Achieving Data Reduction in Space by Applying Unsupervised Machine Learning to Mass Spectrometric Data

S. Gruchola[1], M. Tulej[1], P. Keresztes Schmidt[1], N. J. Boeren[1,2], L. N. Knecht[1], A. Riedo[1,2], P. Wurz[1,2]

[1]Physics Institute, Space Research and Planetary Sciences, University of Bern, Switzerland, [2]NCCR PlanetS, University of Bern, Switzerland

With the ever-increasing volume of data generated by space missions, the limitations in data downlink rates to Earth have become apparent. As the demand for scientific insights into other planetary bodies in our Solar System continues to rise, it becomes crucial to explore methods, which ensure that the data that are sent back are pertinent to the research objectives. Through optimisation of on-spacecraft data preselection, scientific returns can be maximised to ensure an efficient use of limited bandwidth resources.

In this contribution, we introduce a data reduction method based on unsupervised machine learning and clustering applied to mass spectrometric data. The input data are clustered into groups of mass spectra with similar chemical composition, corresponding to different compounds that are present in the dataset. The chemical and/or mineralogical identification of the compounds is performed on ground after the reception of representative mass spectra of each cluster. Two different yet related clustering algorithms called UMAP and densMAP are compared, as well as different levels of data pre-processing. As the method is to be applied on spacecrafts, where computing resources are limited, pre-processing should be minimal. A similar version of this clustering method was previously applied to a 1.88 Ga Gunflint sample (Ontario, Canada) to separate mass spectra recorded from the host (chert) from mass spectra containing signatures of organic matter from fossilised microbes [1].

To study the effectiveness and reliability of our machine learning procedure, data were collected using a space-prototype mass spectrometric instrument. The instrument is a compact and lightweight laser ablation and ionisation mass spectrometer (LIMS) with a femtosecond laser ion source and a reflectron-type time-of-flight (RTOF) mass analyser for the separation of ions. The instrument reaches mass resolutions exceeding 600 and lateral spatial resolutions down to 10 µm. The chemical composition of solids can be determined down to the level of mg/kg concentrations over the full mass scale [2].

The sample selected for the study is a 2.06 Ga apatite crystal obtained from an ultramafic phoscorite rock from the Phalaborwa Carbonatite Complex (Limpopo Province, South Africa). Being an accessory mineral in igneous and other rocks, apatite commonly contains a range of rare earth elements (REE), which provide valuable information for investigating physical and chemical conditions in igneous rocks and the volatile evolution of magmas [3]. By applying the clustering algorithms to the collected mass spectrometric data, three major phases corresponding to fluorapatite, forsterite and calcite were found. Additionally, several micrometre sized inclusions of baddeleyite, uranothorianite, rutile and others were detected. Given that the data were recorded using a miniature mass spectrometer designed for space flight, this analysis demonstrates the analytical capabilities of our LIMS system that could be achieved *in-situ* on other planetary bodies in our Solar System, for example on the Moon or on Mars.

[1] R. A. Lukmanov et al., *Frontiers in Space Technologies*, **2022**, 3, 10.
[2] A. Riedo et al., *Journal of Mass Spectrometry*, **2013**, 48, 1.
[3] M. Tulej et al., *Universe,* **2022**, 8, 410.