# On the decomposition of Mars hyperspectral data by ICA and Bayesian positive source separation

Saïd Moussaoui, Hafrun Hauksdottir, Frédéric Schmidt, Christian Jutten, Jocelyn Chanussot, David Brie, Sylvain Douté, Jon Atli Benediktsson

# On the Decomposition of Mars Hyperspectral Data by ICA and Bayesian Positive Source Separation

S. Moussaoui [a], H. Hauksdóttir [b,c], F. Schmidt [d], C. Jutten [c], J. Chanussot [c], D. Brie [e], S. Douté [d], J. A. Benediksson [b]

[a] *Institut de Recherche en Communications et Cybernétique de Nantes, 1 rue de la Noë, BP 92101, F-44321 Nantes Cedex 3, France*
[b] *Dept. of Electrical and Computer Engineering, University of Iceland, Hjardarhaga 2-6, 107, Reykjavík, Iceland*
[c] *Laboratoire GIPSA, INPG, 46 avenue Félix Viallet, F-38031 Grenoble Cedex*
[d] *Laboratoire de Planétologie de Grenoble, Batiment D de Physique, BP 53, F-38041 Saint-Martin d'Hères Cedex 9, Grenoble*
[e] *Centre de Recherche en Automatique de Nancy, B.P. 239, Boulevard des Aiguillettes, F-54500, Vandoeuvre-lès-Nancy Cedex, France*

**Abstract**

The surface of Mars is currently being imaged with an unprecedented combination of spectral and spatial resolution. This high resolution, and its spectral range, give the ability to pinpoint chemical species on the surface and the atmosphere of Mars moreaccurately than before. The subject of this paper is to present a method to extract informations on these chemicals from hyperspectral images. A first approach, based on Independent Component Analysis (ICA) [1], is able to extract artifacts and locations of $CO_2$ and $H_2O$ ices. However, the main independence assumption and some basic properties (like the positivity of images and spectra) being unverified, the reliability of all the independent components (ICs) is weak. For improving the component extraction and consequently the endmember classification, a combination of spatial ICA with spectral Bayesian Positive Source Separation (BPSS) [2] is proposed. To reduce the computational burden, the basic idea is to use spatial ICA yielding a rough classification of pixels, which allows selection of small, but relevant, number of pixels and then BPSS is applied for the estimation of the source spectra using the spectral mixtures provided by this reduced set of pixels. Finally, the abundances of the components is assessed on the whole pixels of the images. Results of this approach are shown and evaluated by comparison with available reference spectra.

*Key words:* Mars Express mission, hyperspectral data, source separation, independent component analysis, Bayesian source separation, positivity constraint.

## 1. Introduction

The Mars Express (European Space Agency) on board instrument OMEGA [3] (Observatoire pour la Minéralogie, l'Eau, les Glaces et l'Activité) is an imaging spectrometer, which provides hyperspectral images of planet Mars, with a spatial resolution from 300 m to 4 km, on 256 frequency channels in the near infrared spectral band and 128 channels in the visible range. This high spatial resolution coupled with its wide spectral range, give the ability to pinpoint chemical species on the surface and the atmosphere of Mars more accurately than before.

As solar light - incident to the planet - is partially transmitted, reflected and diffused back by interaction with the different constituents of the atmosphere and the surface, the analysis of reflectance spectra may allow the identification and the quantification of the chemical species present along the light path. The simplest model of this physics claims that each measured spectrum at the ground is a mixture of the pure component spectra. In this model, the planetary surface area associated to each pixel, due to the spatial resolution, can contain a few constituents, and the reflected light is a linear mixture - called geographical - of chemical species present on the surface. Consequently, recovering the endmembers and their concentrations, is equivalent to solving a source separation and spectral unmixing problems, for which, at first glance, independent component analysis (ICA) algorithms seem to be attractive candidates.

However, in this problem as in any actual applications of blind source separation, we don't know exactly the sources. Since ICA always provides *independent* [1] *com-*

[1] in fact, only as independent as possible

*ponents* (ICs), we have to pay attention to their physical interpretation. In fact, additional knowledges can be used. In this purpose: $(i)$ synthetic reference spectra of the main endmembers obtained after inversion [4]; $(ii)$ a supervised classification using wavalet transform called *wavanglet* which is in accordance with Mars physical knowledge [5]. Finally, a last difficulty is to check the relevance of the linear mixture model as well as the hypothesis on which the algorithm is based.

From a methodological point of view, the objective of this paper is to point out that, when source independence assumption is not fully satisfied, an ICA algorithm can provide spurious ICs and one has to prefer semi-blind methods which relax partially independence assumption and accounts for additional informations. Especially, in hyperspectral imaging an evident prior concerns the positivity of the images and the component spectra.

The paper is organized as follows. Section 2 presents the simplified observation model in the case of a geographical mixture and the possible decomposition models. Section 3 recalls briefly the source separation problem. Section 4 presents the results when applying ICA to hyperspectral data, and discuss the relevance of the separation. Section 5 introduces the Bayesian framework and shows how Bayesian methods can ensure the positivity of the sources and of the mixing coefficients. The results on Mars hyperspectral data are then discussed. Section 7 recalls the main results and gives some perspectives of this research work.

## 2. Hyperspectral Data Modeling

The OMEGA spectrometer, carried by Mars Express spacecraft on an elliptical orbit, has a spatial resolution range from 300 m to 4 km. This instrument has three channels, a visible channel and two near infrared channels. We will focus in this work only on the near infrared channels since the behavior between major chemicals can be discriminated in this spectral range. The analysis is focused on a data set consisting in a single hyperspectral data cube obtained by looking to the South Polar Cap of Mars in the local summer where $CO_2$ ice, water ice and dust were previously detected [5, 6]. This data cube is made up with 2 channels: 128 spectral planes from 0.93 $\mu m$ to 2.73 $\mu m$ with a resolution of $0.013\mu m$ and 128 spectral planes from $2.55\mu m$ to $5.11\mu m$ with a resolution of $0.020\mu m$. After calibration, the dimensionless physical unit used to express the spectra is the "reflectance", which is the ratio between the irradiance leaving each pixel toward the sensor and the solar irradiance at the ground. Interactions between photons coming from the sun and the planet Mars, through its atmosphere and surface, allows us to identify the different compounds present in the planet. Those compounds are mixed and usually different chemical species can be identified in each measured spectra. Two kinds of physical mixing at the ground can be observed [7]:

– Geographic mixture: each pixel is a patchy area made of several pure compounds. This type of mixture, sometimes called "sub-pixel mixture", happens when the spatial resolution is not large enough to observe the complex geological combination pattern. The total reflectance in this case will be a weighted sum of the pure constituent reflectances. The weights (abundance fractions) associated to each pure constituent are surface proportions inside the pixel.
– Intimate mixture: each pixel is made of one single terrain type which is a mixture at less than the typical mean-path scale (typically the order of 1mm scale). The total reflectance in this case will be a nonlinear function of pure constituent reflectances.

The case of intimate mixtures, which needs nonlinear source separation methods and further development, is not addressed here. In this paper, we perform our analysis with hypothesis of a geographical mixtures and hence linear mixing models.

### 2.1. *Observation Model*

The hyperspectral images can be modeled by examining all the factors that contribute to the radiance signal reaching the sensor after interaction of the sunlight with a planetary surface. An analytical expression of the measured radiance factor in a case of a Lambertian surface [2] with a homogeneous atmosphere has been proposed in [8], under the following assumptions: $(i)$ the multiple diffusion term $r$ and the diffusion terms $E(\mu)$ are negligible, $(ii)$ the path through the atmosphere is equivalent for all pixels, $(iii)$ the direct atmospheric contribution only depends on the wavelength, $(iv)$ the emergence direction is always the same. Thus, based on this model and using the geographic mixture assumption, the radiance factor at location $(x, y)$ and at wavelenght $\lambda$ satisfies the following observation model:

$$L(x,y,\lambda) = \left(\rho_a(\lambda) + \Phi(\lambda)\sum_{p=1}^{P}\alpha_p(x,y)\,\rho_p(\lambda)\right)\cos\left[\theta(x,y)\right] \quad (1)$$

where $\Phi(\lambda)$ is the spectral atmospheric transmission, $\theta(x, y)$ the angle between the solar direction and the surface normal (solar incidence angle), $P$ the number of endmembers in the region of coordinates $(x, y)$, $\rho_p(\lambda)$ the spectrum of the $p$-th endmember, $\alpha_p(x, y)$ its weight in the mixture and $\rho_a(\lambda)$ the radiation that did not arrive directly from the area under view. This mixture model can also be written as:

$$L(x,y,\lambda) = \sum_{p=1}^{P}\alpha'_p(x,y)\cdot\rho'_p(\lambda) + E(x,y,\lambda) \quad (2)$$

where

[2] a surface that reflects the light independently of both incidence and emergence directions

$$\begin{cases} \alpha'_p(x,y) = \alpha_p(x,y)\cos\left[\theta(x,y)\right], \\ \rho'_p(\lambda) = \Phi(\lambda)\,\rho_p(\lambda), \\ E(x,y,\lambda) = \rho_a(\lambda)\cos\left[\theta(x,y)\right]. \end{cases} \quad (3)$$

As it can be seen in equation (3) the true endmember spectra are affected by the atmospheric attenuation and the abundance fractions are corrupted by the solar angle effect. It is not attempted here to remove the atmospheric effect, and thus the spectra obtained are ideally the spectra of the endmembers with atmospheric contribution. On the contrary, it is attempted to correct the solar angle effect to give a map of the constituent proportions in the observed area. In fact, since the abundance fraction is proportional to the quantity of each constituent in the geographical mixture, it can be deduced from the mixture model (2) and equation (3) that the abundance fractions are not altered by the geometrical effect since:

$$\begin{aligned} c_p(x,y) &= \frac{\alpha'_p(x,y)}{\sum_{j=1}^{P}\alpha'_j(x,y)}, \\ &= \frac{\alpha_p(x,y)\cos[\theta(x,y)]}{\sum_{j=1}^{P}\alpha_j(x,y)\cos[\theta(x,y)]}, \\ &= \frac{\alpha_p(x,y)}{\sum_{j=1}^{P}\alpha_j(x,y)}. \end{aligned} \quad (4)$$

### 2.2. *Data Size*

Practically, the data sets used here are collected in the infrared region ranging from 0.96 $\mu m$ to 4.16 $\mu m$, total of 174 wavelengths where noisy, hot and dead spectels [3] have been excluded. Those OMEGA channels correspond to the deterioration of the detector during the time in the space environment. Those channels can be determined during the calibration process. The spatial size of the data sets varies: $323 \leq N_x \leq 751$ and $N_y \in \{64, 128\}$, for a total [4] pixel number: $41344 \leq N_x \cdot N_y \leq 56832$. Each data set (one hyperspectral image) is then a data cube of size $N_x \times N_y \times N_f$ which contains about $10,000,000$ data values, in fact between $7,193,856$ and $9,888,768$ according to the image size.

### 2.3. *Decomposition Models*

Let us now consider a hyperpectral data cube with $N_f$ images of $N_z = (N_x \times N_y)$ pixels obtained from $N_f$ frequency bands. For simplicity, assume raw vectorized images $I(n, \lambda_k)$, with $1 \leqslant n = (i-1)N_y + j \leqslant N_z$ (where $i$ and $j$ are the initial row and column image indices) is the *spatial index* and $k$, $k = 1, \ldots, N_f$, is the *spectral index* for wavelength $\lambda_k$. Consequently, two representations of the hyperspectral data can be considered: spectral and spatial mixture models.

*Spectral Mixture Model:* Each pixel of spatial index $n$ gives an observed spectrum of $N_f$ frequency samples, which is represented by the linear approximation:

$$I_n(\lambda_k) \approx \sum_{p=1}^{N_c} a_{(n,p)}\psi_p(\lambda_k), \quad \forall n = 1, ..., N_z, \quad (5)$$

where $\psi_p(\lambda_k)$, for $p = 1, \ldots, N_c$, are the constituent reflectance spectra, and the number $N_c$ is chosen according to the desired accuracy of the approximation. Denoting the vectorized image (of dimension $N_x \times N_y$) $\mathbf{I}(\lambda_k)$, the $(N_z \times N_c)$ mixing matrix $\boldsymbol{A}$ and $\boldsymbol{\Psi}(\lambda_k) = [\psi_1(\lambda_k), \ldots, \psi_{N_c}(\lambda_k)]^T$, this spectral mixture model is then expressed as:

$$\mathbf{I}(\lambda_k) \approx \boldsymbol{A} \cdot \boldsymbol{\Psi}(\lambda_k). \quad (6)$$

Practically, this spectral model intends to approximate the spectrum of each pixel as a sum of $N_c$ component spectra of the area corresponding to this pixel coordinates. If ICA is used for the estimation, then the $N_c$ basis spectra $\psi_p$, $p = 1, \ldots, N_c$, should be statistically independent. Moreover, the $p$-th column of the matrix $\boldsymbol{A}$ is the unfolded image associated to the basis spectrum $\psi_p$. According to this model, we have $N_x \cdot N_y \approx 50,000$ sensors and a small number of samples $N_f = 174$, for estimating the large matrix $\boldsymbol{A}$ which has $N_z \times N_c \approx 250,000$ parameters (taking $N_c = 5$).

*Spatial Mixture Model:* This model assumes that for each wavelength $\lambda_k$, the measured image $I_{\lambda_k}(n)$ is a weighted sum of $N_c$ basis images, denoted $II_p(n)$, $p = 1, \ldots, N_c$:

$$I_{\lambda_k}(n) \approx \sum_{p=1}^{N_c} b_{(\lambda_k,p)} II_p(n), \quad \forall k = 1, ..., N_f. \quad (7)$$

In vector notations, denoting the $N_f \times N_c$ matrix $\boldsymbol{B}$ and $\mathbf{II}(n) = [II_1(n), \ldots, II_{N_c}(n)]^T$, one can write:

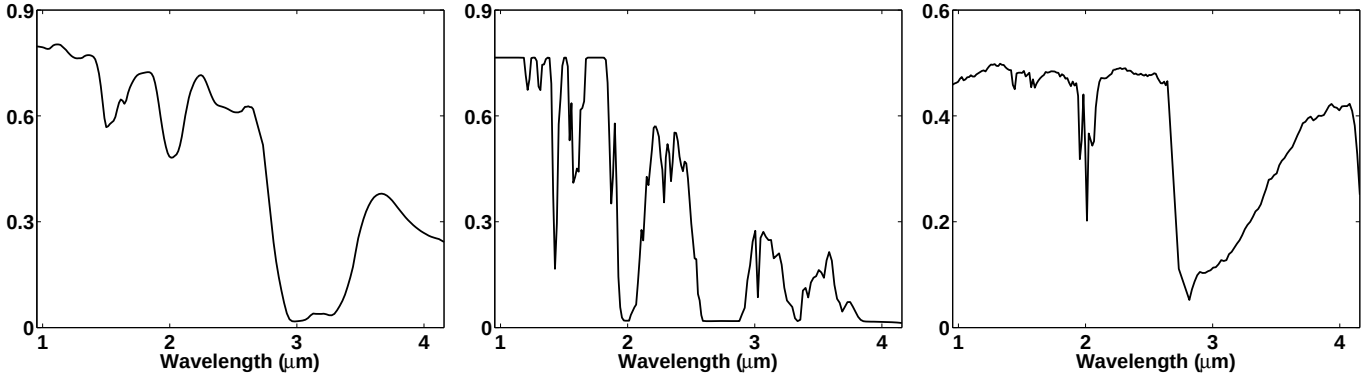
$$\mathbf{I}(n) \approx \boldsymbol{B} \cdot \mathbf{II}(n). \quad (8)$$

Practically, this spatial model intends to approximate the whole image at each frequency as a sum of $N_c$ basis images. If ICA is used for the estimation, then the $N_c$ basis images $II_p$, $p = 1, \ldots, N_c$, should be statistically independent. Moreover, the $k$-th column of the matrix $\boldsymbol{B}$ is the spectrum associated to the basis image $II_k$. According to this model, we have $N_f = 174$ sensors, and a very large number of samples $N_z \approx 50,000$ for estimating the matrix $\boldsymbol{B}$ which has $N_f \times N_c < 900$ parameters (taking $N_c = 5$).

*Comment on the notations:* For limiting notation complexity, in the two models, the hyperspectral dataset is always denoted $\mathbf{I}$, but, in the spatial model, one considers the dataset like a $N_f \times N_z$ matrix while it is a $N_z \times N_f$ matrix in the spectral model. The two matrices contain exactly the same entries, but are in fact transposed from a model to the other one.

[3] respectively, detector with unusual high noise level, detector in a wrong high level and detector in failure

[4] all the $\{N_x, N_y\}$ combinations are not used: the smallest image size is $373 \cdot 128 = 41344$, and the largest is $444 \times 128 = 56832$.




(a) Synthetic $H_2O$ ice spectrum (b) Synthetic $CO_2$ ice spectrum (c) Observed dust spectrum with atmospherical transmission


(d) $CO_2$ ice (e) $H_2O$ ice (f) dust

Figure 1. (a,b,c) Reference spectra and (d,e,f) detection masks produce by the wavanglet classification.

2.4. *Reference Data and Classification*

Because no ground truth is possible on Mars, excepted at some areas where the Mars Exploration Rovers stand, we need some reference informations. To validate the results, two kinds of reference data are used (see figure 1), following one supervised classification study [5] made on the same observations: first are the reference spectra of $CO_2$ ice, $H_2O$ ice and dust, which will be noted as $\psi_{\mathrm{CO_2}}(\lambda)$, $\psi_{\mathrm{H_2O}}(\lambda)$ and $\psi_{\mathrm{dust}}(\lambda)$ respectively. Both reference spectra of $CO_2$ ice and $H_2O$ ice are simulations produced by a radiative transfer model in typical physical conditions of the Permanent South Polar Cap of Mars [4]. These two reference spectra are atmosphere free simulations. The dust reference spectrum is an OMEGA's observation endmember selected with a PCA method. This last reference spectrum contains the atmospheric transmission, in particular in region around 2 $\mu m$ and 2.9 $\mu m$. The second reference information is the results of the wavanglet classification [5], which will be noted as $I_{\mathrm{CO_2}}(n)$, $I_{\mathrm{H_2O}}(n)$ and $I_{\mathrm{dust}}(n)$. This method is based on the correlation between observed and reference spectra in a wavelet filtered space using three main steps: (*i*) keeping only the wavelet at lower scale in order to minimize the effect of the observation geometry ; (*ii*) choosing the wavelet that maximize the difference between reference spectra in order to provide a better classification ; (*iii*) using a correlation coefficient (or angle) in order to discard the scale factor effect. This pattern recognition approach is then supervised because it requires a prior knowledge about the reference spectra. The wavanglet classification method produces classification masks (Fig. 1, (d) to (f)) which are neither unique nor complete, that is pixels can be in more than one class and not all the pixels are classified.

## 3. Source Separation

Source separation consists in retrieving unknown signals, $\boldsymbol{s}(t) = (s_1(t), \ldots, s_n(t))^T$, which are observed through unknown mixtures of them [9, 10]. Denoting the observations $\boldsymbol{x}(t) = (x_1(t), \ldots, x_p(t))^T$, one can write:

$$\boldsymbol{x}(t) = \mathcal{A}(\boldsymbol{s}(t)), \tag{9}$$

where $\mathcal{A}(.)$ denotes the unknown mixture, a function from $\mathbb{R}^n$ to $\mathbb{R}^p$. If the number of observations $p$ is greater than or equal to the number of sources, $n$, the main idea for separating the sources is to estimate a transform $\mathcal{B}(.)$ which inverses the mixture $\mathcal{A}(.)$, and provides estimates of the unknown sources. Of course, without other assumptions, this problem cannot be solved. Basically, it is necessary to have priors about

– the nature of the mixtures: it is very important to choose a separating transform $\mathcal{B}(.)$ suited to the mixture transform $\mathcal{A}(.)$,

– the sources: sources properties - even weak - are necessary for the estimation of $\mathcal{B}(.)$.

In the purpose of hyperspectral data and according to the spatial and spectral approximation models, a first prior implicitly assumed is the linearity of the mixing which can be accepted in the case of geographical mixture. However, as introduced before, this model is not valid in the case of intimate mixtures. A second prior information is the positivity of the reflectance spectra of the chemicals. Thus, the positivity is a hard constraint that should be satisfied to get meaningful solutions. In that respect, two approaches can be applied: source separation by independent component analysis and source separation with positivity constraints.

## 4. Hyperspectral Data Analysis by ICA

If the assumptions on sources are restricted to their statistical independence, the above problem is referred to as blind source separation (BSS), and the method based on source independence property has been called independent component analysis (ICA) [1]. In the simplest case, the model is assumed to be linear and memoryless, *i.e.* $\mathcal{A}$ reduces to a mixing matrix $\boldsymbol{A}$ with scalar entries. This problem has been intensively studied in the last two decades, and many methods and algorithms are available, based on 4th-order statistics, entropic criteria, characteristic functions, etc. For this reason, the ICA principles as well as the ICA algorithms will not be developed in this paper. For more details, we recommend the reader to refer to [1, 11–13].

In the framework of hyperspectral data, although there is some evidence for a mixture model, one considers a sparse representation of the data using a sparse basis, with special properties. ICA provides such a model where the special property is mutual independence of the estimated sources. However, a scale indeterminacy can not be avoided when using ICA. Without imposing positivity constraints on the independent components (ICs) or on the mixing matrix results at least in a sign ambiguity. In this Section, we consider spatial mixture model (7) and we use the well known ICA algorithm, JADE, based on the joint approximate diagonalization of cumulant matrices [14].

### 4.1. *Spatial ICA*

In this experiment, we use two sets of data. The first one is the original data set, while the second one is a preprocessed data set [5] , obtained from the original data set by canceling the geometrical effect, the atmospheric effect and a few known defects of the sensors.

#### 4.1.1. *Source Number Determination*

The number, $N_c$, of independent component (IC) is related to the number of sources present in the mixtures. Of course, if $N_c$ increases, the accuracy of the approximation (7) increases too. A first step is then to choose this number. This is done using principal component analysis (PCA): with 7 principal components, 98.58% of the variance of the initial image is preserved. Then, we run JADE and obtain the estimated ICs. In figure 2, we show the 7 ICs obtained with the original data set.

A second step consists in evaluating the relevance of each component $IC_k$ in the approximation. This is done by measuring the relative quadratic loss:

$$\epsilon_k = -10 \log_{10} \left( \frac{P_{\hat{I}_{N_c}} - P_{\hat{I}_{N_c|N_k}}}{P_{\hat{I}_{N_c}}} \right), \tag{10}$$

obtained when replacing the $N_c$-order approximation $\hat{I}_{N_c}$ by with the $(N_c - 1)$-order, denoted $\hat{I}_{N_c|N_k}$ obtained by canceling the $IC_k$, and where the energies of the approximated images are computed as:

$$P_{\hat{I}_{N_c}} = \sum_{m=1}^{N_f} \sum_{n=1}^{N_z} \left( \sum_{p=1}^{N_c} b_{(\lambda_m,p)} II_p(n) \right)^2 \tag{11}$$

$$P_{\hat{I}_{N_c|N_k}} = \sum_{m=1}^{N_f} \sum_{n=1}^{N_z} \left( \sum_{p=1,p\neq k}^{N_c} b_{(\lambda_m,p)} II_p(n) \right)^2 . \tag{12}$$

These values, computed for the ICs estimated with the two sets of data, are given in table 1.

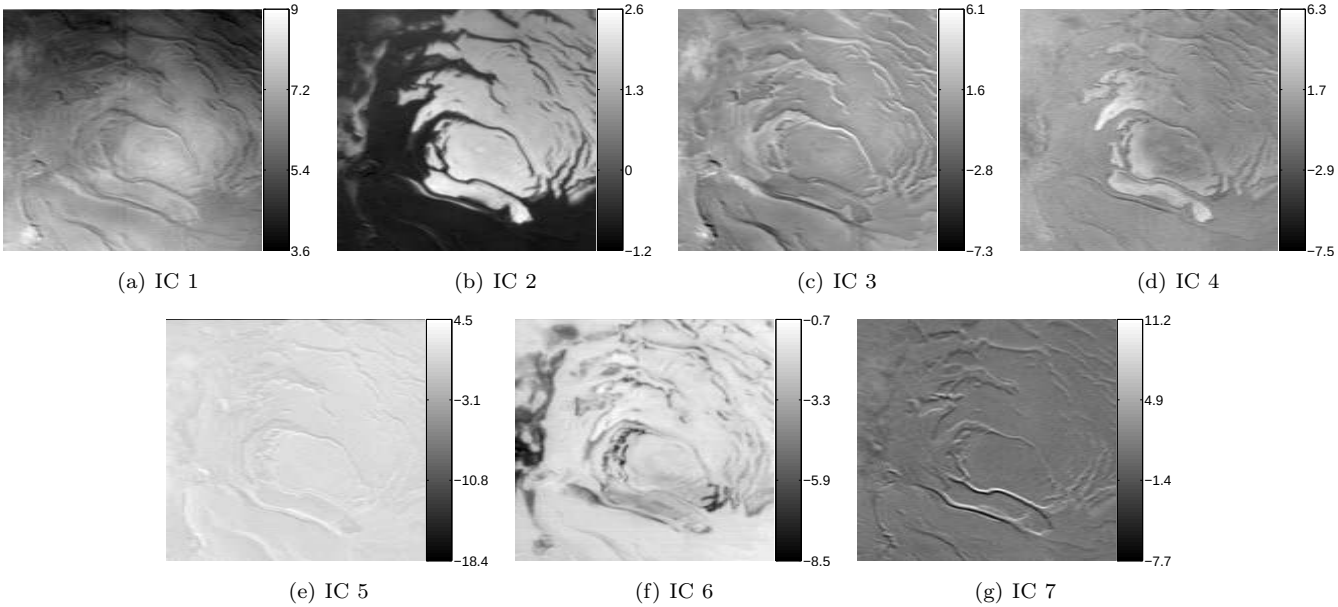
Finally, one has to wonder if the ICs are relevant and especially if they are robust with respect to the ICA algorithm, to similar images and to the number of ICs. In this purpose, we did three sets of experiments:

(i) we both consider two sets of data, data without preprocessing (RDS, for raw data set) and data set with preprocessing (PDS, for preprocessed data set) which remove a few physical artifacts: luminance gradient effect, sensor shift and atmospheric effect. We checked that a high quality reconstruction is achieved with 4 ICs for PDS and with 7 ICs for RDS. If more than 4 (or 7) ICs are used, one always get the same main (4 or 7) ICs, the others have very small contributions on the image reconstruction and cannot be interpreted. For this reason, we chose 4 ICs for PDS and 7 ICs with RDS.

(ii) in both data sets, images are resized by line subsampling (one line over two is kept). Then, this provides two different but very similar images that we call odd image and even image. We can check that ICA (with different IC numbers) of matched odd and even image leads to almost rigorously equal results, which again shows the reliability of the extracted ICs.

(iii) we run three algorithms: FastICA [12] with various non-linearities in the symmetric or deflation versions and JADE [14]. In all the experiments, one obtained very similar results (ICs and reconstruction performance) with JADE and symmetric FastICA, while performance of deflation FastICA was worse. JADE is preferred since it has a very weak computational

[5] the preprocessing is done by astrophysicists




(a) IC 1 (b) IC 2 (c) IC 3 (d) IC 4

(e) IC 5 (f) IC 6 (g) IC 7

Figure 2. Independent components of the RDS hyperspectral image of south polar cap using JADE with 7 ICs.

load, even with a large number of ICs, and required any parameter, except a stopping criterion.

| | | RDS data | PDS data |
|---|---|---|---|
| $k$ Identification | Figure | $\epsilon_k$ [dB] | $\epsilon_k$ [dB] |
| 1 Solar angle effect | 2(a) | 32.6 | 1.3 |
| 2 $CO_2$ ice | 2(b) | 16.3 | 10.7 |
| 3 Atmospheric effect | 2(c) | 12.2 | 0.88 |
| 4 Intimate mixture | 2(d) | 6.8 | 6.6 |
| 5 Corrupted line | 2(e) | 6.2 | - |
| 6 $H_2O$ ice | 2(f) | 7.1 | 5.9 |
| 7 Channel shift | 2(g) | 2.0 | 0.1 |

Table 1
Independent Components estimated with JADE. First column indicates the number of the IC. IC interpretation (see text for details) is given in column two. The third column refers to the Figure number (from a to g). The fourth and fifth column are the loss in $dB$ ($\epsilon_k$) obtained if $IC_k$ is not used in the approximation, for initial data or for preprocessed data, respectively. Note that $IC_2$ is in fact inverted, due to the scale indeterminacy with a negative value.

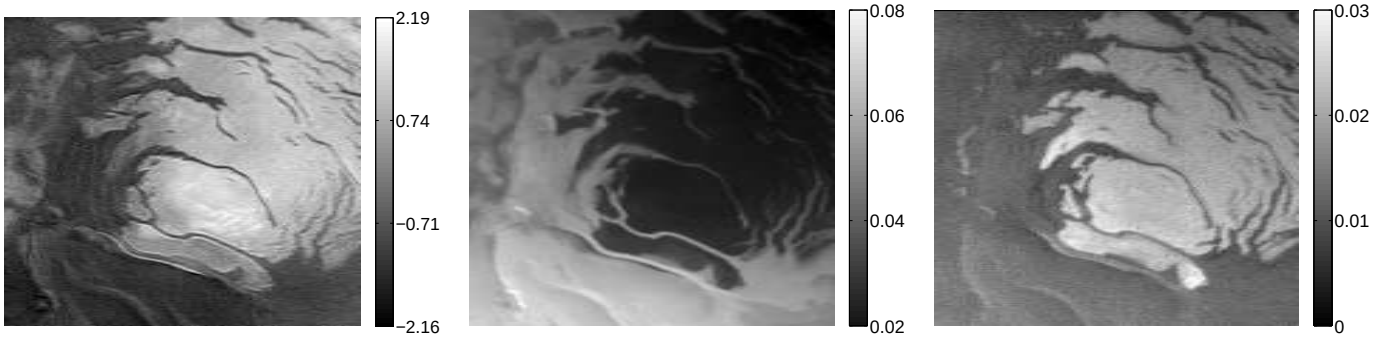
4.1.2. *IC Interpretation*

In this spatial approximation, each $IC_k$ can be viewed as an image, while the column $k$ of the mixing matrix is the spectrum related to $IC_k$. Comparing the IC spectra to the reference spectra, the components $IC_2$ and $IC_6$ can be easily identified (table 1) to respectively $CO_2$ ice and $H_2O$ ice. Conversely, the spectrum associated to $IC_4$ has typical bands of both dust, $CO_2$ and $H_2O$ ices. We could interpret this IC as a nonlinear intimate mixture effect or a non independent distribution of those components.

The other four components ($IC_1$, $IC_3$, $IC_5$ and $IC_7$) are difficult to interpret with spectral informations. We remark that the energies of these ICs are very small for the preprocessed data (Table 1, column 5). Consequently, one can assume that the importance of these ICs has been strongly reduced by the preprocessing, and they are related to phenomena cancelled by the preprocessing. In fact, the first IC (figure 2(a)) has a luminance gradient which is characteristic of the solar angle effect which should be the $E(x, y, \lambda)$ term in equation (3). The $IC_7$ (figure 2(g)) looks like a high-pass filter mainly on the $y$ direction of the image. This along track direction maximizes the instrument shift between the two near IR detectors. This effect is independent of the spectra model and thus it is detected as a separate IC and can be used to assess the quality of the preprocessing. $IC_3$ could be associated to the transmission in the atmosphere effect because it is similar to a map of topography. However, if this interpretation is correct, our model hypothesis about the equivalent path in the atmosphere for all the pixels is wrong. At first glance, $IC_5$ (figure 2(e)) was not recognized. But at a closer look, the first line in the image has a very low response, and corresponds to a corrupted line in the dataset. This IC is then due to a sensor failure, known by the astrophysicists.

4.1.3. *Classification*

From $IC_2$ and $IC_6$ interpretation, we deduce classification masks of $CO_2$ and $H_2O$ ices with an easy criteria : if the IC has a positive value then $CO_2$ ice - respectively $H_2O$ ice - is detected. The classification results compared with the reference classification is seen in table 2. *False alarm* and *missing* indicate the differences between the classification based on the ICs with respect to that based on wavanglet classification. However, keep in mind that the reference images, although pertinent, are not the ground truth and thus the false classifications in this case are not necessarily false.




(a) IC corresponding to $CO_2$ ice (b) 1.9833 $\mu m$ (c) 2.9833 $\mu m$

Figure 3. The mixture of $CO_2$ ice and dust.

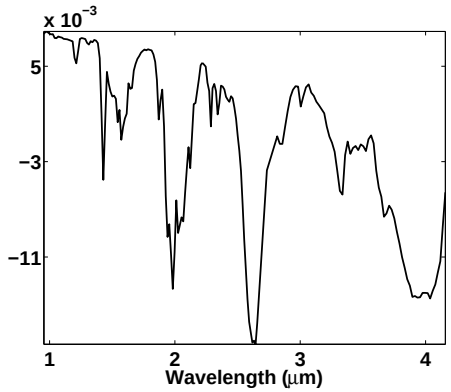



Figure 4. Estimated mixing coefficients associated to $CO_2$ ice

An interesting result is that the classification of the original dataset is slighly better than the classification of the preprocessed dataset. This is in accordance to the SNR gained in table 1, since the atmospheric removal seems to take some information from the $CO_2$ and $H_2O$ ice components.

### 4.2. *Dependence of Spatial IC*

Clearly, looking at the reference spectra and reference images, it appears that the fundamental assumption of independence is not at all satisfied neither in the spectral nor in the spatial dimensions. The spatial independence can be tested by looking at the covariance between the reference classification masks (figure 1).

$$R_s\{I_{dust}, I_{CO_2}, I_{H_2O}\} = \begin{pmatrix} 1 & -0.61 & -0.24 \\ -0.61 & 1 & -0.25 \\ -0.24 & -0.25 & 1 \end{pmatrix}. \quad (13)$$

When using spatial ICA, components of $CO_2$ and $H_2O$ ices are retrieved but dust does not appear as a separate component as seen on figure 2. The negative correlation seen in (13) is not surprising when looking at figures 1 (d, e and f). Spatially, dust and $CO_2$ ice are strongly complementary, and even more than the reference images indicate. As a result, dust is not retrieved when using spatial ICA, but is frequently recovered from the ICs as the negative of $CO_2$ ice.

The IC which corresponds to $CO_2$ ice can be seen on figure 3(a), and the corresponding mixture. As can be seen in the entries [6] of $\boldsymbol{A}$ (Fig. 4), this IC is both used as a positive and negative component in the mixture. From the graph in figure 3, it can be seen that when the mixing coefficients take a positive value, for example at 2.98 $\mu m$, the dominating element for this wavelength is $CO_2$ ice as seen on figure 3(c). When the mixing coefficients take a negative value as at 1.98 $\mu m$ the dominating element is dust as seen of figure 3(b). Taking a closer look at the mixing matrix reveals that the IC seen above is by far the strongest component in the mixing matrix for the layers where dust and $CO_2$ ice are most visible. It is thus evident that this IC presents in fact both $CO_2$ ice and dust.

| | Correct | Missing | False alarm |
|---|---|---|---|
| Original Data | | | |
| $CO_2$ | 97.1 % | 1.1 % | 1.8 % |
| $H_2O$ | 91.3 % | 8.7 % | 1.0 % |
| Preprocessed Data | | | |
| $CO_2$ | 95.9 % | 2.0 % | 2.1 % |
| $H_2O$ | 90.0 % | 8.1 % | 1.9 % |

Table 2
Classification of $CO_2$ and $H_2O$ ices compared with the reference classification

### 4.3. *Discussion*

At first glance, spatial ICA provides ICs which can be interpreted as artifacts or endmembers. Concerning artifact ICs, the results are very interesting, and suggest that the data preprocessing could be avoided and done using ICA results. In fact, the classification performance seems [7] even

[6] we recal that the column $k$ of $\boldsymbol{A}$ can be interpreted as the spectrum related to $IC_k$

[7] remember that we do not have ground truth, but only a reference classification, designed by astrophysicists. Consequently, "bet-

a little bit better on original data than in preprocessed data. Conversely, only two endmembers ($CO_2$ and $H_2O$ ices) are associated to $IC_2$ and $IC_6$: the third one, dust, only appears as the negative on $CO_2$, which results in a very poor classification of dust. Moreover, the IC statistical independence, the main hypothesis on which ICA is based, is not satisfied. Especially, in the endmember classification, it appears that dust and $CO_2$ ice are strongly correlated. Thus, the reliability of ICA is not sure, and the relevance of the extracted ICs is poor. Consequently, other methods, based on priors satisfied by the data, must be used.

## 5. Source Separation with Positivity Constraint

### 5.1. *Problem Statement*

The main constraint in hyperspectral mixture data decomposition is the positivity of both the mixing coefficients and the source signals. Unfortunately, this constraint alone does not lead to a unique solution unless under some particular conditions [15,16]. Thus, in general cases additional assumptions are required to select a particular solution among the admissible ones. The estimation can then be performed using either a constrained or a penalized least square estimation. In the constrained approach, there are mainly two methods: *alternating least squares* (ALS) [17] and *non-negative matrix factorization* (NMF) [18]. While ALS method performs a decomposition of the mixture by minimizing alternatively the least square criterion with respect to source signals and mixing matrix, under the non-negativity constraint [19, 20], the NMF method minimizes this criterion with gradient descent algorithm over this objective function by updating iteratively mixing coefficients and source signals using a multiplicative learning rule that ensures the estimates to remain non-negative. The key point is that the solution using only non-negativity is not unique, therefore the results provided by ALS and NMF methods depend on the initialization. In a penalized least square method, a regularization criterion is added to the weighted mean squares criterion in order to select a particular solution fulfilling the additional assumptions. This is the basis of methods such as *positive matrix factorization* (PMF) [21] and *non-negative sparse coding* (NNSC) [22,23]. Unlike the constrained least squares methods, penalized approaches lead to an unconstrained optimization problem and ensures the uniqueness of the solution for a fixed set of the regularization parameters. However, to also estimate the regularization parameters it is more suitable to address the separation problem in a Bayesian framework.

ter performance" only means "better correlation" with the reference classification.

### 5.2. *Bayesian Positive Source Separation*

In a Bayesian source separation approach, one can ideally incorporate every prior knowledge as long as this prior can be stated in probabilistic terms. The approach is founded on the likelihood $p\left(\boldsymbol{x}|\boldsymbol{A},\boldsymbol{s}\right)$ and prior distributions of the source signals and mixing coefficients. Applying Bayes' theorem leads to:

$$p(\boldsymbol{A},\boldsymbol{s}|\boldsymbol{x}) \propto p(\boldsymbol{x}|\boldsymbol{A},\boldsymbol{s}) \times p(\boldsymbol{A}) \times p(\boldsymbol{s}). \tag{14}$$

From this posterior law, both the mixing matrix $\boldsymbol{A}$ and sources $\boldsymbol{s}$ can be estimated using various Bayesian estimators. A complete discussion on Bayesian approach to source separation can be found in [24–26]. However, its application to the case of positive sources and mixing has only received a few attention [27–29]. In this purpose, a recent contribution consists of the method termed by *Bayesian positive source separation* (BPSS) [2, 30], which allows to jointly estimate source signals, mixing coefficients and regularization parameters in an unsupervised framework.

#### 5.2.1. *Bayesian Model*

The probabilistic model for source separation in the case of a spectral mixture model is based on the hypothesis of independent and identically distributed Gaussian errors and Gamma distributions priors on source signals and mixing coefficient distributions. The spectral mixing model being expressed by:

$$I_n(\lambda_k) = \sum_{p=1}^{N_c} a_{(p,n)}\psi_p(\lambda_k) + E_n(\lambda_k) \quad \forall n = 1, ..., N_z, \tag{15}$$

where $E_n(\lambda_k)$ is a noise term which models errors due to the simplified model (2), the restricted number of components, $N_c$ and measurement noise, the Bayesian model is then summarized as:

$$\left(E_n(\lambda_k)|\sigma_n^2\right) \sim \mathcal{N}(E_n(\lambda_k); 0, \sigma_n^2), \tag{16}$$
$$\left(\psi_p(\lambda_k)|\alpha_p, \beta_p\right) \sim \mathcal{G}(\psi_p(\lambda_k); \alpha_p, \beta_p), \tag{17}$$
$$\left(a_{(p,n)}|\gamma_p, \delta_p\right) \sim \mathcal{G}(a_{(p,n)}; \gamma_p, \delta_p), \tag{18}$$

where $\mathcal{N}(z; 0, \sigma^2)$ represents a Gaussian distribution of the random variable $z$ with zero mean and variance $\sigma^2$ and $\mathcal{G}(z; \alpha, \beta)$ stands for a Gamma distribution of the random variable $z$ with parameters $(\alpha, \beta)$. The Gamma law takes into account explicitly the positivity constraint since it leads to a posterior distribution with positive support (the probability distribution function is zero for negative values of the sources and the mixing coefficients). In addition, its two parameters give a flexibility to adapt its shape to that of spectral source signals. According to this probabilistic model and Bayes' theorem, with the hypothesis of statistical independence of the source signals and the mixing coefficients, the joint *a posteriori* distribution can be calculated:

$$p\left(\boldsymbol{\Psi},\boldsymbol{A}|\boldsymbol{I},\boldsymbol{\theta}\right)\propto \prod_{k=1}^{N_f}\prod_{n=1}^{N_z}\mathcal{N}\left(I_n(\lambda_k);\sum_{p=1}^{N_c}a_{(n,p)}\,\psi_p(\lambda_k),\sigma_n^2\right) \times\prod_{p=1}^{N_c}\prod_{k=1}^{N_f}\mathcal{G}((\psi_p(\lambda_k);\alpha_p,\beta_p)\times\prod_{n=1}^{N_z}\prod_{p=1}^{N_c}\mathcal{G}(a_{(p,n)};\gamma_p,\delta_p). \tag{19}$$

The criterion corresponding to the logarithm of this posterior distribution can be decomposed into three parts

$$\begin{aligned}J(\boldsymbol{\Psi},\boldsymbol{A}|\boldsymbol{I},\boldsymbol{\theta}) &= -\log p\left(\boldsymbol{\Psi},\boldsymbol{A}|\boldsymbol{I},\boldsymbol{\theta}\right)\\ &= Q(\boldsymbol{\Psi},\boldsymbol{A}|\boldsymbol{I},\boldsymbol{\theta})+R_S(\boldsymbol{\Psi}|\boldsymbol{\theta})+R_A(\boldsymbol{A}|\boldsymbol{\theta}),\end{aligned} \tag{20}$$

where $Q, R_S$ and $R_A$ are given by

$$Q=\sum_{k=1}^{N_f}\sum_{n=1}^{N_z}\frac{1}{2\,\sigma_n^2}\Big(I_n(\lambda_k)-\sum_{p=1}^{N_c}a_{(p,n)}\psi_p(\lambda_k)\Big)^2, \tag{21}$$

$$R_S=\sum_{k=1}^{N_f}\sum_{p=1}^{N_c}\Big((1-\alpha_j)\log\psi_p(\lambda_k)+\beta_j\psi_p(\lambda_k)\Big), \tag{22}$$

$$R_A=\sum_{n=1}^{N_z}\sum_{p=1}^{N_c}\Big((1-\gamma_j)\log a_{(p,n)}+\delta_j a_{(p,n)}\Big). \tag{23}$$

The first part $Q$ of the objective function is the mean square criterion, while the last two parts $R_A, R_S$ are regularization terms. This criterion may be connected with those of previously proposed methods:

- The case where $\{(1-\alpha_p)=\alpha\}_{p=1}^{N_c}$, $\{(1-\gamma_p)=\gamma\}_{p=1}^{N_c}$, $\{\beta_j=\beta\}_{p=1}^{N_c}$ and $\{\delta_j=\delta\}_{p=1}^{N_c}$, allows to get a criterion similar to that in PMF [21]. Therefore, it may be interpreted as a particular case of the proposed criterion, in which the same prior parameters are assigned to all the source signals and the same prior parameters are assigned to all the mixing coefficients.
- The case where $\{\alpha_p=1\}_{p=1}^{N_c}$, $\{\beta_p=\beta\}_{p=1}^{N_c}$, $\{\gamma_p=1\}_{p=1}^{N_c}$, $\{\delta_p=0\}_{p=1}^{N_c}$ corresponds to assigning an exponential distribution prior to the source signals and a uniform positive prior to the mixing coefficients, leading to the non-negative sparse coding (NNSC) criterion [22].

Thus, the proposed Bayesian model with Gamma prior has the advantage of using a more flexible prior model and offers a well stated theoretical framework for estimating the hyperparameters $\left\{\sigma_n^2\right\}_{n=1}^{N_z},\{\alpha_p,\beta_p,\gamma_p,\delta_p\}_{p=1}^{N_c}$ which are also included in the Bayesian model with appropriate prior distributions [2, 31]. Thus, by using Bayes' theorem and assigning appropriate *a priori* distributions to these hyperparameters, the whole *a posteriori* distribution, including the hyperparameters, is expressed as

$$p(\boldsymbol{\Psi},\boldsymbol{A},\boldsymbol{\theta}|\boldsymbol{I})\propto p(\boldsymbol{\Psi},\boldsymbol{A}|\boldsymbol{I},\boldsymbol{\theta})\times p(\boldsymbol{\theta}). \tag{24}$$

The joint estimation of the pure spectra, of the mixing coefficients and of the hyperparameters is then performed from this *a posteriori* distribution.

5.2.2. *Estimation Algorithm*

The estimation of the source signals and of the mixing coefficients is performed using marginal posterior mean estimator and Markov Chain Monte Carlo (MCMC) methods. These stochastic methods are extensively documented in the statistical literature (see the books [32, 33] and the references therein). In short, these iterative methods (Gibbs sampling [34] and Metropolis-Hastings [35, 36]) generate random numbers from the posterior distribution and use these simulated realizations to approximate expectations with respect to the posterior distribution by empirical averages.

To simulate $p\left(\boldsymbol{\Psi},\boldsymbol{A},\boldsymbol{\theta}|\boldsymbol{I}\right)$, at each new iteration $r$ of the algorithm, the main steps consist in simulating

1. the source signals $\boldsymbol{\Psi}^{(r+1)}$ from
$$p\left(\boldsymbol{\Psi}|\boldsymbol{I},\boldsymbol{A}^{(r)},\boldsymbol{\theta}^{(r)}\right)\propto p\left(\boldsymbol{I}|\boldsymbol{\Psi},\boldsymbol{A}^{(r)},\boldsymbol{\theta}^{(r)}\right)\times p\left(\boldsymbol{\Psi}|\boldsymbol{\theta}^{(r)}\right), \tag{25}$$
2. the mixing coefficients $\boldsymbol{A}^{(r+1)}$ from
$$p\left(\boldsymbol{A}|\boldsymbol{I},\boldsymbol{\Psi}^{(r+1)},\boldsymbol{\theta}^{(r)}\right)\propto p\left(\boldsymbol{I}|\boldsymbol{\Psi}^{(r+1)},\boldsymbol{A},\boldsymbol{\theta}^{(r)}\right)\times p\left(\boldsymbol{A}|\boldsymbol{\theta}^{(r)}\right), \tag{26}$$
3. the hyperparameters $\boldsymbol{\theta}^{(r+1)}$ from
$$p\left(\boldsymbol{\theta}|\boldsymbol{I},\boldsymbol{\Psi}^{(r+1)},\boldsymbol{A}^{(r+1)}\right)\propto p\left(\boldsymbol{I}|\boldsymbol{\Psi}^{(r+1)},\boldsymbol{A}^{(r+1)},\boldsymbol{\theta}\right)\times p\left(\boldsymbol{\theta}_1\right). \tag{27}$$

After a random initialization, the MCMC sampler is run to have $M$ realizations of $\boldsymbol{\Psi}$, $\boldsymbol{A}$ and $\boldsymbol{\theta}$ from the posterior distribution (24). The first $L$ realizations corresponding to the burn-in of the Markov chain being discarded, the Monte Carlo approximation for marginal posterior mean estimation is achieved by

$$\hat{\boldsymbol{X}}\approx\frac{1}{M-L+1}\sum_{r=L+1}^{M}\boldsymbol{X}^{(r)}, \tag{28}$$

where $\boldsymbol{X}\in\{\boldsymbol{\Psi},\boldsymbol{A},\boldsymbol{\theta}\}$. All the stochastic simulation steps including the expressions of the conditional posterior distributions and their simulation techniques are detailed in [2], where this method is termed *Bayesian Positive Source Separation* (BPSS).

## 6. Hyperspectral data analysis by BPSS

A practical constraint in the case of hyperspectral data is the high resolution of the acquisition instrument which provides a data cube of large size. In such case, the computation load of the BPSS approach becomes very important and even the computation becomes impossible with a standard computer (need of a huge memory space, high computation time). To illustrate this problem table 3 shows the computation time per iteration, with a standard PC (3 GHz and 512 MO RAM) , for different image sizes with

$N_f = 200$ spectral bands and $N_c = 4$ components. For example, it can be seen that for a data cube where the image size is $128 \times 128$ pixels, the computation time for $10^4$ iterations is about 4 days and 14 hours. In that respect, a reduction of the dimension is necessary before applying the BPSS approach to the hyperspectral data provided by the OMEGA instrument. Of course, to reduce the computation burden one can also use deterministic optimization methods to jointly estimate the unknown parameters and hyperparameters (see [37] and [38] for two examples in signal deconvolution/restoration with non-negativity and sparsity constraints).

| Image size (pixels) | $(16 \times 16)$ | $(32 \times 32)$ | $(64 \times 64)$ | $(128 \times 128)$ |
|---|---|---|---|---|
| Time per iteration (s) | 0.58 | 2.39 | 9.89 | 39.85 |

Table 3
Computation time of BPSS for different image sizes

### 6.1. *Combination of Spatial ICA with Spectral BPSS*

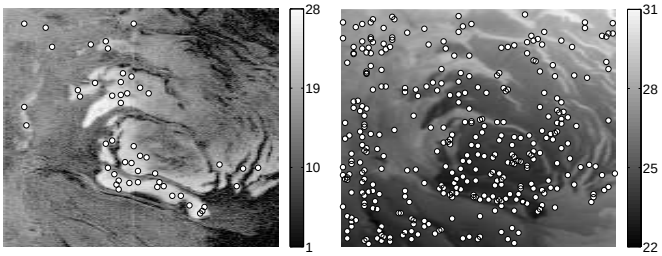
A more realistic solution to handle the computation time limitation is to process a smaller data set which is representative of the whole hyperspectral images. That is a selection of pixels corresponding to areas where all the chemicals present in the whole image are present. Therefore, their pure spectra can be estimated by BPSS with a reduced computation load. Our proposal is to exploit the spatial independent component analysis results (Section 4) for selecting a few number of pixels in independent areas of the spatial coordinates, *i.e.* areas classified as $H_2O$ ice, $CO_2$ ice or dust. We define the relevant pixels associated to each source as those where the contribution of this source is important. To measure this contribution at each pixel $n$, we use the spatial SNR loss which is defined as the variation of the spatial SNR when one particular source is removed from the mixture. Thus, removing the component $II_j$, one gets:

$$SNR_j(n) = SNR(n) - 10\log_{10}\left(\frac{\sum_{k=1}^{N_f} I_{\lambda_k}(n)^2}{\sum_{k=1}^{N_f}\left(I_{\lambda_k}(n) - \sum_{p=1,p\neq j}^{N_c} a_{(\lambda_k,p)} II_p(n)\right)^2}\right), \quad (29)$$

where $SNR(n)$ is the spatial SNR of the approximation with $N_c$ components, defined by:

$$SNR(n) = 10\log_{10}\left(\frac{\sum_{k=1}^{N_f} I_{\lambda_k}(n)^2}{\sum_{k=1}^{N_f}\left(I_{\lambda_k}(n) - \sum_{p=1}^{N_c} a_{(\lambda_k,p)} II_p(n)\right)^2}\right). \quad (30)$$

For each chemical, we define the most relevant pixels as the first 15% with the highest spatial SNR loss and then we select randomly a fixed number of pixels among this set. In the sequel, we select 50 pixels from each independent component image.




(a) Spatial SNR loss and selected pixels from IC 4

(b) Selected pixels from the whole image.

Figure 5. Illustration of the pixel selection step.

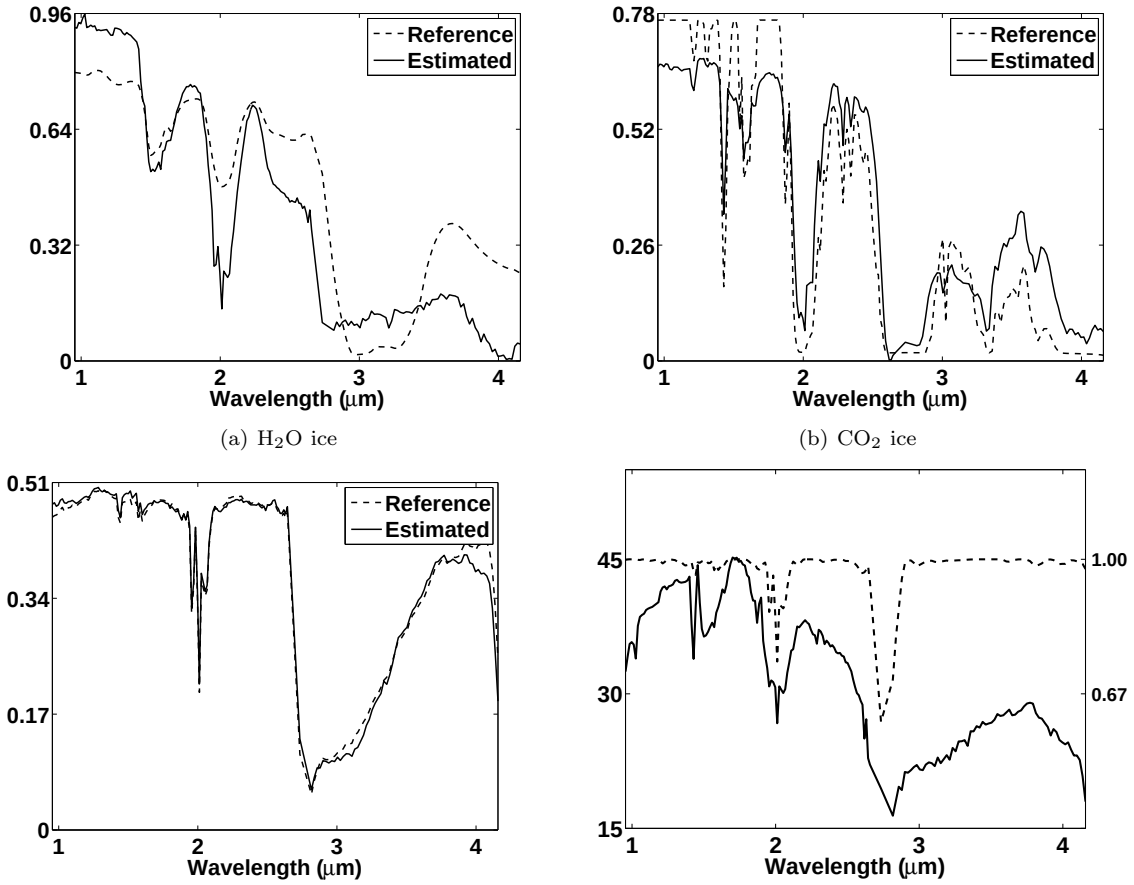
### 6.2. *Experimental Results*

The proposed approach is applied to the RDS hyperspectral images presented in section 2. Figure 5 illustrates the spatial SNR loss and the selected pixels from the fourth independent component image and the selected pixels after processing of all the independent components.The results of the separation using BPSS with the mixture spectra provided by the selected pixels are post-processed to correct scale and ordering ambiguities and deduce abundance fractions. The identification of the spectra is straightforward from the correlation with the reference spectra as seen from equation (31) where the matrix entries $r_{ij}$ of the matrix $\mathbf{R}$ are the correlation coefficients between the reference spectra $\psi_i \in \{\psi_{H_2O}, \psi_{CO_2}, \psi_{dust}\}$ and the estimated pure spectra $\hat{\psi}_j \in \{\hat{\psi}_1, \hat{\psi}_2, \hat{\psi}_3]$.

$$\mathbf{R} = \begin{pmatrix} \mathbf{0.91} & 0.79 & 0.87 \\ 0.72 & 0.65 & \mathbf{0.99} \\ 0.89 & \mathbf{0.96} & 0.55 \end{pmatrix} \quad (31)$$

One can note that the correlation coefficient is high (0.99) for the dust endmember and lower for $CO_2$ ice and $H_2O$ ice. After scaling and permutation of the identified spectra, the reference spectra are plotted together on figure 6. It can be noted the similarity between the estimated spectra and the references ones. The similarity is lower for the both $CO_2$ ice and $H_2O$ ice in the spectral region near 2 $\mu m$ because of the presence of a deep atmospherical band. On the contrary, the dust source is in relative better agreement with the reference spectra (see equation (3)) because both contain the atmospherical transmission. These similarities lead to an easy identification of the chemical component associated to each estimated source. Furthermore, the spectral information in the estimated sources is rich enough to provide knowledge about the physical properties of the surface constituents (such as : grain size, rugosity, density, etc.).

On the other hand, one can compute the abundance fractions in each pixel of the hyperspectral images, which are shown in figures 7 as soon as the simplest physical model is true. It can be seen that the abundance fractions not only agree with the results of the wavanglet classification shown in figure 1, but also give information about the amount of the chemicals in the south polar cap of Mars - which was




(a) $H_2O$ ice

(b) $CO_2$ ice

(c) dust

(d) Spectral SNR

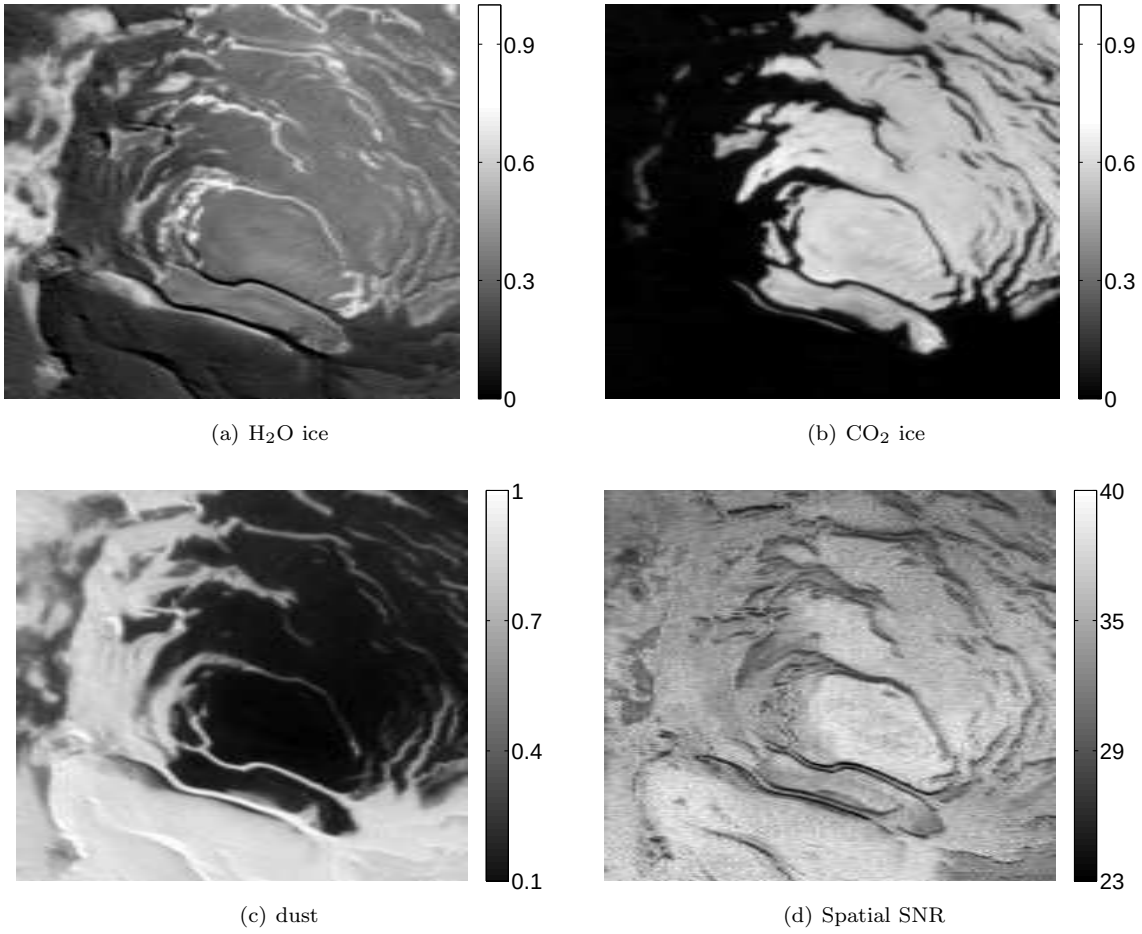
Figure 6. Estimated and reference spectra (a,b,c) and (d) spectral reconstruction error in dB (continuous line, scale on the left), atmospheric transmission (dotted line, scale on the right)

not available with the wavanglet method since the classification only provides binary masks. This unsupervised detection and classification based on source separation is an original result which has not been obtained before.

In order to estimate robustness and quality of these result, we analyze both spatial and spectral reconstruction errors. Spectral SNR is relatively low in all atmospherical absorption bands (2 and 2.9 $\mu m$) because our modeling is not valid for the atmosphere (figure 6-d). As previously exposed in figure 2 and in table 1, spatial $IC_3$ also show this effect. The atmosphere spectra is mainly controlled by a factor depending on the altitude. Altitude can change in our scene from 1000 m to more than 2000 m. Second the spectral bands around 2 and 3 $\mu m$ are present everywhere and implies strong absoptions (figure 6 for reference spectra of $CO_2$ ice and $H_2O$ ice and dust). At the larger wavelength, near 4 microns, the thermal emission of the planet is superposed to the reflectance spectra for temperature higher than 140 K which can be the case for some unfrosted area. Spatial SNR estimate the reconstruction in the spatial domain (figure 7-d). The general pattern is similar to spatial $IC_7$ (figure 2) provided for the same RDS and the interpretation of the channel shift remain the best one (table 1). The simple correction applied to avoid the shift of the two near infrared detectors is the number of pixel in the y direction that best fit the overlapping spectral domain. However, a more accurate correction (leading to a real-valued shift) can be done by resampling the data, but this cannot be done without loosing information. We choose here to use the integer shift correction. Therefore, all relative disagreement in both spatial and spectral domain can be interpreted as non-linearities justified by specific physical phenomena that occur on planet Mars and in the OMEGA detector.

## 7. Conclusion and further works

In this paper we have proposed an application of source separation methods to the problem of hyperspectral data analysis. It appears that ICA can be usefull for providing efficient preprocessing of the observation data without information on the accurate geometry, measurement errors and sensor defects. However, classification based on ICA seems neither reliable nor accurate, especially since it is based on a wrong assumption: source spectra and locations of endmembers in the observed hyperspectral cubes are not independent. In a second step, enforcing positivity of spectra and mixing coefficients, we proposed a semi-blind




(a) $H_2O$ ice (b) $CO_2$ ice

(c) dust (d) Spatial SNR

Figure 7. Estimated abundance fractions of water ice, $CO_2$ ice and dust (a, b and c) and spatial reconstruction error (d)

approach allowing to jointly estimate the pure spectra of the chemicals composing the area under view and compute their abundances. In this second step, ICA classification, although approximate, is used for selecting a restricted set of pixels, representative of the constituents: this reduces the BPSS computational load without penalizing the performance. The spectral accuracy allows the user to easily identify the chemical species, and the estimated sources are precise enough to provide knowledge about their physical properties. Finally, the linked abundance are in agreement with the reference wavanglet masks.

Current works include two tasks. First, the study concerning the result confidence is very important for astrophysicists. Secondly, since the approximation done by ICA as well as BPSS is based on a linear model of $N_c$ components, the spatial reconstruction error $(\hat{I}_{\lambda_k}(n) - I_{\lambda_k}(n))$ should inform on the local quality of the approximation. A too large error could be associated either to a wrong local model, *i.e.* due to intimate (nonlinear) mixture, or to the local presence of another endmember, which does not belong to the $N_c$ extracted components. On the other hand, future works should be done in the case of OMEGA hyperspectral images with more ambiguous spectral endmembers. Especially with minerals because associated spectral signatures are thinner than the ice one. Furthermore, the new generation of hyperspectral imaging instruments will soon incorporate a new dimension: the observation angle. An instrument, like CRISM on the NASA'mission Mars Reconnaissance Orbiter, will observe the same field of view according to different angles. Semi-blind approaches, inspired of these presented in this paper, seem straightforward candidates for analyzing this type of dataset, which is a real challenge for the next years.

*Acknowledgments.* This work was supported in part by the Research Fund of the University of Iceland and in part by the Jules Verne Program of the French and Icelandic governments (PAI EGIDE).

## References

[1] P. Comon, Independent component analysis, a new concept?, Signal Processing 36 (3) (1994) 287–314.

[2] S. Moussaoui, D. Brie, A. Mohammad-Djafari, C. Carteret, Separation of non-negative mixture of non-negative sources using a Bayesian approach and MCMC sampling, IEEE Trans. Signal Processing 54 (11) (2006) 4133–4145.

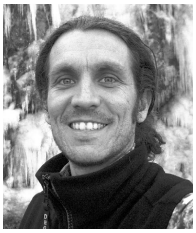
[3] J.-P. Bibring et *al.*, OMEGA: Observatoire pour la Minéralogie, l'Eau, les Glaces et l'Activité, ESA SP-1240, 2004, Ch. Mars Express: the Scientific Payload, pp. 37–49.
[4] S. Douté, B. Schmitt, Y. Langevin, J.-P. Bibring, F. Altieri, G. Bellucci, B. Gondet, F. Poulet, the MEX OMEGA team, South Pole of Mars: Nature and composition of the icy terrains from Mars Express OMEGA observations, Planetary and Space Sience 55 (2007) 113–133.
[5] F. Schmidt, S. Douté, B. Schmitt, Wavanglet : an efficient supervised classifier for hyperspectral images, accepted to IEEE Trans. on Geoscience and Remote Sensing, (2007).
[6] J.-P. Bibring et *al.*, Perennial water ice identified in the south polar cap of Mars, Nature 428 (2004) 627–630.
[7] N. Keshava, J. F. Mustard, Spectral unmixing, IEEE Signal Processing Magazine 19 (1) (2002) 14–57.
[8] D. Tanre, M. Herman, P. Y. Deschamps, A. de Leffe, Atmospheric modeling for space measurements of ground reflectances, including bidirectional properties, Applied Optics 18 (1979) 3587–3594.
[9] C. Jutten, J. Hérault, Blind separation of sources, Part I: An adaptive algorithm based on neuromimetic architecture, Signal Processing 24 (1991) 1–10.
[10] P. Comon, C. Jutten, J. Hérault, Blind separation of sources, Part II: Problems statement, Signal Processsing 24 (1991) 11–20.
[11] J.-F. Cardoso, Blind signal separation: statistical principles, Proc. of the IEEE 9 (10) (1998) 2009–2025.
[12] A. Hyvärinen, J. Karhunen, E. Oja, Independent component analysis, Adaptive and Learning Systems for Signal Processing, Communications, and Control, John Wiley, New York, 2001.
[13] A. Cichocki, , S.-I. Amari, Adaptive Blind Signal and Image Processing- Learning Algorithms and Applications, John Wiley, 2002.
[14] J.-F. Cardoso, A. Souloumiac, Blind beamforming for non Gaussian signals, IEE Proceedings-F 140 (6) (1993) 362–370.
[15] D. Donoho, V. Stodden, When does non-negative matrix factorization give a correct decomposition into parts?, in: Advances in Neural Information Processing Systems (NIPS'2003), Cambridge, United States, 2003.
[16] S. Moussaoui, D. Brie, J. Idier, Non-negative source separation: Range of admissible solutions and conditions for the uniqueness of the solution, in: proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'2005), Philadelphia, USA, 2005, pp. 289–292.
[17] R. Tauler, B. Kowalski, S. Fleming, Multivariate curve resolution applied to spectral data from multiple runs of an industrial process, Analytical Chemistry 65 (1993) 2040–2047.
[18] D. Lee, H. Seung, Learning the parts of objects by non–negative matrix factorization, Nature 401 (1999) 788–791.
[19] C. Lawson, R. Hanson, Solving Least-Squares Problems, Prentice-Hall, 1974.
[20] R. Bro, S. De Jong, A fast non-negativity constrained least squares algorithm, Journal of Chemometrics 11 (1997) 393–401.
[21] P. Paatero, U. Tapper, Positive matrix factorization: a nonnegative factor model with optimal utilization of error estimates of data values, Environmetrics 5 (1994) 111–126.
[22] P. O. Hoyer, Non-negative sparse coding, in: Proceedings of IEEE Workshop on Neural Networks for Signal Processing (NNSP'2002), 2002, pp. 557–565.
[23] S. A. Abdallah, M. D. Plumbley, Polyphonic transcription by non-negative sparse coding of power spectra, in: Proc. International Conference on Music Information Retrieval (ISMIR'2004), Barcelona, Spain, 2004, pp. 318–325.
[24] A. Mohammad-Djafari, A Bayesian approach to source separation, in: J. T. Rychert, G. J. Erickson, C. R. Smith (Eds.), MaxEnt, Vol. 567, AIP Conference Proceedings, Boise, Idao, 1999, pp. 221–244.
[25] D. Rowe, Multivariate Bayesian Statistics: Models for Source Separation and Signal Unmixing, CRC Press, Boca Raton, Florida, USA, 2003.
[26] H. Snoussi, J. Idier, Bayesian blind separation of generalized hyperbolic processes in noisy and underdeterminate mixtures processes in noisy and underdeterminate mixtures, IEEE Trans. Signal Processing 54 (9) (2006) 3257–3269.
[27] M. F. Ochs, R. S. Stoyanova, F. Arias-Mendoza, T. R. Brown, A new method for spectral decomposition using a bilinear Bayesian approach, Journal of Magnetic Resonance 137 (1999) 161–176.
[28] J. Miskin, D. MacKay, Ensemble learning for blind source separation, in: S. Roberts and R. Everson, editors, Independent Component Analysis: Principles and Practice, Cambridge University Press, 2001, pp. 209–233.
[29] S. Roberts, R. Choudrey, Data decomposition using independent component analysis with prior constraints, Pattern Recognition 36 (2003) 1813–1825.
[30] S. Moussaoui, D. Brie, C. Carteret, A. Mohammad Djafari, Application of Bayesian non-negative source separation to mixture analysis in spectroscopy, in: R. Fischer, R. Preuss, U. von Toussaint (Eds.), MaxEnt, Vol. 735, AIP Conference Proceedings, Garching, Germany, 2004, pp. 237–244.
[31] S. Moussaoui, C. Carteret, D. Brie, A. Mohammad-Djafari, Bayesian analysis of spectral mixture data using Markov chain Monte Carlo methods, Chemometrics and Intelligent Laboratory Systems 81 (2) (2006) 137–148.
[32] C. Robert, Monte Carlo Statistical Methods, Springer-Verlag, 1999.
[33] M.-H. Chen, Q.-M. Shao, J. G. Ibrahim, Monte Carlo Methods in Bayesian Computation, Springer-Verlag, New York, 2000.
[34] S. Geman, D. Geman, Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, IEEE Trans. Pattern. Anal. Mach. Intell. 6 (6) (1984) 721–741.
[35] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Taller, E. Taller, Equation of state calculation by fast computing machines, J. Chem. Phy. 21 (1953) 1087–1092.
[36] W. K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, Biometrika 57 (1) (1970) 97–109.
[37] I.-T. Hsiao, A. Rangarajan, G. Gindi, Bayesian image reconstruction for transmission tomography using mixture model priors and deterministic annealing algorithms, Journal of Electronic Imaging 12 (1) (2003) 7–16.
[38] Y. Lin, D. Lee, Bayesian regularization and nonnegative deconvolution for room impulse response estimation, IEEE Trans. Signal Processing 11 (3) (2006) 839–847.

**Saïd Moussaoui** received the State engineering degree from Ecole Nationale Polytechnique, Algiers, Algeria, in 2001, and the Ph.D. degree in 2005 from Université Henri Poincaré, Nancy, France. He is currently Assistant Professor at Ecole Centrale de Nantes. Since september 2006, he is with the Institut de Recherche en Communications et Cybérnétique de Nantes (IRCCYN, UMR CNRS 6597). His research interests are in statistical signal and image processing methods, including source separation, non-negative matrix factorization and their applications to real data analysis.

**Hafrún Hauksdóttir** received the M.S. degree from the University of Iceland. She did her M.S. thesis, Blind and Semiblind Source Seperation of Hyperspectral Data from the Planet Mars, at the Laboratory of Images and Signals (LIS) in Grenoble France. She is currently working at IceTec technological Institute of Iceland. Her research interests mainly concern statistical signal processing.

**Frédéric Schmidt** received in 2001 the Two-years degree in Physics from Université Louis Pasteur, Strasbourg, France, in 2003 the B.S. in Earth Science from Ecole Normale Supérieure, Lyon, France and in 2004 the M.S. in Earth Science from Université Joseph Fourier (Grenoble, France). He is currently completing his Ph.D. at the Laboratoire de Planétologie de Grenoble (CNRS-UJF), Grenoble, France. His research interests are analysis of hyperspectral data, statistics in geophysical fields, ices on planet Mars and Martian climate. In particular he is working on hyperspectral images produced by the OMEGA imaging spectrometer (MEX, ESA).

**Christian Jutten** received the PhD degree in 1981 and the Docteur és Sciences degree in 1987 from the Institut National Polytechnique of Grenoble (France). He taught as associate professor in the Electrical Engineering Department from 1982 to 1989. He was visiting professor in Swiss Federal Polytechnic Institute in Lausanne in 1989, before to become full professor in University Joseph Fourier of Grenoble, more precisely in the sciences and technologies department: PolytechGrenoble. He is currently associate director of the Grenoble images, speech, signal and control laboratory (GIPSA, 300 peoples) and head of the Department Images-Signal (DIS) of this laboratory. For 25 years, his research interests are blind source separation, independent component analysis and learning in neural networks, including theoretical aspects (separability, source separation in nonlinear mixtures) applications in signal processing (biomedical, seismic, speech) and data analysis. He is author or co-author of more than 40 papers in international journals, 16 invited papers and 100 communications in international conferences. He has been associate editor of IEEE Trans. on Circuits and Systems (1994-95), and co-organizer the 1st International Conference on Blind Signal Separation and Independent Component Analysis (Aussois, France, January 1999). From January 2007, he is a member of the Machine Learning for Signal Processing Technical Committee of the IEEE Signal Processing Society. He is a reviewer of main international journals (IEEE Trans. on Signal Processing, IEEE Signal Processing Letters, IEEE Trans. on Neural Networks, Signal Processing, Neural Computation, Neurocomputing, etc.) and conferences in signal processing and neural networks (ICASSP, ISCASS, EUSIPCO, IJCNN, ICA, ESANN, IWANN, etc.). He is a Senior Member of the IEEE.



**Jocelyn Chanussot** received the degree in electrical engineering from the Grenoble Technical Institute (INPG), Grenoble, France, in 1995, and the Ph.D. degree from Savoie University, Annecy, France, in 1998. In 1999, he was with the Geography Imagery Perception Laboratory for the Delegation Générale de l'Armement (French National Defence Department). From 1999 to 2005, he has been an Assistant Professor of signal and image processing at INPG and has been working at the Grenoble Images sPeech Signals and Automatics Lab (GIPSA-Lab), Grenoble, where he now is an Associate Professor. His research interests include statistical modeling, multicomponent image processing, nonlinear filtering, remote sensing and data fusion. Dr. Chanussot has been an Associate Editor of the IEEE Geoscience and Remote Sensing Letters (2005-2007). He is now an Associate Editor for the IEEE Transactions on Geoscience and Remote Sensing (2007-2010) and for Pattern Recognition (2006-2008). He is the Co-Chair of the IEEE Geoscience and Remote Sensing Society Data Fusion Technical Committee (2005-2007) and a member of the Machine Learning for Signal Processing Technical Committee of the IEEE Signal Processing Society (2006-2008). He serves as a regular reviewer for various conferences (IEEE ICASSP, IEEE ICIP, ACIVS). He is a Senior Member of the IEEE.

**David Brie** received the Ph.D. degree in 1992 and the *Habilitation à diriger des Recherches* degree in 2000, both from the Henri Poincaré University, Nancy, France. He is currently Professor at the telecommunication and network department from the *Institut Universitaire de Technologie*, Henri Poincaré University. Since 1990, he has been with the *Centre de Recherche en Automatique de Nancy* (CRAN), UMR CNRS 7039. His research interests mainly concern non stationary signal analysis and inverse problems, including deconvolution, source separation and system identification.

**Sylvain Douté** received the B.S. and M.S. degrees from Université Joseph Fourier (UJF), Grenoble, France, in 1991 and 1994, respectively, and the Ph.D. degree from Université Denis Diderot, Paris, France, after working 3 years at the Laboratoire de Glaciologie et Géophysique de l'Environnement, Grenoble, France. He held a post-doctoral position for 2 years at the Institute of Geophysics and Planetary Physics (UCLA) to analyze images of the Near Infrared Mapping Spectrometer (Galileo, NASA). He is currently a researcher at the Laboratoire de Planétologie de Grenoble (CNRS-UJF), Grenoble, France. His research interests include remote sensing of planetary surfaces by hyperspectral imaging, theoretical as well as experimental study of solar light radiative transfer in geophysical media. In particular he is working on images produced by the OMEGA imaging spectrometer (MEX, ESA).

**Jon Atli Benediktsson** received the Cand.Sci. degree in electrical engineering from the University of Iceland, Reykjavik, and the M.S.E.E. and Ph.D. degrees from Purdue University, West Lafayette, IN, in 1984, 1987, and 1990, respectively. He is currently a Professor of electrical and computer engineering at the University of Iceland. He has held visiting positions at University of Trento, Italy (2002-present), Kingston University, Kingston upon Thames, U.K. (1999-2004), the Joint Research Centre of the European Commission, Ispra, Italy (1998), Denmark's Technical University (DTU), Lyngby (1998), and Purdue University (1995). He was a Fellow at the Australian Defence Force Academy (ADFA), Canberra, in August of 1997. From 1999 to 2004, he was Chairman of the energy company Metan Ltd. His research interests are in remote sensing, pattern recognition, neural networks, image processing, and signal processing, and he has published extensively in those fields. Dr. Benediktsson received the Stevan J. Kristof Award from Purdue University in 1991 as outstanding graduate student in remote sensing. In 1997, he was the recipient of the Icelandic Research Council's Outstanding Young Researcher Award, and in 2000, he was granted the IEEE Third Millennium Medal. He is Editor of IEEE Transcactions on Geoscience and Remote Sensing (TGARS) and Associate Editor of the IEEE Geoscience and Remote Sensing Letters. He was Associate Editor TGARS from 1999 to 2002. In 2002, he was appointed Vice President of Technical Activities in the Administrative Committee of the IEEE Geoscience and Remote Sensing Society (GRSS) and (with Paolo Gamba and Graeme G. Wilkinson) a Special Issue on Urban Remote Sensing from Satellite (October 2003). From 1996 to 1999, he was the Chairman of GRSS' Technical Committee on Data Fusion and was elected to the Administrative Committee of the IEEE Geoscience and Remote Sensing Society (GRSS) for the term 2000 to 2002, and in 2002, he was appointed Vice President of Technical Activities of GRSS. He was

the founding Chairman of the IEEE Iceland Section and served as its Chairman from 2000 to 2003. He was the Chairman of the University of Iceland's Science and Research Committee (1999-2005), a member of Iceland's Science and Technology Council (2003-2006), and a member of the Nordic Research Policy Council (2004). He was a member of a NATO Advisory Panel of the Physical and Engineering Science and Technology Sub-Programme (2002-2003). He is a member of Societas Scientiarum Islandica and Tau Beta Pi. He is a Fellow of the IEEE.