SOIL CLASSIFICATION WITH SPACEBORNE MULTI-TEMPORAL HYPERSPECTRAL IMAGERY USING SPECTRAL UNMIXING AND IMAGE FUSION

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

EYLEM KABA

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
GEODETIC AND GEOGRAPHIC INFORMATION TECHNOLOGIES

MARCH 2023

Approval of the thesis:

SOIL CLASSIFICATION WITH SPACEBORNE MULTI-TEMPORAL HYPERSPECTRAL IMAGERY USING SPECTRAL UNMIXING AND IMAGE FUSION

submitted by **EYLEM KABA** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy** i**n Geodetic and Geographic Information Technologies, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. Sevda Zuhal Akyürek
Head of the Department, **Geodetic and Geographic Information Technologies, METU** _____

Prof. Dr. Sevda Zuhal Akyürek
Supervisor, **Civil Eng., METU** _____

Prof. Dr. Uğur Murat Leloğlu
Co-Supervisor, **Astronautical Eng., THK University** _____

Examining Committee Members:

Prof. Dr. Mehmet Lütfi Süzen
Geology Engineering, METU _____

Prof. Dr. Sevda Zuhal Akyürek
Civil Engineering, METU _____

Prof. Dr. Afşar Saranlı
Electric Engineering, METU _____

Prof. Dr. Nevsan Şengil
Astronautical Engineering, THK University _____

Assoc. Prof. Dr. Emre Sümer
Computer Engineering, Başkent University _____

Date: 24.03.2023

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : Eylem Kaba

Signature :

ABSTRACT

SOIL CLASSIFICATION WITH SPACEBORNE MULTI-TEMPORAL HYPERSPECTRAL IMAGERY USING SPECTRAL UNMIXING AND IMAGE FUSION

Kaba, Eylem
Doctor of Philosophy, Geodetic and Geographic Information Technologies
Supervisor: Prof. Dr. S. Zuhal Akyürek
Co-Supervisor: Prof. Dr. Uğur Murat Leloğlu

March 2023, 93 pages

Soil maps are important data sources for many agricultural or environmental studies. Satellites and airborne platforms carrying hyperspectral sensors provide new possibilities for the estimation of soil properties. However, the main obstacle in soil classification with remote sensing methods is the vegetation whose spectral signature mixes with that of the soil. The objective of this thesis is to detect soil texture properties after eliminating the effects of vegetation using hyperspectral imaging data.

First, the endmembers common to all images and their abundances are estimated. Then the endmembers are classified as stable ones (soil, rock, etc.) and unstable ones (green vegetation, dry vegetation, etc.). The method eliminates vegetation from the images with orthogonal subspace projection and fuses multiple images with weighted mean for better signal-to-noise-ratio. Finally, the fused image is classified with the random forest technique to obtain the soil maps.

The method is tested on synthetic and real images in an area in Texas, USA. With three synthetic images, individual classification results are 81.78%, 79.84%, and

86.33%. After OSP, the rates increase to 85.70%, 88.21%, and 91.78%, respectively, while it increases to 91.85% with fusion. With real images from the dates 22/06/2013, 25/09/2013, and 24/10/2013, the classification accuracies increase from 70.51%, 68.87%, and 63.18% to 71.96%, 71.78%, and 64.17%, respectively. Fusion provides a better improvement in classification with 75.27% accuracy. The results show that the method can improve classification accuracy with the elimination of vegetation contribution. The approach is promising and can be applied to various other classification tasks.

Keywords: Soil Classification, Hyperspectral, Random Forest, Unmixing, Image Fusion

ÖZ

SPEKTRAL AYRIŞTIRMA VE GÖRÜNTÜ KAYNAŞTIRMA KULLANARAK UYDU-TABANLI ÇOK-ZAMANLI HİPERSPEKTRAL UZAKTAN ALGILAMA İLE TOPRAK SINIFLANDIRMASI

Kaba, Eylem
Doktora, Jeodezi ve Coğrafi Bilgi Teknolojileri
Tez Yöneticisi: Prof. Dr. S. Zuhal Akyürek
Ortak Tez Yöneticisi: Prof. Dr. Uğur Murat Leloğlu

Mart 2023, 93 sayfa

Toprak haritaları pek çok tarımsal ve çevresel çalışma için önemli bir veri kaynağıdır. Hiperspektral algılayıcılar taşıyan uydu ve hava platformları toprak özelliklerinin belirlenmesi için yeni olasılıklar sağlamıştır. Fakat uzaktan algılama yöntemleriyle toprak sınıflandırmasında en büyük zorluk spektral imzası toprakla karışan bitki örtüsüdür. Bu nedenle öncelikle yeşil ve kuru bitkilerin etkilerini minimize etmek gerekmektedir. Bu tezin ana amacı hiperspektral görüntü verisiyle toprak özelliklerinin, bitki örtüsünün elimine edilmesinden sonra belirlenmesidir.

İlk olarak, tüm görüntülerde ortak olan son üyeler ve onların çokluk haritaları bulunmuştur. Sonra, son üyeler stabil olanlar (toprak, kaya vb.) ve stabil olmayanlar (yeşil ve kuru bitki vb.) olarak ayrılmıştır. Bu yöntem ortogonal alt uzay izdüşümü ile bitki bileşenlerini görüntüden çıkarır ve daha iyi bir sinyal-gürültü oranı için ağırlıklı ortalama ile görüntüleri birleştirir. Sonuç olarak, toprak haritaları elde etmek için birleştirilmiş görüntü rassal orman tekniğiyle sınıflandırılır.

Bu yöntem sentetik görüntülerle ve Texas/USA yakınlarında bir alandan alınan gerçek görüntülerle test edilmiştir. Üç sentetik görüntüyle doğrudan yapılan

sınıflandırmada sınıflandırma sonuçları 81.78%, 79.84% ve 86.33% olarak bulunmuştur. OSP işleminden sonra ise sonuçlar füzyon için 91.85%, OSP için sırasıyla 85.70%, 88.21% ve 91.78% olarak artmıştır. Gerçek görüntülerle yapılan uygulama sonucunda da OSP ve füzyon işlemleri sınıflandırma sonuçlarında artış sağlamıştır. Sınıflandırma doğrulukları 22/06/2013, 25/09/2013 ve 24/10/2013 yılında alınan görüntüler için sırasıyla 70.51%, 68.87% ve 63.18%'den 71.96%, 71.78% ve 64.17%'e yükselmiştir. Füzyon işleminden sonra sınıflandırma sonuçlarında daha da iyileşme sağlanarak 75.27% sınıflandırma doğruluğu elde edilmiştir. Sonuçlar göstermiştir ki bitkilerin elimine edilmesi sınıflandırma doğruluğunda iyileşme sağlayabilmektedir. Bu yöntem ümit verici olup, diğer farklı sınıflandırma işlemlerinde de uygulanabilir.

Anahtar Kelimeler: Toprak Sınıflandırması, Hiperspektral, Rassal Orman, Ayrıştırma, Görüntü Birleştirme

To My Mom and Dad

ACKNOWLEDGMENTS

I would like to thank my family for their love, patience, and everything they gave me, particularly my brother M. Devrim Kaba, whose support and guidance always brighten my way. I especially thank Prof. Dr. Zuhal Akyürek, Prof. Dr. Afşar Saranlı, Prof. Dr. Yasemin Yardımcı Çetin, Prof. Dr. İlkay Ulusoy for their valuable suggestions and time, and I would like to thank examining committee members Prof. Dr. Nevsan Şengil, Assoc. Prof. Dr. Emre Sümer and Prof. Dr. Mehmet Lütfi Süzen for their valuable comments and time spared for me. I feel in debt to the Geodetic and Geographic Information Technologies Department professors of METU for great and valuable lectures. Last but not least, I would like to thank my supervisor Prof. Dr. Uğur Murat Leloğlu, for his guidance, advice, support, and encouragement throughout the research and for teaching me the notion of scientific thinking that I will keep and apply all fields in my life and for my lifetime.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

ABBREVIATIONS

AMEE: Automatic Morphological Endmember Extraction

CoNMF: Collaborative Nonnegative Factorization

C-SUnSAL: Constrained SUnSAL

DSMW: Digital Soil Map of World

DWT: Discrete Wavelet Transform

FAO: Food and Agriculture Organization of the United Nations

LDA: Linear Discriminant Analysis

MESMA: Multiple Endmember Spectral Mixture Analysis

MVSA: Minimum Volume Simplex Analysis

NDVI: Normalized Difference Vegetation Index

OSP: Orthogonal Subspace Projection

PV: Photosynthetic Vegetation

PPI: Pixel Purity Index

RF: Random Forest

RMSE: Root Mean Square Error

SAM: Spectral Angle Mapper

SCM: Spectral Correlation Mapper

SISAL: Simplex Identification via Split Augmented Lagrangian

SMACC: Sequential Maximum Angle Convex Cone

SNR: Signal-to-Noise-Ratio

SUnSAL: Sparse Unmixing by Variable Splitting and Augmented Lagrangian

SVM: Support Vector Machine

UNESCO: United Nations Educational, Scientific and Cultural Organization

USDA: United States Department of Agriculture

UTM: Universal Transverse Mercator

VCA: Vertex Component Analysis

VNIR: Visible and Near Infrared

VNIRA: Visible and Near Infrared Analysis

CHAPTER 1

INTRODUCTION

Classifying and mapping soil types provide very important data for a wide range of agricultural and environmental studies. There are many different methods and parameters which are used to classify soils. Traditionally, soil classification is realized by sample collection in the field and analysis in laboratories. Such costly processes, which require substantial time and effort, cause the studies to be held only in small areas or to generalize sparse samples to large areas.

Developments in remote sensing technology have enabled the use of information obtained by this technology to generate soil maps. In this way, producing soil maps on a global scale has been possible nowadays. Nevertheless, thus far, many of these maps are mainly based on multi-spectral images which have a low spectral resolution. However, the satellites and airborne platforms carrying hyperspectral sensors have higher spectral resolution with respect to multi-spectral imagery. Hyperspectral imagery has provided a new possibility to produce more qualified soil classification with this property. Yet, a main obstacle, even with hyperspectral data, is the vegetation cover of the interest area, which significantly reduces the performance of classification. In this study, an algorithm is described to classify soils on hyperspectral images by handling this problem.

1.1 Problem Definition

Soil maps are useful resources for many different purposes, such as land evaluation, agricultural planning, and environmental protection (Basil, 1971). Sometimes, even global maps are needed for earth system studies, such as the evaluation of the worldwide effects of a phenomenon. Local maps are generally created using field

surveys, and their classification system and their quality vary. Global scale maps, on the other hand, require many improvements. In global meaning, the Soil Map of the World at 1:5 000 000 scale was prepared with the efforts of the Food and Agriculture Organization of the United Nations (FAO) and the United Nations Educational, Scientific and Cultural Organization (UNESCO) (FAO, 2022h). Later, another project developing Soil and Terrain (SOTER) databases (ISRIC, 2022) at the 1:1 million scales was initiated by the FAO. But, this project succeeded only in some regions and countries, not globally. Nachtergaele and Van Ranst (2003) showed that less than half of the World has a complete, systematically published 1:1 million scale soil map. The rest of the area, which is equal to two-thirds of the land of the World, has not been mapped even at the 1:1 million scales. Over time, world soil maps started to have more complex soil patterns. Ongoing soil classification systems affect soil mapping methods (Hartemink et al., 2013). With the increasing temporal, spatial, and, most importantly, spectral resolution of satellite imagery, remote sensing is very promising for obtaining reliable global soil maps.

Hyperspectral imagery provides a better solution for soil classification with hundreds of spectral channels. NASA's Hyperion (2000-2017), HICO (2009-2014), and ESA's CHRIS (2001-cont.) are spaceborne hyperspectral sensors that have provided valuable data to researchers. HySIS (Hyper Spectral Imaging Spectrometer), an Indian earth observation (EO) satellite, was launched on 29 November 2018. PRISMA, an Italian earth observation satellite, was launched on 22 March 2019. HISUI, a Japanese spaceborne hyperspectral Earth imaging system, was launched on 5 December 2019, and EnMAP, a German hyperspectral satellite, was launched on 1 April 2022. Indian startup Pixxel is planning to launch a hyperspectral satellite, and more are in development, like NASA's HyspIRI (Hyperspectral Infrared Imager).

Although the spectral resolution of hyperspectral imagery brings advantages, vegetation is the biggest challenge in most parts of the World to classify soils. A method to "see through" vegetation cover is necessary for soil and possibly geological classification in semi-arid and humid parts of the World. One can benefit

from the fact that the vegetation cover changes with the phenology, and spaceborne imagers can periodically image any area to minimize the effect of vegetation.

1.2 Aim of the Thesis

Vegetation is considered as the main problem in the classification of soil. It is a necessity to remove vegetation effects from the images when using satellite imagery techniques in soil classification. Because vegetated areas cover the top of the soil by disabling soil reflection or causing mixed pixel reflection. The aim of this study is to remove vegetation effects from the hyperspectral images as much as possible and to fuse information from multiple images before soil classification and hence to improve the classification performance of soils.

1.3 Approach of the Thesis

In this study, an algorithm to detect and eliminate vegetation components from the hyperspectral image is described. The SMACC (Sequential Maximum Angle Convex Cone) unmixing method is used to detect endmembers. First of all, a large number of endmembers are detected using the pixels from images of the study area taken at different times; so that all images are described in terms of the same basis. Then, dominantly pure pixels for major land cover types, such as soil, green vegetation, dry (non-photosynthetic) vegetation, and water, are selected manually from appropriate images. Each endmember is associated with one of the chosen material types using the spectral angle mapper (SAM) distance metric. The endmembers are considered in two different groups: stable and unstable material types. The unstable components are material types, like green and dry vegetation, because vegetation spectra change throughout the year. Stable components are, for

example, soil and rock, which have almost constant spectra during the year. Although the relative abundances of stable components are almost constant, their absolute abundances change along with the unstable components. The images are projected to a subspace perpendicular to that defined by the unstable components. Since the remaining parts can be noisy, multiple images are fused for a better signal-to-noise ratio (SNR). Resultant images are used for soil classification using the random forest (RF) algorithm. Then the label relaxation algorithm and a post-filter are applied to the classification result. The algorithm is first tested on synthetic images; then, it is tested in an area in Texas, where Hyperion images and soil maps are available. The results are promising. The method can be used for the accomplishment of accurate global soil maps with oncoming hyperspectral satellites.

1.4 Contributions of the Thesis

The contributions of this thesis to the state-of-the-art can be summarized as follows:

- Elimination of vegetative components from hyperspectral images by
    - Spectral unmixing,
    - Classification of endmembers as stable and unstable, and
    - Orthogonal subspace projection.
- A fusion method for combining multiple filtered images to obtain a single image with a better signal-to-noise ratio.
- An overall method for classification of soils using the contributions above.

1.5 Organization of the Thesis

This thesis is organized as follows. Section 2 explains the background of soil properties, spectral properties of soil, and common land covers. Section 3 examines

previous work on hyperspectral image classification of soils. Section 4 describes the synthetic data, the study area, and other data used in this paper. Section 5 explains the proposed method in detail, while Section 6 presents the results and discussion. Finally, Section 7 exhibits the conclusions of the thesis.

CHAPTER 2

BACKGROUND

2.1 Properties of Soils

Soil is formed by disintegrated parts of rocks and decomposed organic materials (Young, R., 2012). According to FAO (FAO, 2022e), soil has three groups of properties: physical, chemical, and biological. All these properties of the soils determine their quality level for agricultural purposes.

Physical properties are identified with soil structure, soil depth, soil water characteristics, available soil water, soil texture, color, consistency, porosity, density, and water flows (FAO, 2022d). When single particles of soil components (sand, silt, and clay) assemble, they form larger particles called aggregate. These aggregates can have different patterns forming different soil structures (FAO, 2022f). Soil depth is about the thickness of the solum. Soil water characteristics are related to soil moisture storage. Available soil water gives the soil ability to hold water. Soil texture is determined by the ratio of silt, sand, and clay in it. Soil texture is related to soil behavior, such as its capacity of holding nutrients and water. Color provides auxiliary information about soil. For example, it helps to distinguish soil boundaries in a profile. Because the color of soil changes with organic content, drainage level, and oxidation degree. Consistency shows the soil's ability to resist fragmentation and shows how it is successful in sticking together. Porosity determines space in the soil that is filled by water or air. The ideal ratio is 50% in total within the soil. Density defines the density of mineral particles after removing pore space and organic material in the soil. Water flows are determined by soil drainage class. It is a concept affected by gravity, osmosis, and capillarity (FAO, 2022d).

Chemical properties are identified with Cation Exchange Capacity (CEC), soil pH, base saturation percentage, plant nutrients, organic soil carbon, soil nitrogen, soil salinity, soil sodicity, calcium carbonate content, and calcium sulfate (Gypsum) content (FAO, 2022b). CEC is related to the soil's capability of holding cations. It works as the measurement of fertility, nutrient containment ability, and ability of groundwater protection from cation contamination. Soil pH explains soil properties as being acidity or alkalinity. Soils having a large amount of acidity property can be toxic because of the aluminum level in them. Soils having a large alkalinity can come apart easily. Base saturation percentage determines the ratio of base-forming cations like sodium, calcium, potassium, and magnesium in the soil. Plant nutrients include two categories: one is macro-nutrients, and the other is micro-nutrients. These two groups cover 16 vital nutrients which are essential for plants and organisms in the soil. Organic soil carbon is the leftovers of dead plants' roots and leaves. It provides an important contribution to soil health by improving soil physical properties and by enhancing cation exchange and water holding capacity. Soil nitrogen is an important factor in the soil. It contributes to plant growth. If it is deficient, it affects plant growth in a negative manner. Soil salinity affects soil quality and plants. When salinity rises in the soil, it may cause soil disintegration and plant decay. Soil sodicity causes weakness in soil's physical and chemical properties. Sodicity is mainly seen in the soils of arid or semi-arid regions. Excess sodium clogs soil pores and disables water-holding capacity, which affects plant growth. Calcium carbonate content is another factor for soil quality. If calcium carbonate is in fair amounts in the soil, it is good for the soil structure. Because a moderate amount provides correction for soil pH. If calcium carbonate is too much, it may not be absorbed by plants either. Soils having a high amount of gypsum can be seen in places having dry climate. The plants like ephemeral grasses and xerophytic shrubs are among the most seen plants in those places (FAO, 2022b).

Soil biological properties determine the microbial and faunal activities that occur in soil. It has an important effect on soil characteristics. (FAO, 2022a).

2.2 Soil Texture

Basically, the soil is made up of clay, silt, sand, and gravel (Gardner et al., 1999). Those are formed from rock particles over the years under various climatic and environmental conditions. The main difference between clay, silt, sand, and gravel is their particle sizes. According to the USDA, clay particle size is less than 0.002 mm, silt particle size is between 0.002 and 0.05 mm, sand particle size is between 0.05 and 2 mm, and gravel particles are bigger than 2 mm (FAO, 2022g). United States Department of Agriculture (USDA) describes soil texture in twelve major classes, which are sands, loamy sands, sandy loams, loam, silt loam, silt, sandy clay loam, clay loam, silty clay loam, sandy clay, silty clay, and clay. A soil sample is determined in one of these soil texture classes according to the proportions of sand, silt, and clay in it, which are located along each of the three axes (Figure 2.1).

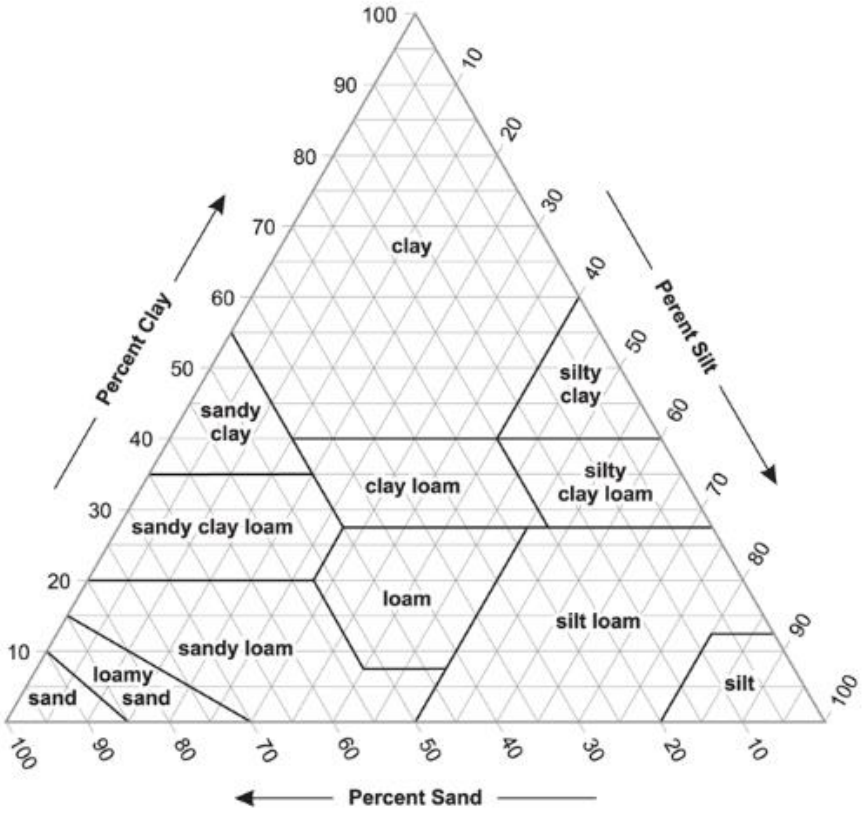

Figure 2.1 USDA textural triangle (Soil Science Division Staff. 2017)

In the sands texture class, there should be more than 85% sand, and the sum of the percentage of silt and 1.5 times the percentage of clay should be less than 15. Sub-classes of sand texture class are coarse sand, sand, fine sand, and very fine sand.

The loamy sands texture class has 70% - 90% of sand; the sum of the percentage of silt and 1.5 times the percentage of clay should be 15 or more, and the sum of the percentage of silt and twice the percentage of clay should be less than 30. Sub-classes of the loamy sands class are loamy coarse sand, loamy sand, loamy fine sand, loamy very fine sand.

The sandy loams texture class covers between 7% and less than 20% clay and more than 52% sand, and the sum of the percentage of silt and twice the percentage of clay should be 30 or more, or it should have less than 7% clay and less than 50% silt, and the sum of the percentage of silt and twice the percentage of clay should be 30 or more. Sub-classes of this texture class are coarse sandy loam, sandy loam, fine sandy loam, very fine sandy loam.

The loam texture class covers clay between 7 and less than 27% in amount, silt between 28 and less than 50% in amount, and 52% or less sand.

The silt loam texture class covers 50% or more silt and clay between 12% and less than 27% in amount, or it should have silt between 50% and less than 80% in amount and clay less than 12%.

The silt texture class includes 80% or more silt and less than 12% clay.

The sandy clay loam texture class includes clay between 20% and less than 35% in amount, less than 28% silt, and more than 45% sand.

The clay loam texture class includes clay between 27% to less than 40% in amount and more than 20% to 45% sand.

The silty clay loam texture class includes clay between 27% and less than 40% in amount and 20% or less sand.

The sandy clay texture class includes 35% or more clay and more than 45% sand.

The silty clay texture class covers 40% or more clay and 40% or more silt.

Finally, the clay texture class includes 40% or more clay, 45% or less sand, and less than 40% silt (Soil Science Division Staff. 2017).

2.3 Spectral Properties of Common Land Covers

Land covers have different spectral characteristics. Many different properties, such as structure, texture, and water capacity of them, create these differences that help us to differentiate land covers (Figure 2.2). This section discusses the spectral properties of the most common materials, like soil, vegetation, water, and snow. Besides them, cloud spectra, which are confused with snow, are examined.

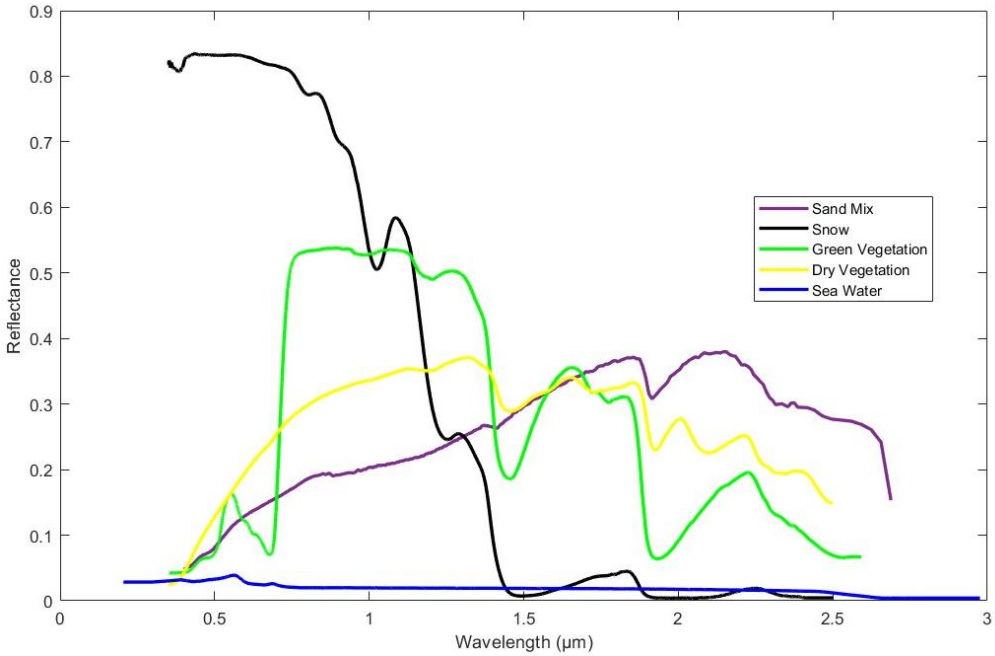

Figure 2.2 Spectral reflectance signatures of some common land covers (data is from USGS Spectral Library V7, Kokaly et al., 2017)

2.3.1 Soil spectra

Soil spectral reflectance is affected by many properties of the soil. Surface roughness, moisture content, organic matter content, texture, mineralogical composition, carbonate content, total iron, and water-soluble salts are among these properties (Prudnikova et al., 2019). Soil chemical elements like organic or mineral ingredients, in addition to soil water, can also absorb the incident light at specific wavelengths. The soil particle size is another feature that affects the soil spectral reflection in the visible and near-infrared (VNIR) band. Generally, when particle size is decreased, it causes an increase in scattering, and because of that, soil reflectance increases (Figure 2.3). Therefore, different soil textures cause different spectra (Soltani et al., 2019).

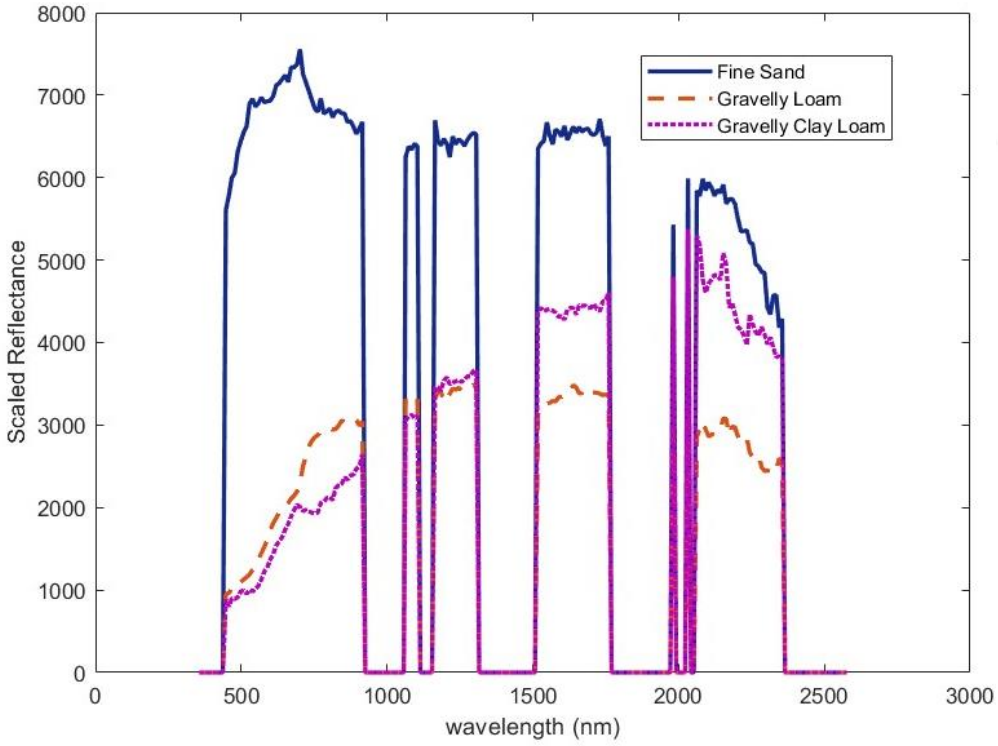

Figure 2.3 Reflectance spectra of different soil textures from the study area described in Section 4.1

Another factor is iron oxides in soil. Iron oxides can be detected around 900 nm in the spectrum because of absorption (Figure 2.4).

Roughness also highly affects soil spectra. Because micro-shades cause a decrease in reflectance (Escribano et al., 2017).

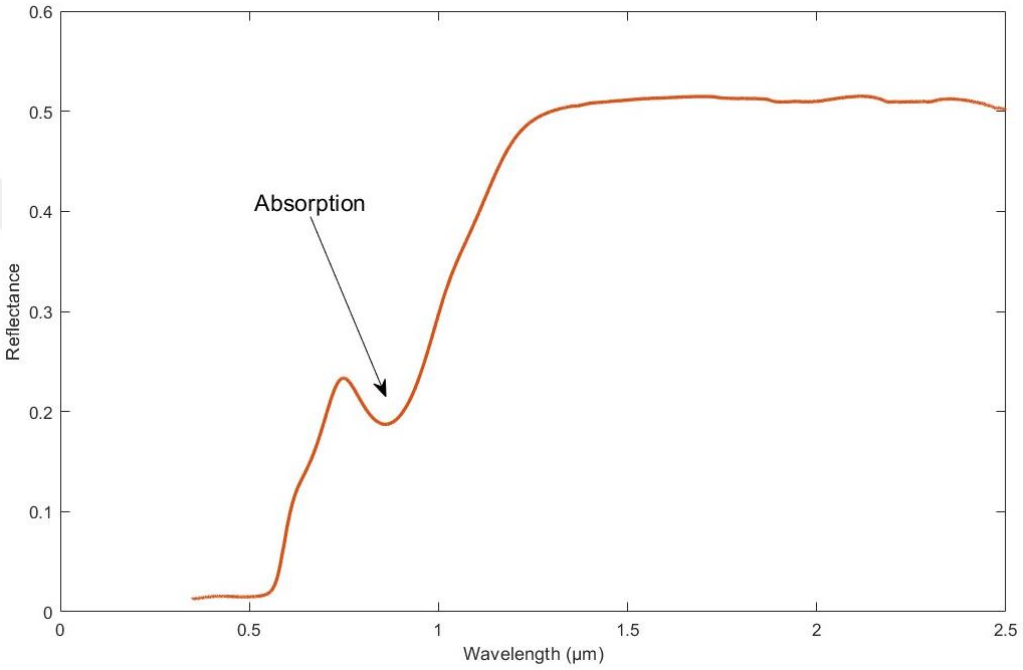

Figure 2.4 Absorption in Iron Oxide spectra around 900 nm (data is from USGS Spectral Library V7, Kokaly et al., 2017)

The presence of water absorption bands at around 1400 nm and 1900 nm; and weaker absorption bands at around 950 nm and 1200 nm also affects soil reflectance (Figure 2.5). Water reduces the reflectance in these regions (Ben-Dor et al., 1999).

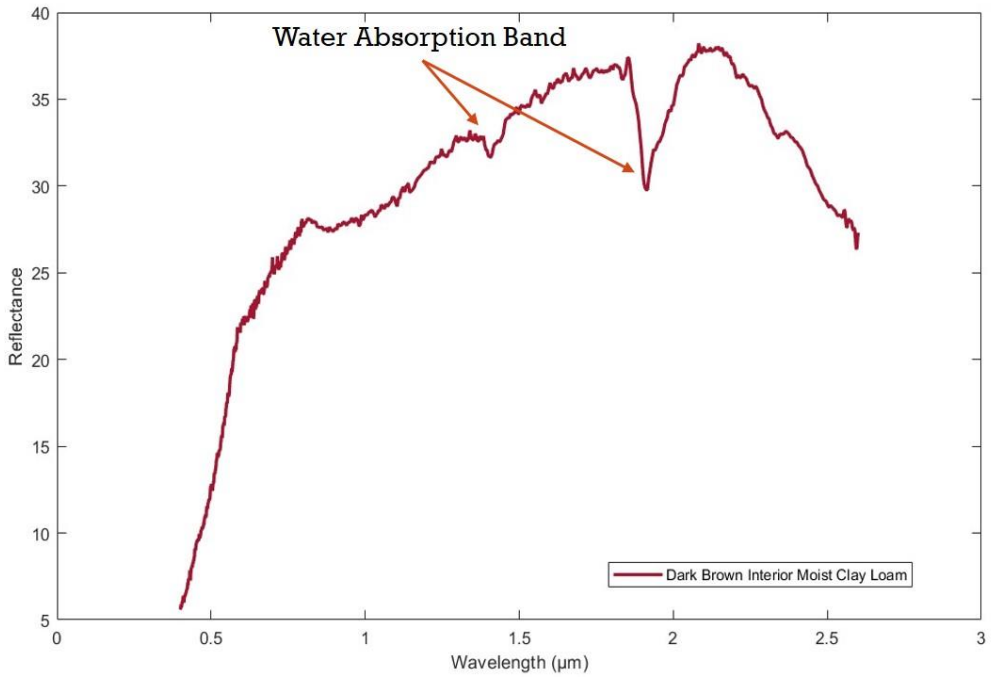

Figure 2.5 Reflectance of the clay loam soil with moisture (data is from ECOSTRESS spectral library, Baldridge et al., 2009, Meerdink et al.,2019)

All these properties also help to distinguish the spectra of different soil types. For example, clay (<0.002 mm) has lower reflectance than sand or silt (>0.002 mm) and appears darker than others in general. The water-holding capacity and higher content of organic matter in clays can cause this (Summers et al., 2011).

2.3.2 Vegetation Spectra

Vegetation has a unique spectral signature that can be differentiated easily from the other land cover types. There are three important features that affect the vegetation spectra in the visible, near-infrared, and shortwave-infrared regions of the spectrum: leaf pigments, cell structure, and water content (Navalgund, R.R., 2001).

Because of the absorption of chlorophyll in the visible region around blue (400-500 nm) and red (600-700 nm), vegetation has low reflectance. Reflection in the green region (500-600 nm) has a peak (Figure 2.6).

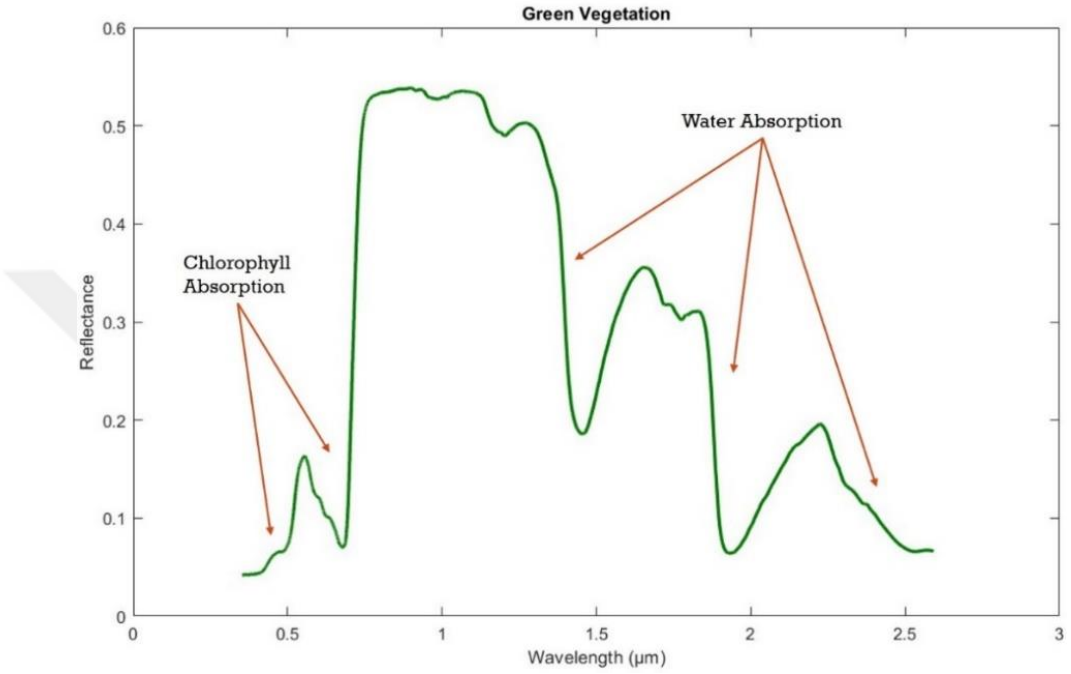

Figure 2.6 Reflectance of green vegetation (data is from USGS Spectral Library V7, Kokaly et al., 2017)

The cellular structure of leaves causes much higher reflectance in the near-infrared region than in the visible region (Ashraf et al., 2011). While leaves are growing, inter-cellular air spaces also expand, causing an increase in reflection. If vegetation gets older or stressed, chlorophyll absorption and inter-cellular air spaces decrease, causing a decrease in reflectance in the near-infrared band (Navalgund, R.R., 2001). The leaf water absorption is powerful in the shortwave-infrared region (1300–2500 nm). When leaf water content decreases, the reflectance increases; when water content increases, the reflectance decreases, especially around 1400 and 1900 nm. Because green vegetation spectra in shortwave infrared region decrease.

Non-photosynthetic vegetation, like grass litter, also has unique spectral signatures related to its decomposition stage. While the chlorophyll pigments are decreasing, the vegetation becomes yellowish, and the reflection in the red region increases. When degradation gets higher, the blue reflectance increases and causes a gray appearance in grass litter (Huete, 2004).

2.3.3 Water Spectra

Water has a high transmittance, and it has low absorbance and reflectance in the visible region of the spectrum. While the absorbance of water increases in the near-infrared region, the reflectance and transmittance of water decrease (Figure 2.7) (Jain & Singh, 2003).

Water reflectance is also affected by the water surface, bottom material, and suspended materials found in it. The turbidity of water is a reason for an increase in reflectance. When the chlorophyll amount gets higher, greater absorption can be seen in the blue and red regions of the spectrum (Navalgund, R.R., 2001).

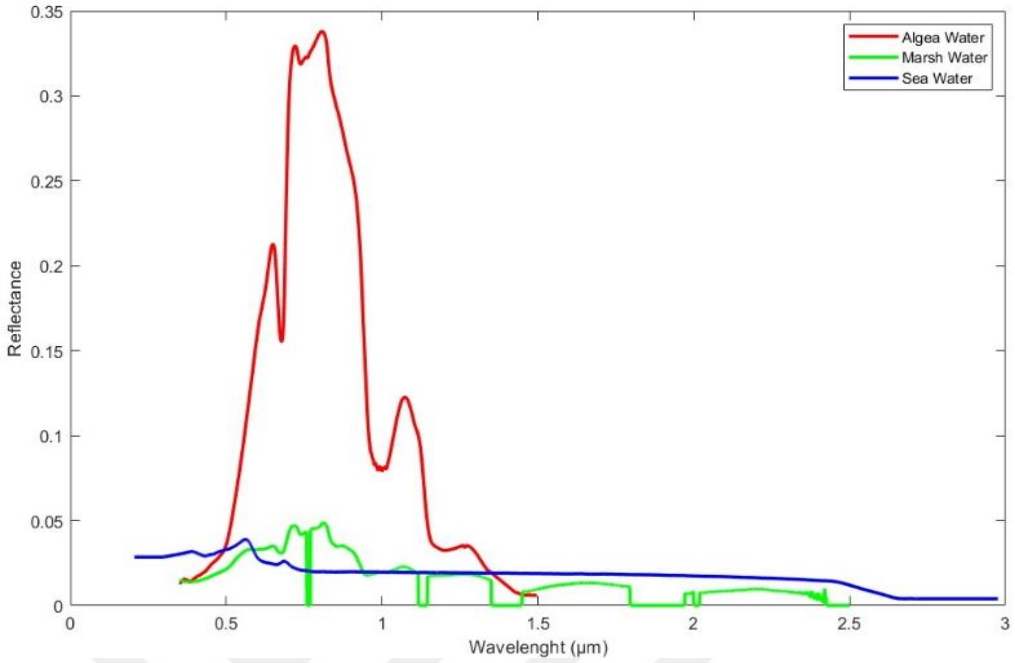

Figure 2.7 Water reflectance spectra (data is from USGS Spectral Library V7, Kokaly et al., 2017)

2.3.4 Snow and Cloud Spectra

Snow shows high reflectance until 800 nm wavelength in the spectrum, and then it starts to decrease. Because of non-selective scattering, the clouds appear bright between 300 nm to 3000 nm wavelengths.

Atmospheric windows in the shortwave infrared regions of 1550 nm to 1750 nm and 2100 nm to 2350 nm make snow and cloud separation possible (Navalgund, R.R., 2001).

CHAPTER 3

LITERATURE REVIEW

Hyperspectral sensors provide a valuable data source for many applications, such as mining, geology, environment, and agriculture. In this perspective, hyperspectral images are also used for soil classification. Recent studies on soil classification and vegetation cover from hyperspectral images are mainly based on images acquired by airborne platforms such as DAIS 7915, PROBE, HyMap, and AVIRIS.

Evaluating hyperspectral images mostly starts with unmixing algorithms, which is required to solve the mixed pixel problem. Many methods have been suggested for the solution to the problem.

This section examines the literature in two parts: the first is on reviewing studies about unmixing, and the second is about studies on soil classification.

3.1 Spectral Unmixing

Analyzing hyperspectral images generally begins with spectral unmixing. Because, most of the time, a pixel includes reflections from multiple materials. Many unmixing methods have been suggested during the last decades. Nascimento & Dias (2005) proposed vertex component analysis (VCA) method that performed unsupervised endmember extraction from hyperspectral data. The method iteratively projected data to an orthogonal direction of the subspace of endmembers until all endmembers were used up. Martínez et al. (2006) compared the pixel purity index (PPI), N-FINDR, and automatic morphological endmember extraction (AMEE) algorithms to find out appropriate endmembers for real hyperspectral images. Gillis et al. (2007) studied a linear unmixing model. In this model, hyperspectral spectra

were separated as a sum of components. As a different concept in the model, each material in the component was represented by multidimensional subspaces rather than a single vector. Gillis et al. (2008) continued to enhance the previous study using a combination of multidimensional affine endmember subspaces instead of the common linear mixing model with a new demixing algorithm. Bioucas-Dias J. (2009) suggested a linear hyperspectral unmixing method of the minimum volume class, called simplex identification via split augmented Lagrangian (SISAL). The hyperspectral linear unmixing is equal to the detection of the minimum volume simplex, including the hyperspectral vectors, which is actually a nonconvex optimization problem with convex constraints. The positivity constraints were replaced by soft constraints using a sequence of augmented Lagrangian optimizations. Bioucas-Dias & Figueiredo (2010) studied convex optimization problems in hyperspectral unmixing. In the study, two new algorithms were developed to solve the optimization problems efficiently depending on the alternating direction method of multipliers (a method from the augmented Lagrangian family). The algorithms were named sparse unmixing by variable splitting and augmented Lagrangian (SUnSAL) and constrained SUnSAL (C-SUnSAL).

Some of the unmixing methods built algorithms depending on the existence of pure pixels on the image and some of them depend on the assumption of the absence of pure pixels on the image. Li et al. (2015) developed minimum volume simplex analysis (MVSA) algorithm. The method considers the cases where pure pixels do not exist in the hyperspectral data. The algorithm fits a minimum volume simplex to the hyperspectral data and constrains the abundance fractions to the probability simplex to solve the unmixing problem. Contrary to the previous method, Li et al. (2016) developed the collaborative nonnegative factorization (CoNMF) algorithm, which assumes the existence of pure pixels in the scene. The method uses a mixed-norm regularizer. The main goal of this method is to provide the actual endmember contribution to abundances.

Besides linear mixture models, nonlinear algorithms were also used for unmixing problem. The need for nonlinear methods emerges from real-world situations like a scene having complex vegetated surfaces with multiple light scattering (Nascimento et el., 2019). Many methods were suggested for nonlinear unmixing, such as kernel methods, neural networks, support vector machine (SVM), etc. A detailed review on frequently used methods was published by Heylen et al. (2014).

3.2 Literature on Soil Classification

Hyperspectral imagery is used for the evaluation of different land cover materials. There are many recent reviews about the studies, which include the methods using hyperspectral imaging technology. Peyghambari & Zhang (2021) reviewed current techniques for geological information extraction with hyperspectral imagery. Atmospheric correction, dimensionality processing, endmember estimation, and image classification were identified as major steps. In this perspective, hyperspectral images were also used for soil classification: Diaz-Gonzalez et al. (2022) reviewed recent studies on crop yield prediction by analyzing remote sensing images with machine learning techniques to find chemical, physical, and biological soil quality indicators. Tziolas et al. (2021) reviewed the research, which covered a three-year period (2019–2021), on soil mapping and monitoring with spaceborne and aerial earth observation techniques. Scaling, resolution, data characteristics, and modeling approaches in recent articles were collected and encapsulated in the study to summarize earth observation methods and their usage in soil monitoring.

Recent studies on soil classification and vegetation cover from hyperspectral images have mostly been based on images acquired by airborne platforms such as DAIS 7915, PROBE, HyMap, and AVIRIS. Hyperspectral image classification with Orthogonal Subspace Projection (OSP) was implemented in (Harsanyi & Chang, 1994). Basically, the suggested method projects each pixel vector onto an orthogonal

subspace to undesired signatures. The method reduces data dimensionality according to the number of desired endmembers and produces a new sequence of images that reveals the presence of each signature of interest. These new component images represent class maps for each of the signatures of interest. The OSP method in the study was used on two sets of simulated data and also one real image from AVIRIS at Lunar Crater Volcanic Field in Northern Nye County, Nevada, USA. Okin et al. (2001) aimed at evaluating hyperspectral remote sensing data for detecting vegetation cover, vegetation type, and soil type in areas of low vegetation cover. Multiple endmember spectral mixture analysis (MESMA) was applied to AVIRIS data which were acquired over Manix Basin on April 30, 1998. The study showed that MESMA mapped soil types successfully, although vegetation types could not be found by the algorithm for the areas having low vegetation cover. Margate & Shrestha (2001) used HyMap data to classify the desert-like soil surface of their study area at Tabernas, Spain, employing the SAM technique and comparing the image reflectance spectra with reference spectra that were extracted from the images or obtained from a spectral library. Asner & Heidebrecht (2002) aimed to retrieve photosynthetic vegetation (PV), non-photosynthetic vegetation, and bare soil with spectral mixture analysis. A probabilistic spectral mixture model, AutoMCU, was developed for this study and tested with different schemes from AVIRIS data in the Chihuahuan Desert, New Mexico, USA. Nanni et al. (2021) studied to find particle size and soil organic matter of tropical soils in São Paulo/Brazil by using airborne hyperspectral imagery in the field and non-imaging hyperspectral sensors for soil samples in the laboratory. Sand, silt, clay, and soil organic matter were predicted by partial least square regression method.

Portable spectrophotometers were also used for the research on soil properties. Dhawale et al. (2022) compared two portable spectrophotometers according to the prediction of main soil properties like texture (sand and clay) and soil organic carbon (SOC) with 282 soil samples from four different Canadian provinces.

Soil properties were also the interest of hyperspectral remote sensing studies. Ben-Dor et al. (2002) produced quantitative soil property maps of organic matter, soil

field moisture, soil saturated moisture, and soil salinity using DAIS-7915 airborne hyperspectral sensor data. The visible and near infrared analysis (VNIRA) approach was used to produce a prediction model for each soil property. The results were validated with data from several sampling points. Then, a procedure was developed to map the study area covering soils under vegetated areas. Forkuor et al. (2017) used data from RapidEye and Landsat satellites, terrain/climatic measurements, and results of soil sample analysis in the laboratory to expose the spatial distribution of cation exchange capacity, clay, silt, sand, soil organic carbon, and nitrogen in an area in Burkina Faso. Support vector machine, RF regression, multiple linear regression, and stochastic gradient boosting were used to map the six properties, and then the results were compared. Meng et al. (2021) used Gaofen-5 hyperspectral images for predicting soil organic matter from spectral indices and image bands via the RF algorithm after denoising and recursive feature selection. Pandey et al. (2021) classified soil types depending on their physical and chemical properties by satellite images using the deep convolutional neural network method. LANDSAT-8 satellite images were used to search for four soil types: alluvial, black, desert, and red soils.

Many methods were suggested for soil classification using hyperspectral images. Lumme (2004) compared multiple methods to classify soil from hyperspectral images, namely, SAM, spectral correlation mapper (SCM), and spectral unmixing. Also, minimum distance and maximum likelihood classifiers are involved in the comparison of classifications. The data were obtained from the AISA airborne imaging spectrometer in Southern Finland. The classification methods maximum likelihood, SAM, and SCM yield good results. The classification result of the minimum distance method is poor. Spectral unmixing classification only worked on the parts where essential spectra are found. Because it is difficult to find suitable reference spectra for mixed pixels. Zhang et al. (2003) developed a soil texture classification system in the hyperspectral domain using the discrete wavelet transform (DWT) for feature extraction and the linear discriminant analysis (LDA) for feature vector reduction and optimization. The algorithm, aimed at detecting three main texture components of soil: sand, silt, and clay, was tested on the

hyperspectral data collected by a spectroradiometer at the Mississippi University North Farm. 25 Sample spectra of three components are collected as training data. The linear mixture model is used to determine soil textures having different abundance values of the three components. The experimental results showed that the suggested method is useful, but it needs more training data to become robust. Zhang et al. (2005) presented an automatic soil texture classification system with the help of hyperspectral soil signatures and wavelet-based statistical models. The maximum likelihood and hidden Markov models were used for the classification of data collected by a spectroradiometer. Vibhute et al. (2015) classified soils using the support vector machine method. Hyperion Hyperspectral satellite images from Maharashtra/India were used to classify five different soil types. Li et al. (2021) aimed to find the applicability of the convolutional neural network algorithm for land cover classification with small samples from six different places in Qingdao, China. Spectra of soil samples were collected using a spectrometer. Meng et al. (2022) used digital elevation maps and Landsat time series along with Gaofen-5 satellite data for classifying soils into three major groups, namely Phaeozems, Chernozems, and Cambisols, via RF classification. Rajitha et al. (2022) conducted a survey on the studies on the classification of soil types obtained by hyperspectral imaging to expose the benefit of that technology. Samadi & Samadi (2022) worked with machine learning algorithms for soil classification. For this purpose, components of clayey fine sand and coarse sand soils having 34 soil samples were used and classified with two machine learning classifiers: The Naïve Bayes and the artificial neural network model. The suggested models improved classification accuracy and also helped to reduce the cost of classification. Uddin & Hassan (2022) suggested a computer vision-based method for an automated soil classification system to predict soil types. The algorithm combined quartile histogram-oriented gradients, the most frequent φ-Pixels, and a new feature selection method.

The studies about soil classification of hyperspectral images focus on different properties of soil, such as chemical properties, organic matter or texture, etc. However, one of the critical obstacles to the classification of soil on a local or global

scale is vegetation covering the soil. An automated method that detects and eliminates the vegetation and only uses the target component for classification is a necessity.

CHAPTER 4

STUDY AREA AND DATA

4.1 Synthetic Data

The algorithm is tested using three different synthetic images. Endmembers, which are obtained from real data, are used to create the synthetic data (Figure 4.1). The endmembers include three types of soil (fine sand, gravelly loam, and gravelly clay loam), green vegetation, and dry vegetation. For more realistic results, synthetic images are created for three seasons: spring, summer, and autumn. Seasonal effects are applied to the images considering vegetation cover. For example, spring has more green vegetation cover; summer has more dry vegetation cover than the other seasons; and bare soil is more visible in the autumn season. Abundance maps are created taking vegetation cover of seasons into account (Figure 4.2). The possibility of pure pixels of soil/non-soil components in the images is also considered. One small pure pixel area for each endmember is placed for in all the abundance images. Besides all seasonal effects, endmember variability is also considered to create spectral differences of an endmember spectrum throughout the images. Environment and illumination of materials, atmospheric conditions, and temporal situations cause variability of spectral behavior throughout a hyperspectral image. Considering this effect on the images provides more accurate endmember estimation (Zare & Ho, 2013). All endmembers, which are used to create synthetic images, are multiplied by a polynomial in each pixel to create the variability. The polynomial is structured as:

$$\boldsymbol{Pl}(\boldsymbol{\lambda}) = (\boldsymbol{a\lambda} + 1 + \boldsymbol{b}) \tag{4.1}$$

where $\boldsymbol{a}$ is a normally distributed zero mean random variable with standard deviation of 3,

$\boldsymbol{b}$ is a normally distributed zero mean random variable with standard deviation of 0.3, and $\boldsymbol{\lambda}$ is the wavelength

Endmember in each pixel after added variability is:

$$\boldsymbol{E'} = \boldsymbol{E}(x, y) * \boldsymbol{Pl} \tag{4.2}$$

Each image is formed using endmember variability:

$$\boldsymbol{A} = \boldsymbol{A} + \boldsymbol{ab} * \boldsymbol{E'} \tag{4.3}$$

where $\boldsymbol{ab}$ contains the abundance values of the endmembers.

Figure 4.3 (a) and (b) show the variability of soil endmember gravelly clay loam and green vegetation throughout the synthetic images.

Also, noise is added to the images to see the effect of the method:

$$\boldsymbol{A} = \boldsymbol{A} + \boldsymbol{n}(x, y) \tag{4.4}$$

where $\boldsymbol{n} \sim \boldsymbol{N}(0,\ 5^2)$.

Figure 4.4 and Figure 4.5 exhibit the images before and after the noise is added to the images.

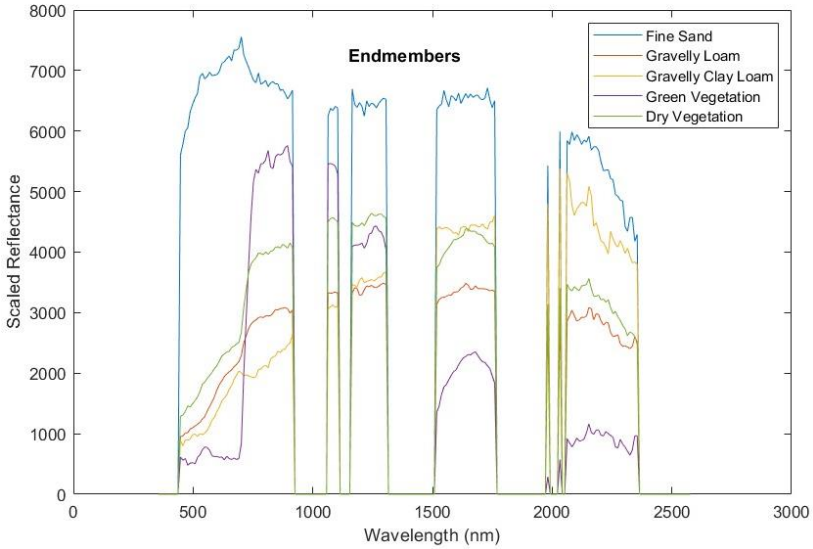

Figure 4.1 Endmembers are used to create synthetic images

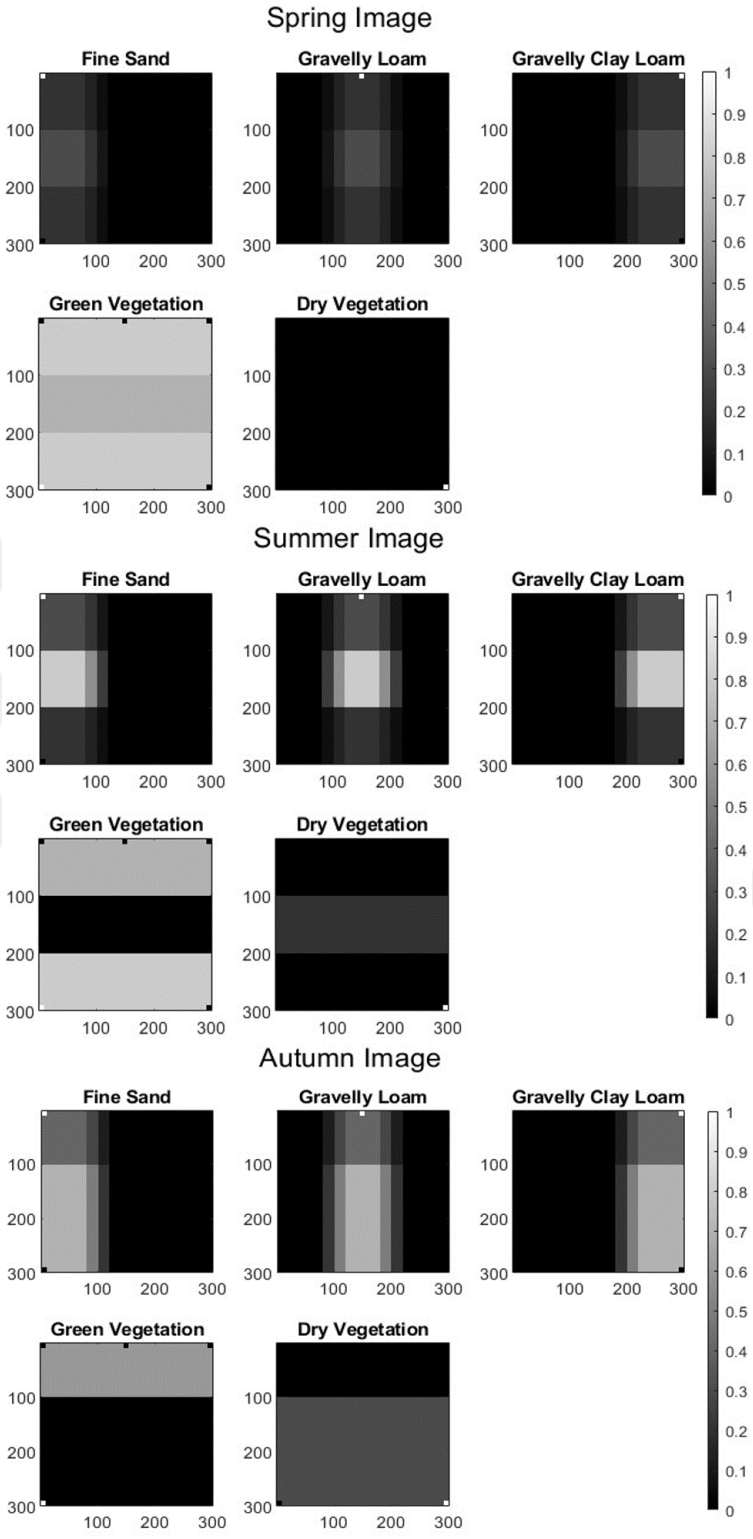

Figure 4.2 Initial abundance maps of synthetic images

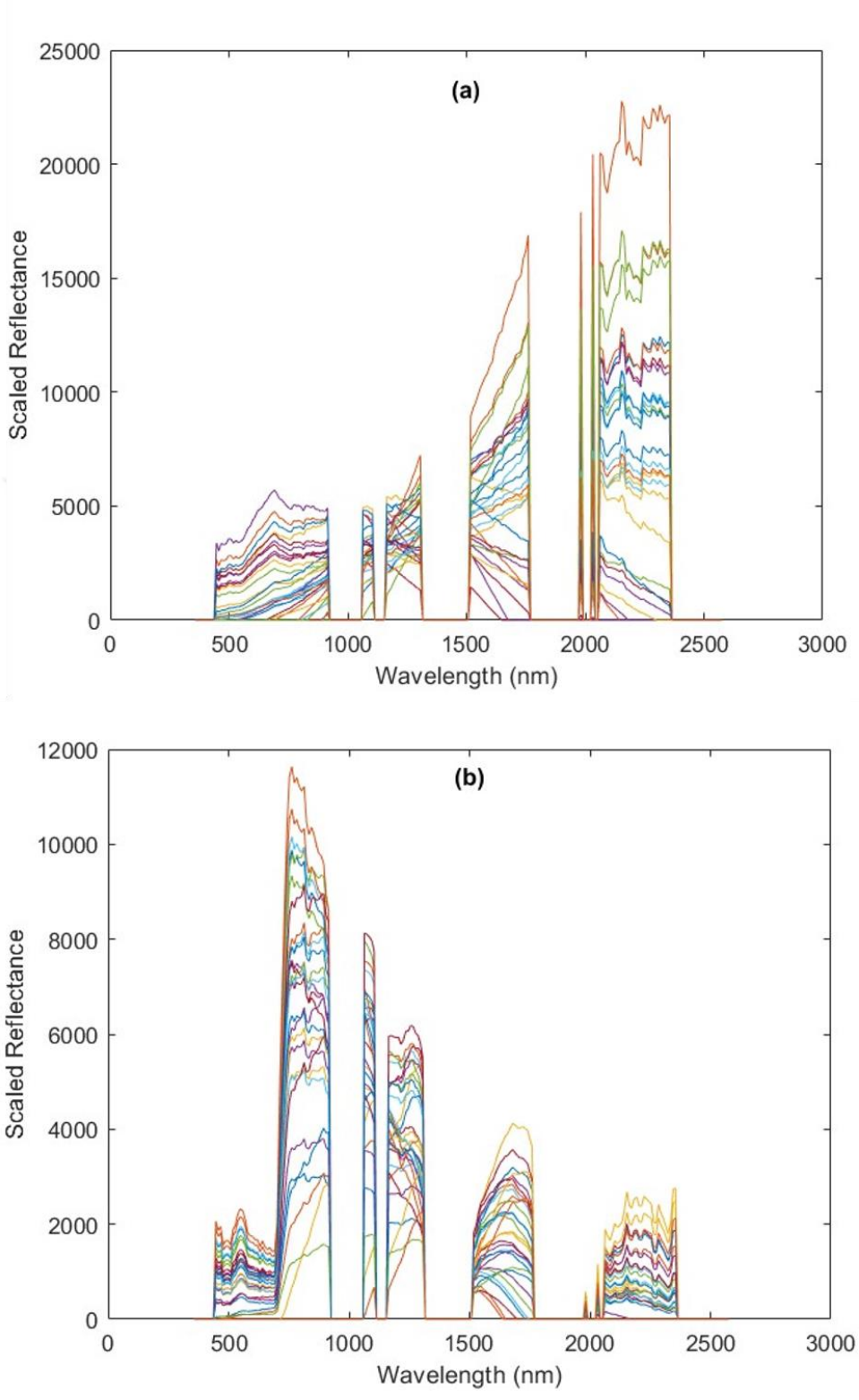

Figure 4.3 (a) Variability of gravelly clay loam endmember (b) Variability of green vegetation endmember

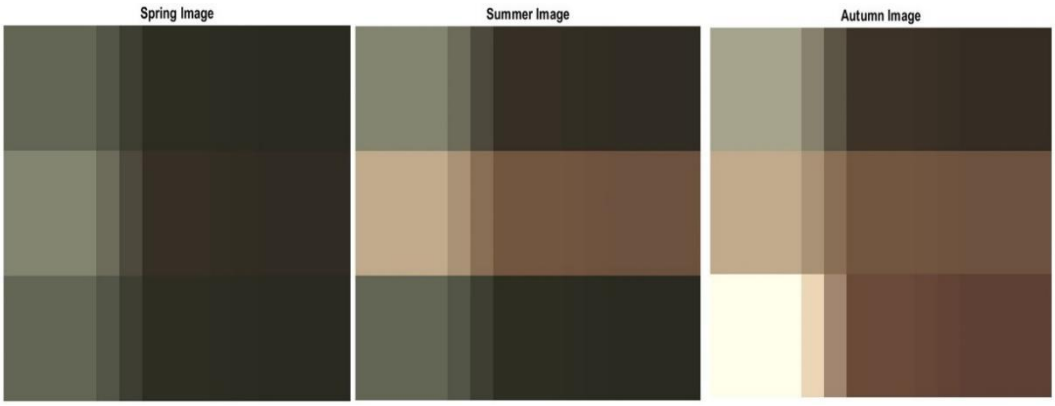

Figure 4.4 Synthetic images without endmember variability and noise (Bands 20, 11 and 3 are used to represent RGB)

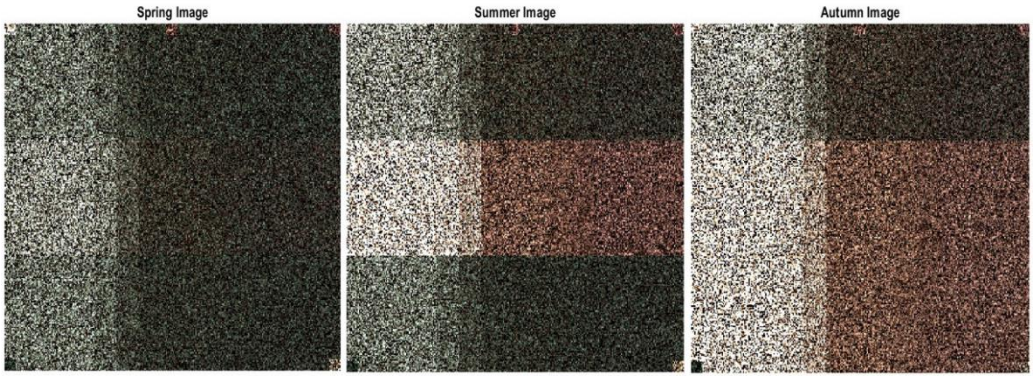

Figure 4.5 Synthetic images with endmember variability and noise (Bands 20, 11 and 3 are used to represent RGB)

4.2 Real Data

An area in Texas State (Figure 4.6) in the US is chosen for testing the algorithm, where soil maps and Hyperion images are all available, and there is considerable vegetation cover. Three Level 1R images from the dates 22/06/2013, 25/09/2013,

and 24/10/2013 in Texas were chosen for further analysis. The data is served in Geo-TIFF format by USGS to the users.

Hyperion was one of the sensors of NASA's Earth-Observing One (EO-1) satellite. The satellite was launched on November 21, 2000, and continued its mission until 2017. The Hyperion provides high-resolution hyperspectral images with 220 spectral bands ranging from 0.4 to 2.5 μm and with a 30-meter ground resolution and 10 nm spectral resolution. One image which is obtained from the instrument covers a 7.5 km by 100 km land area, and it also provides detailed spectral data throughout all 220 channels with high radiometric accuracy.

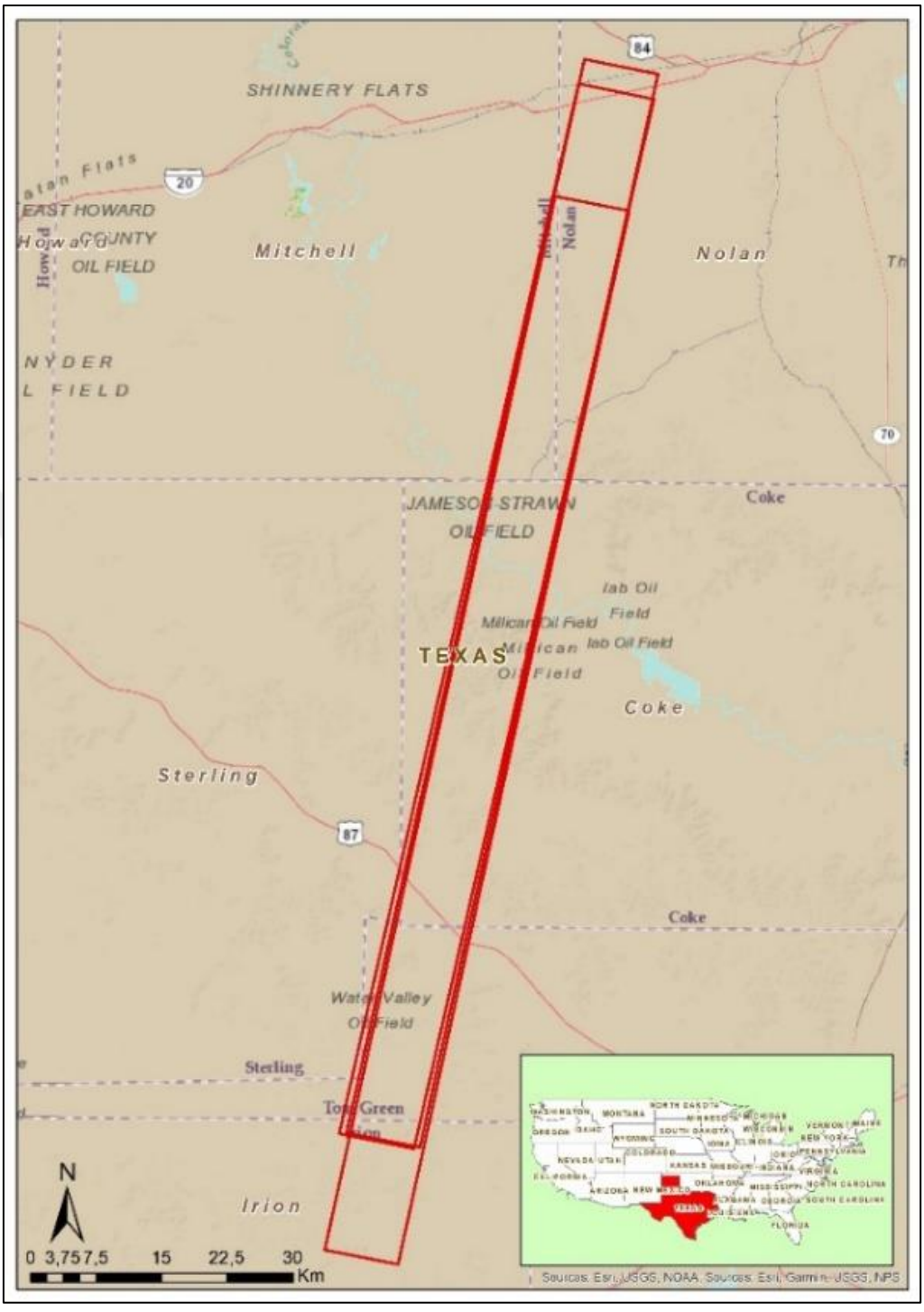

Figure 4.6 The study area in Texas (Base map source: U.S. Census Bureau, Esri, DOC, NOAA, NOS, NGS)

4.2.1 Reference Spectra

Region of Interest (RoIs) are collected from images with different dates for different land cover types such as dry vegetation, green areas, and soil types (Figure 4.7). Note that there is the liberty of using any image, whichever is wanted, to mark pure pixels because all images are represented with the same basis of endmembers. These reference spectra (Figure 4.8) are used to label the endmembers with the SAM method.

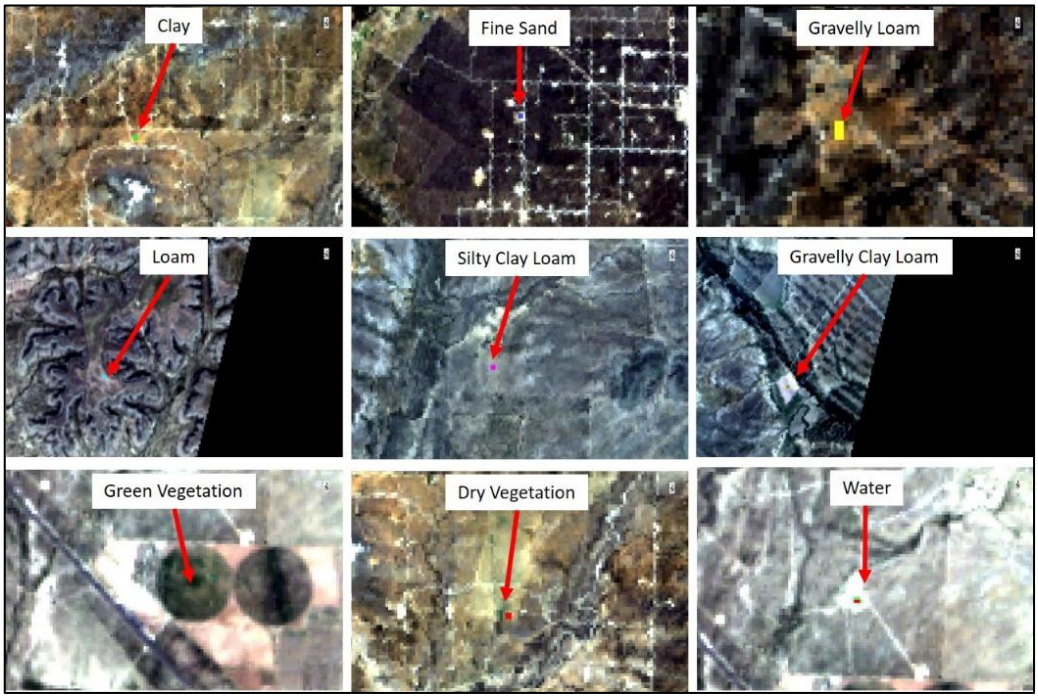

Figure 4.7 RoIs of different materials

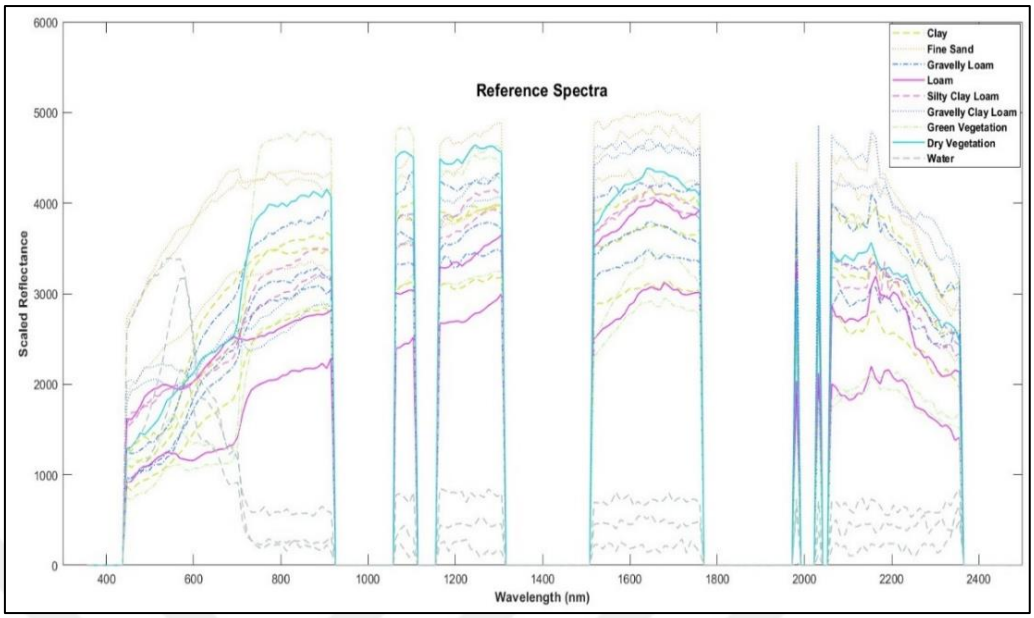

Figure 4.8 Reference spectra of different land covers

The spectral behavior of different land covers differs from each other. Also, the same material yields spectral variability throughout the images. Figure 4.9 shows fine sand, gravelly loam, gravelly clay loam, and green vegetation spectral graphs. Apart from the noise, the shape of the graphs for the same material changes revealing the endmember variability.

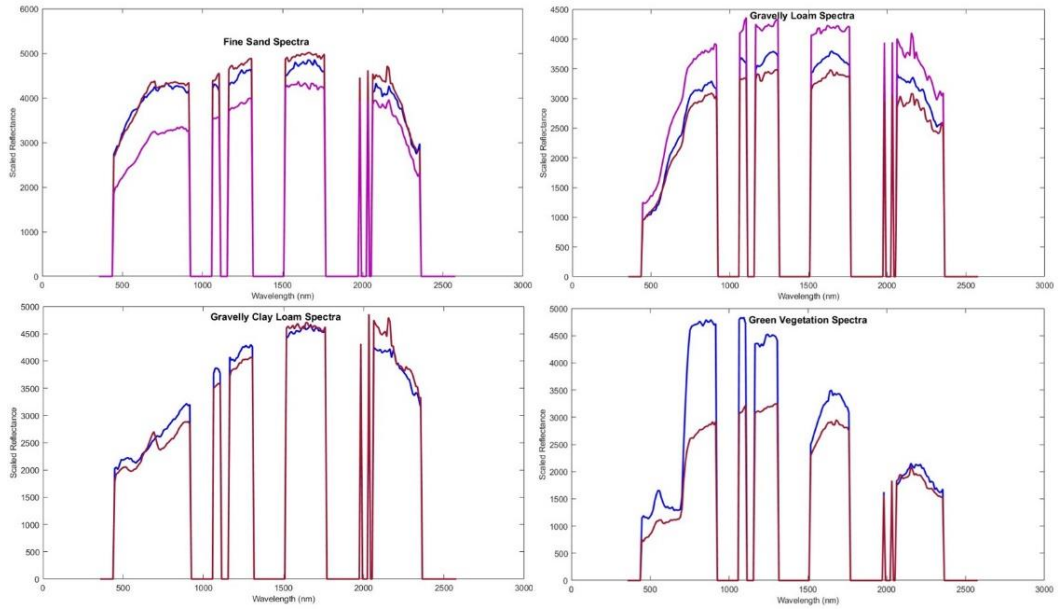

Figure 4.9 Spectral variability of materials

4.3 Ground Truth

The Digital General Soil Map of the United States (STATSGO2) is available online for users from the US Department of Agriculture website. The soil database (STATSGO2) is the output of the generalization of detailed soil survey maps. It is an outcome of the combination of data on geology, topography, vegetation, and climate in the areas where more detailed soil survey maps did not exist. In those areas, the maps also were formed by Land Remote Sensing Satellite (LANDSAT) images. Ground truth for soil types (Figure 4.10) of the Texas study area is determined from the Digital General Soil Maps of the United States.

This ground truth provides more detailed information than other online digital sources. FAO also provides Digital Soil Map of World (DSMW), produced by digitizing the paper map of FAO-UNESCO Soil Map of the World at 1:5.000.000. But, this version is more generalized than the STATSGO2 version. Soil textures are

grouped under three classes: Coarse (1), Medium (2), and Fine (3), according to the percentage of silt, sand, clay, and loam in DSMW(Figure 4.11).

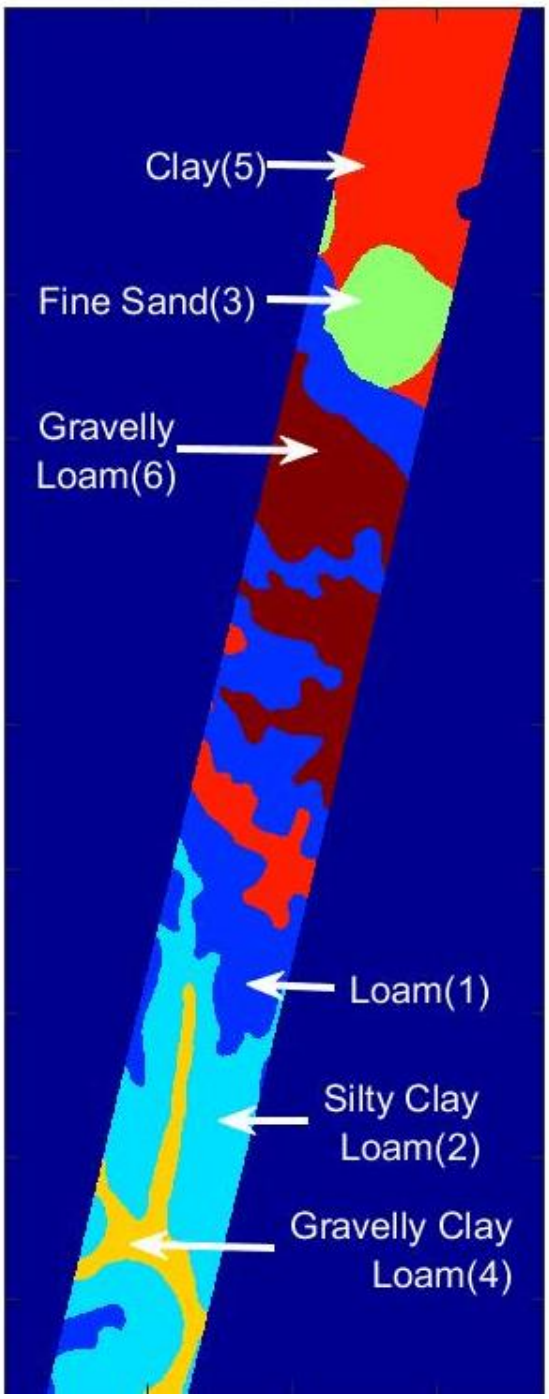

Figure 4.10 The soil ground truth image of the Texas test area (Soil Survey Staff, 2022)

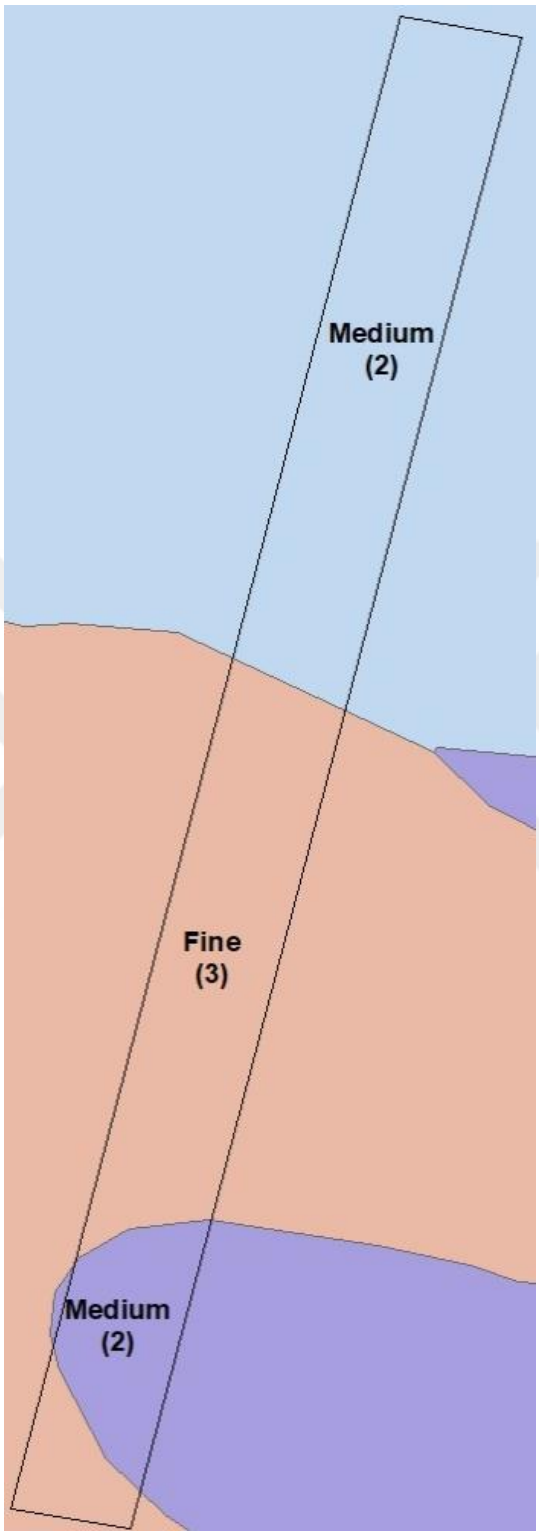

Figure 4.11. FAO's Digital Soil Map of World (DSMW) for the study area (FAO, 2022h).

CHAPTER 5

THE PROPOSED ALGORITHM

The first step of the algorithm is the pre-processing of the data. Then, spectral unmixing of hyper-spectral images is applied to extract the endmembers. Later, the endmembers are labeled using the SAM method, comparing their spectra with the reference spectra. SAM results give the labels of endmembers according to the soil or green/dry vegetation reference spectra in the study area. These endmembers are also used to form synthetic data to test unmixing algorithms and to test the suggested method. The endmembers are grouped as stable and unstable. Next, unstable land covers, such as green and dry vegetation, are extracted from the images with OSP. Later, soil images without vegetation cover are formed by fusing of remaining parts of the images. The final step is the classification of the fused soil image using the RF classifier. Figure 5.1 shows an overall flowchart of the proposed algorithm.

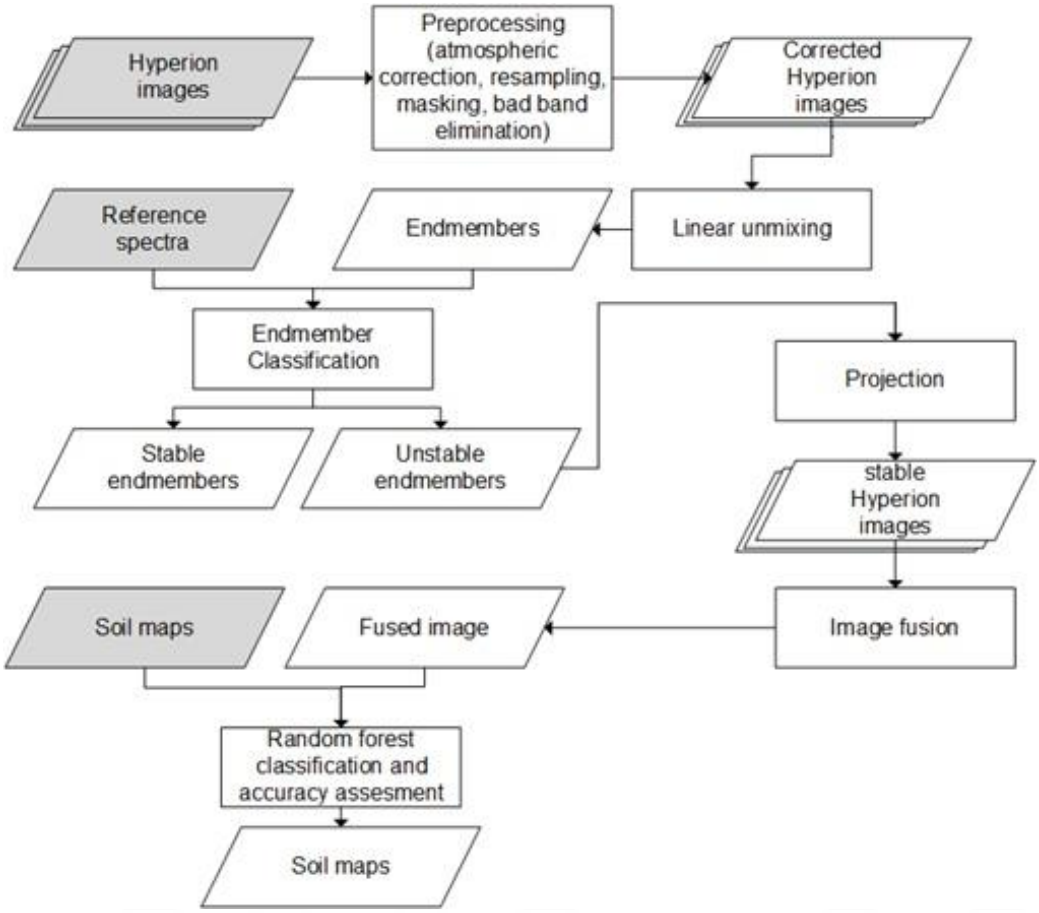

Figure 5.1 The flowchart of the proposed algorithm.

Each part of the algorithm is explained in detail in the following sub-sections.

5.1 Pre-Processing

This procedure covers de-stripping, radiometric calibration, atmospheric correction, geo-registration of images, resampling, masking of the irrelevant area, and bad band elimination.

5.1.1 De-Stripping

Hyperion has some problems, and one of them is the stripping effect on the images. The gain of the CCD detector elements changes with time because of thermal effects. These changes cause vertical stripes on the images. The first 12 VNIR bands and many bands in short-wave infrared are affected by the striping (Scheffler & Karrasch, 2013).

ENVI-THOR De-Stripping module is used to remove periodic scan line striping on the images. The algorithm is simple. It calculates the mean value of every n'th vertical line and normalizes each one to its respective mean (Harris Geospatial Solutions, Inc., 2020).

5.1.2 Radiometric Calibration

Radiometric calibration is applied to the images to set the pixel value according to the radiance gains and offsets. Basically, it changes pixel values to radiance.

$$Radiance = Gain * Pixel\,Value + Offset \quad (5.1)$$

It is equivalent to dividing each pixel's value by 40 in VNIR and by 80 in SWIR for a Hyperion image. Because the offset is zero, and the gain value is 0.025 for the visible and the NIR bands and 0.0125 for the SWIR bands (Harris Geospatial Solutions, Inc., 2020).

5.1.3 Atmospheric Correction

Quick Atmospheric Correction (QUAC) is developed to be used in multispectral and hyperspectral images (Harris Geospatial Solutions, Inc., 2020). The method works

on visible, near-infrared, and shortwave infrared (VNIR-SWIR) wavelengths. QUAC uses an in-scene approach. The method only needs sensor band locations and their radiometric calibration. QUAC algorithm assumes that at least ten or more spectrally diverse materials can be found in every image so that the average reflectance spectrum of them can be accepted as a universal reference. This information is used by the algorithm to find the average endmembers of the scene. Average endmembers, which are found from the images, and endmembers from spectral libraries are used to estimate the gain and the offset. No additional information is needed about the images because the method obtains necessary information from the scene. The method is faster than physics-based methods and provides a more approximate result (Bernstein, 2012).

5.1.4 Geo-Registration

Level 1R images do not have coordinates, which are necessary to find their places on Earth according to the national or global coordinate reference systems. L1R images are registered by image to map registration method with the help of L1T images, which have UTM coordinates. As a result, all the images are registered to the same UTM projection.

5.1.5 Resampling

Since the images are not co-registered, all images are transformed to an arbitrary reference image and resampled by bicubic interpolation.

5.1.6 Bad Band Elimination

Not all Hyperion bands are well-calibrated. Some of the bands are zeroed (1-7, 58-76, and 225-242), some VNIR bands overlap with the bands in the SWIR region, and some of the bands have zero atmospheric transmission. These bands are eliminated, and 131 bands (10:56, 85:96, 102:116, 137:161, 183, 188, 191:220) are used in total.

5.1.7 Masking

Footprints of Hyperion images are overlapped, and then the intersection of the time-series is determined. For the Texas test data set, an area of 1432.85 square kilometers, having different soil types and vegetated areas, is used because of computer memory limitations, and the rest are masked out.

5.2 Endmember Extraction

Unmixing algorithms are used to find endmembers, which are the spectra of distinct materials found on the images, and their abundance maps from hyperspectral images. Endmembers are defined based on the soil texture; however, the soil's composition and mineralogy also determine the reflectance. Hence, in a different part of the world, the endmembers of the soil texture types may change considerably. Linear unmixing algorithms, which work with hyperspectral imagery, are used to solve the mixed pixel problem on the images (Figure 5.2).

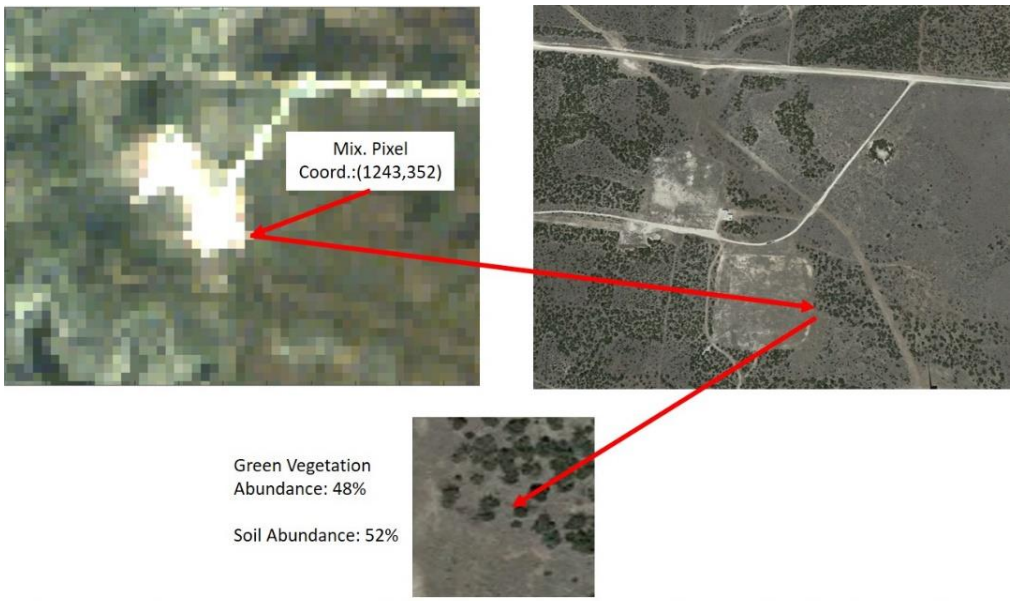

Figure 5.2 An example of the mixed-pixel problem

SMACC unmixing algorithm is used for endmember extraction in this study. The method uses the convex cone model. The first endmember is found by a convex cone, which is determined by extreme points in the scene. The next endmember is found by a constrained oblique projection applied to the existing cone. It is repeated until an existing endmember is found in the convex cone or desired endmember numbers are reached (Gruninger et al., 2004).

5.3 Abundance Estimation

Besides proper estimation of endmembers, proper estimation of abundance amounts is essential for this study because they are used as weights in calculating weighted means in the fusion step. SMACC produces abundance maps, which determine the fractions of endmembers, during the endmember extraction procedure, in addition to the endmembers.

5.4 Endmember Classification

After the endmember extraction step, the main problem is labeling the endmembers according to the basic material types on the image. There are lots of endmembers, and it is difficult to say whether or not two endmembers represent the same type of material or not. For this purpose, the reference spectra of manually selected RoIs are compared with the endmember spectra with the help of the SAM algorithm to label the endmembers. Then, the endmembers are separated into two groups according to the material types, such as stable land covers (soil, rock) and unstable land covers (green/dry vegetation, snow, and water).

The SAM algorithm builds the spectral similarity of two spectra by calculating the angle between them in n-dimensional space. The algorithm considers the spectra as vectors in a space with dimensionality equal to the number of bands to find spectral similarity (Rashmi et al., 2014). SAM is insensitive to the scaling effects.

5.5 Orthogonal Subspace Projection

The next step of the study is to separate endmembers according to the stability of materials. Soil and rock are considered stable endmembers since their change are very slow over the years, while green or dry vegetation are considered unstable endmembers because they change considerably throughout the year or sometimes from year to year.

Unwanted materials are eliminated by projecting images into a space orthogonal to the unstable endmembers. The linear mixing is modeled as follows (Harsanyi & Chang, 1994):

$$\boldsymbol{p}(x,y) = \boldsymbol{E}\,\boldsymbol{a}(x,y) + \boldsymbol{n}(x,y) \tag{5.2}$$

where

$\boldsymbol{p}$, the hyperspectral vector, is a tx1 vector, t being the number of bands,

$\boldsymbol{E}$ is composed of vertical endmember vectors,

$\boldsymbol{a}$, the abundances, is an cx1 vector, c being the number of endmembers and

$\boldsymbol{n}$ is the noise at the pixel (x, y).

Without loss of generality, one can arrange the order of the endmembers as follows:

$$\boldsymbol{E} = (\boldsymbol{e}_1, \dots, \boldsymbol{e}_r, \boldsymbol{e}_{r+1}, \dots, \boldsymbol{e}_c) \tag{5.3}$$

where $\boldsymbol{E}$ is composed of c endmembers, the first r corresponding to unstable and the rest to the stable endmembers. Let us denote $\boldsymbol{E}$ as

$$\boldsymbol{E} = (\boldsymbol{E}_u, \boldsymbol{E}_s) \tag{5.4}$$

The rejecter of the unstable components is (Harsanyi & Chang, 1994)

$$\boldsymbol{F} = \boldsymbol{I} - \boldsymbol{E}_u(\boldsymbol{E}_u{}^T \boldsymbol{E}_u)^{-1} \boldsymbol{E}_u{}^T \tag{5.5}$$

This operator is applied to all the images to get rid of the unstable components. So the pixel after the projection is

$$\boldsymbol{p}'(x, y) = \boldsymbol{F}\, \boldsymbol{p}(x, y) \tag{5.6}$$

5.6 Image Fusion

After the elimination of the unstable materials, the remaining are rocks and soil, which show limited variability over time, probably except for the variability caused by the moisture in them. Each image is a noisy measurement of the soil scaled down by the unstable components. Rescaling and fusion of multiple images should lead to an image with a better signal-to-noise ratio than any of the individual images. Images are fused with the weighted mean of images using stable abundance values as weights for each pixel in an image. Equation (5.7) shows the abundance of stable and unstable components. Equation (5.8) represents total unstable abundance, and

Equation (5.9) represents total stable abundance. Pixel values are arranged according to the stable abundance value in Equation (5.10). Each image is composed of these rearranged pixels (5.11). Then mean of the images is calculated in Equation (5.12).

$$\boldsymbol{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_r \\ a_{r+1} \\ \vdots \\ a_c \end{bmatrix} \tag{5.7}$$

$$w_u = \sum_{i=1}^{r} a_i \tag{5.8}$$

$$w_s = \sum_{i=r+1}^{c} a_i \tag{5.9}$$

$$\widehat{p_s} = \frac{1}{m \sum_{j=1}^{m} w_{s,j}} \sum_{j=1}^{m} p'_j \tag{5.10}$$

$$A = \begin{bmatrix} \widehat{p_{s_1}} \\ \vdots \\ \widehat{p_{s_r}} \\ \widehat{p_{s_{r+1}}} \\ \vdots \\ \widehat{p_{s_c}} \end{bmatrix} \tag{5.11}$$

$$fused\ image = (\sum_{i=1}^{n} A_i)/n \tag{5.12}$$

5.7 Soil Classification

The soil classification algorithm is applied to the synthetic images as well as to the real images of the study area in Texas. The fused image is classified with supervised classification techniques using the RF algorithm. RF is an ensemble learning method for classification and regression. The method builds many decision trees at training time and produces the class from the mode of the individual classes during the classification process (Ho, 1995). In RF classification, there is no need for cross-

validation or a separate test set to get an unbiased estimate of the test set error. It is estimated during the procedure as out-of-bag error estimate. Each tree is formed with a different bootstrap sample from the original data. A third of the cases are left out of the bootstrap sample, and they are not used in the formation of a tree. In this way, a test set classification is obtained for each case for one-third of the trees (Breiman & Cutler, 2022).

As a final step, label relaxation (Wilde & Bardsley, 2010) using class probabilities and then post-filtering (Wilde & Bardsley, 2010) performing a 3x3 mode filter is applied to real images for better classification result.

CHAPTER 6

RESULTS AND DISCUSSION

The suggested algorithm is first applied to the synthetic images, and after successful results, it is applied to real images from the dates 22/06/2013, 25/09/2013, and 24/10/2013 of the study area in Texas, USA.

6.1 Results on Synthetic Images

The suggested method is first tested on synthetic images. Three synthetic images representing three seasons are used to test the method.

6.1.1 Endmember Extraction

SMACC unmixing algorithm is used on the synthetic data. Synthetic images, which are created for three different seasons: spring, summer, and autumn, are used to test the suggested algorithm. All three images are concatenated to form one large image. The unmixing algorithm is applied to find the endmembers. Because five endmembers are used to create synthetic images, it is decided to set the parameter to find five endmembers during the unmixing procedure (Figure 6.1).

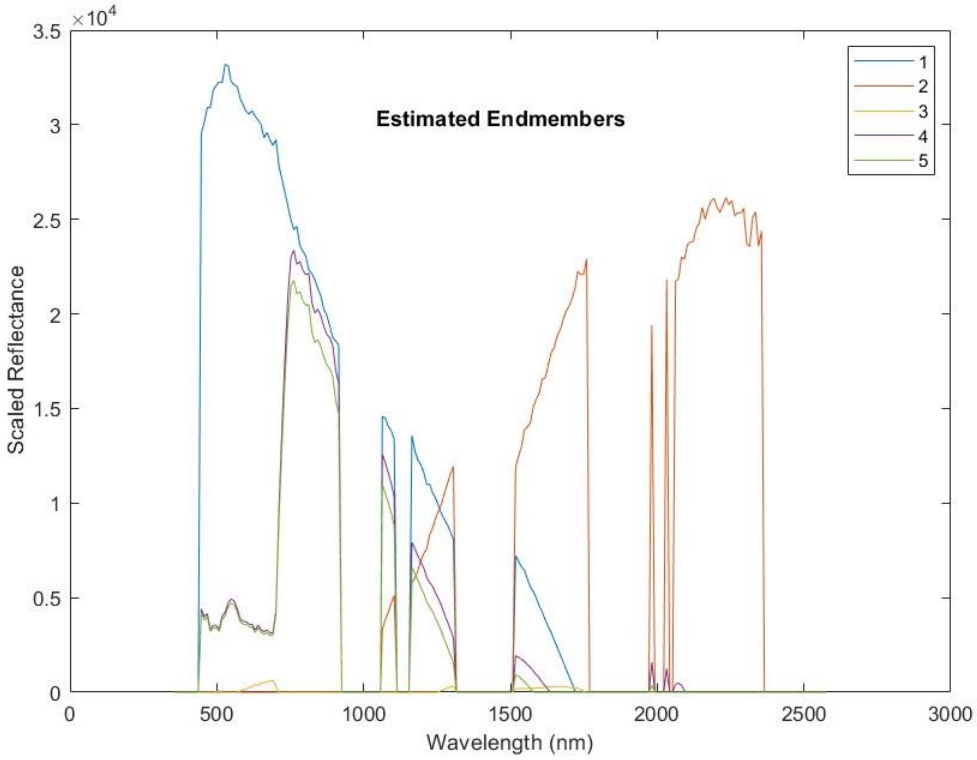

Figure 6.1 Endmembers which are found by SMACC algorithm from the combined synthetic image

It can be seen on the endmember spectral graph (Figure 6.1) that there is a huge gap between some of the endmembers when it is compared with the spectral plot of the initial endmembers (Figure 4.1). The reason is because of the noise that is added to the synthetic images.

6.1.2 Abundance Estimation

Abundance maps are produced during the endmember extraction procedure (Figure 6.2).

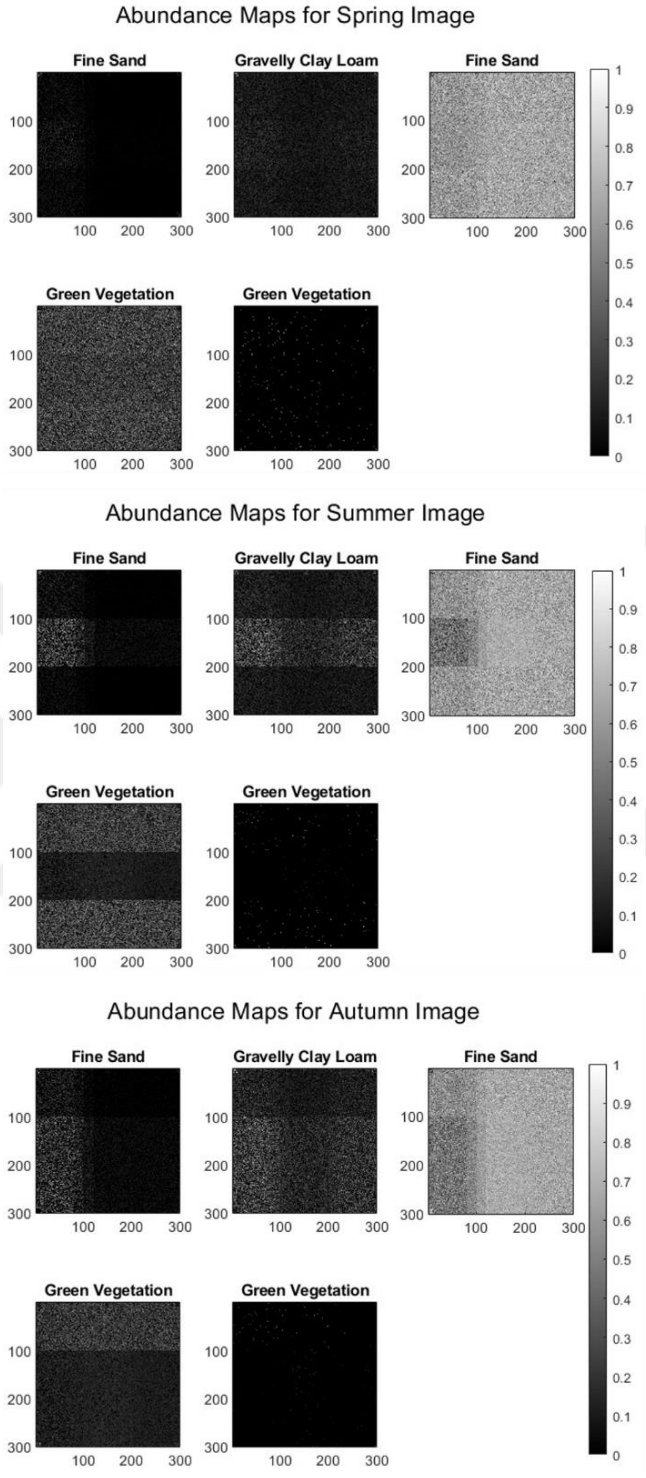

Figure 6.2 Abundance maps of the endmembers for each synthetic image

When the initial abundance maps with the abundance maps are compared after unmixing, the resemblance between them is clear in spite of the noise that is added to the images before analysis. R-Square is estimated to see the similarity between two abundance maps. Higher values of R-Square represent good similarity. When the abundance maps are examined more closely, the similarity between the initial and the estimated abundance maps of green vegetation for the Summer image is detected (Figure 6.3). R-Square is 25% between the reference abundance map and the estimated abundance map.

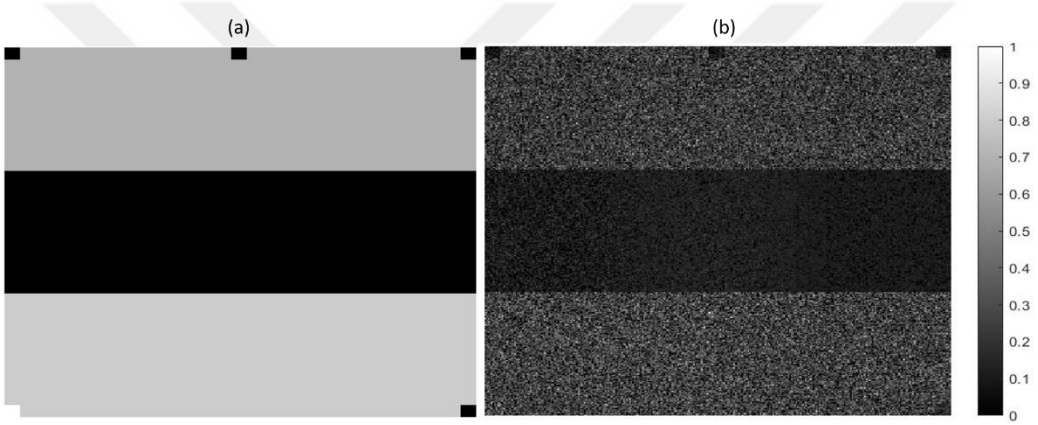

Figure 6.3 Comparison of abundance maps of green vegetation for Summer image. (a) Initial abundance map of green vegetation component, (b) Estimated abundance map of green vegetation component

Also, abundance maps of green vegetation are similar for Autumn (Figure 6.4). R-Square is 27% between the reference abundance map and the estimated abundance map of green vegetation for Autumn image.

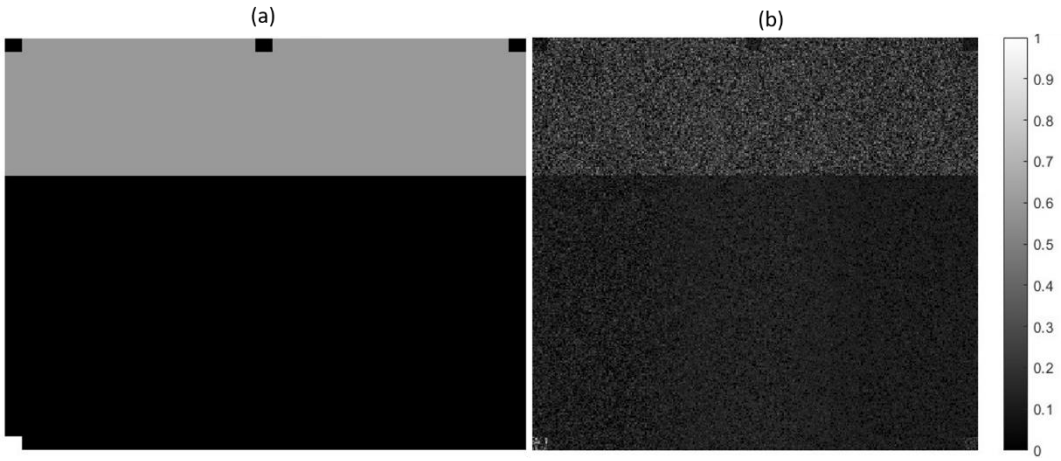

Figure 6.4 Comparison of abundance maps of green vegetation for Autumn image. (a) Initial abundance map of green vegetation component, (b) Estimated abundance map of green vegetation component

The comparison of abundance maps shows that the unmixing algorithm at least achieved finding abundance values of green vegetation for two synthetic images despite the noise.

6.1.3 Endmember Classification

After endmember extraction, the endmembers are labeled according to the material types in the image. For this purpose, with the help of the SAM algorithm, estimated endmember spectra are compared to the spectra of initial endmembers, which are used to create the synthetic images (Figure 6.5).

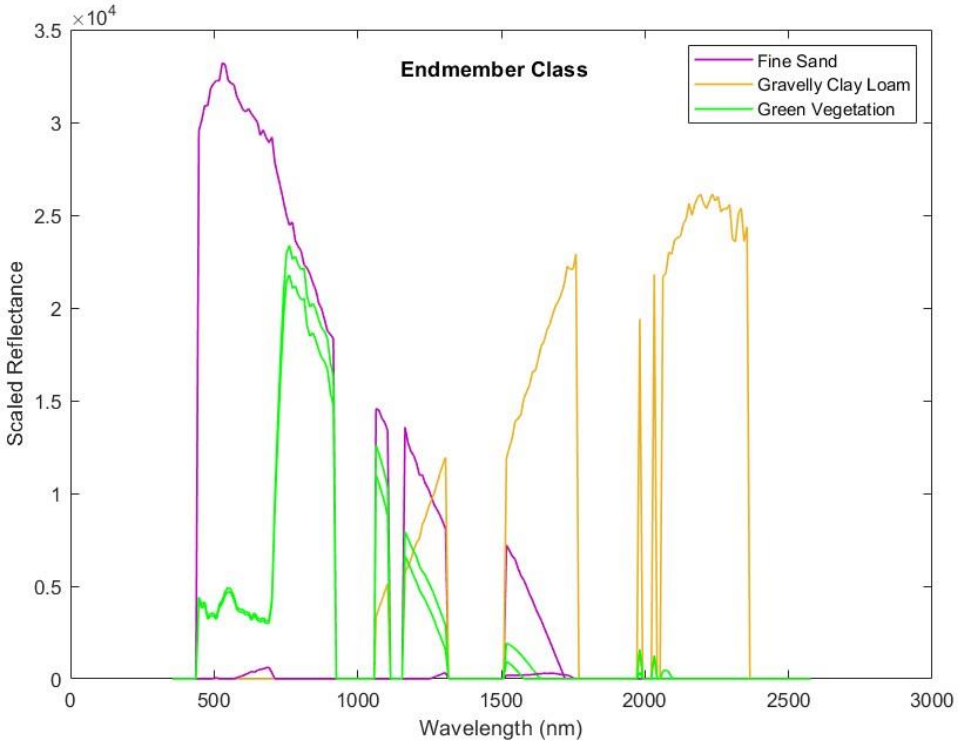

Figure 6.5 Label of endmembers after SAM procedure

Then, endmembers are separated into two groups according to the material types, such as stable land covers (soil, rock) and unstable land covers (green/dry vegetation, snow, and water) (Figure 6.6). Fine sand and gravelly clay loam endmembers are in the soil group. Green vegetation is in the non-soil group.

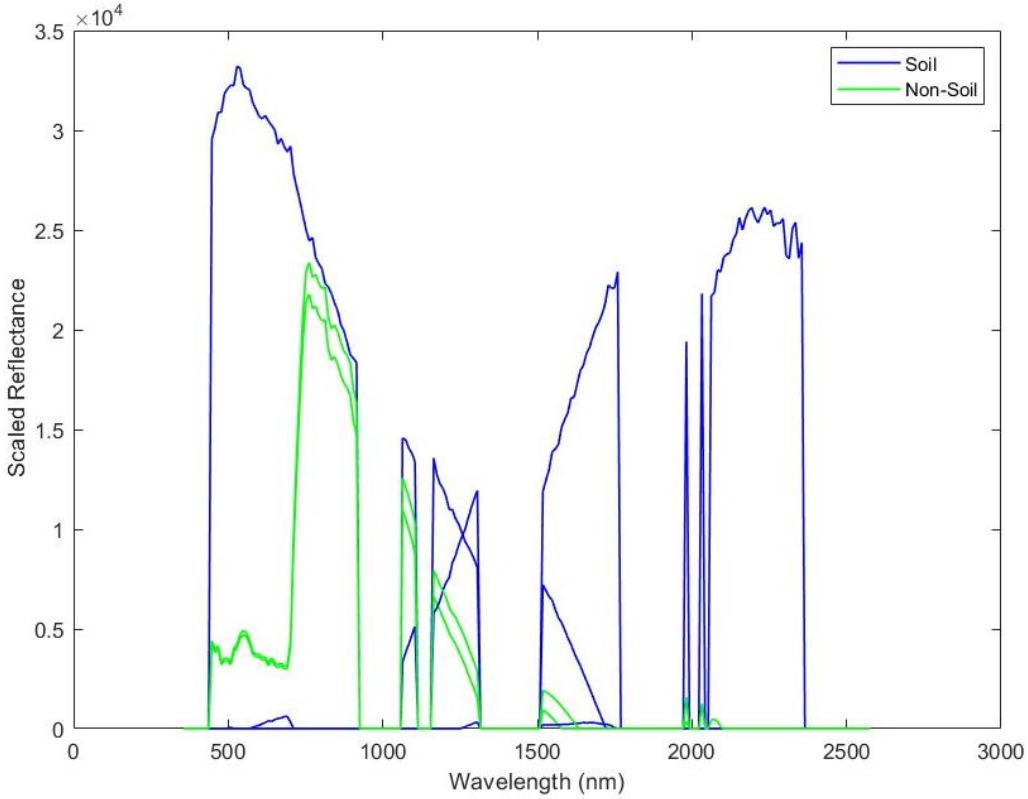

Figure 6.6 Soil/non-soil endmember groups

Separation of groups is important because the aim is to remove non-soil materials from the images using this information in further analysis. Although the method could not find the complete match of whole endmembers from such noisy images, it still can be useful to eliminate the green component from the images in the next step.

6.1.4 Orthogonal Subspace Projection

In the previous step, endmembers were separated according to the stability of materials. Figure 6.7 shows the result of the OSP procedure for the image from the Summer season. When the images are examined for comparison of the results, it is

expected to see that non-soil components are eliminated from the image after OSP. But, it is difficult to see the expected result clearly on the images because of the noise. Also, images have different soil abundance ratios based on the season. For the summer image, it seems that the algorithm is successful in removing non-soil components when green areas are examined.

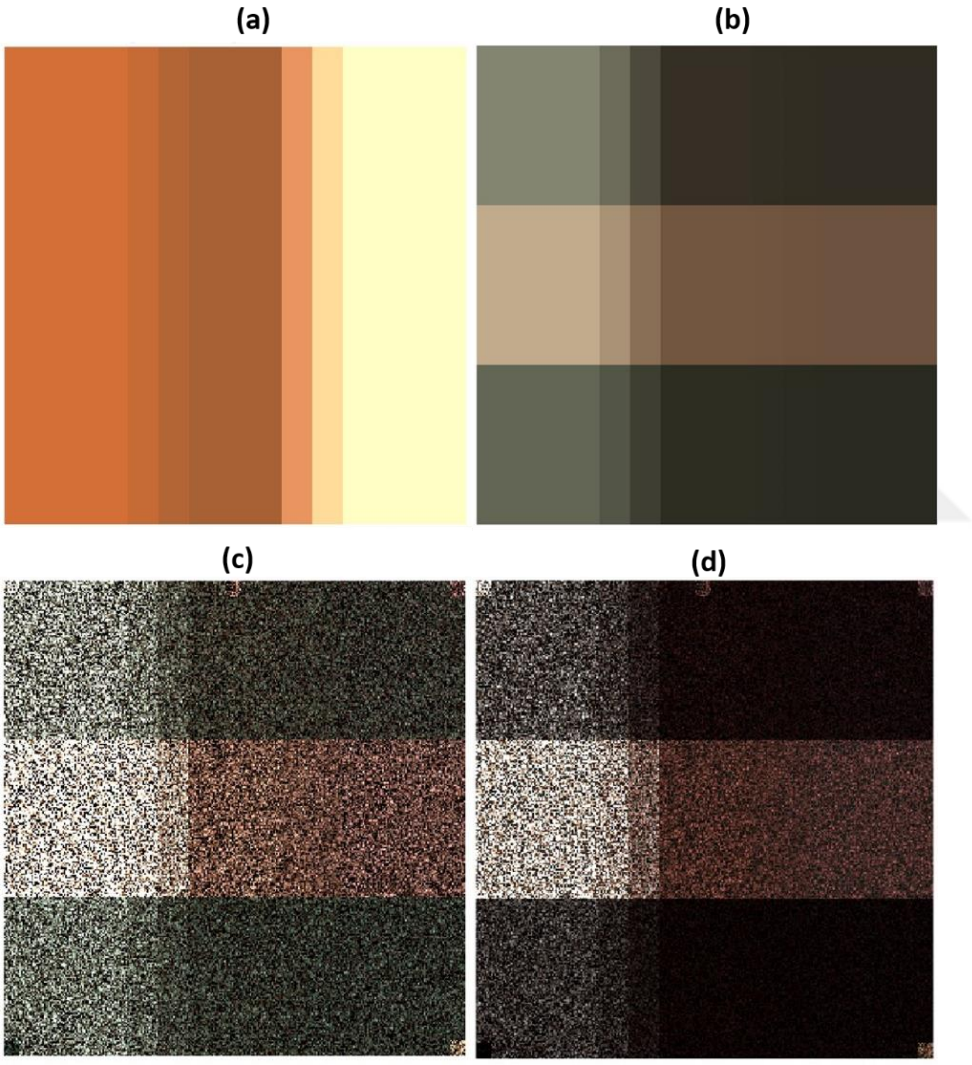

Figure 6.7 (a) Soil image without vegetation component (b) Synthetic image for the Summer season (c) Soil image after OSP for the Summer image (Bands 20, 11 and 3 are used to represent RGB)

6.1.5 Image Fusion

After the OSP procedure, the unstable materials are eliminated; and the remaining are soil and noise. Each image is fused by the weighted mean algorithm, which uses abundance values as the weights (Figure 6.8). The fusion of multiple images provides an image with a better signal-to-noise ratio than any of the individual images. Table 6.1 shows the SNR values of the synthetic images and the values after fusion. Also, when the results are compared visually in Figure 6.8, it can be seen that there is less greenness on the fused image despite the noise, which spoils the appearance of the image. That means fusion is successful in revealing the soil components on the image.

Table 6.1 SNR values for the synthetic images and the fused image

Image	*SNR Value of Synthetic Images*	*SNR Value After Fusion*
Spring	-59.69	2.25
Summer	-39.09	1.96
Autumn	-36.22	1.78

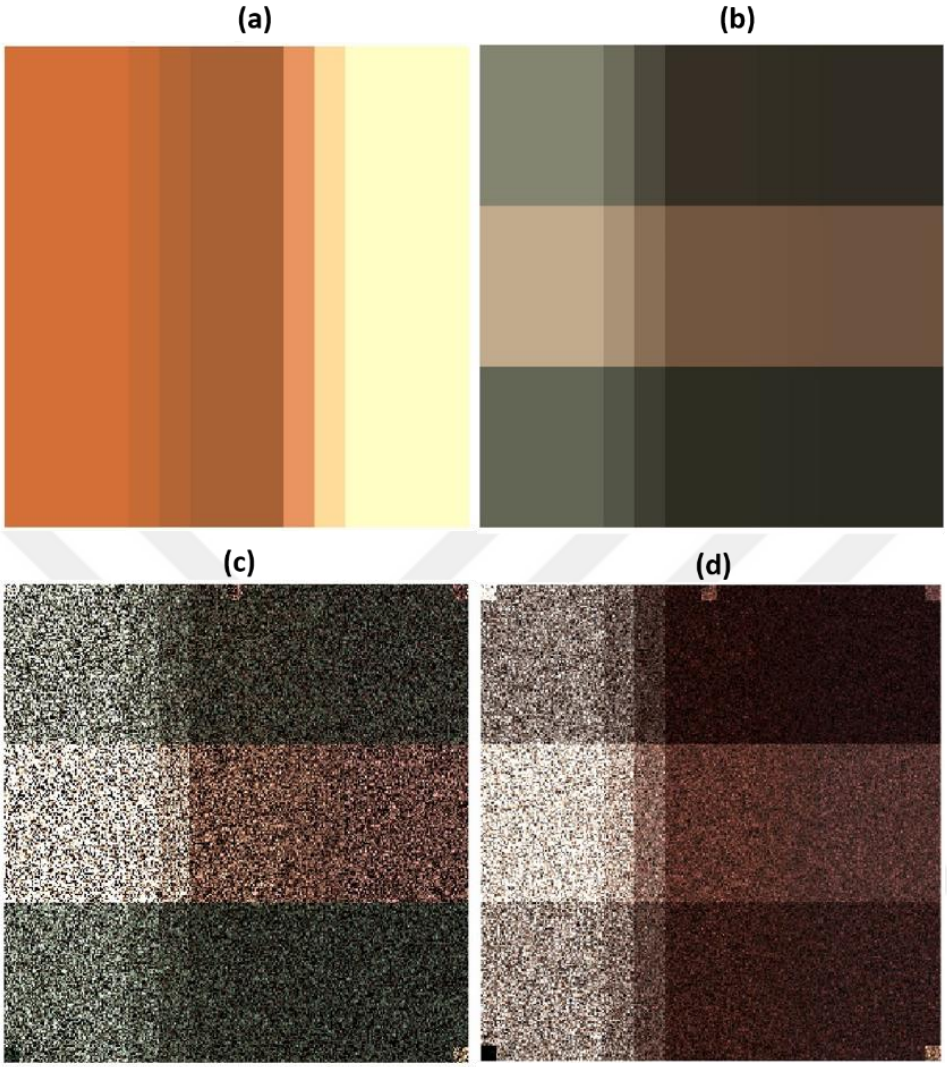

Figure 6.8 (a) Soil image without vegetation component (b) Synthetic image for the Summer season (c) Synthetic image after fusion (Bands 20, 11 and 3 are used to represent RGB)

6.1.6 Soil Classification

The fused image is classified using the RF classifier as a final step. Two hundred decision trees are used, and 1000 training pixels are chosen randomly for each class.

Three different soil types are chosen for the synthetic image: fine sand, gravelly loam, and gravelly clay loam. So, ground truth for classification is formed according to these soil types (Figure 6.9 (a)).

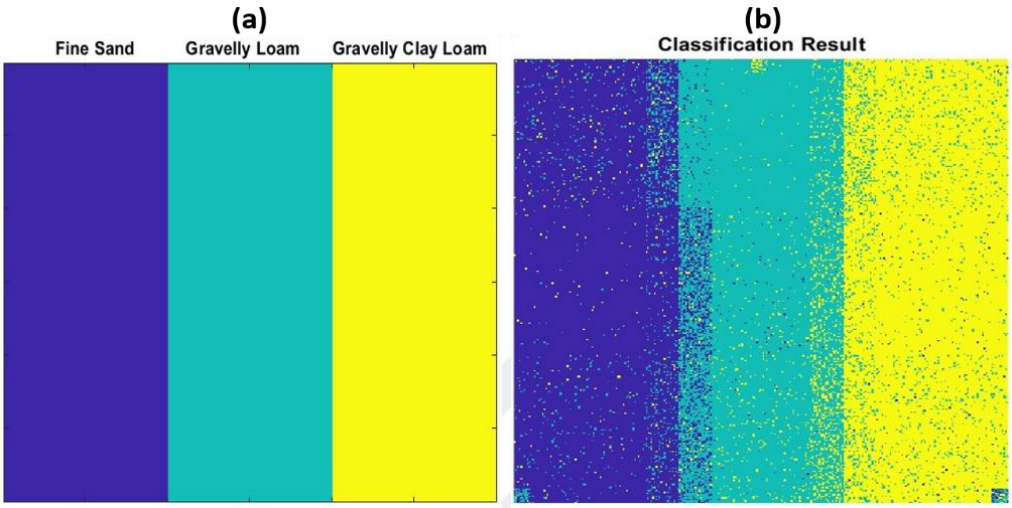

Figure 6.9 (a) Ground truth for soil classification of the synthetic image. The blue area is for fine sand, the green area is for gravelly loam and the yellow area is for gravelly clay loam (b) Classified image

The fused synthetic image is classified by the RF algorithm with 91.85% accuracy (Figure 6.10).

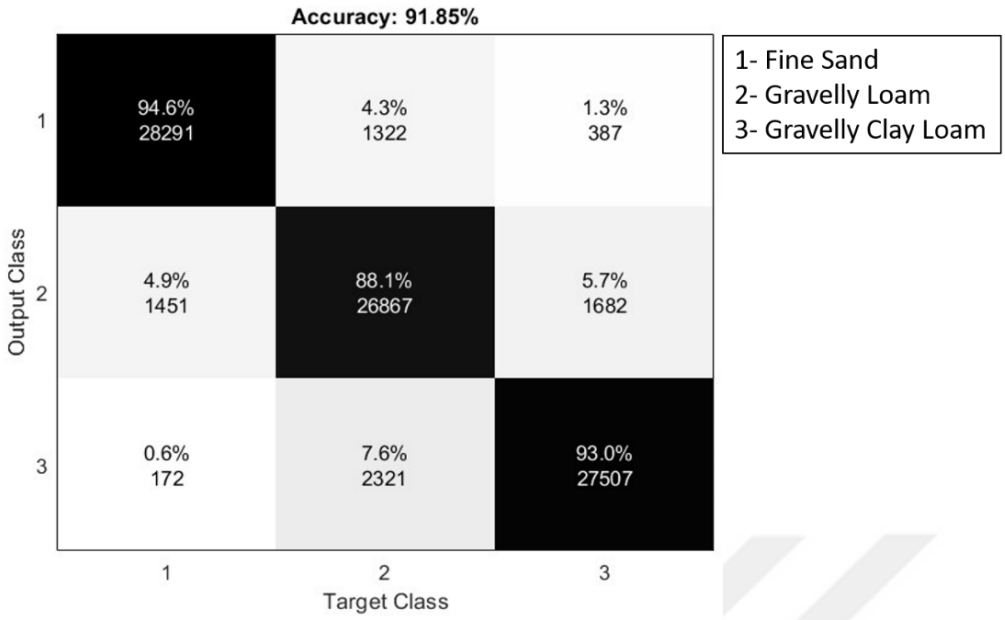

Figure 6.10 Classification result for synthetic image classification

The same procedure is repeated on the combination of two synthetic images, and classification is applied to those with the same parameters. The classification with the same parameters was also applied to images one by one after the OSP procedure. Initial synthetic images are also classified to make a comparison and see the suggested method's efficiency after classification. All results are summarized in Tables 6.2-6.6. The first columns of the tables are the classification results of synthetic images before analysis. The second columns are the classification results of the arithmetic mean of synthetic images. The third columns are the classification results of single images after the OSP procedure. The fourth columns of the tables are the classification results of images after fusion.

Table 6.2, Table 6.3, and Table 6.4 show the classification results when the suggested algorithm is applied to the synthetic image pairs. It can be seen that even using only two images gives better classification results than the classification results of raw images. Raw synthetic images have 81.78%, 79.84%, and 86.33% accuracy. After OSP, for Spring and Summer pair, the OSP classification result can have 82.35% and 82.63% accuracy; for Spring and Autumn pair, the OSP classification result can

have 85.96% and 90.47% accuracy; for Summer and Autumn pair, the OSP classification result can have 81.57% and 87.39% accuracy. When the classification result of the arithmetical mean of raw image pairs with fusion results are compared, it can be seen that the results of fused image classification are better than their results. For the Spring and Summer image pair, the classification accuracy of the mean image is 80.13%, while the fusion image accuracy is 85.08%. For Spring and Autumn pair, the classification accuracy of the mean image is 82.68%, while the fusion image accuracy is 94.34%. For Summer and Autumn pair, the classification accuracy of the mean image is 83.62%, while the fusion image accuracy is 87.86%. The classification results of images after OSP is better than single raw image classification, and fused image classification is better than the classification results of images after OSP. For the pair of Summer and Autumn, the results are close to the fusion result. However, fusion results for Spring-Summer and Spring-Autumn pairs are better. While the pair of Spring and Summer has 82.63% accuracy, the fusion result has 85.08% accuracy. While the pair of Spring and Autumn has 90.47% accuracy, the fusion result has 94.34% accuracy. The method provides an improvement in the OSP step and then the Fusion step for synthetic images, even when only two images are used.

Table 6.5 shows the results when the algorithm is applied to three images. It can be said that algorithm is successful when the results are examined closely. The classification accuracy of the synthetic image before the application of the algorithm is 86.33% at best. After the application of the algorithm, OSP-applied images have better classification accuracy than previous results: images can have 91.78% classification accuracy after OSP. Moreover, the fusion of three images has 91.85% classification accuracy. When the classification result of the mean of raw images and fusion result are compared, the algorithm gives better results when the fused image is used. The mean of raw images has 84.17% accuracy, while the fused image has 91.85% classification accuracy.

Table 6.2 Comparison of classification accuracy of synthetic images based on the fusion of Spring and Summer images

Image	*Clasf. Accuracy of Raw Image*	*Clasf. Accuracy of Mean Raw Image*	*Clasf. Accuracy of OSP*	*Clasf. Accuracy of Fusion*
Spring	0.8178	0.8013	0.8235	0.8508
Summer	0.7984		0.8263	

Table 6.3 Comparison of classification accuracy of synthetic images based on the fusion of Spring and Autumn images

Image	*Clasf. Accuracy of Raw Image*	*Clasf. Accuracy of Mean Raw Image*	*Clasf. Accuracy of OSP*	*Clasf. Accuracy of Fusion*
Spring	0.8178	0.8268	0.8596	0.9434
Autumn	0.8633		0.9047	

Table 6.4 Comparison of classification accuracy of synthetic images based on the fusion of Summer and Autumn images

Image	*Clasf. Accuracy of Raw Image*	*Clasf. Accuracy of Mean Raw Image*	*Clasf. Accuracy of OSP*	*Clasf. Accuracy of Fusion*
Summer	0.7984	0.8362	0.8157	0.8786
Autumn	0.8633		0.8739	

Table 6.5 Comparison of classification accuracy of synthetic images based on all three image fusion

Image	*Clasf. Accuracy of Raw Image*	*Clasf. Accuracy of Mean Raw Image*	*Clasf. Accuracy of OSP*	*Clasf. Accuracy of Fusion*
Spring	0.8178		0.8570	
Summer	0.7984	0.8417	0.8821	0.9185
Autumn	0.8633		0.9178	

Figure 6.11 summarizes the experiment on synthetic images. Classification results differ significantly before and after the algorithm is applied.

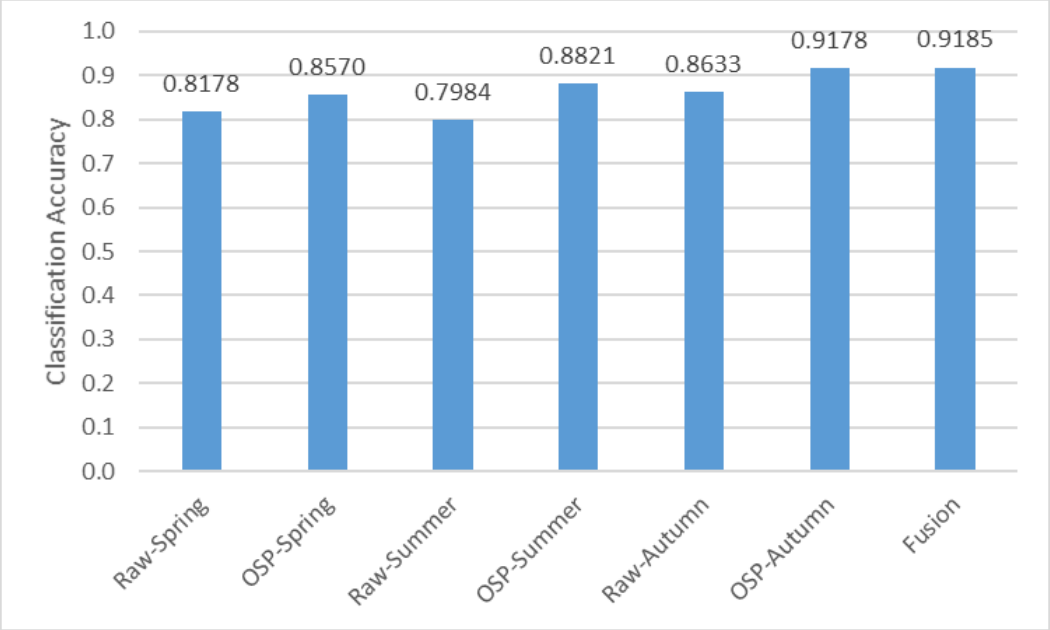

Figure 6.11 Graph of all classification accuracy results for synthetic image

6.2 Results with Real Images

The suggested algorithm is applied to the Hyperion images of the study area in Texas, USA, from different months of the year 2013 (Figure 6.12).

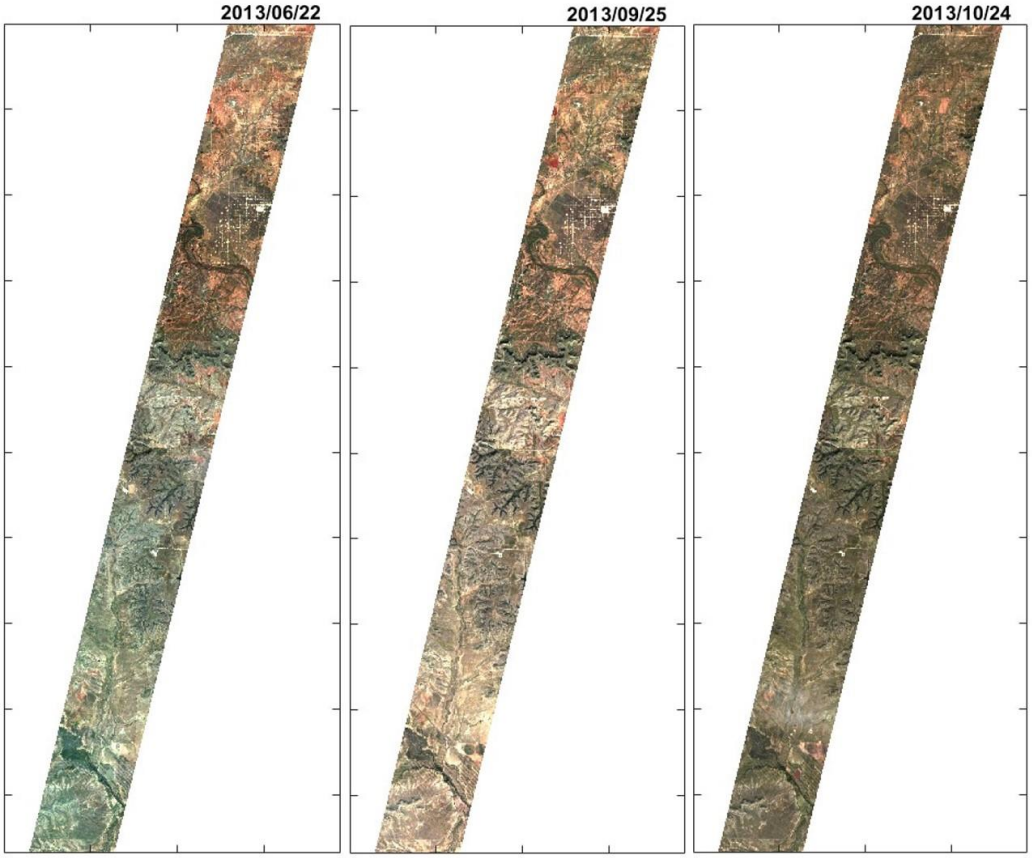

Figure 6.12 Hyperion images from the year 2013 of study area in Texas, USA

Images are taken in different months: June, September, and October of the same year. The images are chosen according to the vegetation cover variability. Conditions might be different and change from month to month. For example, the image from June and October seems greener than the image from September. A rainier season of these months might be the reason for the increase in the green areas.

6.2.1 Pre-Processing

Preprocessing steps are performed on all images at first. Preprocessing step covers: (a) de-stripping to correct vertical stripes, which is caused by the CCD detector, (b) radiometric calibration to convert pixel values into radiance, (c) atmospheric correction to eliminate atmospheric effects by modeling atmospheric conditions for the acquisition date, (d) geo-registration of images to correct them geometrically by converting pseudo coordinates to UTM coordinates, (e) resampling to make the images compatible with each other, (f) masking of the irrelevant area to handle the computer memory limitations and (g) bad band elimination to remove uncalibrated and overlapping bands.

6.2.1.1 De-Stripping

ENVI-THOR De-Stripping module is used to remove periodic scan line striping on the selected Hyperion images of the study area. Figure 6.13 shows the result of the de-stripping algorithm.

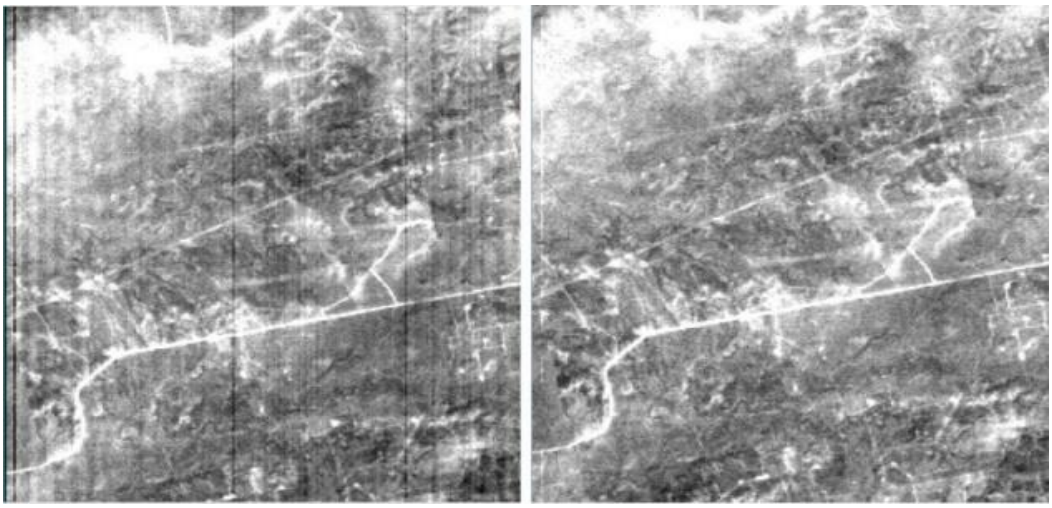

Figure 6.13 Image with vertical stripes (left) and image after de-stripping (right)

6.2.1.2 Geo-Registration

Because Hyperion L1R images do not have assigned Earth coordinates, images are registered to UTM coordinates to make them compatible with other sources (Figure 6.14).

Figure 6.14 Image before (left) and after (right) geo-registration

6.2.1.3 Masking

Using an image covering large areas can be challenging for computer memory capacities. It may not be possible to execute analysis on that large areas. So, some of the areas from the edges are masked to handle this kind of problem (Figure 6.15).

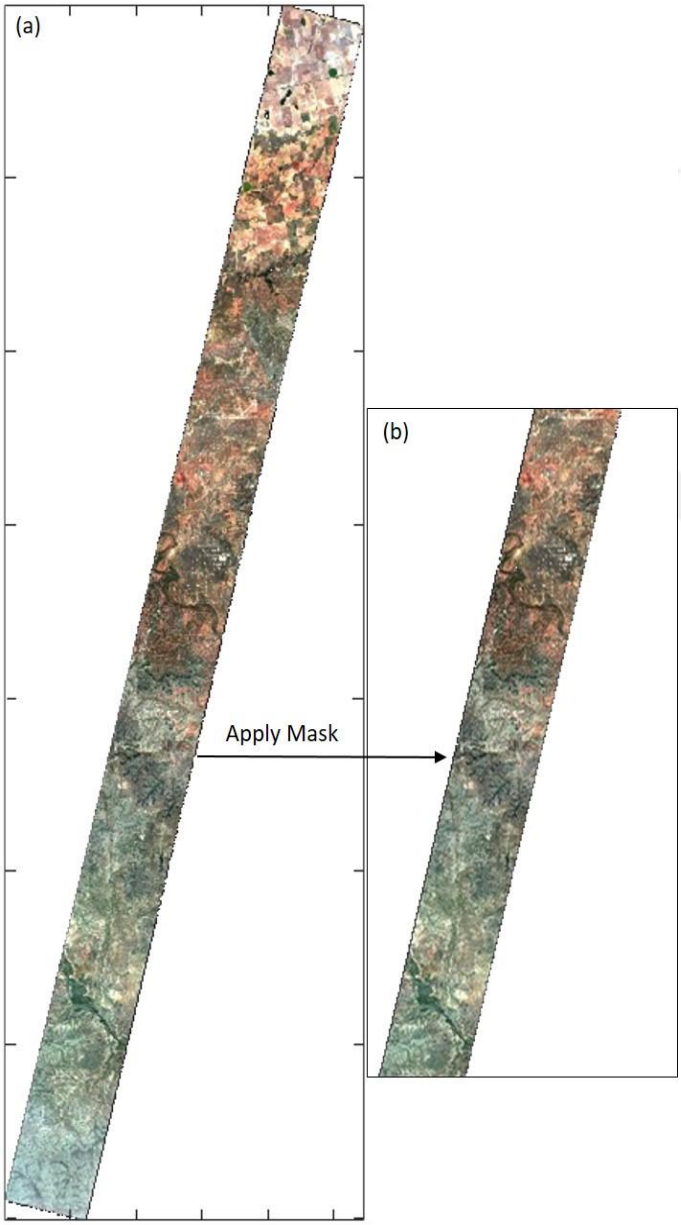

Figure 6.15 The study area before (a) and after (b) masking

6.2.2 Endmember Extraction

All three Hyperion images are concatenated before unmixing analysis. In this way, endmembers will be determined on a common base. SMACC algorithm is used to find eight endmembers (Figure 6.16). The spectral plot of endmember-3 has behavior similar to green vegetation, and endmembers with lower spectral reflectance behave like water spectra.

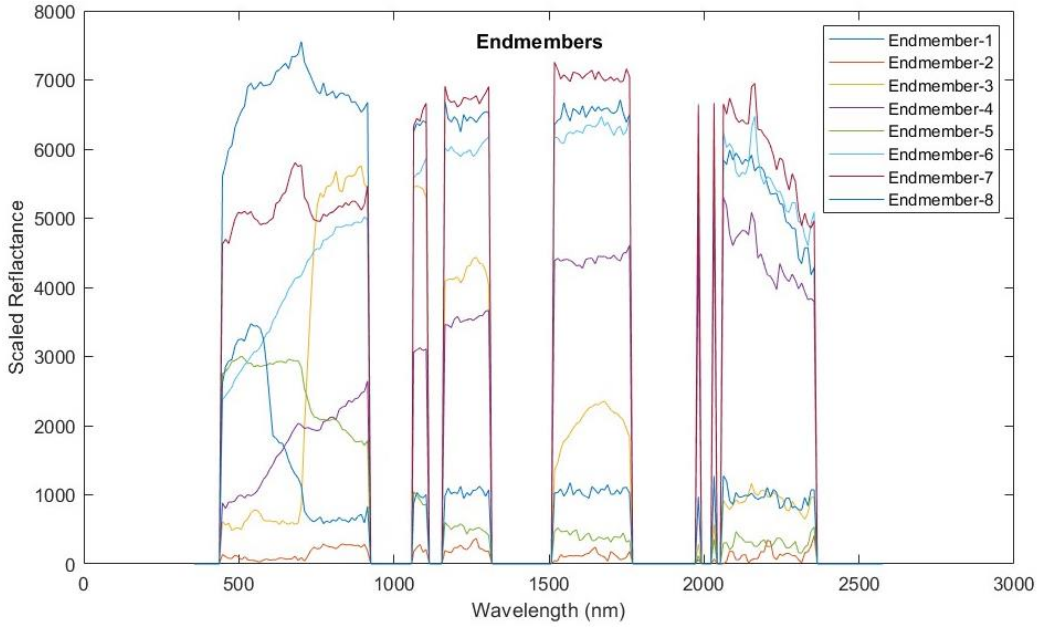

Figure 6.16 Endmembers found by SMACC algorithm on concatenated image

6.2.3 Abundance Estimation

Abundance values are also found during the unmixing procedure. SMACC algorithm calculates abundance amounts, and abundance maps are created for each endmember on each image (Figure 6.17).

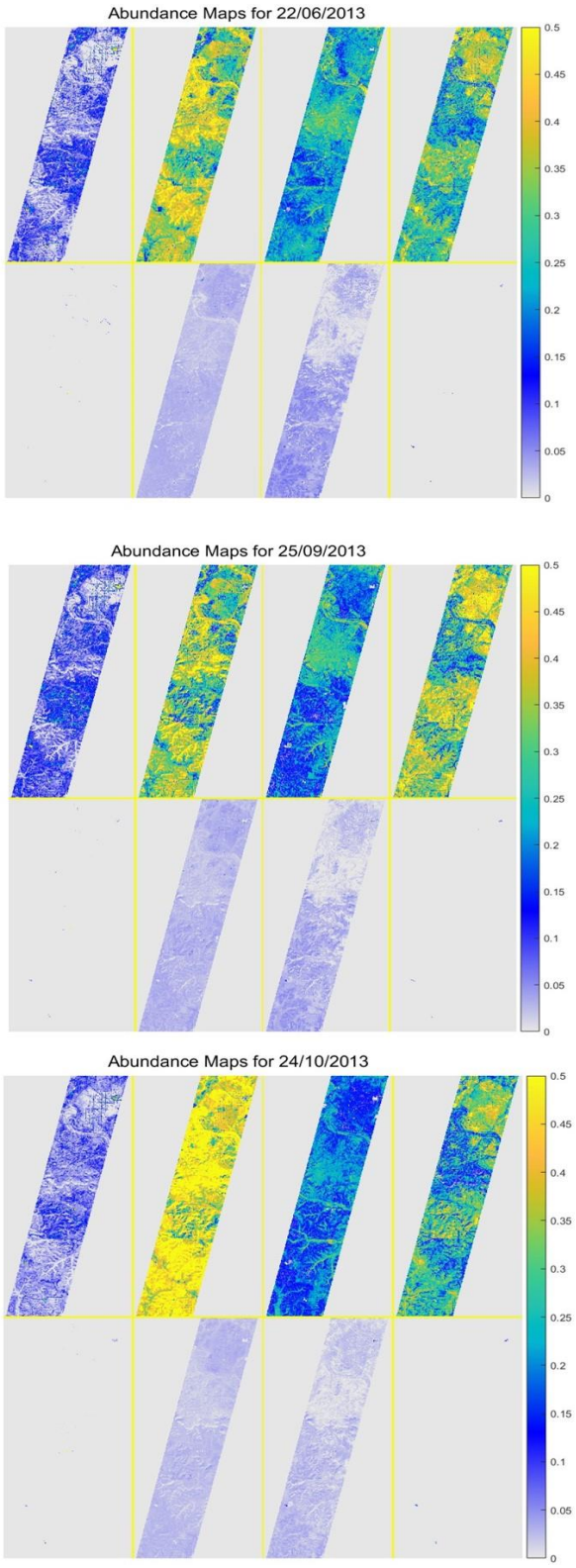

Figure 6.17 Abundance maps of endmembers for each image

When abundance maps are examined, it can be seen that endmembers 5 and 8 have limited value. However, their endmembers are found distinct by the unmixing algorithm. When its places are checked on images, it is seen that these places belong to small ponds having water components in the study area. This limited information can be used as a reference to see whether these endmembers are classified as water or not. If they are classified as water class, the components having these spectra already will be eliminated from the images by the suggested method.

6.2.4 Endmember Classification

Endmembers are labeled according to the material types on the image after the unmixing step. For this purpose, reference spectra (Figure 6.18) are collected on the images to represent soil types, green vegetation, dry vegetation, and water components.

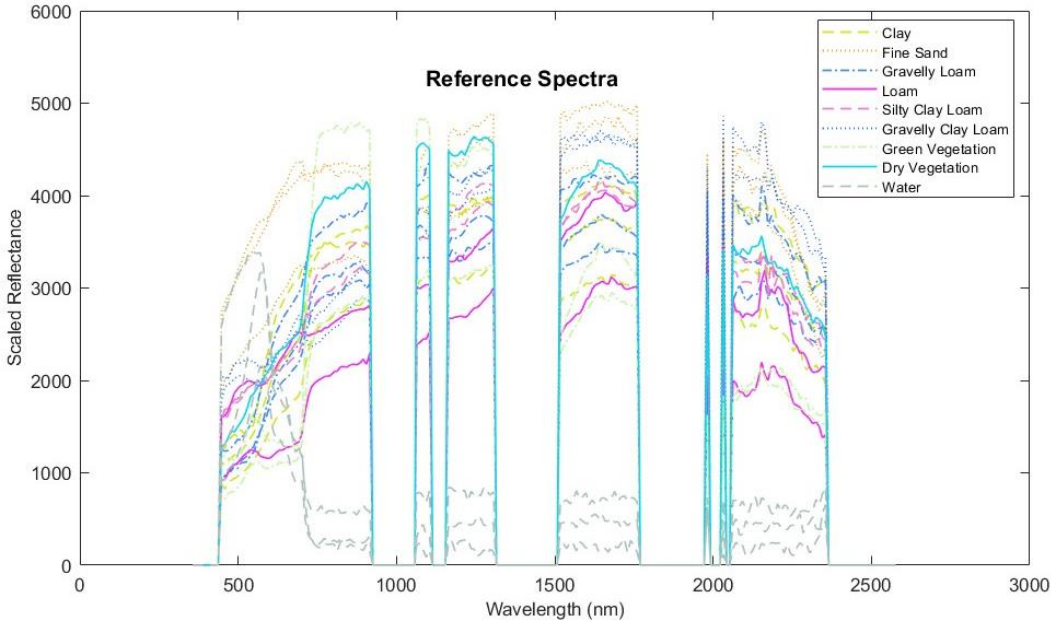

Figure 6.18 Reference spectra of materials found on the images

Estimated endmember spectra are compared with the reference spectra with the help of the SAM algorithm (Table 6.6), and endmembers are labeled (Figure 6.19).

Table 6.6 SAM results for real image analysis (Degrees)

Endmember	*Clay*	*Fine Sand*	*Gravelly Loam*	*Loam*	*Silty Clay Loam*	*Gravelly Clay Loam*	*Green Vegetation*	*Dry Vegetation*	*Water*
1	14.36	6.57	15.08	13.24	13.54	16.11	19.17	15.44	34.12
2	25.18	28.90	26.24	28.25	26.70	30.49	20.84	24.79	55.52
3	29.37	34.73	31.42	32.17	31.54	37.81	15.05	27.92	62.03
4	13.86	15.29	14.40	15.70	15.65	7.74	21.72	18.56	50.97
5	51.32	43.99	51.91	50.29	50.81	53.54	51.25	51.38	27.70
6	3.88	5.12	4.75	6.06	5.05	3.78	12.84	8.42	44.81
7	8.25	4.68	9.37	6.50	7.04	6.98	15.78	11.38	39.19
8	41.93	34.75	42.84	38.64	39.87	39.76	45.70	43.05	8.95

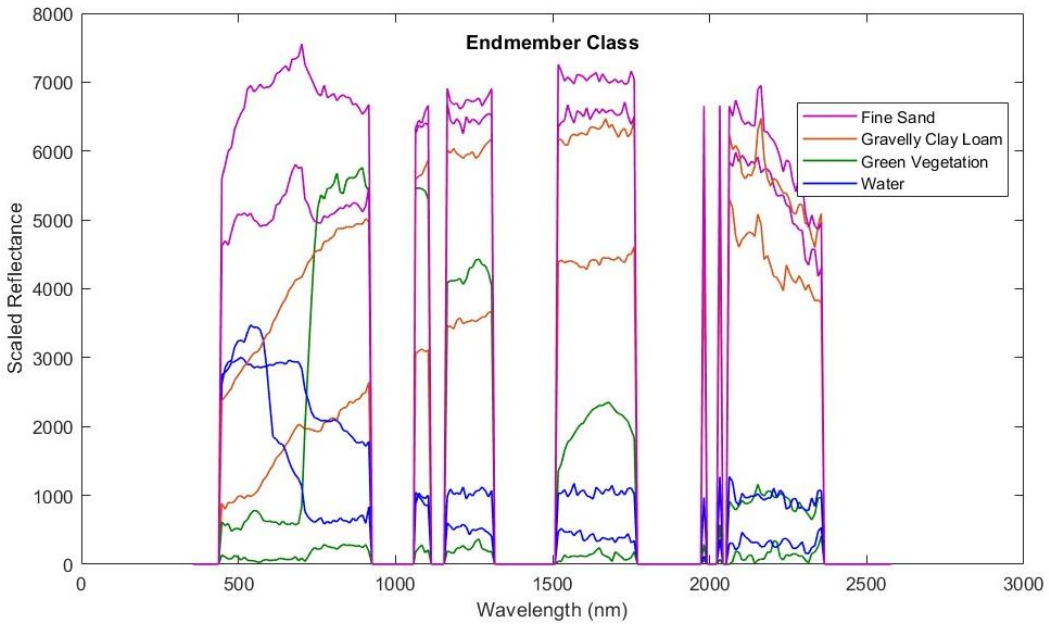

Figure 6.19 Label of endmembers

SAM results separate the endmembers into two groups: (a) stable land covers (soil, rock) and (b) unstable land covers (green/dry vegetation and water) (Figure 6.20). Fine sand and gravelly clay loam endmembers are in the soil group. Green vegetation and water are in the non-soil group.

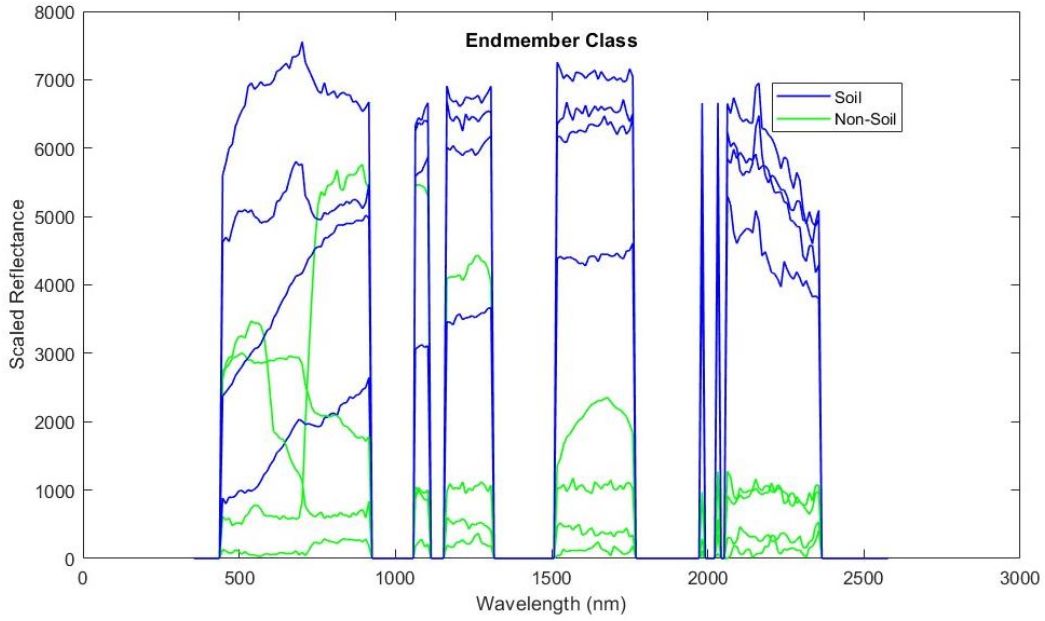

Figure 6.20 Soil and non-soil endmember groups

The critical part of the classification of endmembers is to identify non-soil spectra correctly because the study aims to remove vegetation from the images to get bare soil. Although none of the endmembers are grouped under clay, gravelly loam, loam, silty clay loam, and dry vegetation class, it can be said that other endmembers are grouped correctly according to the distance matrix of the SAM result. The results showed that the angle between the two spectra is small enough, and the differences in angles among the others are distinguishable. Endmember 1 and 7 has 6.57° and 4.68° angle with the spectral vector of fine sand. Endmember 4 and 6 has 7.74° and 3.78° angle with the spectral vector of gravelly clay loam. Endmember 2 and 3 has 20.84° and 15.05° angle with the green vegetation spectra. Endmember 5 and endmember 8 have 27.70° and 8.95° with water spectra.

6.2.5 Orthogonal Subspace Projection

Since the focus of the research is to find and remove the non-soil endmembers correctly, it does not matter to find soil endmembers correctly. To accomplish that purpose, unwanted materials like green vegetation and water are removed from the images with the OSP method by using the information about endmember groups, which is found in the previous section (Figure 6.21).

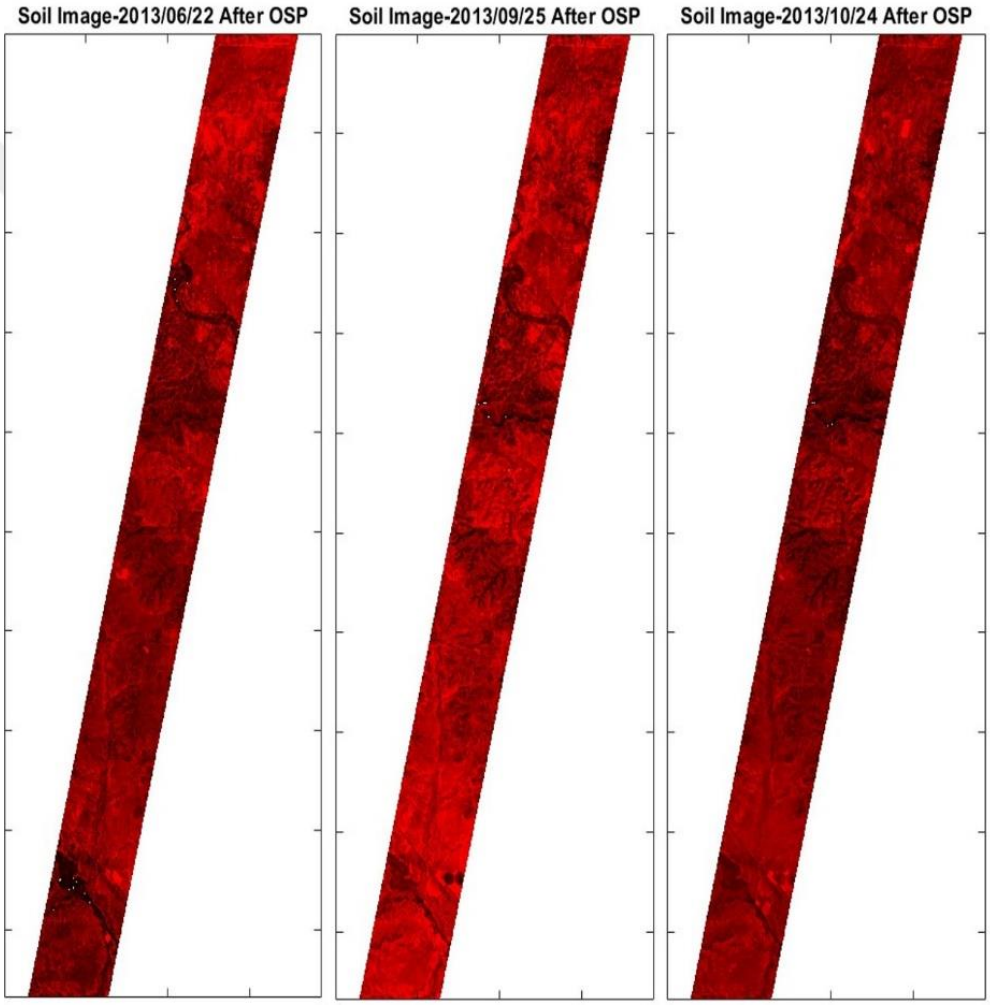

Figure 6.21 Images after OSP (Bands 20, 11, and 3 are used to represent RGB)

When the rejection vector removes unwanted components from the images during the OSP procedure, it alters whole pixel values. As a consequence, it yields RGB images in different color than expected brown color.

6.2.6 Image Fusion

While the OSP procedure removes the unstable materials, such as green vegetation and water, essentially, soil, and noise are left on the images. Each image is fused with the weighted mean algorithm using abundance values as weights to get a better signal-to-noise ratio (Figure 6.22).

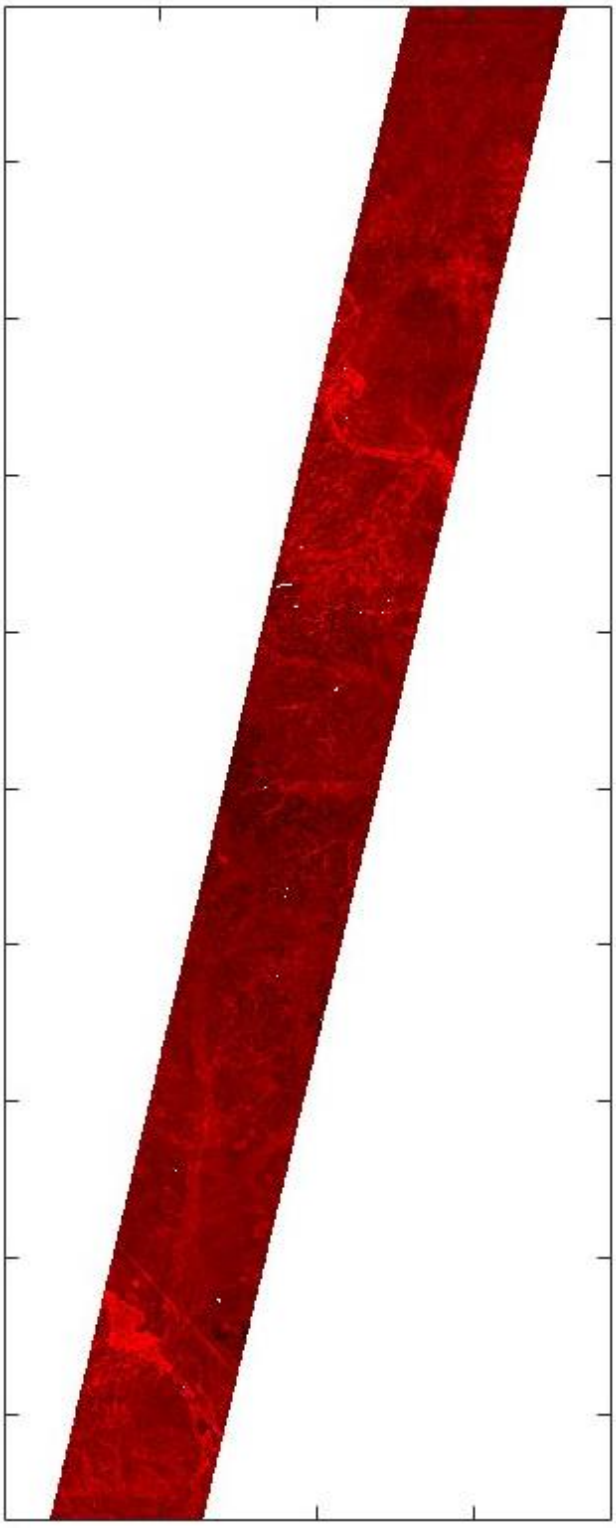

Figure 6.22 Image after the fusion process (Bands 20, 11, and 3 are used to represent RGB)

It can be seen on the fused image that the places having non-soil components, especially the places having small ponds with high abundance values, are extracted from the images.

6.2.7 Soil Classification

The RF algorithm is selected as a classifier with 200 decision trees and 1000 training pixels for each class as classification parameters, and the fused image is classified using it. One thousand training pixels are randomly chosen from each soil type.

Soil map from Digital General Soil Maps of the United States is used as the ground truth, and six different soil types are searched on the fused image (Table 6.7).

Table 6.7 Soil types in Study area

Class Number	Soil Type
1	Loam
2	Silty clay loam
3	Fine sand
4	Gravelly clay loam
5	Clay
6	Gravelly loam

The fused image is classified with 75.27% accuracy after label relaxation and post-filtering (Figure 6.23 and Figure 6.24).

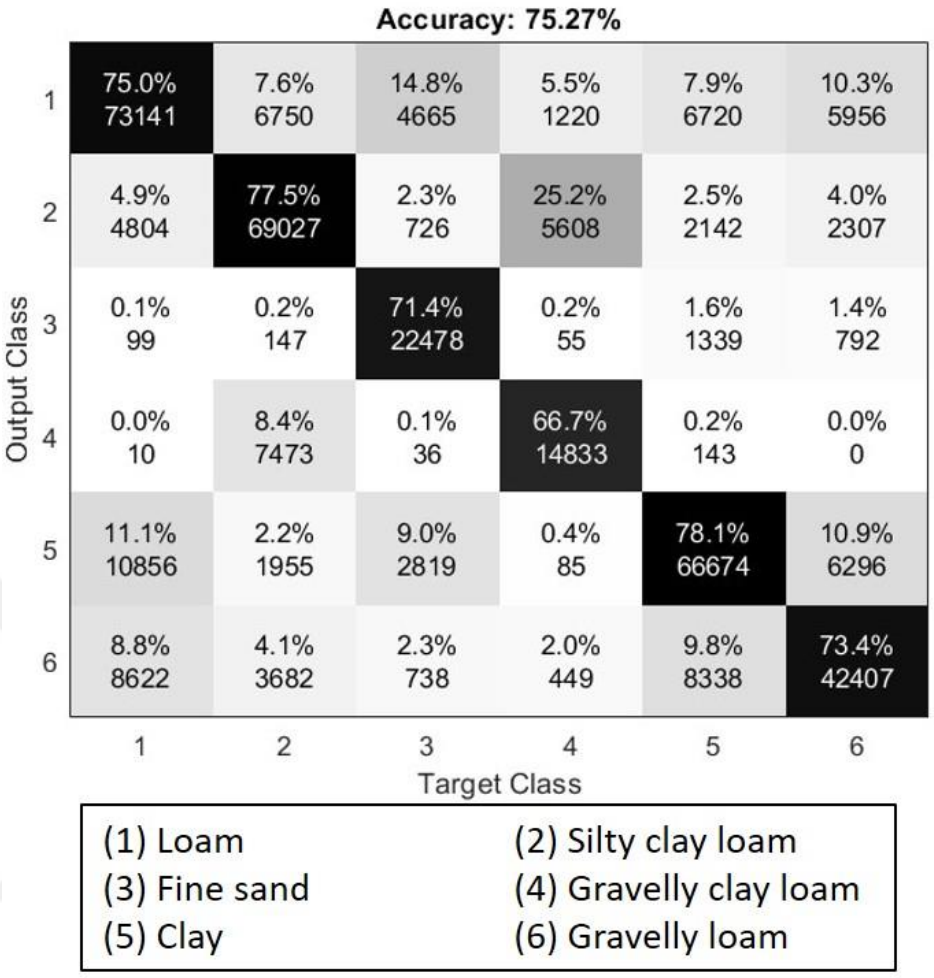

Figure 6.23 Confusion matrix of classification result

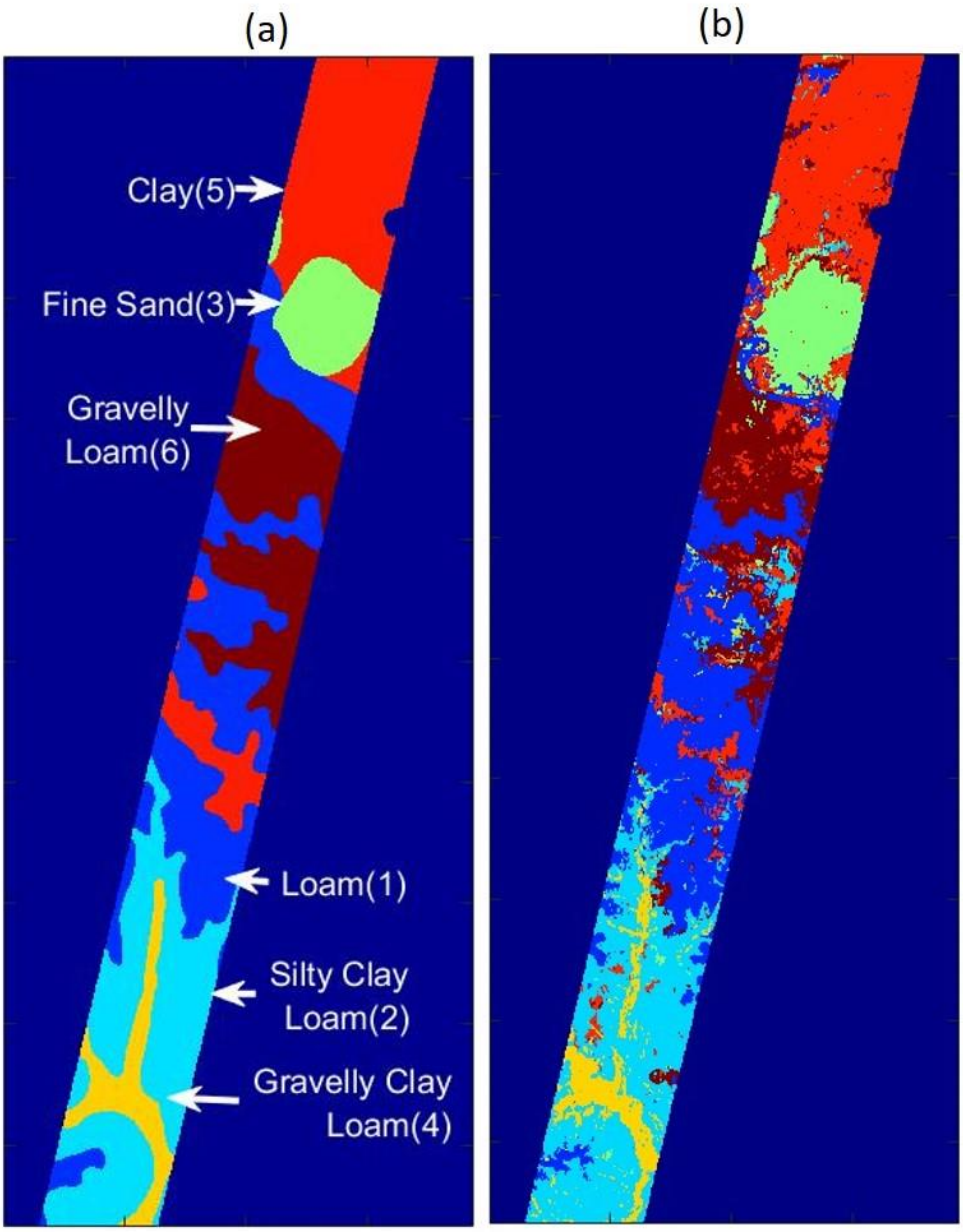

Figure 6.24 (a) Soil ground truth (b) classification result of the fused image

The classification is repeated on the three images individually after preprocessing and after OSP separately. Moreover, the same procedure is applied to the combination of two images, and classification is applied to those with the same parameters.

All results are shown in the tables (Table 6.8-6.11) below. The first and second columns of the tables are the classification results of preprocessed images before analysis. The third columns of the tables are the classification results of single images after the OSP procedure. The fourth columns of the tables are the classification results of images after fusion.

When the classification results of image pairs in Table 6.8, Table 6.9, and Table 6.10 are examined, it can be seen that the method provides better results with the suggested algorithm, even when it is applied to only two images. The accuracies after OSP are better than the results of preprocessed images, but the accuracy after fusion is better than all results.

Table 6.11 shows the results when the algorithm is applied to three images. When the results are examined closely, the classification after OSP is more successful than preprocessed image results. Classification accuracies after OSP are 71.96%, 71.78%, and 64.17%, while accuracies before OSP are 70.51%, 68.87%, and 63.18%. Fusion of images has 75.27% classification accuracy. The classification accuracy is improved when OSP is applied, and the improvement continues to increase with fusion. Also, the simple mean of preprocessed images is used for classification to understand if the improvement is due to reduced SNR from averaging or not. The result shows that the classification of the fusion of images is better than the classification of the mean of the preprocessed images.

The percentage of the improvements in Table 6.12 shows that fusion yields better improvement than other results. The mean of the preprocessed images provides a 0.93% improvement according to the mean of the classification accuracy of preprocessed images, while OSP provides a 1.78% improvement and fusion provides a 7.75% improvement. If we compare the accuracy of the mean of the preprocessed image with the mean accuracy after OSP and with the accuracy after fusion, OSP provides a 0.85% improvement, while fusion provides a 6.82% improvement. Finally, fusion provides a 5.97% improvement according to the mean of the classification accuracies after OSP.

Table 6.8 Comparison of classification accuracies in the case of fusion of two images from June and September of 2013

Image	*Clasf. Accuracy of Preprocessed Image*	*Clasf. Accuracy of Mean of Preprocessed Image*	*Clasf. Accuracy of OSP Applied Image*	*Clasf. Accuracy of Fusion*
22.06.2013	0.7051	0.7031	0.7067	0.7549
25.09.2013	0.6887		0.7238	

Table 6.9 Comparison of classification accuracies in the case of fusion of two images from June and October of 2013

Image	*Clasf. Accuracy of Preprocessed Image*	*Clasf. Accuracy of Mean of Preprocessed Image*	*Clasf. Accuracy of OSP Applied Image*	*Clasf. Accuracy of Fusion*
22.06.2013	0.7051	0.6846	0.7139	0.7232
24.10.2013	0.6318		0.6493	

Table 6.10 Comparison of classification accuracies in the case of fusion of two images from September and October of 2013

Image	*Clasf. Accuracy of Preprocessed Image*	*Clasf. Accuracy of Mean of Preprocessed Image*	*Clasf. Accuracy of OSP Applied Image*	*Clasf. Accuracy of Fusion*
25.09.2013	0.6887	0.6571	0.7115	0.7155
24.10.2013	0.6318		0.6624	

Table 6.11 Comparison of classification accuracies of real images in the case of three image fusion for year 2013

Image	*Clasf. Accuracy of Preprocessed Image*	*Clasf. Accuracy of Mean of Preprocessed Image*	*Clasf. Accuracy of OSP Applied Image*	*Clasf. Accuracy of Fusion*
22.06.2013	0.7051		0.7196	
25.09.2013	0.6887	0.6845	0.7178	0.7527
24.10.2013	0.6318		0.6417	

Table 6.12 Improvements of classification accuracy

Classification Accuracy	*Clasf. Accuracy of Mean of Preprocessed Image*	*Clasf. Accuracy of OSP Applied Image*	*Clasf. Accuracy of Fusion*
Improvement to Preprocessed Image	0.93%	1.78%	7.75%
Improvement to Mean of Preprocessed Image		0.85%	6.82%
Improvement to OSP Applied Image			5.97%

The suggested method is also applied to different triple images from the year 2016. Similar results also are found when the method is applied to them (Table 6.13).

Table 6.13 Comparison of classification accuracies of real images in the case of three image fusion for year 2016

Image	*Clasf. Accuracy of Preprocessed Image*	*Clasf. Accuracy of Mean of Preprocessed Image*	*Clasf. Accuracy of OSP Applied Image*	*Clasf. Accuracy of Fusion*
17.02.2016	0.5707		0.5899	
12.08.2016	0.6281	0.6408	0.6393	0.6902
16.09.2016	0.6380		0.6633	

To see the effect of the algorithm on the parts of the image having dense vegetation levels, the normalized difference vegetation index (NDVI) is calculated. The places having NDVI greater than 0.4 are found. These places have highly dense vegetation or forest. Classification results on these parts are examined (Table 6.14 and Table 6.15). The results showed that the suggested method is successful on these parts of the images. Classification accuracies are considerably improved. Especially the result of the fused image from 2016 has remarkable improvement. While overall accuracy for the whole part of the image is 69.03%, the areas having NDVI greater than 0.4 have 73.06% classification accuracy. Similarly, the fused image from 2013 has an improvement in classification accuracy at these parts of the image.

Table 6.14 Classification results for the places where NDVI is greater than 0.4 for 2013

Image	*Original Clasf. For Pixels NDVI is Greater Than 0.4*	*OSP Clasf. For Pixels NDVI is Greater Than 0.4*	*Fused Im. Claf.For Pixels NDVI is Greater Than 0.4*
22.06.2013	0.7054	0.718	
25.09.2013	0.6888	0.7043	0.7578
24.10.2013	0.6405	0.6642	

Table 6.15 Classification results for the places where NDVI is greater than 0.4 for 2016

Image	*Original Clasf. For Pixels NDVI is Greater Than 0.4*	*OSP Clasf. For Pixels NDVI is Greater Than 0.4*	*Fused Im. Claf.For Pixels NDVI is Greater Than 0.4*
17.02.2016	0.5417	0.5754	
12.08.2016	0.6409	0.6627	0.7306
19.09.2016	0.6754	0.6929	

The suggested algorithm is applied to images from 2013, and after successful results, the experiment is tested on the images from 2016. The results show that the method is successful, and it improves classification results after the OSP step and then after the fusion step with real images.

CHAPTER 7

CONCLUSION

Accurate and reliable soil maps are vital for many applications. However, vegetation covering the study area is the main obstacle to obtaining local or global soil maps from spaceborne hyperspectral images. To handle this problem, an automated method to eliminate unwanted components from the image that exploits the time series images is suggested. Unmixing algorithms are applied to the concatenated images to detect endmembers common to all images. After classifying the endmembers as stable and unstable components, unstable components were projected and extracted from the images. The remaining images were fused into one image and classified according to the soil properties. The accuracy of the classification, which is measured using existing soil maps, is very promising. The experiment on the synthetic images gives successful results. When the suggested method is applied to the real hyperspectral images, the same result is obtained. It can be observed that the method improves classification accuracy after the OSP step and then after fusion for the concatenated images selected from 2013 and 2016. It is possible that if endmembers are estimated correctly, the method provides an enhancement in the classification accuracy with OSP and fusion. Another critical point for better classification accuracy is the radiometric differences between the images. Because of the fact that the images from different dates are used in the suggested method, the radiometric differences should be modeled before the application of the technique.

In the future study, the algorithm should be repeated on the hyperspectral images from different platforms and the hyperspectral images using better atmospheric correction methods.

New generation hyperspectral satellite sensors will have an essential role in soil classification by promising detailed monitoring of Earth. Effective and automated methods for the extraction of soil targets will provide more accurate information about soil properties.

REFERENCES

Ashraf, M.A., Maah, M.J. and Yusoff, I., 2011. Introduction to remote sensing of biomass. In Biomass and remote sensing of biomass (pp. 129-170). IntechOpen.

Asner, G.P. and Heidebrecht, K.B., 2002. Spectral unmixing of vegetation, soil and dry carbon cover in arid regions: comparing multispectral and hyperspectral observations. International Journal of Remote Sensing, 23(19), pp.3939-3958.

Baldridge, A. M., S.J. Hook, C.I. Grove and G. Rivera, 2009. The ASTER Spectral Library Version 2.0. Remote Sensing of Environment, vol 113, pp. 711-715.

Basile, R.M., 1971. A geography of soils. WCB/McGraw-Hill.

Ben-Dor, E., Irons, J.R. and Epema, G.F., 1999. Soil reflectance. Remote sensing for the earth sciences: Manual of remote sensing, 3, pp.111-188.

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. International Journal of Remote Sensing, 23(6), pp.1043-1062.

Bernstein, L.S., Jin, X., Gregor, B. and Adler-Golden, S.M., 2012. Quick atmospheric correction code: algorithm description and recent upgrades. Optical engineering, 51(11), p.111719.

Bioucas-Dias, J.M., 2009, August. A variable splitting augmented Lagrangian approach to linear spectral unmixing. In 2009 First workshop on hyperspectral image and signal processing: Evolution in remote sensing (pp. 1-4). IEEE.

Bioucas-Dias, J.M. and Figueiredo, M.A., 2010, June. Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing. In 2010 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (pp. 1-4). IEEE.

Breiman, L. and Cutler A. Random Forests. [online] Available at: <https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm#ooberr> [Accessed 8 August 2022].

Dhawale, N. M., Adamchuk, V. I., Prasher, S. O., Rossel, R. A. V., and Ismail, A. A., 2022. Evaluation of two portable hyperspectral-sensor-based instruments to predict key soil properties in Canadian soils. Sensors, 22(7), p.2556.

Diaz-Gonzalez, F. A., Vuelvas, J., Correa, C. A., Vallejo, V. E., and Patino, D., 2022. Machine learning and remote sensing techniques applied to estimate soil indicators–review. Ecological Indicators, 135, p.108517.

Escribano, P., Schmid, T., Chabrillat, S., Rodríguez-Caballero, E. and Garcia, M., 2017. Optical remote sensing for soil mapping and monitoring. In Soil mapping and process modeling for sustainable land use management (pp. 87-125). Elsevier.

FAO, 2022a. Biological properties. [online] Available at: <https://www.fao.org/soils-portal/data-hub/soil-properties/biological-properties/en/> [Accessed 5 August 2022].

FAO, 2022b. Chemical properties. [online] Available at: <https://www.fao.org/soils-portal/data-hub/soil-classification/numerical-systems/chemical-properties/en /> [Accessed 5 August 2022].

FAO, 2022c. FAO/UNESCO Soil Map of the World. [online] Available at: <https://www.fao.org/soils-portal/data-hub/soil-classification/numerical-systems/chemical-properties/en/> [Accessed 5 August 2022].

FAO, 2022d. Physical properties. [online] Available at: <https://www.fao.org/soils-portal/data-hub/soil-properties/physical-properties/en/> [Accessed 5 August 2022].

FAO, 2022e. Soil properties. [online] Available at: <https://www.fao.org/soils-portal/data-hub/soil-properties/en/> [Accessed 5 August 2022].

FAO, 2022f. Soil structure. [online] Available at: <https://www.fao.org/fishery/docs/CDrom/FAO_Training/FAO_Training/General/x6706e/x6706e07.htm > [Accessed 5 August 2022].

FAO, 2022g. Soil texture. [online] Available at: <https://www.fao.org/fishery/docs/CDrom/FAO_Training/FAO_Training/General/x6706e/x6706e06.htm > [Accessed 5 August 2022].

FAO, 2022h. FAO/UNESCO Soil Map of the World. [online] Available at: <https://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/faounesco-soil-map-of-the-world/en/> [Accessed 5 August 2022].

Forkuor, G., Hounkpatin, O. K., Welp, G., and Thiel, M., 2017. High resolution mapping of soil properties using remote sensing variables in south-western Burkina Faso: a comparison of machine learning and multiple linear regression models. PloS one, 12(1), e0170478.

Gardner, C. M., Laryea, K. B., and Unger, P. W., 1999. Soil physical constraints to plant growth and crop production (Vol. 11). Rome, Italy: Land and Water Development Division, Food and Agriculture Organization.

Gillis, D., Bowles, J., Ientilucci, E.J. and Messinger, D.W., 2007, September. Linear unmixing using endmember subspaces and physics based modeling. In Imaging Spectrometry XII (Vol. 6661, p. 66610E). International Society for Optics and Photonics.

Gillis, D., Bowles, J., Ientilucci, E.J. and Messinger, D.W., 2008, April. A generalized linear mixing model for hyperspectral imagery. In Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIV (Vol. 6966, p. 69661B). International Society for Optics and Photonics.

Gruninger, J. H., Ratkowski, A. J., and Hoke, M. L. 2004, August. The sequential maximum angle convex cone (SMACC) endmember model. In Algorithms and technologies for multispectral, hyperspectral, and ultraspectral imagery X (Vol. 5425, pp. 1-14). SPIE.

Harrisgeospatial.com. 2020. Bad Data Mitigation Tools. [online] Available at: <https://www.harrisgeospatial.com/docs/THORBadDataTools.html> [Accessed 5 May 2020].

Harrisgeospatial.com. 2020. Radiometric Calibration. [online] Available at: <https://www.harrisgeospatial.com/docs/RadiometricCalibration.html> [Accessed 5 May 2020].

Harsanyi, J.C. and Chang, C.I., 1994. Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach. IEEE Transactions on geoscience and remote sensing, 32(4), pp.779-785.

Hartemink, A.E., Krasilnikov, P. and Bockheim, J.G., 2013. Soil maps of the world. Geoderma, 207, pp.256-267.

Heylen, R., Parente, M., and Gader, P., 2014. A review of nonlinear hyperspectral unmixing methods. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7(6), pp.1844-1868.

Ho, T.K., 1995, August. Random decision forests. In Proceedings of 3rd international conference on document analysis and recognition (Vol. 1, pp. 278-282). IEEE.

Huete, A.R., 2004. Remote sensing for environmental monitoring. In Environmental Monitoring and Characterization (pp. 183-206). Academic Press.

ISRIC 2022. Soter Databases. [online] Available at: <https://www.isric.org/explore/soter> [Accessed 9 September 2022].

Jain, S.K. and Singh, V.P., 2003. Reservoir sedimentation. In Developments in Water Science (Vol. 51, pp. 681-741). Elsevier.

Kokaly, R.F., Clark, R.N., Swayze, G.A., Livo, K.E., Hoefen, T.M., Pearson, N.C., Wise, R.A., Benzel, W.M., Lowers, H.A., Driscoll, R.L., and Klein, A.J.,

2017, USGS Spectral Library Version 7: U.S. Geological Survey Data Series 1035, 61 p., https://doi.org/10.3133/ds1035.

Li, J., Agathos, A., Zaharie, D., Bioucas-Dias, J. M., Plaza, A., & Li, X. 2015. Minimum Volume Simplex Analysis: A Fast Algorithm for Hyperspectral Endmember Extraction. IEEE Transactions on Geoscience and Remote Sensing, pp.5067-5082.

Li, J., Bioucas-Dias, J.M., Plaza, A. and Liu, L., 2016. Robust collaborative nonnegative matrix factorization for hyperspectral unmixing. IEEE Transactions on Geoscience and Remote Sensing, 54(10), pp.6076-6090.

Li, X., Fan, P., Li, Z., Chen, G., Qiu, H., & Hou, G. 2021. Soil classification based on deep learning algorithm and visible near-infrared spectroscopy. Journal of Spectroscopy, 2021, pp.1-11.

Lumme, J.H., 2004, July. Classification of vegetation and soil using imaging spectrometer data. In Geoscience and remote sensing symposium (p. 83).

Margate, D.E. and Shrestha, D.P., 2001, November. The use of hyperspectral data in identifying 'desert-like'soil surface features in Tabernas area, southeast Spain. In The 22nd Asian Conference on Remote Sensing (pp. 5-9).

Martínez, P.J., Pérez, R.M., Plaza, A., Aguilar, P.L., Cantero, M.C. and Plaza, J., 2006. Endmember extraction algorithms from hyperspectral images. Annals of Geophysics, 49(1).

Meerdink, S. K., Hook, S. J., Roberts, D. A., and Abbott, E. A., 2019. The ECOSTRESS spectral library version 1.0. Remote Sensing of Environment, 230, p.111196.

Meng, X., Bao, Y., Ye, Q., Liu, H., Zhang, X., Tang, H., and Zhang, X., 2021. Soil organic matter prediction model with satellite hyperspectral image based on optimized denoising method. Remote Sensing, 13(12), p.2273.

Meng, X., Bao, Y., Liu, H., Zhang, X., and Wang, X., 2022. A new digital soil mapping method with temporal-spatial-spectral information derived from multi-source satellite images. Geoderma, 425, p.116065.

Nachtergaele, F.O. and Van Ranst, E., 2003. Qualitative and quantitative aspects of soil databases in tropical countries. Evolution of tropical soil science: Past and future. Koninklijke Academie voor Overzeese Wetenschappen, Brussels, pp.107-126.

Nanni, M. R., Demattê, J. A. M., Rodrigues, M., Santos, G. L. A. A. D., Reis, A. S., Oliveira, K. M. D., Cesar, E., Furlanetto, R. H., Crusiol, L. G. T., and Sun, L., 2021. Mapping particle size and soil organic matter in tropical soil based on hyperspectral imaging and non-imaging sensors. Remote Sensing, 13(9), p.1782.

Nascimento, J.M. and Dias, J.M., 2005. Vertex component analysis: A fast algorithm to unmix hyperspectral data. IEEE transactions on Geoscience and Remote Sensing, 43(4), pp.898-910.

Nascimento, J., and Martin, G., 2019. Nonlinear spectral unmixing. Data Handling in Science and Technology, Elsevier, Vol. 32, pp. 151-166.

Navalgund, R.R., 2001. Remote sensing. Resonance, 6(12), pp.51-60.

Okin, G.S., Roberts, D.A., Murray, B. and Okin, W.J., 2001. Practical limits on hyperspectral vegetation discrimination in arid and semiarid environments. Remote Sensing of Environment, 77(2), pp.212-225.

Pandey, A., Kumar, D., and Chakraborty, D. B., 2021, July. Soil Type Classification from High Resolution Satellite Images with Deep CNN. In 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS (pp. 4087-4090). IEEE.

Peyghambari, S. and Zhang, Y., 2021. Hyperspectral remote sensing in lithological mapping, mineral exploration, and environmental geology: an updated review. Journal of Applied Remote Sensing, 15(3), pp. 031501-031501.

Prudnikova, E., Savin, I., Vindeker, G., Grubina, P., Shishkonakova, E. and Sharychev, D., 2019. Influence of soil background on spectral reflectance of winter wheat crop canopy. Remote Sensing, 11(16), p.1932.

Rajitha, A., Bhargavi, P., and Jyothi, S., 2022. A Survey On Soil Classification Using Hyperspectral Images. Webology (ISSN: 1735-188X), 19(2).

Rashmi, S., Addamani, S., Venkat and Ravikiran S., 2014. Spectral Angle Mapper Algorithm for Remote Sensing Image Classification", IJISET-International Journal of Innovative Science, Engineering & Technology, 1, pp.201-205.

Samadi, L., and Samadi, H., 2022. Soil Classification Modelling Using Machine Learning Methods. database, 9, p.11.

Scheffler, D. and Karrasch, P., 2013, October. Preprocessing of hyperspectral images: a comparative study of destriping algorithms for EO1-hyperion. In Image and Signal Processing for Remote Sensing XIX (Vol. 8892, p. 88920H). International Society for Optics and Photonics.

Soil Science Division Staff. 2017. Soil survey manual. C. Ditzler, K. Scheffe, and H.C. Monger (eds.). USDA Handbook 18. Government Printing Office, Washington, D.C.

Soil Survey Staff. 2022. Natural Resources Conservation Service, United States Department of Agriculture. Web Soil Survey. [online] Available at: <https://websoilsurvey.nrcs.usda.gov/> [Accessed 1 March 2023].

Soltani, I., Fouad, Y., Michot, D., Bréger, P., Dubois, R. and Cudennec, C., 2019. A near infrared index to assess effects of soil texture and organic carbon content on soil water content. European journal of soil science, 70(1), pp.151-161.

Summers, D., Lewis, M., Ostendorf, B. and Chittleborough, D., 2011. Unmixing of soil types and estimation of soil exposure with simulated hyperspectral imagery. International journal of remote sensing, 32(21), pp.6507-6526.

Tziolas, N., Tsakiridis, N., Chabrillat, S., Demattê, J. A., Ben-Dor, E., Gholizadeh, A., Zalidis, G., Van Wesemael, B., 2021. Earth observation data-driven cropland soil monitoring: A review. Remote Sensing, 13(21), p.4439.

Uddin, M., and Hassan, M. R., 2022. A novel feature based algorithm for soil type classification. Complex & Intelligent Systems, 8(4), pp. 3377-3393.

Vibhute, A. D., Kale, K. V., Dhumal, R. K., and Mehrotra, S. C., 2015, December. Soil type classification and mapping using hyperspectral remote sensing data. In 2015 International Conference on Man and Machine Interfacing (MAMI) (pp. 1-4). IEEE.

Wilde, M., and Bardsley J., 2010. Label relaxation source code, (final version) [Source code].

Young, R., 2012. Soil properties and behaviour (Vol. 5). Elsevier.

Zare, A., and Ho, K. C. 2013. Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing. IEEE Signal Processing Magazine, 31(1), pp. 95-104.

Zhang, X., Vijayaraj, V. and Younan, N.H., 2003, October. Hyperspectral soil texture classification. In IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data, 2003 (pp. 182-186). IEEE.

Zhang, X., Younan, N.H. and O'Hara, C.G., 2005. Wavelet domain statistical hyperspectral soil texture classification. IEEE Transactions on geoscience and remote sensing, 43(3), pp.615-618.

CURRICULUM VITAE

Surname, Name: Kaba, Eylem