

CONCEPTS OF LIBS AND RAMAN DATA FUSION FOR IN-SITU MEASUREMENTS OF MARS RELEVANT SALTS WITH COMPACT INSTRUMENTATION K. Rammelkamp^{1,2}, S. Schröder^{1,2}, S. Kubitzka¹, D. S. Vogt¹, S. Frohmann¹, P. B. Hansen¹, U. Böttger¹, F. Hanke¹ and H.-W. Hübers^{1,3}; ¹German Aerospace Center (DLR), Institute for Optical Sensor Systems, Berlin, Germany; ²Institut de Recherches en Astrophysique et Planétologie, Toulouse, France; ³Humboldt-Universität zu Berlin, Institute of Physics, Berlin, Germany; kristin.rammelkamp@irap.omp.eu

Introduction: LIBS and Raman spectroscopy are promising techniques for the in-situ robotic exploration of extraterrestrial bodies [1,2]. The first LIBS instrument in space is the ChemCam instrument on NASA's Mars science laboratory (MSL) which is successfully measuring the Martian surface in Gale crater since landing in 2012 [3]. Partly due to the great success of ChemCam, its enhanced follow-up instrument SuperCam got selected for NASA's Mars 2020 mission and will combine for the first time LIBS and Raman spectroscopy besides other techniques in one space exploration instrument [4]. Both techniques return valuable information on their own but their combination is even more promising as their data is complementary: With LIBS, the elemental composition of a target can be derived while from Raman data, information about the molecular and lattice structures can be obtained.

Due to the complementary nature of LIBS and Raman data [5], they are ideal candidates for the application of data fusion. One can distinguish different levels of data fusion [6] and two of them will be compared in this study: Low-level data fusion refers to a combination on the data level as the spectra are concatenated and then analyzed in this form. High-level data fusion, on the other hand, already extracts information from the individual datasets and combines these for the final identity declaration. Previous studies have already shown that data fusion in general can increase the scientific return compared to the LIBS and Raman only datasets for terrestrial applications [7,8] as well as for extraterrestrial scenarios [9,10]. In this study, we present results of low-level and high-level data fusion from LIBS and Raman spectral data of Mars relevant salts which were measured with miniaturized setups simulating the performance of lightweight (≈ 2 kg) space exploration instruments. In order to evaluate the benefit of data fusion, outcomes from the individual data analysis were compared with those from the combined analysis.

Samples and experimental setup: Two sets of samples were investigated for this study. The first one contains samples of five pure sulfates (BaSO_4 , $\text{CaSO}_4 \times 2\text{H}_2\text{O}$, K_2SO_4 , $\text{MgSO}_4 \times 7\text{H}_2\text{O}$, Na_2SO_4) and binary mixtures of them in 50:50 wt% and 30:70 wt% mixing ratios. For the other set, four types of salts (carbonates, chlorides, perchlorates, sulfates) each with the cations Ca, Mg, and Na were mixed with the high-alkali basalt GBW 07105 in 30:70 wt % ratios. All samples were

measured in form of pellets (14 mm diameter) which were pressed from powders with a pressure of 5 t.

Data was taken with LIBS and Raman setups at DLR Berlin which are built of miniaturized components. For LIBS, a Nd:YLF laser (1053 nm, 2 ns pulse length, 2 mJ/pulse, *neoLASE*) and four (UV, VIS1, VIS2, NIR) small sized spectrometers ($95 \times 68 \times 20$ mm, 175 g, $\Delta\lambda \approx 0.15$ nm, *Avantes, AvaSpec-Mini*) are employed. As the LIBS plasma is sensitive to atmospheric conditions the samples were measured in a small chamber in simulated Martian atmospheric conditions (7 mbar of mainly CO_2). The Raman setup consists of a frequency doubled Nd:YAG laser (532 nm, continuous wave, 22 mW output, *Oxxius, LCX-532*) and one spectrometer ($89 \times 64 \times 32$ mm, 265 g, $\Delta\nu \approx 25$ cm^{-1} , *Ocean Optics, Flame-T*). All spectrometers are Czerny-Turner spectrometers with uncooled detectors.

Data Analysis: For the data analysis, mainly multivariate data analysis (MVA) techniques such as principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA) were applied. For the low-level data fusion, the MVA models were evaluated by means of their cluster quality in PCA space derived from distance and spread calculations, and their predictive capabilities. For the latter, two approaches were investigated: First, the data of the binary sulfate mixtures were projected into PCA models which were built using only the data of the pure sulfates. From distance calculations between the scores of the projected spectra and the pure sulfate clusters, the two constituents of the mixtures were predicted and it was counted how often and to what extent the predictions were correct. Second, in PLS-DA of the salt/basalt mixtures, the root mean square error (RMSE) also known as prediction error was evaluated for different models as it serves as a figure of merit for predictive capabilities. For the high-level data fusion, PCA models of the data from selected samples from the salt/basalt mixtures were computed and the data of the not considered samples was projected into the models in order to derive the cation of the salt by the LIBS data and the anion by the Raman data, respectively. Furthermore, similarity rates from Pearson coefficients were derived as another approach for the ion identification.

Results from low-level data fusion: In this section, we will briefly summarize the outcomes from the low-level data fusion approaches for which more details can

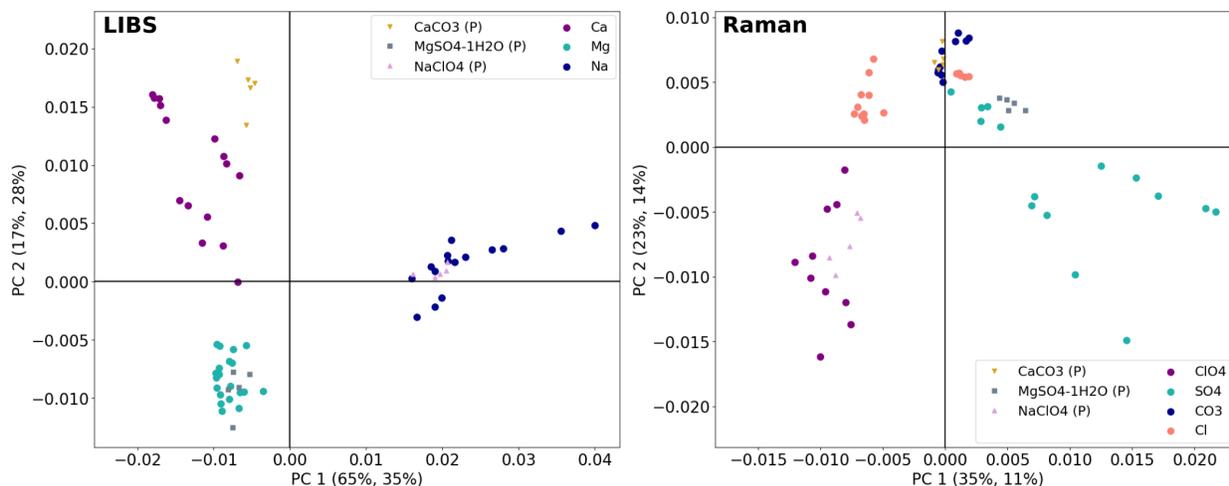


Figure 1: PC 1/PC 2 score plots of LIBS (*left*) and Raman (*right*) PCA models from salt/basalt mixtures with projected spectra (*P*) and explained variance in brackets. It has to be noted that PC 1-4 were used for calculations. For the LIBS model, the scores of the projected spectra locate close to the clusters from samples with the same cation. In the Raman model, clear clusters cannot be identified, however, the scores of the projected spectra are close to those of spectra from samples with the same anion.

be found in [11]. First to mention is that more correct identifications of the constituents of the binary sulfate mixtures were found with the combined LIBS and Raman model. However, the best cluster quality of the pure sulfate clusters was never given in the combined model, most likely because larger signal fluctuations in the LIBS data increased the spread of the clusters compared to the Raman only model. From the data of the salt/basalt mixtures, PCA as well as PLS-DA models were computed. On average, the cluster quality was best in the PCA model of the combined data. There was, however, a variation of the cluster quality among the samples and for several samples, the Raman only model revealed the best cluster quality. Thus, the gain due to low-level data fusion is not significant especially when compared to the Raman only model. Promising results were obtained from the PLS-DA that was implemented to predict class membership of different classes, which were defined by the anion, the cation, and the single salts. As expected, the LIBS data is best suited to derive the cation of the salts while Raman data can be best used for anion identification. For the single sample groups as response, the combined data is best suited for predicting the class membership. In general, it has to be mentioned that the comparability of the individual models with the combined model was limited due to different models featuring different optimum dimensions. However, the overall observations reveal that the gain of low-level data fusion is small compared to the larger amount of processed data in the combined models.

Results from high-level data fusion: The investigated high-level data fusion approaches for the identification of the salt in the basalt have shown that the

cation identification with the LIBS data is straightforward with either calculating distances in PCA models or with deriving similarity rates. Regarding the anion identification with Raman data, a few wrong assignments were observed, which is most likely due to less defined clusters in the Raman PCA models and larger differences between Raman spectra of the same anion. Nevertheless, Raman spectra might be better suited to identify single samples which means that more samples are needed for calibration. Exemplarily, score plots of LIBS and Raman PCA models are shown in Figure 1.

Summary: The present investigation has shown that the level of data fusion for the combination of LIBS and Raman data has to be chosen carefully. Although low-level fused data contains more information than the individual datasets, the scientific return did not increase significantly in this study. Regarding the salt identification, it was observed that a separate identification of the cation and the anion by the LIBS and the Raman data, respectively, is better suited. Since the best way of data combination depends on the specific data and the scientific objective, a dedicated and individual analysis approach should be carefully developed particularly for a specific instrument.

References: [1] Knight et al. (2000), *Appl. Spec.*; [2] Wang et al. (1995), *Journ. of Geophys. Res.*; [3] Maurice et al. (2016), *Journ. of Analy. Atom. Spectrom.*; [4] Wiens et al. (2017), *Spectroscopy*; [5] Wiens et al. (2005), *Spectrochim. Acta Part B*; [6] Borrás et al. (2015), *Analy. Chimica Acta*; [7] Hoehse et al. (2012), *Analy. and Bioanaly. Chem.*; [8] Moros and Laserna (2015), *Talanta*; [9] Sobron et al. (2014), *45th LPSC*; [10] Moros et al. (2018), *Analy. Chem.*; [11] Rammelkamp et al. (2019), *Journ. of Raman Spec.*