CHEMCAM TARGET CLASSIFICATION: WHO'S WHO FROM CURIOSITY'S FIRST NINETY SOLS.

O. Gasnault[1], O. Forni[1], P.-Y. Meslin[1], S. Maurice[1], R. Wiens[2], R. Anderson[3], G. Berger[1], S. Clegg[2], A. Cousin[1], C. D'Uston[1], J. Lasue[1], E. Lewin[4], N. Melikechi[5], H. Newson[6], P. Pinet[1], and the MSL Science Team. [1]Institut de Recherche en Astrophysique et Planétologie (9 av. Roche, 31028 Toulouse, France – Olivier.Gasnault@irap.omp.eu). [2]Los Alamos National Lab., USA. [3]USGS Astrogeol. Sci. Center, USA. [4]Inst. Sciences Terre, France. [5]Delaware State Univ., USA. [6]Univ. New Mexico, USA.

Introduction: The ChemCam instrument onboard the Curiosity rover contains a Laser-Induced Breakdown Spectroscopy (LIBS) experiment [1; 2]. LIBS spectra are rich in information, with numerous atomic emission lines per element [3]. ChemCam spectra are processed using multivariate analyses to extract the composition and characteristics of the targets [4; 5; 6].

Dataset: We study data from Sols 13 to 92 and a traverse of nearly 400 m. With 11,810 spectra on about 50 martian targets, the ChemCam dataset addresses the need for a survey of the landing site composition [7]. This large volume of data also requires tools to develop synthetic summaries as proposed here. The density plots of the abundances of the various elements detected by ChemCam reveal multimodal distributions, which indicates the possibility of several groups of more uniform composition. This is also supported by several trends seen between the elements [4; 8].

The data are pre-processed to remove the noise, the continuum, and the ambient light signal; They are normalized by the total intensity in each of the three spectrometers, and the instrument response is partly corrected. A wavelength calibration is also applied. Then, the data are processed through an Independent Component Analysis (ICA) to estimate the signal sources, i.e. proxies of elemental abundances [4]. Here we retain the ICA components corresponding to Al, Ca (elemental lines), CaO (molecular band), Fe, H, K, Mg, Na, O, Si, and Ti. The components serve as input to a clustering algorithm.

Target nomenclature: Hereafter we refer to the informal names given by the MSL Science Team to the rocks and soils along the traverse to facilitate the discussion. The number after the name corresponds to the observation point sampled by ChemCam, since several locations were observed per target. For each of these points, a few tens of laser shots are used, providing several spectra per point. However here, we use only the average spectrum for each point, after removing those corresponding to the five first laser shots, the spectra from which are contaminated by dust.

Method: To encompass all the (useful) dimensions of the compositional information extracted from our dataset, a clustering analysis is conducted to highlight both the diversity and the relationship between the targets. It is expected that (compositional) signal similarities are indicative of similarity with respect to more fundamental geologic properties.

We adopt the divisive algorithm described by *Kaufman and Rousseeuw* [9, chap. 6], which starts with the entire dataset followed by iterative binary subdivisions, to search for the main data structures. This hierarchical method is also appropriate to find the optimal number of clusters and to reveal relations between a parent unit and its sub-units.

Results: The dendrogram in Figure 1 shows distinctive structures up to 10 groups, and a few outliers (singleton Epworth5 by its CaO signature [5]; Stark and Preble2 by their high-Si signatures).

The first division separates more felsic compositions seen mainly at the beginning of the mission (e.g. Mara seen on Sol 15 [10; 11]) from other compositions (more numerous, but this is a sampling bias) found all along the traverse and showing on average higher Fe and Ti contents.

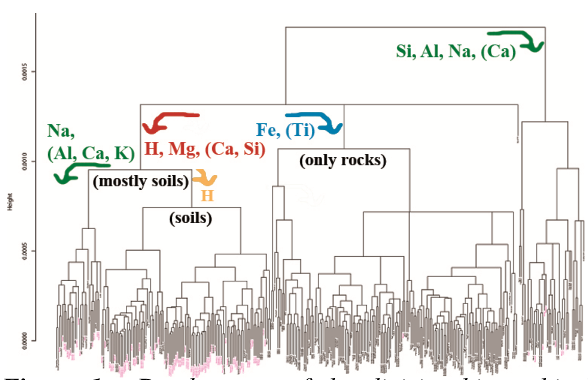

Figure 1: *Dendrogram of the divisive hierarchical clustering analysis (from top to bottom). The height corresponds to the cluster size before its division; The cluster size is defined as the largest distance between any two of its observations. The divisive analysis can be stopped by drawing a horizontal line, and the number of clusters is the number of intercepted branches.*

The felsic group: Based on histograms of the ICA components, this group is made of samples relatively enriched in Si and Al, and to some extent in Na, Ca and K. The rocks and the soils are generally mixed in this group (e.g. Mara and Kam9; Kilian9 and Murky4; Mara1 and Beechey5), except for a sub-group richer in Al and made almost exclusively of rocks (ThorLake2, -3 and -4, and the conglomerates Goulburn and Link [12]: Goulburn7b1, -7b3 and -7b5, Link4).

Note that the first terrains encountered by the rover, where we found 95% of the samples of this felsic group, were rather free of fine sand with abundant gravels (Fig. 2), which certainly influences the tentative measurements of soils with LIBS there [10], while Si shows a range of values in the rocks [13]. At least one observation of that group (Kilian9 on Sol 72), was obtained later in the mission, at the Rocknest station, and may perhaps suggest that felsic-like compositions can be found (here, in the soil) at other locations than the Bradbury landing area.

The basic group: The lower abundance of Si in this group corresponds more to basic compositions. Most of the samples in this group are also relatively enriched in Ti and Fe with respect to the felsic group. They have a similar distribution of H for low abundances, but the group also include high-H samples compared to the felsic group.

As shown in Figure 1, the group of basic compositions is subdivided clearly into rocks enriched in Fe (and Ti on average) dominated by samples from the Rocknest area on one side, and another subgroup dominated by soils enriched in H and Mg (as well as Ca and Si on average) on the other side. In the latter subgroup, the soils are found essentially in the branch enriched in H, while a mixture of rocks and soils is found in the other branch characterized by higer Na and other elements similar to the first felsic group but richer in Mg rather than in Si and Al (the quality criterion for that cluster is not strong; not shown). Indeed, we found rocks split between the two extremes of the dendrogram, such as some observations on Jake Matijevic (Jake_1) or Beaulieu (which hit both soils and pebbles), revealing the different components sampled with the laser and forming these targets.

To reach the soils (Akaitcho, Beechey, Crestaurum, Epworth, Kenyon, Schmutz), we followed the branch enriched in Mg and then in H, in comparison to the targets dominated by Fe and then Na (Fig. 1), which is consistent with an ICA component identified by *Forni et al.* [4] respectively correlated and anticorrelated to these elements and providing a good proxy to chemically identify the soils in the ChemCam data collection. This soil subgroup, enriched in H and depleted in Si and alkaline elements, is described as Type 1 in the soil survey by *Meslin et al.* [10]. Crestaurum2-1 is a good representative of the soil subgroup, and Bathurst is the only unambiguous rock falling in this subgroup, although it is not particularly enriched in hydrogen.

Further subdivisions are possible, some involving oxygen that may partly reflect residual physical effects (e.g. laser coupling with the target or distance) since part of the laser-induced emission line is produced by the atmosphere [14].

The targets that are poorly classified in the soil group are generally occurrences when ChemCam was pointing at a rock but the laser hit soils, or more probably a mixture of soils and rocks, either because some points missed a rock (e.g. top of Rocknest3) or because the rock was partly covered with soils (e.g. Pearson).

Conclusions: While this first unsupervised classification of ChemCam targets primarily reaches the same conclusions as some other analyses, it provides an important tool for rapid assessment of new observations: are we seing the same old stuff or new materials? How many clusters describe our dataset?

This approach can be improved in many ways, for example by considering clustering quality criteria, or more elements such as minor elements [15], and more importantly by giving a geological meaning to the groups. One expected outcome is to highlight the relationships between targets that are physically different or geographically separated, but are chemically similar. One exemple is the rock Mara (Sol 19) and the “soil” Kam (Sol 43), which fall into the same category in our analysis, even at a fine level and with good confidence. It is essential to understand these surficial materials to make the link between ground truth and remote observations from space [16].

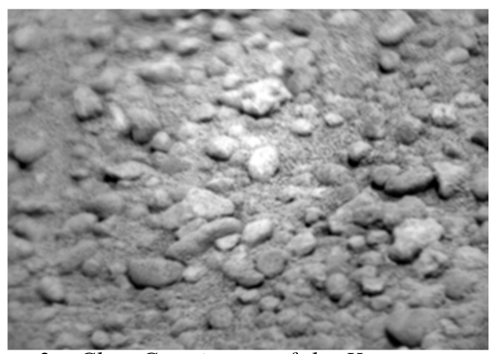

Figure 2: *ChemCam image of the Kam target with a mixture of coarse grains and pebbles, defining a distinctive type of soils [10; 13].*

References: [1] Maurice S. et al. (2012) *Space Sci. Rev., 170*, 95-166. [2] Wiens R. et al. (2012) *Space Sci. Rev., 170*, 167-227. [3] Cremers D. and Radiemski L. (2006) Handbook of laser-induced breakdown spectroscopy, Wiley (ed.) [4] Forni O. et al. (2013) *LPS XLIV*. [5] Clegg S. et al. (2013) *LPS XLIV*. [6] Lasue J. et al. (2013) *LPS XLIV*. [7] Maurice S. et al. (2013) *LPS XLIV*. [8] Anderson R. et al. (2013) *LPS XLIV*. [9] Kaufman L. and Rousseeuw P. (2005) Finding Groups in Data: An Introduction to Cluster Analysis. Wiley (ed.) [10] Meslin P.-Y. et al. (2013) *LPS XLIV*. [11] Bridges N. et al. (2013) *LPS XLIV*. [12] Mangold N. et al. (2013) *LPS XLIV*. [13] Wiens R. et al. (2013) *LPS XLIV*. [14] Gasnault O. et al. (2012) *LPS XLIII,* Abstract #2888. [15] Ollila A. et al. (2013) *LPS XLIV*. [16] Karunatillake S. et al. (2007) *JGR, 112*, E08S90