Partial Least Squares sensitivity analysis and improvements for CHEMCAM LIBS data analysis on Mars, J. Lasue1, O. Forni1, R. B. Anderson2, G. Berger1, S. M. Clegg3, A. Cousin1, M. D. Dyar4, C. Fabre3, O. Gasnault1, E. Lewin6, P.-Y. Meslin1, S. Maurice1, R. L. Tokar2, R. C. Wiens2, and the MSL Science team.

1IRAP-OMP, CNRS-UPS, Toulouse, France (jlasue@irap.omp.eu) 2US Geological Survey, Flagstaff, AZ, USA 3Los Alamos National Laboratory, NM, USA, 4Mount Holyoke College, South Hadley, MA, USA, 5G2R-Georessources, Nancy, France, 6ISTerre, Grenoble, France, 7Planetary Science Institute, Tucson, AZ, USA

Introduction: ChemCam is a Laser-Induced Breakdown Spectroscopy (LIBS) instrument on-board MSL that can analyze the chemical composition of geological samples at a distance by detecting the light emission of constituting elements [1, 2]. Up to sol 100, the instrument successfully acquired > 14000 spectra. Each spectrum consists of intensity signals distributed over 6144 channels. The structure of the spectral data can be handled by multivariate data analysis techniques for classification [3] as well as quantification purposes. The ChemCam team uses a Partial least squares (PLS) algorithm, regressed against a database of 66 standards measured by ChemCam prior to flight, to provide rapid elemental abundances on the tactical (day-to-day) timeline. PLS can correct for chemical matrix effects which can obscure a linear relationship between abundance and peak areas [e.g. 4, 5, 6]. In this work, we report on the current status of the PLS technique used to quantify the elemental composition of ChemCam’s targets. We describe the current capabilities of the method in predicting the elemental composition of the target, its sensitivity to the different parameters affecting the data, its validation using other methods, and further improvements to be investigated.

PLS current status: One of the most used and most accurate techniques for LIBS instruments is the PLS regression [4, 5]. Using a database of standards of known compositions and analyzed under laboratory conditions replicating Mars, it is possible to apply PLS regression to the spectral data to obtain quantitative predictions and determine the accuracy of the instrument [6]. In order to determine the accuracy, a classic “leave-one-out” (LOO) cross-validation procedure was used, where one standard is removed at a time and predicted using the remaining standards as training set. The accuracy of the model is obtained using the root mean square of error prediction: RMSEP = \sqrt{ \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 }, where n is the number of standards, y_i the composition of the sample for a given element and \hat{y}_i is the prediction of that composition for the LOO model. Given in units of the predicted quantity, it is related to the standard deviation of the errors of predictions and gives an indication of its distribution. The global RMSEP value (including all elements) typically decreases with the number of principal components taken for the model before reaching a minimum and an asymptotic behavior. This curve is used to select the best training model for the predictions by taking the number of components for which the RMSEP value is about 1σ above the minimum [e.g. 7]. The RMSEP values obtained for the current PLS algorithm, applied to the 66 standards in the current spectral database [6], and used to analyse ChemCam data are given in Table 1. The best number of components (NC) is 9 – though this could differ from element to element.

Table 1: Comparison between Portage (sol 89) chemical compositions from APXS and ChemCam using PLS. RMSEP values obtained at the best number of components NC=9. All values in wt%.

Validation of the PLS results on Mars: Several techniques have been used to validate the PLS algorithm used to quantify ChemCam data. First, the loadings used to predict the elements have been checked to be correlated with the emission lines present in the spectra. The PLS components compare also qualitatively well with the ICA separation of the multivariate classification algorithm [3]. Second, it is known that LIBS emission lines’ intensity usually increases with the proportion of a given element present in the sample. The PLS predictions have been shown to correlate very well with the intensity of emission lines for the major elements of the rocks in the training dataset as well as those detected on the martian targets.

Third, the MSL rover also has a set of 10 calibration targets (CCT) on-board that can be used to verify the PLS predictions [8, 9]. This is all the more useful since some of those calibration targets are not included in the training set.

Figure 1 presents the prediction curve of the training model for SiO₂ and the prediction curve obtained for the CCT on Mars with this model. The linear regression lines have been indicated as well as the y=x line where compositions agree. Note that the CCT with the highest SiO₂ content (macusante 73.7 wt%) is well predicted though not present in the training set. Similar
curves have been generated for all major elements. The calibration curves follow well the actual compositions of the samples. For some elements for which few high content samples exist or with strong self-absorption, the calibration curves tend to underestimate the highest contents (Fe$_2$O$_3$, TiO$_2$ and alkali).

Finally a comparison with APXS [10, 11] has been performed on martian targets as much as possible, keeping in mind that both instruments probe different surfaces of the targets (few 100 $\mu$m$^2$ for ChemCam vs. few cm$^2$ for APXS) and that ChemCam data analysis removes the first few shots taken on a target as these correspond to the omnipresent dust covering the rocks. In table 1, we present a comparison between soils measured by APXS and ChemCam (Portage, sol 89). Considering the errors mentioned in Table 1 and possible underestimation for Fe$_2$O$_3$ and TiO$_2$, the results are quite comparable. See also [12, 13] for other ChemCam to APXS comparisons.

All these tests give us confidence in the fact that the PLS results give reasonable predictions for the chemical content of the martian targets, and that in the few cases of absolute value biases, at least the trends calculated do follow an actual composition change between targets.

**Conclusion:** We have demonstrated that PLS is a technique that generates good predictions of the chemical composition of Mars targets from the ChemCam LIBS data. In the near future, we are intending to improve both the accuracy and precision of the predictions by modifying the algorithm and databases used.

**Acknowledgments:** The instrument was developed in the US under the NASA Mars Program Office support to MSL. Work on PLS was also supported by LANL internal funding, in the instrument was supported in France at IRAP, CNES, and Thales under funding from CNES.

**References:**