EPSC Abstracts
Vol. 7 EPSC2012-577-1 2012
European Planetary Science Congress 2012

Application of multivariate analysis techniques for the identification of sulphates from Raman spectra – Implications for ExoMars

G. Lopez-Reyes (1), P. Sobron (2), B. Lafuente (1) and F. Rull (1)
(1) Unidad Asociada UVA-CSIC-Center of Astrobiology, Valladolid, Spain. (2) Space Science and Technology. Canadian Space Agency, St. Hubert, Canada (guillermo.lopez.reyes@cab.inta-csic.es / Fax: +34-983-140514)

Abstract

We present results of the application of Principal Component Analysis (PCA), Partial Least-Squares Regression (PLSR) and Artificial Neural Networks (ANN) for the identification of sulphates in pure and mixed samples. The studied samples include Fe-, Mg-, Ca-, and Na sulphates.

1. Introduction

The Raman instrument (RLS) onboard the 2018 ExoMars rover will determine the structural and compositional features of martian surface and subsurface samples. In its current configuration, the samples will be crushed by the sample preparation and distribution system and delivered to the analytical laboratory. In such a scenario, with the loss of geological context, unambiguous identification of the mineral phases present in the geological targets and accurate quantification of their abundance become a priority task.

Prior attempts to use multivariate techniques for the quantitative analysis of the Raman spectra of different materials show that PCA is capable of clearly differentiating mineral species in geological samples [1]. PLS has been used to determine the quality of biodiesel fuels [2]. ANN have been designed for the identification and quantification of inorganic salts in water solutions [3]. The aim of this study is to evaluate these multivariate techniques for the analysis of a set of mineral samples in the framework of the operation of the RLS instrument. We used sulphates because they are one of the two major types of secondary minerals found on Mars that may provide potentially habitable environments. Due to the association of sulfate salts with ancient aqueous environments in which life might have thrived, it is expected that sulphates will become priority targets for the ExoMars mission.

2. Spectra set

A total set of 17 spectra of sulphates were used as input for training all the three techniques. The set was divided in four subsets, each grouping sulphates with different hydration state: $FeSO_4$ (1, 4 and 7w), $MgSO_4$ (1-7w, and 12w), $CaSO_4$ (0, ½ and 2w) and Na_2SO_4 (0 and 10w). A set of mixed spectra was synthetically generated by computing linear combinations of the original spectra, parameterized with the expected proportion of the mixture and the cross-section of the mixed materials. Random noise was added to the synthetic spectra to guarantee differentiation. This set of mixtures was used to validate the models.

3. PCA model

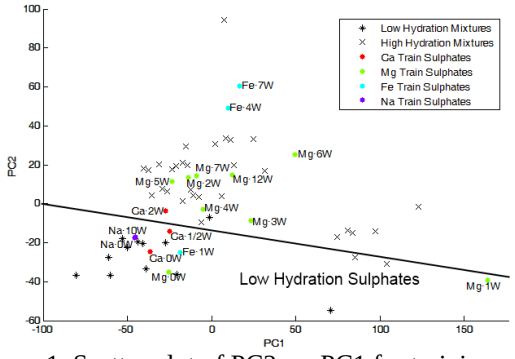

Figure 1. Scatter plot of PC2 vs. PC1 for training (colored) and validation sets (black) with PCA.

PCA analysis of the samples showed that the first two components, PC1 and PC2, represent ~80% of the variance in the data set: 61% and 17%, respectively. The scores are plotted in Figure 1. Dehydrated or poorly hydrated salts are somewhat differentiated from the highly hydrated salts, for both the training and test samples.

4. PLSR model

The PLSR responses were chosen to indicate the hydration level and the relative abundance of the

various cations. The criterion to optimize the model behavior was to evaluate the Mean Square Error (MSE) of the difference between the expected regression responses and those fitted by the model.

The optimum number of components in the PLSR model was found to be 7. Good prediction (linear unitary slope and 93% mean correlation coefficient for all the responses) was found between the expected and fitted responses of the training samples (pure salts).

However, the validation tests showed a non-unitary slope linear correlation between the expected and the fitted responses. The averaged correlation coefficient for all the responses was 86%. We conclude that while the model does not predict hydration states and cation abundances for the mixed salts, it provides a good linear calibration curve which allows computing these values (Figure 2).

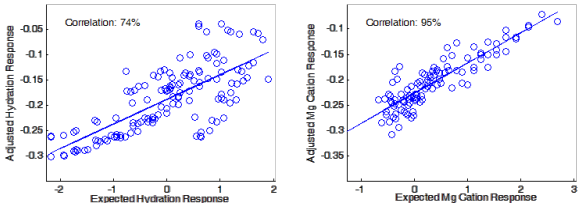

Figure 2. Calibration curves for the hydration state and Mg abundance responses. Values are scaled and centred.

5. ANN model

We have designed a three-layer Feed-Forward Back-Propagation network with 31 neurons in the hidden layer and log sigmoid neuron transfer function.

The input data for the ANN consisted on a relatively small number of Raman intensities at selected wavenumbers. Three different input sets were fed into the network: (1) pure sulphates for training, (2) pure sulphates for validation, and (3) pure and binary mixtures spectra, in different proportions for testing. Results show that the network is able to achieve very high levels of accuracy in the detection of materials. The pure sulphates were detected with 100% accuracy, while the mixtures show varying results depending on the relative abundances (Figure 3). These preliminary results allow defining a detection window in which we achieve a 75% positive identification index when the sulphates are mixed in a 10-90% ratio.

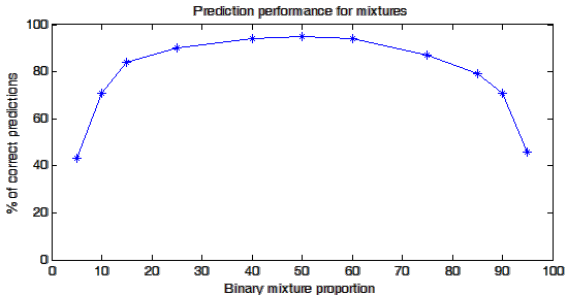

Figure 3. Prediction accuracy for the ANN model.

6. Summary and Conclusions

The use of multivariate techniques as a tool for the interpretation of Raman spectra in the context of the RLS instrument has been investigated. The preliminary results presented here are encouraging, and demonstrate that these techniques can provide critical information for the identification and quantification of mineral phases in geological samples. Future work will focus on including quantification routines for ANN and on incorporating additional mineral samples (phyllosilicates, carbonates, igneous rocks, etc.) to our models.

The application of unsupervised multivariate techniques for the processing of RLS products is a must in order to provide science support to the ExoMars mission. These techniques will provide fast identification of materials and quantification of mineral species during the tactical operations of the rover-based mission.

Acknowledgements

A. Wang provided the Raman spectra used in this work. C. Lefebvre and A. Koujelev provided helpful insights on the implementation of ANN.

References

[1] Lafuente, B. et al: Multivariate analysis of Raman spectra for geological classification and identification: Application to Exomars Mission, Georaman X, 11-13 June 2012, Nancy, France, 2012.

[2] Ghesti et al: FT-Raman Spectroscopy Quantification of Biodiesel in a Progressive Soybean Oil Transesterification Reaction and Its Correlation with 1H NMR Spectroscopy Methods, Energy & Fuels, 21 (5), pp. 2475-2480

[3] Burikov et al: Application of ANNa to solve problems of identification and determination of concentration of salts in multi-component water solutions by Raman spectra, Optical Memory & Neural Networks, Vol. 19, pp. 140-148.