PRELIMINARY MULTIVARIATE CALIBRATION OF MAJOR ELEMENTS USING A CHEMCAM LABORATORY INSTRUMENT UNDER LUNAR EXPERIMENTAL CONDITIONS.
N. Klein¹, A. Ollila¹, P. J. Gasda¹, N. Lanza¹, E. C. Sklute¹, B. Gifford¹, K. Ganguly¹, C. Legett¹, K. L. Williams¹, R. K. Martinez¹, A. E. Nelson¹, Z. Wolf³, R. B. Anderson², J. Lasue³, O. Forni³, W. Rapin³, S. Maurice³, A. Cousin³, R. C. Wiens⁴, C. Bedford⁴, S. Connell⁴, and S. M. Clegg⁴.
¹Los Alamos National Laboratory, ²USGS Flagstaff, ³IRAP Toulouse, France, ⁴Purdue University.

Introduction: The ChemCam instrument onboard the NASA Curiosity rover enables the measurement of geological composition via atomic emission spectroscopy using laser-induced breakdown spectroscopy (LIBS) [1]. Statistical models known as calibration models are produced from laboratory spectra to predict composition as oxide weight percent [2]. In this work, we evaluate two classes of calibration models on lab ChemCam data collected under lunar conditions: partial least squares (PLS) and Dirichlet regression (DR). In addition, we construct uncertainty intervals on the test set predictions via conformal prediction and Bayesian inference. We evaluate the calibration models with respect to both prediction accuracy and properties of the uncertainty intervals.

Sample Suite and Data Collection: For this preliminary study, we selected a limited sample suite composed of lunar-relevant silicates. Twenty-two targets (10 basalts; 1 anodesite; 1 gabro; 1 anorthosite; 3 forsterite olivines; 1 fayalite-rich olivine; 3 pyroxenes: augite, diopside, and hedenbergite; and 2 alkali feldspars) where analyzed in a vacuum chamber using the ChemCam lab instrument. We used standard spectral data collection parameters and laser energy (~14 mJ) for the lab unit. The lunar environment chamber registered a pressure of ~1.4x10⁻⁵ hPa, well below what was previously determined to be suitable for lunar-environment LIBS plasmas [3]. We collected data at three locations with 30 laser pulses per location at distances of 1.6m and 6.34m. We use spectra comprising of the average of 25 single shot (first 5 shots removed to avoid surface contamination). Fig. 1 shows the average spectra at both distances for BHVO2 (basalt). Chemical compositions are given by oxide weight percent values (wt. %) for SiO₂, TiO₂, Al₂O₃, FeO_T (FeO total), MgO, CaO, Na₂O, and K₂O. The composition wt. % ranges are (35.85, 74.38), (0.002, 3.39), (0.36, 27.42), (0.79, 21.51), (0.07, 56.14), (0.08, 20.87), (0.02, 4.65), and (0.01, 4.06), respectively.

Calibration Models: We fit two types of statistical calibration models: PLS with conformal prediction intervals and Bayesian DR. For each model, we construct predictions and intervals via leave-one-out analysis; the prediction for a given target is calculated by holding out all spectra from that target during fitting.

PLS with conformal prediction intervals. We fit separate PLS1 models for each oxide [4] using nested leave-one-out cross-validation to select the number of components (min: 2, max: 10), applying the one-standard-error procedure to select the simplest model (i.e., fewest components) that achieved within one standard error of the lowest cross-validation error [5].

For prediction intervals, we use conformal prediction, a model-agnostic framework for constructing prediction intervals with good statistical properties. In essence, the intervals are constructed to contain values that would be “typical” according to the distribution of absolute residuals observed for the other targets. Specifically, we apply the jackknife+ procedure [2] to construct 95% prediction intervals for each test set prediction.

Bayesian DR. For certain geologic targets, it is reasonable to assume that the measured oxides should sum to nearly 100%, with any residual attributed to unmeasured trace elements. Because this sample suite focused on silicate minerals, the assumption that the sum total should approximately be 100% is reasonable. Therefore, this data can be treated as compositional data which has special structure due to the constraint that all components (including a residual category) sum to 1 [6]. Specifically, we denote the measured compositions as (y_1,...,y_{K-1}) and define y_K = 1 - \sum_{k=1}^{K-1} y_k. Then a DR model for y = (y_1,...,y_K) incorporates the sum-to-one constraint and allows for missing y values to be incorporated into the residual category. We use the common parameterization [7] which leads to the a probability density function for the data y conditional.
on the covariates \( \mathbf{x} \): 
\[
p(y \mid x) \propto \prod_{k=1}^{K} Y_k^{(a_k(x)-1)}.
\]

The functions \( a_k(x) \) for component \( k \) are specified as linear functions of the covariates \( \mathbf{x} \) with log link functions to achieve nonnegative values. As covariates, we use projections of the LIBS spectra based on PLS2 models. We use Bayesian inference to infer the posterior distribution over the parameters given the data, with zero-mean Gaussian priors on the regression coefficients. The resulting algorithm makes use of Markov Chain Monte Carlo to obtain samples from the posterior predictive distribution, from which we take the mean and quantiles to achieve predictions and 95% prediction intervals, respectively.

**Evaluation Metrics:** We measure predictive accuracy with the root mean squared error (RMSE), evaluated using the squared difference between the true and predicted oxide weight percent on the held-out test target averaged across targets. We use two metrics for prediction intervals: coverage and average width. Coverage refers to the proportion of times that a prediction interval contains the true value; for a 95% prediction interval, we would hope to achieve 0.95 coverage. We evaluate coverage as the average number of intervals containing the true values (on the held-out test target predictions). Intervals should achieve the nominal 0.95 coverage, but we also want intervals to be short if possible, as high coverage can always be achieved with extremely wide intervals that would not be useful in practice. Therefore, we also evaluate the average width of the prediction intervals (in wt. %).

<table>
<thead>
<tr>
<th></th>
<th>SiO(_2)</th>
<th>TiO(_2)</th>
<th>Al(_2)O(_3)</th>
<th>FeO(_2)</th>
<th>MgO</th>
<th>CaO</th>
<th>Na(_2)O</th>
<th>K(_2)O</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PLS 1.6m</strong></td>
<td>3.81</td>
<td>0.99</td>
<td>4.90</td>
<td>3.00</td>
<td>3.89</td>
<td>1.54</td>
<td>1.46</td>
<td>0.50</td>
</tr>
<tr>
<td><strong>DR 1.6m</strong></td>
<td>3.53</td>
<td>0.52</td>
<td>2.70</td>
<td>2.55</td>
<td>3.80</td>
<td>1.56</td>
<td>0.89</td>
<td>2.23</td>
</tr>
<tr>
<td><strong>PLS 6.34m</strong></td>
<td>3.88</td>
<td>0.68</td>
<td>5.74</td>
<td>3.77</td>
<td>4.97</td>
<td>1.64</td>
<td>1.09</td>
<td>0.81</td>
</tr>
<tr>
<td><strong>DR 6.34m</strong></td>
<td>8.94</td>
<td>0.79</td>
<td>5.74</td>
<td>4.51</td>
<td>5.57</td>
<td>2.22</td>
<td>1.57</td>
<td>0.70</td>
</tr>
</tbody>
</table>

**Table 1:** Leave-one-out RMSE (wt. %) for each model and distance for all eight major oxides (best are bolded).

**Results:** Table 1 gives the RMSE values for both models and both distances. We find that for 1.6m data, the DR model outperforms the PLS model for nearly all oxides; we note that K\(_2\)O for one target (with the highest observed K\(_2\)O in the data set) is particularly poorly predicted by the DR model, skewing the RMSE. Fig. 2 shows the reference versus predicted values for CaO for both models at 1.6m. In contrast, for 6.34m data, the PLS model generally outperforms the DR model.

The prediction intervals from the DR model tend to be smaller than those for PLS (DR average width 6.55 wt. %, compared to 12.83 wt. % for PLS). While the DR intervals achieve coverage of 0.81, the PLS intervals achieve coverage of 0.92, which is closer to the nominal coverage of 0.95. We note that by construction, the DR model predictions and intervals must fall in [0, 1].

**Discussion and Conclusions:** In this work, we demonstrate two preliminary multivariate calibration models, PLS and DR, on ChemCam data collected under lunar conditions. Compared to previous lunar LIBS analysis with PLS models [3], our models exhibit higher RMSE, but our samples span a wider range of compositions which likely explains this discrepancy. We find some indication that DR models, which jointly model the composition of all major oxides with the constraint that the measured compositions plus a residual category should sum to 1, could outperform PLS models, as evidenced by generally lower RMSE for the 1.6m data. Intuitively, some components may be better predicted when knowledge of the other components is available. However, further work is needed in the design and tuning of these models, as the 6.34m presented challenges during the model selection and fitting process that could have prevented the 6.34m DR model from performing better. Furthermore, while conformal prediction intervals for PLS exhibit good coverage, they tend to be longer and more homogeneous in width than the intervals from Bayesian inference in the DR models, suggesting that the Bayesian intervals may be more indicative of per-sample confidence in the predictions and perhaps more useful in practice. In future work, we plan to expand upon these models and to compare to other model classes. Furthermore, we plan to determine whether models fitted jointly on all data collected at 1.6m and 6.34m could outperform the separate models demonstrated here.

**Acknowledgments:** Research presented here was supported by the Laboratory Directed Research and Development program of Los Alamos National Laboratory under project number 20240065DR.