

MULTIVARIATE ANALYSIS OF CHEMCAM FIRST CALIBRATION SAMPLES. O. Forni¹, S. Clegg², R. C. Wiens², S. Maurice¹ and O. Gasnault¹; ¹Centre d'Etude Spatiale des Rayonnements-CNRS (31028 Toulouse Cedex 4, France; Olivier.Forni@cesr.fr, Sylvestre.Maurice@cesr.fr, Olivier.Gasnault@cesr.fr) ²Los Alamos National Laboratory (Los Alamos, NM 87545 USA; sclegg@lanl.gov, rwiens@lanl.gov)

Introduction: ChemCam is an active remote sensing instrument suite being built for MSL [1,2]. It uses laser pulses to remove dust and to profile through weathering coatings of rocks up to 9 m away. Laser-induced breakdown spectroscopy (LIBS) obtains emission spectra of materials ablated from the samples in electronically excited states over an analysis spot < 1 mm in diameter at the focal point of the incident laser beam. The flight model (FM) of the instrument composed of the body unit including the spectrometers and their electronics, and of the mast unit including the laser and the microcamera (RMI) and their electronics, went through intensive testing and calibration at CESR and LANL before delivery to JPL. The samples and calibration procedures are described in more detail in another abstract [3]. LIBS, with its narrow emission peaks and large number of channels, lends itself very well to multivariate analysis (MVA). Here we describe investigations of several MVA techniques aimed at rock classification using the ChemCam calibration data. The elemental analysis is not treated here per se.

Samples and Calibration Details : This abstract reports on a subset of the 65 standards used for ChemCam calibration. The standards include, in pressed powder form, basalts, andesites, gypsums, olivines, limestones, dolomites, a graphite (for carbon), sediments of marine, river, and stream origin, and glass and ceramic standards to simulate Mars basalts and sedimentary rocks, respectively.

Most of the details of the calibration are reported in another abstract at this meeting [3]. Briefly, the samples were analysed by ChemCam LIBS in 7 Torr of CO₂ using 50 laser pulses per analysis spot and 5 spots per standard at various distances. The data used in this study were taken during thermal-Mars-atmosphere testing up to 7 meters distance, a relatively long reach for ChemCam.

Methodology: In spectroscopy, it is considered that the observed spectra are linear combinations of pure sources or poles that we want to estimate. We used two different multivariate analysis methods to achieve this goal with the LIBS spectra: the well-known Principal Component Analysis (PCA) and the Independent Component Analysis (ICA). PCA calculates principal components (PCs) which are linear combinations of the variables that convey the major part of the information contained in the global dataset. Spectra are decomposed in linear combinations of the

PCs. Their coordinates along the PCs (scores) can be represented to visualize similarities between spectra and to identify classes of spectra. PCA maximises the dispersion, and consequently, decorrelates the signals assuming that all the sources follow a Gaussian probability function. For a complex ensemble of data in which many signals contribute to the observed data, the Gaussian hypothesis of PCA will not be valid in all cases. The extraction of pure and statistically independent spectra from a mixture cannot be reliably obtained with the PCA.

ICA is a different technique that comes from developments in the Blind Source Separation (BSS) research. The goal of ICA is to estimate h source signals, assumed to be stationary and independent, using n observed signals that are independent unknown mixing of the source signals ([4], [5]). ICA, like PCA, tries to transform the original representative space by searching for directions in a new space, so that the resulting vectors are independent, and not only uncorrelated. ICA is then a method of linear transformation in which the representation is the one that minimise the statistical dependence of the components. This is achieved, using a criterion, related to the information entropy theory that gives back the statistical independence assuming that the data follow a non-Gaussian distribution. As in [6] and [7] we have used the Joint Approximate Diagonalization of Eigenmatrices method (JADE) algorithm ([8], [9]), which is an orthogonal ICA method and, like most mainstream ICA techniques, it exploits higher order statistics related to the fourth order cross cumulants to maximise the non-gaussianity of the sources.

Results: PCA was applied to the 7 meters data set, as shown in Figure 1.. PC1, on the x-axis, represents 48% of the variance and is weighted strongly towards calcium. PC2 is weighted towards Na and K, and anti-correlated with Mg. Higher number PCs are also very relevant, as important classification elements such as Si and Fe have weaker LIBS responses than the alkalis and alkali earths. Clustering analysis tools can be used to classify the samples with more than two PCs.

ICA was also applied to the 7 meters data. One drawback of ICA is that the number of useful independent components is a priori unknown. We have imposed the number of independent components such that a carbon component with molecular Swan bands appears specific to graphite. This constraint results in

imposing 11 components from which 8 have a spectral significance, the remaining being continuum and/or noise related. Figures 2 and 3 show two examples of these independent components.

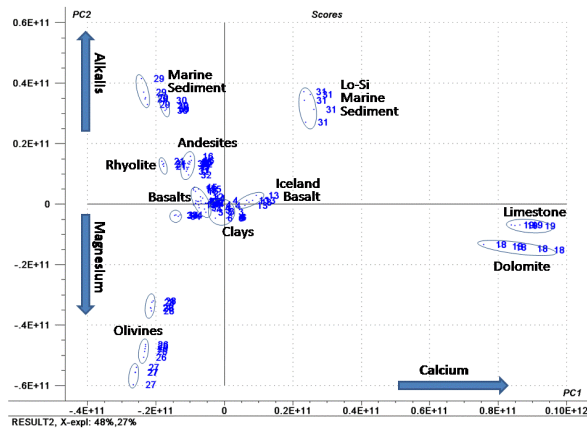


Fig. 1. Principal components 1 vs. 2 for ChemCam LIBS data taken at 7 m.

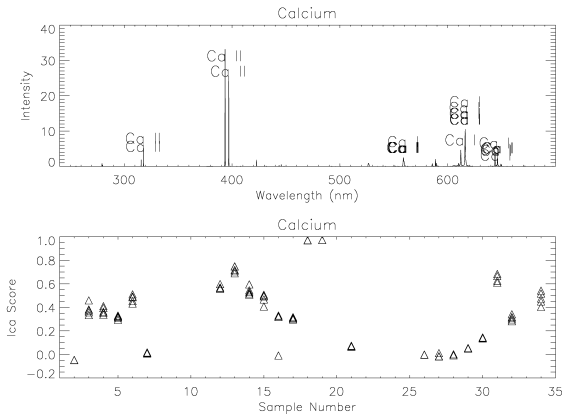


Figure 2: The Calcium component diagnostics of the Dolomite and Limestone.

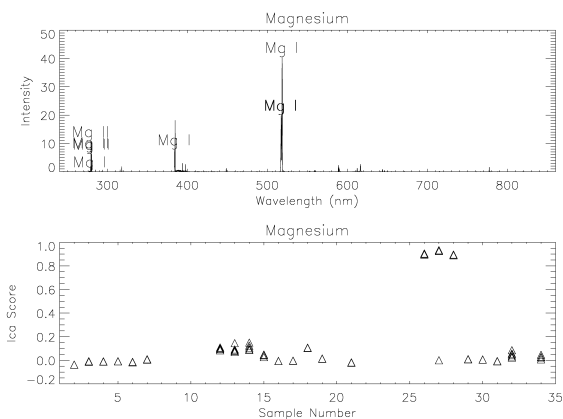


Figure 3: The Magnesium component exhibiting the olivine samples.

Figure 2 is a Calcium dominated component and is diagnostic of the Dolomite, and Limestone samples. Figure 3 represents the Magnesium components in which the Olivine samples are clearly isolated. The other dominant components are driven by Potassium, Sodium, Lithium, Oxygen associated with Aluminium, Titanium and finally Carbon.

In Figure 4 we present a particular combination of these components, namely Ca vs. (Na + K) - Mg. This combination was chosen to mimic the PCA tendencies. It illustrates the ability of this method to apparently better separate the various types of rocks with respect to the PCA.

Work on comparing the rich LIBS data sets between different MVA methods is just beginning.

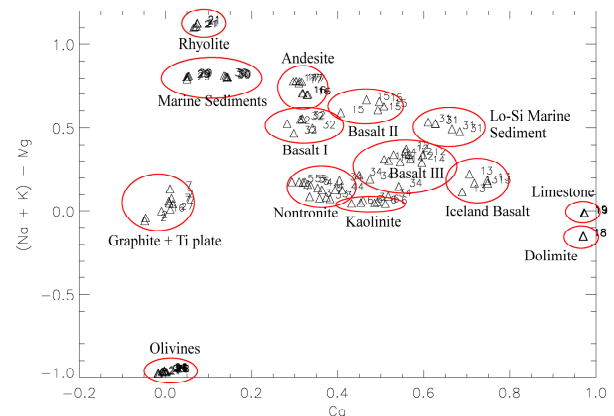


Figure 4: Diagram showing the combination of the Ca vs. (Na + K) - Mg ICA components illustrating the classification of the different group types.

References:[1] Maurice S. et al. (2005) *LPS*. XXXVI, Abstract #1735. [2] Wiens R.C. et al. (2005) *LPS* XXXVI, Abstract #1580. [3] Wiens R.C. et al. (2009), this meeting [4] Comon P., (1990), *Traitement du Signal*, 7, 435-450. [5] Comon P., (1994), *Signal Processing*, 36, 287-314. [6] Forni O. et al; (2005) *LPS* XXXVI, Abstract #1623. [7] Erard S. et al. (2008) *JGR accepted* [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.