

UNDERSTANDING THE EFFECT OF SPATIAL RESOLUTION
ON UNMIXING ALGORITHMS FOR HYPERSPECTRAL
IMAGERY

by

Andrea Santos García

A thesis submitted in fulfillment of the requirements for the degree of

MASTER OF SCIENCE

in

ELECTRICAL ENGINEERING

University of Puerto Rico
Mayagüez Campus

2010

Approved by:

Yahya M. Masalmah, Ph.D
Member, Graduate Committee

Date

Skip Van Bloem, Ph.D
Member, Graduate Committee

Date

Shawn D. Hunt, Ph.D
Member, Graduate Committee

Date

Miguel Vélez-Reyes, Ph.D
President, Graduate Committee

Date

Paul A. Sundaram, Ph.D
Representative of Graduate Studies

Date

Hamed Parsiani, Ph.D
Chairperson of the Department

Date

Abstract of THESIS OR DISSERTATION Presented to the Graduate School
of the University of Puerto Rico in Partial Fulfillment of the
Requirements for the Degree of MASTER OF SCIENCE

UNDERSTANDING THE EFFECT OF SPATIAL RESOLUTION
ON UNMIXING ALGORITHMS FOR HYPERSPECTRAL
IMAGERY

by
Andrea Santos García
2010

Chair: Miguel Vélez Reyes
Major Department: Electrical and Computer Engineering

This research work presents an experimental study on unmixing of hyperspectral images at different spatial resolutions. The study looks at how the assumptions made in the unmixing procedure affects the results of the inversion. Different algorithms such as Constrained Positive Matrix Factorization (cPMF), Sequential Maximum Angle Convex Cone (SMACC), and Maximum distance (MaxD) are studied and compared. We also study the problem of estimating the number of endmembers using the rank and the positive-rank of a matrix. The rank is estimated by the Scree test and the positive rank is estimated with a search procedure. Hyperspectral Imagery from the Guanica Dry Forest at one and four meters and Enrique Reef Area in Parguera at one, two, and eight meters, were used in the experiments. The data was collected using the AISA sensor during the 2007 Puerto Rico Hyperspectral Mission. Field work in Guanica area and previous class maps of Enrique Reef area were used in validation of the results.

Resumen de TESIS O DISERTACION Presentado a Escuela Graduada
de la Universidad de Puerto Rico como requisito de los
Requerimientos para el grado de MAGISTER EN CIENCIA

UN MARCO DE TRABAJO PARA EL MONITOREO DE ANUROS,
CLASIFICACIÓN DE ESPECIES Y ESTUDIOS BIOACUSTICOS

por
Andrea Santos García
2010

Consejero: Miguel Vélez Reyes
Departamento: Ingeniería Eléctrica y de Computadoras

Este trabajo de investigación presenta un estudio experimental acerca del desmezcado de imágenes hiperspectrales a diferentes resoluciones espaciales. El estudio muestra de que forma los supuestos establecidos durante el proceso de desmezcado afectan los resultados del proceso de inversión. Diferentes algoritmos como la Factorización matricial positiva con restricciones (cPMF), Cono convexo secuencial de máximo ángulo (SMACC) y Distancia máxima (MaxD) son estudiados y comparados. Además se estudia el problema de la estimación del número de firmas espectrales clase, utilizando la aproximación del rango y el rango positivo. El rango se estima usando un método llamado Test Scree y el rango-positivo se estima hallando un valor de error adecuado en los resultados obtenidos con la Factorización matricial positiva con restricciones con respecto al cambio en el número de las firmas espectrales clase usadas. Imágenes del área de Guánica a uno y cuatro metros y del área del Cayo Enrique en Parguera, a uno, dos y ocho metros, son usadas para experimentar. Los datos fueron coleccionados por el sensor AISA durante la misión hiperspectral de Puerto Rico en 2007. Trabajo de campo en el área de Guánica y mapas previos de clases del área del Cayo Enrique fueron utilizadas para realizar un reconocimiento de las características principales de las áreas de estudio.

*To my parents Rosalba and Manuel
for teaching me to believe in my dreams and strive to achieve them,
to my husband to hold my hand every time that I need it,
to my sisters Mónica and Jimena, and my new family for their unconditional
support*

ACKNOWLEDGEMENTS

Thanks to Puerto Rico for making possible that my dreams come true.

Thanks to CenSSIS to support me during these 2 and a half years. Thanks to Dr. Miguel Velez-Reyes to be an excellent advisor. Thanks to the members of my committee: Dr. Shawn Hunt, Dr. Yahya Masalmah and Dr. Skip Van Bloem. Thanks to Sandy just for everything. Thanks to CENSSIS and IRISE staff: Marie, Samuel, Victor, Maribel, Vanessa, Yairanise , Soliris and Evelyn . Thanks to my partners at LARSIP, specially for all the coffee cups that we shared, and my good friends Ineris, Héctor and Jacqueline. Thanks to my new and old family. And finally thanks to my husband , my unconditional support everyday .

TABLE OF CONTENTS

	<u>page</u>
Abstract English	ii
Resumen	iii
ACKNOWLEDGEMENTS	v
LIST OF TABLES	viii
LIST OF FIGURES	ix
1 INTRODUCTION	1
1.1 Justification	1
1.2 Objectives	2
1.2.1 General Objective	2
1.2.2 Specific Objectives	2
1.3 Contribution of this work	3
1.4 Outline	3
2 BACKGROUND AND LITERATURE REVIEW	5
2.1 Hyperspectral images	5
2.1.1 Examples of Hyperspectral Sensors	6
2.2 Linear Mixing Model	9
2.3 Spectral Unmixing Problem	10
2.3.1 Estimation of the Number of Endmembers	10
2.3.2 Endmembers Extraction Methods	11
2.3.3 Abundance Estimation	17
2.3.4 Full Unmixing Algorithms	20
2.3.5 Constrained Approaches for Positive Matrix Factorization	22
3 UNMIXING ALGORITHMS	25
3.1 Linear Mixing Model as Forward Problem	25
3.2 Unmixing as an Inverse Problem	26
3.2.1 Estimation of the number of Endmembers and its relation with the spatial resolution	28
3.2.2 Endmembers Extraction and Abundance Estimation and its relation with the spatial resolution	30

4	EXPERIMENTS AND ANALYSIS OF THE RESULTS	33
4.1	Methodology	33
4.2	Experiment 1 - Guanica Dry Forest Area	34
	4.2.1 Imagery	34
	4.2.2 Results and Discussion	38
4.3	Experiment 2 - Enrique Reef Area	56
	4.3.1 Imagery	56
	4.3.2 Results and Discussion	58
5	CONCLUSIONS AND FUTURE WORK	78
5.1	Conclusions	78
5.2	Future Work	79

LIST OF TABLES

<u>Table</u>	<u>page</u>
3-1 Comparison between Endmembers Extraction Methods	32
4-1 Comparison of results with 3 endmembers at different spatial resolution using cosine angle distance	43
4-2 Cosine Angle distance for results with 3 endmembers at 4 meters. . . .	45
4-3 Cosine Angle distance for results with 3 endmembers at 1 meter	45
4-4 Comparison of results with 5 endmembers at different spatial resolution using cosine angle distance	50
4-5 Cosine Angle distance for results with 5 endmembers at 4 meters	50
4-6 Cosine Angle distance for results with 5 endmembers at 1 meter	51
4-7 Comparison of the fraction areas covered by the endmembers in Guanica Dry Forest	51
4-8 Cosine Angle distance for 4 endmembers results, for Enrique Reef at 1 meter	67
4-9 Cosine Angle distance for 4 endmembers results, for Enrique Reef at 2 meters	67
4-10 Cosine Angle distance for 4 endmembers results, for Enrique Reef at 8 meters	67
4-11 Cosine Angle distance for 5 endmembers results, for Enrique Reef at 1 meter	74
4-12 Cosine Angle distance for 5 endmembers results, for Enrique Reef at 2 meters	74
4-13 Cosine Angle distance for 5 endmembers results, for Enrique Reef at 8 meters	75

LIST OF FIGURES

<u>Figure</u>	<u>page</u>
2-1 Graphical representation of a pixel in a hyperspectral image	6
2-2 Graphic representation of the ideal Scree test model based on [15] . . .	12
2-3 Graphical representation of the simplex with maximum volume based on the graphics of [23]. Points A, B, and C form a simplex but it does not have the maximum volume. The points A, B, D are the three true endmembers for the image and form the simplex with the maximum volume.	13
2-4 Graphical representation of the orthogonal projection for the Maximum Distance method.	14
2-5 Graphical representation of the cocktail party problem. Taken from Encyclopedia of Statistics in Behavioral Science [25]	15
2-6 (a) Reflectances of carnallite, ammonioalunite, and biotite.(b) Two-dimensional scatterplot of mixtures of the three endmembers shown in (a). Taken from Nascimento and Bioucas [21]	16
3-1 Linear Mixing Model as a forward problem	25
3-2 Geometric description of a single detector element in the focal plane of an optical sensor. Taken from [35]	26
3-3 Unmixing Model as an inverse problem	27
3-4 Unmixing Model for images at different spatial resolution	28
3-5 Graphical representation of the Scree model for hyperspectral imagery	30
3-6 Fitting error model curve.	31
4-1 Fitting error model curve.	33
4-2 Flight lines over Guanica area in 2007 Puerto Rico Hyperspectral Mission. (a)Flight lines at 1 meter. (b)Flight lines at 4 meters. . .	34
4-3 Guanica Dry Forest Image. (a)True Color Composite of image at 1 meter. (b) True Color Composite of image at 4 meters. (d)Color Infrared Composite of image at 1 meter (e)Color Infrared Composite of image at 4 meters.	35

4-4	Interpretation of the Guanica Dry forest image based on fieldwork at the area.	36
4-6	Photos and Spectral signature Bourreria Suculenta and Exostema Caribeaum	36
4-5	Photos and Spectral signature of Buffel grass and some images of grass field.	37
4-7	Photos and Spectral signature Prosopis Procera	37
4-8	Scree test for Guanica subset images at one and four meters.	38
4-9	Fitting Error for Guanica Dry Forest Image at 4 meters and 1 meter.	39
4-10	Endmembers obtained with different endmembers extraction algorithms. Results for image at 4 meters and 1 meter of spatial resolution.	41
4-11	Unmixing results using cPMF with 3 endmembers from Guanica Area at one and four meters.	41
4-12	Unmixing results using SMACC with 3 endmembers from Guanica Area at one and four meters	42
4-13	Unmixing results using MaxD with 3 endmembers from Guanica Area at one and four meters	42
4-14	Abundance Maps for endmember 1 using cPMF, SMACC and MaxD.	43
4-15	Abundance Maps for endmember 2 using cPMF, SMACC and MaxD.	44
4-16	Abundance Maps for endmember 3 using cPMF, SMACC and MaxD.	44
4-17	Endmembers obtained with different endmembers extraction algorithms. Results for image at 4 meters and 1 meter of spatial resolution.	46
4-18	Unmixing results using cPMF with 5 endmembers. (a) Results for image at 4 meters and (b) 1 meter.	46
4-19	Unmixing results using SMACC with 5 endmembers. (a) Results for image at 4 meters and (b) 1 meter.	47
4-20	Unmixing results using MaxD with 5 endmembers. (a) Results for image at 4 meters and (b) 1 meter.	47
4-21	Abundance Maps for endmember 1 using cPMF, SMACC, and MaxD	48
4-22	Abundance Maps for endmember 2 using cPMF, SMACC, and MaxD	48

4-23	Abundance Maps for endmember 3 using cPMF, SMACC, and MaxD	49
4-24	Abundance Maps for endmember 4 using cPMF, SMACC, and MaxD	49
4-25	Abundance Maps for endmember 5 using cPMF, SMACC, and MaxD	50
4-26	Abundance estimation using image at 1 meter and endmembers obtained by cPMF for 4 meters	52
4-27	Abundance estimation using image at 1 meter and endmembers obtained by SMACC for 4 meters	53
4-28	Abundance estimation using image at 1 meter and endmembers obtained by MaxD for 4 meters	53
4-30	Unmixing results using PMF with 3 endmembers from Guanica at one and four meters.	54
4-29	Endmembers obtained with PMF and cPMF algorithms. Results for images at 4 meters and 1 meter spatial resolution.	55
4-32	Unmixing results using PMF with 5 endmembers. (a) Results for image at 4 meters and (b) 1 meter.	55
4-31	Endmembers obtained with PMF and cPMF algorithms. Results for images at 4 meters and 1 meter spatial resolution.	56
4-33	Enrique reef image at three different spatial resolutions.	57
4-34	Class map from Enrique Reef at 1 meter spatial resolution.	57
4-35	High definition image from Google Earth.	58
4-36	Scree Test for Enrique Reef image at different spatial resolutions.	58
4-37	Fitting error curves of Enrique Reef Images.	59
4-38	Endmembers Spectral signatures, from Enrique Reef Images, using cPMF, SMACC and MaxD. (4 endmembers case)	60
4-39	cPMF result for Enrique Reef image at 1 meter with 4 endmembers.	61
4-40	cPMF result for Enrique Reef image at 2 meters with 4 endmembers.	62
4-41	cPMF result for Enrique Reef image at 8 meters with 4 endmembers.	62
4-42	SMACC result for Enrique Reef image at 1 meter with 4 endmembers.	63
4-43	SMACC result for Enrique Reef image at 2 meters with 4 endmembers.	63
4-44	SMACC result for Enrique Reef image at 8 meters with 4 endmembers.	63

4-45 MaxD result for Enrique Reef image at 1 meter with 4 endmembers. .	64
4-46 MaxD result for Enrique Reef image at 2 meters with 4 endmembers.	64
4-47 MaxD result for Enrique Reef image at 8 meters with 4 endmembers.	64
4-48 Comparison for endmember 1 between methods for different spatial resolutions.	65
4-49 Comparison for endmember 2 between methods for different spatial resolutions.	65
4-50 Comparison for endmember 3 between methods for different spatial resolutions.	66
4-51 Comparison for endmember 4 between methods for different spatial resolutions.	66
4-52 Spectral signatures obtained for different endmembers, from Enrique Reef images, using cPMF, SMACC and MaxD. Case with 5 end- members.	68
4-53 cPMF result for Enrique Reef image at 1 meter with 5 endmembers. .	69
4-54 cPMF result for Enrique Reef image at 2 meters with 5 endmembers.	69
4-55 cPMF result for Enrique Reef image at 8 meters with 5 endmembers.	69
4-56 SMACC result for Enrique Reef image at 1 meter with 5 endmembers.	70
4-57 SMACC result for Enrique Reef image at 2 meters with 5 endmembers.	70
4-58 SMACC result for Enrique Reef image at 8 meters with 5 endmembers.	70
4-59 SMACC result for Enrique Reef image at 1 meter with 5 endmembers.	71
4-60 MaxD result for Enrique Reef image at 2 meters with 5 endmembers.	71
4-61 MaxD result for Enrique Reef image at 8 meters with 5 endmembers.	71
4-62 Comparison among endmember 1 of each method with different spa- tial resolutions.	72
4-63 Comparison among endmember 2 of each method with different spa- tial resolutions.	72
4-64 Comparison among endmember 3 of each method with different spa- tial resolutions.	73
4-65 Comparison among endmember 4 of each method with different spa- tial resolutions.	73

4-66 Comparison among endmember 4 of each method with different spatial resolutions.	74
4-67 Enrique reef high resolution photo and spectral signatures associated with water, reef and sea grass classes.	76
4-68 Manual endmembers selected by the average of six pixel of each class for different spatial resolution	76
4-69 Abundance maps obtained for different spatial resolution with manual endmembers	77

CHAPTER 1

INTRODUCTION

1.1 Justification

In hyperspectral imaging, the reflected or emitted radiation represented by a single pixel in the remotely sensed image rarely comes from the interaction with a single homogeneous material. However, the high spectral resolution of imaging spectrometers enables the detection, identification, and classification of subpixel objects from their contribution to the measured spectral signal.

Unmixing is an approach for subpixel information extraction where the measured spectral signature is decomposed into a collection of constituent spectra, or endmembers, and a set of corresponding fractions or abundances which correspond to the fractional area occupied by the particular endmember in that pixel. Nevertheless, Unmixing is an ill-posed inverse problem and as such the solution computed with different unmixing algorithms depends on the underlying assumptions for the inverse problem.

At present many authors have developed different algorithms to find endmembers based on linear mixing model. Different approaches such as geometric approximations and matrix algebraic approximations, have been exploited to choose the best set of endmembers. One important issue is that the endmembers obtained

could be or not be pixels of the image. Sequential maximum angle convex cone (SMACC) and Maximum distance (Max D) are examples of algorithms in which their endmembers are pixels of the image. These algorithms are fast but the optimization of their results is limited by the presence of pure pixels in the image. On the other hand cPMF obtain endmembers that are not present in the image and makes it possible for better results with longer processing time.

Proposed and existing hyperspectral remote sensors provide information about the scene of interest at resolutions ranging from few meters to few kilometers in terrestrial and space applications. Understanding the type of information extracted with image exploitation algorithms and how it relates to actual spectra on the ground are important problems when algorithms that perform unmixing of hyperspectral images for subpixel analysis are studied. Ideally one would expect similar solutions for unmixing a hyperspectral image of different spatial resolutions of the same scene.

For this reason, we study the results of unmixing different images of the same area at different spatial resolutions using different unmixing algorithms. We also compare the estimation of the number of endmembers by using different approaches.

1.2 Objectives

1.2.1 General Objective

Study how the spatial resolution affects the type of information that is possible to extract from a scene using unmixing techniques.

1.2.2 Specific Objectives

- Study the dependency of abundance estimation algorithms on spatial resolution.

- Study the dependency of endmembers extraction algorithms on spatial resolution
- Study the dependency of algorithms to estimate the number of endmembers on spatial resolution.

1.3 Contribution of this work

Our work represents a review of the performance of different endmembers extraction methods where additionally the impact of the changes in the spatial resolution of the images is analyzed. This is an important issue because it is useful to know where the data from hyperspectral sensor with low spatial resolution can be used. We also analyze the differences between endmembers extraction methods that obtain virtual endmembers and image pixels as endmembers.

The work also introduces the use of the rank estimation and positive-rank estimation as a tools to estimate the number of endmembers.

1.4 Outline

The thesis is presented in five chapters. The first one, is the introductory chapter. The second one, is dedicated to present the theoretical background and literature review that are necessary for the development of this thesis, in this chapter the fundamental concepts of hyperspectral images, the linear mixing model and the spectral unmixing problem are described, including the review of topics such as estimation of the number of endmembers, abundance estimation and different approaches about extraction of endmembers. The third chapter one presents our problem where the linear mixing model is analyzed as a Forward problem and the unmixing problem as an inverse problem, evaluating the impact of spatial resolution changes in the general idea of the problem. This chapter also presents how the problem is going to be studied. Chapter four presents the experiments and the analysis

of the results obtained with the hyperspectral data from Guanica and Enrique Reef area. Finally the conclusions are presented in chapter five.

CHAPTER 2

BACKGROUND AND LITERATURE REVIEW

2.1 Hyperspectral images

A hyperspectral image (HSI) is the result of the collection of a hyperspectral sensor. In remote sensing, the sensor can be on an aircraft or a satellite [1]. The image is built from the energy radiated by the earth which is collected by the sensor. The smallest element in a HSI is called pixel. The reflectance associated with each pixel is the result of interaction of different physical factors but in specially of the constituent materials of the area covered by the spatial field of view of the sensor. Three aspects define in a general way a hyperspectral image: the spatial resolution that determines the spatial size of the pixel, the spectral resolution that is the wavelength width of the different frequency bands recorded by the sensor and the radiometric resolution that is the number of different intensities of radiation that the sensor is able to discriminate. The spectral resolution of an image is also related to the number of spectral bands, typically, the hyperspectral images have more than 100 bands. Each hyperspectral pixel is a vector, where the number of components depends on the number of bands in the image and represents a spectral signature. (See Figure 2-1)

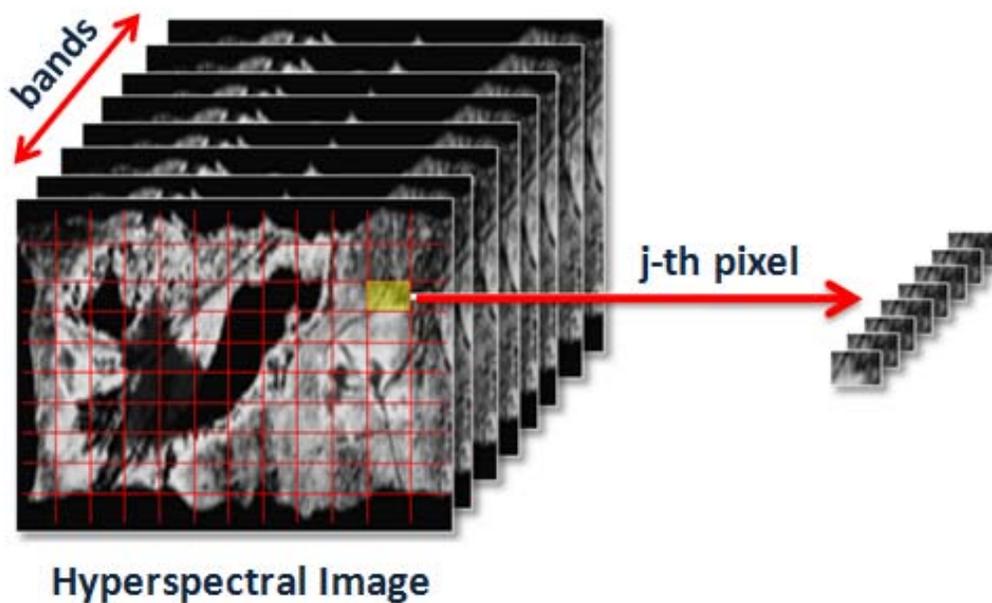


Figure 2–1: Graphical representation of a pixel in a hyperspectral image

2.1.1 Examples of Hyperspectral Sensors

AISA

The AISA Eagle is a hyperspectral sensor normally used in aircrafts. The sensor has 244-channels between the 400 to 970 nm [2], resolving spectral differences as fine as 2-4 nm. The spectral resolution of the sensor is 2.9 nm that works with a progressive scan CCD camera. The system is capable of collecting around 1,000 pixels per scanning line using push broom scanning and it is able to image at sub-meter resolution. Besides a high performance sensor head, all AISA systems integrate the following basic system components:

- Real time acquisition computer with a user-friendly interface
- GPS/INS sensor
- Power supply
- Galileo post-processing software

The sensor has established its ability in a range of commercial, research and public service applications. Some of the applications that AISA Eagle has been involved in are forestry management, vegetation cultivation, environmental investigations, precision farming, target identification, water assessment and land use planning. The sensor is operated and distributed by Galileo Group and manufactured by SPECIM [3].

HYPERION

The Hyperion sensor is part of the primary instruments on the EO-1 spacecraft. The EO-1 orbit is sun-synchronous, with an altitude of 705 kilometers, inclination of 98.2 degrees, a period of 99 minutes, and a repeat cycle of 16 days. The Hyperion provides a high resolution hyperspectral images with 220 spectral bands (from 0.4 to 2.5 μm) with a 30-meter resolution. The instrument can image a 7.5 km by 100 km land area per image, and also has a high radiometric accuracy.

The sensor has 70 spectral channels in the VNIR (visible and near-infrared) interval (356 nm - 1058 nm) and 172 channels in the SWIR (Short-wavelength infrared) interval (852 nm - 2577 nm). The spectral bandwidth is equal to 10nm.

Hyperspectral imaging has different applications in mining, geology, forestry, agriculture, and environmental management. All the Hyperion products are distributed by the U.S. Geological Survey (USGS) Earth Resources Observation and Science (EROS)

AVIRIS

AVIRIS is an acronym for the Airborne Visible InfraRed Imaging Spectrometer [4]. The sensor works across the 400 to 2500 nanometers (nm) range allowing

complete VIS-NIR-SWIR spectrum. The AVIRIS instrument contains 224 different detectors, each with a wavelength sensitive range of approximately 10 nanometers (nm), that provide 224 contiguous spectral channels. The system uses “Whisk broom” scanning with a scanning rate equal to 12 hertz (Hz).

AVIRIS is a sensor used on aircraft platforms. It has been flown on four aircraft platforms:

- NASA’s ER-2 jet: It flies at approximately 20 km above sea level, collecting pixels on the ground with an area approximately 20 meters in diameter, and yielding a ground swath about 11 kilometers wide.
- Twin Otter International’s turboprop: It flies at 4km above ground level with a speed of 130km/hr, collecting pixels with an area approximately of 4 meters square, and with a swath equal to 2km wide.
- Scaled Composites’ Proteus.
- NASA’s WB-57.

AVIRIS has flown all across the US, plus Canada and Europe.

HyspIRI

HyspIRI is part of a new NASA mission that is currently at the study stage[5]. The mission was recommended in the National Research Council Decadal Survey requested by NASA, NOAA, and USGS. This mission includes two instruments mounted on a satellite in Low Earth Orbit. There is an imaging spectrometer measuring from the visible to short wave infrared (VSWIR) and a multispectral thermal infrared (TIR) imager.

The VSWIR will have a spatial resolution of 60 m at nadir with a temporal revisit of approximately 3 weeks. TIR will also have a spatial resolution of 60 m at nadir with a temporal revisit of approximately 1 week. The application of the data

is oriented on a wide variety of studies, two of the most important are the Carbon Cycle and Ecosystem and Earth Surface and Interior focus areas.

2.2 Linear Mixing Model

In hyperspectral imaging, the reflected or emitted radiation represented by a single pixel in the remotely sensed image rarely comes from the interaction with a single homogeneous material. A pixel would be pure if the spatial resolution was smaller than the size of the class portion in the image, but, in real data it is not common to find pure pixels. However, the high spectral resolution of imaging spectrometers enables the detection, identification, and classification of sub-pixel objects from their contribution to the measured spectral signal.

The linear mixing model [6] presents a pixel as the linear combination of the spectral signatures of each material multiplied by its relative abundances. The spectral signature of each pure material is known as endmember.

The model is mathematically presented for each pixel by the equation 2.1.

$$\mathbf{x}_j = \mathbf{S} \mathbf{a}_j + \mathbf{w}_j \quad (2.1)$$

where $\mathbf{x} \in \mathfrak{R}_+^m$ is the measured spectral signature at a pixel, $\mathbf{S} \in \mathfrak{R}_+^{m \times p}$ is the matrix of endmembers, $\mathbf{a} \in \mathfrak{R}_+^p$ is the vector of spectral abundances, and $\mathbf{w} \in \mathfrak{R}_+^m$ is measurement noise, m is the number of spectral bands, and p is the number of endmembers [6]. In hyperspectral imaging, $m > p$, notice that all elements of \mathbf{S} , \mathbf{a} , and \mathbf{x} are constrained to be positive, and $\sum_{i=1}^m a_{ij} = 1$. For the entire hyperspectral image, the linear mixing model given above can be written in matrix form as

$$\mathbf{X} = \mathbf{S} \mathbf{A} + \mathbf{W} \quad (2.2)$$

where $\mathbf{X} = [\mathbf{x}_1 \dots \mathbf{x}_N]$ is the matrix containing all image pixels, $\mathbf{S} = [\mathbf{s}_1 \dots \mathbf{s}_p]$ is the matrix of endmembers, $\mathbf{A} = [\mathbf{a}_1 \dots \mathbf{a}_N]$ is the matrix of abundances of all image pixels, $\mathbf{W} = [\mathbf{w}_1 \dots \mathbf{w}_N]$ is the noise matrix, and N is the number of pixels in the image.

2.3 Spectral Unmixing Problem

The spectral unmixing problem is the determination of the endmembers and their abundances. Most state of the art methods solve the unmixing problem in two stages. In the first stage, the number of endmembers are determined and spectral signatures of the endmembers are extracted by searching for “pure” pixels in the image or by using spectral signatures from a library, or field spectroradiometer data [6, 7]. In the second stage, the abundances are estimated in many cases by solving a constrained linear least squares problem. [8–10]. The most endmembers extraction methods considers the number of endmembers as an input for the algorithm, for this reason the estimation of the number of endmember also have to be consider in the analysis of the unmixing problem solution.

2.3.1 Estimation of the Number of Endmembers

x The estimation of the number of the endmembers is a problem that many researchers have tried to study. Virtual Dimensionality (VD) [11] is an example of this kind of attempt and it is being used in many publications. However, recently the effectiveness of this method has been disputed and the concepts associated with it. Bajorski [12] shows us that the method has some inconsistencies and for this reason we do not consider this method. We decided to explore methods oriented into the estimation of the rank of the matrix such as the Scree Test.

Scree Test

This test is a graphical method that was developed by Cattell [13]. This method is based on the idea of identifying the “elbow” of a correlation factors curve. The term “scree” comes from geology [14]; it means the rubble at the bottom of a cliff. In our context the curve is built with eigenvalues of the normalized correlation matrix. The correlation matrix of the image is obtained by

$$\mathbf{R}_x = \frac{1}{n} \mathbf{X} \mathbf{X}^T \quad (2.3)$$

where, $\mathbf{R}_x \in \mathfrak{R}^{m \times m}$ is the correlation matrix, n is the number of pixels of the image and $\mathbf{X} \in \mathfrak{R}^{m \times n}$ is the image matrix. For rank determination, it is better to work with the normalized correlation matrix presented in equation 2.4

$$\tilde{\mathbf{R}}_x = \tilde{\mathbf{D}}^{-1/2} \mathbf{R} \mathbf{D}^{-1/2} \quad (2.4)$$

where $\mathbf{D} = \text{diag}(\mathbf{R}_x)$. The elements of the diagonal of $\tilde{\mathbf{R}}_x$ are normalized to 1 and the off diagonal elements have magnitude less than or equal to one. The Scree test is performed with the eigenvalues of the $\tilde{\mathbf{R}}_x$ matrix. A graphical representation of the ideal Scree curve is shown in Figure 2-2. The number of eigenvalues above the knee of the curve is used as an estimate of the linear rank. In practice, finding the elbow is not an easy task with real data.

2.3.2 Endmembers Extraction Methods

At present many authors have developed diverse algorithms to find endmembers based on the linear mixing model. Different approaches ([16–18]) such as geometric approximations, statistical methods, and matrix algebraic approximations, have been exploited to choose the best set of endmembers. Some of the most popular algorithms are N-FINDR [19], Maximum Distance (Max D) [20], Independent

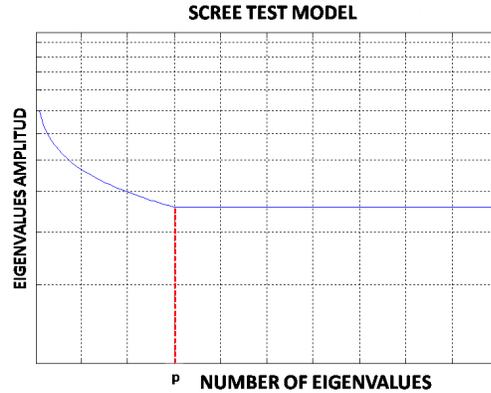


Figure 2–2: Graphic representation of the ideal Scree test model based on [15]

Component Analysis - Endmember Extraction Algorithm (ICA-EEA) [16], Vertex Component Analysis (VCA) [21], among others.

NFINDR

This method developed by Winter [19] and modified by Du et al. [22], looks for a simplex with the maximum volume that could be inscribed into the data based on $\sum_{i=1}^M a_{ij} = 1$ and imposing the assumption that there exists at least one pure pixel per endmember species within the image. They define a pure pixel as a pixel that has one of its abundance coefficients approximately equal to one.

They initialize the process with a random set of pixels as endmembers. The set of endmembers estimated is refined by evaluating each pixel to find new “pure” pixels. The endmembers are allocated in the matrix \mathbf{S} and the volume of the simplex is calculated. If the new pixel increases the volume, the endmember is replaced by the pixel. This procedure is repeated until there are no more replacements. The matrix $\tilde{\mathbf{S}}$ (See equation 2.5) is the matrix of the endmembers \mathbf{S} augmented with a row of ones.

$$\tilde{\mathbf{S}} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ s_1 & s_2 & s_3 & \dots & s_M \end{bmatrix} \quad (2.5)$$

The volume V of the simplex depends on the following relation:

$$V(\tilde{\mathbf{S}}) = \frac{1}{M-1} |\tilde{\mathbf{S}}| \quad (2.6)$$

In the Figure 2-3 the graphical representation of the simplex is showed.

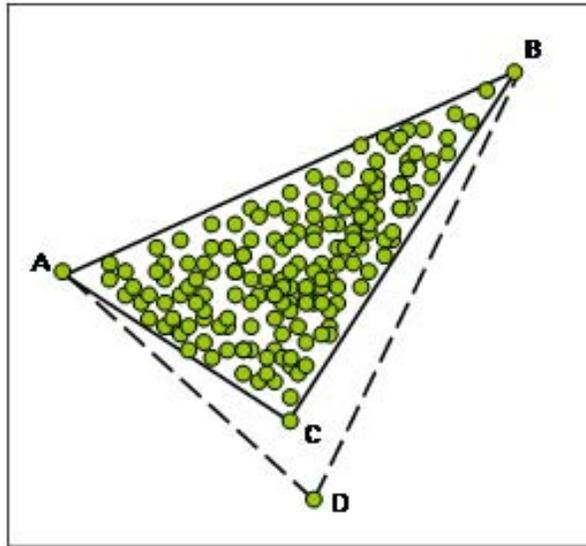


Figure 2-3: Graphical representation of the simplex with maximum volume based on the graphics of [23]. Points A, B, and C form a simplex but it does not have the maximum volume. The points A, B, D are the three true endmembers for the image and form the simplex with the maximum volume.

Maximum Distance (MaxD)

This is a method proposed by Lee [20]. MaxD is a geometric method that has the objective of finding the vertices of a simplex that involves all the data. The method starts by finding two vectors with special characteristics; it chooses the vectors with the largest and smallest magnitude in the data. These two vectors are

the two initial endmembers and the two initial vertices of the simplex. Let us call these vectors v_1 and v_2 respectively. The next step is to build the difference vector between v_1 and v_2 and project all pixels of the image in the orthogonal plane to the vector difference called v_{12} . In the orthogonal plane, we find the maximum distance between the pixel projections and the vector distance projection and the vector with the maximum distance is the new endmember and it is called v_3 . In the second step, we build a new vector distance and the process is repeated until the desired number of endmembers is obtained. This process is illustrated in the Figure 2–4.

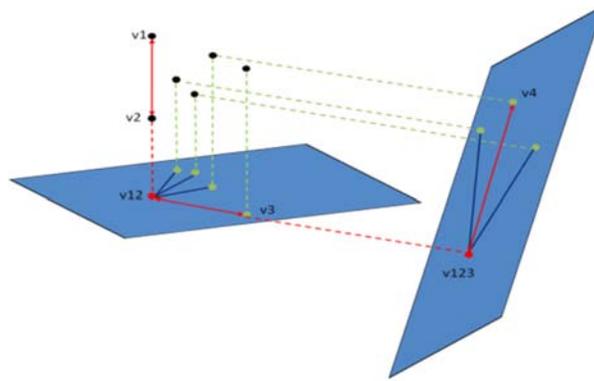


Figure 2–4: Graphical representation of the orthogonal projection for the Maximum Distance method.

Independent Component Analysis - Endmember Extraction Algorithm (ICA-EEA)

ICA-EEA is an algorithm that chooses pure endmembers based on ICA (Independent Component Analysis). Originally ICA is an algorithm that tries to obtain a solution for the classic cocktail-party problem in signal processing that presents a room where two different people are talking at the same time when we have a mixed signal and want to reconstruct the two original sources. The ICA decomposes each x_j pixel of an image into their k independent components using the following expression:

$$\mathbf{x}_j = a_{j1}\mathbf{s}_1 + a_{j2}\mathbf{s}_2 + \cdots + a_{jk}\mathbf{s}_k, \forall_j \quad (2.7)$$

Assuming that s_i is an independent random variable that symbolizes an independent component or endmember and the $a_{j1}, a_{j2}, \dots, a_{jk}$ are the coefficients of abundance of each independent component in the image, k is the number of bands and theoretically the algorithm could find independent components but ICA-EEA only keeps m of these k endmembers where m is a number with physical meaning. The independent components are ordered and scored [16] and the highest components are kept. In practice ICA-EEA uses an improved algorithm of ICA known as fast ICA [24]. In fact, the ICA model does not apply to the unmixing problem since endmembers are linearity dependent due to the sum to one constraint normally used in linear mixing model. A graphical representation of the model is presented in Figure 2-5.

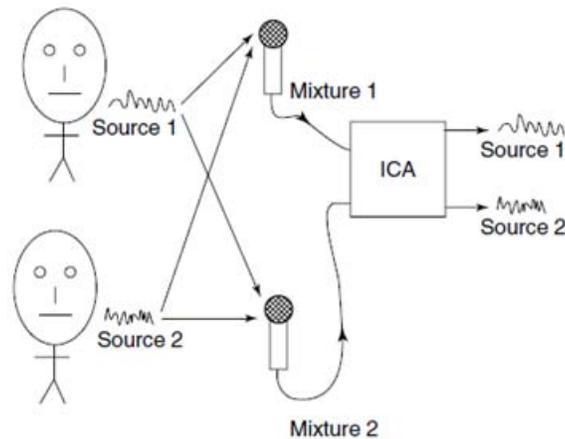


Figure 2-5: Graphical representation of the cocktail party problem. Taken from Encyclopedia of Statistics in Behavioral Science [25]

Vertex Component Analysis (VCA)

Nascimento [21] describes the model adopted for the image in a manner similar to the above algorithms but adding a new coefficient γ that is a scale factor

modeling illumination variability due to surface topography. In this model a pixel is represented as in the equation 2.8:

$$\mathbf{x}_j = \mathbf{S}(\gamma \mathbf{a}) \quad (2.8)$$

where, \mathbf{S} is the matrix of endmembers and \mathbf{a} is the vector of abundances and γ is the scale factor defined previously. Each element of $\mathbf{a} \in \mathfrak{R}^+$ meets the sum to one constraint, and forms a simplex. This approach explains that each pixel can be presented as a vector in a k -dimensional Euclidean space, where each channel is assigned to one axis of the space and k is the number of bands. Figure 2-6 (b) presents the two-dimensional scatterplot of mixtures of the three endmembers shown in Figure 2-6 (a). The circles denote pure materials.

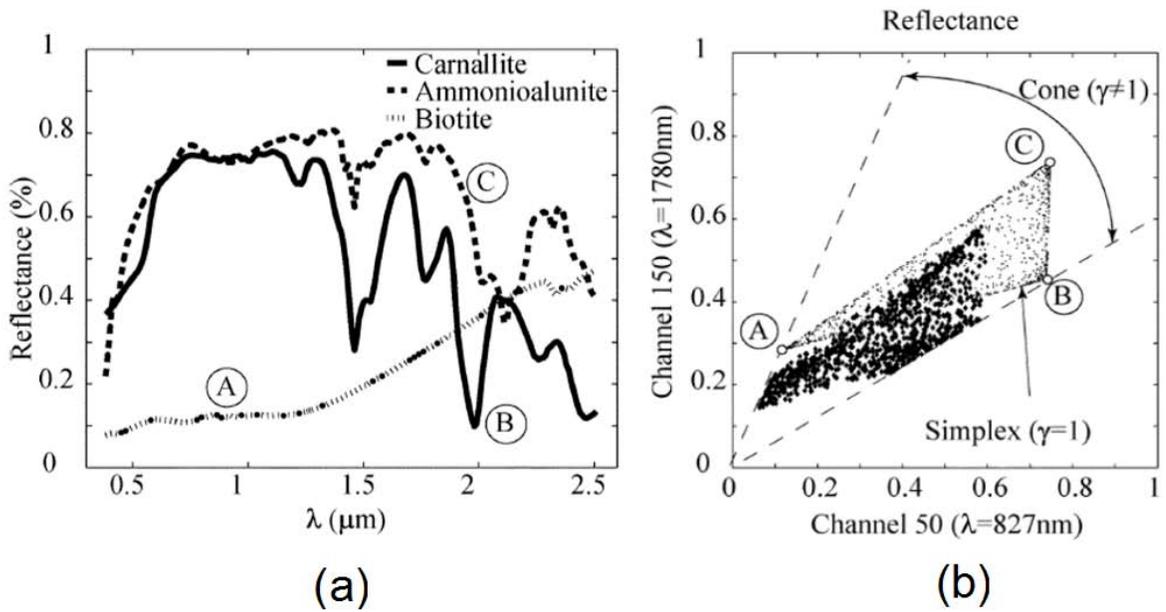


Figure 2-6: (a) Reflectances of carnallite, ammonioalunite, and biotite.(b) Two-dimensional scatterplot of mixtures of the three endmembers shown in (a). Taken from Nascimento and Bioucas [21]

2.3.3 Abundance Estimation

Abundance estimation is the process through the abundances associated with the endmembers are determined for each pixel in the image. Velez-Reyes and Rosario [10] proposed a method to solve the abundance estimation as a least distance. In this way, they present an approach based on the QR decomposition that permits an exact solution to the fully constrained problem, where:

$$\begin{aligned} \tilde{\mathbf{a}} &= \underset{\mathbf{a}}{\operatorname{arg\,min}} \|\mathbf{S}\mathbf{a} - \mathbf{x}\|_2^2 \\ &\text{subject to } \mathbf{a} \geq 0 \text{ and } \sum_{i=1}^m a_i = 1 \end{aligned} \quad (2.9)$$

The solution proposed for the above equation is based on the solution of the problem proposed in the equation 2.10, where the condition of the abundances is relaxed to sum less than or equal to one. For this reason, we will analyze first, the solution to equation 2.10 that is called Non Negative Least Squares Algorithm (NNLS).

$$\begin{aligned} \tilde{\mathbf{a}} &= \underset{\mathbf{a}}{\operatorname{arg\,min}} \|\mathbf{S}\mathbf{a} - \mathbf{x}\|_2^2 \\ &\text{subject to } \mathbf{a} \geq 0 \text{ and } \sum_{i=1}^m a_i \leq 1 \end{aligned} \quad (2.10)$$

The algorithm to solve 2.10 can be explained in some steps:

- Compute QR decomposition of S

$$\mathbf{Q}^T \mathbf{S} = \begin{bmatrix} \mathbf{R} \\ 0 \end{bmatrix}; \quad \mathbf{c} = \mathbf{Q}^T \mathbf{x} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \quad (2.11)$$

- Rewrite the equation 2.10 with the QR decomposition in equation 2.11, obtaining:

$$\tilde{\mathbf{a}} = \underset{A}{\operatorname{arg\,min}} \|\mathbf{R}\mathbf{a} - \mathbf{c}_1\|_2^2 \quad (2.12)$$

subject to $\mathbf{a} \geq 0$ and $\sum_{i=1}^m a_i \leq 1$

- Compute $\tilde{\mathbf{a}}_{ULS} = \mathbf{R}^{-1}\mathbf{c}_1$

- Define:

$$\mathbf{z} = \mathbf{R}\tilde{\mathbf{a}} - \mathbf{c}_1 \Rightarrow \tilde{\mathbf{a}} = \mathbf{R}^{-1}(\mathbf{z} + \mathbf{c}_1) = \mathbf{R}^{-1}\mathbf{z} + \tilde{\mathbf{a}}_{ULS} \quad (2.13)$$

- Substitute equation 2.13 in the equation 2.12, obtaining:

$$\tilde{\mathbf{a}} = \underset{z}{\operatorname{arg\,min}} \|\mathbf{z}\|_2^2 \quad (2.14)$$

subject to $\mathbf{G}\mathbf{z} \geq \mathbf{g}$

where $\mathbf{G} = \begin{bmatrix} \mathbf{I} & -1 \end{bmatrix}^T$, $\mathbf{g} = \begin{bmatrix} -\mathbf{a}_{ULS}^T & 1^T \mathbf{a}_{ULS} - 1 \end{bmatrix}^T$, $\mathbf{I} \in \Re^{m \times m}$ is the identity matrix and $1^T = (1 \ 1 \ \dots \ 1)$. The problem presented in the equation 2.14, finally can be solved as a Non Negative Least Square Problem.

- Define:

$$\underset{u}{\operatorname{arg\,min}} \|\mathbf{E}\mathbf{u} - \mathbf{f}\|_2^2 \quad (2.15)$$

subject to $\mathbf{u} \geq \mathbf{0}$

where $\mathbf{E} = \begin{bmatrix} \mathbf{G}^T & \mathbf{g}^T \end{bmatrix}^T$, $\mathbf{f} = \begin{bmatrix} \mathbf{0} & 1 \end{bmatrix}^T$, and $\mathbf{0} = (0 \ 0 \ \dots \ 0)$ is a vector of m zeros.

- Define:

$$\mathbf{r} = \mathbf{E}\mathbf{u} - \mathbf{f} \quad (2.16)$$

Note: if $\|\mathbf{r}\| = 0$ there are no solutions because it would be inconsistent with the constraints imposed.

- The unique solution for equation 2.14 is

$$\tilde{\mathbf{a}} = \tilde{\mathbf{a}}_{ULS} - \mathbf{R}^{-1}\mathbf{z} \quad (2.17)$$

where, $\mathbf{z}_i = -r_i/r_{n+1}$, for $1 \leq i \leq m$. Then, the estimation of $\tilde{\mathbf{a}}$ can be obtained by the equation 2.13

Now, having the solution for the equation 2.10, we can obtain the solution for equation 2.9, that is the problem of our interest.

First, equation 2.9 is redefined as:

$$\begin{aligned} & \underset{\mathbf{a}}{\operatorname{arg\,min}} \|\mathbf{S}\mathbf{a} - \mathbf{x}\|_2^2 \\ & \text{subject to } \mathbf{a} \geq 0 \text{ and } a_1 = 1 - \sum_{i=1}^m a_i \end{aligned} \quad (2.18)$$

Then, the equation 2.18 is transformed into a inequality constrained problem:

$$\begin{aligned} & \underset{\bar{\mathbf{a}}}{\operatorname{arg\,min}} \|\bar{\mathbf{S}}\bar{\mathbf{a}} - \bar{\mathbf{x}}\|_2^2 \\ & \text{subject to } \bar{\mathbf{a}} \geq 0 \text{ and } \sum_{i=2}^m \bar{a}_i \leq 1 \end{aligned} \quad (2.19)$$

where, $\bar{\mathbf{a}} = [a_2, \dots, a_m]$, $\bar{\mathbf{S}} = [a_2 - a_1, \dots, a_m - a_1]$, and $\bar{\mathbf{x}} = \mathbf{x} - a_1$. Now we apply the NNLS algorithm to estimate $\bar{\mathbf{a}}$ and $a_1 = 1 - \sum_m^{i=2} \bar{\mathbf{a}}$

2.3.4 Full Unmixing Algorithms

The full unmixing algorithms bring a complete solution for the unmixing problem, extracting endmembers and solving abundances in the same process. Some of these algorithms are presented below.

Sequential Maximum Angle Convex Cone(SMACC)

The method solves linear unmixing using a convex factorization technique that simultaneously generates the set of endmembers and their abundances [26]. When the abundance coefficients are constrained to sum to one, the convex cone reduces to a convex hull and the extreme vectors form a simplex.

The endmembers and abundance coefficients are determined sequentially. To obtain the endmembers, SMACC first selects a group of pixels that are extreme vectors in the data. They become a basis and form a convex cone within their subspace. The data that is outside of the cone is called residual. A test is used to determine the next basis vector or endmember based on the length of its residual in the actual model. The length of the residual is the distance that the vector lies outside the convex cone defined by the current endmembers [26]. After an endmember has been identified, its contribution to the residual is removed by oblique projections. The sequence is repeated until the desired number of endmembers is obtained. This method is available as part of the ENVI/IDL software from ITT VIS.

Positive Matrix Factorization (PMF)

Positive Matrix Factorization is as a factorization problem with non-negativity constraints which, given a matrix $X \in \mathfrak{R}_+^{m \times n}$ is desired to find two matrices that satisfies:

$$\mathbf{X} = \mathbf{S}\mathbf{A} \tag{2.20}$$

where \mathbf{S} is a matrix $m \times p \in \mathfrak{R}^+$ and \mathbf{A} is a matrix $p \times n \in \mathfrak{R}^+$. The minimum value of p for which an exact solution to the PMF problem exist is called the positive rank [27] of \mathbf{X} .

In unmixing, \mathbf{X} is the matrix containing all image pixels, \mathbf{S} is the matrix of endmembers, and \mathbf{A} is the matrix of abundances for all pixels in the image.

PMF of a matrix \mathbf{X} can be approximated by finding solution to an optimization problem defined as follows:

$$\begin{aligned} \tilde{\mathbf{S}}, \tilde{\mathbf{A}} = \underset{\tilde{\mathbf{S}}, \tilde{\mathbf{A}}}{\operatorname{arg\,min}} \|\mathbf{X} - \mathbf{S}\mathbf{A}\|_F^2 \\ \text{subject to } s_{ij} \geq 0 \text{ and } a_{ij} \geq 0 \text{ and } \sum_{i=1}^m a_{ij} = 1 \end{aligned} \tag{2.21}$$

Based on the approximation used to find the solution of the PMF problem, the minimum value of p is found when $\|\mathbf{X} - \mathbf{S}\mathbf{A}\|_F^2$ is approximately equal to zero.

Paatero [28] and Lee and Seung [29] have proposed approaches that had been the basis of many subsequent studies. Initially, Paatero worked with alternating least squares (ALS). The method is initialized with different random matrices to obtain a global optimum. Paatero in later studies presented improvements to this technique as PMF2 and PMF3 for two and three dimensions respectively [30]. One

of the most popular contributions of Lee and Seung bring an algorithm based on multiplicative updating rules that are applied to solve this kind of optimization problem.

In general, the positive matrix factorization had been explored to find solutions for problems in different applications of image processing such as classification, face recognition, and target detection among others. In the process to obtain better solutions for these applications, many authors have found that imposing additional constraints could improve the behavior of the algorithm [8, 9, 31–33].

2.3.5 Constrained Approaches for Positive Matrix Factorization

Masalmah and Velez-Reyes [9] and Jia and Qian [8] have shown that the unmixing problem can be related to the computation of a constrained positive matrix factorization (cPMF). In addition to the positivity constraint, a very common practice is to add the sum to one constraint to the columns of \mathbf{A} making this a constrained PMF or cPMF. The cPMF has been shown to be a powerful approach for hyperspectral unmixing ([8, 9]). Jia and Qian [8] also build up the idea to include the sparse and smooth concepts in the constraints based on the studies of Pascual-Montano et al. [32] and Hoyer [33].

Sum to one

Masalmah and Velez-Reyes [9] proposed an iterative algorithm based on an optimization problem described by Lee and Seung [29] but adding the sum to one constraint to the columns of \mathbf{A} :

$$\begin{aligned} \tilde{\mathbf{S}}, \tilde{\mathbf{A}} = \underset{\tilde{\mathbf{S}}, \tilde{\mathbf{A}}}{\operatorname{arg\,min}} \|\mathbf{X} - \mathbf{S}\mathbf{A}\|_F^2 \\ \text{subject to } s_{ij} \geq 0 \text{ and } a_{ij} \geq 0 \text{ and } \sum_{i=1}^m a_{ij} = 1 \end{aligned} \tag{2.22}$$

The initial endmembers are obtained using SVDSS algorithm in order to find the more independent pixels in the image. This method has two principal iterative steps. In the first step, the endmember matrix will be fixed and the consequent abundances will be estimated. In this phase, we are applying that the $\sum_{i=1}^m a_{ij} = 1$. This constraint has a physical significance because a material cannot have an associated abundance above 100%. This constraint also could be substituted using $\sum_{i=1}^m a_{ij} \leq 1$ opening space to a shadow endmember to take into consideration topography as described by Nascimento [21]. In the second step the abundance matrix will be fixed and the endmember matrix is estimated. The estimation of each matrix on each step is done by solving a nonnegative linear least squared problem. The method iterates between the two steps until convergence is reached.

One of its key features is its capability to extract endmembers from the image that are not present as pure pixels. This is an important feature since many objects of interest are only present in a sub-pixel level.

Sparseness

In the image processing environment, the sparseness concept refers to an image where only a few pixels can effectively represent the typical data vectors of the scheme. It implies that high quantities of pixels are zero or close to zero and only few pixels are contributing with considerable information [33]. It also suggests that the number of basis components required to represent X is minimized.

Actually, PMF produces a sparse representation of the data, with the disadvantage that we cannot control the degree to which representation is sparse. The sparseness measure permits us to identify how much energy of a vector is crowded into only few components.

The sparseness criterion is presented including the minimization of the matrix of abundances using the 1 norm [34]. The complete minimization problem is transformed as follows:

$$\begin{aligned} \tilde{\mathbf{S}}, \tilde{\mathbf{A}} = \underset{\tilde{\mathbf{S}}, \tilde{\mathbf{A}}}{\operatorname{arg\,min}} \|\mathbf{X} - \mathbf{S}\mathbf{A}\|_2^2 + \beta \|\mathbf{A}\| \\ \text{subject to } s_{ij} \geq 0 \text{ and } a_{ij} \geq 0 \end{aligned} \quad (2.23)$$

where, β is regularization parameter.

A solution for this optimization problem is presented by Kim and Koh [34]. This work suggests an algorithm based on an interior-point method for solving large-scale l1 regularized least squares.

CHAPTER 3

UNMIXING ALGORITHMS

3.1 Linear Mixing Model as Forward Problem

Linear mixing is a forward problem that can explain the formation of a hyperspectral image. (See Figure 3-1). Even though our work is to analyze the inverse problem, so-called Unmixing problem, it is important to understand how the forward problem changes when the spatial resolution is modified.

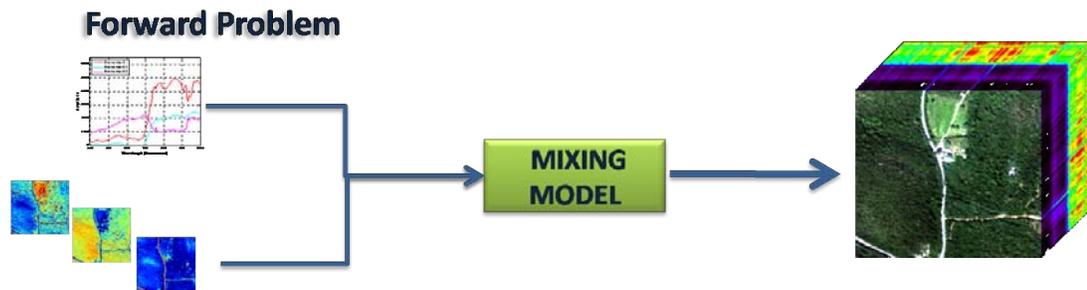


Figure 3-1: Linear Mixing Model as a forward problem

Each pixel in the image covers an area that is approximately defined by the spatial resolution of the sensor. Each detector in the sensor has an Instantaneous Field of View (IFOV) that defines its angular coverage[35]. The altitude between

the optics of the detector element and the Earth's surface establishes the Ground-projected Instantaneous Field of View (GIFOV) that defines the spatial resolution of the pixel [35]. (See Figure 3–2). The pixel is formed with the radiation collected by the detector over the area delimited by the GIFOV. It means the spatial resolution of an image changes when the altitude has been modified, as the resolution changes the level of mixing in the pixel may increase. For example, if we have two images of the same scene with different spatial resolution, although we will have the presence of the same materials, the level of mixing in the pixels is going to be different because the area covered by the pixel is not exactly the same.

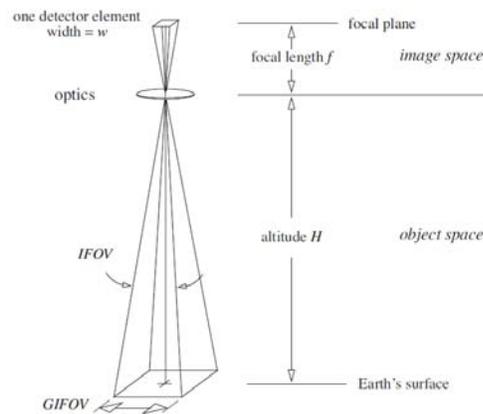


Figure 3–2: Geometric description of a single detector element in the focal plane of an optical sensor. Taken from [35]

3.2 Unmixing as an Inverse Problem

In the inverse problem our input is the output of the forward problem, and the objective is try and extract the correct endmembers and their abundances. (See Figure 3–3). Unmixing is an ill-posed inverse problem such as the solution computed with different unmixing algorithms depends on the underlying assumptions for the inverse problem. The performance of the unmixing process is also influenced by the number of endmembers, because if the number is lower some of the endmembers solution are going to be the mixture of different spectral classes or

there will be missing classes; on the other hand, if the number is higher we can “repeat” endmembers, “add” non existing classes or “divide” the same class.



Figure 3–3: Unmixing Model as an inverse problem

In conclusion, the complete unmixing process requires paying attention to three important aspects:

- Number of Endmembers
- Estimation of the Endmembers, and
- Abundances

Ideally one would expect similar solutions for unmixing a hyperspectral image of different spatial resolutions of the same scene. (See Figure 3–4). This means that the case of two images of the same scene with different spatial resolutions would obtain the same endmembers matrix response with different abundance estimation matrices as a result. However, this is only an assumption of the solution to the problem when it is exposed to changes in the spatial resolution of the images. For this reason, we study the results of unmixing different images of the same area at different spatial resolution using different unmixing algorithms.

Unmixing of images at different spatial resolutions

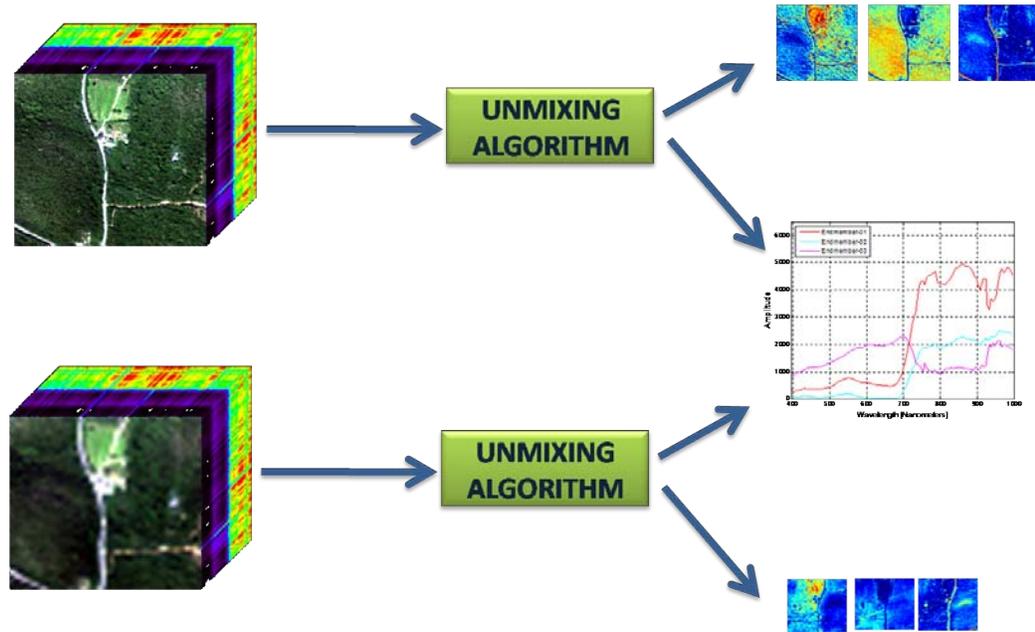


Figure 3–4: Unmixing Model for images at different spatial resolution

Based on the interest of analyzing the performance of the unmixing techniques according with changes in the spatial resolution of the images, we decided to work with three geometric algorithms SMACC, MaxD and cPMF. The estimation of the number of endmembers will be completed using the estimation of the Rank and Positive Rank. The abundance estimation will be made with NNSTO algorithm developed by Velez-Reyes and Rosario. [10]

3.2.1 Estimation of the number of Endmembers and its relation with the spatial resolution

An important component of estimating the parameters of the LMM is determination of the number of endmembers. Many approaches use the rank of the endmember matrix \mathbf{X} or of a matrix related to it such as the correlation matrix $\mathbf{X}^T \mathbf{X}$. A more reasonable approach is to use the positive rank [36] which is the

minimum value of p in equation 2.21 for which the approximation holds with equality. The rank is a lower bound on the positive rank. No direct method is known to estimate the positive rank. Only search methods are used [36]. In the following experiments, the rank is estimated using the Scree Test [13]. The positive-rank is estimated by plotting the fitting error in equation 2.22 as a function of the number of endmembers and using the point at which the normalized fitting error flattens as the estimate of the positive rank.

Rank Estimation

As mentioned earlier, the Scree Test will be use to estimate the rank of the image matrix. This method was explained in Chapter 2. The following chart presents the procedure graphically.

Positive-Rank Estimation

There is no direct method to calculate the positive-rank (p-rank) of a matrix. To determine an estimate of the p-rank, we will look at a plot of the fitting error in the approximation showed in equation 2.22 as a function of the number of endmembers and choose as an estimate the value after which no significant reduction in fitting error occurs. The fitting error is given by equation 3.1.

$$\mathbf{Fitting\ Error} = \frac{\|\mathbf{X-SA}\|_F^2}{\|\mathbf{X}\|_F^2} \quad (3.1)$$

where, $\|\cdot\|_F$ represents the Frobenious norm.

A model of the curve obtained using this approach is shown in the Figure 3-6, the graphic showed is an example of one of the experimental result obtained with real data.

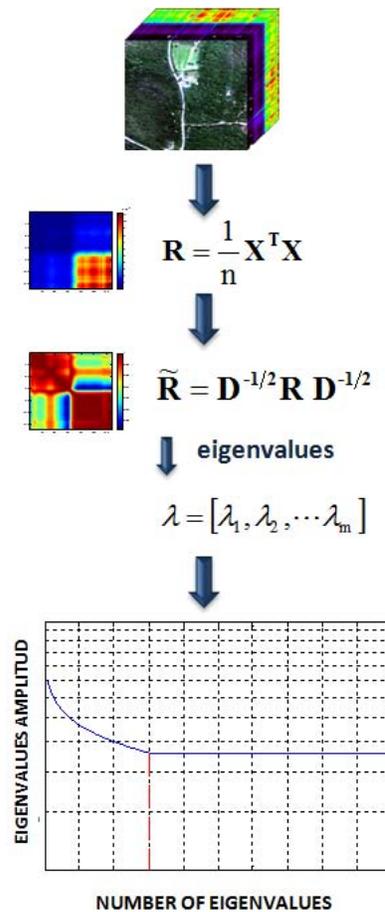


Figure 3-5: Graphical representation of the Scree model for hyperspectral imagery

3.2.2 Endmembers Extraction and Abundance Estimation and its relation with the spatial resolution

As we showed in chapter 2, different approaches have been developed to try to find the best solution to the unmixing problem. In fact, the extraction of endmembers is the part of the problem with most advances at the moment. The methods are generally classified in two main ways: Geometric Algorithms (SMACC, MaxD, VCA, NFIDR, and cPMF) and Statistical Algorithms (ICA-EEA).

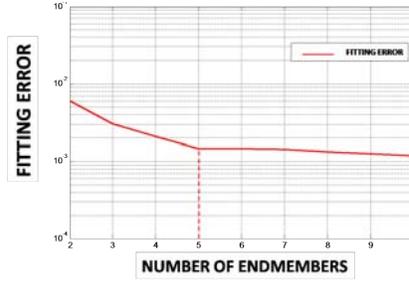


Figure 3-6: Fitting error model curve.

Approaches employed to make the Endmembers Extraction

In our work, we are concentrated in the analysis of Geometric Algorithms. We selected SMACC, MaxD and cPMF to make the comparison among different endmembers extraction methods and different spatial resolutions. SMACC and MaxD are methods that are based on projections, the principle of the methods are based on the presence of pure pixels in the image. On the other hand, cPMF is an algorithm that obtains virtual endmembers; it means, that the method does not require pure pixels in the image. SMACC and MaxD are sequential algorithms and cPMF is an iterative method. For this reason, cPMF has a heavier computational cost compared to SMACC and MaxD. Despite of SMACC is also a full unmixing algorithm, we only use the endmembers extracted by the method.

The same abundance estimation is applied to the results of all the endmember extraction methods. The abundance estimation algorithm is NNSTO. An interesting detail of cPMF is that it is a full unmixing algorithm and it has implicit the NNSTO algorithm for the abundance estimation, in fact, the result of abundances in each iteration is also employed by the algorithm to refine the group of endmembers. Table 3-1 shows the comparison of the main characteristics of the three algorithms.

Characteristic	Constrained Positive Matrix Factorization (cPMF) [Masalmah,2006]	Sequential Maximum Angle Convex Cone (SMACC) [Gruninger,2004]	Maximum Distance method (MaxD) [Lee,2003]
Endmembers	Not necessary pixels in the image	Pixels in the image	Pixels in the image
Type of algorithm	Iterative	Sequential	Sequential
Computational Resource	Very high	Low	Low

Table 3-1: Comparison between Endmembers Extraction Methods

CHAPTER 4

EXPERIMENTS AND ANALYSIS OF THE RESULTS

4.1 Methodology

Two experiments were performed to explore the estimation of the number of endmembers, extraction of endmembers and abundance estimation.

The methodology for experimental work is shown in Figure 4–1.

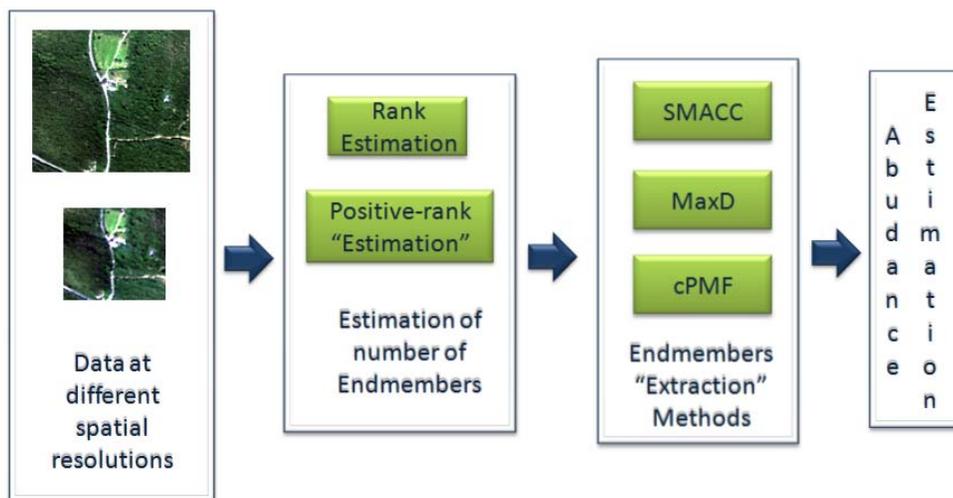


Figure 4–1: Fitting error model curve.

The experiments are divided in three main parts. The first one, is the estimation of the number of endmembers using the estimation of the rank and the positive-rank.

The rank will be estimated using the Scree Test and the positive-rank by fitting error analysis. These two approaches were explained in Chapters 2 and 3 respectively.

The second part is endmember extraction. SMACC, MaxD and cPMF are the methods studied. For the abundance estimation NNSTO will be used and the same method is going to be employed in all cases.

4.2 Experiment 1 - Guanica Dry Forest Area

4.2.1 Imagery

Guanica Dry forest images with spatial resolution of one and four meters are used. These two images are part of the data collected with AISA sensor during the 2007 Puerto Rico Hyperspectral Mission [37]. The images have spectral information from 397 to 995 nm and 128 bands. The complete area covered by the flights at one and four meters is shown in the Figure 4–2.

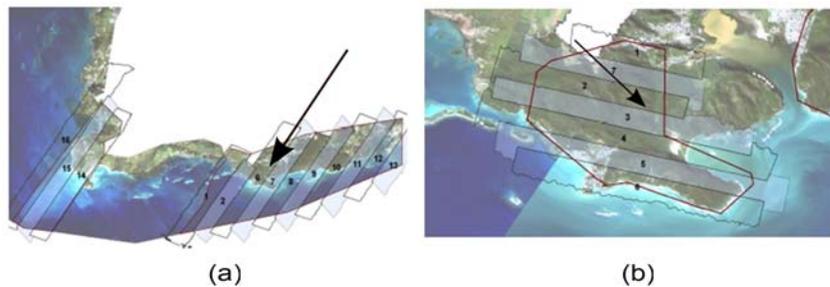


Figure 4–2: Flight lines over Guanica area in 2007 Puerto Rico Hyperspectral Mission. (a)Flight lines at 1 meter. (b)Flight lines at 4 meters.

Figure 4–3 shows the true and the color infrared composites of the image subset selected. The scene is mostly covered by vegetation and a road crosses from north to south. Three general classes of vegetation are observable: tall vegetation, low vegetation and a grass field as shown in Figure 4–4. In the center part of the scene, is a building near a road.

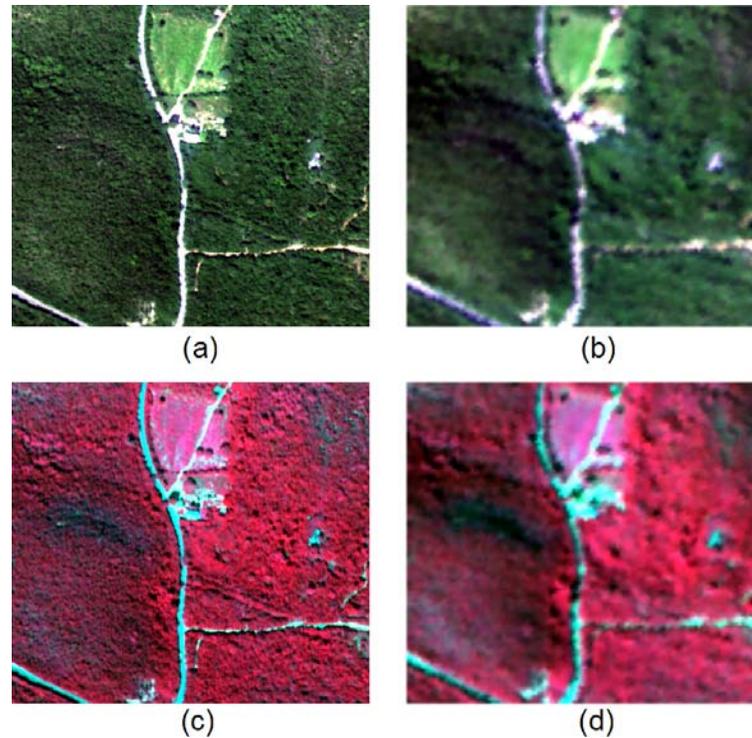


Figure 4-3: Guanica Dry Forest Image. (a) True Color Composite of image at 1 meter. (b) True Color Composite of image at 4 meters. (c) Color Infrared Composite of image at 1 meter (e) Color Infrared Composite of image at 4 meters.

During the field work, RGB images of the main vegetation classes of the area were collected. In Figures 4-5, 4-6, and 4-7, Buffel grass, *Bouyeria Suculenta* and *Exostema Caribeum*, and *Prosopis Procera* are presented. In these figures, the spectral signatures of the species are also shown. These spectral signatures were collected during the mission in 2007. It is important to mention that *Bouyeria Suculenta* and *Exostema Caribeum* species are presented in the same figure, because their spectral signatures are very similar. This is a common situation with vegetation classes, and for this reason we do not specify the classes in the interpretation figure by species, but instead identify general classes.



Figure 4-4: Interpretation of the Guanica Dry forest image based on fieldwork at the area.

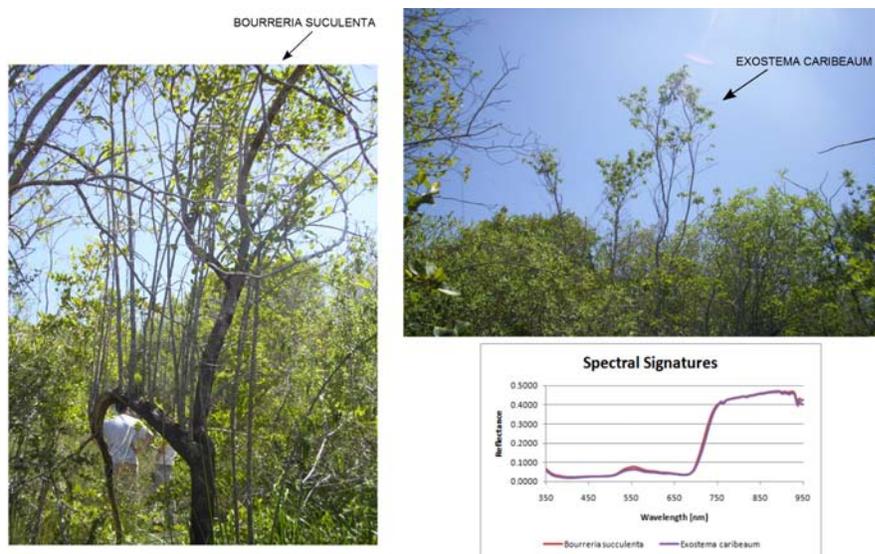


Figure 4-6: Photos and Spectral signature *Bourreria succulenta* and *Exostema caribeum*

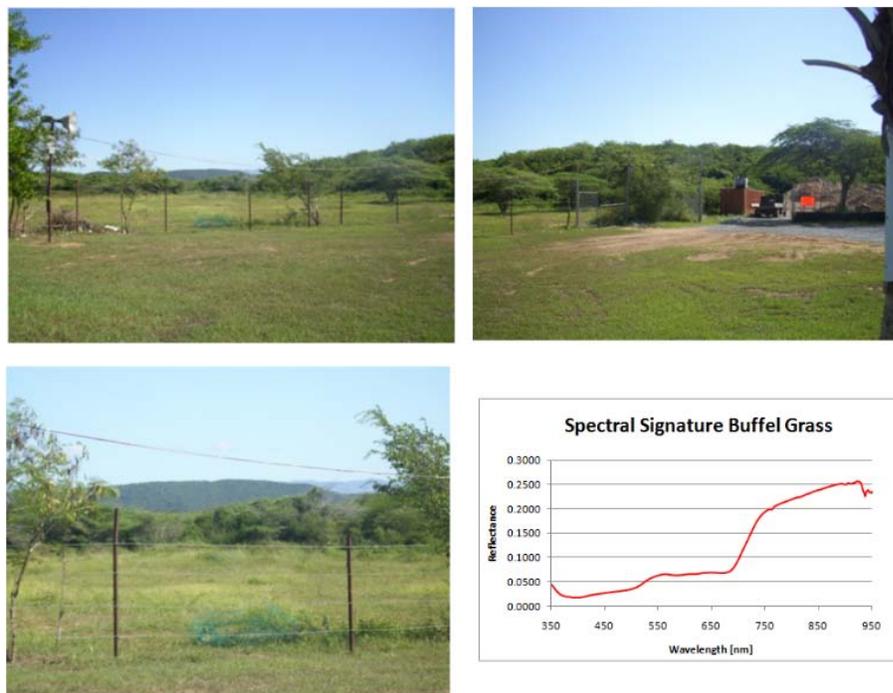


Figure 4-5: Photos and Spectral signature of Buffel grass and some images of grass field.

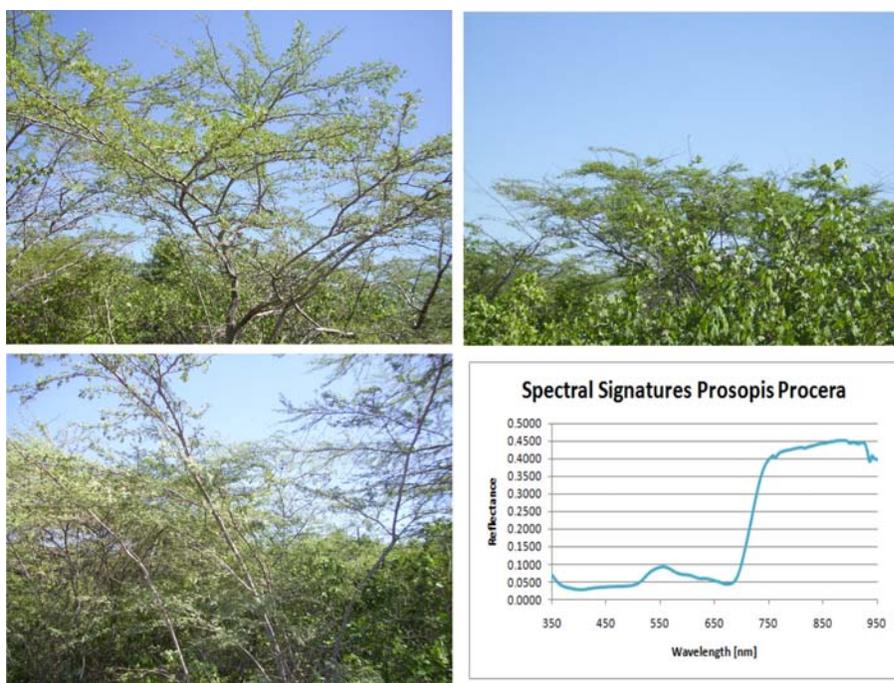


Figure 4-7: Photos and Spectral signature Prosopis Procera

4.2.2 Results and Discussion

Estimation of the Number of Endmembers

- Rank Estimation

The estimation of the rank using Scree test is presented in Figure 4–8. Application of the Scree test is sometimes difficult because it is difficult to identify the “elbow” of the curve. This point is found by visual inspection of the curve and for this case the rank estimated was three.

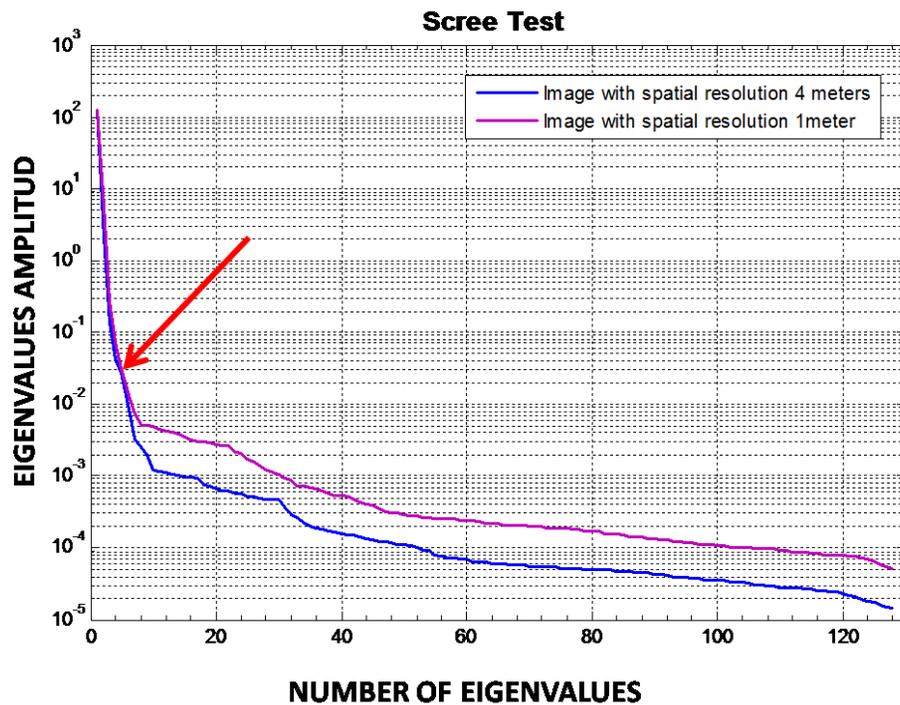


Figure 4–8: Scree test for Guanica subset images at one and four meters.

- Positive-Rank Estimation

The positive rank was estimated by calculating the fitting error (See equation 3.1) across the results obtained with cPMF between 2 and 10 endmembers. In this case we also show the fitting error obtained for the PMF results. In Figure 4–9, the results of the fitting error are shown. The estimate of the positive-rank according to Figure 4–9 is equal to five. In the Figure, we can identify that the error decreases

faster and is lower for the image at 4 meters. This is due to the number of pixels in each image, the image with higher spatial resolution has more pixels. In addition, the results for PMF present a lower error for both cases. This is expected since PMF does not need to satisfy the sum to one constraint.

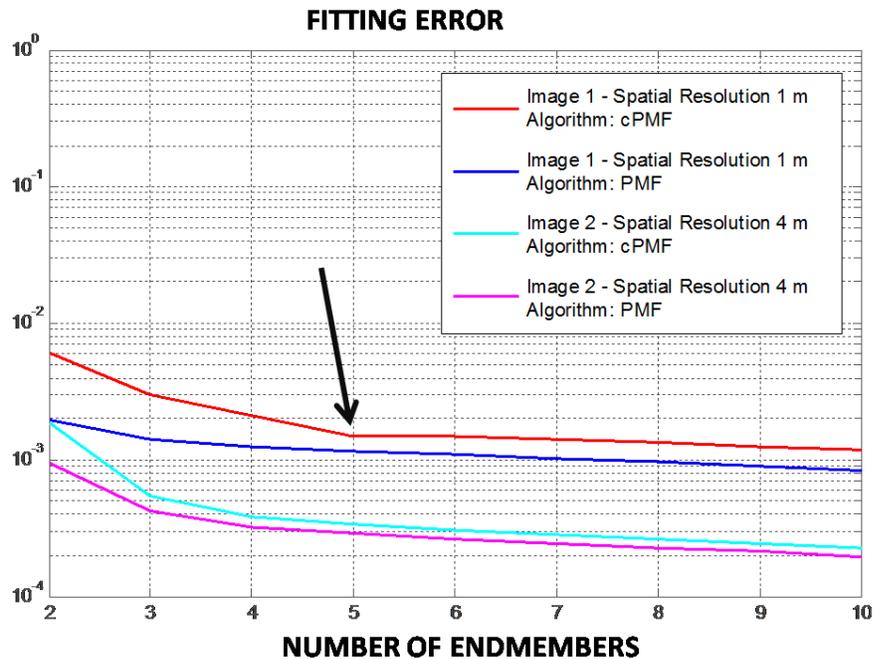


Figure 4–9: Fitting Error for Guanica Dry Forest Image at 4 meters and 1 meter.

Endmembers Extraction and Abundance Estimation

Based on the estimate of the number of endmembers obtained in the previous section for the Guanica image subset, we compared the unmixing results for 3 and 5 endmembers. Figure 4–10 presents the spectral signatures obtained for three and five endmembers by using cPMF, SMACC and MaxD. The results obtained with SMACC and MaxD are very similar at the same spatial resolution. However, the cPMF results are more consistent across scale changes.

When comparing the abundances it is clear that there are certain features that all algorithms easily retrieve such as the asphalt road from top to bottom. However the different types of vegetation such as those on the left and right of the road and the grass field on the top center area as well as unpaved road running from left to right and the small rock formation near the center right are better captured by the cPMF abundance maps than the SMACC or MaxD maps.

The results for the experiment with 3 endmembers present an interesting features. For the image at 4 meters, the abundances maps obtained with the endmembers extracted by the three methods identify in the first map a mixture between grass field and tall vegetation, in the second map low vegetation and in the last one a mixture between roads and building. For the image at 1 meter the abundance maps for SMACC and MaxD have the same performance at 4 meters. Meanwhile, the results with cPMF show another allocation. In the first abundance map a triangle-shaped area is shown in the north part of the image. This area was identified in the field work as an area covered by low grass. The second abundance map shows a mixture between tall and low vegetation, and roads and building are shown in the third abundance map. We found that the results obtained with cPMF at 1 meter, better extract the information presented in the scene, despite of this it is clear that we have more than 3 classes, the distribution of the classes agree better in comparison with the field work (See Figures 4–11, 4–12 and 4–13).

Table 4–1 shows the angle distance between the endmembers obtained with the same method at different spatial resolutions. Results show that the shapes of retrieved endmembers are similar while Figure 4–10 shows that their intensities might differ significantly.

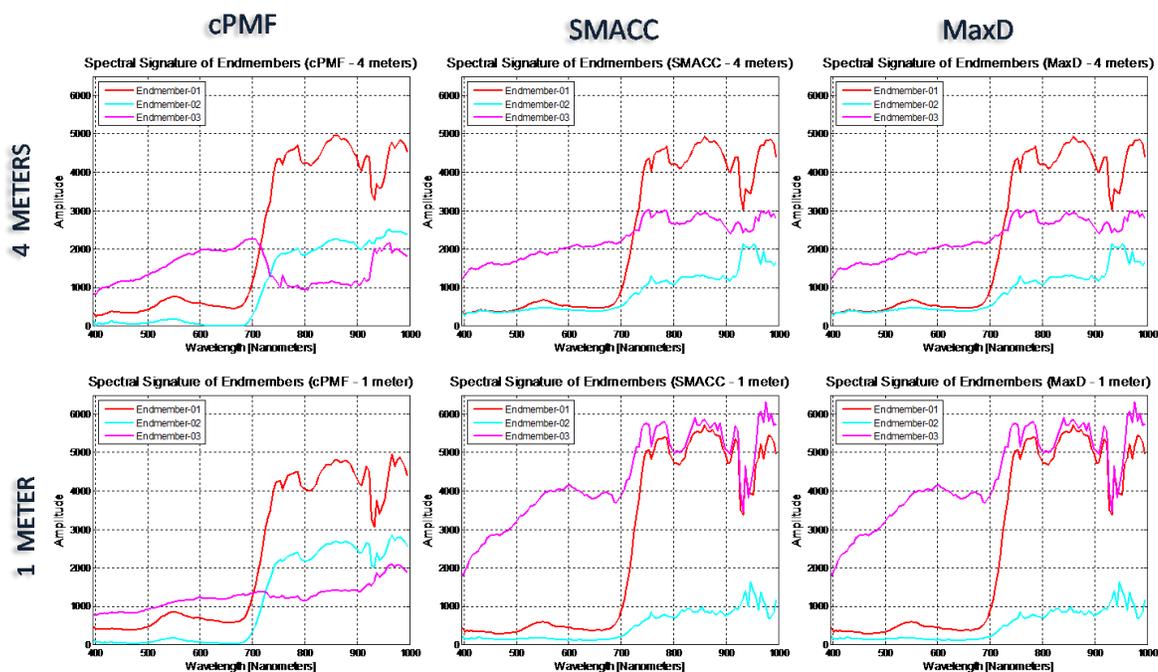


Figure 4-10: Endmembers obtained with different endmembers extraction algorithms. Results for image at 4 meters and 1 meter of spatial resolution.

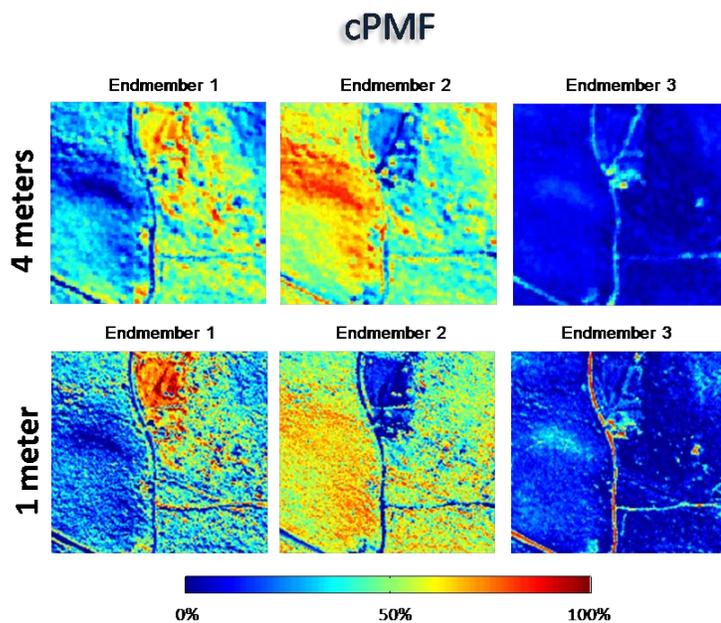


Figure 4-11: Unmixing results using cPMF with 3 endmembers from Guanica Area at one and four meters.

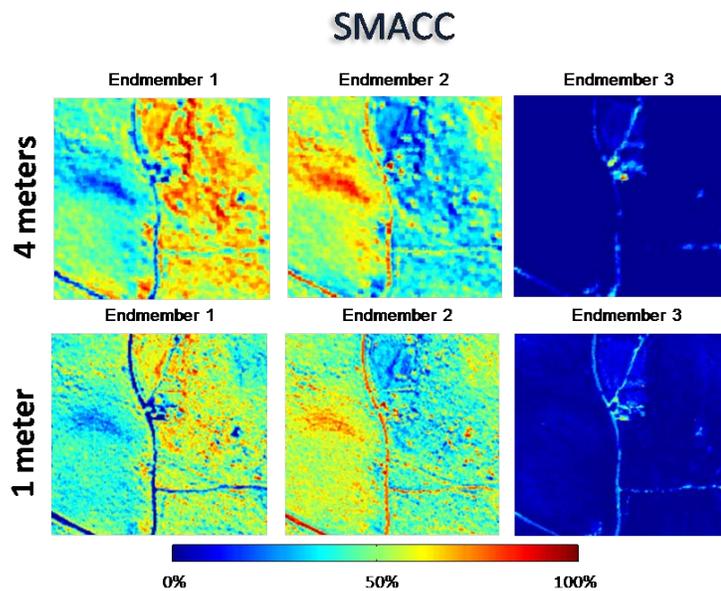


Figure 4–12: Unmixing results using SMACC with 3 endmembers from Guanica Area at one and four meters

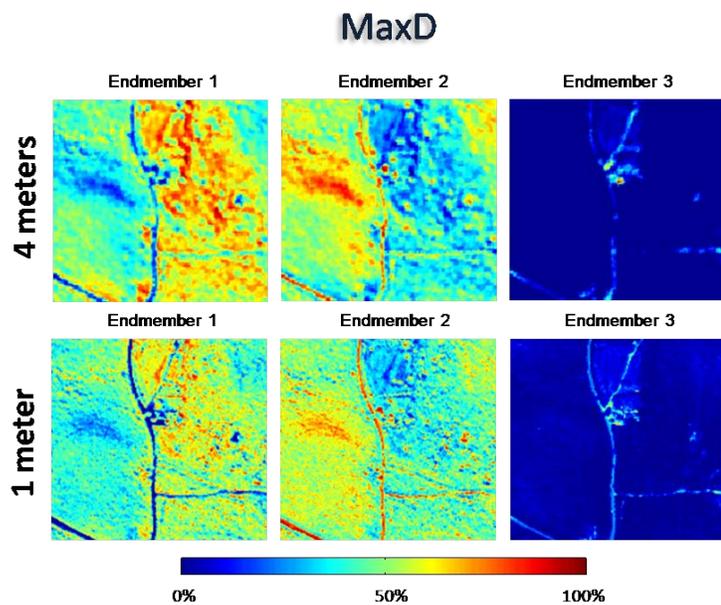


Figure 4–13: Unmixing results using MaxD with 3 endmembers from Guanica Area at one and four meters

Cosine Angle distance for results with 3 endmembers				
Endmembers Extraction Method	Comparison between:	Endm 1	Endm 2	Endm 3
cPMF	1m/ 4m	1.00	1.00	0.96
SMACC	1m/ 4m	1.00	0.98	1.00
MaxD	1m/ 4m	1.00	0.98	1.00

Table 4–1: Comparison of results with 3 endmembers at different spatial resolution using cosine angle distance

In Figures 4–14 to 4–16, we show abundance maps for similar endmembers obtained with the different methods. Tables 4–2 and 4–3, present the cosine of angle among the similar endmembers that produce the abundances maps shown in Figures 4–14 to 4–16.

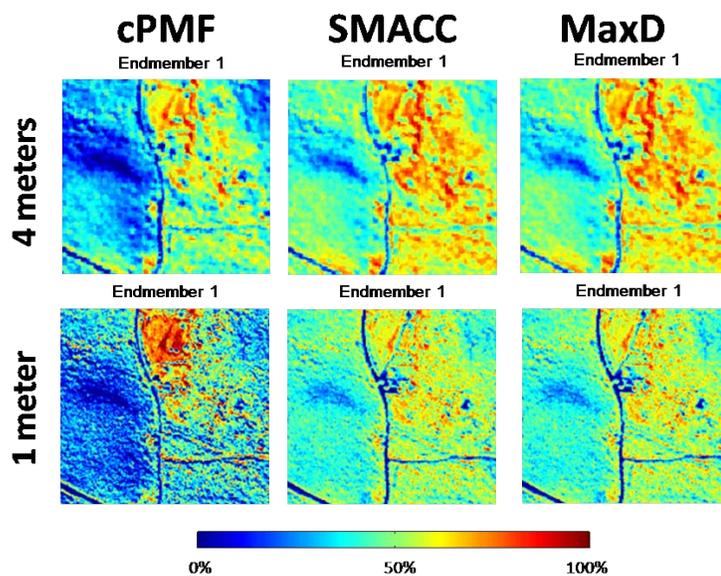


Figure 4–14: Abundance Maps for endmember 1 using cPMF, SMACC and MaxD.

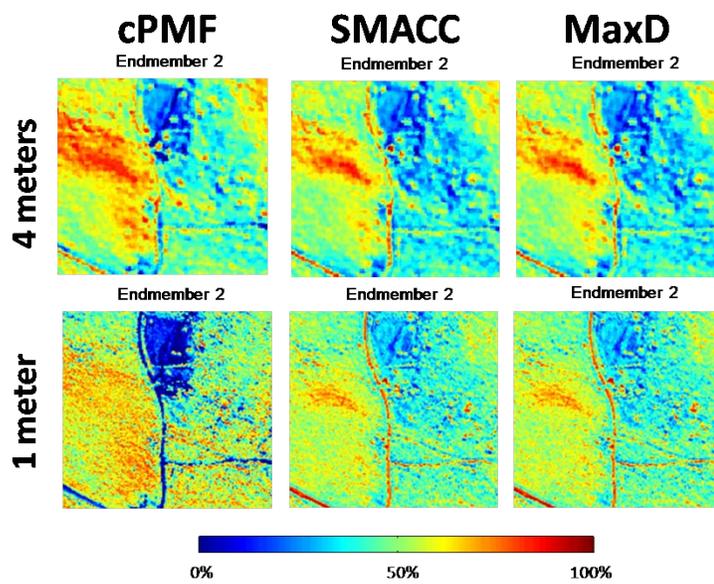


Figure 4–15: Abundance Maps for endmember 2 using cPMF, SMACC and MaxD.

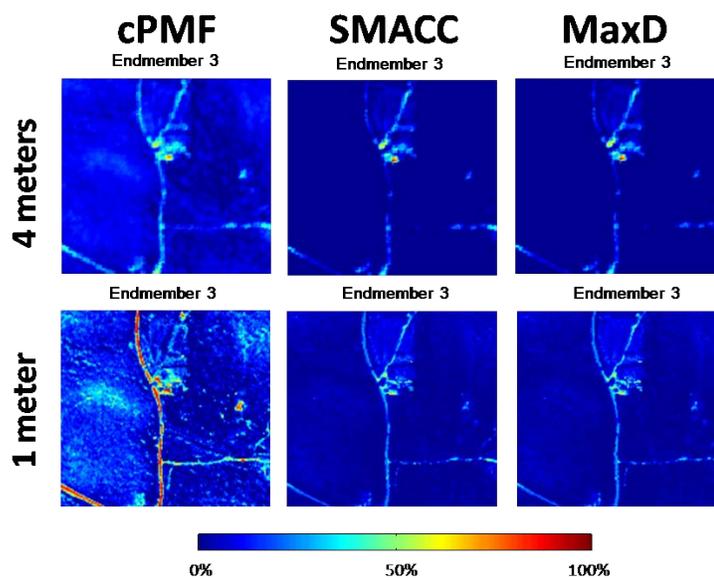


Figure 4–16: Abundance Maps for endmember 3 using cPMF, SMACC and MaxD.

Cosine Angle distance for results with 3 endmembers at 4 meters			
Comparison between:	Endm 1	Endm 2	Endm 3
cPMF/SMACC	1.00	0.96	0.98
cPMF/MaxD	1.00	0.96	0.98
SMACC/MaxD	1.00	1.00	1.00

Table 4–2: Cosine Angle distance for results with 3 endmembers at 4 meters.

Cosine Angle distance for results with 3 endmembers at 1 meter			
Comparison between:	Endm 1	Endm 2	Endm 3
cPMF/SMACC	1.00	0.96	0.98
cPMF/MaxD	1.00	0.96	0.98
SMACC/MaxD	1.00	1.00	1.00

Table 4–3: Cosine Angle distance for results with 3 endmembers at 1 meter

For the five endmembers case, the results for SMACC and MaxD do not extract information significantly different from the already extracted with 3 endmembers (See Figures 4–12, 4–13, 4–19, and 4–20). However, for the image with spatial resolution of 4 meters, the first two maps for each case (cPMF, SMACC and MaxD), retrieve the same areas, but the results for the last three abundance maps for SMACC and MaxD are difficult to interpret. For cPMF, results at 4 meters, the results are coherent; the maps extract the following classes: the small area of grass, two vegetation areas at each side of the image, a road class and in the last one recover the buildings. The same classes are retrieved with cPMF as with 1 meter, but for this case the areas are better delimited. For SMACC and MaxD at 1 meter, the results are similar to those obtained with the same algorithms at 4 meters. The best result was the obtained for cPMF with 5 endmembers (See Figure 4–18). We also present abundances maps for similar similar endmembers among methods in Figures 4–21, 4–22, 4–23, 4–24, and 4–25.

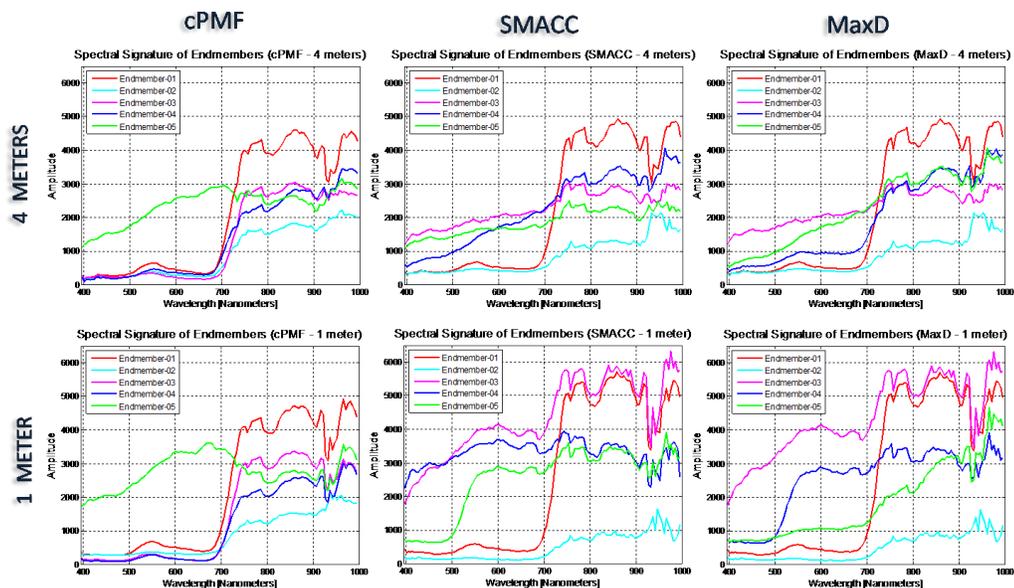


Figure 4-17: Endmembers obtained with different endmembers extraction algorithms. Results for image at 4 meters and 1 meter of spatial resolution.

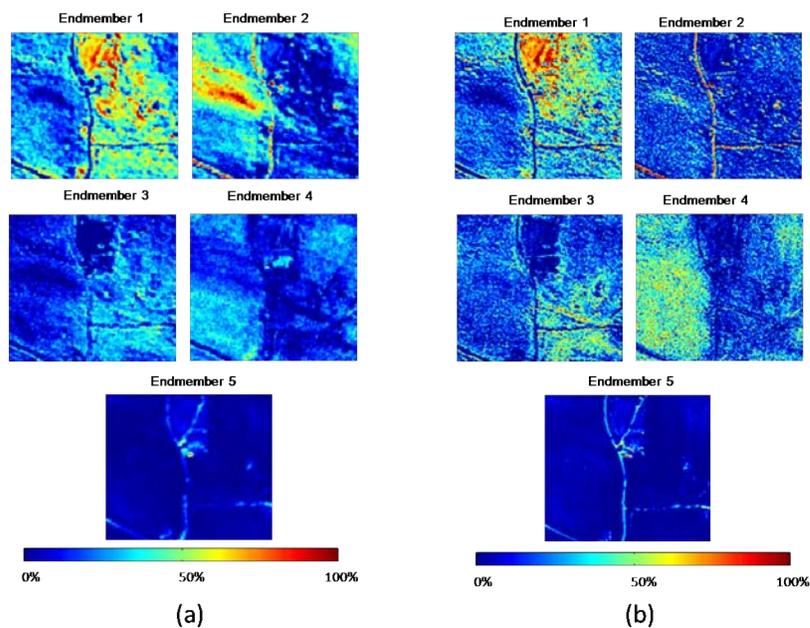


Figure 4-18: Unmixing results using cPMF with 5 endmembers. (a) Results for image at 4 meters and (b) 1 meter.

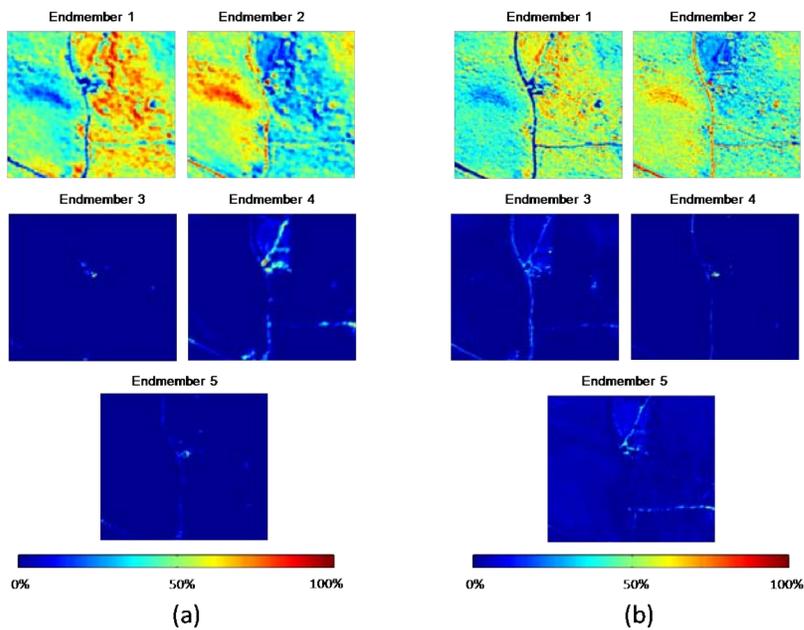


Figure 4–19: Unmixing results using SMACC with 5 endmembers. (a) Results for image at 4 meters and (b) 1 meter.

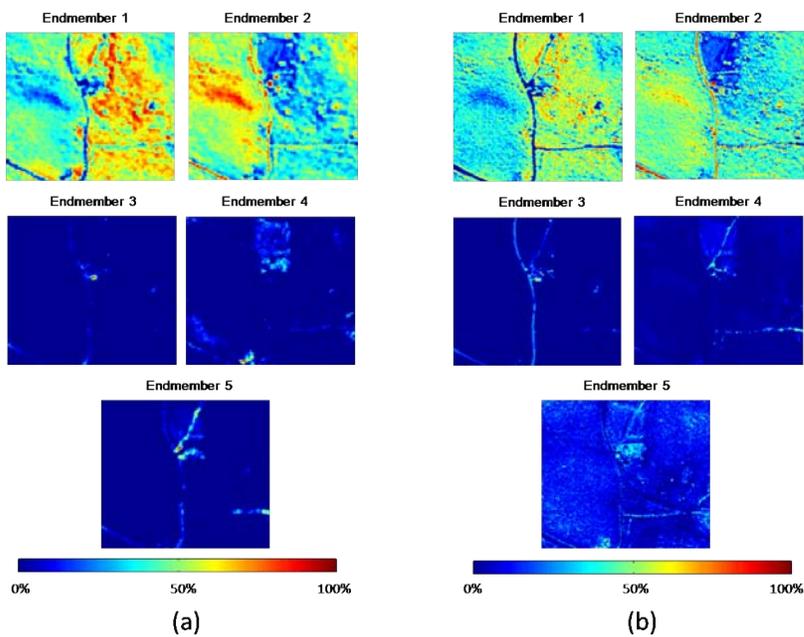


Figure 4–20: Unmixing results using MaxD with 5 endmembers. (a) Results for image at 4 meters and (b) 1 meter.

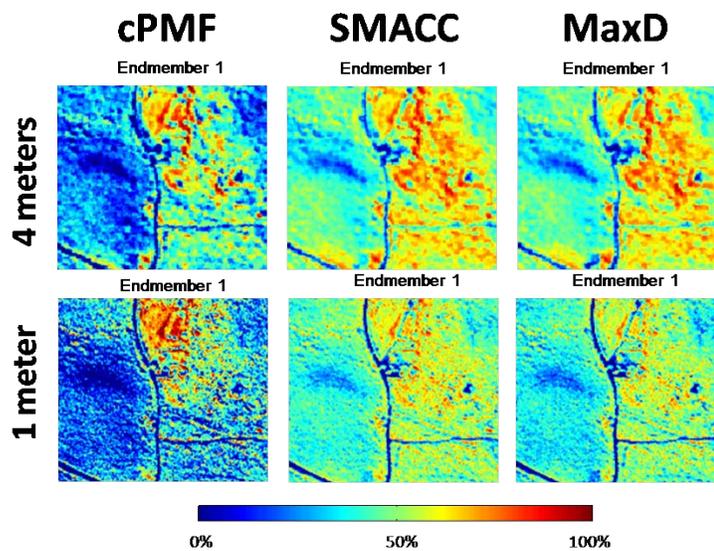


Figure 4–21: Abundance Maps for endmember 1 using cPMF, SMACC, and MaxD

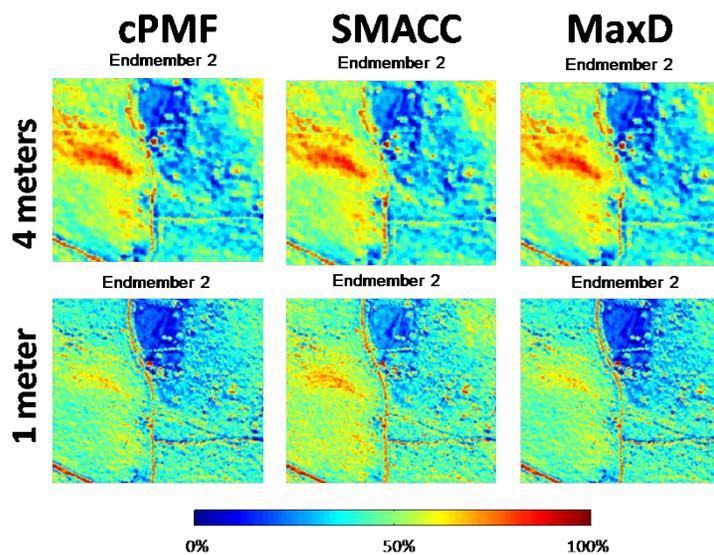


Figure 4–22: Abundance Maps for endmember 2 using cPMF, SMACC, and MaxD

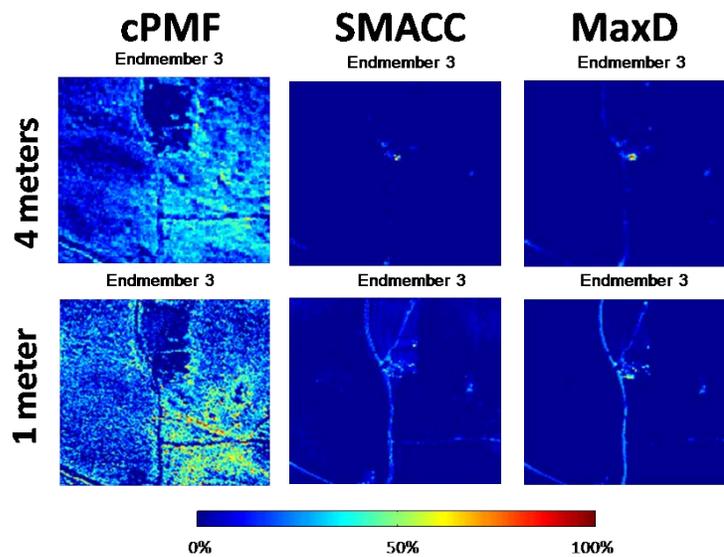


Figure 4-23: Abundance Maps for endmember 3 using cPMF, SMACC, and MaxD

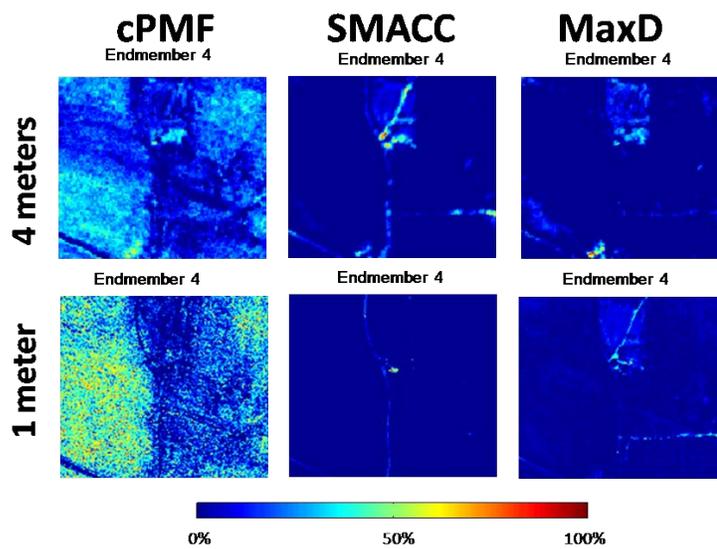


Figure 4-24: Abundance Maps for endmember 4 using cPMF, SMACC, and MaxD

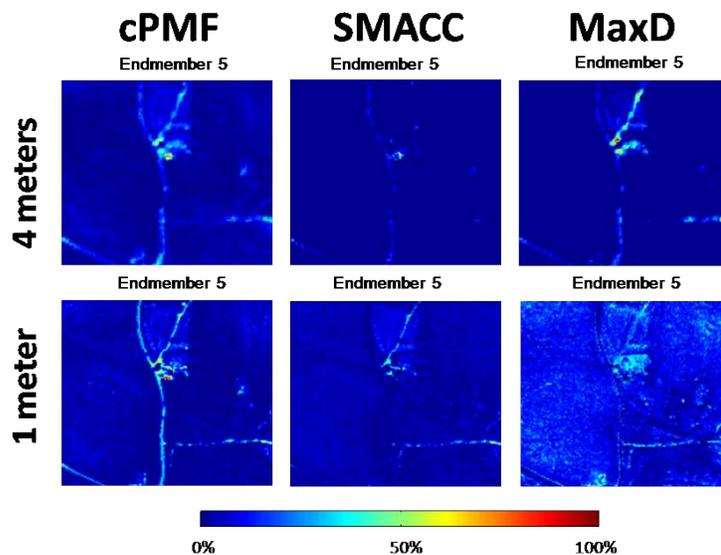


Figure 4–25: Abundance Maps for endmember 5 using cPMF, SMACC, and MaxD

Cosine Angle distance for results with 5 endmembers						
Endmembers Extraction Method	Comparison between:	Endm 1	Endm 2	Endm 3	Endm 4	Endm 5
cPMF	1m/ 4m	1.00	1.00	1.00	1.00	0.99
SMACC	1m/ 4m	1.00	0.98	1.00	0.92	0.97
MaxD	1m/ 4m	1.00	0.98	1.00	0.93	0.98

Table 4–4: Comparison of results with 5 endmembers at different spatial resolution using cosine angle distance

Cosine Angle distance for results with 5 endmembers at 4 meters					
<i>Comparison between:</i>	Endm 1	Endm 2	Endm 3	Endm 4	Endm 5
cPMF/SMACC	1.00	0.98	0.85	0.95	0.99
cPMF/MaxD	1.00	0.98	0.85	0.99	0.96
SMACC/MaxD	1.00	1.00	1.00	0.98	0.97

Table 4–5: Cosine Angle distance for results with 5 endmembers at 4 meters

Cosine Angle distance for results with 5 endmembers at 1 meter					
Comparison between:	Endm 1	Endm 2	Endm 3	Endm 4	Endm 5
cPMF/SMACC	1.00	0.99	0.85	0.74	0.96
cPMF/MaxD	1.00	0.99	0.85	0.85	0.86
SMACC/MaxD	1.00	1.00	1.00	0.95	0.93

Table 4–6: Cosine Angle distance for results with 5 endmembers at 1 meter

Table 4–4 presents the comparison between endmembers retrieved with each method across different spatial resolution changes using cosine angle distance. Table 4–5, and 4–6 present the comparison between different endmember extraction algorithms. The results agree with the discussion in above paragraphs. Table 4–7 presents the comparison of the total area abundance of the different endmembers in the scene for each method at each spatial resolution. Notice that total fraction of area are closer for cPMF than for SMACC and MaxD. Endmember 1 is the most consistent across methods. Endmember 5 is also consistent in cPMF.

Comparison of the fraction areas covered by the endmembers in Guanica Dry Forest						
Method:	cPMF		SMACC		MaxD	
Resolution:	1m	4m	1m	4m	1m	4m
Endmember 1	0.33	0.36	0.47	0.52	0.44	0.46
Endmember 2	0.16	0.26	0.47	0.46	0.42	0.51
Endmember 3	0.21	0.17	0.02	0.00	0.01	0.00
Endmember 4	0.26	0.17	0.00	0.02	0.02	0.02
Endmember 5	0.04	0.04	0.04	0.00	0.11	0.01
Sum	1.00	1.00	1.00	1.00	1.00	1.00

Table 4–7: Comparison of the fraction areas covered by the endmembers in Guanica Dry Forest

An additional analysis was included in this experiment, we use the endmembers obtained with the image at 4 meters and obtain the abundance associated in the image at 1 meter. The results are shown in Figures 4–26, 4–27, 4–28. Despite of

the result with the original endmembers for the image at 1 meter are better, these results can extract the main classes in a good way.

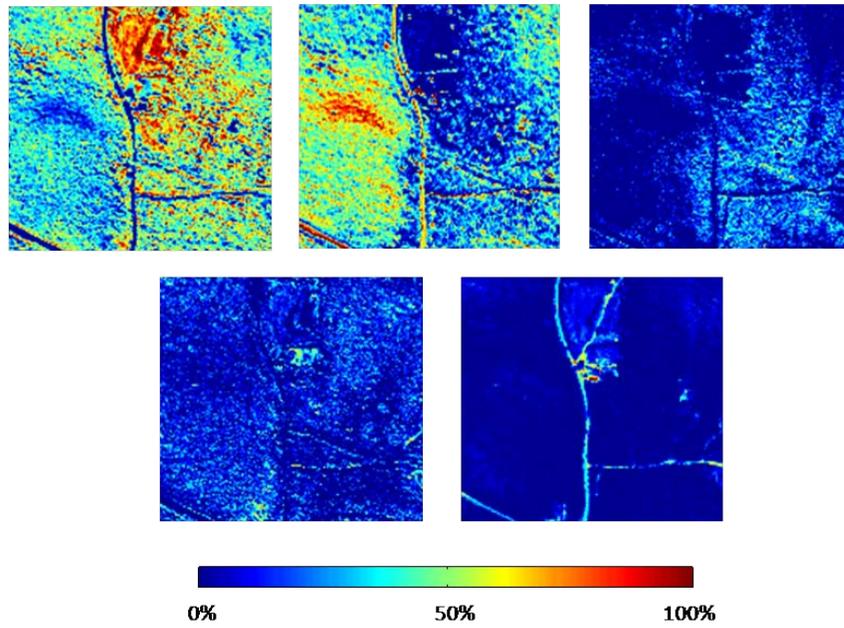


Figure 4-26: Abundance estimation using image at 1 meter and endmembers obtained by cPMF for 4 meters

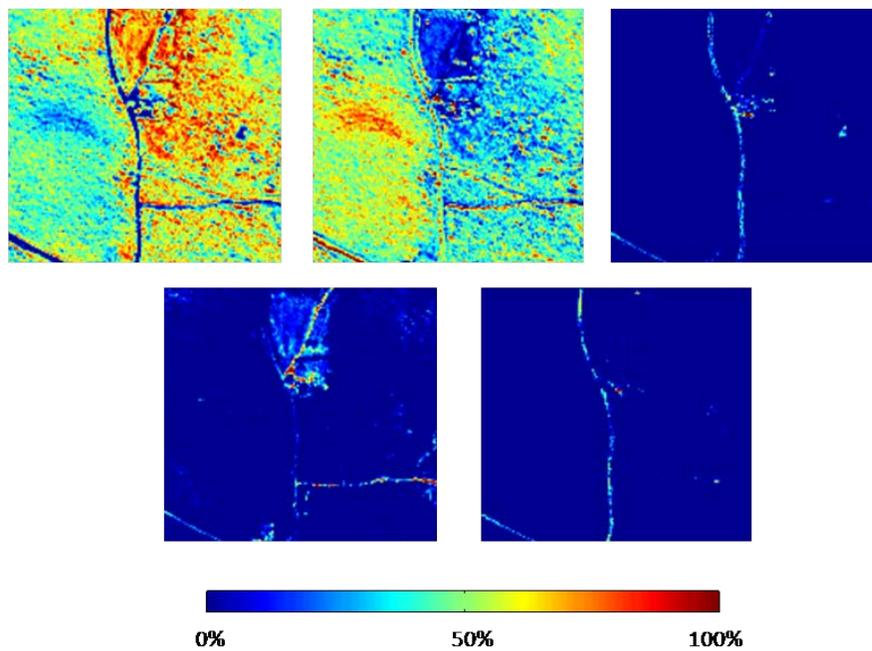


Figure 4–27: Abundance estimation using image at 1 meter and endmembers obtained by SMACC for 4 meters

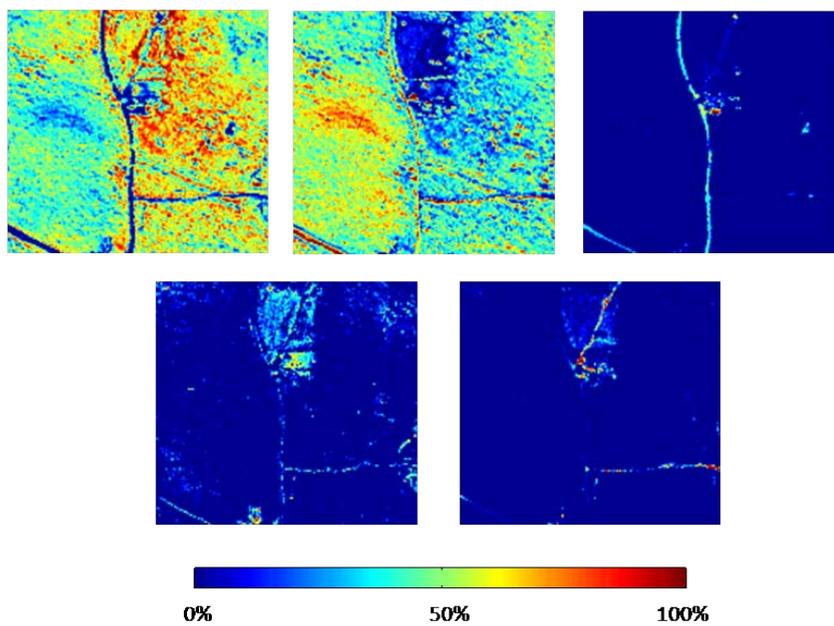


Figure 4–28: Abundance estimation using image at 1 meter and endmembers obtained by MaxD for 4 meters

cPMF is more consistent in well identified classes with point to the robustness of this approach.

In addition, unmixing results were obtained using PMF algorithm with the objective of evaluating the impact of the sum-to-one constraint imposed on the abundance coefficients. The endmembers obtained by PMF with 3 and 5 endmembers for images at 1 and 4 meters are compared with the obtained with cPMF in Figures 4–29 and 4–31. Their abundances maps are presented in Figures 4–30, and 4–32. PMF results for both resolutions are capable to extract the general spatial distribution of the classes, it is important to point out that the sum to one constraint helps to introduce coherency in the scale of the abundances maps.

It is important to mention that the number of endmember is a fact that modifies significantly the results of unmixing. Experimentally the best estimation that we found was the positive rank, we also tested the algorithms with higher number of endmembers and the results showed us for values higher than the positive rank estimate. the new endmembers have very small abundances across the image. The results points to a differncy of our method that could miss small targets.

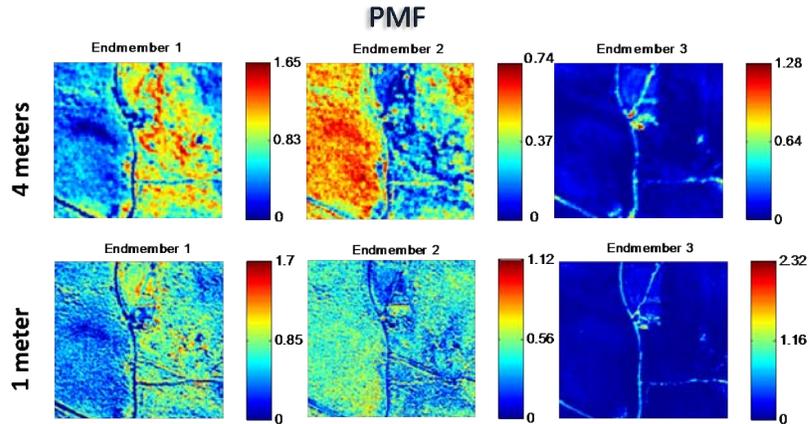


Figure 4–30: Unmixing results using PMF with 3 endmembers from Guanica at one and four meters.

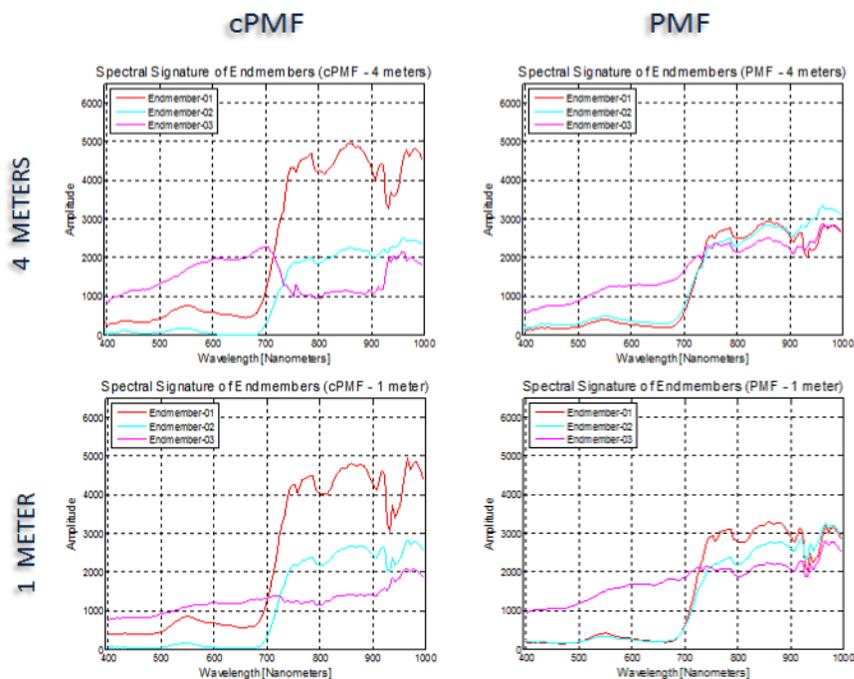


Figure 4-29: Endmembers obtained with PMF and cPMF algorithms. Results for images at 4 meters and 1 meter spatial resolution.

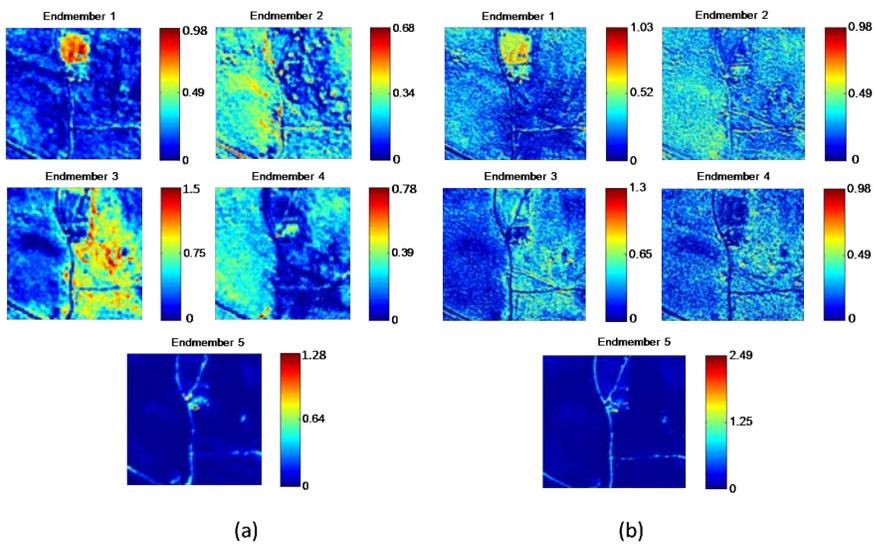


Figure 4-32: Unmixing results using PMF with 5 endmembers. (a) Results for image at 4 meters and (b) 1 meter.

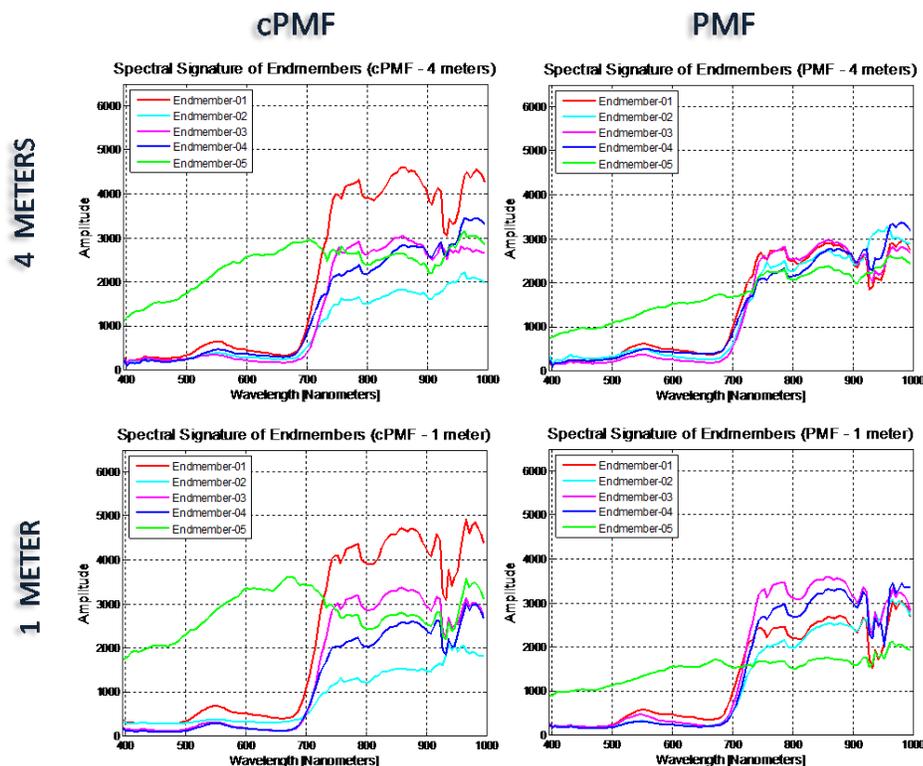


Figure 4–31: Endmembers obtained with PMF and cPMF algorithms. Results for images at 4 meters and 1 meter spatial resolution.

4.3 Experiment 2 - Enrique Reef Area

4.3.1 Imagery

The Enrique Reef data used in this experiment is also part of the data collected during the 2007 Puerto Rico Hyperspectral Mission[37]. We use data at three different spatial resolutions: one, two, and eight meters. Figure 4–33 presents a true color composite of Enrique Reef image at one, two, and eight meters.

The Enrique Reef area is a widely studied area, Figure 4–34 shows a class map published in 2008 [38], where spatial classes can be observed. In addition, we

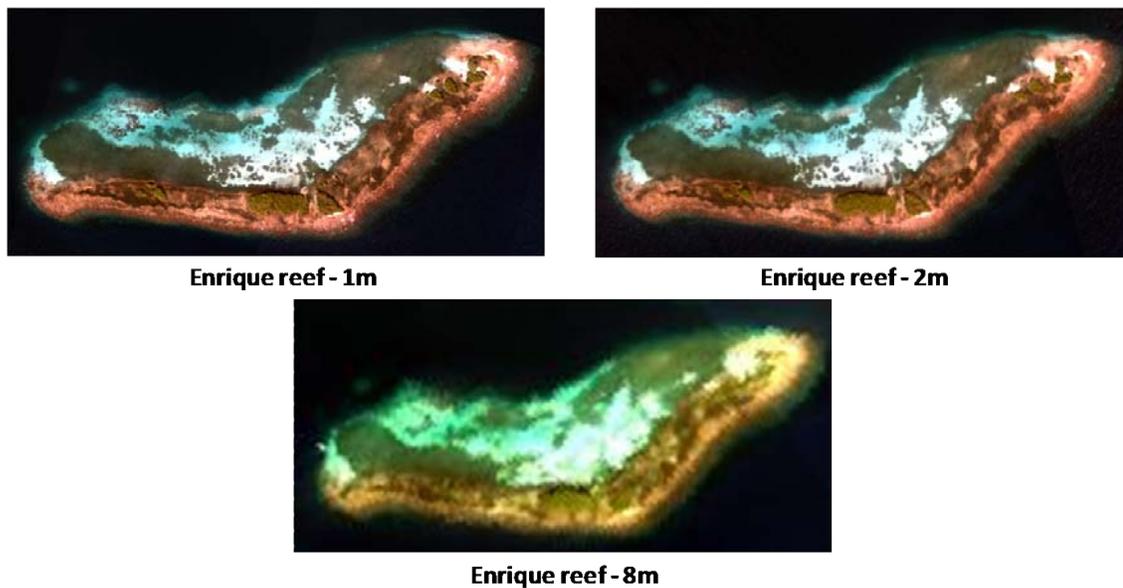


Figure 4–33: Enrique reef image at three different spatial resolutions.

present a high definition image taken from Google Earth where the main areas are observable (See Figure 4–35).

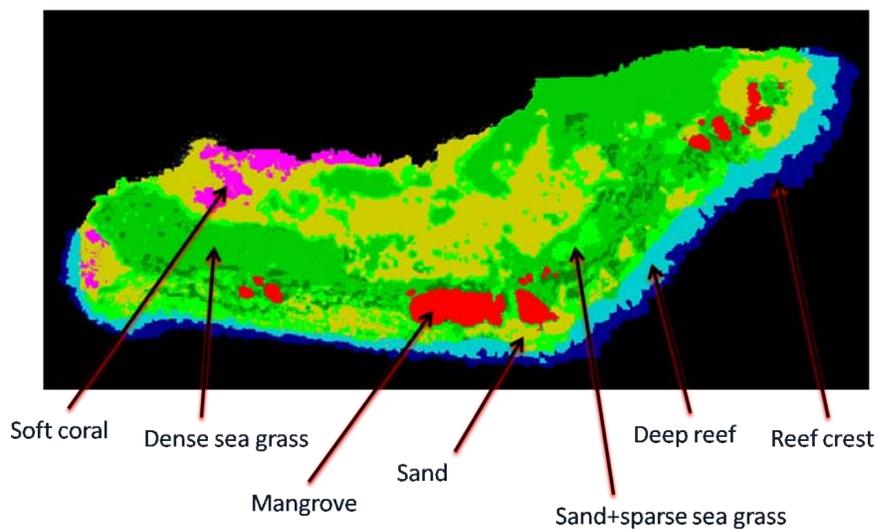


Figure 4–34: Class map from Enrique Reef at 1 meter spatial resolution.



Figure 4–35: High definition image from Google Earth.

4.3.2 Results and Discussion

Estimation of the Number of Endmembers

As in the first experiment, the rank and the positive-rank were estimated using the Scree test and the Fitting error analysis respectively.

- Rank Estimation

Figure 4–36 shows the results for the Scree test applied to Enrique Reef images at different spatial resolutions. The number of endmembers estimated was four.

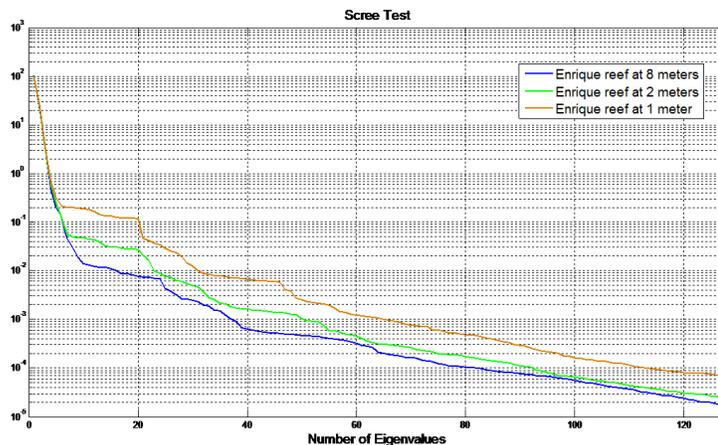


Figure 4–36: Scree Test for Enrique Reef image at different spatial resolutions.

- Positive-Rank Estimation

The fitting error analysis was applied to the cPMF results of the Enrique Reef images at one, two, and eight meters of spatial resolutions. cPMF was run across 2 to 10 endmembers. The estimation of positive-rank was found equal to five. Figure 4–37 presents the results.

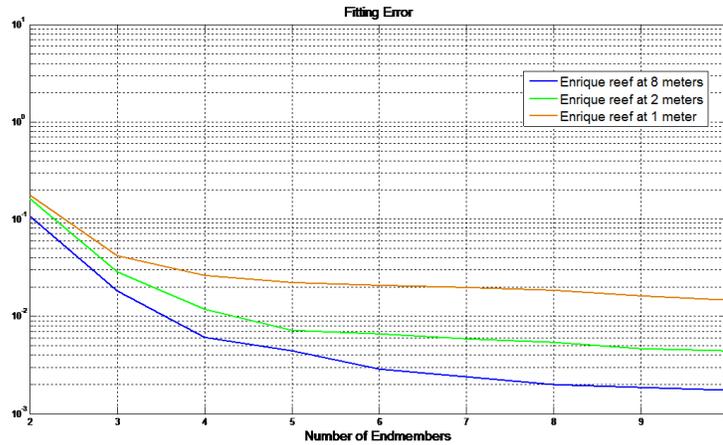


Figure 4–37: Fitting error curves of Enrique Reef Images.

Endmembers Extraction and Abundance Estimation

Owing to the results obtained in the rank and positive-rank estimation, the endmember extraction methods were run for 4 and 5 endmembers. First, we will show the results for 4 endmembers.

Figure 4–38 presents the four endmembers extracted by cPMF, SMACC and MaxD for each spatial resolution. In addition, the abundances maps associated with each endmember were obtained. Abundance maps for cPMF endmembers are given in Figures 4–39 to 4–41. Abundance maps for SMACC results are shown in Figures 4–42 to 4–44. Finally, abundances maps for MaxD results are presented in Figures 4–45 to 4–47.

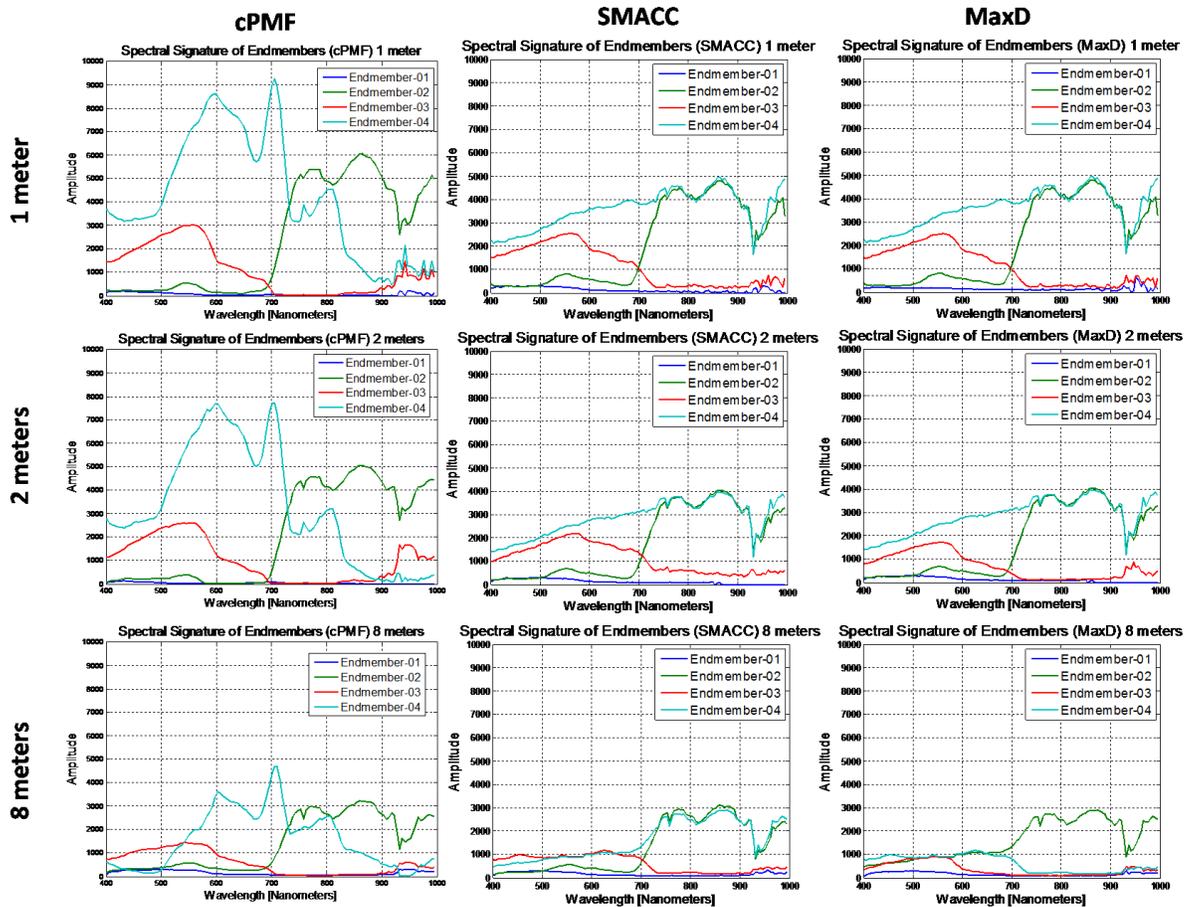


Figure 4–38: Endmembers Spectral signatures , from Enrique Reef Images, using cPMF, SMACC and MaxD. (4 endmembers case)

The endmembers obtained were organized by similarity among the spatial resolution and methods. The first endmember, in dark blue (see Figure 4–38), is associated with sea water. In Figure 4–48, the abundances maps associated with the first endmember in all the methods and resolutions are shown.

The second endmember is associated with the mangrove that is clearly shown in the class map of Figure 4–34 in red. Despite the fact that the spectral signatures are diverse in scale over the different methods and spatial resolutions, in all cases

the general form of a vegetation spectral signature is preserved. Figure 4–49 shows the result for mangrove at different spatial resolution with the different endmembers extraction methods.

The third endmember represents the area primarily covered by sand (see Figure 4–50). The area with highest abundances agree with sand area in the class map presented in Figure 4–34.

The last endmember is associated with the reef as well. Figure 4–51 shows the abundances maps associated with the reef crest and deep reef area.

In general, the results with 4 endmembers in all methods, identified similar classes associated with similar areas. The performance of the three methods was very good and the results were consistent across different spatial resolutions. However notice the confusion between water and sea grass area.

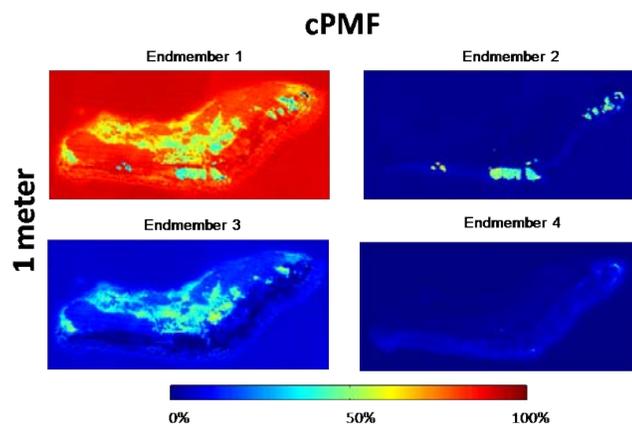


Figure 4–39: cPMF result for Enrique Reef image at 1 meter with 4 endmembers.

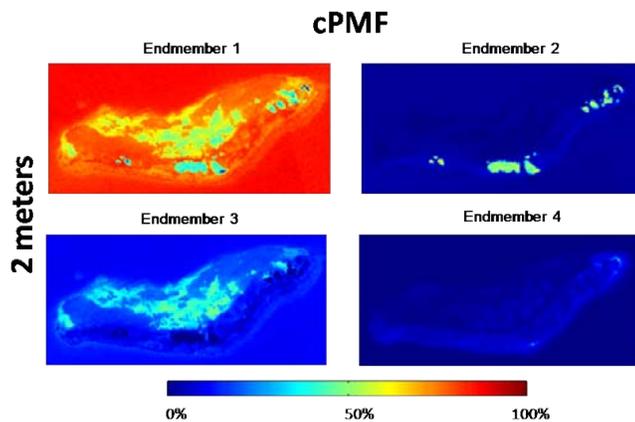


Figure 4–40: cPMF result for Enrique Reef image at 2 meters with 4 endmembers.

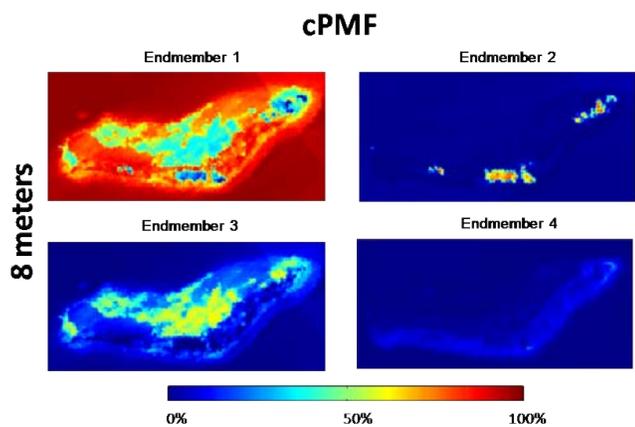


Figure 4–41: cPMF result for Enrique Reef image at 8 meters with 4 endmembers.

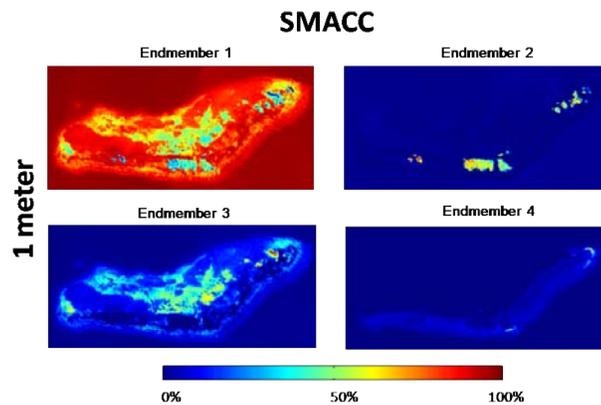


Figure 4–42: SMACC result for Enrique Reef image at 1 meter with 4 endmembers.

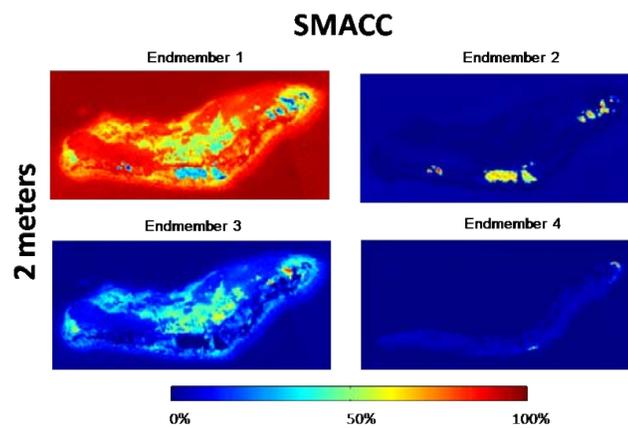


Figure 4–43: SMACC result for Enrique Reef image at 2 meters with 4 endmembers.

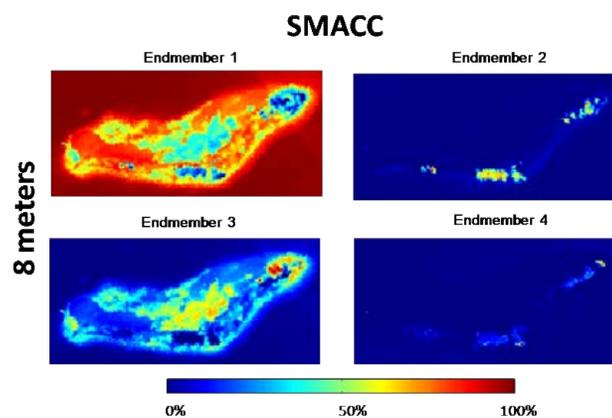


Figure 4–44: SMACC result for Enrique Reef image at 8 meters with 4 endmembers.

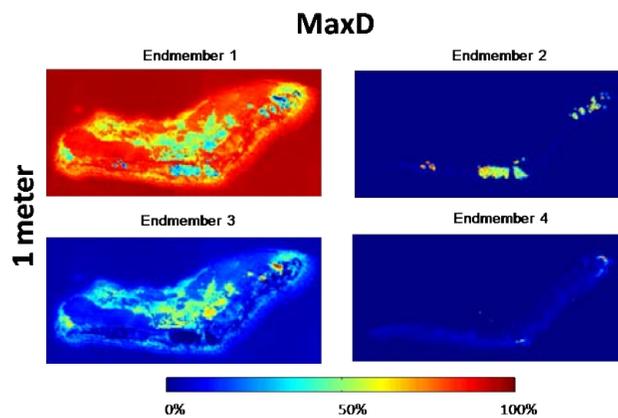


Figure 4–45: MaxD result for Enrique Reef image at 1 meter with 4 endmembers.

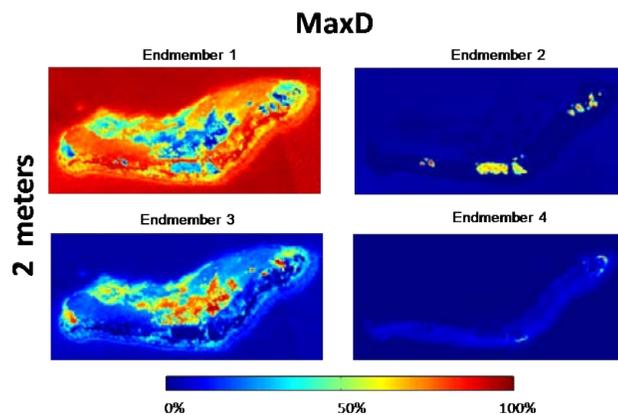


Figure 4–46: MaxD result for Enrique Reef image at 2 meters with 4 endmembers.

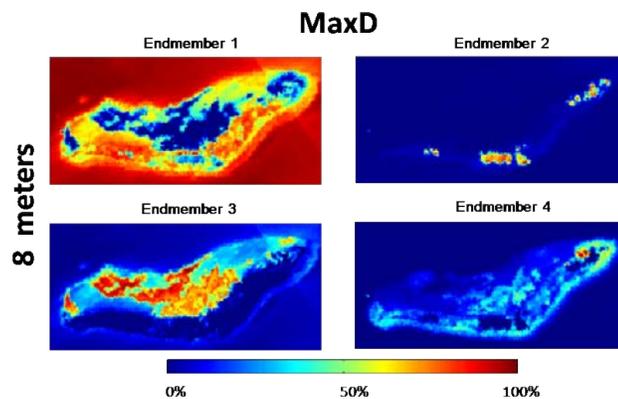


Figure 4–47: MaxD result for Enrique Reef image at 8 meters with 4 endmembers.

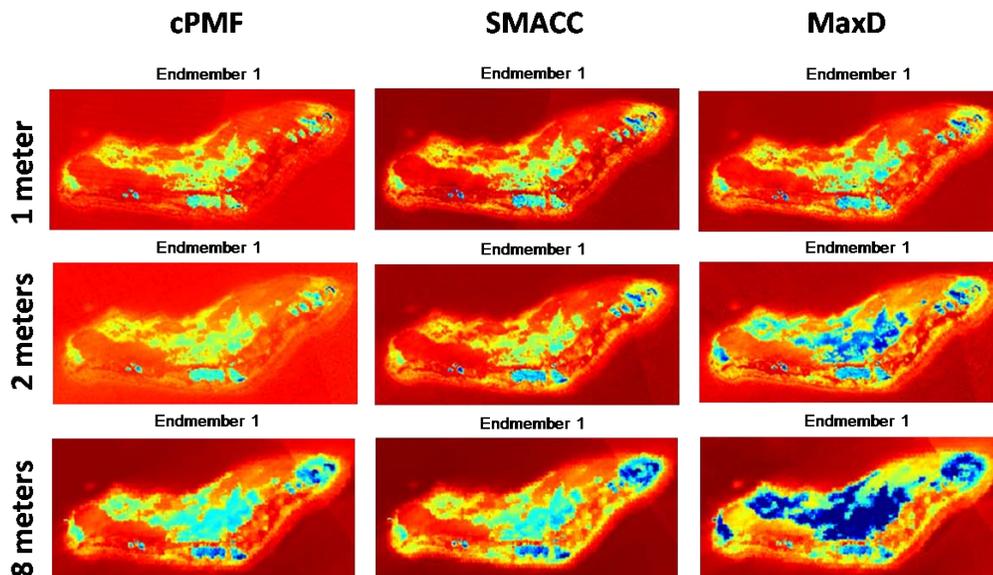


Figure 4–48: Comparison for endmember 1 between methods for different spatial resolutions.

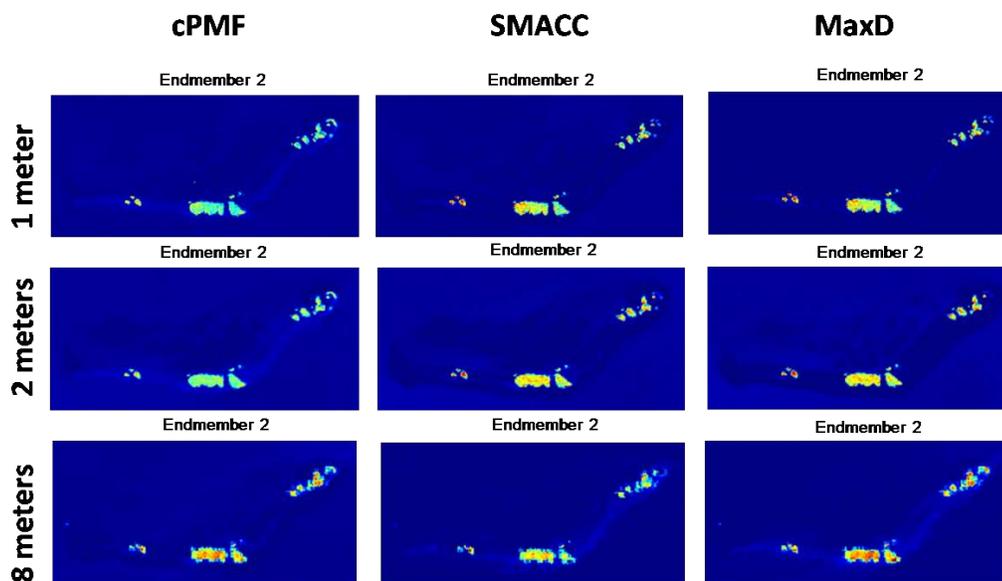


Figure 4–49: Comparison for endmember 2 between methods for different spatial resolutions.

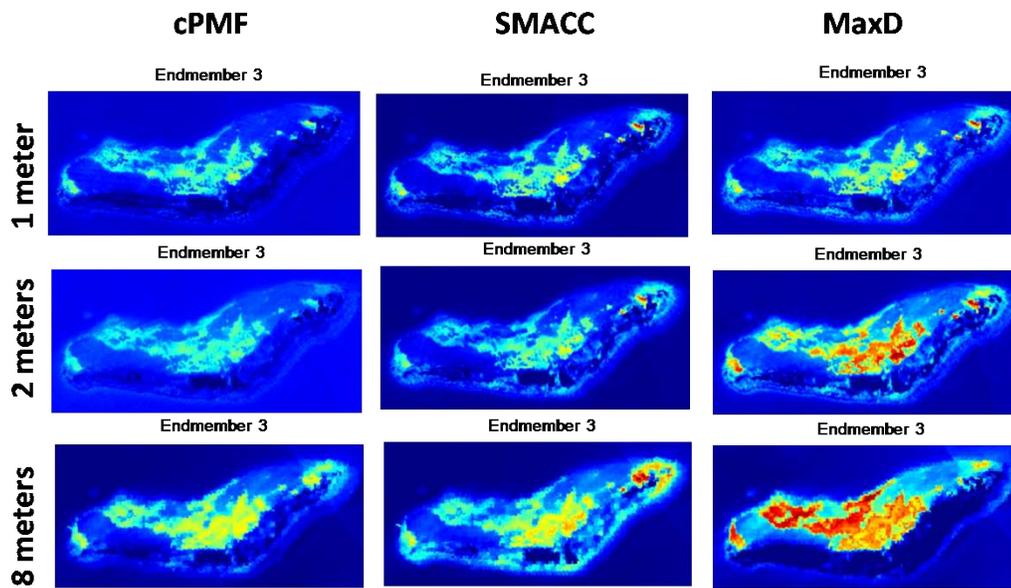


Figure 4–50: Comparison for endmember 3 between methods for different spatial resolutions.

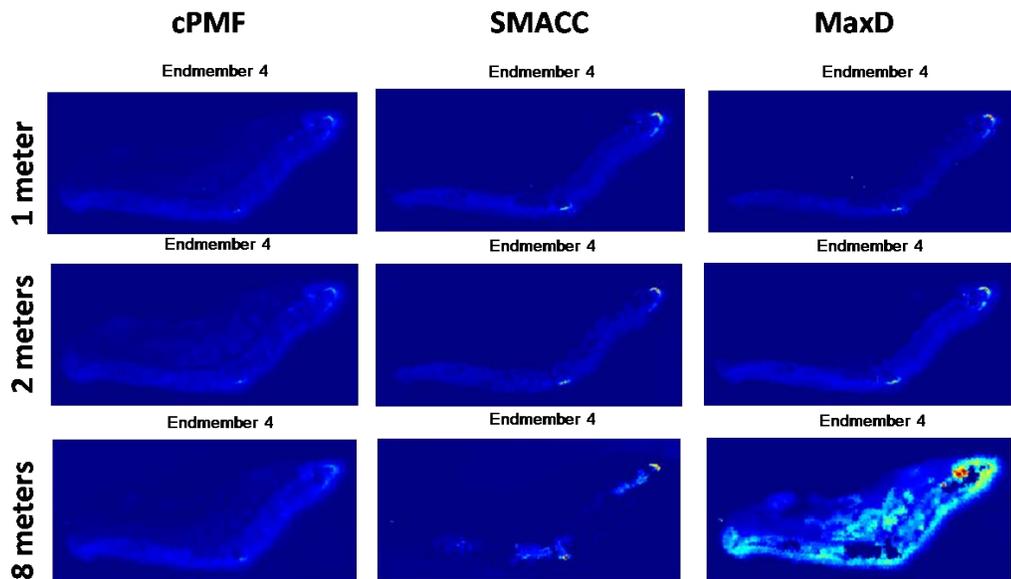


Figure 4–51: Comparison for endmember 4 between methods for different spatial resolutions.

Cosine Angle distance for results with 4 endmembers at 1 meter				
<i>Comparison between:</i>	Endm 1	Endm 2	Endm 3	Endm 4
cPMF/SMACC	0.90	1.00	0.97	0.82
cPMF/MaxD	0.77	1.00	0.97	0.82
SMACC/MaxD	0.84	1.00	1.00	1.00

Table 4–8: Cosine Angle distance for 4 endmembers results, for Enrique Reef at 1 meter

Cosine Angle distance For results with 4 endmembers at 2 meters				
<i>Comparison between:</i>	Endm 1	Endm 2	Endm 3	Endm 4
cPMF/SMACC	0.71	0.99	0.88	0.74
cPMF/MaxD	0.71	0.99	0.98	0.74
SMACC/MaxD	1.00	1.00	0.95	1.00

Table 4–9: Cosine Angle distance for 4 endmembers results, for Enrique Reef at 2 meters

Cosine Angle distance For results with 4 endmembers at 8 meters				
<i>Comparison between:</i>	Endm 1	Endm 2	Endm 3	Endm 4
cPMF/SMACC	0.99	1.00	0.92	0.71
cPMF/MaxD	0.98	0.97	0.98	0.76
SMACC/MaxD	0.99	0.97	0.88	0.57

Table 4–10: Cosine Angle distance for 4 endmembers results, for Enrique Reef at 8 meters

Tables 4–8 to 4–10, present the cosine angle distance among the endmembers obtained with the different methods. We found that the results between SMACC and MaxD are closer than the results from cPMF. However, the results with cPMF are similar across different spatial resolutions. The fourth endmember was the endmember that presented the most differences between cPMF and the other two methods. In addition, it is important to mention that the good results obtained with SMACC

and MaxD idealize that the image has “pure” pixels that permit to make good identification of the main classes.

The following results were obtained using five endmembers. Figure 4–52 presents the endmembers obtained using cPMF, SMACC, and MaxD across the different spatial resolutions. The results obtained using SMACC and MaxD are very similar, which is supported by Tables 4–11 to 4–13, where the cosine angle distance among endmembers are compared. However, the first four classes present a behavior similar to those with four endmembers. Again, the first endmember represents water (see Figure 4–62), the second one describes mangrove areas (see Figure 4–63), the third endmember represents sand (see Figure 4–64), and the fourth endmember describes the reef (see Figure 4–65). The fifth endmember difficult to relate to a particular class (see Figure 4–66).

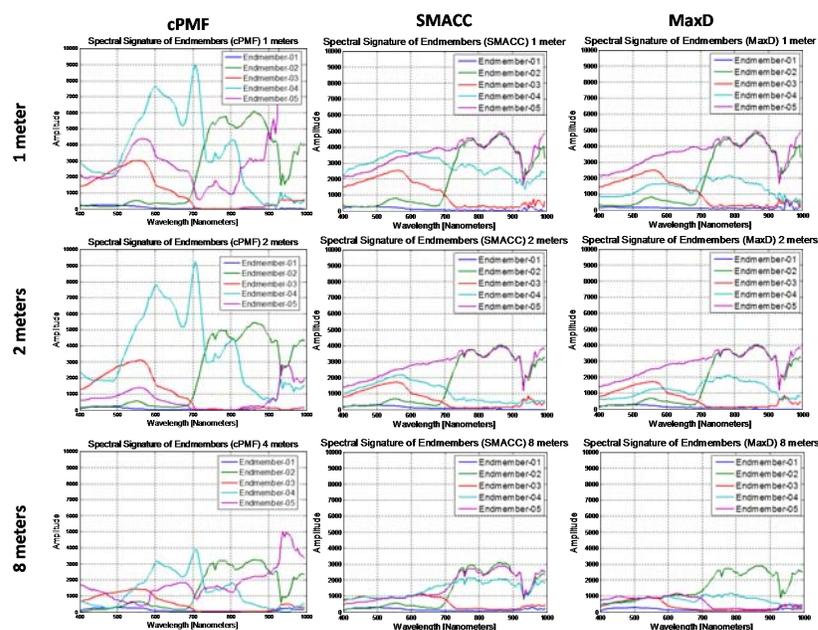


Figure 4–52: Spectral signatures obtained for different endmembers, from Enrique Reef images, using cPMF, SMACC and MaxD. Case with 5 endmembers.

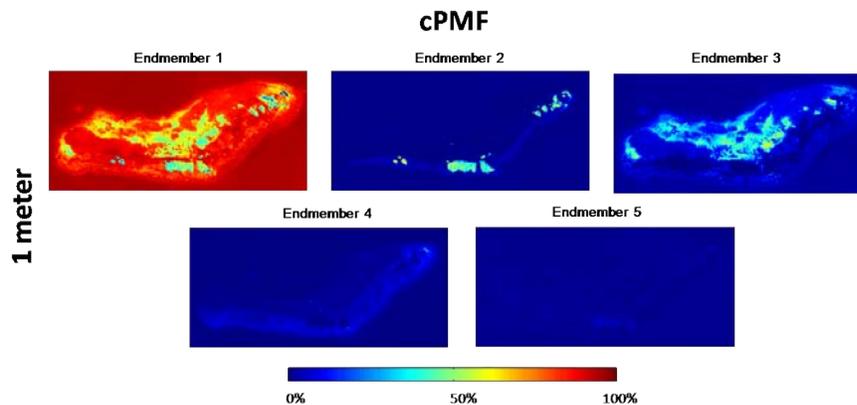


Figure 4-53: cPMF result for Enrique Reef image at 1 meter with 5 endmembers.

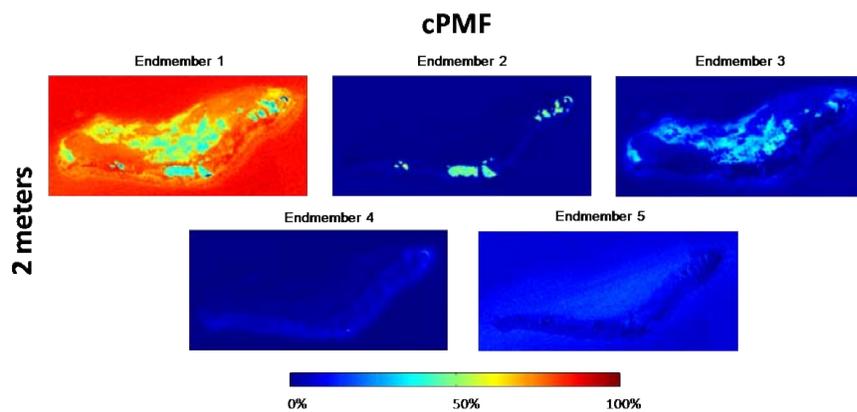


Figure 4-54: cPMF result for Enrique Reef image at 2 meters with 5 endmembers.

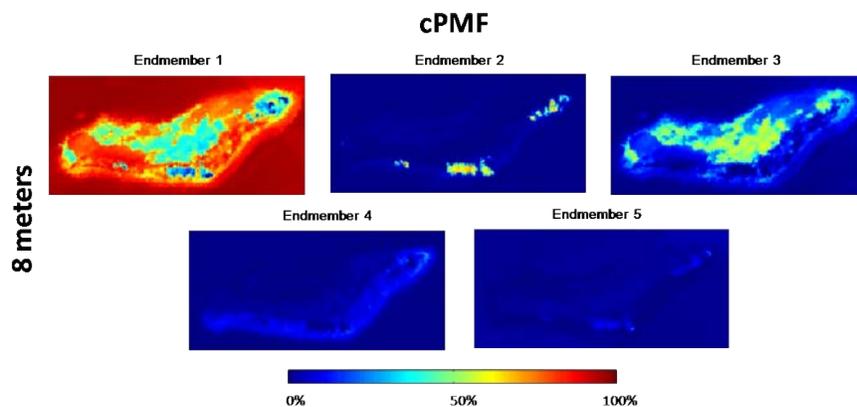


Figure 4-55: cPMF result for Enrique Reef image at 8 meters with 5 endmembers.

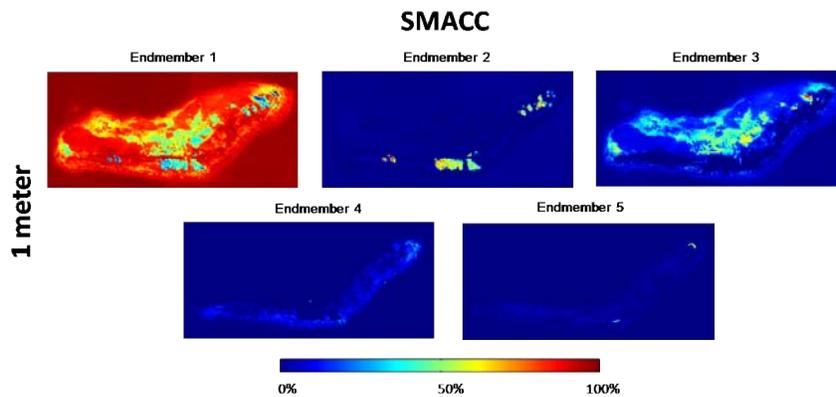


Figure 4–56: SMACC result for Enrique Reef image at 1 meter with 5 endmembers.

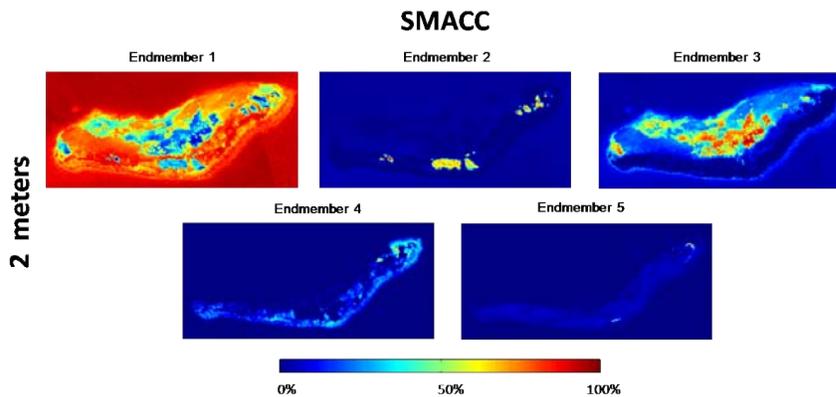


Figure 4–57: SMACC result for Enrique Reef image at 2 meters with 5 endmembers.

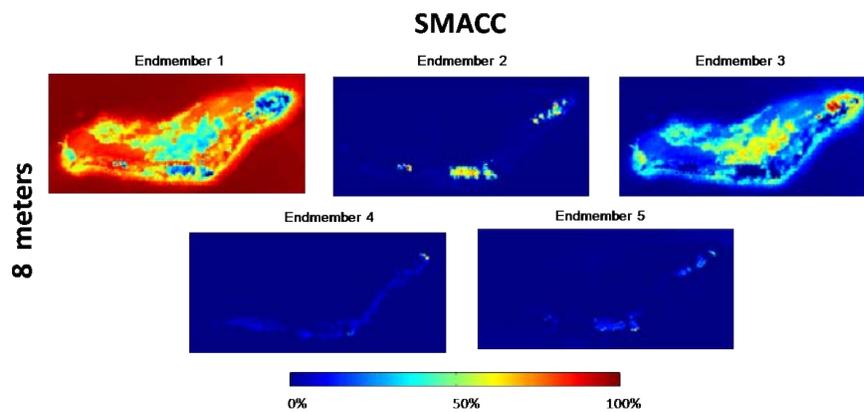


Figure 4–58: SMACC result for Enrique Reef image at 8 meters with 5 endmembers.

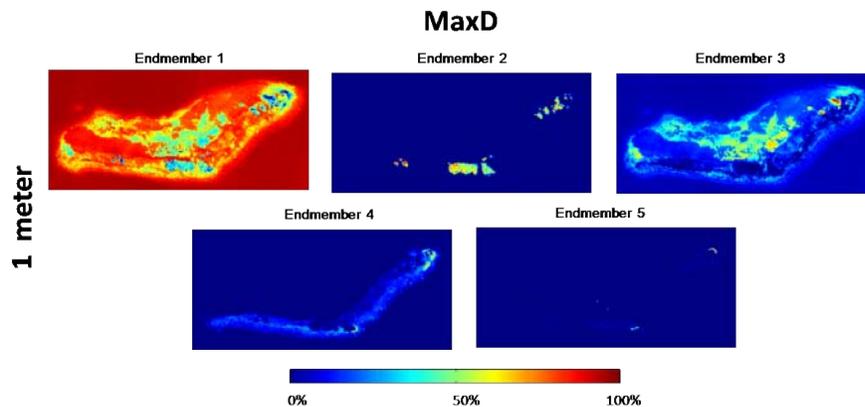


Figure 4–59: SMACC result for Enrique Reef image at 1 meter with 5 endmembers.

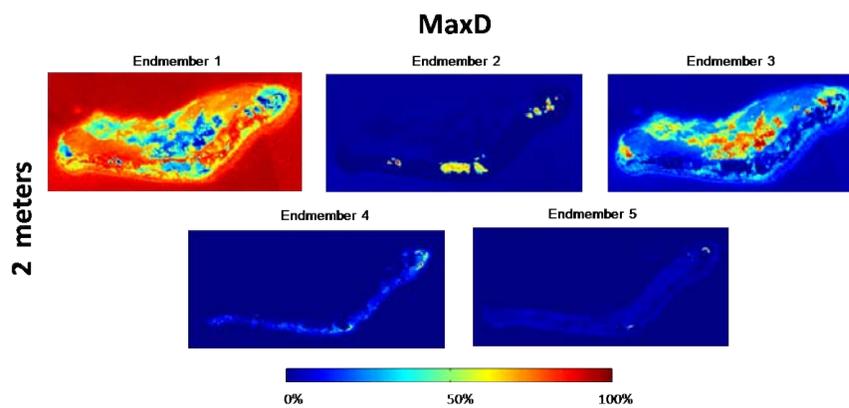


Figure 4–60: MaxD result for Enrique Reef image at 2 meters with 5 endmembers.

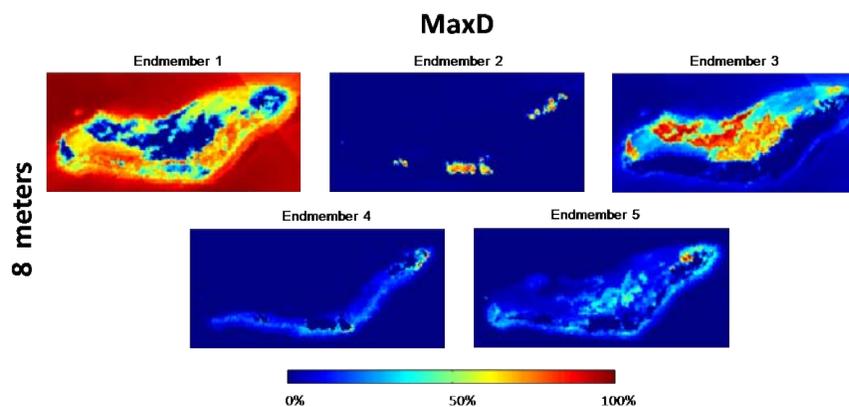


Figure 4–61: MaxD result for Enrique Reef image at 8 meters with 5 endmembers.

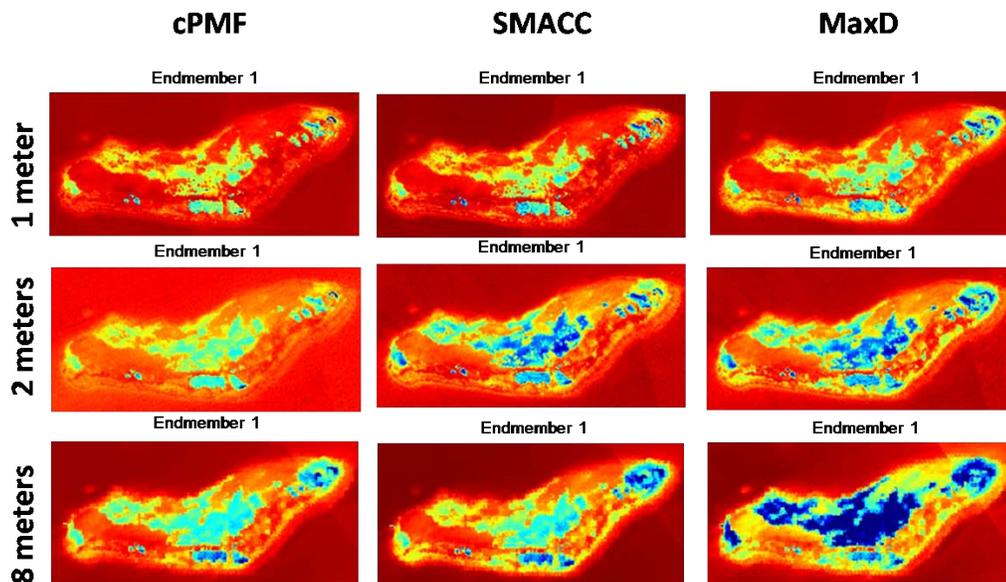


Figure 4–62: Comparison among endmember 1 of each method with different spatial resolutions.

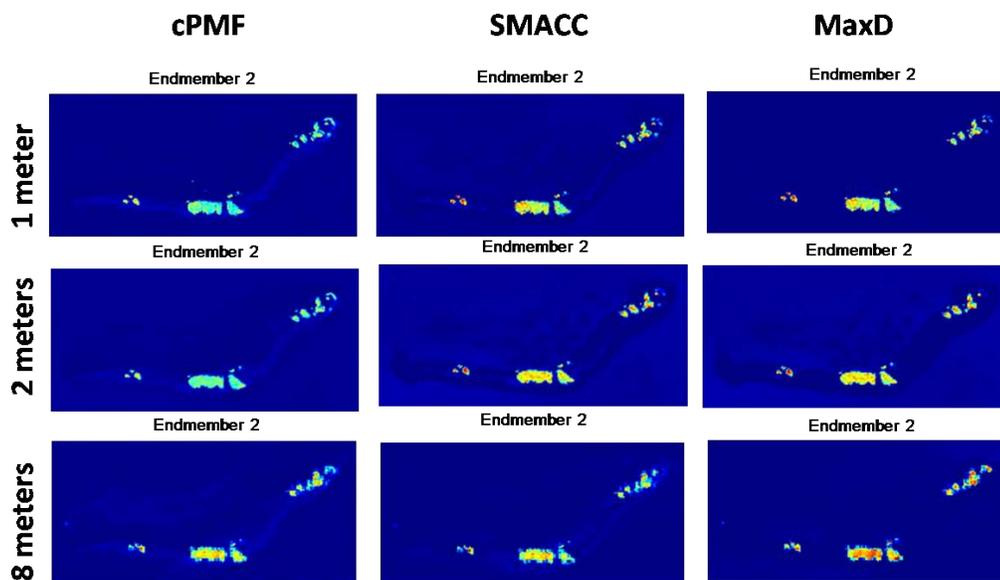


Figure 4–63: Comparison among endmember 2 of each method with different spatial resolutions.

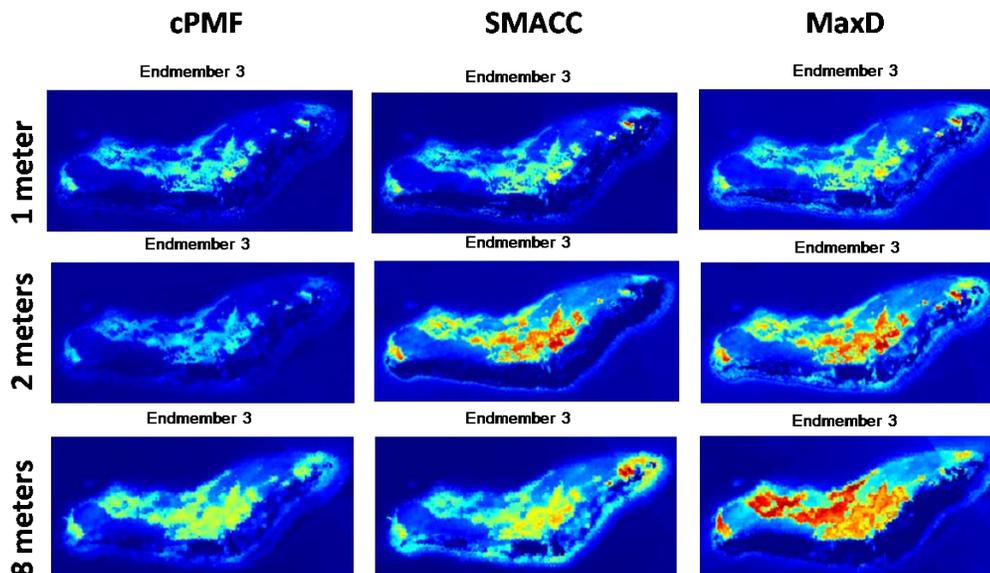


Figure 4–64: Comparison among endmember 3 of each method with different spatial resolutions.

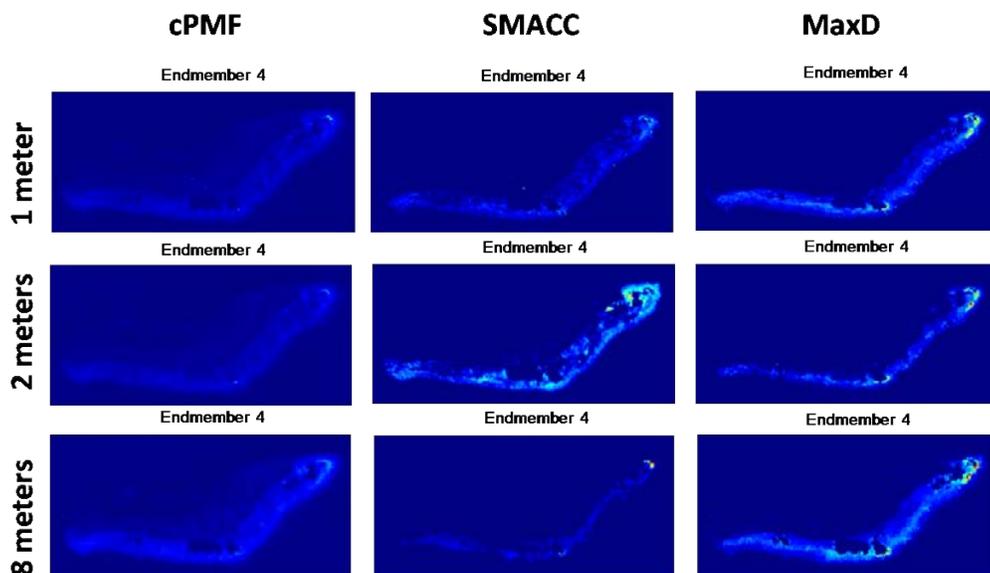


Figure 4–65: Comparison among endmember 4 of each method with different spatial resolutions.

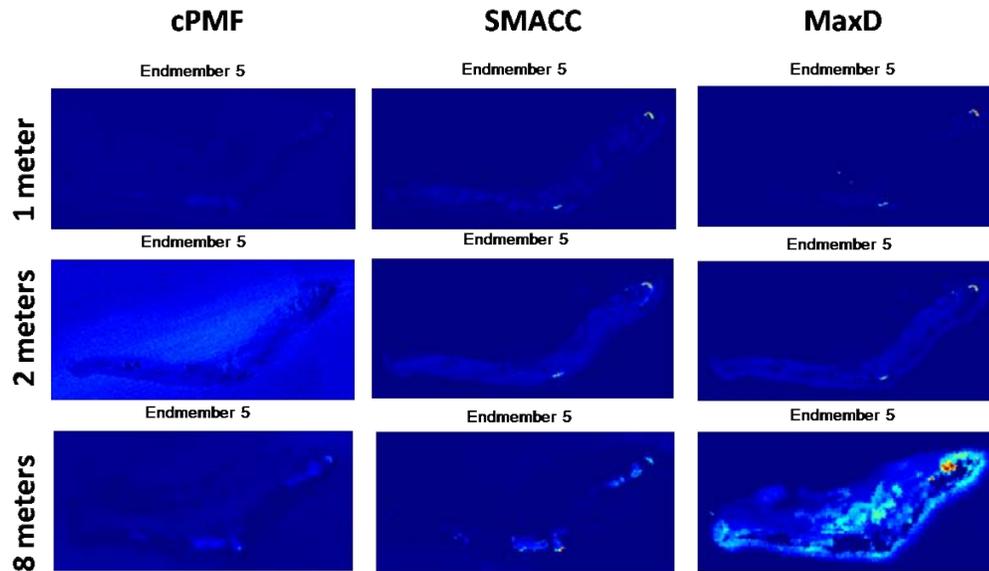


Figure 4–66: Comparison among endmember 4 of each method with different spatial resolutions.

Cosine Angle distance For results with 5 endmembers at 1 meter					
Comparison between:	Endm 1	Endm 2	Endm 3	Endm 4	Endm 5
cPMF/SMACC	0.95	0.99	0.97	0.89	0.68
cPMF/MaxD	0.70	0.99	0.97	0.87	0.68
SMACC/MaxD	0.84	1.00	1.00	0.95	1.00

Table 4–11: Cosine Angle distance for 5 endmembers results, for Enrique Reef at 1 meter

Cosine Angle distance For results with 5 endmembers at 2 meters					
Comparison between:	Endm 1	Endm 2	Endm 3	Endm 4	Endm 5
cPMF/SMACC	0.93	1.00	0.98	0.90	0.63
cPMF/MaxD	0.92	1.00	0.98	0.83	0.63
SMACC/MaxD	1.00	1.00	1.00	0.77	1.00

Table 4–12: Cosine Angle distance for 5 endmembers results, for Enrique Reef at 2 meters

Cosine Angle distance For results with 5 endmembers at 8 meters					
<i>Comparison between:</i>	Endm 1	Endm 2	Endm 3	Endm 4	Endm 5
cPMF/SMACC	0.98	1.00	0.90	0.74	0.85
cPMF/MaxD	0.98	0.96	0.98	0.85	0.64
SMACC/MaxD	0.99	0.97	0.88	0.95	0.57

Table 4–13: Cosine Angle distance for 5 endmembers results, for Enrique Reef at 8 meters

Although results in terms of endmembers correlation across scale and methods are similar to those for Guanica, it is important to point out that, as with four endmembers, we could not separate water from sea grass. We were not capable to separate after many trials and varying parameters.

To understand this problem, we decided to analyze the spectral signatures of pixels that we understand that are part of the water, sea grass and reef classes. Figure 4–67 show some of these signatures. The similarity of pixels of sea grass with water and reef is very high and point to do the confusion. We selected endmembers manually and are shown in Figure 4–68 and computed the corresponding abundance map shown in figure 4–69. The abundance maps show that with manual endmembers we were capable of separating all classes in the map. This result shows that the errors in the Enrique reef experiment was in the endmember extraction stage because the methods used cannot separate these spectral signatures.

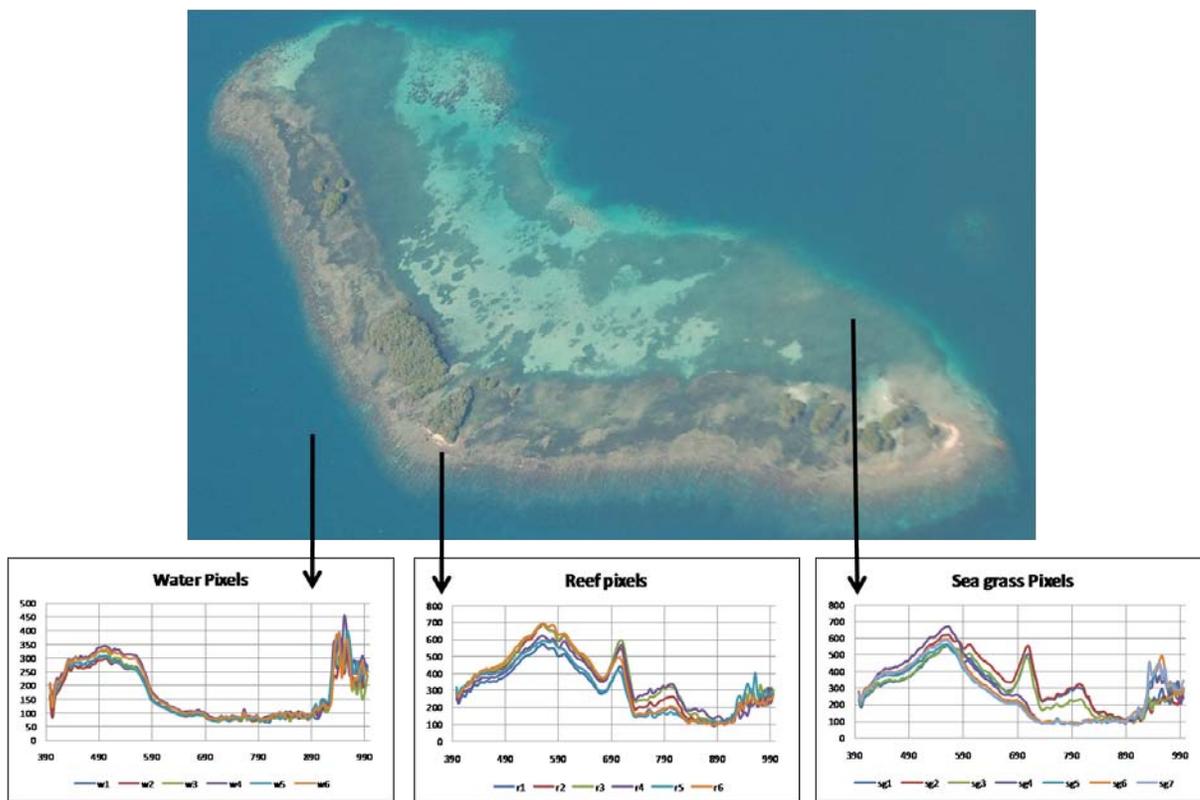


Figure 4-67: Enrique reef high resolution photo and spectral signatures associated with water, reef and sea grass classes.

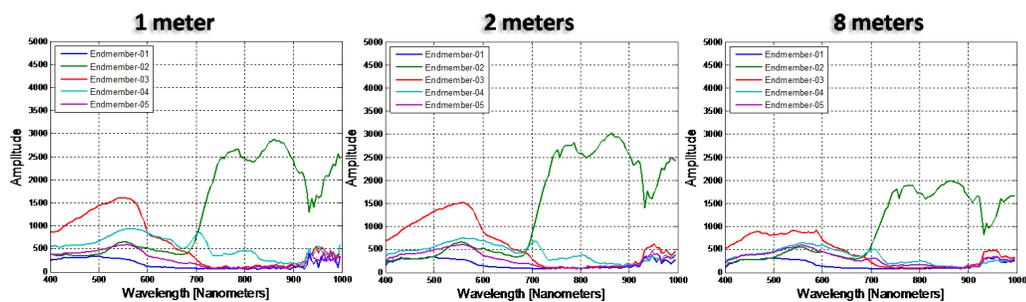


Figure 4-68: Manual endmembers selected by the average of six pixel of each class for different spatial resolution

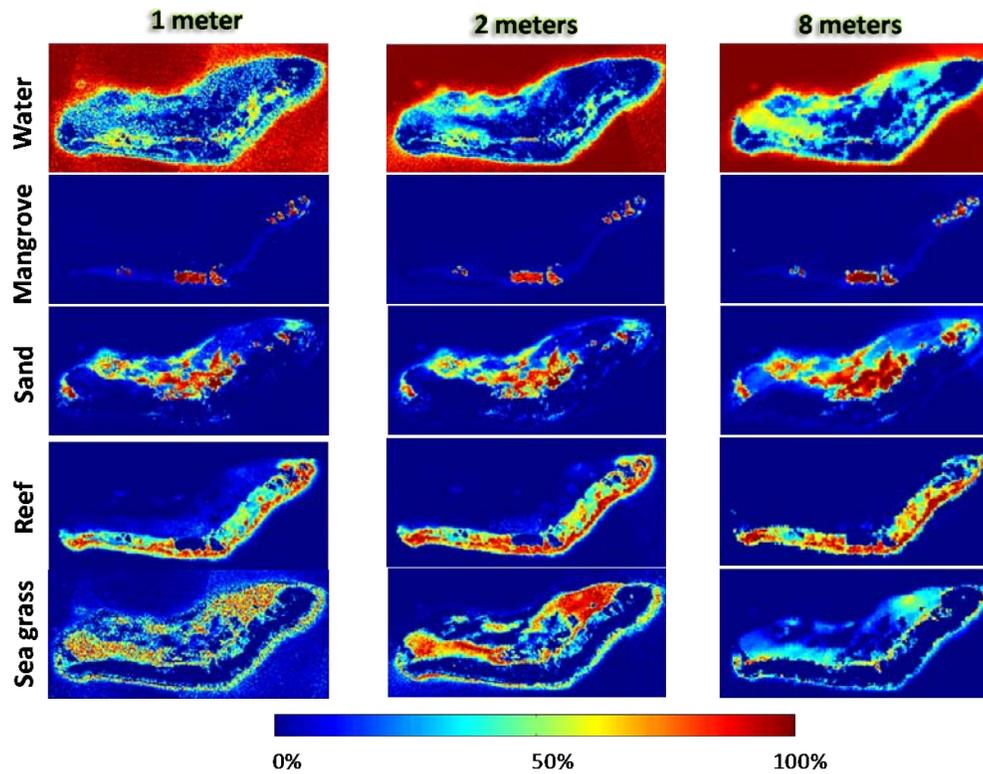


Figure 4-69: Abundance maps obtained for different spatial resolution with manual endmembers

CHAPTER 5

CONCLUSIONS AND FUTURE WORK

5.1 Conclusions

Unmixing results for algorithms that assume pure pixels present in the image did not improve when the number of endmembers is larger than the rank of the training data matrix. cPMF had similar results to the other approaches with the number of endmembers was equal to the matrix rank.

For the Guanica area, the cPMF extracted abundance distributions that agree more with field knowledge of the area under study when using the positive rank as the estimate of the number of endmembers. Endmembers extracted with cPMF agree more across spatial scales as opposed to those with MaxD or SMAACC which agree more between themselves for a given spatial scale. It is also interesting to point out that cPMF with 3 endmembers at 1 meter resolution identifies a distribution more correlated with the broad classes present in the picture: grass, tall/low vegetation, roads.

For the Enrique reef area, the presence of pure pixels is clear, and we can see similar results for the three methods when the rank estimation is being used as a number of endmembers. The positive rank estimation does not give us an improvement of the results in any case for this image.

Unmixing results depend on the assumption made in the development of the unmixing algorithms. cPMF produces better results than PMF. However results with PMF also agree well with our knowledge of the area for the case of Guanica experiment.

The SMACC and MaxD are fast algorithm but need presence of pure pixels in the image, cPMF is significantly slower than the other two algorithms but can obtain virtual endmembers of the image.

Estimating the number of endmembers is a hard problem and we have identified that this number has a high impact in the results of the endmembers extraction algorithms and abundance estimators.

For good performance of any algorithm used, it is important to verify that the image has an appropriate atmospheric correction.

5.2 Future Work

Further work is needed in estimating of number of endmembers. This is very important to continue because of the impact that it has over the performance of the unmixing algorithms is very significant.

It would be interesting compare the results of the endmembers extraction algorithm based on geometrical approaches with algorithms based on statistical principles.

We need to work in the improvement of the atmospheric correction of the Enrique reef images.

REFERENCES

- [1] John A. Richards and Xiuping Jia. *Remote Sensing Digital Image Analysis*. Springer-Verlag Berlin Heidelberg, 2006.
- [2] Galileo Group Inc. http://www.galileo-gp.com/aisa_eagle.html. Copyright. All rights reserved., 2007.
- [3] Specim. <http://www.specim.fi/>. Spectral Imaging Ltd. - All Rights Reserved, 2007.
- [4] NASA. <http://aviris.jpl.nasa.gov/html/aviris.overview.html>. Jet Propulsion Laboratory - California Institute of Technology, Last Updated: October 6, 2008.
- [5] NASA. <http://hyspirc.jpl.nasa.gov/>. ALL RIGHTS RESERVED Copyright California Institute of Technology U.S. Government Sponsorship Acknowledged under NAS7-1260 CL 08-2417 Plone and the Plone logo are registered trademarks of the Plone Foundation. Distributed under the GNU GPL license., 2010.
- [6] N. Keshava and J.F. Mustard. Spectral unmixing. *Signal Processing Magazine, IEEE*, 19(1):44–57, jan 2002.
- [7] A. Lucas, R. Mitchell and P. Bunting. *Hyperspectral data for assessing carbon dynamics and biodiversity of forest*, chapter Chapter 3, Hyperspectral Remote Sensing of Tropical and Sub-Tropical Forests. CRC Press, Taylor and Francis, Boca Raton, FL., 2008.
- [8] S. Jia and Y. Qian. Constrained nonnegative matrix factorization for hyperspectral unmixing. *Geoscience and Remote Sensing, IEEE Transactions on*, 47(1):161–173, jan. 2009.
- [9] Yahya M. Masalmah and Miguel Vélez-Reyes. A full algorithm to compute the constrained positive matrix factorization and its application in unsupervised

- unmixing of hyperspectral imagery. In Sylvia S. Shen and Paul E. Lewis, editors, *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIV*, volume 6966, page 69661C. SPIE, 2008.
- [10] M. Velez-Reyes and S. Rosario. Solving abundance estimation in hyperspectral unmixing as a least distance problem. In *Geoscience and Remote Sensing Symposium, 2004. IGARSS '04. Proceedings. 2004 IEEE International*, volume 5, pages 3276 –3278 vol.5, sept. 2004.
- [11] Chein-I Chang and Qian Du. Estimation of number of spectrally distinct signal sources in hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 42(3):608 – 619, march 2004.
- [12] Peter Bajorski. Does virtual dimensionality work in hyperspectral images? In Sylvia S. Shen and Paul E. Lewis, editors, *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XV*, volume 7334, page 73341J. SPIE, 2009.
- [13] Raymond B. Cattell. The scree test for the number of factors. *Multivariate Behavioral Research*, 1(2):245–276, 1966.
- [14] Monica Chagoyen, Pedro Carmona-Saez, Concha Gil, José María Carazo, and Alberto D. Pascual-Montano. A literature-based similarity metric for biological processes. *BMC Bioinformatics*, 7:363, 2006.
- [15] A. Umana and M. Vlez-Reyes. Determining the dimensionality of hyperspectral imagery for unsupervised band selection. In *Proceedings of SPIE*, volume 5093, 2003.
- [16] I. D. Gerg. An evaluation of three endmember extraction algorithms: Atgp, ica-eea, and vca. Master of science thesis, The Pennsylvania State University, 2008.
- [17] Nicolas Dobigeon and Veronique Achard. Performance comparison of geometric and statistical methods for endmembers extraction in hyperspectral imagery.

- In Lorenzo Bruzzone, editor, *Image and Signal Processing for Remote Sensing XI*, volume 5982, page 13. SPIE, 2005.
- [18] E. J. Ientilucci Bajorski, P. and J. R. Schott. Comparison of basis-vector selection methods for target and background subspaces as applied to subpixel target detection. In 5425, editor, *Proceedings of SPIE*, 2004.
- [19] Michael E. Winter. N-findr: an algorithm for fast autonomous spectral end-member determination in hyperspectral data. In Michael R. Descour and Sylvia S. Shen, editors, *Imaging Spectrometry V*, volume 3753, pages 266–275. SPIE, 1999.
- [20] K. Lee. *A Subpixel Scale Target Detection Algorithm for Hyperspectral Imagery*. PhD thesis, Rochester Institute of Technology, Center for Imaging Science, 2003.
- [21] J.M.P. Nascimento and J.M.B. Dias. Vertex component analysis: a fast algorithm to unmix hyperspectral data. *Geoscience and Remote Sensing, IEEE Transactions on*, 43(4):898 – 910, april 2005.
- [22] Qian Du, Nareenart Raksuntorn, Nicolas H. Younan, and Roger L. King. Variants of n-findr algorithm for endmember extraction. In Lorenzo Bruzzone, Claudia Notarnicola, and Francesco Posa, editors, *Image and Signal Processing for Remote Sensing XIV*, volume 7109. SPIE, 2008.
- [23] J.W. Boardman. Geometric mixture analysis of imaging spectrometry data. In *Geoscience and Remote Sensing Symposium, 1994. IGARSS '94. Surface and Atmospheric Remote Sensing: Technologies, Data Analysis and Interpretation., International*, volume 4, pages 2369 –2371 vol.4, aug 1994.
- [24] Ella Bingham, Aapo Hyvriinen, and Aapo Hyvriinen. A fast fixed-point algorithm for independent component analysis of complex valued signals. *International Journal of Neural Systems*, 10:1–8, 2000.

- [25] JAMES V. Stone. *Encyclopedia of Statistics in Behavioral Science*, volume Independent Component Analysis, Volume 2. John Wiley & Sons, Ltd, Chichester, 2005.
- [26] John H. Gruninger, Anthony J. Ratkowski, and Michael L. Hoke. The sequential maximum angle convex cone (smacc) endmember model. In Sylvia S. Shen and Paul E. Lewis, editors, *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery X*, volume 5425, pages 1–14. SPIE, 2004.
- [27] Ngoc-Diep Ho. *Nonnegative Matrix Factorization Algorithms And Applications*. PhD thesis, Universit Catholique De Louvain, cole Polytechnique De Louvain, Dpartement Dingnierie Mathmatique, 2008.
- [28] Pentti Paatero and Unto Tapper. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, 5(2):111–126, 1994.
- [29] Daniel D. Lee and H. Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, October 1999.
- [30] Pentti Paatero. Least squares formulation of robust non-negative factor analysis. *Chemometrics and Intelligent Laboratory Systems*, 37(1):23 – 35, 1997.
- [31] Daniel D. Lee and H. Sebastian Seung. Algorithms for non-negative matrix factorization. In Todd K. Leen, Thomas G. Dietterich, and Volker Tresp, editors, *Advances in Neural Information Processing Systems 13*, pages 556–562, The MIT Press, 55 Hayward Street Cambridge, MA 02142-1493 USA, April 2001. MIT Press.
- [32] A. Pascual-Montano, J.M. Carazo, K. Kochi, D. Lehmann, and R.D. Pascual-Marqui. Nonsmooth nonnegative matrix factorization (nsnmf). *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(3):403 –415, march 2006.

- [33] Patrik O. Hoyer. Non-negative matrix factorization with sparseness constraints. *Journal of Machine Learning Research* 5, Aug 2004.
- [34] Seung-Jean Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky. An interior-point method for large-scale l_1 -regularized least squares. *Selected Topics in Signal Processing, IEEE Journal of*, 1(4):606–617, dec. 2007.
- [35] Robert A. Schowengerdt. *Remote Sensing: Models and Methods for Image Processing*. Elsevier Inc., 2007.
- [36] J.M. van den Hof and J.H. van Schuppen. Positive matrix factorization via extremal polyhedral cones. *Linear Algebra and its Applications*, 293:171–186, 1999.
- [37] J. Goodman. 2007 puerto rico hyperspectral mission: Image acquisition and field data collection, november 28 - december 21, 2007. Technical report, University of Puerto Rico at Mayagez, 2009.
- [38] Carlos Rivera-Borrero, Samuel Rosario, Shawn Hunt, Carmen Zayas, Adrienne Mundorf, and Suhaily Cardona. Ground truth data collection for unmixing algorithm evaluation. In Sylvia S. Shen and Paul E. Lewis, editors, *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIV*, volume 6966, page 69661D. SPIE, 2008.