Retrieval of Mars surface physical properties from OMEGA hyperspectral images using Regularized Sliced Inverse Regression.

Caroline Bernard-Michel, Sylvain Douté, Laurent Gardes, Stéphane Girard

Retrieval of Mars surface physical properties from OMEGA hyperspectral images using Regularized Sliced Inverse Regression

C. Bernard-Michel,[1] S. Douté,[2] L. Gardes,[1] and S. Girard[1]

C. Bernard-Michel, L. Gardes and S. Girard, MISTIS, INRIA Rhône-Alpes, Montbonnot, Inovallée, 655 avenue de l'Europe, 38 334 Saint Ismier Cedex, France.

S. Douté, Laboratoire de Planétologie de Grenoble, Bât. D de Physique, B.P. 53, 38041 Grenoble Cedex 9, France. (sylvain.doute@obs.ujf-grenoble.fr)

[1]MISTIS, INRIA Rhône-Alpes,

Montbonnot, France.

[2]Laboratoire de Planétologie de Grenoble,

Saint Martin-d'Hères, France.

Abstract. Hyperspectral remote sensing, also known as imaging spectroscopy, is a promising space technology regularly selected by agencies with regard to the exploration and observation of planets, to earth's geology or to the monitoring of the environment. It allows to collect for each pixel of a scene, the intensity of light energy reflected from planets as it varies across different wavelengths. More than one hundred spectels in the visible and near infra-red are typically recorded, making it possible to observe a continuous spectrum for each image cell. Usually, in space exploration, the analysis of these spectral signatures allows to retrieve the physical, chemical or mineralogical properties of surfaces and of atmospheres that may help to understand the geological and climatological history of planets.

We propose in this paper a statistical method to evaluate the physical properties of surface materials on Mars from hyperspectral images collected by the OMEGA instrument aboard the Mars express spacecraft. The approach we develop is based on the estimation of the functional relationship F between some physical parameters and observed spectra. For this purpose, a database of synthetic spectra is generated by a physical radiative transfer model and used to estimate F. The high dimension of spectra is reduced by using Gaussian regularized sliced inverse regression (GRSIR) to overcome the curse of dimensionality and consequently the sensitivity of the inversion to noise (ill-conditioned problems). Compared with a naive spectrum matching approach such as the k-nearest neighbors algorithm, estimates are more accurate and realistic.

1. Introduction

Visible and near infrared imaging spectroscopy is a key remote sensing technique to study and monitor planets. It allows the detection, mapping and characterization of minerals, as well as volatile species, that often constitute the first step toward the resolution of key climatic and geological issues [*Murchie et al.*, 2007; *Bibring et al.*, 2004a; *Brown et al.*, 2004; *Carlson et al.*, 1992]. These tasks are carried out by the spectral analysis of solar light reflected in different directions by the materials forming the top few millimeters or centimeters of the ground. Physical properties of the surface, such as chemical composition, granularity, texture, and physical state are some of the most important parameters that characterize the morphology of the hundred thousand spectra that typically constitute an image. Modeling the direct link between these parameters and observable spectra is usually called the forward problem in classical physics. It can be evaluated numerically by radiative transfer models, simulating the propagation of solar light through the atmosphere and surface to the sensor [*Douté et al.*, 2007; *Shkuratov et al.*, 1999; *Douté and Schmitt*, 1998; *Hapke*, 1993]. It allows us, for given values of the model parameters, to simulate the spectra that should be observed. Conversely, deducing the physical model parameters from the observed spectra is called an inverse problem. It generally cannot be solved analytically, and the use of optimization or statistical methods becomes necessary. From a general point of view, solving inverse problems requires that the fundamental physics are adequately understood, such that a function, G, may be specified between the spectra X and the parameters Y such that:

$$X = G(Y). \tag{1}$$

Once this relationship G is known, different methods can be used to deduce the parameters Y from the observable X. They can be roughly divided in three categories (for further details and comparisons, see [*Kimes et al.*, 2000; *Pragnère et al.*, 1999]):

• *Optimization algorithms* are the most traditional approaches. They consist of minimizing an objective function that expresses the similarity between an observed and a simulated spectrum. For example, this function may be the mean square errors function, where the errors are the differences between observed and simulated reflectances at each wavelength. These methods involve numerical optimization techniques (Powell's method, Simplex method, quasi-Newton method...) starting with an initial guess of the parameters and searching for the optimum parameters via an iterative process that minimizes the objective function. In fact, estimations are unstable since inverse problems are often ill-posed (solutions are not unique and a small change in the data can lead to enormous differences in the estimations). A probabilistic formalism can then be used to regularize inverse problems by introducing a prior distribution on model parameters [*Aster et al.*, 2005; *Tarantola*, 2005; *Mosegaard and Tarantola*, 2002]. These approaches are computationally expensive since they iteratively simulate new spectra. Therefore, they cannot be applied in the case of the inversion of an image with several hundred thousand pixels. Moreover, they can sometimes fall into local minima if the objective function is not convex.

• *Look-up table (LUT) / k-nearest neighbors approach (k-NN)*: In order to replace a heavy runtime computation with a simpler look-up operation, a large database (look-up table) is generated by radiative transfer for many parameter values and stored on disks. To inverse a hyperspectral image, the spectrum at each pixel is then compared with the

look-up table spectra in order to find the best match (the nearest neighbor), according to an objective function minimization. Parameters are then deduced from the look-up table best match spectrum [*Carlson et al.*, 2005; *Philpot et al.*, 2004; *Combal et al.*, 2002; *Weiss et al.*, 2000; *Douté et al.*, 2001]. In comparison with traditional optimization methods, the speed gain is significant, since retrieving a value from memory is often faster than undergoing an expensive computation. The main disadvantages of this approach are the multiplicity of solutions and their instability (see section 3).

- *Training approaches* assume that there exists a functional relationship

$$Y = F(X) \tag{2}$$

between spectra and parameters. This relationship corresponds to the inverse of the physical model G in the forward problem (1). The idea is to use a look-up table to estimate the underlying mathematical relationship F. This relationship then allows us to estimate the parameters of new spectra. Computation time can then be very competitive. Neural networks [*Kamgar-Parsi and Gualtieri*, 1990] or support vector machines (SVM) [*Scholkopf and Smola*, 2002] can be used to learn the function F. The advantage of a such a training approach is that, once the relationship has been established, it can be used for very large sets and for all new images with the same physical model.

Inverse problems occur in many branches of sciences such as geophysics, hydrology, medical imaging, pedology, remote sensing, astronomy or planetology. The choice of one methodology among the others is generally determined according to the degree of knowledge of the studied physical phenomenon. It is also determined by the quality and quantity of measurements. In geophysics or meteorology, probabilistic approaches combined to optimization techniques are most of the time preferred for the following main

reasons: the inverse problem can be linearized, observations are numerous and physicists do have an *a priori* about the parameters variations. In planetary sciences, massive inverse problems have only been tackled recently since they intervene in the last step of a physical study of the huge collections of hyperspectral images. Moreover, inverting hyperspectral imaging on planets has to be designed with the following constraints:

• Dealing with extremely large datasets and various modeling require simple and fast methodologies,

• Inverting high dimensional spectra comes up against the curse of dimensionality,

• Observed spectra always contain some noise.

In that context, LUT and training approaches are generally preferred since they are fast and allow to deal with non linear inverse problems:

• The LUT/k-NN methodology is currently used by physicists to study planetary bodies [*Carlson et al.*, 2005; *Philpot et al.*, 2004; *Weiss et al.*, 2000; *Douté et al.*, 2001]. It provides an instinctive, simple and fast way to retrieve the model parameters. A good match between a simulated and an observed spectrum also reinforces the idea that the chosen physical modeling is realistic. However, we will show in section 3 that LUT/k-NN approaches lead to very unstable estimates.

• Training approaches have been recently considered in the study of terrestrial vegetation. Among them, neural networks [*Combal et al.*, 2002; *Pragnère et al.*, 1999] seem promising but the underlying learning process still remains time consuming. Since the study of Mars is not focused on a particular type of soil, many physical models are explored, and a faster methodology is required. Support vector machines [*Durbha et al.*,

2007] could be used but their disadvantage is that they are difficult to interpret statistically and physically.

In this paper, we propose to develop a new training approach: the Gaussian regularized sliced inverse regression (GRSIR). It is based on a dimension reduction technique: the sliced inverse regression [*Li*, 1991]. To the best of our knowledge, this method has never been used in the context of inverse problems in planetology. It has the advantage to be fast, stable, statistically and physically interpretable and can also help to select an appropriate look-up table for inversion. For the sake of validation, this approach is compared all along with the naive but currently used k-NN approach. In section 2, we first introduce the datasets we are working with. Then, the limits of the LUT/k-NN approach are discussed in section 3. The Gaussian regularized sliced inverse regression (GRSIR) [*Bernard-Michel et al.*, 2007b] is introduced and detailed in section 4. The choice of an appropriate look-up table is discussed in section 5. Two applications of GRSIR methodology are then presented:

- A validation on simulations, in section 6, allows to quantify the relevance of GRSIR according to two criteria,
- In section 7 the retrieval of some physical parameters for the south permanent polar cap of Mars: proportions of CO_2 ice, dust, water ice and grain sizes of CO_2 ice and water ice.

2. Data

The datasets used in the next sections can be divided in three categories: The hyperspectral images observed on Mars, a look-up table simulated by radiative transfer models

according to the physical modeling of these images and a test data in order to quantify the estimates accuracy. In any case, these datasets consist of:

• Some spectra $X \in \mathbb{R}^d$, where d is the number of considered wavelengths and will be called the dimension of the problem. In the presented application d is equal to 184 and and consequently leads to the resolution of a high-dimensional problem.

• Some associated parameters $Y \in \mathbb{R}^p$, where p is the number of free parameters in the physical model. In this paper, five parameters ($p = 5$) will be considered.

2.1. Hyperspectral Images from Mars

We concentrate on datasets collected by the spectro-imaging instrument: OMEGA (observatoire pour la minéralogie, l'eau, la glace et l'activité). OMEGA is one of the seven scientific instruments aboard the European spacecraft Mars express mission, sent in orbit around Mars in 2003. It has been developed by IAS and LESIA (Observatoire de Paris) with the support of CNES, and a participation of IFSI (Italy) and IKI (Russia). This visible and infrared mapping spectrometer should scan most of Mars from orbits in order to observe the gases and dust in the atmosphere and to look for signs of specific materials such as silicates, carbonates, ices at the surface. It records the visible and infrared light reflected from the planet in the wavelength range 0.5-5.2 microns and with a ground resolution varying from 350 m to 10 km. Our study will refer to three OMEGA hyperspectral images acquired during orbits 41, 61 and 103. They cover the high southern latitudes of Mars. The spatial resolution is about 2km per pixel and we considered 184 wavelengths in the range 0.95-4.15. For each image, a preprocessing aiming at correcting the atmospheric gaseous contribution in the spectra has been applied [*Douté et al.*, 2007]. After a first analysis, these OMEGA observations revealed [*Bibring et al.*, 2004b] that

the south polar region of Mars mainly contains water ice, carbon dioxide ice and dust. A detailed qualitative mapping of H_2O and CO_2 ices during the local summer shows that the permanent south polar region is dominated by superficial CO_2 ice on the bright cap except at its edges where water ice appears in extended areas. Examining the coexistence modes (geographical or granular) between H_2O, CO_2 ice and dust that best explain the morphology of the spectra has then led to the implementation of a physical modeling of individual spectra with a surface reflectance model [*Douté et al.*, 2007]. This model allows the generation of synthetic spectra with the corresponding pairs of parameters that constitute a look-table [*Douté et al.*, 2007]. We will focus on the terrain unit of strong concentration of CO_2: the bright permanent south polar cap. This unit has been carefully mapped by a classification method based on wavanglets [*Schmidt et al.*, 2007]. For each image, the CO_2 ice areas contain about 10000 to 20000 spectra.

2.2. Look-up table

Because we do not have any *a priori* about the range of values in which each parameter is varying, we chose to simulate a large look-up table, containing parameters values that are probably not found in our three considered images. The look-up table has been simulated for five parameters varying spatially: the grain sizes of water and CO_2 ices, the proportions of water ice, CO_2 ice and dust. The sum of the proportions is constrained to one. The other parameters intervening in the physical model, such as the incidence and emergence angles or the grain size of dust have been fixed to a constant. Details about the sampling strategy are given in Table 1. We will denote by:

- n the number of simulated spectra ($n = 31500$),
- $x_i \in \mathbb{R}^d, i \in \{1, \ldots, n\}$ the spectra,

158 • $y_i \in \mathbb{R}^p, i \in \{1, \ldots, n\}$ the p associated parameters.

2.3. Test data

159 For the validation, the use of a test dataset is required that we will denote by "Tdata".
160 The one we chose has been simulated by radiative transfer modeling for random values of
161 the proportions of dust and CO_2 ice and the grain sizes of CO_2 and water ices. They have
162 been simulated according to the same physical model than the one used for generating
163 the look-up table. Details about the ranges of variation can be found in Table 1. In
164 order to work in realistic conditions, a multiGaussian noise of dimension 184 has been
165 added to all the spectra of the test dataset. The noise has been simulated with a mean
166 fixed to zero for all wavelengths and with a covariance matrix determined experimentally
167 from the OMEGA image acquired during orbit 41. A small portion of the image, very
168 homogeneous in terms of composition and physical properties, is chosen so that we can
169 assume that much of the variability comes from the noise. The latter is then evaluated
170 by a statistics after applying a shift difference on the selected portion. Afterward, we will
171 denote:

172 • n_T the number of spectra from Tdata ($n_T = 3500$),

173 • $x_i^T \in \mathbb{R}^d, i \in \{1, \ldots, n_T\}$ the spectra from Tdata,

174 • $y_i^T \in \mathbb{R}^p, i \in \{1, \ldots, n_T\}$ the associated parameters.

3. Limits of the *k*-NN approach

Let $o \in \mathbb{R}^d$ be the spectral reflectance of one observed pixel on Mars.

The nearest neighbor algorithm (also known as the k-NN algorithm) consists of searching the k nearest spectra of o in the learning database minimizing the mean square errors

between the observed spectrum o and the simulated spectra. The k nearest spectra of the learning database are selected by sorting in ascending order the values of the objective function $\Phi(i)$ defined by:

$$\Phi(i) = \|o - x_i\|^2. \tag{3}$$

where $i \in \{1, \ldots, n\}$.

If $k = 1$, the estimated parameters for the pixel are then the ones associated to the nearest spectrum selected in the learning database. If k is greater than 1, then one can choose to estimate parameters by the mode, or the mean of the k best matches, but also to keep all the possible estimates. Most of the time, only the best match is retained.

The problem of k-NN approach is that it usually leads to very unstable estimates and it is then difficult to choose a judicious k. Let us consider the example of two spectra S1 and S2 for which parameters are known and let us add a reasonable noise to these spectra. Applying the k-NN methodology ($k = 1$) leads to estimations given in Table 2. We can notice here that estimation errors are relatively small for spectrum S1 whereas they are much greater for spectrum S2. One first thinks that a good idea to improve results would be to consider more than 1 neighbor. In fact, figure 1 shows how unstable the estimates are. It presents the relative estimation error as a function of k, where k is the kth neighbor. First, we can see that considering the kth neighbor or the $(k+1)$th neighbor generally leads to very different estimations. Moreover, for spectrum S1, errors are increasing when k is increasing whereas for spectrum S2 errors are decreasing when k is increasing. That means that for spectrum S1, very few neighbors would be necessary to estimate the parameters properly whereas with spectrum S2, a lot of neighbors would

be required. In fact, for spectrum S2, more than one hundred neighbors are necessary to have at least a realistic estimation of the grain size of CO_2 ice. In these conditions, one can see that estimating parameters by a k-NN approach can lead to significant errors especially when there is no *a priori* on the solutions.

4. Proposed approach

We propose here to estimate the functional relationship F between the spectra $X \in \mathbb{R}^d$ and each physical parameter $Y^j = F(X) \in \mathbb{R}$ with $j \in \{1, \ldots, p\}$. Since each parameter is studied individually, we will consider in the next sections that $p = 1$ and we will always omit j for the sake of simplicity. Because of the curse of dimensionality, estimating a d-variate function from the data is not realistic and dimension reduction techniques have to be applied. They rely on the assumption that the predictor X can be replaced by its projection onto a subspace of smaller dimension L, called the effective dimension reduction space (EDR), without loss of information. Denoting by $\beta_1, \ldots, \beta_L$ the basis of this subspace, the functional relationship $Y = F(X)$ can be rewritten as $Y = f(\beta_1^t X, \ldots, \beta_L^t X)$ where f is now a L-variate function. In most applications, L is smaller than 3 making the estimation of f realistic. The estimation of the projection axes $\beta_1, \ldots, \beta_L$ is discussed in the next paragraph.

4.1. Dimension reduction step: Gaussian regularized sliced inverse regression

Principal component analysis (PCA) is the most current technique that could be used to choose the β_i's [*Bishop*, 2006, chapter 12]. The basic idea in PCA is to define the orthogonal projection of the spectra onto a lower dimensional linear space such that the total variance of the projected spectra is maximized. It amounts to diagonalizing the

spectra covariance matrix

$$\Sigma = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^t \text{ where } \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{4}$$

or equivalently to calculating the eigenvectors of Σ. However, in the case of a regression problem, PCA is generally not satisfying since only the explanatory variables X are considered and the dependent variable Y is not taken into account. Specific dimension reduction techniques have been developed for regression problems, and among them Sliced Inverse Regression (SIR) is the most popular [*Li*, 1991]. It consists first of sorting the parameter values in increasing order and to divide them into a chosen number H of non overlapping slices $S_h, h \in \{1, \ldots, H\}$. For each slice S_h, a mean spectrum m_h is calculated by averaging the spectra for which the parameter values are varying in the slice (see equation (5)). The basic idea in SIR is then to define the orthogonal projection of the mean spectra onto a lower dimensional linear space such that the variance of the projected mean spectra is maximized under the constraint that the total projected variance is 1. It amounts to diagonalizing $\Sigma^{-1}\Gamma$ where Γ is the mean spectra covariance matrix given in equation (6). This principle is illustrated on figure 2 showing that GRSIR consists of maximizing the between-slice variance of the reduced spectra or equivalently in minimizing the within-slice variance.

Since inverse problems are generally ill-posed [*Aster et al.*, 2005; *Tarantola*, 2005], Σ is ill-conditioned making its inversion difficult. We thus propose to compute a Gaussian regularized version of Sliced Inverse Regression (GRSIR). Theoretical details can be found in [*Bernard-Michel et al.*, 2007a]. The concept of this method is to incorporate some prior information on the projections in order to dampen the effect of the noise in the input data, and to make the solution more regular or smooth. The ill-posed problem is then

replaced by a slightly perturbed well-posed problem that depends on a parameter δ, called the regularization parameter. Similarly, GRSIR consists of computing the L eigenvectors corresponding to the L largest eigenvalues of $(\Sigma^2+\delta I_p)^{-1}\Sigma\Gamma$ where I_p is the $d\times d$ identity matrix. In practice, GRSIR then requires the four following computational steps:

Step 1: Sorting $y_i, i \in \{1,\ldots,n\}$ in increasing order and dividing them into H non overlapping slices S_h, $h=1,\ldots,H$. If the Look-up table has been simulated for random values of Y, then the slices are chosen such that each slice contains the same number of observations. If the look-up table has been simulated for a fixed number of distinct parameters, slices are these distinct values.

Step 2: Computing the slice means. For each slice:

$$m_h = \frac{1}{n_h}\sum_{y_i\in S_h} x_i \tag{5}$$

where n_h denotes the number of observations in the slice S_h.

Step 3: Computing the "between slices" means covariance matrix:

$$\Gamma = \sum_{h=1}^{H}\frac{n_h}{n}(m_h-\bar{x})(m_h-\bar{x})^t \tag{6}$$

Step 4: Estimating the GRSIR axes $\beta_1(\delta),\ldots,\beta_L(\delta)$ by computing the eigenvectors of $(\Sigma^2+\delta I_p)^{-1}\Sigma\Gamma$. The first GRSIR axis $\beta_1(\delta)$ is determined by the eigenvector corresponding to the largest eigenvalue, the second GRSIR axis $\beta_2(\delta)$ is determined by the eigenvector corresponding to the second largest eigenvalue and so on.

In section 6.3, we will give a criterion to choose the dimension L. We will see that in practical cases, only the first GRSIR axis is sufficient to retrieve parameters. Let us notice that this first axis can be viewed as a weight function of the wavelengths, giving

some important weights to the wavelengths that carry much more information about the parameter. Afterward, we will denote by reduced spectra the projections of the spectra on this axis $\beta_1(\delta)$.

4.2. Estimation of the functional relationship

Once a relationship has been revealed between reduced spectra and the parameter (see for example figure 2), then the question comes to estimate this relation f. We propose to use a piecewise linear interpolation on the set of data points $(m_h^{proj}, m_h^{param}), h = 1, \ldots, H$ where H denotes the number of slices S_h, $m_h^{proj} =< x_i, \beta_1(\delta) >$ denotes the average of the projection of the spectra x_i, $y_i \in S_h$ for slice S_h and $m_h^{param} = \frac{1}{n_h} \sum_{y_i \in S_h} y_i$ denotes the average parameter value for slice S_h.

For each new spectrum x with a projection $t =< \beta_1(\delta), x >$, the estimated parameter value $\hat{y}$ is then given by:

$$\hat{y} = \begin{cases} m_1^{param} & \text{if } t \in]-\infty, m_1^{proj}] \\ \sum_{h=1}^{H-1} \left[m_h^{param} + (t - m_h^{proj}) \left(\frac{m_{h+1}^{param} - m_h^{param}}{m_{h+1}^{proj} - m_h^{proj}} \right) \right] & \text{if } t \in]m_h^{proj}, m_{h+1}^{proj}] \\ m_H^{param} & \text{if } t \in]m_H^{proj}, +\infty[\end{cases}$$

An example of the application of GRSIR to the look-up table is given in Figure 2. It shows the relationship between reduced spectra and the grain size of CO_2 ice and its estimation by a linear interpolation.

5. Choice of the Look-up table

In the simulated look-up table, we can imagine that many spectra are not necessary to estimate properly the parameters of an observed image. For example, if the studied image is the test data Tdata, we should remove from the look-up table all the spectra with parameters values that are not included in the range of values of Tdata's parameters.

We would then obtain a new look-up table that leads to better estimations for both k-NN and GRSIR [*Bernard-Michel et al.*, 2007a]. The difficulty is that in practical cases, we do not have any *a priori* about the range of variations of the parameters. The best approach in that case is then to take a huge look-up table and to reduce it after the first estimations by GRSIR or k-NN. This approach suffers from the fact that if the database is too important, then estimations are strongly deteriorated. In the case of k-NN, these deteriorations are due to the similarity of some spectra with very different parameters that lead to multiple solutions. In the case of GRSIR, these deteriorations are mainly due to the spectra that are on the opposite too different from the observed one. In order to visualize the adequacy of the chosen look-up table with the observed spectra, we propose to use principal component analysis. We consider that we can retrieve the parameters of the observed spectra from a look-up table only if their projections on the m first PCA axes (deduced from the application of PCA to the look-up table) coincide with the projection of the look-up table itself. For the sake of visibility and because in our applications two components explain until 88% of the total variance, we chose $m = 2$. In some cases, a greater value for m could be required. To select the most appropriate spectra for inversion in the look-up table, our idea is to retain the spectra from the look-up table whose projections on PCA axes are close to projections of the spectra from the observed image. In the plane spanned by the m first PCA axes, we calculate the distance of each projected spectrum from the look-up table with its nearest neighbor in the observed image. The histogram of these distances allows to distinguish a mixture of $nclass$ Gaussian densities $\mathcal{N}(\mu_i, \Sigma_i)$ characterized by parameters such as the proportions of the mixture $\pi_i, i \in \{1, \ldots, nclass\}$, the expectation μ_i and the covariance matrix Σ_i

of each class $i, i \in \{1, \ldots, nclass\}$. The number of classes is chosen by the user and the parameters can be estimated by the Expectation-Maximization algorithm (EM) after being initialized randomly or by the K-means algorithm (for further details about mixture models and EM, see *Bishop* [see 2006, chapter 9]). Roughly summarizing, EM algorithm is a simple iterative optimization process for computing the maximum likelihood estimates of the parameters. One can then calculate and maximize the posterior probability that a spectrum belongs to a class observing the distances. This leads to a classification of the spectra into *nclass* classes, generally two classes: the class of spectra from the look-up table that are far from the observed spectra, and conversely the class of spectra that are close. A statistics calculated separately for the spectra belonging to each class allows to refine the estimation of π_i, μ_i, and $\Sigma_i, i \in \{1, \ldots, nclass\}$. An example is given in Figure 3 with the image observed from orbit 41. The histogram allows to distinguish 3 classes from which the third class corresponds quite well to the class we are looking for.

We can also notice that some of the observed spectra are outside the projected look-up table. We can wonder if we should keep these spectra for the inversion step, because either the simulated look-up table is too small, or the chosen physical model is not relevant for them. We propose to select and remove these spectra using the previous methodology. In the PCA plane, the histogram of the distance of each observed spectrum with its nearest neighbor in the selected sub-look-up table allows to distinguish two classes: the invertible spectra and the non invertible spectra (figure 3). The non invertible spectra correspond to pixels at the boundary of the CO_2 bright area and we can then suppose that another physical model should be used to inverse them. Other examples can be found in the work

of *Bernard-Michel et al.* [2007a]. It is also shown that, on simulations, estimations are more accurate when applying such a selection.

6. Analyzing results on simulations

In this section, we propose to quantify the accuracy of GRSIR estimations by using the test data presented in section 2. We also show how to choose the regularization parameter and the EDR dimension. At last, results are presented comparing GRSIR and k-NN methods.

6.1. Validation criteria

To assess the relevance of GRSIR methodology, we focus on two aspects of the validation: the accuracy of the estimates and the quality of the relationship between reduced spectra and parameters. To this end, we define two validation criteria. The first one, denoted Normalized Root Mean Square Errors (NRMSE), quantifies the importance of estimation errors, that are the differences between the estimations $\hat{y}_i^T$ and the real values y_i^T. The second one, denoted SIR Criterion (SIRC) is the ratio between the "between-slices" variance of GRSIR projections and the total variance. It quantifies the quality of the relationship between projected spectra and parameters. A third validation criterion that could have been introduced is the RMSE between the test spectrum x_i^T and the spectrum that can be reconstructed by running the radiative transfer model with the estimated parameters. However, this criterion would have faced the ill-posed problem that two spectra can be very close in terms of RMSE even with very different parameters values. The consequences are that a good reconstruction does not always mean that parameters are well estimated.

316 ### 6.1.1. the NRMSE

The Normalized Root Mean Square Errors (NRMSE) criterion is defined as:

$$NRMSE = \sqrt{\frac{\sum_{i=1}^{n_T}(\hat{y_i}^T - y_i^T)^2}{\sum_{i=1}^{n_T}(y_i{}^T - \overline{y}^T)^2}} \text{ with } \overline{y}^T = \frac{1}{n_T}\sum_{i=1}^{n_T} y_i^T.$$

317 This measure is invariant with respect to position and scale and therefore it is possible to
318 deduce if one parameter is better estimated than another. The NRMSE is close to zero
319 when the predicted values are accurate and conversely becomes larger when predictions
320 are poor.

321 ### 6.1.2. the SIRC

The SIR criterion for an axis $\beta_\ell(\delta)$ is defined as the ratio between the "between-slices" variance $\beta_\ell^t(\delta)\Gamma\beta_\ell(\delta)$ of the projections of $x_i, i \in \{1,\ldots,n\}$ on $\beta_\ell(\delta)$ and the total variance $\beta_\ell^t(\delta)\Sigma\beta_\ell(\delta)$ of these same projections:

$$SIRC = \frac{\beta_\ell^t(\delta)\Gamma\beta_\ell(\delta)}{\beta_\ell^t(\delta)\Sigma\beta_\ell(\delta)} \tag{7}$$

322 It indicates the quality of the functional relationship between projections of the spectra
323 on the ℓth GRSIR axis and the parameters. In the SIR criterion, the total variance can
324 be interpreted as the sum of the between slices variance and the within slice variance. In
325 this case, one can see the quality of the relationship is perfect when the within variance
326 is close to zero, or equivalently when SIRC is close to 1. Finally, the closer SIRC is to 1,
327 the better the relationship is.

6.2. Choice of the regularization parameter

328 In the case of Regularized Sliced Inverse Regression, estimations $\hat{y_i}^T$ and consequently
329 NRMSE criterion as well as the SIRC score depend on a regularization parameter δ.

When the regularization parameter increases, the functional relationship between projected spectra and parameters is getting worse and consequently estimation errors are increasing. From another point of view, if the regularization parameter is too small, then estimation errors are huge in presence of noise because we are dealing with an ill-posed problem. So, there is a compromise to reach between deteriorating the functional relationship and improving estimations by regularizing. In fact, this compromise lies in the choice of the regularization parameter. We propose here to choose this regularization parameter, for each parameter individually, minimizing the NRMSE criterion calculated between the parameter values from a test data and their estimations. The test data can be the look-up table itself spoiled by a multiGaussian noise supposed representative of the one affecting the OMEGA hyperspectral images. Let us notice that if there is no noise in the data, or in other words, if the observed data exactly corresponds to spectra that could be simulated by radiative transfer model, then no regularization is required and minimizing the NRMSE criterion for GRSIR would yield a value δ close to zero. Figure 4 shows how the choice of the regularization parameter changes according to the noise:

- the stronger the noise is, the greater the NRMSE are,
- when the noise is increasing, the chosen regularization parameter, that corresponds to the minimum of the NRMSE, is also increasing.

One can easily see that making a mistake on the estimation of the noise leads to a mistake in the estimation of the regularization parameter and consequently can lead to much more uncertain estimations.

6.3. Choice of the EDR dimension

In order to choose the EDR dimension, we propose to calculate the SIRC for each GRSIR axis $\beta_\ell(\delta)$ with $\ell \in \{1, \ldots, d\}$. As ℓ is increasing, the SIRC decreases very rapidly and is almost zero after very few iterations of the index ℓ. In the next sections, we will fix the dimension of the EDR to the index L from which the SIRC is significantly decreasing. For example, If we apply GRSIR to the look-up table, the SIRC calculated for the proportion of CO_2 ice on the first GRSIR axis is 0.985. This SIRC, very close to one, indicates that the relationship of interest is well revealed by the first axis. Calculated with the second axis, the SIRC drops to 0.1799 and then is close to zero for other axes. In this case, we will consider that taking into account these axes in the regression step do not bring sufficient information and is not necessary, and will fix the EDR dimension to one. In fact, analyzing the SIRC for all parameters and for two different physical models reveals that only one dimension is sufficient to retrieve the model parameters. We found only one situation for which two GRSIR axes have to be kept: the thickness of the compact slab of nearly pure CO_2 ice making the seasonal deposits of the south pole of Mars during spring [*Douté et al.*, 2008]. For this parameter, the SIRC is 0.89 for the first axis, 0.81 for the second axis and close to zero for the third one. Because the necessity to keep more than one GRSIR axis seems rare, L is set to one in the next sections.

6.4. Analyzing the GRSIR axis

The spectral variability of the individuals x_i constituting the look-up table results from a complex interplay of the different input parameters y_i of the model. In particular two parameters can have a similar influence on the spectra for a certain range of values. As a consequence it is crucial to understand how the GRSIR method can untangle the dependencies in order to find independent one to one functional relationships between spectra

and parameters. We thus try to correlate each axis $\beta_1(\delta)$ - which is a vector of spectral weights - and the variability specifically induced by the variation of the corresponding physical parameter. The other parameter values are kept constant at mean values. The weights are individually applied to very different level of reflectance depending on wavelength when calculating the reduced spectrum $\beta_1^t x_i$. Hence we find it more illustrative to represent the componentwise multiplication (wavelength by wavelength) of each axis with a representative spectrum. An illustration is given in figure 5 for the grain size of CO_2 ice. Additionally we overplot in figure 6 the absolute value (for the sake of clarity) of the proportions of water ice and dust and of the grain sizes of water and CO_2 ice. From our analysis we make the following observations: (i) The criteria applied when seeking an axis (maximizing the between-slice variance while minimizing the within-slice variance) implies that GRSIR does put weight on key spectral points that do not display necessary the highest variability when we change the value of the corresponding parameter. (ii) In the case of the physical model that we consider in this paper (intimate mixture of H_2O, CO_2, and dust), the spectral range from 0.92 to 2.75 microns is the most determinant for evaluating the parameters. (iii) The axis for H_2O abundance, dust abundance, and grain size of CO_2 ice are quite similar: we have a strong weight put on wavelength 1.43 micron (the bottom of a very reliable and narrow CO_2 ice absorption band) and, to a lesser extent, on the interval between 1.47 and 1.54 microns (H_2O ice feature), and on wavelengths 1.87 and 2.29 microns (weak CO_2 ice bands but very distinct). Besides the three axes have in common a moderate value at 2.00 microns, at the bottom of a strong CO_2 ice band. The latter is not very diagnostic because it often saturates. Thus, it is satisfying that GRSIR does not put too much emphasis on it. The three axes also shows

differences that allow differentiation between the functional relationships. A relatively strong weight at 2.38 microns on the right wing of the 2.34 micron band is a specificity of the proportion of water ice. We can note in figure 5 that the reflectance at this wavelength is greatly variable when water ice abundance or CO_2 ice grain size varies. The axis for the proportion of dust stresses very much the variation of the spectrum continuum around 1 micron and 1.77 micron that indeed strongly changes with dust concentration. No other parameter affects so much the continuum. The axis for the proportion of dust minimizes the influence of the small plateau at 2.62 microns contrary to the axis for the grain size of CO_2 ice. This is the solution GRSIR finds to differentiate axes of the proportion of dust and the grain size of CO_2 ice because the plateau level varies very much with both dust concentration and CO_2 ice grain size. Furthermore the axis for grain sizes of CO_2 ice induces less sensitivity to the spectrum continuum than the axis for the proportion of dust. In addition, it does not put a noticeable weight on the right wing of the 2.34 micron band contrary to the axis for the proportion of water ice. (iv) Axis 5 which is linked to H_2O grain size is very different from the others since it emphasizes, on the one hand, the continuum level around 1 micron, but not around 1.75 microns and, on the other hand, it emphasizes the whole spectral range between 2.35 and 2.65 microns. The principal conclusions we can draw from our study is that, at least for the model presented in this paper, the GRSIR method succeeds in decorrelating the evaluation of the different parameters. Indeed it has the ability to find a unique set of wavelengths where the variability of the spectrum is the most pertinent (but not necessary the highest) for the evaluation of a given parameter. One must note that these key wavelengths not only fall on the bottom of specific absorption bands for both H_2O and CO_2 ices but also on specific

parts of the continuum as well as on specific band wings. The GRSIR method thus reveals and uses all the pertinent information contained in the spectra that allows to estimate the parameters. This information is more complete than if we were focusing only on band depths as one could expect in the first place. Finally, knowing at the key wavelengths the characteristics (signal to noise ratio, accuracy of the absolute radiometric calibration, etc.) of the instrument that acquires the spectra (the imaging spectrometer OMEGA in our case) is crucial to assess the quality of the solutions.

6.5. Results

In this section, we finally compare the naive k-NN approach with the GRSIR methodology on the test data Tdata. We quantify the gain of one approach from the other according to the SIRC and NRMSE criteria but we also compare estimations to real values with the help of graphic tools such as scatter plots or histograms. For both methods, the look-up table defined in section 2.2 has been used except that a selection of the spectra described in section 5 has been performed before applying the GRSIR methodology. It has no interest to apply this selection to the k-NN methodology because in this selection, most of the spectra leading to multiple solutions for parameters are kept. The results, presented in Table 3, show that for most parameters, estimations are much more accurate with GRSIR. Only one parameter, the grain size of water ice, is better estimated on average by k-NN. This may be due to the fact that too few values have been considered for simulations in the look-up table (only 5). In all other cases, we can see an important gain on the NRMSE which can be 3 times smaller with GRSIR than with k-NN. An illustration of this gain is highlighted in Figure 7 which shows the scatter plot of the estimated proportions of CO_2 ice for both methods versus the real proportions. In this graphic, we

can notice that the range of variation of the estimated proportions of CO_2 ice is much larger with k-NN than with GRSIR. In fact, we can easily see, plotting the associated densities, that GRSIR reproduces better the density of the original proportions of CO_2 ice simulated in Tdata than k-NN (see figure 8). It is interesting to point out that k-NN does not give any information about the range of variation of the parameters of the test data and that estimations often vary in the same ranges of variation that the simulated values from the look-up table. On the other hand, GRSIR, as a first step, allows to evaluate in which ranges the parameters of an observed image are varying. Then a more appropriate look-up table can be built in the estimated ranges with a higher density of parameter values. Applying GRSIR with this new extracted look-up table leads to even more accurate estimations given in Table 4. Finally, the SIRC, always close to 1, is very satisfying.

7. Retrieval of the physical parameters for the south polar cap of Mars

In this section, we present the first model inversions obtained by k-NN and GRSIR for hyperspectral images acquired by OMEGA during orbits 61, 103 and 41 (see section 2). The study of these images by k-NN and GRSIR leads to 30 different maps that we will not present in this paper but that can be found in the work of *Bernard-Michel et al.* [2007a]. We draw up here the main conclusions.

The model inversion on the image observed from orbit 103 by GRSIR and k-NN shows that GRSIR gives very smooth mappings for all sets of parameters. Moreover, GRSIR estimations vary continuously while k-NN estimations assume a small number of values. An example is given in Figure 9, where we see that the proportion of dust is nearly always estimated to a constant equal to 0.0003 with k-NN. With GRSIR, the map is more detailed. In the same way, figure 10 shows the proportion of dust estimated by k-NN and

GRSIR on the portion of the polar cap observed during orbit 41. With k-NN, estimations assume now 8 different values and seem to indicate that, at the very center of the bright cap, very few dust is observed. The abundance of dust increases significantly when getting closer from the boundaries. The estimated map with GRSIR is more detailed and leads to slightly different conclusions. In particular, the area presenting a poor proportion of dust is more extended than with k-NN. Nevertheless if, globally, maps are much smoother and detailed with GRSIR, they never differ entirely from k-NN's.

Another interesting remark concerns the estimation of the grain size of CO_2 ice in images 61 and 103. In fact, images 61 and 103 represent approximately the same portion of the polar cap and should consequently give close estimations for each studied parameter. But we observed that estimations of the grain size of CO_2 ice are strongly different for both images and more especially with k-NN methodology (see figure 11). Observations 61 and 103 were acquired approximately 12 Mars days apart during a period of the year when the solar illumination was declining over the south pole. Consequently, neither the state of the atmosphere (e.g. the dust content) nor the acquisition geometry by the sensor were strictly equivalent but likely comparable. These differences translate to small changes on the overall level of reflectance of the spectra and on the CO_2 ice bands intensity and shape. This is very much comparable to the introduction of noise in a relative sense between observations 61 and 103. Since our remote sensing problem is partly ill-conditioned, a non regularized inversion method such as the k-NN will be much more sensitive to this noise than a regularized one such as the GRSIR. This is exactly what one can notice. The discrepancy that remains between the densities of the estimated grain sizes for orbit 61 and 103 is mainly due to the conjugated effect of an increasingly grazing illumination

over a slightly dusty atmosphere that diminishes the spectral contrast of the observations. Such an effect was not taken into account in the modeling.

Finally, for the sake of validation, we compared estimated maps with the Wavanglet approach developed in [*Schmidt et al.*, 2007]. "Wavanglet" is a supervised automatic detection method that identifies in hyperspectral images spectral features and thus produces distribution maps of chemical compounds. It uses three steps: A. Selection of a library composed of reference spectra (the signature of the compounds to be detected); B. Application of a Daubechies wavelet transform to referenced spectra and determination of the wavelet subspace that best separates all referenced spectra; C. In this selected subspace, calculation of the spectral angle between each spectrum of an observation and a given reference spectrum. In particular this angle, that we will denote the wavanglet angle, allows to quantify in a relative sense the spatial variations of the different compound abundances at the surface. We present in figure 12 the cosine of the wavanglet angle that makes each spectrum of the images 41 and 103 with a reference spectrum of martian dust. The closer it is from one, the greater the dust proportion is. We have a similar map for the water proportion. Globally, estimates of the latter quantities are more noisy with wavanglet than with GRSIR and k-NN, especially for the dust proportion but they generally are in agreement with some exceptions. For example, in the image observed during orbit 103, both k-NN and wavanglet methodologies display an area with strong proportion of dust in the lower right part of figure 12(B), that is absent with GRSIR. In order to see if this area really contains more dust than other areas, we selected the spectrum (denoted by A in figure 12) corresponding to the greatest proportion of dust. We compared it to two other spectra from the cap. The first spectrum (B) has been chosen in a pure CO_2

area with very few dust. The second one (C) has been chosen in a area containing dust. According to k-NN and wavanglet results, this area should however contain less dust than the area where we extracted spectrum A. The spectra A, B and C are presented in figure 13. We can see that as, expected, the spectrum B is really different from the spectrum A. On the other hand, spectra A and C are really similar showing that the area of interest does not contain as much dust as the wavanglet methodology predicts. On the opposite, estimations given by GRSIR are more coherent with this spectral analysis.

8. Conclusion

In this paper, we propose a regularized version of Sliced Inverse Regression in order to retrieve the physical parameters that best explain the spectra observed on Mars by the OMEGA imaging-spectrometer. To the best of our knowledge, this methodology has never been used in the domain of remote sensing and more particularly in planetary sciences. Results on simulations are promising showing that estimations are accurate and most of the time better than the ones given by the k-nearest neighbors algorithm. On a real data, maps are much smoother than with k-NN and seem to give a coherent mapping if we compare the inversion of different hyperspectral images of the same portion of surface of Mars. Moreover, GRSIR is a fast algorithm that calculates once and for all the relationship between spectra and parameters for a determined physical model. Thus, it is then really easy to inverse each new observed spectrum if it is coherent with the model. The main limit of GRSIR approach is that we currently do not give any uncertainties of our estimations when inverting a real image. We could calculate experimental uncertainties based on simulations, but it supposes that the noise in the spectra has been well evaluated. If not, uncertainties will probably be underestimated. Some improvements could also be

proposed to choose the regularization parameter and a more complete analysis of the influence of the noise in the GRSIR methodology would be interesting. A comparison with other training approaches such as SVM or neural networks would allow to better validate results. Finally, the development of a multivariate regularized GRSIR under constraint is conceivable in order to estimate proportions simultaneously.

Acknowledgments. This work is supported by a contract with CNES through its Groupe Système Solaire Program and by INRIA.

References

Aster, R., B. Borchers, and T. C.H. (2005), *Parameter Estimation and Inverse Problems*, Elsevier Academic Press.

Bernard-Michel, C., S. Douté, L. Gardes, and S. Girard (2007a), Estimation of mars surface physical properties from hyperspectral images using sliced inverse regression, *Tech. rep.*, INRIA, http://hal.inria.fr/inria-00187444/.

Bernard-Michel, C., L. Gardes, and S. Girard (2007b), Gaussian regularized sliced inverse regression, *Tech. rep.*, INRIA, http://hal.inria.fr/inria-00180496/.

Bibring, J.-P., et al. (2004a), *OMEGA: Observatoire pour la Minéralogie, l'Eau, les Glaces et l'Activité*, pp. 37–49, ESA SP-1240: Mars Express: the Scientific Payload.

Bibring, J.-P., et al. (2004b), Perennial water ice identified in the south polar cap of mars, *Nature*, *428*, 627–630.

Bishop, C. (2006), *Pattern Recognition and Machine Learning*, Springer.

Brown, R. H., et al. (2004), The Cassini Visual And Infrared Mapping Spectrometer (Vims) Investigation, *Space Science Reviews*, *115*, 111–168, doi:10.1007/s11214-004-

552 1453-x.

553 Carlson, R., M. Anderson, R. Mehlman, and R. Johnson (2005), Distribution of hydrate
554 on europa: Further evidence for sulfuric acid hydrate, *Icarus*, *177*(2), 461–471.

555 Carlson, R. W., P. R. Weissman, W. D. Smythe, J. C. Mahoney, the NIMS Science, and
556 E. Teams (1992), Near infrared spectrometer experiment on Galileo, *Space Sci. Rev.*,
557 *60*, 457–502.

558 Combal, B., F. Baret, M. Weiss, A. Trubuil, D. Macé, A. Pragnère, R. Myneni,
559 Y. Knyazikhin, and L. Wang (2002), Retrieval of canopy biophysical variables from
560 bidirectional reflectance using prior information to solve the ill-posed inverse problem,
561 *Remote Sensing of Environment*, *84*, 1–15.

562 Douté, S., and B. Schmitt (1998), A multilayer bidirectional reflectance model for the anal-
563 ysis of planetary surface hyperspectral images at visible and near-infrared wavelengths,
564 *J. Geophys. Res.*, *103*(12), 31,367–31,390.

565 Douté, S., B. Schmitt, R. M. C. Lopes-Gautier, R. W. Carlson, L. Soderblom, and
566 J. Shirley (2001), Mapping SO_2 frost on Io by the modeling of NIMS hyperspectral
567 images, *Icarus*, *149*, 107–132.

568 Douté, S., E. Deforas, F. Schmidt, R. Oliva, and B. Schmitt (2007), A Comprehen-
569 sive Numerical Package for the Modeling of Mars Hyperspectral Images, in *Lunar and*
570 *Planetary Institute Conference Abstracts*, *Lunar and Planetary Institute Conference*
571 *Abstracts*, vol. 38, p. #1836.

572 Douté, S., B. Schmitt, J.-P. Bibring, Y. Langevin, F. Altieri, G. Bellucci, B. Gondet, and
573 the Mars Express Omega Team (2007), Nature and composition of the icy terrains from
574 mars express omega observations, *Planetory and Space Science*, *55*, 113–133.

Douté, S., F. Schmidt, B. Schmitt, Y. Langevin, M. Vincendon, J.-P. Bibring, and Omega Team (2008), Physical Characterization of the South Seasonal Cap of Mars During Recession from OMEGA Observations, in *Lunar and Planetary Institute Conference Abstracts*, *Lunar and Planetary Inst. Technical Report*, vol. 39, p. #1736.

Durbha, S., R. King, and N. Younan (2007), Support vector machines regression for retrieval of leaf area index from multiangle imaging spectroradiometer, *Remote Sensing of Environment*, *107*, 348–361.

Hapke, B. (1993), *Theory of reflectance and emittance spectroscopy*, Topics in Remote Sensing, Cambridge, UK: Cambridge University Press, 1993.

Kamgar-Parsi, B., and J. Gualtieri (1990), Solving inversion problems with neural networks, *International Joint Conference on Neural Networks*, *3*, 955–960.

Kimes, D., Y. Knyazikhin, J. Privette, A. Abuegasim, and F. Gao (2000), Inversion methods for physically-based models, *Remote Sensing Reviews*, *18*, 381–439.

Li, K. (1991), Sliced inverse regression for dimension reduction, *Journal of the American Statistical Association*, *86*, 316–327.

Mosegaard, K., and A. Tarantola (2002), Probabilistic approach to inverse problems, *International Handbook of Earthquake and Engineering Seismology (Part 1)*, pp. 237–265.

Murchie, S., et al. (2007), Compact reconnaissance imaging spectrometer for mars (crism) on mars reconnaissance orbiter (mro), *Journal of Geophysical Research (Planets)*, *112*(E11), E05S03, doi:10.1029/2006JE002682.

Philpot, W., et al. (2004), Bottom characterization from hyperspectral image data, *Oceanography*, *17*(2), 76–85.

Pragnère, A., F. Baret, M. Weiss, R. Myneni, Y. Knyazikhin, and L. Wang (1999), Comparison of three radiative transfer model inversion techniques to estimate canopy biophysical variables from remote sensing data, *Geoscience and Remote Sensing Symposium, 1999. IGARSS '99 Proceedings. IEEE 1999 International*, *2*, 1093–1095.

Schmidt, F., S. Douté, and B. Schmitt (2007), WAVANGLET: An efficient supervised classifier for hyperspectral images, *Geoscience and Remote Sensing, IEEE Transactions*, *45*(5), 1374–1385.

Scholkopf, B., and A. J. Smola (2002), *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, MIT Press, Cambridge, MA.

Shkuratov, Y., L. Starukhina, H. Hoffmann, and G. Arnold (1999), A model of spectral albedo of particulate surfaces: implications for optical properties of the moon, *Icarus*, *137*, 235–246.

Tarantola, A. (2005), *Inverse problem theory and model parameter estimation*, SIAM.

Weiss, M., F. Baret, R. Myneni, A. Pragnère, and Y. Knyazikhin (2000), Investigation of a model inversion technique to estimate canopy biophysical variables from spectral and directional reflectance data, *Agronomie 20*, pp. 3–22.

Table 1. Sampling strategy for the simulation of the look-up table. The range of variation is given for each varying parameter as well as the number of distinct values simulated.

Parameters	Look-up table		Test data
	range	# distinct values	range
Proportion of water	[0.0001 0.0029]	15	[0.0006 0.002]
Proportion of CO_2	[0.9942 0.9998]	29	[0.996 0.9988]
Proportion of dust	[0.0001 0.0029]	15	[0.0006 0.002]
Grain size of water	[50 450]	5	[100 400]
Grain size of CO_2	[30000 165000]	28	[40000 105000]

Table 2. k-NN estimates for spectra S1 and S2 (described in section 3)

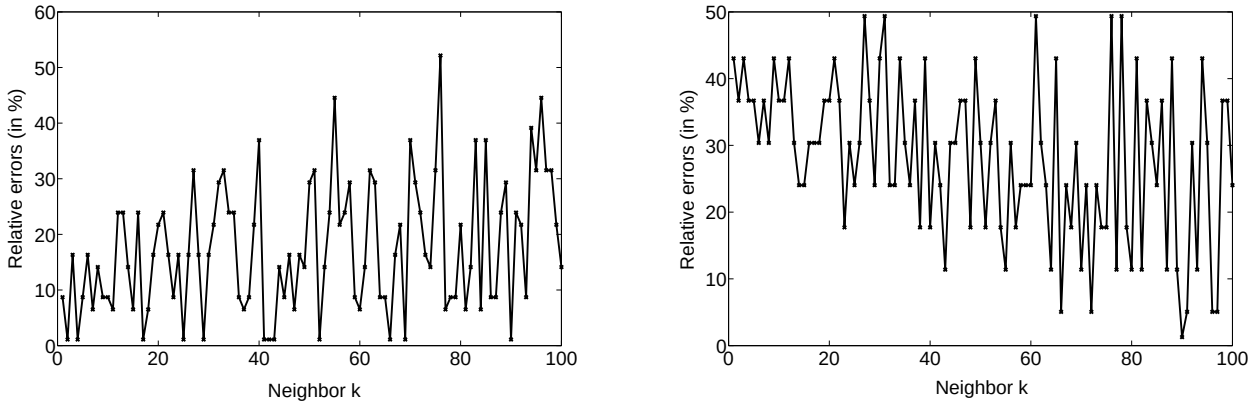
	Spectrum S1		Spectrum S2	
Parameters	Real values	k-NN	Real values	k-NN
Proportion of water	0.0019	0.0021	0.0013	0.0029
Proportion of CO_2	0.9969	0.9966	0.9969	0.9942
Proportion of dust	0.0012	0.0013	0.0017	0.0029
Grain size of water	156	150	109	150
Grain size of CO_2	65721	60000	78993	45000

Table 3. Validation criteria calculated on Tdata with GRSIR and k-NN methodology.

Parameters	k-NN	GRSIR	
	NRMSE	NRMSE	SIRC
Proportion of water	0.86	0.40	0.90
Proportion of CO_2	0.88	0.30	0.98
Proportion of dust	0.44	0.17	0.99
Grain size of water	0.43	0.54	0.84
Grain size of CO_2	0.53	0.22	0.95

Table 4. Validation criteria calculated on Tdata with GRSIR. Here the spectra that constitute the look-up table have been selected after a first inversion by GRSIR.

Parameters	GRSIR	
	NRMSE	SIRC
Proportion of water	0.27	0.92
Proportion of CO_2	0.22	0.99
Proportion of dust	0.13	0.99
Grain size of water	0.37	0.92
Grain size of CO_2	0.19	0.98



S1 S2

Figure 1. Relative errors on the grain size of CO_2 ice for the 100 nearest neighbors of spectra S1 and S2 in the lookup table. Horizontally: kth neighbor. Vertically: Relative errors on the grain size of CO_2 ice.

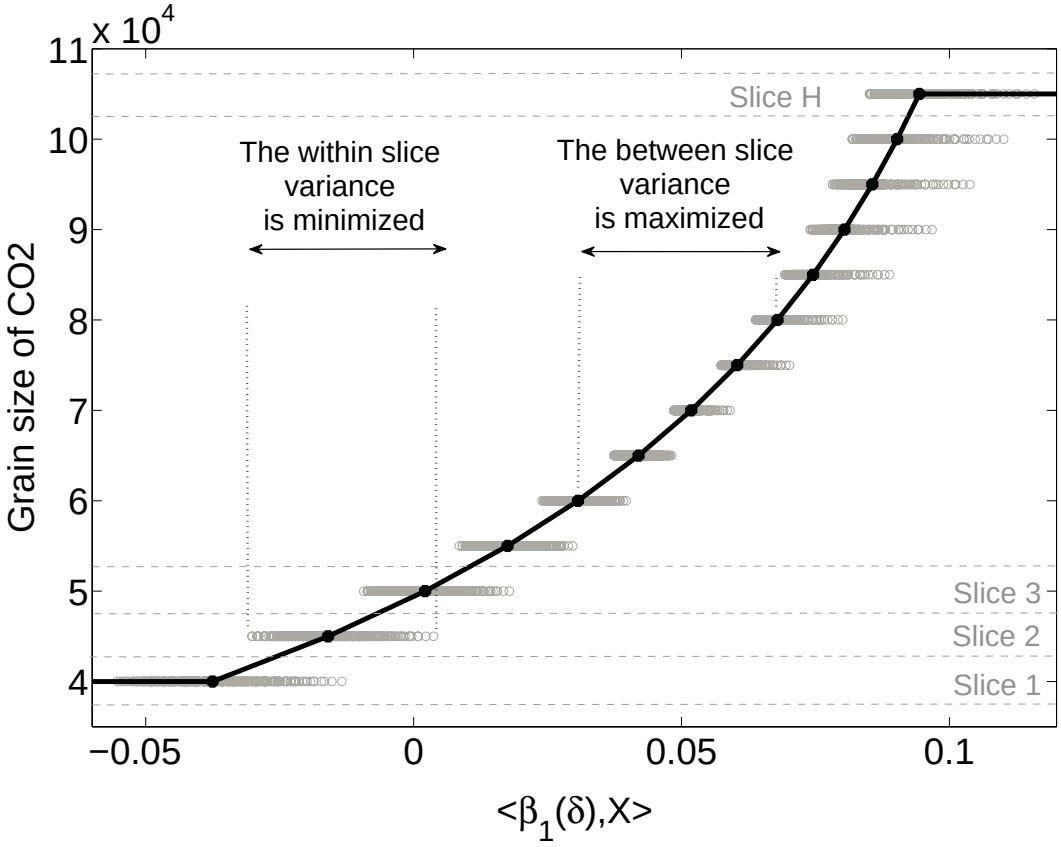


Figure 2. Functional relationship between reduced spectra on the first GRSIR axis and the grain size of CO_2 ice. This graphics illustrates the GRSIR methodology showing that the relationship is the best when the within slice variance is minimized or equivalently when the between slice variance is maximized. Horizontally: reduced spectra from the learning database on the first GRSIR axis. Vertically: Grain size of CO_2 ice.

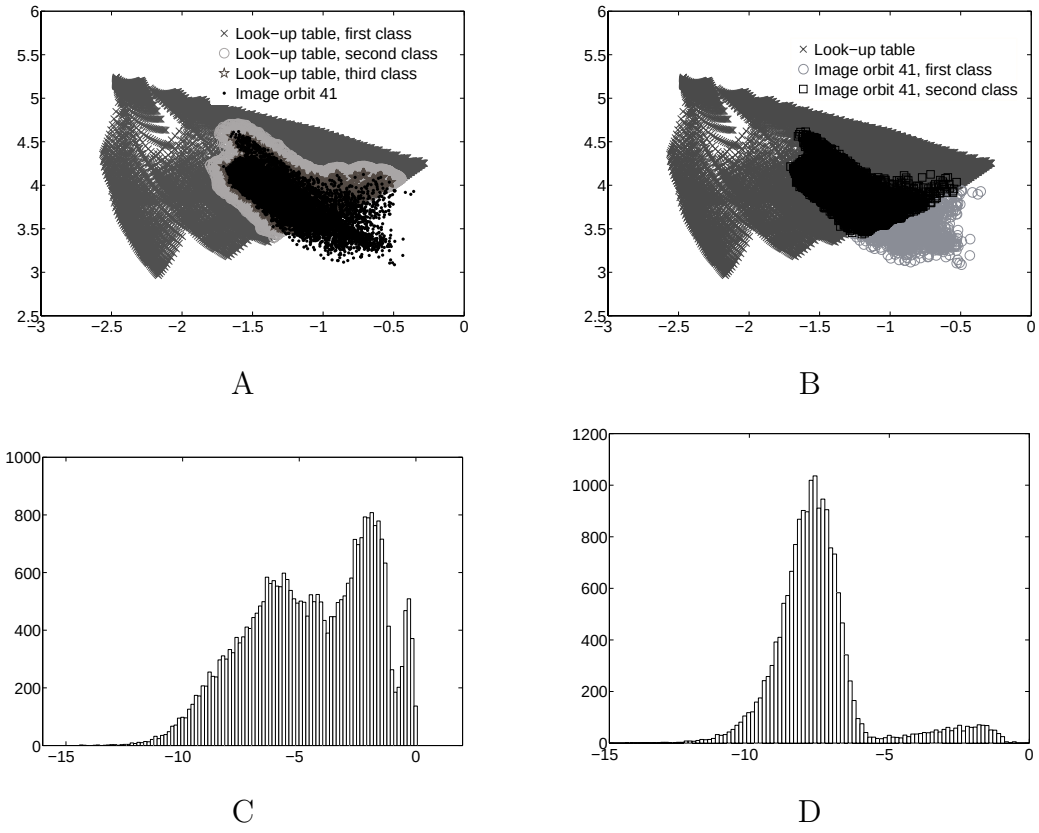


Figure 3. Selection of the "useful" spectra in the look-up table and of the "invertible" spectra in the observed data. A: Selection of a look-up table. Projections of the observed spectra from orbit 41 and the look-up table on the 2 first PCA axes (PCA applied to the look-up table). We distinguish 3 classes. The third class is the retained look-up table. B: Selection of invertible spectra in the image from orbit 41. The second class is retained for inversion. C: histogram of the distances between each spectrum of the look-up table and its nearest neighbor in the image from orbit 41. D: histogram of the distances between each spectrum of the image from orbit 41 and its nearest neighbor in the selected sub-look-up table.

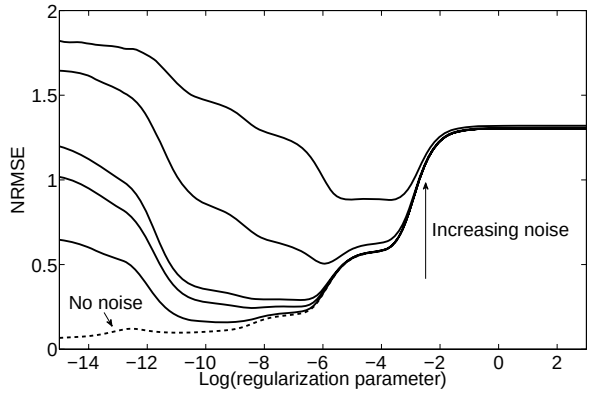


Figure 4. Influence of the noise in the choice of the regularization parameter. Horizontally: logarithm of the regularization parameter. Vertically: NRMSE for the grain size of CO_2 ice. A multiGaussian noise has been introduced in the test data with a diagonal covariance $C = \nu I_p$ proportional to the identity matrix I_p. The experience has been repeated five times with increasing values of ν. One test has been also realized without any noise in the data.

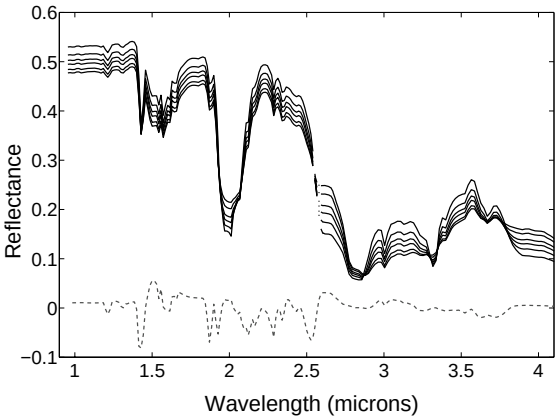


Figure 5. Plain line: series of spectra extracted from the lookup table for CO_2 ice grain size varying between 30 and 165 millimeters, the other parameters being kept constant at mean values. Dotted-dash line: Componentwise multiplication between GRSIR axis for the grain size of CO_2 ice and a representative spectrum (see section 6.4)

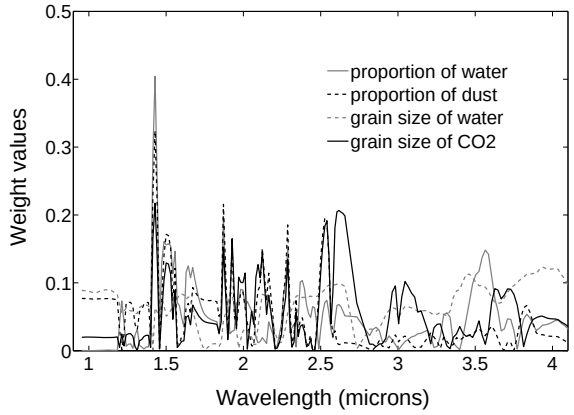


Figure 6. Horizontally: wavelengths. Vertically: GRSIR weights for the proportions of water ice and dust and the grain sizes of water and CO_2 ices.

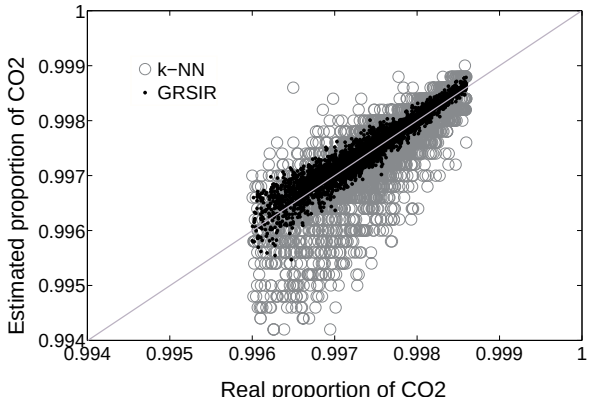


Figure 7. Scatter plot of the proportion of CO_2 ice from the test data versus estimated values. Horizontally: proportion of CO_2 ice. Vertically: estimated proportions of CO_2 ice by GRSIR and k-NN.

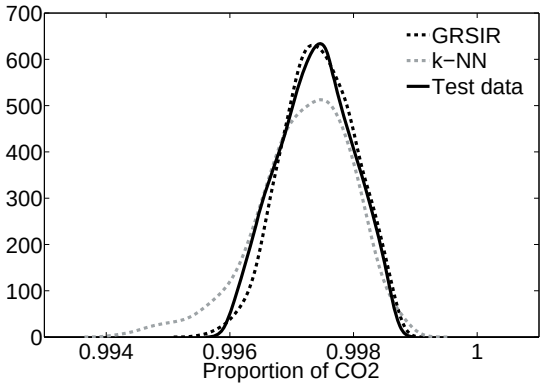


Figure 8. Density of the estimated proportions of CO_2 ice from the test data by k-NN and GRSIR compared to the density of real proportions.

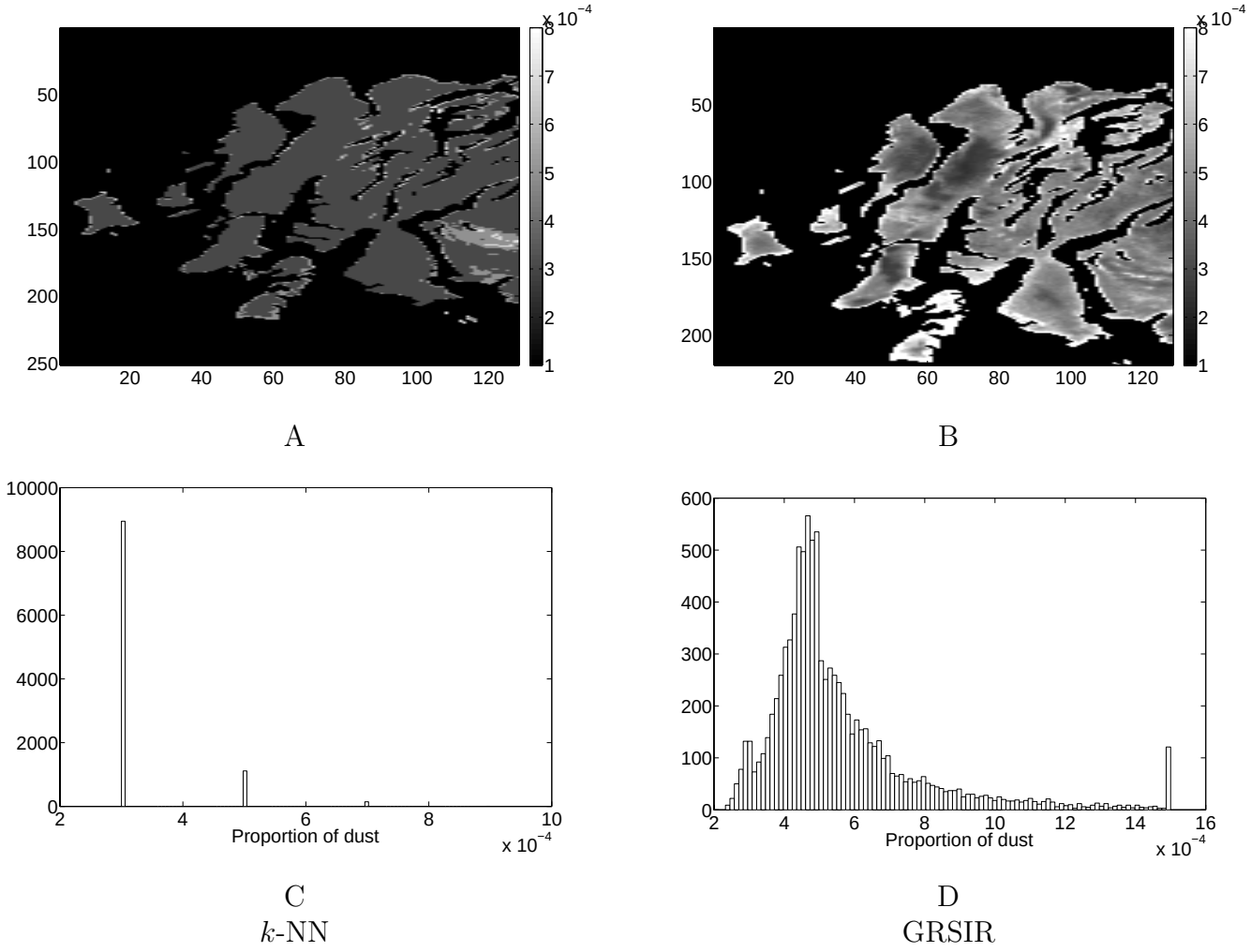


Figure 9. Proportion of dust estimated by k-NN (A) and GRSIR (B) from the hyperspectral image observed from orbit 103. The histograms of the estimations for the entire image are presented in (C) for k-NN and (D) for GRSIR.

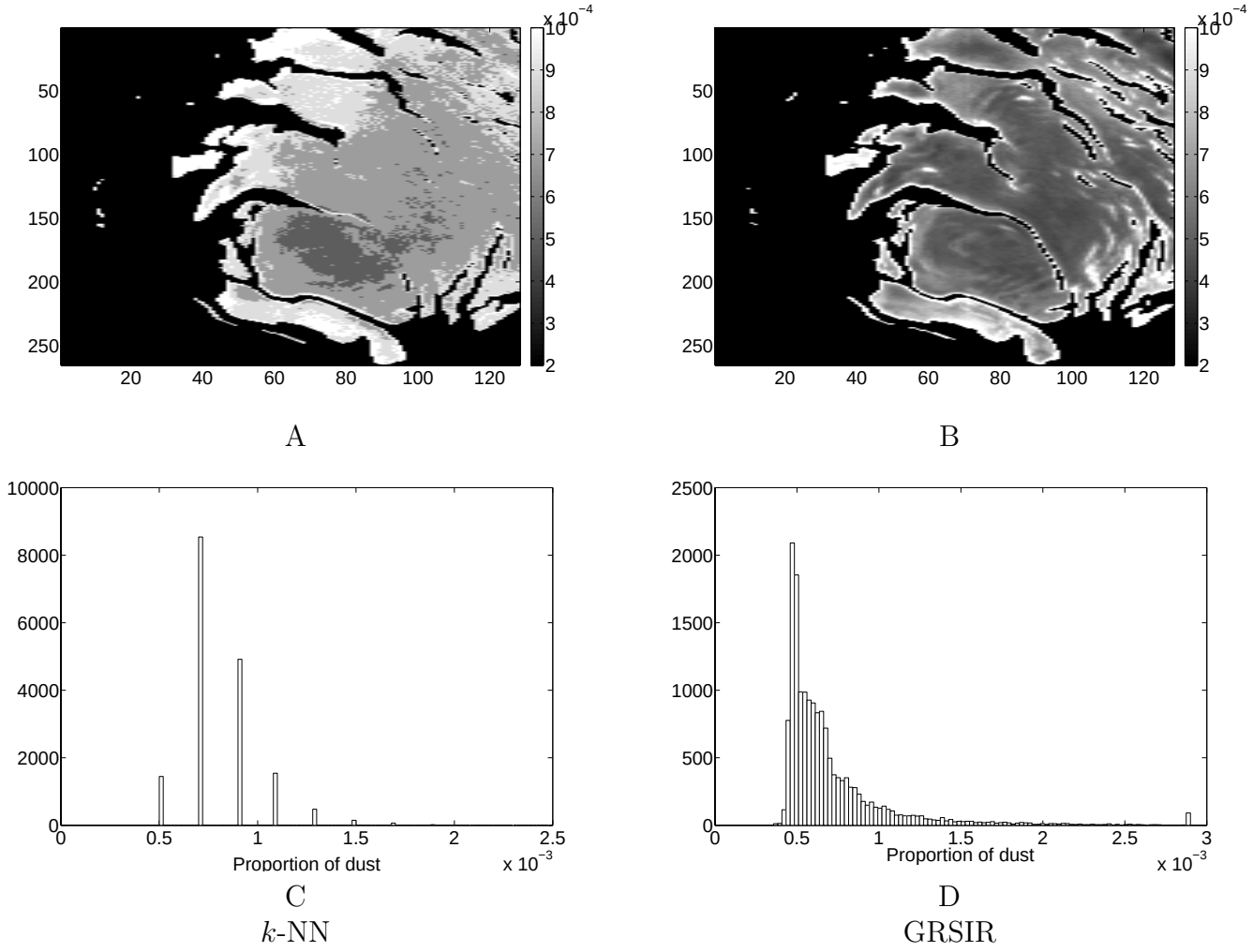


Figure 10. Proportion of dust estimated by k-NN (A) and GRSIR (B) from the hyperspectral image observed from orbit 41. The histograms of the estimations for the entire image are presented in (C) for k-NN and (D) for GRSIR.

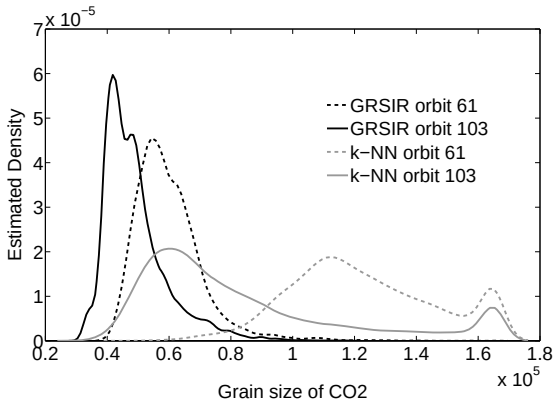


Figure 11. Comparison between the densities of the estimated grain sizes of CO_2 ice by k-NN and GRSIR in images from orbit 61 and 103.

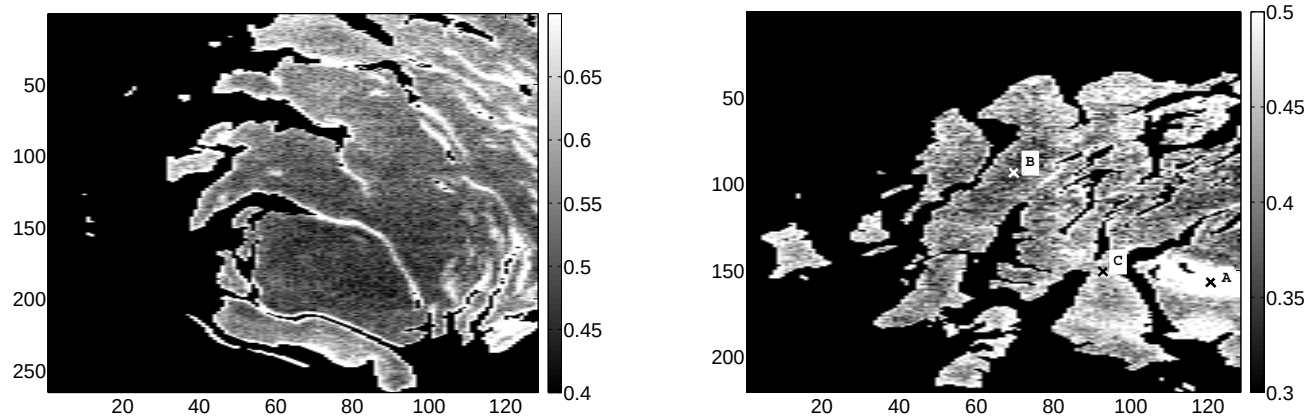


Figure 12. Cosinus of the Wavanglet angle between the observed spectra and a reference spectrum of martian dust for the hyperspectral images acquired during orbit 41 (A) and during orbit 103 (B).

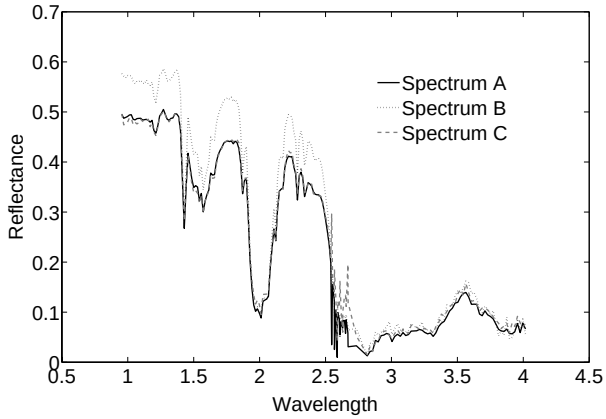


Figure 13. Spectra extracted from the locations labelled A, B and C in the image 103 of figure 12 (B)