RECOGNISING PLANETARY ROCKS AND MINERALS BY COMBINING A CUSTOM MINERALOGICAL DATABASE WITH DEEP LEARNING BASED MULTI-SPECTRAL UNMIXING.

I. Drozdovskiy[1], F. Sauro[2], S.J. Payler[1,3], S. Hill[4], P. Jahoda[4], K. Jaruskova[4], F. Venegas[4], A. Angellotti[4], M. Franke[4], P. Lennert[4], G. Ligeza[4], P. Vodnik[4], L. Turchi[1,5], L. Bessone[1] [1]Directorate of Human and Robotics Exploration, European Astronaut Centre (EAC) - European Space Agency (Linder Höhe, D-51147 Cologne, Germany; igor.drozdovskiy@esa.int), [2]Geological and Environmental Sciences, Italian Institute of Speleology - Bologna University, [3]Agenzia Spaziale Italiana, Rome, Italy, [4]ESA-EAC, CAVES & PANGAEA interns, [5]Spaceclick Srl, Milan, Italy.

Introduction: The ESA-PANGAEA Mineralogical Toolkit is a set of data analytics tools aiming to enhance the recognition of planetary minerals. It includes a custom structured database called the PANGAEA Mineralogical Database, which contains information on all known minerals found on the Moon, Mars and other planetary bodies [1]. This database then serves as the basis for a set of spectral classification methods using machine learning designed to perform in-situ spectroscopic identification of minerals [2]. Developed and tested together in the context of ESA's astronaut field science training using analogue environments, PANGAEA, the mineral library and recognition software are conceived as a real-time decision support tools for future planetary surface exploration missions.

PANGAEA Mineralogical Database: The Mineralogical database [1] can be viewed as two distinct products: a catalogue of petrographic information and an analytical library. The catalogue consists of petrographic information on all currently known minerals identified on the Moon, Mars, and found primarily, or exclusively, within meteorites. The catalogue is envisioned to provide essential analytical in-field information for each mineral to assist in rapid identification and understanding of significance in real time geological exploration. Each mineral entry includes: IMA recognized Name, Chemical Formula, Mineral Group, surface abundance on planetary bodies, geological significance in context of planetary exploration (e.g. occurrence, environmental conditions, marker for important processes), number of collected VNIR and Raman spectra, their spectral discoverability and the possible spectral features. In addition, supplementary characteristics for each mineral that may help with its identification are included, such as chemical abundances calculated from known empirical chemical formula, as well as basic physical properties such as hardness, specific gravity, crystal system. The database was compiled through systematic literature research, followed by the careful cross-validation ("out-of-sample" testing) of all mineral characteristic information, including flagging of doubtful or erroneous data. The second major contribution, provided by the PANGAEA Mineralogical Database, is a customized library of analytical data from all known planetary terrestrial analogue minerals Fig. 1). This covers four analytical methods: reflective Visual-to-Near- & Shortwave-Infrared (VNIR), Raman vibrational (molecular) spectroscopy, Laser-Induced-Breakdown (LIBS), and X-Rays Fluorescence (XRF) atomic spectroscopy. This library also includes a set of standard spectra, which is used for evaluating the detectability of minerals with different analytical methods.

Part of the archive consists of spectra collected from available open access on-line catalogues, such as RRUFF (Raman), and USGS, RELAB and ECOSTRESS (VNIR). It also includes our/our collaborators own bespoke spectroscopic measurements (VNIR, Raman, LIBS & XRF) of planetary analogue minerals taken from different collections, and synthetic spectral libraries, such as LIBS NIST; see [2] and references therein. Only high-quality spectra of confirmed mineral samples were included, determined by the quality flag in the original database or by our own statistical evaluation of the within-class spectra. We also removed outliers from all single-method spectra of each mineral by finding the weighted average spectrum for each class and removing those that significantly deviated from the average [2]. This was to ensure the set was not skewed by random instrumental artefacts or sample misclassification. This multispectral library is designed to be used for the recognition of planetary materials, and acts as a training dataset for our mineral recognition software.

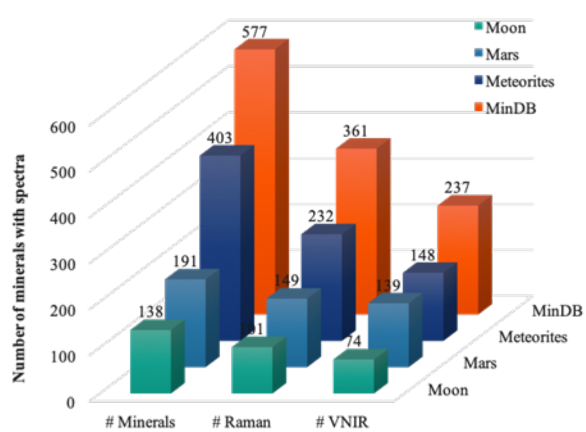

Figure 1: The current census of the minerals with archived Raman or VNIR spectra in the PANGAEA mineralogical database.

Machine Learning (ML) software for recognition of minerals from multispectral data: To utilize the PANGAEA Mineralogical Database for identifying minerals in the field from the output of analytical tools, we also developed identification methods that combine types of material characteristics, mineral structure (obtained with VNIR and Raman spectra) and its chemical composition (derived from XRF and LIBS spectra). To achieve this, we chose to use a ML-based approach. This was for several reasons: ML is fast and accurate when developed properly, can handle multimethod datasets, and the accuracy can be progressively improved by adding new training data to the classification models without losing the recognition speed. To maximise the accuracy of the ML methods, we evaluated various Machine Learning approaches used to identify mineral species from single analytical methods (Raman, VNIR or LIBS), and developed a flexible and modular algorithm that can classify minerals either from standalone spectroscopic methods, or using a combination thereof. The flow diagram detailing this methodology is shown in Figure 2. Our new approach was then evaluated using our own internal archive of analytical data, as well as in some cases, other publicly available spectroscopic datasets. Our cross-validation tests show that multi-method spectroscopy paired with ML paves the way towards rapid and accurate characterization of minerals [2] (see Figure 3), as well improving the quantification of mineral abundances in rocks and soils using ML-based spectral unmixing.

PANGAEA Mineralogical Tookit as an Analytical Toolset for Moon Surface Exploration: The PANGAEA Mineralogical Toolkit is envisioned as a part of the PANGAEA Electronic Fieldbook Suite (EFB) [3]. The EFB is a deployable system that enhanced scientific outcome of field mission operations, enabling scientific documentation of field traverses, sampling and interaction with remote science support teams. The EFB can interface with handheld spectrometers intended for planetary exploration, simultaneously feeding their measurements into the embedded Mineralogical Toolkit. Combined within the EFB Tool Suite with various spectral analytical tools, and benefiting from its instrument agnostic nature the PANGAEA Mineralogical Toolkit will enable fast and reliable in-situ recognition of rocks and minerals, thus becoming a crucial decision support tool for future human and robotic planetary surface exploration missions.

References: [1] Drozdovskiy, I. et al. 2020, *Data in Brief*, 31, 105985. [2] Jahoda, P. et al. 2020, *The Analyst*, 146(1), pp.184-195. [3] Turchi L. et al. *Planetary Space Science*, 197, p.105164.

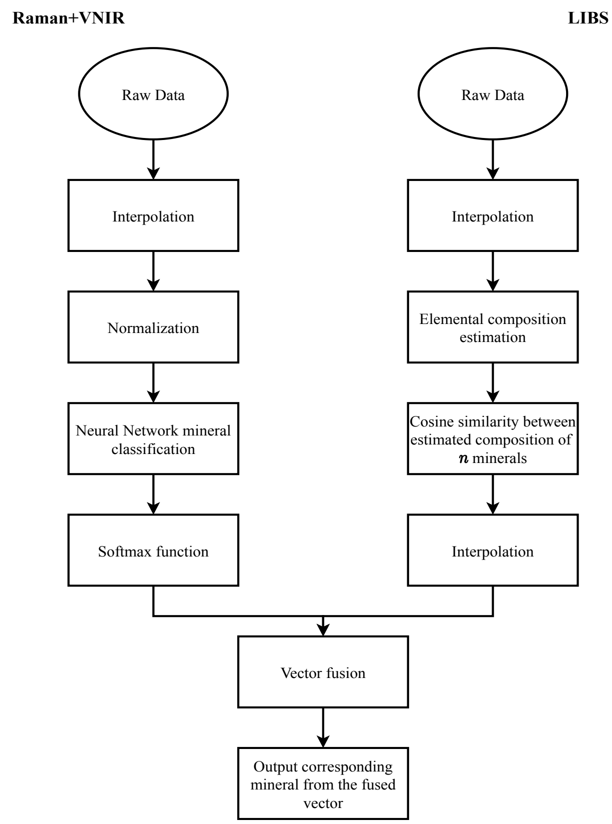

Figure 2: Simplified flow diagram showing our method for recognizing minerals from combined Raman/VNIR and LIBS spectra.

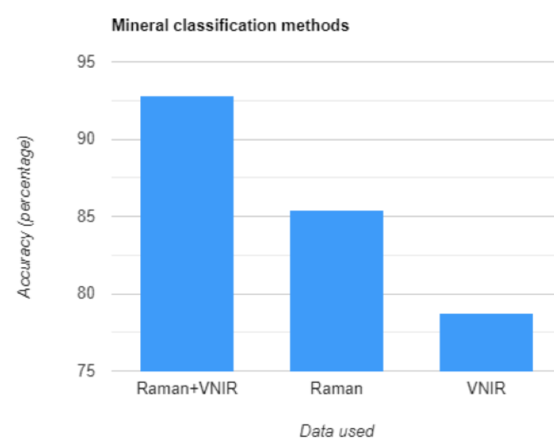

Figure 3: Better mineral predictions rates from combined Raman and VNIR spectra than from single-method spectroscopy.