Independent Component Analysis classification for ChemCam remote sensing data. A. Cousin[1], O. Forni[1], S. Maurice[1], J. Lasue[2], O. Gasnault[1], R. Wiens[2] and the ChemCam team, [1]Institut de Recherche en Astrophysique et Planétologie, Toulouse (agnes.cousin@cesr.fr), [2]Los Alamos National Laboratory, New Mexico.

Introduction: ChemCam is part of the Mars Science Laboratory (MSL) rover payload which is scheduled for launch in December 2011. This instrument uses the Laser Induced Breakdown Spectroscopy (LIBS) technique to investigate the geochemistry of the Martian surface [1,2]. ChemCam is composed of two parts: the Body Unit (located in the rover) including three spectrometers and their electronics, and the Mast Unit (on top of the mast) including the laser, telescope, Remote Micro Imager (RMI) and their electronics. ChemCam performs an active small-footprint (0.2-0.6 mm dia.) chemical analysis by remote sensing: the LIBS technique creates a plasma of ablated material in electronically excited states, and then analyses the plasma light with its spectrometers to obtain emission spectra. The Flight Model (FM) underwent rigorous calibrations before to delivery, with a wide range of samples and conditions [3]. The calibration datasets are necessary also to understand how ChemCam is able to classify rocks using multivariate analysis such as Independent Componant Analysis (ICA). The objectives of this study are (1) to classify a set of spectra obtained at 3 m comparing several parameters in the ICA model, (2) to understand if sample distance plays a role in the classification of the rocks.

Samples and data acquisition: This study reports on a set of 14 standards used during the ChemCam calibration in Thermal Vacuum Chamber. Samples consist in a wide range of basalts and andesites, one dolomite, oil-rich shale, graphite, and one Mars sedimentary simulant from the on-board calibration target [1,2]. These samples were placed at 1.5, 3, 5 and 7 m distances, and under a simulated Martian atmosphere (7 Torr of CO2). For more details about the samples see [3]. The ChemCam FM was placed in a chamber at -45°C and 7 Torr. 200 spectra per sample were acquired (50 shots per burst, at 4 locations). The 3 meter data were averaged by groups of 10 in order to have 20 spectra per target, which allows a good sampling. Each spectrum is corrected for noise and a continuum has been substracetd. Finally, all spectra have been calibrated for channels to wavelengths.

Methodology: In LIBS technique, multivariate analysis is an efficient tool to classify the samples. The ICA technique was chosen because of its capability to classify very similar rocks better than the PCA (Principal Component Analysis) technique [4-5]. This method derives from the Blind Source Separation (BSS) technique. The ICA method is a decomposition of a vector in linear independent components. For more details about ICA, see [6, 7]. Since with the ICA method, the optimal number of components is not known, we need to try several components to find the best model. We use the Joint Approximate Diagonalization of Eigenmatrices method (JADE) algorithm which is an orthogonal ICA method [8-9]. We use also a Jackknife test to classify each spectrum [10]:this method is used to estimate the bias of the prediction by deleting one spectrum from the total data set, and to recalculate the prediction from the remaining data. At the end, the assignment to a group is performed by computing the minimum distance between the spectra to be classified and all the clusters issued of the data set. This distance is computed using the Mahalanobis distance, which takes into account the covariance among the variables [5].

Results: Results presented here consist of the JADE model and Jackknife test performed using data obtained at 3 meters, and comparing several parameters for the model. Then 1.5 meter data are classified also using this ICA model.

3 m data : Before looking at the number of components to choose and their spectral significance, we tried to freeze several parameters such as the normalization and the size of the input signal. First, we make an ICA model using the entire spectra (all of the 6144 channels for the 280 groups of spectra), and we compare the results thus obtained to classify each spectrum, with and without normalization. This normalization was performed by dividing each spectrum by its standard deviation. Secondly, we made an ICA model with the same data, but using only the most important lines for each spectrum. To perform this, we take into account only signal greater than the median of the averaged spectra. In this case, instead of 6144 channels there are only 3071 channels for each of 280 spectra. We then compared the results for the Jackknife test with and without normalization .

Figure 1 shows that the best results are obtained using all the channels of the spectra: the rate of 100% well-classified spectra is reached with fewer components (7 or 5 components depending if there is a normalization or not, respectively) than using only the most important lines. Therefore we decided to continue this study using the entire signal of each spectrum, and with no normalization. All of the five components have a spectral significance, and each corresponds to one or several elemental compositions namely Ca, Ti, Fe or Mg. The last component is associated to the lithium and identifies theMars sedimentary simulant targets (in which Li was used in a binder).

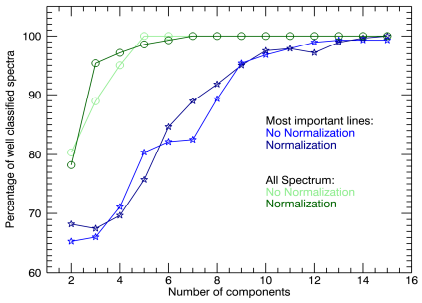

Figure 1 : Percentage of spectra well-classified depending on parameters used in the model.

1.5 m data : We tried to classify the data obtained at 1.5 meters using this 3 meter model and the parameters described above. The objective is to understand if distance can play a significant role classifying the targets. We took only samples having a corresponding class in the 3 meter data set, which corresponds to a set of 14 spectra.

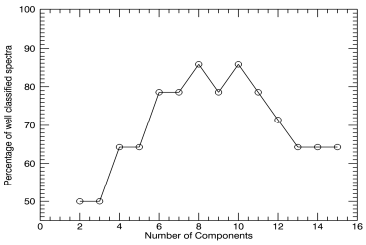

Figure 2 : Jackknife test results for data taken at 1.5 meters, using the ICA model performed with data taken at 3 meters.

The best obtained classification rate is 85.7% of well-classified spectra, which is reached using 8 and 11 components (figure 2). This means that only two spectra are poorly classified. Each of the eight components used to best classify these 1.5 meters data has an elemental signification. In particular, the Li component is specific to Mars sedimentary simulant samples, whereas the H dominated component is specific to Graphite and Oil shale, and the Ca component to Dolomite.

The two poorly-classified spectra are JB3 and BHVO, which are basalts. They are classified as JB2 and BT2, respectively, which are also basalts. The two wrongly classified basalts (JB3 and BHVO) have small differences in composition with the group where they are assigned, by only ~ 3wt% in Al_2O_3 and ~ 2wt% in Fe_2O_3. The spatial representation computed by the Sammon's map projection (figure 3) [11] shows the several clusters and the classified 1.5 meter spectra. This Sammon's map representation was computed using the K component versus (Fe + Ti + Na) from the JADE algorithm.

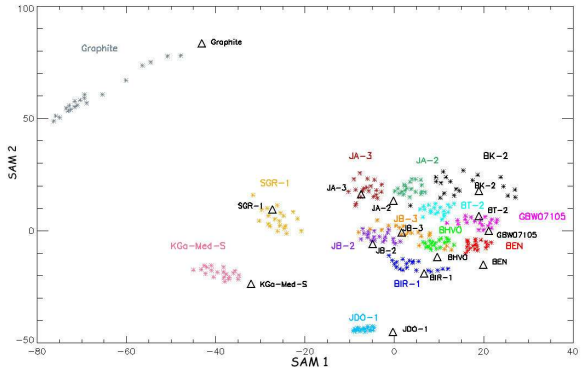

Figure 3 : Spatial representation computed by the Sammon's map of the unknown spectra (black triangles, at 1.5 meters) and clusters, which correspond to the spectra from samples at 3 meters.

Discussion: The first ICA classification analysis with these datasets is still in progress: we want to study the inverse, computing the model with data at 1.5 m and then classify the 3 m data. We would like also to re-done this study with no graphite which is an oultier because of its particular composition (100% C). Moreover, there are other ways to explore, like the distance measure to assign a group to an unknown spectrum. In this study, we use the Mahalanobis distance, but it could be interesting to do the same work using the Euclidian or the Manhattan distances, which considers the absolute difference between two points. We also don't use distance thresholds to assign a spectrum to a group. This threshold is necessary to avoid wrong classification, as we observed with BHVO and JB3 basalts at 1.5 m, and to assign the spectra as unclassified.

References: [1] Wiens et al. (2005), *LPSC 36th*, #1580. [2] Maurice et al. (2005), *LPSC 36th*, #1735. [3] Wiens et al. (2010), *LPSC 41th*, #2205. [4] Forni et al. (2009) *LPSC 40th*, #1523.[5] Forni et al (2011), *Analytical & Bioanalytical Chemistry*, submitted. [6] Comon P., (1990), *Traitement du Signal*, *7*, 435-450. [7] Comon P., (1994), *Signal Processing*, *36*, 287-314. [8] Cardoso J.-F., (1997*), IEEE Letters on Signal Processing*, *4*, 112–114. [9] Cardoso J.-F.,(2003*), Journal of Machine Learning Research* , 1177-1203. [10] Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap. Springer-Verlag*, Inc. pp. 281. [11] Lasue J. et al (2011), *Analytical & Bioanalytical Chemistry*, submitted.