



**HAL**  
open science

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

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

## ► To cite this version:

Caroline Bernard-Michel, Sylvain Douté, Mathieu Fauvel, Laurent Gardes, Stéphane Girard. Retrieval of Mars surface physical properties from OMEGA hyperspectral images using Regularized Sliced Inverse Regression.. [Research Report] 2008, pp.43. inria-00276116v2

**HAL Id: inria-00276116**

**<https://inria.hal.science/inria-00276116v2>**

Submitted on 6 Nov 2008

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

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

**1 Retrieval of Mars surface physical properties from**  
**2 OMEGA hyperspectral images using Regularized**  
**3 Sliced Inverse Regression**

C. Bernard-Michel<sup>1</sup>, S. Douté<sup>2</sup>, M. Fauvel<sup>1</sup>, L. Gardes<sup>1</sup>, and S. Girard<sup>1</sup>

---

C. Bernard-Michel, M. Fauvel, 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)

<sup>1</sup>MISTIS, INRIA Rhône-Alpes,  
Montbonnot, France.

<sup>2</sup>Laboratoire de Planétologie de Grenoble,  
Saint Martin-d'Hères, France.

4 **Abstract.** In this paper, a statistical method is proposed to evaluate the  
5 physical properties of surface materials on Mars from hyperspectral images  
6 collected by the OMEGA instrument aboard the Mars express spacecraft.  
7 The approach is based on the estimation of the functional relationship  $F$  be-  
8 tween some observed spectra and some physical parameters. To this end, a  
9 database of synthetic spectra is generated by a physical radiative transfer  
10 model and used to estimate  $F$ . The high dimension of spectra is reduced by  
11 Gaussian regularized sliced inverse regression (GRSIR) to overcome the curse  
12 of dimensionality and consequently the sensitivity of the inversion to noise  
13 (ill-conditioned problems). Compared with other approaches, GRSIR has the  
14 advantage of being very fast, interpretable and accurate.

## 1. Introduction

15 Visible and near infrared imaging spectroscopy is a key remote sensing technique used  
16 to study and monitor planets. It allows the detection, mapping and characterization of  
17 minerals, as well as volatile species, that often constitute the first step toward the res-  
18 olution of key climatic and geological issues [Murchie *et al.*, 2007; Bibring *et al.*, 2004a;  
19 Brown *et al.*, 2004; Carlson *et al.*, 1992]. These tasks are carried out through spectral  
20 analysis of solar light reflected by the material forming the top few millimeters or cen-  
21 timeters of the ground. Physical properties of the surface, such as chemical composition,  
22 granularity, texture, and physical state are some of the most important parameters that  
23 characterize the morphology of spectra. Modeling the direct link between these parame-  
24 ters and observable spectra is called the *forward problem* in classical physics. It can be  
25 evaluated numerically by radiative transfer models, simulating the propagation of solar  
26 light through the atmosphere and reflected back to the sensor [Douté *et al.*, 2007b; Hapke,  
27 2002; Shkuratov *et al.*, 1999; Douté and Schmitt, 1998; Hapke, 1993]. Such a model al-  
28 lows the simulation of spectra from given values of the model parameters. Conversely,  
29 deducing the physical model parameters from the observed spectra is called an *inverse*  
30 *problem*. Since it generally cannot be solved analytically, the use of optimization or statis-  
31 tical methods is necessary. Solving inverse problems requires an adequate understanding  
32 of the fundamental physics, so that a relation  $X = G(Y)$  may be specified between the  
33 spectra  $X$  and the parameters  $Y$ . Given  $G$ , different methods can be used to deduce the  
34 parameters  $Y$  from the observation  $X$ . Current solutions to the inverse problem can be

35 divided into three main categories (for further details and comparisons, see [*Kimes et al.*,  
36 2000; *Pragnère et al.*, 1999]):

37 1. *Optimization algorithms*: These methods minimize an objective function that ex-  
38 presses the quality of the estimations, and include numerical optimization techniques such  
39 as Powell’s method, Simplex method and the quasi-Newton method. Estimation can be  
40 unstable since inverse problems are often ill-posed (solutions are not unique, and a small  
41 change in the data can lead to large differences in the estimations). A probabilistic for-  
42 malism can be used to regularize inverse problems by introducing a prior distribution on  
43 model parameters [*Aster et al.*, 2005; *Tarantola*, 2005]. These approaches are computa-  
44 tionally expensive since they independently invert new spectra. Therefore, they cannot  
45 be used to invert an image with several hundred thousand pixels. Moreover, they can  
46 sometimes fall into local minima if the objective function is not convex.

47 2. *Look-up table (LUT) / k-nearest neighbors approach (k-NN)*: This methodology is  
48 currently used by physicists to study planetary bodies [*Carlson et al.*, 2005; *Philpot et al.*,  
49 2004; *Weiss et al.*, 2000; *Douté et al.*, 2001]. The previous heavy runtime computation  
50 is replaced by a simpler look-up operation. A large database (LUT) is generated by  
51 radiative transfer for many parameter values. To invert an hyperspectral image, the  
52 pixel’s spectrum is then compared with the LUT spectra in order to find the best match  
53 (the nearest neighbor), according to an objective function minimization. Parameters  
54 are then deduced from this best match. The speed gain is significant in comparison to  
55 traditional optimization methods, since retrieving a value from memory is often faster  
56 than undergoing an expensive computation. The main disadvantages of this approach are  
57 the multiplicity of solutions and their instability [*Bernard-Michel et al.*, 2007].

3. *Training approaches*: They have been recently considered in the study of terrestrial vegetation. A functional relationship

$$Y = F(X) \tag{1}$$

between spectra and parameters is assumed. This relationship corresponds to the inverse of the physical model  $G$  in the forward problem. The idea is to use a LUT to estimate the underlying mathematical relationship  $F$ . This relationship then allow us to estimate the parameters of new spectra. The advantage of 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. Among training approaches, neural networks [*Hastie et al.*, 2003, chapter 11] or support vector machines (SVM) [*Hastie et al.*, 2003, chapter 12] seem promising but the underlying learning process remains time consuming [*Combal et al.*, 2002; *Durbha et al.*, 2007; *Pragnère et al.*, 1999].

Hyperspectral images on planets must be inverted with the following constraints: 1. Working with large data-sets and various models require fast methodologies, 2. When dealing with very high-dimensional data, one is faced with the ‘curse of dimensionality’ and the associated sparsity issues and, 3. Observed spectra always contain some noise.

In this paper, a new training approach is proposed: The Gaussian Regularized Sliced Inverse Regression (GRSIR)[*Bernard-Michel et al.*, to appear, 2008]. It is based on a dimension reduction technique first proposed by [*Li*, 1991] and similar to Partial Least Squares regression (PLS), [*Hastie et al.*, 2003, chapter 3]. It has the advantage of being 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 with  $k$ -NN, PLS and SVM.

78 In Section 2, real and simulated data-sets are presented, then the principle of GRSIR  
79 is recalled in Section 3. The choice of an appropriate LUT is discussed in Section 4.  
80 Experimental results are presented on simulated data-sets in Section 5 and on real Mars  
81 images in Section 6. Conclusions are drawn in Section 7.

## 2. Data

82 The data-sets used in the next sections can be divided in three categories: The hy-  
83 perspectral images observed on Mars, a LUT simulated by radiative transfer algorithms  
84 according to some physical modeling of these images, and test data in order to quantify  
85 the according of the estimation. All data-sets consist of some spectra  $X \in \mathbb{R}^d$  with as-  
86 sociated parameters  $Y \in \mathbb{R}^p$ ,  $d$  being the number of wavelengths and  $p$  the number of  
87 parameters.

### 2.1. Hyperspectral images from Mars

88 The data-sets were collected by the imaging spectrometer OMEGA (Observatoire pour  
89 la Minéralogie, l'Eau, la Glace et l'Activité) [Bibring et al., 2004a]. OMEGA is one of the  
90 seven scientific instruments aboard the European spacecraft Mars Express mission, sent to  
91 orbit Mars in 2003. It was developed by IAS and LESIA (Observatoire de Paris) with the  
92 support of CNES, and with the participation of IFSI (Italy) and IKI (Russia). This visible  
93 and infrared instrument can scan most of Mars from orbit in order to observe gas and  
94 dust in the atmosphere and look for signs of specific materials such as silicates, carbonates  
95 and ice at the surface. It records the visible and infrared light reflected from the planet  
96 in the 0.5-5.2 micron wavelength range and with a ground resolution varying from 350 m  
97 to 10 km. Three OMEGA hyperspectral images acquired during orbits 41, 61 and 103

98 will be used. They cover the high southern latitudes of Mars. The spatial resolution is  
99 approximately 2km per pixel and 184 wavelengths have been considered in the range 0.95-  
100 4.15 (*i.e.*  $d = 184$ ). For each spectrum of the image, the atmospheric gaseous contribution  
101 has been removed [Douté *et al.*, 2007b]. After a first analysis, these OMEGA observations  
102 revealed that the south polar region of Mars exhibits mainly water ice, carbon dioxide ice  
103 and dust at the surface [Bibring *et al.*, 2004b]. A detailed qualitative mapping of the ice  
104 during the local summer shows that the permanent bright cap is dominated by superficial  
105 CO<sub>2</sub> ice. Nevertheless, its spectral signature is slightly contaminated by the contribution  
106 of dust and water ice. Among the possible coexistence modes of the latter components,  
107 a granular mixture of H<sub>2</sub>O, CO<sub>2</sub> ice and dust is the most appropriate to reproduce the  
108 morphology of the spectra [Douté *et al.*, 2007a]. This conclusion is based on numerical  
109 experiments that aimed at reproducing the spectra by different surface reflectance models.  
110 The permanent polar cap has been mapped by a classification method based on wavelets  
111 [Schmidt *et al.*, 2007]. For each image, the CO<sub>2</sub> ice areas contain about 10000 to 20000  
112 spectra. The selected model for the polar cap considers the intimate mixture of H<sub>2</sub>O,  
113 CO<sub>2</sub> ice and dust as an optically thick parallel layer without substantial roughness. The  
114 transfer of solar photons through such a material is calculated by a modified version of the  
115 Hapke semi-empirical formulation [Hapke, 2002; Douté and Schmitt, 1998; Hapke, 1993].

## 2.2. Look-up table (LUT)

116 The model implies five spatially varying parameters (*i.e.*  $p = 5$ ): the grain size of  
117 water and CO<sub>2</sub> ice, the proportion of water ice, CO<sub>2</sub> ice and dust. The proportions are  
118 normalized to sum to one. The other parameters involved in the physical model, such as  
119 the incidence and emergence angles or the grain size of dust have been fixed to a constant.



120 Since *a priori* knowledge on the parameters range is unavailable, the simulation of a large  
 121 look-table (LUT) is necessary. Details of the sampling strategy are provided in Table 1.  
 122 The notation used in this article is as follows:  $n$  is the number of simulated spectra  
 123 ( $n = 31500$ ),  $x_i \in \mathbb{R}^{184}$  denotes a spectrum from the LUT and  $y_i \in \mathbb{R}^5$  its associated  
 124 parameters, where  $i \in \{1, \dots, n\}$ .

### 2.3. Test data

125 For validation and comparison, the use of a test data-set (Tdata) is required. Since  
 126 no ground truth is currently available for the physical properties of Mars polar regions,  
 127 we can only rely on synthetic data. The same physical model is used to simulate the  
 128 LUT and Tdata. The range of variation for the parameter is detailed in Table 1. To  
 129 emulate real data, a zero mean multiGaussian noise of dimension 184 is added to all the  
 130 spectra of the test data-set. The covariance matrix is determined experimentally from  
 131 a small, spatially homogeneous, portion of the real image. Assuming that much of the  
 132 variability comes from the noise, the latter is then evaluated using statistics based on a  
 133 shift difference on the selected portion. As in the previous section,  $n_T$  is the number of  
 134 simulated test spectra ( $n_T = 3500$ ),  $x_i^T$  is a spectrum from Tdata and  $y_i^T$  its associated  
 135 parameters, where  $i \in \{1, \dots, n_T\}$ .

## 3. Proposed approach

136 Our goal is to estimate the functional relationship  $F$  between the spectra  $X \in \mathbb{R}^d$   
 137 and each physical parameter  $Y^j = F(X) \in \mathbb{R}$  with  $j \in \{1, \dots, p\}$ . In this work, each  
 138 parameter is studied individually, therefore  $p = 1$  and  $j$  will be omitted for the sake of  
 139 simplicity. The originality of the methodology is to split the difficult problem of learning

140 a  $d$ -variate function into two sub-problems that can be easily solved. First, a dimension  
 141 reduction technique is applied (Section 3.1). Then, estimation is performed in the lower  
 142 dimensional subspace (Section 3.2).

### 3.1. Dimension reduction step: Gaussian regularized sliced inverse regression

143 Dimension reduction methods rely on the assumption that the predictor  $X$  can be  
 144 replaced, without loss of information, by its projection onto a subspace of smaller dimen-  
 145 sion  $L$ , called the effective dimension reduction space (EDR). The basis of the subspace  
 146 is denoted  $\beta_1, \dots, \beta_L$  and the functional relationship  $Y = F(X)$  can be rewritten as  
 147  $Y = f(\beta_1^t X, \dots, \beta_L^t X)$ , where  $f$  is now a  $L$ -variate function. In most applications,  $L$  is  
 148 smaller than 3 making the estimation of  $f$  tractable.

Principal Component Analysis (PCA) is a classical approach to reduce the dimen-  
 sion [Hastie et al., 2003, chapter 14]. The basic concept of PCA is to define the orthogonal  
 projection of the spectra onto a lower dimensional linear space such that the total vari-  
 ance of the projected spectra is maximized. It amounts to diagonalization of the spectra  
 covariance matrix

$$\Sigma = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^t \text{ where } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad (2)$$

149 or equivalently to calculation of the eigenvectors of  $\Sigma$ . However, in regression problems  
 150 PCA is generally not satisfactory since only the explanatory variable  $X$  is considered  
 151 while the dependent variable  $Y$  is not taken into account. Specific dimension reduction  
 152 techniques have been developed for regression problems, such as PLS and Sliced Inverse  
 153 Regression (SIR) [Li, 1991]. SIR consists of sorting the parameter values in increasing  
 154 order and dividing them into  $H$  non-overlapping slices  $S_h, h \in \{1, \dots, H\}$ . For each slice

155  $S_h$ , a mean spectrum  $m_h$  is calculated as the mean value of the parameter values that  
 156 are varying in the slice. The orthogonal projection of the mean spectra onto a lower  
 157 dimensional linear space is defined so that the variance of the projected mean spectra is  
 158 maximized under the constraint that the total projected variance is one. This is equivalent  
 159 to diagonalization of  $\Sigma^{-1}\Gamma$ , where  $\Gamma$  is the mean spectra covariance matrix given in  
 160 equation (3). The SIR principle is illustrated in Figure 1.

161 Since inverse problems are generally ill-posed [*Aster et al.*, 2005; *Tarantola*, 2005],  $\Sigma$   
 162 is ill-conditioned, making its inversion difficult. To solve this problem, it is proposed to  
 163 compute a Gaussian Regularized version of Sliced Inverse Regression (GRSIR). Theoret-  
 164 ical foundations can be found in [*Bernard-Michel et al.*, to appear, 2008]. The concept of  
 165 this method is to include some prior information on the projections in order to dampen  
 166 the effect of noise in the input data. The ill-posed problem is then replaced by a slightly  
 167 perturbed well-posed problem that depends on a regularization parameter  $\delta$ . There-  
 168 fore, GRSIR computes the  $L$  eigenvectors corresponding to the  $L$  largest eigenvalues of  
 169  $(\Sigma^2 + \delta I_p)^{-1}\Sigma\Gamma$  where  $I_p$  is the  $d \times d$  identity matrix. In practice, GRSIR then requires  
 170 the three following computational steps:

171 **Step 1:** Sort  $y_i, i \in \{1, \dots, n\}$  into increasing order and divide into  $H$  non-overlapping  
 172 slices  $S_h, h = 1, \dots, H$ . If the LUT has been simulated for random values of  $Y$ , then  
 173 slices are chosen such that each slice contains the same number of observations. If the  
 174 LUT has been simulated for a fixed number of distinct parameters, the slices are chosen  
 175 to coincide with these discrete values.

**Step 2:** Compute the "between slices" covariance matrix of the means:

$$\Gamma = \sum_{h=1}^H \frac{n_h}{n} (m_h - \bar{x})(m_h - \bar{x})^t \text{ with } m_h = \frac{1}{n_h} \sum_{y_i \in S_h} x_i, \quad (3)$$

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

**Step 3:** Estimate the GRSIR axes  $\beta_1(\delta), \dots, \beta_L(\delta)$  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 5.4, a criterion to choose the dimension  $L$  is given. Interestingly, the first axis can be viewed as a weighted function of the wavelengths, giving some knowledge about the wavelengths that carry information on the parameter (Section 5.5). The projections of the spectra on the axis  $\beta_1(\delta)$  is termed "reduced spectra".

### 3.2. Estimation of the functional relationship

Once the spectra have been reduced, the relationship  $f$  has to be estimated. In the proposed approach, a piecewise linear interpolation is performed on the set of data points  $(m_h^{proj}, m_h^{param})$ ,  $h = 1, \dots, H$ , where  $m_h^{proj} = \langle x_i, \beta_1(\delta) \rangle$  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  $t = \langle \beta_1(\delta), x \rangle$ , the estimated parameter value  $\hat{y}$  is then given by:

$$\hat{y} = \begin{cases} m_1^{param} & \text{if } t \in ]-\infty, m_1^{proj}] \\ m_h^{param} + (t - m_h^{proj}) \left( \frac{m_{h+1}^{param} - m_h^{param}}{m_{h+1}^{proj} - m_h^{proj}} \right) & \text{if } t \in ]m_h^{proj}, m_{h+1}^{proj}] \\ m_H^{param} & \text{if } t \in ]m_H^{proj}, +\infty[ \end{cases}, \quad h = 1, \dots, H \quad (4)$$

190 An example of the application of GRSIR to the LUT is given in Figure 1. It shows the  
 191 relationship between reduced spectra and the grain size of CO<sub>2</sub> ice and its estimation by  
 192 a piecewise linear interpolation.

#### 4. Choice of the Look-up table

193 In the simulated LUT, many spectra could be superfluous to the proper estimation of the  
 194 parameters. For instance, considering Tdata, spectra with parameter values not included  
 195 in the Tdata's parameter range should be removed from the LUT. Better estimations  
 196 can be obtained with the reduced LUT [*Bernard-Michel et al., 2007*]. The difficulty is  
 197 that, in practical cases, the range of variation of the parameters is unknown. A good  
 198 strategy is to assumed a large LUT and to reduce it after the first estimation. However,  
 199 the effectiveness of this technique in practical situations depends strongly on the quality  
 200 of the first estimation.

201 In this paper, another approach is proposed that makes use of PCA to visualize the  
 202 adequacy of the chosen LUT for a given set of observed spectra. The parameters of the  
 203 observed spectra can be estimated only if their projections onto the first  $m$  PCA axes  
 204 (deduced from the application of PCA to the LUT), coincides with the projection of the  
 205 LUT itself. Conversely, to select the most appropriate spectra from the LUT for inversion,  
 206 the strategy is to retain spectra from the LUT whose projections onto the PCA axes are  
 207 close to projections of the spectra from the observed image (see Figure 2.(A)). In the space  
 208 spanned by the first  $m$  PCA axes, the distance between each projected LUT spectrum  
 209 and its nearest neighbor from the projected spectra is computed. The histogram of these  
 210 distances is considered as a mixture of  $n_c$  Gaussian densities  $\mathcal{N}(\mu_i, \Sigma_i)$  characterized by  
 211 the proportions of the mixture  $\pi_i, i \in \{1, \dots, n_c\}$ , the expectation  $\mu_i$  and the covariance

212 matrix  $\Sigma_i$  of each class  $i, i \in \{1, \dots, n_c\}$ . The number of classes is chosen by the user, and  
213 the parameters can be estimated by the Expectation-Maximization algorithm (EM) after  
214 random or by the K-means initialization (for further details about mixture models and  
215 EM, see *Hastie et al.* [see 2003, Sections 6.8 and 8.5]). The algorithm is a simple iterative  
216 optimization process for computing the maximum likelihood estimate of the parameters.  
217 One can then calculate and maximize the posterior probability that a spectrum belongs to  
218 a particular class by observing the aforementioned distances. This leads to a classification  
219 of the spectra into  $n_c$  classes: The class of LUT spectra that are far from the observed  
220 spectra, and those that are close. An example is given in Figure 2.(C) with the image  
221 observed from orbit 41. The histogram allows us to distinguish 3 classes. The third class  
222 corresponds well to spectra that belongs to both LUT and observed data.

223 Note that in Figure 2.(B) some of the observed spectra are outside the projected LUT.  
224 Since the chosen physical model is not relevant to them, these spectra are removed from the  
225 observed data. Hence, in the PCA space, the histogram of the distance of each observed  
226 spectrum with its nearest neighbor in the selected sub-LUT facilitates the distinction  
227 of two classes: The invertible spectra and the non invertible ones (Figure 2.(D)). The  
228 non invertible spectra from orbit 41 correspond to pixels at the boundary of the CO<sub>2</sub>  
229 bright area and thus to pixels with a linear geographical mixture of terrains. Thus, the  
230 physical model used to simulate the data is not valid. Other examples can be found in  
231 the work of *Bernard-Michel et al.* [2007]. Experimental results in the next section show  
232 that estimations are more accurate when applying such a selection.

## 5. Analyzing results on simulations

233 In this section, the competing methods are first presented. Then, the validation criteria  
234 and the parameters selection are detailed. Finally results on simulated data-sets are given.

### 5.1. Competing methods

235 ***k*-NN:** From the LUT one finds the  $k$  nearest spectra and fixes the estimated  $y$  as the  
236 mean parameter value of  $k$  nearest spectra parameter. The distance between two spectra  
237 is taken as the Euclidean distance:  $\|o - x_i\|^2$ , where  $o$  is a spectrum from the image and  
238  $x_i$  a spectrum from the LUT. In the experiments,  $k$  is fixed to one, so only one neighbor  
239 is used for the estimation.

240 **Partial Least Squares regression:** The PLS method is closely related to PCA (or-  
241 thogonal projection onto lower dimensional space) and GRSIR ( $y$  is accounted for). PLS  
242 searches for the projection of the explanatory variable  $x$  onto a lower dimensional space  
243 that maximizes the covariance between  $x$  and  $y$ :  $\max_{\beta} \text{cov}^2(y, \beta^t x)$ . PLS is parametrized  
244 by the size  $L$  of the subspace spanned by  $\beta$ . The regression in the subspace is necessarily  
245 linear and may limit the efficiency of the method if a non linear relationship exists between  
246  $x$  and  $y$ .

247 **Support Vector Machines regression:** SVM approximates the functional  $F$ :  
248  $y = F(x)$  using a solution of the form  $F(x) = \sum_{i=1}^n \alpha_i K(x, x_i) + b$ , where  $x_i$  are sam-  
249 ples from the training set,  $K$  a kernel function and  $\left( (\alpha_i)_{i=1}^n, b \right)$  are the parameters of  
250  $F$  which are estimated during the training process [Scholkopf and Smola, 2002]. The  
251 kernel  $K$  is used to produce a non-linear function. One widely used kernel is the *Gaus-*  
252 *sian Kernel*:  $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$ . The SVM training entails minimization of  
253  $\left[ \frac{1}{n} \sum_{i=1}^{\ell} l(F(x_i), y_i) + \lambda \|F\|^2 \right]$  with respect to  $\left( (\alpha_i)_{i=1}^n, b \right)$ , and with  $l(F(x), y) = 0$  if

254  $|F(x) - y| \leq \epsilon$  and  $|F(x) - y| > \epsilon$  otherwise. Prior to running the algorithm, the following  
 255 parameters need to be fitted:  $\epsilon$  which controls the resolution of the estimation,  $\lambda$  which  
 256 controls the smoothness of the solution and the kernel parameters ( $\gamma$  for the Gaussian  
 257 kernel).

## 5.2. Validation criteria

258 To assess the relevance of GRSIR methodology, two aspects of the estimation were  
 259 investigated: The accuracy of the estimates and the quality of the relationship between  
 260 the reduced spectra and the parameters. To this end, two validation criteria were used

1. The Normalized Root Mean Square Errors (NRMSE):

$$NRMSE = \sqrt{\frac{\sum_{i=1}^{n_T} (\hat{y}_i^T - y_i^T)^2}{\sum_{i=1}^{n_T} (y_i^T - \bar{y}^T)^2}} \text{ with } \bar{y}^T = \frac{1}{n_T} \sum_{i=1}^{n_T} y_i^T \quad (5)$$

261 The NRMSE quantifies the difference between the estimations  $\hat{y}_i^T$  and the real values  
 262  $y_i^T$ . This measure is normalized enabling direct comparisons between several parameter  
 263 estimations. The closer NRMSE is to zero, the more accurate are the predicted values.

2. The SIR Criterion (SIRC)

$$SIRC = \frac{\beta_\ell^t(\delta)\Gamma\beta_\ell(\delta)}{\beta_\ell^t(\delta)\Sigma\beta_\ell(\delta)} \quad (6)$$

264 The SIRC is the ratio between the "between-slices" variance  $\beta_\ell^t(\delta)\Gamma\beta_\ell(\delta)$  of the projections  
 265 of  $x_i, i \in \{1, \dots, n\}$  on  $\beta_\ell(\delta)$  and the total variance  $\beta_\ell^t(\delta)\Sigma\beta_\ell(\delta)$  of these same projections.  
 266 It quantifies the quality of the relationship between projected spectra and parameters.  
 267 The closer SIRC is to 1, the better is the relationship.



### 5.3. Choice of the parameters

For GRSIR, the quality of the estimation  $\hat{y}_i^T$  depends on a regularization parameter  $\delta$  (see Section 3.1). When  $\delta$  increases, the functional relationship between projected spectra and parameters gets worse and consequently estimation errors increase. On the contrary, if  $\delta$  is too small, then the estimation errors are considerable in the presence of noise because the problem is ill-posed. So, the choice of  $\delta$  dictates to a compromise between improving the functional relationship and increasing estimation accuracies by regularization. In this article,  $\delta$  is chosen for each parameter individually, the NRMSE criterion calculated between the parameter values from a validation data and their estimations is minimized. The validation data is the look-up table itself, perturbed by multiGaussian noise representative of that affecting the OMEGA hyperspectral images (see Section 2.3). It is noted that, if there is no noise in the data, *i.e.* 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 yields a value  $\delta$  close to zero.

For  $k$ -NN, since  $k$  is fixed to one, no parameter tuning is required. For PLS, the dimension  $L$  of the subspace is chosen to minimize the NRMSE on the validation set. For SVM, three parameters are fixed.  $\epsilon$  is fixed to 0.01, while  $\lambda$  and  $\gamma$  are selected to minimize the NRMSE on the validation set. As it is typically done with SVM, the range of each component of  $Y$  has been stretched between 0 and 1 for the regression.

### 5.4. Choice of the EDR dimension

In order to choose the EDR dimension, the SIRC is calculated for each GRSIR axis  $\beta_\ell(\delta)$  with  $\ell \in \{1, \dots, d\}$ . As  $\ell$  increases, the SIRC decreases rapidly and is almost zero after a few iterations. For instance, when applying GRSIR to the LUT, the SIRC for the

289 proportion of CO<sub>2</sub> ice on the first GRSIR axis is 0.985. It indicates that the relationship  
290 of interest is captured by the first axis. Calculated the second axis, the SIRC drops to  
291 0.1799 and then is close to zero for other axes. In this case, one can consider incorporating  
292 these axes in the regression step does not bring extra information and is unnecessary, and  
293 finally fixes the EDR dimension to one. Analyzing the SIRC over all parameters and two  
294 different physical models reveals that in most situations only one dimension is necessary  
295 to retrieve the model parameters. Hence, in this article  $L$  is fitted to one.

### 5.5. Analyzing the GRSIR axis

296 The spectral variability of the  $x_i$  constituting the LUT arises from a complex interplay  
297 of the different input parameters  $y_i$  of the model. In particular, two parameters can have  
298 a similar influence on the spectra for a certain range of values. Consequently, it is crucial  
299 to understand how the GRSIR method can untangle the dependencies in order to find an  
300 independent one-to-one functional relationship between spectra and parameters. Accord-  
301 ingly, attempt to correlate each axis  $\beta_1(\delta)$  (which is a vector of spectral weights) with the  
302 variability induced by the variation of the corresponding physical parameter. The other  
303 parameters are kept constant at mean values. The weights are individually applied to  
304 different levels of reflectance, depending on the associated wavelength, to calculate the  
305 reduced spectrum  $\beta_1^t(\delta)x_i$ . Hence, we find it more illustrative to represent the compo-  
306 nent wise multiplication (wavelength by wavelength) of each axis with a representative  
307 spectrum. An illustration is given in Figure 3 for the grain size of CO<sub>2</sub> ice. Our analysis  
308 shows that the axis for H<sub>2</sub>O abundance, dust abundance, and grain size of CO<sub>2</sub> ice are  
309 quite similar. A strong weight is assigned to 1.43 micron wavelength (the bottom of a  
310 very reliable and narrow CO<sub>2</sub> ice absorption band) and, to a lesser extent, to the interval

311 between 1.47 and 1.54 microns ( $\text{H}_2\text{O}$  ice feature), and wavelengths of 1.87 and 2.29 mi-  
312 crons (weak  $\text{CO}_2$  ice bands but very distinct). The three axes also show differences that  
313 allow differentiation between the functional relationships. A relatively strong weight at  
314 2.38 microns on the right wing of the 2.34 micron band is a specificity of the proportion  
315 of water ice. The axis for the proportion of dust stresses very much the variation of the  
316 spectrum continuum around 1 micron and 1.77 micron that indeed strongly changes with  
317 dust concentration. No other parameter affects the continuum as much. The axis for the  
318 grain size of  $\text{CO}_2$  ice maximizes the influence of the small plateau at 2.62 microns that  
319 varies very much with the latter parameter. The axis that is linked to  $\text{H}_2\text{O}$  grain size  
320 is unique in the fact that it emphasizes, on the one hand, the continuum level around 1  
321 micron, but not around 1.75 microns and, on the other hand, the entire spectral range  
322 between 2.35 and 2.65 microns. The main conclusions that can be drawn from this study  
323 are that, at least for the model presented in this paper, the GRSIR method has the abil-  
324 ity to find a unique set of wavelengths where the variability of the spectrum is the most  
325 pertinent (but not necessary the highest) for the evaluation of a given parameter. One  
326 must note that these key wavelengths not only fall on the bottom of specific absorption  
327 bands for both  $\text{H}_2\text{O}$  and  $\text{CO}_2$  ices but also on specific parts of the continuum as well as  
328 on specific band wings.

## 5.6. Results

329 In this section, the experimental results are compared in terms of the NRMSE and  
330 SIRC (Table 2), and also with scatter plots of real and estimated parameters (Figure 4).  
331 Except for  $k$ -NN, a spectra selection in the look-up table was performed before applying  
332 the different algorithms, as mentioned in Section 4.

333 From Table 2, it is clear that the worst results are obtained with  $k$ -NN. GRSIR outper-  
334 forms  $k$ -NN and PLS for most parameters. Problems with the estimation of the grain size  
335 of water ice may be due to the fact that not enough values are considered for simulations  
336 in the LUT (only 5). SVM provides the best estimation in terms of the NRMSE, but with  
337 a dramatic increase in processing time: Approximately 15 hours against about 1 minute  
338 versus  $k$ -NN or GRSIR.

339 Figure 4 shows the scatter plot of the estimated proportions of CO<sub>2</sub> ice with each method  
340 versus the real proportions. It appears that the range of the estimated proportions of  
341 CO<sub>2</sub> ice is much larger with  $k$ -NN than with GRSIR, PLS and SVM. PLS's scatter plot  
342 is slightly curved: PLS could not handle the non linear relationship between  $X$  and  $Y$ ,  
343 while GRSIR was able to model it.

344 It is interesting to point out that GRSIR, as a first step, can determine the ranges the  
345 parameters of an observed image are varying. Then, a more appropriate LUT can be  
346 built in the estimated ranges with a higher density of parameter values. Applying GRSIR  
347 with this new extracted LUT leads to more accurate estimations comparable to SVM, see  
348 Table 3. Finally, the SIRC, always close to 1, is very satisfying. The PLS benefits from  
349 the new LUT while the SVM does not. However, for the SVM, the processing time is  
350 decreased since the number of samples in the LUT is reduced.

## 6. Retrieval of the physical parameters for the south polar cap of Mars

351 This section summarizes the first model inversions obtained by the different algorithms  
352 for hyperspectral images acquired by OMEGA during orbits 61, 103 and 41 (see Section  
353 2). The study of these images leads to 30 different maps. All the maps can be found  
354 in [Bernard-Michel et al., 2007]. For brevity, only some are reported in the article.

355 The model inversion on the image observed from orbit 103 by GRSIR, PLS, SVM and  $k$ -  
356 NN shows that GRSIR and SVM give very smooth mappings for all sets of parameters. An  
357 example is given in Figure 5, where it appears that the proportion of dust is nearly always  
358 estimated at 0.0003 with  $k$ -NN, whereas with GRSIR and SVM, the map is more detailed.  
359 Some problems are encountered, with PLS: Negative values are estimated, making the map  
360 difficult to interpret.

361 Figure 6 shows the proportion of dust estimated by the different algorithms on the  
362 portion of the polar cap observed during orbit 41. With  $k$ -NN, estimations now assume 8  
363 different values and seem to indicate that, at the very center of the bright cap, little dust is  
364 observed. The abundance of dust increases significantly with proximity to the boundaries.  
365 The estimated map with GRSIR is more detailed and leads to slightly different conclusions.  
366 In particular, the area presenting a poor proportion of dust is more extended than with  
367  $k$ -NN. Nevertheless if, globally, maps are much smoother and detailed with GRSIR, they  
368 never differ entirely from  $k$ -NN's.

369 Another interesting remark concerns the estimation of parameters in images 61 and  
370 103 that represent approximately the same portion of the polar cap. The analysis should  
371 consequently give close estimations for each studied parameter regardless of the method.

372 Figure 7 shows the distribution of CO<sub>2</sub> grain size values obtained respectively by GRSIR,  
373 PLS on the left side (A) and SVM,  $k$ -NN on the right side (B) for both observations.

374 Figure 8 focuses on the distribution of the estimated proportion of water and dust and  
375 on the distribution of the estimated grain size of water and CO<sub>2</sub> with GRSIR and SVM.

376 First we consider the evolution of these histograms between observations 61 and 103 for  
377 a given method. For CO<sub>2</sub> grain size and, to a lesser extent for H<sub>2</sub>O grain size, a shift in

378 the distribution of values can be noted that is increasing when using GRSIR, then SVM,  
379 and finally  $k$ -NN. The histograms of abundances for H<sub>2</sub>O ice and dust are relatively stable  
380 between 61 and 103 with SVM and GRSIR. The abundance for CO<sub>2</sub> ice would present a  
381 similar behavior since it is fully constrained (sum of abundances equals 1).

382 Second we note that the distributions of parameter values obtained by SVM are always  
383 shifted between to observation. The shift is always noticeable, sometimes considerable,  
384 and in addition systematically positive.

385 We explain these facts as follows. Observations 61 and 103 were acquired approximately  
386 12 Mars days apart during a period of the year when the solar illumination was declining  
387 over the south pole. The conjugated effect of an increasingly grazing illumination over a  
388 slightly dusty atmosphere diminishes the spectral contrast of the observations, *i.e.* the  
389 overall level of reflectance of the spectra and on the CO<sub>2</sub> ice bands intensity and shape.  
390 This is very much comparable to the introduction of noise in a relative sense between  
391 observations 61 and 103. Since our remote sensing problem is partly ill-conditioned, a  
392 non regularized inversion method such as the  $k$ -NN will be much more sensitive to this  
393 noise than a regularized one such as the GRSIR or SVM. Even though regularization  
394 is thus applied to SVM, we note degraded performance when compared to GRSIR. For  
395 both methods, the estimation of the optimal value for the regularization parameter  $\delta$  or  
396  $\lambda$  is sensitive to the noise statistics applied to Tdata. In turn, this value determines the  
397 estimation of  $Y$ . As a consequence, uncertainties on the noise statistics propagate to bias  
398 on the estimation. We suspect that the choice of  $\lambda$  is less robust in the case of SVM than  
399 the choice of  $\delta$  in the case of GRSIR.

400 Finally, for the sake of validation, the estimated maps are compared with the Wavan-  
401 glet approach developed in [*Schmidt et al.*, 2007]. "Wavanglet" is a supervised automatic  
402 detection method that identifies in hyperspectral images spectral features and thus pro-  
403 duces distribution maps of chemical compounds. It uses three steps: 1. Selection of a  
404 library composed of reference spectra (the signature of the compounds to be detected); 2.  
405 Application of a Daubechies wavelet transform to referenced spectra and determination  
406 of the wavelet subspace that best separates all referenced spectra; 3. In this selected  
407 subspace, calculation of the spectral angle between each spectrum of an observation and  
408 a given reference spectrum. In particular, this angle called the wavanglet angle, allows to  
409 quantify in a relative sense the spatial variations of the different compound abundances  
410 at the surface. The cosine of the wavanglet angle between each spectrum of the images  
411 and a reference spectrum of martian dust is presented Figure 9 . The closer it is to one,  
412 the greater the dust proportion is. A similar map is observed for the water proportion.  
413 Globally, estimates of the latter quantities are more noisy with wavanglet than with GR-  
414 SIR and k-NN, especially for the dust proportion but they generally are in agreement  
415 with some exceptions. For example, in the image observed during orbit 103, *k*-NN, SVM,  
416 PLS and wavanglet methodologies display an area with strong proportion of dust in the  
417 lower right part of Figure 9(B), that is absent with GRSIR. In order to check if this area  
418 really contains more dust than other areas, the spectrum denoted by A in Figure 9 is  
419 selected, corresponding to the greatest proportion of dust, and compared to two other  
420 spectra from the cap. The first spectrum (B) has been chosen in a pure CO<sub>2</sub> area with  
421 very few dust. The second one (C) has been chosen in a area containing dust. According  
422 to *k*-NN, SVM, PLS and wavanglet results, this area should however contain less dust

423 than the area containing spectrum A. The spectra A, B and C are presented in Figure 10.  
424 One can see that, as expected, the spectrum B is really different from the spectrum A.  
425 On the other hand, spectra A and C are really similar showing that the area of interest  
426 does not contain as much dust as the wavanglet methodology predicts. On the contrary,  
427 estimations given by GRSIR are more coherent with this spectral analysis.

## 7. Conclusion

428 In this paper, a regularized version of Sliced Inverse Regression has been proposed to  
429 retrieve the physical parameters that best explain the spectra observed on Mars by the  
430 OMEGA imaging-spectrometer.

431 Results on simulations are promising, showing that estimations are accurate and most  
432 of the time better than the ones given by the  $k$ -nearest neighbors algorithm or PLS and  
433 close to those given by SVM. On a real data, maps are much smoother than with  $k$ -NN  
434 or PLS and seem to give a coherent mapping by comparison to the inversion of different  
435 hyperspectral images of the same portion of surface of Mars. If best results in terms  
436 of the NRMSE were obtained with SVM, the latter presents two drawbacks. First, the  
437 choice of the regularization parameter is less robust and, second, the processing time is  
438 900 times more than GRSIR, thus making GRSIR approach more suitable for practical  
439 situations. The inversion of each new observed spectrum is really fast as well as the  
440 selection of regularization parameter. The main limitation of the proposed approach is  
441 that currently no uncertainties of the estimations are given when inverting a real image.  
442 Experimental uncertainties could be computed based on simulations, but it supposes that  
443 the noise in the spectra has been well evaluated. If not, uncertainties will probably be  
444 underestimated. Some improvements could also be proposed to choose the regularization



445 parameter and a more complete analysis of the influence of the noise in the GRSIR  
446 methodology would be interesting. Finally, the development of a multivariate regularized  
447 GRSIR under constraint is conceivable in order to estimate proportions simultaneously.

448 **Acknowledgments.** The authors would like to thank the reviewers for their many  
449 helpful comments. They also appreciate the contribution of Mr. Senan Doyle to this  
450 paper. This work is supported by a contract with CNES through its Groupe Système  
451 Solaire Program and by INRIA. It would not be possible also without the financial support  
452 of the "Agence Nationale de la Recherche" (French Research Agency) through its MDCO  
453 program ("Masse de Données et COonnaissances"). The Vahiné project was selected in  
454 2007 under the reference ANR-07-MDCO-013.

## References

- 455 Aster, R., B. Borchers, and T. C.H. (2005), *Parameter Estimation and Inverse Problems*,  
456 Elsevier Academic Press.
- 457 Bernard-Michel, C., S. Douté, L. Gardes, and S. Girard (2007), Estimation of Mars surface  
458 physical properties from hyperspectral images using sliced inverse regression, *Tech. rep.*,  
459 INRIA, <http://hal.inria.fr/inria-00187444/>.
- 460 Bernard-Michel, C., L. Gardes, and S. Girard (to appear, 2008), Gaussian regularized  
461 sliced inverse regression, *Statistics and Computing*, doi:doi:10.1007/s11222-008-9073-z.
- 462 Bibring, J.-P., et al. (2004a), *OMEGA: Observatoire pour la Minéralogie, l'Eau, les Glaces*  
463 *et l'Activité*, pp. 37–49, ESA SP-1240: Mars Express: the Scientific Payload.
- 464 Bibring, J.-P., et al. (2004b), Perennial water ice identified in the south polar cap of Mars,  
465 *Nature*, 428, 627–630.

- 466 Brown, R. H., et al. (2004), The Cassini Visual And Infrared Mapping Spectrometer  
467 (Vims) Investigation, *Space Science Reviews*, *115*, 111–168, doi:10.1007/s11214-004-  
468 1453-x.
- 469 Carlson, R., M. Anderson, R. Mehlman, and R. Johnson (2005), Distribution of hydrate  
470 on Europa: Further evidence for sulfuric acid hydrate, *Icarus*, *177*(2), 461–471.
- 471 Carlson, R. W., P. R. Weissman, W. D. Smythe, J. C. Mahoney, the NIMS Science,  
472 and E. Teams (1992), Near infrared spectrometer experiment on Galileo, *Space Science*  
473 *Reviews*, *60*, 457–502.
- 474 Combal, B., F. Baret, M. Weiss, A. Trubuil, D. Macé, A. Pragnère, R. Myneni,  
475 Y. Knyazikhin, and L. Wang (2002), Retrieval of canopy biophysical variables from  
476 bidirectional reflectance using prior information to solve the ill-posed inverse problem,  
477 *Remote Sensing of Environment*, *84*, 1–15.
- 478 Douté, S., and B. Schmitt (1998), A multilayer bidirectional reflectance model for the anal-  
479 ysis of planetary surface hyperspectral images at visible and near-infrared wavelengths,  
480 *Journal of Geophysical Research (Planets)*, *103*(12), 31,367–31,390.
- 481 Douté, S., B. Schmitt, R. M. C. Lopes-Gautier, R. W. Carlson, L. Soderblom, and  
482 J. Shirley (2001), Mapping SO<sub>2</sub> frost on Io by the modeling of NIMS hyperspectral  
483 images, *Icarus*, *149*, 107–132.
- 484 Douté, S., B. Schmitt, J.-P. Bibring, Y. Langevin, F. Altieri, G. Bellucci, B. Gondet, and  
485 the Mars Express OMEGA Team (2007a), Nature and composition of the icy terrains  
486 from Mars express OMEGA observations, *Planetary and Space Science*, *55*, 113–133.
- 487 Douté, S., E. Deforas, F. Schmidt, R. Oliva, and B. Schmitt (2007b), A Comprehen-  
488 sive Numerical Package for the Modeling of Mars Hyperspectral Images, in *Lunar and*

489 *Planetary Institute Conference Abstracts, Lunar and Planetary Institute Conference*  
490 *Abstracts*, vol. 38, p. #1836.

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

494 Hapke, B. (1993), *Theory of reflectance and emittance spectroscopy*, Topics in Remote  
495 Sensing, Cambridge, UK: Cambridge University Press, —c1993.

496 Hapke, B. (2002), Bidirectional reflectance spectroscopy: 5. the coherent backscatter op-  
497 position effect and anisotropic scattering, *Icarus*, 157(2), 523–534.

498 Hastie, T., R. Tibshirani, and J. Friedman (2003), *The Elements of Statistical Learning:*  
499 *Data Mining, Inference, and Prediction*, Springer.

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

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

504 Murchie, S., et al. (2007), Compact reconnaissance imaging spectrometer for Mars  
505 (CRISM) on Mars reconnaissance orbiter (MRO), *Journal of Geophysical Research*  
506 *(Planets)*, 112(E11), E05S03, doi:10.1029/2006JE002682.

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

509 Pragnère, A., F. Baret, M. Weiss, R. Myneni, Y. Knyazikhin, and L. Wang (1999), Com-  
510 parison of three radiative transfer model inversion techniques to estimate canopy bio-  
511 physical variables from remote sensing data, *Geoscience and Remote Sensing Sympo-*

- 512 *sium*, 1999. *IGARSS '99 Proceedings. IEEE 1999 International*, 2, 1093–1095.
- 513 Schmidt, F., S. Douté, and B. Schmitt (2007), WAVANGLET: An efficient supervised  
514 classifier for hyperspectral images, *Geoscience and Remote Sensing, IEEE Transactions*,  
515 *45*(5), 1374–1385.
- 516 Scholkopf, B., and A. J. Smola (2002), *Learning with Kernels: Support Vector Machines,*  
517 *Regularization, Optimization, and Beyond*, MIT Press, Cambridge, MA.
- 518 Shkuratov, Y., L. Starukhina, H. Hoffmann, and G. Arnold (1999), A model of spectral  
519 albedo of particulate surfaces: implications for optical properties of the moon, *Icarus*,  
520 *137*, 235–246.
- 521 Tarantola, A. (2005), *Inverse problem theory and model parameter estimation*, Society for  
522 Industrial and Applied Mathematics.
- 523 Weiss, M., F. Baret, R. Myneni, A. Pragnère, and Y. Knyazikhin (2000), Investigation of  
524 a model inversion technique to estimate canopy biophysical variables from spectral and  
525 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. The increment between two values is constant.

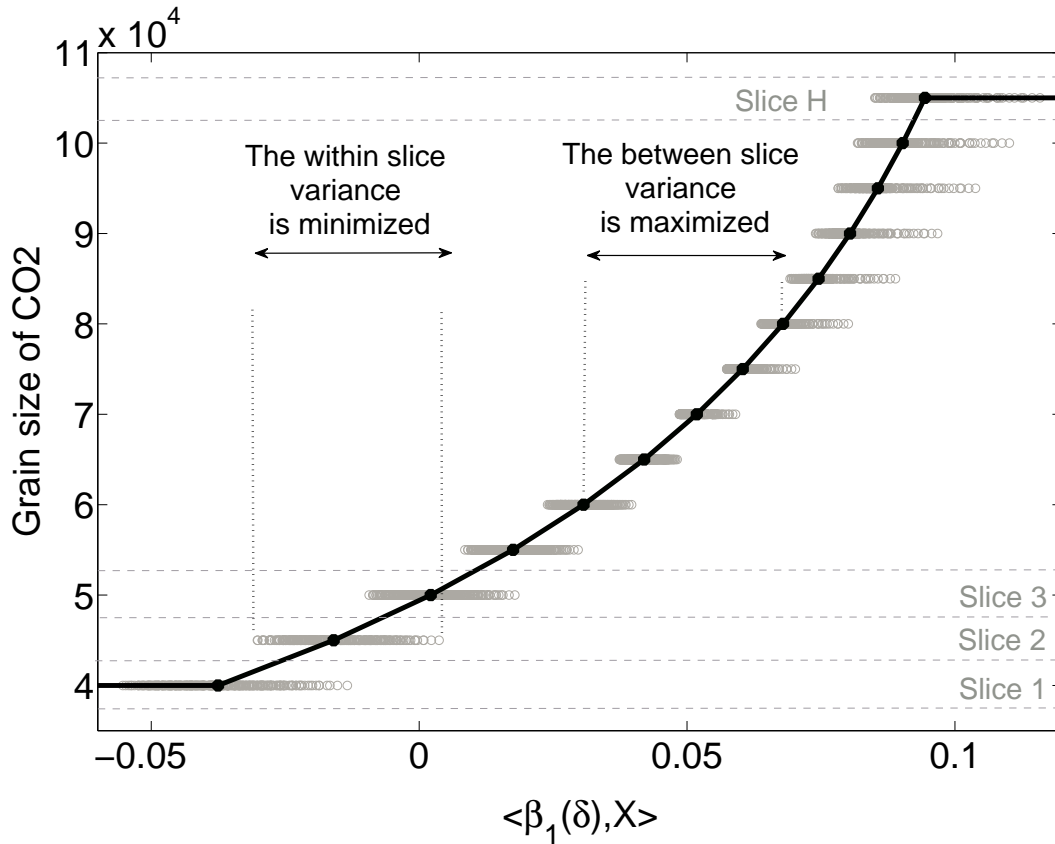
Parameters	Look-up table		Test data
	range	# distinct values	range
Proportion of water	[0.0001 0.0029]	15	[0.0006 0.002]
Proportion of CO <sub>2</sub>	[0.9942 0.9998]	29	[0.996 0.9988]
Proportion of dust	[0.0001 0.0029]	15	[0.0006 0.002]
Grain size of water	[50 450]	5	[100 400]
Grain size of CO <sub>2</sub>	[30000 165000]	28	[40000 105000]

**Table 2.** Validation criteria calculated on Tdata with GRSIR,  $k$ -NN, PLS and SVM. Note that the emphasized proportion has been fixed such that the sum of the proportions is one. For  $k$ -NN inversion the LUT contains 31500 spectra. For other methods, the LUT after selection contains 15407 spectra.

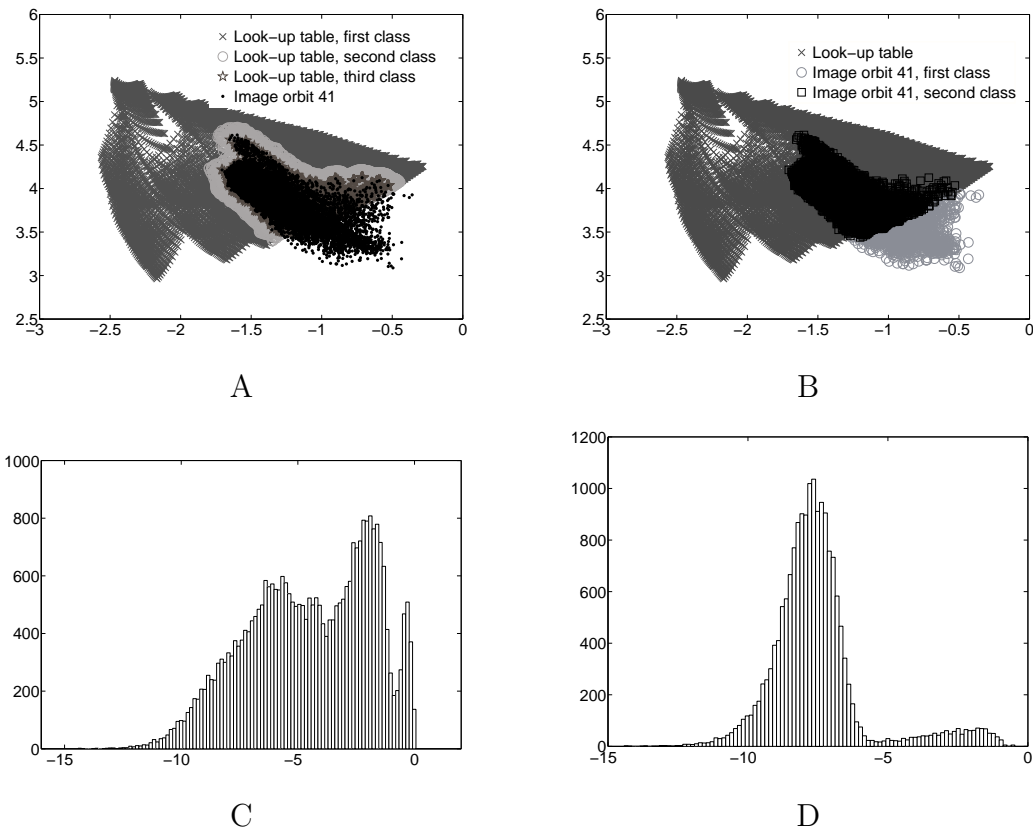
Parameters	$k$ -NN	PLS	SVM	GRSIR	
	NRMSE	NRMSE	NRMSE	NRMSE	SIRC
Proportion of water	0.86	0.52	<i>0.17</i>	<i>0.40</i>	0.90
Proportion of CO <sub>2</sub>	0.88	<i>0.56</i>	0.18	0.30	0.98
Proportion of dust	0.44	0.36	0.11	0.17	0.99
Grain size of water	0.43	0.44	0.17	0.54	0.84
Grain size of CO <sub>2</sub>	0.53	0.47	0.14	0.22	0.95
CPU time	60s	181s	54,992s	58s	

**Table 3.** Validation criteria calculated on Tdata with GRSIR, PLS and SVM. The 3584 spectra that constitute the look-up table have been selected after a first inversion by GRSIR.

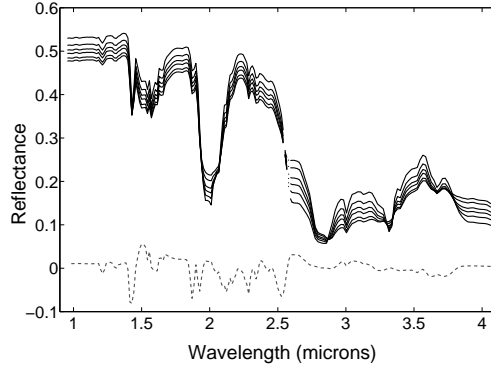
Parameters	PLS	SVM	GRSIR	
	NRMSE	NRMSE	NRMSE	SIRC
Proportion of water	0.35	<i>0.18</i>	<i>0.27</i>	0.92
Proportion of CO <sub>2</sub>	<i>0.32</i>	0.17	0.22	0.99
Proportion of dust	0.23	0.13	0.13	0.99
Grain size of water	0.45	0.23	0.39	0.92
Grain size of CO <sub>2</sub>	0.28	0.15	0.19	0.98
CPU time	832s	2,750s	19s	



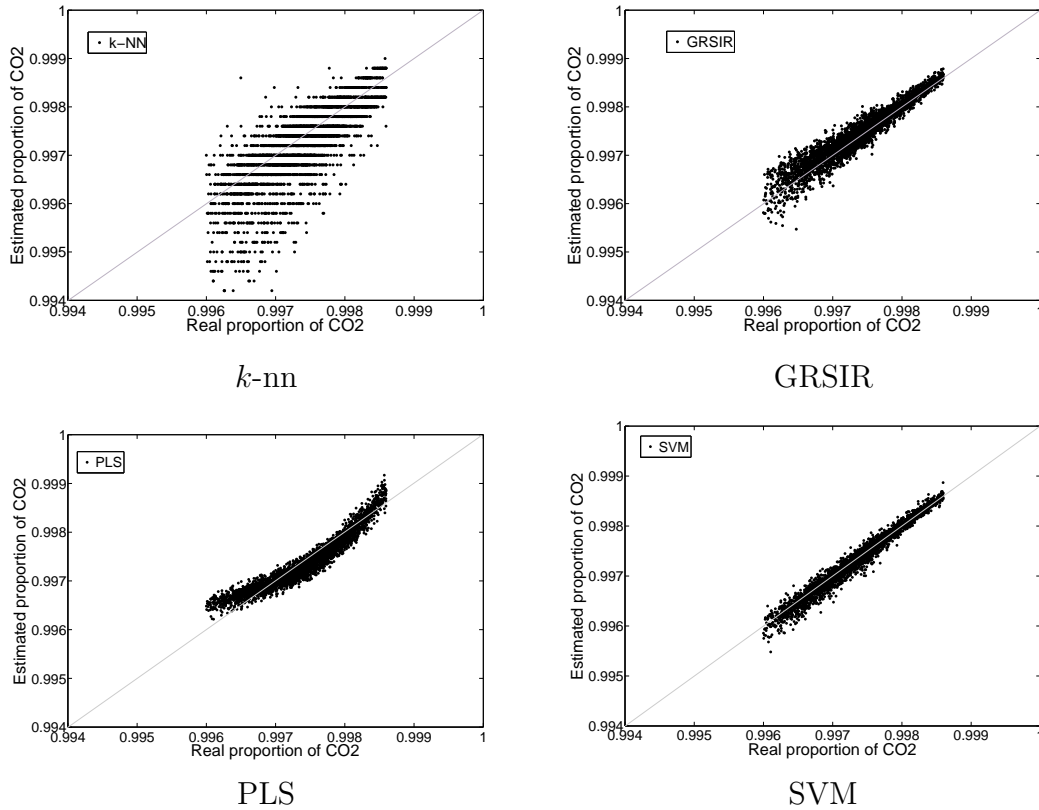
**Figure 1.** Functional relationship between reduced spectra on the first GRSIR axis and the grain size of CO<sub>2</sub> ice. This graphics illustrates the GRSIR methodology showing that the relationship is the best when the within slice variance is minimized or equivalently when the between slice variance is maximized. X-axis: reduced spectra from the learning database on the first GRSIR axis. Y-axis: Grain size of CO<sub>2</sub> ice.



**Figure 2.** 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). 3 classes are identified. 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.

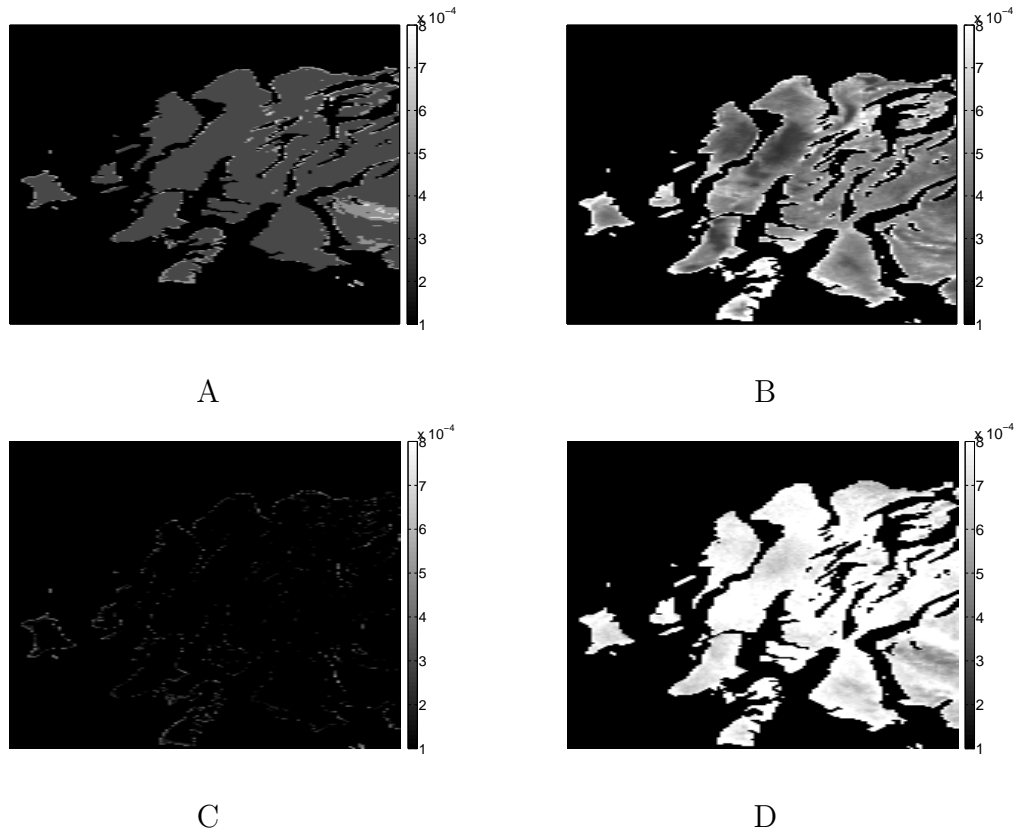


**Figure 3.** Plain line: Series of spectra extracted from the look-up table for CO<sub>2</sub> ice grain size varying between 30 and 165 millimeters, the other parameters being kept constant at mean values. Dotted-dash line: Component wise multiplication between GRSIR axis for the grain size of CO<sub>2</sub> ice and a representative spectrum (see Section 5.5).

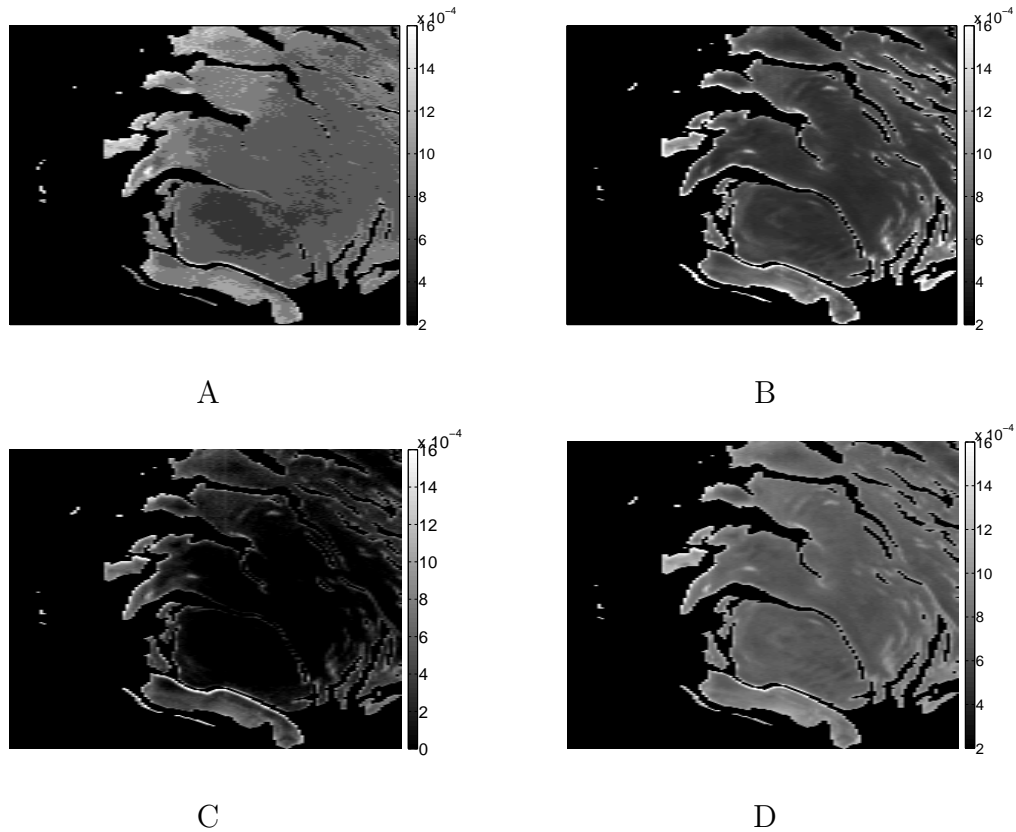


**Figure 4.** Scatter plots of the proportion of CO<sub>2</sub> ice from the test data versus estimated values. Horizontally: Proportion of CO<sub>2</sub> ice. Vertically: Estimated proportions of CO<sub>2</sub> ice by *k*-NN, GRSIR, PLS and SVM.

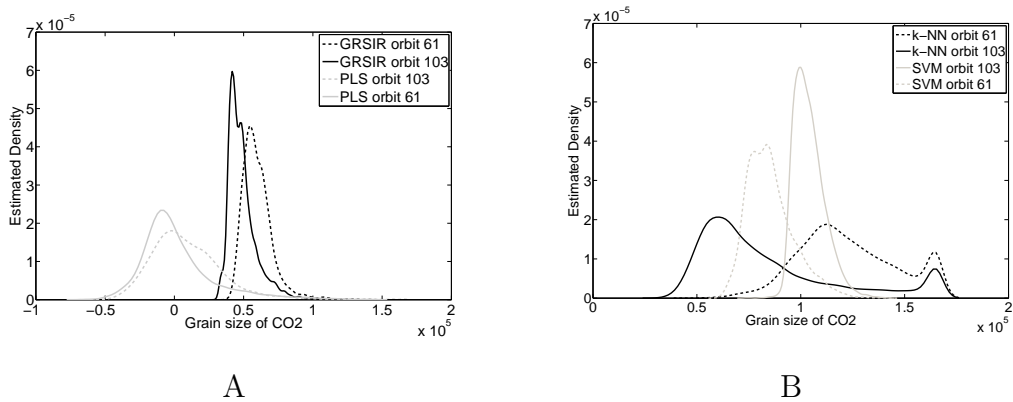




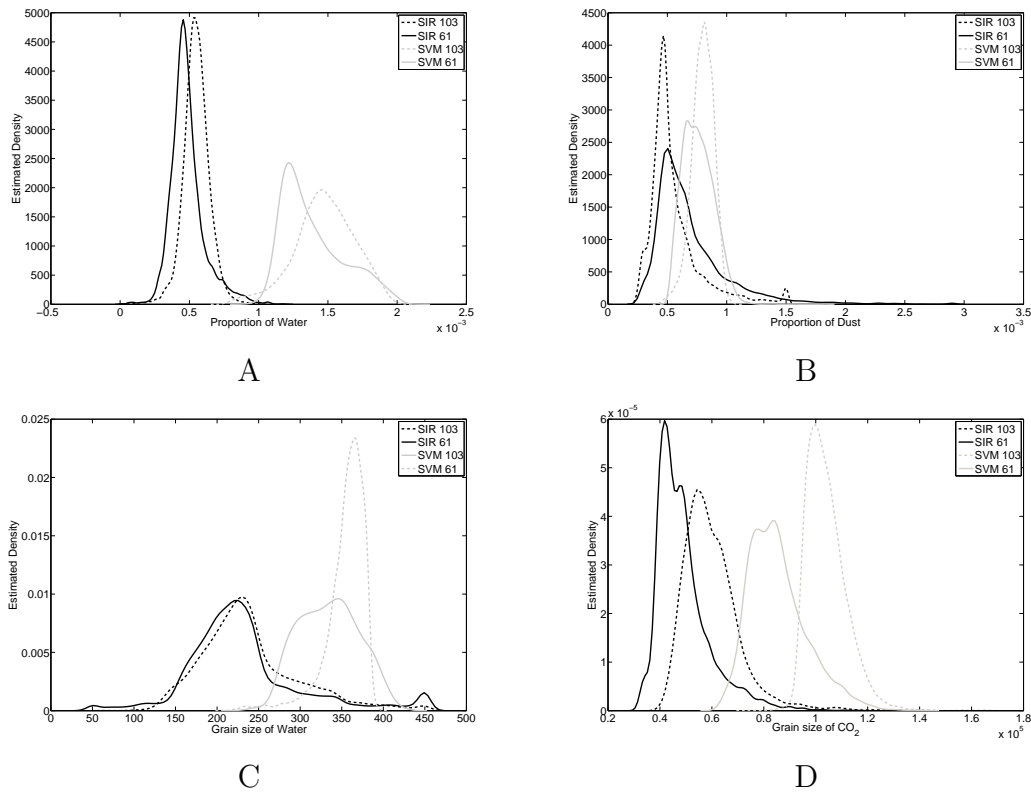
**Figure 5.** Proportion of dust estimated by  $k$ -NN (A), GRSIR (B), PLS (C) and SVM (D) from the hyperspectral image observed from orbit 103.



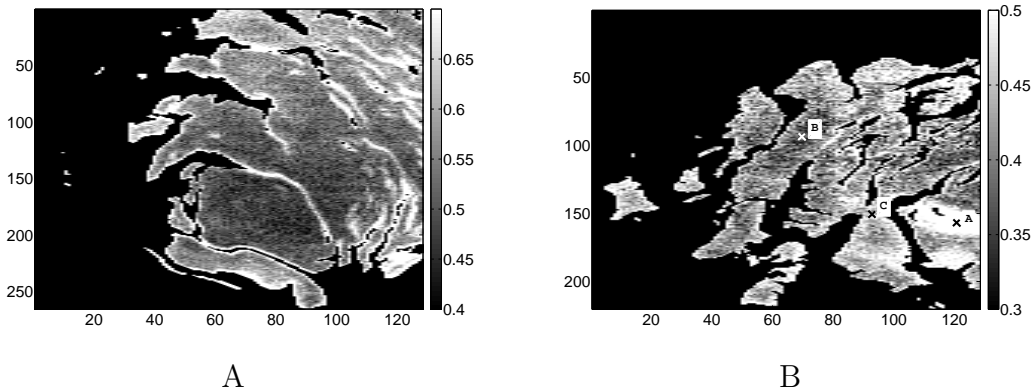
**Figure 6.** Proportion of dust estimated by  $k$ -NN (A), GRSIR (B), PLS (C) and SVM (D) from the hyperspectral image observed from orbit 41.



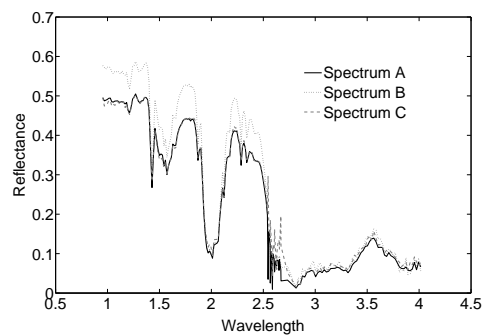
**Figure 7.** Comparison between the densities of the estimated grain sizes of  $\text{CO}_2$  ice with PLS and GRSIR (A) and  $k$ -NN and SVM (B) in images from orbit 61 and 103.



**Figure 8.** Distribution of values obtained for all parameters respectively by GRSIR and SVM for observations 61 and 103.



**Figure 9.** Cosines 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 10.** Spectra extracted from the locations labeled A, B and C in the image 103 of figure 9 (B).