CALIBRATION TRANSFER FOR SPECTROSCOPY IN SPACE SCIENCE. Thomas Boucher[1], M. Darby Dyar[2], CJ Carey[1], Stephen Giguere[1], and Sridhar Mahadevan[1], [1]College of Information and Computer Sciences, University of Massachusetts Amherst, Amherst MA 01003, USA, boucher@cs.umass.edu, [2]Dept. of Astronomy, Mount Holyoke College, South Hadley, MA 01075.

Introduction: Calibration transfer (CT) is the process of transferring a calibration curve from one instrument to another or from one set of environmental conditions to another without the need to resample all calibration standards. This task is especially critical in space science where extraterrestrial instruments calibrated with terrestrial data sets are used. For example, the current calibration target set for the ChemCam LIBS instrument on the Curiosity rover [1] includes only 10 standards, so Earth-based laboratory data (>326 standards [2]) are being used to develop prediction algorithms for Mars data that assume identical instrument performance. The addition of CT has the potential to improve the accuracy of Mars predictions by aligning laboratory and Mars spectra. This problem is not unique to ChemCam; CT can be used to augment any instrument calibration after deployment.

There are two primary ways of performing CT: by changing the weights of the calibration model or by directly modifying lab-recorded spectra of the standards to align them with those on Mars. This research uses the latter method, where CT is used to calculate a transformation function that maps spectra from one condition to another. There are a variety of methods for calculating the transfer function, but the most common approach relies on a set of overlapping standards that have been recorded under both conditions (or on both instruments). This overlapping subset is the linking set between the two data sets. For future flight instrument design, it is important to know the minimum size of an effective linking set.

So here we examine the effectiveness of using CT while varying the size of the linking set. To answer the more general question of how best to utilize the linking set, three meta-algorithms that do not use CT are also included in our comparison. Lastly, we explore implementation details of how best to use CT.

Experiments and Results: For all experiments, a large LIBS data set composed of 280 samples recorded under Mars-analog conditions at Mount Holyoke College [3] was used. Data were recorded at laser energies of 3.2% power (1.6 mJ) and 7% power (3.5 mJ), referred to here as *low power* and *high power*. Spectra of identical samples recorded at these two varying powers differ because the ionization states that populate LIBS plasmas are temperature-dependent.

All of the experiments followed the same basic task structure: given a large set of low power standards and a smaller overlapping linking set of high power standards, all with known wt.% SiO_2 composition, predict the wt.% SiO_2 of an unknown set of high power spectra. Root mean squared error was calculated using cross validation, where each experiment was repeated 50 times with random shuffling of samples per trial to ensure an accurate estimate of the error.

Without CT: In the first experiment, we test the hypothesis that using a series of preprocessing steps is sufficient; CT is often unnecessary and sometimes harmful [4]. Three meta-algorithms were compared (**Figure 1**). The term meta-algorithm is used because all models tested used the same underlying calibration algorithm, partial least squares (PLS), and the same set of preprocessing was applied to all of the data, but each model was fitted using a different set of training data. The task for all models was the same: predict wt.% SiO_2 in a set of unknown high power spectra. The same set of preprocessing steps was applied to all data (including unknowns) at the start of the experiment. These included wavelength and intensity alignment, baseline/continuum removal, and normalization. The number of PLS components was tuned individually for each model using the high power linking set to prevent models from overfitting their varied training data, especially *No Link*, which was most susceptible.

Figure 1 shows that *No Link*, in which low power is used to predict high power with no common data, shows constant but poor performance because the model disregards one of the most basic rules in statistical forecasting; the testing data ought to be drawn from a similar distribution as the training data. *Small Train* performs very well when the size of the linking set is

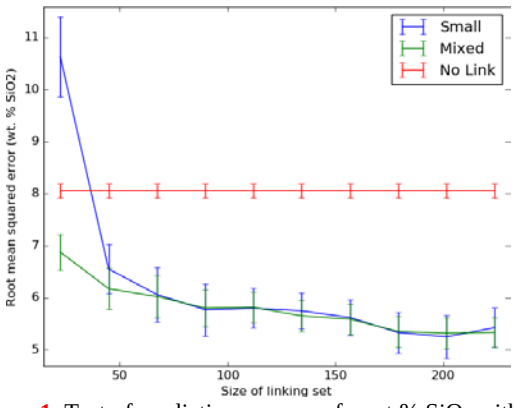

Figure 1. Test of prediction accuracy for wt.% SiO_2 without using CT. *No Link* was a model trained using exclusively the low power standards. *Small Train* was a model trained using only a small subset of high power linking standards. *Mixed* was a model trained on both the low power and the high power linking standards.

large, but its performance is poor and unstable when the size of the linking set is <50 standards. The *Mixed* model yields the lowest prediction error due to two effects: the high power linking set prevents PLS from overfitting the large low power standards, and the linking set encourages PLS to find components that are well-suited to predicting both power settings.

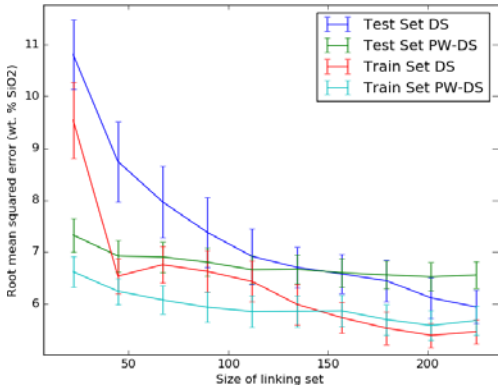

Figure 2. Test of prediction accuracy for wt.% SiO_2 using CT and testing two classic methods of calibration transfer for each meta-algorithm: direct standardization (DS) and piecewise direct standardization (PW-DS) [5].

With CT: In the second experiment, the same procedure was repeated to explore the usage details of CT. Two different CT meta-algorithms were tested (**Figure 2**), and a window size of three was tuned as optimal for PW-DS. In the first CT meta-algorithm, the PLS calibration was unmodified and a transfer function was learned to transfer the high power unknown testing spectra into a native low power representation, so the same calibration could be used. This method is referred to here as *Test Set* CT because the unknown testing spectra are those being transferred. In the second meta-algorithm, a transfer function was learned from low power to high power and was used to transfer the known low power standards to a high power representation. The PLS calibration was then fit on the high power representation and used to predict the unmodified high power unknowns. This method is referred to as *Train Set* CT because the whole training set was transferred and the PLS model refitted.

Figure 2 shows that PW-DS was nearly universally better than DS, except when the linking set size grew very large. It was expected that PW-DS would outperform DS when the number of linking standards remained small because it imposes additional constraints, forcing the model only to consider a small window of source channels when transferring each target channel, that better condition the pseudo-inverse problem. As the size of the linking set increases, the additional window constraints perform feature selection and prevent the model from overfitting, which DS is prone to do. Although DS uses all source channels when predicting each target channel, this global information did not prove to be beneficial. We suspect this is due to the non-smooth nature of LIBS spectra.

For both DS and PW-DS, the *Train Set* model outperformed the Test Set model. Initially, we believed this was due to the high power spectra being an intrinsically better predictor of SiO_2, but the experiment was repeated with high and low power swapped and a similar result was observed. Thus, this improvement in performance results because PLS selects only those components that are well transferred. If a channel was poorly transferred using the *Test Set* model and the PLS calibration preferred this channel in its components, then performing CT can hinder the performance. However, if a channel was poorly transferred using the *Train Set* model, then when the PLS calibration was refit, it could be trained ignore these channels.

Full comparison: **Figure 3** compares the CT and non-CT methods, showing all three non-CT methods along with the best performing CT model, focusing on smaller linking sets.

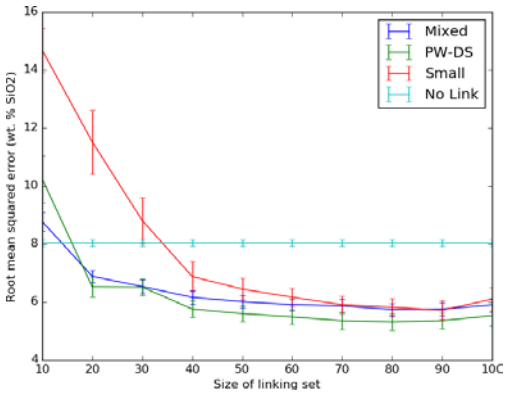

Figure 3. Test of prediction accuracy for wt.% SiO_2 using CT and testing showing all three non-CT methods (Figure 1) along with the best performing CT model (Figure 2).

Conclusion: When the size of the overlapping linking set was small (≤10 samples), training on the large homogeneous but foreign *No Link* set was the optimal choice. With a larger linking set (>10 samples), significant performance improvement was gained by including the overlapping set in its original form (*Mixed* model) or in its calibration transferred form (PW-DS model). As the size of the overlapping set increased, the benefit of using CT increased as well. These results suggest that use of a mixture of Mars calibration target data and laboratory training set data might improve the accuracy of ChemCam predictions on Mars.

Acknowledgments: Work supported by NSF grant CHE-1306133 and NASA grant NNX12AK84G.

References: [1] Wiens R. C. et al. (2013) *Spectroch. Acta B, B82*, 1-27. [2] Clegg S. M. (2014) *LPS XLV*, Abstract #2378. [3] Dyar M. D. et al. (2015) *LPS XLVI*, Abstract #1510. [4] Feudale R. N. et al., (2002) *Chemometr Intell Lab., 64*, 181-192. [5] Wang Y. et al. (1991) *Anal. Chem., 63*, 2750-6.