SUPERCAM LASER INDUCED BREAKDOWN SPECTROSCOPY (LIBS) DATA PROCESSING, CALIBRATION, AND FIRST RESULTS.

R.B. Anderson[1], O. Forni[2], S.M. Clegg[3], A. Cousin[2], J. Frydenvang[4], P. Pilleri[2], C. Legett[3], R.C. Wiens[3], S. Maurice[2], G. Arana[5], O. Beyssac[6], B. Bousquet[7], B. Chide[2], E. Clavé[7], D. Delapp[3], A. Essunfeld[3], T. Fouchet[8], C. Garcia-Florentino[5], O. Gasnault[2], E. Gibbons[9], J. Laserna[10], J. Lasue[2], J.A. Manrique[11], J.M. Madariaga[5], R. Newell[2], A. Ollila[3], S. Sharma[12], J. Simon[13], P. Sobron[14], [1]USGS Astrogeology (rbanderson@usgs.gov), [2]IRAP, [3]LANL, [4]U. of Copenhagen, [5]Univ. Basque Country, [6]U. Pierre and Marie Curie, [7]U. Bordeaux, [8]Observatory de Paris, [9]McGill U. [10]U. Malaga, [11]Unviersidad de Valladolid, [12]U. Hawaii, [13]NASA JSC, [14]SETI Institute

Introduction: The SuperCam instrument suite [1,2,3] on the Perseverance rover is a versatile tool for stand-off and remote sensing observations. In addition to Laser Induced Breakdown Spectroscopy (LIBS), which is the focus of this abstract, SuperCam capabilities include: 1) color Remote Microscopic Imager (RMI), 2) Raman and time-resolved luminescence spectroscopy (TRLS) up to 12 m from the rover, and 3) passive spectroscopy in the 1.3-2.6 micron range.

This abstract focuses on SuperCam LIBS, providing information about the instrument, the data processing steps, and initial quantitative calibration. Similar to ChemCam [4,5], SuperCam LIBS uses a pulsed 1064 nm laser to ablate small amounts of target material at distances <7 m from the rover. The atomic emission spectrum of the plasma spark is collected and used to determine the target elemental composition.

SuperCam LIBS data: SuperCam collects LIBS spectra across 5 spectral ranges: Ultraviolet (UV; 245-340 nm), Violet (VIO; 385-465 nm), Green (534-625 nm), Orange (613-722 nm), and Red (718-853 nm). The Green, Orange, and Red spectra are concatenated into a continuous VNIR spectrum (536-853 nm). SuperCam's spectral-resolution requirement for the Raman spectral range resulted in significantly improved LIBS resolution in the green spectral range [2], yielding better detection for sulfur emission peaks. SuperCam can also take advantage of the intensifier in the VNIR region [2] to study weaker emission lines or time-dependent phenomena.

Raw data are subjected to a series of preprocessing steps: 1) remove anomalous spikes due to cosmic rays (not applied to laboratory data); 2) subtract "dark" spectra of the target (acquired when there is no plasma spark) from the LIBS spectra; 3) denoise the spectra using a wavelet transform algorithm, 4) convert spectra from counts to photons using the instrument response; 5) stitch the Green, Orange, and Red ranges together; 6) wavelength calibrate the spectra; 7) subtract the smoothly-varying continuum to isolate signal from atomic emission lines; 8) convert to units of radiance.

SuperCam LIBS observations on Mars will be similar to ChemCam observations, consisting of a series of "points" on the target organized into a line or grid. At each point, we acquire multiple spectra (typically 30 laser pulses, one spectrum per pulse). The first ~5 single shot spectra on Mars targets tend to be contaminated by surface dust, so these are excluded when calculating the "statistic spectra" (average, median, and standard deviation) for each point. All spectra from each point are stored in a single FITS file.

LIBS Calibration: LIBS provides information about the elemental composition of the target. SuperCam will quantify the following elements as oxides: SiO_2, TiO_2, Al_2O_3, FeO_T, MnO, MgO, CaO, Na_2O, and K_2O. Trace/minor elements (e.g. Ba, Cr, Li, Rb, Sr, etc.) will also be quantified, but are not discussed here.

Quantifying the composition is non-trivial: the intensities of the atomic emission lines are influenced by physical and chemical properties of the sample, e.g., an emission line from one element may be influenced by the presence of other elements. Because of this, a "univariate" approach based on plotting the intensity of a diagnostic line against composition is often not adequate. Instead, multivariate regression approaches that develop statistical models using all of the available spectral channels (not just those from one element) to predict an element's concentration tend to be more successful [6].

Database: Multivariate regression models must be trained on spectra from a suite of samples for which the elemental composition is known, and this data set must be large and diverse enough to span the range of compositions expected on Mars. For the initial SuperCam calibration effort, we use a suite of 1198 spectra of 334 different samples acquired using the Flight body unit [2] and the Engineering Qualification Model (EQM) of the mast unit [1]. The compositions in the database range from "typical" (basalts, olivine, pyroxene), to "unusual" (e.g. apatites, manganese oxides). The bulk of these data were acquired at a distance of 3 m (the typical observation distance on Mars) but a subset were acquired at 1.5 m (the distance to the calibration targets), and 4.25 m.

An important part of the calibration process has been to carefully assess this data set to remove outliers. These include bad spectra (e.g. missed targets or low signal to noise) as well as valid spectra of targets with compositions that are too different from the majority of the targets to be accurately modeled. The latter case typically occurs for pure mineral end-members that have very high concentrations of one or more element (e.g. Mn ores, apatites, etc.). These samples can have an outsized influence on a model, overriding the model's ability to

accurately predict more "typical" compositions. However, these unusual samples can be useful when using methods that are more adaptable (e.g. local regression, ensemble methods, submodels) and in such cases they are kept in the data set.

Additional preprocessing: In addition to the preprocessing applied to all raw data, additional preprocessing steps can be applied to the spectra prior to quantification. The spectra are typically masked near the edges of the spectral ranges (<245.5 nm, 712.17-713.7 nm, >848.3 nm) to account for the strong variations in the instrument response function there.

Peak binning can be useful for developing models of elements with relatively weak emission lines [7] and also significantly improves the robustness of models to small shifts in wavelength calibration. It also reduces the number of input variables, speeding up model cross validation and training. To apply peak binning, we first calculate the average spectrum of the training set and identify local minima and maxima of that spectrum. Then, for each individual spectrum, we sum the signal between each pair of local minima and assign it to the bin corresponding to the local maximum.

Normalization helps reduce intensity fluctuations between spectra of the same target and mitigates some of the effects of distance. Typically spectra are either divided by the sum of all signal, or by the sum of the signal for each of the 5 spectrometers, such that the intensity values across the full spectrum sum to 1 or 5, respectively. We have also found that in some cases, standardizing the spectra by subtracting the mean and dividing by the standard deviation on a per-spectral element basis can result in improved regression results.

Model Selection: We subdivide the full suite of data into five subsets ("folds") for each element, with a similar range of compositions within each fold. All spectra from a given sample are included in the same fold. One fold is defined as a "test set" and is excluded from analysis until the regression model parameters have been optimized. The rover calibration targets are explicitly assigned to the test set. Model parameters are optimized using cross validation: each of the 4 training set folds is withheld in turn and predicted using a model trained on the remaining folds. By repeating this process for a range of model parameters, we determine the optimal parameter settings for each method. Each optimized model is then used to predict the test set, and the test set results are compared to determine which regression method gives the best accuracy. The performance on the test set at 3 m will be the primary factor in choosing a model, but the model robustness to wavelength shifts (+/- 1 pixel) and distance will also be considered.

A wide variety of multivariate regression methods are currently being considered, including: Partial Least Squares (PLS), LASSO, Elastic Net, Random Forest, Gradient Boosting, and local regression. Many of these are implemented via the scikit-learn library [8] and/or the PyHAT library [9]. In addition, we are considering combinations of "sub-models" trained on restricted composition ranges, similar to the approach used for ChemCam [9, 10, 11].

LIBS on Mars: Once the best regression model is determined for each element based on test set RMSE and robustness to expected variations in the data, the models will be incorporated into an automated pipeline that will produce quantitative elemental abundances almost immediately upon receipt of data from Mars.

The regression models will be based on data collected with the EQM mast unit under laboratory "Mars chamber" conditions rather than with the flight unit under actual martian conditions. We will therefore collect calibration target data as soon as possible after landing on Mars, to determine whether an "Earth-to-Mars" correction similar to that used by ChemCam [10] needs to be applied to make the data sets compatible.

Although the calibration database samples were selected to be realistic but diverse, it is possible that minerals will be identified on Mars that are not well represented in our calibration database. This may require collection of more database spectra and development of updated regression models to better handle the data received from Mars. Specialized models that are more accurate for certain key composition types encountered on the mission (e.g. Fe-Ni meteorites, carbonates) at the expense of generalizability are also a possibility.

This abstract is being written prior to the Perseverance rover landing, but by the time of the conference, initial LIBS data from Jezero crater may be available. Experience from ChemCam shows that the ability to rapidly assess the elemental composition of targets at a small scale from a distance results in a rich geochemical data set and also provides vital information for directing more resource-intensive contact science and drilling activities. With the addition of Raman/TRLS and VNIR spectroscopy, SuperCam will play an important role in guiding the Perseverance rover's efforts to collect well-characterized samples for eventual return to Earth.

References: [1] Maurice S. et al. (2020) SSR, SPAC-D-20-00069R1. [2] Wiens R.C. et al. (2020) SSR 10.1007/s11214-020-00777-5. [3] Manrique J.-A. et al. (2020) SSR 216, 138. [4] Wiens R.C. et al. (2012) SSR 170, 167. [5] Maurice S. et al. (2012) SSR 170, 95. [6] Clegg S.M., et al.(2009) Spectrochim. Acta B, 64, 79. [7] Clegg S.M. et al. (2020), 51st LPSC, #2561. [8] https://scikit-learn.org/ [9] https://github.com/USGS-Astrogeology/PyHAT/tree/master [10] Anderson R.B. et al. (2017) Spectrochim. Acta B 129, 49 [11] Clegg S.M. et al. (2017) Spectrochim. Acta B, 129, 64.