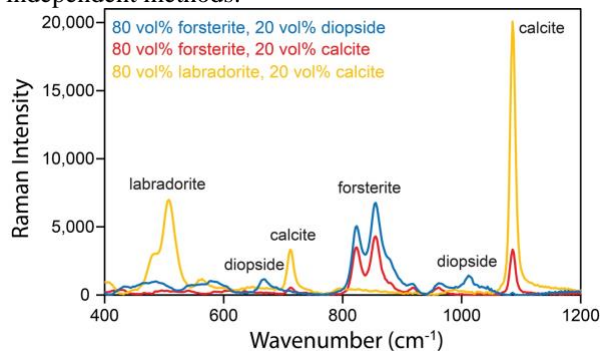


**QUANTIFICATION OF MINERAL ABUNDANCES IN MIXTURE USING RAMAN SPECTROSCOPY: APPLICABILITY TO MARS ROVER MEASUREMENTS.** L. B. Breitenfeld<sup>1</sup>, M. D. Dyar<sup>2</sup>, T. D. Glotch<sup>1</sup>, A. D. Rogers<sup>1</sup>, and M. Eleazer<sup>3</sup>, <sup>1</sup>Department of Geosciences, Stony Brook University, Stony Brook, NY 11794, laura.breitenfeld@stonybrook.edu, <sup>2</sup>Department of Astronomy, Mount Holyoke College, South Hadley, MA 01075, <sup>3</sup>Department of Astronomy, University of Massachusetts Amherst, Amherst, MA 01003.

**Introduction:** Two Raman instruments were recently deployed to the surface of Mars on the Mars 2020's Perseverance Rover. The Scanning Habitable Environments with Raman and Luminescence for Organics & Chemicals (SHERLOC) utilizes deep ultraviolet resonance for the analysis of organics, chemicals, and surface mineralogy [1]. SuperCam analyzes surface materials with a 532 nm laser from long ranges [2]. A third Raman spectrometer, the Raman Laser Spectrometer (RLS) on the ExoMars mission, will arrive years from now. It will use a 532 nm laser to probe powdered samples obtained by a drill [3].

For Raman spectrometers with large spot sizes, multiple mineral phases can be interrogated in the collection of a single Raman spectrum. However, currently there is not a reliable methodology for quantifying mineral species in mixture due to non-linear mixing effects in Raman Spectroscopy (**Figure 1**). Here, we report machine learning multivariate unmixing models and introduce Raman scattering coefficients, which are numerical metrics associated with specific Raman features for the quantification of common rock-forming minerals. Utilizing these models and Raman scattering coefficients simultaneously allows users to quantify mineral contents in mixtures using two independent methods.



**Figure 1.** Raman spectra of three binary mixtures consisting of forsterite, calcite, labradorite, and diopside that exhibit non-linear mixing behavior.

**Background:** Quantitative relationships between peak areas and mineral abundances are obscured by many complicating factors in Raman spectroscopy: the exciting laser wavelength, the Raman cross-section of the minerals (dependent on the strength of covalent bonding and polarizability of the molecule), crystal orientation relative to laser polarization, and long-range chemical and structural ordering in the crystal lattices [4]. Even if the Raman laser interrogates a broad area to avoid crystal orientation effects, variations in Raman

cross-sections of different mineral species prevent quantitative assessment of mineral abundances in mixtures. Until a theoretical model for unmixing of Raman data is developed, empirical methods such as machine learning multivariate analysis and/or use of Raman scattering coefficients are needed.

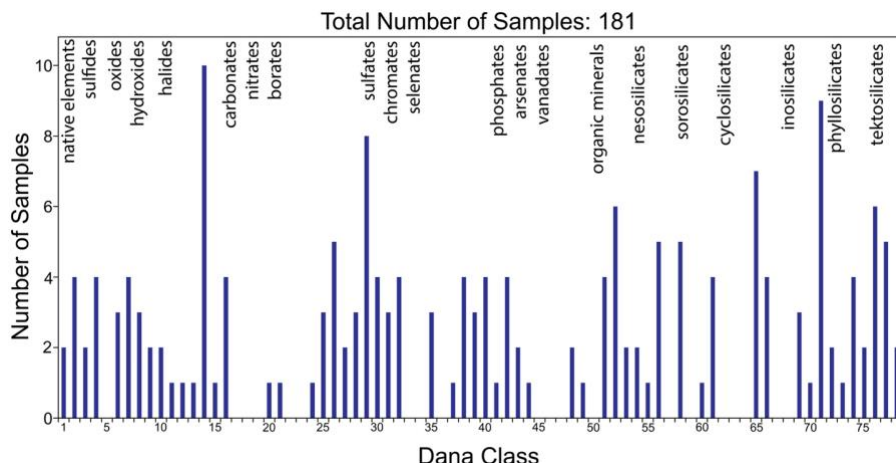
Due to the complexities regarding Raman mineral quantification in the context of planetary surfaces, a combination of utilizing the collection of narrow-beam spectra and a point-count method [4] can theoretically solve this problem. However, despite the use of spectral rasters and point-counting, the ubiquity of mixed-phase spectra presents difficulties for quantification. The 1.5–5.0 mm beam size of SuperCam results in mixed-mineral spectra [2]. This can also be true for the smaller ~100  $\mu\text{m}$  beam size of SHERLOC [1].

**Methods:** We quantified the Raman scattering coefficients of 181 samples (**Figure 2**) and used 37 of those pure mineral samples to develop machine learning multivariate analysis models. Electron microprobe chemical analyses were used to calculate the density of each mineral. This allowed mixtures to be made in volume percentages, analogous to mineral abundances reported from remote sensing measurements. Samples were crushed, handpicked for purity, and sieved.

Spectra of end-members and mixtures were acquired on a Bruker Optics BRAVO Raman spectrometer using dual laser excitation and their fluorescence mitigation strategy employing successive laser heating [5]. Three scans/spectra were obtained using an integration time of 10 seconds covering a wavenumber range of 300–3200  $\text{cm}^{-1}$ .

**Machine Learning Multivariate Analysis:** A total of 640 binary mineral mixtures were made from pairings of 37 mineral samples. Samples were weighed into ratios of 50:50, 20:80, and 95:5 volume % depending on the Raman cross-sections of the paired phases. Mixtures were customized for each pair to ensure that each phase would have characteristic bands with sufficient peak areas for multivariate analyses.

Partial least squares (PLS) [6], a type of machine learning multivariate analysis, was utilized in building models for the prediction of mineral abundances in each mixture. Multivariate model performance is mineral-dependent. We found that depending on the mineral species of interest, the cross-validated root mean square error can vary (e.g., 3.7–15.5 volume%). These errors are small enough to be promising for evaluating mineral mixtures in field and extraterrestrial applications.



**Figure 2.** Histogram of all 181 pure mineral samples organized by Dana Class. The 37 mineral samples utilized in the machine learning multivariate analysis models include hematite, ilmenite, spinel, magnetite, calcite, magnesite, siderite, aragonite, anhydrite, gypsum, rozenite, melanterite, alunite, jarosite, garnet, two forsterites, cordierite, clinopyroxene, orthopyroxene, enstatite, diopside, augite, amphibole, tremolite, ripidolite, serpentine, phlogopite, montmorillonite, nontronite, saponite, quartz, potassium feldspar, albite, labradorite, bytownite, and chabazite.

**Raman Scattering Coefficients:** Raman scattering coefficients were calculated by making 95:5 volume % mixtures of the 181 individual minerals with diamond powder as a reference material. Synthetic diamond was chosen as a reference for the Raman scattering coefficients because it has a simple spectrum with a strong peak at 1332 cm<sup>-1</sup> that does not overlap with common rock-forming minerals. Peak areas for diagnostic features in each mineral species’ spectrum were compared to those of the prominent diamond peak. The ratio of those areas is termed a “Raman scattering coefficient” (**Equation 1**). A normalization parameter is included within the equation to account for the volume percentage of diamond given that different proportions could be chosen when making a diamond-mineral mixture.

$$Coefficient = \left( \frac{mineral\ area}{diamond\ area} \right) \times \left( \frac{\%diamond}{100} \right) \quad \text{Equation 1}$$

We plan to test the performance of the Raman scattering coefficients in unmixing unseen data using the binary mineral mixture training set from the multivariate analysis models.

**Applications:** This project is of importance to terrestrial field work and laboratory analyses of rocks and mixtures. Additionally, workers can utilize these methods for in situ analyses of Mars. This work holistically addresses the problem of mineral abundance quantification for mixtures using Raman spectroscopy. In the future, we anticipate creating specialized multivariate analysis models customized to the mineralogy of specific geologic units. For example,

for the Perseverance Rover landing site, we will develop models for specific geologic units like Mááz and Séítah in Jezero Crater (e.g., olivine, pyroxene, carbonate, and hematite) [7]. We expect that this change will improve model accuracy further. Through this work, we lay the foundation for quantifying the relevant mineralogy using two independent methods.

**Conclusions:** Both Raman scattering coefficients and machine learning multivariate analysis can be used to account for non-linear mixing effects and quantify mineral abundances using Raman spectroscopy. The pros and cons of the two Raman

quantification techniques are summarized in **Figure 3**. Different scenarios may favor the use of one of these methods over the other. We expect the most confident and accurate prediction results when utilizing the two quantitative solutions simultaneously.

The chief liability of these techniques lies in the lack of available well-characterized individual pure minerals suitable for making mixtures, and in the labor-intensive nature of creating those mixtures. Availability of mixtures is of paramount importance in creating the fundamental data needed to improve the accuracy of these techniques.

**Acknowledgments:** We thank the Massachusetts Space Grant Consortium for initial funding of this project, and NASA SSERVI funding from the RIS<sup>4</sup>E and RISE2 nodes for subsequent support.

**References:** [1] Bhartia, R., Beegle, L. W., DeFlores, L., et al. (2021) *SSRv*, 217(4), 1-115. [2] Wiens, R. C., Maurice, S., Robinson, S. H., et al. (2021) *SSRv*, 217(1), 1-87. [3] Rull, F., Maurice, S., Hutchinson, I., et al. (2017) *Astrobiology*, 17(6-7), 627-654. [4] Haskin L. A. et al. (1997) *JGR: Planets*, 102, 19293-19306. [5] Cooper J. B. et al. (2014) *Spectroscopy*, 29, 38-42. [6] Wold, S., Sjöström, M., and Eriksson, L. (2001) *Chemom. Intell. Lab. Syst.*, 58(2), 109-130. [7] Bell III, J. F., Maki, J. N., Alwmark, S., et al. (2022) *Sci. Adv.*, 8(47), eabo4856.

Raman Scattering Coefficients	Multivariate Analysis Machine Learning
(+) efficient model construction	(-) time consuming model construction
(+) small sample requirement	(-) large sample requirement
(-) mineral identification required	(+) mineral identification not required
(+) flexibility encountering unexpected materials	(-) difficulty encountering unexpected materials
(+) fast application	(+) fast application

**Figure 3.** Pros (+, green) and cons (-, red) of the Raman scattering coefficients and the multivariate analysis technique.