

DATA FUSION IN PLANETARY LIBS+RAMAN SPECTROSCOPY. P. Sobron,^{1,2} G. Lopez-Reyes,³ A. Sansano³, J. A. Manrique³, F. Rull³. ¹SETI Institute, 189 Bernardo Ave., Suite 100, Mountain View, CA 94043, ²MalaUva Labs, St. Louis, MO 63104 (psobron@seti.org), ³Unidad Asociada UVA-CSIC-Center of Astrobiology, Valladolid, Spain.

Advantages of combined LIBS/Raman analysis:

Laser-induced breakdown spectroscopy (LIBS) is as a powerful tool for geochemical investigations of rocks and regolith, and for exploratory surveys in Mars with MSL's ChemCam instrument suite [1-3].

Laser Raman Spectroscopy (LRS) is one of the most powerful tools available for in-situ molecular identification and mineralogical characterizations: (i) a Raman/Fluorescence instrument was listed in strawman payloads for the MAXC mission concept study and for the Mars2020 mission; (ii) an in-situ laser Raman instrument (RLS) remains a priority science instrument in ESA's 2018 ExoMars mission, surviving substantial payload rearrangements; (iii) the MatISSE program is currently funding the maturation of a Compact Integrated Raman Spectrometer (CIRS).

The advantages of combining LIBS and LRS techniques for the analysis of a given sample are evident: (1) Both can share most hardware components [4]; (2) LIBS can reveal the relative concentration of major (and often trace) elements present in a bulk sample; and (3) Raman yields information on the individual mineral species and their chemical and structural nature. Thus combining the data from both tools enables definitive mineral phase identification with precise chemical characterization of most major and minor and some trace mineral species.

In planetary surface exploration, a combined LIBS/LRS instrument can provide a rapid mineralogical/chemical evaluation of the target that will be critical for selecting samples to be eventually collected for sample return purposes, for selecting sample sites to be drilled in the search for other species (*e.g.*, organics), or for enabling the implementation of more advanced, unsupervised exploration and operational routines in landed spacecrafts.

Prototypes of combined LIBS/LRS instruments have already been tested [4]. However, when LIBS and LRS spectral data are used separately, the information that can be retrieved from a target is often limited, particularly in planetary and planetary analogue applications, where targets are often complex geological materials [5,6].

Data fusion: New approaches for data processing are needed in order to find tools for processing large amounts of LIBS+LRS data efficiently and, by means of data fusion, obtain a single result. Data fusion often refers to the acquisition, processing, and combination of data. Data is handled in a way that optimizes the

utilization of the information provided by two or more sensors. This synergistic usage of data from a target aims to providing a more informed interpretation of the target than that achieved by any of the sensors were those used individually.

The use of data fusion for boosting the performance of LIBS and LRS for the detection of explosives has been reported [7]. Here we explore the use of low-level data fusion (spectral concatenation) for improving the analytical capabilities of LIBS and LRS in planetary exploration. Specifically, we evaluate the performance of three principal-component analysis (PCA) models to determine the absolute abundance of two sulfate salts in binary mixtures: a model developed with LIBS data; a model developed with LRS data; and a model developed with fused LIBS/LRS data. We demonstrate that a PCA-model that combines fused LIBS and LRS spectral data yields a more robust and accurate quantitative analysis of mixtures of sulfates than models that utilize LRS or LIBS data in isolation.

Methods:

Samples: MgSO₄ and NaSO₄ were synthesized in the laboratory. A set of 9 binary mixtures of these salts were prepared. The mixtures proportions are 1:0, 0.99:0.01, 0.9:0.1, 0.75:0.25, 0.5:0.5, 0.25:0.75, 0.1:0.9, 0.01:0.99 and 0:1. Throughout the experiments, the samples were checked for transformation through LRS and XRD – no sample transformation was observed.

LIBS spectral acquisition: LIBS spectra were recorded using an Andor Mechelle 5000 spectrometer equipped with an iStar ICCD. The laser source is a Litron NANO SG 120-20 Nd-YAG "Q-switched" with 50mJ a 532nm, 25Hz with 8ns pulse. The spectral range analyzed with our instrument is 170-1100 nm. For statistical purposes, a total of five spectra were recorded from each sample.

LRS spectral acquisition: LRS spectra were recorded using a Nikon Eclipse E600 microscope coupled to a KOSI Holospec f/1.8i spectrometer. The excitation source is a laser REO LSRP-3501, He-Ne 632.8 nm laser. The detection was performed with a CCD Andor DV420AOE-130. For convenience, the spectral range was limited to the 130-1980 cm⁻¹ region; this region contains all of the spectral peaks associated with sulfate vibrational modes. A total of five LRS spectra were recorded for each of the samples.

PCA models: All of the three PCA models are trained based only on the pure samples, and validated with the mixtures, as described in [8,9].

Low-level fusion: LIBS and LRS spectra were baseline-corrected and normalized. For each of the samples, LIBS and LRS spectra were concatenated.

Software: All of the routines for spectral processing and PCA modeling were implemented in Mathworks MATLAB.

Results and discussion: Figures 1 to 3 show the scores representation in one dimension (arbitrary units) for the three different PCA models. The PC1 distribution in the figures shows that PCA is capable of differentiating different proportions when only trained with pure endmembers, in agreement with [8,9]. Qualitatively, cluster distribution and separation is better for the fused data model (Figure 3) than for LRS-only and LIBS-only models. Furthermore, the correlation between the clusters distribution and the relative proportion of the corresponding mixture is clearly better for the fused data model – increasing PC1 values correspond to increasing Mg-sulfate relative concentration. This difference is critical for the quantification of salt abundance at low concentrations.

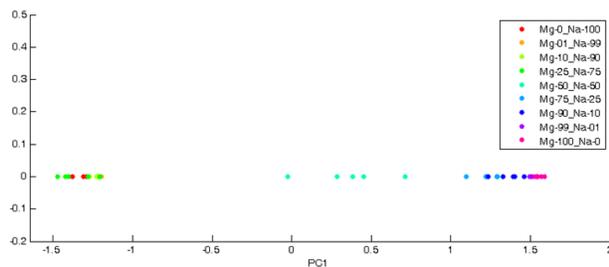


Figure 1 – PC1 score distribution for LRS-only model

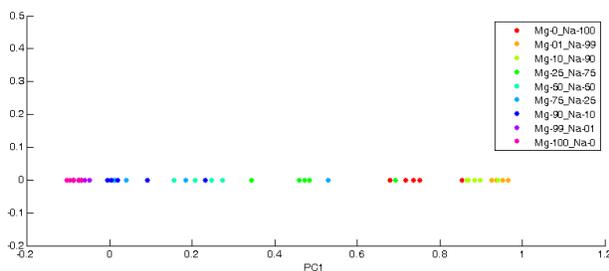


Figure 2 – PC1 score distribution for LIBS-only

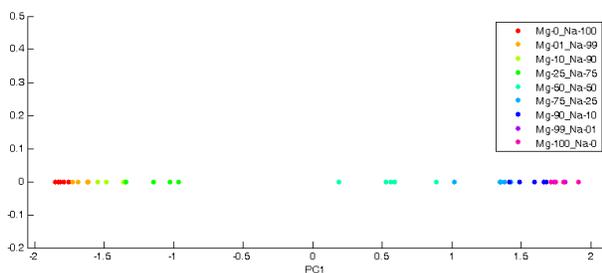


Figure 3 – PC1 score distribution for the fused data model

In all these observations indicate a better predictive power of the fused data PCA model; the use of fused LIBS and LRS spectral data clearly enhances the behavior of the models, reducing the number of PCs and the model complexity – more components would be needed in both LIBS-only and LRS-only in order to achieve comparable predictive power, thus reducing model parsimony, *i.e.*, we risk retaining variables that are unnecessary for prediction.

A caveat to this discussion is that we have used just two endmembers, MgSO_4 and NaSO_4 , thus, the models are rather simplistic. For the data used in this work, the difference in the number of components needed by the three models for accurate prediction is not great, and the three models seem to pick the same variable to predict the composition of our binary mixture. That may not be true for ternary, quaternary, etc. mixtures.

Conclusions and implications: Low-level data fusion (spectra concatenation) is the simplest way to perform data fusion; no features were extracted from the individual streams of data, thus these had to be preprocessed prior to applying chemometric methods (PCA). A limitation to low-level data fusion is that non-reduced data may become redundant, which may lead to overfitting or poor model performance.

Despite these limitations, we have demonstrated that the quantification of the abundance of salts in binary mixtures was improved through using low-level data fusion. Qualitatively, the PCA model developed with fused data is more robust for predicting salt abundance in our binary mixtures.

These are encouraging results, and warrant further development of data fusion strategies for LIBS and LRS data. These strategies may be available for the Mars2020 mission and, through the synergistic utilization of LIBS and LRS data we propose, may contribute to a better understanding of Mars than would be possible using both techniques in isolation.

Presently, we are adding additional salts to our binary mixtures and exploring mid- and high-level data fusion for advanced processing of LIBS+LRS data.

Acknowledgements: this work was funded by NASA Mars Fundamental Research Program under grant # NNX13AJ18G.

References: [1] Maurice S. et al. (2013) *EGU General Assembly Conference Abstracts*, p 14161. [2] Wiens R.C et al. (2012) *Space Science Reviews*, 170, 167-227. [3] Williams R.M.E., et al. (2013) *Science*, 340, 1068-1072. [4] Sharma S.K. et al. (2009) *Spectrochimica Acta Part A*, 73, 468-476. [5] Sobron P., and Wang A. (2012). *Journal of Raman Spectroscopy*, 43, 212-227. [6] Sobron P. et al. (2010) *LPSC*, p2585. [7] Moros J., and Laserna J.J. (2011) *Anal Chem*, 83, 6275-85. [8] Lopez-Reyes G. et al. (2014) In *Mathematics of Planet Earth*, p. 135-139. [8] Lopez-Reyes, G., et al. (2013) *LPSC*, p2135.