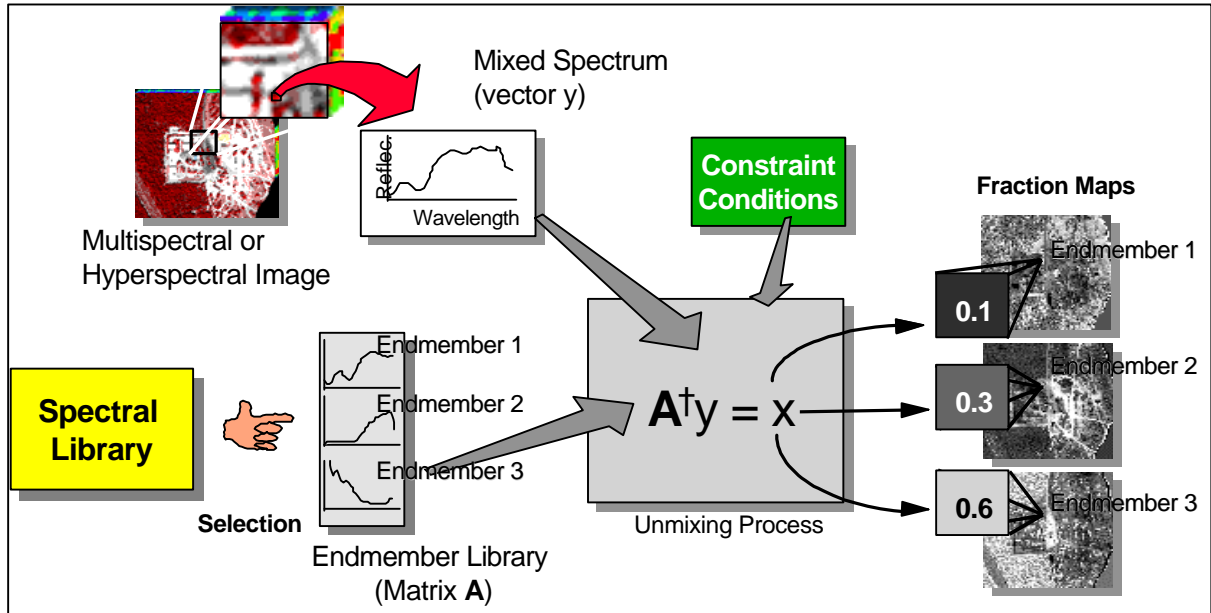
Figure 2-7: Schematic of Unmixing Process (Konno, 1999)

weighted sum of the signal from each endmember. The linear mixing model depicted in equation 2-5 can also be represented in matrix format of $\mathbf{Ax} = \mathbf{y}$, where $\mathbf{A}$ is the endmember spectra organised by column, $\mathbf{x}$ contains the fractions and $\mathbf{y}$ contains the observed radiance or reflectance. However, a simple inversion of $\mathbf{A}$ to solve for $\mathbf{x}$ may result in a solution that makes little or no sense (Konno, 1999). In this sense, constraints must be placed on the linear unmixing model in equation 2-5.

Solving equation 2-5 as it is presented is referred to *unconstrained* unmixing. The resulting fractions may assume negative values and are not constrained to sum to unity – which makes physical sense. Applying the condition that all the resulting fractions must sum to unity (i.e. $\Sigma F_{em} = 1$) is referred to *partially constrained* unmixing. This unit-sum constraint is then added to the system of simultaneous equations in the unmixing inversion process. However, fraction values which are negative or greater than one are still possible. These infeasible fractions indicate erroneous endmembers. *Fully constrained* unmixing dictates an additional condition in that all determined endmember fractions must be between 0 and 1. Obviously, fully constrained unmixing best represents the physical world but is most computationally intensive. ENVI linear

spectral unmixing has two constraint options: unconstrained or a partially constrained unmixing (RSI, 1998). The partially constrained unmixing model will be used in this thesis study.

It should be noted that the endmembers chosen for linear spectral unmixing should "explain the spectrally distinct materials that form the convex hull of the *spectral volume*" (Pinzon et al, 1998). That is, only the endmembers that explain the majority of the variance in the image should be used in selecting the reference endmembers for unmixing. This fact was underlined by Konno (1999) in which he found that traditional linear spectral unmixing worked best with 6 to 8 reference endmembers. It should also be noted that due to the mathematics behind the unmixing algorithm that the final results are dependent upon the type and number of input endmembers. That is, any changes made to the reference endmembers will alter the system of equations used to perform unmixing (equation 2-5) and therefore change the final results.

2.4.3 Target Detection Algorithms

Typically, the final product of target detection algorithms is a series of grey-scale target maps, one for each selected endmember/target. The grey-scale values are determined via floating point results from the target detection algorithm. These floating point numbers could represent the relative degree of match of the pixel to the reference spectrum and approximate sub-pixel abundance. Obviously, 1.0 is a perfect match. These floating point numbers could also be a degree of scale and measure the absorption feature depth, which is related to material abundance; and thus material presence. In either case the final product is an image that indicates some confidence level in predicting a user defined target at each image pixel location. Target detection algorithms are used in a variety of applications ranging from military reconnaissance to mining. It should be noted that the two target detection algorithms described here could also be used for some thematic mapping, but for the purpose of this study these algorithms are primarily used to detect user-input targets. As stated previously, the algorithms used are those supplied with ENVI (RSI, 1998).

Spectrally Matched Filter (SMF)

One can think of matched filtering as a means of performing a partial unmixing. That is, we determine the abundance of only the user-defined endmembers – also known as targets. The algorithm, as it is implemented in ENVI maximizes the response of the known endmember/target

and suppresses the response of the composite unknown background, thereby "matching" the known signature (RSI, 1998). Although the exact details of how this algorithm is implemented in ENVI are not explicitly stated, ENVI refers to the literature regarding the Orthogonal Subspace Projection (OSP) algorithm used for target detection (Harsanyi and Chang, 1994). The basics behind the OSP algorithm are discussed here.

The OSP algorithm (Harsanyi and Chang, 1994) begins in a similar fashion to the unmixing algorithm in that each pixel may encompass several different constituents and their respective spectral signatures. A mixed pixel at (x, y) that contains p distinct endmembers can be expressed as:

$$\vec{r} = \overline{M}\overline{a} + \overline{n}$$

(2-6)

where **r** is an nb x 1 vector where nb is the number of spectral bands. **M** is an nb x p matrix with the columns representing the endmembers spectral signature and in α is a p x 1 vector of endmember fractions. The nb x 1 vector **n** is random noise. If we are interested in the presence of only one endmember/target in the image, **d**, then we may rewrite the above equation as:

$$\vec{r} = \vec{d}\boldsymbol{a} + \overline{U}\boldsymbol{g} + \overline{n}$$

(2-7)

where α is the fraction of the target in the pixel, **U** is the nb x $(p – 1)$ matrix of all the other scene endmembers except **d,** with γ being a $(p – 1)$ x 1 vector equal to the fraction of the backgrounds. The goal of the OSP algorithm is to suppress the background effects, represented by **U,** prior to the second step of using a matched filter.

The first step is to suppress the background effects by "projecting **r** onto a subspace that is orthogonal to the columns of **U**" (Harsanyi and Chang, 1994). The second step is to perform a matched filter to find the target of interest, **d**. The overall operator that is applied to equation 2-7 in order to make this happen is:

$$\overline{q} = \overline{d}^{\,T}\overline{P}$$

(2-8)

where **P** = **I** - **UU#** and **U#** is the pseudo-inverse of **U.** The nb x nb matrix **P** suppresses the background by orthogonal projecting **r** as discussed earlier. Harsanyi and Chang (1994) provide

the complete mathematical proof and a more in-depth discussion regarding the mechanics of this algorithm.

The result of the SMF algorithm as implemented in ENVI is a series of grey-scale images, one for each selected target. As discussed earlier, the grey-scale images are based on floating point numbers that represent the relative degree of match of the pixel to the reference spectrum where 1.0 is a perfect match.

Spectral Feature Fitting (SFF)

The previously discussed SMF algorithm indicates how similar the material in the image pixel/spectrum is in comparison to reference spectra. The Spectral Feature Fitting (SFF) algorithm is an absorption based method that matches image spectra to reference spectra based on specific spectral features. Although not explicitly stated the SFF implemented in ENVI is analogous to the TRICORDER algorithm (Clark et al., 1991) and ENVI refers to this literature as background material to its SFF algorithm.

The SFF/TRICORDER algorithm requires that the image is calibrated into units of reflectance and that a continuum removal be conducted on both the image and reference spectra prior to absorption feature fitting. Fitting straight-line segments between the high points of the spectra forms a continuum. This corresponds to the background signal unrelated to the spectra absorption features of interest. Dividing the original spectrum by the continuum itself is the process of continuum removal, as depicted in Figure 2-8.

Once continuum removal has been conducted each spectrum is subtracted from one, thereby making the continuum line zero. The reference spectra are now multiplied by a scaling factor that "matches" the reference spectra absorption features to the unknown image spectra at each pixel location. "Large scaling factors are equivalent to a deep spectral feature while small scaling factors indicate weak spectral features" (RSI, 1998). A least-squares fit is calculated *band-by-band* between each reference endmember and image spectra, by utilizing the following equation. The symbol l indicates the spectral band dependence within this equation.

(2-9)

$$O_{cl} = a_l + b_l L_{cl}$$

where O_c is the continuum removed image spectra, L_c is the continuum removed reference spectra, *a* is a constant and *b* is the scaling factor discussed earlier (Clarke et al., 1991). Computing this regression fit for each image pixel produces a measure of band depth and a root-mean-square (RMS). The measure of band depth revealed by the scaling factor indicates the
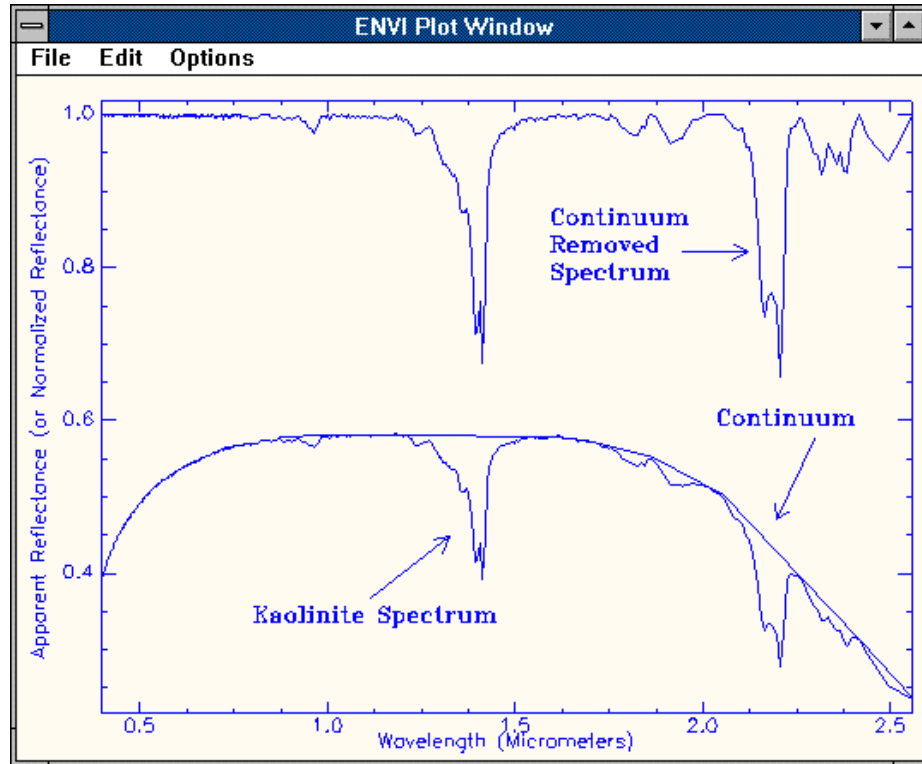
Figure 2-8: Continuum Removal of Kaolinite (RSI, 1998)

abundance and presence of a specified material in a given image pixel. The final product of this algorithm, as implemented in ENVI, is a series of scale images, one for each target spectrum, that are a measure of absorption feature depth. As previously stated, absorption feature depth is related to material abundance and presence. Also produced is an RMS image for each target endmember, which is a measure of the goodness of fit in the regression model and therefore provides some confidence as to the presence of materials in the image (RSI, 1998).

2.5 Endmembers, Ground-Truth and Reference Images

One of the most important philosophical questions to answer before beginning this thesis study was "what is truth?" That is, if we are attempting to degrade an image spectrally, spatially and by adding noise, we require references to draw quantitative comparisons about algorithm performance. Assuming that the initial images are of the highest spatial and spectral resolution, we choose the original images as the references.

Another aspect to this problem is answering the question of what constituents or endmember materials make up a hyperspectral scene. It can be easily seen in section 2.4 that many of the hyperspectral algorithms require input spectra to begin processing. There are two approaches in determining what endmembers exist within a remotely sensed scene and determining their respective spectral signatures.

The first approach is by far the simplest and most accurate. However, it is also the most costly in terms of time and resources. This approach involves acquiring knowledge through ground-truth collection about the scene. This approach involves measuring the spectral signature of the various materials in the scene using a hand-held spectrometer, usually at the same time the remotely sensed data is being acquired. In essence, a spectral library can be built from this "truth" data. Similarly, spectral libraries are formed without a particular image in mind and are the result of laboratory measurements of numerous materials. A spectral library allows the user to later select endmember spectra for use in classification and advanced spectral analysis techniques (RSI, 1998). Another method of knowing exactly what materials a scene is composed of is by using synthetic or simulated imagery, which will be discussed in more detail later on.

Similarly, if we know the type of material in a scene at a given location, we can use the spectral signature from that pixel location as an endmember's spectral signature. This *in-scene* determination of endmembers is the second approach. This technique is also straightforward as long as the exact location of recognized materials is known. ENVI 3.2 provides the user the ability to match the spectrum from a given location to library spectrum values for comparison purposes (RSI, 1998). This method is quite effective if a good knowledge of the area is used when selecting the endmembers from within the scene. If the locations and types of materials that constitute a scene are unknown then the complexity of endmember collection increases.

In the case of unknown material locations and unknown endmembers it is possible to derive the endmembers of a scene by employing different algorithms. These work under the postulate that spectra can be represented as points in an n-dimensional scatterplot, where n is the

number of spectral bands. The distribution of these points in n-space can be used to estimate the number of spectral endmembers and their pure spectral signatures (RSI, 1998). Thus, endmember selection is equivalent to finding the vertices of a simplex that enclose all the spectra of a scene in this n-dimensional space (Tsang et al, 1998). These derived endmember spectra are more pure and extreme than any other spectra in the image since every observed spectrum in the image is a linear combination of these vertices or endmembers.

The Pixel Purity Index (PPI) algorithm (RSI, 1998) uses the above theory of discovering the spectrally pure or spectrally extreme pixels. The Pixel Purity Index is found by repeatedly projecting n-dimensional scatterplots onto random unit vectors. The number of times a pixel falls at the extreme end of these projections is recorded. Obviously, the maximum number of times a pixel appears as an extreme projection point indicates its likelihood as an endmember in the scene. The ENVI user is then able to interactively select endmembers using this information (RSI, 1998). More details on how these tools were used in the selection of scene endmembers from the real imagery can be found in Chapter 3.

2.5.1 Real Imagery

Two scenes acquired from NASA's instrument AVIRIS (Airborne Visible Infrared Imaging Spectrometer) are used in this thesis study. This instrument covers the spectral range of 0.4 to 2.5 μm with an average spectral resolution (sampling interval) of 10 nm over the 224 spectral bands. The instantaneous field of view is 1 mrad which provides a ground spot size (pixel size) of 20 m when the instrument is flown at an altitude of 20 km (Vane, Green et al. 1993). The instrument boasts a spectrally averaged signal-to-noise ratio equal to 500 with reference to a 50% reflector. The reasons that images are used from this sensor are twofold. Firstly, this sensor offers very clean hyperspectral imagery with respect to noise and calibration issues at a moderate spatial resolution. The second reason stems from the excellent performance of AVIRIS in that it has become an industry standard for hyperspectral imagers.

The two scenes used can be seen in Figure 2-9. The first scene was acquired over Lake Ontario near Rochester, New York in May 1999. This image is complex and spectrally diverse in that there are many endmembers in the scene covering many possible classes (urban, water, vegetation) and respective subclasses. There is also a lot of spatial information within this scene. The second scene was acquired over Rogers Dry Lake, California in June 1998. This scene is somewhat homogeneous both spectrally and spatially speaking.
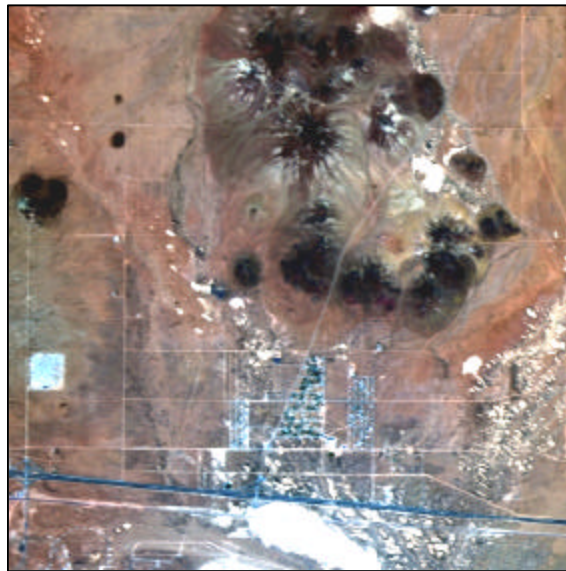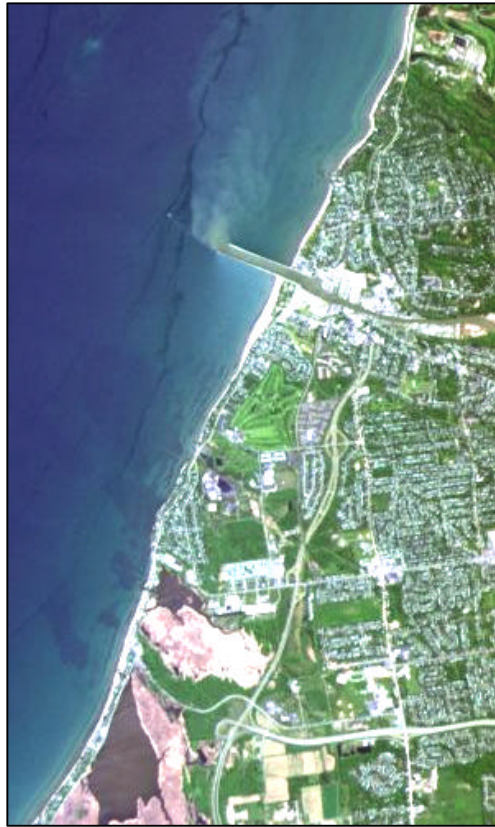
Figure 2-9: AVIRIS Images used – Rochester, NY (top) and Rogers Dry Lake, CA (bottom)

2.5.2 Synthetic Imagery

The use of simulated or synthetic imagery is extremely advantageous in the study of the imaging chain (Schott, 1997). The Digital Imaging and Remote Sensing Image Generation (DIRSIG) model is a computerized model used for image simulation and image chain modelling at RIT's DIRS laboratory. More detailed information regarding DIRSIG and its various interrelated components that model the image chain can be read in Brown (1999).

The primary reason for the use of synthetic imagery in image chain analysis is that all details of the constructed image are known.   These details include the geometry of the scene and the spatial relationships of objects in the scene. Once the scene and all of its components are constructed we can then assign material identifications to every item in the scene.   Linked to the material identification is all of the physical characteristics of that material including the spectral signature as a function of wavelength (Schott, 1997). In addition to being able to simulate the operating parameters of the sensor, we can also model radiation propagation to the sensor by incorporating MODTRAN.   It is readily apparent that all stages along the image chain are controllable when using synthetic imagery.   The DIRSIG produced truth material map reveals the exact material contained within each image pixel and is analogous to using an image with 100% complete ground-truth. Additionally, a spectral library is made for each scene constructed in DIRSIG thereby making the selection of reference spectra for processing effortless.

The DIRSIG image used is a desert scene entitled Western Rainbow. It is relatively homogeneous in the sense that the majority of the scene consists of either desert pavement or desert wash. However, there are deciduous trees and military targets (tanks and missile carriers) scattered throughout the scene. The image is 400 x 400 pixels in size with a ground spot size of approximately 2 m. The spectral range of 0.4 to 2.5 µm is covered with a spectral resolution of 5 nm. This spectral resolution is then degraded to 10 nm and the reason for this over-sampling approach will be discussed more thoroughly in Chapter 3. The DIRSIG scene used can be seen in Figure 2-10.
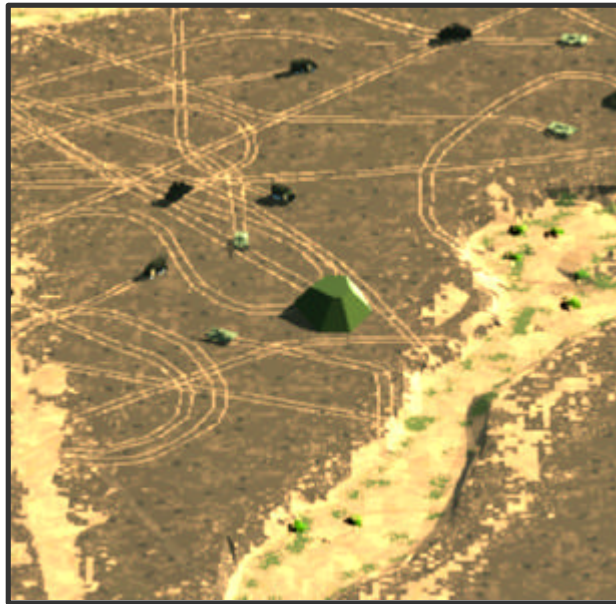
Figure 2-10: DIRSIG Image used – Example of Western Rainbow desert scene

2.6   Spatial Resolution, Spectral Resolution and Noise

As previously mentioned in Chapter 1, the spectral resolution, spatial resolution and noise characteristics of an electro-optical imaging system share a strong relationship with one another. This inter-relationship is based on the law of conservation of energy (Landgrebe, 1978b) which dictates that energy can be neither created nor destroyed but just transferred from one state to the other. Therefore, the radiance reaching the sensor can only be separated in ways that improve spatial resolution, spectral resolution or SNR - but not improving the performance of all three parameters simultaneously.   To better observe this trade-off space look at Figure 2-11 (Konno, 1999). If high spectral resolution is desired then the size of the detector is increased in order to satisfy SNR requirements since only a small amount of energy is allowed through the spectral filter. This increases the size of the sensor and a degradation of spatial resolution results. Conversely, if high spatial resolution is desired then the size of the detector is minimized. In order for SNR requirements to be met we can either choose a longer integration time (which will lead to blur for moving airborne/space imagers) or widen the spectral filter. This ultimately means coarse spectral sampling when high spatial resolution is desired.
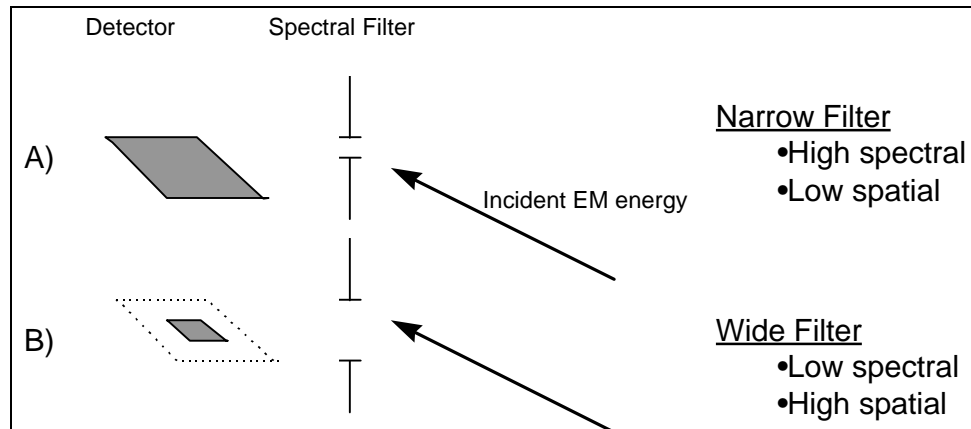
Figure 2-11: Spectral versus Spatial resolution trade-offs (Konno, 1999)

An example of this trade-off space as it relates to algorithm performance is that finer spatial resolution will result in higher purity pixels with regard to endmembers. In this case fewer spectral bands may be needed to separate the scene using a classification algorithm. However, the converse to this is also true in that there will be less spectral information to perform detailed material identification (Bowles et al, 1996). This type of spectral/spatial resolution trade-off study was conducted in the assessment of an unmixing algorithm (Pinzon et al, 1998). One of the recommendations of that study was a call for a better understanding of spatial/spectral tradeoffs. These comments are echoed by Kerekes (1996) in which he emphasizes the importance to better comprehend the relationships between spectroradiometric hardware specifications and the performance of processing algorithms.

The following subsections will discuss the spatial, spectral and noise characteristics of remote sensing systems in more detail. The first portion of these subsections includes a brief outline of how these operating characteristics effect system performance. Since a major piece of this thesis will be to measure the effectiveness of hyperspectral algorithms under differing values of spectral and spatial resolution and noise, it is necessary that methods to change these parameters within an image be established. The second portion of these subsections will outline the theory behind image degradation methods.

2.6.1 Spatial Resolution

The spatial resolution properties of digital images, including hyperspectral images, can be evaluated by an examination of the Modulation Transfer Function (MTF). The MTF is used to characterize the performance of an imaging system and is defined as the modulus of the Optical Transfer Function (OTF) or System Transfer Function (Gaskill, 1978). In essence, the MTF describes "how well sinusoidally varying brightness of a given spatial frequency will be reproduced by the imaging system" (Schott, 1997). In other words the MTF represents how well an imaging system can duplicate the spatial detail of an object as it defines how much an imaging system attenuates spatial detail. A more complete discussion regarding MTFs and the mathematics of Linear Systems and Fourier Transforms necessary for their calculation can be found in numerous sources (Gaskill, 1978; Gonzalez and Woods, 1992; Easton, 1998). The following paragraphs will use concepts from Linear Systems Mathematics (from the listed sources) in describing how the spatial resolution of an image and its degradation will pertain to this thesis study.

The Point Spread Function (PSF) of an imaging system is the response of a system to an impulse of light called a point source (Gaskill, 1978). Assuming that the imaging chain can be adequately modelled by a linear shift invariant system, we can derive the MTF directly from the PSF using the relationship:

$$\text{MTF}(\boldsymbol{x},\boldsymbol{h}) = \Im\{PSF(x, y)\} \tag{2-10}$$

In this simple equation, the PSF is in the spatial domain, MTF is in the spatial frequency domain and they are related by the Fourier Transform operator, $\Im$ (Gaskill, 1978). The symbols ξ and η denote the horizontal and vertical spatial frequencies respectively. Each component of the image chain will have an individual PSF associated with it. This is also called the impulse response, h (x, y), and can be viewed as a filter which acts on the spatial frequencies within an image. If f (x, y) represents the brightness of the original image at spatial location (x, y) and h (x, y) is the filter function (a.k.a. impulse response) then the output image g (x, y) is defined by a convolution operation (Gaskill, 1978; Easton, 1998):

$$g(x, y) = h(x, y) * f(x, y) \tag{2-11}$$

Taking the Fourier Transform of g, h and f will yield G, H and F in the spatial frequency domain and through the filter/convolution theorem (Gaskill, 1978; Easton, 1998) these functions are related by multiplication as seen in the equation below,

$$G(\pmb{x},\pmb{h}) = H(\pmb{x},\pmb{h})F(\pmb{x},\pmb{h})$$

(2-12)

where H (ξ,η) is called the transfer function and its magnitude is the MTF (Gonzalez and Woods, 1992; Easton, 1998).

Each component of the imaging chain will have a characteristic MTF. That is, each component of the imaging system helps in attenuating (i.e. "blurring") the spatial frequencies in the final image (Schott, 1997). The MTF of the system can be found by application of equation 2-13, where N represents the number of components in the imaging chain (Easton, 1998).

$$MTF(\pmb{x},\pmb{h})_{system} = \prod_{i=1}^{N} \left| H_i(\pmb{x},\pmb{h}) \right|$$

(2-13)

In order to comprehend how an imaging system maintains the spatial integrity, we must understand how each component of the system degrades spatial information. Schott (1997) provides a thorough example of this type of analysis for remote sensing systems. The final results of how each component of the image chain effects spatial attenuation can be seen in Figure 2-12.
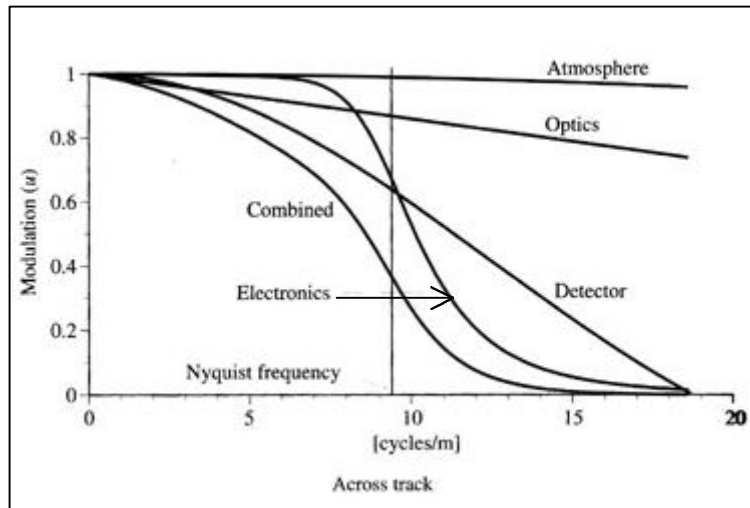
Figure 2-12: Cascaded MTFs of the atmosphere, optics, detector, and electronics for a hypothetical remote sensing system (Schott 1997).

31

From Figure 2-12 it is evident that the effects of the atmosphere, and to a certain degree the optics of the system, act as an all-pass filter. There is minimal attenuation of the higher spatial frequencies by these components in comparison to the detector. Typically, the detector is the limiting factor in image collection because its dimensions are greater than the PSF of the optics (Schott, 1997). In this detector limited case, spatial resolution is typically reported as the size of the sensor pixel projected onto the ground below and refers to the smallest spatial feature that can be resolved. As we have seen previously, this term is known as Ground Instantaneous Field of View (GIFOV), ground sampling distance (GSD) or ground spot size and is calculated by the following equation:

$$GIFOV = H \bullet \frac{l_o}{f}$$

(2-14)

where H is the altitude of the imaging sensor, l_O is the side dimension of the square detector /pixel and f is the focal length of the system (Schott, 1997).

Spatial Resampling

One of the main objectives of this thesis study is to examine the effectiveness of hyperspectral processing algorithms under differing values of spectral and spatial resolution and noise. A method to change the spatial resolution of an image is required. The method will be based on the assumption that the detector is the limiting factor of spatial resolution. Stemming from this first assumption and the application of linear systems mathematics (Gaskill, 1978; Easton, 1998) is the premise that the image, f (x, y) is sampled by a COMB function in the construction of a sampled function f_s (x, y). Finite sampling can be represented mathematically by a *window* or 2-D RECT function (Gaskill, 1978; Gonzalez and Woods, 1992) which will be denoted h(x, y) and be considered the spatial response function of the detector. In essence, the measured signal now becomes a weighted average of the input over the detector area (Easton, 1998). Mathematically the sampling process can be seen in equation 2-15 below:

$$f_s(x, y) = [f(x, y) * h_{\text{detector}}(x, y)] \frac{1}{\Delta x \Delta y} COMB\left[\frac{x}{\Delta x}, \frac{y}{\Delta y}\right]$$

(2-15)

where Δx represents the detector spacing and Δy can be the sample interval for a push-broom system (Schott, 1997). As stated already, the spatial response function of the detector, h(x, y), can be modelled by a RECT function and this is seen in equation 2-16.

$$h_{\text{detector}}(x, y) = RECT\left[\frac{x}{\Delta x}, \frac{y}{\Delta y}\right]$$ (2-16)

It should be noted that in equation 2-16 Δx is the width of the detector and we are assuming that the width of the detector is equal to detector spacing. Therefore we use Δx and Δy interchangeably. This width of the detector, Δx, is the same as l_O or the side dimension of the square detector/pixel in equation 2-14 used to determine the GIFOV of a remote sensing system. Therefore to perform spatial resampling we develop a new impulse function reflecting the degraded GIFOV, $h_{\text{detector-new}}$ (x, y). This is an *n x m x k* convolution kernel, where n x m is the spatial dimensions of the kernel and k is the spectral band dimension (Gonzalez and Woods, 1992; RSI, 1998). This new spatial filter or kernel, $h_{\text{detector-new}}$ (x, y), is then convolved with the image in the spatial domain and the result of this is re-sampled with a nearest neighbour operation to produce a "spatially resampled" image. This process is represented by equation 2-17 and the net result is an image that appears as if were taken with a lower spatial resolution sensor.

$$f_{new}(x, y) = RESAMPLED \ [f_{old}(x, y) * h_{\text{detector -new}}(x, y)]$$ (2-17)

The variables of sensor altitude and focal length remain constant given that the ground swath of the sensor is not changing. For example, if we change the GIFOV from 1m to 2m for an image that is initially 400x400 pixels, the final spatially degraded image will be only 200x200 pixels given that the ground swath does not change.

The convolution and sampling process described above is fine but computationally expensive. Using the premise that the measured signal at each pixel is the average of the input over the detector area, we adopt a neighbourhood averaging approach to spatial degradation in this thesis. That is, we resize or "shrink" the image to dimensions that are based upon the new-GIFOV and old-GIFOV given that the ground swath is constant. The new dimensions that the image is to be "shrunk" to are found by (x-dimension is only shown here):

33

$$new\ xsize = \frac{old\ xsize \bullet initial\ GIFOV}{\left(new\ GIFOV\right)} \qquad (2\text{-}18)$$

The pixels in the degraded image are the aggregate of the initial pixels in a 2x2 or 4x4 manner (and so forth) depending on the specified GIFOV that the image is to be degraded to. This type of method was used in a parameter trade-off study of target and anomaly detection algorithms (Keller et al, 2000) and can be termed an aggregate or "boxcar" approach to spatial degradation.

2.6.2   Spectral Resolution

As explained previously, hyperspectral remote sensing systems sample the spectrum (0.4 µm to 2.5 µm) at intervals, or spectral resolutions of 10 to 20 nm. The width of the spectral band, typically at full-width-at-half-maximum (FWHM) is called the spectral resolution (Wolfe, 1997) and this is usually presented in units of nm or cm^{-1}. As the resolution decreases in magnitude one can expect finer detail in the spectral information obtained from the imaging system. This allows us not only to differentiate between materials but also to accurately identify materials based on characteristic absorption features at specific wavelengths. These types of differences can be easily seen in Figure 2-13 for the mineral kaolinite. At 80 nm and 40 nm it may be difficult to tell the difference between this mineral and any other. However, as the spectral resolution improves we can better see the doublet at 2.2µm that is characteristic of kaolinite (RSI, 1998).
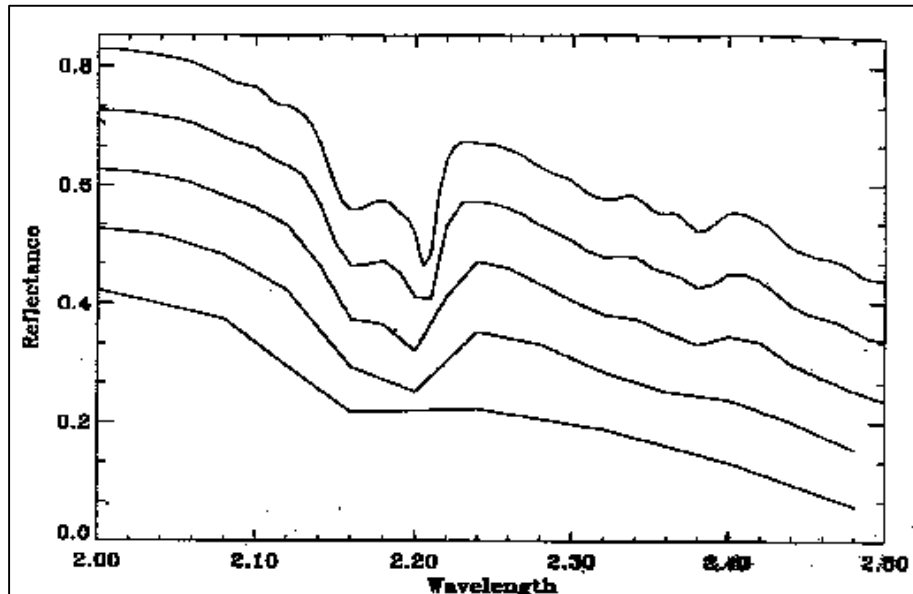
Figure 2-13: Kaolinite at different spectral resolutions. Spectral resolution from top to bottom: 5, 10, 20, 40, and 80-nm resolution. The spectral curves have been offset from one another to better visualise the differences attributed to changes in spectral resolution.

It is important that the terms spectral resolution and spectral sampling are not confused. Spectral resolution is the width of the spectral band-pass while spectral sampling refers to the band spacing or "the quantization of the spectrum at discrete steps" (RSI, 1998). Wolfe (1997) refers to band spacing as "free spectral range" or the spectral interval between resolution "peaks." Despite their difference in meaning these terms are often interchanged because the majority of spectrometers are designed so that the band spacing is about equal to the band FWHM (Wolfe, 1997).

The ultimate question is how many spectral bands of information are actually needed in the processing of hyperspectral images to obtain accurate information. This becomes a very complex question to answer that involves the topics of are of application, data dimension reduction and spectral band trade-off studies.

It is readily apparent that as the spectral resolution improves, there is a high correlation of information between adjacent bands. This means that the sensor is actually taking the repeated measurements of the same quantity (Green et al, 1988). Most of the information about a scene can described by less than ten dimensions (Harsanyi and Chang, 1994). There are several methods, based on Principle Component (PC) transform, that reduce the dimensionality of hyperspectral data (Green et al., 1988; Lee et al., 1990; Johnson and Wichern, 1998). For example, the final product of the Maximum Noise Fraction (MNF) transform (Green et al., 1988) is a series of images in descending order of SNR. Although the MNF transform will be discussed in more detail in section 2.6.3, it can be mentioned briefly here that the first few images from this transformation contain the majority of information base on SNR derived from the calculated eigenvalues of each band. Another advantage of these types of transformations is that noisy bands are separated and in essence eliminated from the data set prior to processing (Lee et al, 1990). However, these types of transforms are a linear combination of information from several spectral classes since the spectral signatures are correlated (Harsanyi and Chang, 1994). This results in newly found PC dimensions that often lack intuitive interpretability.

In addition to studies that reduce the dimensionality of hyperspectral imagery, there have been numerous investigations into the number of bands required in the analysis of hyperspectral data. These studies have typically been done with respect to general tasks such as target detection or thematic mapping. When similar materials are grouped together, such as man-made objects and naturally occurring objects, it is found that the spectral signatures of the two different groups are uncorrelated. However, the spectral signatures within each individual group (i.e. the

35

man-made group may contain asphalt and concrete) are highly correlated and these correlation properties are true for signatures taken over more than two bands (Haskett and Sood, 1998). Applying this knowledge to target detection algorithms reveals that increasing the number of bands may not necessarily improve detection performance. Improving spatial resolution and retaining fewer spectral bands will ultimately improve detection performance (Haskett and Sood, 1998). This is because a higher spatial resolution ensures that the pixels will contain more purely distinct and less mixed spectra. These results where echoed in a study in which it was empirically determined that matched filter target detection algorithms showed a lack of dependence on spectral resolution (Keller et al., 2000).

Somewhat similar results were found in a study that examined the effects of spectral resolution in separating a hyperspectral scene into its constituents (Bowles et al., 1996). This study involved an algorithm that calculated matched filters from image derived endmembers to separate the scene. The conclusions of this study show that general thematic mapping through the use of classification algorithms can be accomplished adequately with 10 or more wavelengths and not hundreds. However, for a more detailed analysis and improved discrimination between specific materials the study concludes the obvious - more wavelengths are needed. Unlike the study by Haskett and Sood (1998), this study by Bowles (1996) admitted that degrading the spectral resolution eventually resulted in a loss of spectral contrast and "therefore the ability to discern targets was compromised" (Bowles et al., 1996).

It is apparent from the above discussion that the level of spectral resolution is both application and algorithm specific. It is also evident that certain algorithms, such as target detection processing, may place a higher dependence on the spatial information within an image over spectral information. Again, this points to the requirement for a study of the spectral-spatial resolution trade-off space with respect to algorithm performance.

Spectral Resampling

To study the effects of spectral resolution on the performance of mapping and target detection algorithms, a method to change the spectral resolution of the original images is needed. The following paragraphs outline a few of the available methods to perform spectral resampling of hyperspectral images.

The first method is a spectral binning process (Haskett and Sood, 1998; Keller, 2000) whereby neighbouring bands of the original spectral signatures are added together to create a new

signature appearing as if it where taken over fewer bands.   In this way, the bands of the new image represent a wider spectral bandwidth. For example, if two adjacent bands are binned together then an original 210-band image/spectrum will now become a 105-band image/spectrum. This works but it does not realistically simulate how a sensor operates spectrally.

The second and more precise method is to perform spectral resampling. ENVI v3.2 (RSI, 1998) allows the user to resample spectral libraries and more importantly hyperspectral image files.   There are different methods available to the user to perform spectral resampling which include using the response spectral response function of a given hyperspectral sensor, wavelength files or user-defined response function/wavelengths. In all cases ENVI assumes critical spectral sampling and uses a Gaussian model with a FWHM equivalent to the band spacing to perform spectral resampling (RSI, 1998).   This method more realistically simulates sensor operation.

2.6.3   Noise

Noise can be defined as "any source or effect that occurs in a system that is not information bearing or degrades the desired information of the output" (Landgrebe and Kerekes, 1987). In this manner, we can see that noise not only measures the quality of the signal (Schott, 1997) but noise can be viewed as dependent upon what the "signal" will be used for.   One of the objectives of this thesis is to study the effects of noise on hyperspectral algorithm performance. Although it is beyond the scope of this thesis to categorize and describe each noise source in a remote sensing system, a very brief discussion may serve useful.

As mentioned in section 2-2, the image chain consists of three components – the scene, the sensor and processing. Each on of these components introduces data variation or noise to the final signal or image.   Kerekes and Landgrebe (1987) present a fully detailed taxonomy of noise sources for remote sensing systems.   Figure 2-14 provides this list of possible noise sources and how they are categorized according to where they are introduced along the image chain.
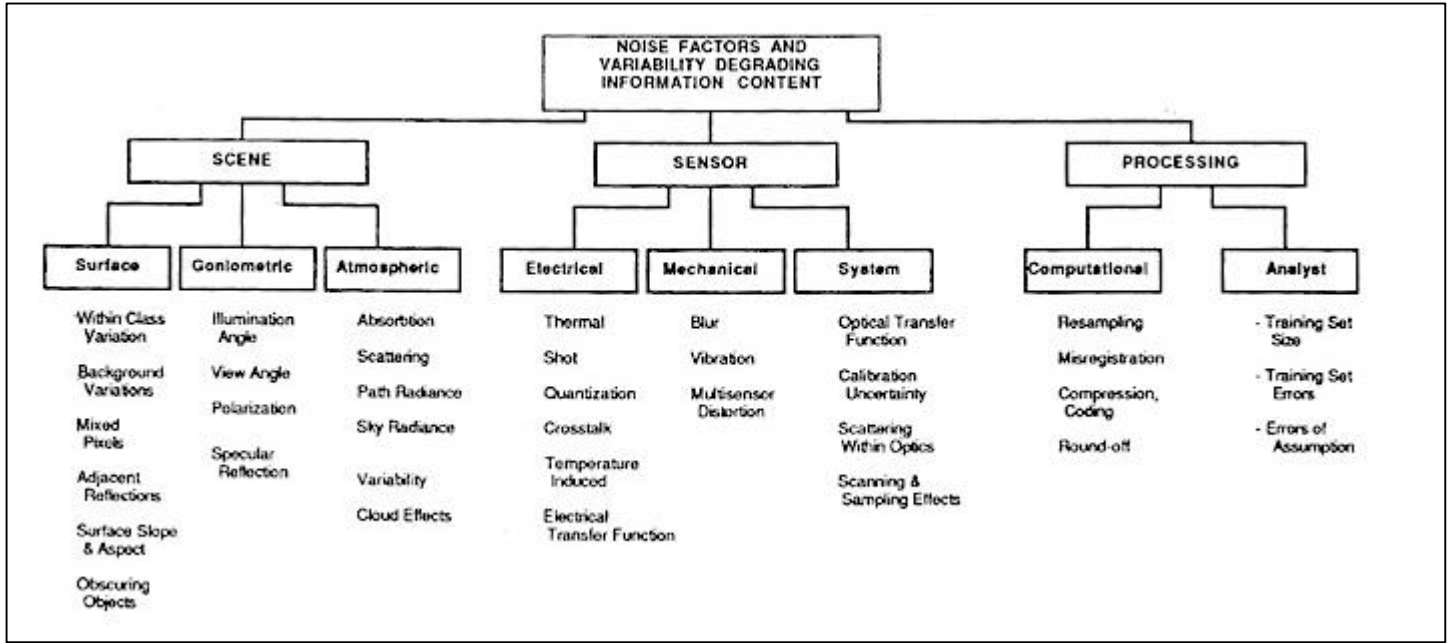
Figure 2-14: Noise and Information degrading effects in a remote sensing system

(Kerekes and Landgrebe (1987))

Noise can also be seen as the random variations about the mean signal level and is measured by the root-mean-square (RMS) variation of the instantaneous signal, S_i, with respect to the average signal, S_{avg}, as seen in equation 2-19 (Schott, 1997).

$$N_{RMS} = \left[\frac{\sum_{i=1}^{n}\left(S_i - S_{avg}\right)^2}{n} \right]^{\frac{1}{2}} \quad volts \tag{2-19}$$

Noise is usually a more useful term when it is measured with respect to the signal as done when expressing a signal-to-noise ratio (SNR). Typically, when a SNR is provided as part of sensor specifications it is with reference to a certain degree of reflector or source of flux. As an example, the SNR of AVIRIS is given with respect to a 50% reflector. It should also be noted that the SNR of a hyperspectral sensor is also a function of wavelength. Other times when noise is discussed with respect to detector performance specifications it is provided in terms of radiometric input units (watts) vice the output units of the signal (volts). In this manner a detector performance metric is the noise-equivalent-power (NEP) which is wavelength dependent (Schott, 1997).

$$NEP(\lambda) = \frac{N_{RMS}}{R(\lambda)} \qquad watts$$

<div align="right">(2-20)</div>

In this equation $R(\lambda)$ is the spectral responsivity of the sensor. The NEP(λ) is a level that the incoming signal must be above in order to be detected (Schott, 1997). From the calculation of the NEP of the stystem we can also compute many other sensor performance specifications. These include noise equaivalent radiance (NER) of a sensor, noise equivalent reflectance (NE$\Delta\rho$) and noise equivalent temperature (NEΔT). These are similar in definition to NEP and a more detailed discussion of these metrics can be found in Schott (1997).

One of the primary interests of this thesis study is to investigate the effects of noise and spatial-spectral resolution changes on hyperspectral algorithm performance. Similar studies have been done in the past that have studied only the effects of noise. These studies not only result in many intuitively expected conclusions, but also demonstrate the interrelationships of several noise factors. The following paragraphs will highlight some of the results of these studies.

One of the most obvious results is that it was determined that adding increasing amounts of white noise decreased classification accuracy (Kerekes and Landgrebe, 1987). A study of periodic noise that results from either striping between detectors, electronic coupling or induced by outside power supplies was performed with Landsat TM (Wrigley et al, 1984). They found that periodic noise added unwanted spatial frequencies into the final image and periodic noise components obscured detail in low contrast areas of certain bands. Although this study worked primarily with multispectral imagery, the results may be extended to hyperspectral imaging systems.

Kerekes and Landgrebe (1987) further illustrate the interrelationship of noise effects on classification accuracy. The addition of higher additive noise levels was found to have a more adverse effect on images acquired under poor visibility than when acquired under excellent remote sensing conditions. Similar to this study, Landgrebe and Malaret (1986) investigated the interrelated effects of atmospherically introduced noise, sensor noise, pre-amplification noise and quantization noise on classification performance. The results of their study show that the impacts of shot noise and thermal noise are similar with respect to classification performance. Most importantly, their study demonstrates that "the impact of the atmosphere makes the impact of the other noise sources more significant as atmospheric conditions deteriorate" (Landgrebe and Malaret, 1986). One could say that the atmosphere amplifies the other noise sources. Another

study found that classification accuracy is dependent upon the type of noise added to the imagery (Lee and Landgrebe, 1993). Their investigation concluded that uncorrelated noise added to each spectral band has a more adverse effect with respect to classification accuracy than adding larger values of correlated noise. Finally, a study of target and anomaly detection algorithms revealed no surprises in that the probability of detection monotonically decreases as more white noise is added to the initial image data (Keller et al., 2000). Although all of the above studies were well investigated they all have one fundamental problem in that either the type of noise added is not mentioned or it is added white noise. In the case that the type of noise addition is not mentioned it most likely that it is white noise or some type of random Gaussian noise. Indeed this type of noise addition will degrade the desired information of the algorithm output. However, the type of noise addition conducted in the above studies does not necessarily reflect the true behaviour of hyperspectral sensors. It was previously mentioned that as the spectral resolution of a sensor improves, there is a high correlation of information between adjacent bands. This is just the same for the noise – it is structured and has correlation.

Noise Modelling and Addition to Imagery

In its most basic form, the imaging process can be represented mathematically by equation 2-21 (Gonzalez and Woods, 1992, Easton, 1998). In this equation $\mathbf{f}$ (x, y) represents the object's spectral distribution at spatial location (x, y) and $\mathbf{g}$ (x, y) represents the final image. Also, in this equation h (x, y) is the impulse response of the entire system and $\mathbf{n}$ (x, y) is the added noise to the entire process. This added noise comes from the many sources already discussed. The bold face (and vector notation) of these functions indicates that they are vectors representing the spectral information at location (x, y).

$$\vec{g}(x, y) = \vec{f}(x, y) * h(x, y) + \vec{n}(x, y)$$

(2-21)

This type of simple additive model has been used as an overall system noise model (Landgrebe and Maralet, 1986) to assess the effects of types of noise on classification routines. In this type of modelling the total noise, $\mathbf{n}$ (x, y), was further broken down into the noise contributions of the system components. This can be seen in equation 2-22, where k represents the total number of system components that the noise model will include.

$$\overrightarrow{n_{total}}(x, y) = \sum_{i=1}^{k} \overrightarrow{n_i}(x, y) \qquad\qquad (2\text{-}22)$$

The noise contribution of each system component, $\mathbf{n}_i$ (x, y), is an independent random vector that can be statistically modelled. For example, a Gaussian distribution whose variance is proportional to the signal level can model shot noise. It is also known that thermal noise is independent of signal level, white over a large bandwidth and also Gaussian. However, quantization noise is modelled most accurately using a uniform distribution (Landgrebe and Maralet, 1986). One can model every type of noise using the correct statistical distribution with different standard deviations and add this to the initial image. However, the bookkeeping involved in this approach will soon prove cumbersome. Further investigation of equation 2-21 reveals the simple fact that some simple algebra may be applied to find a suitable spectrally correlated noise model. If the final image $\mathbf{g}$ (x, y) is the "sum" of the convolved initial spectral distribution of the object $\mathbf{f}$ (x, y) and noise, $\mathbf{n}$ (x, y), then the noise is the simple difference between the final image and the object's spectral distribution. Unfortunately, this mere subtraction is not as elementary as what is presented here and we require an advanced method to separate the noise from the final image.

Maximum Noise Fraction (MNF) Transform

The MNF transform allows noise to be removed from an image by transforming the initial image data into "MNF-space" and then smoothing or removing the noisy components prior to re-transforming the data into the original image space (Green et al, 1988). Let $\mathbf{Z}$(x, y) represent the image spectrum at spatial location x. Using the premise behind equation 2-21 we arrive at:

$$Z(x, y) = S(x, y) + N(x, y) \qquad\qquad (2\text{-}23)$$

where S(x, y) is the signal and N(x, y) is the noise. Now let Σ_S and Σ_N represents the covariance matrices of the signal and noise respectively. If we assume signal and noise are not correlated then the overall covariance Σ is the mere sum of Σ_S and Σ_N. The noise fraction in the ith band is defined as the ratio of the noise variance to the total variance for that i^{th} band (Green et al, 1988). This can be represented mathematically as: $VAR\{N_i(x, y)\} \div VAR\{Z_i(x, y)\}$. Similar to principle components analysis (Johnson and Wichern, 1998), the maximum noise fraction transform

chooses linear combinations of the form $Y_i = \mathbf{a}^T_i \mathbf{Z}(x, y)$ such that the noise fraction for Y_i is the maximum among all linear transformations and orthogonal to Y_{i+1}. That is, the first MNF band contains the most noise and the least information while the last MNF band contains the least noise and most information. This ordering is based upon the descending eigenvalues of $\mathbf{a}$, which are equivalent to the noise fractions (Green et al., 1988). The final results of this algorithm presented in ENVI (RSI, 1998) are the opposite of what is presented by Green et al. The MNF bands in ENVI are ordered by descending order of information so the first band contains the most information and the last band contains the most noise.

The image in MNF space can be divided into two parts. The first portion consists of those MNF-bands that contain information and the second portion includes those MNF-bands that are dominated by noise. Noise can be removed from the image data by performing an inverse MNF transform using only the bands that contain useful information (Green et al, 1988; RSI, 1998) and ignoring the MNF-bands purely characterised by noise. Likewise, the dimensionality of the image data can be reduced for processing algorithms by working in MNF space with only those bands containing useful information and again ignoring the noisy bands. In the next chapter it will be evident how this algorithm is used in adding spectrally correlated noise to images.

2.7   Image Information Utility Metrics

The primary objective of this thesis is to assess the effectiveness of hyperspectral processing algorithms under differing values of spectral resolution, spatial resolution and noise. In essence we wish to measure the information utility of processed hyperspectral imagery. Obviously, any approach in conducting this must include the use of a metric(s). We require a repeatable and reproducible means of objectively quantifying the nature and extent of image utility after algorithm processing. Whatever means are used in quantifying image utility, the metric must be meaningful and relevant to the spectral product being assessed. The metric must account for both spectral and spatial information.

As we will see in the following sections, fulfilling the above criteria for a metric is a challenge. This is because some imaging chain products, discussed in sections 2.3 and 2.4, are expressed in terms of engineering units (e.g. DC, radiance, reflectance) while other product levels are defined in terms of end-user metrics (e.g. accuracy, fractions, probabilities). The following sub-sections outline various possibilities in measuring image fidelity and information utility of degraded and processed hyperspectral imagery. Several metrics are included below, yet as we will see, not all of the metrics meet the above criteria for measuring the effectiveness of hyperspectral algorithms. The result of the next sub-sections will be the selection of useful, meaningful and repeatable metrics for assessing algorithm performance under differing values of spectral resolution, spatial resolution and noise.

2.7.1   Information Theory

The main goal of any imaging system is to gather information and from this perspective it makes intuitive sense that information, itself, be included as a metric. In order for a calculation of information to be carried out by the use of information theory, an analogy must be drawn between an imaging system and a communication channel. In brief, a communication channel consists of a source producing a message, transmitter that encodes the message, a channel to carry it, a receiver to decode the message and a destination. The mathematical theory of communications (Shannon, 1948) forms the basis of information theory and quantifies information as a metric called entropy. Entropy (H) corresponds to statistical variability or uncertainty and has the equation:

$$H = -\sum_i p_i \log p_i \qquad [bits] \qquad (2\text{-}24)$$

where p_i is the probability of a specific numerical value or event "*i*" in a series of measurements (Shannon, 1948). Obviously a higher entropy value means that there is more uncertainty about an event or measurement. Hence, there is more derived information about this specific numerical value since there is more uncertainty surrounding its occurrence. Many other metrics can be derived from this basic entropy definition to numerically describe information (Shannon, 1948).

Using the analogy between a communication channel and the imaging chain it was found that an informationally optimized imaging system will ensure maximum fidelity of restored images with respect to spatial features (Huck et al, 1985). Furthermore, it was found that an informationally optimized design is preferred when processing of the image is to be later conducted (Huck et al, 1985). These results were tested in the study of design trade-offs between detector shape and size and electronic filters of an imaging spectrometer (Feng, 1995). In this study it was again found that the informational optimized system maximizes the fidelity of images and edges.

Evidently information theory is useful in the initial design of an imaging system. However, its use as a metric when working with existing images proves to be limited. Since this thesis study will involve the accuracy measurement of algorithm performance, it will be necessary to calculate the entropy of images. There are a few different approaches in estimating the information content in an image (Gonzalez and Woods, 1992). The first method involves an assumption that the image was produced by a source that emits statistically independent pixels and follows some type of probability distribution. The second method involves determining the frequency of occurrence of the same pixel in an image. It is from this frequency that a probability of occurrence of that pixel can be found and applied to equation 2-24 to determine the entropy (*first-order estimate*). The third method and final method is an extension of the second whereby the relative frequency of pixel blocks is determined (*higher-order estimates*). These blocks are essentially groups of neighbouring pixels and "as the block size approaches infinity, the estimate of information approaches the true entropy of the source" (Gonzalez and Woods, 1992).

There are numerous problems with calculating the information content of hyperspectral imagery - let alone Level 3 products. First-order estimates of entropy assume that pixels are statistically independent. This is not necessary true considering pixel bleed-over during image acquisition and the geometrical interdependence of some scenes (i.e. scene homogeneity). This problem is even extended to the higher-order estimates where blocks are considered statistically independent and as the order of the estimate increases so does the complexity of the calculation.

Another problem is that since hyperspectral imaging is being considered we must take into account the multiple data channels. Information corresponds to interchannel variability and this allows us to eliminate equal readings from two different spectral channels (Price, 1984). This is because new information is not added by repeated readings.   Therefore, an estimate of source entropy from a hyperspectral image must treat each pixel as statistically dependent and the redundancy of channel readings within that pixel must be considered since each pixel is a spectrum.   The complexities of such a calculation make this metric very ugly and we have not discussed how to possibly calculate the information content of a Level 3 product.

The final problem with this method is that entropy calculations of this sort are highly dependent on the scene characteristics.   Since information is associated with variability, any variability in the scene itself will effect the final estimation of entropy. For example, one "expects to find more information per unit area in a heterogeneous region like a city than in a uniform region like a grassy plain" (Price, 1984) since there is more scene variability in an image of a city. In fact, the entropy of a "clean" image will be less than that of a noisy image, again because more variation is introduced by noise addition. This does not fulfil the requirements for a repeatable or meaningful metric of processed images.   It can easily be seen from the above discussion that the use of information as a Level 3 (let alone Level 1 or 2) metric is pointless in this study.   However, from an academic perspective it was a worthy investigation.

2.7.2   RMS Error

A very simple approach is to use the principle that every pixel is a spectral vector and then monitor the change of vector direction and magnitude along the image chain. This can be done by looking specifically at the average spectral signature in regions of interest (ROIs), the changing spectral signatures of known endmembers or the changes in fractional endmember images and other Level 3 products. This type of approach can be quite useful in assessing the degradation of the image. However, adopting this method to assess the accuracy of algorithms by measuring the magnitude and direction of pixels from different Level 3 products is not useful in that it is not readily interpretable.

Although the above method is not completely useful, its basis of expressing information loss as a function of the original image can be extended to other metrics. A good example of this is the root-mean-square (RMS) error between an input and output image (Gonzalez and Woods, 1992).   Let $f(x, y)$ represent an input image and let $f'(x, y)$ represent the same image after it is

degraded by changing the spectral and spatial resolution and adding noise. Obviously, both f and f ' are vectors representing the spectral signature of the pixel at spatial location (x, y). The error, e (x, y), between the input and output image can be expressed as:

$$e(x, y) = f^{'}(x, y) - f(x, y) \tag{2-25}$$

The RMS error is found by taking the square root of the total error between the two images and averaging it over the entire image (image size is M x N), as seen in equation 2-26 (Gonzalez and Woods, 1992).

$$error_{RMS} = \left[\frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \left[f^{'}(x, y) - f(x, y) \right]^2 \right]^{\frac{1}{2}} \tag{2-26}$$

These metrics offer an understandable and repeatable method of assessing image degradation along the image chain. These types of metrics are useful in quantifying the error or information loss between original images and the images produced by varying spectral and spatial resolution and noise. These metrics take into account both spectral and spatial information as each pixel is treated as a spectral vector at a given spatial location. However, the spectral component of this metric is slightly compromised by the mere fact that degrading the image spectrally will result in fewer bands. That is, f(x, y) and f'(x, y) will have different spectral dimensions. The above calculations will not be done by bands but according to wavelength and therefore will not be a true representation of the degradation. However, these metrics could be used to quantify Level 3 products since they are able to quantify the error between any initial image and an output image. For instance, we can produce a series of fraction maps by an unmixing algorithm and then compare these maps to those produced by applying the same algorithm to a degraded version of the initial image cube. Similarly, an error metric can be calculated between an initial classification map and the classification map of a degraded image. Although this type of metric suits all of our criteria, we will see that there are even more meaningful metrics which are related to this one.

2.7.3   NIIRS, the General Image Quality Equation (GIQE) and Q

A discussion that involves image quality with respect to specific applications and task performance would not be complete without reference to the National Image Interpretability Rating Scale (NIIRS). NIIRS defines and measures image quality by rating an image on a 10 level scale that quantifies the image's interpretability (Fiete, 1999 and Leachtenauer, 1997). The scale was initially developed for the intelligence community but its usage has spilled over to the development of a Civil NIIRS Guide (www.fas.org/irp/imint/niirs_c/). NIIRS ratings describe the information that can be extracted from an image based on a pre-defined list of tasks. The tasks associated with each level use military, cultural, agricultural and natural cues in the image. An example of NIIRS scales can be found in Appendix B. A high NIIRS scale rating means that more information can be extracted from that image. For example, if we were able to distinguish between taxiways and runways in an image of an airport this would be defined as NIIRS 1. If we were able to identify aircraft wing configurations and the presence of aircraft servicing equipment this would indicate a NIIRS level of 5 – and so on.   It should be noted that a Multispectral NIIRS scale does exist (www.fas.org/irp/imint/niirs_c/). It rates interpretability based on both the spectral and spatial information within the acquired image. Some tasks associated with certain NIIRS levels are only associated with the spectral character of the image (see Appendix B). For example, some of the tasks include the detection of small boats (sub-pixel in size) on open water and the detection of recently installed minefields. Based on what was presented in previous sections, it is easy to see that such tasks are performed using spectral algorithms and are Level 3 products. It may prove useful to use these types of tasking definitions in this thesis study.

NIIRS proves to be a versatile rating system for image quality. It provides a method of communication amongst image analysts about the information potential of images and a new way to define remote sensing system requirements.   NIIRS, as an "industry standard" has proven useful in validating image-quality prediction models (Leachentauer, 1997).

One of these prediction models is the General Image Quality Equation or GIQE (Leachentauer, 1997). In brief, it predicts the NIIRS rating of an image based on the imaging systems operating parameters. The GIQE is a regression based model that accounts for target, sensor and processing characteristics of the system by including the terms of GSD (ground-sampled distance), SNR, RER (relative edge response - i.e. sharpness) and MTF (modulation transfer function).   It is not worth going into the details of this model since the GIQE model only

assesses the spatial information within an image. A GIQE model that incorporates spectral information has not yet been developed.

A more recent and similar prediction model is the calculation of Q which is "the ratio of the spatial sampling frequency to theoretical bandpass of an incoherent diffraction limited optical system" (Fiete, 1999). This ratio is denoted as:

$$Q = \frac{\lambda FN}{p}$$
(2-27)

where λ is the mean wavelength, FN is the system f/number (focal length divided by aperture) and p is the detector sampling pitch. The value of Q and hence image quality is "sensitive to the system design parameters of modulation transfer function, signal to noise ratio and ground sampled distance" (Fiete, 1999). Again this model deals only with the spatial information within an image. The value of λ is just the mean wavelength for a panchromatic system and this model has not been applied to multispectral or hyperspectral sensors. It is quite possible that a new value of Q be defined which is a weighted sum of the individual Q-values in each spectral bandpass of the hyperspectral sensor. This is just an idea at this point and the investigation of a new definition of Q is beyond the scope of this thesis.

The above discussion about NIIRS, the GIQE and Q models of image interpretability and information demonstrate that it is possible to define image utility in terms of sensor operating parameters. This allows us to vary the operating parameters of the sensor and then observe the effect of these changes on image derived information. Unfortunately, these metrics only account for the spatial information within the image and not the spectral information that Level 3 processing relies on. Therefore, they are of no use in this thesis study other than using NIIRS type tasks in algorithm employment

So far we have many useful metrics to quantify the actual image and its degradation as the spatial resolution, spectral resolution and noise of the image are varied. However, none of these metrics, except perhaps RMS error, properly assess the accuracy and effectiveness of hyperspectral processing algorithms under differing values of spectral and spatial resolution and noise. It is growing apparent that metrics which relate directly to target identification, spectral unmixing and thematic mapping are necessary.

2.7.4   Analysis of Variance (ANOVA) and Multivariate Analysis - KAPPA

<u>Analysis of Variance</u>

In the use of classification algorithms that produce thematic maps we require a method to assess the accuracy of the final Level 3-map product. In general, comparing the processed remotely sensed data to some type of reference is a measure of accuracy. This reference could be another image or ground truth data about the regions depicted in the image.   The metric becomes a set of agreements and disagreements between the classes in the reference and those determined by the classification algorithm used. The binomial probability density function (better yet a multinomial pdf) properly represents the exact number of successful classifications (Rosenfield, 1981; Wallpole, 1982).

Rosenfield (1981) used ANOVA in his study of the effects of changing the scale on classification accuracy. Although he found that scale significantly effected the results of thematic mapping, his use of ANOVA as an analysis metric was not well chosen. Admittedly, ANOVA allows for the study of measurements that depend on different factors operating simultaneously (Rosenfield, 1981), yet it is based on several assumptions. Firstly, it assumes that the data is normally distributed when it has already been established that classification results have a somewhat multinomial distribution. To overcome this problem the binomial results may be transformed into a normal distribution (Rosenfield, 1981). Another assumption is that the variances are homogeneous. Bartlett's Test (Johnson and Wichern, 1998) is used to test for the homogeneity of variance yet it is sensitive to data that is not normal – which essentially the results are prior to the normal transformation. The final assumption, that misses the mark, is that the classification errors with respect to the reference data (i.e. errors) are independent. This assumption is not very strong when using remotely sensed data as there may be confusion between categories due to similar spectral signatures, lack of spectral resolution or noise (Congalton et al., 1983). Evidently, the use of ANOVA is limited in the assessment of classification algorithms. However, in a later section the benefits of using ANOVA for analysing the results of a factorial designed experiments will become readily apparent.

<u>Discrete Multivariate Analysis - KAPPA</u>

Another method of accuracy measurement for thematic maps is needed. This leads us to the techniques associated with discrete multivariate analysis. This involves a complete

examination of the error matrix (Congalton et al., 1983, Congalton and Green, 1999). The error matrix, confusion matrix, or contingency table is a square array of numbers in which each cell contains the number of pixels assigned to a certain class by a classification method compared to how they were assigned by another method. Typically the classification method is compared for agreement to some reference as explained previously – this may be another image or ground truth. Table 2-1 shows a typical example of an error matrix from a classification algorithm.


```
                           Reference Data
                                              row
                      D     C     AG    SB    total

                 D    65    4     22    24    115          Land Cover Categories
                                                           ――――――――――――――
                 C    6     81    5     8     100          D = deciduous

  Classified    AG    0     11    85    19    115          C = conifer
  Data
                 SB   4     7     3     90    104          AG = agriculture

             column   75    103   115   141   434          SB = shrub
             total                                         OVERALL ACCURACY =
                                                           (65+81+85+90)/ 434 =
                                                           321/434 = 74%


        PRODUCER'S ACCURACY              USER'S ACCURACY
        ――――――――――――――                  ――――――――――――

        D  = 65/75  = 87%               D  = 65/115 =  57%
        C  = 81/103 = 79%               C  = 81/100 =  81%
        AG = 85/115 = 74%               AG = 85/115 =  74%
        SB = 90/141 = 64%               SB = 90/104 =  87%
```

Table 2-1: Example Confusion/Error matrix

We are able to calculate many useful metrics from a confusion matrix. The first metric is the overall accuracy, which is equivalent to the sum of the diagonal divided by the total number of sample pixels in the image. This measures the correctly classified samples and as we can see from the above table the overall accuracy equals 74%. We are also able to see errors of commission and omission. Classifying a sample into a certain category that it does not belong to entails making commission errors while omission errors are made by failing to include a sample in its proper category (Johnson and Wichern, 1998). Producer and User Accuracy can also be calculated from the error matrix and these accuracies represent individual category accuracies

(Congalton and Green, 1999). The following example taken from Congalton and Greene (1999) demonstrates the utility of these calculations.

In Table 2-1 we see that the overall accuracy of the classification map is 74%. The producer's accuracy is found by dividing the correctly classified number of samples in a specific class by its column total which is the total number of sample units as indicated by the reference data. Likewise, the user accuracy is found by dividing the number of correctly classified samples by its row total which represents the number of samples classified by the algorithm into that category. This results in a producer accuracy of 87% and a user accuracy of 57% for the deciduous tree category in Table 2-1. This means that "although 87% of the deciduous area was correctly identified as deciduous, only 57% of the areas called deciduous on the mapped image is actually deciduous on the ground" (Congalton and Green, 1999). These types of accuracy metrics are quite useful in assigning confidence to the thematic mapping performed by classification algorithms.

These accuracy metrics can be shown in equation form by representing the confusion matrix mathematically, as seen in Table 2-2 (Congalton and Green, 1999).

j = columns (reference)

		1	2	k	n_{i+} - row total
i = rows	1	n_{11}	n_{12}	n_{1k}	n_{1+}
(classification)	2	n_{21}	n_{22}	n_{2k}	n_{2+}
	k	n_{31}	n_{32}	n_{3k}	n_{3+}
n_{+j} - column total		n_{+1}	n_{+2}	n_{+k}	N

Table 2-2: Mathematical Representation of a Confusion Matrix

In this table we see that N samples are assigned into one of k classes. Let n_{ij} represent the number of samples classified into class i and class j of the reference data. Equation 2-28 is the number of samples classified into class i by the classification algorithm (n_{+i}) and equation 2-29 is the number of samples classified into class j of the reference data set (n_{+j}).

$$n_{i+} = \sum_{j=1}^{k} n_{ij} \qquad (2\text{-}28)$$

$$n_{+j} = \sum_{i=1}^{k} n_{ij} \qquad (2\text{-}29)$$

It is from these equations that overall, user and producer accuracies can be easily derived as seen in equations 2-30, 2-31 and 2-32 respectively.

$$overall \quad accuracy \quad = \frac{\sum_{i=1}^{k} n_{ii}}{N} \qquad (2\text{-}30)$$

$$user \quad accuracy \quad = \frac{n_{ii}}{n_{i+}} \qquad (2\text{-}31)$$

$$producer \quad accuracy \quad = \frac{n_{jj}}{n_{+j}} \qquad (2\text{-}32)$$

We can further take the number of samples occurring in a certain class and convert that to a proportion of samples in the i-jth cell. This leads to the following simple equations:

$$p_{ij} = \frac{n_{ij}}{N} \qquad (2\text{-}33)$$

$$p_{i+} = \sum_{j=1}^{k} p_{ij} \qquad (2\text{-}34)$$

$$p_{+j} = \sum_{i=1}^{k} p_{ij} \qquad (2\text{-}35)$$

The use of the above proportions can be used in a discrete multivariate analysis technique known as ***Kappa Analysis***, initially adopted by Congalton at al. (1983) in the assessment of the thematic mapping of remotely sensed data. This technique is a very useful accuracy metric since it can be used to determine if two error matrices are statistically different. This provides a substantial benefit in performing this thesis study and searching for significant differences between algorithm performance at various sensor-operating parameters.

In performing Kappa Analysis, a maximum likelihood estimate of kappa, $\kappa^/$ is determined. The value, $\kappa^/$, is a measure of agreement "based on the difference between actual agreement (error matrix diagonal) and chance agreement indicated by the row and column marginals" (Congalton and Green, 1999). Values of $\kappa^/$ range from 0 to +1 where the closer the number is to unity the stronger the agreement between the classified and reference data. To calculate $\kappa^/$ we first define the actual agreement in equation 2-36 and the chance agreement in equation 2-37. These values for agreement are then combined into the value $\kappa^/$ shown in equation 2-38 – also shown in its expanded form using the notation from the mathematical representation of the error matrix that was discussed previously.

$$p_o = \sum_{i=1}^{k} p_{ii} \tag{2-36}$$

$$p_c = \sum_{i=1}^{k} p_{i+} p_{+j} \tag{2-37}$$

$$\boldsymbol{k}' = \frac{p_o - p_c}{1 - p_c} = \frac{n \sum_{i=1}^{k} n_{ii} - \sum_{i=1}^{k} n_{i+} n_{+i}}{n^2 - \sum_{i=1}^{k} n_{i+} n_{+i}} \tag{2-38}$$

The variance for the estimate of $\kappa^/$ can also be determined from the data within the error matrix. The equations necessary to calculate VAR($\kappa^/$) can be found in Appendix C. It is with the use of these calculated statistics that a number of tests can be conducted with regards to the classification performance. Congalton and Green (1999) outline many of these tests in greater detail than presented in the subsequent paragraphs.

The first test that can be performed is to see how well the classified data agrees with the reference data. This test is performed by a quick examination of the value of κ' and where it falls on the scale of disagreement (0) to complete agreement (1). Using the fact that the κ' statistic is asymptotically normally distributed, confidence intervals for the κ' value can be generated by using the sample variance calculation (Congalton and Green, 1999). Confidence limits for the diagonal values within the error matrix can also be found. This is done by first computing the individual cell probabilities. These probabilities are then incorporated with the marginal proportions of the error matrix.

One can test the significance of the κ' statistic to determine if the agreement between the processed and reference data is better than random chance. The standardised and normally distributed test statistic for significance testing of a single confusion matrix is:

$$Z = \frac{\boldsymbol{k}'}{\sqrt{\text{var}(\boldsymbol{k}')}}$$

(2-39)

The testing follows standard statistical hypothesis testing (Wallpole, 1982) where the null hypothesis is $H_o: \kappa = 0$ and the alternate is $H_1: \kappa \neq 0$. The null hypothesis is rejected if $Z \geq Z_{\alpha/2}$ where $\alpha/2$ is the confidence level and the degrees of freedom are assumed to be infinity (Congalton and Green, 1999).

Similar to the above methods is the ability to statistically compare different confusion matrices. This provides us the ability to compare the performance of different classification algorithms with respect to the same reference/ground truth. It also provides the ability to "track" the accuracy of one algorithm being used under varying conditions of noise, spectral and spatial resolution and observe any significant difference in algorithm performance. The applicability of these kinds of tests to this thesis study is quite apparent. Given that one error matrix produces a kappa estimate of κ_1' and a second matrix is represented by κ_2' the test statistic becomes:

(2-40)

$$Z = \frac{\left|\boldsymbol{k}_1' - \boldsymbol{k}_2'\right|}{\sqrt{\text{var}(\boldsymbol{k}_1') + \text{var}(\boldsymbol{k}_2')}}$$

where the null hypothesis H_o: $(\kappa_1 - \kappa_2)=0$ is rejected if $Z \geq Z_{\alpha/2}$ and the alternate hypothesis of H_1: $(\kappa_1 - \kappa_2) \neq 0$ is accepted (Congalton et al., 1983).

Although the application of Kappa Analysis to assess classification performance proves extremely versatile, it is evidently limited to measuring classification accuracy only. It will only be able to be used as a metric with those Level 3 products associated with classification/thematic mapping. It is not a useful or meaningful metric for target detection or products that involve fractional end-member maps.

Sum of Squared Errors

There is great difficulty in obtaining an error matrix from Level 3 products that consist of fraction endmember maps. Therefore, the application of Kappa Analysis to algorithms such as spectral unmixing and orthogonal subspace projection algorithms is pointless and meaningless. Another statistical metric must be devised. Studies have been previously conducted that compare different unmixing techniques (Konno, 1999). This study by Konno (1999) quantitatively compared traditional unmixing to a stepwise unmixing technique. The RMS error between the measured spectra and the regressed spectra was used as a metric. Yet more importantly, he also used the Sum of Squared Error between the truth image and processed image (i.e. fraction map) as a metric to assess the performance of unmixing algorithms. Level 3 unmixing algorithms produce fraction maps listing the fraction/abundance of each endmember present in the mixed pixels. If truth of these fractions is known as a reference images, the error can be calculated as:

$$SSE = \frac{1}{N} \sum_{j=1}^{J} \sum_{n=1}^{N} \left[image_{jn} - reference_{jn} \right]^2 \tag{2-41}$$

where j=1…J represents the endmember and n=1,…N represents the pixels with N being the total number of pixels. In this way, both "image" and "reference" refer to the fraction of that endmember at a specific pixel. There is an apparent similarity between equation 2-41 and those metrics discussed in section 2.7.2. However, the use of the SSE is more meaningful with respect to measuring unmixing performance.

This metric is quite simple in its calculation but when you consider a large image with several endmembers this calculation would be intensive without computer assistance. Based on the past successful use of this metric in assessing unmixing algorithms, it will be very useful in

this thesis study to evaluate the performance of Level 3 unmixing algorithms that produce fractional endmember maps.

2.7.5   Signal Detection Metrics and ROC Curves

The above SSE metric could also be applied to measuring the performance of target detection algorithms, but the meaning of such a metric would be limited. Additionally, many target detection algorithms offer a RMS error image in conjunction with its regular Level 3 products (RSI, 1998). Target detection algorithms, such as spectral matched filtering (OSP) and spectral feature fitting (TRICORDER), are analogous to signal detection algorithms. In this manner we can adopt the methods and metrics used in signal detection theory to develop a metric for this thesis study. There are many excellent sources covering the theory and metrics of signal detection (DeFatta et al., 1988; Kay, 1998).   The following paragraphs summarize the theory and metrics as they pertain to this study. Signal detection metrics have been used in previous studies assessing detection performance with hyperspectral images and results from these studies will also be cited in the following paragraphs (Zavaljevski et al., 1996; Haskett and Sood, 1998; Tsang et al., 1998; Keller et al., 2000).

When attempting to identify a target spectrum amid background clutter there are several quantities that must be considered. These are the sample size, target size, the signal (i.e. the image), the false-alarm probability, the background density/distribution and the detection probabilities (Kassam, 1988).   Our goal is to detect items with a spectrum or signal of interest (example: army tank) that is mixed within background spectrum or noise (ex. trees and grass). The detection process can easily be looked at as statistical hypothesis test where a choice must be made between H_0 the noise/background and H_1, the signal plus noise/background (DeFatta et al., 1988, Walpole, 1982). The signal/spectrum of interest and noise/background can be modelled and viewed as probability density functions (PDFs) as seen in Figure 2-15.
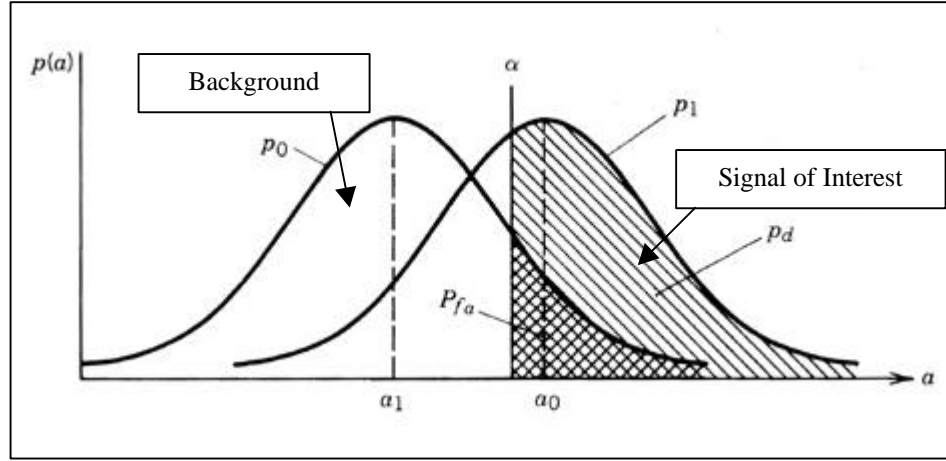
Figure 2-15: PDF distributions of signal of interest + background and background/noise
(DeFatta et al., 1988)

As we can see from Figure 2-15, the probability of false alarm, P_{fa}, is the area under the background/noise PDF (p_o) to the right of the threshold, α. This is represented by equation 2-42. The probability of detection, P_d, found by equation 2-43, is equivalent to the area to the right of the threshold α under the signal + background PDF (p_1). The notation accompanying these equations can be read as the probability of choosing H_X when H_Y is true. Obviously, $P(H_1, H_1)$ would then indicate detection and $P(H_1, H_0)$ is a false alarm. Equation 2-44 is another probability that is of interest, which is the probability of a missed target P_m. The equations below would become summations for the discrete case.

$$P_{fa} = \int_{a}^{\infty} P_0(x)dx = P(H_1, H_0)$$
(2-42)

$$P_{d} = \int_{a}^{\infty} P_1(x)dx = P(H_1, H_1)$$
(2-43)

$$P_{m} = \int_{-\infty}^{a} P_1(x)dx = P(H_0, H_1)$$
(2-44)

Evidently the errors introduced by false alarms and misses must be traded off against each other since they occupy the same space under the P_0 PDF curve. It is impossible to reduce both error probabilities simultaneously (Kay, 1998). Typically the probability of false alarms is constrained at some fixed value, γ (i.e. $P_{fa} = \gamma$) and at this point we wish to maximize the value of detection probability, P_d. In other words, we wish to maximize P_d subject to the constraint of P_{fa} $= \gamma$. This approach to signal detection is referred to the Neyman-Pearson method (DeFatta et al., 1988, Kay, 1998). This method, unlike the Bayes criterion method, does not rely upon costs associated with errors or *a priori* probabilities. The solution to this method is based on the ratio of the signal plus background/noise to just the noise PDF – called the likelihood ratio test and is seen here:

$$L(x) = \frac{P_1(x)}{P_0(x)} \geq \boldsymbol{b} \tag{2-45}$$

The value of β is a function of the threshold setting α which is dependent upon the probability of false alarms desired (DeFatta, 1988). The higher the value of β the more likely it is that proper detection of the signal was obtained.

Values for P_d and P_{fa} for differing threshold values, α, can be plotted to form a Receiver Operating Characteristic curve (ROC). The shape of the sample ROC curve in Figure 2-16 shows that as α increases in value both P_d and P_{fa} will decrease. The ROC curve should lie above the 45° line since a perfectly diagonal line is attained by a detection algorithm that bases its decisions on "the flip of a coin" (Kay, 1998). The ROC curve completely specifies detection performance (DeFatta, 1988). Furthermore, the presentation of this type of curve is far more informative than simple presenting error rates associated with various P_m and P_{fa} at various thresholds. This is because the ROC curve is independent of *a priori* probabilities and the costs of error – all of which will vary from case to case.
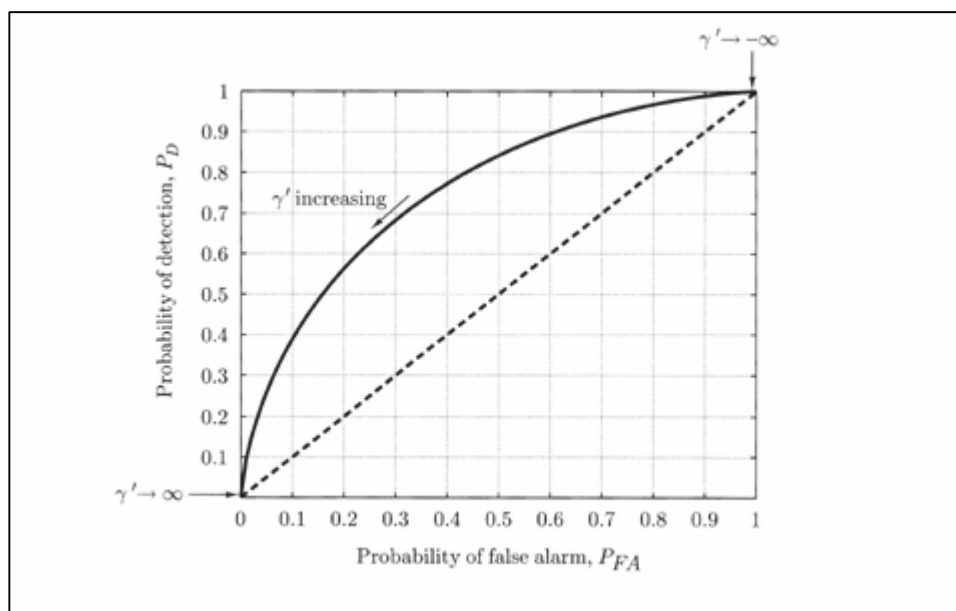
Figure 2-16: Sample ROC curve

Another way of obtaining this type of curve, with specific reference to remotely sensed data, is to plot the probability of detection (P_d) versus false alarm rate (Haskett and Sood, 1998). This approach requires prior knowledge about the scene to derive P_d. Again knowledge of the scene is required to compute the false alarm rate (FAR) which is expressed as the [# False alarms/km^2]. In this study (Haskett and Sood, 1998), the required PDFs where based on the spectral angle between target and background separated by some threshold.

In a study to assess a spectral identification algorithm used on hyperspectral imagery (Tsang et al., 1998) ROC curves were used in order to conduct performance comparisons. In this study detection is defined simply as a spatial pixel on a known target being identified correctly. Likewise, a false alarm is defined as a "highlighted" spatial pixel outside of the known target area and the probability of false alarms is defined on a per pixel basis (Tsang et al., 1998). For example, if the size of the image is 640 x 480 pixels, then one false alarm pixel equates to a probability of false alarm equal to3.25x10^{-6}.

Another sub-pixel target detection algorithm (Zavaljevski et al., 1996) used ROC curves to assess detection performance under varying image acquisition conditions. These conditions included different SNR levels, target sizes, target materials (i.e. spectral signature) and background materials. The results of their study can be summarized into four main points.

Firstly, detection performance improves with increasing SNR as seen in Figure 2-17. Secondly, detection also improves as the physical size of the target increases with respect to the GIFOV. Thirdly, detection characteristics become worse as the number of endmembers increase that are used in the prior processing of spectral unmixing. Finally, detection performance heavily depends on the spectral contrast between target and background. Figure 2-17 below shows the results of increasing SNR. This figure also demonstrates detection performance improvements are indicated by the ROC curve moving towards the upper left-hand corner of the chart – indicating a higher P_d at a lower P_{fa}.
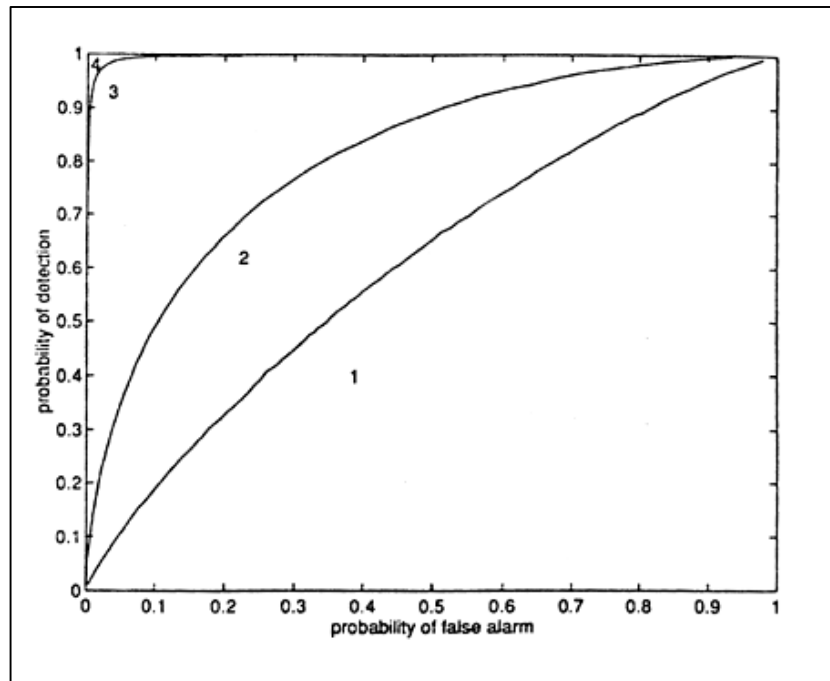
Figure 2-17: ROC curves for SNR of 10, 20, 30 and 40 dB (1 to 4 respectively) with the target size and background remaining constant (Zavaljevski et al., 1996)

It can be easily seen that the use of ROC curves, as a performance metric, is beneficial when assessing target detection algorithms. ROC curves provide an interpretable performance measure for each hyperspectral target detection algorithms tested in terms of varying spectral and spatial resolution and noise. This type of metric has already been proven in the study of hyperspectral processing algorithms under differing operating values of spectral information, target size, endmember selection and noise.

2.8   Experimental Strategy – Factorial Designed Experiment

It has been established that there is a naturally occurring interaction between the spatial resolution, spectral resolution and SNR of a sensor that is based on the law of conservation of energy. It has also been established that varying these parameters will effect the collected data and thus the extracted information from spectral algorithms (Level 3). A factorial designed approach to this experiment is necessary to properly study and analyze each of these main effects (i.e. parameters or factors) and their interaction on algorithm performance.   A one-factor-at-a-time approach does not allow an examination of factor interactions and is a statistically less efficient approach to experimentation (Montgomery, 1997). A fully detailed treatment regarding the designing of experiments is provided in Montgomery's text (1997). For the purposes of this thesis, a simplified description of factorial designed experiments is provided in the following paragraphs.

The **first** step in experimental design is to select an appropriate response variable. An exhaustive discussion of candidate metrics can be found in section 2.7. The **second** step is choosing what factors will be varied to test and observe changes in the response variable. In this case, the factors to be tested are spatial resolution, spectral resolution and noise. In a factorial designed experiment it is necessary to conduct L^k runs/trials of the experiment to investigate all possible combination of factors, where k is the number of factors to be tested and L is the levels that each factor will be tested at.   For example, if we wish to test 2 factors (A and B) at 2 differing levels of each factor (high and low), this experiment will require $2^2 = 4$ runs/trials to fully observe all treatment combinations (i.e. interactions) of the factors. It is only from this that we may draw statistical conclusions about the individual effects and their interactions. Obviously, as the number of factors and levels included in the test increases so does the number of required trials. Many times the number of runs may be reduced by a fractional factorial experiment (Montgomery, 1997). The **third** step is deciding whether the model for experimentation is to be a fixed effects model or a random effects model. Montgomery (1997) states that a random effects model has an infinite number of levels to each factor. In this way statistical conclusions can be made about the entire population from the levels tested. This is not possible with a fixed effects model since conclusions can only be made about those factor levels tested. This decision will ultimately determine the extent and type of statistical analysis, especially with regards to hypothesis testing after raw data collection. Figure 2-18 shows a geometric illustration of a 2^3 factorial designed experiment that tests three factors A, B and C at

high (+) and low (-) levels. The design matrix is the small table included in this figure and is a list of the necessary runs in conducting this hypothetical experiment.

Trial	Factor A	Factor B	Factor C
1	-	-	-
2	+	-	-
3	-	+	-
4	+	+	-
5	-	-	+
6	+	-	+
7	-	+	+
8	+	+	+
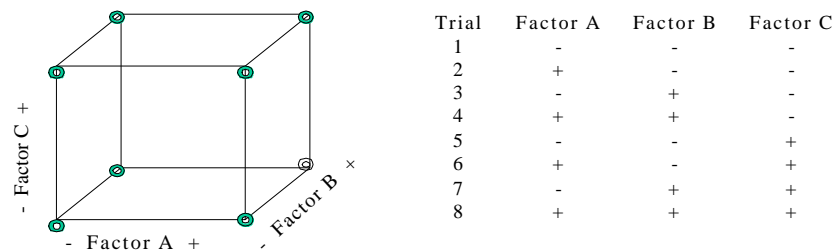
Figure2-18: A hypothetical 2^3 factorial experiment design

At this point the trials/runs of the experiment are ready to be conducted. Once the data is collected a statistical analysis of the data is required.   This typically involves statistical procedures involved in an analysis of variance (ANOVA).   This includes finding the Sum of Squares, Mean Squared values and degrees of freedom for each factor, interactions, and experimental error term.   For example, consider an experiment with two factors A and B. Factor A is tested at *a* different levels (fixed effects model), factor B is tested at *b* different levels and there are *n* replicates of the experiment. The required ANOVA table for this experiment is depicted in Table 2-3.

Source of Variation	Sum of Squares	DF	Mean Square	F_o
Factor A	SS(A)	a-1	MS(A) = SS(A) / (a-1)	F_o = MS(A) / MS(E)
Factor B	SS(B)	b-1	MS(B) = SS(B) / (b-1)	F_o = MS(B) / MS(E)
Interaction of A and B	SS(AB)	(a-1)(b-1)	MS(AB) = SS(AB) / (a-1)(b-1)	F_o = MS(AB) / MS(E)
Error	SS(E)	ab(n-1)	MS(E) = SS(E) / ab(n-1)	
Total	SS(T)	abn-1		

Table 2-3: ANOVA Table for a hypothetical two-factor experiment (Montgomery, 1997)

The calculated ANOVA values allow the experimenter to test the observed results of each factor for statistical significance via an "F-test." That is, the experimenter can see which factors and interactions make a significant impact on the outcome (response variable) of the experiment based on the magnitude and statistical hypothesis testing of F. Large values of F_o, calculated in the ANOVA table, indicate that the source of variation is significant. As alluded to previously, this can be extended into hypothesis testing where the ANOVA calculated value of F_o is compared to an F test statistic with a specified degree of significance (Montgomery, 1997). A more in-depth discussion regarding ANOVA and statistical testing can be found in any decent statistics textbook (Wallpole, 1982; Johnson and Wichern 1998; Montgomery, 1997). Exactly how this hypothesis testing will be applied to our results will be discussed in Chapter 4.

It can be clearly seen that the primary advantage to this type of experimental approach is that it allows an examination of both the main factors and their interactions with respect to a response variable or metric. This can also be done graphically. One simple method is to map the factors and how they varied with respect to the response variable. An example of this is seen in Figure 2-19. In this figure we see that part (X) shows no interaction between the factors while part (Y) shows interaction between factors A and B. That is, the level of the response is dependent upon both the levels of A and B such that a high response will be observed when A is at a high level and B is at a low level. A low response is observed when A and B are operating at the same low level. Intersecting lines indicate interaction.
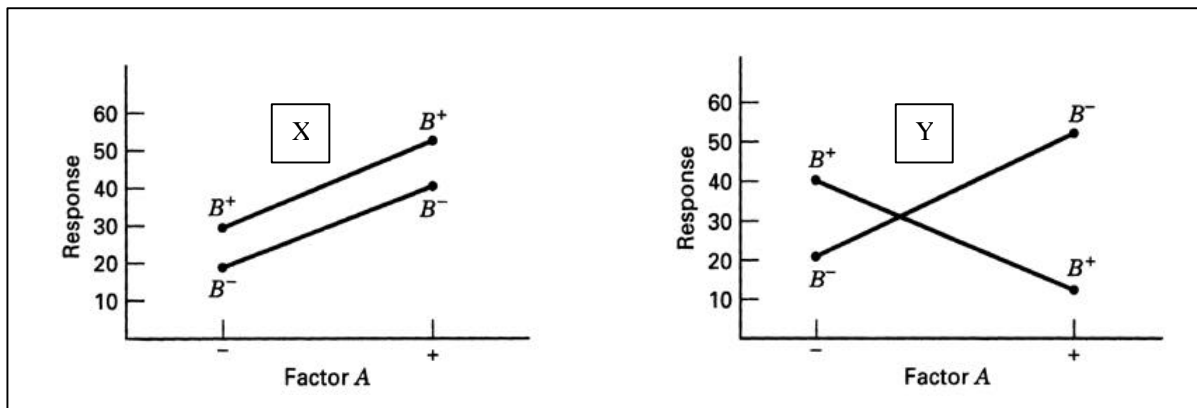
Figure 2-19: Interaction diagrams of two factors in an experiment

Another method to illustrate interaction, along with the effects of individual factors, is by developing an empirical model of the experimental results via regression analysis. Again information regarding linear regression can be easily found in many statistical textbooks (Wallpole, 1982; Johnson and Wichern 1998; Montgomery, 1997). After checking the adequacy of the developed regression model, a response surface and contour plot can be developed. This type of plot allows the experimenter to visualise the interactions of factors and the individual factor effects (Montgomery, 1997). Each of the various analysis methods is a means to the same end in that each method readily illustrates the significance of the main factors and their respective interactions.

Obviously, adopting a factorial designed experiment will help in attaining the main objective of this thesis. This experimental approach will allow an in-depth study of the accuracy and effectiveness of hyperspectral processing algorithms under differing values of spectral resolution, spatial resolution and noise. It will also allow us to effectively witness the joint effects of these remote sensing parameters with respect to spectral algorithm performance.

Chapter 3

Approach and Algorithm

3.1   Overview of the Approach

Prior to a description of the approach and algorithm used in this thesis study, it is useful to review our main objective.   The principal aim of this thesis study is to examine the accuracy and effectiveness of hyperspectral processing algorithms (Level 3 processing) under different image values of spectral resolution, spatial resolution and noise. This will involve the use of various scene types and different Level 3 processing techniques. The Level 3 algorithms that will be tested are Binary Encoding (BE), Spectral Angle Mapper (SAM), Gaussian Maximum Likelihood (GML), Linear Spectral Unmixing, Spectral Matched Filter (SMF) and Spectral Feature Fitting (SFF).   These algorithms were discussed extensively in Chapter 2 and again it should be noted that the algorithms tested are as they are implemented in ENVI (RSI, 1998). Also in Chapter 2, a discussion of many possible algorithm assessment metrics was also conducted. The metrics that will be used in this thesis are *Kappa* for classification algorithms, *Sum of Squared Error* for unmixing algorithms and *ROC curves* for target detection algorithms. By adopting a factorial designed experimental approach we are able to simultaneously analyze the main and joint effects of the remote sensing parameters of spectral resolution, spatial resolution and noise with respect to spectral algorithm performance.

It is appropriate to first provide a brief overview of the adopted approach and then describe each module in more detail.   The basic approach can be best represented schematically as a flowchart, as seen below in Figure 3-1.
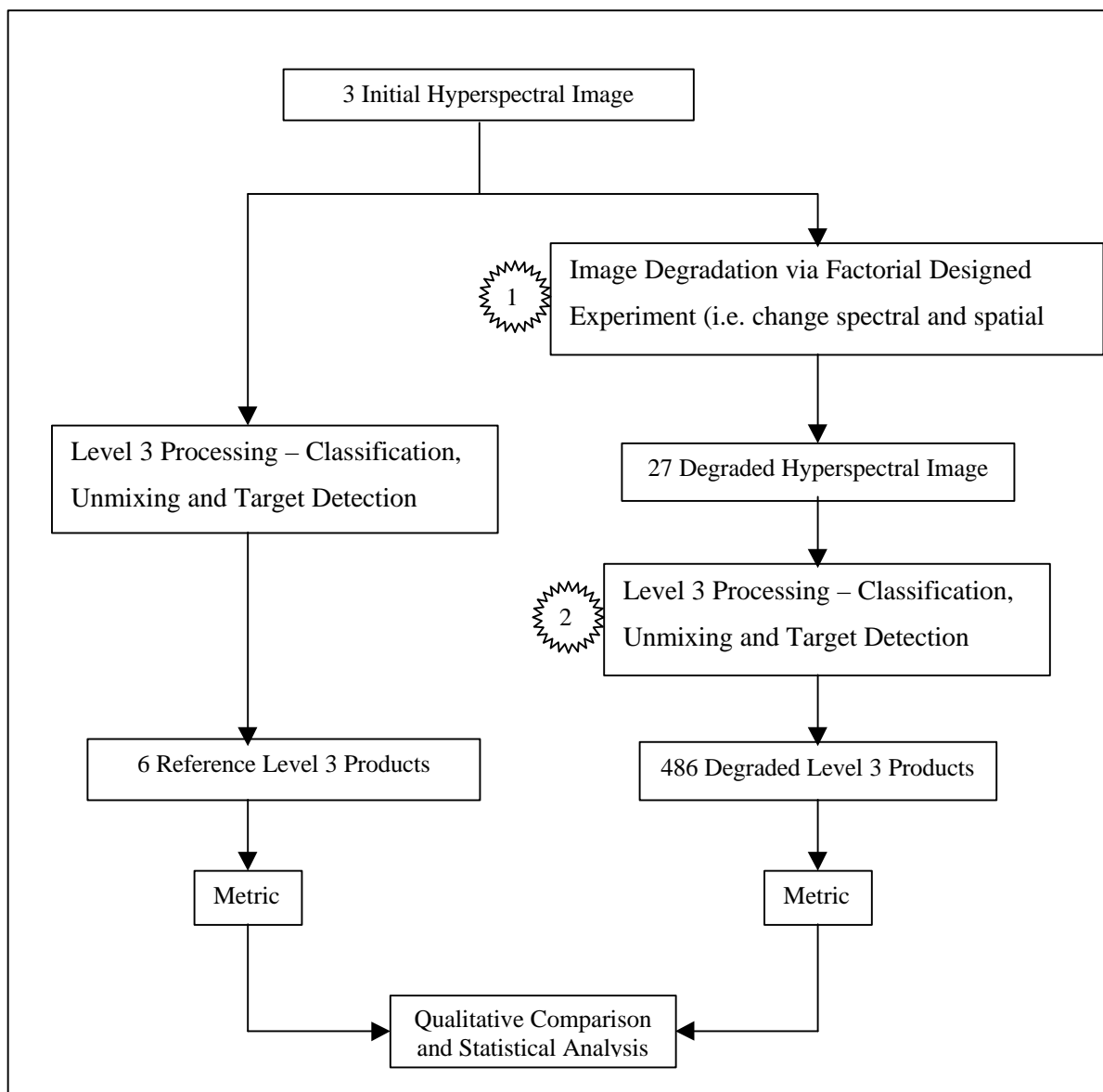
Figure 3-1: Flowchart of Approach to be used in studying effectiveness of Level 3
hyperspectral processing algorithms

As seen in Figure 3-1, we start with various initial hyperspectral images that are
characterized by high spatial and spectral resolution and low noise. The images used were
discussed in Chapter 2 and consist of two AVIRIS scenes (Rochester, NY and Rogers Dry Lake,
CA) and one DIRSIG scene (Western Rainbow). The real images will not have ground truth
associated with them. Ground truth will obviously be available for the synthetic imagery. More

details on establishing groundtruth and the selection of endmembers for inputs into Level 3 processing algorithms will be detailed in section 3.3. As seen on the left-hand side of Figure 3-1, the Level 3 processing results of the initial non-degraded image will serve as references for later comparisons.

The right-hand side of the flow chart follows a similar path with the addition of a few extra steps. The first extra step, labelled as *"star-1"*, is the degradation of the initial image by changing the spectral resolution, spatial resolution and adding spectrally correlated noise. This part of the overall approach will be discussed in section 3.2. The initial image will be degraded by following a factorial designed experiment outlined in Chapter 2 (section 2-8). Spectral resolution, spatial resolution and noise are tested at three different *levels* (high, medium and low). The exact numerical levels/parameters that each initial image will be degraded to is also discussed in section 3.2. However, it is worth mentioning here that since each of the three imaging parameters is tested at three levels the degrees of freedom associated with this experimental approach is equal to 26. The factorial designed experimental approach produces 3^3 or 27 degraded copies of the initial image. Therefore, processing these degraded images through three thematic mapping algorithms, two target detection algorithms and one unmixing algorithm results in a grand total of **486** Level 3 products for quantitative comparisons (27 degraded copies of the initial image x 3 types of image scenes x 6 algorithms).

The same Level 3 algorithms used on the initial image are used to process the degraded images produced by the factorial "degradation tool". This step is labelled as *"star-2"* in Figure 3-1 and will be discussed in section 3.3 since this step is very specific to the Level 3 processing utilized. The results from these degraded images are retained and labelled as "Degraded Level 3 Products." These results are compared to the reference Level 3 products and a metric is computed to indicate the errors introduced by degrading the image. It should be noted that each of the selected metrics – kappa, SSE and ROC curves - are specific to the spectral algorithm employed.

Now that a metric has been calculated for each "degraded" Level 3 product, we analyse the results by employing statistical techniques such as ANOVA and plot the results to visually establish trends. The analysis of the metrics will reveal the significance that the main effects and interactions of the spatial resolution, spectral resolution and noise on spectral algorithm performance. More details regarding the statistical analysis of the data are found in section 3.3. It is expected that the metric results and trends for each individual algorithm will be similar when compared between the different scenes used. Furthermore, it is from these results and their

analysis that we will also determine which algorithm performed best under the given levels of spectral resolution, spatial resolution and noise.

3.2   Image Degradation

The image degradation portion, labelled as *"star-1"* in Figure 3-1, requires further discussion. Image degradation is a separate and crucial first part of this entire thesis. As part of this thesis, programs were written using IDL (Interactive Data Language, RSI, 1998) and ENVI to degrade hyperspectral images with respect to spatial resolution, spectral resolution and the addition of spectrally correlated noise. These programs provide a user-friendly interface for image selection and parameter choices determining how the input image will be degraded. After selection of the initial image the user is prompted for a new GIFOV, a new number of bands to cover the spectral range of the initial image and a new SNR for the degraded image. The three programs and instructions for installation/use are available on the CD enclosed with this thesis. They are entitled *degrade_input.pro*, *degrade_doit_v16.pro* and *noise_covariance5.pro*. The program *degrade_input.pro* provides the user-friendly widgets for initial image and parameter selection. The program *degrade_doit_v16.pro* performs spatial degradation and spectral resampling. This program also calls upon *noise_covariance5.pro* to add spectrally correlated noise to the spatially and spectrally degraded image cube. Once these programs are installed, image degradation may be done through the regular ENVI main menu by selecting "Transforms > Image Cube Degradation."   The code is based on the following image degradation algorithm that is best represented as a flowchart (Figure 3-2).

The algorithm treats all input hyperspectral images as an image cube or 3-D array with the following dimensions – number of samples by number of lines by number of bands (*ns,nl,nb*). Obviously, the first input into this algorithm is the original image cube. The user is then prompted for the GIFOV, number of bands and SNR they wish to have this initial image cube degraded to. The specifics of how each one of these degradations is handled will be discussed in the following sub-sections. Whether the file is interleaved as Band-Interleaved-by-Pixel (BIP), Band-Interleaved-by-Line (BIL) or Band-Sequential (BSQ) format the program *degrade_doit_v16.pro* is able to read it into an image cube for further processing.   The final output of this program is a new image cube that is degraded to the user-specified parameters in the same interleaf format as the initial image file.

Image Cube Degradation Algorithm

<u>INPUTS</u>   <u>Processing</u>

| Original Image Cube | → | Read into a 3-D array |

User Inputs
GIFOV
Number Bands
SNR

Spatially Degrade by averaging multiple pixels ("shrink & expand")

Spectrally Degrade using critical sampling (I.e. FWHM = band spacing)

Noise Cube

Addition of Spectrally Correlated Noise

<u>OUTPUT</u>   **Degraded Image**
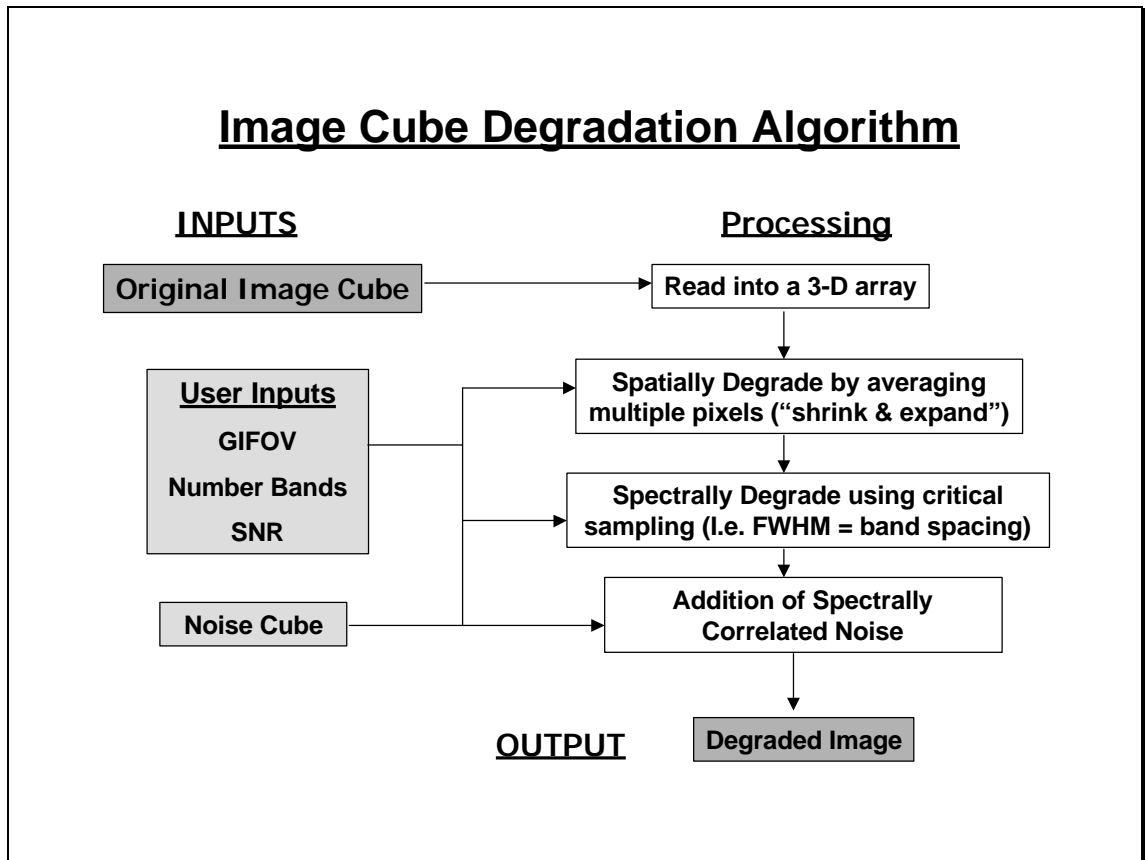
Figure 3-2: Flowchart of Image Cube Degradation Algorithm

3.2.1   Spatial Degradation

It should be noted again that in this thesis spatial resolution changes are synonymous with GIFOV changes. As it was discussed previously in Chapter 2, the correct way of approaching spatial degradation is via a convolution and resampling process. However, this process is computationally expensive and a slightly different means to the same end is available. In this respect, we used a "neighbourhood averaging" approach to spatial degradation. For example, if we change the GIFOV from 1m to 2m for an image that is initially 400x400 pixels, then the spatially degraded image will be 200x200 pixels given that the ground swath remains constant. The pixels in the degraded image are the aggregate of the initial pixels in a 2 x 2 manner. The initial image is "shrunk" to dimensions that are based upon the new-GIFOV and old-GIFOV given that the ground swath is constant. The new dimensions that the image is to be "shrunk" to are found via equation 2-18 in Chapter2. Although this "boxcar" approach to spatial degradation

is a very good approximation of the convolution and resampling method, it does have limitations. The first limitation is that the way this neighbourhood averaging approach is implemented in *degrade_doit_v16.pro* requires that the size of the degraded image be an integer multiple of the initial image size. That is, the user-specified degraded GIFOV must be an integer multiple of the original GIFOV in consideration with initial size. The second problem is that the degraded image is smaller than the original. This is what should physically happen when changing the GIFOV while keeping the ground swath constant. However, as we saw in Chapter 2, many of the metrics used for quantitative comparison are based on image size. In this regard, we require that, although the image is spatially degraded via a neighbourhood averaging process, that it also remain the same size.

To solve this problem, the "shrunken" image produced by neighbourhood averaging is treated as an intermediate product. This smaller image is then "expanded" back to its initial size using nearest-neighbour resampling. The result is a spatially degraded image that is the same size as the initial image and consists of "super-pixels" which are the aggregate of the initial image pixels based on the user-specified GIFOV. The net result of this entire process can be better seen in Figure 3-3.
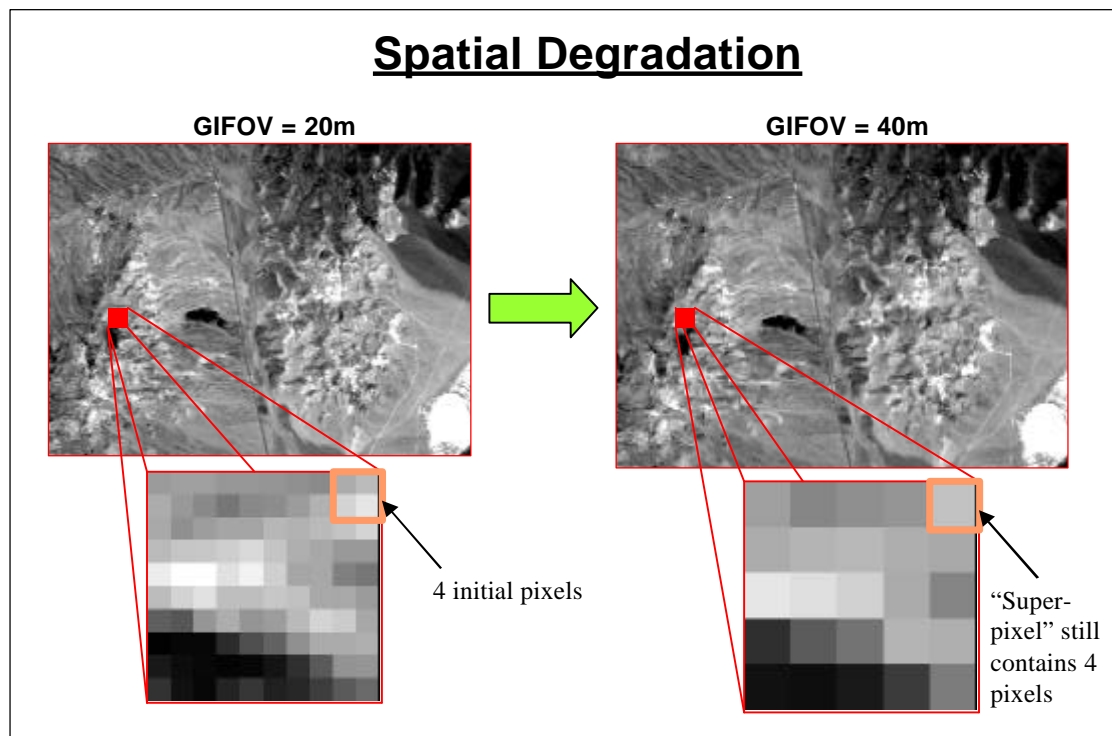
Figure 3-3: Net result of Spatial Degradation Approach

The image on the left is the initial image that is to be degraded from a GIFOV of 20m to 40m. This means that pixels will be averaged in neighbourhoods of 2 x 2 and the resulting average will be placed back into each pixel locations from which they came. Each of the orange boxes contains 4 pixels.

3.2.2   Spectral Resampling

Once the degradation program opens the image file, the initial number of bands is assigned to one variable and the list of wavelengths that these bands represent is stored as a vector. The user is asked for the number of bands that they wish to have the image degraded to. This user-defined number of spectral bands, *degrade_nbspect*, determines the spectral resampling of the image. The dimension of the initial vector of wavelengths is obviously equal to the initial number of bands. This initial vector of wavelengths is resampled via linear interpolation to match the dimensions of *degrade_nbspect*. Special attention is made so that the first and last wavelengths remain same. In this way, the true spectral range of the image does not change and it is just the wavelength values within this range that are interpolated. The net result of this process is a new vector of wavelengths that have dimensions equal to the number of "degraded" spectral bands specified by the user. This new vector of wavelengths represents only the spectral band centres. The ENVI spectral resampling function assumes critical spectral resampling when FWHM values are not provided. It uses a Gaussian model with a FWHM equivalent to the band spacing to perform spectral resampling (RSI, 1998).

The spectral resolution, in nm or μm, of the degraded image can be quickly determined by simple arithmetic. For instance, an initial image covers the spectral range 400 – 2500 nm with 210 spectral bands/channels. The spectral resolution of this initial image is approximately 10 nm (calculation is as follows [2500 – 400] / 210). We wish to degrade this image so that 75 spectral bands represent the spectral range. Following the same simple arithmetic, the spectral resolution of the spectrally degraded image is 28 nm.   Apparent improvements to this part of the program are discussed in Chapter 5.

3.2.3   Addition of Spectrally Correlated Noise

Two main methods were fabricated and tested in an effort to add spectrally correlated noise to a hyperspectral image. The first method, called the *"dark current image approach,"* calculated the covariance matrix of the dark current noise files from AVIRIS flight data. The

eigenvalues of this covariance matrix were determined and scaled to reflect a user-defined SNR for the degraded image. The result is the noise variances for each band in de-correlated space. Subsequently, 2-D arrays of random numbers were created for each band based on a mean of zero and a standard deviation equal to the square root of its respective eigenvalues/variance. A series of these 2-D arrays stacked together form a 3-D cube of de-correlated noise with a size equal to the spatial dimensions of the image and the number of degraded spectral bands. A Principal Components transform was performed on this de-correlated noise cube so that the result of the transform is a correlated noise cube. Although this method did work in producing noise, the covariance and correlation matrices of the noise never matched that presented in Boardman's work (1995), as they theoretically should. It is believed that one of the problems with this approach is that a scaling or conversion factor is missing in the overall calculation. Despite a great length of time and frustration this mystery was never fully solved. A further discussion regarding this approach can be read in Appendix D. The programs needed for this approach are also discussed in this appendix and are included on the accompanying CD.

The second method, which used the difference between a noisy image and a noiseless image, was used to construct a spectrally correlated noise cube. It is best to first represent this approach schematically (Figure 3-4) and then discuss its implementation.
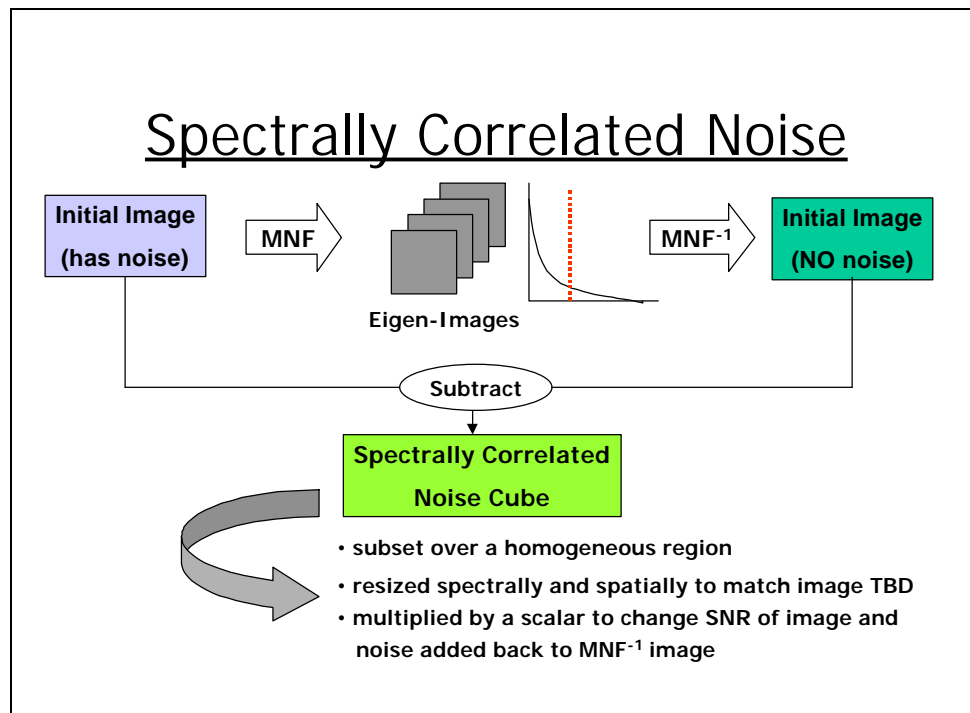
Figure 3-4: Approach used in Constructing Spectrally Correlated Noise

It is understood that through the process of image acquisition a final image contains both signal and noise. In Chapter 2 (section 2.6.3) we were introduced to the MNF transform. After performing an MNF transform on an image in ENVI, the MNF bands are ordered by descending order of information. The first band contains the most information and the last band contains the most noise. In this sense the image in MNF-space can be divided into two parts. The first portion consists of those MNF-bands that contain information and the second portion includes those MNF-bands that are dominated by noise. As it was stated in Chapter 2, noise can be removed from the image data by performing an inverse MNF transform using only the bands that contain information (Green et al, 1988; RSI, 1998) and ignoring the MNF-bands characterised by noise. This is precisely the starting point of this noise approach.

First, an MNF transform was performed on the AVIRIS Rochester image (216 spectral bands, including atmospheric bands, in units of radiance – μwatts/cm^2/nm/sr). A spatial subset of the image over the lake was used to estimate the noise statistics. This noise estimation works under the assumption that each pixel contains both signal and noise, and that adjacent pixels in a homogeneous region (like the deep part of the lake) will contain the same signal, but different noise (RSI, 1998). Once the MNF transform has been performed, the resulting eigen-images are inspected for information and noise content. This inspection is done in conjunction with a look at the percentage of cumulative variance explained by the calculated eigenvalues. It was found that 50 MNF-bands explain ~97% of the cumulative variance or image information. This is confirmed by visual inspection of the eigen-images in which it was found that bands 51 to 216 were predominantly noise. As seen in Figure 3-4, an inverse MNF transform is applied only on the MNF bands containing information – in this case MNF-bands 1 to 50. This theoretically results in an image with little noise. As seen in Figure 3-4, the difference between the initial image with noise and the noiseless MNF transformed image is the noise cube. A 100x100 pixel spatial subset of this resulting noise cube is taken over the water region. This region was selected because it is spatially homogeneous and contains low signal. This small noise cube is then resampled spectrally to match the spectral dimension of the degraded image. Then the small noise cube is "mirrored" to make multiple copies of itself to fill the spatial dimensions of the degraded image that it will soon be added to. This mirroring approach was selected to minimise the spatial effects introduced by mere copying or tiling. This spectrally resampled and spatially resized noise cube is then multiplied by a scalar value so it will reflect the user-specified SNR when added to the degraded image. Finally, the noise cube is made to match the spatial

degradation of the image via the same neighbourhood averaging process that was applied to the image. Now, the noise cube can be added to the image. The statistics of the resulting 100 x 100 noise cube are presented graphically in Figure 3-5.
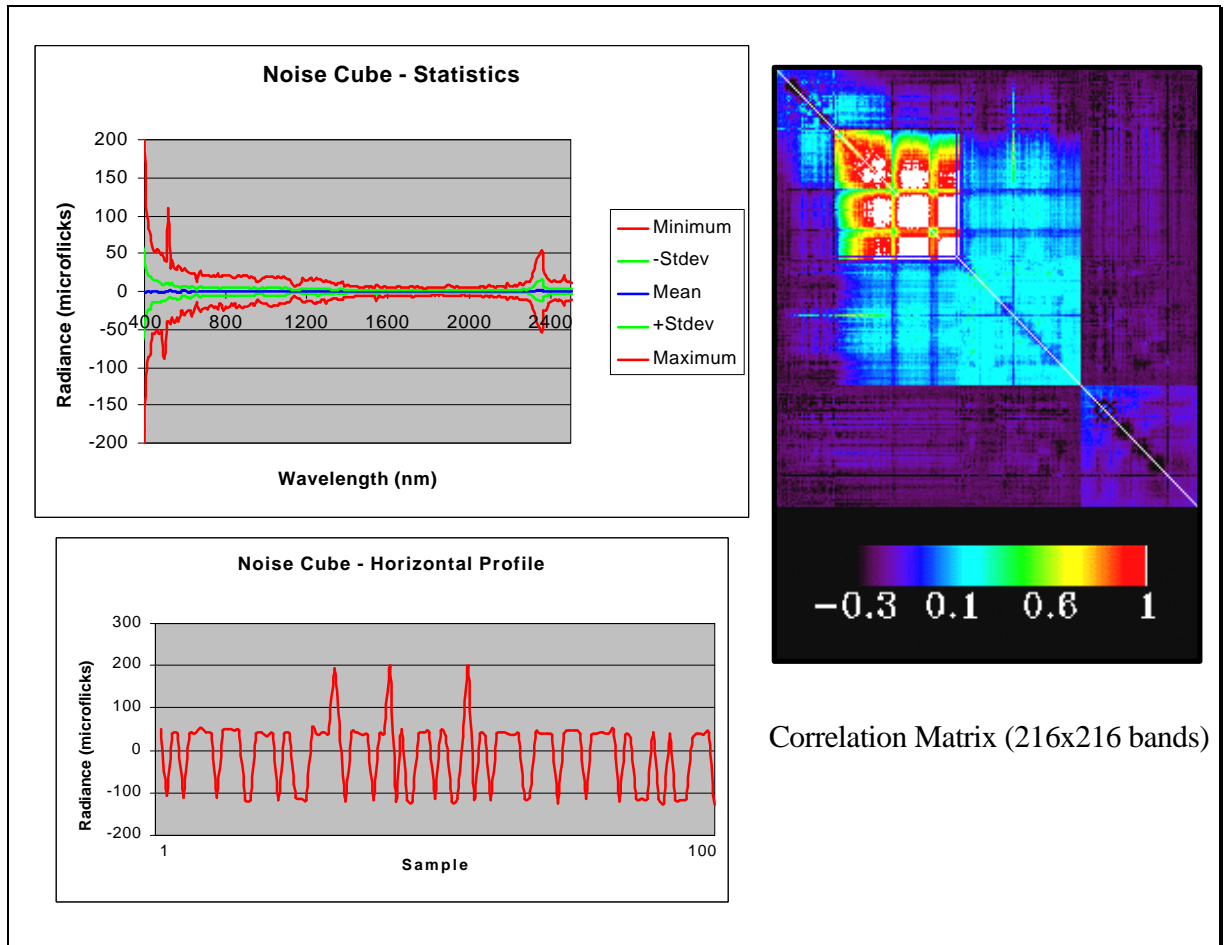
Figure 3-5: Statistics of Noise Cube used in Image Degradation

It can be easily seen from Figure 3-5, that we have attained the desired results for adding spectrally correlated noise to hyperspectral images. The statistics reveal that the noise cube has a desired mean of zero and there is definite correlation between bands. This correlation structure is due to the inherent redundancy of both information and noise characteristic of fine spectral resolution sensors. Close inspection of the corrleation matrix shows the separate spectrometers in AVIRIS. It is believed that this approach, although different from the dark current approach, best represents the noise of the entire image chain since the noise was computed directly from an

image. An example of this can be seen in Figure 3-5 where we see a periodic distribution in the horizontal profile of the noise. This profile chart only represents the first spectral band of the noise cube, yet it is interesting that a similar behaviour is repeated in all bands and is along the "sampling" direction of the acquired image.

It should be noted that the noise was added to all images while in units of radianace and not after the images may have been calibrated to reflectance. This approach best represents how noise is introduced along the image chain. Furthermore, noise added to degrade the SNR was done with reference to a constant signal. That is, all SNR values selected for degradation are with respect to a 30% reflector for the AVIRIS images and a 36% reflector for the DIRSIG image. The 30% reflector in the AVIRIS Rochester scene is the beach and the same levels of SNR degradtion were carried over to the AVIRIS Rogers Dry Lake scene. The 36% reflector in the DIRSIG scene is a ground panel of known reflectance. More information regarding the selection of these SNR levels and ensuring their consistency given that we are also changing the spatial and spectral resolution can be found in Appendix D.

3.2.4   Levels of Degradation

The initial images used to produce any degraded images in this thesis were the noiseless versions made by the MNF transform and removal of noisy bands depicted in Figure 3-4. In this respect, we always start with an image characterised by high spatial and spectral resolution and low noise prior to any degradation. The levels chosen for degradation in this factorial designed experiment are seen in Table 3-1.

AVIRIS IMAGES - Rochester, NY & Rogers Dry Lake, CA			
Spatial Resolution (m)	20	40	80
Spectral Resolution (nm)	10	55	110
SNR @ 30% reflector	225	100	10

DIRSIG IMAGE - Western Rainbow desert scene			
Spatial Resolution (m)	2	4	8
Spectral Resolution (nm)	10	55	113
SNR @ 36% reflector	200 (pure)	100	10

Table 3-1: Levels of Degradation for spatial resolution, spectral resolution and noise

The level of error asscociated with each one of these degradation parameters should be noted. The error associated with changes of spatial resolution is nil. Spectral resolution error levels of $\pm$ 0.50 nm are to be expected. The error associated with SNR levels is approximately $\pm$ 5.0 for SNRs of 225, 200 and 100. The error for SNRs equal to 10 is $\pm$ 1.0. An explanation of the casue of these boundaries upon the SNR can be seen in Appendix D.

3.3   Level 3 Products and Image Utility Metrics

We now have a means of degrading hyperspectral images spatially, spectrally and through the addition of spectrally correlated noise.   The next steps of this experimental approach, as seen in Figure 3-1, are to process the images through algorithms resulting in "degraded" Level 3 products and analyze these in comparison to the reference products. To accomplish these next steps three items needed to be accomplished. These were endmember selection for input into each algorithm, establishing groundtruth or reference products and the quantitative assessment of the collected data. It is these three tasks that will be discussed in the following sub-sections.

3.3.1   Endmember Selection and Algorithm Batch Codes

It was apparent from the discussions in Chapter 2 that each algorithm whether it be for classification, unmixing or target detection, requires input in the form of endmembers. Endmembers are the material constituents (e.g. grass, trees, rocks and water) that make up a scene and are often referred to as "classes".   The spectral signatures of all the endmembers making up a scene are typically stored as a spectral library.

Endmembers were derived from the AVIRIS images using *in-scene* techniques as described in Chapter 2 (section 2.5).   This worked quite well for the Rochester, NY scene as we have working knowledge of that area. We could readily identify known regions and cross-reference the AVIRIS image to a high-resolution aerial photograph of the same land area. Enough regions of interest (ROI) were selected from the image to best describe its spectral content. The mean of the numerous pixels within each of these ROIs was calculated and used to form a spectral library for the image. Furthermore, the PPI algorithm and n-D Visualizer tool available in ENVI (previously described in section 2.5) was used in conjunction with our in-scene technique.   Using these two ENVI tools we were able to find the most "spectrally pure" pixels in a hyperspectral image. Theoretically, these pixels correspond to scene endmembers. The resulting spectrum from these tools was compared to the spectral library formed by the in-scene

techniques. In general, the two methods corresponded very well and the necessary amendments were made to the final spectral library for any intuitive differences found. This same approach was also applied to the AVIRIS Rogers Dry Lake image. Even though we lack knowledge of the area, the in-scene technique used in conjunction with the PPI algorithm produced admirable results for endmember selection. The only problem is that we may not know the exact name of these endmembers/classes. A list of the endmembers for the Rochester, NY and Rogers Dry Lake, CA scenes is available in Table 3-2. The spectra of each of these endmembers can be viewed using ENVI and the spectral library data is included on the enclosed CD.

AVIRIS – Rochester, NY				AVIRIS - Rogers Dry Lake, CA		
Endmember	**Applied to….**	**% OF SCENE**		**Endmember**	**Applied to….**	**% OF SCENE**
trees	S,BE,GML,U	17.2253		highway-1	S, BE, GML, U	0.9304
marsh-1	S,BE,GML,TD	1.0320		urban-1	S, BE, GML, TD	0.0256
marsh-2	S,BE,GML,U	2.0740		urban-2	S, BE, GML, U	2.0840
beach	S,BE,GML	0.5507		playa-1	S, BE, GML, U	14.1324
field-1	S,BE,GML	1.4787		playa-2	S, BE, GML, U	33.9892
field-2	S,BE,GML,U	4.9727		playa-3	S, BE, GML, U	12.8744
grassland	S,BE,GML,U	5.6020		mineral 1-1	S, BE, GML, U	1.3516
deep water	S,BE,GML,U	34.3673		mineral 1-2	S, BE, GML	0.9080
shallow water	S,BE,GMLTD	0.6280		mineral 1-3	S, BE	0.0372
Genesee water	S,BE,GML,U	1.2527		mineral 2-1	S, BE, GML, U	32.0024
bay water	S,BE,GML,U	8.5007		mineral 2-2	S, BE, GML, TD	0.2112
urban-1	S,BE	0.0280		field-1	S, BE, GML, TD	0.2868
urban-2	S,BE,GML	0.0927		field-2	S, BE, GML, U	1.0416
urban-3	S,BE,GML,U	9.1327				
urban-4	S,BE,GML,TD	1.2993				

Table 3-2: Endmembers for AVIRIS Rochester, NY and Rogers Dry Lake, CA images and the Algorithms to which they were applied [Key is: S = SAM, BE = Binary Encoding, GML = Gaussian Maximum Likelihood, U = Linear Spectral Unmixing, TD = Target Detection (both Spectral Matched Filter and Spectral Feature Fitting]. Percentage of scene is from SAM results.

Also included in Table 3-2 is a list of the algorithms to which the endmembers were applied. As we can see, the classification algorithms used all but one of the endmembers. The reason for class exclusion when applied to GML classification will be discussed in Chapter 4. The input spectra for unmixing were the first 8 endmembers that explained the majority of the image. The number eight was chosen based on the results of Konno (1999) in which he found linear spectral unmixing performed best when 6 to 8 endmembers where used as input. Target

detection algorithms were applied on 3 different endmembers for each scene. It deserves to be noted that typically these targets/endmembers were spectrally distinct, were interesting targets and comprised only a small portion of the scene in order to minimise false alarms. The percentage of the scene explained by each endmember is the post classification statistics of the SAM algorithm on the initial non-degraded images. As we will see in section 3.3.2 the results of the SAM algorithm were used as a benchmark for the AVIRIS scenes.

The selection of endmembers for the DIRSIG scene was much simpler because a spectral library is produced by DIRSIG once the scene is rendered. The only amendments made to this spectral library were that the various spectra for "vehicle 1" were combined to form one spectral signature. Additionally, when using the GML classification algorithm all of the ground panels were combined into one class as were the two types of deciduous trees. The reason behind changing the spectral library for GML classification will be explained in Chapter 4. A list of the endmembers used for the DIRSIG Western Rainbow scene is available in Table 3-3 and the spectral library is found on the enclosed CD. The reasons behind selecting certain endmembers for certain algorithms follow the same logic that was presented for the AVIRIS images.

DIRSIG - Western Rainbow Desert Scene					
Endmember	Applied to….	% OF SCENE	Endmember	Applied to….	% OF SCENE
desert wash	S, BE, GML, U	57.5675	rusty tan	S, BE	0.0075
dirt road	S, BE, GML, U	1.1469	vehicle 2	S, BE	0.0088
desert pavement	S, BE, GML, U	37.4613	vehicle 3	S, BE, TD	0.0169
deciduous	S, BE, U	0.4238	black wood	S, BE	0.0050
deciduous - 2	S, BE, U	2.3187	bare wood	S, BE	0.0088
target 1	S, BE, GML, U	0.3187	2% panel	S, BE	0.0306
target 2	S, BE, GML, U	0.1613	4% panel	S, BE	0.0306
target 3	S, BE	0.0325	12% panel	S, BE	0.0306
target 4	S, BE	0.0444	24% panel	S, BE	0.0350
target 5	S, BE	0.0425	36% panel	S, BE	0.0350
desert bush	S, BE, TD	0.0150	48% panel	S, BE	0.0350
vehicle 1	S, BE, GML, U, TD	0.1269	60% panel	S, BE	0.0306
rubber tire	S, BE	0.0056	white card	S, BE	0.0306
aluminium	S, BE	0.0231			

Table 3-3: Endmembers for DIRSIG Western Rainbow Desert image and the Algorithms to which they were applied [Key is the same as table 3-2]

ENVI allows endmembers to be input into its spectral algorithms in the form of ASCII files, spectral libraries, statistics files, or from ROI means. In order to apply consistency in the

experiment, the same approach was used in providing spectral inputs to all the algorithms except GML. The method employed resamples the initial high-resolution spectral libraries to match the spectral resolution of the degraded images outlined in Table 3-1. These newly resampled spectral libraries are made into ASCII files containing only the spectra needed for a specific algorithm. This follows the "Applied to…" column of Tables 3-2 and 3-3. For instance, there will be three spectral libraries for performing unmixing with the DIRSIG scene. Each of these three spectral libraries will be at a different spectral resolution yet contain the exact same materials – those indicated with a "U" in Table 3-3. This method of spectral library input enables us to use ENVI batch codes. These programs were created as part of this thesis and are available on the enclosed CD. The use of batch codes in this thesis allowed us to classify, unmix or target detect all of the degraded images at once and calculate our required metrics after being supplied with ASCII spectral information.

As indicated above, the input into the GML algorithm did not follow the standard method described above. Instead, the ROI file used to form the initial spectral library was overlaid on each degraded image prior to executing the GML algorithm. The mean, standard deviation and covariance matrix was calculated from these ROIs in the degraded imagery. This is typically how an operator performs GML – by selection of ROIs in the scene. Experimental consistency is maintained because the ROI files used remain the same for all degraded images. Additionally, the ROIs used are the original ROIs used to form the spectral library for the first method using spectral libraries described above.

Radiance or Reflectance?

As mentioned above, an in-scene technique was used to construct spectral libraries for the Rochester and Rogers Dry Lake scenes. The spectra of these two images is in radiance space [units of micro-flicks – μwatts/cm^2/nm/sr] with atmospheric absorption bands still intact. Atmospheric inversion was not performed on these two scenes primarily because we lack sufficient groundtruth to build a spectral library of the true reflectance values or perform an ELM inversion. Other methods of inversion could have been performed yet it is strongly believed that the trends we notice in spectral algorithm performance will be the same whether the image is in reflectance or radiance space; especially since we are using scene derived endmembers.

The DIRSIG Western rainbow scene was atmospherically corrected using ELM. Since the spectral library provided by DIRSIG is in units of reflectance this step was mandatory. Even

though scene-derived endmembers could have been used, like the AVIRIS scenes, the fact that true reflectance curves were available allowed us the opportunity to better investigate the entire image chain. The images were first degraded spatially, spectrally and by adding correlated noise while in radiance space prior to ELM inversion. This was done to best represent the image chain. A batch code (available on the CD) was used to perform ELM atmospheric correction for all the degraded images using path radiance and solar irradiance values spectrally resampled to match the spectrally degraded images they would be applied to. This necessary ELM data was acquired at the highest spatial resolution of the image without noise addition so as not to introduce artifacts due to noise or spatial resolution changes into the ELM. Atmospheric bands, for the most part, have been removed from these images.

3.3.2   Establishing Groundtruth and Reference Products

The goal of this thesis is to assess the utility of spectral products derived from degraded imagery. In this regard it is imperative that benchmarks be established so that quantitative comparisons can be made. We require groundtruth. It has been established that there is insufficient ground truth for both AVIRIS scenes. To fill the void, the SAM results of the initial AVIRIS images (no degradation) will serve as "groundtruth" and reference for kappa calculations for SAM, BE, GML and ROC curve determination for SMF and SFF. A fraction map of the initial image (no degradation) will be made as the reference for SSE calculations of the unmixing algorithm. The reason SAM results were chosen as a benchmark/reference is that the SAM result of the Rochester scene was the most accurate given our knowledge of the area and when cross-referenced to the high-resolution aerial photograph. The SAM result for the initial Rogers Dry Lake scene was also picked as groundtruth/reference to maintain experimental consistency.

One advantage of DIRSIG is that we are provided with 100% complete groundtruth. This makes the selection of references for the Western Rainbow scene very simple. In this regard, the material map produced by DIRSIG was used as groundtruth for metric calculations for all algorithms. As a cross-reference, the DIRSIG produced material map was compared to classification results produced by applying the SAM algorithm to the initial Western Rainbow scene. The results between the material map and the SAM results were an 86% match, thereby indicating that the choice of using SAM results as ground truth for the AVIRIS imagery was a sound decision. It is fully realized that better ground truth could have been used given more

resources and the employment of more complex techniques, but the ground truth we established will assist in this thesis proof of concept.

3.3.3 Analysis Plan for the Collected Data

At this point we have constructed spectral libraries, degraded all the images to the parameters dictated by our factorial designed experiment and processed these degraded images using various classification, linear unmixing and target detection algorithms. The benchmarks established above, in section 3.3.2, will serve as references for the calculation of the metrics – kappa, sum of squared error and ROC curve determination. These calculated metrics are a measure of how the information has changed with respect to established groundtruth. A second metric will be calculated for each algorithm to measure how the information changes with respect to the algorithm itself. For example, a binary encoding algorithm is performed on a series of degraded images and a value of kappa is calculated for each BE result with respect to the SAM reference – the established "groundtruth". A second value of kappa is also calculated for each BE result with respect to a BE reference (no degradation) to see how the results of the algorithm change with respect to itself.

The calculation of these metrics still require statistical analysis in order to demonstrate any trends and show the significance of main and joint effects. A statistical analysis plan has been developed to tackle the data.

<u>Basic Analysis</u>

Firstly, the mean and standard deviation of all the metric values are determined with respect to each main effect. That is, a collective mean of the metric will be calculated for each level of spatial degradation, spectral degradation and SNR level – a total of 9 means and standard deviations. For example, consider one image and its associated series of degraded images. A mean for all values of kappa that are associated with a spatial resolution of 20 m, regardless of the other parameters of spectral resolution or noise, is done. These means were plotted to help visualise the data and establish any trends in algorithm performance resulting from a change of one of the main factors. To statistically demonstrate the level of significance of each main factor and their interactions an analysis of variance was performed on the metric values.

Classification

In addition to the basic analysis, the variance and Z-score for each kappa value and its respective confusion matrix was calculated (as per equation 2-39 and Appendix C). This allows us to test the significance of the κ' statistic and determine if the agreement between the processed and reference data is better than random chance. Furthermore, this provided us the required inputs to statistically compare different confusion matrices as per equation 2-40. This gives us the ability to compare the performance of different classification algorithms with respect to the same reference/ground truth and with respect to itself.

Unmixing

Only the basic analysis was conducted with the Sum-of-Squared-Errors calculated from Linear Spectral Unmixing.

Target Detection

Again, just the basic analysis was conducted with the results from the ROC curves determined from the Spectral Matched Filter and Spectral Feature Fitting algorithms. However, it is worth mentioning some of the specifics of how these metrics were calculated. As seen in Tables 3-2 and 3-3, three endmembers for each scene are used as input targets. Each one of these targets will have an individual ROC curve. A probability of detecting (P_d) these targets at a fixed probability of false alarm (P_{fa}) is read from the curve, using interpolation if needed. A weighted average of the three P_d at a common P_{fa} was calculated and used as the final metric. The weights of this calculation are the number of pixels that the target is comprised of in the groundtruth/ reference image. For example, a SMF is used to find targets A, B and C in an image. The groundtruth for this image reveals that target A is comprised of x number of pixels, target B is made up of y pixels and target C is comprised of z pixels. After performing the SMF algorithm and reading the three P_d from the ROC curves at a fixed P_{fa} we determine the weighted average of the detection results. This is the final metric that is reported and is found by:

$$PD = (Ax + By + Cz) / (x + y + z) \tag{3-1}$$

This approach was adopted since it is a better representation of target detection performance than if just merely selecting one target per scene.

The calculations of the selected metrics (kappa, SSE and ROC curve) have been incorporated into the created batch codes. In some cases, it may be necessary that extra processing be involved to extract our needed metrics. These include programs that perform interpolation on the ROC curves, calculate the variance of kappa (as per Appendix C) and statistically compare two values of kappa at various confidence limits. These extra programs/tools are also available on the enclosed CD and their use is detailed in the "*extraprograms_readme.txt*" file.

Chapter 4

Results and Observations

4.1   Overview and Assumptions

Prior to a full examination and discussion of the results it is important that we first outline how these results will be presented and some of the assumptions used in the statistical analysis. The results are grouped according to the spectral algorithms and the metrics they were derived from.   There were three main categories of algorithms tested – classification, unmixing and target detection algorithms. Each one of these categories is discussed in its own separate section. In the respective sections, a data mean and standard deviation of the selected metric at each parameter level will be presented. For example, with a given image all values of a metric (e.g. kappa) that had a 40 m GIFOV associated with it are averaged together - despite what level of noise or spectral resolution was also part of that degraded image product. The metric averages are then plotted to visually help establish any trends of the main factors – spatial resolution, spectral resolution and noise. Finally, an analysis of variance is performed to demonstrate the significance of each factor and interactions on the metric from the degraded spectral product. The results of each one of these analysis steps – means, plots and ANOVA – are compared scene to scene to establish commonalties. All of the raw data is available on the enclosed CD. Summaries of the data are presented in Appendix E.

As discussed in Chapter 2 (section 2.8) the use of ANOVA tables allows us to determine the significance of main and joint effects via hypothesis testing. This hypothesis testing is done by comparing the calculated value F_o (from the ANOVA tables) to a critical region of the F-distribution (Montgomery, 1997). This region is defined by some level of significance, α, and has numerator degrees of freedom (df1) equal to the degrees of freedom of the source of variation

being tested and denominator degrees of freedom (df2) equal to that associated with the error. The set of hypotheses we make for testing all of our results is:

H0: Equal significance of each main effect and no interaction effects

H1: Main effects and interactions are significant

We reject the null hypothesis if $F_o > F_{\alpha, df1, df2}$. The level of significance for each test in this chapter is $\alpha = 0.05$. As we will see shortly the values of the F-test statistics we care about are $F_{0.05, 2, 8} = 4.46$ and $F_{0.05, 4, 8} = 3.84$. These values are listed again for convenience under each ANOVA table.

The layout of the ANOVA tables presented in this chapter are similar to that outlined in Table 2-3. There are two differences worth mentioning. Firstly, there is a sequential and adjusted sum-of-squares calculated. Adjusted sums of squares are the additional sums-of-squares determined by adding each particular term last into the linear ANOVA model. Sequential sums of squares are the sums of squares added by a term with only the previous terms entered in the model. (Minitab, 1998). These sums of squares will differ when the experimental design is not balanced. Secondly, an additional column "P" is added to the ANOVA table presented here. The numbers under this column represent the smallest level of significance that would lead to rejection of the null hypothesis (Montgomery, 1997). This allows us to see if and at what level the factor associated with a certain P-value will become significant other than at the level used in testing.

One of the principal assumptions made in this thesis concerns the statistical analysis of the data using ANOVA techniques. As seen in Chapter 2, the application of ANOVA to the results from a factorial designed experiment allows the user to statistically test the significance of all main factors and their interactions. However, *closer* inspection of Table 2-3 shows that it is necessary that at least two runs ($n \geq 2$) of the experiment be performed in order to properly calculate an error sum of squares SS_E. For instance, one run of the experiment is considered taking a degraded image and applying one of the spectral algorithms. Proper ANOVA techniques, as depicted in Table 2-3, dictate that this will be done at least one more time to properly establish the experimental error. In this thesis study only one run of the experiment is performed. It does not make sense to perform any portion of this experiment twice since the spectral products will not change from run-to-run given that the inputs (i.e. endmembers) remain the same. This means

that experimental error, SS_E, cannot be separated from the effect of the three-way interaction/joint effects of spatial, spectral resolution and noise, SS_{ABC} – as seen in Table 2-3. One could argue that each degraded image set be considered a single run of the experiment. That is, the degraded spectral products from AVIRIS Rochester are one run, products from AVIRIS Rogers Dry Lake are a second run and so forth. This sort of combination would definitely allow us to individually calculate a SS_E and SS_{ABC}. However, the SS_E would then represent the error introduced by different image sets and subsequent statistics (such as the mean square and F_o statistic) would be calculated with respect to a scene dependent error. This is believed to be an incorrect approach. In this regard, we could assume that the three-way interaction effect is zero and then proceed with the usual significance testing of main effects and two-way joint effects. Another way of looking at this assumption is that all hypothesis tests are done with respect to a sum-of-squares that represents the three-way joint effects of spatial resolution, spectral resolution and noise – which is really denoted as the experimental error, SS_E. This is the assumption made in the production of the following ANOVA tables and hypothesis testing.

4.2   Classification/Thematic Mapping

Prior to a presentation of the results obtained by using the metric kappa and applying it to our series of degraded classification products, it is important to first review which algorithms are being tested and the metric being used. The classification or thematic mapping algorithms being tested are the Spectral Angle Mapper (SAM), Binary Encoding (BE) and the Gaussian Maximum Likelihood (GML) algorithms. It should be noted that the GML method was done in conjunction with a Minimum Noise Fraction (MNF) transform of the image data to reduce noise and data dimensionality. The employment of the kappa statistic, as discussed in Chapter 2, is more versatile than merely using a measure of accuracy to grade algorithm performance. It allows us to test significant differences between spectral products and different algorithms.

ENDMEMBERS

	trees (white)
	grassland (cyan)
	marsh1 (red)
	marsh2 (green)
	beach (blue)
	field1 (yellow)
	field2 (brown)
	urban1 (orange)
	urban2 (teal)
	urban3 - roads (orchid)
	urban4 (chartreuse)
	deep water (magenta)
	shore water (maroon)
	shallow/bay water (purple)
	Genesee water (sea green)

GIFOV = 20m

GIFOV = 80m

SAM
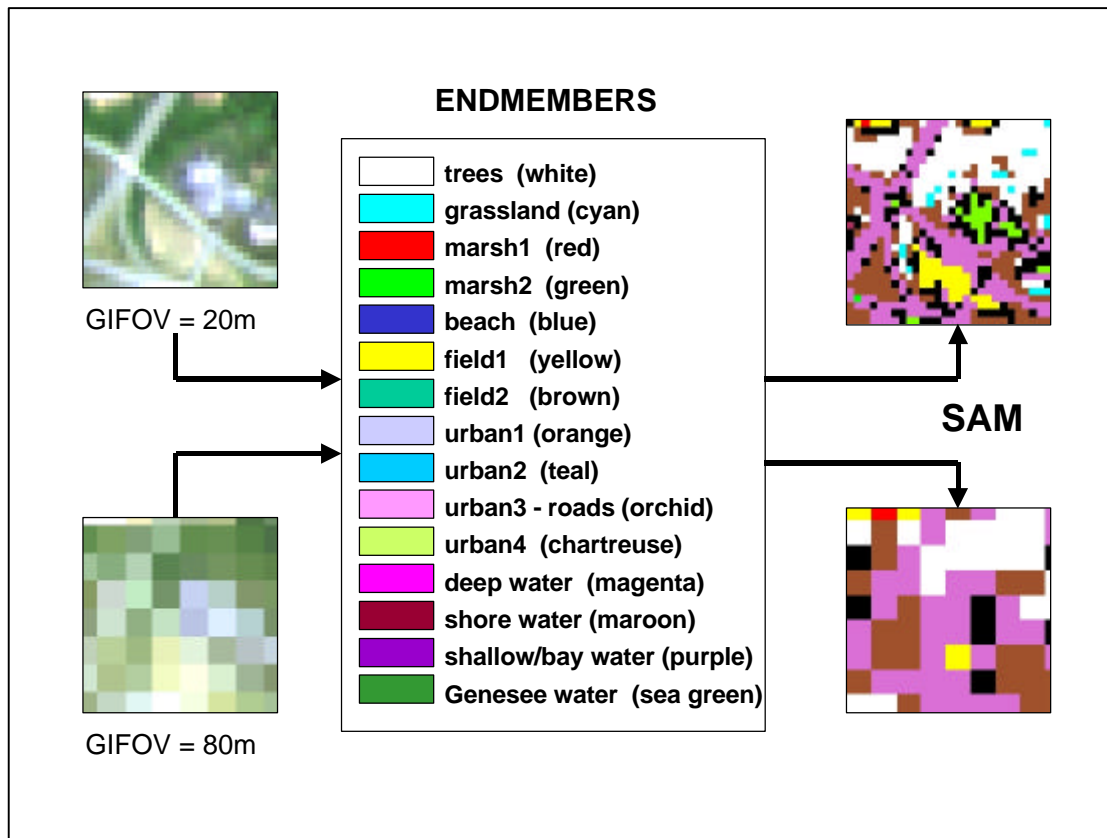
Figure 4-1: Example of Degradation of Classification Map (Level 3 Product)

As seen in Figure 4-1, we can see the expected results of a classification algorithm applied to degraded imagery. In Figure 4-1, we have *only* degraded the spatial resolution of the imagery. By doing this we are essentially forcing the pixels to become "more mixed." It is through this combination of the spectral signatures resident in each pixel that we overlook classifying such land cover classes as grassland, marsh2 and field1 in the 80 m GIFOV image. Obviously for each image that is degraded and subsequently classified we will have a classification map similar to that in Figure 4-1. This is excellent for assessing visual differences qualitatively, however, we are interested in the quantitative metric.

The ground truth references used in the calculation of kappa were the initial SAM results from the non-degraded imagery for the AVIRIS images and the material truth map for the DIRSIG image. Using these classification results as ground truth allows us to see how image utility degrades with respect to selected ground truth. Likewise, a calculation of kappa was conducted for each algorithm with respect to its own respective non-degraded spectral product. This was done to see how the information utility produced by the algorithm behaves with respect

to itself. For instance, a BE was applied to all of the degraded Rochester scenes. Calculations of kappa for each degraded BE spectral product were determined with respect to the initial SAM product as reference ground truth. Subsequently, the BE results of the non-degraded image were used as ground truth for a second calculation of kappa – a value of kappa showing how the spectral product changes with respect to its initial non-degraded BE product.

A summary of the collected raw data is available in Appendix E and summarized below in Table 4-1, 4-2 and 4-3. Undoubtedly, there is an absolute wealth of data here and it is better to visualize the data in graphs. At this point the size of the standard deviations associated with these metrics should be noted. Recall that each statistic is the average or standard deviation of all the values while one parameter is held constant. For example, the mean value of kappa for a spatial resolution of 20 m includes all raw data at all levels of noise or spectral resolution associated with a spatial resolution of 20 m. This includes data points with SNR values of 10, 100 and 225 and spectral resolutions of 10, 55 and 110 nm. Obviously the huge standard deviations (many close to the size of the mean itself) are attributed to the noisy data with low SNR. Despite these large standard deviations and the hindrance they may impose on establishing concrete conclusions – the main aim here is to investigate and observe variable trends with respect to algorithm performance.

spatial (m)	Rogers Dry Lake SAM - Kappa		Rogers Dry Lake BE - Kappa		Rogers Dry Lake BE to SAMref - Kappa		Rogers Dry Lake GML - Kappa		Rogers Dry Lake GML to SAMref - Kappa	
	mean	std	mean	std	mean	std	mean	std	mean	std
20	0.5778	0.4404	0.2509	0.3056	0.1276	0.1422	0.7176	0.2581	0.5577	0.1029
40	0.5166	0.3752	0.2242	0.2639	0.0896	0.1098	0.6130	0.1724	0.5001	0.0920
80	0.4485	0.3309	0.1930	0.2264	0.1135	0.1287	0.2924	0.2821	0.2852	0.2352

spectral (nm)	Rogers Dry Lake SAM - Kappa		Rogers Dry Lake BE - Kappa		Rogers Dry Lake BE to SAMref - Kappa		Rogers Dry Lake GML - Kappa		Rogers Dry Lake GML to SAMref - Kappa	
	mean	std	mean	std	mean	std	mean	std	mean	std
10	0.5471	0.4146	0.4138	0.3127	0.1691	0.1438	0.5819	0.4501	0.3683	0.2776
55	0.5273	0.3908	0.2543	0.1427	0.1616	0.0918	0.5717	0.1747	0.4952	0.1069
110	0.4684	0.3539	0.0000	0.0000	0.0000	0.0000	0.4916	0.1858	0.4794	0.1468

Noise (SNR@30%)	Rogers Dry Lake SAM - Kappa		Rogers Dry Lake BE - Kappa		Rogers Dry Lake BE to SAMref - Kappa		Rogers Dry Lake GML - Kappa		Rogers Dry Lake GML to SAMref - Kappa	
	mean	std	mean	std	mean	std	mean	std	mean	std
10	0.0139	0.0118	0.0381	0.0316	0.0215	0.0183	0.4001	0.2967	0.3446	0.1692
100	0.7058	0.1122	0.2602	0.2033	0.1494	0.1127	0.6201	0.2652	0.5004	0.1936
225	0.8231	0.1032	0.3699	0.3309	0.1599	0.1525	0.6416	0.2845	0.4980	0.1917

Table 4-1: Mean and Standard Deviation of Kappa – Rogers Dry Lake

spatial (m)	Rochester SAM - Kappa		Rochester BE - Kappa		Rochester BE to SAMref - Kappa		Rochester GML - Kappa		Rochester GML to SAMref - Kappa	
	mean	std	mean	std	mean	std	mean	std	mean	std
20	0.5490	0.4292	0.4696	0.2454	0.3504	0.0812	0.7896	0.2058	0.5511	0.0215
40	0.4072	0.3126	0.3983	0.1730	0.3177	0.0695	0.6802	0.1368	0.4785	0.0300
80	0.3237	0.2572	0.3404	0.1385	0.2808	0.0651	0.5706	0.0992	0.4028	0.0344

spectral (nm)	Rochester SAM - Kappa		Rochester BE - Kappa		Rochester BE to SAMref - Kappa		Rochester GML - Kappa		Rochester GML to SAMref - Kappa	
	mean	std	mean	std	mean	std	mean	std	mean	std
10	0.4653	0.3789	0.5707	0.2200	0.3250	0.0752	0.7981	0.1688	0.4645	0.0652
55	0.4310	0.3558	0.3616	0.1210	0.3108	0.0775	0.6445	0.1497	0.4750	0.0744
110	0.3836	0.3214	0.2760	0.0633	0.3131	0.0813	0.5978	0.1492	0.4929	0.0683

Noise (SNR@30%)	Rochester SAM - Kappa		Rochester BE - Kappa		Rochester BE to SAMref - Kappa		Rochester GML - Kappa		Rochester GML to SAMref - Kappa	
	mean	std	mean	std	mean	std	mean	std	mean	std
10	0.0034	0.0048	0.2476	0.0714	0.2238	0.0273	0.5603	0.2078	0.4552	0.0729
100	0.5376	0.1610	0.4495	0.1608	0.3467	0.0357	0.7395	0.1216	0.4978	0.0667
225	0.7388	0.1622	0.5112	0.2122	0.3785	0.0369	0.7406	0.1253	0.4794	0.0642

Table 4-2: Mean and Standard Deviation of Kappa – Rochester

spatial (m)	West Rain SAM to GT		West Rain SAM to SAMref		West Rain BE to GT		West Rain BE to Beref		West Rain GML to GT		West Rain GML to GMLref	
	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std
2	0.5140	0.3859	0.6358	0.4755	0.7305	0.3390	0.8090	0.2642	0.9361	0.0540	0.9573	0.063(
4	0.5336	0.4071	0.4946	0.3757	0.7024	0.2538	0.7306	0.2551	0.8620	0.0564	0.8798	0.062(
8	0.4914	0.3855	0.4510	0.3520	0.6572	0.2784	0.6785	0.2790	0.5220	0.3935	0.5318	0.401;

spectral (nm)	West Rain SAM to GT		West Rain SAM to SAMref		West Rain BE to GT		West Rain BE to Beref		West Rain GML to GT		West Rain GML to GMLref	
	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std
10	0.4659	0.3677	0.4850	0.3936	0.7311	0.2272	0.7730	0.2369	0.6028	0.4541	0.6158	0.464(
55	0.5167	0.3925	0.5304	0.4138	0.6985	0.2684	0.7371	0.2804	0.8785	0.0671	0.8968	0.071;
110	0.5564	0.4129	0.5659	0.4270	0.6606	0.3653	0.7079	0.2920	0.8388	0.0997	0.8562	0.107

Noise (SNR@30%)	West Rain SAM to GT		West Rain SAM to SAMref		West Rain BE to GT		West Rain BE to Beref		West Rain GML to GT		West Rain GML to GMLref	
	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std
10	0.0022	0.0041	0.0030	0.0060	0.3240	0.1448	0.3888	0.0961	0.7244	0.2820	0.7308	0.284(
100	0.7050	0.1196	0.7317	0.1678	0.8812	0.0454	0.9127	0.0533	0.8001	0.3062	0.8206	0.314
225	0.8318	0.0257	0.8467	0.1136	0.8849	0.0501	0.9166	0.0544	0.7956	0.3055	0.8175	0.314;

Table 4-3: Mean and Standard Deviation of Kappa – DIRSIG – Western Rainbow

Key

mean – mean of all kappa values at that parameter level (spatial or spectral resolution, or noise)

std – standard deviation of kappa values at that parameter level (spatial or spectral resolution, or noise)

SAM – Spectral Angle Mapper algorithm results (mean) at that parameter level

BE – Binary Encoding algorithm results (mean) at that parameter level

GML – Gaussian Maximum Likelihood algorithm results (mean) at that parameter level

GT – Ground Truth (SAM results of initial AVIRIS imagery and material map for DIRSIG image) to observe how the algorithm results degrade with respect to established and constant ground truth

SAMref – SAM results of the initial (non-degraded) image used as ground truth for kappa determination to observe how the algorithm results degrade with respect to itself

BEref – BE results of the initial (non-degraded) image used as ground truth for kappa determination to observe how the algorithm results degrade with respect to itself

GMLref – GML results of the initial (non-degraded) image used as ground truth for kappa determination to observe how the algorithm results degrade with respect to itself

Classification - BE - Kappa = f(Spatial Resolution)

Legend:
- mean-BE Rogers
- mean - BE to SAMref Rogers
- mean-BE Rochester
- mean - BE to SAMref Rochester
- mean - BE GT WestRain
- mean - BE Beref WestRain

Y-axis: Kappa
X-axis: Spatial Resolution (GIFOV - m)
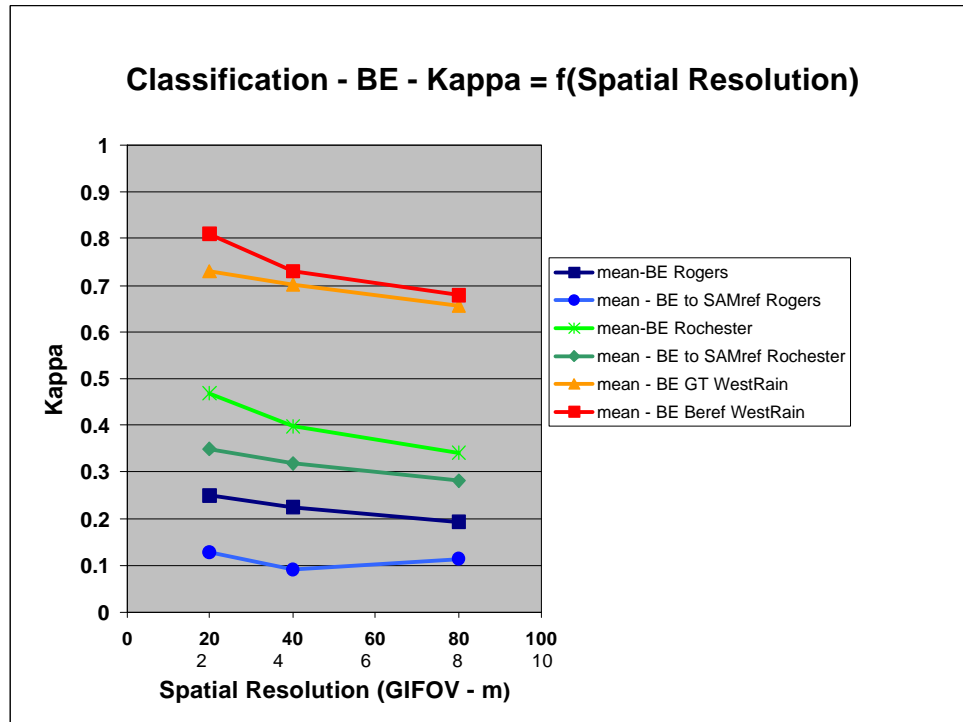
Figure 4-2: Binary Encoding – Kappa as a Function of Spatial Resolution

Immediate inspection of Figure 4-2 reveals that the lines are grouped together according to their scene. This is indicative of scene dependence within the calculation of the metric, which is attributed to the degree of spectral and spatial complexity of the initial images. On first look at the above plots, it appears that BE performs much better in reflectance space than in radiance space. This is mostly due to the fact that DIRSIG's spectral-spatial variability of backgrounds is not as nearly complex as the real world. This fact must be remembered when drawing any conclusions throughout this thesis when comparing performance of the algorithms in reflectance or radiance space – we are comparing apples and oranges. However, when using an image in radiance space we are initially taking the average of the spectra whose overall shape is dictated by the exoatmospheric solar irradiance and atmospheric absorption bands. Therefore BE does not pick up the spectral detail as well as it is able to in reflectance space during the encoding process. The way that ENVI has encoded this algorithm does not allow for "localized" averaging of the spectra prior to encoding. This type of approach would allow the user to focus in on specific absorption features. Additionally, by applying numerous local averages over radiance spectra, whose shape is dictated by atmospheric absorption and exoatmospheric solar irradiance, the effect of this shape would be minimized when encoding the data from the calculation of the spectral mean.

91

Classification - SAM - Kappa = f(Spatial Resolution)
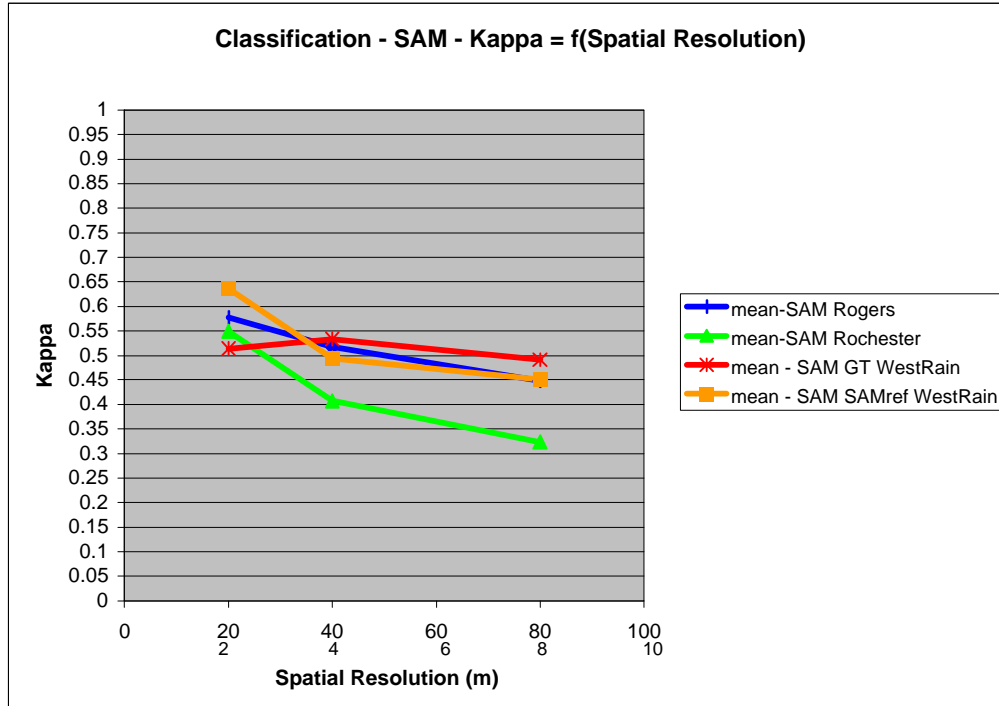
Figure 4-3: Spectral Angle Mapper – Kappa as a Function of Spatial Resolution

There is an obvious difference between Figure 4-3 and Figure 4-2. Immediately we see that there is not as much of a scene dependence associated with the SAM algorithm as there was with the BE algorithm. All of the curves in Figure 4-3 are grouped together. We also see that as spatial resolution degrades so does the value of kappa in a linear fashion. However, there appears to be a strange increase in the performance of the Western Rainbow scene (red), indicated by a small rise in the "curve" at 40 m. This is not a peculiarity if we take into consideration the standard deviations associated with these plotted numbers - as displayed in Table 4-3. It must be remembered that each point in these charts is a mean of all data points with a spatial resolution of 20, 40 or 80 m for the AVIRIS images and 2, 4 or 8 m for the DIRSIG scene. Associated with this mean is a large standard deviation that cannot be ignored. It must also be noted that these graphs were initially plotted with error bars for each data point. However, the inclusion of these error bars quickly made these graphs difficult to interpret and readily identify performance trends.
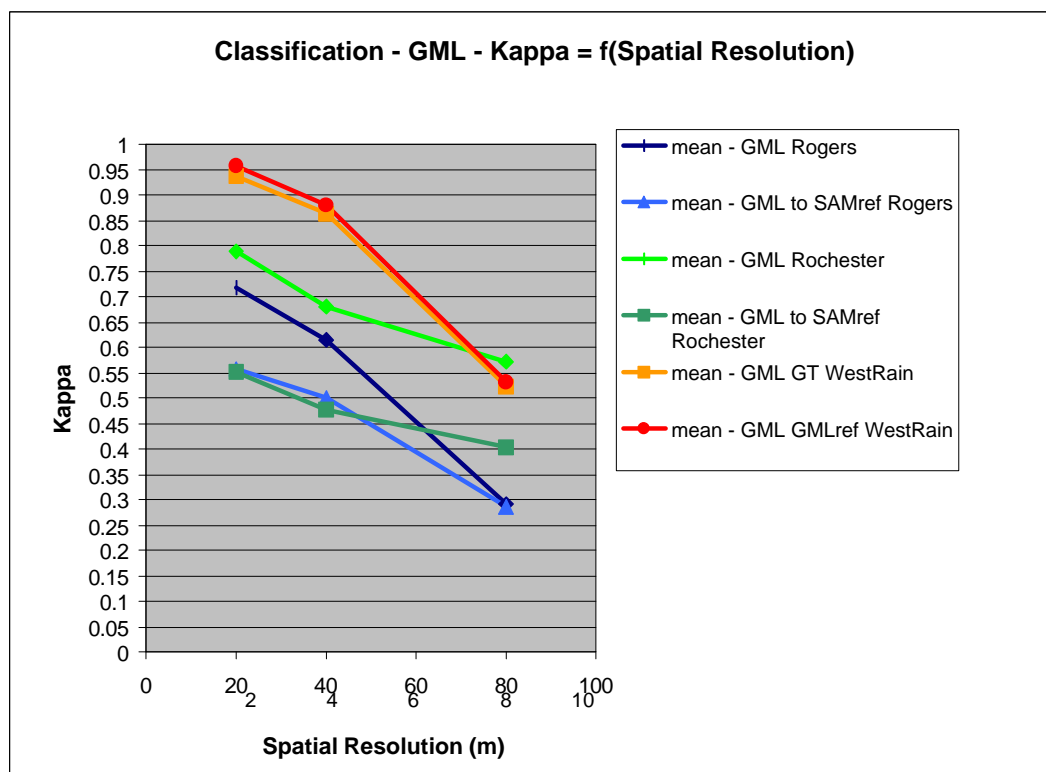
Figure 4-4: GML – Kappa as a Function of Spatial Resolution

Figure 4-4 exhibits a sharper decrease in performance of the GML algorithm as a function of spatial resolution. This can be attributed to the fact that the ROIs used during the training stage of this classification technique remained the same for all of the degraded images - even though more pixel mixing was occurring at the larger GIFOVs. This use of the *same* ROIs was done to maintain experimental consistency, but ultimately has contributed to this sharper degradation. Of special interest is that at the higher spatial resolutions, it appears that GML outperforms the other two algorithms (Figure 4-2 and 4-3) – this is especially seen with the DIRSIG scene. In this scene, the kappa metric was calculated with reference to the material truth map. Again we have to be careful in making this conclusion regarding GML performing better than the other algorithms since these points are a collective mean of all parameters and it is possible that the noise was suppressed by the use of the MNF transform prior to using the GML. Nevertheless, the GML results of the AVIRIS images with respect to the SAM reference image show equal performance between GML and SAM for real imagery.
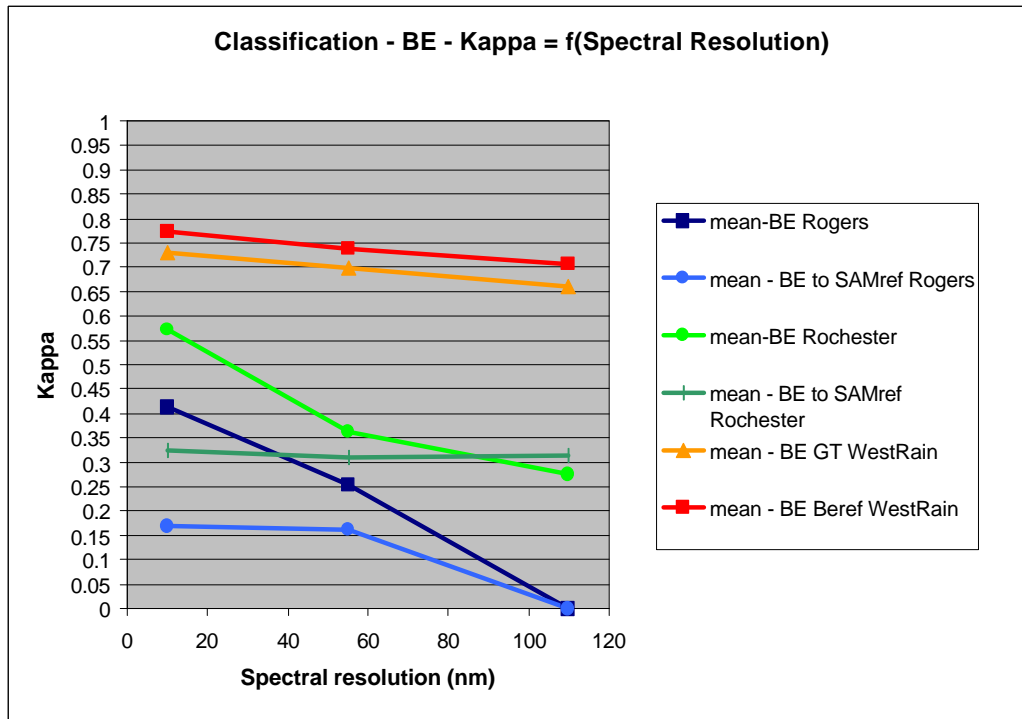
Figure 4-5: Binary Encoding – Kappa as a Function of Spectral Resolution

As seen initially in Figure 4-2, again we witness a noticeable scene dependence associated with the performance of the BE algorithm. This time the scene dependence is seen with respect to spectral resolution, as seen above in Figure 4-5. This is due in part to the spectral complexity of the different scenes. Firstly, the Rogers Dry Lake scene is spectrally homogeneous; thus posing an obstacle when trying to distinguish endmembers within the image. Along these lines, the pixel spectra in the DIRSIG scene are pure and of a smaller GIFOV so less mixing of pixels occurs. That is, there is more spectral seperability in the DIRSIG scene then the other scenes and therefore a better performance. All of these factors contribute to an overall scene dependence of the computed metric.

The difference in BE performance of an image in reflectance space versus radiance space is also noted again. However, we must remember that the spatial-spectral variability of backgrounds in a DIRSIG generated image is not as complex in the real world. Again the idea of a better encoding approach of the BE algorithm using localized averages of spectral features still holds validity.

Classification - SAM - Kappa = f(Spectral Resolution)
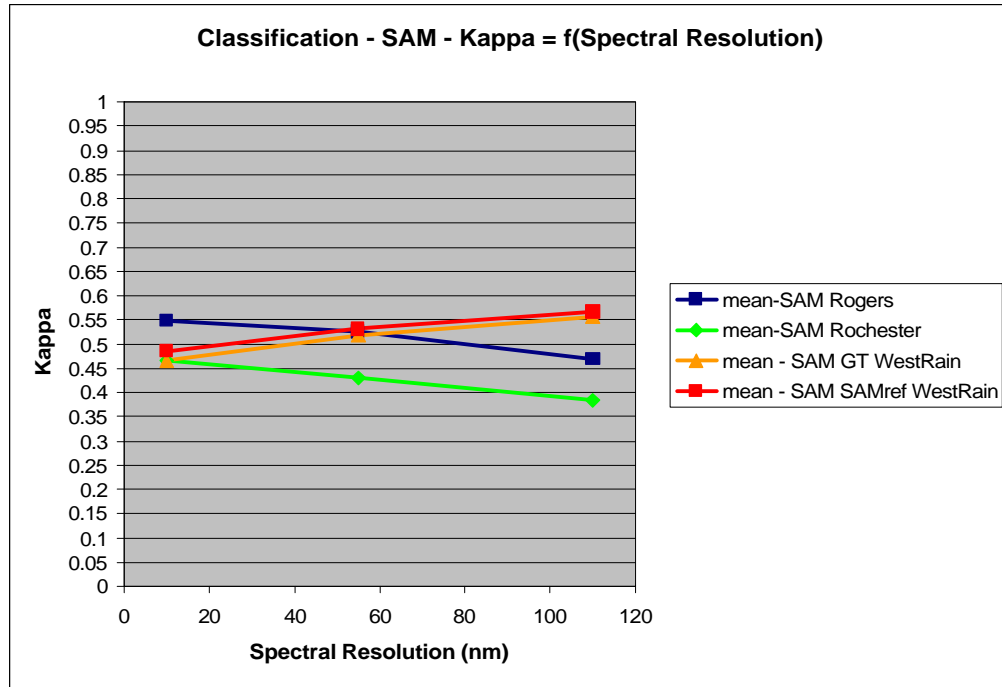
Figure 4-6: Spectral Angle Mapper – Kappa as a function of Spectral Resolution

Immediate inspection of Figure 4-6 reveals an interesting peculiarity of SAM performance. The "phenomena" of the kappa metric actually increasing as spectral resolution worsens (Western Rainbow – red line) is counter intuitive and can be explained by including the standard deviation of these calculations (from Table 4-3). If the above charts included error bars this would be more apparent, since these error bars are quite large. Given that, it would appear the tight grouping of curves would indicate there is little to no scene dependence with respect to spectral resolution changes and the performance of SAM. These lines are also relatively flat – indicating that SAM performance may be independent of spectral resolution. Does this make sense? Yes, since spectral resolution changes will not effect the position and direction of the spectral vectors. Only the number of dimensions (i.e. number of bands) it takes to represent this spectral signature as a vector will change. In other words, all of the spectral vectors still lie in the same direction - they are just represented by fewer dimensions. In this respect, the angle between them has not changed drastically and therefore any changes in spectral resolution do not contribute much of a change to SAM algorithm performance.
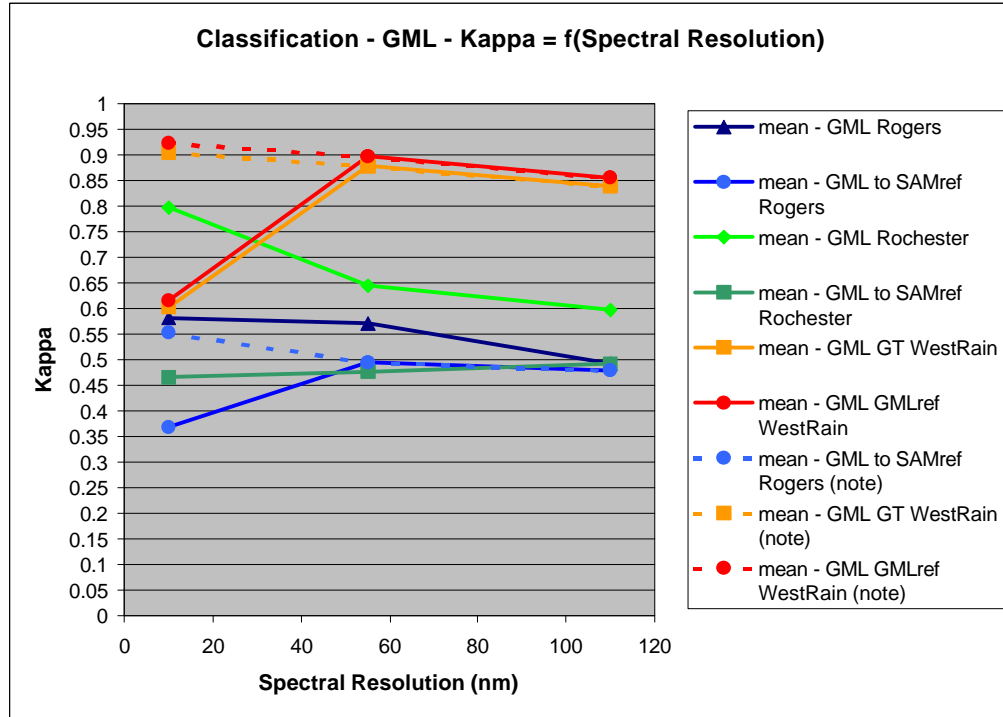
Figure 4-7: GML – Kappa as a Function of Spectral Resolution

At first glance of Figure 4-7, we could say that the performance of GML increases as spectral resolution coarsens. However, closer inspection of the data in Appendix E indicates that three points need to be excluded in the average calculation of a spectral resolution of 10 nm associated with a corresponding spatial resolution of 80m.   As seen in Appendix E, the kappa value at these levels is zero. This is because the ROIs employed at a low spatial resolution of 80m result in many of the collected spectra being the same (via the aggregate process of spatial degradation and using the same ROIs throughout). When the statistics are calculated these equal spectra are carried over into the covariance matrix. Equal or proportional rows or columns in a matrix result in a determinant of zero and therefore the matrix in not invertible. In our case, the covariance matrix cannot be inverted to complete the calculations associated with the GML classifier – and therefore a zero result is delivered. Throwing these data points out leads to a trend that is more readily expected in that the value of kappa decreases with a gradual slope as spectral resolution worsens. Again, scene dependencies on the performance are noted.
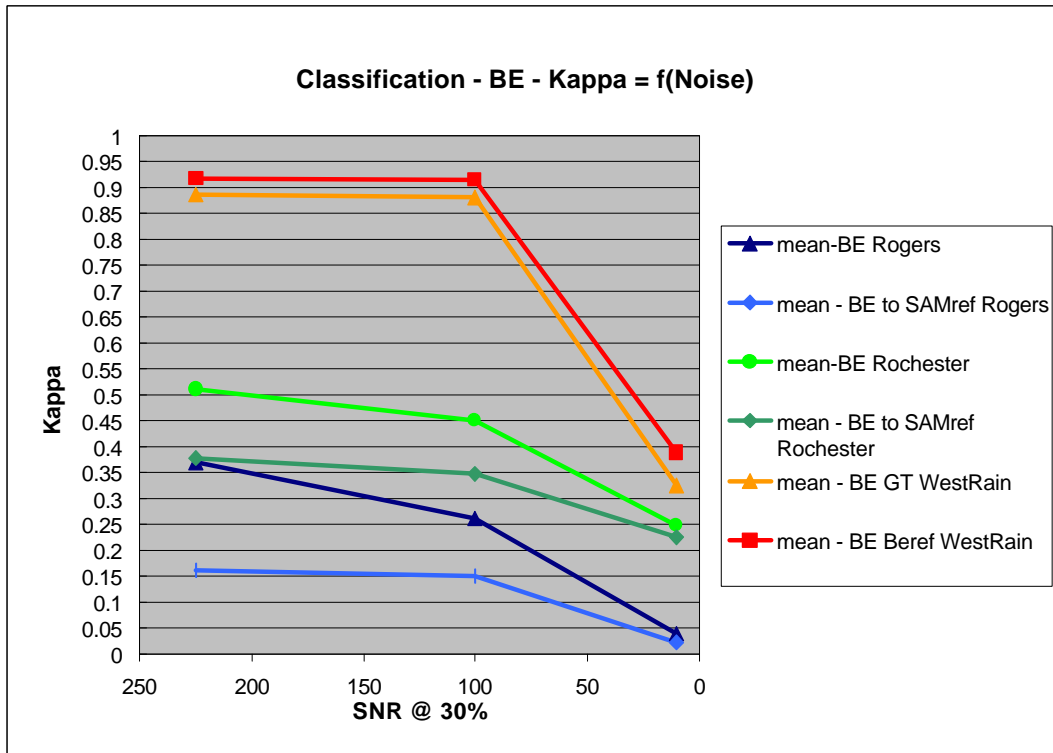
Figure 4-8: Binary Encoding – Kappa as a Function of Noise

As expected, noise has an adverse effect on the spectral products from the BE algorithm. This can be seen in Figure 4-8. It is interesting to note the different effects that noise has on images in reflectance space versus those in radiance space. The images in radiance space have a gentle linear decrease in their performance while the image in reflectance space is quite tolerant of noise up to some threshold between a SNR of 10 to 100. In radiance space, noisy spectra lie above the same average as a noiseless spectrum of the same material since the general shape of the spectra remains the constant with noise addition. Recall that this shape is dictated by atmospheric absorption and exoatmospheric irradiance. The averaging process done prior to encoding is relatively insensitive to noise in radiance space and there seems to be sensitivity to noise in reflectance space at a certain threshold. However, it should also be noted that the images in reflectance space (DIRSIG generated) also correspond to images with a finer spatial resolution and pixels that are more spectrally pure. More investigation is needed.
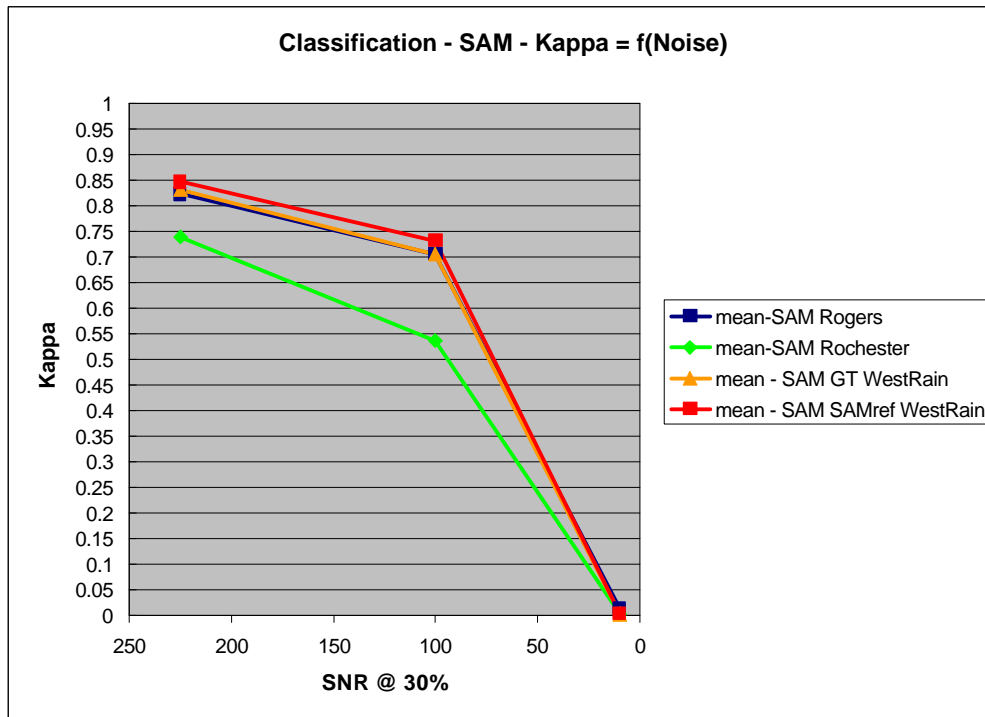
Figure 4-9: Spectral Angle Mapper – Kappa as a Function of Noise

Examination of Figure 4-9 shows that noise has a more profound effect on the performance of SAM. Especially when compared to the effects of noise associated with BE seen in Figure 4-8. With respect to the SAM algorithm, at high SNR values we have excellent performance, yet this drops off dramatically. Essentially noise in a band will change that spectral vector. The spectral vector's position and direction in n-dimensional space (where n is the number of bands) will also change. In that respect, the angle between the reference spectra and pixel spectra will correspondingly change. If enough noise is added then the vector position is drastically changed and the corresponding angle to the reference/library spectra will also alter, thereby leading to a sudden drop in algorithm performance.
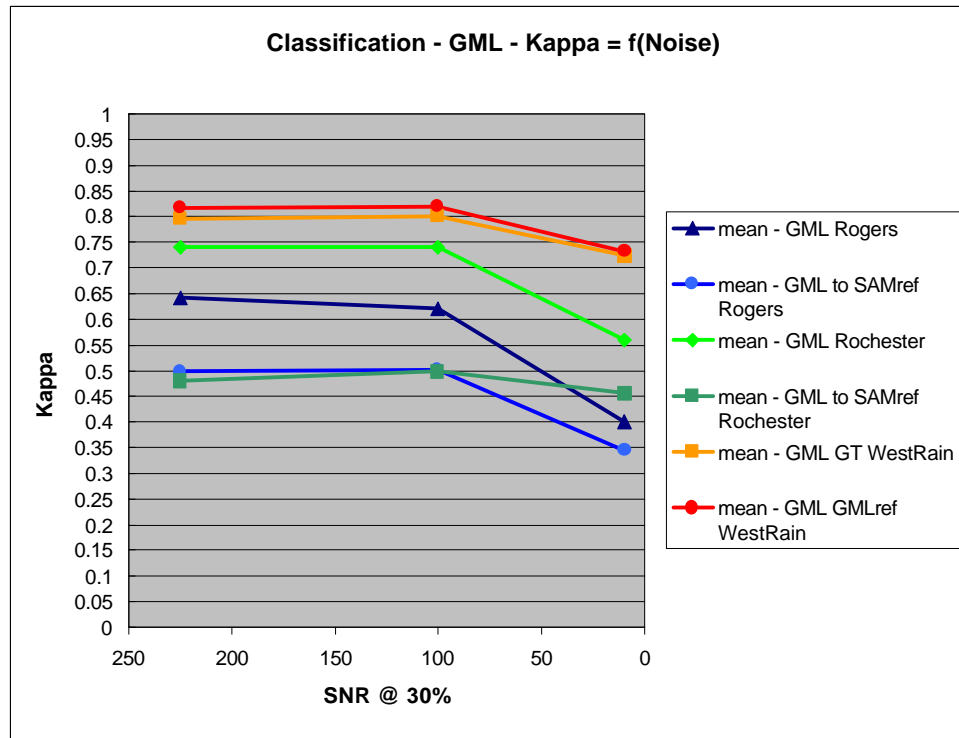
Figure 4-10: GML – Kappa as a Function of Noise

As expected transforming the degraded images into MNF space and ultimately removing the noisy bands prior to GML lessens the harsh effects of noise. This can be seen by the gentle decreasing linear slope in the kappa metric as plotted in Figure 4-10. Although there is a slight demise of performance with the GML algorithm, with respect to noise, it surely is not as drastic as the decrease in performance of the BE and SAM algorithms seen in Figures 4-8 and 4-9. This would suggest that some type of noise removal and data dimensionality reduction is advantageous. Again, we must be careful since the averages plotted include the effect of another parameter (spectral or spatial resolution).

Examination of the ranges of the kappa values corresponding to each algorithm and main parameters requires further discussion. For instance, with reference to SAM – its performance outcome, as indicated by kappa, with respect to spectral resolution has a close range from 0.4 to 0.6. However, examination of the SAM algorithm with respect to noise delivers a wider kappa range from approximately 0 to 0.85. This would indicate that noise is a more significant factor than spectral resolution for the SAM algorithm. This approach is useful in determining which factor is more significant over another in algorithm performance, but this is why the use of

ANOVA techniques are extremely beneficial and employed in this thesis. Further testing of the significance of each main factors and interactions is done using ANOVA in the following paragraphs.

Binary Encoding

```
Analysis of Variance for BE – Kappa - Rochester

Source             DF     Seq SS     Adj SS     Adj MS        F       P
Spatial             2   0.075371   0.075371   0.037686    52.11   0.000
Spectral            2   0.413502   0.413502   0.206751   285.90   0.000
Noise               2   0.342015   0.342015   0.171008   236.47   0.000
Spatial*Spectral    4   0.035004   0.035004   0.008751    12.10   0.002
Spatial*Noise       4   0.010182   0.010182   0.002546     3.52   0.061
Spectral*Noise      4   0.067922   0.067922   0.016981    23.48   0.000
Error               8   0.005785   0.005785   0.000723
Total              26   0.949783
```

Table 4-4: ANOVA for Rochester BE – Kappa ($F_{0.05, 2, 8} = 4.46$ and $F_{0.05, 4, 8} = 3.84$)

As seen in Table 4-4, all main effects of spatial resolution, spectral resolution and noise do not support the null hypothesis and are therefore significant to the outcome of the BE algorithm applied to the Rochester scene. The main factor of spectral resolution is the dominant factor, which is closely followed by noise and then spatial resolution. The interaction between spatial resolution and noise is not significant at this testing level.

```
Analysis of Variance for BE to SAMref - Rochester

Source             DF     Seq SS      Adj SS      Adj MS         F       P
Spatial             2   0.0218305   0.0218305   0.0109153    788.22   0.000
Spectral            2   0.0010510   0.0010510   0.0005255     37.95   0.000
Noise               2   0.1200765   0.1200765   0.0600382   4335.54   0.000
Spatial*Spectral    4   0.0016688   0.0016688   0.0004172     30.13   0.000
Spatial*Noise       4   0.0009857   0.0009857   0.0002464     17.79   0.000
Spectral*Noise      4   0.0014355   0.0014355   0.0003589     25.91   0.000
Error               8   0.0001108   0.0001108   0.0000138
Total              26   0.1471587
```

Table 4-5: ANOVA for Rochester BE to SAM reference – Kappa
($F_{0.05, 2, 8} = 4.46$ and $F_{0.05, 4, 8} = 3.84$)

Table 4-5 shows that all main effects and interactions are still significant when comparing the results of BE to a SAM reference image as ground truth. However, we can see that by comparing the magnitudes of the F-statistics that the interaction of spatial resolution and noise is the least significant. This corresponds to the results in the previous table, in which the trade-space between spatial resolution and noise was also the least significant.

```
Analysis of Variance for BE - Kappa - Rogers Dry Lake

Source              DF      Seq SS      Adj SS      Adj MS        F       P
Spatial              2    0.015143    0.015143    0.007572     6.05   0.025
Spectral             2    0.784066    0.784066    0.392033   313.43   0.000
Noise                2    0.514423    0.514423    0.257211   205.64   0.000
Spatial*Spectral     4    0.016519    0.016519    0.004130     3.30   0.071
Spatial*Noise        4    0.005497    0.005497    0.001374     1.10   0.420
Spectral*Noise       4    0.383551    0.383551    0.095888    76.66   0.000
Error                8    0.010006    0.010006    0.001251
Total               26    1.729205
```

Table 4-6: ANOVA for Rogers Dry Lake BE – Kappa
($F_{0.05, 2, 8} = 4.46$ and $F_{0.05, 4, 8} = 3.84$)

As seen in Table 4-6, none of the main effects support the null hypothesis and are therefore significant to BE performance when applied to the Rogers Dry Lake scene. Similar to the Rochester scene, the main effect of spectral resolution is the dominant factor and the interaction between spatial resolution and noise has no significance in algorithm performance. Scene dependence is evident here since the size of the F-statistic for the main effect of spatial resolution for the Rogers Dry Lake scene is much smaller in comparison to the Rochester scene. This is primarily due to the fact that the Rogers scene is spatially homogeneous.

```
Analysis of Variance for BE to SAMref- Rogers Dry Lake

Source              DF      Seq SS      Adj SS      Adj MS        F       P
Spatial              2    0.006659    0.006659    0.003329     0.78   0.489
Spectral             2    0.164295    0.164295    0.082147    19.31   0.001
Noise                2    0.106866    0.106866    0.053433    12.56   0.003
Spatial*Spectral     4    0.013564    0.013564    0.003391     0.80   0.559
Spatial*Noise        4    0.017304    0.017304    0.004326     1.02   0.454
Spectral*Noise       4    0.054568    0.054568    0.013642     3.21   0.075
Error                8    0.034030    0.034030    0.004254
Total               26    0.397285
```

Table 4-7: ANOVA for Rochester BE to SAM reference – Kappa
($F_{0.05, 2, 8} = 4.46$ and $F_{0.05, 4, 8} = 3.84$)

As seen in Table 4-7, with respect to BE applied to the Rogers Dry Lake image, the only factors which are significant when the BE algorithm is compared to some reference ground truth (SAM results) are the factors of spectral resolution and noise. This compares well with the results from the Rochester scene. Again, due to the lack of spatial detail in the Rogers Dry Lake scene, we see that the main effect of spatial resolution is of no significance.

```
Analysis of Variance for BE - Kappa GT - Western Rainbow


Source            DF      Seq SS     Adj SS     Adj MS       F       P
Spatial            2     0.02465    0.02465    0.01233     1.57   0.266
Spectral           2     0.02244    0.02244    0.01122     1.43   0.294
Noise              2     1.87517    1.87517    0.93759   119.52   0.000
Spatial*Spectral   4     0.00706    0.00706    0.00177     0.23   0.917
Spatial*Noise      4     0.01433    0.01433    0.00358     0.46   0.766
Spectral*Noise     4     0.07289    0.07289    0.01822     2.32   0.144
```

Table 4-8: ANOVA for Western Rainbow BE to Material Map Ground Truth reference – Kappa
($F_{0.05, 2, 8} = 4.46$ and $F_{0.05, 4, 8} = 3.84$)

```
Analysis of Variance for BE - Kappa to SAMref - Western Rainbow


Source            DF      Seq SS     Adj SS     Adj MS         F       P
Spatial            2    0.077606   0.077606   0.038803   103.64   0.000
Spectral           2    0.019133   0.019133   0.009566    25.55   0.000
Noise              2    1.659024   1.659024   0.829512  2215.56   0.000
Spatial*Spectral   4    0.001162   0.001162   0.000290     0.78   0.571
Spatial*Noise      4    0.002969   0.002969   0.000742     1.98   0.190
Spectral*Noise     4    0.016466   0.016466   0.004117    10.99   0.002
Error              8    0.002995   0.002995   0.000374
Total             26    1.779355
```

Table 4-9: ANOVA for Western Rainbow BE to SAMref Ground Truth reference – Kappa
($F_{0.05, 2, 8} = 4.46$ and $F_{0.05, 4, 8} = 3.84$)

Examination of Tables 4-8 and 4-9 reveal some interesting results with respect to the BE algorithm and its application to the DIRSIG generated scene. It is readily seen that the significance of factors is dependent on the selected ground truth. For instance, in Table 4-8, only the main effect of noise is significant to the performance of the BE when the material map is used as ground truth. However, when using the non-degraded SAM results as ground truth, all three main factors carry a large degree of significance in the algorithm's performance. This may suggest that our previous findings, with respect to the significance of main effects and interactions, may be erroneous. Again, more investigation is needed and as stated earlier the need for better ground truth is realized. However, this was non-existent at the time of this thesis study.

The analysis of the collected data and ANOVA tables made for the SAM and GML algorithms are included on the enclosed CD in the "data/minitab" directory. The in-depth analysis of the BE algorithm that was conducted above is similar to that done from the ANOVA tables produced from analysing the two other classification algorithms. However, for the sake of brevity only the highlights from this analysis will be discussed in the following paragraphs. The user is invited to cross reference the ANOVA tables on the CD if so desired.

<u>Spectral Angle Mapper (SAM)</u>

The application of ANOVA techniques to the kappa values found after applying the SAM algorithm to the Rochester and Rogers Dry Lake images exhibit very similar results. With both images, all main factors were significant contributors to algorithm performance. The main factor of noise was the most significant factor, by approximately two orders of magnitude, in the ANOVA tables of both images. This is followed by the main factor of spatial resolution and then spectral resolution. Since the ranking of the main factors is the same for each image, the point made earlier about less scene dependence associated with the SAM algorithm is further validated. Furthermore, the significance ranking of spectral resolution as the last of all main factors completely agrees with the findings arrived at previously from Figure 4-6. It was from Figure 4-6, that the results of the SAM algorithm being somewhat independent of spectral resolution changes were discussed. Additionally from these ANOVA tables (Rochester and Rogers Dry Lake), it is found that the interaction of spatial resolution and noise was significant with the products from both images and ranked as fourth.

Application of ANOVA techniques to the values of kappa generated by applying the SAM algorithm to the Western Rainbow scene immediately reveal that there seems to be a dependence on the selected ground truth. All three main factors of spatial resolution, spectral resolution and noise were significant to SAM algorithm performance. However, similar to the BE results of the Western Rainbow scene, the level and ranking of significant factors changes when we use the material map as ground truth compared to using the non-degraded SAM results as ground truth. When using the material map as ground truth the main factor of noise was most significant (by two orders of magnitude), followed by spectral resolution and spatial resolution. However, when using the initial SAM results as ground truth it was found that, noise is the most significant main factor, followed by spatial resolution and then the interaction between spatial resolution and noise. Despite the fact that spectral resolution is still of statistical significance, in

this "SAM" case, it is the least. This difference in ranking and corresponding significance levels is similar to those discovered from conducting an ANOVA on the BE results from the Western Rainbow scene. This brings to mind a few questions that need further investigation. Firstly, despite the relative match between the material map and the initial SAM results, it is possible that a better selection of reference ground truth than the initial SAM results could be done. However, with a complete lack of ground truth for the AVIRIS images, something had to be chosen as ground truth. More investigation is needed. Secondly, it is possible that the independence of spectral resolution on the value of kappa exhibited with the AVIRIS images is due mostly to the ground truth that was selected – despite the fact that this independence makes theoretical sense. Yet, using the initial SAM results, as in this case, is an attempt at seeing how the information utility produced by the algorithm degrades with respect to itself. Either way, more investigation is required. Of special note when using the material map as ground truth was that the magnitude of the F-statistic for the interaction of spectral resolution and noise was approximately equal to that of the main factor of noise – which was also of very high significance.

Gaussian Maximum Likelihood

When examining the ANOVA results of the GML algorithm applied to the Rogers Dry Lake scene we see similar results whether we use the initial/non-degraded SAM product as ground truth reference or the initial GML product as a reference. The main factor of spatial resolution was most significant to algorithm performance in both ANOVA tables. The second most significant factor, when using the initial GML product as ground truth, was the interaction between spatial and spectral resolution. The third most significant factor in this case was the main factor of noise. The results from the ANOVA of Rogers Dry Lake, using the initial SAM product as reference, closely follow those in which the GML product was used as the reference. Using the initial SAM results as reference yields the trade-space between spatial and spectral resolution as the second most significant factor. Additionally, the value of the F-statistic for spatial-spectral resolution interaction closely matches the F-statistic of noise, which was third. The similar ranking of these results possibly suggests that the selection of ground truth does not make as much of a difference with the GML algorithm as it does with the SAM and BE algorithms. It also suggests that the GML algorithm degrades with respect to itself in a very similar fashion as it does with respect to selected ground truth. Obviously, we need to examine the ANOVA results from the other two scenes before arriving at any further conclusion.

The ANOVA table showing the results of the GML algorithm applied to the Rochester scene, using the initial SAM results as ground truth indicates that the spatial resolution is the most significant main factor to GML performance - by an order of magnitude. This is followed by the main factors of noise and spectral resolution – which share approximately the same value. The ANOVA table of the GML algorithm applied to the Rochester scene, using the initial GML results as ground truth, show that the significance of factors can be ranked as follows: spatial resolution, spectral resolution and noise. Given that the ranking from these two ANOVA tables is the same, one could conclude that the GML algorithm behaves similar with respect to ground truth as it does when looking at how it degrades with respect to itself. This has been previously hinted at already. It should be noted that all main factors share approximately equal F-statistic values when using the initial GML product as ground truth and thereby *could* be considered of equal significance to the outcome of the GML algorithm. Despite what spectral product was used as ground truth reference, in either case, the interaction between spectral resolution and noise was of notable significance and ranked as fourth.

Immediate inspection of the ANOVA tables produced by applying the GML algorithm to the Western rainbow scene show that all of the main factors are significant contributors to GML performance. Regardless of whether the material map or the non-degraded GML product was used as ground truth the ranking of significant factors was the same. This ranking of significant contributors to GML performance is as follows: spatial resolution, spectral resolution and noise. Again, the similar results between selected ground truth references suggest that the GML algorithm degrades with respect to itself in a very similar fashion as it does with respect to selected ground truth. It should also be noted that in <u>all</u> cases and with <u>all</u> images used, the main factor of spatial resolution has consistently been the most significant contributing factor to GML performance. This was initially discovered when looking at the results shown in Figure 4-4. However, it was also discussed previously that this might be attributed to using the same ROIs throughout the experiment. Again, more investigation is needed to see whether in fact spatial resolution is the most significant factor to GML performance as is strongly indicated by this study.

The next portion of our analysis was using kappa to statistically compare the classification products from two different degraded images. This permits us to track the accuracy of an algorithm as the images are degraded with different levels of spatial resolution, spectral resolution and noise. For example, we can statistically test whether there is a difference in the SAM thematic map produced by an image with a spectral resolution of 20 m and spectral resolution of 10 nm and the SAM product from the same image with resolutions of 40 m and 55 nm respectively. This is done by using equation 2-40 from Chapter 2. The results of this statistical comparison are presented in Appendix F.

However it is worth mentioning here that the results from this type of statistical comparison were not as remarkable as expected. It was anticipated that definite patterns would be apparent. Immediately these patterns would be easily attributed to spatial resolution, spectral resolution and noise changes. Patterns were noticed but they were not as prominent as first anticipated. However, what was *not* expected is the most interesting result of all. From an examination of the tables in Appendix F, it can be generally concluded that any degradation of the initial image will result in an immediate difference in the information conveyed by the produced thematic map compared to the original. Likewise, when one "degraded" classification product was statistically compared to another "degraded" classification product the null hypothesis (see section 2.7.4) was rejected the majority of the time at the lowest confidence limit tested. This means that there is an immediate difference between classification maps produced from an image at one level of degradation compared to another. In other words, an image acquired with a certain spectral resolution, spatial resolution and noise characteristics will produce a different thematic map than the same image collected with different acquisition parameters. This makes sense.

4.3   Linear Spectral Unmixing

The Sum-of-Squared Error (depicted here as simply SE) was calculated using equation 2-41 from Chapter 2. The squared error is calculated from the fraction maps produced by unmixing the degraded imagery with respect to some reference fraction map. The references for the AVIRIS scenes are the fraction maps of the untouched initial images.   The reference for the DIRSIG image is the material truth map converted into a fraction map. Obviously the reference never changes for these calculations. Summaries of the SE data for each scene and each level of degradation are available in Appendix E. Presented below, in Table 4-10 are the mean and standard deviations of the SE at each level of degradation. Again, the first number listed under the parameter columns refers to the degradation levels of the AVIRIS images while the second number refers to the DIRSIG scene. If there is only one number than the level of degradation is shared between the two image sets.

	Rogers Dry Lake Unmixing - SE		Rochester Unmixing - SE		DIRSIG - West Rain Unmixing - SE	
spatial (m)	mean	stdev	mean	stdev	mean	stdev
20 / 2	5070.5721	10227.7361	341.2885	691.3684	0.4910	0.7183
40 / 4	5628.4617	11325.8693	381.1470	771.4130	0.5800	0.8364
80 / 8	7142.9845	14297.4494	491.3965	991.2883	0.8828	1.3511

	Rogers Dry Lake Unmixing - SE		Rochester Unmixing - SE		DIRSIG - West Rain Unmixing - SE	
Spectral (nm)	mean	stdev	mean	stdev	mean	stdev
10	11506.4605	17294.7267	835.8456	1259.8336	0.5387	0.8808
55	5684.0691	8568.3167	218.3019	329.1770	0.8028	1.2573
110/ 113	651.4887	982.9718	159.6845	240.9147	0.6123	0.8587

	Rogers Dry Lake Unmixing - SE		Rochester Unmixing - SE		DIRSIG - West Rain Unmixing - SE	
Noise(SNR@30%)	mean	stdev	mean	stdev	mean	stdev
10	17688.2722	14395.9071	1203.2343	996.4574	1.8809	0.7642
100	142.0403	116.3830	9.7231	8.0684	0.0432	0.0071
225 / 200	11.7058	12.1875	0.8744	0.8431	0.0296	0.0052

Table 4-10: Mean and Standard Deviation of Squared Error Metric for Linear Unmixing

Similar to the classification results, we see that the SE associated with low SNR dominates the means and standard deviations in the above table. This can be confirmed by an examination of the raw data available in Appendix E. For example, consider the Rogers Dry Lake image, the mean SE with of image products with a spatial resolution of 20 m is 5070.5721 with a standard deviation of 10227.7361. Looking at the raw data in Appendix E, we see that the calculations of these statistics cover a range of numbers from 0.5394 to 29356.900. The low values of SE are associated with those 20 m images with a high SNR (low noise content) and the high value of SE are associated with the 20 m images with a very low SNR (high noise content). This behaviour is also extended to the calculation of factor means for spectral resolution and is also witnessed in each image used. Obviously, there is an inversely proportional relationship between SNR and SE – that is as SNR decreases the SE resulting from unmixing with respect to some reference will increase. As we would expect, as the spectra within the image are degraded by adding increasing amounts of noise it would become more difficult to discriminate between materials because their spectra are changing. Despite the fact that noise dominates our mean and standard deviation calculations we are still able to note some interesting trends.

As we can see in the first portion of Table 4-10 (SE averages at different spatial resolutions) that as we degrade the spatial resolution of the image the SE increases. In other words, as the GIFOV increases the error (SE) made from unmixing also increases. This makes sense because as we increase the ground spot size/GIFOV we also increase the number of endmembers within each pixel. This was seen in the visual example of spatial degradation provided in Figure 3-3. This obviously leads to more mixed pixels in the image which will inevitably increase the complexity of the unmixing process, thereby leading to higher values of SE. We have already discussed the deteriorating effect of noise on the SE results. The trends discussed above can be better seen in Figures 4-11 4-12 and 4-13 below. The difference in magnitude of the SE between scenes will be discussed shortly.

Most interesting, are the SE results with respect to changes in spectral resolution. It is apparent from Table 4-20 and Figure 4-12 that as we degrade the spectral resolution the SE actually improves by getting smaller. This is counter intuitive and goes against the entire principle behind using high spectral resolution data to better discriminate between different materials. These results do not agree with the previous results of Konno (1999) in which he found that finer spectral resolution leads to an overall decrease in SE. However, a difference may be that he was working with a step-wise unmixing algorithm and not a linear spectral unmixing

algorithm like that tested here. This strange behaviour may be attributed to the "relative insensitivity that linear spectral unmixing has to subtle absorption features that result in quantification errors due to endmember variability in a pixel from linear and non-linear mixtures (e.g. scattering and lighting effects)" (Pinzon et al, 1998). That is, the naturally occuring variability of endmembers due to differing lighting effects and scattering are not accounted for in the simplicity of the linear spectral unmixing algorithm. Another explanation to this peculiarity is due to the approach adopted in the degradation of the image. Changing the spatial resolution or GIFOV by a pixel aggregation process, like adopted here and discussed in Chapter 3, works quite well. However, "there is no guarantee that the initial material fractions are maintained" from the initial image pixels and carried over to the new aggregated pixel (Keller et al, 2000) after spatially resampling the image. Either way, these spectral degradation results are counter intuitive and unreliable for any interpretation.
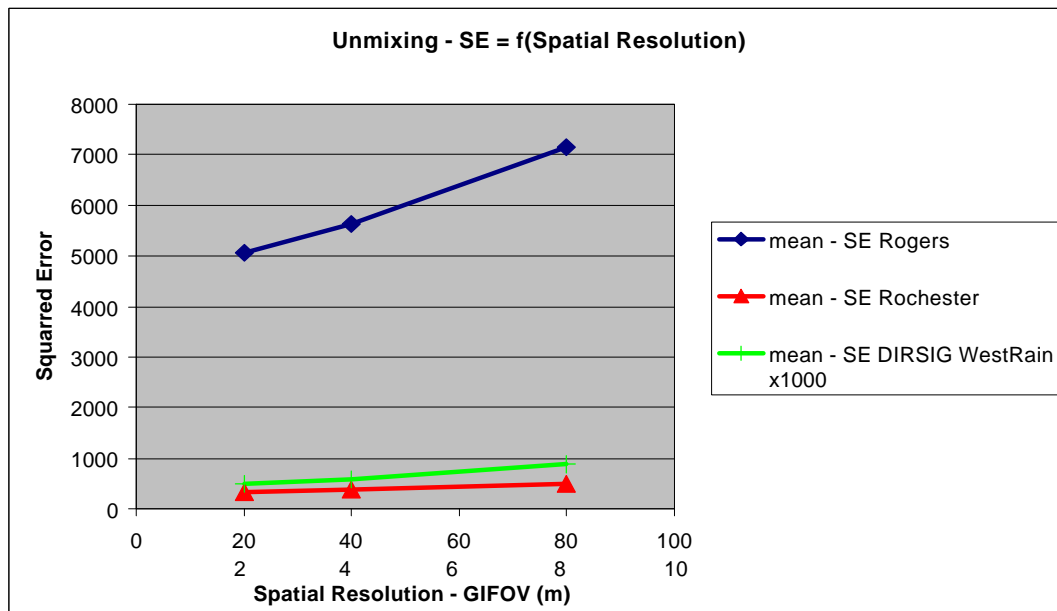
Figure 4-11: Linear Spectral Unmixing Results – Squared Error as a function of Spatial Resolution
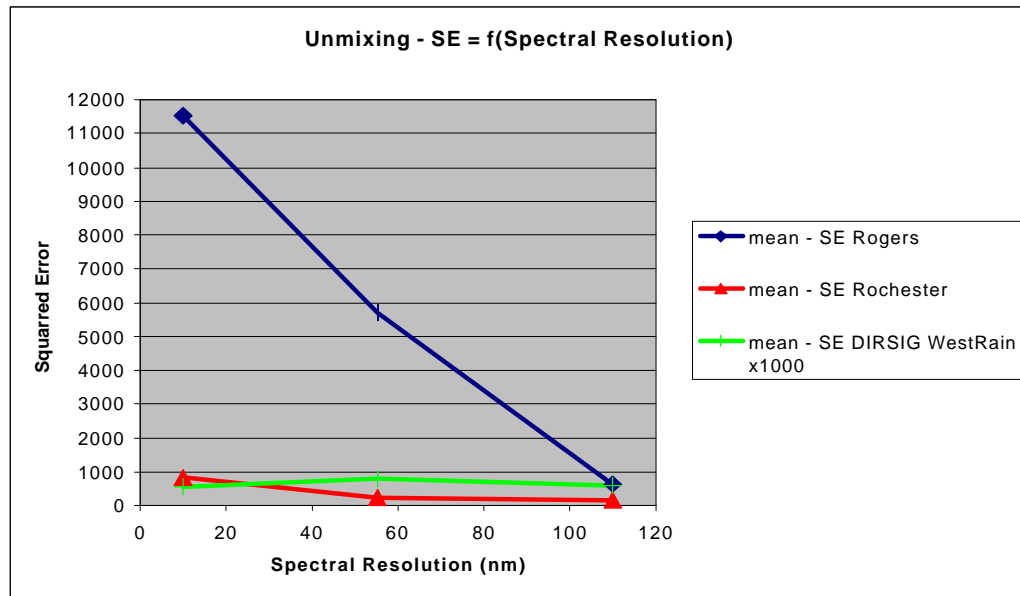
Figure 4-12: Linear Spectral Unmixing Results – Squared Error as a function of Spectral Resolution
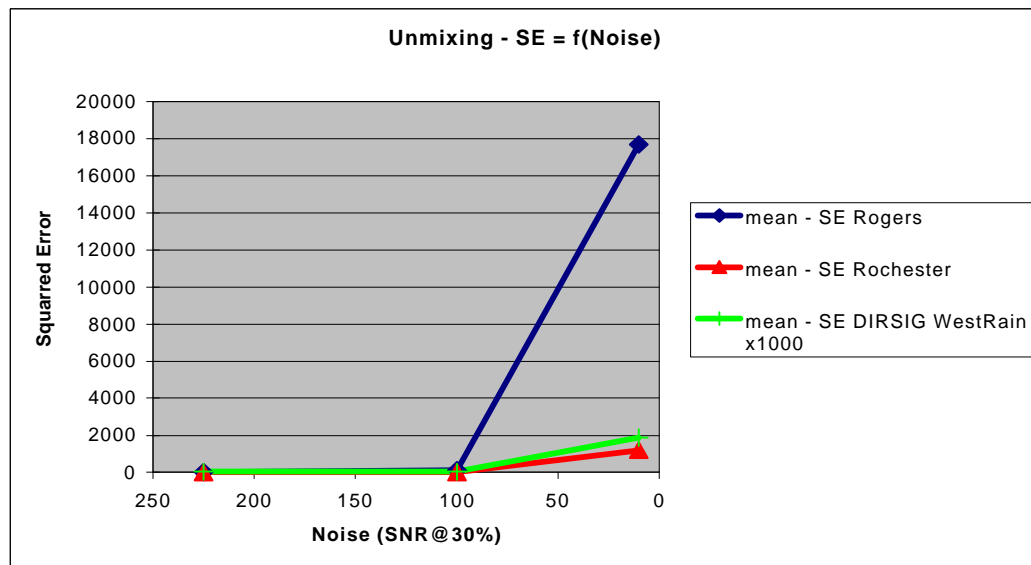
Figure 4-13: Linear Spectral Unmixing Results – Squared Error as a function of Noise

There are two noticeable differences upon inspection of Figures 4-11, 4-12, and 4-13. The first difference is the difference in SE magnitude between the three scenes. For example, the value of SE for the Rogers Dry Lake scene is larger in all three plots in comparison to the other scenes. This is because this scene is spectrally homogeneous and discriminating between its endmembers is more difficult than discriminating between the spectrally diverse endmembers of

the Rochester scene. The value of SE for the DIRSIG scene was multiplied by 1000 so that it could be plotted on the same charts. The SE for the DIRSIG scene is so low because this scene begins with a GIFOV of 2 m and the image is degraded to a maximum GIFOV of 8 m (in comparison with 20 m and 80 m). The very low SE can also be explained by the characteristic nature of DIRSIG in which each pixel only contains one spectral endmember. Nevertheless, the trends are much the same between real and synthetic imagery.   The second main difference is the magnitude of SE contributed by each main effect. This can also be seen in Table 4-20 from which this data was plotted. Using the Rochester scene as an example, we see that the range of SE means attributed to changes to spatial resolution is 341.2885 to 491.3965. However changes in the SE attributed to a change in spectral resolution or noise have a more profound effect as the SE range is 159.6845 to 835.8456 and 0.8744 to 1203.2343 respectively. This shows that although changing the spatial resolution of an image will degrade the information derived from unmixing, it will not have as profound an effect as a change in the spectral resolution or noise because it has a tighter range.   This behaviour is also apparent in the other scenes tested. We can further see the impact of each main factor and their interactions by examining the ANOVA tables.

```
Analysis of Variance for Unmixing – Rogers Dry Lake

Source             DF      Seq SS      Adj SS      Adj MS       F      P
Spatial             2    20699739    20699739    10349869    4.05   0.061
Spectral            2   531172567   531172567   265586284  103.92   0.000
Noise               2  1861044731  1861044731   930522366  364.09   0.000
Spatial*Spectral    4    10587284    10587284     2646821    1.04   0.446
Spatial*Noise       4    40019184    40019184    10004796    3.91   0.048
Spectral*Noise      4  1035121691  1035121691   258780423  101.25   0.000
Error               8    20446221    20446221     2555778
Total              26  3519091416
```

Table 4-11: Analysis of Variance for Unmixing – Rogers Dry Lake
($F_{0.05, 2, 8} = 4.46$ and $F_{0.05, 4, 8} = 3.84$)

Using the statistical hypothesis testing strategy outlined in section 4.1, we see that the main factor of spatial resolution and any interaction it may have with other factors have no significance in determining the outcome of linear spectral unmixing (by measure of SE) at the level of significance tested. This does not mean to say that spatial resolution does not matter, it only says that spatial resolution is not as significant as the other two factors or any other joint effect. This is primarily true because the AVIRIS pixels are initially quite mixed with a 20 m GIFOV. We also see that the main effect of noise carries the most significance (largest F-statistic)

to the outcome of linear unmixing. The main effect of spectral resolution and its joint effect with noise have equally weighted significance.

```
Analysis of Variance for Unmixing – Rochester, NY


Source              DF      Seq SS      Adj SS      Adj MS        F       P

Spatial              2      108828      108828       54414     3.76   0.071
Spectral             2     2525970     2525970     1262985    87.25   0.000
Noise                2     8610650     8610650     4305325   297.42   0.000
Spatial*Spectral     4       59977       59977       14994     1.04   0.445
Spatial*Noise        4      210126      210126       52532     3.63   0.057
Spectral*Noise       4     4923239     4923239     1230810    85.03   0.000
Error                8      115805      115805       14476
Total               26    16554594
```

Table 4-12: Analysis of Variance for Unmixing – Rochester
($F_{0.05, 2, 8} = 4.46$ and $F_{0.05, 4, 8} = 3.84$).

Similar to the Rogers Dry Lake image, we see here with the Rochester scene (Table 4-22) that the main factor of spatial resolution and any interaction it may have with other factors has no significance at this test level ($\alpha = 0.05$). Inspection of the F-statistics shows that the main effect of noise contributes most to the degradation in performance of the linear spectral unmixing algorithm. Again, like the Rogers scene we see here with the Rochester scene that the main effect of spectral resolution and its interaction with noise have nearly equal significance.

```
Analysis of Variance for Unmixing – DIRSIG Western Rainbow


Source              DF      Seq SS      Adj SS      Adj MS        F       P
Spatial              2      0.7595      0.7595      0.3797     3.31   0.090
Spectral             2      0.3344      0.3344      0.1672     1.46   0.289
Noise                2     20.4142     20.4142     10.2071    88.95   0.000
Spatial*Spectral     4      0.4764      0.4764      0.1191     1.04   0.445
Spatial*Noise        4      1.5350      1.5350      0.3838     3.34   0.069
Spectral*Noise       4      0.6497      0.6497      0.1624     1.42   0.312
Error                8      0.9180      0.9180      0.1147
Total               26     25.0872
```

Table 4-13: Analysis of Variance for Unmixing – DIRSIG Western Rainbow
($F_{0.05, 2, 8} = 4.46$ and $F_{0.05, 4, 8} = 3.84$)

Examination of the ANOVA table (Table 4-23) for unmixing the DIRSIG scene shows, that like the AVIRIS scenes, noise is the most significant factor on the degradation of information

from linear spectral unmixing as measured by SE. Interestingly, all other factors are insignificant with respect to this scene. However at a different significance level ($P = \alpha = 0.09$) the main factor of spatial resolution would become significant prior to the main effect of spectral resolution. This is quite different from the previous AVIRIS scenes in which spatial resolution had a very low F-statistic associated with it in comparison to the other factors. This tends to indicate that if the starting resolution is very fine (like 2 m) then any degradation from this resolution has a more profound effect than any other main factor in comparison to degrading an image with a larger GIFOV. This makes perfect sense since the AVIRIS scenes initially begins with mixed pixels, because of a relatively large GIFOV (20 m), and the DIRSIG generated scene is comprised of smaller spectrally pure pixels with a 2 m GIFOV. Furthermore, even if the smaller pixels were spectrally mixed in the DIRSIG scene it follows that with a smaller GIFOV it is more likely that each pixel contains very few endmembers compared to an image with a larger GIFOV. The numbers presented here would tend to indicate this theory holds true, but more testing would be needed to make any valid conclusions. Another suitable, but not as important, reason that this difference may exist is that the DIRSIG image is in reflectance space and not plagued by atmospheric absorption bands like the AVIRIS images. The endmembers in reflectance space are typically more spectrally distinct since their spectral shape is not defined by the atmosphere and the general shape of the exoatmospheric irradiance. In this case, spectral resolution is not as significant as spatial resolution.

4.4   Target Detection Algorithms

The target detection routines tested were the Spectral Matched Filter (SMF) and Spectral Feature Fitting (SFF) algorithms. The metric used to measure the performance of these algorithms was the ROC curve. More specifically the metric employed was the probability of detection read from these curves at a fixed probability of false alarm. As mentioned in Chapter 3 and seen in Tables 3-2 and 3-3, three different endmembers from each scene were used as input targets. Each one of these targets will produce an individual ROC curve. A weighted average of the three P_d's at a common P_{fa} was calculated and used as the final metric. This approach was adopted since this final metric, an average probability of detection, is a better representation of target detection performance for this study than merely one target per scene. Unfortunately, the targets are not the same in each scene. The groundtruth/references for the AVIRIS scenes were the SAM results of the initial untouched image. The groundtruth/references for the DIRSIG image was the material truth map. It should also be noted that the fixed P_{fa} chosen was as low as possible while still enabling meaningful readings from the ROC curves. Readings at two P_{fa}'s were taken so as to better characterize the curve and to also see if the significance of spatial resolution, spectral resolution and noise are dependent on where the ROC curve readings were taken. Summaries of the final weighted P_d's at fixed P_{fa} data for each target detection algorithm, each scene and level of degradation are available in Appendix E.

Spectral Matched Filter (SMF) Results

Presented below, in Table 4-24 are the mean and standard deviations of the P_d at each level of degradation for the SMF results from each scene tested.

spatial (m)	Rogers Dry Lake SMF - PD @ PFA =0.01		Rogers Dry Lake SMF - PD @ PFA =0.001		Rochester SMF – PD @ PFA =0.01		Rochester SMF – PD @ PFA =0.001	
	mean	stdev	mean	stdev	mean	stdev	mean	stdev
20	0.5878	0.0736	0.4135	0.1347	0.5274	0.0910	0.2726	0.0467
40	0.5673	0.0744	0.4008	0.1408	0.5094	0.0952	0.2425	0.0492
80	0.5338	0.0798	0.3780	0.1475	0.4659	0.0853	0.1847	0.0430

spectral (nm)	Rogers Dry Lake SMF - PD @ PFA =0.01		Rogers Dry Lake SMF – PD @ PFA =0.001		Rochester SMF – PD @ PFA =0.01		Rochester SMF - PD @ PFA =0.001	
	mean	stdev	mean	stdev	mean	stdev	mean	stdev
10	0.5958	0.0235	0.4623	0.0105	0.5984	0.0356	0.2837	0.0388
55	0.5671	0.0527	0.3879	0.1219	0.4649	0.0743	0.2173	0.0502
110	0.5260	0.1136	0.3420	0.1929	0.4394	0.0606	0.1988	0.0494

Noise(SNR@30%)	Rogers Dry Lake SMF - PD @ PFA =0.01		Rogers Dry Lake SMF - PD @ PFA =0.001		Rochester SMF – PD @ PFA =0.01		Rochester SMF – PD @ PFA =0.001	
	mean	stdev	mean	stdev	mean	stdev	mean	stdev
10	0.4934	0.0996	0.2586	0.1660	0.4479	0.1177	0.2064	0.0690
100	0.5974	0.0231	0.4679	0.0149	0.5127	0.0694	0.2428	0.0500
225	0.5981	0.0199	0.4657	0.0137	0.5421	0.0545	0.2506	0.0493

spatial (m)	DIRSIG - West Rain SMF - PD @ PFA =0.001		DIRSIG - West Rain SMF - PD @ PFA =0.0001	
	mean	stdev	mean	stdev
2	0.7739	0.3565	0.5885	0.4035
4	0.6922	0.3752	0.3378	0.2607
8	0.5394	0.2876	0.1221	0.1557

spectral (nm)	DIRSIG - West Rain SMF - PD @ PFA =0.001		DIRSIG - West Rain SMF - PD @ PFA =0.0001	
	mean	stdev	mean	stdev
10	0.8708	0.1188	0.5468	0.2948
55	0.5929	0.3743	0.3339	0.3621
113	0.5418	0.3972	0.1678	0.2800

Noise(SNR@30%)	DIRSIG - West Rain SMF - PD @ PFA =0.001		DIRSIG – West Rain SMF - PD @ PFA =0.0001	
	mean	stdev	Mean	stdev
10	0.3189	0.3704	0.1417	0.2820
100	0.8305	0.1553	0.3582	0.3541
200	0.8561	0.1208	0.5486	0.2819

Table 4-14: Mean and Standard Deviation of Weighted P_d for Spectral Matched Filter

Visual inspection of the numbers contained in the above table reveals no surprises or peculiar trends. All results meet our intuitive expectations and may be confirmed with previous studies. It is evident that as any of the main factors are degraded – be it spatial resolution, spectral resolution or noise – the P_d suffers from these degradations. Also according to expectation, a lower selected value of P_{fa} results in a lower P_d. This is caused by the distinctive ROC curve shape discussed in Chapter 2. To better illustrate and discuss these trends further the data from Table 4-24 is plotted below.
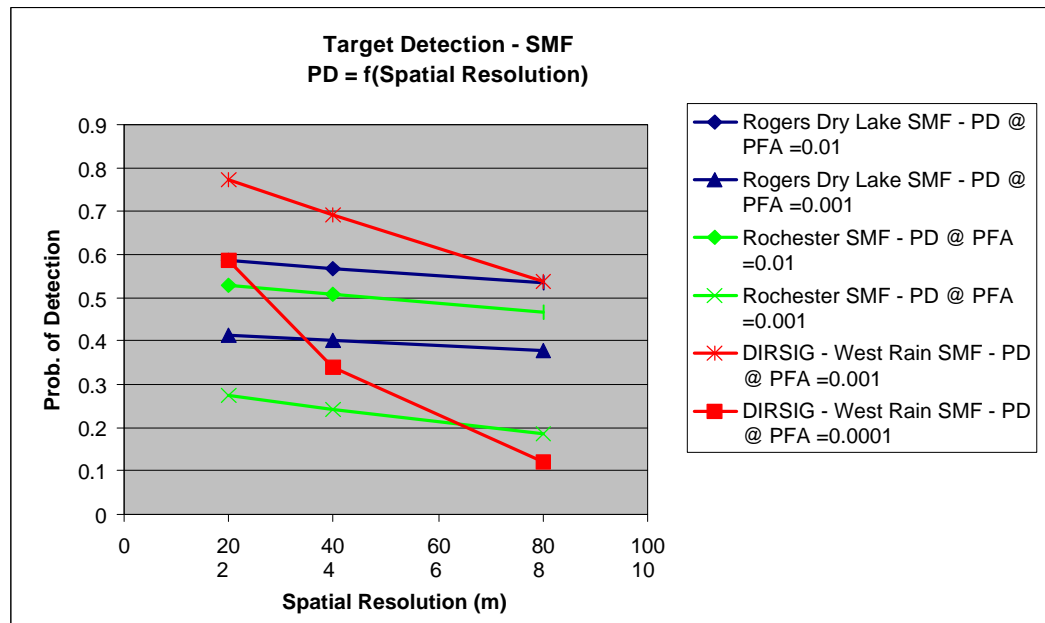
Figure 4-14: Spectral Matched Filter – Probability of Detection as a function of Spatial Resolution

From Figure 4-14, we see that the probability of detecting targets decreases somewhat monotonically as the GIFOV is increased. This agrees with previous studies of this nature (Keller et al., 2000). The probability of detection for the Rogers Dry Lake scene is consistently higher than the Rochester scene. The scene dependence of the result and metric can be attributed to the fact that the Rogers Dry Lake image is spectrally and spatially homogeneous. Therefore we expect better results when attempting to find a distinct spectral target within this image then when posed with a more spatially and spectrally complex image. It should be noted that the selected P_{fa}'s for the two AVIRIS images are the same. *Lower P_{fa}'s (0.001 and 0.0001) were chosen for*

the DIRSIG scene. Retaining the same P_{fa} for the DIRSIG images as the AVIRIS imagery would have resulted in a constant P_d near unity – thereby not allowing any observations. This is to be expected given the smaller GIFOV and spectrally pure pixels characteristic of DIRSIG scenes. Lower values for P_{fa}'s were chosen to hopefully witness the same behaviour as that with real imagery and be able to draw valid conclusions. Nevertheless, the same monotonic demise in P_d is witnessed with the synthetic imagery at an appropriate P_{fa}. It is interesting to note that requiring a lower P_{fa} in the DIRSIG scene results in a sharper P_d decrease with respect to changes in spatial resolution. It is possible that this may have also been the same case with the AVIRIS images, but any lower than a P_{fa} equal to 0.001 would have resulted in many P_d readings equal to zero – and meaningless results. It is also interesting to note the sharper slopes associated with the synthetic image in comparison to the real images. Like the results from unmixing, this may be attributed to the fact that with this synthetic image we are starting with pixels with a much smaller GIFOV, they are essentially spectrally pure without complex spatial-spectral variability and the scene is in reflectance space vice radiance.
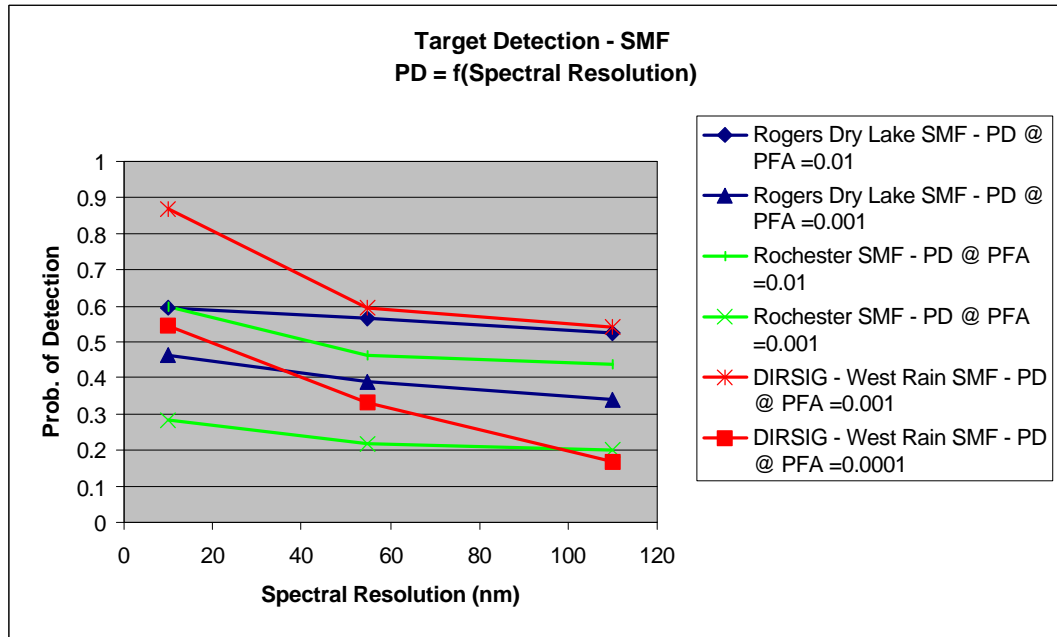
Figure 4-15: Spectral Matched Filter – Probability of Detection as a Function of Spectral Resolution

At first glance it appears that Figure 4-15, depicting probability of detection as a function of spectral resolution, is a duplicate of the chart depicting detection as a function of spatial

resolution (Figure 4-14). There is a monotonic decrease in detection performance, as the spectral sampling of an image becomes coarser. This makes sense since many of the fine absorption features that allow us to identify and discriminate among materials are lost when we degrade the spectral resolution.   Again we note a difference between the spectrally homogeneous scene (Rogers Dry Lake) versus a more complex scene (Rochester). That is, the spectral contrast between target and background is more pronounced in a homogeneous scene – thereby leading to higher P_d. Similar to the results from spatial resolution, the DIRSIG image shows a more dramatic drop in performance as the spectral resolution is widened. Again this can be attributed to a smaller initial GIFOV, the spectral purity of the pixels and the underlying difference between scene-derived endmembers that are in radiance space (like the AVIRIS scene) and endmembers from true reflectance values.

Target Detection - SMF PD = f(Noise)

Legend:
- Rogers Dry Lake SMF - PD @ PFA =0.01
- Rogers Dry Lake SMF - PD @ PFA =0.001
- Rochester SMF - PD @ PFA =0.01
- Rochester SMF - PD @ PFA =0.001
- DIRSIG - West Rain SMF - PD @ PFA =0.001
- DIRSIG - West Rain SMF - PD @ PFA =0.0001

Y-axis: Prob. of Detection
X-axis: SNR @ 30% reflector
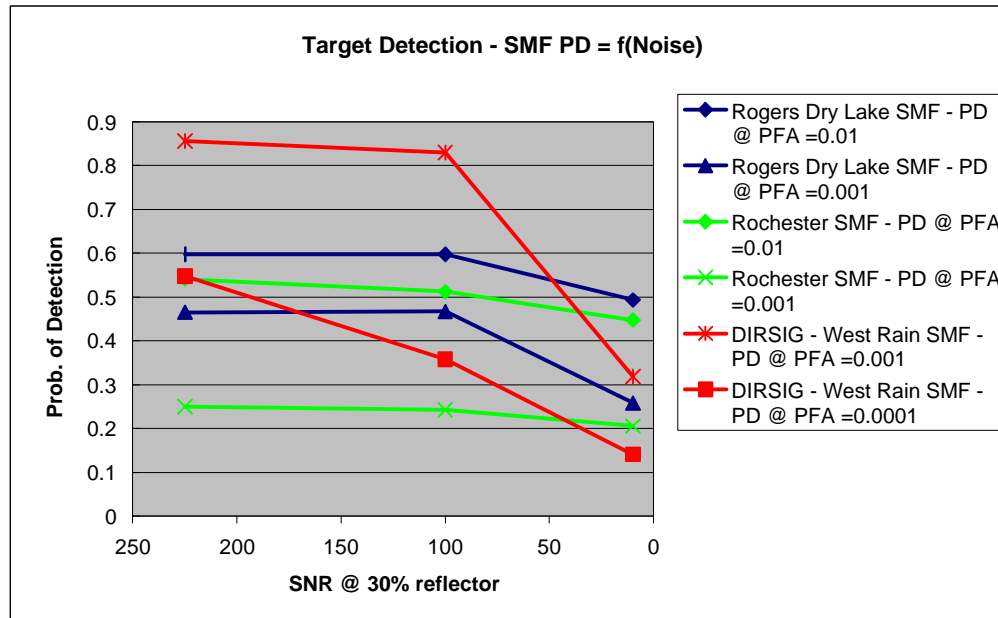
Figure 4-16: Spectral Matched Filter – Probability of Detection as a Function of SNR

It is interesting to note that the probability of detection as a function of SNR does not have a monotonic decrease with respect to its degradation like spatial resolution and spectral resolution did. This can be seen in Figure 4-16 and by confirmation of the numbers in Table 4-24.   Unfortunately, there are only three points on any of the above plots. Notwithstanding this fact, it would appear from the above plot that the probability of detection remains relatively constant for SNR between 225 (or 200 with the DIRSIG scene) and 100. It is only at some point

between a SNR of 100 and 10 that the probability of detection begins to decrease. This type of behaviour is evident with each image tested.   The only exception is the second DIRSIG result with a P_d acquired at a very low P_{fa} of 0.0001. In this regard, it is quite possible that a re-read of P_d at lower P_{fa} is required for these scenes. Conversely, if we remember from section 2.4.3, the basis behind the OSP or SMF is that an operator q is applied to equation 2-7 that represents the contents of the image cube. This operator effectively not only suppresses the background spectral effects, represented by $U,$ via an orthogonal projection but "it also suppresses the original noise by Pn" (Chang and Ren, 2000). It is possible that this noise is effectively suppressed up to a certain SNR. After this point it begins to become more of a crucial factor in the detection performance of the algorithm.

Another interesting observation of noise with respect to SMF performance can be seen by an examination of the collected data in Appendix E (highlighted portions). Looking at the real image data we see that the value of P_d does not change with regard to noise as long as the image is at the highest spectral resolution. For example, looking at the Rogers Dry Lake scene, we see a constant P_d for all images with a spectral resolution of 10 nm - regardless of the noise content. Obviously, there are fluctuations due to changes in spatial resolution with this metric, but at each spatial resolution the constant prevails. This behaviour is also seen in the Rochester image, but not with the DIRSIG image. One could prematurely conclude that an analyst could expect the same results when attempting to find a target within an image regardless of the noise content as long as the image acquisition included very fine spectral sampling (and the spatial resolution remains constant). This observation only holds valid for the real imagery used and needs further testing.

```
Analysis of Variance for SMF PD @ PFA = 0.01 – Rogers Dry Lake

Source            DF    Seq SS      Adj SS      Adj MS        F       P
Spatial            2  0.0134157   0.0134157   0.0067079    93.75   0.000
Spectral           2  0.0222067   0.0222067   0.0111033   155.18   0.000
Noise              2  0.0652817   0.0652817   0.0326409   456.20   0.000
Spatial*Spectral   4  0.0000196   0.0000196   0.0000049     0.07   0.990
Spatial*Noise      4  0.0004301   0.0004301   0.0001075     1.50   0.289
Spectral*Noise     4  0.0500947   0.0500947   0.0125237   175.04   0.000
Error              8  0.0005724   0.0005724   0.0000715
Total             26  0.1520209

Analysis of Variance for SMF PD @ PFA=0.001– Rogers Dry Lake

Source            DF    Seq SS      Adj SS      Adj MS         F       P
Spatial            2  0.005824    0.005824    0.002912    100.27   0.000
Spectral           2  0.066273    0.066273    0.033137   1141.10   0.000
Noise              2  0.259943    0.259943    0.129971   4475.71   0.000
Spatial*Spectral   4  0.000571    0.000571    0.000143      4.91   0.027
Spatial*Noise      4  0.000386    0.000386    0.000097      3.33   0.070
Spectral*Noise     4  0.150458    0.150458    0.037614   1295.29   0.000
Error              8  0.000232    0.000232    0.000029
Total             26  0.48368
```

Table 4-15: Analysis of Variance for SMF – Rogers Dry Lake($F_{0.05, 2, 8} = 4.46$ and $F_{0.05, 4, 8} = 3.84$)

As seen in Table 4-25, the significance of factors (measured by the F-statistic) seems to be dependent upon the required P_{fa}. For instance, if the operator requires a smaller P_{fa} then the significance of both spectral resolution and noise effects increase by an order of magnitude when using the Rogers Dry Lake scene. However, the significance of the spatial resolution remains relatively constant as the P_{fa} is changed. This can be attributed to a general lack of spatial complexity in the Rogers scene. The joint effects between spatial-spectral and spectral-noise are insignificant (or very close) in at both P_{fa} readings.

```
Analysis of Variance for SMF PD @ PFA = 0.01

Source             DF      Seq SS     Adj SS     Adj MS        F       P
Spatial             2    0.017966   0.017966   0.008983    95.17   0.000
Spectral            2    0.131247   0.131247   0.065624   695.21   0.000
Noise               2    0.041819   0.041819   0.020909   221.51   0.000
Spatial*Spectral    4    0.000958   0.000958   0.000240     2.54   0.122
Spatial*Noise       4    0.000488   0.000488   0.000122     1.29   0.350
Spectral*Noise      4    0.021686   0.021686   0.005421    57.43   0.000
Error               8    0.000755   0.000755   0.000094
Total              26    0.214919

Analysis of Variance for SMF PD @ PFA = 0.001

Source             DF      Seq SS     Adj SS     Adj MS        F       P
Spatial             2   0.0359662  0.0359662  0.0179831   812.88   0.000
Spectral            2   0.0358327  0.0358327  0.0179164   809.87   0.000
Noise               2   0.0099963  0.0099963  0.0049982   225.93   0.000
Spatial*Spectral    4   0.0003383  0.0003383  0.0000846     3.82   0.050
Spatial*Noise       4   0.0001931  0.0001931  0.0000483     2.18   0.161
Spectral*Noise      4   0.0050580  0.0050580  0.0012645    57.16   0.000
Error               8   0.0001770  0.0001770  0.0000221
Total              26   0.0875616
```

Table 4-16: Analysis of Variance for SMF – Rochester
($F_{0.05, 2, 8} = 4.46$ and $F_{0.05, 4, 8} = 3.84$)

In table 4-26, we see an underlying dependence of factor significance on the P_{fa}. For example, with a $P_{fa} = 0.01$ the F-statistic for the main effect of spatial resolution is 95.17 and with a $P_{fa} = 0.001$ this same F-statistic jumps to 812.88. This would indicate that as we wish to detect a target with a lower P_{fa}, then the importance of finer spatial resolution dramatically increases when using the Rochester scene. However, one cannot make this conclusion for the all images. The significance of the main effect of spatial resolution did not change much from one P_{fa} to another with the Rogers Dry Lake scene. Again, this scene dependence on the level of the metric and therefore performance may be ascribed to the level of complexity within the scene and the similarity of endmembers. Spectral resolution has approximately the same weighting of importance with respect to detection at both P_{fa}. The interaction between spatial-spectral and spatial-noise is not significant at either P_{fa}. Furthermore, it is interesting in this image that the significance placed on the main effect of noise (221 and 225) and all two-way interactions remain relatively constant with changes in P_{fa}. Scene dependence seems to prevail.

```
Analysis of Variance for SMF PD @ PFA = 0.001 - DIRSIG Western Rainbow

Source             DF    Seq SS    Adj SS    Adj MS        F       P
Spatial             2   0.25507   0.25507   0.12754    24.49   0.000
Spectral            2   0.56437   0.56437   0.28218    54.19   0.000
Noise               2   1.65284   1.65284   0.82642   158.70   0.000
Spatial*Spectral    4   0.01510   0.01510   0.00377     0.72   0.599
Spatial*Noise       4   0.02198   0.02198   0.00549     1.06   0.437
Spectral*Noise      4   0.50925   0.50925   0.12731    24.45   0.000
Error               8   0.04166   0.04166   0.00521
Total              26   3.06026


Analysis of Variance for SMF PD @ PFA = 0.0001 - DIRSIG Western Rainbow

Source             DF    Seq SS    Adj SS    Adj MS        F       P
Spatial             2   0.98079   0.98079   0.49039    16.68   0.001
Spectral            2   0.64987   0.64987   0.32494    11.05   0.005
Noise               2   0.74606   0.74606   0.37303    12.69   0.003
Spatial*Spectral    4   0.09166   0.09166   0.02291     0.78   0.569
Spatial*Noise       4   0.10527   0.10527   0.02632     0.90   0.509
Spectral*Noise      4   0.21228   0.21228   0.05307     1.81   0.221
Error               8   0.23517   0.23517   0.02940
Total              26   3.02109
```

Table 4-17: Analysis of Variance for SMF – Western Rainbow
($F_{0.05, 2, 8} = 4.46$ and $F_{0.05, 4, 8} = 3.84$)

The results displayed in Table 4-27 does not seem to make any intuitive sense at first. However, closer inspection reveals some interesting facts. As seen in the two previous series of ANOVA tables, there is an underlying dependence on the significance of the factors with respect to the P_{fa}. The ANOVA table for the DIRSIG scene is no different. At a $P_{fa} = 0.001$, the main factor of noise is the most significant factor followed by the spectral resolution. The significance of spatial resolution is third. However, if an image analyst changes the desired P_{fa} to 0.0001 then the ranking of these main effects changes, as seen in Table 4-27. At this lower P_{fa} level, the main effect of spatial resolution becomes most significant - followed by spectral resolution and noise. A closer examination of the numbers in Table 4-27 shows that all factors share approximately equal weighting of significance. A similar change to the level of significance for the main effect of spatial resolution was also seen in the AVIRIS Rochester products as the value of P_{fa} changed. It is from this trend that one could draw a conclusion – a weak one based on only three observations, but a conclusion just the same. In general it appears that at certain P_{fa} levels the spectral resolution and noise dictate the performance of the SMF algorithm. However, as we require a lower P_{fa} the importance of fine spatial resolution overrides these other two main factors. This trend obviously needs further investigation to solidify such a conclusion.

Spectral Feature Fitting (SFF)

Prior to an examination of the results from the SFF data it needs to be mentioned that the entire image spectrum was not used in the application of this algorithm (unlike all the other algorithms tested). As we saw in Chapter 2, the SFF algorithm is an absorption-based method whose product results are directly linked to how well the absorption features of reference spectra match the absorption features of image spectra. If this method were applied directly to the AVIRIS real images, which are in radiance space with atmospheric bands still included, the produced detection results for each endmembers/targets selected would all be exactly the same. This is because the algorithm would match just the pre-dominate atmospheric absorption bands which are common to all spectral signatures in radiance space. This theory was tested and was proven correct. Two approaches could be adopted to solve this problem, besides the most obvious which is to convert the image into reflectance. The first solution was to remove the atmospheric bands in both the image and library. This was not done since the spectral shape is still governed by the exoatmospheric solar irradiance curve and the spectra now being discontinuous could effect proper continuum removal. The second, and chosen method, was to chose a common spectral subset from both the image and library that is relatively void of atmospheric bands (except 960 and 1140 nm) and shows the most spectral distinction between endmembers. The spectral range used was 385 to 1240 nm. This method was adopted for both AVIRIS scenes. To maintain experimental consistency with SMF, the entire spectral range is used for the DIRSIG scene since it has been calibrated to reflectance space and each endmember contains only one or two distinct absorption features. This brings up another important point. The algorithm requires that the scene first be calibrated to reflectance prior to continuum removal and scaling. As previously mentioned in here and in Chapter 2, this calibration was not done for any of the real imagery. This may help to explain some of the odd results from this algorithm. Presented below, in Table 4-28 are the mean and standard deviations of the P_d at each level of degradation.

spatial (m)	Rogers Dry Lake SFF - PD @ PFA =0.2		Rogers Dry Lake SFF - PD @ PFA = 0.1		Rochester SFF - PD @ PFA =0.6		Rochester SFF - PD @ PFA = 0.4	
	mean	stdev	mean	stdev	mean	stdev	mean	stdev
20	0.5914	0.1011	0.5168	0.1724	0.3813	0.0908	0.0909	0.0746
40	0.5790	0.1205	0.5036	0.1716	0.3648	0.0845	0.0938	0.0791
80	0.5546	0.1473	0.4690	0.1811	0.3406	0.0910	0.0958	0.0791

spectral (nm)	Rogers Dry Lake SFF - PD @ PFA =0.2		Rogers Dry Lake SFF - PD @ PFA = 0.1		Rochester SFF - PD @ PFA =0.6		Rochester SFF - PD @ PFA = 0.4	
	mean	stdev	mean	stdev	mean	stdev	mean	stdev
10	0.5147	0.1618	0.4329	0.2243	0.4077	0.1041	0.0923	0.0757
55	0.5693	0.1075	0.4869	0.1663	0.3290	0.0861	0.0904	0.0850
110	0.6410	0.0220	0.5697	0.0686	0.3501	0.0512	0.0978	0.0714

Noise(SNR@30%)	Rogers Dry Lake SFF - PD @ PFA =0.2		Rogers Dry Lake SFF - PD @ PFA = 0.1		Rochester SFF - PD @ PFA =0.6		Rochester SFF - PD @ PFA = 0.4	
	mean	stdev	mean	stdev	mean	stdev	mean	stdev
10	0.4493	0.1412	0.2976	0.1611	0.4655	0.0606	0.1965	0.0077
100	0.6366	0.0195	0.5957	0.0172	0.3110	0.0387	0.0462	0.0060
225	0.6391	0.0159	0.5961	0.0200	0.3103	0.0383	0.0379	0.0085

spatial (m)	DIRSIG - West Rain SFF - PD @ PFA =0.2		DIRSIG - West Rain SFF - PD @ PFA = 0.05	
	mean	stdev	mean	stdev
20	0.6185	0.3339	0.2679	0.1052
40	0.4826	0.2522	0.0862	0.0540
80	0.2640	0.1961	0.0302	0.0362

spectral (nm)	DIRSIG - West Rain SFF - PD @ PFA =0.2		DIRSIG - West Rain SFF - PD @ PFA = 0.05	
	mean	stdev	mean	stdev
10	0.2654	0.1686	0.1191	0.1440
55	0.5048	0.3011	0.1698	0.1248
110	0.5948	0.3181	0.0954	0.1025

Noise(SNR@30%)	DIRSIG - West Rain SFF - PD @ PFA =0.2		DIRSIG - West Rain SFF - PD @ PFA = 0.05	
	mean	stdev	mean	stdev
10	0.2101	0.1147	0.1090	0.0964
100	0.5432	0.2848	0.1375	0.1391
225	0.6117	0.2957	0.1377	0.1442

Table 4-18: Mean and Standard Deviation of Weighted P_d for Spectral Feature Fitting

One of the first observations from Table 4-28 is that the probability of detection is read at different probability of false alarms for each image. Common P_{fa} could not be selected for the two AVIRIS images as was done with the assessment of the SMF algorithm. This is primarily due to the fact that the ROC curves associated with each scene were quite different in shape. As seen in Table 4-28, the P_d in the Rochester scene at a $P_{fa} = 0.4$ is very low - numbers less than or around 0.1. Applying this same P_{fa} to the Rogers scene resulted in P_d of 1.0 for all endmembers. A lower P_{fa} for the Rogers Dry Lake image was applied to better observe its performance. One could say that the use of different P_{fa}'s between scenes does not allow valid observations with respect to scene dependence. While this is somewhat true, the converse to this is also very true. The simple fact that a higher P_{fa} was needed to acquire P_d readings is a direct reflection of the algorithms poor performance as spectral and spatial complexity of the scene increase. The same P_{fa}'s used to assess the SMF algorithm were applied to the SFF algorithm and this resulted in all P_d's equal to zero. Obviously, SMF outperforms the SFF when detecting the same targets. Again, the lower P_{fa}'s were used with the DIRSIG scene than the real images. To better illustrate and discuss these trends the following data from Table 4-28 is plotted below.

As seen in Figure 4-17 below, the probability of detecting targets decreases monotonically as the GIFOV is increased. This completely agrees with our expectation and the results of the SMF algorithm. As observed previously with the SMF results, the probability of detection for the Rogers Dry Lake scene is consistently higher than the Rochester scene even with a lower P_{fa}. Like the other algorithms tested, the difference between homogeneous and complex images has an apparent bearing on the performance of this algorithm. Although lower P_{fa}'s are chosen for the DIRSIG scene the same decrease in P_d is witnessed. Similar to the SMF results, the SFF results show a steeper slope for the detection performance in the DIRSIG scene as the spatial resolution is degraded. Again this may attributed to initial GIFOV differences, spectrally pure pixels and working in reflectance space vice in units of radiance.
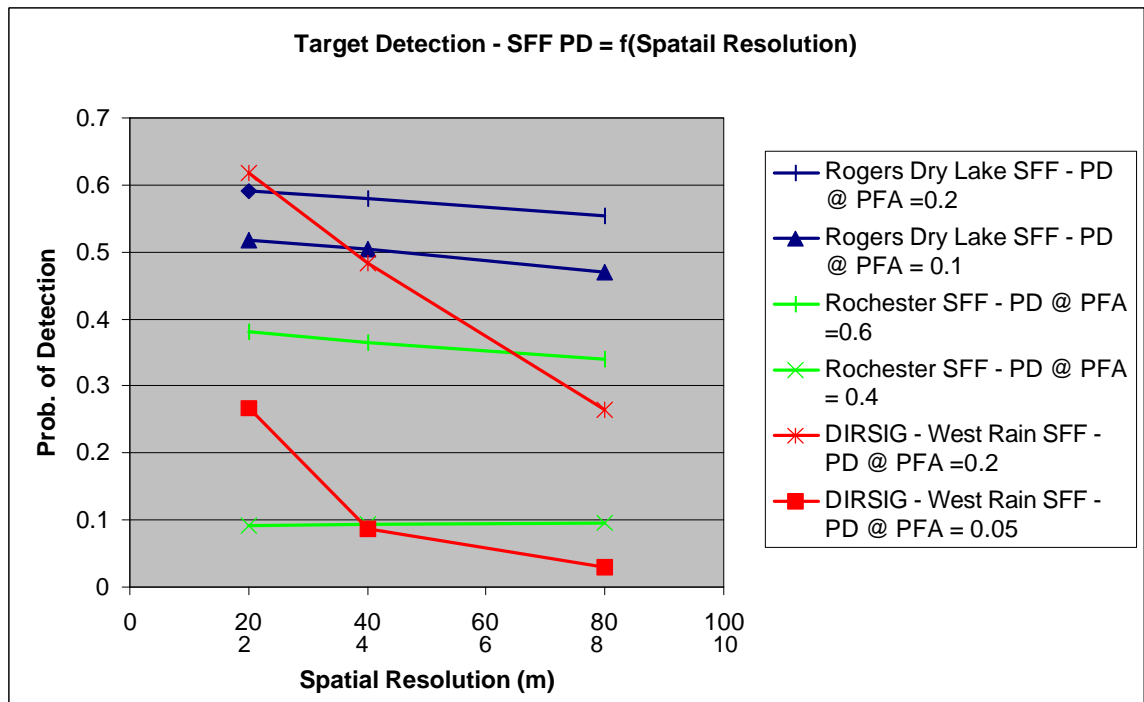
Figure 4-17: Spectral Feature Fitting – Probability of Detection as a function of Spatial Resolution
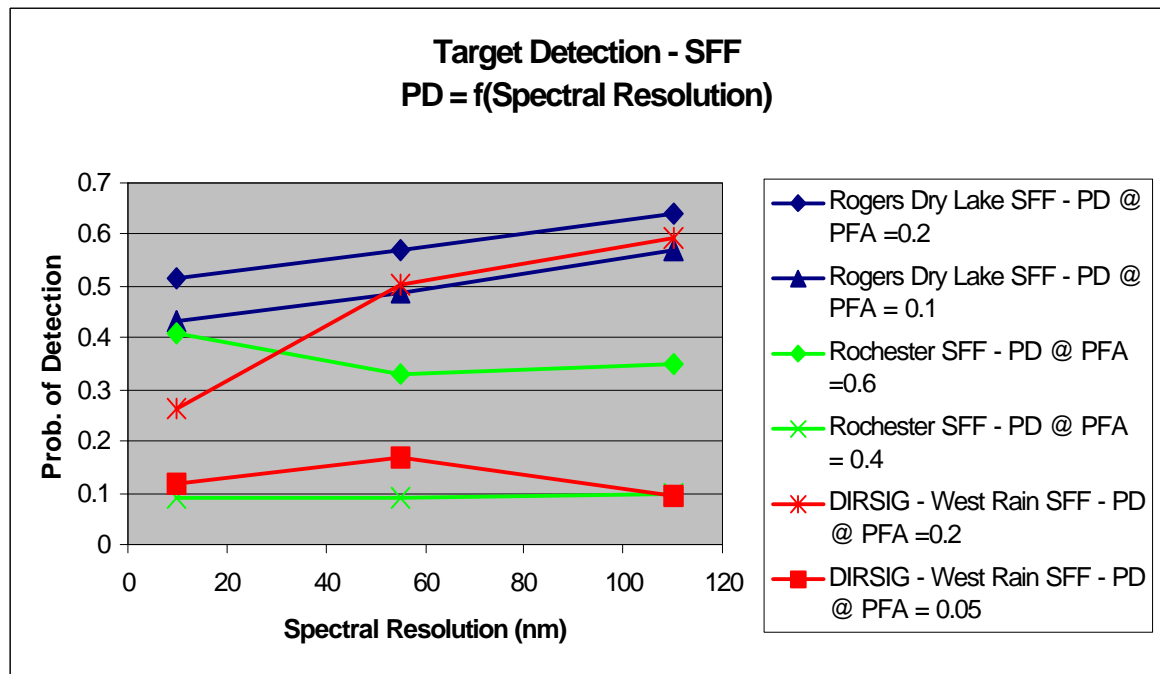
Figure 4-18: Spectral Feature Fitting – Probability of Detection as a Function of Spectral Resolution

The general trend presented in Figure 4-18 would indicate that as spectral resolution worsened the probability of detection increased. This seems to be counter-intuitive but closer investigation reveals that this phenomenon is somewhat valid. For example, when considering the Rogers Dry Lake scene we see that as spectral resolution is degraded the probability of detection increases. This means that better matches of absorption features were attained by the SFF at these lower resolutions. This inherently means that the absorption features must be more distinct after a continuum removal is performed on these lower resolution spectra. As mentioned previously, the spectral range of the real image was reduced to a range of 385-1240 nm. Admittedly, this range still contains two absorption features (at 960 and 1140 nm) but these were left in so a better comparison could be drawn between the SFF and SMF algorithm. All the atmospheric bands were left in place when the SMF was run. It is readily apparent in Figure 4-19 that the absorption features of the specified targets become more distinct with spectral degradation and continuum removal. Many of the small non-influential absorption features are smoothed out and the distinction between mineral and field becomes more prominent especially between 680-980 nm. The result is better detection as spectral resolution is degraded.
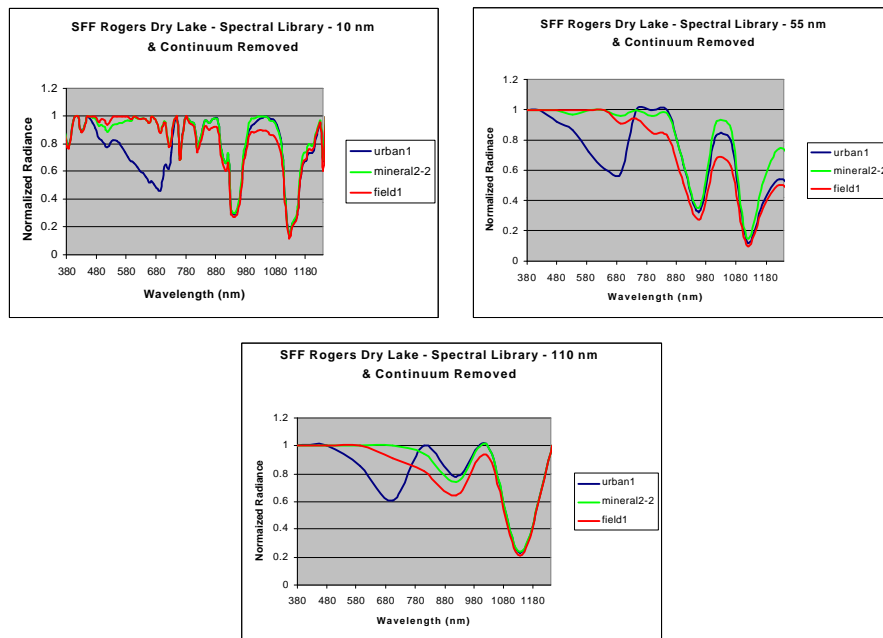
Figure 4-19: Spectral Library used for SFF with Rogers Dry Lake at different Spectral Resolutions

This type of phenomena is also observed (but not shown here) with the DIRSIG scene when the $P_{fa} = 0.2$ and also at $P_{fa} = 0.05$ when we include the associated standard deviation to explain the bend in the curve at 55 nm. Taking standard deviation (from Table 4-28) into account for the Rochester image we see that the probability of detection is somewhat independent of the spectral resolution for this scene.
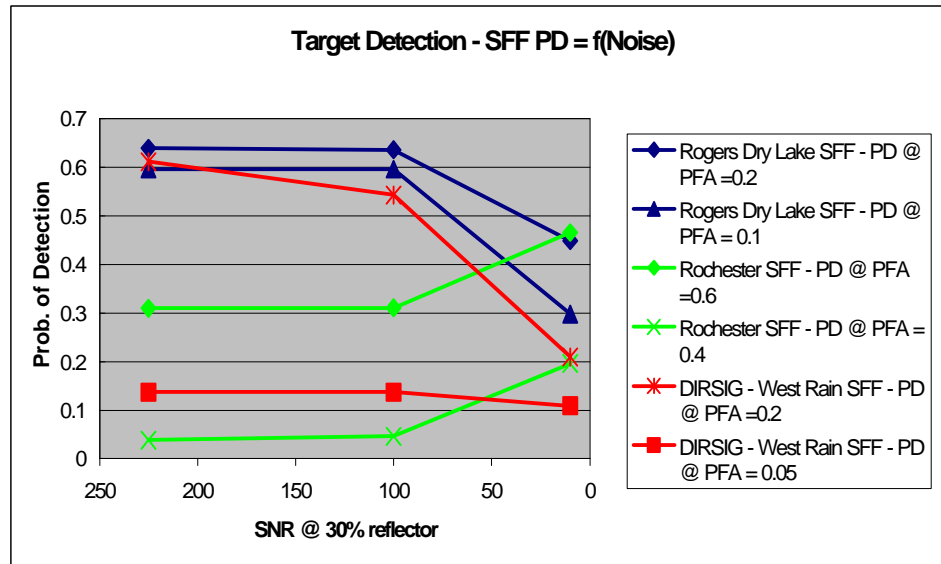
Figure 4-20: Spectral Feature Fitting – Probability of Detection as a function of SNR

The overall trend seen in Figure 4-20 shows that as SNR decreases so does the probability of detection. This agrees with our intuition and the results from testing other algorithms. However, the performance of the Rochester scene with respect to noise degradation seems to contradict the general trend. Is it possible that adding noise to the image spectra makes its absorption features more distinct and better matches to the library spectra? Highly doubtful. As mentioned in Chapter 2, one product of this algorithm is an RMS error image for each specified target. This gives the user some confidence as to the actual presence of materials within a scene by a providing a visible "goodness of fit" of the regression model used. Every RMS error image produced in this thesis, despite the level of degradations assigned or whether it was synthetic or real imagery, had a vertically striped pattern to it. An example of this garbage is seen in Figure 4-21. Although not fully tested, the periodicity of these stripes seems to be a function of the user- specified spatial resolution. In any case, the user cannot place much reliance in this

type of "goodness of fit." Furthermore, if this is the RMS of the regression model used to fit absorption features then one can only wonder how accurate are the scale maps that indicate material abundance. Interestingly enough, this stripping effect is not present in the scale maps that show material abundance. If the regression model is not accurate, which these consistently poor RMS images may indicate, then little faith can be put into the scale maps of material abundance.
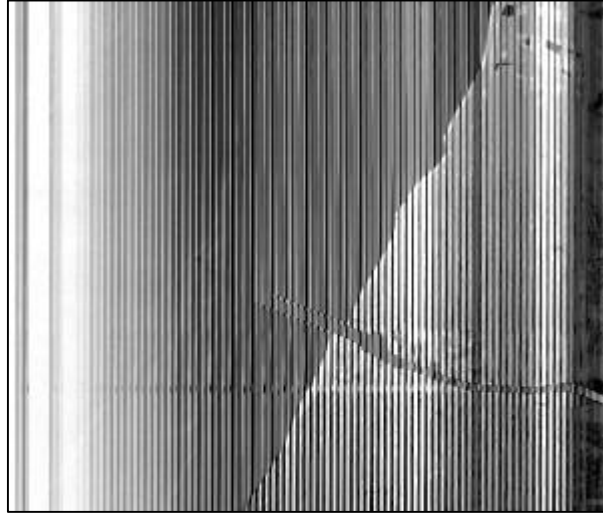
Figure 4-21: RMS Image from SFF Rochester, NY image using Shallow Water
as the Specified Target

It is useful to see if any of these peculiarities are apparent or better explained by conducting an ANOVA on the collected probabilities of detection.

```
Analysis of Variance for SFF PD @ PFA = 0.2 - Rogers Dry Lake

Source              DF    Seq SS     Adj SS     Adj MS        F       P
Spatial              2  0.006321   0.006321   0.003161     6.22   0.023
Spectral             2  0.072253   0.072253   0.036126    71.12   0.000
Noise                2  0.213425   0.213425   0.106713   210.08   0.000
Spatial*Spectral     4  0.001636   0.001636   0.000409     0.81   0.555
Spatial*Noise        4  0.004679   0.004679   0.001170     2.30   0.147
Spectral*Noise       4  0.075508   0.075508   0.018877    37.16   0.000
Error                8  0.004064   0.004064   0.000508
Total               26  0.377887
```

(continued)

```
Analysis of Variance for SFF PD @ PFA= 0.1 - Rogers Dry Lake

Source              DF     Seq SS     Adj SS     Adj MS        F      P
Spatial              2   0.010939   0.010939   0.005469    18.97  0.001
Spectral             2   0.085412   0.085412   0.042706   148.09  0.000
Noise                2   0.533686   0.533686   0.266843   925.33  0.000
Spatial*Spectral     4   0.000973   0.000973   0.000243     0.84  0.535
Spatial*Noise        4   0.002752   0.002752   0.000688     2.39  0.137
Spectral*Noise       4   0.110722   0.110722   0.027680    95.99  0.000
Error                8   0.002307   0.002307   0.000288
Total               26   0.746790
```

Table 4-19: Analysis of Variance for SFF – Rogers Dry Lake
($F_{0.05, 2, 8} = 4.46$ and $F_{0.05, 4, 8} = 3.84$)

Like the results from the SMF, it appears from Table 4-29(above) that the level of significance is a function of the PFA level specified   The ranking between the two P_{fa}'s is consistent and it is only the magnitude of the F-statistic that changes. At both PFA levels, noise remains the most significant contributing single factor with spectral resolution as the second most significant. This makes sense based on the fact that this algorithm is trying to match absorption features and any noise would lessen the chance of a good fit. At both P_{fa} levels, the interactions between spatial-spectral and spatial-noise remain insignificant at the test level of α=0.05.

The trends, indicated in Table 4-30 (below) for the Rochester image, do not follow those previously seen with the Rogers scene. Furthermore they are not as intuitive. As seen in Table 4-30, as the level of required P_{fa} drops, the significance of all the main effects, with the exception of noise, drops. Yet, the significance of noise on detection performance dramatically increases as the required P_{fa} level is reduced. Another interesting result is that the ranking of joint effect significance changes as the P_{fa} level is changed. As seen in Table 4-30 the interaction between spatial-spectral resolution is significant at $P_{fa} = 0.6$ yet it is not at $P_{fa} = 0.4$. The exact opposite of this is true for the interaction between spatial resolution and noise. None of the observations made from the Rochester image agree with observations made from the other AVIRIS scene. Again, this could be attributed to scene dependence or the poor results obtained from the SFF algorithm.

```
Analysis of Variance for SFF PD @ PFA = 0.6 - Rochester

Source               DF      Seq SS      Adj SS      Adj MS       F       P
Spatial               2    0.007541    0.007541    0.003771    34.46   0.000
Spectral              2    0.029909    0.029909    0.014954   136.68   0.000
Noise                 2    0.143803    0.143803    0.071901   657.18   0.000
Spatial*Spectral      4    0.002328    0.002328    0.000582     5.32   0.022
Spatial*Noise         4    0.000988    0.000988    0.000247     2.26   0.152
Spectral*Noise        4    0.011397    0.011397    0.002849    26.04   0.000
Error                 8    0.000875    0.000875    0.000109
Total                26    0.196841


Analysis of Variance for SFF PD @ PFA = 0.4 - Rochester

Source               DF      Seq SS      Adj SS      Adj MS       F       P
Spatial               2   0.0001133   0.0001133   0.0000566    10.86   0.005
Spectral              2   0.0002664   0.0002664   0.0001332    25.53   0.000
Noise                 2   0.1433928   0.1433928   0.0716964  1.4E+04   0.000
Spatial*Spectral      4   0.0000167   0.0000167   0.0000042     0.80   0.557
Spatial*Noise         4   0.0001116   0.0001116   0.0000279     5.35   0.021
Spectral*Noise        4   0.0007973   0.0007973   0.0001993    38.19   0.000
Error                 8   0.0000417   0.0000417   0.0000052
Total                26   0.1447400
```

Table 4-20: Analysis of Variance for SFF – Rochester ($F_{0.05, 2, 8} = 4.46$ and $F_{0.05, 4, 8} = 3.84$)

```
Analysis of Variance for SFF PD @ PFA = 0.2 - Western Rainbow

Source               DF      Seq SS      Adj SS      Adj MS       F       P
Spatial               2     0.57596     0.57596     0.28798    40.60   0.000
Spectral              2     0.52183     0.52183     0.26091    36.78   0.000
Noise                 2     0.83065     0.83065     0.41532    58.55   0.000
Spatial*Spectral      4     0.06144     0.06144     0.01536     2.17   0.164
Spatial*Noise         4     0.09921     0.09921     0.02480     3.50   0.062
Spectral*Noise        4     0.13825     0.13825     0.03456     4.87   0.028
Error                 8     0.05675     0.05675     0.00709
Total                26     2.28408


Analysis of Variance for SFF PD @ PFA = 0.05 - Western Rainbow

Source               DF      Seq SS      Adj SS      Adj MS       F       P
Spatial               2    0.277958    0.277958    0.138979    64.88   0.000
Spectral              2    0.025967    0.025967    0.012983     6.06   0.025
Noise                 2    0.004897    0.004897    0.002448     1.14   0.366
Spatial*Spectral      4    0.006824    0.006824    0.001706     0.80   0.560
Spatial*Noise         4    0.040247    0.040247    0.010062     4.70   0.030
Spectral*Noise        4    0.027350    0.027350    0.006837     3.19   0.076
Error                 8    0.017137    0.017137    0.002142
Total                26    0.400379
```

Table 4-21: Analysis of Variance for SFF – Western Rainbow ($F_{0.05, 2, 8}=4.46$ and $F_{0.05, 4, 8}=3.84$)

Yet again, the results from performing an ANOVA on the Western Rainbow scene (Table 4-31) do not permit any broad conclusion when compared to the other images. At a $P_{fa} = 0.2$ all main effects are significant to the performance of this algorithm. However, as the operator requires a lower P_{fa} the level and ranking of significant factors changes. At a $P_{fa} = 0.05$ the main effect of spatial resolution is by far the most significant. This agrees with the spatial resolution trends seen with the SMF at lower P_{fa}, and its relationship to higher pixel purity. However, the trend witnessed here does not agree with the two AVIRIS scenes tested.

The bottom line is there are problems with the way the SFF algorithm was tested as part of this thesis and in the way it is implemented in ENVI. Firstly, we used scene-derived endmembers from a scene in radiance space as inputs into this algorithm. The SFF algorithm requires that the image be in units of reflectance prior to continuum removal. This may, in part, explain the better performance of this algorithm with the atmospherically corrected DIRSIG image. The second problem is that this method would work better if very specific and characteristic absorption features are sought. In this manner using this algorithm over the entire reflectance image, although it produces results, would have yielded better results if a narrow spectral range of interest (e.g. only covering a range of 50-100 nm) were selected. This was somewhat done with the AVIRIS images, yet this range included atmospheric bands and the spectral range was still to wide. Even without these improvements to our approach one would expect some more commonality between the final SFF products from the image sets used when looking at the ANOVA tables. Despite these improvements to our approach in using the SFF algorithm, full confidence cannot be placed in this algorithm given that the RMS images exhibit a strange and unexplained periodic pattern. As stated earlier, we can only wonder how accurate the scale maps are that indicate the material abundance/presence when our "goodness of fit" does not look so good. The bottom line is that the poor performance of the SFF algorithm reduces our interest in using this algorithm in future work and it brings any delivered results into question. The way this algorithm is implemented within ENVI requires further investigation and possible correction.

Chapter 5

Conclusions and Recommendations

5.1   Conclusions

The primary objective of this thesis study was to conduct an examination into how differing values of spectral resolution, spatial resolution and noise effect the performance of hyperspectral algorithms and the utility of the information derived from them. In other words, the goal was to characterize the error in spectrally based information products by measuring the utility of processed images. This was accomplished by utilizing several meaningful metrics to measure image utility and the employment of a factorial designed experiment. This ultimately has provided us with better insight into the effectiveness of these algorithms under different image acquisition parameters. This type of assessment has also allowed us to make performance comparisons between different types of processing algorithms.

In summary, a tool has been made that degrades hyperspectral images spatially, spectrally and by adding spectrally correlated noise. This has allowed us to "simulate" image acquisition under different sensor collection parameters. Several spectral algorithms were selected as testing candidates. These included Spectral Angle Mapper (SAM), Binary Encoding (BE), Gaussian Maximum Likelihood (GML), Linear Spectral Unmixing, Spectral Matched Filter (SMF) and Spectral Feature Fitting (SFF). A full investigation into image information content/utility metrics has also been conducted. This investigation resulted in the use of three meaningful *algorithm-specific* metrics – kappa for classification/thematic mapping, squared error for unmixing techniques and ROC curves for target detection. Applying these metrics to spectral products derived from degraded imagery has shown that definite trends exist which indicate the effectiveness of spectral algorithms under differing levels of spatial resolution, spectral resolution

and noise.   However, it has also been demonstared that there is an underlying scene dependence on algorithm performance. That is, the effectiveness of the tested algorithms depends on the spectral and spatial complexity of the initail images.

Through the use of a factorial designed experiment an investigation into the joint effects of these three sensor parameters (spatial resolution, spectral resolution and noise) was possible. This revealed that the main factors, by themselves, hold more significance in the outcome of spectral algorithm performance than any of the combined effects of these parameters. It was also discovered from the interpretation of our data that each algorithm tested seems to have one factor that is more significant in determining its performance and information degradation. For instance, it appears from our data that if an operator using the SMF algorithm requires lower probabilities of false alarm, then the significance of spatial resolution increases. At higher values of probability of false alarm it seems that spectral resolution carries more importance.

It has been shown that the use of a factorial designed experiment is a very effective way of testing algorithm performance. Furthermore, this experimental approach allows the investigator to study several different factors, simultaneously, along the image chain. The output from this type of experimental design permits easy identification of trends and readily allows statistical hypothesis testing to establish the significance of the parameters on algorithm outcome.

The results of this thesis study are useful and important in three main areas. Firstly and as previously mentioned, as a proof of concept, we have demonstrated that the use of a factorial designed experiment is an excellent approach for simultaneously studying several factors along the image chain and their impact on image information utility. Secondly, this thesis is the first and necessary step to establishing confidence limits on information derived from spectral products acquired under certain collection parameters. For example, an image analyst is using the SMF algorithm to find a certain target within an image characterized by a specific spatial resolution, spectral resolution and noise content.   Continuing with the example, with this information the image analyst will be 80-85% confident in his/her information from the spectral product. Finally, the information from this thesis may be applied during the first phases of spectral sensor design. Given that a specific imaging system is designed to fulfil a certain role, designers and engineers can better investigate the trade-offs between sensor parameters.

5.2   Recommendations

The following recommendations can be divided into two main categories - changes to the image degradation tool and future work. As stated previously, this thesis has proven that definite trends exist indicating the effectiveness of spectral algorithms under differing levels of spatial resolution, spectral resolution and noise. It has also shown that the factorial designed experiment is an extremely useful tool for investigating factors along the imaging chain. However, it is also realized that this is just a preliminary step towards a much larger goal – improvements and more work are required.

Changes to Image Degradation Tool

Although a tool has been made that degrades hyperspectral images spatially, spectrally and by adding spectrally correlated noise, improvements to this tool are needed prior to any future work. By implementing the improvements listed below the degradation tool will better "simulate" the acquisition of an image under different and "degraded" collection parameters.

With regards to spectral resampling, the user should be allowed to input the desired spectral resolution by entering the spectral bandwidth in nanometers or microns. As explained in Section 3.2.2, the user is presently requested to input the number of bands he/she wishes the initial image to be degraded to. This requires the user to manually calculate the number of bands desired for a certain spectral resolution before input into the degradation tool. This improvement would make the program more user-friendly. It would also be advantageous to incorporate a user-defined list of FWHM for each new/degraded spectral band centre. Currently, the degradation tool assumes critical spectral resampling by establishing a Gaussian model with a FWHM equivalent to the band spacing. Incorporating the optional input of FWHM values would provide more realistic spectral resampling.

It would be also be beneficial to implement the theoretically correct method of spatial degradation using a convolution process and subsequent resampling to maintain image size. The aggregate process or "box car" approach currently used, although very effective, does not truly represent the exact phenomenology of image acquisition. Additionally, the current approach does not maintain consistent material fractions when resampling – as seen in the testing the spectral unmixing algorithm.

The method by which spectrally correlated noise was added to the imagery needs to be revisited for two reasons. Firstly, a further validation of the "reverse MNF" process used here (via image subtraction of noisy image and noiseless image) needs to be done. Secondly, re-investigating the use of dark current images that was first attempted and outlined in Appendix D would also be worthwhile. It is believed that the approach using a dark current image is more theoretically sound and would achieve better results with respect to noise addition. However, the proper implementation proved more difficult and time consuming than ever anticipated. I believe the algorithm and code available on the enclosed CD is close to working and requires a little more attention.

Finally, the degradation tool should be revamped to enable the program to run in batch mode. The present method degrades one image at a time. This improvement would definitely save time and labour when future work is conducted in degrading numerous image sets to new parameters of spatial resolution, spectral resolution and noise.

Future Work and More Testing

The factorial designed experimental approach used in this thesis study has definitely proved itself effective. Yet many of the spectral algorithm performance trends witnessed in this study *cannot* yet be formed into concrete conclusions without further testing. More images need to be tested since there is an obvious scene dependence on the performance of these algorithms. Just the same, conclusions cannot be drawn with merely three points on a curve. More levels of each factor (spectral resolution, spatial resolution and noise) also need to be tested. By using many more scenes and testing at more levels, we will be better able to notice more global trends with respect to algorithm performance and eventually reduce the scene dependence from our statistics. In conducting any further studies, we should entertain the use of a "random effects" factorial designed experiment vice the fixed effects approach adopted here. A random effects factorial approach will allow us to conduct hypothesis tests that will deliver conclusions with respect to a population of possible parameter levels. The testing done in this thesis only allows us to draw conclusions about the levels that were tested.

In conjunction with more images being tested at more parameter levels, a proper regression model could then be developed for each spectral algorithm tested. This would allow a user to input the image parameters of spatial resolution, spectral resolution and noise of an image into the developed regression equation and subsequently be delivered the expected metric result.

That is, the measure of information utility/degradation they may expect with regards to any selected algorithm will be available by the use of a regression model. Obviously, this requires more testing and further study. Additionally, it would be advantageous to re-plot the graphs shown in this thesis to better show the three dimensional trade-space of the parameters of spectral resolution, spatial resolution and noise. Examples of what these plots may look like and ideas with respect to developing a 3-D representation of this trade-space are discussed in Appendix H.

Prior to moving ahead, certain aspects of the work conducted in this study need more investigation. Firstly, the bizarre behaviour of linear unmixing with respect to changes in spectral resolution needs to be revisited. Furthermore, more investigation is needed into the performance of these algorithms in radiance space versus reflectance space. It is clearly evident that spectral algorithm performance is also dependent upon the atmospheric inversion technique used since the type of inversion used may effect retrieved reflectance values. Along these lines, it would be interesting to run the same experiment and algorithms on an atmospherically corrected Rochester image for comparison to the results achieved here. The SFF algorithm also needs further attention by either recoding it or abandoning it. Part of this should include using the algorithm in reflectance space only (as it was designed for) and selecting very narrow and distinct spectral absorption bands vice wide ranges. The details into how the RMS error image is produced within ENVI also needs to be investigated.

In this light, it is imperative that we try implementing these algorithms ourselves by using our own code instead of ENVI. We are not entirely sure what is happening "underneath the hood" of ENVI. By recoding these algorithms ourselves, we would have better control on algorithm testing, be provided with a better understanding of the algorithms and establish another benchmark.

In hindsight, DIRSIG should have been used to produce "AVIRIS-like" scenes by matching the same collection parameters as AVIRIS (spectrally, spatially, swath, etc.). We were unable to model AVIRIS scenes exactly to these parameters at the time of this thesis study. This would have provided us a better benchmark of spectral algorithm performance between real and synthetic images. However, at this point in time it can be concluded that the spectral-spatial variability and spectral purity of DIRSIG images provide significant obstacles when conducting rigorous algorithm testing. In other words, although DIRSIG does an absolute superb job at modelling the image chain and spectral phenomenology, it is not as quite as complex as the real world - yet. Food for thought.

References

Alexander, R.J. and Cheatham, P.S. (1998). Automated Hyperspectral Ground Processing Chain. *(publication unknown)*, 1998.

Author unknown (date unknown). National Image Interpretability Rating Scale (NIIRS). Internet: www.fas.org/irp/imint/niirs_c/

Boardman, J.W. (1994). Geometric Mixture Analysis of Imaging Spectrometry Data. *Proceedings of IGARSS 1994 Symposium*, pages 2369-2371. IGARSS – International Geoscience and Remote Sensing Symposium.

Boardman J. W., and Kruse, F. A. (1994). Automated spectral analysis: A geologic example using AVIRIS data, North Grapevine Mountains, Nevada. *Proceedings, Tenth Thematic Conference on Geologic Remote Sensing*, San Antonio, TX, 9-12 May 1994, pages I-407 - I-418.

Boardman, J.W. (1995). Using Dark Current Data to Estimate AVIRIS Noise Covariance and Improve Spectral Analysis. *Proceedings of the Fifth Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Workshop.* JPL Publication 95-1, 23-26 Jan 1990. Pages19-22.

Bowles, J.H., Palmadesso, P.J., Antoniades, J.A., Baumback, M.M., Grossman, J.M. (1996) Effect of spectral resolution and number of wavelength bands in analysis of a hyperspectral data set using NRL's ORASIS algorithm. *Proceedings of the SPIE*, volume 2821, pages 171-182. SPIE – The International Society for Optical Engineering.

Brown, S.D. (Date unknown). *A brief overview of DIRSIG*. Internet. www.cis.rit.edu/research/dirs/dirsig/overview/index.html.

Chang, C.I. and Ren, H. (2000) An Experiment-Based Quantitative and Comparative Analysis of Target Detection and Image Classification Algorithms for Hyperspectral Imagery. *IEEE Transactions on Geoscience and Remote Sensing,* volume 38, no. 2, pages 1044-1063, March 2000.

Clark, R. N., Swayze, G. A., Gallagher, A., Gorelick, N., and Kruse, F. A. (1991). Mapping with imaging spectrometer data using the complete band shape least-squares algorithm simultaneously fit to multiple spectral features from multiple materials. *Proceedings, 3rd Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Workshop.* JPL Publication 91-28, May 20-21, 1991.pages 2-3.

Congalton, R.C., Oderwald, R.G. and Mead, R.A. (1983). Assessing LANDSAT Classification Accuracy using Discrete Multivariate Analysis Statistical Techniques, *Photogrammetric Engineering and Remote Sensing*, volume 49, number 12, December 1983, pages 1671-1678

Congalton, R.C. and Green, K. (1999). *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. Lewis Publishers, New York.

DeFatta, D.J., Lucas, J.G., Hodgkiss, W.S. (1988). *Digital Signal Processing: A System Design Approach*. John Wiley & Sons, New York.

Easton, R. (1998). *Linear Systems Mathematics for Imaging Science*. Textbook in progress and course notes for SIMG716 and SIMG717 at Rochester Institute of Technology, Rochester, New York.

Feng, X. (1995). *Design and Performance evaluation of a Modular Imaging Spectrometer Instrument*. PhD dissertation – Centre for Imaging Science, Rochester Institute of Technology, Rochester, New York.

Fiete, R.D. (1999). Image Quality and λFN/p for Remote Sensing Systems. *Optical Engineering*, volume 38, no. 7, pages 1229-1240, July 1999.

Gaskill, J.D. (1978). *Linear Systems, Fourier Transforms and Optics*. Wiley Series in Pure and Applied Optics, John Wiley and Sons, New York.

Gonzalez, R.C. and Woods, R.E. (1992). *Digital Image Processing*. Addison-Wesley Publishing Co., New York.

Green, A.A., Berman, M., Switzer P. and 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*, volume 26, no. 1, pages 65-74, January 1988

Harsanyi, J.C. and Chang, C. (1994) Hyperspectral Image Classification and Dimensionality reduction: An Orthogonal Subspace Projection Approach. *IEEE Transactions on Geoscience and Remote Sensing,* volume 32, no. 4, pages 779-785, July 1994

Haskett, H.T. and Sood, A.K. (1998). Trade-off studies of detection performance versus the number of reflective spectral bands in hyperspectral imagery. *Proceedings of the SPIE – Algorithms for Multispectral and Hyperspectral Imagery IV*, volume 3372, pages 26-42. SPIE – The International Society for Optical Engineering.

Huck, F.O., Fales, C.L., Halyo, N., Samms, R.W., Stacy, K. (1985) Image gathering and processing: information and fidelity. *J. Opt. Soc. Am. A*, volume 2, no. 10, pages 1644-1666, October 1985.

Johnson, R.A. and Wichern D.W. (1998). *Applied Multivariate Statistical Analysis*. Prentice Hall, Inc., New Jersey.

Kassam, S.A., (1988). *Signal Detection in Non-Gaussian Noise*. Thomas, J.B., editor, Springer-Verlang, New York

Kay, S.M. (1998). *Fundamentals of Statistical Signal Processing – Volume II: Detection Theory*. Prentice-Hall, New Jersey.

Keller, R.A., Shen, S.S., and Pritt, A.T. *Sensor Design Considerations for HIS Remote Sesning*. Not yet published. The Aerospace Corporation. received May 2000.

Kerekes, J.P. and Landgrebe, D.A. (1987) A Noise taxonomy for Remote Sensing Systems. *Proceedings of IGARSS 1987 Symposium*, Ann Arbor MI, May 1987, pages 903-908. IGARSS – International Geoscience and Remote Sensing Symposium.

Kerekes, J.P. and Landgrebe, D.A. (1991). Parameter Trade-Offs for Imaging Spectroscopy Systems. *IEEE Transactions on Geoscience and Remote Sensing*, volume 29, no. 1, pages 57-65, January 1991.

Kerekes, J.P. (1996). Parameter Impacts on Hyperspectral remote Sensing System Performance. *Proceedings of the SPIE*, volume 2821, pages 195-201. SPIE – The International Society for Optical Engineering.

Kerekes, J.P. (1998). Error Analysis of Spectral Reflectance Derived from Imaging Spectrometer Data. *Proceedings of IGARSS 1998 Symposium*, Seattle, WA, July 1998, volume 5, pages 2697-2701. IGARSS – International Geoscience and Remote Sensing Symposium

Kerekes, J.P. (1999). Analytical Model of Hyperspectral System Performance. *Proceedings of the SPIE – Infrared Imaging Systems: Design, Modelling and Testing X*, Orlando, FL, April 1999, volume 3701 (received prior to publication). SPIE – The International Society for Optical Engineering.

Konno, D. (1999) *Development and Testing of Improved Spectral Unmixing Techniques*. Masters Thesis (Imaging Science), Rochester Institute of Technology, 1999

Kruse, F.A., Lefkoff, A.B. and Dietz, J.B. (1993a).Expert System-Based Mineral Mapping in Northern Death Valley, California/Nevada, using the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). *Remote Sensing of the Environment*, May-June 1993, volume 44, pages 309-336.

Kruse, F. A., Lefkoff, A. B., Boardman, J. W., Heidebrecht, K. B., Shapiro, A. T., Barloon, J. P., and Goetz, A. F. H. (1993b). The Spectral Image Processing System (SIPS) - Interactive Visualization and Analysis of Imaging Spectrometer Data. *Remote Sensing of the Environment*, May-June 1993, volume 44, pages 145 - 163.

Landgrebe, D.A. (1978a). The Quantitative Approach to Remote Sensing: Concept and Rationale. In Swain, P.H. and Davis, S.M., editors, *Remote Sensing: The Quantitative Approach*. pages 1-20, McGraw-Hill Inc., New York.

Landgrebe, D.A. (1978b). Useful Information from Multispectral Image Data. In Swain, P.H. and Davis, S.M., editors, *Remote Sensing: The Quantitative Approach*. pages 1-20, McGraw-Hill Inc., New York.

Landgrebe, D.A. and Maralet, E. (1986) Noise in Remote Sensing Systems: The Effect on Classification Error. *IEEE Transactions on Geoscience and Remote Sensing*, volume 24, no. 2, pages 294-200, March 1986.

Leachtenauer, J.C., Malila, W., Irvine, J., Colburn, L., Salvaggio, N. (1997). General Image Quality Equation – GIQE. *Applied Optics*, volume 36, no. 32, pages 8322-8328, November 1997.

Lee, J.B., Woodyatt, A.S. and Berman, M.(1990) Enhancement of High Resolution Remote-Sensing Data by a Noise-Adjusted Principal Components Transform. *IEEE Transactions on Geoscience and Remote Sensing,* volume 28, no. 3, pages 295-304, May 1990

Lee, C., and Landgrebe, D.A. (1993). Analyzing High-Dimensional Multispectral Data. *IEEE Transactions on Geoscience and Remote Sensing*, volume 31, no. 4, pages 792-800, July 1993.

Minitab, Inc. (1998). Minitab 12.2 for Windows (http://www.minitab.com)

Montgomery, D.C. (1997). *Design and Analysis of Experiments*. John Wiley & Sons, New York.

Pinzon, J.E., Ustin, S.L., Castaneda, C.M., Pierce, J.F., Costick, L.A. (1998). Robust spatial and spectral feature extraction for multispectral and hyperspectral imagery. *Proceedings of the SPIE – Algorithms for Multispectral and Hyperspectral Imagery IV*, volume 3372, pages 199-210. SPIE – The International Society for Optical Engineering.

Price, J.C. (1984). Comparison of the Information Content of data from the LANDSAT-4 Thematic Mapper and the Multispectral Scanner. *IEEE Transactions on Geoscience and Remote Sensing*, volume 22, no. 3, pages 272-281, May 1984.

RSI – Research Systems Incorporated (1998). ENVI (Environment for Visualizing Images) v 3.2 and IDL (Interactive Data Language) v 5.2

Richards, J.A. (1993). *Remote Sensing Digital Image Analysis*. Springer-Verlag Inc., New York.

Roberts, D. A., Yamaguchi, Y., and Lyon, R. J. P.(1985). Calibration of Airborne Imaging Spectrometer data to percent reflectance using field measurements. *Proceedings, Nineteenth International Symposium on Remote Sensing of Environment*, Ann Arbor, MI, October 21-25, 1985.

Rosenfield, G.H. (1981). Analysis of Variance of Thematic Mapping Experiment Data. *Photogrammetric Engineering and Remote Sensing*, volume 47, number 12, December 1981, pages 1685-1692.

Schott, J.R. (1997). *Remote Sensing: The Image Chain Approach*. Oxford University Press, New York.

Shannon, C.E. (1948). A Mathematical Theory of Communication. *Bell Systems Technical Journal*, volume 27, pages 379-423, July 1948.

Shannon, C.E. (1949). Communication in the Presence of Noise. *Proceedings of the IRE*, volume 37, pages 10-21, January 1949.

Stoner, W. and Resmini, R. (1996). Hyperspectral Remote Sensing. Advanced Technology Applications Division – SAIC (Science Applications International Corporation) – internal document.

Swain, P.H., Siegel, H.J., Smith, W.S. (1980). Contextual Classification of multispectral Remotely Sensing Data Using a Multiprocessor System. *IEEE Transactions on Geoscience and Remote Sensing*, volume 18, no. 2, pages 197-203, April 1980.

Tsang, K.Y., Grossman, J.M., Palmedesso, P.J., Antoniades, J.A., Baumback, M.M., Bowles, J.H., Daniel, M., Fisher, J., Haas, D. (1998). Evaluation of endmember Selection Techniques and Performance Results form ORASIS Hyperspectral Analysis. *Proceedings of the SPIE – Algorithms for Multispectral and Hyperspectral Imagery IV*, volume 3372, pages 43-50. SPIE – The International Society for Optical Engineering.

Vane, G., Green, R.O., Chrien, T.G., Enmark, H.T., Hansen, E.G., Porter, W.M. (1993). The Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). *Remote Sensing of the Environment*, volume 44, pages 127-143.

Walpole, R.E. (1982). *Introduction to Statistics*. Macmillan Publishing Company, New York.

Wrigley, R.C., Card, D.H., Hlavka, C.A., Hall, J.R., Mertz, F.C., Archwamety, C., Schowengerdt, R.A. (1984). Thematic Mapper Image Quality: Registration, Noise and Resolution. *IEEE Transactions on Geoscience and Remote Sensing*, volume 22, no. 3, pages 263-217, May, 1984.

Wolfe, W.L. (1997). *Introduction to Imaging Spectrometers*. SPIE Optical Engineering Press, Bellingham, WA. SPIE – The International Society for Optical Engineering.

Zavaljevski, A., Dhawan, A.P., Kelch, D.J., Riddell, J. (1996). Adaptive multilevel classification and detection in multispectral images. *Optical Engineering*, volume 35, no. 10, pages 2884-2893, October 1996.

Appendix A

Spectral Product Levels

The following chart is extracted from Alexander and Cheatham's (1998) proposal for an automated hyperspectral processing system. Essentially, this system splits the image chain into a series of processing levels. As discussed in section 2.2.1, the interface points along this processing chain are called products. Each product has defining characteristics that relate to how it was made and what purpose it fulfils. Examples of different products at each processing level can be seen in Table A-1. The top two rows consist of the level name and a brief level description. Under the double line is a list of different products organized according to their respective level - in a column-wise fashion. The shaded Level 3 Spectral Products are the spectral products being studied as part of this thesis. Classification routines, such as Binary Encoding (BE), Spectral Angle Mapper (SAM) and Gaussian Maximum Likelihood (GML), produce classification maps. Linear Spectral Unmixing produces fractional endmember maps. The final products of the Spectral Matched Filter (SMF) and Spectral Feature Fitting (SFF) algorithms are target probability/detection maps. Those products associated with Level 5 (Reporting and Decisions) are not included in Table A-1. The various reports, and the decisions made from them, rely on the information compiled, organised and properly interpreted from Level 4 products.

PROCESSING LEVELS

LEVEL 0: RAW HYPERSPECTRAL IMAGE CUBE	LEVEL 1: CALIBRATION	LEVEL 2: DATA RESAMPLING	LEVEL 3: SPECTRAL PRODUCTS	LEVEL 4: DATA EXPLOITATION
Initial image cube from down-linked data	*Spectral & Radiometric Calibration of Level 0 raw sensor data*	*Atmospheric Correction and Geospatial Rectification*	*Extracted spectral information, rendered as raster images, tables and parsable text files*	*Customer products for mission-specific needs – hardcopy or softcopy products*
PRODUCTS				
Raw Data Image Cube	Radiometrically Corrected Image Cube	Atmospheric Absorption Profiles	Temperature Map	Annotated Target & Material ID Graphic with text
Telemetry Data describing cube location, acquisition and sensor calibration	Data Cube Header – Lat/Long	Water Vapour and Aerosol Overlays	Scene Characterisation (Classification and Endmember) Maps	Material Mixture Constituents Description & Location Map
	Failed detector artifacts and geometric mis-registration error overlay	Atmospheric Correction Transform	Anomaly Detection Maps	Geospatial & Terrain Feature Vectors & Maps
	Radiometric Saturation Mask	Spectrally Corrected Image/Data Cube	Signatures of Interest – Detection maps	Lines of Communication Vectors & Maps
	Quick Look Image (grey scale or RGB)	Lat/Long Pixel Transform	Spectral Analysis of Objects of Interest	Data Fusion Products
		Pixel Spatial Reference (elevation, slope, normal)	Visual Reference Image	Change Detection Overlays
		Cloud Mask	Annotated Summary Graphics and/or Spreadsheet	Mosaicked Products & Image/Data Cubes
		Atmospherically Corrected Image/Data Cube		

Table A-1: Processing Levels and Respective Products

Appendix B

National Image Interpretability Rating Scales (NIIRS)

This appendix contains a brief description and examples of the NIIRS scale. The NIIRS scale primarily describes the type of information that can be extracted from an image and the extent of image interpretability based on a pre-defined list of tasks. The tables included in this appendix contain some examples of tasks/information extracted from NIIRS level 3, 4, 5 and 6 imagery. A more detailed and complete NIIRS tables can be found at www.fas.org/irp/imint/niirs_c/. As seen in the tables below, the extracted information is dependent upon the type of sensor used to acquire the image. For instance, visible NIIRS tasks rely primarily on the spatial resolution and content of an image while multispectral NIIRS also utilise the spectral content of the image. An example of using spectral content is the ability to detect certain types of camouflage netting against a scattered tree background. The NIIRS system offers a qualitative method of rating image information quality. As mentioned in section 2.7.3, we use only quantitative metrics in this thesis study. However, tasks such as those listed in the NIIRS tables were used, as much as possible, in the employment of the spectral algorithms.

"The aerial imaging community utilizes the National Imagery Interpretability Rating Scale (NIIRS) to define and measure the quality of images and performance of imaging systems. Through a process referred to as "rating" an image, the NIIRS is used by imagery analysts to assign a number that indicates the interpretability of a given image. The NIIRS concept provides a means to directly relate the quality of an image to the interpretation tasks for which it may be used. Although NIIRS has been primarily applied in the evaluation of aerial imagery, it provides a systematic approach to measuring the quality of photographic or digital imagery, the performance of image capture devices, and the effects of image processing algorithms."

Source: www.fas.org/irp/imint/niirs_c/

NIIRS 3 [2.5 - 4.5 m GSD]

Visible NIIRS	Radar NIIRS	Infrared NIIRS	Multispectral NIIRS
Identify the wing configuration (e.g., straight, swept, delta) of all large aircraft (e.g., 707, CONCORD, BEAR, BLACKJACK). Identify radar and guidance areas at a SAM site by the configuration, mounds, and presence of concrete aprons. Detect a helipad by the configuration and markings. Detect the presence / absence of support vehicles at a mobile missile base. Identify a large surface ship in port by type (e.g., cruiser, auxiliary ship, non-combatant/merchant). Detect trains or strings of standard rolling stock on railroad tracks (not individual cars)	Detect medium-sized aircraft (e.g., FENCER, FLANKER, CURL, COKE, F-15). Identify an ORBITA site on the basis of a 12-meter dish antenna normally mounted on a circular building. Detect vehicle revetments at a ground forces facility. Detect vehicles/pieces of equipment at a SAM, SSM, or ABM fixed missile site. Determine the location of the superstructure (e.g., fore, amidships, aft) on a medium-sized freighter. Identify a medium-sized (approx. six track) railroad classification yard.	Distinguish between large (e.g., C-141, 707, BEAR, A300 AIRBUS) and small aircraft (e.g., A-4, FISHBED, L-39). Identify individual thermally active flues running between the boiler hall and smokestacks at a thermal power plant. Detect a large air warning radar site based on the presence of mounds, revetments and security fencing. Detect a driver-training track at a ground forces garrison. Identify individual functional areas (e.g., launch sites, electronics area, support area, missile handling area) of an SA-5 launch complex. Distinguish between large (e.g., greater than 200 meter) freighters and tankers.	Detect vegetation/soil moisture differences along a linear feature (suggesting the presence of a fenceline). Identify major street patterns in urban areas. Identify golf courses. Identify shoreline indications of predominant water currents. Distinguish among residential, commercial, and industrial areas within an urban area. Detect reservoir depletion.

NIIRS 4 [1.2 - 2.5 m GSD]

Visible NIIRS	Radar NIIRS	Infrared NIIRS	Multispectral NIIRS
Identify all large fighters by type (e.g., FENCER, FOXBAT, F-15, F-14). Detect the presence of large individual radar antennas (e.g., TALL KING). Identify, by general type, tracked vehicles, field artillery, large river crossing equipment, wheeled vehicles when in-groups.	Distinguish between large rotary-wing and medium fixed-wing aircraft (e.g., HALO helicopter versus CRUSTY transport). Detect recent cable scars between facilities or command posts. Detect individual vehicles in a row at a known motor pool. Distinguish between open	Identify the wing configuration of small fighter aircraft (e.g., FROGFOOT, F- 16, and FISHBED). Detect a small (e.g., 50 meter square) electrical transformer yard in an urban area. Detect large (e.g., greater than 10 meter diameter) environmental domes at an	Detect recently constructed weapon positions (e.g. tank, artillery, self-propelled gun) based on the presence of revetments, berms, and ground scarring in vegetated areas. Distinguish between two-lane improved and unimproved roads. Detect indications of natural surface airstrip

Detect an open missile silo door. Determine the shape of the bow (pointed or blunt/rounded) on a medium-sized submarine (e.g., ROMEO, HAN, Type 209, CHARLIE 11, ECHO 11, VICTOR II/III). Identify individual tracks, rail pairs, control towers,	and closed sliding roof areas on a single bay garage at a mobile missile base. Identify square bow shape of ROPUCHA class (LST). Detect all rail/road bridges.	electronics facility. Detect individual thermally active vehicles in garrison. Detect thermally active SS-25 MSV's in garrison. Identify individual closed cargo hold hatches on large merchant ships.	maintenance or improvements (e.g., runway extension, grading, resurfacing, bush removal, vegetation cutting). Detect landslide or rockslide large enough to obstruct a single-lane road. Detect small boats(15-20 feet in length) in open water

NIIRS 5 [0.75 - 1.2 m GSD]

Visible NIIRS	Radar NIIRS	Infrared NIIRS	Multispectral NIIRS
Distinguish between a MIDAS and a CANDID by the presence of refuelling equipment (e.g., pedestal and wing pod). Identify radar as vehicle-mounted or trailer-mounted. Identify, by type, deployed tactical SSM systems (e.g., FROG, SS-21, SCUD). Distinguish between SS-25 mobile missile TEL and Missile Support Vans (MSVS) in a known support base, when not covered by camouflage. Identify TOP STEER or TOPSAIL air surveillance radar on KIROV-, SOVREMENNY-, KIEV-, SLAVA-, MOSKVA-, KARA-, or KRESTA-II-class vessels.	Count all medium helicopters (e.g., HIND, HIP, HAZE, HOUND, PUMA, and WASP). Detect deployed TWIN EAR antenna. Distinguish between river crossing equipment and medium/heavy armoured vehicles by size and shape (e.g., MTU-20 vs. T-62 MBT). Detect missile support equipment at an SS-25 RTP (e.g., TEL, MSV). Distinguish bow shape and length/width differences of SSNS. Detect the break between railcars (count railcars).	Distinguish between single-tail (e.g., FLOGGER, F-16, TORNADO) and twin-tailed (e.g., F-15, FLANKER, FOXBAT) fighters. Identify outdoor tennis courts. Identify the metal lattice structure of large (e.g. approximately 75 meter) radio relay towers. Detect armoured vehicles in a revetment. Detect a deployed TET (transportable electronics tower) at an SA-10 site. Identify the stack shape (e.g., square, round, oval) on large (e.g., greater than 200 meter) merchant ships.	Detect automobile in a parking lot. Identify beach terrain suitable for amphibious landing operation. Detect ditch irrigation of beet fields. Detect disruptive or deceptive use of paints or coatings on buildings/structures at a ground forces installation. Detect raw construction materials in ground forces deployment areas (e.g., timber, sand, and gravel).

NIIRS 6 [0.40 - 0.75 m GSD]

Visible NIIRS	Radar NIIRS	Infrared NIIRS	Multispectral NIIRS
Distinguish between models of small/medium helicopters (e.g., HELIX A from HELIX B from HELIX C, HIND D from HIND E, HAZE A from HAZE B from HAZE C). Identify the shape of antennas on EW/GCI/ACQ radars as parabolic, parabolic with clipped comers or rectangular. Identify the spare tire on a medium-sized truck. Distinguish between SA-6, SA- I 1, and SA- 17 missile airframes. Identify individual launcher covers (8) of vertically launched SA-N-6 on SLAVA-class vessels. Identify automobiles as	Distinguish between variable and fixed-wing fighter aircraft (e.g., FENCER vs. FLANKER). Distinguish between the BAR LOCK and SIDE NET antennas at a BAR LOCK/SIDE NET acquisition radar site. Distinguish between small support vehicles (e.g., UAZ-69, UAZ-469) and tanks (e.g., T-72, T-80). Identify SS-24 launch triplet at a known location. Distinguish between the raised helicopter deck on a KRESTA II (CG) and the helicopter deck with main deck on a KRESTA I (CG).	Detect wing-mounted stores (i.e., ASM, bombs) protruding from the wings of large bombers (e.g., B-52, BEAR, Badger). Identify individual thermally active engine vents atop diesel locomotives. Distinguish between a FIX FOUR and FIX SIX site based on antenna pattern and spacing. Distinguish between thermally active tanks and APCs. Distinguish between a 2-rail and 4-rail SA-3 launcher. Identify missile tube hatches on submarines.	Detect summer woodland camouflage netting large enough to cover a tank against a scattered tree background. Detect foot trail through tall grass. Detect navigational channel markers and mooring buoys in water. Detect livestock in open but fenced areas. Detect recently installed minefields in ground forces deployment area based on a regular pattern of disturbed earth or vegetation. Count individual dwellings in subsistence housing areas (e.g., squatter settlements, refugee camps).

Appendix C

Calculation of the Variance of Kappa – VAR (Kappa)

The same mathematical representation of the confusion matrix (Table 2-2, page 51) that was used in determining kappa (κ - equation 2-38) is used here to determine the variance of kappa. This is denoted below as var(κ) in equation C-1. An approximate large sample variance of kappa is found using the Delta method (Congalton and Green, 1999) by the following equation:

$$\text{var}(k) = \frac{1}{n}\left\{\frac{q_1(1-q_1)}{(1-q_2)^2} + \frac{2(1-q_1)(2q_1q_2 - q_3)}{(1-q_2)^3} + \frac{(1-q_1)^2(q_4 - 4q_2^2)}{(1-q_2)^4}\right\} \qquad \text{(C-1)}$$

where

$$q_1 = \frac{1}{n}\sum_{i=1}^{k} n_{ii}$$

$$q_2 = \frac{1}{n^2}\sum_{i=1}^{k} n_{i+}n_{+i}$$

$$q_3 = \frac{1}{n^2}\sum_{i=1}^{k} n_{ii}(n_{i+} + n_{+i})$$

$$q_4 = \frac{1}{n^3}\sum_{i=1}^{k}\sum_{j=1}^{k} n_{ij}(n_{j+} + n_{+i})^2$$

Appendix D – Noise Approach and SNR Determination

It can said without any hesitation that the work conducted in attempting to add spectrally correlated noise to an image could have been considered a separate thesis on its own. This proved to be a very time consuming and frustrating effort. However, we succeeded in the implementation of a unique approach that utilizes the difference between the initial image and a noiseless image. The noiseless image was produced via an MNF-transform. The difference between these two images is a noise cube, which accounts for <u>all</u> of the noise along the image chain and it is spectrally correlated. This approach to noise was fully discussed win Chapter 3. This method seems quite simple and yields effective results. In arriving at this final method, two other approaches were first attempted. These first two approaches were more robust and theoretical sound yet never delivered expected results. It is necessary that these two approaches be discussed here in the event that more investigation into these methods is conducted and our "oversight" is found. After these methods are discussed, a more in-depth look at the problems associated with the final approach adopted will be discussed.

Prior to a look at these first two approaches, the reader is directed to the enclosed CD. In the directory "noise" one may find the different noise algorithms coded in IDL. In fact, the reader is asked to cross reference these program with the discussion here, since the programs contain a lot of comments with regards to their overall approach.

Common to the two methods is the use of the AVIRIS dark signal that comes with any image data that is ordered. This is the dark current image (224 channels x 1 sample x 512 lines) of the system and it is divided into two files. The first file, "*.drk1" contains the 12 most significant bits of data while "*.drk2" contains the 12 least significant bits of data. The data in these two files is in digital counts. The two dark current files are then combined to form a 24-bit number representing the total dark signal. This total dark signal is then divided by 4096 (2^{12}) and subsequently divided by the AVIRIS channel gains. The result of these steps is the dark current associated with AVIRIS in radiance units [micro-watts/cm^2/nm/sr]. One problem identified at

this step, is that these gains are provided with the AVIRIS data to convert the 16-bit image data from DC to radiance. When each image spectrum is divided by their respective gain factor the 16-bit integers are converted to radiance. The problem is that we are applying these gain factors (the only ones provided) designed for 16-bit numbers to 24-bit noise data. The results could prove to be wrong and this could haunt us later on. However, not having any other gain factors to use, we pressed on.

The next step was to spectrally resample the total dark current noise file to match the user specified spectral resolution of the image file that was to be degraded. Subsequently, the covariance of this total noise data was found along with the corresponding eigenvectors and eigenvalues of this matrix. These above steps are common to the two approaches that will be discussed and can be best seen in Figure D-1.
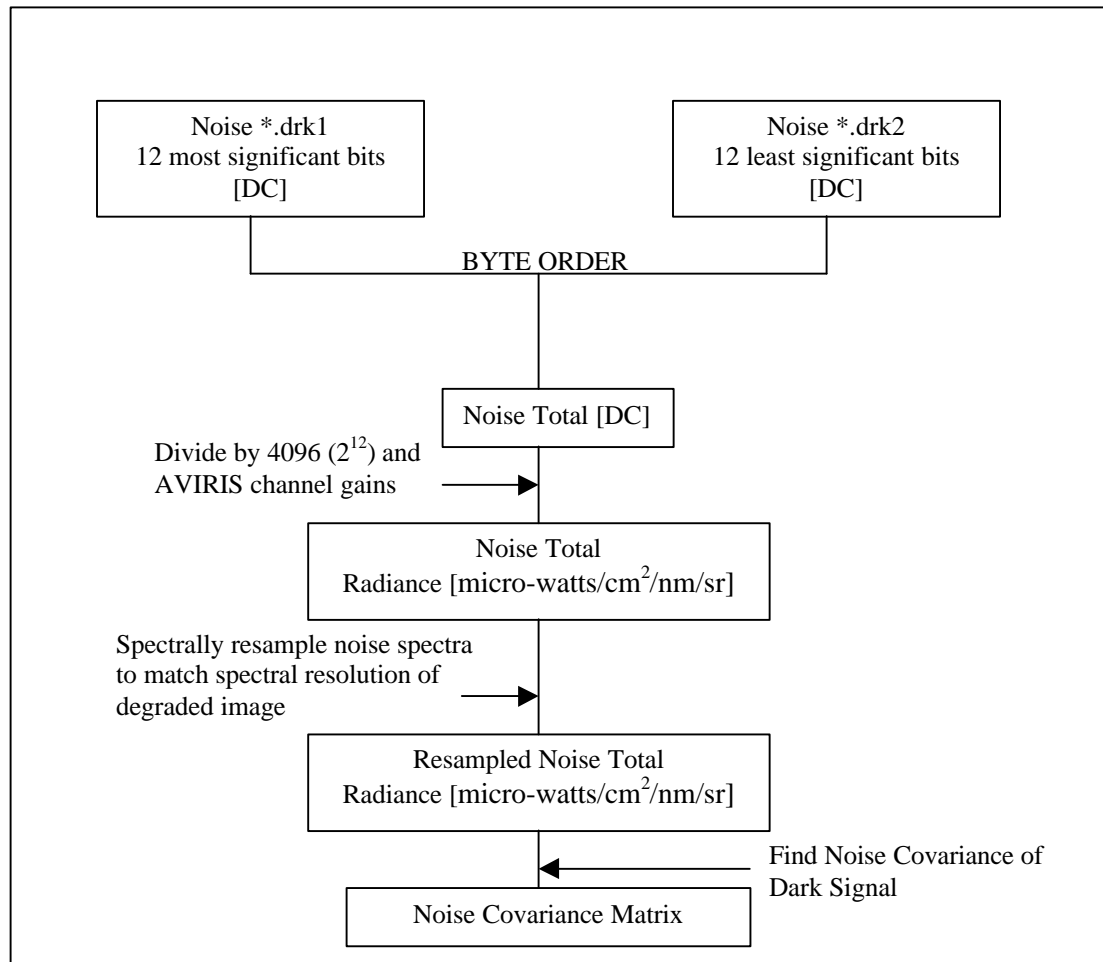
Figure D-1 Combination of AVIRIS *.drk1 and*.drk2 files to form Noise Covariance Matrix

The first real attempt at adding spectrally correlated noise to an image can be entitled the "Principal Components" approach and is depicted in Figure D-2. This method involved finding the covariance matrix of the entire image and adding this to the covariance matrix of the noise. The eigenvectors and eigenvalues of this new covariance matrix, of both image and noise, were then found. Following this addition, the original image was subjected to a forward principal component rotation using the eigenvectors and eigenvalues of the image. Subsequently, an inverse principal component rotation was applied to the image, yet this time using the eigenvalues from the covariance matrix of image and noise. The result of this manipulation was an image with spectrally correlated noise added to it from the AVIRIS dark current signals. If more noise was needed, a scaling factor was applied to the covariance matrix of the noise prior to addition to the image covariance matrix. The problem with this method was the assumption that the covariance of the image plus noise was simply equal to the covariance of the image plus the covariance of the noise as seen here:

$$\Sigma_{I+N} = \Sigma_I + \Sigma_N \qquad\qquad (D\text{-}1)$$

Where Σ denotes a covariance matrix while I and N correspond to image and noise respectively. Equation D-1 holds true for small values of noise but is fundamentally incorrect unless both I and N are completely independent or otherwise uncorrelated (Johnson and Wichern, 1998). This cannot be true for an imaging system since both image and noise are ultimately gathered by the same sensor. The correct equation should read:

$$\Sigma_{I+N} = \Sigma_I + \Sigma_N + 2\Sigma_{I \bullet N}$$

$$(D\text{-}2)$$

where $\Sigma_{I,N}$ is the covariance matrix between the image and the noise covariance matrices. Obviously we run into complications of how this is computed, but this explains why this approach was abandoned.

Other problems encountered with this approach that are found with other approaches, is the problem of scaling the eigenvalues of the noise covariance matrix or the covariance matrix itself so that the noise added to the initial imagery meets some user-defined SNR value. Evidently, the scaling problem added to the 16-bit and 24-bit data scaling problem starts to indicate a theme as to where these algorithms may have gone astray. Additionally, the spatial resampling of the image had to be taken into account as well. The method used here was a boxcar or aggregate process, which averages pixels and their corresponding noise value. A scaling factor for spatial degradation was also included in the process. The programs on the enclosed CD identify where these scaling factors were applied.

```
┌─────────────────────────────────────────────────────────────────────────────┐
│                                                                               │
│   ┌──────────────────────────────┐        ┌──────────────────────────────┐    │
│   │  Covariance Matrix of Image  │        │ Covariance Matrix of Noise - │    │
│   │  -Radiance                   │        │ Radiance                     │    │
│   │  COV(Image)                  │        │ COV(Noise)                   │    │
│   └──────────────────────────────┘        └──────────────────────────────┘    │
│                                                                               │
│   ┌──────────────────────────────┐        ┌──────────────────────────────┐    │
│   │  Find Eigenvalues and        │        │  Normalize COV(Noise) by     │    │
│   │  Eigenvectors of COV(Image)  │        │  dividing matrix by its      │    │
│   │                              │        │  maximum value               │    │
│   └──────────────────────────────┘        └──────────────────────────────┘    │
│                                                                               │
│          ┌──────────────────────────────────────────┐                          │
│          │  Forward PC Transform of Image using      │                          │
│          │  Eigenvalues and Eigenvectors of Image    │                          │
│          └──────────────────────────────────────────┘                          │
│                                                                               │
│       ┌──────────────────────────────────────────────┐                         │
│       │  Add Covariance of Image and Covariance of     │                        │
│       │  Noise to find new Image Covariance Matrix     │                        │
│       │  with Noise addition in decorrelated space     │                        │
│       │                                                │                        │
│       │  COV(New)=COV(Image)+ (LEVEL)*COV(Noise)       │                        │
│       └──────────────────────────────────────────────┘                         │
│                                                                               │
│   ┌──────────────────┐        ┌──────────────────────────────┐                 │
│   │ User-specified   │        │  Find Eigenvectors and       │                 │
│   │ Level of Noise   │        │  Eigenvalues of New          │                 │
│   │ to be added      │        │  Covariance Matrix           │                 │
│   │ LEVEL            │        │  COV(New)                    │                 │
│   └──────────────────┘        └──────────────────────────────┘                 │
│                                                                               │
│                    ┌──────────────────────────────┐                           │
│                    │  Use newly found Eigenvalues │                           │
│                    │  of COV(New) in conducting   │                           │
│                    │  an Inverse PC Transform of  │                           │
│                    │  Image                       │                           │
│                    └──────────────────────────────┘                           │
│                                                                               │
│                    ┌──────────────────────────────┐                           │
│                    │  Result: Image with          │                           │
│                    │  Spectrally Correlated Noise │                           │
│                    │  at scale level specified by │                           │
│                    │  user                        │                           │
│                    └──────────────────────────────┘                           │
│                                                                               │
└─────────────────────────────────────────────────────────────────────────────┘
```

Figure D-2: Principal Components Transform Approach to Correlated Noise Addition

The other abandoned approach, seen in Figure D-3, calculated the covariance matrix of the dark current noise files from the AVIRIS flight data. The eigenvalues of this covariance matrix were determined and scaled to reflect a user-defined SNR for the degraded image. The result is the noise variances for each band in de-correlated space. Subsequently, 2-D arrays of random numbers were created for each band based on a mean of zero and a standard deviation equal to the square root of its respective eigenvalues/variance. A series of these 2-D arrays stacked together form a 3-D cube of de-correlated noise with a size equal to the spatial dimensions of the image and the number of degraded spectral bands. A Principal Components transform was performed on this de-correlated noise cube so that the result of the transform is a correlated noise cube. Although this method also worked in producing noise, the covariance and correlation matrices of the noise never matched that presented in Boardman's work (1995), as they theoretically should. Like the other approach, it is believed that one of the problems here is that a scaling or conversion factor is missing in the overall calculation. Despite a great length of time and frustration this mystery was never fully solved.

In order to figure out the problems with the above two approaches, the covariance produced from the dark current data was compared to the covariance matrix produced from a dark uniform area within an AVIRIS image. Theoretically, these two covariance matrices should be very similar in the magnitude of each matrix element and overall structure. Unfortunately, there was a difference. Additionally, we tried matching the SNR produced via our method to that published in the literature. We did this by propagating a top of the atmosphere irradiance through a MODTRAN atmosphere and reflected this off of a 50% reflector to yield a signal. This signal was then divided by the AVIRIS dark current file that we produced. Theoretically, this should yield a similar SNR curve to that published with the AVIRIS system specifications (Boardman, 1995; Vane et al., 1993).   Once again, no match. It is firmly believed that these two methods are theoretically sound. The problem lies in some mysterious unknown scaling factor that would allow us to convert the dark current files into proper and usable units from which we may construct spectrally correlated noise. Given the time constraints to thesis completion and the lack of response from the AVIRIS office regarding this matter, we pressed on with the "MNF – subtraction" method.

```
┌─────────────────────────────────┐
│   Covariance Matrix of Noise    │
│          COV(Noise)             │
└─────────────────────────────────┘
                │
                ▼
┌─────────────────────────────────┐
│  Calculate Eigenvectors and     │········
│  Eigenvalues of COV(Noise)      │       :
└─────────────────────────────────┘       :
                                          :
┌──────────────────────┐                  :
│ LEVEL User-specified │                  :
│  Level of Noise to be│                  :
│ added by requested SNR│                 :
└──────────────────────┘                  :
          │                               :
          ▼                               :
┌─────────────────────────────────────────┐
│ New Noise Variance = (LEVEL)²*Eigenvalue(COV(Image)) │
└─────────────────────────────────────────┘
                │
                ▼
┌─────────────────────────────────────────┐
│ Generate 2-D Array of Random Numbers for each band. │
│ This 2-D Array will have a Normal Distribution of:  │
│ N(mean = 0, std dev = SQRT(New Noise Variance)).    │
│ There will be one of these arrays for each spectral band. │
└─────────────────────────────────────────┘
                │
                ▼
┌─────────────────────────────────────────┐
│ Stack 2-D Noise Arrays together to form 3-D │
│ Array of De-correlated Noise.               │
└─────────────────────────────────────────┘
                │
                ▼
┌─────────────────────────────────────────┐
│ Conduct Inverse PC Transform to 3-D Array of De- │····
│ correlated Noise using Eigenvectors and          │
│ Eigenvalues determined previously.               │
└─────────────────────────────────────────┘
                │
                ▼
┌─────────────────────────────────────────┐
│ Result: 3-D Array of Spectrally Correlated Noise │
└─────────────────────────────────────────┘
                │
                ▼
┌─────────────────────────────────────────┐
│ Spectrally and Spatially Resample 3-D Array of │
│ Noise to match User-specified Degradation Levels │
└─────────────────────────────────────────┘
```

Figure D-3: Random Numbers Approach to Correlated Noise Addition

One of the major problems encountered with the noise approach that was used in this thesis study (MNF method) was, again, related to scaling. Resampling the noise cube to match the spatial degradation of the subject image was necessary – but this essentially averages the noise. That is, the image had to be degraded spatially as well. In order to properly add the noise to this spatially resampled image, the noise image/array also had to be spatially resampled. The spatially resampling, which follows an aggregate process, averages neighbourhood pixels. This averaging process lowers the noise content of the resampled noise cube in comparison to the initial noise cube. In this respect, the initial noise cube had to be multiplied by a scaling factor that accounts for this spatial degradation. This scaling factor was included with the scaling factor determined from the user-specified SNR. This can be better seen in the programs included in the "noise" directory on the enclosed CD. To verify that the desired SNR was produced in the final degraded image, the average spectrum from an ROI of a bright area (eg. the beach - approximately 30% reflector) was assumed to be signal. This was subsequently divided by the standard deviation of an ROI over a dark portion of the image (water). This yields a SNR for each channel, which was then averaged to produce a mean SNR. This average SNR was compared to that input by the user. In most cases, a higher SNR had to be input by the user to produce the desired SNR in the final degraded image. For instance, if an image was spatially degraded to 80 m, then an input SNR of 290 yielded a SNR of approximately 220. However, at a spatial resolution of 40 m, an input SNR of 380 yields a SNR of 220. This scaling factor problem within the program was never completely solved, yet it was always accounted for by determining the SNR of each finally degraded image to ensure that it matched, as best as possible, the desired SNR levels selected as part of the factorial designed experiment. Sometimes, these SNR values were not *exact* matches, with respect to each other, and this explains for the error limits on SNR presented in Chapter 3. For example, an image with a GIFOV of 20 m would have a SNR of 225, while that of an image degraded to 80 m may have a calculated SNR of 220 – for a desired SNR experimental level of 225.

Appendix E – Summary of Collected Raw Data

AVIRIS - Rogers Dry Lake, CA

Spatial Res. (GIFOV - m)	Spectral Res. (nm)	Noise (SNR @30%)	SAM - Kappa	BE - Kappa	BE to SAMref - Kappa	GML - Kappa	GML to SAMref - Kappa	Unmixing SE	SMF PD @ PFA=0.01 (avg)	SMF PD @ PFA=0.001 (avg)	- SFF PD @ PFA=0.2 (avg)	- SFF PD @ PFA= 0.1 (avg)
20	110	225	0.8950	0.0000	0.0000	0.6601	0.6298	0.5394	0.6188	0.4824	0.6546	0.6182
20	110	100	0.6571	0.0000	0.0000	0.6467	0.6397	12.9222	0.6261	0.4817	0.6558	0.6182
20	110	10	0.0044	0.0000	0.0000	0.3034	0.3428	1630.7000	0.4036	0.1106	0.6278	0.5470
20	55	225	0.9564	0.3794	0.2324	0.7476	0.6117	4.6370	0.6199	0.4792	0.6536	0.6155
20	55	100	0.8146	0.3530	0.2206	0.7192	0.6190	115.6140	0.6169	0.4819	0.6540	0.6130
20	55	10	0.0023	0.0753	0.0454	0.3815	0.4241	14272.2000	0.5438	0.2596	0.4505	0.3085
20	10	225	0.9750	0.8906	0.3867	0.9999	0.5840	11.7397	0.6205	0.4753	0.6258	0.5944
20	10	100	0.8911	0.4915	0.2302	1.0000	0.5840	229.8970	0.6205	0.4753	0.6165	0.5958
20	10	10	0.0042	0.0687	0.0333	0.9999	0.5840	29356.9000	0.6205	0.4753	0.3843	0.1403
40	110	225	0.7849	0.0000	0.0000	0.6307	0.5642	0.9168	0.6017	0.4772	0.6579	0.6157
40	110	100	0.6227	0.0000	0.0000	0.6212	0.5742	12.8951	0.6088	0.4820	0.6586	0.6116
40	110	10	0.0276	0.0000	0.0000	0.2893	0.3121	1829.6300	0.3825	0.0937	0.6154	0.5054
40	55	225	0.8237	0.3628	0.2305	0.7054	0.5550	9.0211	0.5959	0.4680	0.6429	0.6014
40	55	100	0.7299	0.3373	0.2173	0.6788	0.5558	129.1830	0.5965	0.4797	0.6435	0.5997
40	55	10	0.0336	0.0700	0.0437	0.3545	0.3762	15937.5000	0.5196	0.2259	0.4348	0.2823
40	10	225	0.8330	0.7342	0.0363	0.7457	0.5211	19.1010	0.6002	0.4602	0.6244	0.5833
40	10	100	0.7856	0.4667	0.2537	0.7457	0.5211	254.3080	0.6003	0.4602	0.6183	0.5896
40	10	10	0.0084	0.0466	0.0246	0.7457	0.5211	32463.6000	0.6003	0.4602	0.3151	0.1434
80	110	225	0.6831	0.0000	0.0000	0.5807	0.5100	1.8884	0.5743	0.4472	0.6509	0.6011
80	110	100	0.5218	0.0000	0.0000	0.5585	0.5065	19.6861	0.5709	0.4492	0.6504	0.5913
80	110	10	0.0193	0.0000	0.0000	0.2229	0.2354	2354.2200	0.3470	0.0542	0.5980	0.4183
80	55	225	0.7226	0.3454	0.2256	0.6433	0.5064	19.8765	0.5844	0.4503	0.6233	0.5750
80	55	100	0.6409	0.3139	0.2091	0.6112	0.5029	166.1900	0.5700	0.4499	0.6225	0.5721
80	55	10	0.0216	0.0515	0.0297	0.3034	0.3056	20502.4000	0.4568	0.1970	0.3982	0.2142
80	10	225	0.7340	0.6167	0.3274	0.0000	0.0000	37.6322	0.5668	0.4513	0.6186	0.5601
80	10	100	0.6889	0.3790	0.2133	0.0000	0.0000	337.6670	0.5668	0.4513	0.6100	0.5697
80	10	10	0.0039	0.0305	0.0165	0.0000	0.0000	40847.3000	0.5668	0.4513	0.2195	0.1194

AVIRIS - Rochester, NY

Spatial Res. (GIFOV - m)	Spectral Res. (nm)	Noise (SNR @ 30%)	SAM - Kappa	BE - Kappa	BE to SAMref - Kappa	GML - Kappa	GML to SAMref - Kappa	Unmixing - SE	SMF PD @ PFA=0.01 (avg)	SMF PD @ PFA=0.001 (avg)	SFF PD @ PFA=0.6 (avg)	SFF PD @ PFA= 0.4 (avg)
20	110	225	0.8842	0.3392	0.4061	0.7463	0.5671	0.2263	0.5077	0.2739	0.3417	0.0473
20	110	100	0.5713	0.3380	0.3768	0.7434	0.5887	3.2502	0.4765	0.2560	0.3400	0.0542
20	110	10	0.0002	0.2044	0.2216	0.4313	0.5245	398.5200	0.4003	0.1856	0.4064	0.1850
20	55	225	0.9509	0.5348	0.4159	0.8428	0.5433	0.2395	0.5645	0.2905	0.2820	0.0266
20	55	100	0.7296	0.4657	0.3765	0.8180	0.5710	4.4890	0.4962	0.2724	0.2862	0.0406
20	55	10	0.0001	0.2404	0.2429	0.5247	0.5601	546.5260	0.4117	0.2190	0.4445	0.1953
20	10	225	0.9835	0.9367	0.4376	1.0000	0.5350	0.8408	0.6298	0.3187	0.3747	0.0367
20	10	100	0.8210	0.7785	0.4011	0.9999	0.5350	16.4545	0.6298	0.3187	0.3773	0.0431
20	10	10	0.0001	0.3888	0.2753	1.0000	0.5350	2101.0500	0.6298	0.3187	0.5792	0.1891
40	110	225	0.6649	0.3190	0.3760	0.6949	0.5019	0.3393	0.5055	0.2291	0.3245	0.0462
40	110	100	0.4449	0.3127	0.3549	0.7042	0.5312	3.2610	0.4655	0.2198	0.3237	0.0525
40	110	10	0.0136	0.1993	0.2146	0.4080	0.4381	447.3310	0.3623	0.1632	0.4159	0.1958
40	55	225	0.7024	0.4607	0.3766	0.7371	0.4712	0.4515	0.5423	0.2471	0.2786	0.0278
40	55	100	0.5120	0.4025	0.3404	0.7374	0.5041	5.0509	0.4784	0.2363	0.2797	0.0389
40	55	10	0.0077	0.2240	0.2226	0.4895	0.4404	609.8380	0.3900	0.1901	0.4377	0.2091
40	10	225	0.7231	0.7184	0.3828	0.7835	0.4733	1.3353	0.6136	0.2990	0.3406	0.0383
40	10	100	0.5960	0.6117	0.3460	0.7835	0.4733	18.4258	0.6136	0.2990	0.3442	0.0446
40	10	10	0.0001	0.3361	0.2458	0.7835	0.4733	2344.2900	0.6136	0.2990	0.5384	0.1914
80	110	225	0.5481	0.2979	0.3448	0.6374	0.4446	0.6075	0.4643	0.1733	0.2911	0.0486
80	110	100	0.3192	0.2937	0.3326	0.6344	0.4656	4.9703	0.4338	0.1690	0.2887	0.0532
80	110	10	0.0063	0.1801	0.1906	0.3800	0.3744	578.6550	0.3386	0.1195	0.4187	0.1977
80	55	225	0.5831	0.3957	0.3350	0.6127	0.3933	0.9282	0.4993	0.1905	0.2532	0.0291
80	55	100	0.3901	0.3410	0.2963	0.6236	0.4262	6.5745	0.4690	0.1804	0.2532	0.0402
80	55	10	0.0029	0.1897	0.1913	0.4145	0.3654	790.6190	0.3329	0.1296	0.4457	0.2061
80	10	225	0.6093	0.5983	0.3315	0.6110	0.3852	2.9017	0.5518	0.2332	0.3061	0.0403
80	10	100	0.4545	0.5014	0.2956	0.6110	0.3852	25.0320	0.5518	0.2332	0.3063	0.0486
80	10	10	0.0000	0.2661	0.2097	0.6110	0.3852	3012.2800	0.5518	0.2332	0.5027	0.1988

DIRSIG – Western Rainbow

Spatial Res. (GIFOV - m)	Spectral Res. (nm)	Noise (SNR@ 30%)	SAM - Kappa GT	SAM - Kappa SAMref	BE - Kappa GT	BE - Kappa Beref	GML - Kappa GT	GML - Kappa GMLref	Unmixing-SE	SMF PD @ PFA=0.001 (avg)	SMF PD @ PFA=0.0001 (avg)	SFF PD @ PFA=0.2 (avg)	SFF PD @ PFA=0.05 (avg)
2	113	200	0.8106	0.9973	0.9973	0.9695	0.9686	0.9954	0.0411	0.9606	0.8307	0.9880	0.262
2	113	100	0.7987	0.9757	0.9757	0.9642	0.9644	0.9911	0.0536	0.9606	0.1461	0.9511	0.228
2	113	10	0.0122	0.0183	0.0183	0.3789	0.8178	0.8238	1.6971	0.1299	0.0000	0.3763	0.199
2	55	200	0.8098	0.9964	0.9126	0.9907	0.9693	0.9966	0.0315	0.9606	0.8701	0.9356	0.329
2	55	100	0.7409	0.9187	0.9115	0.9810	0.9664	0.9941	0.0434	0.9520	0.8465	0.9266	0.345
2	55	10	0.0001	0.0001	0.4033	0.4386	0.9219	0.9328	1.6385	0.1614	0.0197	0.3626	0.280
2	10	200	0.8074	0.9943	0.9137	0.9947	0.9695	1.0000	0.0282	0.9606	0.9029	0.4725	0.368
2	10	100	0.6462	0.8211	0.9135	0.9908	0.9674	0.9958	0.0351	0.9606	0.8976	0.4645	0.359
2	10	10	0.0000	0.0000	0.5288	0.5722	0.8799	0.8857	0.8505	0.9190	0.7835	0.0895	0.036
4	113	200	0.8688	0.7961	0.8677	0.8873	0.8893	0.9121	0.0332	0.8537	0.3969	0.8876	0.031
4	113	100	0.8471	0.7817	0.8653	0.8854	0.8944	0.9185	0.0477	0.7915	0.0354	0.7368	0.028
4	113	10	0.0046	0.0053	0.2896	0.3136	0.7212	0.7288	1.8914	0.0118	0.0000	0.1714	0.028
4	55	200	0.8653	0.7948	0.8709	0.9056	0.8801	0.9047	0.0265	0.8561	0.5888	0.6812	0.103
4	55	100	0.7580	0.7089	0.8691	0.9019	0.8938	0.9145	0.0417	0.8741	0.3555	0.6049	0.109
4	55	10	0.0001	0.0001	0.3654	0.3905	0.8705	0.8790	1.8540	0.0630	0.0000	0.2282	0.196
4	10	200	0.8496	0.7837	0.8689	0.9023	0.8801	0.9051	0.0262	0.9539	0.5971	0.4066	0.108
4	10	100	0.6093	0.5810	0.8753	0.9074	0.8967	0.9166	0.0334	0.9667	0.6165	0.3378	0.094
4	10	10	0.0000	0.0000	0.4497	0.4815	0.8317	0.8392	1.2655	0.8587	0.4496	0.2891	0.074
8	113	200	0.8406	0.7657	0.8370	0.8482	0.8018	0.8213	0.0294	0.6288	0.1008	0.5979	0.015
8	113	100	0.8222	0.7499	0.8326	0.8436	0.8100	0.8274	0.0422	0.5389	0.0000	0.4745	0.034
8	113	10	0.0029	0.0030	0.2615	0.2809	0.6815	0.6879	1.6748	0.0000	0.0000	0.1697	0.028
8	55	200	0.8317	0.7582	0.8442	0.8726	0.8014	0.8222	0.0273	0.7095	0.3242	0.4607	0.015
8	55	100	0.6440	0.5965	0.8423	0.8685	0.8081	0.8274	0.0528	0.6417	0.0000	0.1831	0.023
8	55	10	0.0000	0.0000	0.2670	0.2843	0.7950	0.8001	3.5095	0.1181	0.0000	0.1608	0.123
8	10	200	0.8022	0.7335	0.8516	0.8782	0.0000	0.0000	0.0234	0.8211	0.3259	0.0755	0.004
8	10	100	0.4787	0.4518	0.8458	0.8715	0.0000	0.0000	0.0388	0.7884	0.3259	0.2094	0.012
8	10	10	0.0000	0.0000	0.3326	0.3587	0.0000	0.0000	2.5471	0.6083	0.0225	0.0439	0.012

Appendix F – Classification Algorithm Comparisons using Kappa

In this section we wish to statistically compare the classification products from two different degraded images. The results presented in this Appendix are those associated with the use of equation 2-40 in Chapter 2. For each image a kappa value is calculated as is a variance of this kappa value. Using equation 2-40 we are able to compute a Z-score that can be used in testing the following hypothesis set.

H_o: (κ_1 - κ_2)=0 (i.e. there is NO difference between classification products)

H_1: (κ_1 - κ_2)≠0 (i.e. there is a difference between classification products)

H_o is rejected if $Z \geq Z_{\alpha/2}$ and the alternate hypothesis of H_1 is accepted.

The results are best presented in a square matrix format with the set of degraded images making up the columns and rows. The notation is as follows : *x* represents the GIFOV, *s* represents spectral resolution and *n* represents the SNR value. For example, each cell represents the results of the statistical hypothesis test done by indicating the level of confidence that the null hypothesis was rejected at. Obviously the diagonal of these matrices should indicate that there is no statistical difference. The colour code displayed in Table F-1 indicates the confidence levels at which the null hypothesis was rejected. For instance, if a cell is coloured blue, this means that there is a difference between the product of degraded image 1 and degraded image 2 at the 95% confidence level. These results were produced by using the program *Zcompare.c* (found on the enclosed CD) and MS Excel.

For example, this sort of hypothesis testing allows us to statistically test whether there is a difference in the SAM product produced by an image with a spectral resolution of 20 m and

spectral resolution of 10 nm and the SAM product from the same image with resolutions of 40 m and 55 nm respectively.

The following tables are grouped first by image and then by classification algorithm. With respect to each of the AVIRIS images, there are two tables for the BE and GML algorithms. One of the tables is derived from using the initial SAM results (SAMref) as ground truth. The other table is derived from using the initial BE and GML results of the non-degraded image as ground truth. These are denoted as BEref and GMLref.

The results discovered from examination of the following tables were not as remarkable as expected. It was anticipated that obvious patterns would be apparent which would be easily attributed to spatial resolution, spectral resolution and noise changes. Patterns are noticed but they were not as earth shattering as anticipated. This can be easily seen in the following tables by the abundance of "white" cells which represent a rejection of the null hypothesis at an 80% confidence limit – the lowest limit tested. This was exactly what was _not_ expected. However, from an examination of these tables, it can be generally concluded that any degradation of the initial image will result in an immediate difference in the information conveyed by the produced thematic map compared to the original. Likewise, when one "degraded" classification product was statistically compared to another "degraded" classification product the null hypothesis (see section 2.7.4) was rejected the majority of the time at the lowest confidence limit tested. This means that there is an immediate difference between classification maps produced from an image at one level of degradation compared to another. In other words, an image acquired with a certain spectral resolution, spatial resolution and noise characteristics will produce a different thematic map than the same image collected with different acquisition parameters. In this regard, a conclusion is drawn from this type of analysis. Furthermore, this type of analysis has proven itself quite useful.

The following tables are now presented with brief comments below each table, where appropriate. An example of this type of commentary is as follows: "As seen in Table F-2, little difference is found between those images associated with a low SNR value. This is seen by the rejection of the null hypothesis with noisy images at higher confidence limits. This is because the kappa values for these noisy images is quite low."

	80%
(white)	80%
(yellow)	85%
(red)	90%
(blue)	95%
(green)	99%
(black)	No difference

Table F-1: Confidence Level Colour Key

	x20s110n225	x20s110n100	x20s110n10	x20s55n225	x20s55n100	x20s55n10	x20s10n225	x20s10n100	x20s10n10	x40s110n225	x40s110n100	x40s110n10	x40s55n225	x40s55n100	x40s55n10	x40s10n225	x40s10n100	x40s10n10	x80s110n225	x80s110n100	x80s110n10	x80s55n225	x80s55n100	x80s55n10	x80s10n225	x80s10n100	x80s10n10
x20s110n225	black																										
x20s110n100		black																									
x20s110n10			black			black			black						red			black			blue			black			black
x20s55n225				black																							
x20s55n100					black																						
x20s55n10			black			black			black						red			black			blue			black			black
x20s10n225							black																				
x20s10n100								black																			
x20s10n10			black			black			black						red			black			blue			black			black
x40s110n225										black																	
x40s110n100											black																
x40s110n10												black			green						red						
x40s55n225													black														
x40s55n100														black													
x40s55n10			red			red			red			green			black			red			black			black			red
x40s10n225																black											
x40s10n100																	black										
x40s10n10			black			black			black						red			black			blue			black			black
x80s110n225																			black								
x80s110n100																				black							
x80s110n10			blue			blue			blue			red			black			blue			black			black			blue
x80s55n225																						black					
x80s55n100																							black				
x80s55n10			black			black			black						black			black			black			black			black
x80s10n225																									black		
x80s10n100																										black	
x80s10n10			black			black			black						red			black			blue			black			black

Table F-2: Z-Comparison of Classification Maps produced by the SAM algorithm – Rochester

3

	x20s110n225	x20s110n100	x20s110n10	x20s55n225	x20s55n100	x20s55n10	x20s10n225	x20s10n100	x20s10n10	x40s110n225	x40s110n100	x40s110n10	x40s55n225	x40s55n100	x40s55n10	x40s10n225	x40s10n100	x40s10n10	x80s110n225	x80s110n100	x80s110n10	x80s55n225	x80s55n100	x80s55n10	x80s10n225	x80s10n100	x80s10n10
x20s110n225	■	■																red					■				
x20s110n100	■	■																■					red				
x20s110n10			■																								
x20s55n225				■																							
x20s55n100					■																						
x20s55n10						■																					
x20s10n225							■																				
x20s10n100								■																			
x20s10n10									■																		
x40s110n225										■																	
x40s110n100											■																
x40s110n10												■															
x40s55n225													■														
x40s55n100														■													
x40s55n10															■												
x40s10n225																■											
x40s10n100																	■										
x40s10n10	red	■																■									
x80s110n225																			■	yellow							
x80s110n100																			yellow	■							
x80s110n10																					■						
x80s55n225																						■					
x80s55n100	■	red																					■				
x80s55n10																								■			
x80s10n225																									■		
x80s10n100																										■	
x80s10n10																											■

Table F-3: Z-Comparison of Classification Maps produced by the BE algorithm compared to BE Ground Truth– Rochester

In Table F-3 it is interesting to note that those images with a spatial resolution of 20 m, a spectral resolution of 110 nm and a SNR value of 225 <u>or</u> 100 show no difference in the kappa value (and corresponding classification maps).

Table F-4: Z-Comparison of Classification Maps produced by the BE algorithm compared to SAMref Ground Truth– Rochester

	x20s110n225	x20s110n100	x20s110n10	x20s55n225	x20s55n100	x20s55n10	x20s10n225	x20s10n100	x20s10n10	x40s110n225	x40s110n100	x40s110n10	x40s55n225	x40s55n100	x40s55n10	x40s10n225	x40s10n100	x40s10n10	x80s110n225	x80s110n100	x80s110n10	x80s55n225	x80s55n100	x80s55n10	x80s10n225	x80s10n100	x80s10n10
x20s110n225	■																										
x20s110n100		■			■					■			■														
x20s110n10			■												■												
x20s55n225				■																							
x20s55n100		■			■					■			■														
x20s55n10						■												R									
x20s10n225							■																				
x20s10n100								■																			
x20s10n10									■																		
x40s110n225		■			■					■			■														
x40s110n100											■																
x40s110n10												■															
x40s55n225		■			■					■			■														
x40s55n100														■					Y								
x40s55n10			■												■												
x40s10n225																■											
x40s10n100																	■		■								
x40s10n10						R												■									
x80s110n225														Y			■		■		■						
x80s110n100																				■		G				■	
x80s110n10																			■		■			■			
x80s55n225																				G		■			Y		
x80s55n100																							■			■	
x80s55n10																					■			■			
x80s10n225																						Y			■		
x80s10n100																				■			■			■	
x80s10n10																											■

5

	x20s110n225	x20s110n100	x20s110n10	x20s55n225	x20s55n100	x20s55n10	x20s10n225	x20s10n100	x20s10n10	x40s110n225	x40s110n100	x40s110n10	x40s55n225	x40s55n100	x40s55n10	x40s10n225	x40s10n100	x40s10n10	x80s110n225	x80s110n100	x80s110n10	x80s55n225	x80s55n100	x80s55n10	x80s10n225	x80s10n100	x80s10n10
x20s110n225	■	red																									
x20s110n100	red	■																									
x20s110n10			■																								
x20s55n225				■																							
x20s55n100					■																						
x20s55n10						■																					
x20s10n225							■		■																		
x20s10n100								■																			
x20s10n10							■		■																		
x40s110n225										■																	
x40s110n100											■																
x40s110n10												■															
x40s55n225													■	■													
x40s55n100													■	■													
x40s55n10															■												
x40s10n225																■	■	■									
x40s10n100																■	■	■									
x40s10n10																■	■	■									
x80s110n225																			■	blue							
x80s110n100																			blue	■							
x80s110n10																					■						
x80s55n225																						■			■	■	■
x80s55n100																							■				
x80s55n10																								■	■		
x80s10n225																									■	■	■
x80s10n100																									■	■	■
x80s10n10																									■	■	■

Table F-5: Z-Comparison of Classification Maps produced by the GML algorithm to GML Ground Truth – Rochester

In Table F-5, there is a lack of significant difference when the image is spectrally sampled at 10 nm despite the level of noise. This is evident with those images having a spatial resolution of 40 m and 80 m. This trend starts to appear at a spatial resolution of 20 m and a spectral resolution of 10 nm.

	x20s110n225	x20s110n100	x20s110n10	x20s55n225	x20s55n100	x20s55n10	x20s10n225	x20s10n100	x20s10n10	x40s110n225	x40s110n100	x40s110n10	x40s55n225	x40s55n100	x40s55n10	x40s10n225	x40s10n100	x40s10n10	x80s110n225	x80s110n100	x80s110n10	x80s55n225	x80s55n100	x80s55n10	x80s10n225	x80s10n100	x80s10n10
x20s110n225	■				🟡																						
x20s110n100		■																									
x20s110n10			■																								
x20s55n225				■																							
x20s55n100	🟡				■																						
x20s55n10						■																					
x20s10n225							■	■	■	🟡																	
x20s10n100							■	■	■																		
x20s10n10							■	■	■	🟡																	
x40s110n225										■				■													
x40s110n100							🟡	🟡	🟡		■																
x40s110n10												■															
x40s55n225													■			■	■										
x40s55n100														■													
x40s55n10													■		■				🟡								
x40s10n225														■		■	■										
x40s10n100																■	■										
x40s10n10																■		■									
x80s110n225															🟡				■								
x80s110n100																				■							
x80s110n10																					■						
x80s55n225																						■					
x80s55n100																							■				
x80s55n10																								■			
x80s10n225																									■		■
x80s10n100																										■	
x80s10n10																									■		■

Table F-6: Z-Comparison of Classification Maps produced by the GML algorithm to SAMref Ground Truth – Rochester

In Table F-6, there is a lack of significant difference when the image is spectrally sampled at 10 nm despite the level of noise. This is most evident with those images having a spatial resolution of 20 m and 40 m.

	x20s110n225	x20s110n100	x20s110n10	x20s55n225	x20s55n100	x20s55n10	x20s10n225	x20s10n100	x20s10n10	x40s110n225	x40s110n100	x40s110n10	x40s55n225	x40s55n100	x40s55n10	x40s10n225	x40s10n100	x40s10n10	x80s110n225	x80s110n100	x80s110n10	x80s55n225	x80s55n100	x80s55n10	x80s10n225	x80s10n100	x80s10n10
x20s110n225	black																										
x20s110n100		black																									
x20s110n10			black			black			black									blue									black
x20s55n225				black																							
x20s55n100					black																						
x20s55n10			black			black			black									yellow									black
x20s10n225							black																				
x20s10n100								black																			
x20s10n10			black			black			black									blue									black
x40s110n225										black						black											
x40s110n100											black																
x40s110n10												black															
x40s55n225													black														
x40s55n100														black													
x40s55n10															black												
x40s10n225										black						black											
x40s10n100								black									black										
x40s10n10			blue			yellow			blue									black									red
x80s110n225																			black								
x80s110n100																				black							
x80s110n10																					black			black			
x80s55n225																						black					
x80s55n100																							black				
x80s55n10																					black			black			
x80s10n225																									black		
x80s10n100																										black	
x80s10n10			black			black			black									red									black

Table F-7: Z-Comparison of Classification Maps produced by the SAM algorithm – Rogers Dry Lake

	x20s110n225	x20s110n100	x20s110n10	x20s55n225	x20s55n100	x20s55n10	x20s10n225	x20s10n100	x20s10n10	x40s110n225	x40s110n100	x40s110n10	x40s55n225	x40s55n100	x40s55n10	x40s10n225	x40s10n100	x40s10n10	x80s110n225	x80s110n100	x80s110n10	x80s55n225	x80s55n100	x80s55n10	x80s10n225	x80s10n100	x80s10n10
x20s110n225	■	■	■							■	■	■							■	■	■						
x20s110n100	■	■	■							■	■	■							■	■	■						
x20s110n10	■	■	■							■	■	■							■	■	■						
x20s55n225				■																						■	
x20s55n100					■																						
x20s55n10						■																					
x20s10n225							■																				
x20s10n100								■																			
x20s10n10									■						■												
x40s110n225	■	■	■							■	■	■							■	■	■						
x40s110n100	■	■	■							■	■	■							■	■	■						
x40s110n10	■	■	■							■	■	■							■	■	■						
x40s55n225													■														
x40s55n100														■													
x40s55n10									■						■												
x40s10n225																■											
x40s10n100																	■										
x40s10n10																		■									
x80s110n225	■	■	■							■	■	■							■	■	■						
x80s110n100	■	■	■							■	■	■							■	■	■						
x80s110n10	■	■	■							■	■	■							■	■	■						
x80s55n225																						■					
x80s55n100																							■				
x80s55n10																								■			
x80s10n225																									■		
x80s10n100			■																							■	
x80s10n10																											■

Table F-8: Z-Comparison of Classification Maps produced by the BE algorithm to BE Ground Truth – Rogers Dry Lake

Interesting results from Table F-8 show that there is no difference in kappa, and the corresponding spectral product, whenever a coarse spectral resolution of 110 nm is used for the BE algorithm. This lack of difference is *independent* of the spatial resolution and the noise content of the image. From cross-referencing the results shown here to Figure 4-5, we see that a spectral resolution of 110 nm produced very low values of kappa – thereby explaining this "checkerboard" appearance.

	x20s110n225	x20s110n100	x20s110n10	x20s55n225	x20s55n100	x20s55n10	x20s10n225	x20s10n100	x20s10n10	x40s110n225	x40s110n100	x40s110n10	x40s55n225	x40s55n100	x40s55n10	x40s10n225	x40s10n100	x40s10n10	x80s110n225	x80s110n100	x80s110n10	x80s55n225	x80s55n100	x80s55n10	x80s10n225	x80s10n100	x80s10n10
x20s110n225	B	B	B							B	B	B							B	B	B						
x20s110n100	B	B	B							B	B	B							B	B	B						
x20s110n10	B	B	B							B	B	B							B	B	B						
x20s55n225				B				G					B														
x20s55n100					B									R													
x20s55n10						B									B												
x20s10n225							B																				
x20s10n100				G				B																			
x20s10n10									B							Y											
x40s110n225	B	B	B							B	B	B							B	B	B						
x40s110n100	B	B	B							B	B	B							B	B	B						
x40s110n10	B	B	B							B	B	B							B	B	B						
x40s55n225				B									B														
x40s55n100					R									B												Y	
x40s55n10						B									B												
x40s10n225									Y							B											
x40s10n100																	B										
x40s10n10																		B									
x80s110n225	B	B	B							B	B	B							B	B	B						
x80s110n100	B	B	B							B	B	B							B	B	B						
x80s110n10	B	B	B							B	B	B							B	B	B						
x80s55n225																						B					
x80s55n100																							B			Y	
x80s55n10																								B			
x80s10n225																									B		
x80s10n100														Y									Y			B	
x80s10n10																											B

Table F-9: Z-Comparison of Classification Maps produced by the BE algorithm to SAMref
Ground Truth – Rogers Dry Lake

Again similar results here in Table F-9 as those previously seen in Table F-8. This again shows that there is no difference in kappa, and the corresponding spectral product, whenever a coarse spectral resolution of 110 nm is used for the BE algorithm. This lack of difference is independent of the noise content of the image and the spatial resolution.

	x20s110n225	x20s110n100	x20s110n10	x20s55n225	x20s55n100	x20s55n10	x20s10n225	x20s10n100	x20s10n10	x40s110n225	x40s110n100	x40s110n10	x40s55n225	x40s55n100	x40s55n10	x40s10n225	x40s10n100	x40s10n10	x80s110n225	x80s110n100	x80s110n10	x80s55n225	x80s55n100	x80s55n10	x80s10n225	x80s10n100	x80s10n10
x20s110n225	■																										
x20s110n100		■																				🟨					
x20s110n10			■																					■			
x20s55n225				■												🟩	🟩	🟩									
x20s55n100					■																						
x20s55n10						■																					
x20s10n225							■		■																		
x20s10n100								■																			
x20s10n10							■		■																		
x40s110n225										■																	
x40s110n100											■																
x40s110n10												■															
x40s55n225													■														
x40s55n100														■													
x40s55n10															■												
x40s10n225				🟩												■	■	■									
x40s10n100				🟩												■	■	■									
x40s10n10				🟩												■	■	■									
x80s110n225																			■								
x80s110n100																				■							
x80s110n10																					■						
x80s55n225		🟨																				■					
x80s55n100																							■				
x80s55n10			■																					■			
x80s10n225																									■	■	■
x80s10n100																									■	■	■
x80s10n10																									■	■	■

Table F-10: Z-Comparison of Classification Maps produced by the GML algorithm to GML Ground Truth – Rogers Dry Lake

The results in Table F-10 demonstrate that there is no difference in the values of kappa, and their corresponding thematic maps, whenever a fine spectral resolution (like 10 nm) is used – regardless of the noise level within the image. This corresponds to results presented in Chapter 4 where the level of noise associated with the images did not effect the performance of the GML algorithm as drastically as noise effected the SAM and BE algorithms. This can be attributed to the prior application of the MNF transform to the image data and use of fine spectral resolution.

Table F-11: Z-Comparison of Classification Maps produced by the GML algorithm to SAMref Ground Truth – Rogers Dry Lake

The results in Table F-11 demonstrate the same concepts as those discussed for Table F-10. The differences between these two tables may be attributed to use of a different spectral product as ground truth.

Table F-12: Z-Comparison of Classification Maps produced by the SAM algorithm to Ground Truth – Western Rainbow

Of note in Table F-12 is that there is little to no difference in the kappa values associated with images characterized by a low SNR value (SNR =10). In fact these cells correspond to very low values of kappa that were initially charted in Figure 4-9. It was in Figure 4-9 that the adverse effects of noise with respect to SAM performance were first observed.

	x2s113n200	x2s113n100	x2s113n10	x2s55n200	x2s55n100	x2s55n10	x2s10n200	x2s10n100	x2s10n10	x4s113n200	x4s113n100	x4s113n10	x4s55n200	x4s55n100	x4s55n10	x4s10n200	x4s10n100	x4s10n10	x8s113n200	x8s113n100	x8s113n10	x8s55n200	x8s55n100	x8s55n10	x8s10n200	x8s10n100	x8s10n10
x2s113n200	■			■			🟥																				
x2s113n100		■																									
x2s113n10			■																								
x2s55n200	■			■																							
x2s55n100					■																						
x2s55n10						■			■			🟩									■			■			■
x2s10n200	🟥						■																				
x2s10n100								■																			
x2s10n10						■			■			🟩									■			■			■
x4s113n200										■			■														
x4s113n100											■					■											
x4s113n10						🟩			🟩			■						🟩			■			🟩			🟩
x4s55n200										■			■														
x4s55n100														■													
x4s55n10															■												
x4s10n200											■					■											
x4s10n100																	■										
x4s10n10						■			■			🟩						■			■			■			■
x8s113n200																			■								
x8s113n100																				■							
x8s113n10						■			■			■						■			■			■			■
x8s55n200																						■					
x8s55n100																							■				
x8s55n10						■			■			🟩						■			■			■			■
x8s10n200																									■		
x8s10n100																										■	
x8s10n10						■			■			🟩						■			■			■			■

Table F-13: Z-Comparison of Classification Maps produced by the SAM algorithm to SAMref Ground Truth – Western Rainbow

The discussion regarding the behaviour seen here in Table F-13 is equivalent to the discussion regarding Table F-12.

Table F-14: Z-Comparison of Classification Maps produced by the BE algorithm to Ground Truth
Western Rainbow

In Table F-14 a trend begins to occur within the red circled area. It seems that at a spatial resolution of 4m, there is little difference between the kappa values associated with 55 nm or 113 nm spectral resolution. This means that there is little difference in the classification maps produced at these levels. This trend is similar to the indicated by the blue circle in that there is a little to no difference in those products with the original 2 m spatial resolution, relatively high SNR (200 and 100) and a spectral resolution of either 55 or 11m nm. This can be attributed to the initial spectral purity of the pixels with the DIRSIG scene and lack of mixed pixels

	x2s113n200	x2s113n100	x2s113n10	x2s55n200	x2s55n100	x2s55n10	x2s10n200	x2s10n100	x2s10n10	x4s113n200	x4s113n100	x4s113n10	x4s55n200	x4s55n100	x4s55n10	x4s10n200	x4s10n100	x4s10n10	x8s113n200	x8s113n100	x8s113n10	x8s55n200	x8s55n100	x8s55n10	x8s10n200	x8s10n100	x8s10n10
x2s113n200	■																										
x2s113n100		■																									
x2s113n10			■																								
x2s55n200				■				■																			
x2s55n100					■																						
x2s55n10						■																					
x2s10n200							■																				
x2s10n100					■			■																			
x2s10n10									■																		
x4s113n200										■																	
x4s113n100											■																
x4s113n10												■															
x4s55n200													■	🟨		🟨	■										
x4s55n100													🟨	■		■											
x4s55n10															■												
x4s10n200													🟨			■											
x4s10n100													■				■										
x4s10n10																		■									
x8s113n200																			■								
x8s113n100																				■							
x8s113n10																					■			🟦			
x8s55n200																						■	🟨		■		
x8s55n100																						🟨	■			🟥	
x8s55n10																					🟦			■			
x8s10n200																									■		
x8s10n100																						■	🟥			■	
x8s10n10																											■

Table F-15: Z-Comparison of Classification Maps produced by the BE algorithm to BEref
Ground Truth – Western Rainbow

	x2s113n200	x2s113n100	x2s113n10	x2s55n200	x2s55n100	x2s55n10	x2s10n200	x2s10n100	x2s10n10	x4s113n200	x4s113n100	x4s113n10	x4s55n200	x4s55n100	x4s55n10	x4s10n200	x4s10n100	x4s10n10	x8s113n200	x8s113n100	x8s113n10	x8s55n200	x8s55n100	x8s55n10	x8s10n200	x8s10n100	x8s10n10
x2s113n200	■			■	yellow		■	green																			
x2s113n100		■			yellow																						
x2s113n10			■																								
x2s55n200	■			■			■	yellow																			
x2s55n100	yellow	yellow			■			■																			
x2s55n10						■																					
x2s10n200	■			■			■	yellow																			
x2s10n100	green			yellow	■		yellow	■																			
x2s10n10									■				■			■											
x4s113n200										■																	
x4s113n100											■			■			blue										
x4s113n10												■															
x4s55n200										■			■			■											
x4s55n100											■			■			yellow										
x4s55n10															■												
x4s10n200										■			■			■											
x4s10n100											blue			yellow			■										
x4s10n10																		■									
x8s113n200																			■			■					
x8s113n100																				■			■				
x8s113n10																					■						
x8s55n200																			■			■					
x8s55n100																				■			■				
x8s55n10																								■			
x8s10n200																									■	■	■
x8s10n100																									■	■	■
x8s10n10																									■	■	■

Table F-16: Z-Comparison of Classification Maps produced by the GML algorithm to Ground Truth – Western Rainbow

The same discussion regarding Table F-14 holds true, to a certain extent, for Table F-16 above.

	x2s113n200	x2s113n100	x2s113n10	x2s55n200	x2s55n100	x2s55n10	x2s10n200	x2s10n100	x2s10n10	x4s113n200	x4s113n100	x4s113n10	x4s55n200	x4s55n100	x4s55n10	x4s10n200	x4s10n100	x4s10n10	x8s113n200	x8s113n100	x8s113n10	x8s55n200	x8s55n100	x8s55n10	x8s10n200	x8s10n100	x8s10n10
x2s113n200	■							■																			
x2s113n100		■																									
x2s113n10			■																blue	yellow		■	yellow				
x2s55n200				■																							
x2s55n100					■																						
x2s55n10						■																					
x2s10n200							■																				
x2s10n100	■							■																			
x2s10n10									■																		
x4s113n200										■				red													
x4s113n100											■						blue										
x4s113n10												■															
x4s55n200													■			■											
x4s55n100										red				■			blue										
x4s55n10															■												
x4s10n200													■			■											
x4s10n100											blue			blue			■										
x4s10n10																		■									
x8s113n200			blue																■			■					
x8s113n100			yellow																	■			■				
x8s113n10																					■						
x8s55n200			■																■			■					
x8s55n100			yellow																	■			■				
x8s55n10																								■			
x8s10n200																									■	■	■
x8s10n100																									■	■	■
x8s10n10																									■	■	■

Table F-17: Z-Comparison of Classification Maps produced by the GML algorithm to GMLref
Ground Truth – Western Rainbow

Appendix G – Contents of Enclosed CD

Enclosed with this thesis is a CD (see back cover envelope) which complements the written portion of this thesis study. The enclosed CD is readable on PC and UNIX platforms. The CD contains several different items including programs used to degrade the images, batch programs to process the imagery, initial copies of the images used in this study, all of the data collected, statistical analysis of this data, a softcopy of this written report and defence presentations. The directory structure of the CD closely follows that depicted in Figure G-1. It is important that the file entitled "READ_ME_FIRST" is opened when using the contents from this CD.
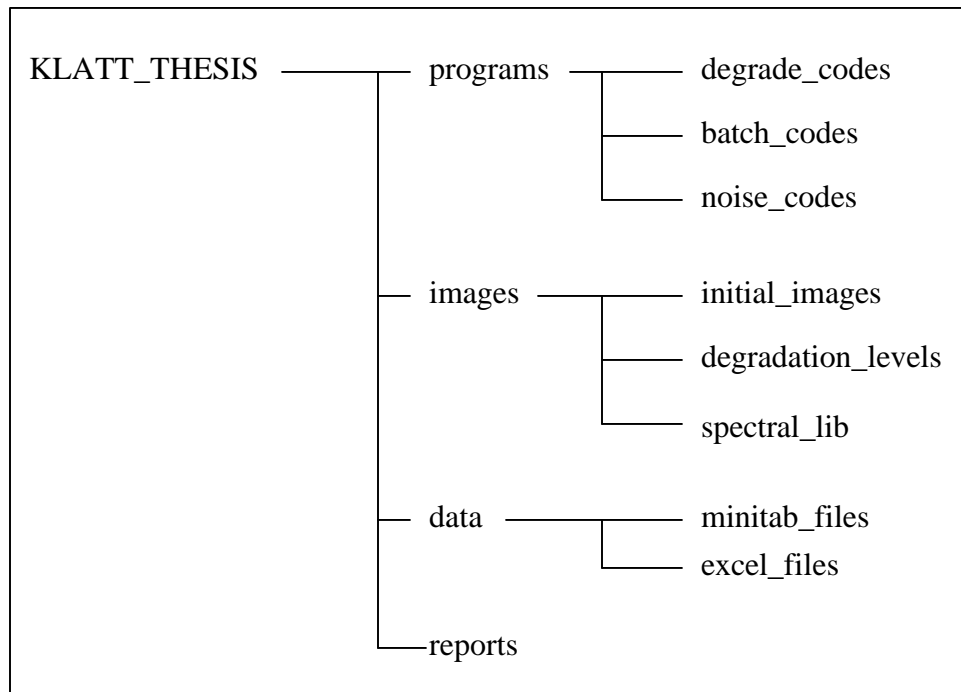
Figure G-1: Directory Structure of Enclosed Thesis CD

Appendix H – Representation of the Three – Dimensional Trade Space of Spectral Resolution, Spatial Resolution and Noise

As previously mentioned in Chapter 5, it would be advantageous to re-plot the graphs shown in this thesis to better show the trade-space shared by the parameters of spectral resolution, spatial resolution and noise. Examples of what these plots may look like and preliminary ideas with respect to developing a full 3-D representation of this trade-space are discussed here.

In this thesis study, plots depicting algorithm performance with respect to each individual parameter of spatial resolution, spectral resolution and noise were constructed. Statistical analysis of the collected data was performed by the use of ANOVA techniques. This analysis included a study of the effects of both individual factors and combined factors on algorithm performance. However, three dimensional or surface plots were not constructed. These would help portray the overall trade-space that represents algorithm performance with respect to the three tested parameters. In Chapter 5, one of the recommendations made was that a further analysis of the collected data is required. This should include plotting the collected data to form surface plots to better visualize the spectral resolution, spatial resolution and noise trade-space. Discussed below is how this further analysis could be conducted.

With the collected data it is first necessary to construct matrices that chart one factor against another. An example of these types of matrices can be seen in Table H-1. As seen in the first matrix, in Table H-1, we are examining the trade space of spatial resolution and spectral resolution with respect to the performance of the SAM algorithm applied to the Rochester scene. We run into the same problem that was encountered with our previous analysis. That is, each cell within the matrix is an average of all the data points with that specific combination of spectral resolution and spatial resolution from the collected raw data (Appendix E). For example, the grey

box in the first matrix of Table H-1 is the combined average of all kappa values associated with a spectral resolution of 110 nm and a spatial resolution of 20 m. This includes each result with SNR values of 10, 110 and 225. Again, because of this averaging approach we can expect a high standard deviation to be associated with the calculated means. Nevertheless, this approach will allow us to plot one parameter against another in order to gain a better understanding of the trade-space. It is fully realized that this only accomplishes charting the behaviours of just two image collection parameters. Attempting to plot the trade-space of all three parameters at once is a much more difficult problem and will be discussed later.

Spatial Resolution (m)	Spectral Resolution (nm)		
	110	55	10
20	0.4852	0.5602	0.6015
40	0.3745	0.4073	0.4397
80	0.2912	0.3254	0.3546

Spectral Resolution (nm)	Noise (SNR @ 30%)		
	225	100	10
110	0.6991	0.4451	0.0067
55	0.7455	0.5439	0.0036
10	0.7719	0.6238	0.0001

Spatial Resolution (m)	Noise (SNR @ 30%)		
	225	100	10
20	0.9395	0.7073	0.0001
40	0.6968	0.5176	0.0072
80	0.5801	0.3879	0.0030

Table H-1: Trade-space statistics of the SAM algorithm applied to the AVIRIS Rochester scene

These values once plotted, as seen in Figures H-1 through H-3, reveal expected and interesting behaviours. Examination of Figure H-1 shows the expected result of the highest value of kappa being associated with the finest spectral and spatial resolution. Likewise, the lowest value for kappa in Figure H-1 occurs at the coarsest values of spectral and spatial resolution. It is from these points that a sloped and contoured surface occurs. Of note is the contours (colour-coded) that depict similar values of kappa. For instance, as seen by the " third-blue" range of 0.400-0.500, similar values of kappa result at 40 m spatial resolution regardless of the level of spectral resolution.
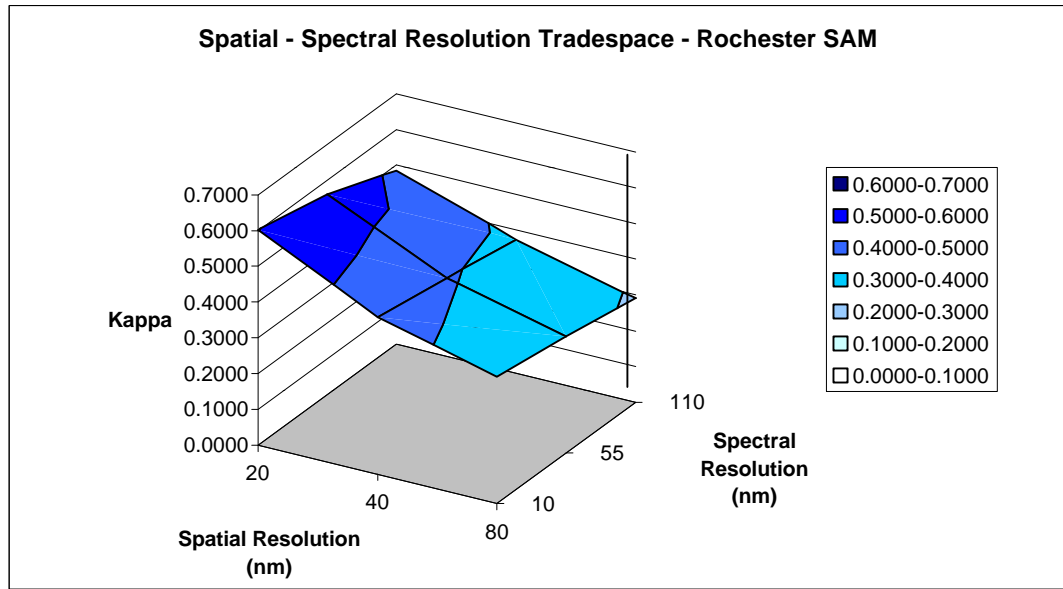
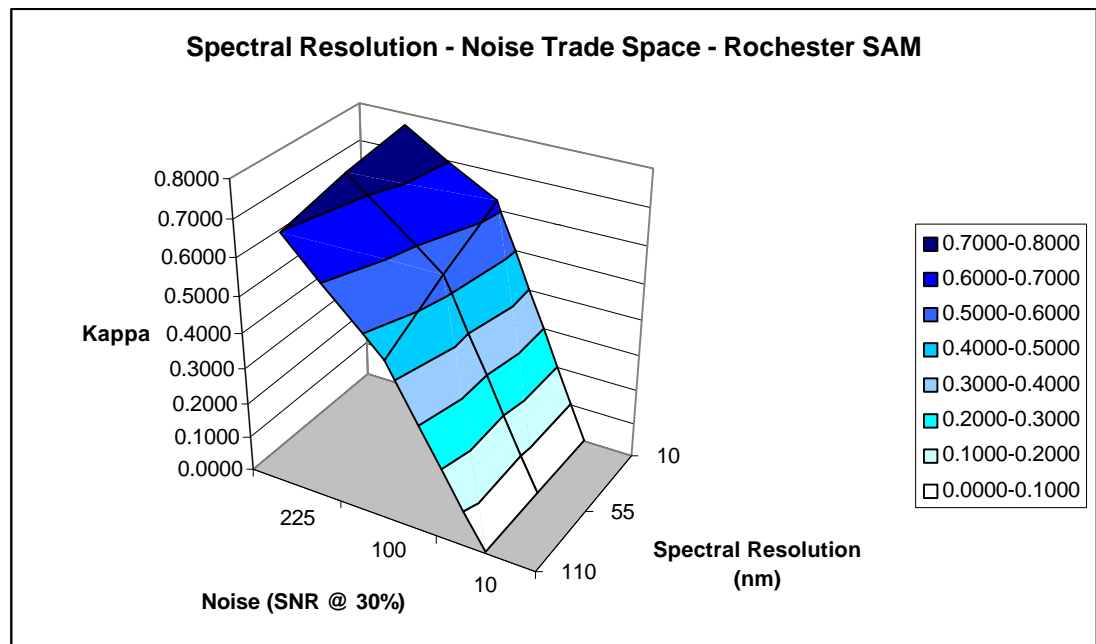Figure H-1: Spatial-Spectral Resolution Trade Space – Rochester SAM

Figure H-2: Spectral Resolution – Noise Trade Space – Rochester SAM
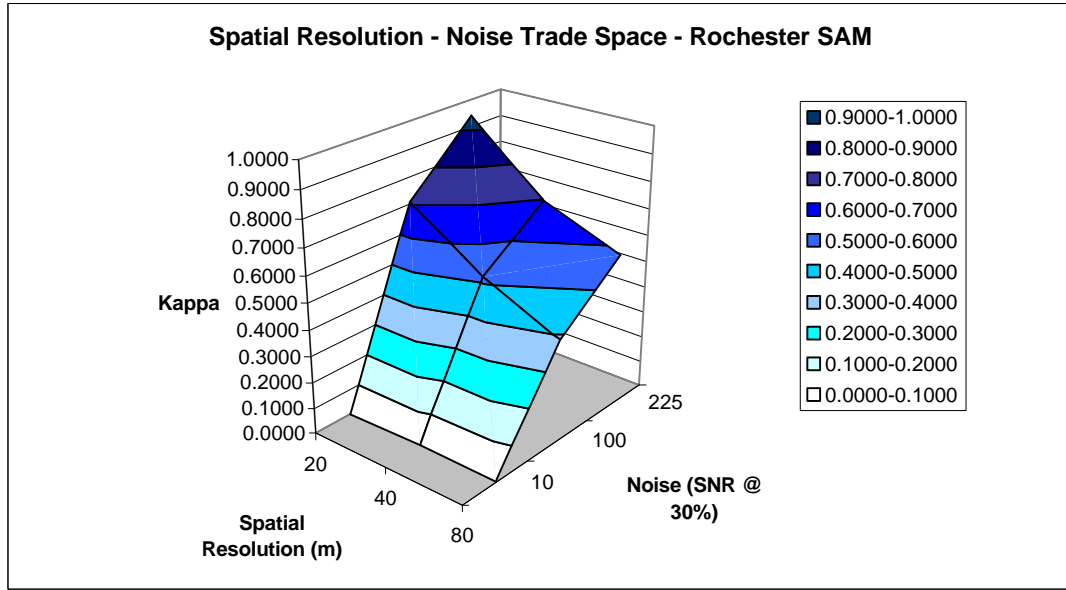
Spatial Resolution - Noise Trade Space - Rochester SAM

Legend:
- 0.9000-1.0000
- 0.8000-0.9000
- 0.7000-0.8000
- 0.6000-0.7000
- 0.5000-0.6000
- 0.4000-0.5000
- 0.3000-0.4000
- 0.2000-0.3000
- 0.1000-0.2000
- 0.0000-0.1000

Kappa axis: 1.0000, 0.9000, 0.8000, 0.7000, 0.6000, 0.5000, 0.4000, 0.3000, 0.2000, 0.1000, 0.0000

Spatial Resolution (m): 20, 40, 80

Noise (SNR @ 30%): 225, 100, 10

Figure H-3: Spatial Resolution – Noise Trade Space – Rochester SAM

The most interesting observation seen in Figures H-2 and H-3 is the stratification of the surface plot due to changes in SNR. The adverse effects of noise were also witnessed in Chapter 4. This plot better shows the dependence of kappa on SNR values independent of the other parameters it is plotted with in the trade-space. Also of interesting note, from Figures H-2 and H-3, is the relatively constant contours of the surface plots with respect to spectral resolution (Figure H-2) and spatial resolution (Figure H-3). For example, the eighth interval (0.200 – 0.300) in Figure H-3 does not change with respect to spatial resolution – as do the other intervals.

Similar surface plots were produced for the SMF algorithm when applied to the Western Rainbow scene. Table H-2 is the compilation of the parameter trade-space matrices. It is from this table that we are able to plot the surface plots seen in Figures H-4 through H-6.

As mentioned previously, another analysis of the collected raw data should be conducted to produce surface plots - such as those shown here. Any future work and testing should result in the production of both individual factor plots, as seen in the main body of this thesis, and trade-space plots like those seen here.

However, the problem of accounting for scene dependence still remains. Each image tested has a unique spectral and spatial complexity associated with it. In order to produce a single "global" surface plot depicting parameter trade-spaces independent of scene spectral and spatial complexity, these scene complexities must be somehow accounted for. In order to do this, a weighted average could be applied to the surface plots from each individual scene. For example, a

4

Spatial Resolution (m)	Spectral Resolution (nm)		
	110	55	10
2	0.6837	0.6913	0.9468
4	0.5523	0.5977	0.9265
8	0.3892	0.4898	0.7393

Spectral Resolution (nm)	Noise (SNR @ 30%)		
	225	100	10
110	0.8144	0.7637	0.0472
55	0.8421	0.8226	0.1142
10	0.9119	0.9053	0.7953

Spatial Resolution (m)	Noise (SNR @ 30%)		
	225	100	10
2	0.9606	0.9577	0.4035
4	0.8879	0.8774	0.3112
8	0.7198	0.6563	0.2421

Table H-2: Trade-space statistics of the SMF algorithm applied to the Western Rainbow scene
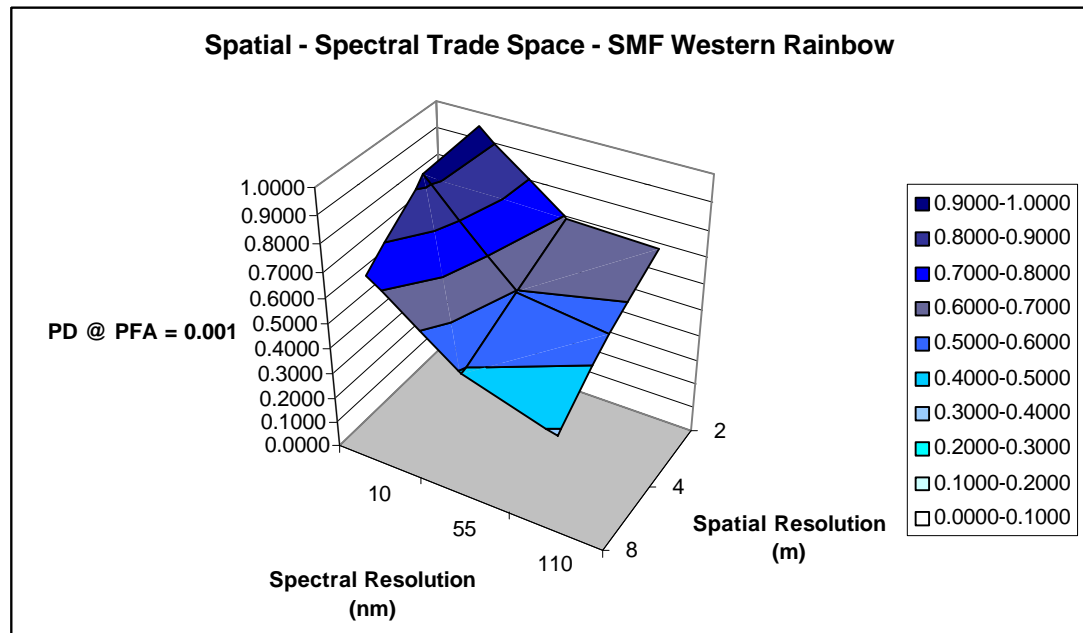
Figure H-4: Spatial – Spectral Trade Space – SMF Western Rainbow
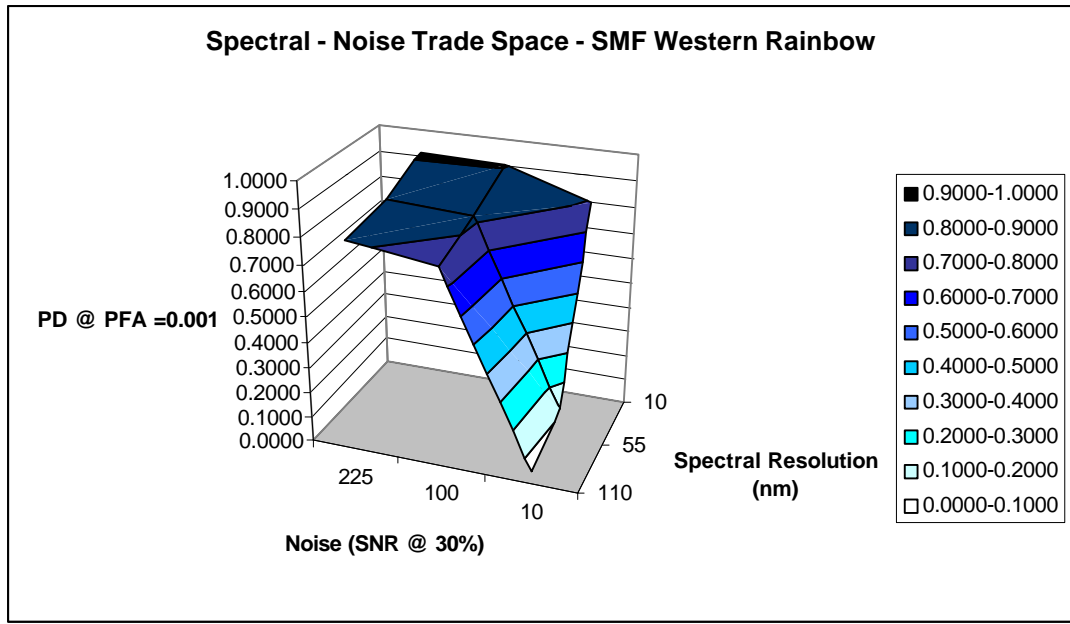
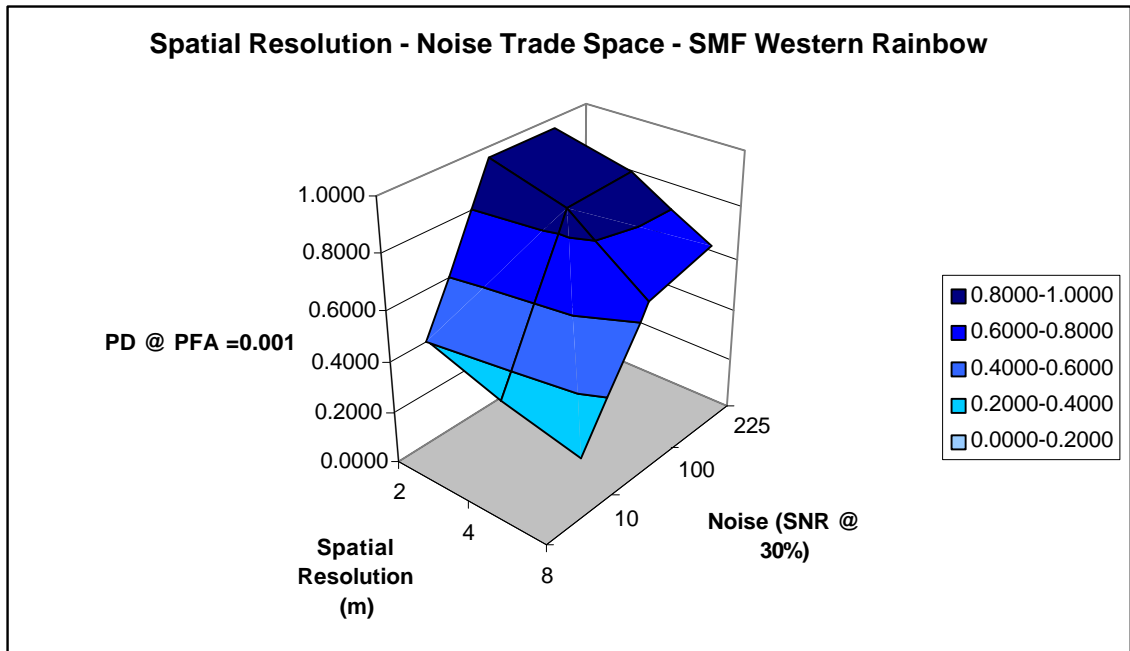Figure H-5: Spectral Resolution – Noise Trade Space – SMF Western Rainbow

Figure H-6: Spatial Resolution – Noise Trade Space – SMF Western Rainbow

surface plot depicting the trade space between spatial and spectral resolutions will be produced for both the SAM results of the AVIRIS Rochester and Rogers Dry Lake scene. These two surface plots of kappa share common axis and therefore the surfaces can be combined. The way by which they are combined is through the use of a weighted average that takes into account the spectral and spatial complexity of each individual scene. Metrics to arrive at these weighting functions may include such things as scene spectral seperability for measuring spectral complexity and metrics derived from the spatial frequency of the image calculated by taking the Fourier transform. This approach at eliminating scene dependence is merely an idea at this point in time and has never been tested. However, an investigation into eliminating scene dependence from the calculated image utility metrics is required and lies in the correct direction of arriving at *one* global surface plot for each algorithm tested – independent of scene complexities.

As seen in the above discussion, the task of producing a surface plot depicting the trade space of all three tested parameters simultaneously (spectral resolution, spatial resolution and noise) has not yet been approached. Producing a chart that demonstrates the performance of a spectral algorithm with respect to all three parameters simultaneously is a more difficult problem. However, extending the above approach of using coloured-contours to chart the metric values in a three-dimensional space seems appropriate. As seen, in Figure H-7, the three axis are spectral resolution, spatial resolution and noise. The coloured contours of the "triangular" surface plot represent a range of metric values – be it kappa, squared error or probability of detection. This approach is in the conceptual stage and has not been applied to the data. Having said that, it is believed that this approach, in concert with scene dependence normalization, would produce an easily interpretable trade space of spectral algorithm performance with respect to spatial resolution, spectral resolution and noise characteristics.
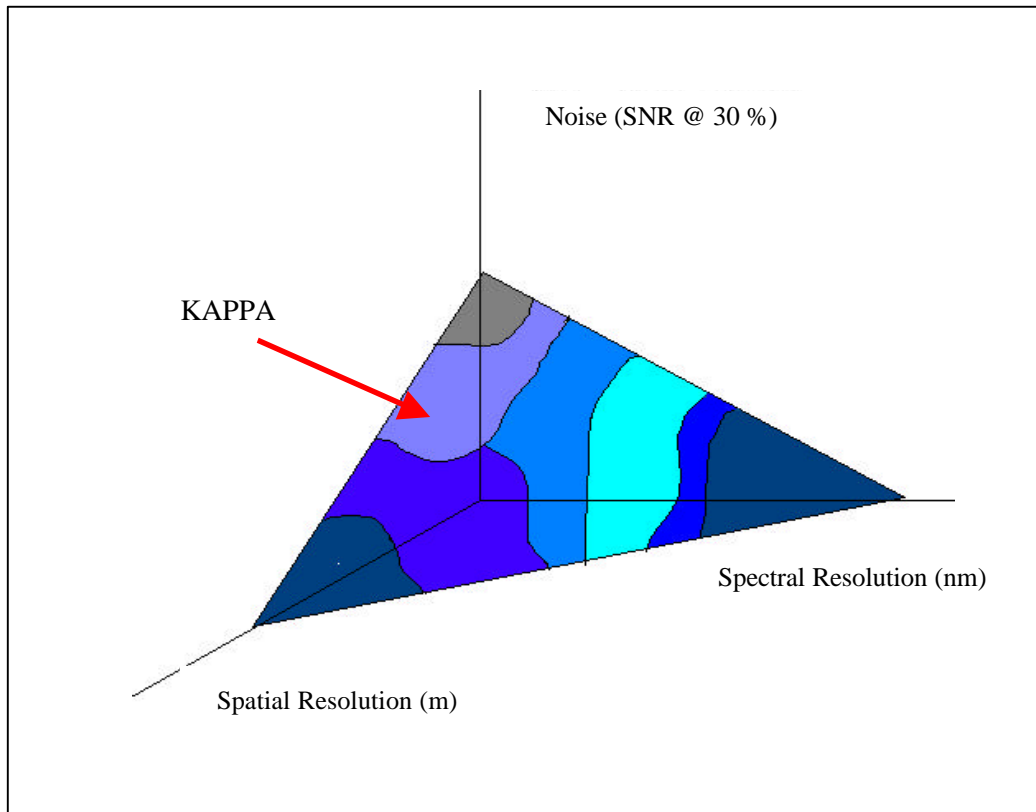
Figure H-7: Conceptual Three-Dimensional Trade Space of Kappa with respect to Spectral
Resolution, Spatial Resolution and Noise