

**Spectral unmixing of hyperspectral and  
multispectral images for predictive mapping of  
surface soil organic matter**

Berhanu Kefale Alemie  
March, 2005

# **Spectral unmixing of hyperspectral and multispectral images for predictive mapping of surface soil organic matter**

By

Berhanu Kefale Alemie

Thesis submitted to the International Institute for Geo-information Science and Earth Observation in partial fulfilment of the requirements for the degree of Master of Science in *Geoinformatics*.

## **Thesis Assessment Board**

Chairman: Prof.Dr. M.J. Kraak

External examiner: M. Schaepman

First supervisor: Prof.Dr.Ir. A. Stein

Second supervisor: Ir. B.G.C.M. Krol



**INTERNATIONAL INSTITUTE FOR GEO-INFORMATION SCIENCE AND EARTH OBSERVATION  
ENSCHEDE, THE NETHERLANDS**

**I certify that although I may have conferred with others in preparing for this assignment, and drawn upon a range of sources cited in this work, the content of this thesis report is my original work. Signed .....**

**Disclaimer**

**This document describes work undertaken as part of a programme of study at the International Institute for Geo-information Science and Earth Observation. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the institute.**

## Abstract

---

Remote sensing is an efficient tool for the assessment and monitoring of natural resources. Mixed pixels, however may lead to inaccurate classification results in most conventional image classification algorithms. Spectral unmixing may solve this problem by resolving those into separate components. The main objective of this research was to explore the applicability of spectral unmixing in hyperspectral and multispectral images to identify potential areas of surface soil organic matter (SSOM). The study was carried out using hyperspectral (DAIS) and multispectral (Landsat TM) in the semi-arid environment of Antequera area in Spain. The DAIS image covers only a small part of the study area (a strip running NW to SE) whereas the TM image covers the whole study area.

Optimum numbers of DAIS and Landsat TM bands for discriminating SSOM levels were selected for the strip area. A relationship has been established between these bands using regression analysis ( $R^2=0.5$ ). The established relationship was used to create a modeled DAIS image. A comparison of the three images in the strip area showed that DAIS, modeled DAIS and Landsat TM images discriminate SSOM levels with an accuracy of 70%, 66.7% and 64.4% respectively. This shows that integration of remote sensing images by means of statistical models help to improve discrimination of SSOM. Next the modelled DAIS image was extrapolated to the whole study area where only the TM image was available.

Linear spectral unmixing was applied to the three images in the strip area. The RMSE values obtained from unmixing analyses of DAIS and modeled DAIS were 0.02 and 0.1 respectively where as for TM it was 0.56. Mapping of potential areas of SSOM was done using endmember abundance images derived from the unmixing analysis. The classified SSOM results of the three images were cross-checked with ancillary data sources such as geological and topographical maps of the study area. The SSOM classified results of DAIS and modeled DAIS images were promising whereas, TM was not. Therefore, the spectral unmixing analysis was only applied on the modeled DAIS image for the whole study area and we mapped potential areas of SSOM.

Key words: *RMSE, SAM, SDA, SSOM, endmembers, linear spectral unmixing, modeled DAIS*

*Dedicated*  
*TO*  
*My loving and affectionate parents*

*My father, Kefale Alemie, who guided me to the right path; for his prayers and support*  
*My mother Azagn Muluneh, who wanted me to have the best, for her love, prayers, pains and troubles*  
*My beloved sisters, for every good thing a sister would offer*  
*My wonderful brothers, who wish to see my successes.*

## Acknowledgements

---

First and foremost I would like to give Glory to God the Almighty for: the unconditionally loving, bringing me this far in my educational career, reviving my inner strength by giving hope and comfort in my work every time when I was feeling like giving-up on the enormous pressure that came with this work. My God... still I need your preservation!

A single man cannot surround an anthill! This African saying is confirmed by the support that I received from several individuals and institutions during my MSc study time.

I am very much grateful to the government of Netherlands for fellowship grants. Equally I am also indebted to my organization (Sustainable Environment and Agricultural Rehabilitation Commission) for provision and facilitation of this opportunity. Especially Mr. Bantegezie in the Amhara civil service commission is highly acknowledged and appreciated for facilitated my applications taking into consideration the limited time I had. Without his unflagging effort, I couldn't have managed to come here.

I would like to express my gratitude to my first supervisor Prof. Dr. Alfred Stein for his guidance throughout this research; his comments and constructive criticisms greatly enhanced this thesis. His door was open for discussion all the time when I needed despite a pressing time schedule. Further word of thanks goes to my second supervisor Ir.B.G.C.M.Krol for his critical comments and suggestions, introducing me about the case study area, and giving and indicating me useful material sources. I had constructive discussions with you that were useful to sharpen this work.

I also would like to extend this acknowledge to Mr. Martin Yemefack, who opened my eyes how to do science during our two days discussion. I would, especially, like to express my gratitude to ITC International Hotel administration for making me feel at home all the time. Many thanks also go to Ethiopian community here in Enschede especially the former committee members for their efforts to feel us at home during our colourful traditional holidays. Especial thanks to our Orthodox Tewahido Sunday school members, my spiritual life have grown up because of you. I am also thankful to my colleagues in the GFM2, 2003; especially my cluster mates who shared all the happiness and pains I felt.

My most sincere gratitude goes to my dear friends back home for encouraging and helping me in one way or another to pursue my study abroad. Among these Molla, Astede, Ashenafi, Hunegnaw, Derese, Enatfenta, Alemayehu, Yeshambel, Mehari, Negash, Genet, Getachew, and Metalign are highly appreciated and acknowledged. I thank you all for your kind cooperation in bringing my dream real.

Finally, no words of thank could describe my debt to my lovely parents for their moral support, although miles away, in spirit they were always with me every single moment of all my graduate studies. You were a pillar of strength, moral and hope. Your blessings and praying are always with my long journey. Your love and affection helped me to go through this long and time demanding journey. Your words of encouragement always reminds me the Romans Philosopher, Cicero (106BC-43BC), "The greater the difficulty, the greater the glory". To you all (AMESEGINALEHU!), thank you!

# Table of contents

---

Abstract .....	i
Acknowledgements .....	iii
Table of contents .....	iv
List of figures .....	vi
List of tables .....	vii
Abbreviations .....	viii
<b>1. Introduction .....</b>	<b>1</b>
1.1. Background .....	1
1.1.1. RS and soil.....	1
1.1.2. RS and Surface Soil Organic Matter (SSOM).....	2
1.1.3. SSOM in Semi-arid regions.....	3
1.2. Research Problem.....	3
1.3. Previous Work.....	4
1.4. Research Objective.....	5
1.5. Research Questions .....	5
1.6. Available datasets.....	5
1.6.1. Earth Observation Data .....	5
1.6.2. Field Reference data.....	7
1.7. Research Approach .....	8
1.8. Thesis Organization .....	9
<b>2. Basic concepts .....</b>	<b>11</b>
2.1. RS data sources .....	11
2.1.1. DAIS imagery.....	12
2.1.2. Band selection technique.....	13
2.2. Pixels in an image .....	16
2.2.1. Mixed Pixels.....	16
2.3. Spectral Mixture Analysis (SMA) .....	17
2.3.1. Comparison between Models .....	20
2.4. Endmember selection .....	21
2.4.1. Endmember selection techniques .....	21
2.5. Linear spectral unmixing.....	22
2.5.1. What we get after unmixing? .....	23
2.6. Summary and conclusion .....	24
<b>3. Study area and Data preparation .....</b>	<b>25</b>
3.1. Study Area discription.....	25
3.1.1. Selection Criteria.....	26
3.2. Data preparation .....	26
3.2.1. Methods .....	26
3.2.2. DAIS Image preparation.....	27
3.2.3. Landsat TM preparation.....	28
3.2.4. Reference data preparation.....	28
3.3. Summary and conclusion .....	32

<b>4. Band selection analysis and results.....</b>	<b>33</b>
4.1. Introduction .....	33
4.2. Methods.....	33
4.3. Image processing .....	34
4.3.1 DAIS band elimination.....	34
4.3.2 Landsat TM band elimination .....	35
4.4. Band selection .....	35
4.4.1. Stepwise Discriminant Analysis (SDA).....	35
4.5. Establishing Relationship between selected bands .....	39
4.6. Comparison of the three images in $A_1$ .....	40
4.7. Summary and conclusion .....	42
<b>5. Linear spectral unmixing analysis and results .....</b>	<b>43</b>
5.1. Introduction .....	43
5.2. Methods.....	43
5.3. Endmember selection .....	44
5.3.1. Spectral Angle Mapper (SAM) .....	44
5.3.2. Other endmember selection techniques.....	49
5.4. Spectral unmixing classification .....	49
5.4.1. Spectral collection .....	50
5.4.2. Spectral unmixing results .....	50
5.5. Surface Soil Organic Matter (SSOM) mapping .....	54
5.6. Summary and conclusion .....	60
<b>6. Discussion.....</b>	<b>61</b>
6.1. About SDA results .....	61
6.2. Establishing relationships between the DAIS and TM bands.....	62
6.3. Unmixing classification .....	62
6.4. Potential SSOM area identification.....	63
6.5. Benefits from the research .....	64
<b>7. Conclusions and Recommendations .....</b>	<b>65</b>
7.1. Conclusions .....	65
7.1.1. From the perspective of the research problems.....	65
7.1.2. From the perspective of the research questions: .....	66
7.1.3. From the perspective the main objective:.....	66
7.2. Recommendations .....	66
<b>Bibliography .....</b>	<b>68</b>
<b>Appendix .....</b>	<b>72</b>

## List of figures

---

Figure 1-1 Reflectance spectrum of three land-cover types (Clark, 1999) .....	3
Figure 1-2 DAIS image covering part of the study area, (26, 15, 4 band combination).....	6
Figure 1-3: False colour combination of Landsat TM (432) .....	7
Figure 1-4: Site map of the study area and field sample points distribution .....	7
Figure 1-5 Box plot showing SSOM ranges in different landscapes of the study area .....	8
Figure 1-6 Conceptual methodological framework.....	9
Figure 1-7 Linkage of the chapters .....	10
Figure 2-1 The DAIS 7915 scan mechanism (Richter, 1996).....	12
Figure 2-2 Examples of mixed pixels occurrence on the ground (Fisher 1997).....	17
Figure 3-2 Data preparation procedure .....	26
Figure 3-3 Different strata locations on DAIS image .....	30
Figure 3-4 Normality check of the artificially generated samples.....	31
Figure 4-1 Band selection and modeling procedure.....	33
Figure 4-2 The three images in $A_1$ .....	40
Figure 4-3 Comparison of Wilk's $\lambda$ values of the three images in $A_1$ .....	41
Figure 4-4 Modeled DAIS extrapolated to $A$ .....	41
Figure 5-1 Spectral unmixing procedure.....	43
Figure 5-2 SAM classified images in $A_1$ of the three images, a) DAIS b) modeled DAIS c) TM .....	46
Figure 5-3 Endmembers spectral curve extracted from DAIS used in image decomposition .....	50
Figure 5-4 Endmember abundance images derived from unmixing of the three images in $A_1$ ;a) DAIS, b) modeled DAIS, c) Landsat TM. (Bright pixels on these images represent higher abundances and vice versa).....	52
Figure 5-5 Error images derived from spectral unmixing in $A_1$ : a) DAIS b) modeled DAIS c) TM.....	54
Figure 5-6 SSOM potential areas in $A_1$ : a) DAIS b) modeled DAIS c) Landsat TM .....	56
Figure 5-7 SSOM difference images (white color shows non-correlated areas): a) DAIS-modeled DAIS, b) DAIS-Landsat TM .....	57
Figure 5-8 SAM classified images of $A$ .....	58
Figure 5-9 SSOM potential areas of $A$ using modeled DAIS .....	59

## List of tables

---

Table 1-1 Spectral characteristics of Landsat TM (Lillesand, 1994.) .....	6
Table 2-1 Characteristics of DAIS image spectrometer (Strobl et al., 1996) .....	12
Table 2-2: Applicability of mixture models for different applications: Adapted from (Charles and Arnon, 1996).....	20
Table 4-1 Wilk’s $\lambda$ values and selected wavelengths in DAIS .....	36
Table 4-2 Standardized discriminant coefficients derived from SDA of DAIS .....	36
Table 4-3 Unstandardized weights showing the nature of discrimination for each discriminant function derived from DAIS .....	37
Table 4-4 Fisher’s linear discriminant classification result of DAIS image.....	37
Table 4-5 Wilk’s $\lambda$ values and selected wavelengths of Landsat TM .....	38
Table 4-6 Standardized discriminant coefficients of Landsat TM.....	38
Table 4-7 Fisher’s linear discriminant classification result of Landsat TM.....	38
Table 4-8 Regression function .....	39
Table 4-9 SDA Classification Results of modeled DAIS image .....	40
Table 5-1 SAM class statistics of candidate endmembers $A_1$ .....	45
Table 5-2 SAM classification accuracy results in $A_1$ : a) DAIS, b) modeled DAIS, c) Landsat TM. (N.C.P=number of classified pixels, N.G.P=number of ground truth points).....	48
Table 5-3 Statistics of fractional abundance and RMSE images obtained through unmixing using image spectra of the three images in $A_1$ : a) DAIS b) modeled DAIS c) TM. (EM = endmembers, Stdv = standard deviation, NA = not available) .....	53
Table 5-4 Classification functions used in the raster calculator .....	55
Table 5-5 Classification accuracy results.....	58
Table 5-6 Statistics of endmember abundance and RMSE values obtained through unmixing using image spectra of modeled DAIS image .....	59

## Abbreviations

---

**A** – the whole study area  
**A<sub>1</sub>** – Part of study area covered by DAIS image  
**DAIS** - Digital Airborne Imaging Spectrometer  
**DN** – Digital Numbers  
**EMR** – Electromagnetic Radiation  
**ENVI** - Environment for Visualizing Images  
**GPS** – Global Positioning System  
**IFOV** – Instantaneous Field Of View  
**Landsat TM** – Landsat Thematic Mapper  
**MNF** – Minimum Noise Fraction  
**NDVI** – Normalized Difference Vegetation Index  
**NIR** – Near Infrared Region  
**PCA** – Principal Component Analysis  
**PPI** – Pixel Purity Index  
**RMSE** – Root Mean Square Error  
**ROI** – Region of Interest  
**RS** – Remote Sensing  
**SAM** – Spectral Angle Mapper  
**SDA** – Stepwise Discriminant Analysis  
**SMA** – Spectral Mixture Analysis  
**SSOM** – Surface Soil Organic matter  
**SWIR** – Short Wave Infrared  
**TIR** – Thermal Infrared  
**USDA** – US Department of Agriculture  
**USGS** – US Geological Survey  
**UTC** – Universal Time Counter  
**UTM** – Universal Transverse Mercator  
**VIS** – Visible

# 1. Introduction

## 1.1. Background

Earth observation from space or Remote Sensing (RS), offers unrivalled capabilities for deep insight, monitoring, forecasting, managing and decision making processes about the environment. This can be done by mounting current and anticipated satellite and airborne sensors, which provide information in the form of digital array of pixels forming images. These images have been identified as a valuable improvement for understanding site-specific variation at the surface of the earth.

In RS studies different types of images are often used. Among these, advanced hyperspectral images provide details of ground features because they have narrow and continuous bands. Under laboratory conditions, the spectral information of the visible, near-infrared and short wave infrared (VIS-NIR-SWIR; 0.4–2.5  $\mu\text{m}$ ) spectral regions provide a promising capability to identify soil, vegetation, rock and mineral materials (Stoner, 1981).

Limitations in resolutions of RS images however, are still not fully solved. This is in particular the case for pixels being mixed up: an area corresponding to a single pixel at a RS image may contain various different materials. These materials are mixed, and the pixel reflectance observed by sensors is a combination of the reflectances of the individual materials. To get more detailed information from a single pixel the proportions of these materials can be approximated using a spectral mixing model (Boardman, 1994). Using this spectral unmixing model, the individual components present in the pixel can be reconstructed from known spectra in the image. In other words, the mixed pixels can be divided into its constituting components.

### 1.1.1. RS and soil

Soil is one of the most valuable resources. It is a base for all life processes. Good quality information from soil and land resource surveys is necessary for better management and wise soil use. Soil inventory is often carried out as part of a regional planning and development process in order to determine the location and extent of various soil types and/or soil variables. RS techniques have been commonly used for these activities. However, the spatial and temporal variability of surface processes makes soil properties variable and therefore, makes it difficult to measure directly from their reflectance spectra even under controlled laboratory condition (Ben-Dor and Banin, 1994). This capability could be even more problematic under RS domain (Peng, 1998) because variations occur on the scale of a few meters or less (Asner and Lobell, 2000). Unlike the vegetation spectra, which acquire advantage of peak-and-valley configurations, the shapes of reflectance spectra obtainable from soils are mainly invariant in the spectral regions (0.4-1.2 $\mu\text{m}$ ) (Clark, 1999). This may be due to a combined effect of different factors such as organic matter that can affect surface spectral reflectance of soils and make it non consistent through the spectrum region.

### 1.1.2. RS and Surface Soil Organic Matter (SSOM)

Soil organic matter is a key indicator of soil quality. Organic matter helps soils to hold water and supplies nutrients that are crucial for crop production; it also protects soils against erosion and helps to support a healthy and diverse set of microscopic plants and animals (Cambridge University, 2002). Concentration of organic matter in the surface soil enables to infer parameters such as soil color and lithology which are variables most often used for soil-landscape predictions (Joachim and Brigitta, 2000). According to Joachim and Brigitta, (2000) SSOM is also an important variable for assessing land degradation processes. Therefore, a detail investigation of SSOM is important.

Identifying potential distribution of SSOM in space and time is important for understanding landscape processes for a better management. One way of doing this is by using RS techniques. However, values of organic matter are hidden within the soil and are highly variable and quickly respond to external factors. This makes detection of SSOM from remote sensors difficult (Clark, 1999). In addition the scale of the spatial resolution of satellite sensors coupled with a continuous variation of organic matter on surface soil makes it more difficult. Even direct determination of SSOM in the laboratory from soil samples is also difficult. As an alternative, SSOM determination can be done indirectly from organic carbon, where organic carbon values multiplied by 2 resulted into realistic values of SSOM data (Nelson, 1982). We have used the same approach here.

However, as an alternative several researchers have made significant contributions for the prediction of SSOM from RS reflectance spectra, in situ measurements and from reflectance data recorded by satellite sensors (Daniel et al., 2002; Udelhoven et al., 1997). In addition, potential distribution of SSOM can be mapped using inferred variables like soil color (Mathieu et al., 1998), soil types and vegetation (Joachim and Brigitta, 2000). The relationship between visible reflectance and SSOM is revealed as darker soils contain more organic matter than lighter ones assuming other factors such as moisture content remains constant. Reports by Galvao and Vitorello, (1998) suggested that greater amount of SSOM is associated with greater absorption of incident energy with a lower spectral reflectance and in the image these areas appeared as dark tone. Joachim and Brigitta, (2000) used inferred variables such as soil types and vegetations to map potential area of SSOM distributions for assessing land degradation scales of Mediterranean Spain area using spectral mixture analysis. Finally they concluded that spectral mixture analysis derived qualitative indicators are corroborated in comparison to the qualitative estimates of soil organic carbon which are suited to provide a more systematic basis for defining soil degradation maps by defining SSOM levels.

Organic matter is abundant on surface soils and small amounts are known to substantially influence scattering by reducing the overall brightness and masking absorption features. The amount of SSOM to substantially affect the spectrum has been the subject of several studies. SSOM content and its composition have a strong influence on soil reflectance, especially when the SSOM content exceeds 2.0 % (Baumgardner et al., 1985). As the SSOM content drops below 2.0 %, it becomes less effective in masking the effects of other soil constituents and reflectance curve tends to be convex. The general observation is that soil reflectance decreases throughout the 0.4-2.5 $\mu$ m wavelength range with increasing soil organic matter content (Hoffer and Johannsen, 1969).

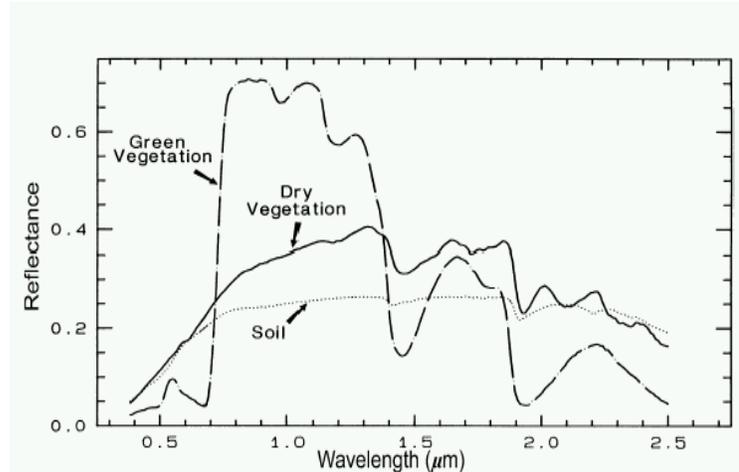


Figure 1-1 Reflectance spectrum of three land-cover types (Clark, 1999)

### 1.1.3. SSOM in Semi-arid regions

In semi-arid regions, the spatial heterogeneity of SSOM is a factor of soil type, vegetation cover, and topography (Grunwald, 1999). The effect of these factors on the existence of SSOM is summarized below.

- Organic matter is governed by soil types; under similar condition fine soils are expected to have relatively high organic matter as compared to coarser soils. Sandy soils are well aerated and tend to have low soil moisture content, which are environmental conditions favour for low organic matter content. Vice versa, clayey soils are less aerated with a high amount of fine micro-pores and are tend to have higher soil moisture content than medium and fine textured soils; hence, they tend to have high organic matter content (Grunwald, 1999)
- Even if vegetation cover is sparse in most semi-arid areas, the general trend shows vegetation affects SSOM. The more vegetative production the greater are the inputs of carbon. Cropping practice in semi-arid areas also usually influences the spatial distribution of SSOM because N fertilizer and rotation of different crops tended to increase organic matter (Biederbeck et al., 2002).
- Soils at the bottom of slopes generally have higher carbon because these areas are generally wetter and have higher clay contents than mountainous areas. High SSOM is expected in these area (Grunwald, 1999).

## 1.2. Research Problems

This research considers two problems:

In every remotely sensed image, a large number of mixed pixels are usually present. A mixed pixel is a picture element representing an area occupied by more than one ground cover type. Basically, two situations exist in which mixed pixels occur (Maurice, 1998). The first case concerns the pixels that are located at the edges of large objects such as agricultural fields. The second case arises when objects are imaged that are relatively small compared to the spatial resolution of the sensor scanner. For a given scanner, the number of mixed pixels depends on the landscape that is being imaged. Proportions of mixed pixels for example in TM-images range from 29.6% for the category water to

68.3% for grass patches (Irons et al., 1985), where as in some Mediterranean countries of the EU, where the average field size is small, the proportion of mixed pixels can be as high as 30% (Schoenmakers, 1995). These figures indicate that mixed pixels have a significant influence on the information that can be derived. Therefore, conventional classification of mixed pixels leads to errors that make the subsequent feature identification inaccurate since they do not acknowledge the compositional nature of natural surfaces in the landscape (Adams et al., 1985). These errors are caused by the premise of classification that all pixels are pure, i.e. consisting of a single ground cover type, whereas in fact they are not. Spectral unmixing may solve this problem of taking surface heterogeneity into account.

- The problem we address here is how to deal with mixed pixels when applying hyperspectral and multispectral sensor systems to identify SSOM potential areas. The subsequent use of hyperspectral data is assumed to overcome the problems of existing mixed pixels.

Hyperspectral images are suited for providing detailed ground surface information since they have narrow and continuous bands. However, the limited availability of these images data and correspondingly high cost makes them difficult to be accessed by a large number of users. According to Koch, (2000) multispectral sensor is most commonly used for land-cover and land-use studies. This might be due to their widely availability and relatively low cost, despite the low spectral resolution nature make them unsuccessful in identifying detailed surface features.

- This research attempts to use extra advantage from hyperspectral DAIS image to improve or enhance SSOM identification of Landsat TM image in a small part of the study area.

### **1.3. Previous Work**

Several studies have been conducted in the area of RS to handle the problem of pixel mixing. As a result, a number of promising results and techniques are investigated. Among these techniques, spectral unmixing (Adams et al., 1985) is recently developed technique for such applications. Spectral unmixing techniques have been used in different arena of natural resource studies. Among these Neville et al., (2003); Van der Meer, (1995) applied spectral unmixing on different RS images for mineral exploration application and as a result they successfully identified the proportions of pure mineral endmembers after unmixing. A similar study has been conducted by Koch, (2000) for land degradation detection assessment and she proved that spectral unmixing is a successful technique in identifying land degradation changes. Relationship between spectral reflectance and soil organic carbon concentration was studied by Joachim and Brigitta, (2000) using spectral mixture analysis and produced differentiated maps of soil organic carbon for their land degradation study. A curvilinear exponential function relationship between SSOM and reflectance values were demonstrated by Daniel et al., (2002; 1977) and Schreier, (1977). Another study showed that SSOM has been better estimated after synthesizing the IRS image using band cloning method after establishing a relationships through mathematical regression between IRS bands and spectrometer channels (Daniel et al., 2000).

## 1.4. Research Objective

The aim of this research is to evaluate the applicability of spectral unmixing of hyperspectral and multispectral images to identify potential areas of SSOM. The Antequera area in southern Spain is considered as a test area.

The following specific objectives are considered:

- Identify the optimum number of DAIS and Landsat TM bands that are suitable for SSOM level discrimination;
- Establish a mathematical relationship between the DAIS and Landsat TM images;
- Estimate endmembers abundance in pixels using spectral unmixing classification;
- Identify potential areas of SSOM in the study area.

## 1.5. Research Questions

The following research questions help us to achieve our objective stated above.

- What are the optimum numbers of bands of DAIS and TM images suitable for SSOM level identification?
- Can integrated dealing of DAIS and Landsat TM images improves the discrimination of SSOM levels in the study area?
- What is the influence of different band combinations on classification results of spectral unmixing?
- How can we evaluate the accuracy of spectral unmixing results?
- Can endmember abundance images derived from unmixing analysis be used for mapping potential areas of SSOM in the study area?

## 1.6. Available datasets

Data used in this research can be categorized into two types as: Earth observation data and field reference data. The nature and characteristics of these data are discussed below.

### 1.6.1. Earth Observation Data

In this research, two image data were used. These are hyperspectral (DAIS) and multispectral (Landsat TM) as explained briefly below.

#### DAIS Image

The DAIS sensor is mounted on DLR Dornier 228 aircraft and flown over southern Spain, Antequera area on 20<sup>th</sup> June, 2001 from an altitude of 10,000 feet (providing a pixel size of about 5m×5 m) and spectral resolution (24 nm to 1.0 μm). The image scene covered only part of the study area of 21 km by 3 km; a strip area running North-west to South-east direction of the study area ( $A_1$ ) as shown in figure 1.2. In contrast to normal panchromatic images, which only record the visible part of the electromagnetic spectrum, DAIS sensor records a broad part of the visible and infrared part of the spectrum in different wavelengths and spectral resolution. In this study, only the refractive portion of the electromagnetic radiation was considered covering the VIS-NIR-SWIR (0.4–2.5 mm) spectral region with 72 spectral bands.



Figure 1-2 DAIS image covering part of the study area, (26, 15, 4 band combination)

### Landsat TM Image

The Landsat programme is the oldest space-borne Earth Observation programme. Thematic Mapper data has a swath width of approximately 185 km from an altitude of 705 km (Jensen, 1996). Its detector records the EMR in seven bands. The Landsat platforms operate following a sun-synchronous, near polar orbit. Its radiometric resolution is 8 bit, and its temporal resolution is 16 days, with an image overlap that varies from 7 percent at the Equator to nearly 84 percent at 81° North or South latitude (Lauer et al., 1997). It has been served and used for many applications including land cover mapping, land use mapping, soil mapping, geological mapping, and sea surface temperature mapping.

A Landsat TM image of the Antequera area which was acquired on 11<sup>th</sup> June, 1994 is also used in this research. The spectral characteristics of Landsat TM image is summarized in the table below.

Band	Spectral range (µm)	Nominal spectral location	Spatial resolution
1	0.45 – 0.52	Blue	30
2	0.52 – 0.60	Green	30
3	0.63 – 0.69	Red	30
4	0.76 – 0.90	Near Infrared	30
5	1.55 – 1.75	Mid Infrared	30
6	10.4 – 12.5	Thermal Infrared	60
7	2.08 – 2.35	Mid Infrared	30

Table 1-1 Spectral characteristics of Landsat TM (Lillesand, 1994.)



Figure 1-3: False colour combination of Landsat TM (432)

### 1.6.2. Field Reference data

The available ground truth data include 78 'full pit' containing soil profile and detail SSOM measurement data. These ground truth points are spread over the area of the topo-map sheet 1:50,000 that cover the study area (see the site map below). Out of the 78 pits only 8 pits are falling within  $A_1$ . The ground truth data were taken in May, June, and July in different years from 1984 to 1991. In this study we considered the organic matter content of the top soil, disregarding the available sub-surface organic matter data to avoid any complications.

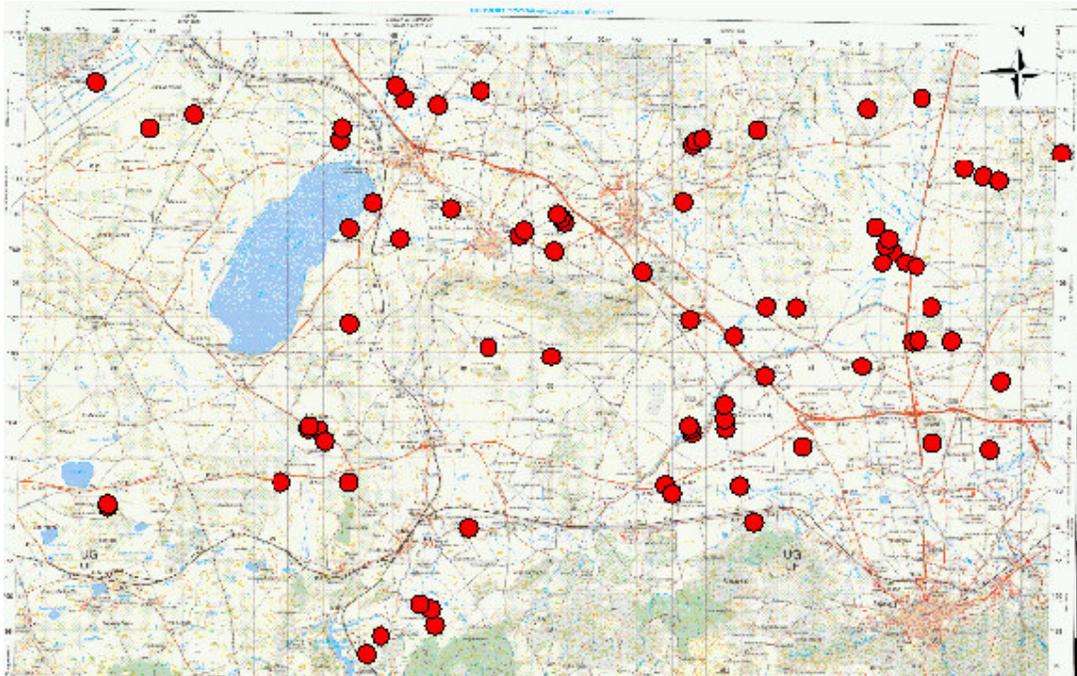


Figure 1-4: Site map of the study area and field sample points distribution

The minimum value of the measured SSOM is 0.04% it was recorded at hilly landscape areas and the maximum value is 4.54% that was recorded on the alluvial deposit area. The average and standard deviation values are 2.12% and 0.82 respectively. The measured SSOM values on different landscape units (we call it strata) in the study area are shown in the box diagram below. Stratum A2 refers the lacustrine deposit dominant area, stratum A3, refers the undulating topography in central part of the strip area, and stratum A4 refers alluvial deposit areas.

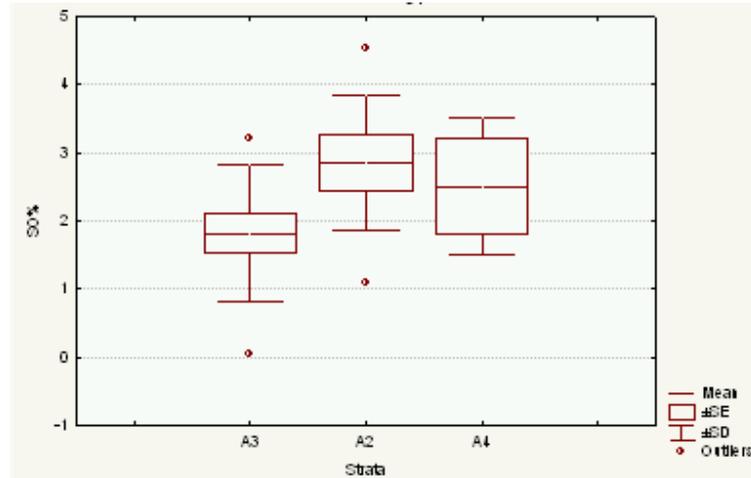


Figure 1-5 Box plot showing SSOM ranges in different landscapes of the study area

## 1.7. Research Approach

Multi-source RS data including DAIS and Landsat TM images (figures 1-2 and 1-3 respectively) and ground truth reference measurement data have been used in this research as discussed above. The TM image covers the whole study area (A) where as DAIS image covers only part of the study area (A<sub>1</sub>) i.e., a strip running at the North-west to South-east part of the study area.

Three main issues are addressed in this research: selection of optimum number of bands that are suitable for discriminating SSOM levels, spectral unmixing analysis to derive endmember abundance images and mapping of potential areas of SSOM in the study area using endmember abundance images derived from spectral unmixing. Consequently, a suitable methodology must be able to:

- Select optimum number of bands useful for SSOM discrimination from both images;
- Establish a relationship between the selected bands of TM and DAIS;
- Based on the relationship, create a modeled image in A<sub>1</sub> and extrapolate it to A;
- Derive endmember abundance images by means of unmixing
- Mapping potential areas of SSOM in the study area

We developed a conceptual framework that helps to address the objectives of this research. It consists of eight basic parts. The parts are linked with each other and each part is input into the next part. Data preparation is essential to prepare the available satellite and ground truth data for further analysis and processing. Therefore, it was carried out in part one of our methodology. The second step was selection of optimum number of bands for SSOM level discrimination. Many bands in hyperspectral images are overwhelmed with significant amounts of “redundant” data leading to great strain on

computers (Prasad, 2000). Thus, band selection is needed. This reduces processing time and storage volume and would help in avoiding handling “redundant” wavebands. We did the same band selection for TM image as well since all the seven bands is not expected to be useful for SSOM level identification. The third step was establishing relationship between the selected TM and DAIS bands in  $A_1$ . In this way, we could create a modeled DAIS image. Part five of our methodology is comparison of the quality of the three images in  $A_1$ . The sixth step consists of extrapolating the modeled DAIS image to  $A$  where we have only Landsat TM image. The seventh step consists of spectral unmixing classification in both  $A_1$  and  $A$ . The eighth part is about mapping potential areas of SSOM based on endmember abundance images derived from the unmixing analysis. The detail methodology is shown in the flow chart below.

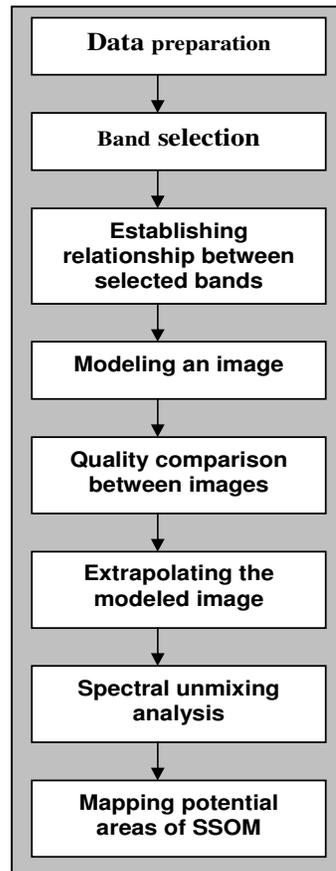


Figure 1-6 Conceptual methodological framework

## 1.8. Thesis Organization

This thesis has been developed as a series of chapters which are connected each other (figure 1.2). **Chapter 2. Basic concepts** - explores different Spectral Mixture Analysis modelling techniques, stepwise discriminant band selection techniques, endmembers selection and spectral unmixing classification techniques. It also gives an overview of DAIS system and technology. **Chapter 3. Study area and Data preparation** – starts with explaining the case study area and explains briefly some basic data preparations carried out on both the images and field reference data. **Chapter 4. Band selection analysis and results** – this part briefly presents how optimum numbers of bands are selected for SSOM level identification, it also gives an overall view how we integrated the two images

for a better SSOM discrimination. **Chapter 5. Spectral unmixing analysis and results** – this chapter explains how spectral unmixing analysis was implemented to derive endmember fraction images which will be used as inferred variables for SSOM mapping. Mapping of potential SSOM areas using endmember fraction images is also elaborated in this chapter. **Chapter 6. Discussion** – gives a brief discussion based on the results achieved and the analysis carried out. **Chapter 7. Conclusions and recommendations** - draw some conclusions from the study results obtained and forwarding some recommendations for further work.

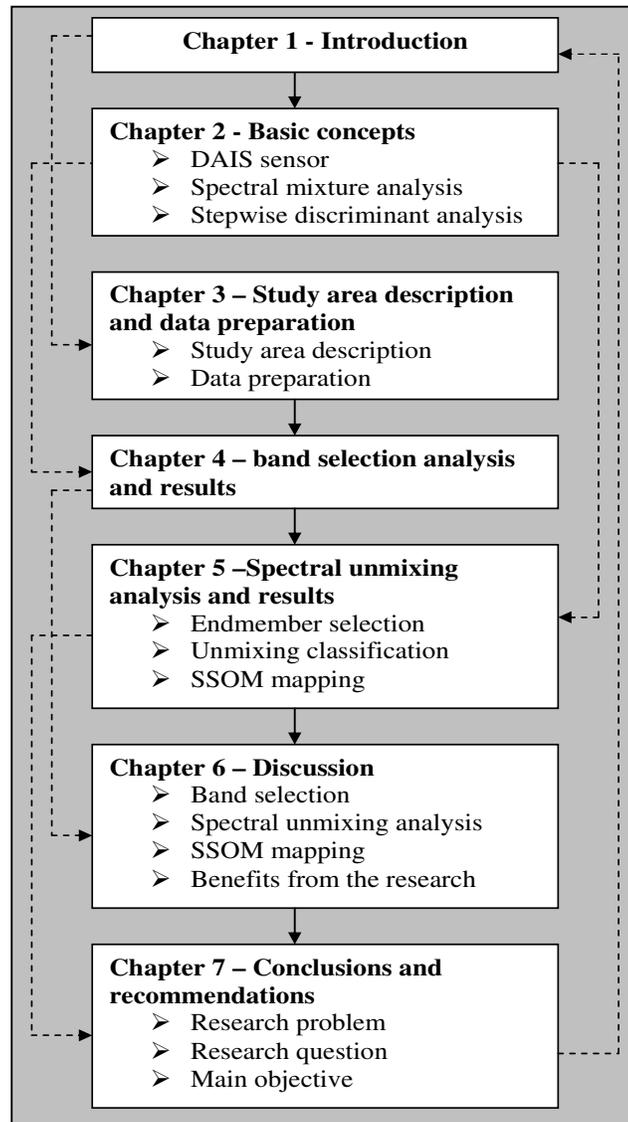


Figure 1-7 Linkage of the chapters

## 2. Basic concepts

In this chapter, four basic concepts are presented and consist of five parts. Section 2.1 explains RS data sources. In this section hyperspectral data sources specifically DAIS instrument technology and characteristics are briefly explained. Selection of optimum number of hyperspectral bands for a particular application beforehand is useful. Stepwise Discriminant Analysis (SDA) has been implemented for band selection purposes in most literatures and it is reviewed in section 2.1.2. Mixed pixels are common in most RS images. The causes and nature of mixed pixels are explained in section 2.2 and the subsequent subsections. One of the possible solutions for solving problems related with mixed pixels is by means of spectral mixture analysis (SMA). Thus, different SMA models are elaborated and compared in section 2.3 and 2.3.1 respectively. Endmember selection is the first step in dealing with SMA and spectral unmixing analysis. Hence, endmember selection techniques are briefly summarized in section 2.4. Section 2.5 explains how spectral unmixing techniques are used to derive endmember abundance images. Finally a summary and conclusion of the chapter is presented in section 2.6.

### 2.1. RS data sources

RS makes use of electromagnetic energy for acquiring information about an object or a phenomenon. Each kind of object has its own specific spectral characteristics: types of spectral curves and different values of the spectral reflectance values in different bands. These characteristics are determined by the physical and chemical properties of the objects. The resulting data are contained in a digital array of numbers in 2D consisting of pixels, also called picture elements, all together forming an image.

Information obtained from RS images is records of surface reflectance values for each material imaged inside the FOV of the sensor system. The detail of information that can be derived from images thus depends on the nature of sensors used. As a result different images for various applications have different spatial and spectral resolution ranging from hyperspectral to multispectral.

To provide greater spectral resolution, hyperspectral imagery is often used to model a remotely sensed scene. In hyperspectral imagery, the electromagnetic spectrum is sampled at tens, hundreds or even thousands of wavelength ranges in the VIS and NIR region (Kyu-Sung et al., 2004). The result is a very detailed view of the spectral signature of the scene represented by a particular pixel since the narrow and contiguous spectral channels that constitute hyperspectral sensors enable the detection of small ground cover features and local variations that might otherwise be masked within the broader bands of multispectral scanner system. Works based on laboratory and field spectral reflectances has demonstrated their potential to provide information on different soil properties among which are quantitative estimation of moisture (Galvao and Vitorello, 1998) and organic matter (Ben-Dor et al., 1997). Airborne high spectral resolution sensors such as Airborne Visible Infrared Spectrometer (AVIRIS) and Digital Airborne Imaging Spectrometer (DAIS), have also proven to be useful in such

estimations on soil surface compositional information to obtain several soil properties (Ben-Dor et al., 2002). DAIS is among the existing hyperspectral images that have been used for various applications in RS study. The DAIS instrument and image processing techniques are discussed below.

### 2.1.1. DAIS imagery

#### The DAIS 7915 instrument

DAIS-7915 is a whisk broom sensor, manufactured by the GER Inc., USA and upgraded by the DLR Germany (Muller, 1997). The sensor is sensitive to the VIS-NIR-SWIR-TIR spectral regions (0.4–14  $\mu\text{m}$ ) depending on the carrier aircraft flight altitude (Muller, 1997). It consists of 79 channels across, with a bandwidth ranging from 0.9 nm to 60nm. The IFOV is 3.3 mrad, and the FOV is 52°. The DAIS scan mechanism is of Kennedy type, where a cubic polygon mirror scans the terrain below through the opened hatch in the bottom of the aircraft. The scan mirror rotates anti clockwise with respect to the aircraft heading to provide a ground element cross track scanning motion while the forward motion of the aircraft provides a requested line-by-line scan. During the scan, the up welling energy collected within the FOV of the scanner is reflected by the polygon to the fixed flat folding mirror and focused by the parabolic mirror on the aperture stop at the entrance of the spectrometer.

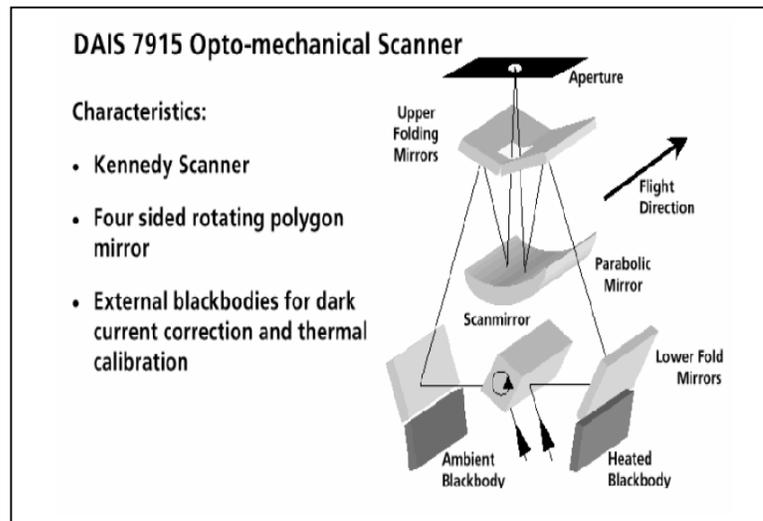


Figure 2-1 The DAIS 7915 scan mechanism (Richter, 1996)

The main characteristics of the DAIS-7915 bands are summarized in the table below.

Number of bands	Bandwidth ( $\mu\text{m}$ )	Spectral location	Wavelength range ( $\mu\text{m}$ )
32	12 -35 nm	VIS/NIR	0.4 – 1.0
8	36 - 56 nm	SWIR1	1.5 – 1.8
32	20 - 40 nm	SWIR2	2.0 – 2.5
1	2 $\mu\text{m}$	MIR	3.0 – 5.0
6	0.6–1.0 $\mu\text{m}$	TIR	8.7 – 12.3

Table 2-1 Characteristics of DAIS image spectrometer (Strobl et al., 1996)

DAIS-7915 is used for various applications since spring 1995 such as environmental monitoring of land and marine ecosystems, vegetation status and stress investigations, agriculture and forestry resource mapping, soil and geological mapping, mineral exploration (Lehmann, 1995).

### **DAIS data processing**

DAIS image undergo image calibration and atmospheric correction during image production. Calibration of the DAIS imagery for the VIS, NIR, SWIR and TIR was carried out in the laboratory at DLR before and after the flights. Due to the optomechanical and electrical characteristics of the instrument, all data must undergo a special pre-processing and calibration procedure before any quantitative analysis (Strobl et al., 1996). For this purpose a software package was designed, which is intended to fulfil the requirements of an operational and mostly automatical pre-processing of the airborne imaging spectrometer data. Data processing takes place in three steps. During the first step the raw data are read from tape, converted to BSQ format and written to hard disk. At the same time auxiliary data are extracted. A raw data report is issued which contains quick looks and histograms of all channels as well as a summary of auxiliary parameters. Thereafter the calibration coefficients may be applied as the third and last step of the radiometric pre-processing.

The atmospheric correction is based on the MODTRAN radiative transfer code (Berk et al., 1989). Image data in the reflective channels of the DAIS sensor are converted to ground reflectance. From the thermal channels emissivities and surface temperatures are derived. Special features include calculation of water leaving radiance, radiation fluxes and heat fluxes (Richter, 1996). To assess the stability of the sensor calibration, interactive user-friendly software was developed to carry out an in-flight calibration based on measurements of selected ground targets and observed atmospheric data. The geometric correction at the current stage uses aircraft attitude data (roll, pitch, yaw) recorded by the scanner to rectify the data. This processing step reduces the number of ground control points needed for the final geocoding of the scene. All these are pre-processing steps carried out during DAIS image production. Different image processing algorithms are used for extracting information from hyperspectral images. Among these SAM and spectral unmixing techniques are briefly explained in sections 2.4.1 and 2.5 respectively.

#### **2.1.2. Band selection technique**

While hyperspectral data are very rich in information, processing them poses several challenges regarding computational time, information redundancy removal, relevant information identification, and modeling accuracy. Therefore, selection of optimum number of bands for the intended application should be the first step in hyperspectral image analysis (Velez-Reyes, 2003). Stepwise Discriminant Analysis (SDA) has been widely used by many researchers for selecting optimum number of bands for different applications. For example Prasad et al., (2004) used SDA for discriminating different forest types from hyperspectral and multispectral images. We used the same approach (SDA) for discriminating SSOM levels in the study area.

### Stepwise Discriminant Analysis (SDA)

Discriminant analysis (Fisher, 1936), is usually performed to investigate differences among multivariate groups, to determine which attributes discriminate between the groups, and to determine the most parsimonious way to distinguish among groups. It is a priori technique in that the groups are defined beforehand (the opposite of cluster analysis where it was used to form the groups). In this study we used field priority knowledge for clustering  $A_1$  of the study area into different strata. The predictors' (reflectance values in our case) characteristics are related to form groups based upon similarities of distribution in n-dimensional space which are then compared to groups that are input by the user as truth. This enables the user to test the validity of groups based upon actual data, to test groups which have been created, or to put objects into groups.

When there are more than two groups, we can estimate more than one discriminant function; number of discriminant function equals number of groups minus one. Once the discriminant functions are calculated, each group is given a discriminant function score, which is the value resulting from applying a discriminant function formula to the data for a given case. These scores are called standard discriminant function scores or canonical correlations (Tabachnick and Linda, 2001). Under the assumptions that the observations are random and is normally distributed, and the variance/covariance matrix for each group is the same, the technique/discriminant analysis involves deriving linear combinations of two or more discriminating variables that will best discriminate among the priori defined groups. The first discriminant function defines the specific linear combination of variables that maximize the ratio of the among-group to the within group variance in any single dimension. The second discriminant function explains the remaining variance and so forth as explained by the eigenvalues result. For our application the discriminant function can be expressed mathematically as:

$$Strata_k = a + b_1.R_{\lambda_1} + b_2.R_{\lambda_2} + \dots + b_m.R_{\lambda_m} \quad (1)$$

Where  $Strata_k$  stands for the grouped property i.e. different strata in our cases,  $a$  is a constant coefficient for the current population,  $b_1$ -  $b_m$  are standard discriminant coefficients for each wavelength,  $R$  is the reflectance of a given wavelengths and  $\lambda$  stands for wavelength.

The relative importance of the independent variable ( $R$ ) in the discriminant function can be explained in the standardised discriminant (canonical) coefficients. As clearly seen in the above equation, standardised discriminant coefficients are useful to look at each factor relative to the others. Therefore, if one coefficient is twice as large as another, it is twice as good a discriminator as the other.

The eigenvalues of each discriminant function reflect the relative importance of the functions because they reflect the percentage of variance explained in the dependent variable, cumulating to 100% for all functions. If there are more than one discriminant function, the first will be the most important, the second next most important in explanatory power, and so on.

The output of discriminant analysis include classification table also called confusion matrix, is used to assess the performance of discriminant analysis. These are not to be confused with the discriminant functions. The classification functions can be used to determine to which strata each case most likely

belongs. This is simply a table in which the rows are the observed categories of the dependent and the columns are the predicted categories of the dependents. When prediction is perfect, all cases will lie on the diagonal. The percentage of cases on the diagonal is the percentage of correct classifications. The ultimate test of the function's discriminatory power is the rate of correct classification from both classification accuracy and the cross-validated results (Tabachnick and Linda, 2001). When more than two groups are considered, the classification formula or most of the time known as Fisher's Linear Discriminant function have the following form:

$$CS_j = c_{j0} + c_{j1}x_1 + \dots + c_{jp}x_p \quad (2)$$

Classification score ( $CS_j$ ) for group  $j$  is found by multiplying the row score on each predictor ( $x_p$ ) by its associated classification function coefficient ( $c_j$ ), summing over all predictors and adding a constant,  $c_{j0}$ . The coefficients are found by taking the inverse of the within covariance matrix  $W$  and multiplying it by the predictor means ( $M_j$ ):

$$C_j = W^{-1}M_j \quad (3)$$

The intercept is found by:

$$c_{j0} = \left(-\frac{1}{2}\right)C_jM_j \quad (4)$$

In addition to the classification result, Wilk's Lambda ( $\lambda$ ) test in SDA can also be used for assessing the quality of the model (Prasad, 2000); Wilk's  $\lambda$  is the test statistic preferred for multivariate analysis of variance and is found through a ratio of the determininants (Webster, 1974). It can be mathematically expressed as:

$$\lambda = \frac{|S_{error}|}{|S_{effect} + S_{error}|} \quad (5)$$

where:  $\lambda$  is Wilk's Lambda and  $S$  is a matrix which is also known as Sum of Square

The values of Wilk's  $\lambda$  is indicative of separability or discriminatory power of independent variables or spectral bands in our case i.e., the less the value of Wilk's  $\lambda$ , the greater spectral differentiation between the strata and close to one indicates the strata means are not different (Prasad et al., 2004).

To summarize, when interpreting multiple discriminant functions, which arise from analyses with more than two groups and more than one variable, one would first test the different functions for statistical significance, and only consider the significant functions for further examination. Next, it is good to look at the standardized  $b$  coefficients for each variable for each significant function. The larger the standardized  $b$  coefficient, the larger is the respective variable's unique contribution to the discrimination specified by the respective discriminant function. In general Discriminant Analysis is a very useful tool: for detecting the variables that allow discriminating between different (naturally occurring) groups and for classifying cases into different groups with a better than chance accuracy. The theory is readily accessible in Webster and Oliver, (2001). They generalized the theory of linear

discriminant analysis to a fuzzy linear discriminant analysis since it considers membership of each individual to each of the classes.

## **2.2. Pixels in an image**

It is important to give some consideration to the mechanism by which a scanner forms an image and to the idea of the IFOV, which is a measure of the ground area viewed by a single detector element in a given instance of time. It is convenient and common practice, to construct an image from square or rectangular picture elements or pixels (Cracknell, 1998). The simplest idea that one often find in many literatures, when describing the operation of a scanner, is that the instrument receives all the radiation from a certain area on the ground (IFOV) and generates a response that is proportional to the amount of radiation received. For convenience that FVO is commonly regarded as a rectangular (or square) piece of the Earth's surface so that, from the array of scanner data, an image can be constructed where the pixels fall a two dimensional plane surface as a scale reduction representing the ground. It would be nice if these were true and all the objects on the ground were: exactly the same size as this hypothetical rectangular or square IFOV, all aligned with their edges parallel or perpendicular to the scan direction, and exactly located so that their centres coincided with the centres of the pixels. Even if the scanner were responding uniformly to all point sources within this geometrical IFOV then the chances of all the objects on the ground is patently near to zero; cars, houses, cows, etc. are just not like that (Cracknell, 1998).

### **2.2.1. Mixed Pixels**

Resolution of an image is constrained by the pixel size. The pixel size is determined by the IFOV of the sensor's optical system. Even this simple idea what we have discussed above fails to represent exactly what is in the ground because geographic features on the earth's surface are heterogeneous by nature. More than one land cover type or feature may be included in an IFOV during scanning. As a result, reflected radiation from a pixel as observed in RS imagery has rarely interacted with a volume composed of a single homogeneous material (Adams et al., 1985; Van der Meer, 1999). These pixels usually consist of a combination of one or more surface materials (mixed pixels) because the radiance value detected by the sensor comes from a heterogeneous ground surface material. The larger the pixel footprint the less likely it is that it consists of pure material. Even when the FOV of a sensor is filled by a single target material, mixing of radiance from adjacent pixels can occur because of atmospheric scattering (Settle and Drake, 1993). According to ENVI Tutorial, (2000) the possible sources for pixel mixing summarized as:

- All materials present in the FOV contribute to the mixed reflectance sensed at a pixel;
- Variable illumination conditions due to topographic effects result in spectrally mixed signals, a very thin volume of material interacts with incident sunlight and;
- Contribution from the optical system of the sensor system itself.

The nature and characteristics of the ground surface also contributes for the mixing of pixels in an image. These ground situations according to Fisher, (1997) include:

- Boundaries between two or more mapping units (e.g. field woodland boundary);
- The integrate between central concepts of the mappable phenomenon;
- Linear sub-pixel objects;
- Small sub-pixel objects (e.g. a house or a tree);

These situations are demonstrated in the following diagrams.

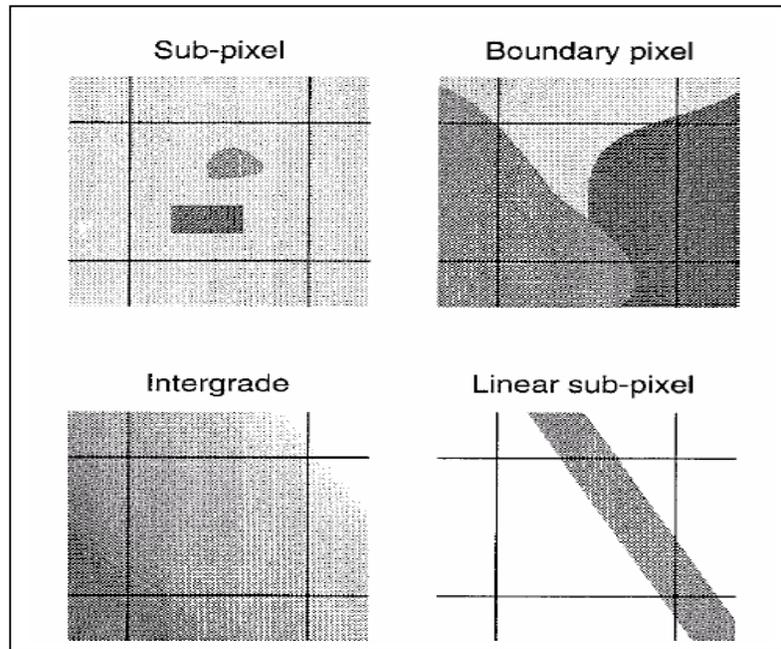


Figure 2-2 Examples of mixed pixels occurrence on the ground (Fisher 1997)

### 2.3. Spectral Mixture Analysis (SMA)

The extraction of land cover information from images traditionally is viewed as a classification problem which labels each pixel in the image as one of only a few possible classes. However, in reality, all degrees of mixing of pure land cover classes within pixels can be found due to the continuum variation found in the landscape (Foody, 1996b) and the intrinsic mixed nature of most land covers (Schowengerdt, 1997). Hence, discretization of land cover into a limited number of categories contributes to a loss of information. Alternatively, mixture modeling in RS predicts the respective fractions/abundance of land cover classes within pixels and characterizes land cover more accurately by decomposing a pixel into a small number of “pure” classes. The resulting mixture map represents the fractions of pure land covers within pixels. For example, mixture fractions - 80% clayey soil, 10% sandy soil, and 10% calcareous soils, a pixel may be denoted in such maps by the clay by majority rule. Therefore, mixture analysis is a necessary component for capturing sub-pixel heterogeneity in the characterization of surface features from remotely sensed images.

During the last couples of decades, scientists have been researching on ways of unmixing pixels to determine the proportions of their components through modeling of spectral mixtures (Charles and Arnon, 1996). Several types of mixture models are proposed including, linear, probabilistic, geometric, and stochastic geometric models. A detailed review of these models is discussed in Charles and Arnon, (1996). It is summarized as follows.

#### Linear Model

A simple and commonly used method of mixture analysis is the linear model (Adams et al., 1985). In the linear model, the spectrum is considered as a linear combination of the "pure" spectra of the

materials located in the pixel area, weighted by their fractional abundance. According to Settle and Drake, (1993) the following assumptions apply:

- No significant amounts of multiple scattering exist between different cover types: each photon that reaches the sensor has interacted with only one cover type;
- Each endmember must be known, implying that for each pixel in an image, the land cover equals unity.

In linear mixture modeling the resulting pixel reflectance spectrum is a summation of the individual material reflectance functions multiplied by the surface fraction they constitute (Singer and McCord, 1979). That means with known number of endmembers and giving the spectra of each pure component, the observed pixel value in any spectral band is modelled by the linear combination of the spectral response of component within the pixel. This linear mixture model can be mathematically described for a pixel in band  $i$ , with the observed pixel reflectance  $r_i$  as a linear equation,

$$r_i = f_1 a_{i1} + f_2 a_{i2} + \dots + f_j a_{ij} + \varepsilon_i \quad (6)$$

Where:  $f_j$  is the  $j^{th}$  fraction of endmember in a pixel,  $j$  is total number of endmembers in the scene,  $a_{i,j}$  is the pure reflectance from the  $j^{th}$  endmembers in the  $i^{th}$  pixel,  $\varepsilon_i$  is an error term.

Equation (6) above can be simplified as:

$$r_i = \sum_{j=1}^n f_j a_{ij} + \varepsilon_i \quad (7)$$

Equation (7) represents a system of linear equations that can be expressed in matrix notation as:

$$r = Af + \varepsilon \quad (8)$$

Where, the observed reflectance  $r$  is a product of pure endmember spectra  $A$  (an  $m \times n$  matrix) and endmember fraction  $f$  plus an error (vector)  $\varepsilon$ . Generally the number of unknowns should be less than or equal to the number of equations for there to be a convenient solution. This implies that the number of components  $n$  should be less than or equal to the number of bands.

### **Probabilistic Model**

In addition to the assumption of linear model, reflectance is likely to vary across space and time, even for a narrowly defined endmembers (David et al., 2004). For example a spectrum of soil does not appear identical for all locations and years, but varies according to a range of environmental and management factors, such as temperature and ploughing condition. Hence, rather than defining endmembers with a single spectrum, as in most approaches do, it is also possible to define endmembers as a set of spectra which represent the whole range of potential variability (Asner and Lobell, 2000). A typical type of probability model is that employing the approximate maximum likelihood techniques as reported by Marsh et al. (1980). The uncertainty in endmember fraction arising from endmember variability can then be quantified in a spectrum by randomly selected from each endmember set, fractions as estimated, and this is repeated a large number of times to derive a distribution of fractions for each endmember.

Thus endmember fractions are not estimated as a single value, but rather as a probability distribution that can be used to construct confidence intervals appropriate to the desired application. This is achieved employing a function which transforms a multivariate set of measurements on a sample onto the linear discriminant line, which in this case is the line joining the multivariate means for the two classes of homogeneous pixels in the multispectral space. Squared distance measured along the discriminant function line are called Mahalanobis or M-distances (Marsh et al., 1980).

The approximate maximum likelihood technique is based on the assumption that the multivariate position of the spectral signature of a mixed pixel relative to those of the means of its homogeneous components gives an approximate maximum likelihood estimate of the proportions of these components under multivariate normally conditions. If two components need to be discriminated denoted by  $x$  and  $y$ , then the relationship for estimating the proportion of one of them in a pixel is given by (marsh et al., 1980):

$$P_y = 0.5 + 0.5 \frac{d(m,x) - d(m,y)}{d(x,y)} \quad (9)$$

Where:

- $P_y$  = proportion of component  $y$  in the mixed pixels;
- $d(x,y)$  = M-distance between means of homogeneous components  $x$  and  $y$ ;
- $d(m,x)$  = M-distance between mixed pixel  $m$  and mean  $Y$ ;
- $d(m,y)$  = M-distance between mixed pixel  $m$  and mean  $Y$ ;
- $P_y = 0$ , if result is negative; and
- $P_y = 1$ , if the result is greater than 1.

A major drawback of this model is that it is limited to determine proportions up to only two endmembers in a mixed model.

### **The Geometric-optical model and stochastic geometric model**

In this model, the landscape is conceived as a plane ground containing trees which cast their shadows on the ground and some times on neighboring trees. Thus, only four ground cover types are usually considered, namely sunlit canopy (C), shadow canopy (T), sunlit background (G), and shadow background (Z). The reflectance  $R$  of a pixel is expressed as a linear combination of the reflectances of these components each weighted by the area covered by it as seen from the direction of observation (Charles and Arnon, 1996). Thus,

$$R = (A_C R_C + A_T R_T + A_G R_G + A_Z R_Z) / A \quad (10)$$

Where:  $R_C$   $R_T$   $R_G$   $R_Z$  denotes the reflectances of the four components as indicated by their subscripts, the subscripted  $A$ 's represent the area occupied by them in the pixel, and the unsubscript  $A$  is the area of the pixel itself.

Geometric optical model is more complex and needs more input than the linear and probability models (Charles and Arnon, 1996).

The stochastic geometric model is a special case of geometric model in which the scene geometric parameters are treated as a random variety in order to absorb the random variability in their spatial structure.

### 2.3.1. Comparison between Models

In a general sense, linear, probabilistic, are relatively simple, where as geometric and stochastic models are complicated, involving the incorporation of parameters of scene geometry. All models discussed above have common in the way that all of them express the reflectance of a pixel as a function of both the characteristics reflectance and the areal proportions of its component land cover. There are some differences in the number and nature of components that can be resolved with the different model. In linear model, the scene variability is accounted for by means of random residual while in geometric-optical and stochastic models, it is based on the analysis of the scene geometry. In probabilistic model it is done by means of some statistical methods like maximum likelihood technique. A brief summary of the applicability of different models to each land cover classes are reported by (Charles and Arnon, 1996). This is summarized in the table below (“yes” is used to designate models applicable to each of the stated problem while “no” is used for models not applicable).

Applications	Model Applicability			
	Linear	Probabilistic	Geometric	Stochastic
Estimation of ...				
Vegetation versus bare ground proportions in dense forest	Yes	Yes	No	No
Vegetation versus bare ground proportions in a spare forest	Yes	Yes	Yes	Yes
Proportion of different plant communities	Yes	Yes	No	No
Proportion of area coverage of different crops	Yes	Yes	No	No
<b>Proportion of different soil or rock types</b>	<b>Yes</b>	Yes	No	No
Proportion of different minerals	Yes	Yes	No	No

Table 2-2: Applicability of mixture models for different applications: Adapted from (Charles and Arnon, 1996)

The table above is giving an overview of different spectral mixture models in terms of their applicability in different fields of applications. The table content suggests that for our application, which is to identify SSOM levels, linear mixture model (high lighted) can be used.

## 2.4. Endmember selection

Pure features in mixed pixels are referred to as endmembers of that pixel. Selection and identification of spectral endmembers in an image is the key to success of a linear mixture model (Dennison and Roberts, 2003). A set of endmembers should allow the description of all spectral variability for all pixels.

Two different approaches have generally been used to define endmembers in a mixing model:

- the use of a library of reflectance spectra (Adam et al., 1986);
- the use of purest pixels extracted from the image data itself

Endmembers resulting through the first option are denoted as known endmembers where as the second option results in deriving endmembers. The maximum number of endmembers that can be derived from an image equals number of bands minus one. Typically, three to seven endmembers are appropriate for most applications, depending on the number of channels used and the spectral variability of the scene components (Mustard et al., 1999). Painter et al., (1998) used image spectra and prior field knowledge to select endmembers for their analysis. Therefore, as no library spectra collected from field exists for this study, the second approach with the help of field reference data is adapted.

### 2.4.1. Endmember selection techniques

Different techniques have been used for selecting endmembers from an image. These includes: Pixel Purity Index (PPI), Principal Component Analysis (PCA), Spectral Angle Mapper (SAM). All these are explained as follows.

#### **Pixel Purity Index (PPI)**

PPI function finds the most spectrally pure or "extreme" pixels in multispectral and hyperspectral data (Boardmann et al., 1995). These correspond to the materials with spectra that combine linearly to produce all of the spectra in the image. The PPI is computed by using projections of n-dimensional scatter plots to 2-D space and marking the extreme pixels in each projection. PPI algorithm uses an array called Systolic array consists of computing a very large number of dot-products and all these dot-products can be performed simultaneously. Interested reader can refer the article by Dominique et al., (1999). The output is an image (PPI Image) in which the DN of each pixel in the image corresponds to the number of times that pixel was recorded as extreme. Thus bright pixels in the image show the spatial location of spectral endmembers. Image thresholding is used to select several thousand pixels for further analysis, thus significantly reducing the number of pixels to be examined.

#### **Principal Component Analysis (PCA)**

PCA is a coordinate transformation typically associated with multi-band imagery (Schowengerdt, 1997). PCA can be applied to compact the redundant data into fewer layers. This can be done by transforming a set of image bands, as that the new layers also called components are not correlated with one another. According to Schowengerdt, (1997) the mathematical expression can be given as:

$$PCA = T \cdot DN' \quad (11)$$

Where:  $DN'$  is the digital number matrix,  $T$  is the transformation matrix expressed by the eigenvectors  $e_k$  as:

$$T' = \begin{bmatrix} e_{11} & \cdot & \cdot & e_{1k} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ e_{k1} & \cdot & \cdot & e_{kk} \end{bmatrix} \quad (12)$$

The resulting components are often more interpretable than the original images. Because of this, a new component carries new information. The relevance of the components are explained by the percentage of variance, the first two or three components carry most of the real information of the original data set, while the later components describe only minor variations (some times noise). When visualizing the first components in a 2D plot, spectrally pure pixels are found at the vertices of the scatter plot that bounds the data space of the principal components (ENVI Tutorial, 2000).

### Spectral Angle Mapper (SAM)

SAM is an automatic method that calculates the spectral similarity between a test reflectance spectrum and a reference reflectance spectrum (ENVI Tutorial, 2000). The spectral similarity between the test (or pixel) spectrum and the reference (laboratory/field) spectrum is expressed in terms of the average angle between the two spectra. In this approach the spectra are treated as vectors in space with dimensionality equal to the number of bands. This method is insensitive to illumination. The spectral mapping algorithm determines the similarity between a test spectrum (endmember spectrum)  $t$  and a reference spectrum (target)  $r$  for  $n$  number of bands from the arc-cosine of the dot-product of the two spectra defined as:

$$\alpha = \cos^{-1} \left[ \frac{\sum_{i=1}^n t_i r_i}{\left[ \sum_{i=1}^n t_i^2 \right]^{\frac{1}{2}} \left[ \sum_{i=1}^n r_i^2 \right]^{\frac{1}{2}}} \right] \quad (13)$$

Therefore the output of SAM for each pixel is an angular distance between the two spectra in radians (ranging from 0 to  $\pi/2$ ). Output of SAM will be a new set of data with as many bands as the number of unknown (endmember) spectra given as input in the algorithm. The other output of SAM classifier is multi-layer images, which are called “rule” images showing the best SAM match between the reference and test spectrum at each pixels. Darker pixels in the rule images represent smaller spectral angles, and thus spectra that are more similar to the endmember spectrum and vice versa.

## 2.5. Linear spectral unmixing

In most conventional image classification algorithms, thematic information (giving a fixed class to a pixel) is obtained disregarding the compositional nature of surface materials. The assumption that a pixel can be assigned only to a single cover type has been unacceptable in many investigations due to the heterogeneity of the earth’s surface. The idea of acknowledging the compositional nature of

natural surfaces has been done by spectral unmixing (Adam et al, 1985). Using these techniques it is possible to derive the relative or absolute abundances of a number of spectrally pure components, together termed as endmembers, contributing to the observed reflectance of the image by deriving the fractions of pure endmembers from the mixed pixel signal. Therefore, the fractions at each pixel (the unmixing result) can be solved by taking the inverse of equation 8.

$$f + \varepsilon = A^{-1}r \quad (14)$$

The residual error  $\varepsilon$  is the difference between the measured and the modelled spectrum in each band.

Additionally, one can impose constraints upon the solutions of equation 14. One set of constraints requires the fractions within a pixel to sum to unity. But these constraints are only meaningful when considering the scientific field of applications (e.g. the earth science perspective), from a statistical point of view, it is meaning less to force mixture models to constraining to the data (Van der Meer and de Jong, 2000).

$$\sum_{j=1}^n f_j = 1 \quad (15)$$

A fully constrained set would also require each individual fraction to lie between 0 and 1:

$$0 \leq f_j \leq 1 \quad (16)$$

Although fully constrained fractions seem intuitive, values outside the bounds have some physical meaning (Adams et al.1993). However, negative fractions can also occur due to an improper numbers of endmember selections since the number of endmembers is restricted by the number of image bands used i.e. number of endmembers should be less than the number of bands at least by one, or because of spectral ambiguity (Gross, 1998).

The RMS error is used to assess the fit of the spectral unmixing model. If an image has  $m$  pixels, the RMS error is calculated for all image pixels. Larger RMS error is an indication of poorly fit model. So, the error image, which is produced during unmixing process, can be used to assess whether the endmembers are properly selected and whether the number of selected endmembers is sufficient or not.

$$\text{RMS} = \sqrt{\frac{\sum_{i=1}^m (\varepsilon_i^2)}{m}} \quad (17)$$

### 2.5.1. What we get after unmixing?

In most conventional image classification techniques such as maximum likelihood, the end product is a single thematic map. Where as spectral unmixing technique decomposes the image spectra into a series of endmember fractions resulting end products that are a series of images showing fractional abundance of the endmembers i.e. showing areal proportions (quantitatively) of each endmember present in each pixel of the input image. These fraction abundance images have the same extent as the original image. The second out put of the unmixing classification is an error image that gives a

quantitative error distribution in each pixel of an image. This helps to assess the fitness of the model. Small RMSE values imply a good unmixing model.

## **2.6. Summary and conclusion**

From the literature review given in this chapter, the following basic concepts are found to be relevant for this research work.

Handling large numbers of bands in hyperspectral images is a problem during image processing. Many of the bands may deliver redundant information. Therefore, selecting optimum number of bands for the given application is found to be essential. For this purpose we found SDA statistical technique is useful since we want to discriminate different SSOM levels.

Causes of mixed pixels in an image, different SMA models, and various techniques used for endmember selection are reviewed in detail. From the available SMA modeling techniques, linear spectral unmixing model is found to be relevant for this study according to the explanation given in section 2.3.

### 3. Study area and Data preparation

In this chapter a brief description of the study area and basic data preparation processes are explained.

#### 3.1. Study Area discription

##### Location and climate

The study area is located 45 km North of Antequera town in the south of Spain between latitude  $4^{\circ}31'10''E$  and  $4^{\circ}51'10''E$  and longitudes  $37^{\circ}00'04''N$  and  $37^{\circ}10'04''N$ . It encompasses a region of approximately 30 by 21 km. The area is characterized by continental Mediterranean type of climate and has a dry summer followed by a cool wet winter. The overall annual temperature is  $17.6^{\circ}C$  (Kariuki, 2004). Main rainfall occurs between October and April. The average annual rainfall in the area is about 470mm. The highest rainfall with a maximum of 770mm occurs in December, while the minimum is 43mm (Ahmadou et al., 1999).

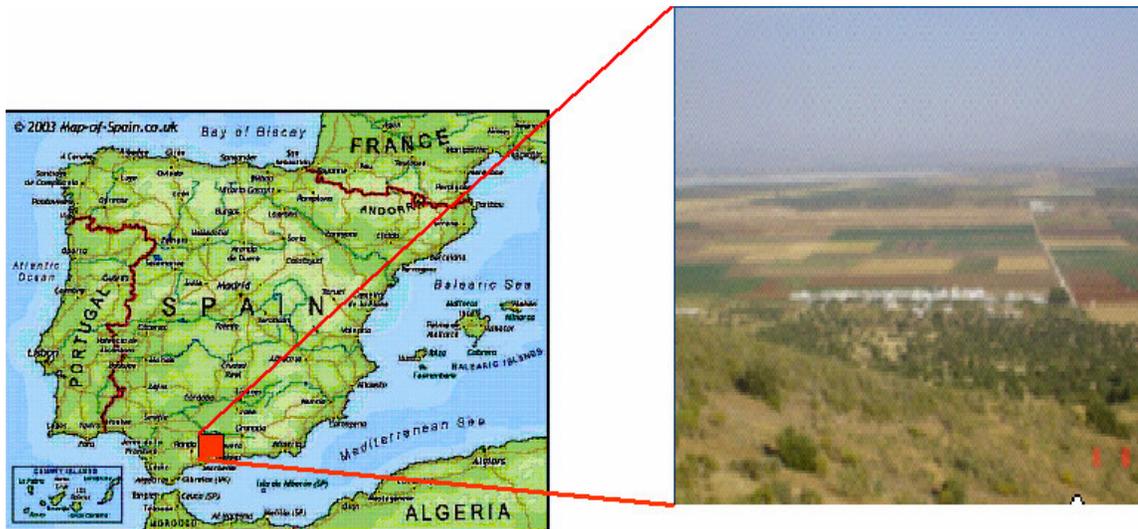


Figure 3-1 Location map of the case study area

##### Geomorphology and Soils

The present geomorphologic configuration of the area is the result of several processes such as tectonism, denudation and deposition (Kariuki, 2004). The main landscapes in the area are hill land and valley. Hill landscape is characterized by isolated hills, and ridges that have been derived from sandstone and marls. The soils in the hill land area are shallow, clay loam and sandy loam in texture. The parent materials of these soils are sediment consisting of gypsum, limestone and marls. The

valley landscape in the area has three main important relief forms based on the nature of deposition. These include lacustrine flat, terraces and floodplain. Topographically the valley regions are gently to gently sloping. The soils in these areas are moderately deep, clayey and sandy texture.

**Land use**

Most of the study area (~75%) is under agricultural cropping, mainly olives, barely and sunflower (Ahmadou et al., 1999). Olive, which is under drip irrigation, is the most common perennial crop. Barley and sunflower are the main annual crops growing under rain-fed condition. They are cultivated on moderately steep to very gently sloping areas in almost every landscape. The agricultural practices are dependent on socio-economic conditions, soil condition, availability of water for irrigation and climatic factors(Ahmadou et al., 1999).

**3.1.1. Selection Criteria**

The availability of RS data from two sensors (DAIS, and Landsat TM) and the availability of ground truth data for validating the results provide a unique opportunity to characterise and model the soil-landscape characteristics of the semi-arid area of Antequera. For this study these two factors were the main criteria for selecting the case study area that helps to characterize the SSOM dimensions, soil-landscape relationships of the study area.

**3.2. Data preparation**

**3.2.1. Methods**

Raw RS image and field reference data should undergo data preparation processes before using them for real applications. Therefore, all earth observation data need some pre-processing procedures before it can be used for analysis. A detailed procedure for data preparation is shown in the figure below.

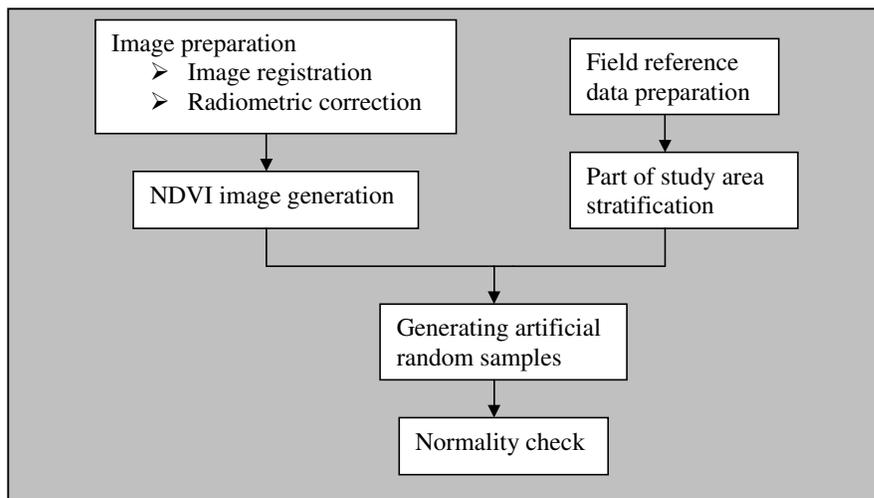


Figure 3-2 Data preparation procedure

The pre-processing procedures are considered as part of basic data preparation and can be divided into two categories as: image registration and radiometric correction. These are explained in sections 3.2.2 and 3.2.3. As explained in section 1.6.2, the available ground truth data in  $A_1$  were very limited (only eight test pits). To solve this limitation, we need to generate artificial random samples. It is explained in section 3.2.4.

### **3.2.2. DAIS Image preparation**

#### **Image Registration**

Information is collected on a pixel-by-pixel basis and therefore geometric corrections must be made to the data to insure that each pixel can be referenced to a real-world coordinate system and used in combination with other spatial datasets. Geometric corrections are needed for geometric deformation due to the variation in altitude, velocity of the sensor platform, for variation in scan speed and in the sweep of the sensor field of view, earth curvature, and relief displacement (Donat-P.Hader, 1992). Random distortions need to be corrected by the analyst through selection of sufficient number of ground control points (GCP) with correct coordinates, usually from a map or GPS points which can be localized in the satellite image. Therefore, before one can begin with analysis of the images they need to be georeferenced. That means the image coordinates should be adjusted to comply with a known coordinate system. If all images have the same geographic coordinate system it will be possible to overlay, analyse and compare the images with each other. These can be done by using scanned topographic maps with a known coordinate system. Features standing out on both the map and the image are used to reference them to each other; usually these are road intersections, buildings, and land use boundaries.

The DAIS image was previously georeferenced into UTM projection using 1909 ellipsoid and European1950 datum based on 1:50,000 north oriented topo-map of the study area. The interpolation method used was nearest neighbourhood. This type of interpolation preserves the spectral integrity of the image pixels. Other resampling techniques such as bilinear and cubic interpolations perform spectral averaging of neighbouring pixels, yielding pixels with spectral properties that are less likely to provide optimal results.

#### **Radiometric correction**

Radiometric correction is required because of factors such as changes in scene illumination, atmospheric conditions, viewing geometric and instrumental response characteristics of which influences the radiance measured by any given system over a given object (Donat-P.Hader, 1992). It considers both haze (due to scattered path radiance and it reduces the image contrast) and sun elevation correction, which accounts for seasonal position of the sun relative to the earth. Radiometric correction is used to change raw DN values to surface reflectance values (ATCOR2 User Manual, 1997).

Radiometric and atmospheric calibration of DAIS has been done by DRL, the supplier as explained in section 2.1.1. DAIS image with extension ending .atm-BSQ, which means it is both atmospherically and radiometrically calibrated during image production. Detail information can be found in Lehmann et al., (1995). The atmospheric correction is based on the MODTRAN radiative transfer code (Berk et

al., 1989). This algorithm helps to convert image data in the reflective channels of the DAIS sensor to ground reflectance. Both the atmospheric and radiometric corrections of DAIS image was done previously. Therefore, radiometric calibration was not carried out.

### **3.2.3. Landsat TM preparation**

#### **Image registration**

To use the two image data in the same environment for further processing, the Landsat TM was georeferenced into UTM projection, 1909 ellipsoid and datum European1950 using a 1:50,000 north oriented topo-map of the study area. We used nearest neighbourhood interpolation method since we want to preserve spectral values of the image pixels.

#### **Radiometric correction**

The derivation of physical quantities like, ground reflectance is a current research topic in RS (Geotz, 1992). ATCOR2 is a spatially-adaptive fast atmospheric correction algorithm working with atmospheric database (Richter, 1996). ATCOR2 has been developed mainly for satellite sensors with small field of view such as Landsat TM. This is essential in improving imagery data like to remove or reduce the influence of atmosphere and solar illumination, and also important for extracting ground reflectance data.

Since a basic prerequisite of our further analysis is spectral information of the image pixels, the TM image needs to be corrected for atmospheric effects. This was achieved by applying atmospheric correction using ATCOR2 algorithm in ERDAS 8.7 software. Sun position is one parameter for ATCOR2. It was not possible to know the exact time of image acquiring; however from cross-checking with USGS Glovis website, it was possible to approximate it. Most of the time, the satellites are taking images in the Antequera area of Spain between 10:14 and 10:15 UTC time. Hence we used the average value 10:1433343 for our analysis. Then the calculated solar zenith (in degree) is 21.1. This value was used for atmospheric correction.

### **3.2.4. Reference data preparation**

Spectral data is used in RS study to relate and model scene reflectance values with surface features and phenomenon. Different materials have different spectral characteristics that are dependent on their physical and chemical property. Reflectance data can be extracted either in the field using field spectrometer or directly from the images itself.

There was no available field spectral reflectance data for this study. We used GPS X and Y coordinates of the available ground truth points for extracting reflectance spectra from the images. For this purpose, we used two soil profile description database tables of the study area. The first database table (OBSANA) contains detail descriptions of soil types, measured SSOM values and soil horizon description. The other table (OBSITE) contains GPS coordinates of each of the sample points with detail description of landscape, topography and parent materials. The X, Y map coordinates, pit number, SSOM%, landscape, topography information were obtained from the available database tables by running a simple SQL using easting and northing key words for X and Y map coordinates

respectively. GPS coordinates were cross-checked with the Antequera soil profile description manual and some missing records were filled. In this study we considered only the top horizon i.e. surface (hor\_no =1 in the database tables). In the SQL below “hi” represents hill (stratum1) and for other landscapes different representations were used according to their description in the database tables:

```
Select OBS_ID, SSOM, topography, easting, northing  
From OBSANA, OBSITE  
Where OBSITE.OBS_ID=OBSAN.OBS_ID and  
Hor_no=1 and landscape="hi";
```

The above query also helped us to see the variation of SSOM in the different landscapes that will be useful in our next task. The available ground truth samples in A<sub>1</sub> were very limited. To minimize this limitation, we generated artificial random samples after stratifying the study area into four strata (stratum 1, stratum 2, stratum 3 and stratum 4) based on SSOM levels. This was done based on prior field knowledge. The descriptions of the four strata are given below.

**Stratum 1** – this stratum concentrates on the two extremes of the DAIS image as shown in figure 3.3. Rock exposures are dominant in this stratum. The soil thickness is thin. SSOM concentration is expected to be low. There were no test pit data collected from this stratum in the available ground truth data (might be due to rocky nature which made excavation impossible).

**Stratum 2** – it is located next to the upper stratum1 towards the salty lake side (see figure 3.3). This area is described as characteristics of ancient lacustrine (lake deposits) landscape according to the soil profile description manual. The topography is low-laying and flat. It is annually cultivated with olive plantation (Arnaldo, 2001). According to the soil profile description manual, this part is well drained and run-off is slow. The soil varies from brown to dark silty clay loam. In general from field prior knowledge and available test pit data, the SSOM in this stratum is expected to be better than any other strata. For more understanding see the box diagram of the available SSOM data of the strata in figure 1.5.

**Stratum 3** – this stratum covers more than half of the A<sub>1</sub>. It is located in the central part of A<sub>1</sub>. Topographically it is gentle and some times undulating. High SSOM is expected on the flat regions and less on the relatively hilly areas because SSOM can be governed by topographical situation of the area (Grunwald, 1999). It is well drained, however it rarely gets saturated. Parent materials are characteristics of residual colluvium sandstone as referred from the soil profile manual. The soil is dark reddish brown sandy clay loam. When other factors kept constant, sandy soils are poor in SSOM according to the explanation given in section 1.1.2. Thus, low SSOM is expected in sandy soil areas.

**Stratum 4** – it is located between stratum 3 and the lower stratum1. It is dissected by a river running from SW-NE directions as seen from the site map in figure 1-4. The landscape and topographic characteristics in this stratum is not uniform. The central part is almost flat and alluvial deposits are dominant. The river carries sediments from the upper part and flooded it out when it reached the plain area. A part near lower stratum 1 is a transition area that is more of weathered nature being dominated by conglomerate fragments, where as in other areas terraces and hillocks topography are dominant.

The parent material in the central part is fine textured alluvium and sunflower plantations with sprinkled irrigation practices are common in this part of the stratum. Better SSOM is expected in the alluvial deposit area.

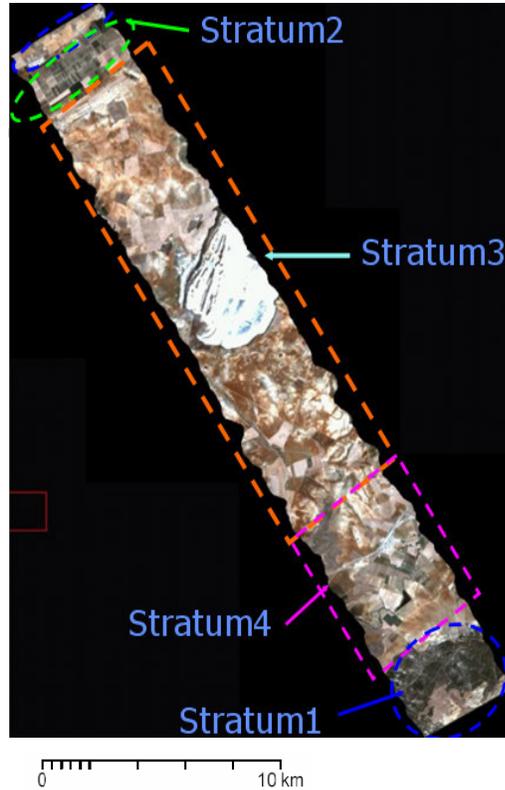


Figure 3-3 Different strata locations on DAIS image

There is vegetation cover in some parts of the study area which could have influence on the soil reflectance values since the reflectance coming from the bare soil surface could have been masked by vegetation. This was overcome by application of masking based on reported discriminating indices for high vegetation cover and percentages at which a vegetation cover can be assumed to have little or no influence on the soil reflectance. At vegetation covers as high as 15% surfaces have been described to appear as soils (Girard-Ganneau, 1975) where as vegetation cover in excess of 40% makes the spectral behaviour that of vegetation. Therefore, to avoid the problem coming from vegetation masking effect, we prepared NDVI image of the study area (see appendix 2). Orueta, et al, (1998) reported a threshold of NDVI of 0.5 as enough to separate areas of little or no vegetation cover from those in which vegetation would interfere with the establishment of soil surface properties. Therefore the image was classified into high NDVI area and low NDVI area after slicing the NDVI map using a threshold of 0.5.

### Generating random samples

The area of each stratum was used to decide how many number of artificial samples to be taken per stratum. The area in square meter of each stratum is:

- Stratum 1 –total area of 7409917 m<sup>2</sup>
- Stratum 2- a total area of 5557437 m<sup>2</sup>
- Stratum 3 – a total area 32603634 m<sup>2</sup> (excluding the lake)
- Stratum 4 – a total area of 10003387 m<sup>2</sup>

The whole DAIS scene covers a total area of 55574376 m<sup>2</sup> excluding the salty lake. It was decided to take 300 random sample points. The number of sample points in each stratum was calculated simply by dividing the area of each stratum by the total area covered by DAIS image and multiplying the result by three hundred. Therefore, the number of artificial samples taken in stratum 1, stratum 2, stratum 3, and stratum 4 respectively are 40, 30,176 and 54. After doing so, the artificial samples locations were sensibly selected in areas where NDVI values are less than 0.5.

Once the artificial samples are sensibly generated, we checked for normality of the samples. The Shapiro-Wilk *W* test was used in testing for normality. If the *W* statistic is significant, then the hypothesis that the respective distribution is normal should be rejected. The Shapiro-Wilk *W* test is the preferred test of normality because of its good power properties as compared to a wide range of alternative tests (Statistica 6.0, 2002). In the following figures histogram with normal curves are plotted. As shown in the figures below all the histograms show normal or near normal distribution.

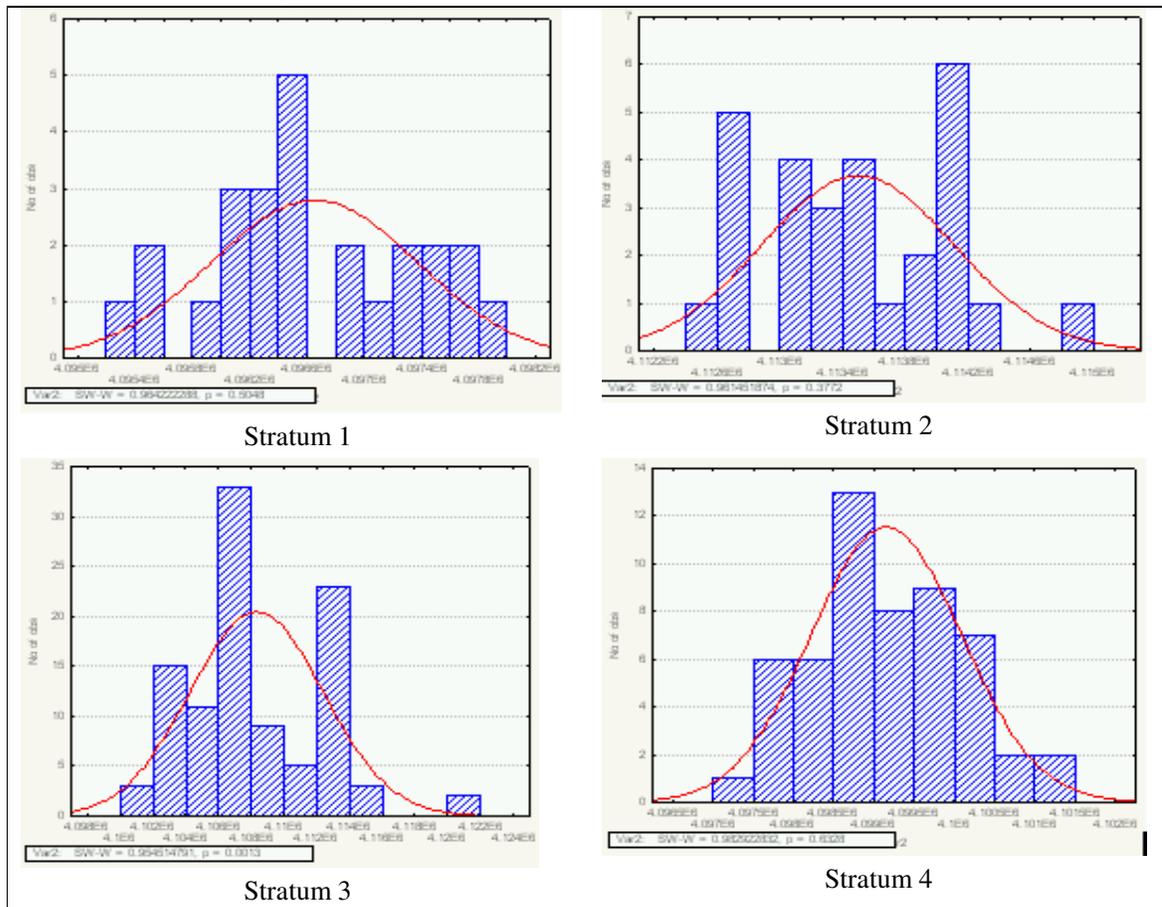


Figure 3-4 Normality check of the artificially generated samples

### **3.3. Summary and conclusion**

In this chapter we briefly explained the detailed data preparation carried out both on the images and reference data. We applied atmospheric correction on the TM image to extract surface reflectance values. To avoid limitation of data in  $A_1$ , we generated 300 artificial randomly distributed samples. Before generating these samples, the study area was stratified into four SSOM levels based on field priori knowledge. We also prepared NDVI image for identifying vegetation and bare soil areas. Then we located the artificial sample points in areas where NDVI values are less than 0.5 since the masking effect of vegetation reflectance is expected to be low below 0.5 as referred from different literatures. The distribution of the generated 300 artificial samples was normal.

## 4. Band selection analysis and results

### 4.1. Introduction

Despite the useful information contained in many bands of hyperspectral images, some of the bands contain redundant information leading to great strain on resources (Prasad, 2000). Thereby, the need to determine optimal hyperspectral bands containing useful information for particular application is an important pre-processing step. This would help to avoid in handling redundant wavebands. The problem then becomes: how to determine which bands to use? A specific goal in this research is to determine optimal DAIS and Landsat TM bands that would help to discriminate SSOM levels in the study area. Therefore, the optimum band selection for SSOM level identification was the first step.

### 4.2. Methods

In this chapter, the band selection analyses and results are explained. A detailed procedure is shown in figure 4-1.

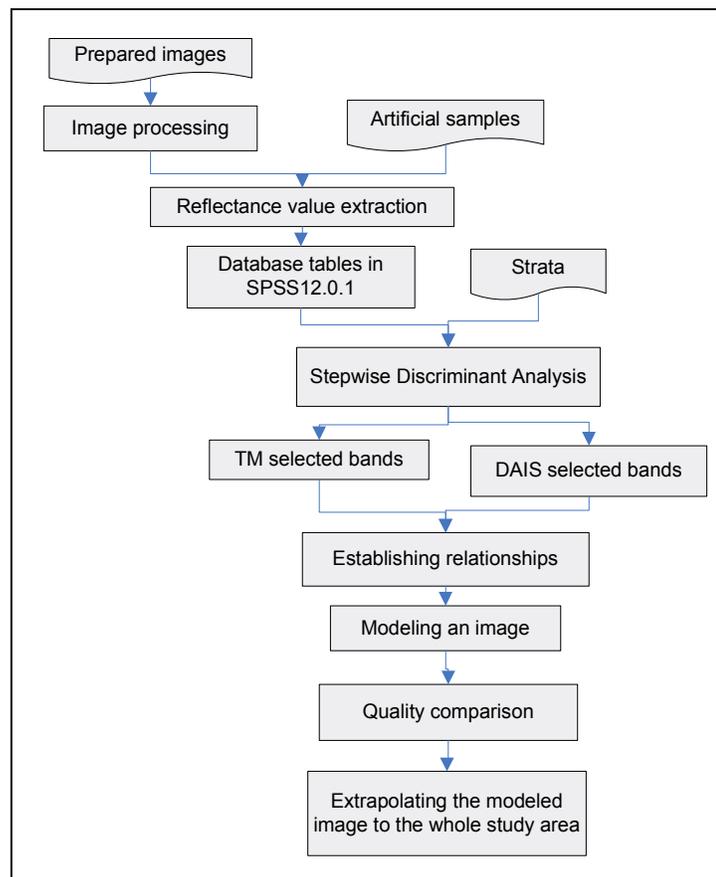


Figure 4-1 Band selection and modeling procedure

We started the band selection analysis by performing some image processing. This is presented in section 4.3. Next the optimum numbers of bands that are suitable for discriminating SSOM levels are selected by integrating the images data, strata and artificial samples by means of SDA. This is presented in section 4.4 and the subsequent subsections. A relationship between selected DAIS and Landsat TM bands has been developed; by doing so we created a modeled image. This is presented in section 4.5. The developed relationship has been extrapolated to the whole study area (A) after checking the quality of the modeled DAIS. This is explained in section 4.6.

### **4.3. Image processing**

We started image processing by eliminating some of bands that are less-useful for SSOM discrimination.

#### **4.3.1 DAIS band elimination**

Recent researches e.g. Thenkabail et al. (2000) have shown that bands located in specific portions of the spectrum have the ability to provide useful information for a given application. Therefore, eliminating wavebands with noisy data or providing little predictive ability reduces the number of bands to be handled. Hence eliminating suspected bands were the first step in our analysis.

Five wavebands were removed from the original 79 spectral bands of the DAIS before processing began. These include wavebands outside the 0.4 -2.50  $\mu\text{m}$  spectrum region: no previous research had identified any correlations between SSOM and reflectance value here. In addition, previous studies have reported that under laboratory conditions, the spectral information of the visible, near-infrared and short wave infrared (0.4–2.5  $\mu\text{m}$ ) spectral regions provide a promising capability to identify soil, vegetation, rock and mineral materials Stoner, (1981).

A visual inspection of DAIS showed that wavebands from 1.948-2.179  $\mu\text{m}$  and 2.317-2.395  $\mu\text{m}$  were contaminated with stripe and noise respectively. Reports by Strobl et al., (1996) have indicated that the striping in some SWIR region of DAIS is probably caused by vibration of the Dornier DO228 prop-engines from detector changeover during spectrophotometer operation. Baumgardner et al. 1985 also reported that strong water absorption bands and vibrational mode of the hydroxyl ion in DAIS bands (2.317-2.395  $\mu\text{m}$  as clearly seen in our DAIS image) causes noise.

Before eliminating these wavebands, we tried to reduce both stripping and noise effects by applying Minimum Noise Fraction (MNF) (Green et al., 1988). This transformation algorithm can be seen as a cascaded Principal Component transformation (PCA). The MNF transformation decorrelates and rescales the noise in the data. The output of the MNF transformation is a given eigenvalue for each MNF band. When the eigenvalue of an MNF band is small ( $<1$ ), the image is dominated by noise Green et al., (1998).

The MNF analysis did not show any improvement for wavelengths 1.948-2.179  $\mu\text{m}$ . We also applied 3X3 smoothing filter that resulted insignificant improvement. Previous studies by Kariuki and van der Meer, (2003) reported that these wavelengths region of DAIS image are less-useful. We also found them less-useful. Therefore, we excluded them from our further analysis.

Some of the bands in 2.317-2.395  $\mu\text{m}$  were cleaned from the noise after MNF. Noise from channels 2.342-2.395  $\mu\text{m}$  however, could not be removed. Thus, we also omitted them from our analysis. These bands displayed large dips in this region of the spectrum. More detail information about this spectral region of the DAIS bands can be found in Baumgardner et al., (1985). After band elimination was completed, 64 DAIS bands were left for further analysis.

#### **4.3.2 Landsat TM band elimination**

Landsat TM is a sun-synchronous satellite with seven spectral bands in visible and thermal regions, for different applications. Landsat TM is the only non-meteorological satellite with a thermal band. This is useful for studying energy processes at the earth surface. For instance, it can be used for crop temperature variability within irrigated areas, for vegetation stress analysis and for other thermal mapping applications. The purpose of the current work, however, is to find useful wavebands for discriminating SSOM levels. Therefore, we did not consider this TM band.

#### **4.4. Band selection**

After eliminating some of less-useful bands, still all the remaining bands are not expected to be useful for SSOM level discrimination. Thus, there is a need to select specific bands where SSOM matter is best discriminated or identified. There is no single best approach is available to determine the optimal number of bands required that provide the best discrimination of SSOM characteristics (Prasad et al., 2004). In this research we used stepwise discriminant analysis.

##### **4.4.1. Stepwise Discriminant Analysis (SDA)**

This section consisted in making an analysis using SDA on the various spectral bands in attempt to determine wavelengths that are useful in discriminating SSOM levels. To determine the most useful spectral bands for SSOM discrimination, SDA techniques were employed using Statistical Package for the Soil Sciences software (SPSS 12.0.1). The technique helps to filter out a few but important independent spectral bands from DAIS and Landsat TM images. As explained in chapter three,  $A_1$  of the study area was stratified into four SSOM levels and 300 artificial samples were sensibly generated. Reflectance values at the 300 points were extracted from both images. A database of 64 columns and 300 rows for DAIS and 6 columns and 300 rows for Landsat TM were developed and used in the SDA.

##### **i) SDA result of DAIS**

Seven DAIS bands are selected after discriminant analysis. The SDA outputs Wilk's  $\lambda$  and standardized discriminant coefficient values. The values of Wilk's  $\lambda$  are indicative of separability or discriminatory power of spectral wavebands. Lesser the value of Wilk's  $\lambda$ , greater is the spectral separability between the strata and vice versa (Prasad, 2000). At each step if a waveband already in the model fails to meet the criterion to stay (decided based on significance level of F test analysis of covariance), the worst such variable is removed (Statistica 6.0, 2002), otherwise, the waveband that contributes most to the discriminatory power of the model is entered. Table 4-1 shows Wilk's  $\lambda$  values of the selected wavebands.

Steps	wavelengths	$\lambda$
1	2.304	0.884
2	2.010	0.788
3	2.026	0.684
4	0.554	0.641
5	0.939	0.596
6	1.020	0.515
7	0.990	0.478

Table 4-1 Wilk's  $\lambda$  values and selected wavelengths in DAIS

The over all discrimination between the four strata is significant ( $\lambda = 0.478$ ,  $f=21$  and  $p=0.000$ ). (See appendix 3).

Table 4-2 shows the eigenvalues as well as the factor structure matrix from the discriminant analysis. The standardized discriminant coefficients represent the correlation between the variables and the discriminant functions. Discriminant analysis produces functions that help to define the groups (strata in our case); the maximum number of functions that can be defined is one less than the number of groups. Therefore, we can have three functions since we have four strata. The functions first distinguish the first group from the other three, then the second group from the other two, and so on. These are identified by the eigenvalues on the output, showing what percentage of variance is accounted for by each function. The largest proportion of the explained variance (60.4%) is contained in the first discriminant function. The highest factor structure coefficients are contained at wavelength 1.020 0.990, 0.939, 2.026, and 2.010 $\mu\text{m}$  (coefficients > 1). In the second discriminant function large contribution is contained at 1.020, 0.990, and 0.939  $\mu\text{m}$  and less at 0.554 and 2.304 $\mu\text{m}$  and the third discriminant function shows the largest contribution at 0.990, 1.020, and 2.01  $\mu\text{m}$  and less at 0.554, 2.026 and 2.304  $\mu\text{m}$ .

	Function1	Function 2	Function 3
Eigenvalues	0.53(60.4%)	0.283(92.6%)	0.065(100%)
0.554 $\mu\text{m}$	-0.836	1.303	1.237
0.939 $\mu\text{m}$	3.174	-4.462	-0.373
0.990 $\mu\text{m}$	-5.087	-6.407	4.712
1.020 $\mu\text{m}$	4.468	10.006	-4.542
2.010 $\mu\text{m}$	-2.195	-0.674	-3.164
2.026 $\mu\text{m}$	2.595	-0.834	1.121
2.304 $\mu\text{m}$	0.422	1.213	1.081

Table 4-2 Standardized discriminant coefficients derived from SDA of DAIS

SDA output also includes an unstandardized weight that shows the relative significance of each variable. Table 4-3 shows unstandardized weights that indicate the nature of the discrimination for each discriminant function. The results in the table indicate that the first discriminant function discriminates mostly stratum 3 (positive) from other strata (negative) followed by stratum 4 (small

negative as compared to the other values). The second discriminant function discriminates stratum 2 from others followed by stratum 3, and the third discriminant function discriminates stratum 1 and 3 from strata 2 and 4, however, based on the review of eigenvalues in table 4-2, the magnitude of discrimination is relatively smaller for the second discriminant function and much smaller for the third discriminant function as compared with that of the first discriminant function.

	Function 1	Function 2	Function 3
1.00	-1.122	-0.418	-0.471
2.00	-1.238	1.188	0.255
3.00	0.561	0.127	-0.054
4.00	-0.310	-0.765	0.382

Table 4-3 Unstandardized weights showing the nature of discrimination for each discriminant function derived from DAIS

Another output of SDA is a classification table, showing to which the data were categorized (table 4-4). The classification table contains the percentage of cases correctly classified through the prediction of group membership. Following the leave out cross validation, the confusion matrix shows that we could classify sample to their respective groups from DAIS with an overall accuracy of 70%. This classification rate indicates the degree of discrimination achieved by the discriminant function when using DAIS image.

Predicted Group Membership					Total
Strata	1	2	3	4	
1	40.0	5.0	32.5	22.5	100.0
2	10.0	56.7	33.3	0.0	100.0
3	2.8	1.1	93.2	2.8	100.0
4	1.9	5.6	68.5	24.1	100.0

Table 4-4 Fisher’s linear discriminant classification result of DAIS image

**ii) SDA result of Landsat TM**

Even if, the number of bands in multispectral images are few as compared to hyperspectral images, which do not put any pressure on the resources and processing time, all these bands are not expected to be equally important for identifying SSOM levels. Prasad et al., (2004) reported that only a subset of multispectral bands are effective in discriminating a given feature. Therefore, we also applied SDA for selecting Landsat TM bands that are important in discriminating SSOM levels.

Using SDA, four bands are selected that are relatively useful for discriminating SSOM. In comparison to narrow DAIS bands performance, the broad bands of TM performance are relatively poor as seen from the SDA results. The TM band Wilk’s  $\lambda$  values are higher than the DAIS data.

Steps	Wavebands	$\lambda$
1	2.350	0.884
2	0.520	0.808
3	0.760	0.716
4	0.630	0.674

Table 4-5 Wilk's  $\lambda$  values and selected wavelengths of Landsat TM

The overall discrimination between the four treatment strata was significant ( $\lambda=0.674$ ,  $f=12$  and  $p=0.000$ ). (See appendix 5).

Table 4-6 shows the eigenvalues as well as the factor structure matrix from the discriminant analysis. The largest proportion of the explained variance (60.4%) is contained in the first discriminant function and the highest factor structure coefficients are contained at wavelength 0.76, 0.63 and 2.35  $\mu\text{m}$  (coefficients > 1). The second discriminant function gets large contribution at 0.76 and 0.63  $\mu\text{m}$  and the third discriminant function at 0.630, 0.52 and 2.35  $\mu\text{m}$ .

	Function1	Function 2	Function 3
Eigenvalues (%)	0.34(77)	0.08(96.1)	0.027(100)
0.520 $\mu\text{m}$	0.501	0.007	1.946
0.630 $\mu\text{m}$	1.514	-1.336	-3.452
0.760 $\mu\text{m}$	2.636	1.156	-0.208
2.350 $\mu\text{m}$	1.000	0.354	1.943

Table 4-6 Standardized discriminant coefficients of Landsat TM

The SDA classification result shows that 64.7% of original grouped cases correctly classified.

Predicted Group Membership					
Strata	1.00	2.00	3.00	4.00	Total
1.00	35.0	15.0	30.0	20.00	100.0
2.00	5.0	40.0	50.3	4.70	100.0
3.00	2.3	1.0	95.0	3.40	100.0
4.00	16.7	9.3	53.7	18.5	100.0

Table 4-7 Fisher's linear discriminant classification result of Landsat TM

In summary, using SDA, we managed in selecting the optimum numbers of bands from both images that are useful for discriminating SSOM levels. Both the Wilk's  $\lambda$  value and the classification results show that DAIS image discriminates the strata better than Landsat TM. This is because of the continuous and narrow bands in DAIS might help to resolve SSOM differences better than TM. The next step in our analysis is to establish a relationship between the selected bands in  $A_1$ .

#### 4.5. Establishing Relationship between selected bands

Hyperspectral images provide detail ground surface information since the sensor systems designed for hyperspectral images able to capture continuous and narrow spectral records. On the other hand, multispectral images provide less detailed information due to limitations in taking continuous and narrow records of the sensor system. Previous studies carried out by Daniel et al., (2000) showed that integrating RS images are important to improve identification of SSOM. Several other studies such as Palacios-Orueta and Ustin, (1998) and Clark, (1999) confirmed that satellite sensors, which are composed of few but broad spectral bands, can benefit from fine spectral information of hyperspectral images through band integration. In particular, satellite sensors like Satellite SPOT, Landsat TM, IRS and Advanced ASTER, which are endowed with suitable spatial resolutions (<30 m) for precision farming/agriculture, have the major part of their bands within the VNIR region. Therefore, developing techniques to harness the featureless VNIR spectral data would be indispensable to capacitate satellite sensors to detect subtle differences of target soils nutrients after being sharpened by the fine-tuned spectral data (Daniel et al., 2000). Therefore, integrating image data of DAIS and TM images is expected to enhance SSOM levels discrimination in the study area better than Landsat TM image.

Correlation and regression analysis are useful statistical techniques for both identifying related variables and for modeling and predicting the relationship between variables. Multiple linear regression analysis was used for establishing a mathematical relationship between reflectance values of the selected wavelengths of the two images in  $A_1$ . In the regression analysis the seven DAIS bands were considered as dependent variable and the four TM bands considered as independent variable. Therefore, seven equations were derived by means of multiple linear regression models. The  $R^2$  values for all the equations were equal or nearly equal to 0.5. In all the cases p values were significant, i.e.  $p < 0.05$ . The mathematical relationship is given in the table below.

Dep't Variables	Functions	$R^2$	F	Stand. Error
DAIS0.554	$0.02+1.02TM_2+0.05TM_3+0.4TM_4+0.01TM_7$	0.51	76.2	0.026
DAIS0.939	$0.2+0.73TM_2+0.61TM_3+0.31TM_4+0.43TM_7$	0.52	62.7	0.040
DAIS0.990	$0.2+0.5TM_2+0.4TM_3+0.52TM_4+0.52TM_7$	0.48	68.4	0.043
DAIS1.020	$0.2+0.45TM_2+0.4TM_3+0.6TM_4+0.5TM_7$	0.50	74.4	0.041
DAIS2.010	$0.13+0.53TM_2+0.84TM_3+0.91TM_4+0.68TM_7$	0.49	63.3	0.069
DAIS2.026	$0.21+0.53TM_2+0.6TM_3+0.33TM_4+0.4TM_7$	0.50	64.1	0.042
DAIS2.304	$0.12+0.15TM_2+0.74TM_3+0.5TM_4+0.5TM_7$	0.47	61.8	0.046

Table 4-8 Regression function

After establishing the mathematical relationship between the two images, the next step was is modeling an image using the above functions, i.e. the regression functions is applied on a pixel-by-pixel basis. This results an image that simulate spectral characteristics from the two images. This was

done in ERDAS 8.7 software using modeller function. The resulting image is a seven band image and we call it modeled DAIS image. The three images in  $A_1$  are shown below.

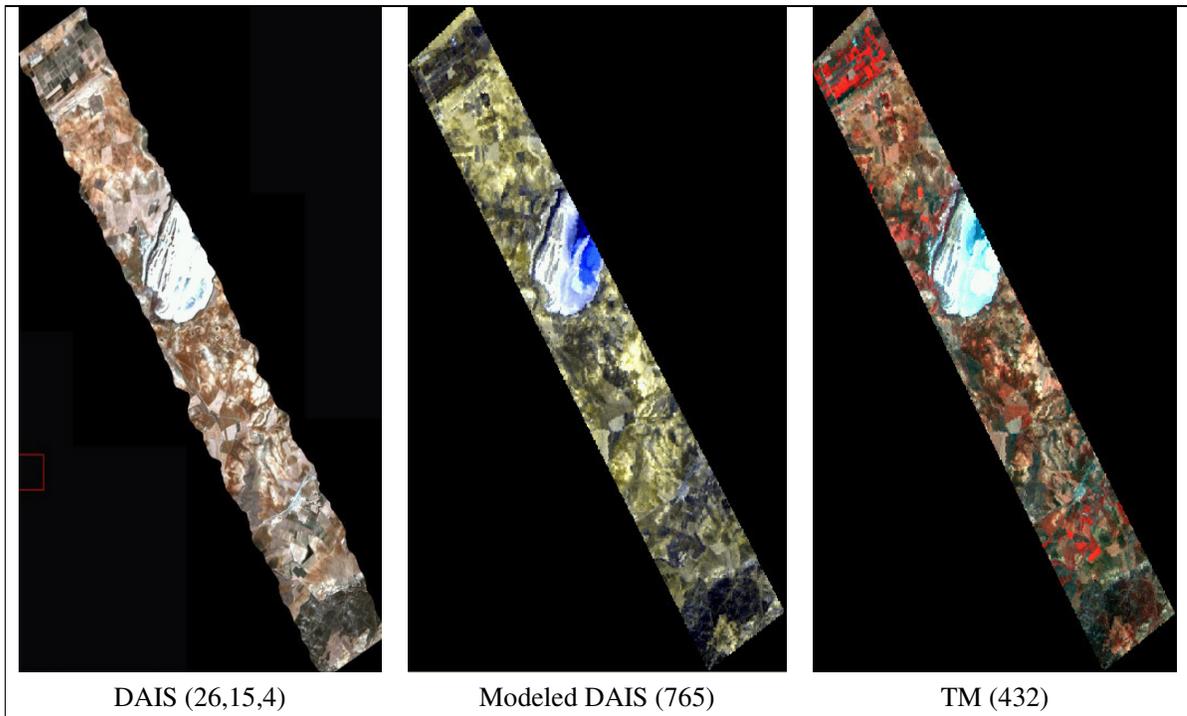


Figure 4-2 The three images in  $A_1$

Checking the quality of modeled DAIS was important since our interest is to get a better SSOM level discrimination as compared to TM. The ultimate test of the function's discriminatory power from the SDA function is the rate of correct classification from both classification accuracy (Tabachnick and Linda, 2001). Thus, we used SDA for quality check. The statistical quality comparison of the three images in  $A_1$  is given in section 4.6. Difference images (DAIS-Modeled DAIS) and (DAIS-TM) comparison is given in the next chapter in section 5.5.

#### 4.6. Comparison of the three images in $A_1$

We extracted reflectance values from modeled DAIS image at the same 300 points as what we did for DAIS and TM images before. Next we applied SDA. From the SDA of modeled DAIS classification result, we got 66.7% of the original strata groups are correctly classified. It is better than Landsat TM by more than 2%.

Strata	Predicted Group membership				Total
	1	2	3	4	
1	42.5	10.0	32.5	15.0	100.0
2	6.7	33.3	50.7	10.0	100.0
3	2.8	1.1	90.3	5.1	100.0
4	7.4	9.3	46.3	37	100.0

Table 4-9 SDA Classification Results of modeled DAIS image

As explained in chapter two and in the previous sections, Wilk's  $\lambda$  values are indicative of discriminating power at a given wavelength. We compared the Wilk's  $\lambda$  values of the three images. The result of the comparison is shown in figure 4-3. From the graph, it is possible to see that  $\lambda$  value of the modeled DAIS lies in between the DAIS and Landsat TM images.

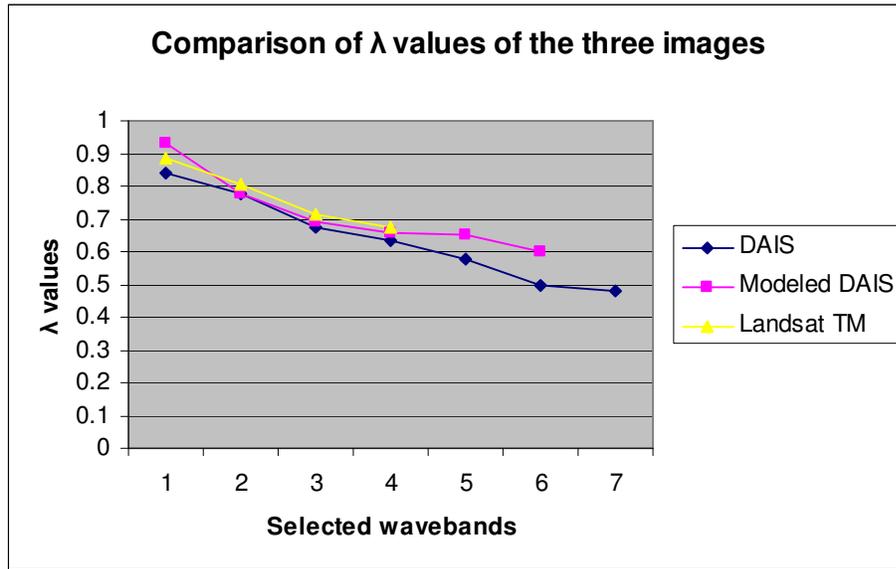


Figure 4-3 Comparison of Wilk's  $\lambda$  values of the three images in  $A_1$

In general, from the SDA classification accuracy results and Wilk's  $\lambda$  values, it is clear that the quality of modeled DAIS is better than TM in  $A_1$ . Therefore, we decided to extrapolate the developed relationship to A, in this case we used the big TM image that covers the whole area.

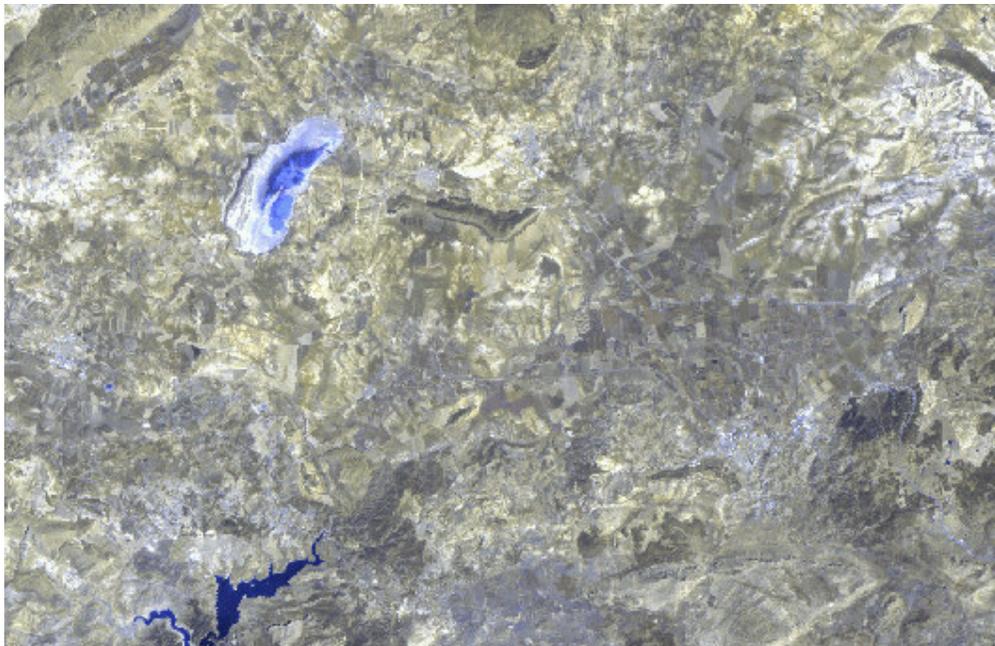


Figure 4-4 Modeled DAIS extrapolated to A (765)

#### **4.7. Summary and conclusion**

The main objective of this chapter was to determine optimum number of DAIS and TM bands that are useful for discriminating SSOM levels in  $A_1$  of the study area. We also applied other two approaches before SDA, which haven't been succeeded in this research. The first attempt was selecting three bands based on literatures. We extracted reflectance values of the three bands at the available eight ground truth points, however the result was unreliable. In our second attempt, we considered the full DAIS spectrum (0.4-2.5 $\mu$ m) and extracted reflectance values at the eight ground truth points for the whole wavelengths. The second attempt also did not improve the result. Unreliability of the results might be due to the limited number of ground truth points.

However, SDA is found to be useful in determining optimum number of bands. A mathematical relationship between seven DAIS and four TM bands has been created by means of multiple regressions. Seven band modeled DAIS image was created using the mathematical relationship. We used SDA to check the discriminating power of the three images. The accuracy of correctly classified groups and Wilk's  $\lambda$  values of modeled DAIS are improved as compared to Landsat TM.

## 5. Linear spectral unmixing analysis and results

### 5.1. Introduction

Image classification is more objective ways of interpreting satellite images. However, the problems associated with mixing up of pixels in images create inaccuracy problem during image classification. This happens most often in conventional classification algorithms. They consider a pixel as consisting of only one surface feature but in reality not. Therefore, acknowledging surface heterogeneity during image classification is important. This can be done by using spectral unmixing technique. It is a technique that has been developed to derive fractions of spectrally pure materials that contribute to observed spectral reflectance characteristics of a mixed pixel using endmember spectra.

### 5.2. Methods

A detailed procedure for the unmixing analysis is shown in figure 5-1.

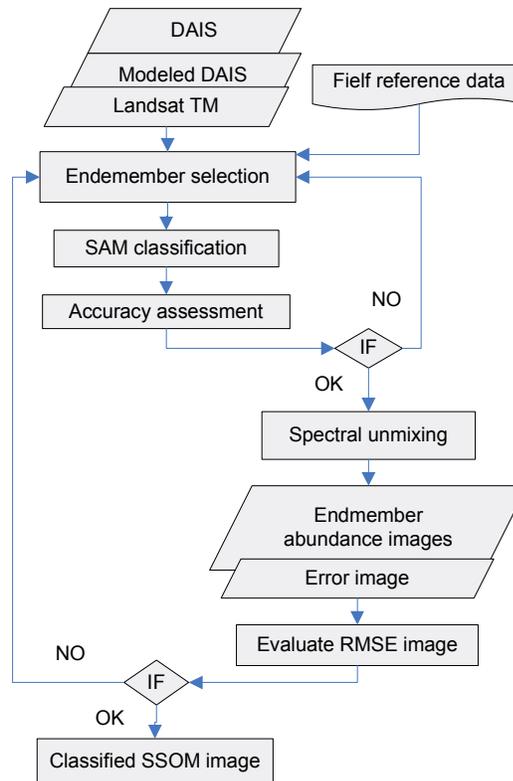


Figure 5-1 Spectral unmixing procedure

As explained in the previous chapter, we have created a modeled DAIS image by integrating DAIS and TM images by means of multiple linear regression models. We performed spectral unmixing classification first for  $A_1$  and then for  $A$ . The first step in implementing spectral unmixing is to select appropriate endmembers. Section 5.3 presents the endmember selection analysis. Next, we implemented the spectral unmixing classification. This is presented in section 5.4. From the abundance images derived by means of unmixing of endmembers, we inferred potential distribution of SSOM in the study area. This is presented in section 5.5. Section 5.6 gives summary and conclusion of the chapter.

### **5.3. Endmember selection**

In the absence of field spectral reflectance data, we used spectra that were extracted from the three images using available ground truth coordinates. Different techniques are explained in chapter 2 for selecting endmembers. These techniques were employed in this research as presented below.

#### **5.3.1. Spectral Angle Mapper (SAM)**

Spectral Angle Mapper (SAM) is a physically-based spectral classification that uses an angle in an  $n$ -dimensional space to match pixels to reference spectra (ENVI Tutorial, 2000). The SAM algorithm determines the spectral similarity between two spectra by calculating the angle between the spectra, treating them as vectors in a space with dimensionality equal to the number of bands. Small angles between the two spectrums indicate high similarity and vice versa. The reference spectra can either be taken from the laboratory, field measurements or extracted directly from the image. We used reference spectra extracted directly from the image in our analysis.

Field reference data were grouped into three soil classes: clayey soil, sandy soil and calcareous soil based on USDA soil classification criteria (Schoeneberger et al., 1998) as: if (clay  $\geq 60$  and sand  $\leq 40$  and silt  $\leq 40$ ) it is grouped as clayey soil and if (clay  $\leq 15$  and sand  $\geq 70$  and silt  $\leq 30$ ) it is grouped as sandy soil. From the available datasets if the laboratory measured calcareous content in the soil is above 60% and if it is none of the above classes, we classified it as calcareous soil. From previous work by Arnaldo, (2001), we selected olive vegetation and bush vegetation endmembers. We also considered water as a candidate endmembers, because the salty nature of the shallow lake in the study area may well contribute to surface reflectance (Kariuki, 2004). Candidate endmembers therefore includes: Clayey soil ( $S_1$ ), sandy soil ( $S_2$ ), calcareous soil ( $S_3$ ), olive vegetation ( $V_1$ ), bush vegetation ( $V_2$ ) and water ( $W$ ).

After assigning candidate endmembers, the next step was applying SAM classification using these six endmembers. We extracted spectra of the candidate endmembers from the three images. The spectra were input into the SAM classifier in ENVI 4.0 software. In SAM classification each pixel will be assigned to the class according to the lowest spectral values. The output of the SAM classifier is a classified image based on the endmembers. This image is one band with nominal values for all the endmembers. More information is presented in chapter 2. We applied SAM classification first for the three images in  $A_1$ . We used the same endmembers in the three images to have a fair comparison of the results.

### SAM classification Results of $A_1$

After generating SAM classified images (figure 5.2), a class statistics was assessed. These statistical values are useful to know the areal proportion of the candidate endmembers in  $A_1$ . The statistics is based upon classification results for any related input candidate endmembers. Endmembers with a low percentage will be excluded from the endmember list, since their contribution to the overall reflectance is expected to be low. The class statistics result is presented in table 5-1. The table show that  $V_1$ ,  $S_1$ ,  $S_2$ ,  $S_3$  and  $W$  are the main endmembers to consider where as,  $V_2$  appears to be a poor candidate. Its proportion is low ( $<10^{-2}$ ) in all the three images because it is concentrated only in some areas of  $A_1$  (Arnaldo, 2001). Therefore, it was discarded as an input endmember for the unmixing analysis.

Candidate Endmembers	SAM class statistics		
	DAIS	Modeled DAIS	Landsat TM
$V_1$	3.5%	2.50%	1.60%
$V_2$	$4.0 \times 10^{-3}\%$	$3.6 \times 10^{-3}\%$	$2.0 \times 10^{-3}\%$
$S_1$	3.0%	0.60%	0.35%
$S_2$	4.3%	2.91%	2.50%
$S_3$	10.0%	9.80%	9.40%
$W$	0.44%	0.41%	0.65%

Table 5-1 SAM class statistics of candidate endmembers  $A_1$

The class statistics shows relatively large numbers (proportion) of endmembers when using DAIS followed by modeled DAIS, where as lowest number occurs when using TM. The DAIS may resolve the spectral signatures of even small features because its narrow and continuous bands helps to decrease the spectral mixture problem (Girouard et al., 2001). The second large proportion occurs for modeled DAIS. This might be attributed to due to simulation of some spectral characteristics of the DAIS image.

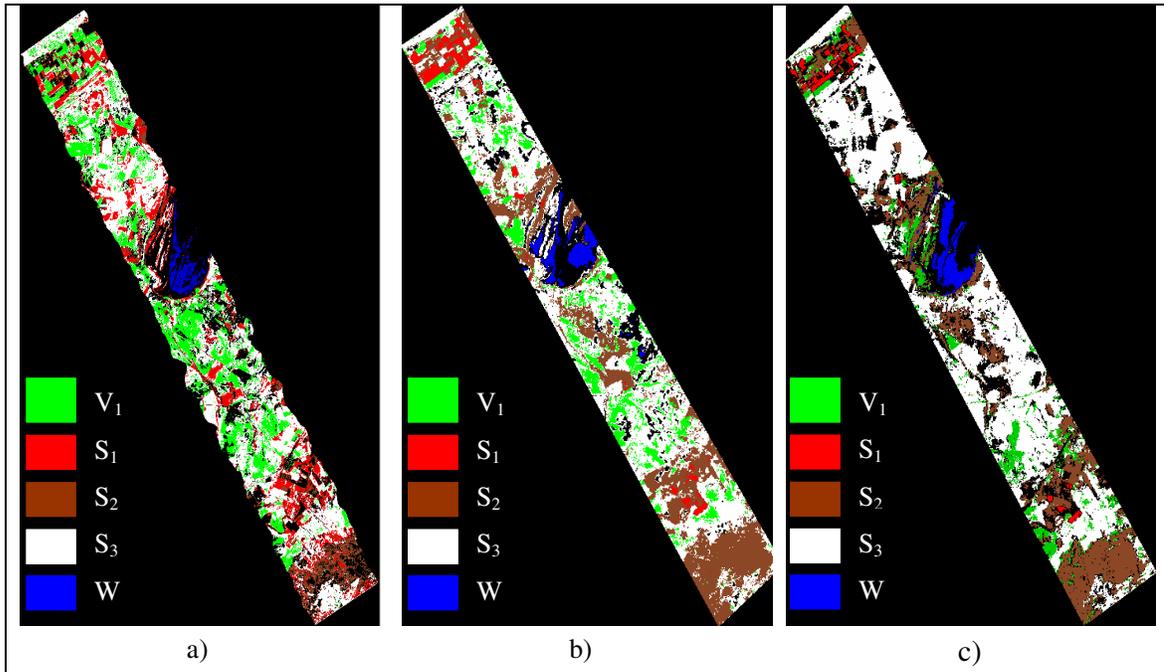


Figure 5-2 SAM classified images in  $A_1$  of the three images, a) DAIS b) modeled DAIS c) TM

#### Accuracy assessment of SAM classifier result

An integral part of image classification is validation of the results based on independent ground truth datasets. A common method for classification accuracy assessment is by means of an error matrix. It compares on a class-by-class bases, the relationship between known reference data and the corresponding results of the classification (Campbell, 1996). From this error matrix the producer's accuracy can be derived, which is the number of correctly classified pixels in each class divided by the number of pixels used for that class.

$$Producer's\ accuracy = \frac{\sum_{i=1}^k n_{ii}}{n_{+i}} \quad (18)$$

$n_{ii}$ : is the number of pixels classified correctly

$n_{+i}$ : is the number of pixels classified into class  $i$  in the ground truth data

$k$ : is the number of endmember class

User accuracy is computed by dividing the number of correctly classified pixels in each class by the total number of pixels that were classified as that class. This is indicative of the probability that a pixel classified on the map actually represents that category on the ground.

The overall accuracy is then the number of pixels correctly classified divided by the total number of pixels checked.

$$\text{Overall accuracy} = \frac{\sum_{i=1}^k n_{ii}}{N} \quad (19)$$

Where:

$n_{ii}$ : is the number of pixels classified correctly

$N$ : is the total number of pixels

$k$ : is the number of endmember classes

According to Campbell (1996), the overall accuracy strongly overestimates the accuracy, because it only incorporates the major diagonal. Therefore, a full picture of variations in accuracies of specific classes, each might be of importance to a particular application, can only be got by examining the full error matrix. Another accuracy measure derived from error matrix is kappa coefficient ( $k$ ), defined as an estimate of a measure of overall agreement between image data and the reference data (Campbell, 1996). According to Congalton, (1991),  $k$  expresses the proportionate reduction in error generated by a classification process compared with completely random classification. This is a widely used technique and it considers within class correlation as well as overall image correlation. Thus,  $k$  analysis is considered as a powerful addition used for analyzing a single error matrix and comparing the difference between different error matrices (Congalton, 1991). Kappa coefficient indirectly incorporates the off-diagonal elements as a product of the row and column marginal as:

$$K = \frac{N \sum_{i=1}^k n_{ii} - \sum_{i=1}^k n_{i+} n_{+i}}{N^2 - \sum_{i=1}^k n_{i+} n_{+i}} \quad (20)$$

Where:

$n_{ii}$ : is the number of pixels classified correctly

$n_{i+}$ : is the number of pixels classified into class  $i$

$n_{+i}$ : is the number of pixels classified into class  $i$  in the truth dataset

$N$ : is the total number of pixels

$k$ : is the number of endmember classes

Kappa is always less than or equal to 1. A value of 1 implies perfect agreement and values less than 1 implies less than perfect agreement. A value of 0 implies that agreement is no better than chance (Congalton, 1991). In rare situation, Kappa can be negative. This is a sign implying the agreement is less than would be expected just by chance. It is rare that we can also get perfect agreement. Kappa can be interpreted differently as to what is a good level of agreement. One possible way of interpreting kappa according to Brennan and Silman, (1992) is: poor = < 0.40, fair = 0.40 -0.59, good = 0.60 – 74, and excellent > 74

As explained in chapter two, there were very small ground truth data in  $A_1$ , and therefore we decided to include some points showing spectra similar to those of individual endmembers. Points on the image showing similar spectra to one of the individual endmembers can be considered to belong to that endmember (Koch, 2000). These points were selected randomly on the DAIS image. Confusion matrices were created in ENVI software, and the results are presented in tables 5-2.

Classified data	Reference data					
	V <sub>1</sub>	S <sub>1</sub>	S <sub>2</sub>	S <sub>3</sub>	W	N.C.P
V <sub>1</sub>	5	1	0	0	0	6
S <sub>1</sub>	1	4	1	0	0	6
S <sub>2</sub>	0	0	3	0	0	3
S <sub>3</sub>	0	0	0	4	0	4
W	0	0	0	1	2	3
N.G.P	6	5	4	5	2	22
Accuracy						
Producer%	83	80	75	80	100	
User%	83	67	100	100	67	
Overall accuracy: 82 <i>k</i> : 0.720						

a)

Classified data	Reference data					
	V <sub>1</sub>	S <sub>1</sub>	S <sub>2</sub>	S <sub>3</sub>	W	N.C.P
V <sub>1</sub>	4	1	0	0	0	5
S <sub>1</sub>	0	3	1	0	0	4
S <sub>2</sub>	1	1	3	0	0	5
S <sub>3</sub>	1	0	0	4	1	6
W	0	0	0	1	2	3
N.G.P	6	5	4	5	3	23
Accuracy						
Producer%	67	60	75	80	67	
User%	80	75	60	67	100	
Overall accuracy: 69.5 <i>k</i> : 0.618						

b)

Classified data	Reference data					
	V <sub>1</sub>	S <sub>1</sub>	S <sub>2</sub>	S <sub>3</sub>	W	N.C.P
V <sub>1</sub>	2	0	2	1	0	5
S <sub>1</sub>	0	2	0	1	0	3
S <sub>2</sub>	1	2	3	1	0	7
S <sub>3</sub>	3	1	1	3	1	9
W	0	0	0	0	2	2
N.G.P	6	5	6	6	3	27
Accuracy						
Producer%	34	40	50	50	67	
User%	40	67	43	34	100	
Overall accuracy: 45 <i>k</i> : 0.36						

c)

Table 5-2 SAM classification accuracy results in A<sub>1</sub>: a) DAIS, b) modeled DAIS, c) Landsat TM. (N.C.P=number of classified pixels, N.G.P=number of ground truth points)

Looking at the classified data in table 5-2 a), very few ground truth points are misclassified. The overall accuracy is 82%. The error matrix shows a high producer's accuracy for W showing that all the validating datasets (only two) are falling on the W class on the classified image. Spectral confusion between endmembers is not occurred i.e. the spectra of individual endmembers are differentiable from one another. The result shows acceptable overall and individual accuracies of endmembers. Where as, in table 5-2 b) the misclassified data are relatively increased in all the classes showing that more spectral confusion between endmember as compared to results in DAIS. This results a relatively decrease in both the overall accuracy and k coefficient as compared to DAIS results. Kappa coefficient gives a more objective quality measure since it takes into account the agreement contributed by chance and considers that the frequency of a sample appearing in a class is proportional to the percentage of number of pixels that the classes cover in the whole image. Therefore, the combination of these measures indicates the quality of the classification results in DAIS and modeled DAIS. According to Brennan and Silman, (1992) specification given above, k coefficients of these two images are within a range of good agreement. In table 5-2 c) the wrongly classified ground truth points are increased especially in  $V_1$  and  $S_1$  endmembers resulting low individual and overall accuracies. The spectra of  $V_1$  and  $S_1$  endmembers are not clearly differentiated from one another in Landsat TM image. Visual inspection of TM shows the two endmembers ( $V_1$  and  $S_1$ ) appeared almost in similar tones. This might be due to the low spatial resolution of the Landsat TM. The overall and k coefficient classification accuracy results are low. According to Brennan and Silman, (1992), the k coefficient is within a range of poor agreement.

Generally, even if it is difficult to draw a general conclusion about the accuracy of the results achieved from the limited number of ground truth data, the results obtained show better accuracy for DAIS and modeled DAIS than TM. This accuracy trend is good indicator of the potential of the three images to identify the five endmembers in  $A_1$ . Therefore, we selected the five endmembers for our unmixing because the classification accuracy of the individual endmembers in the DAIS and modelled DAIS indicate a reasonable accuracy.

### **5.3.2. Other endmember selection techniques**

In the above section, we presented endmember selection results obtained from SAM classification. However, other endmember selection techniques were implemented before we applied SAM classification. These include Pixel Purity Index (PPI) and Principal Component Analysis (PCA). We applied PPI to the Landsat TM image to identify pure pixels. We overlaid the ground truth data on the PPI image. There were no overlap between the PPI image and the ground truth points. More detail information about PPI is presented in chapter two. We also applied PCA transformation on the TM image. PCA transformation enables to extract pixels representing pure spectra from outer lines of the polygon that bounds the data space of the first two principal components (ENVI Tutorial, 2000). We overlaid the ground truth reference data on the image to see overlap between points marked by PCA and ground truth points. These points did not relate to the endmembers in the reference data. Detail explanation about PCA is also given in chapter two. Therefore, these two endmember selection techniques were not important in our analysis.

## **5.4. Spectral unmixing classification**

Two types of spectral unmixing are generally implemented: constrained and unconstrained unmixing. Constraining assumptions that can be implemented are that: the endmember abundances are non-

negative, and the abundance sum to unity. From a statistical point of view, however, it is meaningless to force mixture models to constraining to the data (Van der Meer and de Jong, 2000). As a result, in most literature unconstrained unmixing is preferred over constrained unmixing. Brief explanation for both unconstrained and constrained unmixing is presented in chapter two.

**5.4.1. Spectral collection**

The spectra for the five endmembers were collected from the three images by using the geographic X, and Y coordinates. Spectra of the endmembers extracted from DAIS image are shown in figure 5.3.

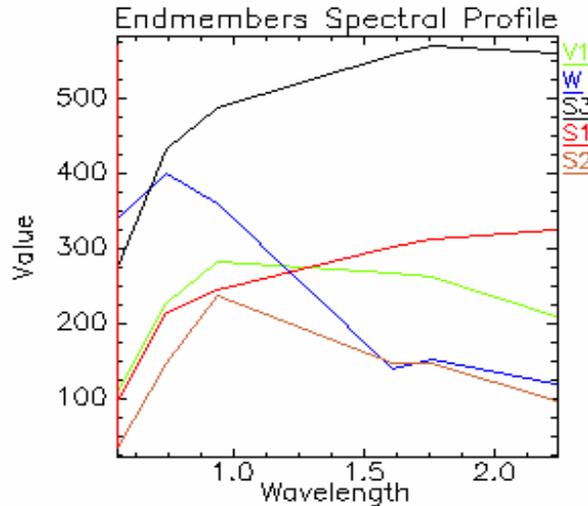


Figure 5-3 Endmembers spectral curve extracted from DAIS used in image decomposition

The figure above shows that the spectra represent distinct surface cover features. This can be distinguished according to their spectral reflectance property within the given range of the spectrum.

**5.4.2. Spectral unmixing results**

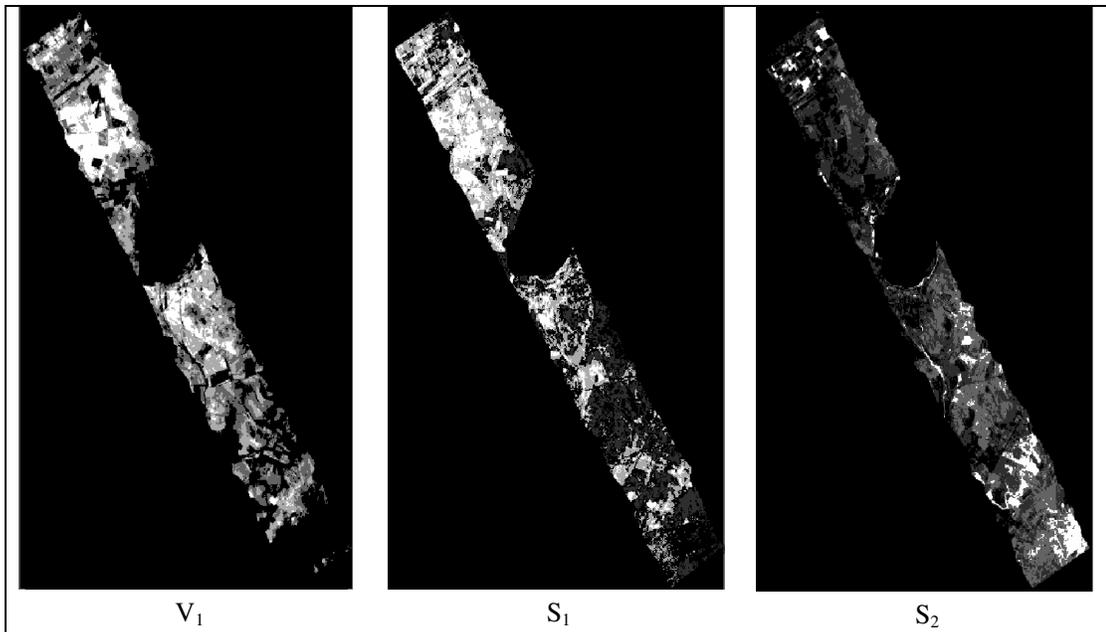
In this section, we presented the results of linear spectral unmixing classification first for  $A_1$  and then for A. Region of interests (ROIs) are portions on the images selected by entering the ground truth X and Y coordinates of the endmembers. They are used to extract statistics for classification and other operations. We created ROIs for the endmembers and entered into the unmixing classification function. Both unconstrained and constrained unmixing were compared. Finally we selected one of them after examining the RMSE value and endmember abundance images. As shown in table 5-1, the SAM class statistics for water is less than one. Thus, we first applied unmixing by excluding water from the endmember list. The RMSE however has over flown above one; meaning that the model could not be well constructed if water is excluded. This occurs because surface reflectance contributed from the salty shallow lake might be important. Thus, we included water in our unmixing analysis.

The output of linear spectral unmixing includes RMSE and endmember abundance images. The validity of the unmixing model can be analyzed by calculating RMSE between modelled and measured reflectance of pixels (Mather, 1999) as explained in chapter 2. The endmember abundance images derived from the unmixing analysis should have a value between 0 and 1. But if the response

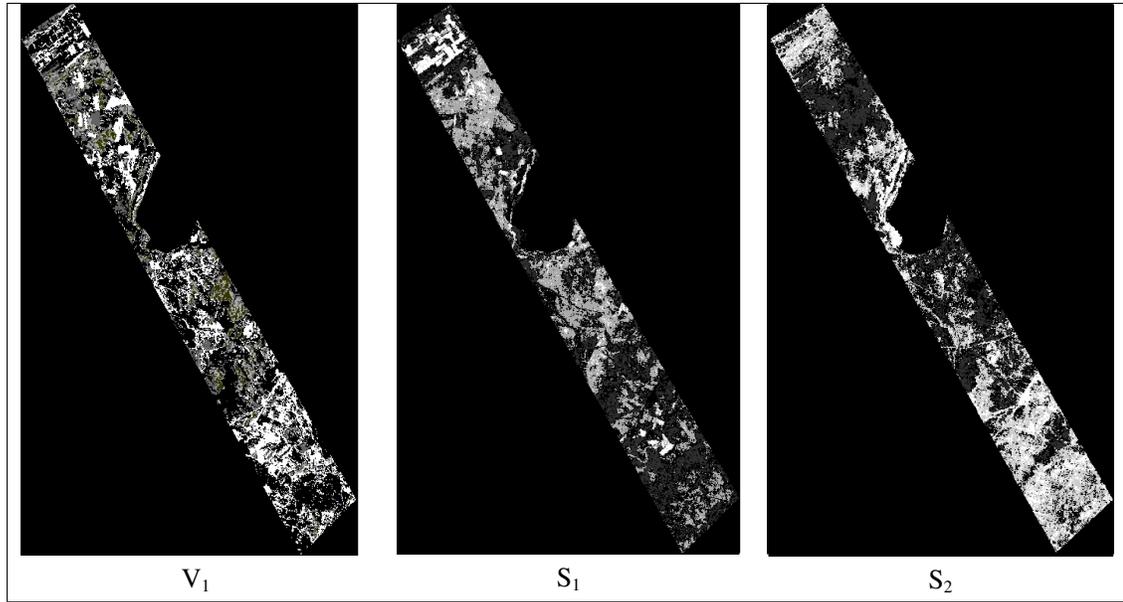
of a pixel is purer than the selected endmembers, abundances above 1 and below 0 may occur (Darvishsefat, 2002). This also happened in our unmixing analysis, meaning that there might be more pure pixels on the image than the selected endmembers.

Different combinations of image bands were checked and the one presented below is the best result obtained. For example, unmixing applied on the spectral subset (bands 2347) of the TM images instead of using all six bands showed a decrease in the RMSE value. In case of DAIS, we first considered all bands between 0.4-2.5  $\mu\text{m}$ . But the RMSE value was over flown above one, where as when using optimal 7 bands selected by means of SDA, the average RMSE value decreased. This leads to the conclusion that in both cases there was redundant information in the data and inclusion does not provide an extra benefit in the analysis. The modeled DAIS using (bands 123457) result low RMSE. This shows that there was no redundant information in these bands, probably due to the fact that it was created from the less-redundant bands of DAIS and TM after SDA selection as explained in chapter four.

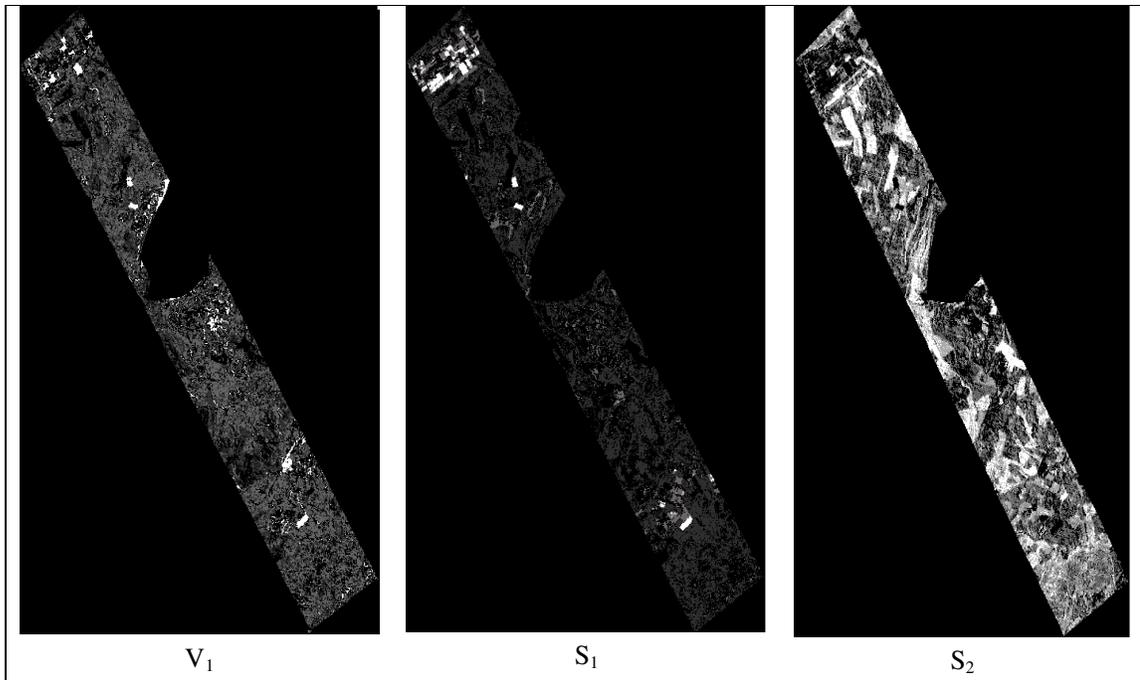
An arbitrarily selected endmember abundance images in  $A_1$  are shown in figure 5-4. According to Arnaldo, (2001) and the soil profile description manual, olive vegetation ( $V_1$ ) covers wide area with mainly bare soil ( $S_1$ ) in between. From the DAIS endmember abundance images below, it is possible to see that: high abundance of  $V_1$  and  $S_1$  endmembers cover a large area. Therefore, the abundance images derived from unmixing of DAIS image is in agreement with the description given in the soil profile manual and Arnaldo, (2001) report. In the modeled DAIS abundance images high abundance of  $V_1$  and  $S_1$  endmembers are identified in a relatively small area as compared to DAIS. Where as, in TM high abundances of  $V_1$  and  $S_1$  are identified in a very small area. This seems unreliable because according to Arnaldo, (2001) report and the available soil profile description manual, these endmembers cover relatively a large area.



a)



b)



c)

Figure 5-4 Endmember abundance images derived from unmixing of the three images in  $A_1$ : a) DAIS b) modeled DAIS c) Landsat TM. (Bright pixels on these images represent higher abundances and vice versa)

Statistics of both unconstrained and constrained unmixing results (abundances, RMSE, and standard deviation values) derived from the three images in  $A_1$  are presented in the tables below.

EM	Unconstrained						Constrained					
	Min	<0%	>1%	Max	Mean	Stdv	Min	<0%	>1%	Max	Mean	Stdv
V <sub>1</sub>	-10.0	7.0	2.0	4.8	-0.10	0.33	-10.0	7.0	2.0	4.8	0.10	0.35
S <sub>1</sub>	-2.51	24.0	4.0	3.46	0.14	0.14	-2.51	23.5	3.5	3.5	0.14	0.15
S <sub>2</sub>	-2.92	5.0	1.8	5.76	0.08	0.35	-2.94	5.0	1.8	5.0	0.08	0.36
S <sub>3</sub>	-4.34	18.0	2.5	6.55	0.03	0.34	-4.31	18.4	2.4	6.5	-0.03	0.36
W	-0.43	10.0	1.5	2.13	0.14	0.30	-0.41	10.0	1.6	2.1	0.130	0.31
RMSE	0.00	NA	NA	0.07	0.02	0.04	0.00	NA	1.2	1.5	0.18	0.12

a)

EM	Unconstrained						Constrained					
	Min	<0%	>1%	Max	Mean	Stdv	Min	<0%	>1%	Max	Mean	Stdv
V <sub>1</sub>	-3.8	23.0	5.0	7.5	-0.17	0.38	-4.0	23.0	5.0	7.5	0.17	0.38
S <sub>1</sub>	-1.9	36.8	1.5	1.6	0.02	0.23	-2.1	36.8	1.5	1.6	0.02	0.23
S <sub>2</sub>	-6.2	32.0	3.0	7.2	0.03	0.17	-6.1	34.0	3.0	7.2	-0.03	0.17
S <sub>3</sub>	-4.6	45.6	3.3	2.5	0.06	0.7	-4.8	45.6	3.3	2.5	0.06	0.77
W	-5.2	62.0	0.5	1.1	0.14	0.65	-0.2	62.0	0.5	1.1	0.14	0.65
RMSE	0.0	NA	2.3	4.0	0.10	0.41	0.0	NA	2.0	3.9	0.39	0.41

b)

EM	Unconstrained						Constrained					
	Min	<0%	>1%	Max	Mean	Stdv	Min	<0%	>1%	Max	Mean	Stdv
V <sub>1</sub>	-0.2	76.8	1.0	1.7	-0.17	0.4	-0.2	77.0	1.0	3.7	-0.16	0.4
S <sub>1</sub>	-2.5	48.0	4.5	2.4	-0.26	0.3	-1.1	48.0	4.5	2.5	-0.26	0.3
S <sub>2</sub>	-0.9	46.3	10.0	4.8	0.40	0.5	-0.9	10.0	10.0	4.8	0.42	0.6
S <sub>3</sub>	-1.4	29.0	4.3	2.7	0.24	0.5	-1.4	29.0	4.0	1.4	0.24	0.5
W	-7.0	63.0	2.5	3.9	-0.11	0.7	-7.0	54.0	2.5	3.8	-0.11	0.6
RMSE	0.0	NA	13.0	5.2	0.56	1.7	0.0	NA	13.0	5.7	0.72	0.9

c)

Table 5-3 Statistics of fractional abundance and RMSE images obtained through unmixing using image spectra of the three images in A<sub>1</sub>: a) DAIS b) modeled DAIS c) TM. (EM = endmembers, Stdv = standard deviation, Max = maximum, Min = minimum, NA = not available)

The overall mean abundances of endmembers in both unconstrained and constrained unmixing are similar as shown in the tables above. When comparing results obtained from the same images (e.g. DAIS unconstrained to DAIS constrained), the differences occur in the RMSE value, which is crucial for evaluating the fitness of the unmixing model. Generally, values of RMSE greatly improve when using unconstrained unmixing. Other observations include:

- mean abundance of the endmembers derived from both DAIS and modeled DAIS are all positive except for one endmember for V<sub>1</sub>, (-0.1 and -0.17 respectively)

- in Landsat TM  $V_1$ ,  $S_1$  and  $W$  endmembers mean abundance values are negative (-0.17, -0.26 and -0.11 respectively)
- the standard deviation values of both the endmembers and RMSE of DAIS and modeled DAIS are much smaller than Landsat TM

In both DAIS and modeled DAIS the performed linear spectral unmixing shows reliable abundance images because most of the endmember mean abundances are positive, and the value of the RMSE image is low. For DAIS, we do not have RMSE values above one (see figure 5-5a). Only a little portion of the modeled DAIS RMSE image has values greater than 1 (see figure 5-5b and appendix 8). Where as, in Landsat TM error image a large number of pixels have RMSE value greater than 1. In the TM error image in figure 5-5, RMSE values greater than 1 is about 35.37% of  $A_1$ . This shows a significant improvement when comparing TM and modeled DAIS, possibly being attributed to a better spectral separability between endmembers when using DAIS and modeled DAIS than Landsat TM.

Generally, all the above observations are indicative of the validation and usefulness of the unconstrained unmixing model in DAIS and modeled DAIS as compared to TM. Therefore, we selected unconstrained unmixing for our further analysis.

Error images derived from the unmixing analysis helps to see the distribution of errors generated by the three images. Error images of the three images are shown below.

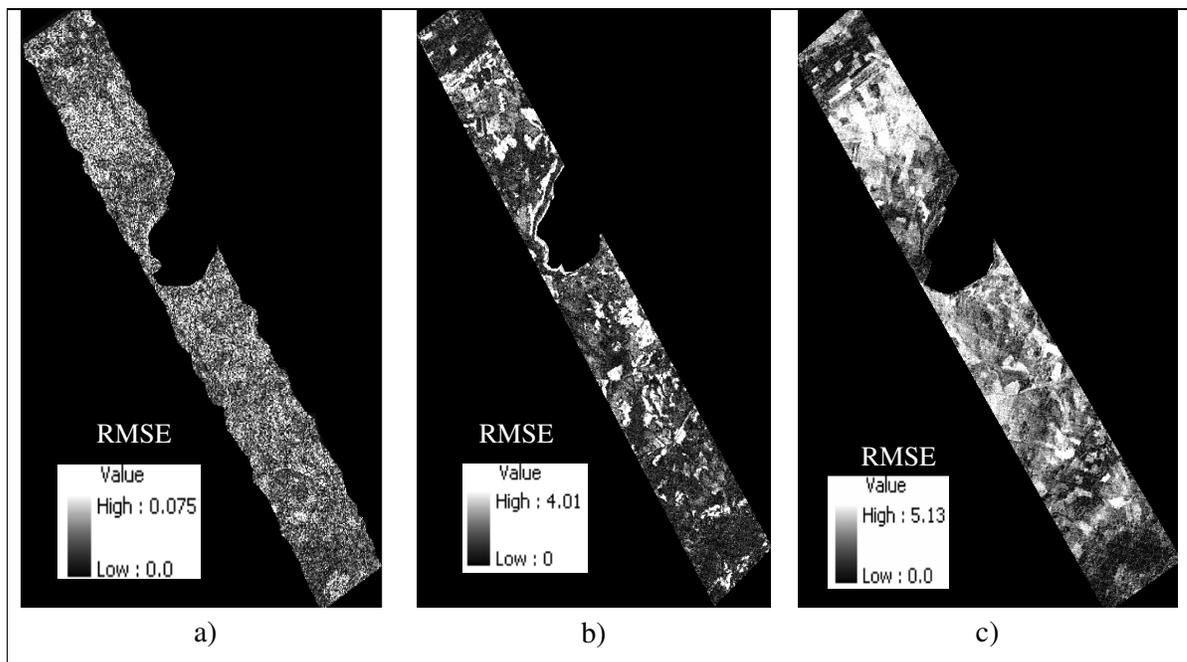


Figure 5-5 Error images derived from spectral unmixing in A1: a) DAIS b) modeled DAIS c) TM

## 5.5. Surface Soil Organic Matter (SSOM) mapping

Land cover mapping is one of the easiest tasks of satellite data interpretation (Schmidt and Schoettker, 2004). Commonly used supervised classification methods such as Maximum Likelihood classifier are not appropriate for a detailed differentiation of certain soil variables classes such as SSOM (Schmidt

and Schoettker, 2004). On the other hand unmixing derived endmember abundance images help to know the proportion of endmembers at a pixel level and thus classifying them may help to solve some of the limitations of conventional classifications if the abundance sum in a pixel. For example, Van der Meer, (1995) produced a classified image from the fractional abundance images by assigning a pixel to the class with the highest abundance of the endmember. He used abundance values contributed greater than 50 percent to classify different minerals and surface features. We also derived a classified potential SSOM images from endmembers abundance derived from spectral unmixing as presented below.

As explained in section 1.1.2, direct mapping of SSOM from RS images is almost impossible because of its varying nature and its quick response to external factors. To overcome this problem we therefore, use inferred variables to identify potential areas of SSOM in the study area. Under normal conditions, high SSOM is concentrated in vegetation and clay soil dominant areas (Cambridge University, 2002). Therefore,  $V_1$  and  $S_1$  endmember abundance images were composed to derive classified SSOM level images. Pixels that present less 0 or above 100 percent abundance sum for the  $V_1$  and  $S_1$  were not classified

Abundance sum of $V_1$ and $S_1$	SSOM levels
0.75 - 1	High
0.50 - 0.75	Relatively high
0.25 - 0.50	Low
0 - 0.25	Very low
Otherwise	Unclassified

Table 5-4 Classification functions used in the raster calculator

We assigned a quantitative value for the above qualitative measures by overlaying the available ground truth SSOM measured values. The classified SSOM maps are shown in the figure below.

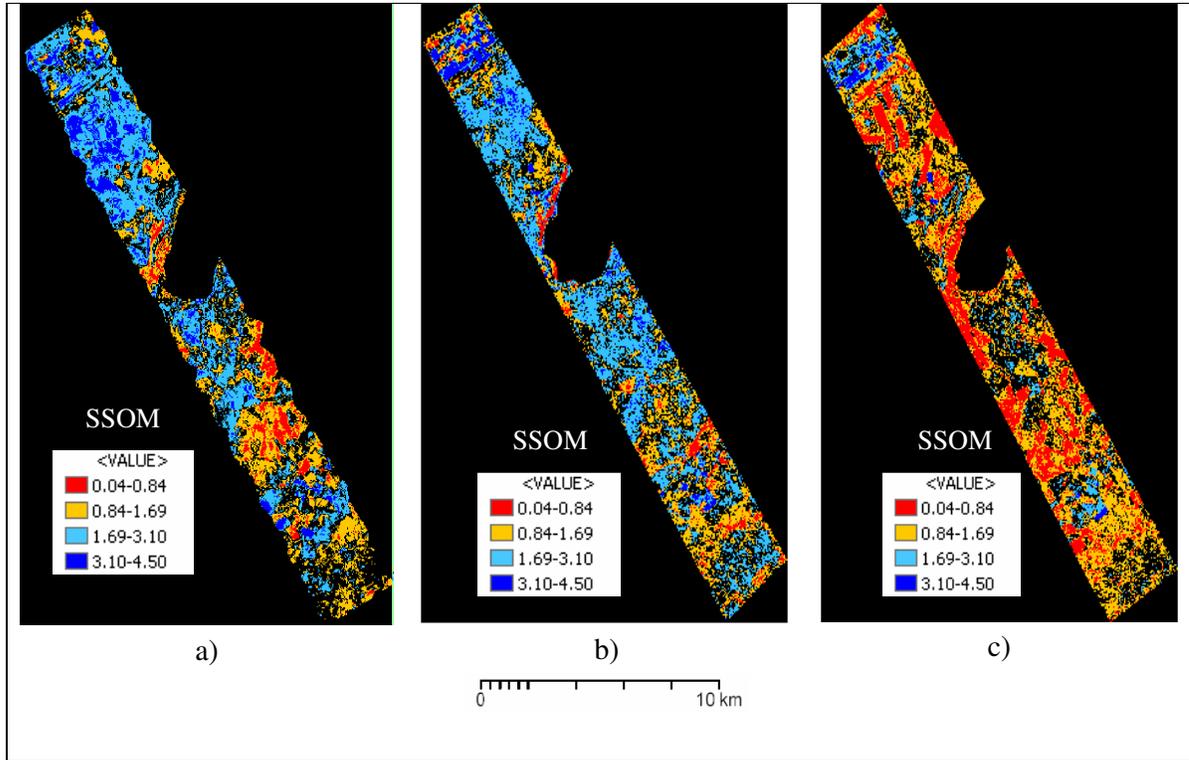


Figure 5-6 SSOM potential areas in A1: a) DAIS b) modeled DAIS c) Landsat TM

We used soil profile description database tables, topographic and geological maps of the study area as ancillary data sources to get an insight into the geology and some geomorphologic characteristics of the study area. These help to interpret the classified SSOM results. We compared the DAIS classified results with the ancillary data sources to see whether the classified results correlate with the geology and geomorphologic characteristics of the study area or not.

The high SSOM areas (dark blue) identified by the classified DAIS are topographically relatively flat, as checked from the topographic maps. This was done by overlaying the classified DAIS on the topographic map of the area. These areas are annually planted with different crops especially olive vegetation according to Arnaldo, (2001) and the soil profile manual. The soils vary from dark clay to silty clay of relatively thick depth according to the soil profile description data. Therefore, the combination of landscape, vegetation condition and soils relationships as discussed in section 1.1.3 are considered to provide favourable conditions for the development of high SSOM in these areas. These areas are clearly identified in DAIS image.

Red and yellow areas on the DAIS SSOM classified image are areas of relatively expected low SSOM. As inferred from the geological and topographical map of the study area, these areas are topographically steep. Sandy soils intercalated with strong and coarse sub-angular blocky, hard, firm and slightly plastic and sticky nature are dominant in these areas according to the soil profile description. According to Grunwald, (1999) and our discussion in section 1.1.3, less SSOM development is expected in such areas. Reports from Utah state university, (1999), showed that salinity in soils inhibits plant growth and fertile soil development. As seen from the geological map,

left-upper part of the lake is gentle as compared to other sides of the lake. When the lake level rises during time of rainfall, the salty lake water is expected to flow this direction because of its gentleness. That might be a possible reason for the low SSOM concentration near the left upper part of the lake in the classified DAIS image. Generally, classification results obtained from DAIS seems promising according to the interpretations made based the ancillary data we have been referring. Therefore, we compared SSOM classified results of the other two images with it, to see whether they are reliable or not.

The classified result of the modeled DAIS image seems to correlate with the classified DAIS except in some high SSOM areas. According to Joachim and Brigitta (2000), high SSOM are found in a spatially less extent areas below trees and clay soils in small accumulation especially in most semi-arid area. Therefore, their spatial extent might not be easily resolved by the modeled DAIS even if we simulate some spectral characteristics of the DAIS by means of the regression models. Where as, the Landsat TM SSOM classified result shows very low reliability as compared with DAIS classified result. Figure 5-7 shows difference images derived from DAIS minus modeled DAIS and DAIS minus TM. From these images, it is evident that SSOM derived from DAIS and modeled DAIS is more correlated than between DAIS and TM.

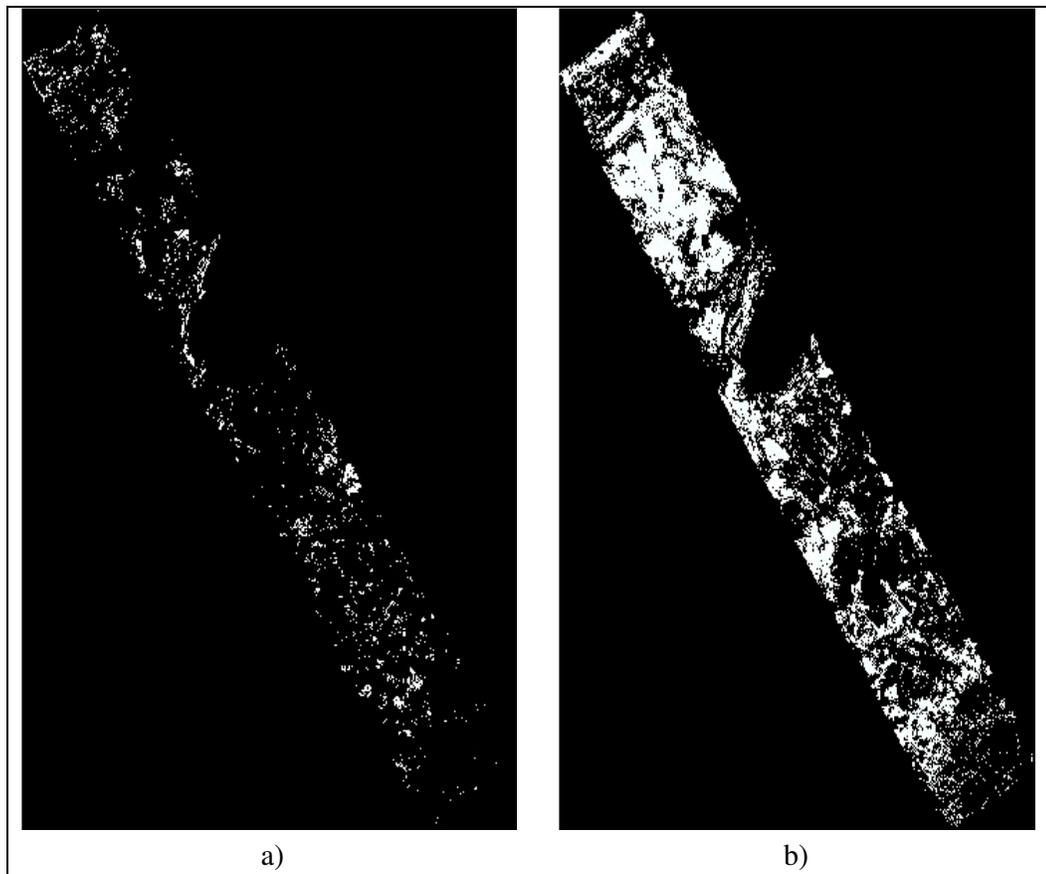


Figure 5-7 SSOM difference images (white color shows non-correlated areas): a) DAIS-modeled DAIS, b) DAIS-Landsat TM

In general, we implemented spectral unmixing in  $A_1$ . From the spectral unmixing derived abundance images, we produced classified SSOM images showing potential areas of SSOM in  $A_1$ . We found a

reliable SSOM classified result from DAIS. The SSOM classified modeled DAIS shows similarity with classified DAIS. Where as the reliability of results derived from TM are under question. Therefore, we decided to use only the modeled DAIS to map SSOM of A. It is explained below

**SAM classification of the large area (A)**

We also applied SAM classification for A using the same endmembers as introduced before. The class statistics showing endmembers areal abundances in the classified SAM image are shown in appendix 9. The SAM classified image of A is shown in the figure 5-8

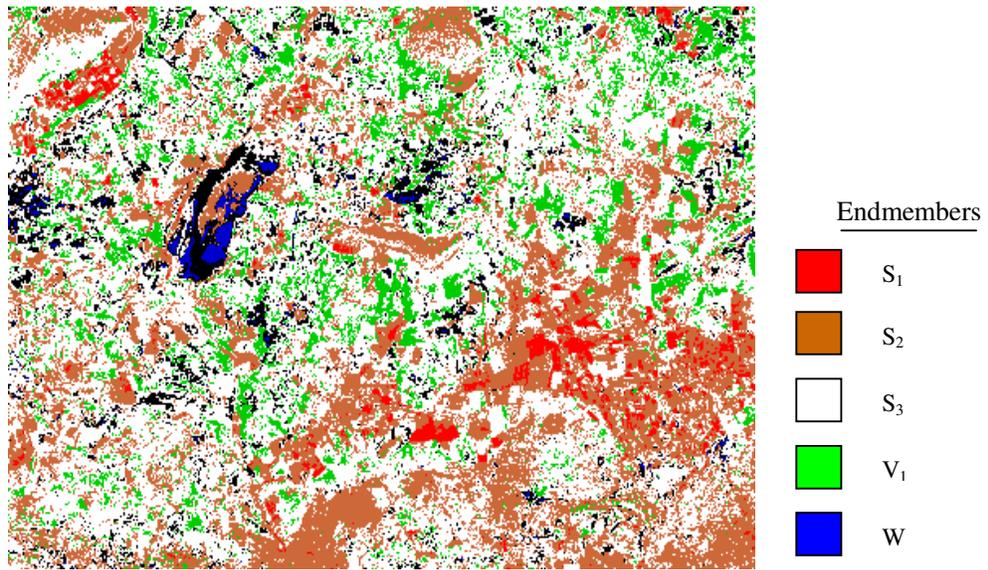


Figure 5-8 SAM classified images of A

We used the available ground truth data for accuracy assessment by excluding those used for endmember selection. The classification accuracy of the SAM result shows an overall accuracy of 66% and kappa coefficient 0.593. It is summarized in the confusion matrix table below.

classified data	Reference data					
	V <sub>1</sub>	S <sub>1</sub>	S <sub>2</sub>	S <sub>3</sub>	W	N.C.P
V <sub>1</sub>	11	1	0	0	0	12
S <sub>1</sub>	1	7	1	0	0	9
S <sub>2</sub>	2	1	12	0	0	15
S <sub>3</sub>	3	1	7	15	2	28
W	1	0	0	4	5	10
N.G.P	16	10	20	20	7	75
Accuracy						
Producer%	69	70	60	75	71	
User%	92	78	80	54	50	
Overall accuracy: 66      k: 0.586						

Table 5-5 SAM Classification accuracy result for A

From SAM classification result in table 5-5 both the individual and overall accuracies of the endmembers show a reasonable accuracy. Hence these endmembers can also be used for unmixing analysis of A. We performed unconstrained unmixing. The statistical results are given in table 5-6.

EM	Min	<0%	>1%	Max	Mean	Stdv
V <sub>1</sub>	-0.17	33	11	1.104	-0.16	0.22
S <sub>1</sub>	-5.37	24.6	7.5	2.788	0.317	0.71
S <sub>2</sub>	-3.56	52	3.0	1.967	0.052	0.30
S <sub>3</sub>	-4.40	42.6	2.3	9.448	0.536	1.36
W	-6.65	62	0.5	12.57	0.129	1.19
RMSE	0.00	NA	2.3	5.036	0.368	0.45

Table 5-6 Statistics of endmember abundance and RMSE values obtained through unmixing using image spectra of modeled DAIS image for A

The tables above shows, SAM classification accuracy is decreasing, while unmixing statistics (standard deviation and RMSE values) are getting larger than the results found in A<sub>1</sub>. In large semi-arid areas with complex landscape, soil reflectances are a function of landscape, moisture, land use management, and topographic characteristics (Ben-Dor et al., 2002). Getting accurate unmixing result for such large area may not be straight forward (Dengsheng and Qihao, 2004). The results found in this study seems to be in line with Dengsheng and Qihao, (2004), who found different unmixing results when applying unmixing techniques in a small and large area.

**SSOM mapping for the whole study area (A)**

We used similar approach to map SSOM in A as well i.e. by composing V<sub>1</sub> and S<sub>1</sub> endmember abundance images. The SSOM classified result is shown below.

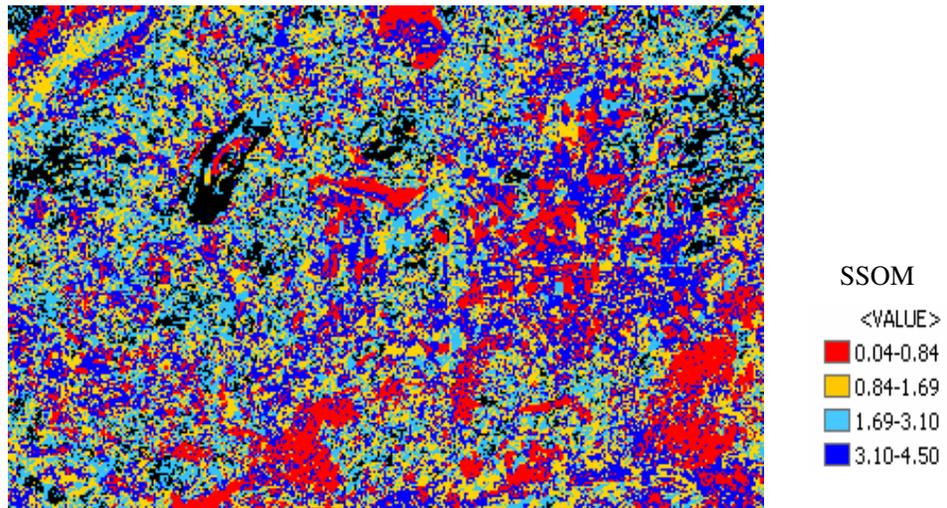


Figure 5-9 SSOM potential areas of A using modeled DAIS

In the above classification result large areas of high SSOM (dark blue color) are identified. According to the soil database table and (Arnaldo, 2001) these areas are covered by  $V_1$  and  $S_1$  endmembers. The soils in these areas are dark brown clay soils. Very low SSOM (red) identified on the classified image are hills and rocky exposures. These areas are well identified in the above classified image.

## **5.6. Summary and conclusion**

In this chapter, we implemented spectral unmixing analysis. Five endmembers for the unmixing analysis was selected using SAM classification. Endmember abundance and RMSE images derived from the unmixing analysis were used to evaluate the fitness of the unmixing model for the three images in  $A_1$ . Using vegetation ( $V_1$ ) and clay soil ( $S_1$ ) abundance images, we identified potential areas of SSOM image in the study area. Ancillary data sources have been used to interpret the results. Classified result of DAIS in  $A_1$  seems promising. The classified result of modeled DAIS correlates the classified DAIS except in some areas, where as the classified TM image does not show close correlation with the DAIS image as shown by the difference images in figure 5-7. Therefore, we used the modeled DAIS image to map potential areas of SSOM in A.

## 6. Discussion

Based on this research, this chapter presents a discussion of the results. The discussion mainly focuses on: the selected optimum number of bands by means of SDA, the established relationships between the selected DAIS and TM bands, the spectral unmixing analysis results and SSOM mapping from unmixing derived endmembers.

### 6.1. About SDA results

Chapter four aimed at discriminating strata based on SSOM levels available from DAIS and Landsat TM images. The motivation was to investigate whether we could discriminate SSOM levels in the study area using DAIS and TM reflectance spectra by reducing the number of bands to an optimum number. To achieve this goal, we used Stepwise Discriminant analysis (SDA). We tested the utility of spectral reflectance values extracted from DAIS and TM. SDA reduced dimensionality of the images into three discriminant functions. It also described and explored the relative importance of individual wavelengths in explaining SSOM differences.

The study has shown that, discriminant functions derived from DAIS and Landsat TM images can discriminate SSOM levels. However, the degree of discrimination is higher in DAIS as compared to TM. This can be seen from the Wilk's  $\lambda$  values (section 4.4.1). Low Wilk's  $\lambda$  values are an indication of a good discrimination and vice versa. Therefore, the Wilk's  $\lambda$  values for DAIS is lower than that for TM, a DAIS image gives a better discrimination of SSOM than a Landsat image. This may be attributed due to the fact that the narrow and continuous nature of hyperspectral bands helps to discriminate even small SSOM differences between strata that are otherwise unattainable when using multispectral images like Landsat TM. This is in agreement with the work of Prasad et al.,(2004), who reported that the discrimination power of hyperspectral data is much better than multispectral images for discriminating different surface features.

The discriminant analysis of the DAIS image showed that the VNIR bands play a significant role in discriminating SSOM differences in the strata (table 4-2). Experiments with VNIR have shown that, this portion of the electromagnetic spectrum is insensitive to atmospheric and background effects (Demetriades-Shah et al., 1990). This might be the reason for their usefulness for discriminating SSOM levels. Also, the discriminating effect of SWIR bands (2.01, and 2.026) is shown clearly (table 4-2). This study has shown, therefore, the successful use of the VNIR and some SWIR portions of the electromagnetic spectrum in accomplishing this task. These results are comparable to the work of Daniel et al.,(2002), who reported that the VNIR can discriminate between different SSOM concentrations in the soil and (Dalal and Henery, 1986), who found sensitivity of some SWIR bands for soil organic matter detection. The contribution in discriminating the strata is high in bands 0.630 and 0.76  $\mu\text{m}$  of Landsat TM. A report by Clevers, (1999) has shown that these regions of the spectrum are sensitive for detecting chlorophyll and nitrogen content in plants. Ben-Dor et al., (2002) reported

that SSOM discrimination can be expected in chlorophyll absorption regions as a result of a strong relationship between chlorophyll and SSOM concentrations. Therefore, this might be the reason for their usefulness in our analysis.

The discriminant classification results of the three images showed that the classification accuracy reported for stratum 4 is much lower than that for the other strata. The soil and landscape conditions in stratum 4 considerably vary as is explained in chapter three. The central part of stratum 4 is flat and it is dominated by sedimentary alluvial deposits. The western part is characterized by terraces, hillocks and slopes, being dominated by conglomerates. The areas near the transition between it and lower stratum 1 are more similar to stratum 1. Therefore, the combination of these differences in landscape and soil conditions as well as the partial similarity with the other strata might make it difficult for the discriminant analysis to discriminate stratum 4 or distinguish it from the other strata.

## **6.2. Establishing relationships between the DAIS and TM bands**

The developed relationship between the selected DAIS and TM bands in table 4-12, shows  $R^2$  Values (0.5). The quality comparison of the three images by the Wilk's  $\lambda$  values gives a better discrimination power of the modeled DAIS as compared to TM. These results show that integrating RS data sources of a different spectral resolution helps to create an image that simulates some spectral characteristics of the two images through statistical methods. This agrees with Daniel et al., (2000), who found better SSOM prediction results after synthesizing IRS images with the spectroscopic bands.

It should be emphasize that neither the discriminant functions nor regression equations for establishing the mathematical relationships between the selected bands of the two images are universal. For every campaign new field samples need to be collected and new SDA and regression equations should be constructed. We developed both the SDA functions and the regression equations from the extracted reflectance values using geographic locations in the study area. According to Agbu and Coleman, (1990), the relationships between soil constituents and reflectance values are geographically dependent and may not be easily extrapolated to other areas.

## **6.3. Unmixing classification**

In spectral mixture analysis, endmember selection is a crucial step. This often involves an iterative process of selecting initial endmembers, refining these endmembers after evaluating the class statistics of SAM classification, evaluating fraction images based on the RMSE and endmember abundance images, and then further refining endmembers. Finally, selected endmembers should be independent from each other. Spectra (figure 5.2) show that the endmembers represent distinctive ground cover features, which can be distinguished according to their spectral reflectance properties in the 0.4-2.5  $\mu\text{m}$  electromagnetic region. The spectral unmixing methodology outlined in figure 5-1 identifies these five endmembers for the unmixing classification after several iterating processes.

The unmixing model can be evaluated using three criteria according to (Roberts et al., 1998): fraction values, standard deviation values and RMSE statistics of the endmembers. The unmixing statistics (table 5-3) show that, the range of data variability as expressed by the standard deviation and RMSE values are low for DAIS and modeled DAIS. This indicates a better spectral separability of the endmembers when using DAIS and modeled DAIS, and thus validating the use of this method for

these images. The unmixing derived standard deviation and RMSE values of the modeled DAIS are becoming large when applied for A as compared to values of A<sub>1</sub> (table 5-6) because spectral unmixing is highly dependent on the illumination condition of the endmembers due to topographical and landscape conditions of the study area (Shrestha et al., 2002). Larger areas may contain differences in topography and landscape features that can affect the illumination condition. Stratification of the large areas into smaller areas of similar topography and landscape may help to overcome such problem.

The negative mean abundance values of V<sub>1</sub> in the unmixing results of the three images (table 5-3) can be explained by high spatial and temporal variability of this endmember. Differences in the sowing dates of the olive plant in the area can cause a spectral variability in the scene; olive vegetation of different age may have different spectral response. Similarly distribution of olive may not be uniformly spread in the scene; it may be dense in a field and sparse in another which will also be received by the sensor as highly variable signals.

#### **6.4. Potential SSOM area identification**

Two endmember fraction images derived from spectral unmixing analysis have been used in identifying potential areas of SSOM in the study area. The results derived from DAIS show the spatial correlation of SSOM to the geology and geomorphologic situation of the study area. The classified result obtained from modeled DAIS shows similarity with the DAIS image. However, the DAIS image resolved a tiny area of high SSOM areas in many parts of A<sub>1</sub> (figure 5-6). This is attributed due to the high spectral resolution nature of the DAIS image. According to Joachim and Brigitta, (2000), in semi-arid areas high organic matter contents are generally found below vegetations or in clay soils in small accumulation areas. Such surface elements are not spatially resolved by the TM systems with its IFOV of 30x30 m<sup>2</sup>. Simulating/mimicking some spectral characteristics of DAIS by means of multiple regression models does not identify these areas in modeled DAIS image.

From the classified SSOM images (figure 5-6), one can see that the high SSOM areas in the North-western part of the study area follows rectangular, elongated land cover pattern. This might be due to agricultural activities; these areas are annually covered dominantly by olive vegetation (Arnaldo, 2001). Cropping practice in semi-arid areas usually influences the spatial distribution and pattern of SSOM because fertilizers and rotation of different crops increase soil organic matter content (Biederbeck et al., 2002). Additionally, management practices also influence the nature of SSOM as indicated by Palacios-Orueta and Ustin, (1998). Therefore, this might be the causes for the recognized pattern.

The results from the DAIS confirmed a strong case for the potential application of hyperspectral RS to SSOM mapping using inferred variables. Classified SSOM results found from DAIS seem promising. They are geologically and geomorphologically correlated after cross checking with ancillary data sources. They also not only provide a clear identification of SSOM potential areas, but the low spectral resolution image can get an extra advantage through the statistical models derived by means of regression analysis. It improved some capability of the TM image through pixel-by-pixel registration of the established mathematical models that has been derived by means of regression analysis. This study indicates a promising future for further implementation of such statistical and mathematical modeling techniques.

## 6.5. Benefits from the research

Timely information on the content and distribution of key soil nutrients is vital to support precision agriculture and agriculture in general. Studying and characterizing the spatial and temporal extent of SSOM plays a vital role to: monitor soil quality, to assess land degradation processes and so on. Therefore, a close investigation of SSOM has great relevance. Currently very few soil nutrients are identifiable from satellite sensors. Satellites with very good spatial resolution are usually composed of few and broad spectral bands, which significantly limits the identification of soil nutrients such as SSOM. Integrating satellite images are the current practices of mapping indiscernible earth objects (Daniel et al., 2000).

The present approach addressed the purpose of identifying potential SSOM areas in DAIS and modeled DAIS images. Generally, this study has demonstrated that integrating hyperspectral and multispectral images can help to improve identification of SSOM. As shown in the SSOM classified results, the identification of potential SSOM areas from modeled DAIS image is improved as compared to TM classified result. Therefore, precision farming or agriculture in general could be benefited from such improved capacity where the potential of remote sensing will be substantially harnessed. The approach can also be useful for identifying other soil nutrients from RS sources. However, hyperspectral images are expensive and not easily available especially in developing countries. The advantage of our approach is that we need hyperspectral data for a small area only. This small area can be manually covered by taking field spectrometer measurements. Although, the cost of the field spectrometer is still may be a problem. However, the current trend in geoinformation encourages outsourcing tools that are expensive, and not often used. Therefore, whenever required it can be hired for a few days from specialised companies at a relatively cheaper cost. The other alternatives are either work in collaboration with international Non-Governmental Organizations (NGO's). There are a number of NGO's in developing countries who are actively participating in monitoring and management of natural resources; or rent field spectrometer from these NGO's because the rent cost may not demand big money since taking spectral measurements of such a small area does not take long time.

This research can be useful in two ways: either it can be directly applied or it will be used as exploratory applications. If it is possible to get the spectrometer or hyperspectral image, it can be easily implemented. Whereas, it can be used as an exploratory research showing that the potential in this area and decision makers will use it in the future when they think it is essential.

## 7. Conclusions and Recommendations

### 7.1. Conclusions

This research was carried out to evaluate the applicability of spectral unmixing of hyperspectral and multispectral images to identify potential areas of SSOM in the semi-arid environment of Antequera area in Spain. The specific objectives were: i) to identify optimum number of DAIS and Landsat TM bands that are suitable for SSOM level discrimination, ii) to establish a relationship between the two images, iii) to identify potential areas of SSOM in the study area. A set of five research questions were put forward to be answered. Within the framework of the objectives, research questions, and problems the following conclusions were derived about the applicability of spectral unmixing to derive endmember abundance images for SSOM mapping.

#### 7.1.1. From the perspective of the research problems

**The problem we proposed to address was how to deal with mixed pixels when applying hyperspectral and multispectral sensor systems to map SSOM potential areas.** This research demonstrated the use of spectral unmixing in combination with spectral angle mapper for the analysis of the different images used in this study. Spectral unmixing yields abundance estimates at a pixel level for all the endmembers. This means that spectral unmixing calculates, at each pixel, the abundance of pixel components that is not possible for conventional classification algorithms. Furthermore, the produced RMSE image as a result of the unmixing provides a means of quantifying the accuracy of the unmixing classification.

This research also showed that selection of endmembers is important for spectral unmixing classification. Choosing the wrong endmembers can make a large difference in classification results. Endmember selection involves an iterative process. Therefore, care has to be taken when applying the techniques, since unmixing is based on the number of endmembers.

**Hyperspectral images have high spectral resolution but also high costs, on the other hand multispectral images have a low spectral resolution and also low costs.** This research attempted to use extra advantage from hyperspectral images to improve or enhance SSOM identification of Landsat TM image in the strip part of the study area. This part also addresses the research question: *Is an integrating DAIS and Landsat TM image improves discrimination of SSOM levels in the study area?* Integrating DAIS and Landsat TM images showed to be able to improve SSOM level discrimination. In doing so, we simulate/mimic spectral characteristics of the DAIS image by means of a statistical regression. The methodology outlined in this thesis accomplished this part of the task. This study has as a limitation that the derived discriminant and multiple regression models could not be universal: for every campaign new field samples need to be collected and new SDA and regression equations should be constructed because soil constituents and reflectance values are geographically dependent.

Limitations associated to the cost problem however can be minimized by different alternatives as mentioned in section 6.5. The advantage of our approach is that we need hyperspectral data for a small area that can be covered by using field spectrometer data collection.

### **7.1.2. From the perspective of the research questions:**

**What are the optimum numbers of bands of DAIS and TM images suitable for SSOM level identification?** The study identified seven DAIS and four Landsat TM bands that are useful for discriminating SSOM levels in the study area in Spain. A Stepwise Discriminant Analysis (SDA) with three separability indicators was used. The discrimination power increased when using the seven DAIS bands as compared to four TM bands. The bands of the two images were related using regression models. This study also showed the usefulness of statistical modeling techniques for integrating RS images. (See chapter four).

**What is the influence of different band combinations on classification results of spectral unmixing?** As discussed in spectral unmixing result discussion section in chapter four, different band combinations should be tried to get the best unmixing results because some bands may contain noise and redundant information and inclusion of them may not help to improve unmixing.

**How can we evaluate the accuracy of spectral unmixing results?** Endmember abundance and RMSE images derived from the unmixing analysis can be used to evaluate the fitness of the unmixing model. (See chapter four).

### **7.1.3. From the perspective the main objective:**

**Evaluate the applicability of spectral unmixing technique to map/identify potential areas of SSOM in the study area;** this part also addresses the research question; *Can endmember abundance images derived from unmixing analysis be used for mapping potential areas of SSOM in the study area?* Images with endmember abundance derived from spectral unmixing of DAIS and modeled DAIS images were used to identify potential areas of SSOM in the strip area. The use of ancillary data such as topographic and geologic maps of the area proved to be useful in interpreting the results. Therefore, spectral unmixing applied on the two images was applicable and potential to derive suitable endmember fraction images that could be used as an inferred variable to map potential areas of SSOM. A spectral unmixing analysis of Landsat TM did not give a reliable result. Therefore, mapping potential areas of SSOM levels was not possible.

In general the methodology applied in this research could offer the prospect for mapping other soil nutrients, which may serve as input nutrient parameters to other models such as crop growth, soil erosion, etc.

## **7.2. Recommendations**

- For selecting optimum number of DAIS and TM bands useful for discriminating SSOM levels, other discriminant analysis techniques such as quadratic, logistic and penalized discriminant analysis also should be further explored.
- For a better SSOM level discrimination in the strip area, purposive stratification of the study area should be done in the field and more test data must be collected.

- Field or laboratory spectral data library should be used to select appropriate and pure endmembers that can describe the spectral character of the study area.

## Bibliography

- Adams, J.B., Smith, M.O., and Johnson, P.E., 1985, Spectral mixture modeling: A new analysis of rock and soil types at the Viking Lander 1 site.: *Journal of Geophysical Research*, v. 91(B8), p. 8090-8112.
- Agbu, P.A., and Coleman, T.L., 1990, Techniques of Digital Image Analysis for Soil Map Unit Discrimination: American Society for Photogrammetry and Remote Sensing, and American Congress on Surveying and Mapping, p. B-1 - B-9.
- Ahmadou, B., Ilaud, D., and Afework, K., 1999, Semi-detailed soil survey of Antequera (Spain): Report on the field work module of SOL3 course, ITC.
- Arnaldo, A.S., 2001, Land cover change detection and monitoring using Remote sensing and geographical information system in Antequera, Southern Spain: Enschede, ITC.
- Asner, G., and Lobell, D., 2000, A biogeophysical approach for automated SWIR unmixing of soils and vegetation: *Remote Sensing of Environment*, v. 74, p. 99–112.
- ATCOR2 User Manual, 1997, ATCOR for EARDAS IMAGINE; Atmospheric and Topographic correction, in 2<sup>nd</sup>, ed.
- Baumgardner, M.F., Silva, L.F., Biehl, L.L., and Stoner, E., 1985, Reflectance properties of soils.: *Adv. Agron.*, v. 38, p. 1–44.
- Ben-Dor, E., and Banin, A., 1994, Visible and near infrared (0.4–1.1  $\mu\text{m}$ ) analysis of arid and semiarid soils.: *Remote Sensing of Environment*, v. 48, p. 261–274.
- Ben-Dor, E., Inbar, Y., and Chen, Y., 1997, The reflectance spectra of organic matter in the visible near infrared and short wave infrared region (400–2500 nm) during a controlled decomposition process: *Remote Sensing of Environment*, v. 61, p. 1–15.
- Ben-Dor, E., PATKIN, K., BANIN, A., and KARNIELI, A., 2002, Mapping of several soil properties using DAIS-7915 hyperspectral scanner data—a case study over clayey soils in Israel: *int. j. remote sensing*, v. 23, p. 1043–1062.
- Biederbecka, V.O., Janzenb, H.H., Campbella, C.A., and Zentnera, P.A., 2002, Labile soil organic matter as influenced by cropping practices in an arid environment: *Soil Biology and Biochemistry*, v. 26, p. 1647-1656.
- Boardmann, J.W., Kruse, F.A., and Green, R.O., 1995, Mapping target signatures via partial unmixing of AVIRIS data: *Proceedings of the fifth JPL Airborne Earth Science Workshop*, v. 1, p. 23-26.
- Brennan, P., and Silman, A., 1992, Statistical methods for assessing observer variability in clinical measures: accessed date 27<sup>th</sup> January 2005: <http://www.cmh.edu/stats/definitions/kappa.htm>.
- Cambridge University, 2002, The state of the nations Ecosystem: Soil organic Matter, Cambridge University.
- Campbell, J.B., 1996, *Introduction to Remote Sensing*, the Guilford Press, New York.
- Charles, I., and Arnon, K., 1996, A Review of Mixture Modeling Techniques for Sub-pixel land cover Estimation: *Remote sensing reviews*, v. 13, p. 161-186.
- Clark, R.N., 1999, *Spectroscopy of rocks and minerals, and principles of spectroscopy*, (New York: John Wiley & Sons), 3–58. p.
- Clevers, J.G.P.W., 1999, The use of imaging spectrometry for agricultural applications: *ISPRS Journal of Photogrammetry & Remote Sensing*, v. 54, p. 299-304.
- Congalton, R.G., 1991, A review of assessing the accuracy of classification of remotely sensed data: *Remote Sensing of Environment*, v. 37, p. 35–46.
- Cracknell, A.P., 1998, Synergy in remote sensing - what's in a pixel?: *International journal of remote sensing*, v. 19, p. 2025± 2047.

- Dalal, R.C., and Henery, R.J., 1986, Simultaneous determination of moisture, organic carbon and total nitrogen by near infrared reflectance spectrophotometry: *Soil Science Society of America Journal*, v. 50, p. 120–123.
- Daniel, K.W., Tripathi, N.K., and Honda, K., 2000, An Approach for Estimating Soil Organic Matter Content Using Synthetic IRS Satellite Data in tropical Soils of Lop Buri, Thailand, *Space Technology Applications and Research (STAR) Asian Institute of Technology (AIT)*.
- Daniel, K.W., Tripathi, N.K., Honda, K., and Apisit, E., 2002, Analysis of VNIR (400–1100 nm) spectral signatures for estimation of soil organic matter in tropical soils of Thailand: *INT. J. REMOTE SENSING*, v. 25, p. 643–652.
- Darvishsefat, A.A., 2002, Application of hyperspectral data for forest stand mapping, *Symposium on geospatial theory, processing and applications: Ottawa*.
- David, B., Lobell, Gregory, and Asner, P., 2004, Cropland distributions from temporal unmixing of MODIS data: *Remote Sensing of Environment*, v. 93, p. 412– 422.
- Demetriades-Shah, T.H., Steven, M.D., and Clark, J.A., 1990, High resolution derivative spectra in remote sensing: *Remote sensing and Environment*, v. 33, p. 55-64.
- Dengsheng, L., and Qihao, W., 2004, Spectral Mixture Analysis of the Urban Landscape in Indianapolis with Landsat ETM+ Imagery: *Photogrammetric Engineering & Remote Sensing*, v. 70, p. 1053-1062.
- Dennison, P.E., and Roberts, D.A., 2003, Endmember selection for multiple endmember spectral mixture analysis using endmember average RMSE: *Remote Sensing of Environment*, v. 87, p. 123–135.
- Dominique, L., Erwan, F., Steven, D., and Charles, W., 1999, Systolic array for computing the pixel purity index (PPI) algorithm on hyper spectral images, IRISA, Campus de Beaulieu, 35042 Rennes cedex, France.
- Donat-P.Hader, 1992, *Image analysis:Methods and applications*, CRB press.
- ENVI Tutorial, 2000, ENVI version 3.4. Environment for Visualizing Images. With software, 200.
- Fisher, P., 1997, The pixel: a snare and a delusion: *International Journal of Remote Sensing*, v. 18, p. 679-685.
- Fisher, R.A., 1936, The use of multiple measurements in taxonomic problems, Volume 7, p. 179-188.
- Foody, G.M., 1996b, Relating the land-cover composition of mixed pixels to artificial neural network classification output models: *Photogrammetric Engineering*, v. 62, p. 491–499.
- Galvao, L.S., and Vitorello, I., 1998, Role of organic matter in obliterating the effects of iron on spectral reflectance and colour of Brazilian tropical soils: *International Journal of Remote Sensing*, v. 19, p. 1969–1979.
- Geotz, A.F.H., 1992, *Imaging Spectroscopy:Fundamentals and Prospective Applications*, Ed's Toselli and Bodechtel: London, Kluwer Acad.Publ., 61-71 p.
- Girard-Ganneau, C., M., 1975, [PhD dissertation thesis], University of Paris.
- Girouard, G., Bannari, A., and El Harti, A., 2001, Validated SAM algorithm for Geological mapping: Ottawa, University of Ottawa.
- Green, A., Berman, A., Switzer, M., and Craig, M.D., 1988, A transformation for ordering multispectral data in terms of image quality with implications for noise removal.: *IEEE T ransactions on Geoscience and Remote Sensing*, v. 26, p. 65–74.
- Gross, H., N., 1998, Application of spectral Mixture Analysis and Image Fusion techniques for Image sharpening: *Remote sensing and Environment*, v. 63, p. 85-94.
- Grunwald, S., 1999, Soil Organic Matter, University of Florida, Last updated: Sept 8, 1999: <http://www.soils.wisc.edu/courses/SS325/organic.htm#composition>, Accessed: 23 September, 2004.
- Hoffer, and Johannsen, C.J., 1969, *Remote sensing in ecology*: Athens, University of Georgia Press.
- Irons, J.R., Markham, B.L., Nelson, R.F., Toll, D.L., Williams, D.L.Latty, R.S., and Stauer, M.L., 1985, The effects of spatial resolution on the classification of Thematic Mapper data.: *International Journal of Remote Sensing*, v. 6, p. 1385-1403.
- Jensen, J.R., 1996, *Introductory Digital Image Processing: A remote sensing perspective*: New Jersey, Prentice-Hall, 572-574 p.

- Joachim, H., and Brigitta, S., 2000, Mapping Complex Patterns of Erosion and Stability in Dry Mediterranean Ecosystems: Remote sensing and Environment, v. 74, p. 557-569.
- Kariuki, P.C., 2004, Spectroscopy and Swelling Soils: An integrated approach [Ph.D thesis], ITC.
- Kariuki, P.C., and van der Meer, F.D., 2003, Swelling Clay Mapping For Characterizing Expansive Soils; Results From Laboratory Spectroscopy and Hysens Dais Analysis, Presented at the 3rd EARSeL Workshop on Imaging Spectroscopy: Oberpfaffenhofen.
- Koch, M., 2000, Geological controls of land degradation as detected by remote sensing:a case study in Los Menegros, north-east Spain: International Journal of Remote sensing, v. 21, p. 457-473.
- Kyu-Sung, L., Warren, B., Robert, E., Thomas, K., and Stith, T., 2004, Hyperspectral versus multispectral data for estimating leaf area index in four different biomes: Remote Sensing of Environment, v. 91, p. 508–520.
- Lauer, D.T., Morain, S.A., and Salomonson, V.V., 1997, The Landsat program: Its origins, evolution, and impacts: Photogrammetric Engineering and Remote Sensing,, v. 63, p. 831-838.
- Lehmann, F., Oertel, D., Richter, R., Hausknecht, P., Rothfuss, H., Strobl, P., and Mueller, A., 1995, The Digital Airborne Imaging Spectrometer DAIS 7915: Hyperspectral applications with a new sensor, Int. Symp. on Spectral Sensing Research: Melbourne.
- Lehmann, F., Oertel, D., Richter, R., Hausknecht, P., 1995, The Digital Airborne Imaging Spectrometer DAIS 7915, Hyperspectral applications with a new sensor, Int. Symp. on Spectral Sensing Research: Melbourne.
- Lillesand, T.M.a.R.W.K., 1994., Remote sensing and image interpretation: Evaluating the usefulness of Landsat: New York, John Wiley.
- Marsh, Switzer, and Kowalik., 1980, Resolving the percentage of component terrains within single resolution elements.: Photogrammetric Engineering and Remote Sensing,, v. 46, p. 1079-1086.
- Mather, P.M., 1999, Computer processing of remotely-sensed images: An introduction, New York: Wiley.
- Mathieu, R., Pouget, M., Cervelle, B., and Escadafal, R., 1998, Relationships between satellite-based radiometric indices simulated using laboratory reflectance data and Typic soil color of an arid environment: Remote Sensing of Environment, v. 66, p. 17–28.
- Maurice, S.G., 1998, Decomposition of mixed pixels in remote sensing images to improve the area estimation of agricultural fields [PhD thesis], Katholieke Universiteit Nijmegen.
- Muller, A., Oertel,D, 1997, DAIS Large-Scale Facility, the DAIS 7915 Imaging Spectrometer in a European Frame., Proceedings, Third International airborne Remote Sensing Conference and Exhibition,, Volume 1: Copenhagen,, ERIM, p. 684-691.
- Mustard, J.F., Sunshine, J.M., and Rencz, A.N., 1999, Spectral analysis for Earth science: Investigations using remote sensing data., New York: Wiley, 251– 307 p.
- Nelson, D.W., Sommers, L.E., 1982, Total carbon, organic carbon, Methods of Soil Analysis, Part 2, p. 539–579.
- Painter, T.H., Roberts, D.A., Green, R.O., and Dozier, J., 1998, The effect of grain size on spectral mixture analysis of snow-covered area from AVIRIS data: Remote Sensing of Environment,, v. 65, p. 320– 332.
- Palacios-Orueta, A., and Ustin, S.L., 1998, Remote sensing of soil properties in the Santa Monica mountains I. Spectral analysis: Remote Sensing of Environment, v. 65, p. 170–183.
- Peng, W., 1998, Synthetic analysis for extracting information on soil salinity using remote sensing and GIS: A case study of Yamggao Basin, China.: Environmental Management, v. 22, p. 153-159.
- Prasad, S., Eden, A., Enclonab, c., Mark, S., Ashtonb, c., Christopher, L., and Minko, J., 2004, Hyperion, IKONOS, ALI, and ETM+ sensors in the study of African rainforests: Remote Sensing of Environment, v. 90, p. 23–43.
- Prasad, S.T., 2000, Optimal Hyperspectral Narrowbands for Discriminating Agricultural crops: USA, Center for Earth Observation (CEO), Yale University.
- Richter, R., 1996, Atmospheric Correction of DAIS 7915 Hyperspectral Image Data: submitted paper to the SPIE AeroSense '96 Symposium.

- Roberts, D.A., Gardner, M., Church, R., Ustin, S., and Scheer, G., 1998, Mapping chaparral in the Santa Monica mountains using multiple endmember spectral mixture models: *Remote sensing and Environment*, v. 65, p. 267–279.
- Schmidt, M., and Schoettker, B., 2004, Sub-pixel analysis in combination with knowledge based decision rules to optimise a land cover classification, Department of Geography, University of Bonn, Germany.
- Schoeneberger, P.J., Wysocki, D.A., Benham, E.C., and Broderson, W.D., 1998, Field book for describing and sampling soils. Natural resources conservation service, USDA, National soil survey center: Lincoln, NE.
- Schoenmakers, R.P.H.M., 1995, Integrated methodology for segmentation of large optical satellite images in land applications of remote sensing. [PhD thesis]: The Netherlands, University of Nijmegen.
- Schowengerdt, A.R., 1997, *Remote sensing: Models and Methods for Image processing*, Academic Press, 389-478 p.
- Schreier, 1977, Quantitative predictions of chemical soil conditions from multispectral airborne, ground and soils, 4th Proc. Can. Symp. Remote Sensing, Volume 38, p. 1-44.
- Settle, J.J., and Drake, N.A., 1993, Linear mixing and the estimation of ground cover proportions: *International Journal of Remote Sensing*, v. 14, p. 1159-1177.
- Shrestha, D.P., Margate, D.E., H.V., A., and Meer, v.D.F., 2002, Spectral unmixing versus spectral angle mapper for land degradation assessment: a case study in Southern Spain, *in* D.P, S., ed., *WCSS: Bangkok*, p. 1141-1151.
- Singer, and McCord, T.B., 1979, Mars: large scale mixing of bright and dark surface materials and implications for analysis of spectral reflectance, *Proceedings of the 10th Lunar and Planetary Science conference: USA*, p. 1835-1848.
- Statistica 6.0, 2002, *Statistica Electronic manual*.
- Stoner, E.R., and Baumgardner, M. F., 1981, Characteristic variation in reflectance of surface soils: *Soil Science Society of American Journal*, v. 45, p. 1161–1165.
- Strobl, P., Richter, R., Mueller, A., Lehmann, F., Oertel, D., Tischler, S., and Nielsen, A., 1996, DAIS System Performance, First Results of the 1995 Evaluation Campaigns, Volume 2758, p. 109-117.
- Tabachnick, B.G., and Linda, S.F., 2001, *Using Multivariate Statistics:chapter 11 covers discriminant analysis: Boston, Allyn and Baco*.
- Udelhoven, T., Hill, J., Imeson, A., and Cammeraat, H., 1997, A neural network approach for the identi. cation of the organic carbon content of soils in a degraded semiarid ecosystem (Guadalentin, SE, Spain) based on hyperspectral data from the DAIS-7915 sensor., *In Proceedings of 1st EARSeL Workshop on Imaging Spectroscopy: Zurich, Switzerland*, p. 437–444.
- Utah state university, 1999, *Understanding your soil test report*.
- Van der Meer, F., 1995, Spectral unmixing of Landsat Thematic Mapper data: *International Journal of remote Sensing*, v. 16, p. 3189-3194.
- Van der Meer, F., and de Jong, S., M., 2000, Improving the reults of spectral unmixing of Landsat TM imagery by enhancing the Orthogonality of end Members: *International Journal of remote Sensing*, v. 21, p. 2781-2797.
- Van der Meer, M., 1999, Spatial statistics for Remote Sensing., *in* Alfred Stein, Van der Meer, F., and Grote, B., eds., *Image classification through spectral unmixing.*, Kluwer Academic, p. 185-193.
- Velez-Reyes, M., 2003, *Iterative Algorithms for Unmixing of Hyperspectral Imagery: Proceeding of SPIE*.
- Webster, R., and Oliver, M., A., 2001, *Geostatistics for environmental scientists.: Chichester, Wiley & Sons*.

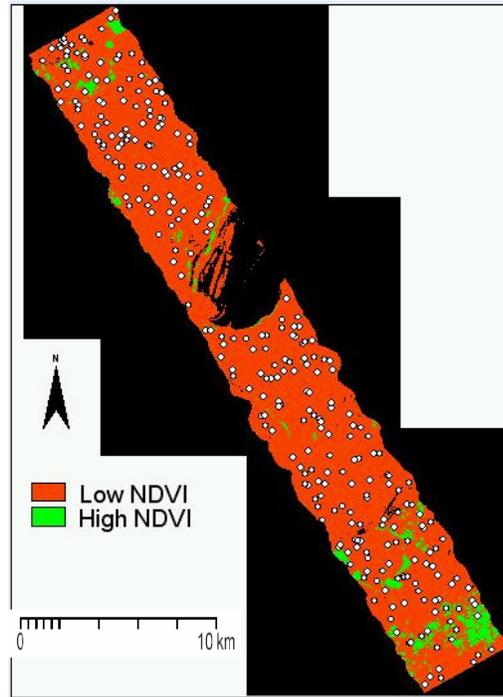
# Appendix

## Appendix 1: Field reference data

Test pit	X	Y	SSOM
ANT01	344448	4112113	0.040
ANT02	346342	4113277	3.880
ANT03	356451	4112403	2.720
ANT04	365186	4111733	3.220
ANT05	337799	4101545	2.740
ANT06	337802	4101619	3.200
ANT07	344700	4102243	1.800
ANT08	353783	4102174	1.880
ANT09	356655	4105285	2.700
ANT10	350518	4105867	1.560
ANT11	348469	4113502	1.740
ANT12	347272	4113094	2.080
ANT13	360878	4106314	2.120
ANT14	361051	4106343	1.840
ANT15	348145	4100942	2.800
ANT16	343862	4103767	0.840
ANT17	359446	4105579	2.320
ANT19	350890	4109739	1.400
ANT20	348724	4106147	1.580
ANT21	350809	4109863	1.840
ANT22	350695	4109965	1.380
ANT23	354564	4111965	1.260
ANT24	354652	4112075	1.300
ANT25	354854	4112146	2.680
ANT26	360030	4108577	2.840
ANT27	360121	4109086	1.100
ANT28	347183	4098149	2.180
ANT29	347079	4098596	2.420
ANT30	346760	4098739	1.980
ANT31	362025	4106337	1.320
ANT32	363425	4105125	1.740
ANT33	363100	4103200	3.120
ANT34	355547	4103813	2.600
ANT35	355522	4104073	3.020
ANT36	355517	4104476	2.700

Test pit	X	Y	SSOM
ANT37	342784	4102233	0.600
ANT38	344037	4103460	2.140
ANT39	343550	4103800	3.840
ANT40	340290	4112834	1.940
ANT41	338996	4112454	2.860
ANT42	337475	4113757	2.260
ANT43	345396	4110304	3.040
ANT44	356693	4107293	1.460
ANT45	357548	4107270	1.480
ANT46	357725	4103288	2.060
ANT47	345217	4097328	2.160
ANT48	345634	4097838	1.060
ANT49	344532	4112454	1.540
ANT51	350615	4108928	1.220
ANT52	359599	4113002	0.800
ANT53	361440	4107278	2.300
ANT54	361475	4103401	2.340
ANT55	355935	4102122	1.680
ANT56	356346	4101108	1.280
ANT57	353972	4101910	2.320
ANT58	347647	4110126	2.100
ANT59	346204	4109272	2.480
ANT60	344720	4109573	1.160
ANT61	354487	4106923	2.020
ANT62	355787	4106467	2.860
ANT63	353155	4108294	2.220
ANT64	359829	4109583	2.100
ANT65	360211	4109257	2.420
ANT66	360687	4108566	2.400
ANT67	349618	4109368	0.800
ANT68	349763	4109509	1.500
ANT69	344727	4106830	2.780
ANT70	346100	4113650	2.000
ANT71	354300	4110317	1.720
ANT72	362931	4111039	2.380
ANT73	362369	4111276	3.140
ANT74	363367	4110906	2.380
ANT75	354548	4103688	2.520
ANT76	354502	4103793	1.360
ANT77	354452	4103880	2.320
ANT78	360999	4108464	4.540
ANT671	361159	4113297	3.460

**Appendix 2: Artificially generated samples and NDVI image**



**Appendix 3 - The Wilk's  $\lambda$  value of DAIS through the three functions**

Test of Function(s)	$\lambda$	$X^2$	df	Sig.
1 through 3	0.478	216.452	21	.000
2 through 3	0.732	91.587	12	.000
3	0.939	18.462	5	.002

**Appendix 4: Fisher's linear discriminant coefficients of DAIS**

Wavelengths	Strata			
	1.00	2.00	3.00	4.00
0.554	-94.821	-12.893	-99.596	-96.803
0.939	52.555	-65.017	43.052	86.430
0.990	-1201.934	-1304.057	-1365.993	-1167.570
1.020	1285.184	1488.856	1469.287	1223.394
2.010	-194.638	-224.496	-246.062	-235.215
2.026	363.519	351.129	429.215	413.781
2.304	-27.028	16.045	2.221	-13.479
Constant	-42.631	-45.311	-56.487	-49.140

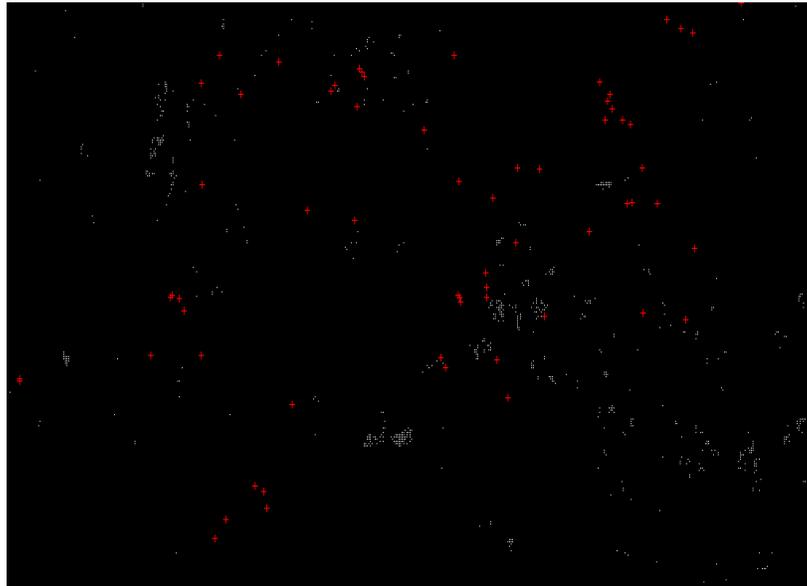
**Appendix 5 – Wilk's  $\lambda$  values of Landsat TM through the three functions**

Test of Function(s)	$\lambda$	$X^2$	df	Sig.
1 through 3	.674	116.266	12	.000
2 through 3	.906	29.145	6	.000
3	.983	5.071	2	.079

**Appendix 6: Fisher's linear discriminant coefficients of Landsat TM**

Wavelengths	Strata			
	1.00	2.00	3.00	4.00
0.520	-166.171	-251.948	-375.069	-246.189
0.630	-129.612	-142.360	-35.859	-72.065
0.760	195.927	262.804	245.060	219.877
2.350	221.779	247.511	258.638	204.566
Constant	-14.162	-19.728	-18.871	-14.955

**Appendix 7: PPI result of Landsat TM**



PPI result of Landsat TM image showing the purest pixels (in white) with an overlay of field reference data (red)

**Appendix 8: RMSE values of each pixel derived from unmixing analysis of modeled DAIS**

RMSE	No of pixels	Total	pixel%	Accumulated%
0.000	2483	2483	0.78%	0.78%
0.018	2372	4855	0.74%	1.52%
0.036	2325	7180	0.73%	2.24%
0.054	2375	9555	0.74%	2.98%
0.073	252442	261997	78.79%	81.78%
0.091	2260	264257	0.71%	82.48%
0.109	2296	266553	0.72%	83.20%
0.127	2198	268751	0.69%	83.88%
0.145	2074	270825	0.65%	84.53%
0.163	2062	272887	0.64%	85.17%
0.182	1995	274882	0.62%	85.80%
0.200	1989	276871	0.62%	86.42%
0.218	1761	278632	0.55%	86.97%
0.236	1804	280436	0.56%	87.53%
0.254	1652	282088	0.52%	88.05%
0.272	1663	283751	0.52%	88.57%
0.290	1598	285349	0.50%	89.06%
0.309	1530	286879	0.48%	89.54%
0.327	1481	288360	0.46%	90.00%
0.345	1417	289777	0.44%	90.45%

RMSE	No of pixels	Total	pixel%	Accumulated%
0.363	1264	291041	0.39%	90.84%
0.381	1153	292194	0.36%	91.20%
0.399	1053	293247	0.33%	91.53%
0.417	1085	294332	0.34%	91.87%
0.436	978	295310	0.31%	92.17%
0.454	975	296285	0.30%	92.48%
0.472	860	297145	0.27%	92.75%
0.490	857	298002	0.27%	93.01%
0.508	813	298815	0.25%	93.27%
0.526	812	299627	0.25%	93.52%
0.544	669	300296	0.21%	93.73%
0.563	631	300927	0.20%	93.93%
0.581	626	301553	0.20%	94.12%
0.599	609	302162	0.19%	94.31%
0.617	553	302715	0.17%	94.48%
0.635	542	303257	0.17%	94.65%
0.653	563	303820	0.18%	94.83%
0.671	458	304278	0.14%	94.97%
0.690	509	304787	0.16%	95.13%
0.708	459	305246	0.14%	95.27%
0.726	493	305739	0.15%	95.43%
0.744	385	306124	0.12%	95.55%
0.762	362	306486	0.11%	95.66%
0.780	408	306894	0.13%	95.79%
0.798	377	307271	0.12%	95.91%
0.817	306	307577	0.10%	96.00%
0.835	294	307871	0.09%	96.09%
0.853	314	308185	0.10%	96.19%
0.871	257	308442	0.08%	96.27%
0.889	222	308664	0.07%	96.34%
0.907	295	308959	0.09%	96.43%
0.925	244	309203	0.08%	96.51%
0.944	266	309469	0.08%	96.59%
0.962	242	309711	0.08%	96.67%
0.980	237	309948	0.07%	96.74%
0.998	226	310174	0.07%	96.81%
1.016	235	310409	0.07%	96.89%
1.034	196	310605	0.06%	96.95%
1.053	209	310814	0.07%	97.01%
1.071	235	311049	0.07%	97.09%
1.089	217	311266	0.07%	97.15%
1.107	212	311478	0.07%	97.22%

RMSE	No of pixels	Total	pixel%	Accumulated%
1.125	175	311653	0.05%	97.27%
1.143	217	311870	0.07%	97.34%
1.161	164	312034	0.05%	97.39%
1.180	167	312201	0.05%	97.45%
1.198	189	312390	0.06%	97.50%
1.216	155	312545	0.05%	97.55%
1.234	172	312717	0.05%	97.61%
1.252	172	312889	0.05%	97.66%
1.270	146	313035	0.05%	97.71%
1.288	130	313165	0.04%	97.75%
1.307	131	313296	0.04%	97.79%
1.325	123	313419	0.04%	97.83%
1.343	158	313577	0.05%	97.88%
1.361	124	313701	0.04%	97.91%
1.379	118	313819	0.04%	97.95%
1.397	121	313940	0.04%	97.99%
1.415	117	314057	0.04%	98.02%
1.434	98	314155	0.03%	98.06%
1.452	109	314264	0.03%	98.09%
1.470	99	314363	0.03%	98.12%
1.488	125	314488	0.04%	98.16%
1.506	130	314618	0.04%	98.20%
1.524	103	314721	0.03%	98.23%
1.542	102	314823	0.03%	98.26%
1.561	93	314916	0.03%	98.29%
1.579	94	315010	0.03%	98.32%
1.597	96	315106	0.03%	98.35%
1.615	97	315203	0.03%	98.38%
1.633	103	315306	0.03%	98.41%
1.651	82	315388	0.03%	98.44%
1.669	78	315466	0.02%	98.46%
1.688	83	315549	0.03%	98.49%
1.706	83	315632	0.03%	98.52%
1.724	95	315727	0.03%	98.55%
1.742	69	315796	0.02%	98.57%
1.760	71	315867	0.02%	98.59%
1.778	66	315933	0.02%	98.61%
1.796	73	316006	0.02%	98.63%
1.815	71	316077	0.02%	98.66%
1.833	86	316163	0.03%	98.68%
1.851	72	316235	0.02%	98.70%
1.869	65	316300	0.02%	98.73%

RMSE	No of pixels	Total	pixel%	Accumulated%
1.887	70	316370	0.02%	98.75%
1.905	64	316434	0.02%	98.77%
1.924	79	316513	0.02%	98.79%
1.942	77	316590	0.02%	98.82%
1.960	63	316653	0.02%	98.84%
1.978	60	316713	0.02%	98.85%
1.996	74	316787	0.02%	98.88%
2.014	59	316846	0.02%	98.90%
2.032	78	316924	0.02%	98.92%
2.051	62	316986	0.02%	98.94%
2.069	67	317053	0.02%	98.96%
2.087	55	317108	0.02%	98.98%
2.105	38	317146	0.01%	98.99%
2.123	46	317192	0.01%	99.00%
2.141	50	317242	0.02%	99.02%
2.159	52	317294	0.02%	99.04%
2.178	53	317347	0.02%	99.05%
2.196	56	317403	0.02%	99.07%
2.214	43	317446	0.01%	99.08%
2.232	46	317492	0.01%	99.10%
2.250	47	317539	0.01%	99.11%
2.268	45	317584	0.01%	99.13%
2.286	54	317638	0.02%	99.14%
2.305	55	317693	0.02%	99.16%
2.323	43	317736	0.01%	99.17%
2.341	39	317775	0.01%	99.19%
2.359	42	317817	0.01%	99.20%
2.377	45	317862	0.01%	99.21%
2.395	45	317907	0.01%	99.23%
2.413	49	317956	0.02%	99.24%
2.432	40	317996	0.01%	99.25%
2.450	47	318043	0.01%	99.27%
2.468	41	318084	0.01%	99.28%
2.486	41	318125	0.01%	99.29%
2.504	44	318169	0.01%	99.31%
2.522	37	318206	0.01%	99.32%
2.540	51	318257	0.02%	99.34%
2.559	42	318299	0.01%	99.35%
2.577	39	318338	0.01%	99.36%
2.595	32	318370	0.01%	99.37%
2.613	41	318411	0.01%	99.38%
2.631	43	318454	0.01%	99.40%

RMSE	No of pixels	Total	pixel%	Accumulated%
2.649	41	318495	0.01%	99.41%
2.667	48	318543	0.02%	99.43%
2.686	42	318585	0.01%	99.44%
2.704	32	318617	0.01%	99.45%
2.722	46	318663	0.01%	99.46%
2.740	51	318714	0.02%	99.48%
2.758	33	318747	0.01%	99.49%
2.776	38	318785	0.01%	99.50%
2.795	28	318813	0.01%	99.51%
2.813	33	318846	0.01%	99.52%
2.831	43	318889	0.01%	99.53%
2.849	40	318929	0.01%	99.55%
2.867	33	318962	0.01%	99.56%
2.885	42	319004	0.01%	99.57%
2.903	52	319056	0.02%	99.59%
2.922	38	319094	0.01%	99.60%
2.940	49	319143	0.02%	99.61%
2.958	36	319179	0.01%	99.62%
2.976	45	319224	0.01%	99.64%
2.994	29	319253	0.01%	99.65%
3.012	45	319298	0.01%	99.66%
3.030	29	319327	0.01%	99.67%
3.049	32	319359	0.01%	99.68%
3.067	45	319404	0.01%	99.69%
3.085	25	319429	0.01%	99.70%
3.103	31	319460	0.01%	99.71%
3.121	38	319498	0.01%	99.72%
3.139	30	319528	0.01%	99.73%
3.157	35	319563	0.01%	99.74%
3.176	37	319600	0.01%	99.76%
3.194	35	319635	0.01%	99.77%
3.212	31	319666	0.01%	99.78%
3.230	35	319701	0.01%	99.79%
3.248	34	319735	0.01%	99.80%
3.266	22	319757	0.01%	99.80%
3.284	28	319785	0.01%	99.81%
3.303	30	319815	0.01%	99.82%
3.321	27	319842	0.01%	99.83%
3.339	31	319873	0.01%	99.84%
3.357	13	319886	0.00%	99.84%
3.375	20	319906	0.01%	99.85%
3.393	40	319946	0.01%	99.86%

RMSE	No of pixels	Total	pixel%	Accumulated%
3.411	18	319964	0.01%	99.87%
3.430	27	319991	0.01%	99.88%
3.448	26	320017	0.01%	99.89%
3.466	8	320025	0.00%	99.89%
3.484	13	320038	0.00%	99.89%
3.502	19	320057	0.01%	99.90%
3.520	13	320070	0.00%	99.90%
3.538	18	320088	0.01%	99.91%
3.557	18	320106	0.01%	99.91%
3.575	16	320122	0.01%	99.92%
3.593	11	320133	0.00%	99.92%
3.611	12	320145	0.00%	99.93%
3.629	16	320161	0.01%	99.93%
3.647	14	320175	0.00%	99.93%
3.666	15	320190	0.00%	99.94%
3.684	14	320204	0.00%	99.94%
3.702	7	320211	0.00%	99.95%
3.720	7	320218	0.00%	99.95%
3.738	10	320228	0.00%	99.95%
3.756	10	320238	0.00%	99.95%
3.774	9	320247	0.00%	99.96%
3.793	7	320254	0.00%	99.96%
3.811	7	320261	0.00%	99.96%
3.829	8	320269	0.00%	99.96%
3.847	8	320277	0.00%	99.97%
3.865	8	320285	0.00%	99.97%
3.883	7	320292	0.00%	99.97%
3.901	10	320302	0.00%	99.97%
3.920	5	320307	0.00%	99.98%
3.938	4	320311	0.00%	99.98%
3.956	5	320316	0.00%	99.98%
3.974	4	320320	0.00%	99.98%
3.992	5	320325	0.00%	99.98%
4.010	4	320329	0.00%	99.98%
4.028	5	320334	0.00%	99.98%
4.047	2	320336	0.00%	99.98%
4.065	1	320337	0.00%	99.99%
4.083	3	320340	0.00%	99.99%
4.101	4	320344	0.00%	99.99%
4.119	4	320348	0.00%	99.99%
4.137	2	320350	0.00%	99.99%
4.155	5	320355	0.00%	99.99%

RMSE	No of pixels	Total	pixel%	Accumulated%
4.174	2	320357	0.00%	99.99%
4.192	1	320358	0.00%	99.99%
4.210	1	320359	0.00%	99.99%
4.246	4	320363	0.00%	99.99%
4.264	2	320365	0.00%	99.99%
4.282	3	320368	0.00%	99.99%
4.301	1	320369	0.00%	100.00%
4.319	4	320373	0.00%	100.00%
4.337	1	320374	0.00%	100.00%
4.373	1	320375	0.00%	100.00%
4.391	2	320377	0.00%	100.00%
4.409	3	320380	0.00%	100.00%
4.446	1	320381	0.00%	100.00%
4.482	2	320383	0.00%	100.00%
4.573	1	320384	0.00%	100.00%
4.609	1	320385	0.00%	100.00%

**Appendix 9:** Endmember class statistics for A using modeled DAIS

Endmembers	SAM class statistics
V <sub>1</sub>	8.7504%
S <sub>1</sub>	2.1344%
S <sub>2</sub>	29.6612%
S <sub>3</sub>	46.7619%
W	0.6758%

**Appendix 10:** Error map for A derived from unmixing of modeled DAIS image

