



Contributions to Hyperspectral Unmixing

Sina Nakhostin

► To cite this version:

Sina Nakhostin. Contributions to Hyperspectral Unmixing. Signal and Image processing. Université de Bretagne Sud, 2017. English. NNT: 2017LORIS477 . tel-01867812

HAL Id: tel-01867812

<https://theses.hal.science/tel-01867812>

Submitted on 4 Sep 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE / UNIVERSITÉ DE BRETAGNE SUD

UFR Sciences et Sciences de l'Ingénieur

sous le sceau de l'Université Européenne de Bretagne

Pour obtenir le grade de :

DOCTEUR DE L'UNIVERSITÉ DE BRETAGNE SUD

Mention : Informatique

École Doctorale MATHSTIC

présentée par

Sina Nakhostin

**IRISA Institut de Recherche en Informatique et
Systèmes Aléatoires**

Contributions to Hyperspectral Unmixing

Thèse soutenue le 13-12-2017,
devant la commission d'examen composée de :

.
/ Président

Prof. Yannick Deville

Université Toulouse III - Paul Sabatier, IRAP / Rapporteur

Prof. Laurence Hubert-Moy

Université Rennes2, laboratoire LETG RENNES COSTEL / Rapporteur

Prof. Sébastien Lefèvre

Université de Bretagne-Sud, IRISA / Examineur

Dr. Harold Clenet

Ecole d'Ingénieurs de PURPAN, INP Toulouse / Examineur

Dr. Nicolas Courty

Université de Bretagne-Sud, Maître de Conférences - HDR / Directeur de thèse

Dr. Thomas Corpetti

Université Rennes2, Directeur de Recherche CNRS, LETG / Co-directeur de thèse

"For the search in knowing what we know"

List of Figures	12
List of Tables	13
1 Spectral Analysis	15
1.1 Remote sensing and its applications	15
1.1.1 Domains of application	15
HSI for Environmental monitoring	17
HSI for medical analysis	18
HSI in ART	19
HSI for Archeology	20
HSI in Astrophysics	20
Ground-based HSI	22
1.1.2 Multispectral vs Hyperspectral: literal vs non-literal	22
HSI from the view of Pigeon-Hole Principle	24
1.1.3 Spectral Unmixing	24
Endmember Variability	25
Noise and Anomaly	26
Contributions of this PhD	27
I State-of-the-art	31
2 Hyperspectral Unmixing	35
2.1 Hyperspectral Data: Mixing & Unmixing	35
2.1.1 (non)Linear mixtures & (Un)supervised methods	36
2.1.2 Estimating endmembers and abundances	38
2.1.3 Endmember variability & Overcomplete dictionaries	39
2.2 Distance Measures for HSI	40
2.2.1 Signature vector-based measures	40
Euclidean Distance (ED)	41
Spectral Angle Mapper (SAM)	41

Orthogonal Projection Divergence (OPD)	41
Spectral Information Divergence (SID)	41
Mixed SID – SAM	41
2.2.2 Correlation-weighted measures	42
HS measures weighted by <i>a priori</i> correlation	42
HS measures weighted by <i>a posteriori</i> correlation	42
2.3 Inversion	43
2.4 Anomaly detection	44

II Kernel methods for Unmixing 47

3 Joint Anomaly detection and spectral Unmixing	51
3.1 Introduction	52
3.1.1 Summary, goals of the chapter	53
3.2 SAGA+ : simplex volume maximization and anomaly detection . . .	54
3.2.1 SAGA principles: Exact Simplex Volume Maximization for endmember extraction	54
3.2.2 Anomaly Detection	58
3.2.3 Sparse Projections onto the Simplex	61
3.2.4 Overall process	62
3.3 Evaluation	64
3.3.1 Experimental setups	65
Evaluation criteria	65
Synthetic data	66
Real data	68
Parameters setting for SAGA and SAGA+	70
3.3.2 Results on synthetic data	71
3.3.3 On the application of Overcomplete Dictionary	76
3.3.4 Results on real data: 4-Vesta Asteroid	78
3.4 Conclusion	82

III	Optimal Transport for Unmixing	83
4	Supervised Planetary Unmixing with Optimal Transport	87
4.1	Introduction	87
4.2	Unmixing using priors and iterative Bregman projections	88
4.2.1	Optimal transport of spectra	88
4.2.2	Entropy regularized Optimal Transport and Bregman projections	89
4.2.3	Unmixing model	90
4.3	Experiments and Analysis	91
4.3.1	4-Vesta properties and dataset	92
4.3.2	Experimental setup	92
4.3.3	Unmixing results and discussion	92
4.4	Discussion/Conclusion	94
IV	Applications of Spectral Unmixing in Urban Context	95
5	CATUT : Spectral Unmixing for Urban surface Temperature	99
5.1	DESIREX 2008 project	99
5.1.1	Description of the urban area	99
5.1.2	AIRBORNE DATA ACQUISITION	100
5.1.3	SPECTRAL CHARACTERIZATION OF URBAN SURFACES	103
5.2	On the valid choice of Endmember dictionary	106
5.2.1	ASD Preprocessing	107
	ASD cut-off	107
	ASD downsampling	108
5.2.2	Expert Endmember Extraction	109
5.2.3	Endmember Extraction with E-SiVM	111
5.3	Unmixing Experiments on Desirex	112
5.3.1	Experiment conditions	112
	K-means clustering for endmember grouping and estimation of the level of sparsity	113
5.3.2	Abundance Estimation with OT-Unmixing	115
5.3.3	Abundance Estimation with SLSQP	118

5.3.4	Endmember Extraction and Abundance Estimation with SAGA+	120
5.4	Conclusion of the chapter	122
6	Conclusion and future work	123
V	Appendices	127
7	Appendix : SAGA+	i
7.1	Spectral Signatures	i
7.2	SOPE maps	ii
7.3	Parameter Setting	iii
7.4	Execution Time	iii
7.5	Abundance Maps for SAGA	v

List of Figures

1.1	Left: Active sensor Right: Passive sensor (Courtesy of NASA)	15
1.2	Electromagnetic wavelength: depending on their presumed utility, different spectrometers are being designed to capture specific portions of the electromagnetic field. In many applications therefore, hyperspectral data combines more than one captor.	16
1.3	Chronological process of water shrinkage of lake Urmia. With a new composite of multi-spectral high resolution satellite observations, it is shown that the area of the Iranian lake has decreased by around 88% in the past decades, far more than previously reported $\sim 25\%$ to 50%. (Courtesy of Kaveh Madani)	16
1.4	Paris area mean temperature map in 2003. (Courtesy of VITO, PLANETEK)	17
1.5	Generated maps based on two different hyperspectral images captured from the same subject with similar tissues depicted with similar colors.	18
1.6	Left: Original painting. Right: Endmember map of the materials used for painting. The red corresponds with egg yolk binder, the blue is consistent with a glue binder, and the green maps the areas of Azurite in a glue binder.	19
1.7	Using an iterative linear spectral unmixing model the global OMEGA dataset has been analyzed. The goal was to retrieve global atlas of the main minerals contributing to the signal recorded by OMEGA on the very top surface of Mars. Top : false color composite emphasizing water ice (in blue). Bottom : false color composite with Red=Olivine, Green= Low-Calcium pyroxene, Blue = High-Calcium pyroxene. (Courtesy of ESA, OMEGA, LPGNantes)	21
1.8	Left : MUSE instrument (property of ESO) installed at Nasmyth focus of UT4. Right : An abstract illustration of image registration by MUSE across the spectral range.	22
1.9	Hyperspectral image (Left) vs. Multipectral image (Right). Hyperspectral Images provide an unprecedented amount of spectral information which requires new approaches of image analysis in order to exploit this amount of information. (Illustration courtesy of [1])	23
1.10	Each pixel can be the mixing result of different endmembers. In this abstract scenario three different materials are being mixed according to the linear mixing model (LMM).	24
1.11	Representation of the northern hemisphere of 4-Vesta asteroid together with some nominal spectral signatures and one signature which is corrupted by anomalies.	26

1.12	DESIREX 2008 Madrid classification map.	28
2.1	In the ideal case, where there is no effect of noise, also with the assumption of pure endmembers being present in the data, the abundance coefficients as the result of inversion part of linear unmixing will lay within a convex polygon (n-1 simplex) whose vertexes are consist of the endmembers. Illustration courtesy of Dobigeon et al. . .	36
2.2	From left to right: linear mixing model(LMM), intimate mixing model (IMM) and bilinear mixing model (BMM). Courtesy of [2].	37
3.1	Overcomplete endmember dictionary for non-linear unmixing	55
3.2	Illustration of difficulties in presence of outliers. First Row: endmember extraction using SAGA without anomalies: in this situation extracted endmembers are a reliable representation of the manifold ; Second Row: endmember extraction using SAGA in presence of anomalies: in this situation extracted endmembers (in red) do not correspond to real borders of the data manifold ; Third Row: endmember extraction and anomaly detection In presence of anomalies using SAGA+: here extracted endmembers (in red) by are in accordance with borders of the manifold embedding data while the anomalies (in cyan) are properly selected.	59
3.3	Evolution of SOPE (Sum Of Projection Errors) when all data in green are represented through a single endmember uniformly represented on a meshgrid. One observes a decrease in SOPE when candidates get closer to the nominal distribution.	60
3.4	Illustration of synthetic data. (a): Some endmembers (top) and the anomaly spectrum (bottom) used to generate the datasets. (b): A 2D Slice of two components of a synthetic dataset of 1010 samples generated based on LMM from the 3 nominal endmembers and 10 samples issued from the single anomaly signature. The geometrical location of the anomalies in the 2D plan makes them visually separable. (c): Similar visualization but mixtures are generated on the basis of BMM. (d): Similar visualization but mixtures are generated on the basis of HCM.	67
3.5	4-VESTA mosaic composed of (842×327) pixels, with 383 spectral bands covering the range $0.55\mu m$ to $2.47\mu m$. Examples of spectra (located on the mosaic with letters) are reported with (a) a normal spectrum, (b) discontinuity at the two detectors boundary and (c) a spectrum at the boundary of a shadow region. The gray wavelength range is where inconsistency are usually observed.	69

3.6	Endmember estimation Averages of SAM values for SAGA, G-SiVM, VCA and N-Finder in function of the number of endmembers. Each plot represents the results for three mixing models.	71
3.7	Endmember estimation Averages of SAM values for SAGA+, G-SiVM, VCA and N-Finder in function of the number of endmembers. Data contains anomalies generated based on 3 random signatures. Each plot represents the results for three mixing models.	72
3.8	Results in presence of anomalies using SAGA+. (a): Representation of the dataset; (b) original endmembers and anomalies (c) estimated endmembers and anomalies with SAGA+	73
3.9	Evaluation of SAGA+: Kappa values as the function of Concentration of the anomalies. The red lines represent the threshold where the Kappa value starts to get over 0.8 which is equal to the concentration level $\alpha = 15$	74
3.10	Effect of Anomaly concentration: In each plot a 2D slice of a dataset generated based on the same endmember signatures and anomaly signatures is illustrated. For construction of anomalies different levels of concentration is set : $\alpha = [5, 15, 50]$	75
3.11	Results in presence of anomalies using SAGA+. (a): original signatures used for generation of dataset; (b): 2D Slice of dataset together with the estimated prototypes (c): signatures of estimated endmembers and anomalies with SAGA+ (d): the 4 weighted averaged endmembers.	76
3.12	Results on real 4-Vesta asteroid. Top: Vesta-4 mosaic with the spatial position of the extracted endmembers. The red points are detected anomalies and the 10 green points with labels are selected endmember prototypes. Bottom: The respective estimated endmember signatures.	78
3.13	Top: laboratory spectra of mixtures involving olivine and orthopyroxene (spectra from the RELAB library with numbers referring respectively to the olivine and orthopyroxene contents (% olivine / % orthopyroxene). Bottom: comparison with the three extracted endmembers P0, P5 and P9. The shift in absorption center at $2\mu m$ is due to an increase in Ca and/or Fe in the pyroxene, compared to the pure orthopyroxene composition.	80
3.14	RGB composition with $R=P9+P5$, $G=P4+P8+P0$, and $B=P4+P8+P0$. Coloring shows the localization of outcrops enriched in olivine (appearing in red)	80

3.15	Abundance maps and respective spectral signatures extracted by SAGA+ (with anomaly detection procedure). Using an anomaly detection procedure, one observes that signatures whit magnitudes higher than 1 and those with sharp peaks are not selected as endmembers.	81
4.1	Mosaic of northern hemisphere of Vesta-4 asteroid. The window in red (61×61) is used as the test bed.	91
5.1	AHS flight lines for urbane surface measurements	100
5.2	Left: Platform C-212-200 property of INTA (PATERNINA). Right: Airborn hyperspectral scanning sensor.	101
5.3	Illustration of the spectral range and spectral resolution of each band for the first three ports of the AHS DESIREX dataset. One observes that the bands of DESIREX dataset are measured using different portions of the electromagnetic field and with different spectral resolutions.	102
5.4	DESIREX hyperspectral images acquired through one north-south flight over Madrid in 2008. (The image is rotated by 90°)	102
5.5	ASD FieldSpec3 portable, general-purpose Spectroradiometer.	103
5.6	Man-made materials considered for endmember dictionary construction.	104
5.7	Spectral signatures of Endmembers which are registered through an in-situ procedure using the ASD equipment.	105
5.8	Illustration of the "wavelength range" and "wavelength sampling rate" of the in-lab measured ASD endmembers across the electromagnetic field. One can see that the bands of these endmembers have uniformly been sampled across the range $.35 - 2.5(\mu m)$, with the sampling rate of $1(nm)$ which makes them incomparable to the spectral characteristic of the Desirex dataset.	105
5.9	Spectral signatures of a sample of 1000 pixels randomly selected from Desirex dataset versus the original ASD endmembers.	107
5.10	Spectral signatures of a sample of 1000 pixels randomly selected from Desirex dataset are plotted versus the clipped ASD endmembers.	108
5.11	Spectral signatures of a sample of 1000 pixels randomly selected from Desirex dataset versus the downsampled ASD endmembers.	109
5.12	Dictionary of the manually extracted endmembers. Based on expert observations, pixel vectors related to highly distinct materials within the scene have been extracted to be used as a dictionary of endmembers. In order to account for material variability, for each material two representatives endmembers were selected in order to build-up a dictionary of 24 signatures.	110

5.13	Using the exact simplex volume maximization (E-SiVM) procedure a dictionary of 24 atoms with maximum variability among their signature profiles has been generated. One can observe even larger variability among endmember signatures while comparing this dictionary to the one extracted manually.	110
5.14	Dictionary of 24 randomly selected signatures from 601×601 window of Desirex dataset.	111
5.15	Random and ESiVM dictionaries compared to the Expert dictionary. The result of ESiVM shows more fidelity to the Expert dictionary in capturing more varieties among the endmember signatures.	112
5.16	Window of 601×601 (pixels) extracted from the main mosaic on which all tests are performed.	113
5.17	Left: Result of Elbow test on a series of k-means clusterings with increasing number of clusters performed on the endmember dictionary of manually selected atoms from the AHS Desirex dataset. Right: Result of Elbow test on a series of k-means clusterings with increasing number of clusters performed on the endmember dictionary generated by performing E-SiVM procedure on the AHS Desirex dataset. . . .	114
5.18	In order to set cost matrix $C0$ in alignment to the characteristics of the dataset, we build $C0(63 \times 63)$ as the square euclidean distance over the spectral values of the AHS Desirex dataset.	115
5.19	$C1$ cost function for three dictionaries. $C1$ provides <i>a priori</i> information regarding the groups that different signatures of the endmember dictionary belong to.	116
5.20	Clustering of two dictionaries. Groups which considered to share similar semantic relationships in two dictionaries are plotted in the same color.	117
5.21	Abundance maps produced by OT-unmixing and according to the first 3 strategies presented in the first column of table 5.2.	118
5.22	Abundance maps produced by SLSQP algorithm and according to the last 3 strategies presented in the last column of table 5.2.	119
5.23	Map distances calculated between pairs of abundance maps for 6 different strategies.	121
5.24	Abundance maps produced by SAGA+ algorithm and according to the strategies 4,5 and 6 presented in the middle column of table 5.2. . .	122

7.1	Example of spectra used for generating the synthetic datasets issued from the 41 spectral signatures of different Lithologies RELAB. The signatures have re-sampled in order to be comparable to the signature characteristics of 4-Vesta dataset according to the wavelength range and sampling range.	i
7.2	Evolution of SOPE map along the endmember extraction process. Each plot shows the map of SOPE values w.r.t the "next" endmember candidate. In each plot, we take all the cells of 26×26 grid one by one as the next endmember candidate and calculate SOPE. In each plot the only difference is the dimensionality of the simplex that we use to calculate the SOPE values. It is observable that the points laying outside the boundaries of the dataset always produce very high values of SOPE (almost close to 1). By setting a correct threshold for SOPE change then SAGA+ distinguishes the good candidates (points on the boundaries of the dataset manifold) from the anomalies (points outside the boundaries of the dataset manifold)	ii
7.3	Abundance maps and respective spectral signatures extracted by SAGA (no anomaly detection). Without an anomaly detection procedure, one observes that signatures with sharp peaks are among selected endmember signatures.	v

List of Tables

3.1	Kappa statistics for SAGA+ in endmember estimation for the three synthetic datasets	74
3.2	Average SAM errors between the original and estimated abundance coefficients. Best results are reported in boldfont	77
5.1	AHS spectral configuration for DESIREX 2008 surveys. Due to the capability of the visible and near-to-mid infrared range of electromagnetic field in representation of the ground compositions and discrimination among different compositional units, only the signatures of the first three ports (PORT 1, PORT 2A and PORT 2) are used for the current study.	101
5.2	Different strategies introduced in section 5.2 which relates different methods of unmixing considered in this chapter to the available end-member dictionaries.	114
5.3	Number of endmembers grouped into the 5 different clusters based on Expert and ESiVM dictionaries.	116
5.4	Statistics of the pairwise distance calculation over the 5 abundance groups produced as the result of each of 6 different strategies. The pairwise distances of each strategy over which we calculated the average value in the last column of this table are illustrated as a separate upper triangular matrix in Figure 5.23.	119
7.1	Parameter setting for synthetic data generation	iii
7.2	Execution Time (s) of the different algorithms for various endmembers and elements in the dataset	iv

Acknowledgement

The Author would like to thank *Conseil Régional de Bretagne* and *Université de Bretagne sud* for their support of the project ARED-RB OPTIMAP. Also thanks goes to ANR ASTERIX project for their support of this work. Many thanks to the authors in [3] for the available package and codes provided along the survey. The same appreciation goes to the authors in [4, 5, 6] for providing the codes of their algorithms, which made the comparisons possible within this work. At the end Special thanks and gratitude goes to *Nicolas Courty* and *Thomas Corpetti* for their availability and supervision of this thesis. The positive and friendly atmosphere within the *Obelix* team of *IRISA* laboratory made the experience of accomplishing this thesis a true pleasure. Particular thanks goes to Sébastien Lefèvre for making it possible.

Spectral Analysis

1.1 Remote sensing and its applications

Hyperspectral (HS) remote sensing is the scientific field which is emerged as the synergy of airborne/satellite imagery and spectroscopy. Hyperspectral sensors are categorized as *passive sensors*. Unlike *active sensors* who perceive and measure the reflection of their own emitted signal, passive sensors perceive the reflected signal of another source. In case of hyperspectral remote sensing, this source is usually the natural light emitted from the sun and reflected by the surface or atmosphere of the astronomical bodies. Figure 1.1 depicts this difference.

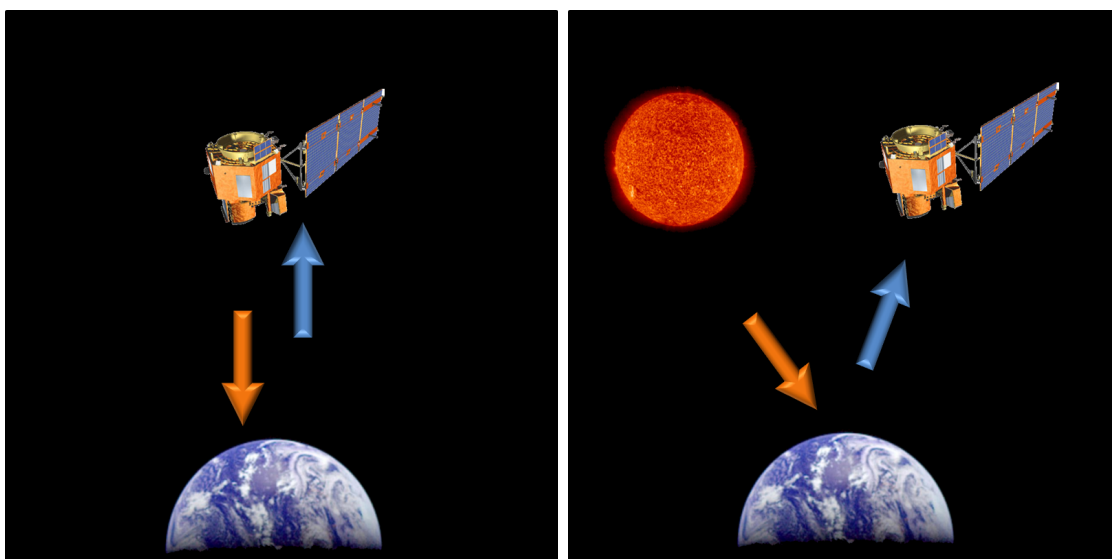


Figure 1.1: Left: Active sensor Right: Passive sensor (Courtesy of NASA)

1.1.1 Domains of application

The range of applications of these sensors is considerably large and concerns for example astronomy, geoscience, agriculture, archeology, surveillance and environ-

ment monitoring where the HS acquisition senses each object not through the three RGB wavelength bands but through hundreds of bands covering large portions of the electromagnetic range, which is depicted in Figure 1.2. Another domain which in recent years has been highly benefited from the advances of hyperspectral images is the medical imaging. In the following a brief look is taken on the diverse domains where this technology can be used.

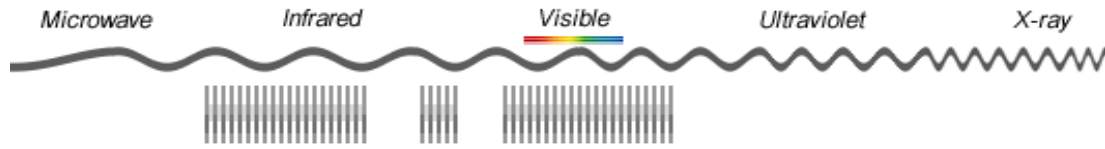


Figure 1.2: Electromagnetic wavelength: depending on their presumed utility, different spectrometers are being designed to capture specific portions of the electromagnetic field. In many applications therefore, hyperspectral data combines more than one captor.

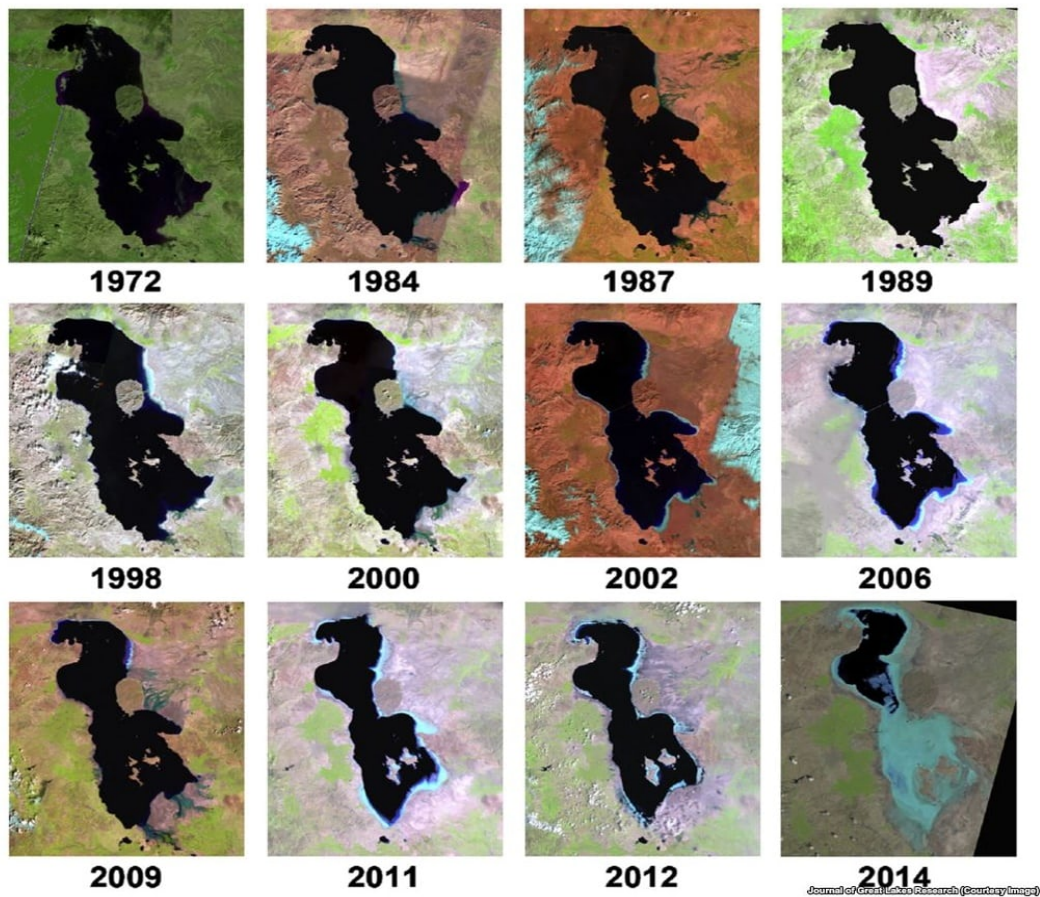


Figure 1.3: Chronological process of water shrinkage of lake Urmia. With a new composite of multi-spectral high resolution satellite observations, it is shown that the area of the Iranian lake has decreased by around 88% in the past decades, far more than previously reported $\sim 25\%$ to 50% . (Courtesy of Kaveh Madani)

HSI for Environmental monitoring

Desertification, land degradation in arid, semi-arid, and dry sub-humid regions, is emerging as a global environmental dilemma [7]. This has called for accurate assessment of the status, change, and trend of desertification for controlling this global threat. Identification, mapping and monitoring of the salt-affected areas have been exploiting remote sensing data during the recent decades[8, 9]. In Figure 1.3 one can follow the chronological process of the shrinkage in the size of a salt lake.

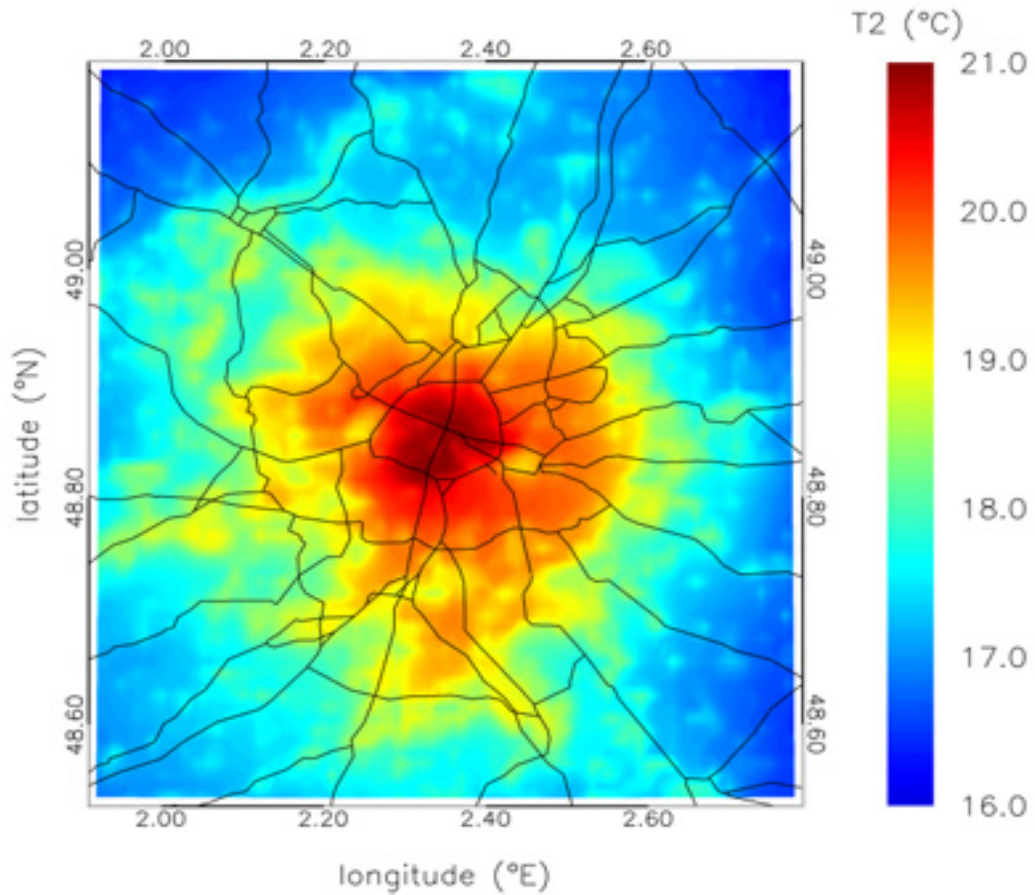


Figure 1.4: Paris area mean temperature map in 2003. (Courtesy of VITO, PLANETEK)

As other HSI applications in environmental monitoring one can name oil spill detection [10], water pollution monitoring [11] and estimation of land surface temperature (LST) [12]. In Figure 1.4 a heat map generated according the satellite remote sensing images is depicted. In chapter 5, a number of algorithms developed during this PhD work are going to be used for generation of urban surface abundance maps which by themselves help to estimate LST maps of urban areas.

HSI for medical analysis

During the recent years hyperspectral imagery has increasingly been used for the early detection of different sorts of cancer using noninvasive or minimally invasive techniques [13]. In order to show the riveting role of HSI in shaping the future of medicine one can refer to a recent European project HELICoiD [14]. The aim of HELICoiD was to demonstrate the use of hyperspectral imaging for real-time identification of tumor margins during surgery, helping the surgeon to extract the entire tumor and to spare as much of the healthy tissue as possible. In [15] and [16] for example new manifold embedding techniques are proposed for delineation and segmentation of the tumor effected tissues of brain, the result of which is presented in Figure 1.5.

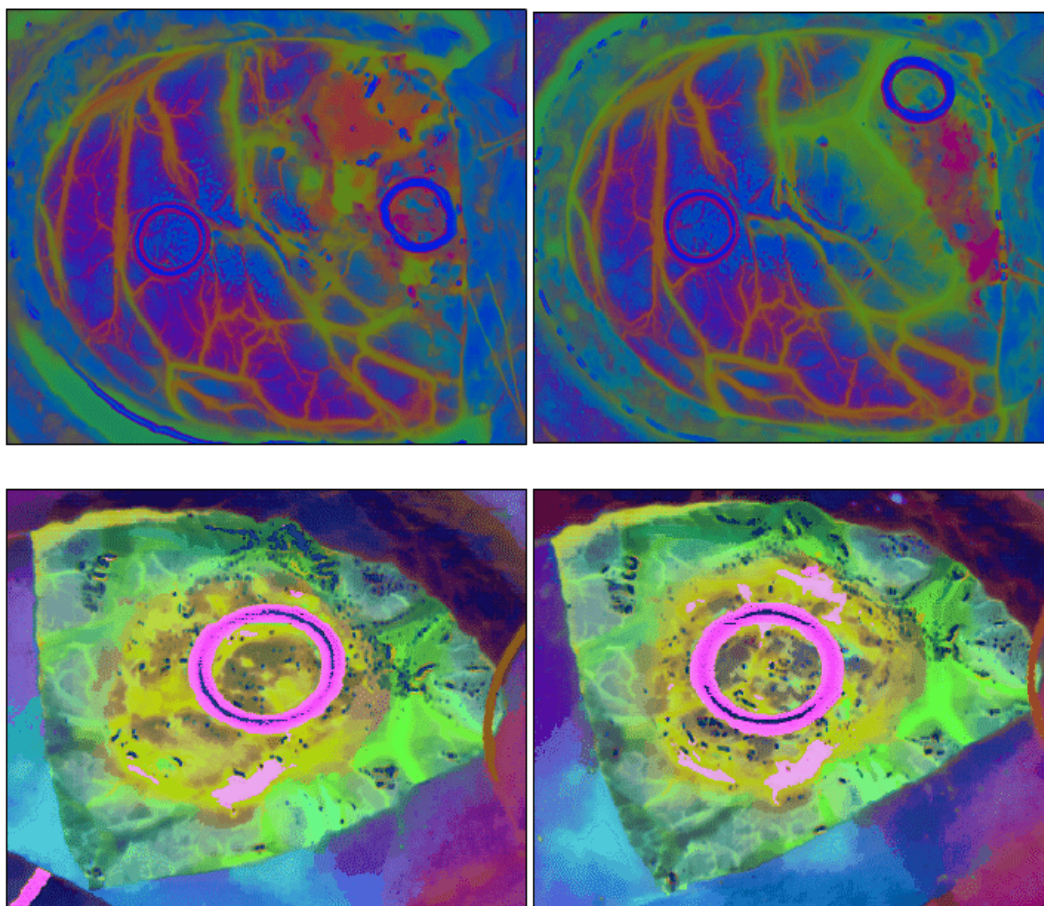


Figure 1.5: Generated maps based on two different hyperspectral images captured from the same subject with similar tissues depicted with similar colors.

By exploiting the spectral depth in HS images of healthy and cancerous tissue, machine learning algorithms are being trained to detect differences in their spectral signature. The learned model then being applied to identify cancerous and healthy tissue in hyperspectral images acquired during surgery. It can be imagined that in the future, this information could be available for surgeon in real time by overlaying

conventional images with a color-map that indicates the likelihood of that particular area being cancerous.

HSI in ART

The need to conserve historical artworks is not new. Many of the methods of art preservation which have developed during 19th century, have been seen to actually contribute to, rather than prevent, the deterioration of paintings. Today it is hoped that by using more advance scientific methods and technology, analysis of paintings, identification of conservation issues and detection of additions and forgeries become faster and more accurate. In general major problems related to art and painting preservation are listed as :

- Pigment Identification and Mapping
- Examination of Underdrawings
- Analysis of 3-D Objects
- Recognition of Organic Materials
- Forgery Detection



Figure 1.6: Left: Original painting. **Right:** Endmember map of the materials used for painting. The red corresponds with egg yolk binder, the blue is consistent with a glue binder, and the green maps the areas of Azurite in a glue binder.

Traditional scientific methods include chemical analysis of paint samples to determine the molecular nature of the pigments and varnishes. X-ray technology is one often utilized method for observation of beneath surface layers and discovering the intentions and processes of the artist. Despite their relative success, many of the methods used either give limited information, are invasive to some degree, or both. HSI turns to be a useful method in material identification and mapping on works of art and paintings [17, 18, 19]. By using hyperspectral imaging systems it has become possible to produce false color infrared images that can show restored zones of the painting and therefore help to guide the choice for the pigment analysis of significant zones of the original paintings. Also in a recent work [20] it showed possible to identify which pigment binders had been used in specific paintings (See Figure 1.6 for an illustration).

HSI for Archeology

Another domain of application for hyperspectral images is Archeology. In a very similar fashion as to the *Underdrawings* of the old paintings, historical sites and landscapes might be palimpsests, or layered, while mixing of past and present occupation layers can obscure imagery analysis from space [21]. Archeological sites, ancient road systems etc. therefore could be well buried under tons of thousands of years of silt deposit or even under the current existing cities and towns. Remote sensing hyperspectral data could be used along other ground-based equipments in order to draw more accurate maps of old archaeological landscapes [22]. In [23] an old road system in northern Mesopotamian has been partially detected using the remote sensing imagery.

HSI in Astrophysics

From planetary surface analysis [24] to atmospheric studies to detection of galaxies, hyperspectral images are being widely used in different astrophysical studies. In [25] and [26] the hyperspectral data of *Observatoire pour la Minéralogie, l'eau, les glaces et l'activité* (OMEGA) which is mount on board of MARS EXPRESS and has been monitoring the surface of the planet Mars since 2003 is analyzed. Using a Multiple-Endmember Linear Spectral Unmixing Model (MELSUM), the existence of different surface materials including CO_2 and H_2O has been studied. For this a reference mineral library composed of 24 different mineral fingerprints (endmember signatures) were used. In Figure 1.7 The global atlas maps produced by this model are illustrated. As for the endmember selection process a systematic exploration of whole set of combinations of four components of the endmember signatures were deployed by manually excluding the maps whit negative values. Although the utilization of this linear method was positively evaluated, it is believed that more subtleties might be captured in case non-linear models of mixing were considered.

In [27] independent component analysis (ICA) and Bayesian Positive Source Separation (BPSS) methods were exploited instead to fuse the spatial and spectral information of the same data and enhance the process of endmember estimation and abundance map estimation. The idea is to do a rough classification of pixels based on ICA which provides a reduced set of signatures, which by itself is being used by BPSS to do endmember estimation and abundance estimation.

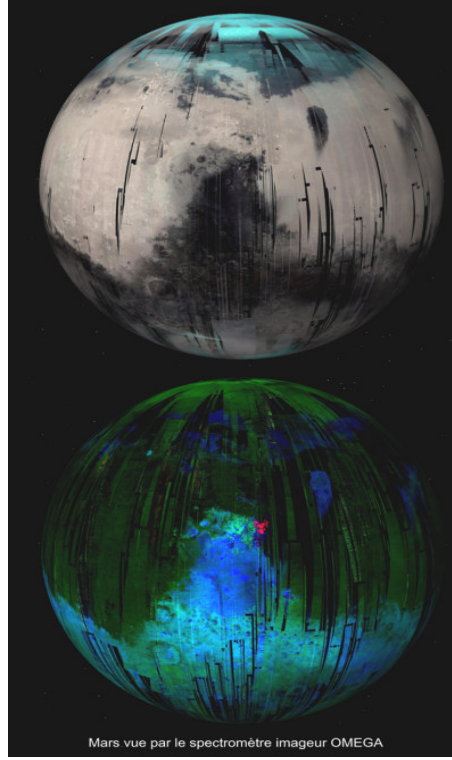


Figure 1.7: Using an iterative linear spectral unmixing model the global OMEGA dataset has been analyzed. The goal was to retrieve global atlas of the main minerals contributing to the signal recorded by OMEGA on the very top surface of Mars. **Top :** false color composite emphasizing water ice (in blue). **Bottom :** false color composite with Red=Olivine, Green= Low-Calcium pyroxene, Blue = High-Calcium pyroxene. (Courtesy of ESA, OMEGA, LPGNantes)

In another study [28] OMEGA data is being used to come up with a processing methodology for converting OMEGA data from radiance-at-sensor to surface reflectance, performing an atmospheric correction and filtering the systematic and random noise. Also the presence of iron-rich clays on some areas is reported within this work. As another example of the applications of HS remote sensing for planetary data analysis one can refer to NASA’s DAWN satellite whose mission is to orbit around two important protoplanets of the asteroid belt (4-VESTA and CERES), for which it utilizes a sensor composed of 432 bands. Through these bands each material acquires a fingerprint which is called *spectral signature* of that material. In chapters 3 and 4 applications of the methods developed during this thesis will be presented on a hyperspectral image from the 4-VESTA asteroid.

Ground-based HSI

The Multi Unit Spectroscopic Explorer (MUSE) instrument installed at the Very Large Telescope (VLT) of the European Southern Observatory (ESO) is a ground based integral-field spectrograph which has combined very high spectral and spatial resolution characteristics in an unprecedented way into one instrument. In Figure 1.8 an illustration of the instrument together with an abstract data cube generated by MUSE is visible. Because of this unique property MUSE is mainly meant to be used for extra-galactic studies including *globular clusters*, *planetary nebulae* and *Lyman-alpha emitters* also *supermassive black holes*. [29] and [30] are among the first efforts in utilizing state-of-the-art spectral analysis procedures in detection and unmixing of MUSE data.

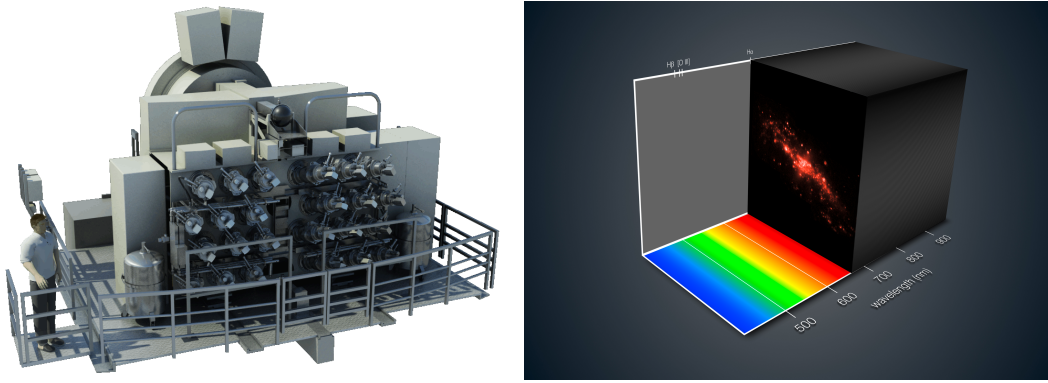


Figure 1.8: Left : MUSE instrument (property of ESO) installed at Nasmyth focus of UT4. Right : An abstract illustration of image registration by MUSE across the spectral range.

In [31] a clustering analysis of a sample of 238 Lyman α emitters is presented at redshift $3 \leq z \leq 6$ from the MUSE-Wide survey. In [32] a new Bayesian method optimized for detection of faint galaxies in MUSE deep fields is presented. The list of publications as the result of research on MUSE data is considerably growing since its down in 2014. One can access an updating list of research related to MUSE in a dedicated web page ¹.

1.1.2 Multispectral vs Hyperspectral: literal vs non-literal

Hyperspectral remote sensors can register images in the format of 3D data cubes. These data cubes are being characterized based on their spatial resolution, spectral resolution and their signal-to-noise ratio. The spatial resolution is the measure of the area being represented through one pixel (thus providing so-called literal information) and spectral resolution is the width of the electromagnetic field being represented through one band of the cube (therefore providing so-called non-literal

¹<http://muse-vlt.eu/science/publications/>

information). The signal-to-noise ratio (SNR) reflects spectrometer's radiance response and relates to its acquired image quality.

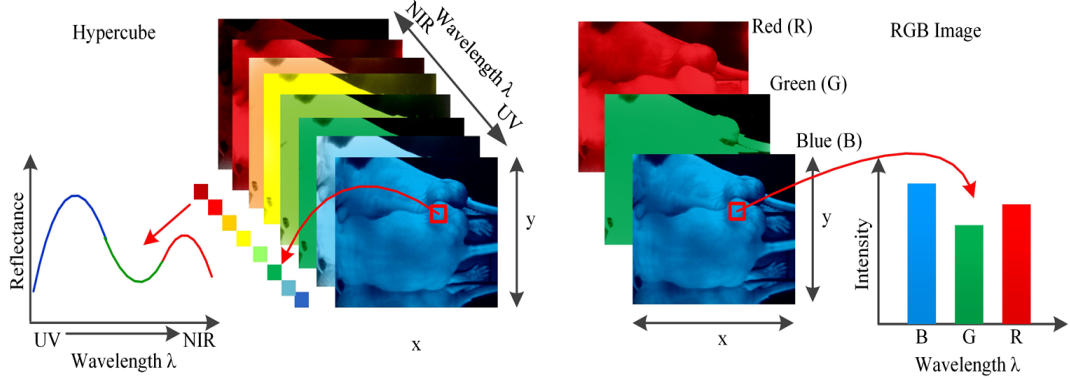


Figure 1.9: Hyperspectral image (Left) vs. Multipectral image (Right). Hyperspectral Images provide an unprecedented amount of spectral information which requires new approaches of image analysis in order to exploit this amount of information. (Illustration courtesy of [1])

The large number of spectral bands in hyperspectral images (HSI) differentiates them from *multispectral* images (MSI) where the spectral representation of pixels never exceeds a few bands. This is illustrated in Figure 1.9. Therefore both the semantics and methodologies associated with HS data are specific and different than the ones used with MSI. In the case of MSI, because of the very low spectral resolution, in practice no or very limited information w.r.t the existing subpixel materials is retrievable, therefore, except a few interesting attempts for MSI subpixel classification ([33] and [34]), processing of these types of data has been historically focused mainly on the spatial information and spatial correlation among pixels. Spatial pattern classification is therefore one of the most studied problems that could be addressed using MSI and correlation-based methods which have been devised upon them. These are statistical inference methods where the *independent and identically distributed (i.i.d)* assumption does not hold upon pixels therefore, by somehow defining a measure of spatial distance between the pixels, one can assign n total pixels of the image to ℓ different classes. The measure of closeness is usually defined through a weighting matrix which then operates on the pixels in the vicinity of the current pixel under study. Thanks to this characteristic, correlation-based methods [35] have been extensively used for different spatial classification problems for MSI analysis.

With hyperspectral images, thanks to their rich spectral information, it is possible to assess the composition of the various materials embedded in a single pixel. Problems including *unmixing associated with anomaly detection* and *endmember extraction* are within this category.

HSI from the view of Pigeon-Hole Principle

In the pigeon-hole principle we have that for D places and ℓ objects, where $0 > D > \ell$, there will be at least one place with more than one occupying object. It is possible to see the relationship between different targets/endmembers and spectral bands of HSI through the pigeon-hole principle. Assume different HSI bands as the places with total number of D , and ℓ different targets of interest as the objects. In case of $D > \ell$, one band can be used to detect one target. Since in HSI there are hundreds of contiguous bands available, we can therefore distinguish hundreds of spectrally distinguished targets. For this however, three requirements must meet: First) The number of spectral bands should be larger than the number of endmembers ($D > \ell$). Second) Number of interesting targets/endmembers (ℓ) should be known and Third) No two distinguished endmember signatures to be accommodated by a single band.

1.1.3 Spectral Unmixing

The current thesis focuses mainly on one specific problem in the field of hyperspectral remote sensing called "Spectral Unmixing". Spectral Unmixing has been an active area of research during the last years and found its application in domains including but not limited to remote sensing, audio signal processing and chemistry. In audio signal processing the term "blind source separation" and in the chemical sciences "chemical agent detection" is used instead to refer to the same problem. Despite their very high spectral resolution, hyperspectral images are indeed known to be of relatively low spatial resolution. This low resolution is a relative notion and is due to technological limitations of the HSI captors, also is negatively proportionates with the distance of the captor from the scene of view.

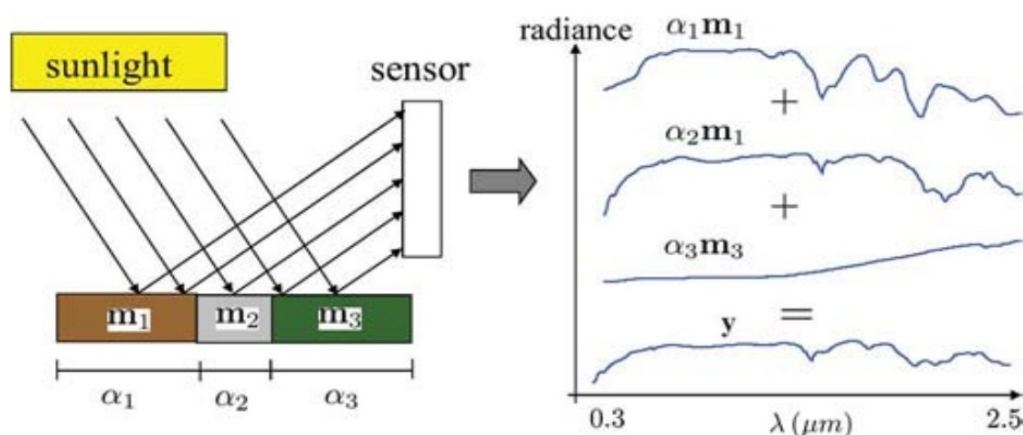


Figure 1.10: Each pixel can be the mixing result of different endmembers. In this abstract scenario three different materials are being mixed according to the linear mixing model (LMM).

As a consequence the values of pixels within an image are likely to be mixtures of different materials in the scene. In Figure 1.10 one can observe the process of mixing of three materials within a sensor (contributing to one pixel of an image) acquisition. Hyperspectral Unmixing then can be defined as an inverse procedure that consists of identifying in each pixel the amount of pure elements that contributed to the pixel's mixture. The total number of pure elements (also called endmembers) and the number of pure elements included in one pixel are then two informations tricky to retrieve.

The simplest situation is when both the total number and type of existing materials within the scene are known, associated with a linear mixing process assumption. Many different unmixing algorithms rely on this prior assumption (detailed in chapter 2). Though efficient in some situations, this linearity assumption however does not necessarily hold in many real world scenarios. Due to many physical factors issued from the interactions of light and the surface materials before it reaches the captor, or characteristics of the imaging instrument, the registered pixels might not be linear mixtures of the existing components. Therefore a number of approaches has been considered to bring this non-linearity into account, yielding to more complex strategies. A major categorization among the unmixing algorithms can be explained as:

- Methods assuming to have a priori the endmembers. These methods are called *Supervised* unmixing and consists in estimating the abundances for each (known) endmember contributing to each pixel.
- Methods that do not know the endmembers. These *Unsupervised* unmixing approaches aim both at estimating endmembers and associated abundances in each pixel.

In addition to these issues, it is important to outline that in practice, we rarely have an exact spectral representation of the pure materials. Each pure material usually has several different representations depending on the conditions of acquisition, its precise composition, etc. Nuances among different spectral signatures of the same endmember can be either issued from the atmospheric condition in which signatures acquired or simply due to different levels of purity of the material.

Endmember Variability

Pure materials are indeed rarely found in the nature and they are usually contaminated with other elements or compounds. In particular case of planetary surfaces, mineral composition, mineral mixing, grain size and moisture content are important factors influencing the spectral signature of the endmembers. This phenomena in the literature is referred to as *endmember variability*. One way to address the issue of endmember variability is through *overcomplete dictionary* learning techniques.

These approaches aim at estimating several spectral signatures for only one end-member. Through constructing dictionaries of overcomplete atoms (endmembers inside dictionary), with the number of atoms much larger than the true number of endmembers within the scene, these techniques try to increase the accuracy of the analysis (unmixing, classification etc.) by having a more robust representations of the actual endmembers.

Noise and Anomaly

In addition to the earlier mentioned challenges, many domains of hyperspectral image analysis are concerned by important defects of noises. The nature of such noise is variable and depends on the atmospheric conditions at the time of image acquisition, captors precision and/or physical malfunctioning of the captors at the onset of image registration (spectrally and spatially). In general the nature of noise inside hyperspectral imagery can be divided in two categories:

- Noises which affect specific bands of the image.
- Noise which affect certain pixels of the image, across the bands of the image cube.

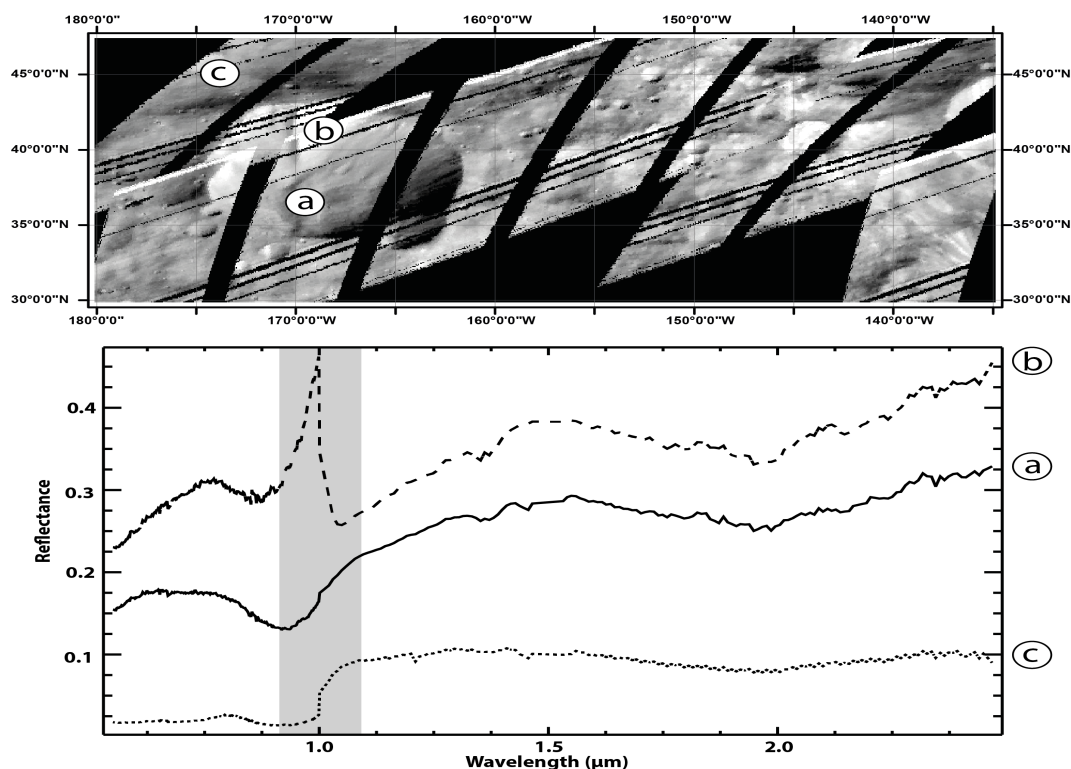


Figure 1.11: Representation of the northern hemisphere of 4-Vesta asteroid together with some nominal spectral signatures and one signature which is corrupted by anomalies.

In this thesis the later type of noise is brought under consideration (henceforth being referred as *anomaly*). Figure 1.11 illustrates the presence of anomaly within a hyperspectral image from 4-Vesta asteroid. Anomaly detection received not much attention during the early years because of :

1. the very poor spatial and spectral resolution of earlier multi-spectral images which made anomalies either embedded or mixed with other materials within a pixel.
2. the very spatial character of the image processing techniques that designed to take advantage of spatial correlation to perform the analysis, brought considerable limitations. Due to the spatial scarcity of anomalies and even worse their existence in a sub-pixel scale, those spatial domain-based methods mostly failed in capturing anomalies.

By rapid improvements in spectral and spatial resolutions of the hyperspectral images, many subtle materials which were unknown and undetectable by visual assessment, became detectable. In many cases including astronomical studies and planetary surface observations, these subtleties are most interesting targets and contain crucial information. According to [36] anomalies are characterized by three characteristic identifiers :

- Their *existence* which usually occur with low probability (w.r.t nominal pixels) and without any *a priori* knowledge.
- Their *presence* which in many cases can be under/over estimated as they can be presented as subpixel targets.
- Their *population* which by nature should not be large compared to the total number of pixels.

Unfortunately no sound number (probability, ratio, ...) is available in practice. Indeed the size of population to be considered as anomaly is not well defined in most scenarios. Moreover, the question of defining a suitable distance between anomaly and nominal data, as well as between anomalies of different nature is tricky. Lastly, how an unsupervised global algorithm can reliably account for these issues while performing other image analysis tasks like spectral unmixing? The design of unmixing approaches robust to anomalies is then a tricky task that is addressed in this PhD thesis.

Contributions of this PhD

One of the contributions of this thesis is a new method that scales favorably with the problem posed by planetary exploration. It performs an unsupervised non-linear

unmixing jointly with anomaly detection capacities, and has a global linear complexity. Non linearities are handled by analyzing the data in a Reproducing Kernel Hilbert Space (RKHS) while endmember variability is managed by decomposing the hyperspectral data on an overcomplete set of spectra, combined with a specific sparse projection. A theoretical study is proposed on synthetic datasets, and results are presented over the challenging 4-Vesta asteroid dataset. In another attempt, a new supervised unmixing algorithm is presented and focused on the exploitation of Optimal Transport. Optimal Transport consists in estimating a plan that transports a spectrum onto another with minimal cost, enabling to compute an associated distance (Wasserstein distance) that can be used as an alternative metric to compare hyperspectral data. This is exploited for spectral unmixing where abundances in each pixel are estimated on the basis of their projections in a Wasserstein sense onto known endmembers. In this work an over-complete dictionary is also used to deal with internal variability between endmembers, while a regularization term, also based on Wasserstein distance, is used to promote prior proportion knowledge in the endmember groups.

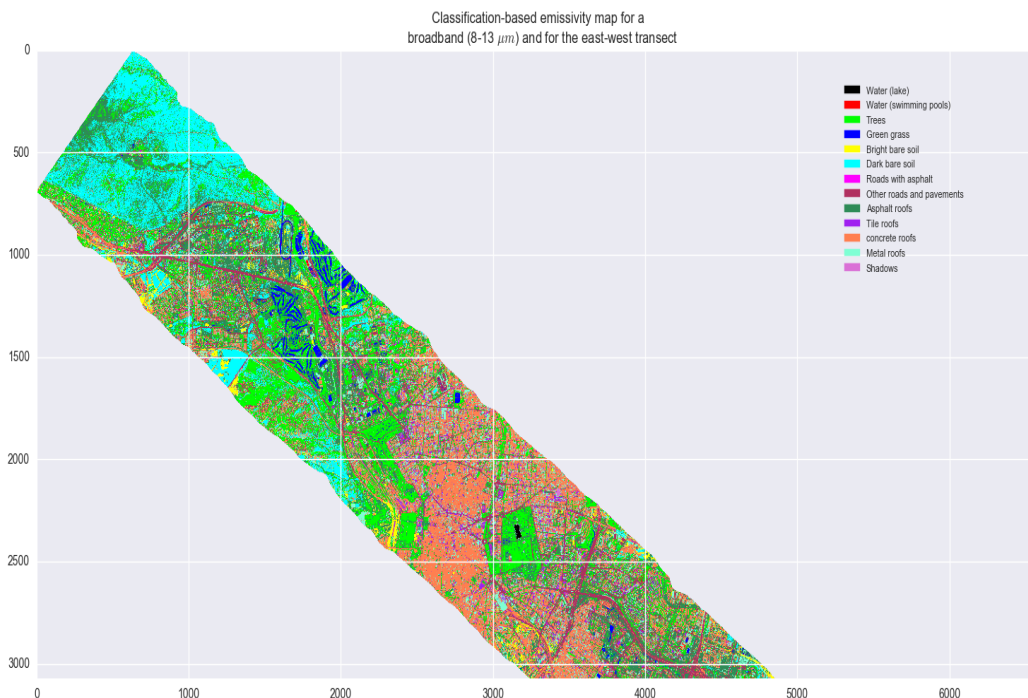


Figure 1.12: DESIREX 2008 Madrid classification map.

The last part of this PhD focuses on the application of the above mentioned algorithms in the framework of the CATUT (French acronym for Urban Temperatures Mapping) project. The aim of this project is to estimate urban land surface temperatures (LST) using mainly hyperspectral data. The long term objective is to specify the characteristics of a future hyperspectral satellite sensor able to estimate LST. The process is first to estimate the nature of the ground and then, to estimate

the associated temperature using emissivity models associated to each material. Because of physical properties, it is impossible to provide such a satellite with the very high spatial resolution required to identify all material inside an urban scene. As a consequence, hyperspectral unmixing is of prime importance here. To this end, LST estimation is mentioned as one of the goals of the Dual-use European Security IR Experiment 2008 (DESIREX 2008) campaign in Madrid Spain which carried out by European space agency (ESA) through several missions between June 23rd and 6th of July. The data used in this report, are the result of this campaign. Figure 1.12 is the classification map of a part of the city Madrid based on one of the airborne images of these series.

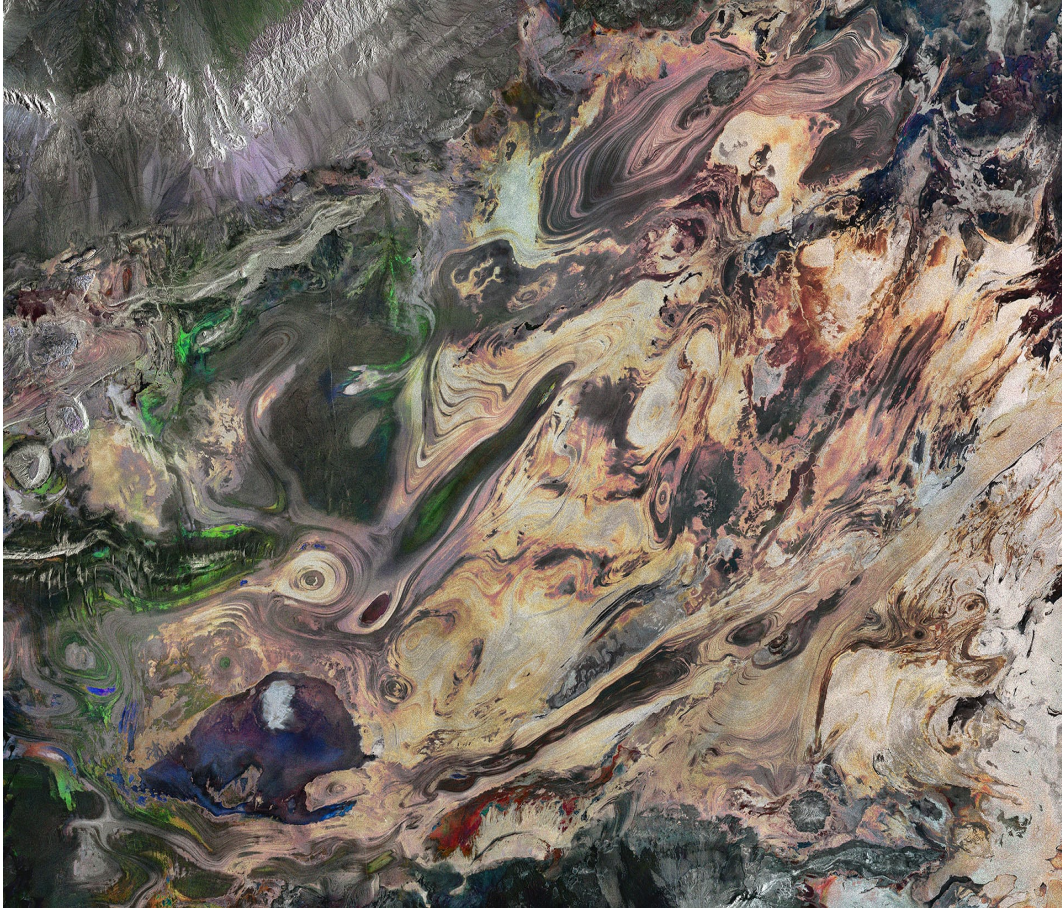
This PhD is structured as follows :

- Chapter two comes with a review of state-of-the-art. First a formal definition of the mixing process is given. Then by introducing the possible existing mixing models (linear and non-linear), different procedures for addressing the spectral unmixing are presented. In particular supervised and unsupervised methods of unmixing are explained. Further the issue of endmember estimation is explained together with the problem of endmember variability are presented. We also discuss about existing hyperspectral measures and their relationships. This chapter will be finished by giving an overview on the state-of-the-art of hyperspectral anomaly detection procedures.
- Chapter three is devoted to an unsupervised kernel-based method of anomaly detection and endmember extraction for unmixing. This is the first main contribution of this PhD.
- Chapter four details a supervised model of unmixing based on optimal transport theory. This is the second main contribution of this PhD.
- Chapter five is dedicated to a detailed application of the introduced methods within the context of CATUT project. Within this chapter we first give an outline of the goals of the project together with the procedure of data acquisition. Then by referring to the characteristics and limitations of the data we explain the data preprocessing steps necessarily to conduct the unmixing procedures. At the end the results of experiments are presented together with some notes on their analysis. This chapter will complete this thesis.
- Chapter Six comes with the conclusion and also gives indicators to the possible future extensions of the current algorithms.

Part I

State-of-the-art

Chapter Summary



Satellite image from a central desert of Iran (*Dasht-e-Kavir*), representing a large mixture of different minerals with different levels of moisture. These areas are salt lakes and the colors show fluctuations in the amount of water present over time. The image is acquired by Landsat-1. (Courtesy of ESA)

As it said earlier due to a number of reasons the values of sensors at the time of acquisition could be mixtures of two or more pure materials. In such cases the classification of hyperspectral pixels (assigning one label to each pixel) turns less representative. This calls for a way to find and retrieve information which is hidden inside each pixel value regarding the types of materials being mixed inside a pixel and their ratio of dominance. Following the introduction provided in 1.1.3 on the issue of *spectral unmixing*, within this chapter we provide a formal definition of mixing processes and introduce the latest achievements in the field of spectral unmixing. In particular linear vs nonlinear and supervised vs unsupervised methods of unmixing will be under focus. Then we will give an overview of the existing spectral measures and distances used for spectral comparison. In the last section of this chapter different approaches for hyperspectral anomaly detection are being introduced.

Hyperspectral Unmixing

2.1 Hyperspectral Data: Mixing & Unmixing

The aim of spectral data analysis is to provide comprehensive descriptions of hyperspectral measurements. In the scope of spectral unmixing it consists of detection of endmember signatures and quantification of their respective abundances over the scene of study. The unmixing of a dataset $\mathbf{X} = [\mathbf{x}_1^\top, \dots, \mathbf{x}_N^\top]$ (where \mathbf{x}_i is of dimension $(1 \times D)$ and represents the spectral signature in a D dimensional space) is an inverse procedure which approximates \mathbf{X} with the product of two (smaller in size) matrices $\mathbf{X} \approx \mathbf{E}\mathbf{G}$ where \mathbf{E} is of dimension $(D \times \ell)$ whose columns contain the endmembers and \mathbf{G} is of dimension $(\ell \times N)$ and represents the projections of each pixel onto this endmember basis. Here we call \mathbf{E} the endmembers matrix and \mathbf{G} the abundance matrix. The concept and definition of endmember is highly dependent on the application. The abundance coefficients usually come with *non-negativity* ($\mathbf{g}_j \geq 0, j = 1, \dots, \ell$) and *additivity* ($\sum_{j=1}^{\ell} \mathbf{g}_j = 1$) principles, due to the general assumption that they quantify the presence of each endmember within the pixels. These two constraints however, are also relative and not necessarily hold for all different types of mixtures. In some cases for example fractional masses of the materials could be used for interpretation of abundance coefficients. We can write the approximation of the value of one hyperspectral pixel \mathbf{x}_i as:

$$\mathbf{x}_i \approx \varphi(\mathbf{E}, \mathbf{g}_i), \quad (2.1)$$

where the approximation sign stands for the measurement noise or errors inherent to the function $\varphi(\cdot, \cdot)$. This way, the solution of the spectral unmixing (SU) will be the following minimization :

$$\min_{\mathbf{E}, \mathbf{G}} \sum_{i=1}^N \mathbf{D}_i[\mathbf{x}_i | \varphi(\mathbf{E}, \mathbf{g}_i)], \quad s.t \quad \mathbf{g}_{i,j} \geq 0, \quad j = 1, \dots, \ell, \quad (2.2)$$

where $\mathbf{D}(\cdot | \cdot)$ measures the discrepancy between the observed measurement \mathbf{x}_i and the prediction function $\varphi(\mathbf{E}, \mathbf{g}_i)$. In many cases the approximation in 2.1 assumes the

existence of additive *i.i.d* noise centered and normally distributed, with covariance matrix Σ . By this assumption in mind we can take an optimization view and solve a weighted least squares problem.

$$D_i(\mathbf{u}|\mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|_{\Sigma}^2, \quad \forall i \quad (2.3)$$

where $\|\mathbf{u}\|_{\Sigma}^2 = \mathbf{u}^T \Sigma^{-1} \mathbf{u}$ is the Mahalanobis norm which is induced by Σ .

Regardless of the type of framework (geometrical, statistical etc.) used for the formulation of the SU problem, the most important issue roots in the choice of $\varphi(\cdot, \cdot)$ which relates the observed pixel with the endmember basis matrix \mathbf{E} . The choice of this function reveals crucial information regarding the underlying mixing processes.

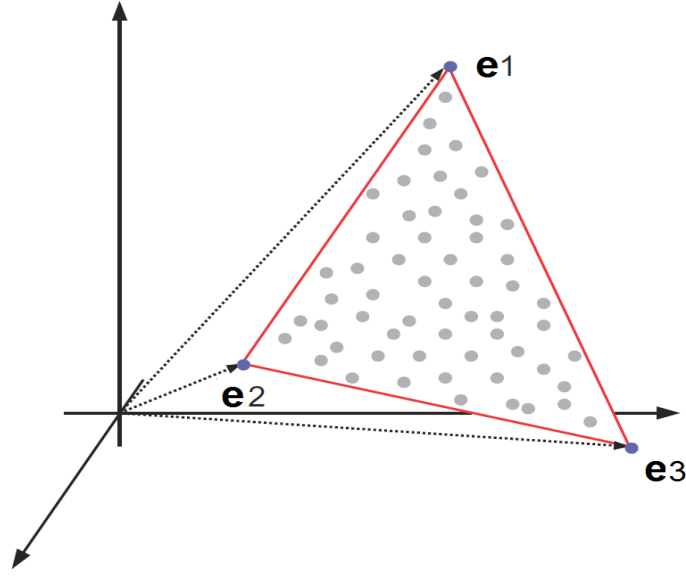


Figure 2.1: In the ideal case, where there is no effect of noise, also with the assumption of pure endmembers being present in the data, the abundance coefficients as the result of inversion part of linear unmixing will lay within a convex polygon ($n-1$ simplex) whose vertexes consist of the endmembers. Illustration courtesy of Dobigeon et al.

2.1.1 (non)Linear mixtures & (Un)supervised methods

A large number of algorithms have been proposed to deal with HS unmixing (see [3, 37] for a comprehensive survey). These algorithms, aim either at estimating endmembers and the associated mixing abundances (unsupervised methods) or they focus only on estimating abundance coefficients on the assumption that endmembers are known (supervised methods). As for unsupervised scenarios the estimation of \mathbf{E} and \mathbf{G} , depending on the underlying approach can be either simultaneous as in [38] or sequential as in [39]. The unmixing methods can also be classified based on their *a priori* assumption on *Linearity* or *Non-linearity* of the mixing procedure.

In linear mixing model (LMM), the spectral signature representing each pixel is assumed to be a linear combination of ℓ endmembers. It gives also a good eye-sight to look SU as a soft multi-class classification, as each spectral signature is linearly associated to the endmembers. In Figure 2.1 one can see the distribution of the abundance coefficients as the result of linear mixing assumption. when the additivity constrain does not hold, the coefficients instead belong to a convex cone defined by the endmembers. One of the earliest unmixing procedures is independent component analysis (ICA) [40] which comes with linearity assumption. In [41], an enhanced version of ICA based on analysis of the spectra issued from wavelet packets transformation is introduced.

Though unmixing algorithms with linear assumption, show relatively good balance between computational complexity and acceptable accuracy, the underlying linear hypothesis is violated in many practical cases [3]. for example because of *multiple scattering* phenomenon [4] that occurs due to the interaction of light with several materials (in macroscopic scale). It can also be due to *intimate mixing* of materials which refers to the simultaneous interaction of light with all the endmembers (in microscopic scale) [2]. Non-linear techniques have thus been developed to cope with this difficulty including *intimate mixtures*, *bilinear models* and other physics-based nonlinear mixing models in order to give an approximate formulation of the real mixing process. In Figure 2.2 three different mixing scenarios are illustrated.

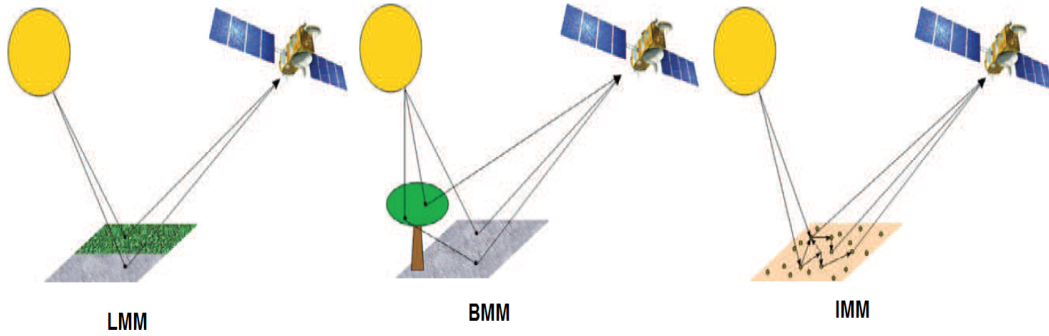


Figure 2.2: From left to right: linear mixing model(LMM), intimate mixing model (IMM) and bilinear mixing model (BMM). Courtesy of [2].

For intimate mixtures, a number of models based on the modeling and inversion of *radiative transfer* [42] have been derived. The main advantage of such approaches is the interpretability of the mixing process, since they involve real physical quantities such as grain size, roughness, single-scattering albedo, etc. However, such physical quantities, most of the time depend on unknown parameters, related to photometry and optical constants which yield a difficult estimation process. To relax the dependency on physical parameters, analytical formulations based on *bilinear models* have been successfully applied to represent the multiple scattering phenomenon which is the interactions of light with several materials. According to BMM, the value of pixel \mathbf{x}_i is being approximated w.r.t the basis matrix \mathbf{E} , this

time using two parameters α as in LMM for calculation of the linear term and β for calculation of the non-linear term as:

$$\mathbf{x}_i \approx \sum_{j=1}^{\ell} \alpha_{(j,i)} \mathbf{e}_j + \sum_{k=1}^{\ell-1} \sum_{h=k+1}^{\ell} \beta_{(k,h,i)} \mathbf{e}_k \odot \mathbf{e}_h, \quad (2.4)$$

where \odot stands for the Hadamard product and \mathbf{e}_j is the j th endmember in \mathbf{E} . Efficient associated techniques can be found in [43], the general bilinear model [44], linear-quadratic mixing [45] where recently in [46] is used for the fusion of high-spectral and high-spatial resolution HSI. Also a comparison of these models in a forestry context can be seen in [47]. To deal with nonlinearity, kernel-based algorithms which project the input data in a feature space of (possibly very) high dimension where linearity is improved also offer interesting alternatives [4, 48].

It worths mentioning that although many unmixing algorithms assume one type of non-linearity in the mixing process, some of them (see [49] as an example) do not take any assumption on the type of nonlinearity in this process (as those working on a *feature space* through the *kernel-trick*, as detailed in the following paragraph). The unsupervised kernel-based method, introduced in this thesis belongs to this family of methods. In the following we discuss about endmember and abundance estimation.

2.1.2 Estimating endmembers and abundances

To estimate the pure elements, also called endmemders, a number of methods assume that at least one pure instance per endmember is included in the data [50, 51, 5]. Various techniques exist to extract them. They are either based on a pixel purity index [50] or on the identification of the simplex that encompasses data [52, 53]. This latter idea exploits the property that a linear combination of some points is included in the simplex, issued from these points; as a consequence extreme points of the simplex correspond to endmembers. N-Finder [51] is for example, issued from this representation since the maximum volume of a simplex is defined by the purest pixels in a dataset. According to the iterative version of this method, it is assumed that the number of endmembers is known *a priori*. By applying a dimensionality reduction technique like PCA [54] or MNF [55], the dimensionality of the original space of data is being reduced to $\ell - 1$. The iterative part of the algorithm starts by randomly selecting ℓ signatures from the data, taking them as endmembers, then within an outer loop of k iteration the algorithm tries to find the volume of the simplex specified by those endmembers. In particular for the ℓ -simplex $\mathbf{S}\{\mathbf{e}_1, \dots, \mathbf{e}_\ell\}$, the volume of \mathbf{S} is being calculated as:

$$\mathbf{V}(\mathbf{e}_1, \dots, \mathbf{e}_\ell) = \frac{\left| \det \begin{bmatrix} 1 & 1 & \dots & 1 \\ \mathbf{e}_1 & \mathbf{e}_2 & \dots & \mathbf{e}_\ell \end{bmatrix} \right|}{(\ell - 1)!}. \quad (2.5)$$

Within an inner loop all the data samples are being added to \mathbf{S} and the volume of the new simplex will be re-calculated in order to reach the maximum volume of the simplex.

Techniques based on non-negative matrix factorization (NMF) [56, 57, 58, 59] exploit also this property and the simplex is sometimes computed through geodesic distances [60] or in *Reproducing Kernel Hilbert Space* (RKHS) [39]. This *kernel-trick* procedure enables to project data in a *feature space* (generally of high dimension), where the linearity assumption is more relevant. This efficient strategy has enabled to tackle a number of unsupervised, non-linear unmixing processes (see [2, 4, 61, 62, 63]). Finally let us point out VCA (Vertex Component Analysis [5]) which is based on the idea that affine transformation of a simplex is also a simplex. According to this method, endmembers are being estimated iteratively by projecting data onto the direction which is orthogonal to the subspace spanned by the previously found endmembers, until the preset number of endmembers is reached.

Similar to N-finder, VCA knows the number of endmembers and starts by dimensionality reduction in the preprocessing step but the initial endmembers are being calculated as $\mathbf{A} = [\mathbf{e}^0, \mathbf{0}, \dots, \mathbf{0}]$, where $\mathbf{e}^0 = \underbrace{(\mathbf{0}, \mathbf{0}, \dots, \mathbf{1})}_{\ell}$. Then in each iteration the volume of a convex hull is being calculated after adding new orthogonal vector to the current basis :

$$\mathbf{f}^k = \left(\left(\mathbf{I} - \mathbf{A}^{(k-1)} (\mathbf{A}^{(k-1)})^\# \right) \mathbf{w}^k \right) / \left(\left\| \left(\mathbf{I} - \mathbf{A}^{(k-1)} (\mathbf{A}^{(k-1)})^\# \right) \mathbf{w}^k \right\| \right). \quad (2.6)$$

Although efficient, all these techniques assume the presence of pure elements in the image which is not guaranteed in practice. By relaxing the assumption of existing pure pixels, the *minimum volume* (MV) algorithms try to estimate the closest simplex encompassing data, making the estimation a much more difficult non-convex optimization procedure. MVC-NMF[64], SISAL[65], MVSA[66] etc. are examples of this trend of SU algorithms. In the next part we consider a new approach for addressing the issue of non-pure pixels.

2.1.3 Endmember variability & Overcomplete dictionaries

The problem of endmember quality has been the focus of several recent studies [67, 68]. As a matter of fact in many cases endmembers are either estimated up to some uncertainties or characterized not only by a single spectrum but rather by a set of different spectra. Using only one spectrum to decompose the information is then hazardous and likely to decrease the efficiency of the unmixing process. In order to cope with these issues, overcomplete dictionaries of hyperspectral signatures have revealed their efficiency [69, 70]. The general idea is to construct a set/dictionary of spectral signatures larger than the effective number of endmembers to deal with the internal heterogeneity inside a particular material (which should correspond to

only one endmember in classical approaches). In some literature, methods where overcomplete dictionaries are used in SU are referred as semi-supervised methods. In order to handle the large number of signatures in dictionaries, the minimization procedure which is used in inversion phase is modified to account for sparsity constraint or penalization while the additivity constraint is being relaxed. Therefore, with a quadratic cost function, the SU consist in solving a sparse LS regression :

$$\min_{\mathbf{g}_i} \|\mathbf{x}_i - \mathbf{M}\mathbf{g}_i\|_2^2 \quad s.t. \quad \|\mathbf{g}_i\|_0 \leq \lambda, \quad \mathbf{g}_{j,i} \geq 0 \quad \forall j, \quad (2.7)$$

where λ is the sparsity level. The $\|\mathbf{g}_i\|_0$ therefore, controls the permitted number of endmembers from the dictionary \mathbf{M} to be involved in the mixture. Greedy algorithms are needed for solving these NP-hard problems. Well-know techniques are including but not limited to Orthogonal matching pursuit [71] and SUNSAL [72].

The overcomplete dictionary principle will be explored in the inversion phase of both supervised and unsupervised methods of unmixing in this thesis. Before embarking upon the inversion part we first introduce some distance measures that are commonly used in unsupervised SU procedures.

2.2 Distance Measures for HSI

Hyperspectral vector-based measures are mainly used in applications like detection, discrimination, classification and identification. In the case of unsupervised scenarios with vectorial data in particular they are the only means of evaluation and comparison. In this section therefore we introduce the major mathematical measures commonly used in analysis of HSI. In the following we introduce a number of *vector-based measures* which share some inherent similarities. Although these measures were not used in this thesis, because they show very interesting and promising properties, also the fact that their utilization is less emphasized in the remote sensing literature, we decided to give the definition of the *sample spectral correlation based measures* and their major subcategories.

2.2.1 Signature vector-based measures

Vector-based spectral measures are mainly used in applications like detection, discrimination and less frequently in classification. In the absence of prior class information, these measures are mainly used to compare two spectral signatures. These measures are effective mostly when the spectral vectors to be compared are believed to be the true representatives of the materials/endmembers which are assigned to. For two spectral signatures \mathbf{x}_i and \mathbf{x}_j of size $(1 \times D)$ we define:

Euclidean Distance (ED)

$$\mathbb{ED}(\mathbf{x}_i, \mathbf{x}_j) = \|\mathbf{x}_i - \mathbf{x}_j\| = \sqrt{\sum_{l=1}^D (x_{il} - x_{jl})^2}, \quad (2.8)$$

which is the most well known distance measure defined over hermitian spaces.

Spectral Angle Mapper (SAM)

$$\text{SAM}(\mathbf{x}_i, \mathbf{x}_j) = \cos^{-1} \left(\frac{\vec{\mathbf{x}}_i \cdot \vec{\mathbf{x}}_j}{\|\vec{\mathbf{x}}_i\| \|\vec{\mathbf{x}}_j\|} \right), \quad (2.9)$$

where $\|\vec{\mathbf{x}}_i\|$ is the norm of \mathbf{x}_i . SAM measures the angle between two vector variables. It is also referred as the cosine distance. When the angle between two signatures is small SAM and ED give similar results.

Orthogonal Projection Divergence (OPD)

$$\text{OPD}(\mathbf{x}_i, \mathbf{x}_j) = \left(\mathbf{x}_i^\top \mathbf{P}_{\mathbf{x}_j}^\perp \mathbf{x}_i + \mathbf{x}_j^\top \mathbf{P}_{\mathbf{x}_i}^\perp \mathbf{x}_j \right)^{1/2}, \quad (2.10)$$

where $\mathbf{P}_{\mathbf{x}_k}^\perp = \mathbf{I} - \mathbf{x}_k(\mathbf{x}_k^\top \mathbf{x}_k)^{-1} \mathbf{x}_k^\top$ and $\mathbf{I}_{(D \times D)}$ is an identity matrix.

This measure works based on the orthogonal subspace projection algorithm [73], and finds the residuals of orthogonal projection of \mathbf{x}_i and \mathbf{x}_j onto each other. From this perspective OPD and SAM also find similarities, as both of them are projective methods.

Spectral Information Divergence (SID)

$$\text{SID}(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{D}(\mathbf{x}_i || \mathbf{x}_j) + \mathbf{D}(\mathbf{x}_j || \mathbf{x}_i), \quad (2.11)$$

where $\mathbf{D}(\mathbf{x}_i || \mathbf{x}_j)$ is the relative entropy (Kullback-Leibler information divergence). SID measures discrepancy between two signature vectors \mathbf{x}_i and \mathbf{x}_j in terms of their corresponding probability mass functions. while OPD calculates the divergence of one signature projected onto the other, SID can be thought of as the stochastic version of OPD where orthogonal projection is replaced by information divergence.

Mixed SID – SAM

Also By fusion of SAM and SID in [74] it is shown that:

$$\text{SID} - \text{SAM} = \text{SID}(\mathbf{x}_i, \mathbf{x}_j) \times \tan(\text{SAM}(\mathbf{x}_i, \mathbf{x}_j)) \quad (2.12)$$

and

$$\text{SID} - \text{SAM} = \text{SID}(\mathbf{x}_i, \mathbf{x}_j) \times \sin(\text{SAM}(\mathbf{x}_i, \mathbf{x}_j)) \quad (2.13)$$

both show better discriminability in some cases [74].

2.2.2 Correlation-weighted measures

The idealistic assumption based on which the vector-based measures operate (presented in 2.2.1) is not true in many cases. In particular in mixed signature *discrimination* and *identification* where spectral vectors are usually mixed with a number of other signatures. Another situation is the subsample target discrimination and identification where the interesting target/endmember is embedded in a single signature and mixed with other signatures. In such situations, the precision of vector-based measures deteriorates. In order to ameliorate this problem *correlation-weighted measures* extend the *vector-based measures* by including the sample spectral correlation to account for spectral variability of the materials. Based on the way these models incorporate the correlation information, the *correlation-weighted measures* are being categorized into measures weighted by *a priori* correlation and measures weighted by *a posteriori* correlation.

HS measures weighted by *a priori* correlation

By incorporation of *a priori* correlation, it is meant that there is a known correlation available prior to the comparison of signatures (identification or discrimination). The orthogonal subspace projection (OSP) approach introduced in [73] is exploiting this *a priori* knowledge. It is possible to decorrelate signature vector \mathbf{x}_i , by orthogonally projecting it onto the basis formed by the columns of \mathbf{E} which means the space $\langle \mathbf{E} \rangle^\perp$ orthogonal to the space linearly spanned by the columns of \mathbf{E} . Different identification and discrimination measures are forged based on the OSP principle

$$P_E^\perp = \mathbf{I} - \mathbf{E}(\mathbf{E}^\top \mathbf{E})^{-1} \mathbf{E}^\top, \quad (2.14)$$

where \mathbf{E} is the set of ℓ endmember signatures.

HS measures weighted by *a posteriori* correlation

As availability of the *a priori* knowledge on the existing endmembers \mathbf{E} is not always guaranteed, an approximation of this knowledge is shown [36] to be achievable by the inverse of the sample spectral correlation matrix \mathbf{R}^{-1} . In this respect sample spectral correlation/covariance can be used to come up with *a posteriori* correlation weighted measures.

2.3 Inversion

Once endmembers known, an optimization procedure is often used to estimate the abundances in each pixel. It is called inversion and usually takes the form $\hat{\mathbf{g}}_i = \arg \min_{\mathbf{g}} \|\mathbf{x}_i - \varphi(\mathbf{E}, \mathbf{g})\|_2^2$ where \mathbf{E} represents the endmembers and $\varphi(\mathbf{E}, \cdot)$ is a (non)linear function that links the vector of abundances \mathbf{g} to the current pixel \mathbf{x}_i to unmix. As we mentioned in 2.1.1 various mixing processes φ and penalization norms have already been proposed (see for example [44, 43, 4]). In cases where the discrepancy measure takes the form of euclidean distance, this optimization procedure boils down to constrained least squares (CLS) regression. The difficulty with the imposed constraints (resulting in impossibility of a closed-form solution) gave raise to other forms of optimization including fully constrained least squares (FCLS) [75] and later on their kernelized versions [76]. In the case of FCLS we formulate the minimization problem as :

$$\hat{\mathbf{g}}_i = \min_{\mathbf{g}} \left\{ (\hat{\mathbf{x}}_i - \hat{\mathbf{E}}\mathbf{g})^\top (\hat{\mathbf{x}}_i - \hat{\mathbf{E}}\mathbf{g}) \right\} \quad s.t \quad \mathbf{g} \geq 0 \quad and \quad \sum_{j=1}^{\ell} g_j = 1 \quad (2.15)$$

Because of the existence of the non-negativity inequality constraints there are no closed-form solution for the optimization of FCLS. In [75] an iterative solution is given which simultaneously enforces the non-negativity and sum-to-one constraints on the abundance values.

For this, in order to account for the sum-to-one constraint a surrogate of the endmember matrix \mathbf{E} and the datum \mathbf{x}_i is calculated as:

$$\hat{\mathbf{E}} = \begin{bmatrix} \delta \mathbf{E} \\ \mathbf{1}^\top \end{bmatrix} \quad (2.16)$$

and

$$\hat{\mathbf{x}}_i = \begin{bmatrix} \delta \mathbf{x}_i \\ \mathbf{1} \end{bmatrix} \quad (2.17)$$

Where $\mathbf{1} = \underbrace{(1, \dots, 1)^\top}_{\ell}$ and δ controls the impact of the sum-to-one constraint.

Since $\mathbf{g} \geq \mathbf{0}$ is a set of inequalities, it is impossible for Lagrange multiplier to reach optimal solutions, therefore a positive constraint constant vector \mathbf{c} is being defined to take care of non-negativity constraint. FCLS can be derived by forming a Lagrangian J as:

$$J = \frac{1}{2} (\hat{\mathbf{x}}_i - \hat{\mathbf{E}}\mathbf{g})^\top (\hat{\mathbf{x}}_i - \hat{\mathbf{E}}\mathbf{g}) + \lambda(\mathbf{g} - \mathbf{c}), \quad (2.18)$$

where $\mathbf{c} = (c_1, \dots, c_\ell)^\top$ is a self defined positive constraint constant vector and

$c_i > 0, \forall i$. Then iterating through the two equations:

$$\hat{\mathbf{g}} = (\hat{\mathbf{E}}^\top \hat{\mathbf{E}})^{-1} \hat{\mathbf{E}}^\top \hat{\mathbf{x}}_i - (\hat{\mathbf{E}}^\top \hat{\mathbf{E}})^{-1} \lambda \quad (2.19)$$

and

$$\lambda = \hat{\mathbf{E}}^\top (\hat{\mathbf{x}}_i - \hat{\mathbf{E}} \hat{\mathbf{g}}), \quad (2.20)$$

which can be used to find the optimal solution $\hat{\mathbf{g}}$ and the Lagrange multiplier vector $\lambda = (\lambda_1, \dots, \lambda_\ell)$.

It should be noticed that in many real-world scenarios the inverse problem is ill-posed and this bounds the efficiency of matrix factorization models like FCLS where the strong assumption of linearity is hold for the mixing process. This is why other discrepancy measures, including KL-divergence are being used instead of the ED. Also care must be taken that least squares does not produce sparse coordinates, while in most cases it is highly probable that the pixel is a mixture of few number of endmembers λ where $\lambda \ll \ell$. Therefore methods which account for this sparsity have been gained attention during the past years. In [77], taking the so-called *manifold assumption*, a graph regularizer is added to sparsity constrained NMF in order to account for the sparsity of solutions as well as the geometry of the manifold embedding data. For this a nearest neighbor graph is created with data points as its vertices together with a weighting matrix based on the heat kernel. The optimization will be an extension of $L_{1/2}NMF$ introduced in [78]:

$$\min_{(\mathbf{E}, \mathbf{G})} = \frac{1}{2} \|\mathbf{X} - \mathbf{E}\mathbf{G}\|_F^2 + \lambda \|\mathbf{G}\|_{1/2} + \frac{\mu}{2} \text{Tr}(\mathbf{G}\mathbf{\Lambda}\mathbf{G}^\top), \quad (2.21)$$

where $\mu \geq 0$ is a regularization parameter and $\mathbf{\Lambda}$ is a symmetric matrix calculated based on the weighting matrix. Similar to FCLS an iterative procedure with the use of Lagrange multipliers solves the minimization. As for the initialization of \mathbf{E} and \mathbf{G} , it is proposed to use VCA (for initialization of \mathbf{E}) and FCLS (for initialization of \mathbf{G}). Similar to FCLS, during the updates, 2.16 and 2.17 will be calculated based on \mathbf{E} and \mathbf{x}_i , in order to guaranty the non-negativity and sum-to-one constraints.

In the next section we give an overview on the current methods of anomaly detection for HSI.

2.4 Anomaly detection

As mentioned in chapter 1, anomalies are likely to be present in hyperspectral data. Although noise can more or less be efficiently managed with some dimensionality reduction and denoising techniques, anomalies, because of their coherent structure, still affect their precision. All techniques, relying on endmember estimation directly

from the image are sensitive to anomalies. For example, strategies based on extraction of a simplex will embed anomalies that are most of the time outside the simplex [79, 80, 5]. The identification of these existing anomalies is then of prime importance.

Many anomaly detection techniques rely on statistical approaches where the goal is to identify among data, points with strongly different statistics w.r.t the majority of data [81, 82], these latter being viewed in a local context (local neighborhood of the current test pixel) or global one (the whole scene). On this basis, a number of approaches have been proposed as for example the use of a Gaussian Mixture Model (GMM) [83] or Gaussian Markov Random Field [84] to model the nominal distribution of data.

Let us finally outline that other approaches try to represent data in other basis to better identify outliers, as in [85] where a Discrete Wavelet Transform (DWT) is used to represent points whereas a kernel-PCA is exploited in [86] and [87].

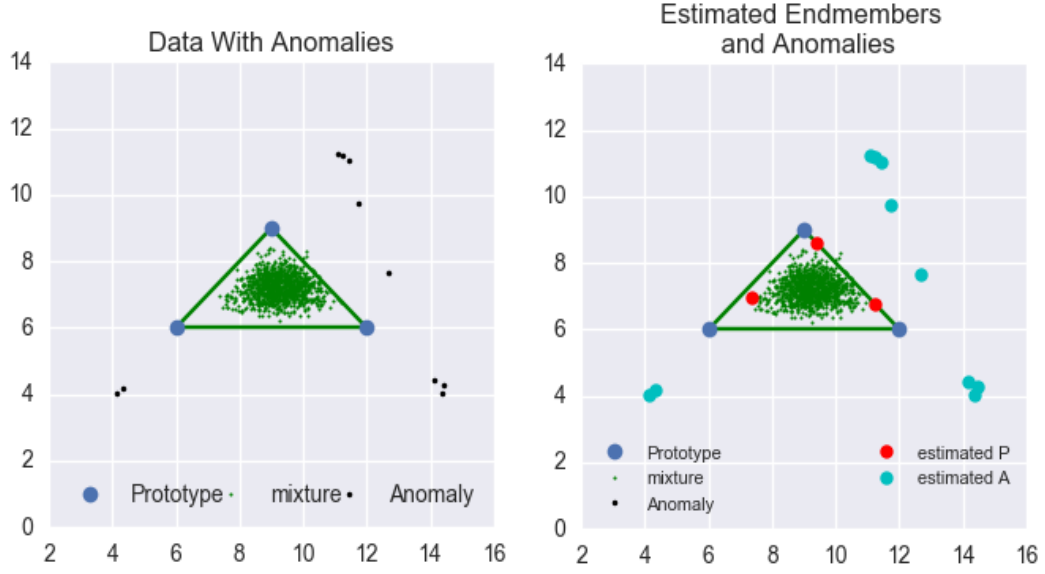
By this overview on different concepts and tools which we encounter in the realm of remote sensing and hyperspectral data analysis we finish this chapter and in the next one an original algorithm of spectral unmixing and anomaly detection which exploits many of the above mentioned concepts and tools is being presented.

In this chapter we first introduced the concept of spectral mixing and the known mathematical models used for explaining different types of mixtures. In particular we made the difference between linear and non-linear mixing models clear. We also provided an introduction to the existing and frequently used methods of spectral unmixing, their assumptions and different properties that make each suitable for addressing certain types of mixtures. Then we introduced different metrics that are frequently used for measuring difference between spectral signatures. In the next part we explained the inversion procedures for calculating the abundance coefficients corresponding to each endmember. At the end we introduced the common problem of anomaly and the existing algorithms for dealing with anomalies within hyperspectral data.

Part II

Kernel methods for Unmixing

Chapter Summary



Endmember extraction and anomaly detection on synthesized 2D data

In this chapter a nonlinear unsupervised kernel-based model of unmixing associated with an anomaly detection procedure is presented. This model is meant to address the ubiquitous problem of noisy datasets where the existence of noise distorts the correct representation of the manifold on which the data exists. The heart of this representation lays on using a kernel to project data from its original space (where data sits on a non-simplicial manifold) onto a feature space where the projected data is embedded by a simplicial manifold. The unmixing of a dataset $\mathbf{X} = [\mathbf{x}_1^T, \dots, \mathbf{x}_n^T]$ (where \mathbf{x}_i is of dimension $(1 \times n)$ and represents the spectral signature in a n dimensional space) is an inverse procedure which approximates \mathbf{X} with the product of two (smaller in size) matrices $\mathbf{X} \approx \mathbf{E}\mathbf{G}$ where \mathbf{E} is of dimension $(D \times \ell)$ whose columns contain the endmembers and \mathbf{G} is of dimension $(\ell \times N)$ and represents the projections of each pixel onto this endmember basis. Here we call \mathbf{E} the endmembers matrix and \mathbf{G} the abundance matrix. In 3.1 we first give a short review of the applications of HSI in planetary science, as well as existing challenges for processing planetary images. We provide the definitions and short overview of the latest approaches for spectral unmixing and mention the importance of anomaly detection for planetary hyperspectral imagery. In 3.2 we give a thorough explanation of the model. First in 3.2.1 we will provide the formal definition for exact simplex volume maximization procedure and then in 3.2.2 we derive the iterative procedure of anomaly detection. In 3.2.3 we mention how the abundance values are being calculated through sparse projections onto the simplex. In 3.2.4 the two major parts of the model are brought in pseudocode form in order to clarify the whole procedure. Then in 3.3 we set up a series of experiments on synthetic and real datasets. First in 3.3.1 we explain the

procedure of generating synthetic datasets and the parameters which are involve in it. Then the results of applying the algorithm on these datasets are demonstrated in 3.3.2 together with their comparison with other unmixing techniques. In 3.3.3 we demonstrate the benefits of exploiting overcomplete dictionaries in estimation of the endmembers and finally in 3.3.4 we demonstrate the performance of our joint anomaly detection and spectral unmixing algorithms on a real dataset.

☞ This chapter corresponds to an article published in IEEE TGRS under the title "Joint anomaly detection and spectral unmixing for planetary hyperspectral images" (Volume: 54, Issue: 12, Dec. 2016).

Joint Anomaly detection and spectral Unmixing

Abstract

Hyperspectral images are commonly used in the context of planetary exploration, especially for the analysis of the composition of planets. As several instruments have been sent throughout the Solar System, a huge quantity of data is getting available for the research community. Among classical problems in the analysis of hyperspectral images, a crucial one is unsupervised non-linear spectral unmixing, which aims at estimating the spectral signatures of elementary materials and determining their relative contribution at a sub-pixel level. While the unmixing problem is well studied for earth observation, some of the traditional problems encountered with earth images are somehow magnified in planetary exploration. Among them, large image sizes, strong non-linearities in the mixing (often different from those found in the earth images) and presence of anomalies are usually impairing the unmixing algorithms. This chapter presents a new method that scales favorably with the problem posed by this analysis. It performs an unsupervised unmixing jointly with anomaly detection capacities, and has a global linear complexity. Non linearities are handled by decomposing the hyperspectral data on an overcomplete set of spectra, combined with a specific sparse projection, which guarantees the interpretability of the analysis. A theoretical study is proposed on synthetic datasets, and results are presented over the challenging 4-Vesta asteroid dataset.

Keywords

Planetary Hyperspectral Unmixing, Anomaly Detection, Manifold Learning, Non-negative Matrix Factorization, Overcomplete Dictionary, Kernel based Learning.

3.1 Introduction

Hyperspectral (HS) remotely sensed images are of prime interest in many scientific fields, since they enable us to assess a dense spectrum (generally composed of several hundreds of contiguous electromagnetic wavelengths) in each pixel, making possible the identification of various materials composing the scene [82]. For this reason, hyperspectral data are vastly being used in earth observation (identifying land cover, crops, ... see for example [88, 89]) as well as planetary exploration (identification of minerals and rocks [90]). This latter application is the one we focus in this chapter. To understand the processes that drive the formation and evolution of planets, it is crucial to characterize their surface's composition. Visible-near infrared reflectance spectroscopy has long been recognized as an extremely powerful tool to achieve this objective as it allows to determine rocks composition at both local and global scales [91]. Imaging spectroscopy therefore, has been rapidly grown in planetary exploration, and very soon dedicated instruments were sent toward other planets in our Solar System [92]. Number of pioneering instruments have followed resulting in huge amount of data available for the scientific community. From the computational point of view however, the processing of hyperspectral images is particularly challenging when considering planetary exploration, since unlike the earth, no ground truth nor field campaigns exist except few particular cases (i.e. Apollo landing sites for the Moon and rovers/landers locations for Mars). Moreover, despite continuous spectral and spatial resolution improvements of HS images, they still suffer from a considerable low spatial resolution to observe every single material. As a consequence a pixel generally contains mixed information of several elements which are of key interest when considering planetary processes, especially magmatic rocks (i.e. rocks made of the minerals *olivine* and *pyroxenes*) that are used to evidence igneous processes typical of the crust or mantle. It should also be outlined that in addition to observation conditions (e.g. illumination, shadows) and to the nature of the surface itself (e.g. roughness, composition) which create non-linearities (as *multiple scattering* [4] or *intimate mixing* [2]), olivine and pyroxenes can have a wide range of chemical compositions, leading for each of them to a non-linear spectral response (e.g. [93] and references therein). Combined to this effect, rocks are themselves a combination of minerals, implying to decompose the different spectra not on a basis of pure materials but rather as a combination of mixtures of materials.

Apart from these problems related to the specificity of hyperspectral data and planetary contexts, let us also note that *anomalies* often occur in the acquisition process of hyperspectral images. Anomaly is referred to any pixel whose signature is considerably different from its surrounding background. Several factors give rise to anomalies : 1) Natural degradation due to time spent in space (especially regarding missions far away from the Solar System) where unrealistic responses can appear progressively and randomly ; 2) Varying natures of sensors used to cover the whole wavelength range which generate misregistrations ; 3) Correction of atmospheric effects where for some planetary objects (e.g. Mars, Titan) residues can affect spectra in specific wavelengths.

Although efficient pre-launch calibration procedures exist, they are not sufficient to perfectly clean data. All these reasons give rise to the necessity of efficient non-linear *unmixing* procedures, robust to outliers, able to decompose mixed pixels into a number of *pure* reflectance spectra called *endmembers*. This problem refers to an inverse procedure to determine endmembers and estimate *fractional abundances* for each pixel [94]. Traditionally, anomaly detection and spectral unmixing have taken as separate subjects. A number of families of unmixing techniques which aim at recovering abundances of endmembers directly from the image (or from a separate dictionary of acquired signatures) have been presented. Despite efficient and sound existing methods (see [3, 37, 2] for an overview), most techniques still suffer from degraded performances in presence of anomalies. In such situations, endmembers are often being incorrectly estimated and consequently, associated fractional abundances are less meaningful. Therefore, together with unmixing, an accurate detection of anomalies is crucial. This constitutes the aim of this chapter.

3.1.1 Summary, goals of the chapter

In this introduction we have discussed the necessity of both unmixing HS images and the detection of anomalies. Also interesting techniques have been cited for solving spectral unmixing. However, to the best of our knowledge, the joint estimation of anomalies and unmixing hyperspectral data has not been proposed yet ¹. In this work we introduce a technique entitled SAGA+ (Sparse And Geometry Aware +) able to jointly detect anomalies and perform spectral unmixing. This *Anomaly-Detection* and *Spectral Unmixing* procedure is unsupervised and defined in RKHS. The idea consists of increasing the volume of the *manifold hull* in feature space. SAGA+, in an iterative way, increases the volume of the simplex by finding the purest pixels within the dataset, while at the same time keeping track of the rate of change of the sum of projection errors (SOPE) in order to exclude anomalies from being taken as simplex vertexes.

The chapter is organized as follows: Section 3.2 introduces the *Anomaly Detection* and *Spectral Unmixing* model mentioned above and in section 3.3, evaluation of the model and comparison results using both synthetic and real datasets are presented. Finally, section 3.4 concludes this work with an overall conclusion and pointers to possible future extensions.

¹Unfortunately, only after submitting this draft we became aware of a very recent work [95], using Bayesian framework for joint anomaly detection and spectral unmixing.

3.2 SAGA+ : simplex volume maximization and anomaly detection

In this section the methodological contributions are presented. The spectral unmixing relies on a matrix factorization technique denoted SAGA (Sparse And Geometry Aware) and introduced in [39]. In the present contribution we add an anomaly detection and removal approach to deal with aforementioned issues, yielding an approach denoted "SAGA+". SAGA is based on the geometrical concept of finding the simplex that embeds data. This simplex is being computed in a feature space associated to a kernel. As for anomalies, they are tracked through their lack of representational capacity in the dataset.

We first present the endmember extraction technique in 3.2.1 and subsection 3.2.2 is devoted to anomaly detection. Once endmembers extracted, abundances are estimated by projecting data onto the endmembers. This procedure is detailed in 3.2.3. The overall process, entitled SAGA+ is summarized in 3.2.4. Before entering into details, let us introduce some general notations:

The hyperspectral image contains D bands and N pixels $\mathbf{x}_i \in \mathbb{R}^D, i = 1, \dots, N$. Data are represented in a $D \times N$ matrix composed of N vectors $\mathbf{X} = [\mathbf{x}_1^T, \dots, \mathbf{x}_N^T]$. The unmixing is formalized as finding a representation of the form : $\mathbf{X} \approx \mathbf{E}\mathbf{G}$ where columns of $\mathbf{E} \in \mathbb{R}^{D \times \ell}$ contain the ℓ endmembers and $\mathbf{G} \in \mathbb{R}^{\ell \times N}$ is the matrix, consisting of projections of each pixel onto this endmember basis. In practice the interpretability of matrix \mathbf{G} is enhanced if each column is positive and sums to 1, since in this case, each value \mathbf{G}_{ij} contains the abundance of endmember i in pixel j . Such a Non-negative Matrix Factorization (NMF) is the basic idea on which we rely. To estimate the endmember matrix \mathbf{E} , we rely on Column Subspace Selection principles: As for each column of \mathbf{E} is selected an input pixel \mathbf{x}_i . This idea suggests that at least one sample of each endmember is present in the data (since this assumption is likely to be violated, in practice we rely on overcomplete strategies, as will be explained). The following section introduces the simplex approach to estimate matrix \mathbf{E} .

3.2.1 SAGA principles: Exact Simplex Volume Maximization for endmember extraction

As mentioned earlier, we rely on a Column Subspace Selection procedure to construct the simplex that embeds the volume generated by the data [96]. To this end we start from the idea that the bigger this simplex is, the more likely it will contain the other elements, and therefore the reconstruction error of any projection onto the new simplex will be small or null. The ℓ vertices of the simplex then correspond to the endmembers. In this study, we also claim that choosing an overcomplete endmembers dictionary will help in performing non-linear unmixing.

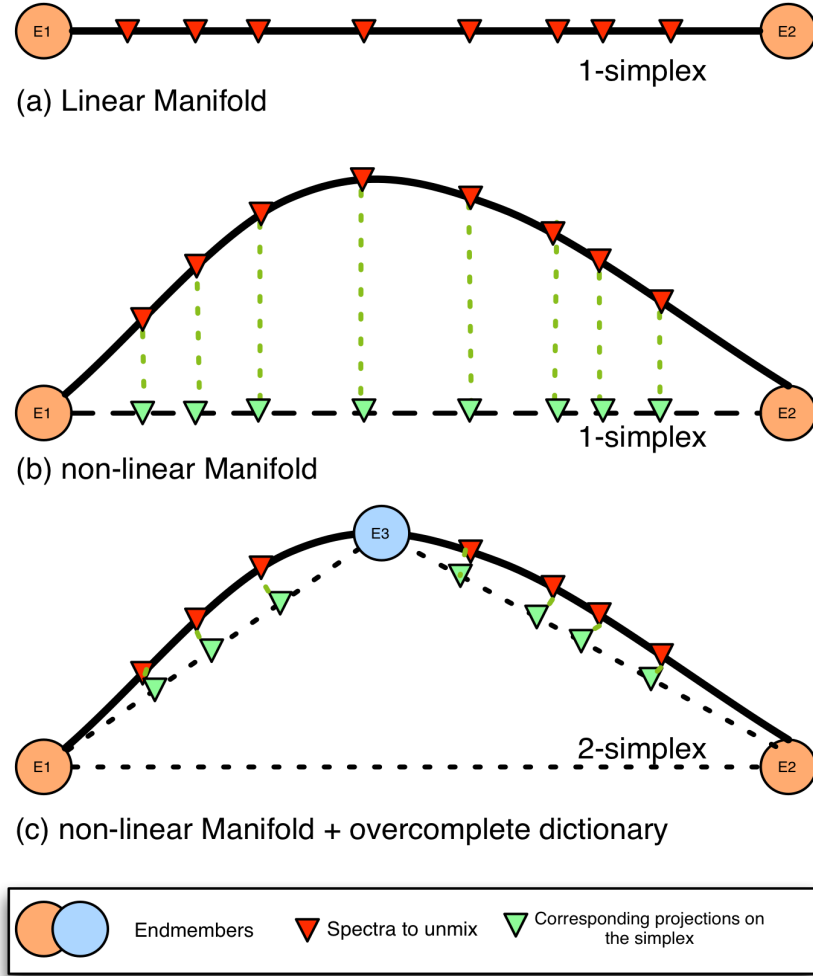


Figure 3.1: Overcomplete endmember dictionary for non-linear unmixing

This point is illustrated in Figure 3.1. In the first case (Figure 3.1(a)), the spectra are living in a perfectly linear manifold. In this case, two endmembers ($E1$ and $E2$) are sufficient and any spectrum can be described as a convex combination of the two endmembers. This is referred to as the linear mixing hypothesis. However, this ideal case is usually different from reality, as illustrated on Figure 3.1(b), where the spectra live on a non-linear manifold. If the linear assumption is nonetheless assumed, then the spectra are projected on the 1-simplex formed by the two endmembers, resulting in an error in the corresponding abundances. By choosing an appropriate additional endmember ($E3$, blue circle in Figure 3.1(c)), one can build a piece-wise linear approximation of the non-linear manifold. The unknown spectra are now projected on the corresponding 2-simplex. Note here that this projection is naturally sparse, *i.e.* for all spectra only two endmembers are required to describe perfectly the projection. However this comes at the price of having one endmember that is in

itself a mixture of the first two ones. There are then two cases:

1. Either one wants only to estimate the mixture coefficients over the two “pure” endmembers $E1$ and $E2$. Assume for the example that $E3$ is a mixture of known coefficients (0.4 and 0.6) of $E1$ and $E2$. Then for any spectrum with decomposition $\mathbf{g}_1 = 0.2$, $\mathbf{g}_2 = 0$ and $\mathbf{g}_3 = 0.8$, the corresponding abundance *w.r.t.* only $E1$ and $E2$ is now $\mathbf{g}_1 = 0.2 + 0.8 \times 0.4 = 0.52$ and $\mathbf{g}_2 = 0 + 0.8 \times 0.6 = 0.48$. More generally, in the case where λ pure endmembers are available and ℓ endmembers are chosen for the overcomplete dictionary, we suppose that we have access to a (generally sparse) linear operator \mathcal{A} of size $(\lambda \times \ell)$, with prescribed sums of rows = 1, that can be applied on any result of unmixing *w.r.t.* the overcomplete dictionary to get only a decomposition over pure endmembers.
2. However, we agree that obtaining this mixing matrix might not be a trivial task even for specialists. Nevertheless in some domains, as the one explored in this work, specialists are used either to work with spectra obtained in laboratory that are already combination of pure chemical elements. The decomposition on the overcomplete basis, provided it can be interpreted, gives in itself interesting insights on the underlying geological process.

In addition of the overcomplete strategy, to deal with complex manifolds issued from nonlinearities in \mathbf{X} , it is possible to project the data in a feature space \mathcal{H} using a projection function $\phi : \mathbb{R}^D \rightarrow \mathcal{H}$ (working in the original space is equivalent to choose the identity as projection function ϕ). We represent this projected data as: $\Phi = \phi(\mathbf{X}) = [\phi(\mathbf{x}_1), \phi(\mathbf{x}_2), \dots, \phi(\mathbf{x}_N)]$. Under the assumption that endmembers are invariant by transformation in the feature space, the endmember extraction therefore can be written as finding an indicator matrix $\mathbf{W}^{(\ell)} \in \{0, 1\}^{N \times \ell}$ where:

$$\phi(\mathbf{X}\mathbf{W}^{(\ell)}) = \Phi\mathbf{W}^{(\ell)}. \quad (3.1)$$

This gives rise to a simplicial convex hull in \mathcal{H} , noted $\Delta^\ell(\phi(\mathbf{E}))$, whose pre-image is a non-simplicial hull (the manifold hull) in the original space. The procedure first extracts the best 1-simplex Δ^1 by finding two instances from Φ that results in the minimum projection error on this basis, then iteratively increases the dimension of the simplex in \mathcal{H} . At the (p) -th iteration the algorithm selects $\phi(\mathbf{x}_i)$ such that the volume of the Δ^{p-1} -simplex is maximized, i.e.

$$i = \arg \max_q Vol(\Delta^{p-1}(\Phi\mathbf{W}^{p-1}) \cup \phi(\mathbf{x}_q)), \quad (3.2)$$

where the volume of Δ^p simplex in this iteration is [39]:

$$\begin{aligned} Vol(\Delta^{p-1}(\Phi \mathbf{W}^{p-1}) \cup \phi(\mathbf{x}_q)) = \\ \frac{Vol(\Delta^{p-1}) \times dist(\phi(\mathbf{x}_q), \Phi \mathbf{W}^{p-1})}{p-1}. \end{aligned} \quad (3.3)$$

In practice the optimal projection ϕ enabling to separate complex data is unknown and potentially of high dimension. Hopefully, thanks to the kernel theory [97], one can rely on kernel functions $k(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$ that allow to rewrite the relation $dist(\phi(\mathbf{x}_q), \Phi \mathbf{W}^{p-1})$ in a closed form:

$$dist(\phi(\mathbf{x}_q), \Phi \mathbf{W}^{(p-1)}) = 1 - (k_{x_q}^T \cdot K_{p-1}^{-1} \cdot k_{x_q}), \quad (3.4)$$

with K_{p-1}^{-1} is the inverse of the kernel matrix of the elements of the manifold:

$$K_{p-1}^{-1} = (\mathbf{W}^{(p-1)\top} \Phi^\top \cdot \Phi \mathbf{W}^{(p-1)})^{-1} \quad (3.5)$$

and k_{x_i} is a vector of length p defined as $k_{x_i} = [k(\mathbf{x}_i, \mathbf{x}_j)]_{\mathbf{x}_j \in \mathbf{X} \mathbf{W}^p}$. It is worth noting that instead of inverting the kernel matrix, in the implementation, for the sake of more numerical stability we chose to apply a Cholesky factorization on the kernel matrix at each iteration and then solve for it.

As for the choice of the kernel function, working on the original space is equivalent to choose k as the usual dot product : $k(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle$. Choosing a different function enables to perform the process in a feature space where the geometry of the manifold embedding \mathbf{X} can be more meaningful. Among existing kernels, the standard Gaussian Radial Basis Function (RBF) is widely used since it enables a projection in an infinite dimensional space :

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp \left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2} \right). \quad (3.6)$$

The overall procedure enables to extract endmembers by finding the optimal manifold hull composed of ℓ spectra in the feature space. This exact simplex volume maximization procedure is the cornerstone of endmember estimation in SAGA [39]. In most methods assuming a linear mixing model, the number of endmembers is exactly defined by the number of pure spectra present in the image, and cannot exceed the spectral dimension. As such, choosing a point which is already a mixture of those pure spectra would lead to a nul increase in the volume of the simplex. In the RKHS corresponding to the Gaussian Kernel, each elements of Φ are orthogonal, and therefore would lead to a positive increase of the volume. In practice, it is therefore possible to have more endmembers than the number of pure materials (and can even exceed the spectral dimension). From this point of view, the process

is a dictionary learning procedure, with an overcomplete dictionary (in the sense of the original spectral space). Apart from handling implicitly the non-linearity of the mixing model, this leads to two important properties of our method:

- The spectra can be decomposed over elements that are also combination of pure elements. This can ease the interpretation of the unmixing for specialists through the obtention of more physically realistic spectra, and alleviate to some extent the purity assumption which states that at least one pixel of each material should be present in the image;
- Choosing an overcomplete endmember set can help in handling the intra-class variability of each element, as similar versions of the same material (but with different spectral signatures, as it is the case with shadows for instance), can be efficiently exploited.

In practice, this also requires some sparsity assumption in the mixing coefficients, which is guaranteed by our projection method (see Section 3.2.3).

3.2.2 Anomaly Detection

As mentioned above, SAGA procedure selects endmembers in a way to increase the volume of the simplex hull in the feature space. According to this strategy, in each iteration, extremal points of the dataset hull are chosen. This criterion makes the algorithm vulnerable to anomalies since they are most likely to appear in the construction of our dictionary \mathbf{E} . This issue is illustrated in the two first rows of Figure 3.2. In this figure, all instances of the first row are generated from a uniform Dirichlet distribution within the 2-simplex identified by blue vertexes (note that these pure points are not included in the dataset). In the second row 10 anomalies have been added to the nominal data. On the right column, the corresponding endmembers extracted by SAGA are depicted (red dots) using both datasets. From this simple experience, it is visible that endmember extraction algorithm fails to identify the correct borders of the nominal data in presence of anomalies.

As soon as extreme points appear, they significantly contribute in maximizing the simplex volume and are thus selected as endmembers. This behavior is observable in all unsupervised endmember extraction procedures which work based on maximization of the geometrical volume of the simplex encompassing data. To cope with this issue, the proposed anomaly detection (AD) procedure relies on the computation of the reconstruction error (in feature space) between a data sample and its perpendicular projection onto a side of the polygon of the simplex. Although similar ideas have already been proposed (see [86] and other AD algorithms), the process here does no attempt for identifying the nominal distribution (for example based on decomposition of the eigen-structure of dataset) but we rather exploit the iterative structure of SAGA where in step p , after selection of the $p + 1$ -th endmember candidate through equation (3.2), we check its validity by computing the

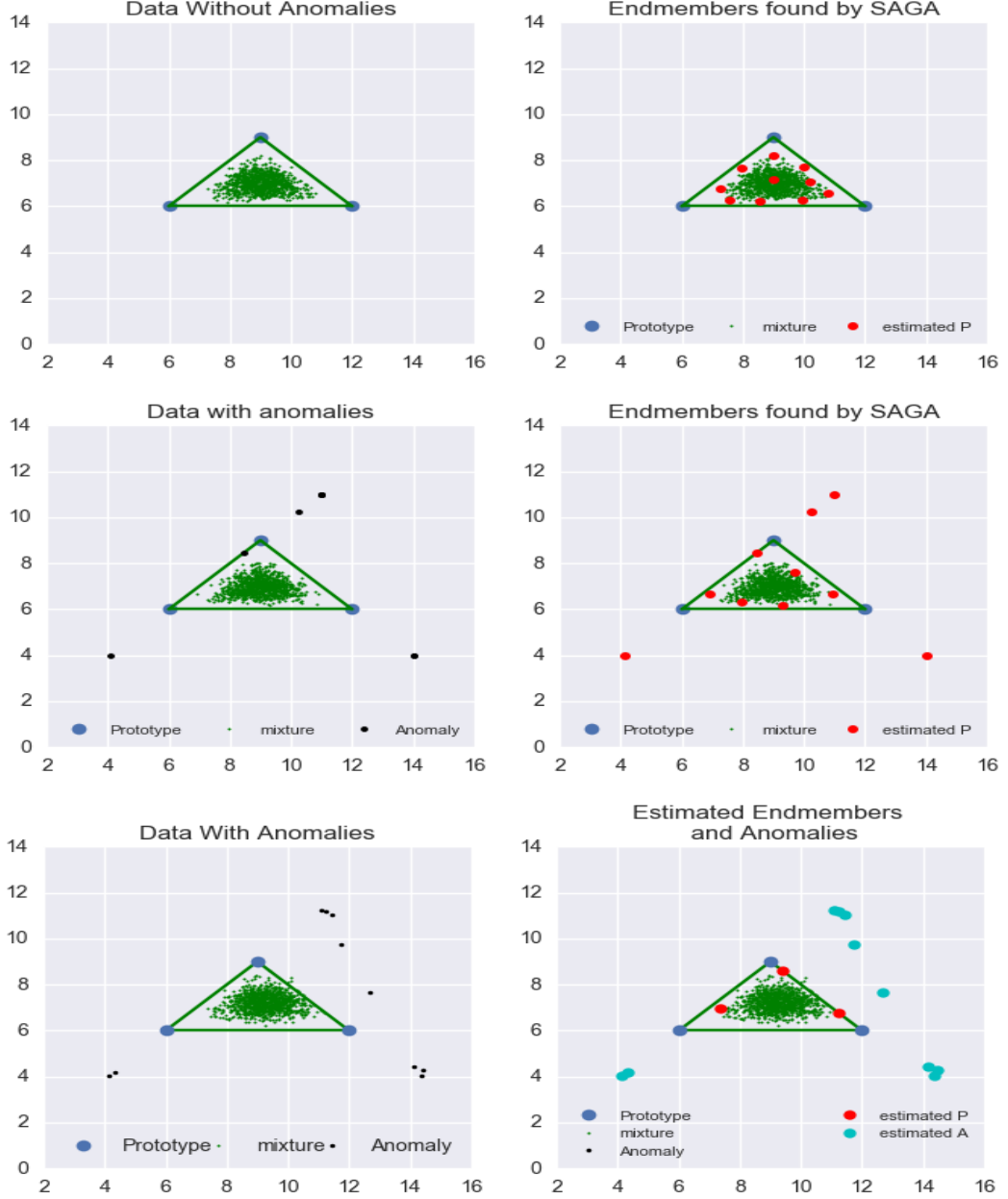


Figure 3.2: Illustration of difficulties in presence of outliers. First Row: end-member extraction using SAGA without anomalies: in this situation extracted endmembers are a reliable representation of the manifold ; **Second Row: endmember extraction using SAGA in presence of anomalies:** in this situation extracted endmembers (in red) do not correspond to real borders of the data manifold ; **Third Row: endmember extraction and anomaly detection In presence of anomalies using SAGA+:** here extracted endmembers (in red) by are in accordance with borders of the manifold embedding data while the anomalies (in cyan) are properly selected.

sum-of-projection-errors (SOPE) of all instances with respect to the new simplex Δ^{p+1} :

$$\text{SOPE}(p+1) = \frac{\left[\sum_{i=1}^N 1 - (k_{x_i}^T \cdot K_{p+1}^{-1} \cdot k_{x_i}) \right]}{N}, \quad (3.7)$$

where N is the number of samples and acts as the normalization constant. A perfect reconstruction gives $\text{SOPE}(p+1) = 0$, whereas its value increases when the quality of the reconstruction decreases (i.e. the simplex is not a good representation of the dataset). During the iterative procedure, when a considerable drop of SOPE is perceived after adding a new vertex to the simplex, this means that previously selected prototypes (endmember candidates of previous steps) could be anomaly and are among instances significantly different from nominal data. By this procedure one can then identify and reject them. In the third row of Figure. 3.2 one can see that this AD procedure, even when data are exposed to anomalies, endmember estimation remains close to a good representation of the data manifold which is in contrast to the second row of Figure.3.2 (without any AD process).

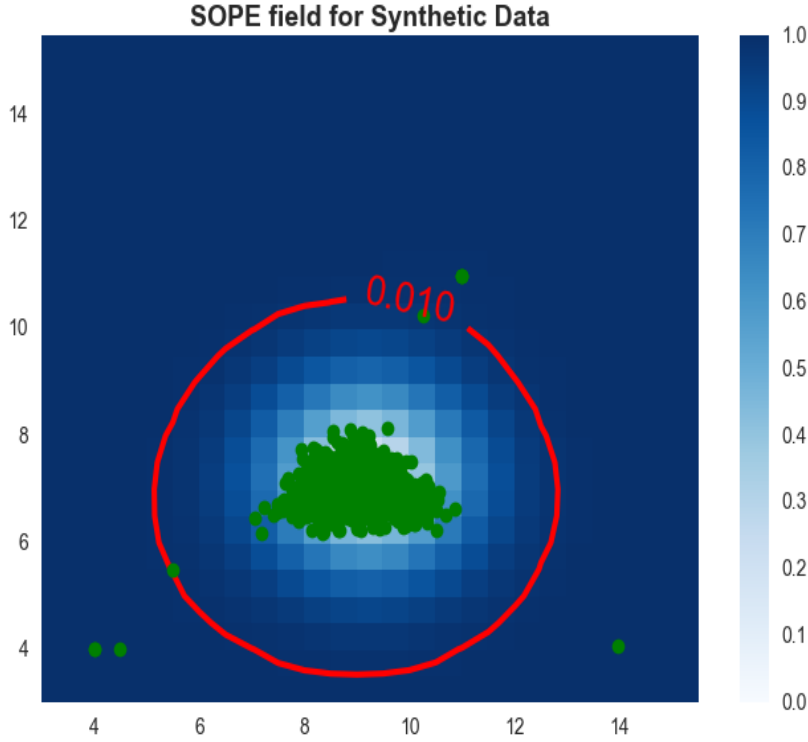


Figure 3.3: Evolution of SOPE (Sum Of Projection Errors) when all data in green are represented through a single endmember uniformly represented on a meshgrid. One observes a decrease in SOPE when candidates get closer to the nominal distribution.

In order to be able to check this drop in each iteration we need to define a threshold τ . This point is not trivial and highly depends both on the distribution of nominal data and the distribution of anomalies. To illustrate this, Figure. 3.3 represents the evaluation of SOPE when all nominal data (green points) are represented on a unique candidate issued from a meshgrid. From this figure, one easily observes that error decreases when endmember candidates get closer to the center of the data distribution². The iso-contour corresponding to $\tau = 0.01$ in SOPE value is also represented to evaluate the associated threshold that enables to remove outlier data. A discussion related to the practical selection of τ is proposed in section 3.3.

Once endmembers are identified (together with anomalies), the projection of each data point onto the new basis has to be performed. This process is discussed in the following.

3.2.3 Sparse Projections onto the Simplex

Let us recall that the unmixing process aims at approximating the data matrix \mathbf{X} through $\mathbf{X} \approx \mathbf{E}\mathbf{G}$, where the estimation of the endmember matrix \mathbf{E} has been presented above. As for the abundance matrix \mathbf{G} , it corresponds to the projection of \mathbf{X} onto a region bounded by \mathbf{E} . This is done, for each row of \mathbf{G} (each row being represented as $\mathbf{G}_{\bullet,i}$ associated to the i th point), with a projected gradient decent :

$$\mathbf{G}_{\bullet,i} = \arg \min_{\mathbf{G}_{\bullet,i}} \|\phi(\mathbf{x}_i) - \Phi \mathbf{W} \mathbf{G}_{\bullet,i}^\top\|^2_{s.t.} \mathbf{G}_{\bullet,i} \in \Delta^\lambda. \quad (3.8)$$

where Δ^λ is the sub-simplex included in Δ^ℓ but composed only with λ vertexes. We indeed want to impose sparsity on each column of \mathbf{G} since in practice, a pixel is likely to be composed only with a subset of existing endmembers. Explicitly, imposing sparsity prevents from solutions where all endmembers have a (even small) contribution. To this end, a greedy selector and sparse projector (**GSSP**)[98] method is used not only to project the data onto the simplex but also to impose a sparsity level (through a number of non null elements given as an input parameter). In practice the coordinate values of each pixel will be sorted based on their magnitude and the λ greatest ones will be projected onto a unit simplex while the rest of the values will be set to zero.

We now have all ingredients for the complete process of anomaly detection and unmixing which is being presented.

²In order to illustrate more into details, various maps of SOPE fields along the endmember extraction process are represented in the supplemental materials provided in the appendix 7.

3.2.4 Overall process

The SAGA+ procedure is presented in Algorithm 1. In a first step we compute the **manifold hull** to derive endmembers together with anomaly detection (lines 4 to 10). To initialize the process, the algorithm selects a random datum \mathbf{x}_i (line 4), and computes the distances between \mathbf{x}_i and all instances in \mathbf{X} (line 5). These distances are then sorted in descending order and will be feed to the Exact Simplex Volume Maximization and Anomaly Detection process to estimate the first endmember $\hat{\mathbf{e}}_i$ associated with a set of anomalies (line 6). This process is repeated ℓ times in order to extract the ℓ endmembers (lines 7 to 9) and to construct the manifold hull \mathbf{E} . In each iteration the volume increase of the simplex is computed in RKHS w.r.t each \mathbf{x}_i . Following equations (3.2)–(3.5), this requires to inverse the kernel matrix \mathbf{K}^{-1} which is done in practice using the Cholesky decomposition. After computation of the volume changes, the anomaly detection algorithm is called to check the trade-off between the increase of the volume of the simplex and the decrease in the sum-of-projection-errors (SOPE). This is represented in Algorithm 2.

Once the manifold hull extracted, the projection matrix \mathbf{G} is calculated using the Greedy Selector and Sparse Projector (GSPP) algorithm. This second step of SAGA+ follows the process described in [98]. The principle of Exact Simplex Volume Maximization and Anomaly Detection algorithm is described in Algorithm 2. Here we check iteratively, on the basis of the SOPE criteria presented in equation (3.7), if successive candidates for the simplex vertexes (sorted based on their distances with the current simplex) are real endmembers or lie among outliers. SOPE is computed by taking into account the presence of already found prototypes and in case no endmembers are already present, the sum of projection error reads $SOPE = [(\boldsymbol{\nu}^\top \cdot \boldsymbol{\nu})/N] - 1$ where $\boldsymbol{\nu}$ represents the vector of the similarities between the current candidate and the rest of the dataset instances.

Algorithm 1 SAGA+

1: **Input:** Data matrix \mathbf{X}

2: **Parameters:**

- Number of required endmembers ℓ
- Threshold τ required for SOPE (see section 3.2.2)
- Sparsity level λ for abundance matrix \mathbf{G}
- Kernel bandwidth σ

3: **Output:**

- Endmember matrix \mathbf{E}
- Abundance matrix \mathbf{G}
- Anomaly matrix \mathbf{A}

▷ **Step 1:** Endmember Estimation (EE) and Anomaly Detection (AD):

4: - Select a random point \mathbf{x}_i .

5: - Compute distances w.r.t \mathbf{x}_i in RKHS (see eq. (3.4)–(3.6)).

6: - Identify the first endmember and first outliers using Exact SiVM and AD algorithm (see description in Figure 2)

7: **for** ($p = 2, \dots, \ell$) **do**

- From current simplex with $p-1$ endmembers, compute volume changes with all instances (see eq. (3.3)).
- Sort the vector of volume changes decreasingly (see eq. (3.2)).
- Since first indexes are possible anomalies, SAGA+ to identify endmembers and outliers (see algorithm in Figure 2)

8: **end for**

9: **return**(\mathbf{A}, \mathbf{E})

10: ▷ **Step 2:** Abundance estimation via sparse projections

11: - Apply eq.(3.8) and extract matrix (\mathbf{G})

12:

13: **return** (\mathbf{G})

Algorithm 2 Exact SiVM and AD

Input:

- Data matrix \mathbf{X}
- Current endmembers $\hat{\mathbf{E}}$
- Current outliers \mathbf{A}
- Vector of Sorted indexes \mathbf{vs}

2:

Output:

- New Endmembers $\hat{\mathbf{e}}$
- New Anomalies \mathbf{A}

```
4: flag = true (Indicator of anomaly)
   i = 0
6: while flag = true do (Loop on all points)
   if ( $\mathbf{vs}_i \notin \{\mathbf{A} \cup \hat{\mathbf{E}}\}$ ) then
8:     Compute SOPE on  $\mathbf{X}[\mathbf{vs}_i]$  Eq (3.7)
   if ( $\text{sope} < \tau$ ) then
10:     flag = false
   else
12:      $i+ = 1$ 
        $\mathbf{A}+ = \mathbf{x}_i$ 
14:   end if
   end if
16: end while
   return ( $\mathbf{A}, \hat{\mathbf{E}} + \mathbf{x}_i$ ) Index of the next endmember
```

3.3 Evaluation

In this section we evaluate on synthetic and real data, SAGA and SAGA+ procedures and compare their performances with a number of state-of-the-art unmixing procedures. In particular we consider two linear unmixing methods N-Finder [51] and VCA [5] introduced above. As mentioned, these are frequently used unsupervised methods of unmixing and their endmember extraction procedures are based on the simplex volume maximization method. This makes them suitable to be compared with the current method. It should be mentioned however that these models assume linear mixing and do not detect outliers, therefore the comparison with our approach would be thought to some extent unfair. Nevertheless it should also be emphasized that they all follow the same logic of extending the encompassing simplex which makes them (to some degree) comparable to the logic behind SAGA. We also decided to compare our approach with G-SiVM [6] which is a non-linear procedure based on simplex volume maximization through shortest-path distances in a nearest-neighbor graph. Non-linearity is performed using kernel trick which makes this approach similar to SAGA and SAGA+ in the sense that they do not

rely on any specific mixing assumption. In order to evaluate and compare the efficiency of the procedures, we use different criteria to assess endmember extraction and anomaly detection processes. These criteria are introduced below together with experimental setups.

3.3.1 Experimental setups

Evaluation criteria

for synthetic data where ground truth is available, a set of criteria are used to measure our performances:

- To evaluate the adequacy of **estimated endmembers** $\hat{\mathbf{E}} = [\hat{\mathbf{e}}_1, \dots, \hat{\mathbf{e}}_\ell]$ with **ground truth endmembers** $\mathbf{E} = [\mathbf{e}_1, \dots, \mathbf{e}_\ell]$, several possibilities are offered as Spectral Information Divergence (SID) or RBF kernel distance, each of which give a measure of similarity between two signals represented as a vector. Though efficient, these criteria are sensible to the difference of endmembers in terms of magnitude while in this application, we are more interested in comparing their shape. Indeed, because of atmospheric attenuation or variation in illumination, spectral energy can be attenuated for two samples even if they belong to the same class. Thus, we rather prefer to rely on Spectral Angle Map (SAM) which is scale invariant. The SAM reads, for spectral vectors $\mathbf{x}_i = [x_{i1}, \dots, x_{iD}]^T$ and $\mathbf{x}_j = [x_{j1}, \dots, x_{jD}]^T$ with positive values:

$$\text{SAM}(\mathbf{x}_i, \mathbf{x}_j) = \cos^{-1} \left(\frac{\vec{\mathbf{x}}_i \cdot \vec{\mathbf{x}}_j}{\|\vec{\mathbf{x}}_i\| \|\vec{\mathbf{x}}_j\|} \right) \quad (3.9)$$

and lives in the range $[0, 1]$ (1 representing maximum dissimilarity and 0 indicating two identical signals). In practice in order to compare two different sets of endmembers, we need a pairing procedure to associate each estimated endmember to a unique real one. This is done as finding the best permutation of estimated endmembers that gives the minimal average SAM values when they are compared to real ones. Therefore, two sets \mathbf{E} and $\hat{\mathbf{E}}$ have a difference $\epsilon(\mathbf{E}, \hat{\mathbf{E}})$ which reads:

$$\epsilon(\mathbf{E}, \hat{\mathbf{E}}) = \min_{\eta} \frac{1}{\ell} \sum_{i=1}^{\ell} \text{SAM}(\mathbf{e}_{\eta^\ell(i)}, \hat{\mathbf{e}}_i) \quad (3.10)$$

where ℓ is the number of endmembers and $\eta^\ell : \{1, \dots, \ell\} \rightarrow \{1, \dots, \ell\}$ is a function corresponding to all possible permutations.

- As for the evaluation of the anomaly detection algorithm, we rely on Cohen's Kappa Statistics [99] to measure the adequacy of the classification between outliers/endmembers. Kappa criterion, noted κ , lies between 0 (nothing is

in accordance) and 1 (perfect accordance) and combines False Positive (FP), False Negative (FN), True Positive (TP) and True Negative (TN) in order to measure the inter-rater agreement of categories.

- In complement we also give an execution-time comparison among the end-member extraction procedures.

We now introduce our datasets.

Synthetic data

In order to generate synthetic datasets, we have selected 41 spectral signatures from RELAB spectral library [93]³. From ℓ spectral signatures $\mathbf{E} = [\mathbf{e}_1, \dots, \mathbf{e}_\ell]$ ($\mathbf{e}_i \in \mathbb{R}^D$), mixed spectra have been generated based on three different mixing assumptions: Linear Mixing Model (LMM), Bilinear Mixing Model (BMM) and Highly Concentrating Model (HCM). These mixing models are well studied in the unmixing literature, though they do not fully represent the variability of mixing processes in real world hyperspectral images. The two first models (LMM, BMM) are implemented according to the procedure in [4]. The usage of HCM model is also mentioned in some earlier works[100]. From the set of endmembers \mathbf{E} , the **Linear Mixture Model** reads :

$$\mathbf{x}_i = \mathbf{E}^T \mathbf{g}_i + \mathbf{w}_i \quad (3.11)$$

where \mathbf{g}_i is the abundance vector related to data \mathbf{x}_i and \mathbf{w}_i is a random noise. Each element in \mathbf{g}_i lives in the range $[0, 1]$ and their sum equals 1. The fractional abundances in LMM have been generated based on Multivariate Symmetric Dirichlet Distribution, uniformly distributed over the standard $\Delta^{\ell-1}$ simplex. More precisely for each sample \mathbf{x}_i^\top , $i = 1, \dots, N$ and $\mathbf{x}_i \in \mathbb{R}^D$ the abundance vector $\mathbf{g}_i = [g_1, \dots, g_\ell]$ is generated through a Dirichlet Distribution. This latter, parametrized by vector $\alpha = [\alpha_1, \dots, \alpha_\ell]$ such that $\alpha_i > 0, \forall i \in [1, \ell]$, reads:

$$\mathbf{g}_i = \frac{\Gamma(\alpha \cdot \ell)}{\Gamma(\alpha)^\ell} \prod_{j=1}^{\ell} \mathbf{e}_j^{\alpha_j - 1} \quad (3.12)$$

where $\Gamma(\cdot)$ is the Gamma function. Dirichlet are generalization of the beta distribution in a multi-variate context. Such distributions are then very useful to simulate mixture models. From equation (3.12), using a parametrization $\alpha = [\alpha_1, \dots, \alpha_\ell]$, we generate mixtures of pure endmemners $[\mathbf{e}_1, \dots, \mathbf{e}_\ell]$ by computing abundance parameters $\mathbf{g}_i = [g_1, \dots, g_\ell]$ as the power-product between endmembers and α . This is written in the right part of equation (3.12). The left part is only a normalization coefficient that involves the Gamma function. In practice α is set as the vector of ones: $\alpha_i = 1, \forall i$. As the result the vector \mathbf{g}_i contains coefficients which are uniformly

³These signatures together with additional information regarding them can be seen in the supplement material provided in the appendix 7.

distributed across the ℓ vertexes of the simplex. To illustrate this, from pure endmembers visible on the top of Figure 3.4(a), blue points of Figure 3.4(b) represent a 2D slice (band 70 vs. band 100) of the synthetic data generated under the LMM model.

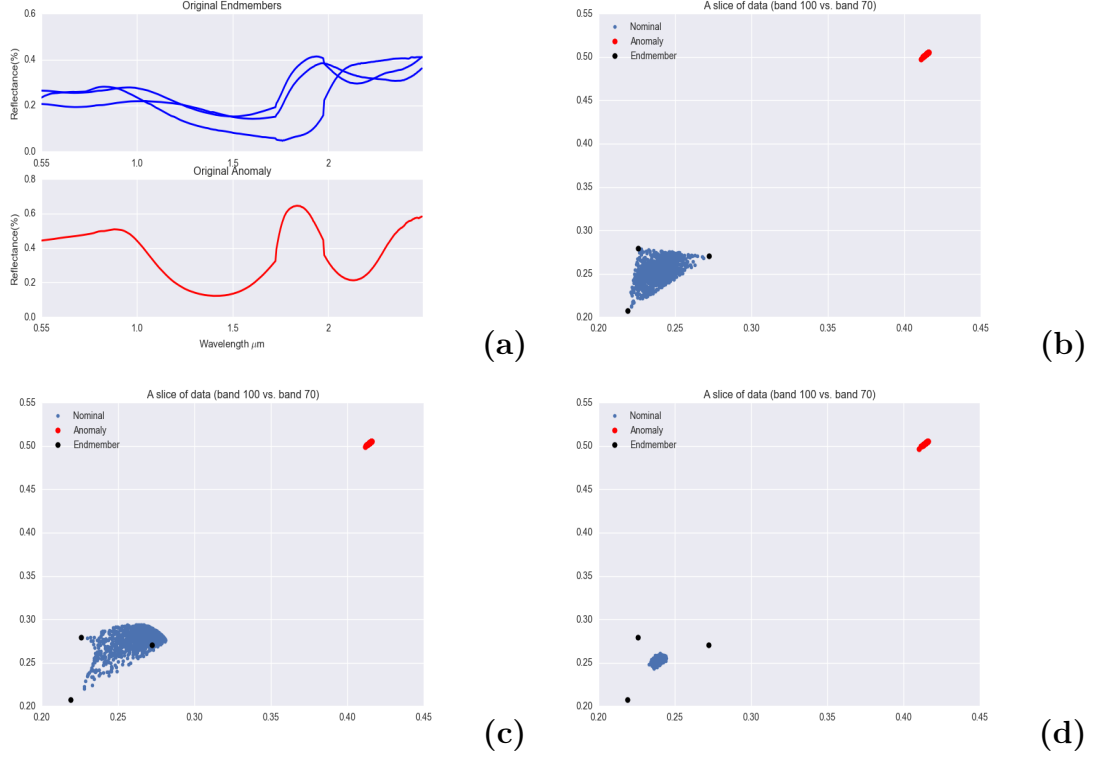


Figure 3.4: Illustration of synthetic data. (a): Some endmembers (top) and the anomaly spectrum (bottom) used to generate the datasets. (b): A 2D Slice of two components of a synthetic dataset of 1010 samples generated based on LMM from the 3 nominal endmembers and 10 samples issued from the single anomaly signature. The geometrical location of the anomalies in the 2D plan makes them visually separable. (c): Similar visualization but mixtures are generated on the basis of BMM. (d): Similar visualization but mixtures are generated on the basis of HCM.

For **Bilinear Mixing Model**, the power products of reflectance is added to the linear mixing model, leading to a model of the form :

$$\mathbf{x}_i = \underbrace{\mathbf{E}^T \mathbf{g}_i}_{\text{LMM}} + \sum_{k=1}^{\ell-1} \sum_{l=k+1}^{\ell} \beta_{k,l,i} \mathbf{e}_k \odot \mathbf{e}_l + \mathbf{w}_i \quad (3.13)$$

where $\beta_{k,l,i}$ are new mixing coefficients and the termwise Hadamard product is defined as:

$$\mathbf{e}_k \odot \mathbf{e}_l = \begin{pmatrix} e_{1,k} \\ \vdots \\ e_{D,k} \end{pmatrix} \odot \begin{pmatrix} e_{1,l} \\ \vdots \\ e_{D,l} \end{pmatrix} = \begin{pmatrix} e_{1,k}e_{1,l} \\ \vdots \\ e_{D,k}e_{D,l} \end{pmatrix}. \quad (3.14)$$

From the same pure endmembers on the top of Figure 3.4(a), blue points of Figure 3.4(c) represent the same 2D slice generated under the BMM model.

As for the **Highly Concentrating Model**, we choose to rely on the same Dirichlet Distribution as in LMM, however with large concentration parameters $\alpha_i \gg 1$ (in practice we used $\alpha_i = 50, \forall i$). This ensures that generated abundance coefficients \mathbf{g}_i are highly concentrated in the middle of the simplex which is formed by the endmembers. This is visible in blue points of Figure 3.4(c).

In order to generate and include a number \mathcal{V} of outliers to the nominal synthetic data, we used three spectral signatures $\mathbf{E}' = [\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3]$ without any similarities to the signatures used to generate nominal data. From ℓ known endmembers in \mathbf{E} and our abnormal spectral signatures in \mathbf{E}' , we use again Dirichlet distribution function to generate fractional abundances for anomalies \mathbf{g}'_i using three more parameters $\alpha' = [\alpha'_1, \alpha'_2, \alpha'_3]$. The generation function reads:

$$\mathbf{g}'_i = \frac{\Gamma(\sum_{j=1}^{\ell+3} \gamma_j)}{\prod_{j=1}^{\ell+3} \Gamma(\gamma_j)} \prod_{j=1}^{\ell+3} \mathbf{e}_j^{\gamma_j-1} \quad (3.15)$$

where $\gamma = [\alpha_1, \dots, \alpha_\ell, \alpha'_1, \alpha'_2, \alpha'_3]$ is the concatenation of parameters α and α' . In practice, we used $\alpha_1, \dots, \alpha_\ell = 1$ and $\alpha'_1, \alpha'_2, \alpha'_3 = 50$ to simulate outliers. These latter are depicted in red in Figures 3.4(b-c-d) and an example of anomaly is visible in the bottom of Figure 3.4(a). As shown on these figures, outliers are lying in a separable distance from the 2-simplex encompassing the nominal data (note that in practice outliers may not necessarily live in a similar area but we prefer to use this simple situation for validation).

In further experiments on synthetic data we use different datasets with various sizes ($N = \{10^3, 10^4, 10^5, 10^6\}$) and endmembers ($\ell = \{3, 5, 7, 10, 12, 15\}$). As for the evaluation of anomaly detection procedures, for all datasets we choose to generate $\mathcal{V} = 20$ anomalies from three different signatures.

Real data

For the application on real data we choose in this study to focus on 4-Vesta asteroid which is the second largest asteroid in the solar system and thus is of prime scientific interest in planetary exploration. 4-Vesta is thought to be a protoplanet, i.e. a building block of terrestrial planets remnant of the early formation of the Solar System [101]. Therefore understanding its properties is of prime importance to understand the initial material that accreted to form the Earth ⁴. The number "4" in the name of the asteroid is due to the fact that 4-Vesta was the fourth asteroid to be discovered. It was observed recently by the Dawn mission [102], providing high resolution hyperspectral images of the whole surface. Moreover a very

⁴In astrophysics, accretion is the growth of particles into a massive object by gravitationally attracting more matter, typically gaseous matter and dust, in an accretion disc.

large number of meteorite samples in our collection come from this asteroid [103], providing additional constraints on the minerals mixtures possibly encountered when processing the HS dataset. As the dataset available for 4-Vesta is concerned by all the limitations described in section 3.1, the use of an improved non-linear unmixing algorithm is essential to progress from basic indexes maps (as done up to now, e.g. [104]) to detailed maps, making the link between meteorites and compositions extracted from remote sensing possible. Because of the large size of the full dataset (as mentioned in section 3.1), here we choose to focus on a small subset located in the northern hemisphere. This region was chosen because Olivine, a mineral indicative of deep magmatic processes, was first detected in this area [105]. This mineral is never found alone in meteorites coming from Vesta-4 [106, 107], implying the use of an algorithm able to decompose spectra as a combination of mixtures of materials.

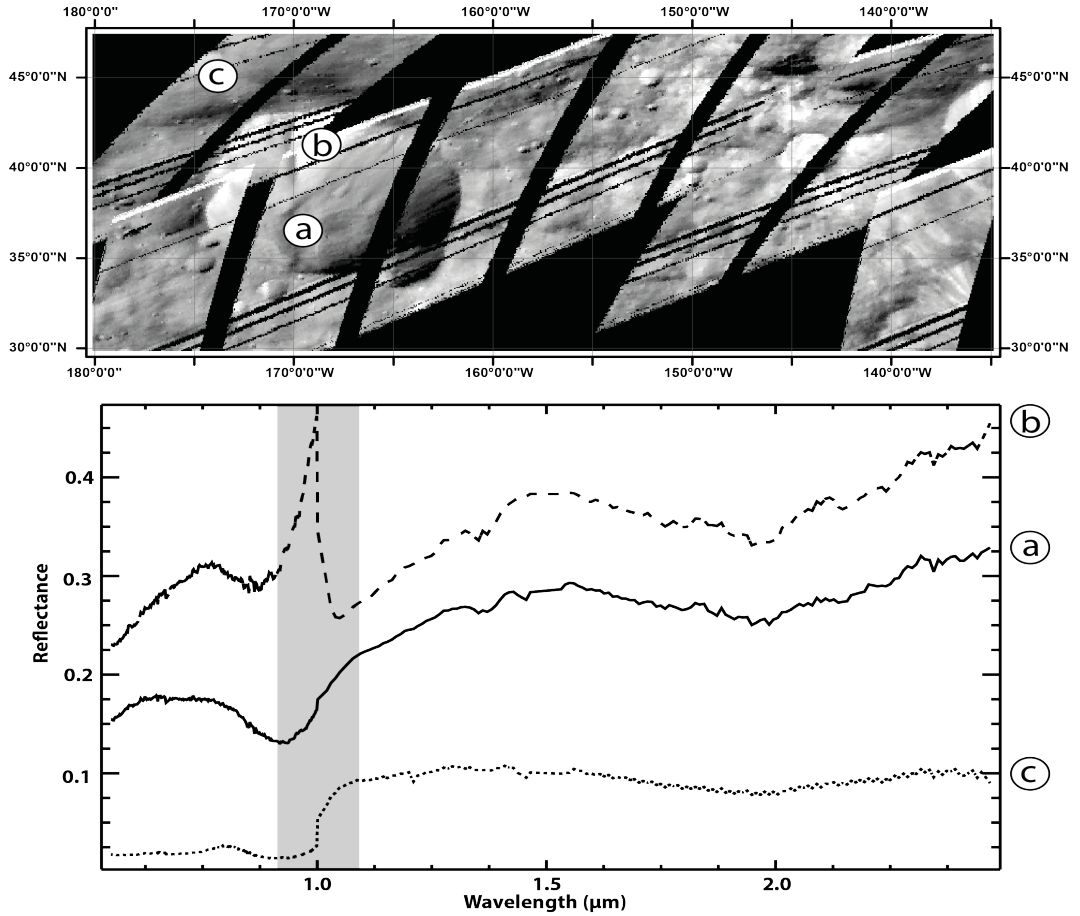


Figure 3.5: 4-VESTA mosaic composed of (842×327) pixels, with 383 spectral bands covering the range $0.55\mu m$ to $2.47\mu m$. Examples of spectra (located on the mosaic with letters) are reported with (a) a normal spectrum, (b) discontinuity at the two detectors boundary and (c) a spectrum at the boundary of a shadow region. The gray wavelength range is where inconsistency are usually observed.

Following the procedure described in [108], we processed Dawn VIR [109] images from the High Altitude Mapping Orbit (HAMO) 1 and 2 and from the Low Altitude Mapping Orbit (LAMO). Raw images were downloaded from PDS in level 1B and were calibrated using ISIS3 pipeline [110] and photometric parameters found in the literature [111, 112, 113]. Additional in-house routines were used to filter bad pixels and to correct the geographical misalignment between the two detectors of the VIR instrument (following the method described in [104]). Finally each image was converted from radiance to I/F and projected using the spacecraft geometry information. Such process remains standard in planetary image analysis.

After removing channels known to be noisy, the final mosaic has a size of (842×327) pixels, with 383 spectral bands covering the range $0.55\mu m$ to $2.47\mu m$. Nevertheless, despite the first-order filtering of bad pixels/spectra, some inconsistency still exist in the dataset and outliers can be observed, as illustrated in Figure 3.5. In this figure the typical shape of an expected spectra is depicted together with highly corrupted spectra due to the misalignment between the visible and near-infrared detectors, either showing a sharp peak when pixels fall on two distinct lithological units or an Heaviside-step shape signature when pixels fall at a boundary between sunlit and shadow.

Parameters setting for SAGA and SAGA+

To guarantee an optimal performance, internal parameters of SAGA and SAGA+ have to be tuned according to dataset characteristics. Apart from the number of atoms/endmembers ℓ and the sparsity level λ , these parameters correspond to kernel bandwidth (σ) and for SAGA+, to the error threshold (τ) of the anomaly detection procedure (see algorithm in Figure 1). For this reason a number of tests have been taken in order to find the best parametrization of the models. As the result of these tests, we observed empirically that the best parameter values for linear mixtures are $\sigma = 0.15$ and $\tau = 0.2$, for bilinear mixtures : $\sigma = 0.15$ and $\tau = 0.04$ and for highly concentrating mixtures : $\sigma = 0.15$ and $\tau = 0.2$. This setting enabled us to detect all true anomalies (true positives) through each run of the algorithm while keeping the number of wrongly selected anomalies (false positives) minimum. These parameters kept fix in all our experiments. It should be noted that these tests are meant to find the best parameters for a model regarding the dataset under consideration and thus not to be thought as generic. For the real application presented later we also use $\sigma = 0.15$ and $\tau = 0.2$ since they are used for LMM and HCM. As will be shown in the next section, this setting produces acceptable results. The question of an optimal parameter selection for real datasets within the unsupervised scenario however remains an open question. We now turn to experiments.

3.3.2 Results on synthetic data

In this part we compare the results from SAGA and SAGA+ with other unmixing procedures on the synthetic datasets, presented in section 3.3.1. We first compare SAGA with N-Finder, VCA, and G-SiVM in order to evaluate their accuracy in endmember estimation in absence of any anomaly. For this we generated various datasets of size $N = 10^3$, using different numbers of signatures $\ell = 3, 5, 7, 10, 12, 15$ and based on the three aforementioned mixing models. Because our approach embeds some randomness in the initialization step, we depict the average of SAM errors computed on 100 runs.

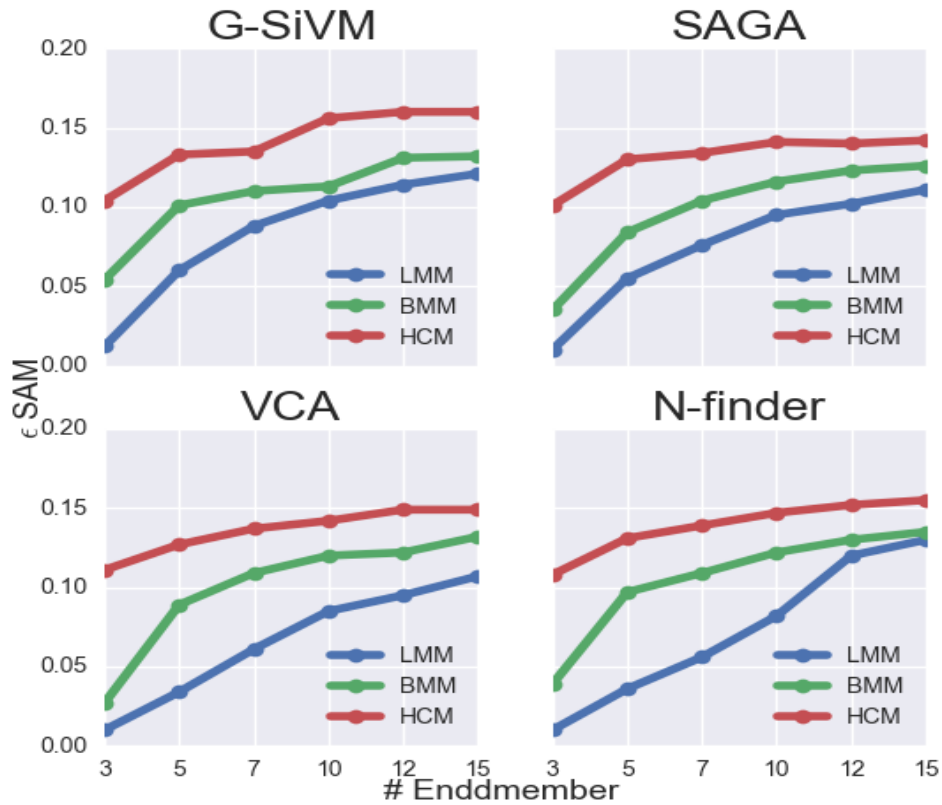


Figure 3.6: Endmember estimation Averages of SAM values for SAGA, G-SiVM, VCA and N-Finder in function of the number of endmembers. Each plot represents the results for three mixing models.

They are visible in Figure 3.6 in function of different number of endmembers for each algorithm. From this figure one observes that SAGA, N-Finder and VCA have similar performances unlike G-SiVMs whose accuracy is generally diminished for the case of intimate mixtures. This first experience where no outliers are involved shows the relative good performances of all techniques to extract consistent endmembers.

In a second experience we ran exactly the same experiments, this time however, we included $\mathcal{V} = 20$ anomalies (generated from 3 extra signatures randomly selected from the dataset as explained in previous section) in the dataset. New results are depicted in Figure 3.7. As expected the overall error increases in all the algorithms except SAGA+ which includes an anomaly detector process. These simple experiments demonstrate the ability of our approach to accurately estimate endmembers while removing outliers.

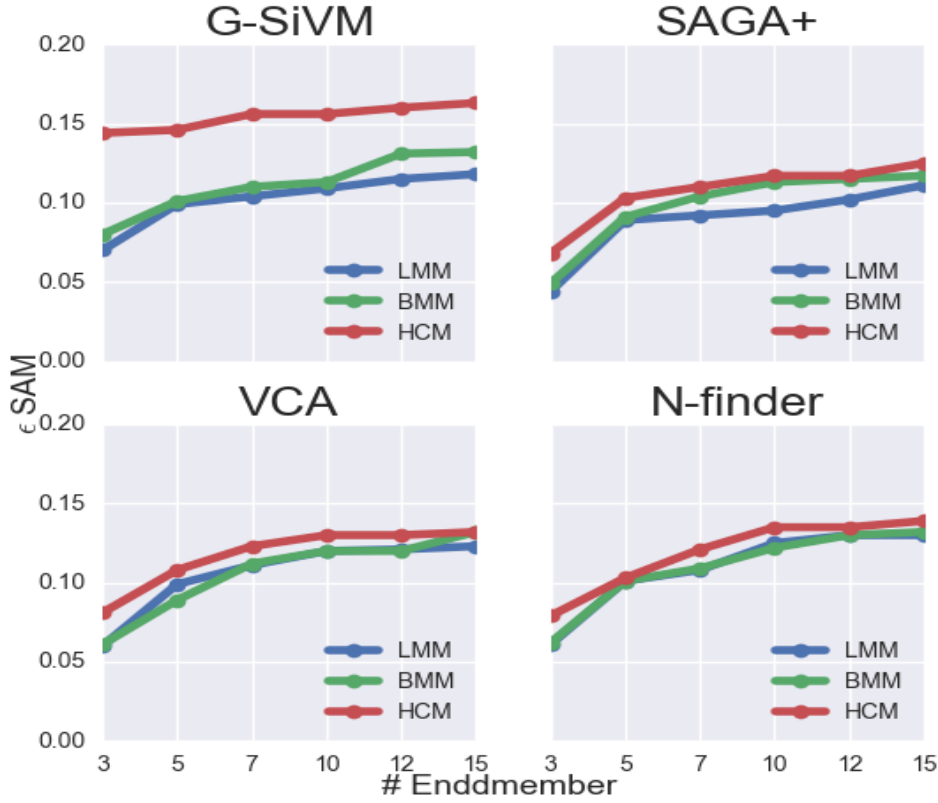
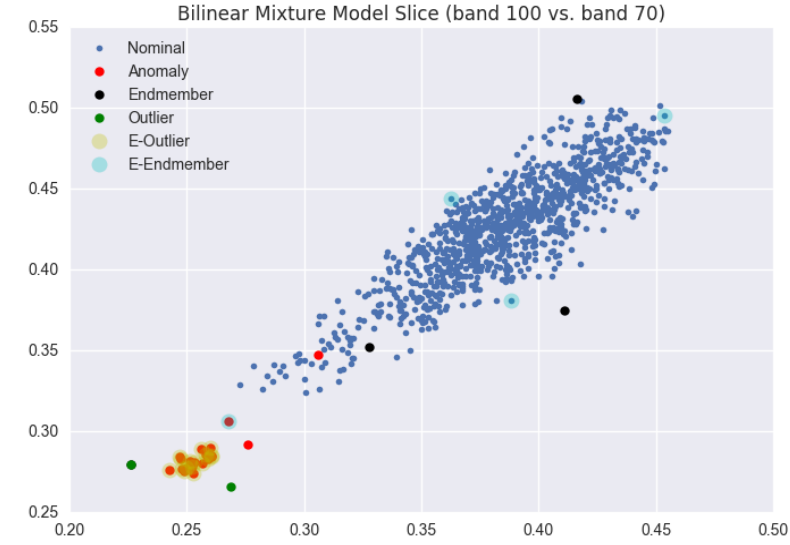


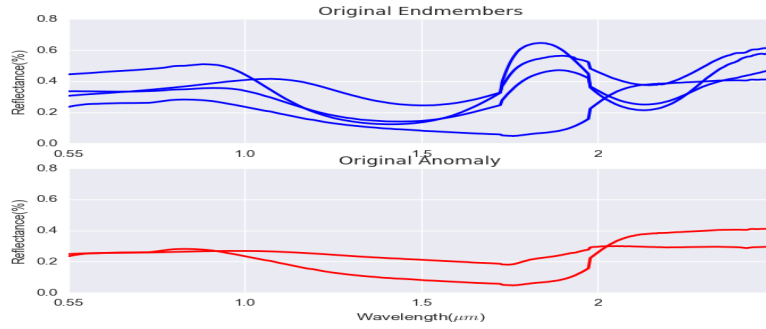
Figure 3.7: Endmember estimation Averages of SAM values for SAGA+, G-SiVM, VCA and N-Finder in function of the number of endmembers. Data contains anomalies generated based on 3 random signatures. Each plot represents the results for three mixing models.

It should be noticed that the 41 signatures used in this setting are not extremely different in their overall profile shape. Nevertheless the results show that the presence of AD even at the situations where the anomalies are not dramatically different (therefore have small distance w.r.t the nominal data) from the nominal signatures can effect considerably. Let us also outline that in the last dataset, outlier signatures were considerably different from the nominal data, making it an ideal condition for the anomaly detection. In order to evaluate the sensitivity of SAGA+ when anomalies are more dispersed w.r.t data, several concentration levels (issued from Dirichlet

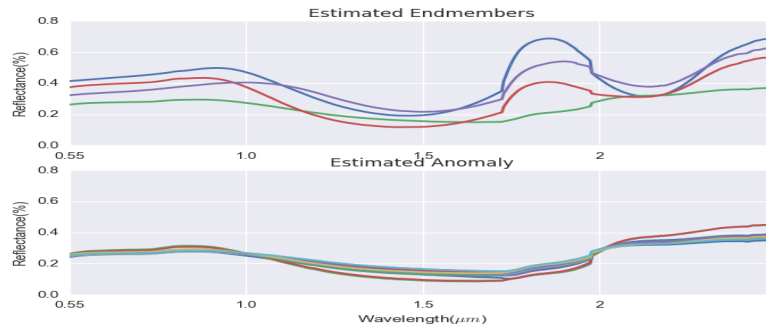
distribution as mentioned in previous section) of anomalies have been generated: the higher the concentration level, the lower the dispersion of anomalies. An example of dataset is shown in Figure 3.8 together with endmember and anomaly estimation with SAGA+. As one observes in this figure, both anomalies and endmembers are accurately detected.



(a)



(b)



(c)

Figure 3.8: Results in presence of anomalies using SAGA+. (a): Representation of the dataset; (b) original endmembers and anomalies (c) estimated endmembers and anomalies with SAGA+

In another test, the anomalies are assigned on the basis of 50 executions, where in each execution, a new concentration level in the range $\alpha' = 1, \dots, 50$ is used to generate anomalies (see equation (3.15)).

Table 3.1: Kappa statistics for SAGA+ in endmember estimation for the three synthetic datasets

	ℓ	3	5	7	9	11	13	15
κ	LMM	0.73	0.84	0.85	0.84	0.84	0.84	0.79
	BMM	0.94	0.94	0.89	0.88	0.85	0.85	0.78
	IMM	0.96	0.96	0.95	0.95	0.94	0.94	0.91

In order to quantitatively evaluate the sensitivity of AD procedure to the level of concentration/dispersion of anomalies. In Table. 3.1 we report the average of Kappa values when comparing true number of anomalies to those detected by SAGA+ over synthetic datasets of size $N = 10^3$, generated with various numbers of endmembers $\ell = 3, 5, 7, 10, 12, 15$ and $\Upsilon = 20$ anomalies generated from three different signatures.

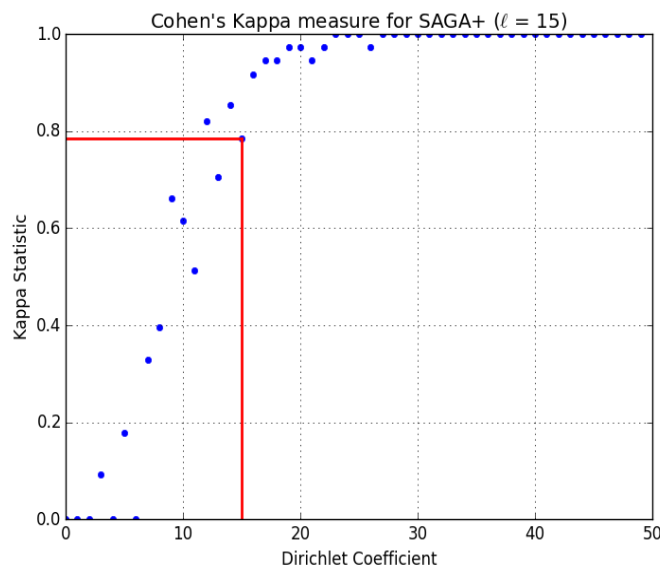


Figure 3.9: Evaluation of SAGA+: Kappa values as the function of Concentration of the anomalies. The red lines represent the threshold where the Kappa value starts to get over 0.8 which is equal to the concentration level $\alpha = 15$.

From this table one observes that when increasing the number of endmembers, the anomaly detection algorithm is still robust since $\kappa > 0.7$ is generally interpreted as a good separation accuracy in two class classification. The way Kappa reacts in function of the concentration of outliers is visually illustrated in Figure 3.9.

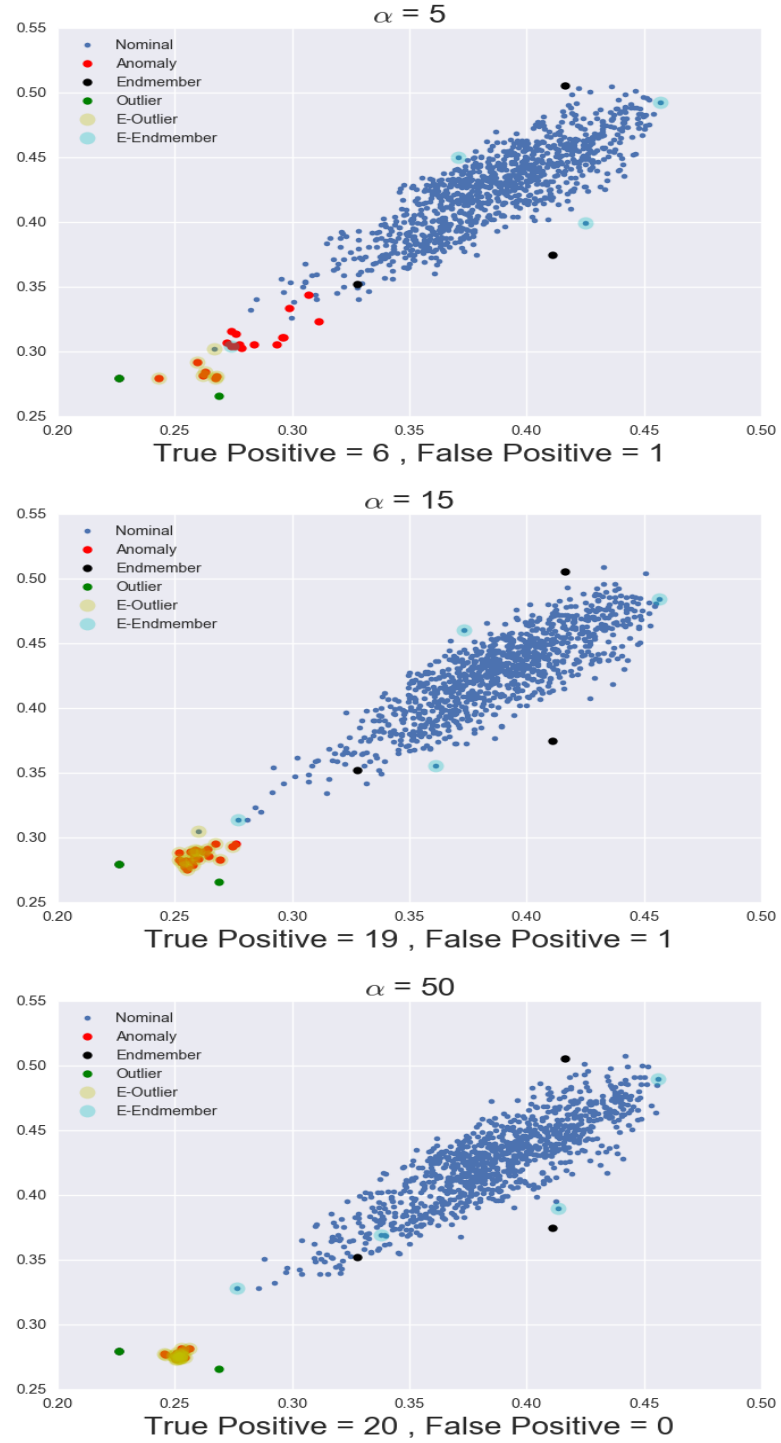


Figure 3.10: Effect of Anomaly concentration: In each plot a 2D slice of a dataset generated based on the same endmember signatures and anomaly signatures is illustrated. For construction of anomalies different levels of concentration is set : $\alpha = [5, 15, 50]$.

Let us remind that when α' grows, the separation between outliers and nominal data is more clear. From this figure one observes that when $\alpha' = 15$, we reach $\kappa \simeq 0.8$ which is a very good accordance. In order to provide a more visible understanding of the effect of anomaly dispersion on the accuracy of the AD algorithm, in Figure 3.10 we generated different datasets using the same endmember signatures and anomaly signatures. Like other tests 1000 samples generated using $\ell = 3$ endmembers and $\mathcal{V} = 20$ anomalies generated using 2 spectral signatures. For generation of anomalies different levels of concentration have set which are reflected in the parameter $\alpha = [5, 15, 50]$. The result of the AD algorithm is then evaluated according to the ratio between *True Positive (TP)* and *False Positive (FP)*. It is visible that how large dispersion (low values of α) can obscure the division line between nominal data and outliers, resulting in low accuracy of anomaly detection. In the contrary high value of α isolates the anomalies and thus it would be easier for the AD algorithm to detect them.

3.3.3 On the application of Overcomplete Dictionary

In order to illustrate the benefits of overcomplete dictionaries in estimating reliable endmembers, we generated a dataset composed of $N = 10^3$ nominal data on the basis of BMM using 4 original signatures.

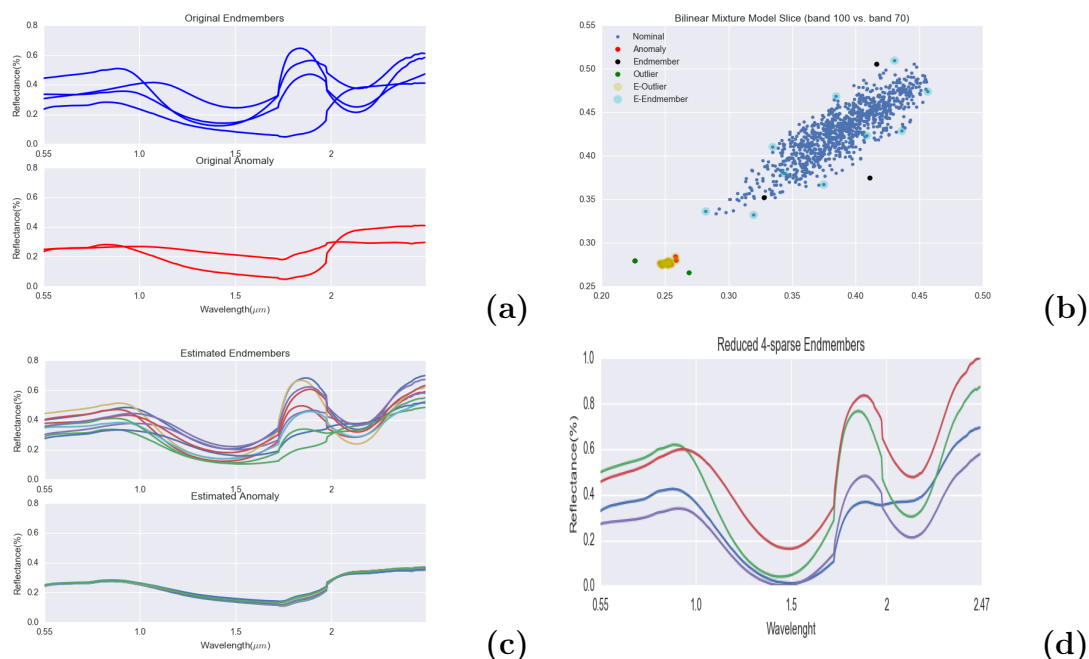


Figure 3.11: Results in presence of anomalies using SAGA+. (a): original signatures used for generation of dataset; (b): 2D Slice of dataset together with the estimated prototypes (c): signatures of estimated endmembers and anomalies with SAGA+ (d): the 4 weighted averaged endmembers.

A number of $\mathcal{Y} = 20$ anomalies have been added from 2 different signatures. When running SAGA+ we choose $\ell = 10$ atoms to represent endmembers on an overcomplete dictionary and we set a sparsity level $\lambda = 4$ since mixing were generated from 4 signatures.

Results are shown in Figure 3.11 where we depicted original endmembers and outliers (Figure 3.11(a)), a 2D slice of the generated dataset (Figure 3.11(b)), the $\ell = 10$ estimated endmembers and associated outliers (Figure 3.11(c)) and the resulting 4 "averaged" endmembers (Figure 3.11(d)). In this last sub-figure, as we know exactly the location of the 4 pure endmembers, each of the ten estimated ones is associated to the closest real one and weighted averages, on the basis of their abundances, are represented. One can see the good adequacy of our approach when comparing these averaged endmembers with real ones in the top of Figure 3.11(a). It is also interesting to see the effect of overcomplete dictionary on the accuracy of the abundance coefficients. For this we generated datasets in a similar way as the previous step. For the projection of dataset on the estimated endmembers by N-finder and VCA algorithms we used the *Constrained Least Squares* (CLS)[3] method. This time we evaluated the closeness of each row of the estimated abundance matrix with its closest counterpart in the original generated abundance matrix using the mean of SAM error calculated as:

$$\epsilon(\mathbf{G}, \hat{\mathbf{G}}) = \min_{\eta} \frac{1}{\ell} \sum_{i=1}^{\ell} \text{SAM}(\mathbf{g}_{\eta^{\ell}(i)}, \hat{\mathbf{g}}_i) \quad (3.16)$$

where ℓ is the number of endmembers and $\eta^{\ell} : \{1, \dots, \ell\} \longrightarrow \{1, \dots, \ell\}$ is a function corresponding to all possible permutations. The result of this comparison is represented in the table 3.2 which shows considerable improvement in the accuracy of the abundance maps in the case of SAGA+ comparing to the other methods.

Table 3.2: Average SAM errors between the original and estimated abundance coefficients. Best results are reported in boldfont

	ℓ	3	5	7	10	12	15
ϵ	VCA	0.237	0.541	0.756	0.912	0.981	1.105
	N-finder	0.155	0.451	0.713	0.918	0.970	1.108
	G-SiVM	0.333	0.592	0.951	1.084	1.121	1.301
	SAGA+	0.096	0.123	0.311	0.549	0.711	0.892

Finally, we report a comparison of execution times in supplementary material provided in the appendix 7.

3.3.4 Results on real data: 4-Vesta Asteroid

In this experiment we consider the performance of SAGA+ on 4-Vesta and try to give a high level interpretation of the results which can be of interest from the planetary observation point of view. In practice, two difficulties arise when manipulating 4-Vesta data : 1) the real number of endmembers is unknown and 2) the assumption that some pure pixels are present in the original image does not hold. In order to cope with these difficulties, as mentioned earlier, we exploit the intrinsic characteristics of *overcomplete* dictionary learning approach which relaxes both mentioned difficulties by selecting out a large number of endmembers (referred to as the dictionary atoms) from the image.

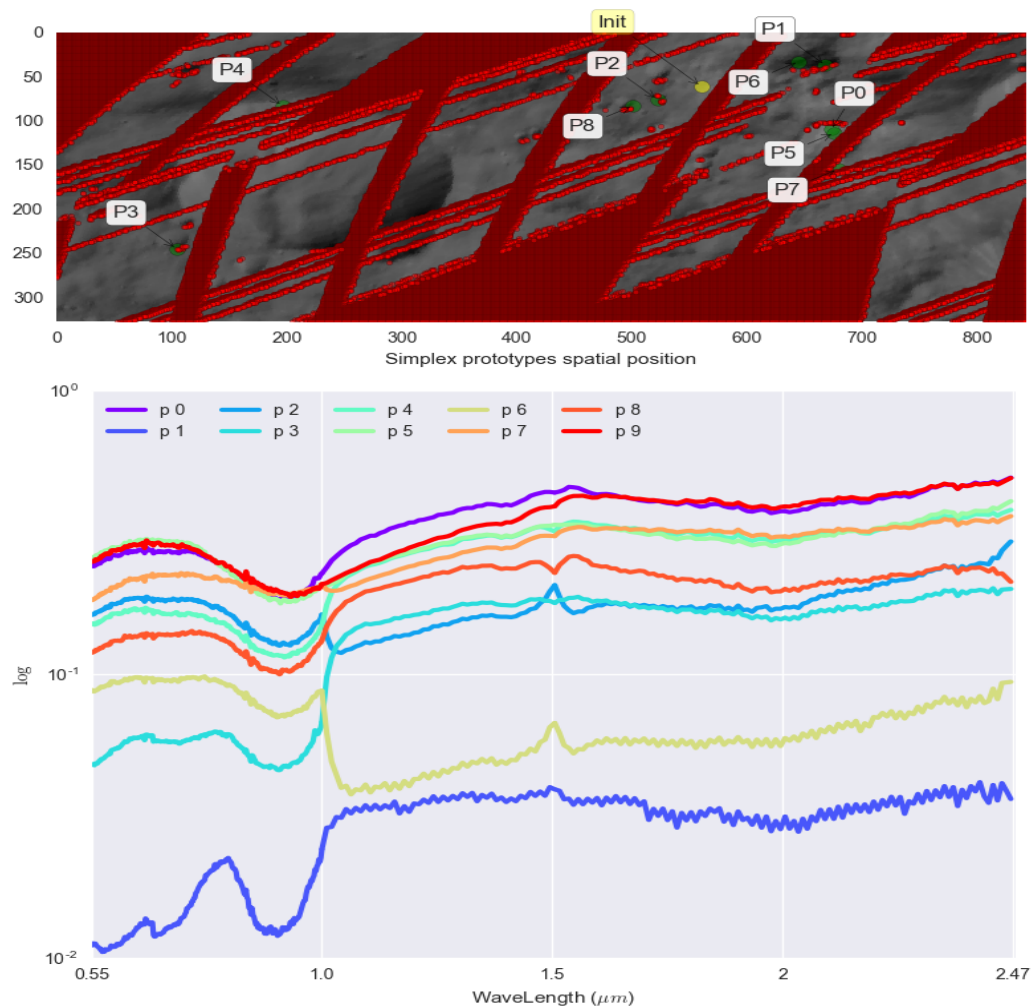


Figure 3.12: Results on real 4-Vesta asteroid. **Top:** Vesta-4 mosaic with the spatial position of the extracted endmembers. The red points are detected anomalies and the 10 green points with labels are selected endmember prototypes. **Bottom:** The respective estimated endmember signatures.

From the geological point of view, the large number of selected endmembers/atoms might be interpreted as different compositions of materials in a scene. In practice for the case of 4-Vesta we choose the dictionary of size $\ell = 10$. Because of the intrinsic characteristic of overcomplete dictionary for re-grouping the similar endmembers, this parameter is not crucial but has to be set such that the number of atoms of the dictionary is larger than the number of expected "pure" materials. The estimated endmembers are represented in the bottom of Figure 3.12 (denoted P0 to P9) while the top of this figure depicts the spatial positions of the respective estimated endmembers (in green) together with the outliers (in red). A specimen of outliers spectra signature (with a sharp peak) is represented in the bottom of Figure 3.5.

By comparing these two plots (Figure 3.12 and Figure 3.5), one can observe that the estimated endmembers signatures are generally in accordance with the expected nominal signature profiles (laboratory spectra related to the lithologies on the northern hemisphere). Although one observes some unwanted signatures (mainly contributing to the shadowing effect), it is of prime interest to note that the absence of signatures corrupted by noise is significant⁵. Indeed, endmembers related to unwanted spectra can be discarded: P1 and P3 have a low reflectance and correlate perfectly with the most dark (shadow) areas, while P2, P6 and P7 show a shift between the two detectors, with a very small amount of noise. The five other endmembers are of interest regarding mineralogical detections. P4 and P8 show characteristic signatures of pyroxenes and represent the surface background. Some outcrops, related to P0, exhibit similar signatures with stronger absorptions, in relation with impact craters (i.e. fresh rocks) and very good illumination conditions. The last two endmembers P5 and P9, detected in small spots in and around craters, show a broader absorption feature in the $1\mu m$ range. One can see the whole set of endmembers and their respective abundance maps in Figure 3.15. From these plots, it is worth noting that the spatial consistency of associated maps matches well with real physical structures in 4-Vesta, which is a very good property.

As a more physical comparison, one can also compare the spectra in Figure 3.12 with the laboratory measured spectra of Olivine and Orthopyroxene which is illustrated in Figure 3.13. As can be observed in this figure, the broader $1\mu m$ absorption seen in P9 and P5 can be correlated to an increased olivine content in the rocks, in accordance with observations made by [105] or [114]. From these observations, SAGA+ is thus able to extract from the hyperspectral image, spectra that are a combination of mixtures of materials, representative of distinct geological units. Additionally, using the extracted endmembers, the outcrops enriched in Olivine (i.e. spots where pixels in which P9 and P5 are needed in the unmixing) can be mapped (see Figure 3.14) and found localization are similar to those obtained with other techniques [114].

⁵Interested reader may find a plot of detected anomaly signatures from this mosaic in the supplement material provided in the appendix 7.

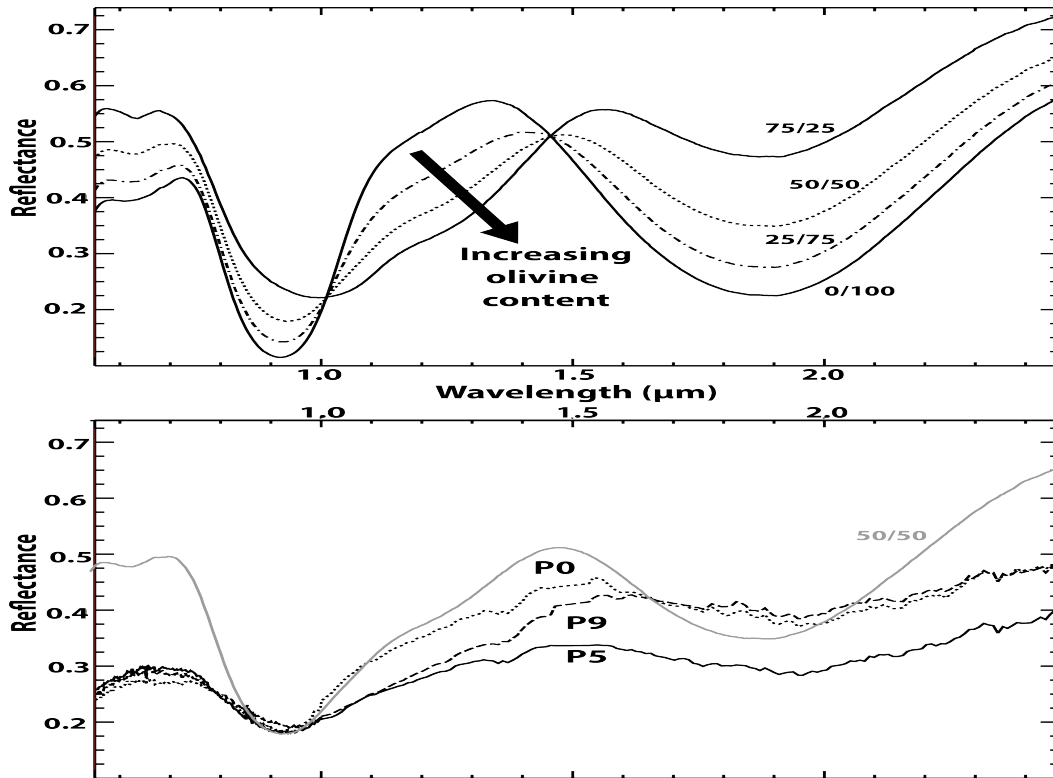


Figure 3.13: **Top:** laboratory spectra of mixtures involving olivine and orthopyroxene (spectra from the RELAB library with numbers referring respectively to the olivine and orthopyroxene contents (% olivine / % orthopyroxene). **Bottom:** comparison with the three extracted endmembers P0, P5 and P9. The shift in absorption center at $2\mu\text{m}$ is due to an increase in Ca and/or Fe in the pyroxene, compared to the pure orthopyroxene composition.

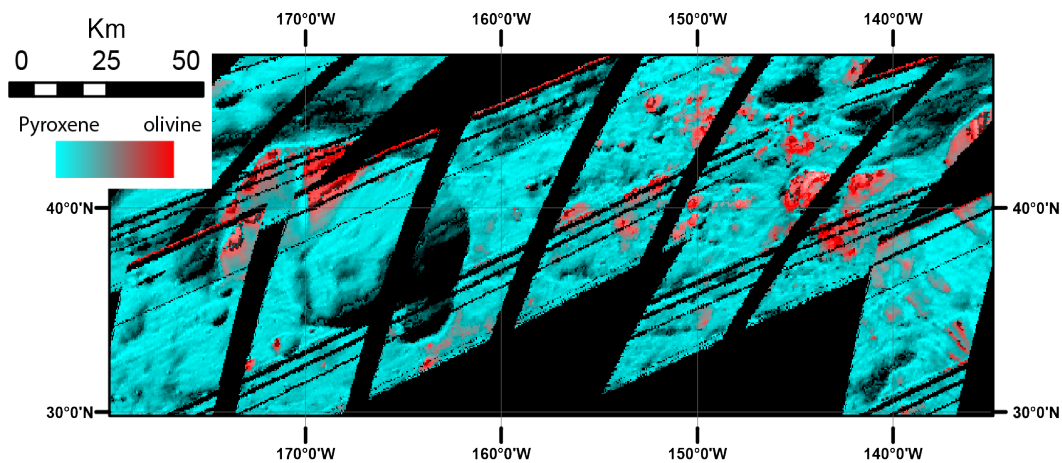


Figure 3.14: RGB composition with $R=P9+P5$, $G=P4+P8+P0$, and $B=P4+P8+P0$. Coloring shows the localization of outcrops enriched in olivine (appearing in red)

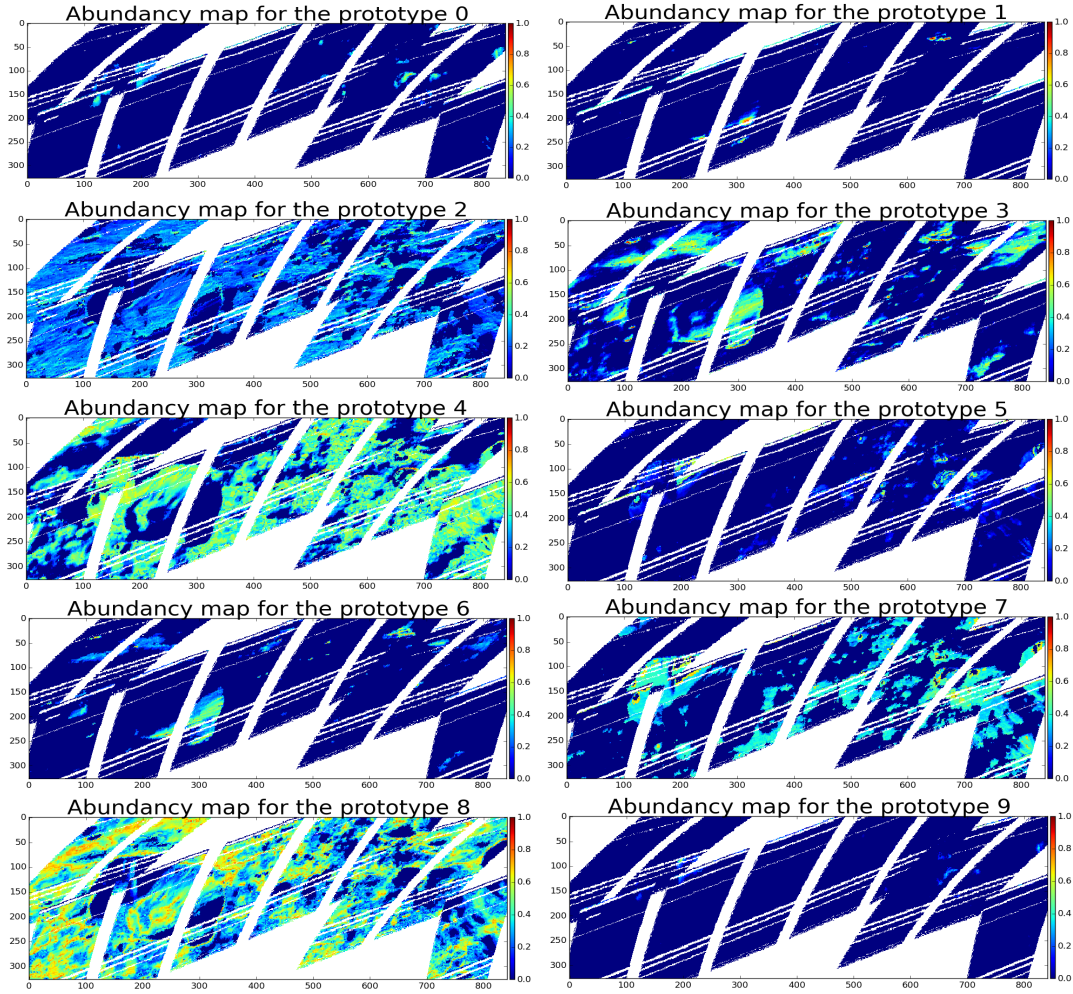


Figure 3.15: Abundance maps and respective spectral signatures extracted by SAGA+ (with anomaly detection procedure). Using an anomaly detection procedure, one observes that signatures with magnitudes higher than 1 and those with sharp peaks are not selected as endmembers.

Let us note that the shift in absorption center at $2\mu\text{m}$ is due to the chemical composition of orthopyroxene (olivine does not have absorptions in this wavelength range) and reflects the fact that the orthopyroxene used as a reference does not have the same chemical composition as the one observed on 4-Vesta's surface. Lastly, in order to represent the benefits of the anomaly detection procedure, we have applied the same method without the outlier removal process. Extracted endmembers and associated abundances are available in the supplement material provided in the appendix 7. By comparing them with Figure 3.15, we clearly observe corrupted endmembers yielding less consistent abundance maps. This experience reveals the importance of the anomaly detection process introduced here.

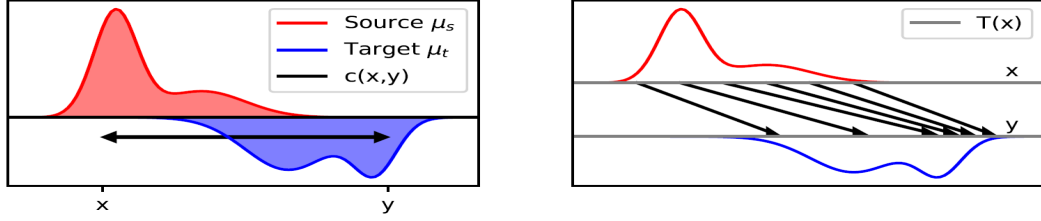
3.4 Conclusion

In this chapter, a joint procedure for anomaly detection and spectral unmixing has been presented for hyperspectral data. Endmembers and anomalies are extracted from the vertexes of the simplex that embeds data, this simplex being computed in a feature space to deal with non-linearities. Abundances are computed on the basis of sparse projections of spectra onto extracted endmembers. To deal both with the internal heterogeneity inside classes and with the fact that pure elements are not necessarily present in the data, overcomplete dictionaries have been used. We evaluated and compared our approach with linear and non linear unmixing methods on synthetic datasets generated from real spectra where linear and nonlinear mixing procedures have been applied. Anomalies have also been introduced in the data. These experiments have revealed the efficiency of our technique to deal with anomalies and endmember detection at the same time. Computational aspects have also been discussed and have revealed the ability of our algorithm to easily scale up to large datasets. Lastly, experiments have been performed on real data issued from the 4-Vesta Asteroid. From a geological point of view, we have shown that anomalies, endmembers and mixing abundances are consistent, demonstrating the efficiency of our approach for real applications.

Part III

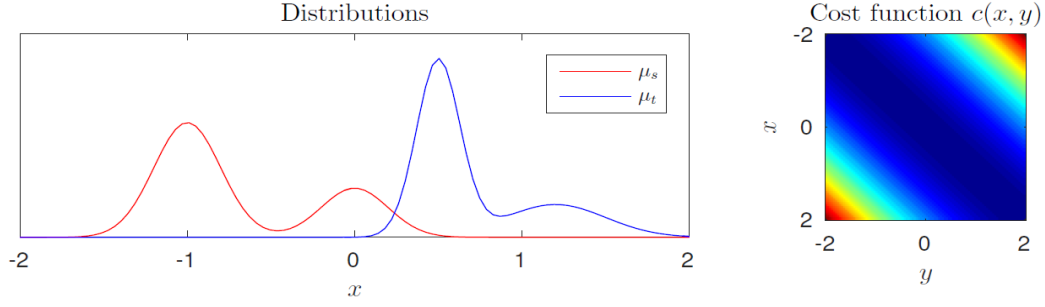
Optimal Transport for Unmixing

Chapter Summary

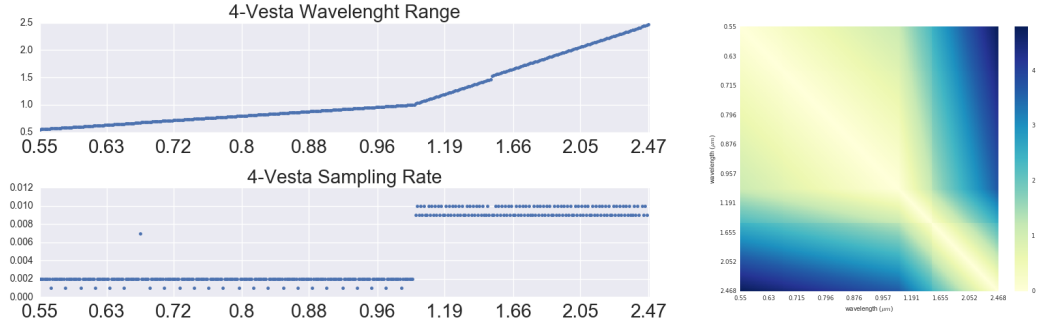


In this chapter a new supervised model for spectral unmixing based on Optimal Transport(OT) theory is presented. This theory, very well adapted to the manipulation of distributions, is exploited here on hyperspectral data viewed as distributions instead of a multivariate vector. Optimal transport consists in searching for a unique plan which transports a distribution \mathbf{x}_1^\top onto \mathbf{x}_2^\top within the minimum cost. The associated distance measure for this is called Wasserstein distance. In order to calculate the abundance values for each pixel in \mathbf{X} , we project the pixel spectrum onto each endmember (also presented as distribution) in our dictionary \mathbf{E} using OT-based projection (called Bregman projection), while an entropy-based regularization term promotes the *a priori* knowledge regarding the inter-endmember relationships. In this framework the information reflecting distance relationships along the spectral dimension (in the case of hyperspectral spectra it will be wavelength values) of data is being used as an initial cost function. Also the information regarding the material dependencies of the endmembers of \mathbf{E} will be provided as another cost function. Our algorithm is also able to deal with endmember variabilities among the values of \mathbf{E} . In 4.1 after a short recap on the notion of spectral unmixing, the necessity of having representative distance measure, being able to reveal the characteristics of data domain is mentioned which consequently follows with the introduction of the Wasserstein distance. In 4.2 Optimal transport, Wasserstein distance and the Entropy based regularization is being explained along with a closed-form solution for bregman projection. In 4.2.3 the application of two tailored cost functions within the optimization procedure is discussed and finally in 4.3 the preliminary results of applying this algorithm on a portion of 4-vesta dataset is represented.

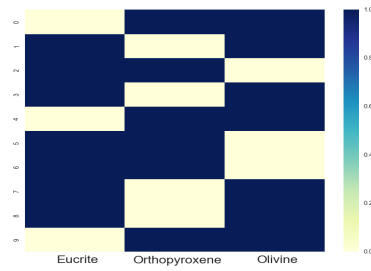
✎ This chapter corresponds to an article published in WHISPERS 8th workshop on Hyperspectral image and Signal Processing: Evolution in Remote Sensing, and presented in Aug 2016, Los Angeles, United States.



OT considers spectra as probability distributions. A pixel spectrum therefore will be normalized along spectral values. Normalization makes the analysis less sensitive to the total power of spectra in each pixel. This improves robustness against shadows or other large radiance changes and thus can prevent degenerate solutions.



In order to tailor a cost matrix C_0 in alignment to the characteristics of the dataset, it is possible to build $C_0(D \times D)$ as the square euclidean distance over the spectral values. This clearly reflects the characteristic of the spectra and the level of (dis)similarity among them.



It is possible to manually construct the cost matrix C_1 in order to incorporate the information regarding the endmembers belonging to the same material group. Two endmembers belonging to the same material share a very low cost with the corresponding material in α_0 , $C_1(i; j) = 0$

Supervised Planetary Unmixing with Optimal Transport

Abstract

This chapter is focused on spectral unmixing and presents an original technique based on Optimal Transport. Optimal Transport consists in estimating a plan that transports a spectrum onto another with minimal cost, enabling to compute an associated distance (Wasserstein distance) that can be used as an alternative metric to compare hyperspectral data. This is exploited for spectral unmixing where abundances in each pixel are estimated on the basis of their projections in a Wasserstein sense (Bregman projections) onto known endmembers. In this work an over-complete dictionary is used to deal with internal variability between endmembers, while a regularization term, also based on Wasserstein distance, is used to promote prior proportion knowledge in the endmember groups. Experiments are performed on real hyperspectral data of asteroid 4-Vesta.

Keywords

Optimal Transport, Spectral Unmixing, Endmember Variability, Wasserstein Distance, Bregman Projection

4.1 Introduction

Unmixing is an active field in hyperspectral data analysis and consists in estimating the abundance of pure materials (named endmembers) that are part of a pixel spectrum. These pure elements are either estimated from the image (unsupervised methods) or directly given in a prior knowledge (supervised approach).

Spectra comparison. & optimal transport In all techniques mentioned in chapter 2, a metric between spectra is required. Several criteria have been used in a hyperspectral context (spectral angle mapper and Euclidean minimum distance being the two most popular ones), however, few approaches consider spectra as probability distributions in the spectral domain. We consider in this work an original metric, designed between probability density functions to compare spectra. This metric, based on optimal transport (OT) is called Wasserstein distance, and can be designed to be mostly sensitive to shifts in the frequency domains. To exploit OT, each spectrum should be normalized (the integration along spectral values should be 1) in order to be viewed as a probability distribution. By doing this normalization we make the analysis insensitive to quantities of materials in each pixel and we rather focus on proportions. This enables for example to improve robustness against shadows or other large radiance changes and as discussed in [115], can prevent degenerate solutions.

Outline of the chapter. In the next section, we explain the basics of OT together with the entropy-based regularization term which is used for calculation of abundance coefficients. Experiments on the challenging 4-Vesta asteroid dataset are then shown and discussed in section 4.3.

4.2 Unmixing using priors and iterative Bregman projections

We first introduce optimal transport, then we present in section 4.2.2 a regularized version more efficient from a computational point of view and section 4.2.3 proposes an unmixing model based on optimal transport.

4.2.1 Optimal transport of spectra

Optimal Transport (OT) can be seen as the search for a plan that moves (transports) a spectrum μ_1 onto another spectrum μ_2 with a minimum cost. This cost is usually related to a metric of the embedding space. In our case, we will generally use a squared Euclidean distance L_2^2 as this cost, but it should be mentioned that other cost functions, specific to a given problem, might be used. In the relaxed formulation of Kantorovitch [116], OT seeks for an optimal coupling that can be seen as a joint probability distribution between μ_1 and μ_2 . In other words, if we define $\Pi(\mu_1, \mu_2)$ the space of probability distributions over \mathbb{R}^2 with prescribed marginals μ_1 and μ_2 , the optimal transport is the coupling $\gamma \in \Pi(\mu_1, \mu_2)$ which minimizes the following quantity:

$$W_c(\mu_1, \mu_2) = \inf_{\gamma \in \Pi(\mu_1, \mu_2)} \int_{\mathcal{R}^2} c(\mathbf{x}_1, \mathbf{x}_2) d\gamma(\mathbf{x}_1, \mathbf{x}_2) \quad (4.1)$$

where $c(\mathbf{x}_1, \mathbf{x}_2)$ is the cost between \mathbf{x}_1 and \mathbf{x}_2 (issued from distributions μ_1 and μ_2 respectively). In the discrete versions of the problem, *i.e.* when μ_1 and μ_2 are sampled by a sensor and expressed as vectors of \mathbb{R}^d , the previous equation reads:

$$W_{\mathbf{C}}(\mu_1, \mu_2) = \min_{\gamma \in \Pi(\mu_1, \mu_2)} \langle \gamma, \mathbf{C} \rangle_F, \quad (4.2)$$

where $\langle \cdot, \cdot \rangle_F$ is the Frobenius dot product and $\mathbf{C} \geq 0$ is a cost matrix (of size $d \times d$), which gathers all the costs for moving from bin i to bin j , and γ is now a matrix of size $d \times d$, with marginals defined as μ_1 and μ_2 . We note that this distance is also known as the Earth Moving Distance in the computer vision community [117]. Solving for equation (4.2) is a simple linear programming problem with equality constraints, but its dimensions scales quadratically with the number of samples in the spectra. This might generate computational problems when the dimension of distributions is high, as in hyperspectral data. For this reason, we now consider a regularized version of the problem which has the extra benefit of being faster to solve.

4.2.2 Entropy regularized Optimal Transport and Bregman projections

The idea is here to put a regularization term over γ that controls the smoothness of the coupling through its entropy. The entropy regularized version of the discrete optimal transport problem reads [118]:

$$W_{\mathbf{C}, \epsilon}(\mu_1, \mu_2) = \min_{\gamma \in \Pi(\mu_1, \mu_2)} \langle \gamma, \mathbf{C} \rangle_F - \epsilon h(\gamma), \quad (4.3)$$

where $h(\gamma)$ is the entropy of γ , reading

$$h(\gamma) = - \sum_{ij} \gamma_{ij} \log \gamma_{ij} = - \langle \gamma, \log \gamma \rangle_F,$$

if \log is applied component-wise to γ . Denoting the Kullback Leibler divergence (KL), *i.e.* $\text{KL}(\gamma, \rho) = \sum_{ij} \gamma_{ij} \log \frac{\gamma_{ij}}{\rho_{ij}} = \langle \gamma, \log \frac{\gamma}{\rho} \rangle_F$ where again the log and the division are taken component wise, one can establish the following link between OT and Bregman projections:

Proposition 4.2.1 [OT as a Bregman projection [119, Eq. (6)]] *The minimizer γ^* of (4.3) is the solution of the following Bregman projection*

$$\gamma^* = \arg \min_{\gamma \in \Pi(\mu_1, \mu_2)} \text{KL}(\gamma, \zeta), \quad (4.4)$$

where $\zeta = \exp(-\frac{\mathbf{C}}{\epsilon})$.

Interestingly enough, the entropy regularized version of OT admits a simple algorithm for the resolution of this problem, based on the successive projections over the two marginal constraints. More specifically, solving (4.4) when one of the two marginals is unknown admits a closed form solution.

Proposition 4.2.2 [Closed form solution of OT with one unknown marginal [119, Eq. (7)]] *For an undefined μ_2 , γ^* is solely constrained by one marginal μ_1 and is the solution of the following closed-form projection:*

$$\gamma^* = \text{diag} \left(\frac{\mu_1}{\zeta \mathbf{1}} \right) \zeta \quad (4.5)$$

where again the division has to be understood component-wise. Consequently, by iterating over successive Bregman projections on the set of the two marginal constraints, one can derive an efficient algorithm to solve for the OT problem. We refer the reader to [119] for a more complete treatment on this subject. Finally, when several OT distances are implied in the same optimization problem, this iterative projection strategy is still applicable. We use this property in our unmixing model, described in the following.

4.2.3 Unmixing model

We start here by recalling our initial assumptions. We assume a linear mixture for each spectrum μ : $\mu = \mathbf{E}\alpha$ where matrix \mathbf{E} of size $d \times q$ contains q (known) endmembers of dimension d and vector $\alpha > 0$ of size $q \times 1$ has to be estimated under the sum-to-one constraint, *i.e.* $\alpha^T \mathbf{1} = 1$. As such, it corresponds to discrete probability distribution function or histogram. From the image analysis, we seek to extract p abundances per pixels of distinct materials of interest. In practice the number of endmembers q exceeds the number of proportions p we aim at estimating ($q \geq p$). This allows to consider several distinct versions of spectra for the same material of interest, thus accounting for endmembers variability. In addition we assume that we have a prior knowledge α_0 (p -dimensional) over the abundances.

The unmixing of a spectrum μ is the solution of the following optimization:

$$\alpha = \arg \min_{\alpha} \underbrace{W_{\mathbf{C}_0, \epsilon_0}(\mu, \mathbf{E}\alpha)}_{\text{data fitting}} + \tau \underbrace{W_{\mathbf{C}_1, \epsilon_1}(\alpha, \alpha_0)}_{\text{prior}}. \quad (4.6)$$

As it can be seen, the problem is subdivided in a data fitting term (which tries to find the best decomposition from observations) and a regularization term (which enforces the compliance of the solution with the prior) balanced by parameter $\tau \in \mathbb{R}^+$. From these two terms, regarding equation (4.3), two cost functions \mathbf{C}_0 and \mathbf{C}_1 are involved. Matrix \mathbf{C}_0 is linked with a metric in the spectral dimensions of the data (and is usually related to the L_2 -norm along frequencies) while \mathbf{C}_1 is a cost matrix of size $q \times p$ with a specific structure which contains information about the classes

of endmembers. Two endmembers related to the same element will share a very low cost with the corresponding material in α_0 but a higher cost with other materials. An easy way to handcraft \mathbf{C}_1 is to set $\mathbf{C}_1(i, j) = 0$ if endmember i is related to material j , and 1 elsewhere. Finally, two regularization parameters ϵ_0 and ϵ_1 related to the regularized version of the transport have to be set. The resolution of the optimization problem in (4.6), inspired from the weighted Wasserstein barycenter problem [119], is done using iterative Bregman projections. Algorithm 3 presents the corresponding procedure in pseudo-code.

Algorithm 3 Supervised unmixing procedure with OT

```

1: Input:  $\mu$ 
2: Parameters:  $\alpha_0, \mathbf{E}, \tau, \mathbf{C}_0, \mathbf{C}_1, \epsilon_0, \epsilon_1, \text{maxIter}$ 
3:  $\text{cpt} \leftarrow 0, \text{err} \leftarrow \infty$ 
4:  $\alpha \leftarrow \mathbf{1}/q$ 
5:  $\zeta_0 \leftarrow \exp(-\frac{C_0}{\epsilon_0}), \zeta_1 \leftarrow \exp(-\frac{C_1}{\epsilon_1})$ 
6: while  $\text{cpt} < \text{maxIter}$  and  $\text{err} > \text{threshold}$  do
7:    $\zeta_0 \leftarrow \text{diag}(\frac{\mu}{\zeta_0 \mathbf{1}}) \zeta_0$ 
8:    $\zeta_1 \leftarrow \text{diag}(\frac{\alpha_0}{\zeta_1 \mathbf{1}}) \zeta_1$ 
9:    $\mu_{tmp} \leftarrow \exp(\log(\zeta_0 \mathbf{1}) + \tau \log(\mathbf{E} \zeta_1 \mathbf{1}))$ 
10:   $\zeta_0 \leftarrow \zeta_0 \text{diag}(\frac{\mu_{tmp}}{\zeta_0^T \mathbf{1}})$ 
11:   $\zeta_1 \leftarrow \zeta_1 \text{diag}(\frac{\mathbf{E}^T \mu_{tmp}}{\zeta_1^T \mathbf{1}})$ 
12:   $\text{err} \leftarrow \|\alpha^{(\text{cpt})} - \alpha^{(\text{cpt}-1)}\|_2$ 
13:   $\text{cpt} \leftarrow \text{cpt} + 1$ 
14: end while
    return  $\alpha$ 

```

4.3 Experiments and Analysis

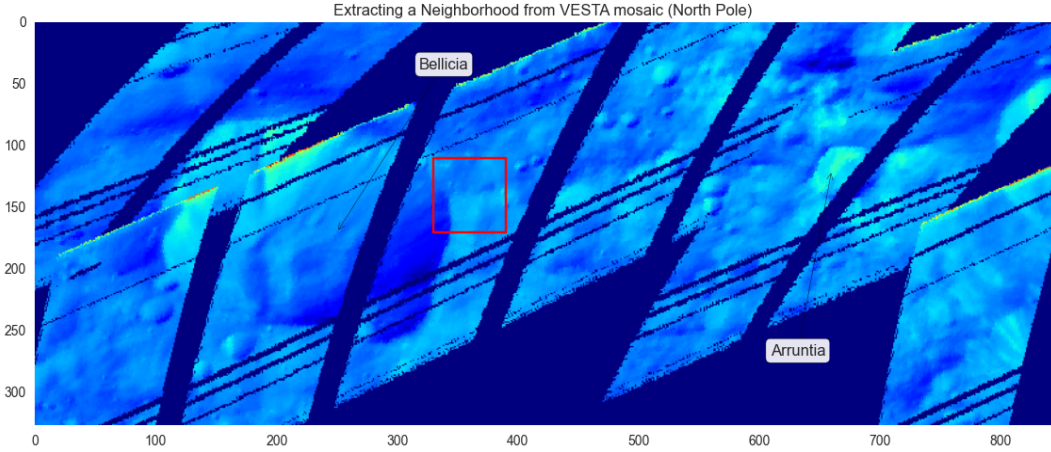


Figure 4.1: Mosaic of northern hemisphere of Vesta-4 asteroid. The window in red (61×61) is used as the test bed.

4.3.1 4-Vesta properties and dataset

A mosaic of Vesta-4 asteroid is visible in Figure 4.1. The image is taken by *visible and infrared mapping spectrometer* (VIR) which comes with two arrays of detectors: One array covering the visible range (between $.25\mu m$ and $1.05\mu m$) with spectral sampling of $1.8nm$ and the other covering the infrared range (between $1.0\mu m$ and $5.0\mu m$) with spectral sampling of $9.8nm$. The total number of considered bands after removing the noisy ones are $d = 383$ covering the range of $.55\mu m$ to $2.4\mu m$. Although there is no exact information about the components of the *crust* and *mantel* of this protoplanet asteroid, some expert knowledge is available (see [120] and [121]). From this we know that the major surface is dominated by three rocks: *Eucrite*, *Howardite* and *Diogenite*. It is believed that roughly the southern hemisphere of 4-Vesta is Deogenite-rich while the northern hemisphere crust is mainly composed of *Eucrite* and different types of *pyroxene* with recently detected clusters of *Olivine*.

4.3.2 Experimental setup

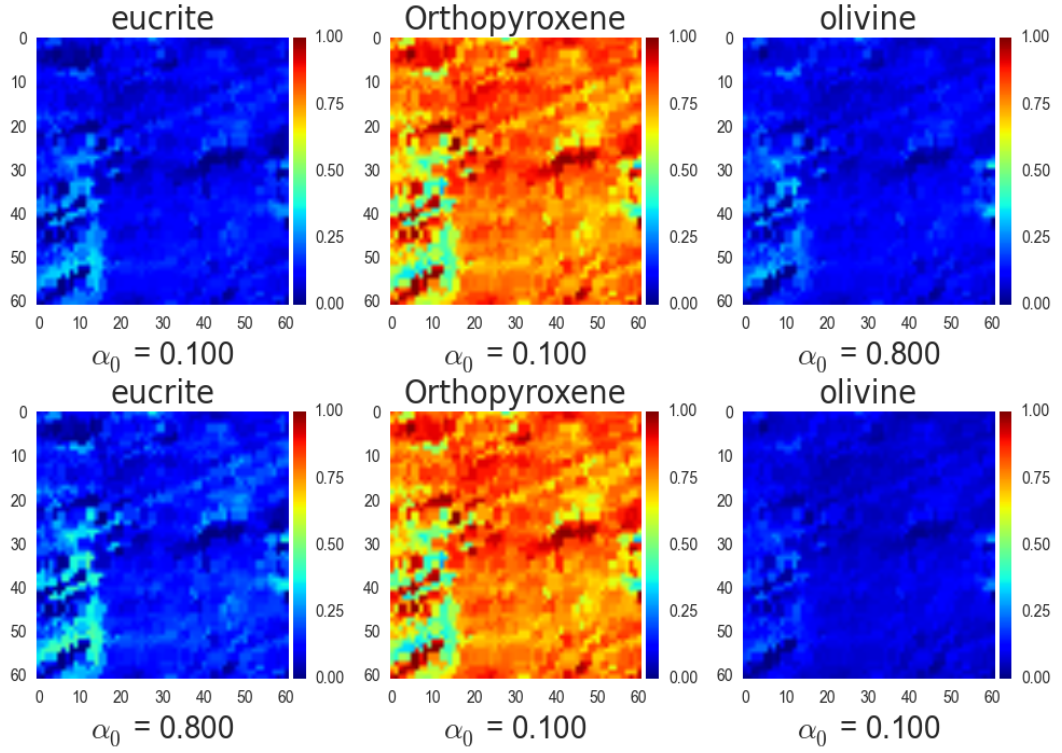
We used a total number of $q = 20$ spectral signatures of different lithologies, found ubiquitously in the composition of the meteorites, collected on Earth. we could categorize these signatures in 3 groups: *Olivine*, *Orthopyroxene*, *Eucrit*. Consequently we set $p = 3$ in order to identify "maximally pure" endmembers which are believed to exist within the scene. As the consequence, matrix \mathbf{C}_1 involved in equation (4.6) will be of dimension 20×3 . As mentioned in previous section, this cost matrix contains information about the classes of endmembers (two endmembers related to the same material will share a very low cost with the corresponding material in α_0 but a higher cost with other materials). In particular we set $\mathbf{C}_1(i, j) = 0$ if endmember i is related to material group j , and 1 elsewhere.

As for the cost matrix \mathbf{C}_0 in relation (4.6), we simply use the L_2^2 norm between each spectral value. The regularization parameters ϵ_0 and ϵ_1 have been set to 1 and 10^3 and the balance parameter is $\tau = 0.9$. This is set manually from data but more advanced approaches with cross-correlation could be used. For setting prior knowledge of the expected endmembers different combinations of non-negative 3-values for α_0 with non-negativity and additivity constraints were tested. By α_0 we reflect our prior expectation of the proportions of the 3 expected groups of lithologies.

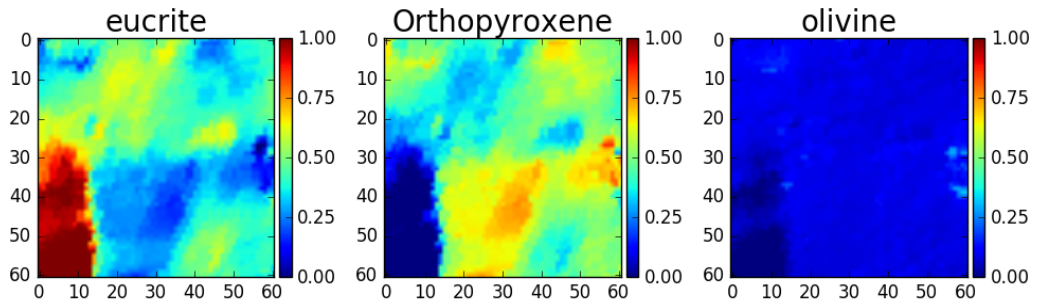
4.3.3 Unmixing results and discussion

We made some experiments with setting different combinations of priors for each group of materials. By this we observe abundance maps with some nuances. These are represented in Figure 4.2(a). It is interesting to observe that in all cases, spatially consistent areas extracted and correspond to existing spatial patterns of rocks. This is in contrast to the unmixing results we achieved by using constrained least squares (CLS) method which is illustrated in Figure 4.2(b) b). As for the quantities of pure

elements, our results are consistent with very recent studies [121] on this part of 4-Vesta. In particular the fact that different compositions of *Orthopyroxene* are dominant while *Olivine* patches are spatially much more isolated and sparse goes well with the latest observations on the composition of the materials of this area. As the role of priors one can see the change in susceptibility/sensitiveness towards a particular material when changing the prior value over that material. This in particular is visible w.r.t the Olivine maps.



(a) Estimated Abundance Maps with different priors α_0 .



(b) Estimated Abundance Maps with Constrained Least Squares.

4.4 Discussion/Conclusion

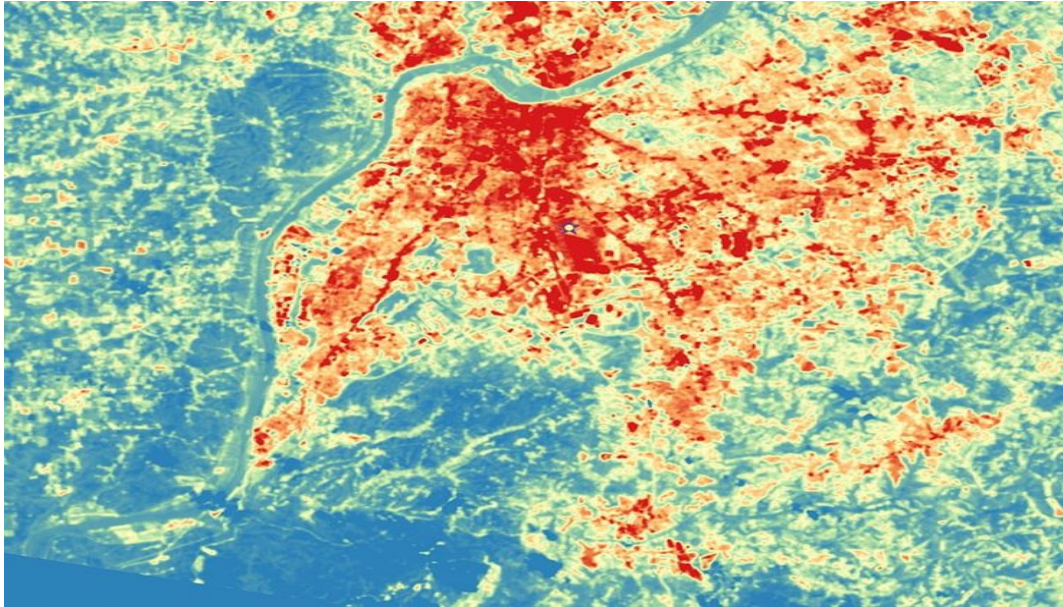
In this work we have proposed a hyperspectral unmixing approach based on optimal transport. This metric, devoted to distributions, is mostly sensitive to shifts in the frequency domains and therefore differs from most usual criteria as spectral angle mapper or Euclidean distance. The problem of spectral variability and the confusion in selecting the best endmembers among the existing dictionary is addressed through an overcomplete dictionary. This results in a specific cost-function to optimize where both observation and regularization terms are based on OT. Our primary experiments on the real 4-Vesta dataset show a good accordance of our results with the unmixing issued from the latest geological findings about the composition of lithologies on this asteroid.

In the next chapter we are going to see the application of the two proposed unmixing procedures on a hyperspectral image for urban monitoring purpose.

Part IV

Applications of Spectral Unmixing in Urban Context

Chapter Summary




The goal of this chapter is to apply and demonstrate the performance of the two spectral unmixing algorithms we introduced in the previous chapters within the context of land surface temperature (LST) estimation. An example of such temperature estimation is visible in above illustration. The global idea of the project associated with our work is to use the abundance maps generated by the unmixing procedures in order to accurately calculate the surface area of different urban components (roads, parks, building roofs with specific materials etc.). This information will be combined with emission models specific to each material to provide accurate maps of surface temperature for different regions of the city. In this chapter and for the sake of this thesis work we focus on the unmixing part of this project. Experiments have been performed on a dataset issued from the city of Madrid (Spain) using the two algorithms introduced in this thesis. The generated abundance maps have also been compared with the results of other state-of-the-art techniques. In the first section we outline the goals of the associated project (entitled CATUT, French acronym for Mapping Urban Temperatures), then we provide a description of the data, (acquisition and characteristics). In the second section we introduce the way three endmember dictionaries which are going to be used for our unmixing task are acquired and preprocessed. We also set six strategies for conducting our unmixing experiments. In the last section we provide our preliminary results.

☞ This chapter is an applicative part under development. First results are introduced here.

CATUT : Spectral Unmixing for Urban surface Temperature

5.1 DESIREX 2008 project

Land surface temperature (LST) estimation is one of the goals of the "Dual-use European Security IR Experiment 2008" (DESIREX 2008) campaign in Madrid Spain, carried out by the European space agency (ESA) through several acquisition missions between June 23rd and 6th of July. Spectrally, geometrically and radiometrically representative datasets were generated and estimating Urban Heat Island (UHI) and Urban Thermography (UT) modeling are the two major objectives. Because of the limited spatial resolution of acquired hyperspectral data (from 10 to 30 square meters in order) estimating LST needs two steps: 1) Spectral unmixing of the pixels to recover abundances of each pure material and 2) Calculation of the area temperature based on the endmember abundances and associated emissivity models for each material. Data were acquired with an airborne hyperspectral scanner (AHS) covering the spectral regions of 445 – 12450nm with varying spectral resolutions (across the spectral range)

 For more details regarding airborne and satellite data acquisition, one can refer to the DESIREX 2008 final report ^a.

^ahttps://earth.esa.int/c/document_library/get_file?folderId=21020&name=DLFE-905.pdf

5.1.1 Description of the urban area

DESIREX measurements were registered over the city of Madrid in two directions north-south and west-east. Figure 5.1 shows the location of these lines in green crossed mosaics.

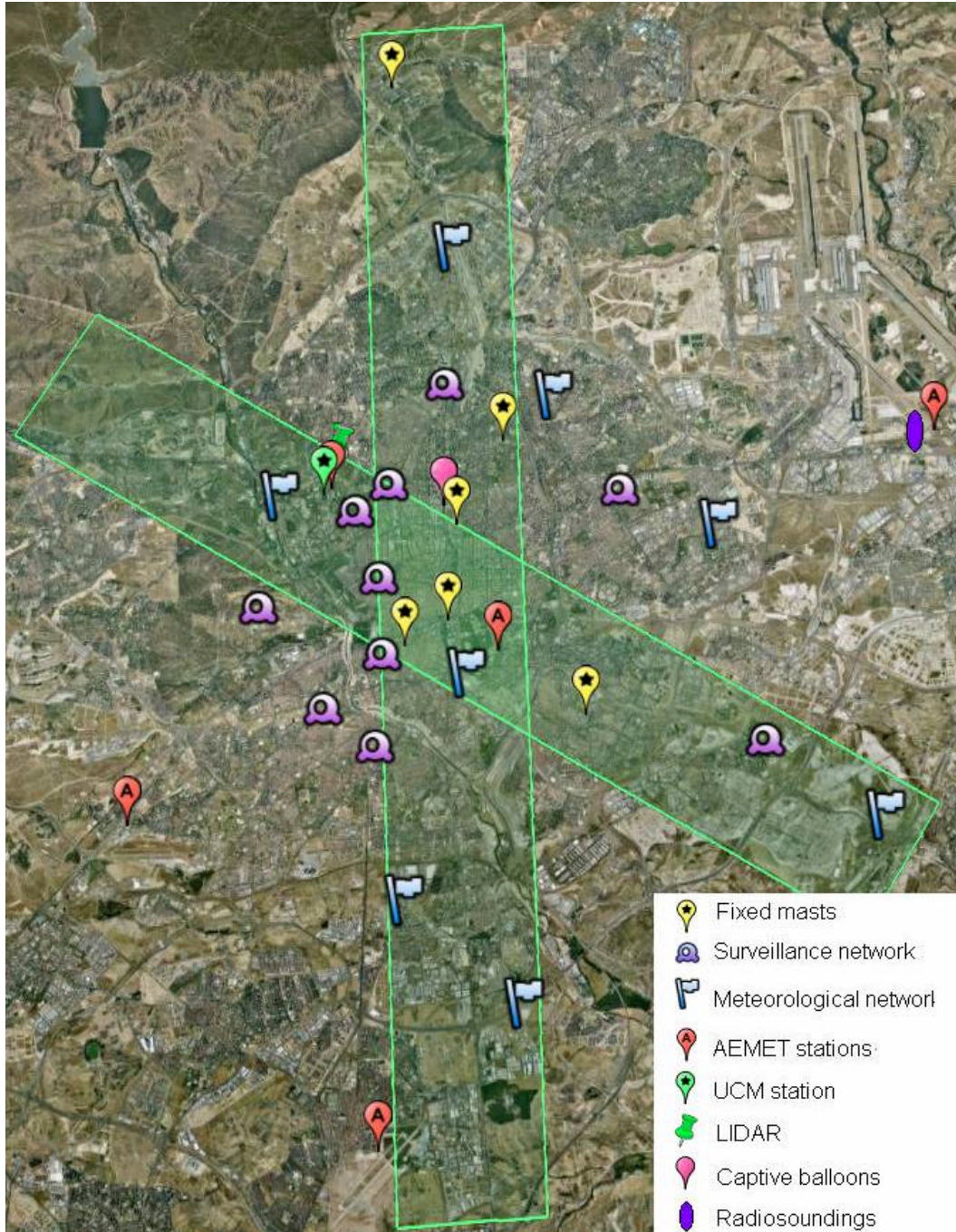


Figure 5.1: AHS flight lines for urbane surface measurements

5.1.2 AIRBORNE DATA ACQUISITION

The aircraft used in the DESIREX 2008 surveys is a C-212-200 property of the Spain Instituto Nacional de Técnica Aeroespacial (INTA). It is a modified aircraft ready to perform remote sensing flight campaigns. This aircraft together with the scanning equipments is depicted in Figure 5.2.



Figure 5.2: **Left:** Platform C-212-200 property of INTA (PATERNINA). **Right:** Airborn hyperspectral scanning sensor.

The AHS sensor is a 80-band radiometer (Figure 5.2), developed and built by SensyTech Inc. (currently Argon ST, and formerly Daedalus Ent. Inc).

AHS spectrometer has been configured with very distinct spectral performances depending on the spectral region considered. It consists of four separate sensors organized in five optical ports. Information regarding the *spectral range* and *spectral resolution* of these PORTs are given in table 5.1.

	PORT 1	PORT 2A	PORT 2	PORT 3	PORT 4
Spectral Coverage (μm)	0.443 - 1.025	1.55 - 1.65	1.9 - 2.55	3.3 - 5.4	8.2 - 13.4
FWHM (nm)(average)	29	90	17	300	450
# Bands	20	1	42	7	10

Table 5.1: AHS spectral configuration for DESIREX 2008 surveys. Due to the capability of the visible and near-to-mid infrared range of electromagnetic field in representation of the ground compositions and discrimination among different compositional units, only the signatures of the first three ports (PORT 1, PORT 2A and PORT 2) are used for the current study.

PORT 1 covers the VIS/NIR range. Its bands have relatively low spectral resolution (28 – 30 nm wide) while the spectral range is continuous from 443(nm) up to 1025(nm).

PORT 2A covers the SWIR range with an isolated band centered at 1.6(μm) with spectral resolution of 90(nm).

PORT 2, has a set of continuous, with relatively high spectral resolution (17 – 18(nm) wide) laying between 1907(nm) and 2558(nm), well suited for soil/geologic studies.

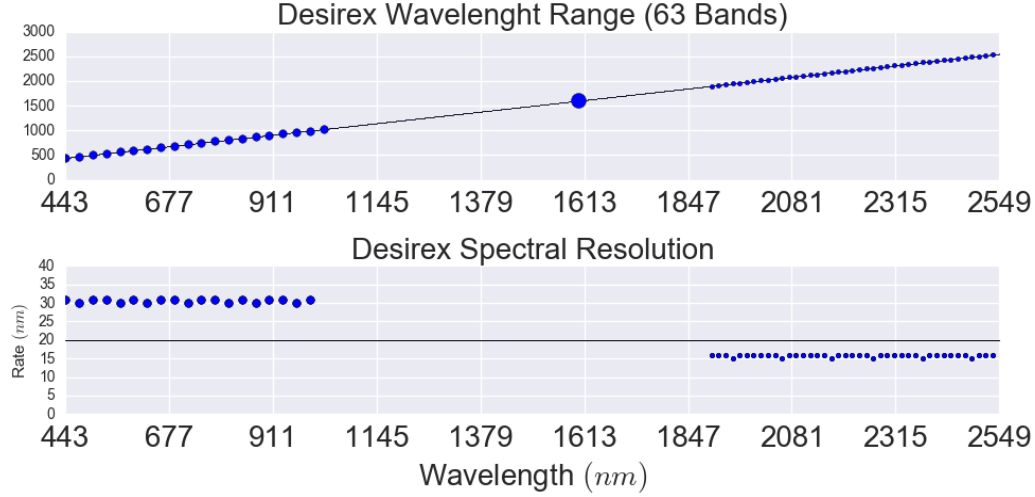


Figure 5.3: Illustration of the spectral range and spectral resolution of each band for the first three ports of the AHS DESIREX dataset. One observes that the bands of DESIREX dataset are measured using different portions of the electromagnetic field and with different spectral resolutions.

PORT 3 covers the MWIR region with 7 bands operating from 3.1 to $5.5(\mu m)$. It gathers the energy through an Indium Antimonide – InSb 7-elements detector array, cooled with liquid nitrogen. The mean bandwidth is $350(nm)$.

Port 4 covers the LWIR photons, from 8.1 to $13.4(\mu m)$, with a 10-element Mercury Cadmium Telluride (MCT) detector. The spectral resolution is $480(nm)$.

Information regarding the spectral range and the *rate-of-change* of spectral resolution across the bands and for the first three ports is visually depicted in Figure 5.3. Within this chapter, we will merely rely on the information provided by these three PORTs for the evaluation of our unmixing algorithms.

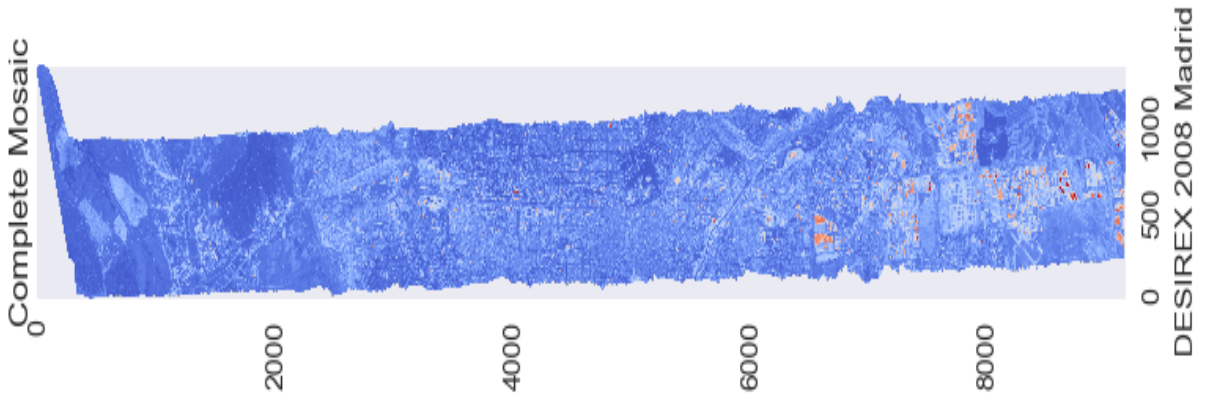


Figure 5.4: DESIREX hyperspectral images acquired through one north-south flight over Madrid in 2008. (The image is rotated by 90°)

All radiometric calibrations have been performed using at-lab measurements. In Figure 5.4 a complete mosaic captured through one north-south flight over the area is depicted.

5.1.3 SPECTRAL CHARACTERIZATION OF URBAN SURFACES

In order to build an urban material spectral library, through the DESIREX 2008 project, spectral signatures over different natural and artificial surfaces were acquired with the ASD FieldSpec-3 by INTA Remote Sensing Area. ASD FieldSpec-3 is a portable and general-purpose spectroradiometer which is illustrated in Figure 5.5.

The Visible/Near Infrared (VNIR) portion of the spectrum, the $350 - 1050(nm)$ wavelength domain, is measured by a 512-channel silicon photodiode array. Each channel is geometrically positioned to receive light within a narrow $1.4(nm)$ bandwidth. The VNIR spectrometer has a spectral resolution (FWHM of a single emission line) of approximately $3(nm)$ at around $700(nm)$. The short-wave Infrared (SWIR), also called Near Infrared (NIR), portion of the spectrum is acquired with two scanning spectrometers.

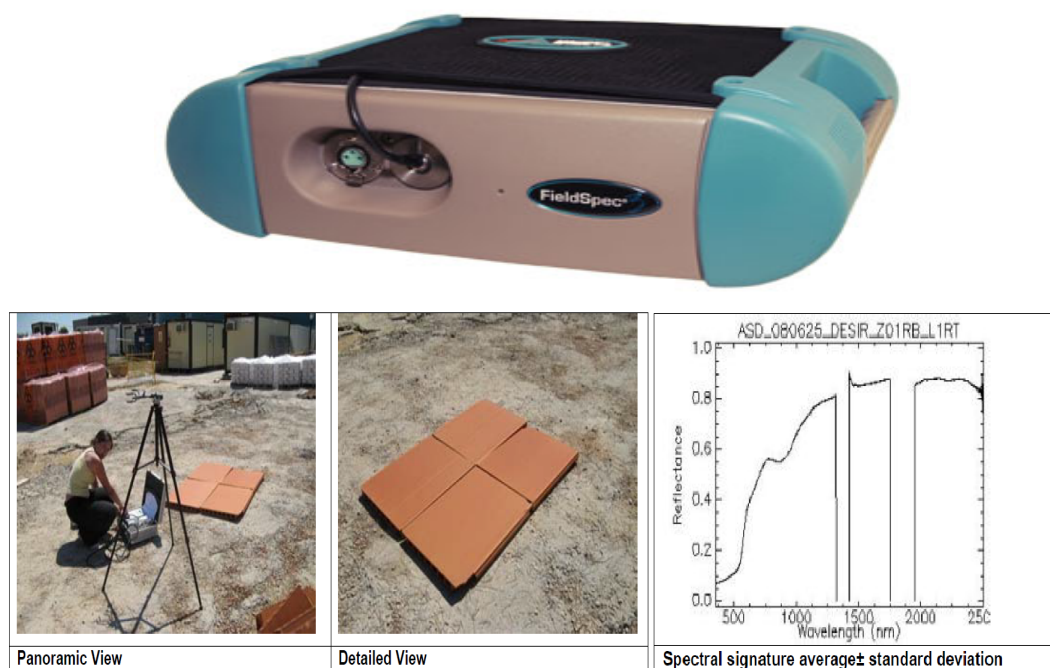


Figure 5.5: ASD FieldSpec3 portable, general-purpose Spectroradiometer.

These differ from the array used in the VNIR in that they measure wavelengths sequentially, rather than simultaneously. Each spectrometer consists of a concave holographic grating and a single thermoelectrically cooled indium gallium arsenide (InGaAs) detector. The gratings are mounted in a common shaft that oscillates

with a period of about 200 milliseconds ($100ms/scan$). Unlike the VNIR, each SWIR spectrometer has only one detector, which is exposed to different wavelengths of light as the grating oscillates. The first spectrometer (SWIR1) measures light between about 900 - 1850(nm); the second (SWIR2) covers the region between about 1700 - 2500(nm). The sampling interval for each SWIR region is about 2(nm), and the spectral resolution varies between 10(nm) and 12(nm), depending on the scan angle at that wavelength.

In particular *Pavement, Asphalt (Parking), Red Brick, Concrete, Asphalt (Composition), Metallic Mesh, Red Tile, Asbestos, Steel, Granite, Green Grass and Bare Soil* were among materials carefully measured during the campaign. Figure 5.6 represents an exemplary image for each one of interesting materials.



Figure 5.6: Man-made materials considered for endmember dictionary construction.

In this study, spectral signatures of 11 distinct materials are considered. The signatures of these materials registered by the ASD equipment are illustrated in Figure 5.7.

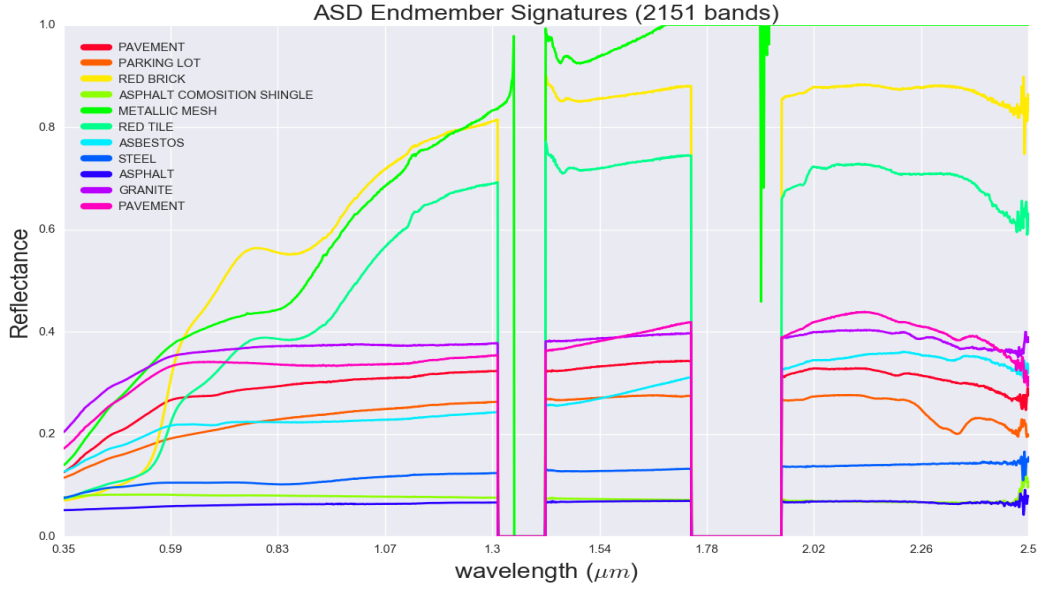


Figure 5.7: Spectral signatures of Endmembers which are registered through an in-situ procedure using the ASD equipment.

These 11 signatures which are registered based on an in-situ acquisition procedure are assumed to be the main materials that cover the urban scene in DESIREX dataset. The spectral range and the *rate-of-change* of spectral resolution (across the bands) of these endmember signatures are illustrated in Figure 5.8, where we can see a considerable difference between them and the characteristics of the pixel signatures of DESIREX dataset represented in Figure 5.3.

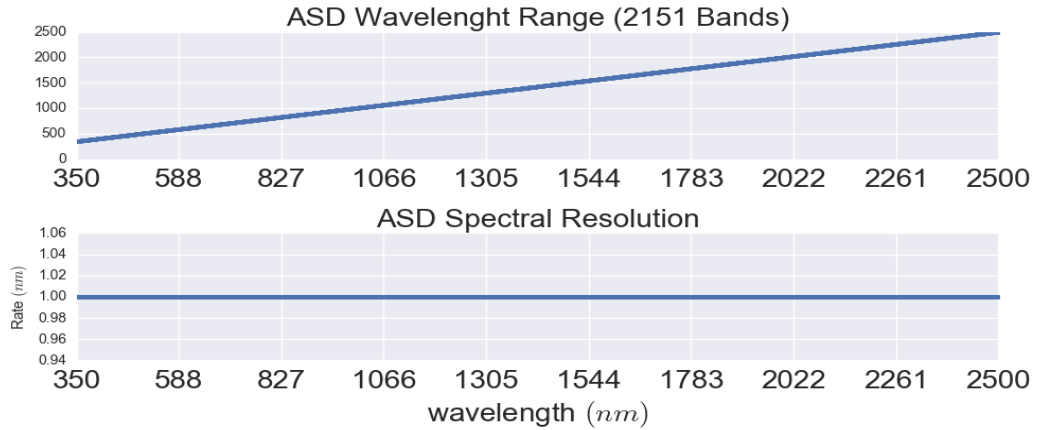


Figure 5.8: Illustration of the "wavelength range" and "wavelength sampling rate" of the in-lab measured ASD endmembers across the electromagnetic field. One can see that the bands of these endmembers have uniformly been sampled across the range $.35 - 2.5(\mu m)$, with the sampling rate of $1(nm)$ which makes them incomparable to the spectral characteristic of the Desirex dataset.

5.2 On the valid choice of Endmember dictionary

Considering the difference between the spectral characteristics of Desirex dataset and the ASD-based endmember dictionary, also due to the fact that for the unmixing of the dataset we are considering supervised and unsupervised procedures, one faces different possible strategies to prepare the required endmember dictionary and conduct the unmixing. In particular one can:

1. Pre-process the ASD dictionary and use it for the supervised OT-unmixing algorithm.
2. Manually extracts endmembers from the Desirex dataset and use them for the supervised OT-unmixing algorithm.
3. Using the Exact endmembers estimation part of SAGA+ to build the endmember dictionary and then use it for OT-unmixing to produce the abundance estimations.
4. Pre-process the ASD dictionary and use it for the SAGA+ algorithm (therefore skipping the endmember extraction part in SAGA+).
5. Manually extracts endmembers from the Desirex dataset and use them for SAGA+ algorithm (therefore skipping the endmember extraction part in SAGA+).
6. Taking a complete unsupervised strategy and let SAGA+ to find the endmembers from the Desirex dataset and estimate the abundances.

Considering the fact that for each of the above mentioned strategies one needs also to set a few number of model parameters (choice of the kernel in SAGA+, level of sparsity of the representations, grouping and weighting values of the endmembers in OT-unmixing etc.), an extensive amount of tests needs to be performed and comparisons to be taken in order to find the best unmixing scenario and validated endmember dictionary. This complete work is in progress and in this PhD, we will represent one unmixing result per strategy using a fix parametrization setup in order to let comparisons between performances of strategies possible. Also, in order to provide a comparison between the proposed methods with other state-of-the-art techniques we consider the performance of the sequential least squares programming (SLSQP) on the same dataset and using three endmember dictionaries presented above.

In the remaining part of this chapter we first briefly introduce the preprocessing procedure for the ASD endmembers and introduce the approach to manually generate dictionary of endmembers under the supervision of human experts in the field of geoprocessing. We also introduce a dictionary of endmembers generated by the endmember estimation part of SAGA+. Having these three endmember dictionaries, in a second part the results of the unmixing procedures (using these three dictionaries) according to the above mentioned strategies are being presented.

5.2.1 ASD Preprocessing

Before using the data and in-lab signatures for conducting unmixing we need to guaranty that our dataset (issued from airborne acquisitions) and the endmember dictionary (issued from lab measurements), are representing the same wavelength ranges. Their spectral resolutions need also to be comparable. When comparing Desired and ASD spectra, we first observe that the wavelength range of DESIREX is slightly superior than $2.5(\mu m)$ while the last wavelength registered for each ASD signature is $2.5(\mu m)$. This forces us to remove the last three bands of the DESIREX dataset, leaving us with 60 spectral bands.

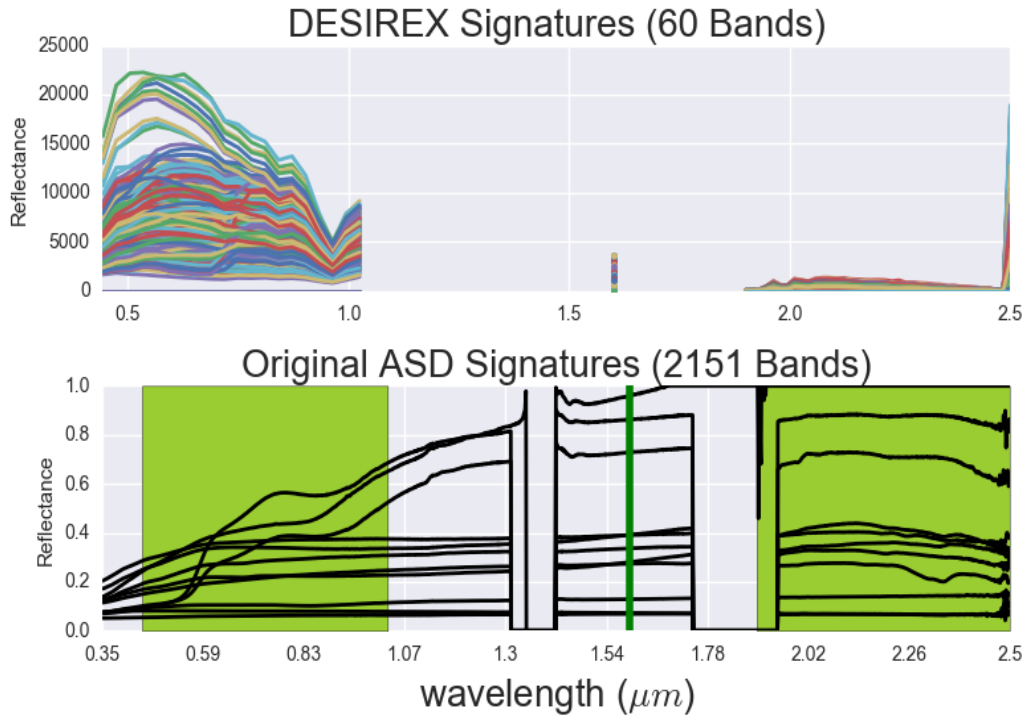


Figure 5.9: Spectral signatures of a sample of 1000 pixels randomly selected from Desirex dataset versus the original ASD endmembers.

ASD cut-off

In Figure 5.9 are depicted examples of spectra issued from DESIREX and ASD. On this figure, one can observe a large difference between their wavelength coverage across the electromagnetic field and also their spectral resolution. Therefore some ranges in ASD (out of the green area in Figure 5.9) must not be considered for the means of unmixing Desirex data. By cutting those excessive ranges off from the endmember signature profiles we will have a new endmember dictionary consisting of endmember signatures with 1185 bands, represented in Figure 5.10.

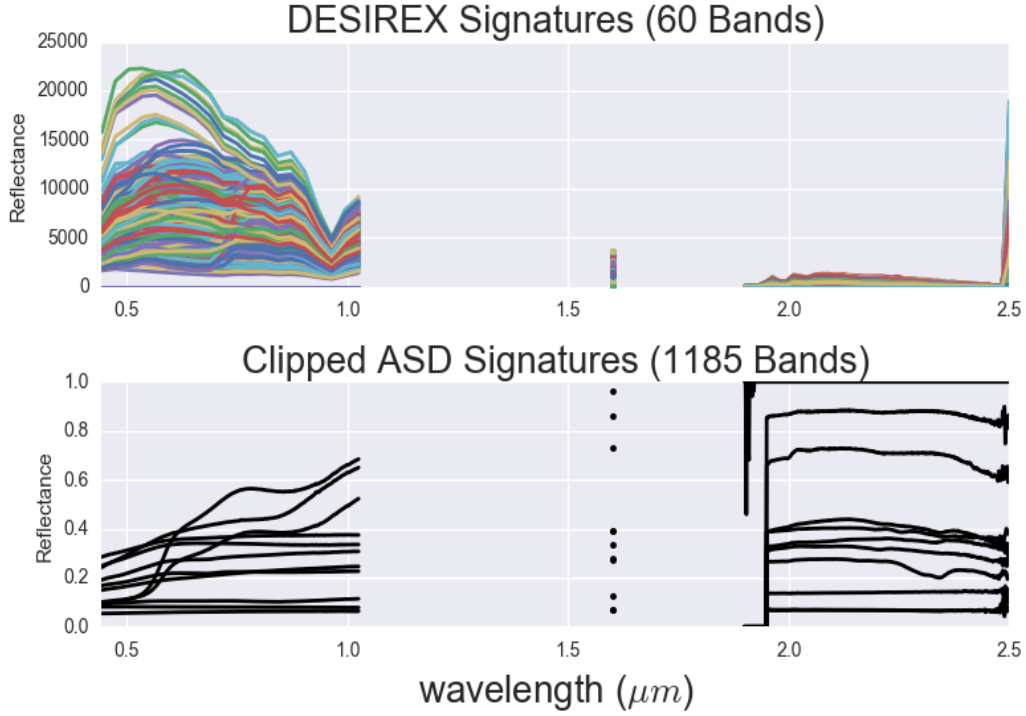


Figure 5.10: Spectral signatures of a sample of 1000 pixels randomly selected from Desirex dataset are plotted versus the clipped ASD endmembers.

ASD downsampling

Once DESIREX and ASD endmember dictionaries are in accordance from the point of view of their wavelength range coverages, we need also to make them comparable in terms of spectral resolution in order to reduce the spectral dimension of the ASD endmembers from 1185 to 60 bands. Figure 5.11 represents the endmember signatures of Figure 5.10 after being downsampled. One can see that by uniformly downsampling the endmember profiles related to PORT1 and PORT2, the overall shape of the endmember signatures remain intact.

Even after performing the above mentioned preprocessing steps, one can see in Figure 5.11 the large level of incompatibility in terms of reflectance, between the overall signature profile of the Desirex dataset across the three PORTs and the overall signature profile of the ASD endmember dictionary. Some scaling parameters related to different sensors have been used which are different from other sensors. This disturbs the shape of all spectra along the wavelength range. Unfortunately in the present study these parameters were not available. By knowing this limitation and assuming that these parameters must be available to bring ASD dictionary and the Desirex in accordance, for the sake of comparison of our unmixing models performances we use the current dictionary of ASD endmembers (adjusted and downsampled) as one of the available sources of knowledge regarding the materials

within the scene and compare their results with other possible dictionaries. Let us now introduce another endmember dictionary, prepared based on expert knowledge.

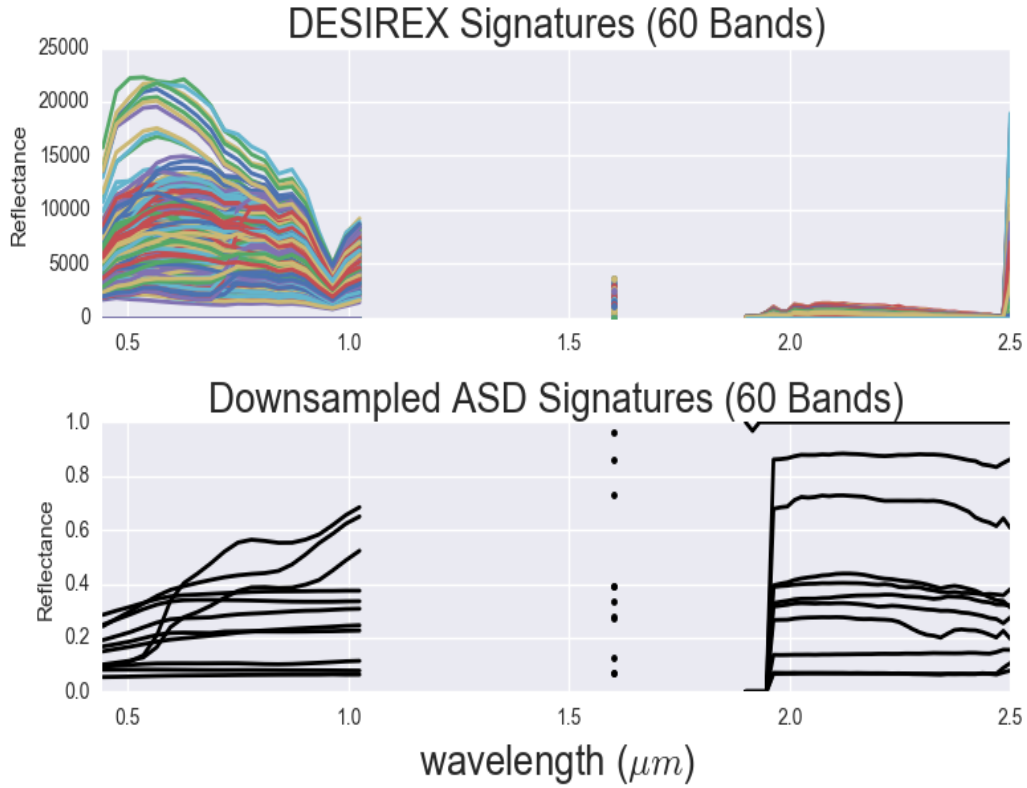


Figure 5.11: Spectral signatures of a sample of 1000 pixels randomly selected from Desirex dataset versus the downsampled ASD endmembers.

5.2.2 Expert Endmember Extraction

Another widely used approach towards hyperspectral unmixing is to find maximally pure pixels from a dataset and take them as endmembers. This approach can be taken manually and with the expert knowledge about the distribution of materials within the scene. Considering the fact that the ASD endmember dictionary, even after the pre-processing procedure does not look completely in accordance to the type of materials that are common in the desirex dataset, we also used a manual approach of endmember extraction based on a human expert observation. Pixel vectors related to highly distinct materials within the scene have been extracted to be used as a dictionary of endmembers. In order to account for material variability, for each material two representatives endmembers were selected in order to build-up a dictionary of 24 signatures. This Endmember dictionary is represented in Figure 5.12.

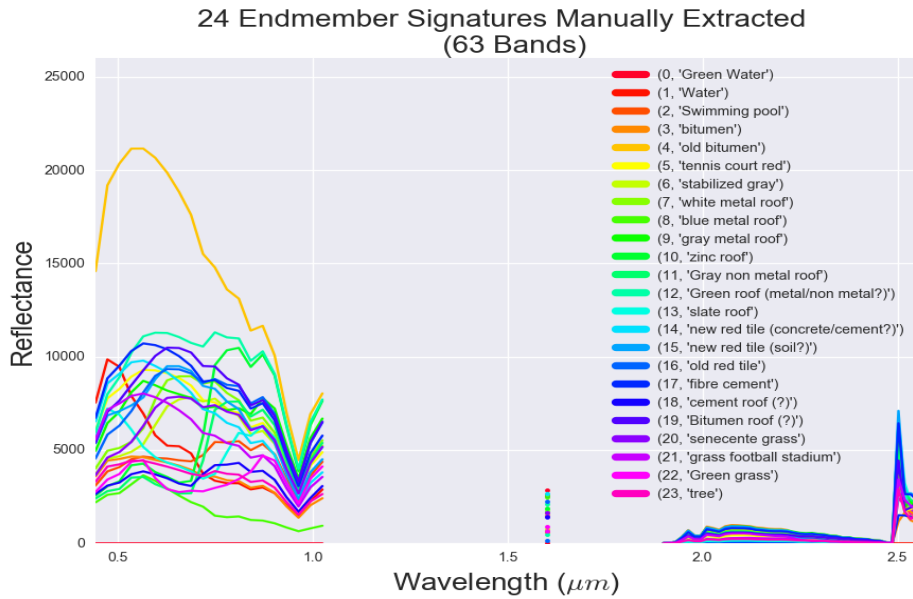


Figure 5.12: Dictionary of the manually extracted endmembers. Based on expert observations, pixel vectors related to highly distinct materials within the scene have been extracted to be used as a dictionary of endmembers. In order to account for material variability, for each material two representatives endmembers were selected in order to build-up a dictionary of 24 signatures.

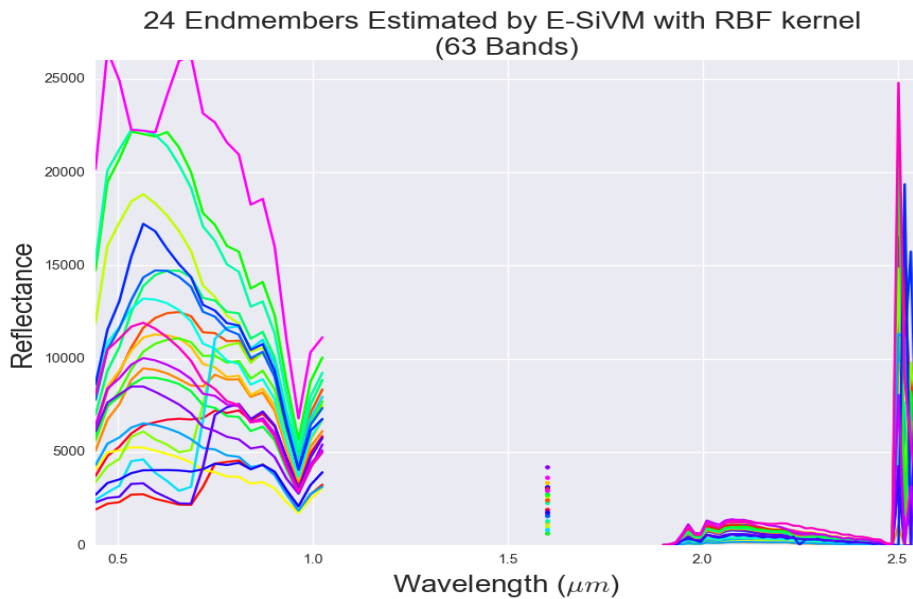


Figure 5.13: Using the exact simplex volume maximization (E-SiVM) procedure a dictionary of 24 atoms with maximum variability among their signature profiles has been generated. One can observe even larger variability among endmember signatures while comparing this dictionary to the one extracted manually.

5.2.3 Endmember Extraction with E-SiVM

In order to evaluate the performance of SAGA+ in automatic estimation of pure endmembers, we use the exact simplex volume maximization (E-SiVM) procedure to generate another dictionary of endmembers. For this we chose an RBF kernel with kernel parameter $\sigma = 2 \times 10^{-5}$ in order to get maximum variability among the signatures of 24 different pixels from a window of size 601×601 pixels. The choice is based on our results on various experimental tests of ESiVM on the data. Of course an automatic selection of these parameters (for example using cross validation) would be better. This is the scope of a current work. This endmember dictionary is represented in Figure 5.13.

By comparing the Figures 5.12 and 5.13 one observes that the estimated endmembers with E-SiVM procedure can represent even larger diversity among the existing materials in comparison to the manually selected dictionary. This difference is more obvious when comparing the performance of **ESiVM** dictionary with a dictionary of randomly selected endmembers from the same dataset (that is shown in Figure 5.14).

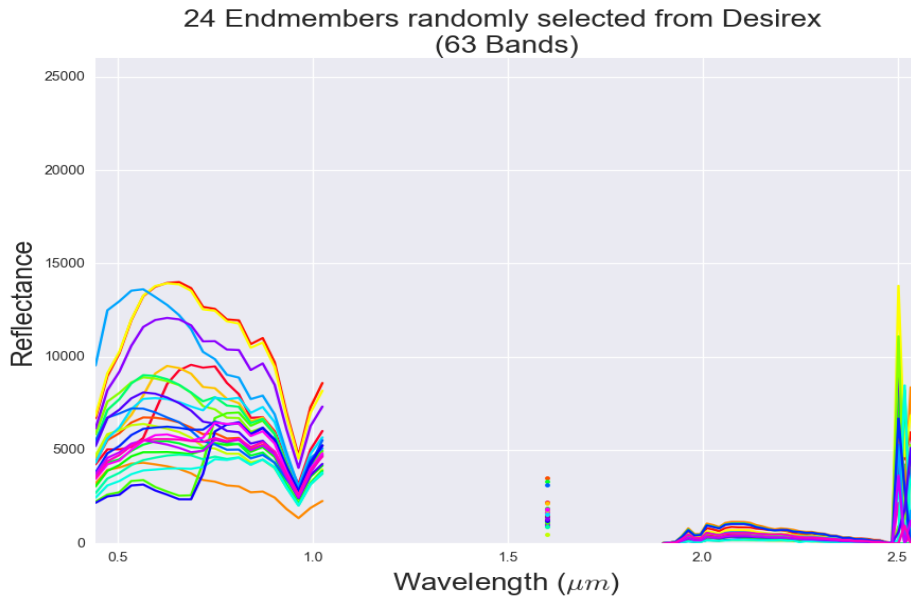


Figure 5.14: Dictionary of 24 randomly selected signatures from 601×601 window of Desirex dataset.

In order to make a visual comparison in Figure 5.15 we provided two illustrations consisting of the superposition of the **Expert** dictionary with both randomly generated dictionary and the one generated by **ESiVM** procedure. In order to have a quantitative analysis, one could select randomly a large number of dictionaries (consisting of thousands of signatures) and compute distances between these dictionaries and endmembers (minimal distance between endmembers and all permutations of

spectra in dictionaries). Unfortunately, with 24 endmembers, this would require $24! \sim 10^{23}$ distance computations between endmembers which is in practice impossible. Therefore only a visual inspection can be done here. As can be seen in Figure 5.15, **ESiVM** dictionary seems to be more in accordance with the **Expert** one (more variability is captured) than random ones. It is indeed obvious that the variability in endmembers extracted by ESiVM is higher.

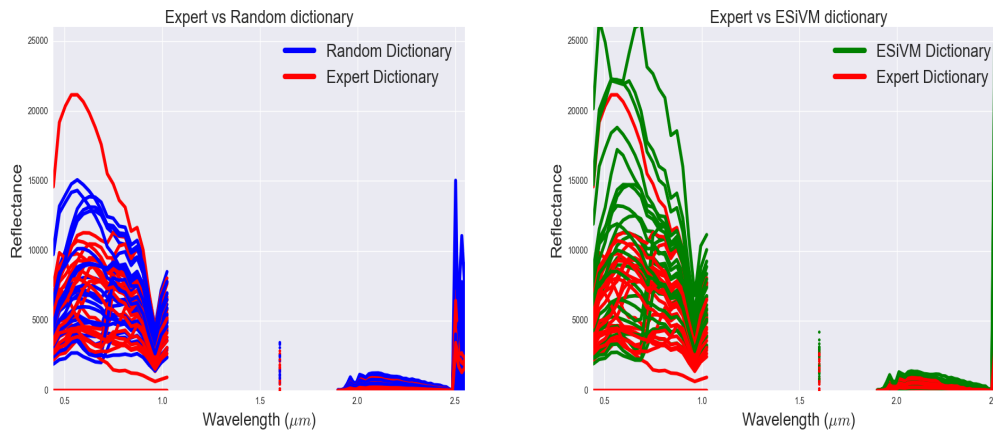


Figure 5.15: Random and ESiVM dictionaries compared to the Expert dictionary. The result of ESiVM shows more fidelity to the Expert dictionary in capturing more varieties among the endmember signatures.

In the next section we introduce preliminary results of our two spectral unmixing algorithms according to the six strategies mentioned above. Results provided from a state-of-the-art procedure called sequential least squares programming (SLSQP [122]) are also included in order to compare the performance of our algorithms from the point of view of computational time and the accuracy of the unmixing.

5.3 Unmixing Experiments on Desirex

5.3.1 Experiment conditions

Following the mentioned strategies in section 5.2, we conduct the abundance estimation tests in three phases. In the first phase (5.3.2) we run a series of unmixing tests by the **OT-unmix** procedure (Strategies 1, 2 and 3) while endmember dictionary is provided through one of the three mentioned ways: 1) The **ASD** dictionary 2) the dictionary manually extracted from the AHS Desirex dataset and 3) The estimated dictionary from AHS Desirex and using E-SiVM. For the sake of clarity from now we refer to them as **ASD**, **Expert** and the **ESiVM** dictionaries. As for the $C1$ cost matrix, we chose the 24 atoms of the dictionary to be grouped based on K-means clustering results into 5 different groups.

In the second phase (presented in section 5.3.3) we repeat the same procedure with the sequential least squares programming (SLSQP), in order to provide a comparison between two supervised scenarios. This leads to Strategies 7, 8 and 9. Finally in the third phase (presented in section 5.3.4) we chose to test the kernel-based SAGA+ while using the three dictionaries. This leads to Strategies 4, 5 and 6. For all the experiments in this phase we choose an RBF kernel with the kernel parameter $\sigma = 2 \times 10^{-5}$ (with which SAGA+ produces mostly distinguishable abundance maps). These strategies are synthesized in table 5.2.

Experiments have been performed on a window of 601×601 pixels extracted from the full Desirex mosaic. This window is assumed to contain a diverse number of interesting urban materials and is illustrated in Figure 5.16.

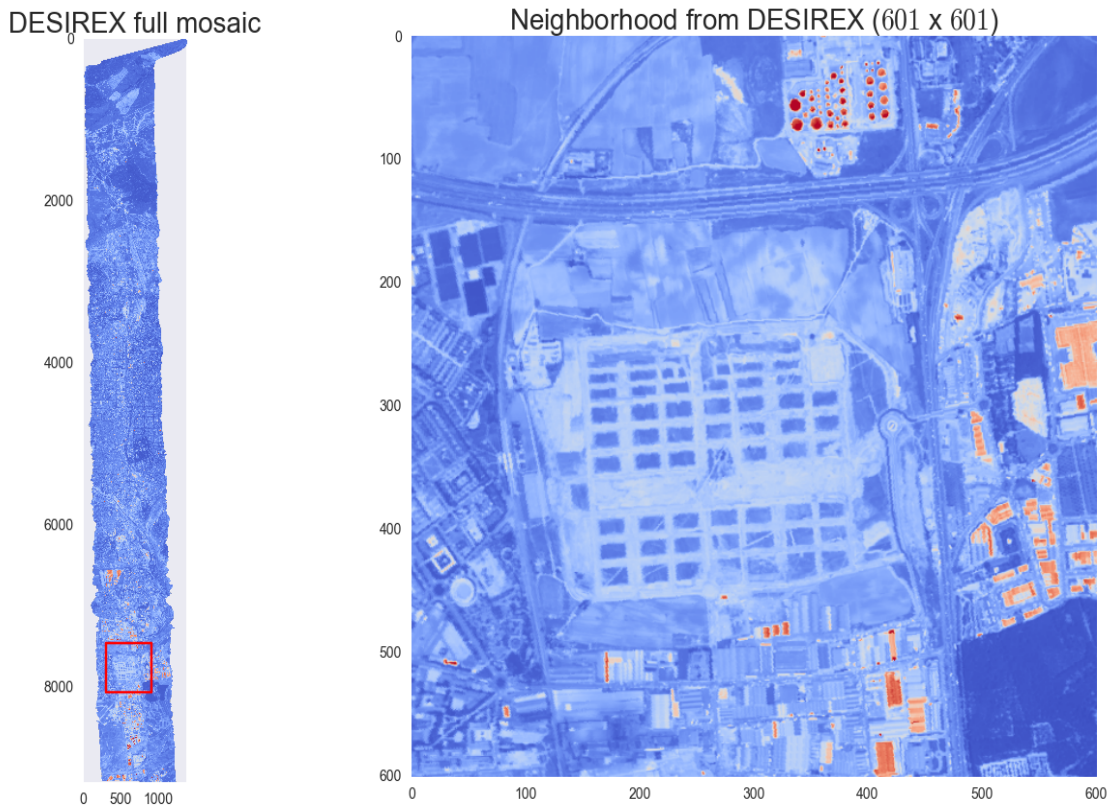


Figure 5.16: Window of 601×601 (pixels) extracted from the main mosaic on which all tests are performed.

K-means clustering for endmember grouping and estimation of the level of sparsity

For these preliminary experiments, we decided first to group the 24 endmembers on several clusters in order to evaluate our capacity to unmix the main urban structures. To be consistent, the level of sparsity λ of the SAGA+ procedure will be set to this

number of clusters.

The choice of the ideal number of clusters is based on the global inertia η_m written for a clustering in m classes as:

$$\eta_m = \sum_{i=1}^m \sum_{j=1}^k w^{(i,j)} \|x^{(i)} - \mu^{(j)}\|_2^2 \quad (5.1)$$

where i is the sample index, j represents the cluster whose centroid is μ . Parameter $w^{(i,j)} = 1$ if the sample $x^{(i)}$ is in cluster j and equals zero otherwise. By increasing the number of clusters m one can observe a decrease in η_m . The principle is to find the optimal number of clusters where the increase in m does not change the distortion in a considerable amount. This principle is usually referred as the Elbow test. In practice we found that a grouping in $m = 5$ clusters was an optimal choice.

The results of this test performed on two endmember dictionaries (**Expert** and **ESiVM**) are visible in Figure 5.17 where one can observe the decreasing rate of inertia (η) when the number of clusters grows.

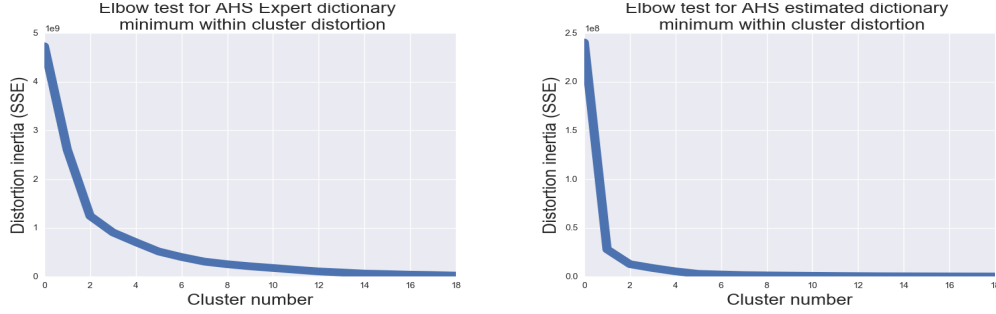


Figure 5.17: Left: Result of Elbow test on a series of k-means clusterings with increasing number of clusters performed on the endmember dictionary of manually selected atoms from the AHS Desirex dataset. **Right:** Result of Elbow test on a series of k-means clusterings with increasing number of clusters performed on the endmember dictionary generated by performing E-SiVM procedure on the AHS Desirex dataset.

Dict \ Model			
	OT-Unmixing	SAGA+	SLSQP
ASD	ST-01	ST-04	ST-07
Expert	ST-02	ST-05	ST-08
ESiVM	ST-03	ST-06	ST-09

Table 5.2: Different strategies introduced in section 5.2 which relates different methods of unmixing considered in this chapter to the available endmember dictionaries.

☞ All tests performed using an intel(R) Core(TM) i7 2.4 GHz platform with 16 GB RAM.

Let us now turn to the analysis of abundance maps.

5.3.2 Abundance Estimation with OT-Unmixing

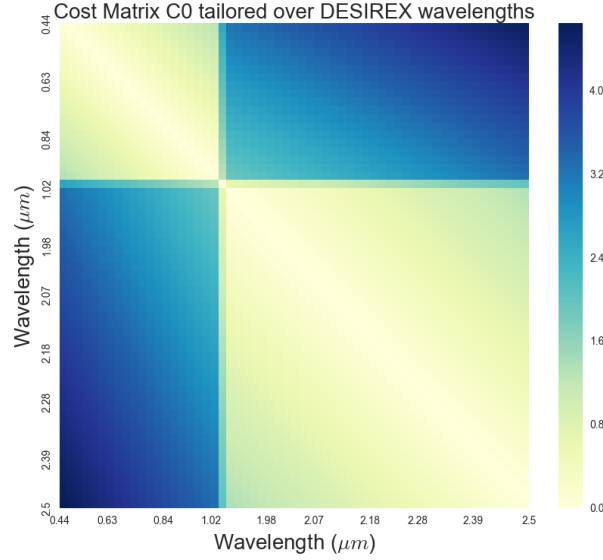


Figure 5.18: In order to set cost matrix $C0$ in alignment to the characteristics of the dataset, we build $C0(63 \times 63)$ as the square euclidean distance over the spectral values of the AHS Desirex dataset.

In the first phase of experiments the supervised OT-unmixing algorithm is used for producing abundance maps using the three mentioned endmember dictionaries. This comprises strategies 1 to 3 represented in the first column of table 5.2. In order to parametrize the model, for these tests we defined the $C0$ cost matrix in alignment to the spectral characteristics of the dataset. In particular we build $C0(63 \times 63)$ as the square euclidean distance over the spectral range of the AHS Desirex dataset. This cost matrix is illustrated in Figure 5.18 has been used in strategies 1 to 3. As we saw in chapter 4 this cost function gathers all the costs for moving one spectral value to another across the signature spectra for computing the differences between two signatures μ_1 and μ_2 .

As for the $C1$ cost function which represents the relationship of the endmembers belonging to the same group (introduced in 4.2.3), we chose to use the clustering information of each endmember dictionary (presented in section 5.3.1). $C1$ provides *a priori* information regarding the groups to which different signatures of the endmember dictionary belong. Here and for each strategy, the total number of signatures within each dictionary has been associated to 5 groups, assumed based on

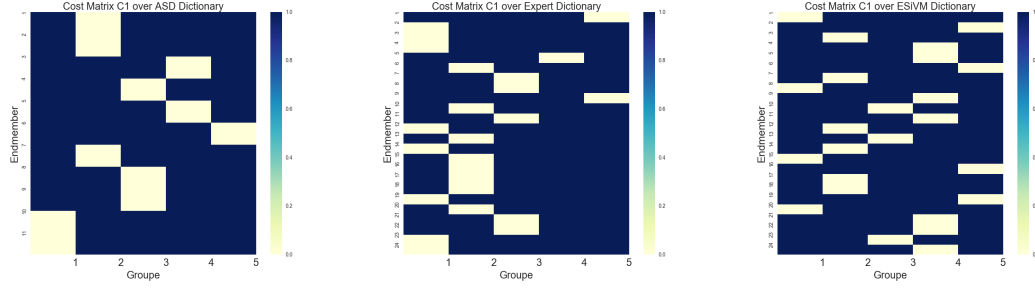


Figure 5.19: $C1$ cost function for three dictionaries. $C1$ provides *a priori* information regarding the groups that different signatures of the endmember dictionary belong to.

the clustering result of K-means algorithm on each dictionary.

Figure 5.19 represents these three cost functions, used as inputs of our unmixing procedure. By comparing the results of clustering on the two endmember dictionaries (**Expert** and **ESiVM**) one can see that both are grouped in a more or less the same ratio. The number of endmembers belonging to each group for each dictionary is given in table 5.3.

	G1	G2	G3	G4	G5
C1 matrix over Expert	8	8	5	1	2
C1 matrix over ESiVM	4	6	3	7	4

Table 5.3: Number of endmembers grouped into the 5 different clusters based on **Expert** and **ESiVM** dictionaries.

Also in order to give a visual representation of the characteristics of the signatures in each group and comparing similar groups over the two dictionaries **Expert** and **ESiVM**, in Figure 5.20 the result of clusterings of these dictionaries are illustrated. By focusing on the intensity values of the first 21 bands (as the result of PORT1) of the endmember signatures, pairs of groups with maximum similarity among their signature envelopes are illustrated in the same colors.

By matching the information of table 5.3 and Figure 5.20, one observes that the G1 and G2 groups over the **Expert** dictionary embed more signatures. As for **ESiVM**, these groups are G2 and G4. This means that the endmembers of these groups share some common characteristics thus, most probably referring to similar materials. The same analysis goes for the pairs (G4,G3) and (G5,G5) across the two dictionaries.

As for the vector of priors a_0 , we chose to set their value as $\frac{g_i}{t}$ where g_i is the total number of endmembers belonging to the i th group and t is the total number of atoms in each dictionary. As mentioned in section 4.2.3, this prior vector defines weights on the relationship of the endmembers belonging to the same group. Having these parameter settings fix for each strategy we made three experiments on the mentioned

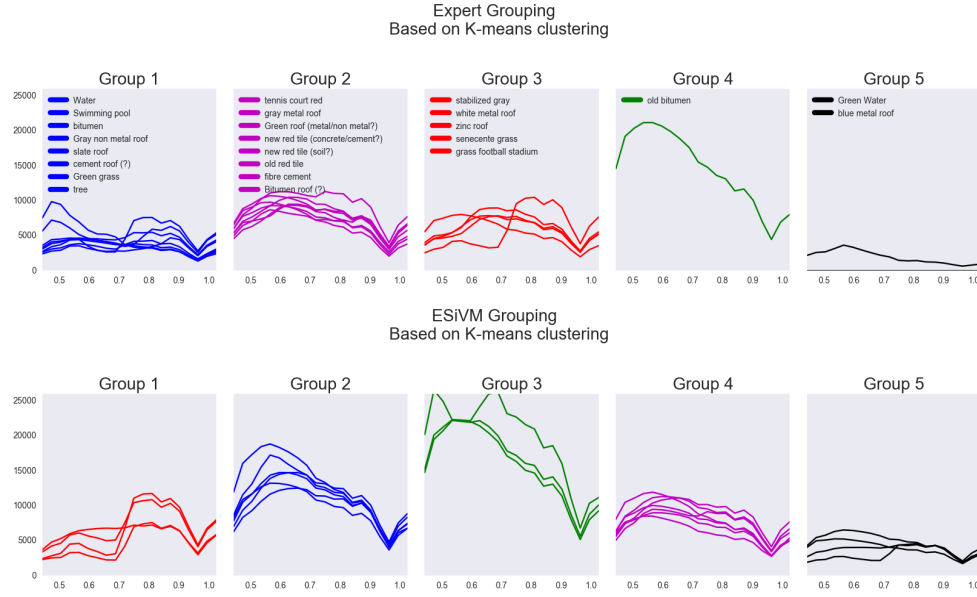


Figure 5.20: Clustering of two dictionaries. Groups which considered to share similar semantic relationships in two dictionaries are plotted in the same color.

area (Figure 5.16) of Desirex. The abundance maps as the result of these tests are depicted in Figure 5.21.

As expected, the estimations based on the **ASD** dictionary are not very informative, though they exhibit interesting spatial structures. Unfortunately, because of the lack of consistency between **ASD** and **DESIREX** spectra, the quality of such abundances is limited. Regarding the other two experiments using the **Expert** and **ESiVM** dictionaries, interesting patterns are visible in each case with different degrees of emphasis on particular structures depending mainly on different ways of grouping the atoms of each dictionary. It should be notified that in general the abundance maps produced based on the **ESiVM** dictionary, which is the result of automatic endmember estimation of our **SAGA+** procedure, seems to outperform the results of unmixing based on the **Expert** dictionary. It is interesting also to notice the relative similarity between the couple (G1, G2) over the **Expert** dictionary and (G2, G4) over the **ESiVM** dictionary, which indeed share similar patterns in representing the intensities.

In the absence of expert interpretation of the maps, one could conclude that our abundance maps seem to reveal consistent structures in this difficult context, which is a good starting point for further investigations. These results will be compared with state-of-the-art techniques in the next phases.

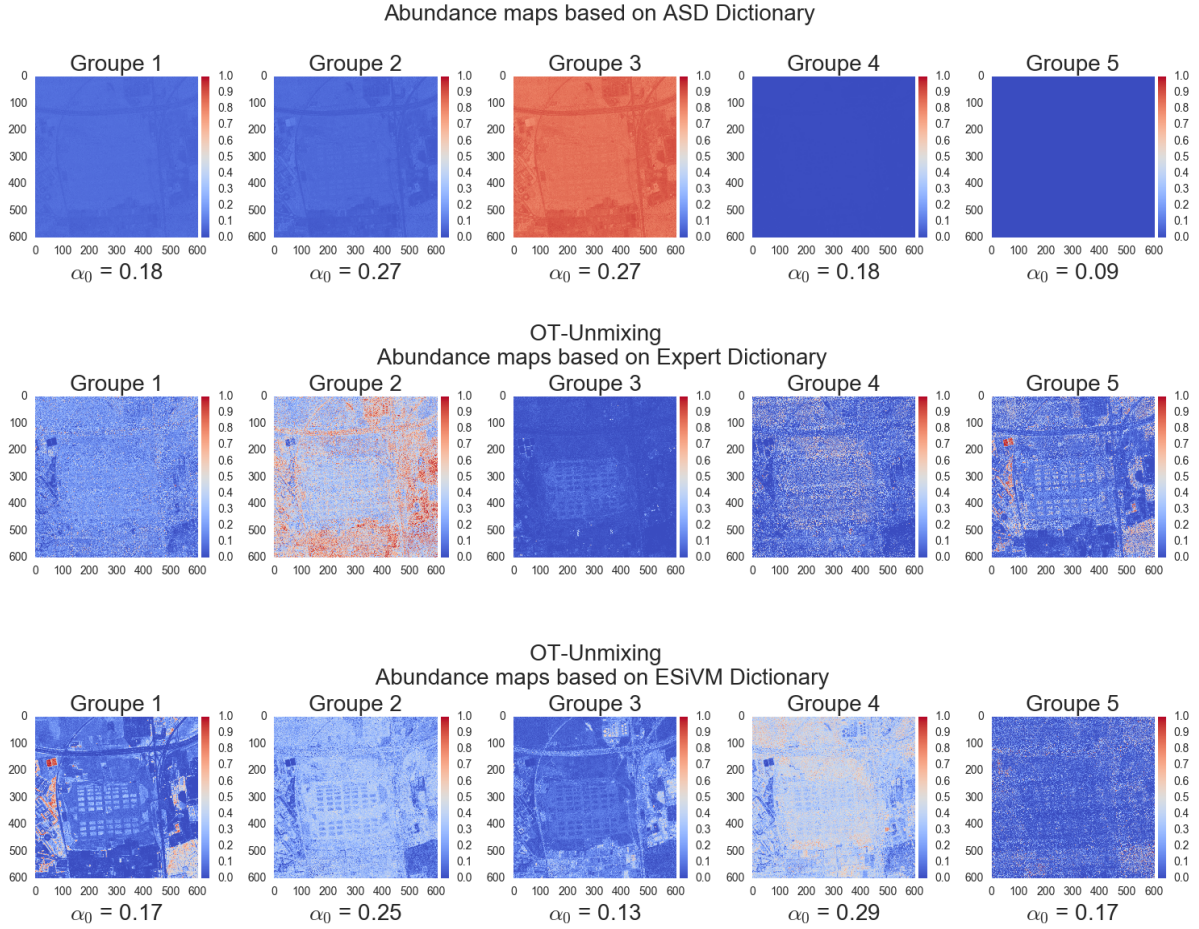


Figure 5.21: Abundance maps produced by OT-unmixing and according to the first 3 strategies presented in the first column of table 5.2.

5.3.3 Abundance Estimation with SLSQP

In the next phase and in order to provide a comparison between the performance of OT-unmixing with a common state-of-the-art technique, we compare our technique with the sequential least squares programming (SLSQP) to perform the inversion procedure on the same dataset and using the same three available dictionaries. The results of these tests are depicted in Figure 5.22.

Comparing the results with those in Figure 5.21 one can easily observe that the abundance maps produced by OT-unmixing are of much higher spatial information. It should also be considered that the computation time of SLSQP on this data was higher than OT-unmixing by the order of 3, which means from about 1.5 hours in the case of OT-unmixing to more than 4 hours in SLSQP.

Though it is complicated to make a quantitative interpretation of these results, when comparing Figures 5.22 and 5.21, it appears clearly that spatial maps issued from our proposed technique reveal more consistent structures. This is to our opin-

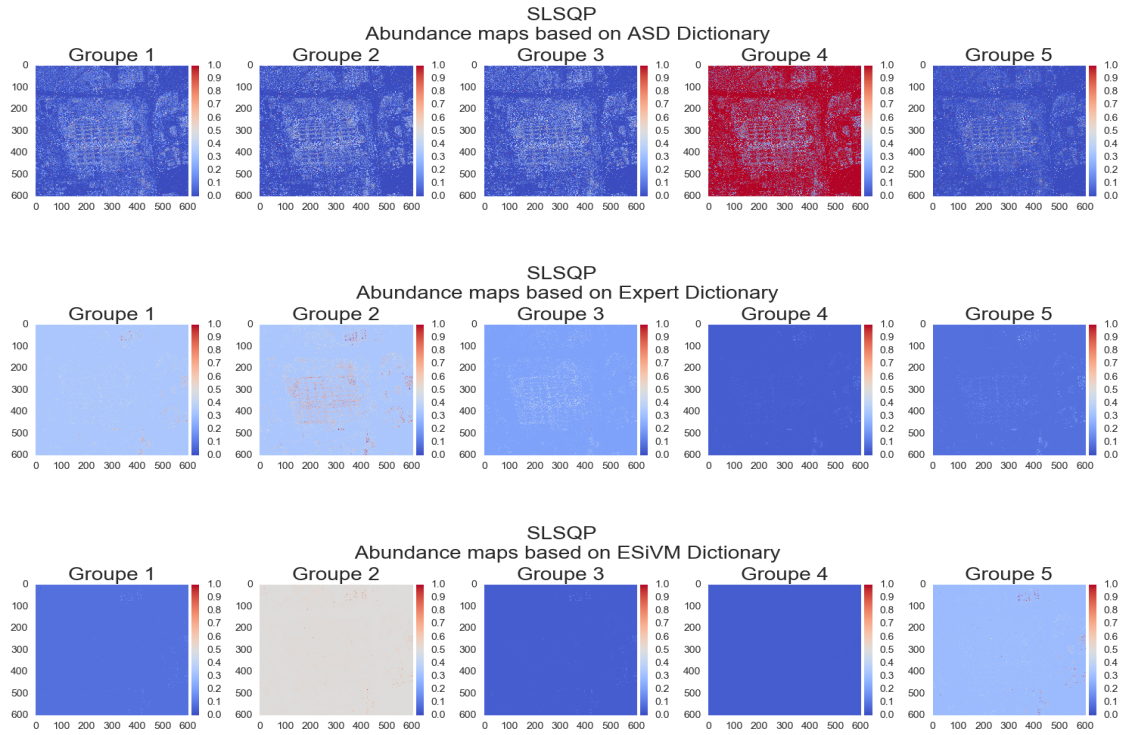


Figure 5.22: Abundance maps produced by SLSQP algorithm and according to the last 3 strategies presented in the last column of table 5.2.

ion a reliable behavior. In addition, the different groups extracted in Figure 5.21 are also more heterogeneous using OT-unmixing, which tends to promote a better separability of our approach.

Strategy \ Distance			
	Minimum Dist	Maximum Dist	Average Dist
ST-02	48444	158785	93369
ST-03	52933	117689	78718
ST-05	50255	122602	82926
ST-06	34169	150426	94135
ST-08	5636	112494	63311
ST-09	124	60196	30227

Table 5.4: Statistics of the pairwise distance calculation over the 5 abundance groups produced as the result of each of 6 different strategies. The pairwise distances of each strategy over which we calculated the average value in the last column of this table are illustrated as a separate upper triangular matrix in Figure 5.23.

This is probably due to the non-linear property of our approach that enables to deal with the complex mixing phenomena generated by urban structures. To justify

this point, we have computed the average *pairwise difference* between all groups estimated with OT-Unmixing and SLSQP techniques. More precisely, for each unmixing strategy $ST_i, i = \{2, 3, 5, 6, 8, 9\}$ where i refers to the strategy used (represented in table 5.2), we have computed the average difference $D_i = \frac{1}{|C_5^2|} \sum_{k, \ell} \mathcal{D}(A_{i,k}, A_{i,\ell})$ such that $k < \ell$ and $\mathcal{D}(A_{i,k}, A_{i,\ell})$ is the distance between abundance map of group k and technique i and the abundance map of group ℓ and technique i . The number of pairwise combinations among the 5 groups $|C_5^2| = \frac{5!}{2!3!} = 10$.

This measures the heterogeneity between the different groups and this is therefore an indicator of the ability of each technique to separate various kind of mixtures (the higher is this difference, the more diverse mixtures are separated).

Of course we would like to point out here that this criteria is not *strong* in the sense that even with reliable values, this does not state that the unmixing is correct. This only quantifies the heterogeneity between abundances and therefore this relies on the assumption that all abundance maps are spatially different (which is quite rational for 5 classes). These results are given in table 5.4 as well as a visual representation of the *pairwise difference* matrices related to 6 strategies (the results of those strategies over the **ASD** dictionary were excluded because of their poor performance) are illustrated in Figure 5.23. As one can observe, this difference is considerably higher for OT-related strategies (ST-02 and ST-03) comparing to the SLSQP-related ones (ST-08 and ST-09). This simple quantitative interpretation strengthens the fact that our approach seems to have a better accuracy in unmixing urban patterns. There is no doubt however, that a more complete comparison with ground truth data could validate this. This is the scope of a current work.

Let us now turn to the experiments using the complete SAGA+ technique (end-member extraction and abundance estimation).

5.3.4 Endmember Extraction and Abundance Estimation with SAGA+

In the last phase of tests, we used the sparse projection part of SAGA+ in order to produce abundance maps using the same three dictionaries.

As for this test we set the RBF kernel parameter $\sigma = 2 \times 10^{-5}$ and the sparsity $\lambda = 5$. This choice of λ is made to make the 5-sparse maps as the result of SAGA+ comparable to the 5-groupings of the OT-Unmixing and the other method.

The 5-sparse abundance maps based on the greedy selector and sparse projector algorithm are represented in Figure 5.24. Here again, the same kind of conclusion than using the OT-unmixing procedure (Figure 5.21) holds. Even if abundances are inconsistent using **ASD** endmembers, the spatial groups extracted using **Expert** and **ESiVM** dictionaries are consistent, visually even less noisy than using OT-based results. We also have computed the average of pairwise distances between endmembers. The values are depicted in table 5.4 and Figure 5.23 respectively. We

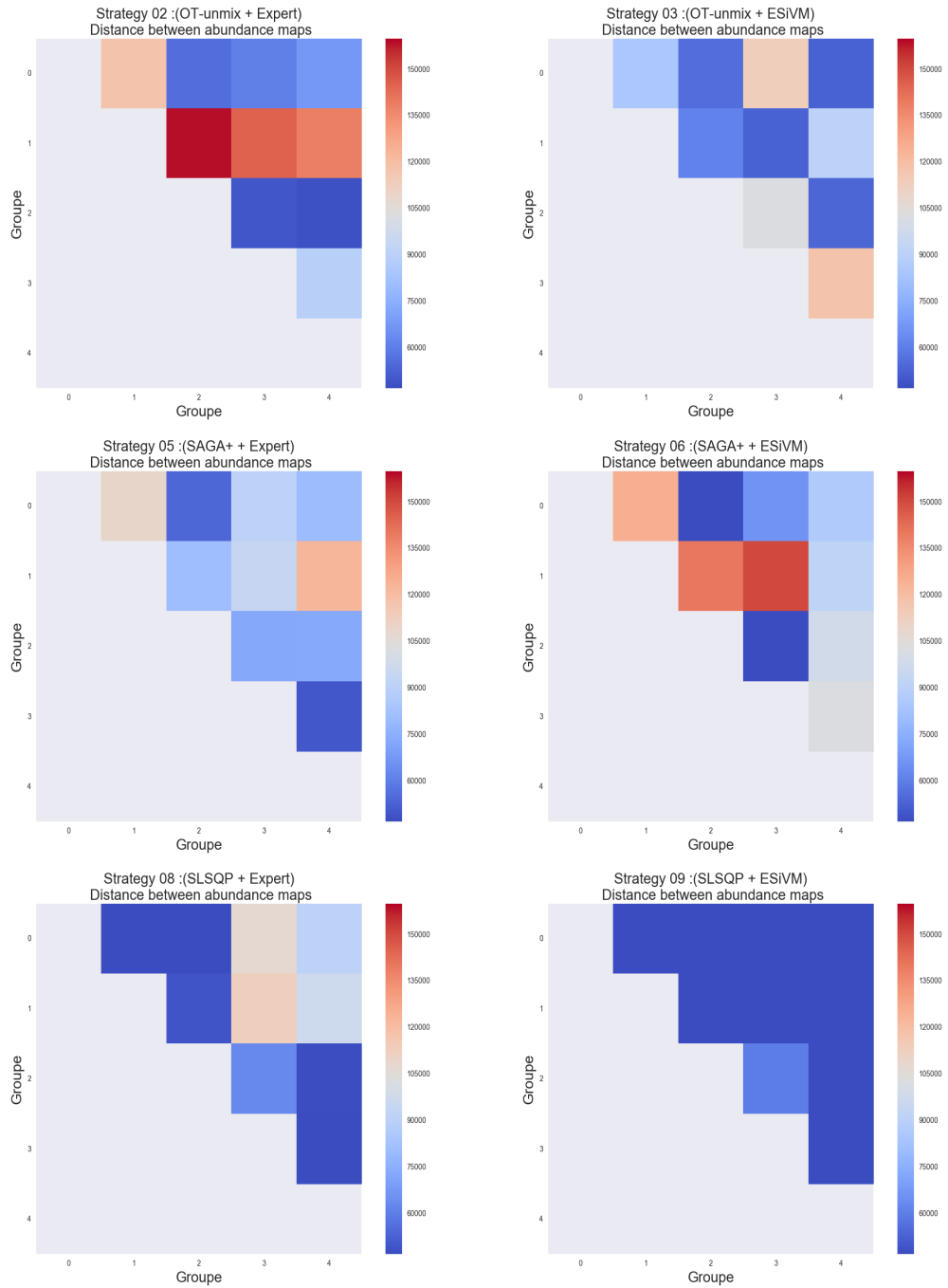


Figure 5.23: Map distances calculated between pairs of abundance maps for 6 different strategies.

can also conclude that the various materials of the scene appear to be more consistent here. Moreover using SAGA+, performances in terms of computational cost are lower since the same calculation took place in about 15 minutes. Ongoing works are focusing on ground truth validations to perform more complete quantitative comparisons.

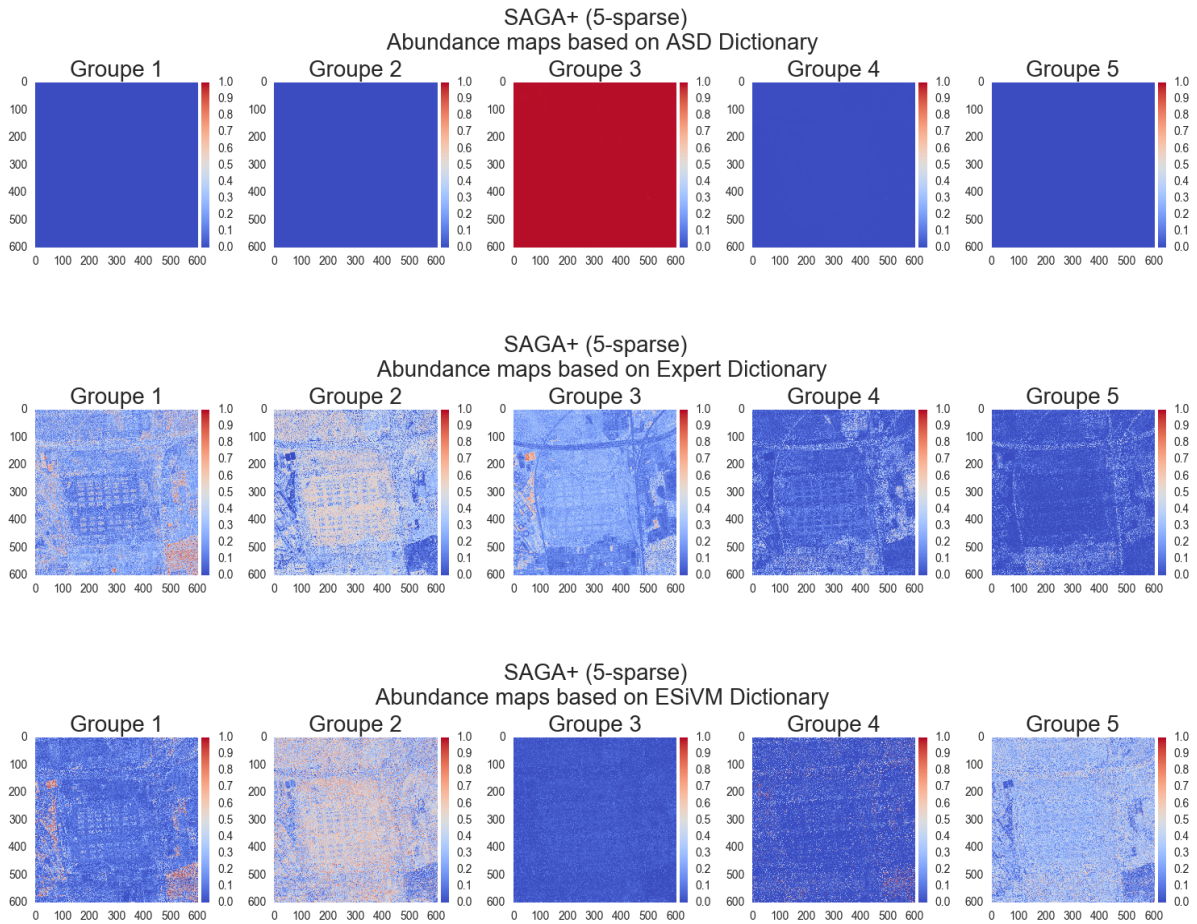


Figure 5.24: Abundance maps produced by SAGA+ algorithm and according to the strategies 4,5 and 6 presented in the middle column of table 5.2.

5.4 Conclusion of the chapter

In this last part of this PhD thesis we provided some preliminary results on a specific challenges of spectral unmixing for surface temperature estimation in an urban context. Preliminary results were presented and some comparisons were with a state-of-the-art approach seems to promote the use of the two algorithms developed through this work. Indeed stronger validations have now to be done. This is the scope of a current campaign with ground truth data.

Conclusion and future work

Spectral unmixing has gained a considerable amount of attention during the recent years within the remote sensing community. As the aftermath of continuously improving spectral resolution of the hyperspectral sensors, more accurate measurements of materials have become feasible.

The major issue in hyperspectral unmixing is the way existing endmembers have mixed at the onset of pixel registration. The formation process of endmembers within a pixel is formally called the *mixing model*. Depending on the context, different mixing models are being assumed while studying remotely sensed data. Mixing models can be split in three major categories, namely *linear*, *non-linear* and *intimate* mixing models. Having a correct assumption of the mixing model, while analyzing hyperspectral data has a great impact on the accuracy of the analysis. Another important issue related to unmixing is the issue of *endmember variability* where spectral signature of the same material shows subtle differences from one sample to the other. Yet another common issue within the field of spectral analysis is the presence of *anomalies* which can either be thought of as unwanted artifacts that spoil the true value of the pixels or on the opposite, to be seen as highly important but spatially scarce information hidden within the scene. In this thesis, the three aforementioned issues were mainly brought into light and by introducing two original algorithms, defined within different mathematical frameworks, solutions to these open problems have been provided.

- The first contribution of this PhD thesis, brings all the three mentioned problems of *mixture modeling*, *endmember variability* and *anomaly detection* into focus and using the applications of kernel theory proposes a new unsupervised algorithm (SAGA+) for representation of the non-linear manifold embedding the data while through a simultaneous anomaly detection procedure makes sure that the representation of the *manifold hull* is not being distorted at the presence of anomalies. In fact SAGA+ can be seen as a kernel-based non-negative matrix factorization (NMF) which is coupled with an anomaly detection. It operates by coding data w.r.t the local neighbors with non-linear weights. For this the solution of NMF is divided in two steps of estimations of ℓ data points characterizing the boundaries of the manifold hull \mathbf{E} embedding

data and then using this as the endmember dictionary to estimate the coefficients of each endmember contributing to the reconstruction of each pixel value of the original dataset. The locality is the aftermath of the simultaneous imposition of sparsity and convexity constraints. The set of abundance maps produced by this procedure therefore, is the convex surrogate of the original data which is achievable through a controllable level of sparsity. The matrix of abundances \mathbf{G} , encodes the convex combinations and the ℓ columns of the endmember dictionary \mathbf{E} characterizes a *hull* of dataset \mathbf{X} in \mathbb{R}^D , as data points lying within are perfectly reconstructed. The Anomaly detection procedure is formulated in order to play the role of a supervisor through the first part of SAGA+. The idea consists of iteratively increasing the volume of the manifold hull in feature space while at the same time keeping track of the rate of change of the sum of projection errors (SOPE) in order to exclude anomalies from being taken as simplex vertexes, which is being formed based on the previously selected endmembers. In the second step of solving NMF, in order to account for the *endmember variability* among ℓ endmembers (as ℓ can be chosen large even to outnumber the dimensionality of the data), by manually imposing a sparsity value λ , each reconstructed point within \mathbf{G} will be a convex combination of λ -selected atoms within the ℓ -endmember dictionary.

This is what makes this algorithm unique in addressing all the three mentioned problems within a unified framework.

- The second major contribution of this PhD focuses mainly on the issue of *endmember variability* and by exploiting the notion of *overcomplete dictionary* tries to address this problem. The main assumption is that a dictionary of endmembers, possibly larger than the effective number of pure materials is available to deal with the internal heterogeneity inside a particular material (which should correspond to only one endmember in classical approaches). This supervised algorithm (OT-unmixing) which is based on the optimal transport theory is comparable to the second step of SAGA+, as it solves an inversion problem and calculates the sparse representation of the original datum/pixels through generation of the abundance maps. For this OT-Unmixing tries to minimize a transition plan between the datum (represented as a probability distribution in the spectral domain) and the known atoms of the dictionary (also in the form of probability distributions). An original metric, designed between probability density functions is considered to compare spectra. This metric (Wasserstein distance), is also based on optimal transport and can be designed to be mostly sensitive to shifts in the frequency domains. Although this problem is solvable through linear programming with equality constraints, considering that the computational cost scales quadratically with the dimension of the spectra, a regularization term is used to help increasing the speed of the calculation. This regularization term which controls the smoothness of the transition map is based on the entropy of the map. Having the discrete version of the OT with an Entropy regularization term, it is possible to establish

a link between OT and Bregman projection which by itself admits a simple closed-form solution. Therefore for each datum/pixel signature, OT-Unmixing uses an iterative Bregman projection to project the signature onto the atoms of the dictionary and calculating the Wasserstein distance in each iteration. It worths mentioning that for producing sparse abundance maps the *a priori* information regarding the λ groups representing endmember variabilities, also the respective weights are being provided through a separate matrix and a weight vector.

The mentioned unmixing algorithms have been used on synthesized and real datasets in different settings in order to be evaluated and compared with existing state-of-the-art procedures. In particular SAGA+ has shown its robustness to highly corrupted by anomalies planetary dataset, letting the expert to identify small patches of otherwise difficult to detect lithologies on the surface of 4-Vesta asteroid. Besides, to be the third major contribution of this thesis, both algorithms were deployed in the context of CATUT project where, SAGA+ and OT-unmixing were used in providing abundance maps of an urban scene. The utilization and comparing different vector-based spectral distance measures was among the subtasks of this thesis. In particular to compare the *endmember estimation* performance of SAGA+ with other endmember extraction techniques. In a current work a more complete evaluation and comparison of distance measures is considered by including other correlation weighted measures along the well-known vector-based measures. Also from the computational time point of view the algorithms are at least as good as their state-of-the-art counterparts. In the case of SAGA+ having an overall linear computational complexity, the algorithm is faster than many non-linear spectral unmixing procedures. This is an asset while processing large volumes of data like planetary datasets. In the case of OT-Unmixing depending on the maximum iterations that it takes for the minimization to converge, the results can be up to twice more expensive.

As for the future possible continuation of the current work one can mention the potential capacity of the current models to be used on multi-angular and multi-temporal hyperspectral datasets. The new multi-angular sensors have been contributing to many astronomical and astrophysical projects (CRISM, OMEGA etc.) providing additional (angular) measurement dimension, therefore a better separation of the signals coming from the atmosphere and the surface become possible. This allows for more accurate characterization of the planetary materials. These hyperspectral multi-angular (HMA) observations, available in large quantities, offer new perspectives but require extensive methodological research and development because of their very nature and the complexity of natural scenes.

SAGA+ seems quite a proper method to be extended for the exploitation of this angular dimension. It can simply expanded by dedicating several kernels to simultaneously unmix the pixels contributing to the same spatial position but representing different angular/temporal perspectives. This will requires us also to devise a fusion mechanism after the value of the current pixel being projected to the feature space

through L kernels (where L represents the number of angular datasets). By using a segmentation strategy (Metropolis-within-Gibbs sampler for example) we can partition the data space into convex regions where separate units represented by these regions are being modeled with not all the dictionary atoms but a limited (hopefully not more than a handful) number of estimated endmembers. Also by considering the possibility of having different mixing paradigms underlying different areas of a scene, it is possible to investigate to what degree segmentation of the scene into separated units and unmixing each unit using a dedicated kernel can help increasing the overall accuracy of the unmixing procedure. In such scenarios one can think of dedicating the result of separate endmembers estimation per segmentation. Besides that by availability of datasets covering vast homogeneous areas of land surface, it will be effective to exploit the potential of domain adaptation techniques for generalizing the unmixing result of a well-studied, but limited-in-spatial-size area to a large scene while improving the computation time. Also the natures of the presented algorithms, are such that make the parallelization of the procedures feasible, thus suitable for exploiting the computation power of parallel-processing units and GPGPUs. As for the OT-unmixing it is possible to think about exploitation of other regularization techniques instead of the entropy-based regularizer presented in this work. Among other regularization methods one can mention LASSO.

Part V

Appendices

7.1 Spectral Signatures

A total number of 41 spectral signatures used in order to generate synthetic datasets for this paper is represented in Figure 7.1. The signatures are related mostly to the

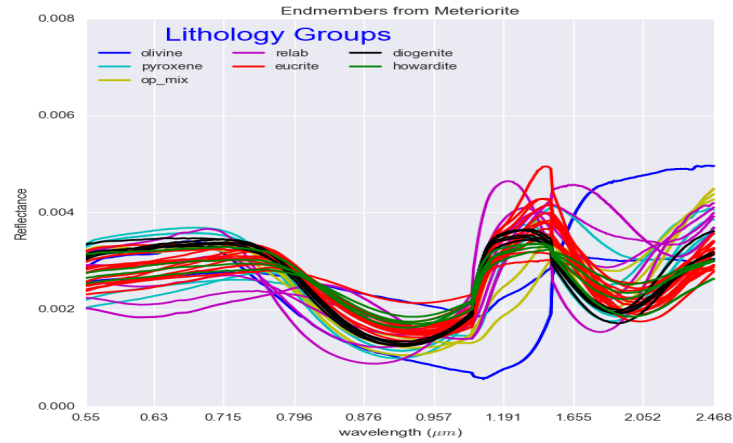


Figure 7.1: Example of spectra used for generating the synthetic datasets issued from the 41 spectral signatures of different Lithologies RELAB. The signatures have re-sampled in order to be comparable to the signature characteristics of 4-Vesta dataset according to the wavelength range and sampling range.

lithologies which are found on the surface of meteorites from 4-Vesta asteroid which have hit the earth surface. Based on their compositions, the signatures were roughly categorized into 7 groups, each representing specific families of magmatic rocks. The original signatures are registered based on 461 contiguous electromagnetic bands covering the range of $0.3 - 2.6\mu m$. The signatures went through a re-sampling process in order to be approximately in comparison to the wavelength range and sampling rate of the 4-Vesta dataset. Therefore after re sampling each signature has 383 bands covering the range $0.55 - 2.46\mu m$.

7.2 SOPE maps

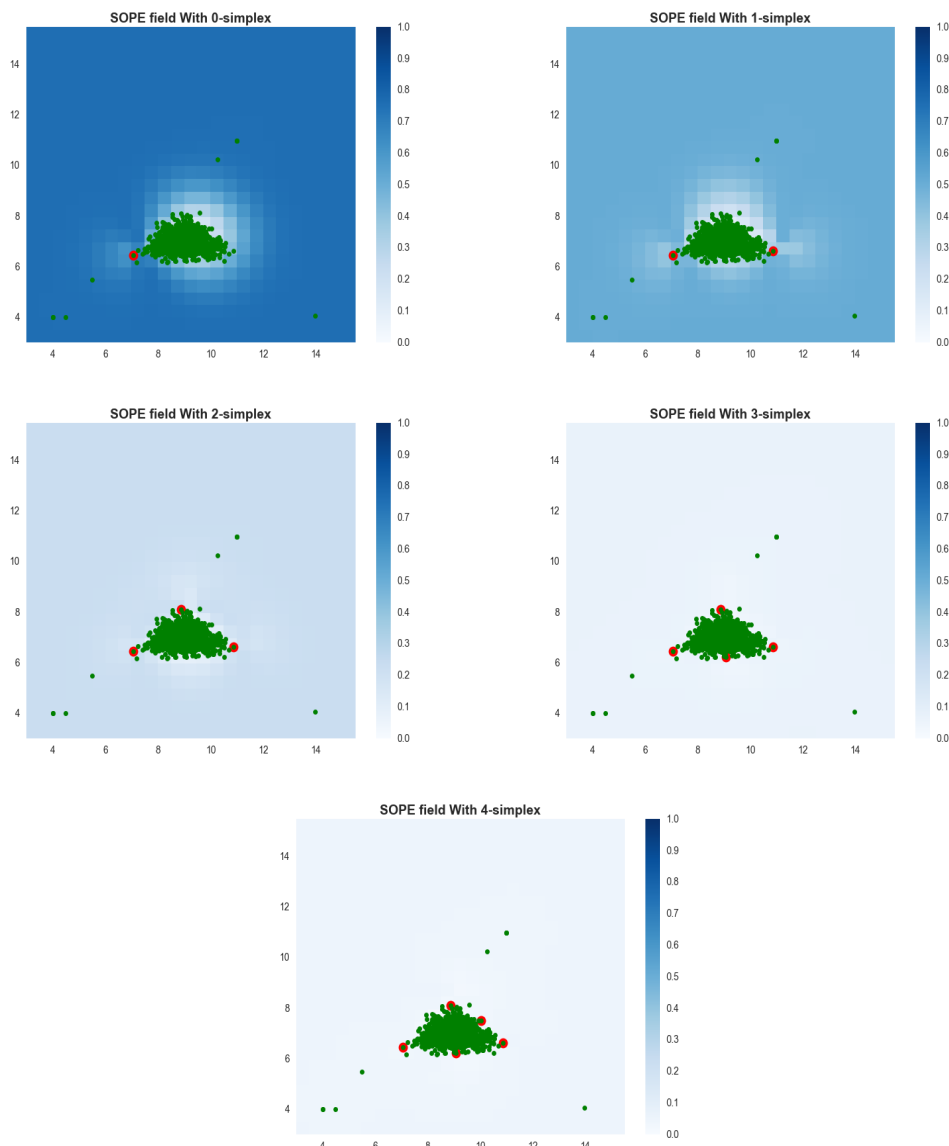


Figure 7.2: Evolution of SOPE map along the endmember extraction process. Each plot shows the map of SOPE values w.r.t the "next" endmember candidate. In each plot, we take all the cells of 26×26 grid one by one as the next endmember candidate and calculate SOPE. In each plot the only difference is the dimensionality of the simplex that we use to calculate the SOPE values. It is observable that the points laying outside the boundaries of the dataset always produce very high values of SOPE (almost close to 1). By setting a correct threshold for SOPE change then SAGA+ distinguishes the good candidates (points on the boundaries of the dataset manifold) from the anomalies (points outside the boundaries of the dataset manifold)

In each plot of Figure 7.2, all points of a meshgrid of size (26×26) are taken as the next prototype and the SOPE value is calculated. The only difference is the number of Endmembers which have already selected to construct the representative manifold. As one can see by selecting endmember candidates from the borders of the nominal data, in each iteration the overall SOPE value drops, until the time when adding an extra simplex does not improve the SOPE significantly. In this particular case we can see that by selecting the fifth endmember the overall SOPE does not change significantly which means that the 3-simplex is almost as good as the 4-simplex in representing this dataset.

7.3 Parameter Setting

Table 7.1: Parameter setting for synthetic data generation

	Parameter Values
Sample Size (N)	$10^3 \dots 10^6$
Endmember (ℓ)	3 ... 15
Sparsity (λ)	3-5
Kernel bandwidth (σ)	0.15
Threshold (τ)	0.04
Concentration LMM (α)	1
Concentration IMM (α)	50
Anomalies (γ)	20
Dimension (D)	383

In the table 7.1 the notations represented different model parameters are gathered together with their range of values which they used for synthetic tests.

7.4 Execution Time

In table 7.2 the execution times of the tested algorithms (SAGA+, G-SiVM, N-finder and VCA) using different endmembers ℓ and instance numbers N have been compared. From this comparison we observed that VCA outperformed any other algorithm while (G-SiVM) has a systematic larger computation time which makes it unusable for real applications. This large computation time is explained by its inherent complexity of the computation of Geodesic distances which scales quadratically with the size of data. As for SAGA+ and N-finder, only minor differences can

be observed.

Table 7.2: Execution Time (s) of the different algorithms for various endmembers and elements in the dataset

		# Samples (N)			
SAGA+		10^3	10^4	10^5	10^6
# (ℓ)	$\ell = 3$.022	.255	3.058	31.432
	$\ell = 5$	0.039	0.586	4.312	42.316
	$\ell = 7$	0.044	0.462	20.716	54.562
	$\ell = 9$	0.065	0.687	10.872	53.666
	$\ell = 11$	0.096	0.727	12.276	75.676
	$\ell = 13$	0.085	0.833	12.012	87.943
	$\ell = 15$	0.097	0.960	9.871	98.948
		# Samples (N)			
G-SiVM		10^3	10^4	10^5	10^6
# (ℓ)	$\ell = 3$	2.382	235.815	$\approx 3 \times 10^3$	—
	$\ell = 5$	2.405	"	"	—
	$\ell = 7$	2.507	"	"	—
	$\ell = 9$	2.674	"	"	—
	$\ell = 11$	2.958	"	"	—
	$\ell = 13$	3.174	"	"	—
	$\ell = 15$	4.480	"	"	—
		# Samples (N)			
N-finder		10^3	10^4	10^5	10^6
# (ℓ)	$\ell = 3$	0.072	0.528	4.641	47.315
	$\ell = 5$	0.103	0.832	7.796	86.957
	$\ell = 7$	0.135	1.135	10.651	369.269
	$\ell = 9$	0.160	1.417	13.831	147.192
	$\ell = 11$	0.248	2.091	20.158	225.753
	$\ell = 13$	0.320	2.501	24.849	276.560
	$\ell = 15$	0.364	3.010	29.752	496.775
		# Samples (N)			
VCA		10^3	10^4	10^5	10^6
# (ℓ)	$\ell = 3$	0.012	0.028	0.306	2.414
	$\ell = 5$	0.015	0.027	0.257	2.501
	$\ell = 7$	0.009	0.028	0.253	3.015
	$\ell = 9$	0.010	0.032	0.260	2.819
	$\ell = 11$	0.010	0.032	0.302	3.184
	$\ell = 13$	0.010	0.036	0.281	2.820
	$\ell = 15$	0.013	0.030	0.292	3.684

As SAGA+ is a non-linear procedure able to capture more subtleties in the variations of signatures, this low computational time makes it a very appealing solution.

7.5 Abundance Maps for SAGA

Abundance maps of 4-Vesta based on 10 estimated endmembers are presented in Figure 7.3. The selected signatures by SAGA shows that without an anomaly detection procedure this method can select corrupted signatures instead of real signatures. The produced maps therefore can be effected by the poor result of endmembers estimation.

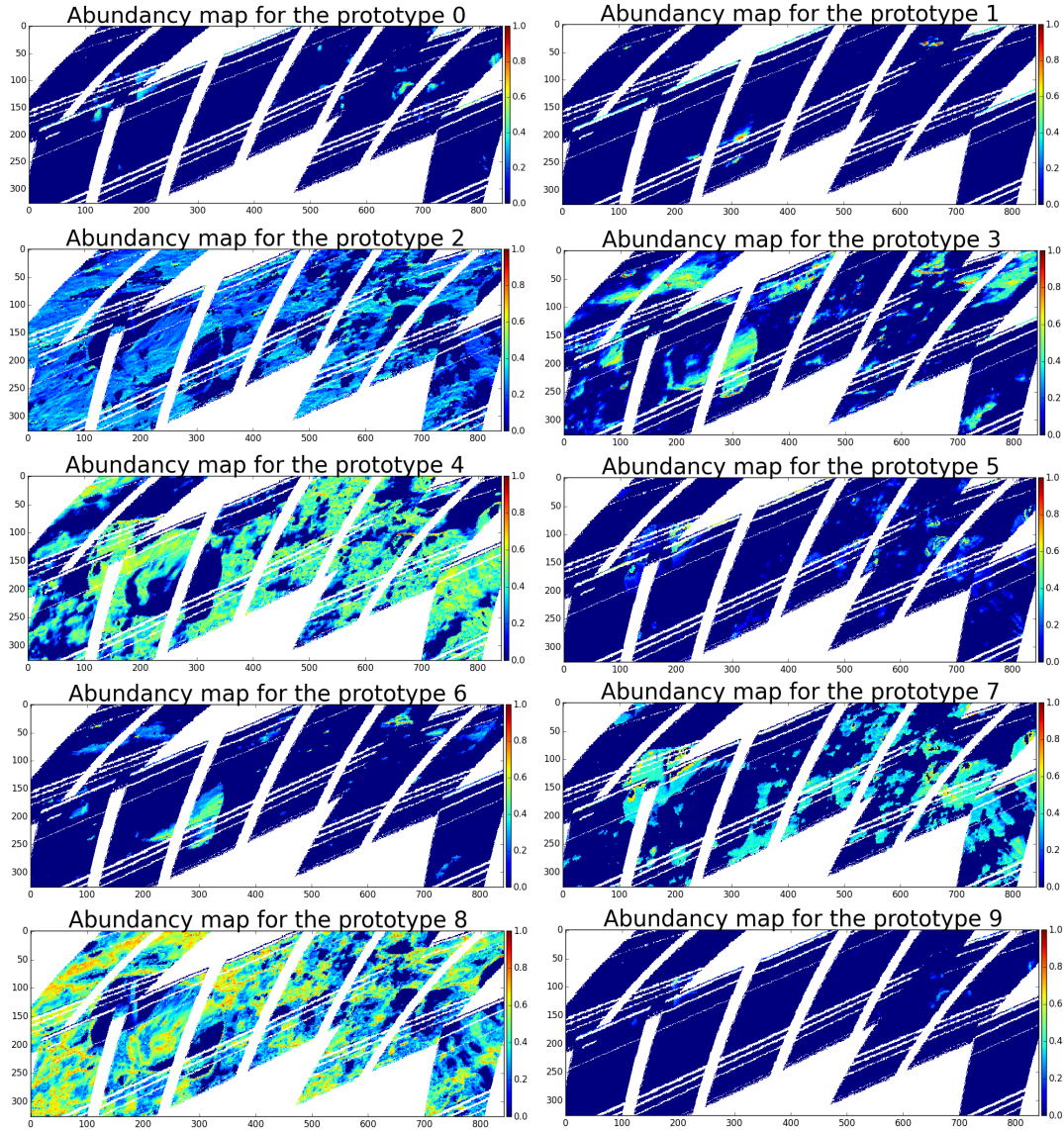


Figure 7.3: Abundance maps and respective spectral signatures extracted by SAGA (no anomaly detection). Without an anomaly detection procedure, one observes that signatures with sharp peaks are among selected endmember signatures.

Bibliography

- [1] Guolan Lu and Baowei Fei. Medical hyperspectral imaging: a review. *Journal of biomedical optics*, 19(1):010901–010901, 2014.
- [2] Nicolas Dobigeon, Jean-Yves Tournet, Cédric Richard, Jose Bermudez, Steve McLaughlin, and Alfred O Hero. Nonlinear unmixing of hyperspectral images: Models and algorithms. *Signal Processing Magazine, IEEE*, 31(1):82–94, 2014.
- [3] J.M. Bioucas-Dias and A. Plaza. An overview on hyperspectral unmixing: Geometrical, statistical, and sparse regression based approaches. In *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International*, pages 1135–1138, July 2011.
- [4] Jie Chen, C. Richard, and P. Honeine. Nonlinear unmixing of hyperspectral data based on a linear-mixture/nonlinear-fluctuation model. *Signal Processing, IEEE Transactions on*, 61(2):480–492, Jan 2013.
- [5] J.M.P. Nascimento and J.M. Bioucas Dias. Vertex component analysis: a fast algorithm to unmix hyperspectral data. *Geoscience and Remote Sensing, IEEE Transactions on*, 43(4):898–910, April 2005.
- [6] R. Heylen, D. Burazerovic, and P. Scheunders. Non-linear spectral unmixing by geodesic simplex volume maximization. *Selected Topics in Signal Processing, IEEE Journal of*, 5(3):534–542, June 2011.
- [7] X. Yang, K. Zhang, B. Jia, and L. Ci. Desertification assessment in china: An overview. *Journal of Arid Environments*, 63(2):517 – 531, 2005.
- [8] J. Farifteh, A. Farshad, and R.J. George. Assessing salt-affected soils using remote sensing, solute modelling, and geophysics. *Geoderma*, 130(3):191 – 206, 2006.
- [9] Amir AghaKouchak, Hamid Norouzi, Kaveh Madani, Ali Mirchi, Marzi Azarderakhsh, Ali Nazemi, Nasrin Nasrollahi, Alireza Farahmand, Ali Mehran, and Elmira Hasanzadeh. Aral sea syndrome desiccates lake urmia: Call for action. *Journal of Great Lakes Research*, 41(1):307 – 311, 2015.
- [10] Camilla Brekke and Anne HS Solberg. Oil spill detection by satellite remote sensing. *Remote sensing of environment*, 95(1):1–13, 2005.
- [11] Sabine Thiemann and Hermann Kaufmann. Lake water quality monitoring using hyperspectral airborne data—a semiempirical multisensor and multitemporal approach for the mecklenburg lake district, germany. *Remote Sensing of Environment*, 81(2):228–237, 2002.

- [12] Qihao Weng, Dengsheng Lu, and Jacquelyn Schubring. Estimation of land surface temperature–vegetation abundance relationship for urban heat island studies. *Remote sensing of Environment*, 89(4):467–483, 2004.
- [13] Hamed Akbari, Kuniaki Uto, Yukio Kosugi, Kazuyuki Kojima, and Naofumi Tanaka. Cancer detection using infrared hyperspectral imaging. *Cancer Science*, 102(4):852–857, 2011.
- [14] Rubén Salvador, Samuel Ortega, Daniel Madroñal, Himar Fabelo, Raquel Lazcano, Gustavo Marrero, Eduardo Juárez, Roberto Sarmiento, and César Sanz. Helicoid: Interdisciplinary and collaborative project for real-time brain cancer detection: Invited paper. In *Proceedings of the Computing Frontiers Conference*, CF’17, pages 313–318, 2017.
- [15] S Kabwama, D Bulters, Harry Bulstrode, Himar Fabelo, S Ortega, Gustavo Marrero Callico, Bogdan Stanciulescu, Ravi Kiran, Daniele Ravì, A Szolna, and J.F. Piñeiro. Intra-operative hyperspectral imaging for brain tumour detection and delineation: Current progress on the helicoid project. 36:S140, 12 2016.
- [16] Daniele Ravì, Himar Fabelo, Gustavo Marrero Callico, and GuangZhong Yang. 04 2017.
- [17] Maria Kubik. Chapter 5 hyperspectral imaging: A new technique for the non-invasive study of artworks. 2:199–259, 12 2007.
- [18] Aurélie Mounier, Gwénaëlle Le Bourdon, Christian Aupetit, Colette Belin, Laurent Servant, Sylvain Lazare, Yannick Lefrais, and Floréal Daniel. Hyperspectral imaging, spectrofluorimetry, fors and xrf for the non-invasive study of medieval miniatures materials. *Heritage Science*, 2(1):24, Oct 2014.
- [19] Hyperspectral imaging applied to the analysis of goya paintings in the museum of zaragoza (spain). *Microchemical Journal*, 126(Supplement C):113 – 120, 2016.
- [20] Kathryn Dooley, Suzanne Lomax, Jason G Zeibel, Costanza Miliani, Paola Ricciardi, Ann Hoenigswald, Murray Loew, and John Delaney. Mapping of egg yolk and animal skin glue paint binders in early renaissance paintings using near infrared reflectance imaging spectroscopy. 138, 06 2013.
- [21] Sarah H Parcak. *Satellite remote sensing for archaeology*. Routledge, 2009.
- [22] Rosa Maria Cavalli, Francesca Colosi, Angelo Palombo, Stefano Pignatti, and Maurizio Poscolieri. Remote hyperspectral imagery as a support to archaeological prospection. *Journal of Cultural Heritage*, 8(3):272–283, 2007.
- [23] Jason Ur. Corona satellite photography and ancient road networks: a northern mesopotamian case study. *Antiquity*, 77(295):102–115, 2003.

- [24] Prakash Chauhan, Prabhjot Kaur, N Srivastava, Rishitosh K Sinha, Nirmala Jain, and SVS Murty. Hyperspectral remote sensing of planetary surfaces: an insight into composition of inner planets and small bodies in the solar system. *Current Science*, 108(5):915–924, 2015.
- [25] J-Ph Combe, S Le Mouelic, C Sotin, A Gendrin, JF Mustard, L Le Deit, P Launeau, J-P Bibring, B Gondet, Y Langevin, et al. Analysis of omega/mars express data hyperspectral data using a multiple-endmember linear spectral unmixing model (melsum): Methodology and first results. *Planetary and Space Science*, 56(7):951–975, 2008.
- [26] Stéphane Le Mouelic, J-Ph Combe, V Sarago, N Mangold, M Masse, JP Bibring, B Gondet, Y Langevin, and Christophe Sotin. An iterative least squares approach to decorrelate minerals and ices contributions in hyperspectral images: Application to cuprite (earth) and mars. In *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, 2009. WHISPERS'09. First Workshop on*, pages 1–4. IEEE, 2009.
- [27] Saïd Moussaoui, Hafrun Hauksdottir, Frédéric Schmidt, Christian Jutten, Jocelyn Chanussot, David Brie, Sylvain Douté, and Jon Atli Benediktsson. On the decomposition of mars hyperspectral data by ica and bayesian positive source separation. *Neurocomputing*, 71(10):2194–2208, 2008.
- [28] WH Bakker, FJA van Ruitenbeek, HMA van der Werff, TE Zegers, JHP Oosthoek, SH Marsh, and FD van der Meer. Processing omega/mars express hyperspectral imagery from radiance-at-sensor to surface reflectance. *Planetary and Space Science*, 90:1–9, 2014.
- [29] Sébastien Bourguignon, Hervé Carfantan, Eric Slezak, and David Mary. Sparsity-based spatial-spectral restoration of muse astrophysical hyperspectral data cubes. In *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2011 3rd Workshop on*, pages 1–4. IEEE, 2011.
- [30] Yu-Shiuan Shen, Tsung-Han Chan, Sébastien Bourguignon, and Chong-Yung Chi. Spatial-spectral unmixing of hyperspectral data for detection and analysis of astrophysical sources with the muse instrument. In *IEEE WHISPERS*, 2012.
- [31] C Diener, L Wisotzki, KB Schmidt, Edmund C Herenz, T Urrutia, T Garel, J Kerutt, RL Saust, R Bacon, S Cantalupo, et al. The muse-wide survey: detection of a clustering signal from lyman α emitters in the range $3 < z < 6$. *Monthly Notices of the Royal Astronomical Society*, 471(3):3186–3192, 2017.
- [32] Céline Meillier, Florent Chatelain, Olivier Michel, Roland Bacon, Laure Piqueras, Raphael Bacher, and Hacheme Ayasso. Selfi: an object-based, bayesian method for faint emission line source detection in muse deep field data cubes. *Astronomy & Astrophysics*, 588:A140, 2016.

- [33] Ivica Kopriva and Andrzej Cichocki. Blind multispectral image decomposition by 3d nonnegative tensor factorization. *Optics letters*, 34(14):2210–2212, 2009.
- [34] Moussa Sofiane Karoui, Yannick Deville, Shahram Hosseini, and Abdelaziz Ouamri. Blind spatial unmixing of multispectral images: New methods combining sparse component analysis, clustering and non-negativity constraints. *Pattern Recognition*, 45(12):4263–4278, 2012.
- [35] Yue Gao and Tat-Seng Chua. *Hyperspectral Image Classification by Using Pixel Spatial Correlation*, pages 141–151. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.
- [36] Chein-I Chang. *Hyperspectral data processing: algorithm design and analysis*. John Wiley & Sons, 2013.
- [37] J.M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 5(2):354–379, 2012.
- [38] Chris HQ Ding, Tao Li, and Michael I Jordan. Convex and semi-nonnegative matrix factorizations. *IEEE transactions on pattern analysis and machine intelligence*, 32(1):45–55, 2010.
- [39] Nicolas Courty, Xing Gong, Jimmy Vandel, and Thomas Burger. Saga: sparse and geometry-aware non-negative matrix factorization through non-linear local embedding. *Machine Learning*, 97(1-2):205–226, 2014.
- [40] Aapo Hyvärinen, Juha Karhunen, and Erkki Oja. *Independent component analysis*, volume 46. John Wiley & Sons, 2004.
- [41] M Lennon, G Mercier, MC Mouchot, and L Hubert-Moy. Spectral unmixing of hyperspectral images with the independent component analysis and wavelet packets. In *Geoscience and Remote Sensing Symposium, 2001. IGARSS’01. IEEE 2001 International*, volume 6, pages 2896–2898. IEEE, 2001.
- [42] Bruce Hapke. *Theory of reflectance and emittance spectroscopy*. Cambridge University Press, 2012.
- [43] José MP Nascimento and José M Bioucas-Dias. Nonlinear mixture model for hyperspectral unmixing. In *SPIE Europe Remote Sensing*, pages 74770I–74770I. International Society for Optics and Photonics, 2009.
- [44] Abderrahim Halimi, Yoann Altmann, Nicolas Dobigeon, and Jean-Yves Tourneret. Nonlinear unmixing of hyperspectral images using a generalized bilinear model. *Geoscience and Remote Sensing, IEEE Transactions on*, 49(11):4153–4162, 2011.

- [45] Ines Meganem, Ph Deliot, Xavier Briottet, Yannick Deville, and Sepehr Hosseini. Linear-quadratic mixing model for reflectances in urban environments. *Geoscience and Remote Sensing, IEEE Transactions on*, 52(1):544–558, 2014.
- [46] Fatima Zohra Benhalouche, Moussa Sofiane Karoui, Yannick Deville, and Abdelaziz Ouamri. Hyperspectral and multispectral data fusion based on linear-quadratic nonnegative matrix factorization. *Journal of Applied Remote Sensing*, 11(2):025008–025008, 2017.
- [47] Wenyi Fan, Baoxin Hu, John Miller, and Mingze Li. Comparative study between a new nonlinear model and common linear model for analysing laboratory simulated-forest hyperspectral data. *International Journal of Remote Sensing*, 30(11):2951–2962, 2009.
- [48] Gustavo Camps-Valls and Lorenzo Bruzzone. Kernel-based methods for hyperspectral image classification. *Geoscience and Remote Sensing, IEEE Transactions on*, 43(6):1351–1362, 2005.
- [49] Javier Plaza, Antonio J Plaza, Pablo Martinez, and Rosa M Perez. Nonlinear mixture models for analyzing laboratory simulated-forest hyperspectral data. In *Remote Sensing*, pages 480–487. International Society for Optics and Photonics, 2004.
- [50] Chein-I Chang and A. Plaza. A fast iterative algorithm for implementation of pixel purity index. *Geoscience and Remote Sensing Letters, IEEE*, 3(1):63–67, Jan 2006.
- [51] Michael E Winter. N-findr: An algorithm for fast autonomous spectral end-member determination in hyperspectral data. In *SPIE’s International Symposium on Optical Science, Engineering, and Instrumentation*, pages 266–275. International Society for Optics and Photonics, 1999.
- [52] M.D. Craig. Minimum-volume transforms for remotely sensed data. *Geoscience and Remote Sensing, IEEE Transactions on*, 32(3):542–552, May 1994.
- [53] Milkos Hollósi, Gerald D Fasman, et al. Convex constraint analysis: a natural deconvolution of circular dichroism curves of proteins. *Protein Engineering*, 4(6):669–679, 1991.
- [54] Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987.
- [55] Andrew A Green, Mark Berman, Paul Switzer, and Maurice D Craig. A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Transactions on geoscience and remote sensing*, 26(1):65–74, 1988.

- [56] S. Jia and Y. Qian. Constrained nonnegative matrix factorization for hyperspectral unmixing. *Geoscience and Remote Sensing, IEEE Transactions on*, 47(1):161–173, 2009.
- [57] Lidan Miao and Hairong Qi. Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization. *Geoscience and Remote Sensing, IEEE Transactions on*, 45(3):765–777, 2007.
- [58] P. Sajda, S. Du, and L.C. Parra. Recovery of constituent spectra using non-negative matrix factorization. In *Optical Science and Technology, SPIE’s 48th Annual Meeting*, pages 321–331. International Society for Optics and Photonics, 2003.
- [59] Naoto Yokoya, Jocelyn Chanussot, and Akira Iwasaki. Generalized bilinear model based nonlinear unmixing using semi-nonnegative matrix factorization. In *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*, pages 1365–1368. IEEE, 2012.
- [60] Rob Heylen and Paul Scheunders. Calculation of geodesic distances in nonlinear mixing models: Application to the generalized bilinear model. *Geoscience and Remote Sensing Letters, IEEE*, 9(4):644–648, 2012.
- [61] Joshua Broadwater, Amit Banerjee, and Philippe Burlina. Kernel methods for unmixing hyperspectral imagery. *Kernel Methods for Remote Sensing Data Analysis*, pages 249–270, 2009.
- [62] J. Broadwater and A. Banerjee. A comparison of kernel functions for intimate mixture models. In *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, 2009. WHISPERS ’09. First Workshop on*, pages 1–4, Aug 2009.
- [63] J. Broadwater, R. Chellappa, A. Banerjee, and P. Burlina. Kernel fully constrained least squares abundance estimates. In *Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International*, pages 4041–4044, July 2007.
- [64] Lidan Miao and Hairong Qi. Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization. *IEEE Transactions on Geoscience and Remote Sensing*, 45(3):765–777, 2007.
- [65] José M Bioucas-Dias. A variable splitting augmented lagrangian approach to linear spectral unmixing. In *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, 2009. WHISPERS’09. First Workshop on*, pages 1–4. IEEE, 2009.
- [66] Jun Li and José M Bioucas-Dias. Minimum volume simplex analysis: A fast algorithm to unmix hyperspectral data. In *Geoscience and Remote Sensing*

- Symposium, 2008. IGARSS 2008. IEEE International*, volume 3, pages III–250. IEEE, 2008.
- [67] Alina Zare and KC Ho. Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing. *IEEE Signal Processing Magazine*, 31(1):95–104, 2014.
 - [68] Charlotte Revel, Yannick Deville, Véronique Achard, and Xavier Briottet. Inertia-constrained pixel-by-pixel nonnegative matrix factorisation: a hyperspectral unmixing method dealing with intra-class variability. *CoRR*, abs/1702.07630, 2017.
 - [69] J. Bieniarz, E. Aguilera, X.X. Zhu, R. Muller, and P. Reinartz. Joint sparsity model for multilook hyperspectral image unmixing. *Geoscience and Remote Sensing Letters, IEEE*, 12(4):696–700, April 2015.
 - [70] J. Bieniarz, R. Muller, Xiaoxiang Zhu, and P. Reinartz. On the use of over-complete dictionaries for spectral unmixing. In *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2012 4th Workshop on*, pages 1–4, June 2012.
 - [71] Yagyensh Chandra Pati, Ramin Rezaiifar, and Perinkulam Sambamurthy Krishnaprasad. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*, pages 40–44. IEEE, 1993.
 - [72] José M Bioucas-Dias and Mário AT Figueiredo. Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing. In *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2010 2nd Workshop on*, pages 1–4. IEEE, 2010.
 - [73] Joseph C Harsanyi and C-I Chang. Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach. *IEEE Transactions on geoscience and remote sensing*, 32(4):779–785, 1994.
 - [74] Yingzi Du, Chein-I Chang, Hsuan Ren, Chein-Chi Chang, James O Jensen, and Francis M D’Amico. New hyperspectral discrimination measure for spectral characterization. *Optical Engineering*, 43(8):1777–1786, 2004.
 - [75] D Heinz, C-I Chang, and Mark LG Althouse. Fully constrained least-squares based linear unmixing [hyperspectral image classification]. In *Geoscience and Remote Sensing Symposium, 1999. IGARSS’99 Proceedings. IEEE 1999 International*, volume 2, pages 1401–1403. IEEE, 1999.
 - [76] Joshua Broadwater, Rama Chellappa, Amit Banerjee, and Philippe Burlina. Kernel fully constrained least squares abundance estimates. In *Geoscience and*

- Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International*, pages 4041–4044. IEEE, 2007.
- [77] Xiaoqiang Lu, Hao Wu, Yuan Yuan, Pingkun Yan, and Xuelong Li. Manifold regularized sparse nmf for hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 51(5):2815–2826, 2013.
 - [78] Yuntao Qian, Sen Jia, Jun Zhou, and Antonio Robles-Kelly. Hyperspectral unmixing via $l_{\frac{1}{2}}$ sparsity-constrained nonnegative matrix factorization. *IEEE Transactions on Geoscience and Remote Sensing*, 49(11):4282–4297, 2011.
 - [79] Jun Li and José M. Bioucas-dias. Minimum volume simplex analysis: A fast algorithm to unmix hyperspectral data, 2008.
 - [80] Tsung-Han Chan, Wing-Kin Ma, A. Ambikapathi, and Chong-Yung Chi. A simplex volume maximization framework for hyperspectral endmember extraction. *Geoscience and Remote Sensing, IEEE Transactions on*, 49(11):4177–4193, Nov 2011.
 - [81] S. Matteoli, M. Diani, and G. Corsini. A tutorial overview of anomaly detection in hyperspectral images. *Aerospace and Electronic Systems Magazine, IEEE*, 25(7):5–28, July 2010.
 - [82] D.W.J. Stein, S.G. Beaven, L.E. Hoff, E.M. Winter, A.P. Schaum, and A.D. Stocker. Anomaly detection from hyperspectral imagery. *Signal Processing Magazine, IEEE*, 19(1):58–69, Jan 2002.
 - [83] Tiziana Veracini, Stefania Matteoli, Marco Diani, and Giovanni Corsini. Fully unsupervised learning of gaussian mixtures for anomaly detection in hyperspectral imagery. In *Intelligent Systems Design and Applications, 2009. ISDA '09. Ninth International Conference on*, pages 596–601. IEEE, 2009.
 - [84] S.M. Schweizer and J.M.F. Moura. Hyperspectral imagery: clutter adaptation in anomaly detection. *Information Theory, IEEE Transactions on*, 46(5):1855–1871, Aug 2000.
 - [85] Mohsen Zare Baghbidi, Kamal Jamshidi, Ahmad Reza Naghsh Nilchi, and Saeid Homayouni. Improvement of anomaly detection algorithms in hyperspectral images using discrete wavelet transform. *arXiv preprint arXiv:1201.2025*, 2012.
 - [86] Laetitia Chapel and Chloé Friguet. Anomaly detection with score functions based on the reconstruction error of the kernel pca. In Toon Calders, Floriana Esposito, Eyke Hüllermeier, and Rosa Meo, editors, *Machine Learning and Knowledge Discovery in Databases*, volume 8724 of *Lecture Notes in Computer Science*, pages 227–241. Springer Berlin Heidelberg, 2014.

- [87] Heiko Hoffmann. Kernel-PCA for novelty detection. *Pattern Recognition*, 40(3):863 – 874, 2007.
- [88] J. M Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N.M. Nasrabadi, and J. Chanussot. Hyperspectral remote sensing data analysis and future challenges. *Geoscience and Remote Sensing Magazine, IEEE*, 1(2):6–36, 2013.
- [89] Michael Theodore Eismann. *Hyperspectral remote sensing*. SPIE Press Bellingham, 2012.
- [90] X. Ceamanos, S. Douté, B. Luo, F. Schmidt, G. Jouannic, and J. Chanussot. Intercomparison and Validation of Techniques for Spectral Unmixing of Hyperspectral Images: A Planetary Case Study. *IEEE Transactions on Geoscience and Remote Sensing*, 49(11):4341–4358, 2011.
- [91] John B Adams. Interpretation of visible and near-infrared diffuse reflectance spectra of pyroxenes and other rock forming minerals. In C Karr, editor, *Infrared and Raman spectroscopy of lunar and terrestrial materials*, pages 91–116. New York Academic Press, 1975.
- [92] Jean-Pierre Bibring, M Combes, Yves Langevin, Alain Soufflot, C Cara, Pierre Drossart, Thérèse Encrenaz, Stéphane Erard, Olivier Forni, Brigitte Gondet, L Ksanfomalfy, E Lellouch, Ph Masson, V Moroz, Francis Rocard, J Rosenqvist, and Christophe Sotin. Results from the ISM experiment. *Nature*, 341:591–593, 1989.
- [93] Harold Clenet, Patrick C Pinet, Yves Daydou, Frédéric Heuripeau, Christine Rosenberg, David Baratoux, and Serge Chevrel. A new systematic approach using the Modified Gaussian Model: Insight for the characterization of chemical composition of olivines, pyroxenes and olivine–pyroxene mixtures. *Icarus*, 213:404–422, 2011.
- [94] P. Boccacci. *Introduction to Inverse Problems in Imaging*. CRC Press, 1998.
- [95] Yoann Altmann, Steve McLaughlin, and Alfred Hero. Robust linear spectral unmixing using anomaly detection. *IEEE Transactions on Computational Imaging*, 1(2):74–85, 2015.
- [96] Christian Thurau, Kristian Kersting, Mirwaes Wahabzada, and Christian Bauckhage. Descriptive matrix factorization for sustainability adopting the principle of opposites. *Data Mining and Knowledge Discovery*, 24(2):325–354, 2012.
- [97] G. Camps-Valls and L. Bruzzone. *Kernel methods for remote sensing data analysis*. John Wiley & Sons, 2009.
- [98] Anastasios T. Kyrillidis, Stephen Becker, and Volkan Cevher. Sparse projections onto the simplex. *CoRR*, abs/1206.1529, 2012.

- [99] Jacob Cohen et al. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1):37–46, 1960.
- [100] Frédéric Schmidt, Albrecht Schmidt, Erwan Tréguier, Maël Guiheneuf, Saïd Moussaoui, and Nicolas Dobigeon. Implementation strategies for hyperspectral unmixing using bayesian source separation. *IEEE Transactions on Geoscience and Remote Sensing*, 48(11):4003–4013, 2010.
- [101] Christopher T Russell, Carol A Raymond, A Coradini, H Y McSween, M T Zuber, A Nathues, M C De Sanctis, R Jaumann, a S Konopliv, F Preusker, S W Asmar, R S Park, R Gaskell, H U Keller, S Mottola, T Roatsch, J E C Scully, D E Smith, P Tricarico, M J Toplis, U R Christensen, W C Feldman, D J Lawrence, T J McCoy, T H Prettyman, R C Reedy, M E Sykes, and T N Titus. Dawn at Vesta: testing the protoplanetary paradigm. *Science*, 336(6082):684–686, May 2012.
- [102] Christopher T Russell and Carol A Raymond. The Dawn Mission to Vesta and Ceres. *Space Science Reviews*, 163(1-4):3–23, October 2011.
- [103] Harry Y McSween, Richard P Binzel, Maria Cristina De Sanctis, Eleonora Ammannito, Thomas H Prettyman, Andrew W Beck, Vishnu Reddy, Lucille Le Corre, Michael J Gaffey, Thomas B McCord, Carol A Raymond, and Christopher T Russell. Dawn; the Vesta-HED connection; and the geologic context for eucrites, diogenites, and howardites. *Meteoritics & Planetary Science*, 48(11):2090–2104, November 2013.
- [104] Eleonora Ammannito, Maria Cristina De Sanctis, Fabrizio Capaccioni, Maria Teresa Capria, Francesco Carraro, Jean-Philippe Combe, Sergio Fonte, Alessandro Frigeri, Steven P Joy, Andrea Longobardo, Gianfranco Magni, Simone Marchi, Thomas B McCord, Lucy A McFadden, Harry Y McSween, Ernesto Palomba, Carlé M Pieters, Carol A Polanskey, Carol A Raymond, Jessica M Sunshine, Federico Tosi, Francesca Zambon, and Christopher T Russell. Vestan lithologies mapped by the visual and infrared spectrometer on Dawn. *Meteoritics & Planetary Science*, 48(11):2185–2198, September 2013.
- [105] Eleonora Ammannito, Maria Cristina De Sanctis, Ernesto Palomba, Andrea Longobardo, David W Mittlefehldt, Harry Y McSween, Simone Marchi, M Teresa Capria, Fabrizio Capaccioni, Alessandro Frigeri, Carlé M Pieters, Ottaviano Ruesch, Federico Tosi, Francesca Zambon, Francesco Carraro, Sergio Fonte, Harald Hiesinger, Gianfranco Magni, Lucy A McFadden, Carol A Raymond, Christopher T Russell, and Jessica M Sunshine. Olivine in an unexpected location on Vesta’s surface. *Nature*, 504:122–125, 2013.
- [106] Andrew W Beck and Harry Y McSween. Diogenites as polymict breccias composed of orthopyroxenite and harzburgite. *Meteoritics & Planetary Science*, 45(5):850–872, August 2010.

- [107] Jean-Philippe Combe, Thomas B McCord, Lucy A McFadden, Simone Ieva, Federico Tosi, Andrea Longobardo, Alessandro Frigeri, Maria Cristina De Sanctis, Eleonora Ammannito, Ottaviano Ruesch, et al. Composition of the northern regions of Vesta analyzed by the Dawn mission. *Icarus*, 259:53–71, 2015.
- [108] Harold Clenet, Martin Jutzi, Jean-Alix Barrat, Erik I. Asphaug, Willy Benz, and Philippe Gillet. A deep crust-mantle boundary in the asteroid 4 Vesta. *Nature*, 511(7509):303–306, July 2014.
- [109] Maria Cristina De Sanctis, Angioletta Coradini, Eleonora Ammannito, G Filacchione, M T Capria, S Fonte, G Magni, A Barbis, A Bini, M Dami, I Ficai-Veltroni, and G Preti. The VIR Spectrometer. *Space Science Reviews*, 163(1-4):329–369, October 2010.
- [110] J Anderson, Kris J Becker, Timothy N Titus, Maria Cristina De Sanctis, Andreas Nathues, Federico Tosi, S Schroeder, Lucille Le Corre, Michael S Kelley, Carol A Raymond, and Christopher T Russell. Isis cartographic tools for the Dawn Framing Camera and Visual and Infrared Spectrometer. In *AGU Fall Meeting*, pages #U31A–0009. American Geophysical Union, 2011.
- [111] Vishnu Reddy, Andreas Nathues, Lucille Le Corre, Holger Sierks, Jian-Yang Li, Robert Gaskell, Timothy J McCoy, Andrew W Beck, Stefan E Schröder, Carlé M Pieters, Kris J Becker, Bonnie J Buratti, Brett W Denevi, David T Blewett, Ulrich Christensen, Michael J Gaffey, Pablo Gutierrez-Marques, Michael Hicks, Horst Uwe Keller, Thorsten Maue, Stefano Mottola, Lucy A McFadden, Harry Y McSween, David W Mittlefehldt, David P O’Brien, Carol Raymond, and Christopher T Russell. Color and albedo heterogeneity of Vesta from Dawn. *Science*, 336(6082):700–704, 2012.
- [112] Maria Cristina De Sanctis, Eleonora Ammannito, M Teresa Capria, Fabrizio Capaccioni, Jean-Philippe Combe, Alessandro Frigeri, Andrea Longobardo, Gianfranco Magni, Simone Marchi, Thomas B McCord, Ernesto Palomba, Federico Tosi, Francesca Zambon, Francesco Carraro, Sergio Fonte, Y J Li, Lucy A McFadden, David W Mittlefehldt, Carlé M Pieters, Ralf Jaumann, Katrin Stephan, Carol A Raymond, and Christopher T Russell. Vesta’s mineralogical composition as revealed by the visible and infrared spectrometer on Dawn. *Meteoritics & Planetary Science*, 48(11):2166–2184, July 2013.
- [113] J-Y Li, Bonnie J Buratti, F Cappaccioni, M Teresa Capria, Lucille Le Corre, Brett W Denevi, Maria Cristina De Sanctis, H Hoffmann, Michael D Hicks, L Jorda, Horst Uwe Keller, N Mastrodemos, Stefano Mottola, Andreas Nathues, Carlé M Pieters, Vishnu Reddy, Carol A Raymond, T Roatsch, Christopher T Russell, Stefan E Schröder, M V Sykes, and Timothy N Titus. Photometric properties of Vesta. In *Asteroids, Comets, Meteors*, page 6387, 2012.

- [114] Ottaviano Ruesch, Harald Hiesinger, Maria Cristina De Sanctis, Eleonora Ammannito, Ernesto Palomba, Andrea Longobardo, Francesca Zambon, Federico Tosi, Maria Teresa Capria, Fabrizio Capaccioni, Alessandro Frigeri, Sergio Fonte, Gianfranco Magni, Carol a. Raymond, and Christopher T. Russell. Detections and geologic context of local enrichments in olivine on Vesta with VIR/Dawn data. *Journal of Geophysical Research: Planets*, 119:1–31, aug 2014.
- [115] Jakob Sigurdsson, Magnus O Ulfarsson, and Johannes R Sveinsson. End-member constrained semi-supervised hyperspectral unmixing. In *Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, page 11, 2014.
- [116] L. Kantorovich. On the translocation of masses. *C.R. (Doklady) Acad. Sci. URSS (N.S.)*, 37:199–201, 1942.
- [117] Y. Rubner, C. Tomasi, and L.J. Guibas. A metric for distributions with applications to image databases. In *Proc. ICCV*, pages 59–66, Jan 1998.
- [118] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, pages 2292–2300, 2013.
- [119] Jean-David Benamou, Guillaume Carlier, Marco Cuturi, Luca Nenna, and Gabriel Peyré. Iterative bregman projections for regularized transportation problems. *SIAM Journal on Scientific Computing*, 37(2):A1111–A1138, 2015.
- [120] Ammannito E. et all. Olivine in an unexpected location on Vesta’s surface, 2013.
- [121] Jean-Philippe Combe, Thomas B McCord, Lucy A McFadden, Simone Ieva, Federico Tosi, Andrea Longobardo, Alessandro Frigeri, Maria Cristina De Sanctis, Eleonora Ammannito, Ottaviano Ruesch, et al. Composition of the northern regions of vesta analyzed by the dawn mission. *Icarus*, 259:53–71, 2015.
- [122] *Sequential Quadratic Programming*, pages 529–562. Springer New York, New York, NY, 2006.

Order No 477