Performance comparison of geometric and statistical methods for endmembers extraction in hyperspectral imagery

Nicolas Dobigeon and Véronique Achard
ONERA, 2 avenue Edouard Belin, BP 4025, 31055 Toulouse cedex 4, France
Dobigeon@enseeiht.fr, Veronique.Achard@onecert.fr

ABSTRACT

Spectral unmixing decomposes an hyperspectral image into a collection of reflectance spectra of the macroscopic materials present in the scene, called endmembers, and the corresponding abundance fractions of these constituents. The purpose of this paper is to compare the performance of several algorithms that process unsupervised endmember extraction from hyperspectral images in the visible and NIR spectral ranges. After giving an analytical formulation of the observations, two significantly different approaches have been described. The first one exploits convex geometry the problem answers to. The second one is based on statistical principles of Independent Component Analysis, which is a classical resolution of the Blind Source Separation issue. First, the performance of the algorithms are compared on synthetic images and sensibility to noise is studied. Then the best methods are applied on part of a HyMap image.

Keywords: Hyperspectral imagery, spectral unmixing, unsupervised endmember extraction, Independent Component Analysis.

1. INTRODUCTION

Hyperspectral sensors provide images which are usually acquired in a few hundred wavelengths simultaneously. After atmospheric correction, the visible and near infrared spectral radiances are converted into pixel reflectance spectra. Consequently, each pixel is characterised by a reflectance measurement vector. Generally, several macroscopic components, called endmembers, are present in the scene: soil, vegetation, water, human buildings, etc. Their spectral signatures contribute to the observed spectra of the pixels. *Spectral unmixing* is the procedure by which the spectrum of each pixel is decomposed into a series of pure spectra and their respective abundances. The analytical model which is chosen to describe the observations intrinsically determines the techniques that can be implemented to perform *spectral unmixing*. The most commonly used model is linear [5]. It constitutes a good approximation in this context, that is to say the reflective spectral domain ranging from 0.4 μm to 2.5 μm. It assumes that the spectrum X_p of a pixel p in L bands is a linear combination of the spectra S_i of the M pure materials that appear in the scene. It can be written:

$$X_p = \sum_{i=1}^{M} a_{pi} S_i$$

where

- $X_p = [x_p(1),..., x_p(L)]$ is a $1 \times L$ vector which denotes a pixel with L spectral bands,
- $S_i = [s_i(1),..., s_i(L)]$ is the $1 \times L$ spectrum of the i^{th} material $(i = 1,..M)$,
- a_{pi} is the fraction of the i^{th} material in the p^{th} pixel.

With obvious matricial notations, it follows:

$$X = AS \qquad (1)$$

Since coefficients a_{pi} represent proportions, the model has two following additional constraints:

$$\begin{cases} a_{pi} \geq 0, \quad i = 1,..M \\ \sum_{i=1}^{M} a_{pi} = 1 \end{cases} \qquad p = 1,..N$$

Image and Signal Processing for Remote Sensing XI, edited by Lorenzo Bruzzone
Proceedings of SPIE Vol. 5982 (SPIE, Bellingham, WA, 2005)
0277-786X/05/$15 · doi: 10.1117/12.626563

To solve equation (1) where both A and S are unknown, two kinds of approaches have been already suggested. The first one calls upon convex geometry principles the problem answers to and is going to be described in the second paragraph. The second one, dealt in the third paragraph, is funded on blind source separation, a technique that proved reliable in the fields of acoustics, biomedical, etc.

2. THE SPECTRAL UNMIXING, A CONVEX GEOMETRY ISSUE

The spectra of observed pixels can be represented in an L-dimensional scatterplot, the *hyperspectral space*, as points whose coordinates are L-tuples of reflectance values. If the previously defined mixing model is correct, it can be shown that the dataset of the observations does not occupy all this *hyperspectral space*. More precisely, there is a link between the number of endmembers which appears in the image and the dimension of the subset really occupied by the data: the pixels of an image where M materials are present form a M-simplex, that is to say the simplest geometric shape that encloses a space of dimension $M-1$. For instance, a 2-simplex is a segment, a 3-simplex is a triangle (Fig. 1), a 4-simplex is a tetrahedron, etc. Moreover, the vertices of this simplex are the sought endmember spectra.

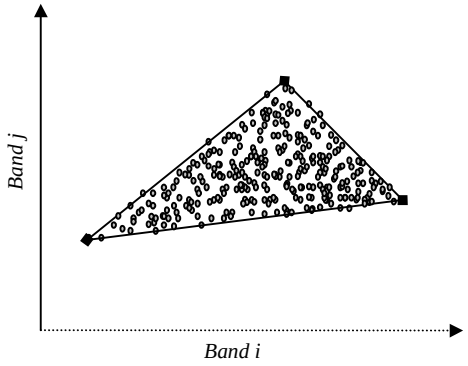


Figure 1 : A simplex in 2-dimension space

According to the considerations of the last paragraph, to determine the number of constituents present in the image, unknown *a priori*, it is sufficient to find the dimension of the space really occupied by the dataset.
A dimension reduction via Principal Components Analysis (PCA), which consists in looking for the high variance axes, allows decreasing the number of bands in an image to be one less than the number of endmembers. This classical pre-processing step allows projecting the data vectors onto the Principal Component Space. Let X be the $L \times N$ matrix which collects the spectra of the N pixels of the image in L spectral bands. The data are centered:

$$Y_i = X_i - \overline{X_i}$$

where $\overline{X_i} = \frac{1}{L}\sum_{k=1}^{L} X_{k,i}$ is the empirical average of each observation.

The variance/covariance matrix of the centered data is defined by : $C_Y = YY^T$.

Let Λ be the diagonal matrix formed by the eigenvalues of C_Y : $\Lambda = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_L \end{pmatrix}$ with $\lambda_1 \geq \ldots \geq \lambda_L$

and V the $L \times L$ matrix of the eigenvectors V_i ($1 \times L$ vector): $V = \begin{pmatrix} \leftarrow & V_1 & \rightarrow \\ & \vdots & \\ \leftarrow & V_L & \rightarrow \end{pmatrix}$.

The spherised data are: $Z = \Lambda^{-\frac{1}{2}} VY$.

Several approaches can be used to determine the number of bands to be retained after the PCA. We chose the following one. Since the eigenvectors are sorted according to the magnitudes of the eigenvalues, most of the information is contained in the first components which are thoses with largest variance. Taking bands in decreasing order of their eigenvalues until their sum be higher than a predefined threshold η, for instance 99%, the number of bands are reduced significantly to $M-1$:

$$\eta = \frac{\lambda_1 + \cdots + \lambda_{M-1}}{\lambda_1 + \cdots + \lambda_L} \times 100 .$$

2.1 N-FINDR algorithm

The vertices of the simplex defined by the dataset are the sought spectra. Two algorithms have been tested to estimate them. The first algorithm, called N-FINDR [9], looks for the "largest" simplex that could be inscribed inside the data. It assumes that the endmembers are present among the mixed pixels. Let $\hat{S}^+$ be the augmented matrix of the pure components:

$$\hat{S}^+ = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \hat{S}_1 & \hat{S}_2 & \cdots & \hat{S}_M \\ \downarrow & \downarrow & & \downarrow \\ 1 & 1 & \cdots & 1 \end{pmatrix}$$

where $\hat{S}_i = S_i^T = [\hat{s}_{1i},...,\hat{s}_{Li}]^T = [s_i(1),...,s_i(L)]^T$ is the column-vector that contains the spectrum of the i^{th} component in the $M-1$ bands retained by the PCA. It can be noticed that $\hat{S}^+$ is a squared matrix.

The volume of the simplex formed by these pure constituents is a function of the determinant of $\hat{S}^+$:

$$V(\hat{S}^+) = \frac{1}{(M-1)!} \left| \det(\hat{S}^+) \right| .$$

Therefore, the procedure consists in searching the set of M pixels with the largest possible volume. The algorithm has a finite complexity since the number of iterations is the combinatory number C_M^N if N is the number of pixels in the image.

2.2 ICE algorithm

The second algorithm, "Iterative Constrained Endmembers" [2], has a more statistical nature. It is based on estimations $\tilde{S}_i$ and $\tilde{a}_{pi}$ of the parameters S_i and a_{pi} . A natural idea is to minimize a quadratic error criteria defined by :

$$J^2 = \sum_{p=1}^{N} {J_p}^2 = \sum_{p=1}^{N} \left\| \tilde{X}_p - X_p \right\|^2$$

where $\tilde{X}_p = \sum_{i=1}^{M} \tilde{a}_{pi} \tilde{S}_i$ and ${J_p}^2 = \left\| \tilde{X}_p - X_p \right\|^2 = \sum_{j=1}^{L} \left(\tilde{X}_p(j) - X_p(j) \right)^2$.

It deals with a least square minimisation issue. The solutions of this problem are all the M -simplex which encloses the dataset formed by the observations. It is necessary to introduce a regularisation term to take into account the size of the simplex. Instead of using the volume of the simplex, which would require too much computation time due to the computation of the determinant, another measure of the size is used : it is the sum of the squared distances between the vertices of the simplex. Then, the quantity to be minimised is :

$$RSS_{reg} = \sum_{j=1}^{L} \left\{ \left(X(j) - A \times S(j) \right)^T \left(X(j) - A \times S(j) \right) + \lambda S(j)^T \times D \times S(j) \right\}$$

3. THE SPECTRAL UNMIXING, A BLIND SOURCE SEPARATION ISSUE

For a decade, a lot of publications have focused on Independent Component Analysis, and more generally Blind Source Separation, because it represents a efficient means to resolve problem of *mixing/unmixing*. The issue can be formulated in the following way [3] :

N stochastic processes or *observations*, noted $(x_i)_{i=1..N}$, result from a linear mixing of M stochastic processes or *sources*, noted $(s_i)_{i=0..M}$. The linear transformation is noted A. It can be written :

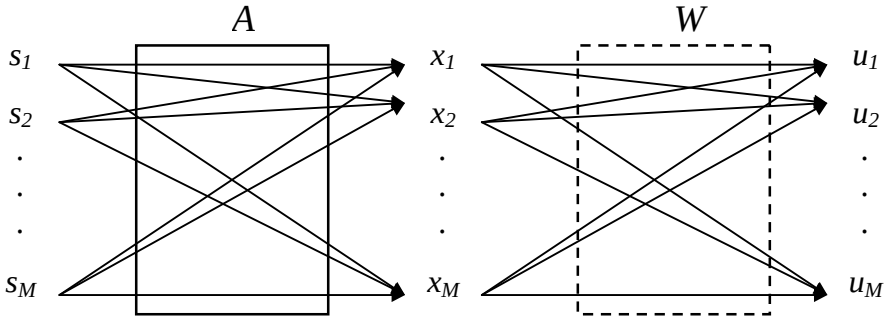
$$\begin{pmatrix} x_1(1) & \cdots & x_1(L) \\ \vdots & & \vdots \\ x_N(1) & \cdots & x_N(L) \end{pmatrix} = \begin{pmatrix} a_{11} & \cdots & a_{1M} \\ \vdots & & \vdots \\ a_{N1} & \cdots & a_{NM} \end{pmatrix} \begin{pmatrix} s_1(1) & \cdots & s_1(L) \\ \vdots & & \vdots \\ s_M(1) & \cdots & s_M(L) \end{pmatrix}$$

$$X(t) = AS(t) \quad \text{or} \quad X = AS ,$$

where $X(t) = [x_1(t),..., x_N(t)]^T$ and $S(t) = [s_1(t),..., s_M(t)]^T$.

The aim of the blind source separation is, starting from the known observations $(x_i)_{i=1..N}$, to determine M random variables $(u_i)_{i=0..M}$ which are estimations of the unknown sources $(s_i)_{i=0..M}$.

Under three additional assumptions – the sources are mutually independent, there are more observations than sources, at the most one source has Gaussian distribution –, the problem consists in finding a linear transformation W to be applied on the observations in order to obtain the most statistically independent estimated sources.



As it deals with exploiting the independent feature of the sources, a criterion that measures this independence has to be introduced. In the information theory framework, the mutual information $I(u_1,...,u_M)$ between several random variables is a positive quantity that becomes null when the processes are mutually independent. If f_i are the probability density functions of the estimations $(u_i)_{i=0..M}$, the derivative of the information is :

$$\frac{\partial I(u_1,...,u_M)}{\partial W} = (W^T)^{-1} + \left(\frac{f_i'(u_i)}{f_i(u_i)} \right)_i X^T .$$

Hence, the learning rule with the natural gradient is :

$$\Delta W \propto \left[I - \left(\varphi(u_i)\right)_i U^T \right] W$$

where $U = [u_1 \quad \cdots \quad u_M]^T$ and $\left(\varphi(u_i)\right)_i$ is the column-vector such as :

$$\varphi_i(u_i) = -\frac{f_i'(u_i)}{f_i(u_i)} .$$

Of course, f_i are unknown functions. Bayliss proposes a parametric model to estimate them [1,8]. A Matlab implementation of this approach was tested with sources different from those presented. However, the algorithm too often diverges.
Finally, we chose to implement the "extended infomax" algorithm as it is described by Lee et al. [7]. It allows to separate mixed sources with subgaussian and supergaussian distributions, that is to say respectively with negative and positive kurtosis values. The learning rule is :

$$\Delta W \propto \left[I - \left(K \tanh(U) + U \right) U^T \right] W$$

where $K = (k_{ij})$ is a $M \times M$ diagonal matrix that offers the choice between subgaussian and supergaussian estimated sources :

$$k_{jj} = \begin{cases} +1 & \text{if the } j^{th} \text{ source is supergaussian} \\ -1 & \text{if the } j^{th} \text{ source is subgaussian} \end{cases}$$

At this point, it is imperative to make a parallelism between ICA and spectral unmixing. Precisely, we have to answer the following questions. What are *observations* called? What are *sources* called? In the literature, it can be found two ways of answering these questions.
The first intuitive approach (ICA1) consists in considering pixel spectra as observations. Therefore, the sources are the endmember spectral signatures. The mixing equation introduced by spectral mixing is identically the ICA one.

$$\begin{pmatrix} x_1(1) & \cdots & x_1(L) \\ \vdots & & \vdots \\ x_N(1) & \cdots & x_N(L) \end{pmatrix} = \begin{pmatrix} a_{11} & \cdots & a_{1M} \\ \vdots & & \vdots \\ a_{N1} & \cdots & a_{NM} \end{pmatrix} \begin{pmatrix} s_1(1) & \cdots & s_1(L) \\ \vdots & & \vdots \\ s_M(1) & \cdots & s_M(L) \end{pmatrix}.$$

The second approach (ICA2) appears when the last equation is transposed:

$$\begin{pmatrix} x_1(1) & \cdots & x_N(1) \\ \vdots & & \vdots \\ x_1(L) & \cdots & x_N(L) \end{pmatrix} = \begin{pmatrix} s_1(1) & \cdots & s_M(1) \\ \vdots & & \vdots \\ s_1(L) & \cdots & s_M(L) \end{pmatrix} \begin{pmatrix} a_{11} & \cdots & a_{N1} \\ \vdots & & \vdots \\ a_{1M} & \cdots & a_{NM} \end{pmatrix}.$$

Now, one observation is the reflectance values in one band for all the pixels. Consequently, one source is the abundances of one material in all the pixels. However, in this configuration, the sources are not statically independent as the linear mixing model imposes that, for each pixel, the sum of all the material abundances is equal to 1. In this constrained case, it can be demonstrated that ICA will generate estimated sources that will be variations with an average abundances [6]. The minimum and the maximum of these sources indicate pure pixels. By picking these extreme values in the estimated source matrix, the original endmember spectra can be found. Therefore, with this approach, only two sources can be recovered.

4. TEST ON SYNTHETIC IMAGES

To test the two algorithms on synthetic images, two aspects have to be chosen: firstly, the pure components involved in the mixture and, secondly, their spatial repartitions and respective abundances in the image. The chosen endmembers are representative of urban or suburban environment. The spectra of the components are shown Fig. 2.
The spatial repartitions are displayed Fig. 3. On these simulated images, several areas are defined where one or several components are distributed randomly with different averages.

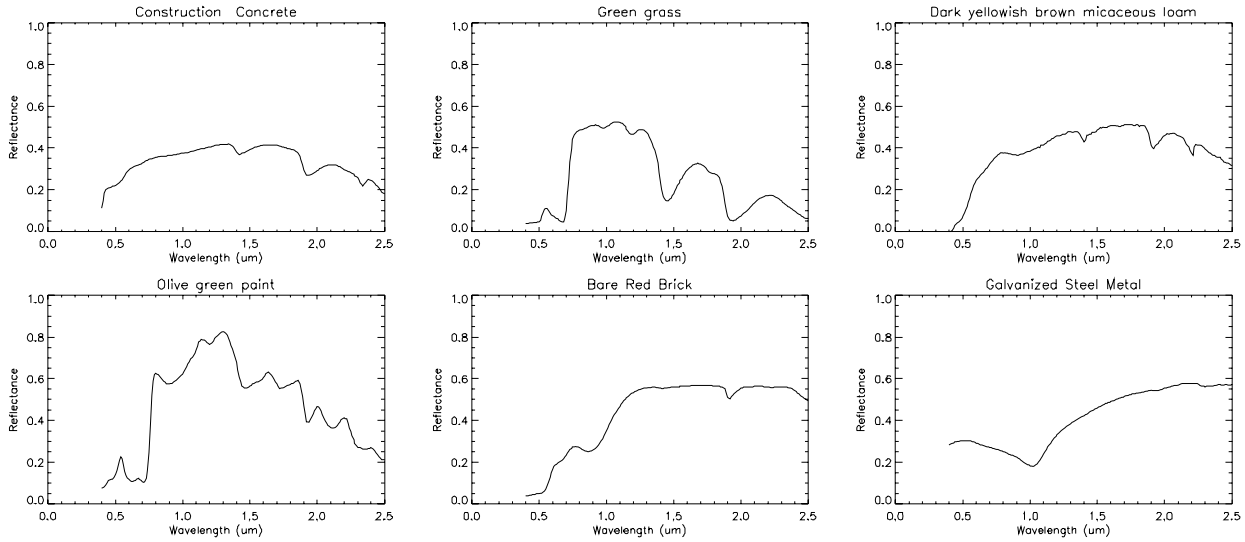


Figure 2 : Spectra of the endmembers involved in the mixture

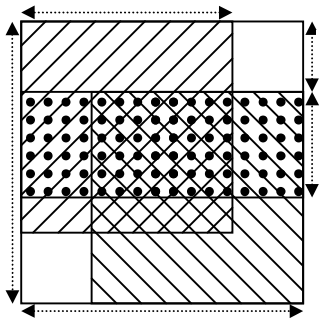


Figure 3 : Spatial repartition of the original materials

To evaluate the robustness of the algorithms, an additive uniform noise is added to the observations in all the bands and for each pixel. The mixing equation becomes :

$$X = AS + B$$

where B is a $N \times L$ matrix if N pixels compose the image and if the reflectances are acquired in L spectral bands with b_{ij} uniformly distributed in [-β ;+ β] . The tests are done with values of β equal to 0, 0.15 and 0.40.

To measure the quality of the estimation, a quantity which gauges the likelihood between the real spectrum x and its estimation $\hat{x}$ has to be introduced. In the field of hyperspectral imagery, the comparison of signals resorts to the variable called *spectral angle* which measures the angle between two spectra in the hyperspectral space [4]:

$$\theta(x,\hat{x}) = \arccos\left(\frac{\langle x,\hat{x}\rangle}{\|x\|_2\|\hat{x}\|_2}\right)$$

where $\langle .,. \rangle$ is the usual scalar product and $\|.\|_2$ the corresponding norm. Afterwards, the cosine of this angle is used :

$$SA(x,\hat{x}) = \cos\left(\frac{\langle x,\hat{x}\rangle}{\|x\|_2\|\hat{x}\|_2}\right)$$

When $x = \hat{x}$, $SA(x,\hat{x}) = 1$. $SA(x,\hat{x}) > 0.99$ means that x and $\hat{x}$ have very similar shapes.

The four algorithms are applied on mixtures of two, three and four constituents. For ICA algorithms, results are based on 200 Monte Carlo runs. SA are averaged for each mixture and plotted as a function of the noise magnitude (Fig. 4 to 6).

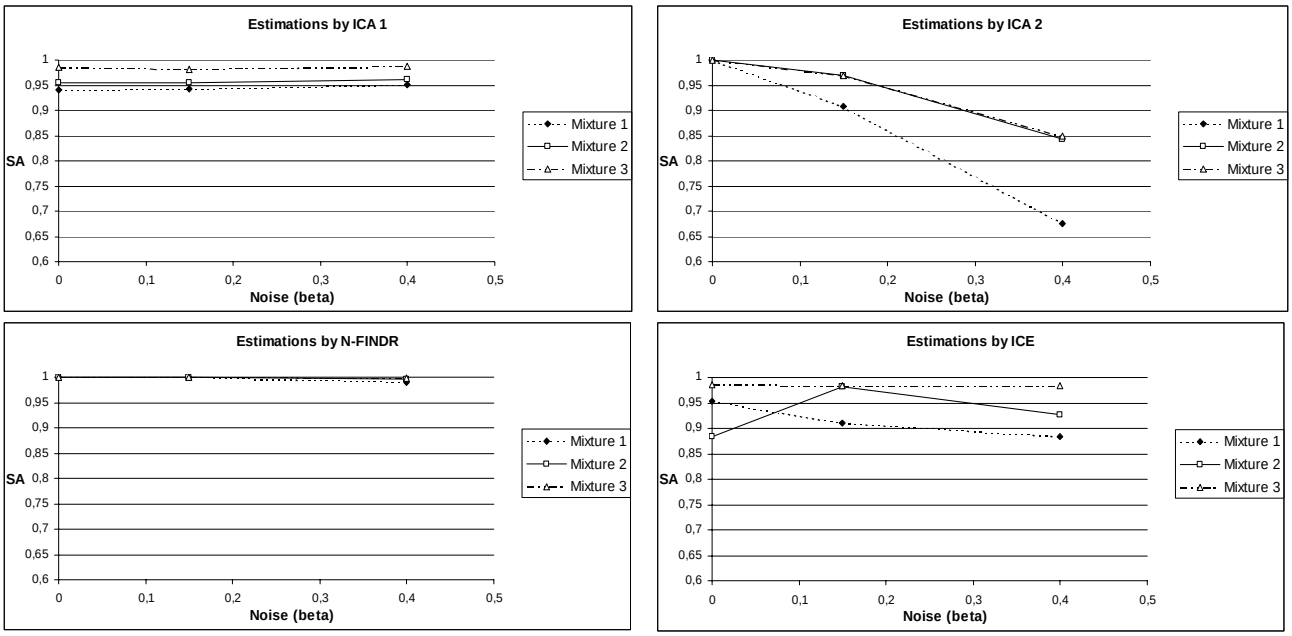


Figure 4 : Cosines of the Spectral Angles between real endmembers spectra and endmembers spectra estimated by the four algorithms for three different mixtures of two components

The ICA 2 algorithm appears to be very sensitive to the level of noise (Fig. 4). For a mixture of two components, N-FINDR performs an accurate endmember extraction. The two other algorithms provide approximately the same results, less accurate than those obtained with N-FINDR and they seem to be sensitive to the sources nature.

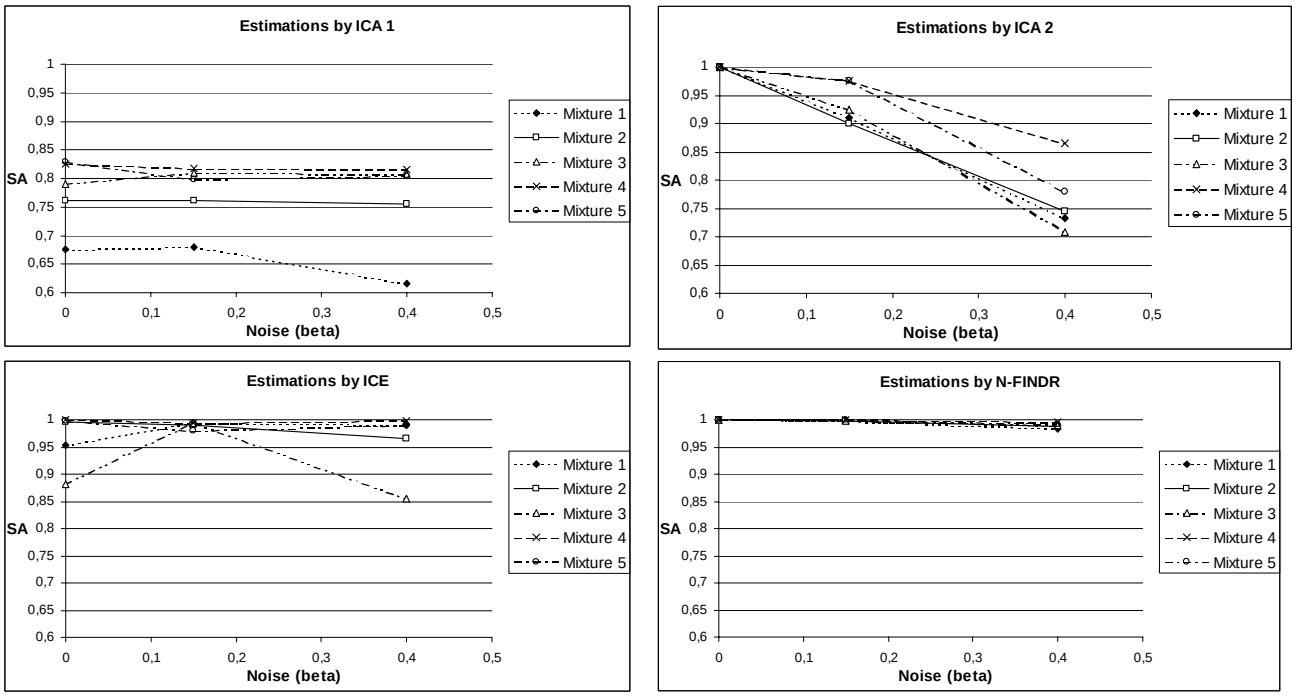


Figure 5 : Cosines of the Spectral Angles between real endmembers spectra and endmembers spectra estimated by the four algorithms for five different mixtures of three components

As previously, for a mixture of three components, N-FINDR is the most accurate. The two algorithms based on Independent Component Analysis are not able to estimate the spectra of the endmembers reliably (Fig. 5). At this stage of the simulations, it was decided to carry on with the two most performing algorithms based on geometric considerations, that is to say the N-FINDR and the ICE algorithms.

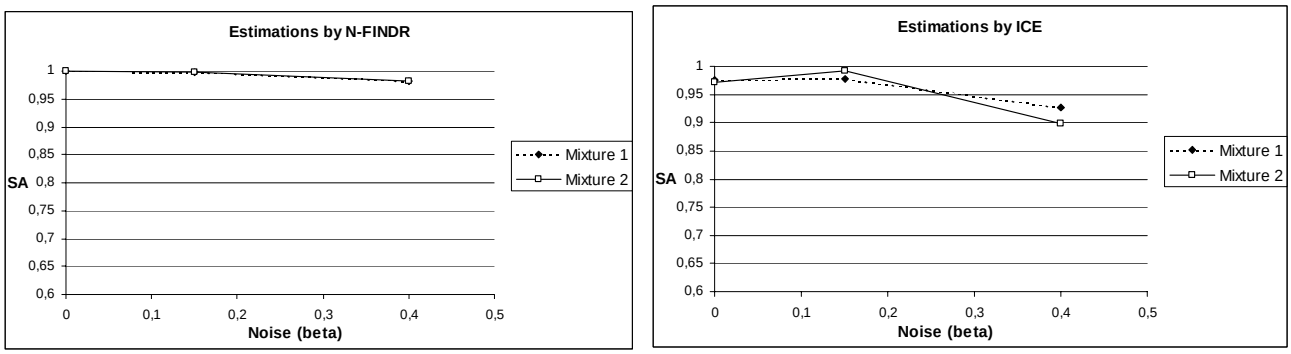


Figure 6 : Cosines of the Spectral Angles between real endmembers spectra and endmembers spectra estimated by N-FINDR and ICE algorithms for two different mixtures of four components

When four components are present in the image (Fig. 6), N-FINDR still gives an accurate estimation of the endmembers, even with moderately noisy spectra (β=0.15). For higher (β=0.4), the quality of the retrievals begins to decrease, but is still acceptable.

At the issue of these tests on synthetic images, N-FINDR seems to provide the most accurate endmember estimates. ICE results are less satisfying, as the $SA(x,\hat{x})$ values are less than 0.99 in most cases. The ICA 1 algorithm fails for more than two sources; the ICA 2 algorithm is very sensitive to the noise.

5. TEST ON HYMAP IMAGE

The two "geometric" algorithms are tested on a hyperspectral image acquired by the HyMap instrument over the Hartheim conifer forest. On Fig. 7, the 66x66 pixels of the studied image are represented at the wavelength λ=0.701 µm (spectral band 19). The forest is crossed by grass tracks. Two water points appear in black on the image. We defined regions of interest (ROIs), characteristic of each constituent, where the spectra of the pixels are averaged.

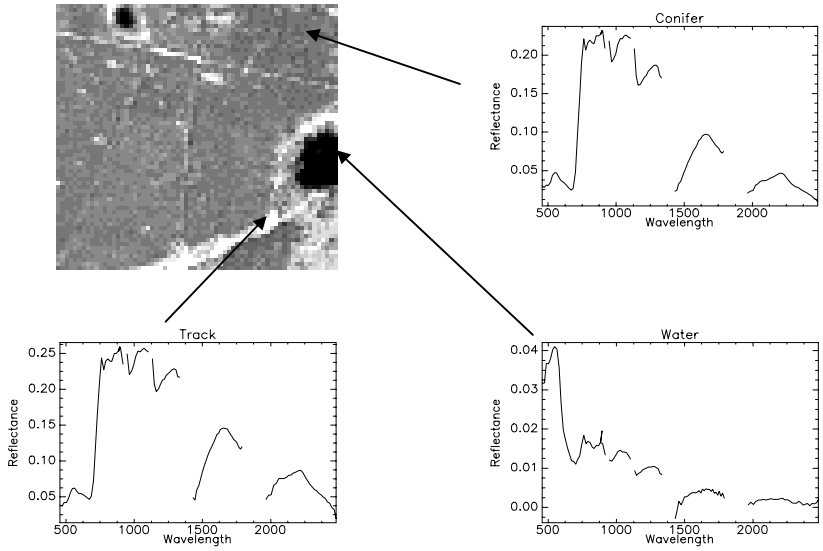


Figure 7 : Image of Harheim and mean spectra of ROIs

On Fig. 8, the estimated endmembers spectra (red curve) are superposed with the ground measurements over conifer (blue curve) taken from a 10 meters pylon, and with the extreme airborne spectra in the corresponding ROIs. Spectral angles between endmembers and ROIs mean spectra are computed.

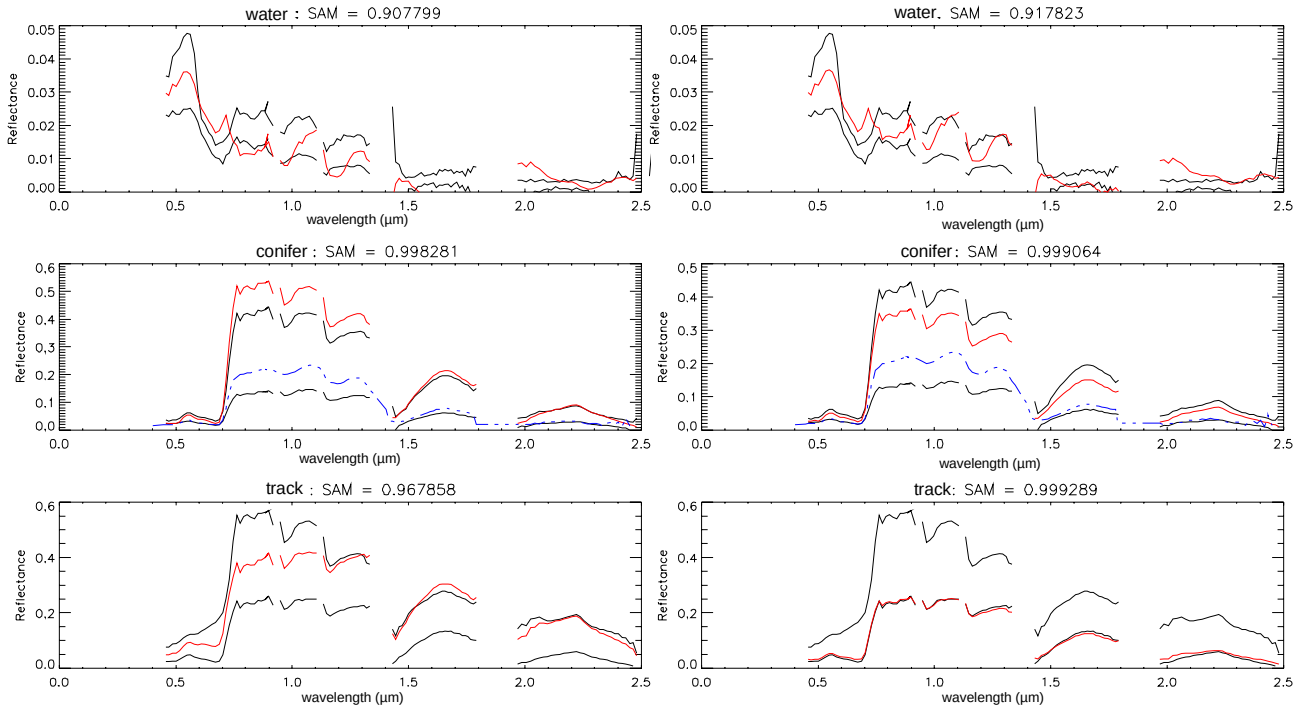


Figure 8 : Estimated endmember spectra in Hartheim image by N-FINDR (left) and ICE (right) - Endmembers spectra (red), field measurements, when available (blue), and ROIs airborne spectra (mimimum and maximum in black) - (SAM : cosine of spectral angles).

The extraction of the water is difficult because of its low reflectance. The conifer spectrum is well estimated by the two algorithms and the third endmember is quite representative of spectra collected on the track composed of a mixture of more or less dry and green grass.

6. CONCLUSION AND FINAL REMARKS

This paper consisted in comparing performances of four algorithms dedicated to resolving the problem of endmembers extraction in hyperspectral images. A description of the observations by a linear model allowed introducing two different approaches. The first one, based on stochastic notion of Independent Component Analysis, does not seem to provide an efficient mean to estimate the sought spectra reliably as soon as more than two components are present in the image or as soon as the observations are very noisy. The second approach exploits geometric characteristics the problem answers to. At the issue of tests on synthetic images, N-FINDR seems to provide the most accurate endmember estimates. ICE results are less satisfying, as the values of the spectral angles cosine between real and estimated endmembers spectra are less than 0.99 in most cases. On the other hand, when applied to a part of HyMap image, the “geometric” algorithms, are able to provide good estimates of endmembers spectra, with slightly better results for the ICE algorithm.

ACKNOWLEDGMENT

DAISEX’99 data including HyMap image on Hartheim is courtesy of Patrick Wurteisen, European Spatial Agency; many thanks to Marc-Phillipe Stoll and Françoise Nerry, LSIIT, for the supplying of the data.

REFERENCES

[1] J. Bayliss, J.A. Gualtieri, R.F. Cromp, « Analyzing Hyperspectral Data with Independent Component Analysis », *Proc. SPIE*, vol. 3240, pp. 133-143, 1998.

[2] M. Berman, H. Kiiveri, Ryan Lagerstrom, A. Ernst, R. Dunne, J. Huntington, « ICE : An Automated Statistical Approach to Identifying End-members in Hyperspectral Images », *Proc. IEEE Geosciences and Remote Sensing Symposium, IGARSS'03*, vol. 1, pp. 279-283, 2003.

[3] A. Hyvarinen, J. Karhunen, E. Oja, *Independent Component Analysis*, Wiley Interscience, New York, 2001.

[4] N. Keshava, « A Survey of Spectral Unmixing Algorithms », *Lincoln Laboratory Journal*, Vol. 14, n°1, Massachusetts Institute of Technology, pp. 55-78, 2003.

[5] N. Keshava, J.F. Mustard, « Spectral Unmixing », *IEEE Signal Processing Mag.*, vol. 19, pp. 44-57, 2002.

[6] C.Y. Kuan, G. Healey, « Using Source Separation Methods for End-member Selection », *Proc. SPIE*, vol. 4725, pp. 10-17, 2002.

[7] T.W. Lee, M. Girolami, T.J. Sejnowski, « Independent Component Analysis Using an Extended Infomax Algorithm for Mixed Subgaussian and Supergaussian Sources », *Neural Computation*, vol.11, n°2, pp. 417-441, 1999.

[8] B.A. Pearlmutter, L.C. Parra, « Maximum Likelihood Blind Source Separation : A Context-Sensitive Generalization of ICA », *Proc. Neural Information Processing Systems, NIPS'96*, vol. 9, pp. 613-619, 1997.

[9] M.E. Winter, « Fast Autonomous Spectral End-member Determination in Hyperspectral Data », *Proceedings of the 13th International Conference on Applied Geologic Remote Sensing*, Vancouver, vol. 2, pp. 337-344, 1999.