**CORRELATIONS BETWEEN MULTISPECTRAL IMAGING AND COMPOSITIONAL DATA FROM THE MARS EXPLORATION ROVERS AND IMPLICATIONS FOR MARS SCIENCE LABORATORY (MSL) DATA ANALYSIS.**

R. B. Anderson[1] and J. F. Bell III[2], [1]U.S. Geological Survey, Astrogeology Science Center, Flagstaff, AZ 86001 (rbanderson@usgs.gov), [2]Arizona State University, Tempe, AZ.

**Introduction:** The Mars Exploration Rovers have provided multispectral imaging and in-situ compositional measurements of the martian surface, which in turn have provided valuable insight into the planet's geologic history. While multispectral imaging can be used to survey the rover's surroundings, compositional information is typically restricted to locations near the spacecraft. It is therefore desirable to seek relationships between multispectral imaging and compositional data to allow inferences to be made about the composition of objects visible in the distance but inaccessible to contact instruments such as the alpha-particle x-ray spectrometer (APXS) and Mössbauer.

**Data Sets:** We restricted our study to the list of Pancam 13-filter observations that contained spots disturbed by the rock abrasion tool (RAT) or the Mössbauer nose-plate. Restricting the dataset in this way ensured that the Pancam spectra correspond as closely as possible to the measured APXS and Mössbauer compositions. Perfect correspondence is not possible because Pancam collects photons that have interacted with the upper several µm of the surface, while the APXS and Mössbauer measurements collect information from depths of tens to hundreds of µm [1,2]. In addition, the APXS and Mössbauer measurements are averages over the instrument's field of view (~2.5 cm for APXS and ~1.5 cm for Mössbauer) [1,3].

The Multi-mission Image Processing Laboratory (MIPL) at the Jet Propulsion Laboratory, coregistered the left and right eye images from Pancam [4], allowing the images to be digitally stacked so that individual pixel spectra could be extracted using the same coordinates in both the left and right eye images.

**Correlations**: Prior to conducting multivariate analysis of the data, we investigated the relationships between the MER Pancam spectra (and spectral parameters), APXS, and Mössbauer data by calculating their correlation coefficients (R)[5]. These calculations were done using the average Pancam spectra for each region of interest (ROI).

*Results:* The correlations are generally weak, as shown in Table 1. The higher absolute correlation coefficient values for Meridiani data are caused by several soils and Bounce Rock, which were spectrally and compositionally distinct from the Meridiani bedrock. The higher correlations for Meridiani data may also reflect the lower amount of dust cover at Meridiani.

**Table 1:** Maximum absolute correlation (R) value

| *Site* | *Pancam & APXS* | *Pancam & Mössbauer* |
|---|---|---|
| Gusev | 0.48 | 0.59 |
| Meridiani | 0.85 | 0.84 |

The correlation between ferric phases and the red-blue ratio across all Gusev ROIS was poor (R=0.21), but was improved when only RAT-abraded spots were considered (R=0.55). This suggests that the sampling depth difference between the instruments and/or the presence of dust coatings or weathering rinds may be responsible for the lack of correlation.

**Multivariate Methods:** After examining the correlation coefficients between Pancam spectra and APXS and Mössbauer data, we used two multivariate methods to search for any relationships between the data sets not captured by the simple correlation coefficient. We used Partial Least Squares regression (PLS) [6], commonly used in multivariate calibration [7, 8, 9], to predict target composition based on Pancam spectra. We also implemented "soft independent modeling of class analogy" (SIMCA), a common classification method [e.g., 8, 10, 11, 12], to classify the Pancam spectra of the individual pixels in each ROI.

Gusev data were classified according to their APXS class, but for Meridiani data, there were not previously-defined APXS classes. We therefore defined classes based on k-means clustering [13] of the PDS-released Meridiani APXS and Mössbauer results. To ensure that all APXS oxide or Mössbauer phases were considered equally, the values for each oxide or phase were mean-centered and normalized to the standard deviation of that variable. To determine the optimum number of clusters, we implemented the cluster validity parameter described by [14].

The resulting six clusters correspond to distinct materials encountered during Opportunity's traverse: (1) Endurance crater bedrock, (2) Auk (dark soil), (3) Sol 373 wheel trench soil, (4) Escher (brushed), (5) Bounce Rock, and (6) Meridiani bedrock.

In our analysis, we designated some of the Pancam observations of each class as training samples and the remaining observations as test samples. This designation was random, although brush mosaics were restricted to the test set because of their variable dust coverage. The training samples were used to generate the SIMCA and PLS1 models, which were then used to classify and predict the composition of the test samples. The PLS1 and SIMCA calculations were run in the Unscrambler software v9.8 (Camo, Inc.), which calculates a modified, squared correlation coefficient ($\bar{R}^2$) and root mean squared error (RMSE) as metrics to evaluate the PLS performance.

*Results*: The results for most of our PLS1 calculations are poor ($\bar{R}^2$=0.01 to 0.34 for Gusev), with relatively high RMSE values. The PLS1 results from Meridiani Planum are somewhat better than those from Gusev crater (*e.g.*, up to $\bar{R}^2$ of 0.86 for pyroxene), but this is again likely caused by differences in dust coatings between the sites and/or outliers such as the pure pyroxene Bounce Rock and the dark soils in clusters 2 and 3. To test the Meridiani PLS1 pyroxene model, we used it to predict the pyroxene content of the Gusev targets based on their Pancam spectra. The results were poor (RMSE ~24%), indicating that the model is not generalizable to datasets other than the training set.

SIMCA classification at Gusev crater showed mixed results. Of the 24 Gusev test ROIs with known classes, 11 had >30% of the pixels in the ROI classified correctly, while others were mis-classified or were not classified at all. These errors reflect the disconnect between Pancam spectra and the sample composition: in some cases nearly identical spectra have very different compositions (Figure 1), while in other cases samples in the same APXS class have very different Pancam spectra. The variation in the spectra of samples within the same APXS class may indicate isochemical weathering at a low water-to-rock ratio.

The SIMCA classification results for the data from Meridiani Planum were better than those from Gusev, with >30% of pixels correctly classified in 9 of the 11 ROIs with known classes. It is possible that the improvement in Meridiani results was because the classes were automatically defined by clustering based on both APXS and Mössbauer results, while the Gusev crater classes were manually determined based on APXS only. Additionally, the smaller number of classes and the predominance of Meridiani bedrock in the data set may have led to more frequent correct classifications.

**Implications for MSL:** The MSL mission lacks remote mineralogic detection capability and will rely on multispectral Mastcam images to identify distant targets, followed by elemental analysis from <7 m with ChemCam and in-situ measurements with APXS and CheMin. Our work indicates that extrapolating the results of chemical analyses to distant targets will likely be difficult. However, the compositions derived from the first few ChemCam shots on a target provide information about the upper several microns, and may show a closer correlation with the Mastcam spectra than APXS measurements, which are averaged over a larger volume.

The methods used in this study are useful for the analysis of multispectral and compositional data sets, even if the correlations between the data sets are weak. K-means clustering can be used to define classes within any data set and, as shown above with the Meridiani data, the derived clusters can correspond to meaningful geologic differences in the targets. Once classes within a data set have been defined, SIMCA can be used to formalize the classification of new samples and the identification of novel classes encountered as the mission progresses.

Although our studies using MER data indicate that it is challenging to draw conclusions from inter-data set correlations, it is important to search for these correlations using the MSL payload to maximize the ability to identify targets of interest from a distance.

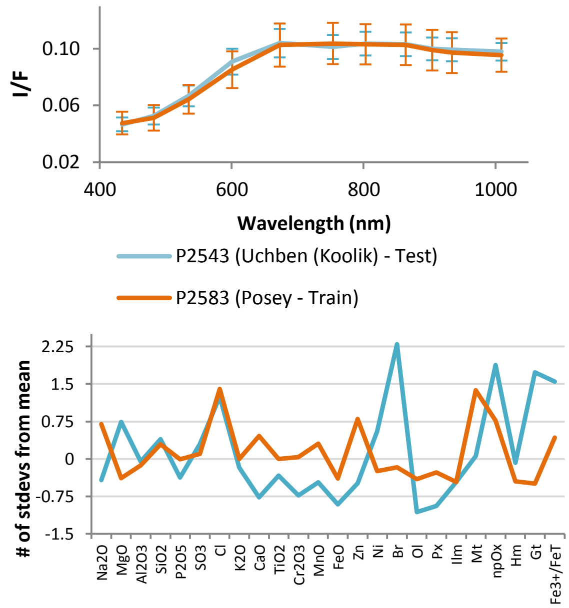


**Figure 1**: (a) Nearly identical average Pancam spectra of the Clovis-class Uchben (Koolik) region of interest (ROI) and the Barnhill-class Posey ROI. (b) Plots of the mean-centered and standard deviation-scaled APXS and Mössbauer values for Uchben Koolik and Posey.

**References:** [1] Brückner, J. et al., (2008) *The Martian Surface: Composition, Mineralogy, and Physical Properties*, Bell III, J. F. Ed., 58-101. [2] Klingelhöfer, G. et al., (2003) *6th Intl. Conf. on Mars*, 3132 [3] Klingelhöfer, G. & Morris, R. V. (2008) in *The Martian Surface: Composition, Mineralogy, & Physical Properties*, Bell III, J. F. Ed. 339-365. [4] Deen, R., Alexander, D. & Maki, J. (2004) *Proc. 8th SpaceOps*, 1. [5] Rodgers, J.L., Nicewander, W.A. (1988) *Am. Stat*., 42, 59–66. [6] Geladi, P. (1986) *Anal. Chim. Acta*, 185, 1-17. [7] Clegg, S.M. et al. (2009), *Spectrochim. Acta B*, 64, 79–88. [8] Anderson, R.B. et al., (2011) *Icarus*, 215, 608-627. [9] Adams, M.J. & Allen, J.R. (1998) *Analyst*, 123, 537-541. [10] de Oliveira, F.S. et al., (2004) *Fuel*, 83, no. 7-8, 917-923. [11] Vogt et al., N.B. (1987) *Env. Sci. & Tech*., 21, 35-44. [12] Munson, C. et al., (2005) *Spectrochim. Acta B*, 60, 1217-1224. [13] MacKay, D. J. C. (2003) *Information Theory, Inference & Learning Algorithms*, 284-292. [14] Ray, S. & Turi, R.H. (1999) *Proc. 4th ICAPRDT*, 137–143.