INVESTIGATING CHEMCAM CALIBRATION METHODS WITH AUGMENTED DATA. R. A. Morris¹, P. J. Gasda², N. Klein³, A. Essunfeld², A. Huynh¹, J. Comellas¹, A. Ollila¹, E. Sklute¹, D. Delapp¹, S. Clegg¹, O. Gasnault⁴, N. Lanza¹. ¹LANL, ²Yale University, ³University of Hawai‘i at Mānoa, ⁴IRAP, Toulouse

Introduction: NASA’s Curiosity rover’s ChemCam instrument uses laser-induced breakdown spectroscopy (LIBS) to measure martian geochemistry [1]. Multivariate techniques are then applied to these spectra to predict elemental abundances [2, 3]. Current quantification methods (Major Oxide Calibration; MOC) use two multivariate methods: Partial Least Squares (PLS) and Independent Component Analysis (ICA) to predict the major rock-forming elements (presented as oxides): Al₂O₃, CaO, FeO₉, K₂O, MgO, MnO, Na₂O, SiO₂, and TiO₂. PLS is designed to reduce the dimensionality of the input data while retaining maximal covariance between the predictor and response variables and fits a specialized linear regression to this reduced space [4]. For this investigation, it is not necessary to understand ICA.

Each of the current models were trained on unique databases that include single-oxide abundances with corresponding LIBS data. While the performance of these models can be quantified in a number of ways, we use the test set Root Mean Squared Error (RMSE).

Methods: We updated the ChemCam database to ensure the accuracy and reliability of this data. We reviewed all standards to update values and fix errors. We increased the precision of modeled abundances, updated values with newly published data [e.g., 5], and verified ‘zero’ values as either true zero, less than quantification limit, or not measured and checked each spectra for the presence/absence of a line.

We used the Python Hyperspectral Analysis Tool (PyHAT) to mask and normalize the data [6]. We used a stratified k-fold procedure to propagate an equal distribution of data into five sets, with one set held out of the training model as a test set.

We introduced a two-stage data augmentation process to the training dataset to enhance generalizability. 1) synthesizing ten noisy spectra. For each target in the training set, we averaged the spectral intensities across all five occurrences and then introduced noise generated using the standard deviation of the target’s spectra. 2) we simulated +/- 1 channel pixel shifts likely in Mars data. To maintain the original channel range, we applied a mask and added zeros to the newly formed edge channels.

We tested two linear models and two non-linear models: PLS, Least Absolute Shrinkage and Selection Operator (LASSO) regression using the Least Angle Regression (LARS) algorithm for efficiency, Extremely Randomized Trees (XT), and a Convolutional Neural Network (CNN). LASSO, leveraging LARS, can efficiently and effectively process high-dimensional data [7, 8]. XT is a random ensemble learning technique that can outperform conventional decision tree regressors by randomly generating decision splits, yielding stronger generalizations in shorter training time [9]. We also experimented with a simple CNN architecture to contrast results with a more complex regression model.

During the training process, each model was tuned using a cross validation grid search, optimizing a single hyperparameter, to improve predictive accuracy. For PLS, we tuned the “number of components,” which controls the dimensionality of the reduced space in which regression is performed. For the LASSO models, we tuned “alpha” to regulate the sparsity of the model, or how many regression coefficients are set to zero. For ExtraTrees, we tuned “max features,” which adjusts the number of spectral channels to consider when making a split in the decision tree.

We employed a stratified k-fold splitting procedure, identical to our initial train/test dataset segregation, to partition the training data for validation. This data was then used to iteratively fit a model initialized with a hyperparameter in a given domain. During each iteration, a different fold was excluded and used as a validation set, consequently allowing us to obtain more representative RMSE values for each hyperparameter.

After this cross validation, we calculated the mean and standard deviation of the RMSEs across all folds for each hyperparameter. Using the minimum mean RMSE, we applied a simple algorithm to determine if a simpler model would achieve comparable results [10]. This algorithm identifies smaller hyperparameter values whose RMSE error bars (standard errors) overlap with the error bars of the minimum mean RMSE. By optimizing predictive accuracy and model complexity, we create a more generalizable model that avoids potential overfitting.

Results: The full range of standards used in this study span the following compositional ranges: Al₂O₃ 0-61.97 wt%, CaO 0-54.9 wt%, FeO 0.0187-92.41 wt%, K₂O 0-15.35 wt%, MgO 0-49.37 wt%, MnO 0-60.84 wt%, Na₂O 0-46.48 wt%, SiO₂ 0-97.81 wt%, and TiO₂ 0-75.26 wt%.

Contemporary cross validation methods select the hyperparameter that results in the smallest error (red, Fig 1). However, to minimize model complexity, we ultimately chose the smallest hyperparameter within one standard error of this value (green, Fig 1).

All of the proposed methods reduced the RMSEs of Al₂O₃ (Fig 2) and TiO₂ predictions. Specifically, XT had a smaller error than both the MOC and PLS1-SM. The CNN further reduced the error for the Al₂O₃ model. The CNN and LASSO MgO and K₂O models outperformed
the MOC, but did not reach the same performance as PLS1-SM. PLS also improved MgO MOC error, surpassing even XT.

**Discussion:** We obtained impressive results using simpler models than currently employed calibrations. Typically, the two non-linear models performed better than the linear models, likely due to their greater complexity. However, there were instances when the linear models could be favorable, namely for CaO, K₂O, and MgO. In these instances, it is possible that XT and CNN are too complex, leading to potential overfitting. While the newly investigated models are not a universal improvement, they provide similar results to current implementations while covering a wider range of test compositions and being robust to noise and slight changes in the wavelength calibration.

It is somewhat difficult to compare previous results with the current models because they fail to accurately reflect relative performance. The current ChemCam model utilizes blended submodels that have individually scored RMSEs, but this nuance is often lost when averaging them to compare with models that do not use submodels, like those we tested.

**Conclusion and Future Work:** This methodology gives way to simple and robust model creation. Going forward, Bayesian neural networks could offer insightful uncertainty quantifications. They have proven relevance when learning from datasets with fewer samples, like the ChemCam calibration database [12]. We also plan to investigate its application in our proposed methodology to enhance predictive accuracy. Additionally, as we explore more complex models, we plan to implement Bayesian Optimization during cross validation as an alternative to the currently employed grid search algorithm. This new approach allows the model to converge to optimal parameters faster by using each iteration to choose the next optimal hyperparameter value to test [13].

Furthermore, we also plan to investigate more realistic alternatives to our proposed data augmentation, such as training models on shot-to-shot data.

**Acknowledgments:** We gratefully acknowledge support from the NASA Mars Exploration Program and CNES.