DEVELOPMENT OF AN OPEN SOURCE TOOL TO PERFORM UNSUPERVISED CLUSTERING OF CRISM/MRO DATA ON MARS: FIRST RESULTS. B. Baschetti[1,2] M. D'Amore[3], C. Carli[2], M. Massironi[1], F. Altieri[2]; [1]Department of Geosciences, University of Padova, Italy (beatrice.baschetti@phd.unipd.it), [2]INAF-IAPS, Rome, Italy. [3]German Aerospace Centre (DLR), Berlin, Germany.

Introduction: Remotely sensed spectral and hyperspectral data provide essential information on the composition of planetary surfaces. On Mars, high resolution hyperspectral data are provided by the CRISM instrument (Compact Reconnaissance Imaging Spectrometer for Mars) [1], onboard NASA's MRO (Mars Reconnaissance Orbiter) spacecraft. CRISM can collect hyperspectral cubes in the 0.4 - 4 micron range, with a spectral sampling of 6.55 nm/channel and a spatial resolution of 18.4 meter/pixel at 300 kilometer altitude on targeted scenes. A CRISM scene is traditionally explored through RGB maps of spectral parameters, such as band depth. To guide the user in this work, the CRISM team provided a set of 60 standard spectral parameters [2], which were identified based on the known spectral variability of the planet. After a first assessment with this method, extraction of single or mean spectra from selected regions of interest is usually performed. This is a solid approach, however, it does not fully exploit the potentials of a hyperspectral dataset, due to its focus on a few portions of the available spectral range at once. Machine learning techniques such as unsupervised clustering can help us explore CRISM datasets more efficiently.

Here we present the results from the development of a Python framework that allows the application of two different unsupervised clustering techniques (k-Means and Gaussian Mixture Models) to the CRISM dataset. Results on specific CRISM data are shown and evaluated, discussing the impact of different processing steps and comparing the performance of the two algorithms used.

Dataset and methods: We used the CRISM Map Projected Targeted Data Records (MTRDRs) [3], corresponding to the most advanced, ready-to-use CRISM dataset available.

The tool has been developed in Python using Jupyter notebooks, which allow great flexibility and interactivity, implementing the literate programming paradigm. The clustering algorithms are from the Machine Learning library Scikit-Learn [4]. A combination of linear (Principal Component Analysis, PCA) and nonlinear (Uniform Manifold Approximation and Projection, UMAP) [5] dimensionality reduction techniques were employed to ensure a correct interpretation of the data's structures and patterns. Several tested combinations are listed in **Table 1**. In some cases, labeled with an asterisk (*), the first component of PCA is discarded. This choice was driven by the fact that the first principal component, by capturing most of the data variance, mainly correlates with the average reflectance of the surface, often dependent on the morphology and topography of the terrain, without bearing significant mineralogical information. Finally, the quality of resulting clusters was assessed using the Silhouette criterion [6]. The silhouette value can vary between -1 and +1, with numbers close to +1 indicating optimal clustering.

Processing combination	Silhouette score
PCA + k-Means	0.211
PCA + GMM	0.184
PCA* + k-Means	0.209
PCA*+ GMM	0.195
PCA* + UMAP + k-Means	0.365
PCA* + UMAP + GMM	0.357

***Table 1**. Data processing combinations tested with the developed tool. The asterisk (*) indicates that the first principal component is not taken into account. Average silhouette score for n° clusters = 11 is shown in the second column.*

Results: The methods described above have been applied to several CRISM scenes of known composition in the area of Meridiani Planum, which is a well-known equatorial region of Mars with a high degree of spectral variability. Both mafic [7] and aqueous mineral phases [7, 8] have been observed in the area. Here, as an example, we show and discuss the results from the FRT00009B5A CRISM image covering the northern portion of Kai crater (Lat 4°20'N; lon 2°50'E). Within this scene, we have some areas which show a mafic composition (pyroxenes) and complex layered sediments which show presence of clay minerals and sulfates. For the different combinations listed in **Table 1**, the Silhouette coefficient indicates that the best clustering is achieved when about 11 clusters are given as input to the algorithms *and* the data is dimensionally reduced with the PCA*+UMAP combination. In this case, the Silhouette is around 0.36 for both k-Means and GMM, while it is around 0.2 for all the other cases. **Figure 1E** shows the results for k-Means clustering with PCA*+UMAP dimensionality reduction technique and compares them to RGB maps of spectral parameters related to either mafic or hydrous phases in the scene (**Figures 1B, C, D**).

Discussion and conclusions: As it can be seen from **Figure 1**, all the main mineralogic phases present in the CRISM scene are segmented in different clusters by both algorithms. Although only the PCA*+UMAP+k-Means is shown here, we have very similar results with Gaussian Mixture Models (GMMs). Only the monohydrated sulfate occurrences, which have a very limited spatial extension throughout the image (yellow areas in **Figure 1B**) are not assigned correctly, in any of the tested cases listed in **Table 1**. However, the algorithm's ability to capture the subtle mineralogical variations within the layered materials at the center of the scene (clusters 9 and 8, shown in lilac and aquamarine colors, in Figure **1E**) is a really interesting result. These materials are known to be composed of sulfates mixed with different percentages of clay minerals [9].

Overall, k-Means and Gaussian Mixture Models clustering algorithms provide an interesting and valid alternative/complement for the analysis of CRISM images. We plan to apply and test the same methods shown here to other areas of Mars as well, in order to validate them on a wider range of spectral features.

Code availability: the code is fully available in the following GitHub repository: https://github.com/beatricebs/CRISM-python-unsupervised-clustering

Acknowledgements: CRISM Data were downloaded through the PDS Geosciences Node Orbital Data Explorer (ODE). This project was possible thanks to the support of Fondazione Aldo Gini (University of Padova) and it is also partially funded by Europlanet RI20-24 GMAP project (research grant agreement No. 871149). Part of the computational resources for this work are kindly provided by INAF computational infrastructure for big data (DATA-STAR).

References: [1] S. Murchie et al. (2007), *JGR, 112 (E5)*, E05S03. [2] C. E. Viviano-Beck et al. (2014), *JGR Planets, 119*, 1403-1431. [3] F. P. Seelos et al. (2016), *LPSC XXXXVII*, Abstract #1783. [4] Pedregosa, F., et al. (2011), *JMLR, 12*, 2825–2830, https://scikit-learn.org/stable/ [5] L. McInnes et al. (2020), https://doi.org/10.48550/arXiv.1802.03426 [6] P. J. Rousseeuw (1987), *J. Comput. Appl. Math.*, *20*, 53-65. [7] F. Poulet et al. (2008) *Icarus*, *195*, 106-130 [8] Flahaut J. et al. (2015) *Icarus, 248*, 269-288. [9] Baschetti et al., "Quasiperiodic Fe/Mg clay enrichment within sulfate beds of Equatorial Layered Deposits in Meridiani Planum" (this conference), *CNSP 2024*.

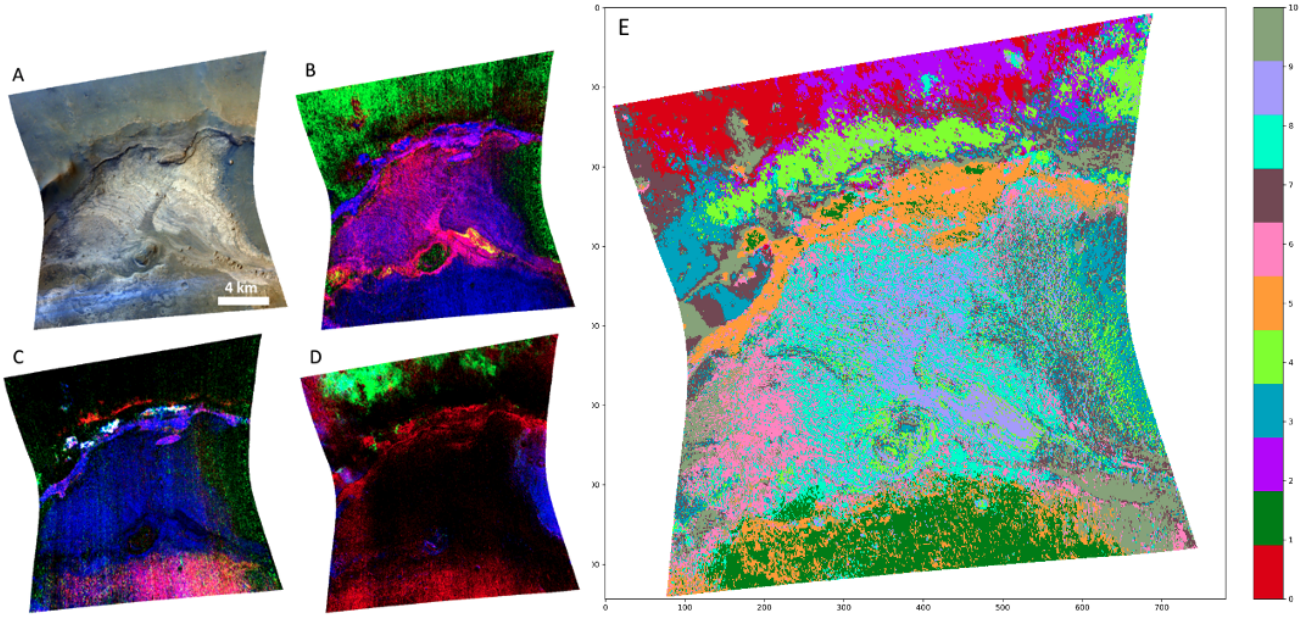

Figure 1: *Comparison between RGB maps of CRISM scene 9B5A and PCA*+UMAP+k-Means clustering.* ***A****: image of the surface at visible wavelength, provided for context;* ***B****: RGB map of hydrous minerals showing monohydrated (yellow) and polyhydrated (magenta) sulfates, mafic minerals (green), and clays (blue);* ***C****: RGB map of hydrous minerals showing different types of clays (white and magenta);* ***D****: RGB map of mafic minerals showing different kinds of pyroxenes (green, purple/blue areas);* ***E****: PCA*+UMAP+k-Means clustering results for 11 clusters.*