UNSUPERVISED CLASSIFICATION OF MARS 2020 SUPERCAM VISIR SPECTRA. A. M. Zastrow[1], A. M. Ollila[1], S. M. Clegg[1], E. Dehouck[2], E. Gibbons[3], J. R. Johnson[4], R. C. Wiens[5], C. Quantin-Nataf[2], A. Brown[6], J. Lasue[7], O. Forni[7], P. Pilleri[7], C. Legett IV[1], T. Fouchet[8], C. Royer[5], A. Cousin[7], and S. Maurice[7]. [1]Los Alamos National Laboratory (a.m.zastrow@lanl.gov), [2]Université Claude Bernard Lyon 1, [3]McGill University, [4]Johns Hopkins University Applied Physics Laboratory, [5]Purdue University, [6]Plancius Research, [7]Institut de Recherche en Astrophysique et Planétologie, [8]Observatoire de Paris.

In this abstract, we take a "data-first" approach to analysis of the SuperCam VISIR spectral dataset from the Mars2020 rover to classify targets based on spectral similarities and determine the main drivers of spectral variance.

SuperCam VISIR: SuperCam collects "passive" VISIR data from ~0.4 µm to 2.6 µm across two wavelength ranges [1]. The VIS portion of the spectrum covers 5925 spectral channels from ~0.38 to 0.84 µm and the IR portion covers 256 spectral channels from ~1.3 to 2.6 µm. In this abstract, we have analyzed 2157 VISIR spectra from sols 11 to 642.

Data preparation. Prior to performing the analysis, several data downselection and pre-processing steps were performed. Long-distance (>10m) and shadowed targets were removed. Beginning with calibrated data [2, 3], we smoothed the VIS spectra with a 51-channel average to reduce spikes. Next, we downsampled the data so that the spectral resolution is 0.01 µm over the entire spectral range and to ensure that our results are not overwhelmed by the VIS spectra, which have ~23x more channels than the IR. Last, we normalized the data using the Euclidean distance ("L2-normalization") to reveal spectral differences that are otherwise hidden by albedo. The VIS and IR portions of the spectra were normalized separately and then concatenated by point.

Principal Components Analysis: PCA is a commonly used method of dimensionality reduction. It mathematically determines the spectral shapes that account for the most variance for a given dataset. In this way, we projected the dataset into "principal components-space" and began to group spectra based on their spectral similarities and determine the dominant spectral and physical features of the dataset.

Results and Discussion: For the SuperCam VISIR dataset analyzed here, the first principal component (PC-1) accounts for 61% of the spectral variance in the dataset, PC-2 accounts for 21%, and PC-3 accounts for 11%. Figure 1 shows these three principal components. 99% of the variance in the dataset is represented by the first 10 PCs. It should be noted that PCs do not actually correspond to "physical" spectra, as they are dimensionless in PC-space and centered around zero. We can still look at the PCs and correlate their shapes with distinct spectral features and use other approaches to derive physical meaning from the PCs.

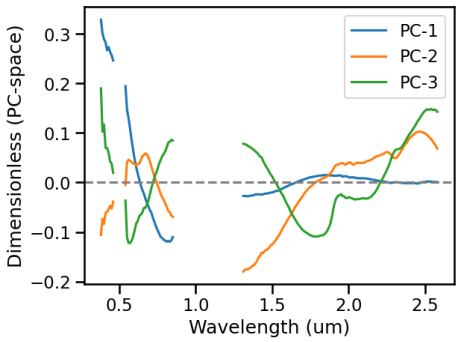

Figure 1. Principal components (or loadings) #1-3. PC shapes correlate with IR spectral features, such as the slope in PC-2 from 1.3-1.9 µm correlating with olivine and the strong feature in PC-3 at ~1.9 µm correlating to hydration.

Figure 2 shows the result of plotting the PC values for each spectrum against each other and color them by "formation". As of now, we have included the Maaz and Seitah formations from the crater floor [3] and labeled the remaining targets as the "Delta" [4].

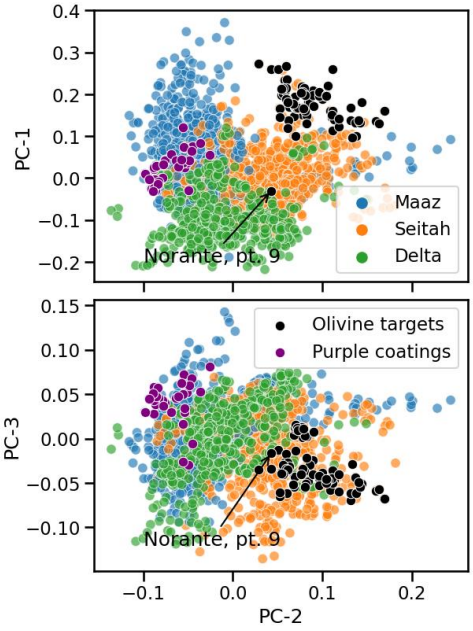

Figure 2. Top: PC-2 versus PC-1, Bottom: PC-2 versus PC-3. Points are colored by formation (blue: Maaz, orange: Seitah, green: Delta). Black points are olivine targets and purple points are targets with purple coatings; target names are noted in the text.

In PC-space, each formation tends to cluster together with some overlap. In general, the Maaz targets have higher PC-1 values, the Delta targets have lower PC-1 values, and the Seitah targets have middle-to-high PC-1 values. In terms of PC-2, the Seitah targets have higher values than the Maaz and Delta targets. There is a small set of Maaz targets with higher PC-2 values that are regolith targets, which may point to the mixing of regolith between the Maaz and Seitah [5].

Next, we compared PC values to a variety of spectral parameters (derived from the VISIR spectra).

The top panel of Figure 3 plots the spectral slope from 600 to 840 nm (S6084) against the ratio of reflectance to 600 to 840 nm (R6084), which highlights a trend of relative dustiness, assuming spectrally dark, flat substrates [6]. This trend correlates well to the PC-1 values, indicating that points with lower PC-1 values are dustier than those with higher PC-1 values. This is seen in the PC-1 shape, which is relatively flat in the IR: indicative of a dustier target.

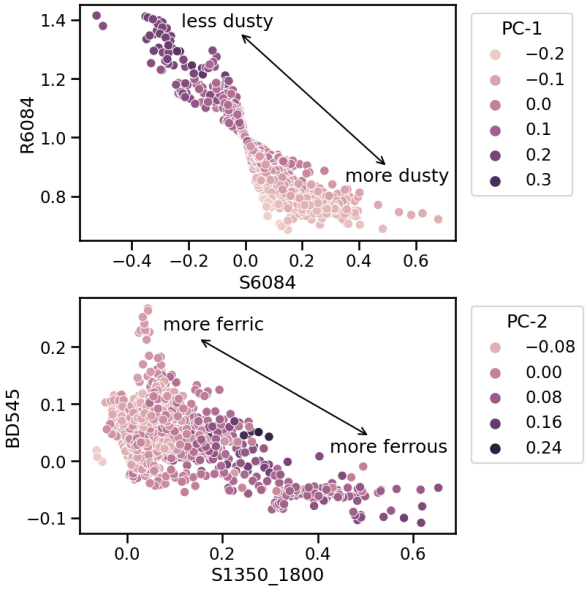

Figure 3. Top: S6048 vs R6048, colored by PC-1. Bottom: S1350_1800 vs BD545, colored by PC-2. Arrows indicate trends based on the compared spectral parameters.

The bottom panel of Figure 3 plots the spectral slope from 1350 to 1800 nm (S1350_1800) against the band depth at 545 nm (BD545). Higher BD545 values are indicative of more ferric materials [6] and higher S1350_1800 values are indicative of more ferrous materials. Here, lower PC-2 values correspond to higher BD545 values and higher PC-2 values correspond to higher S1350_1800 values. This slope is seen in the PC-2 shape in Figure 1. To confirm that PC-2 is correlated with olivine, we highlighted a set of olivine-rich Seitah targets (*Cine, Dourbes, Garde, and Norante* [7]) on Figure 2, and found that they do have higher PC-2 values. Additionally, plotting these targets helped confirm that PC-1 is dependent on the dustiness of a given target: point #9 of *Norante* (Figure 2) is significantly dustier than the rest of the *Norante* points and thus has a lower PC-1 value, but a still high PC-2 value.

Lastly, we attempted to determine what the driving physical factor behind PC-3 is. Throughout the mission, we have seen that many rocks have "purple coatings" [8, 9], whose origin and significance is still under investigation. Plotting some of these targets (*Alk es disi, Naadiin, and Grandes Tours du Lac* [10]) on Figure 2 showed that they cluster together with low PC-2 values and higher PC-3 values.

Figure 4 shows the mean spectra of the olivine and purple coating points from Figure 2. These two spectra share spectral features with the PCs in Figure 1 with which they correlate. For example, the strong slope from 1.3-1.9 µm in the olivine spectra corresponds to the same slope in PC-2.

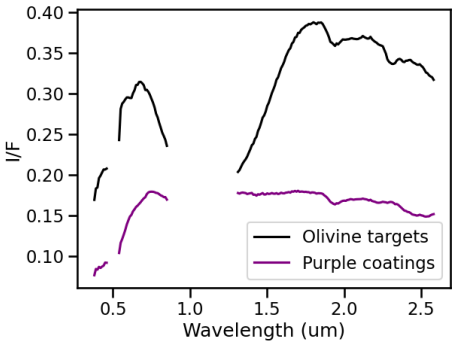

Figure 4. Mean spectra for the points in Figure 2.

In conclusion, we have analyzed the full SuperCam VISIR dataset with PCA and determined the physical properties of targets that dominate the spectral variance in the VISIR for the first 2-3 PCs. The next step in this research will be to further investigate higher PCs, look for correlations with LIBS compositions, and do more detailed unit breakdowns.

Acknowledgments: The authors would like to thank the Mars2020 project and the SuperCam instrument team for their support. Data analysis was completed using the scikit-learn Python package.

References: [1] Fouchet et al. (2022) Icarus, 373, 114733. [2] Royer, C. et al. (2022) JGR-Planets, 127. [3] Mandon, L. et al. (2022) JGR-Planets, 127. [4] Dehouck E. et al., this meeting. [5] Beyssac O. et al., this meeting. [6] Johnson, J.R. (2015) Icarus, 249, 74–92; Johnson, J. (2016) Am. Min., 101, 1501–1514. [7] Johnson, R. et al. (2022) LPSC LIII, #1254. [8] Ollila, A. (2021) AGU Fall Meeting, #P22B-05. [9] Garczynski, B. (2022) LPSC LIII, #2346. [10] Meslin, P.-Y. et al. (2022) LPSC LIII, #2694.