Thermal infrared spectral analysis of fine grained compacted mineral mixtures: assessment of applicability of Partial Least Squares (PLS) methods and implications for spectral interpretations of Martian sedimentary materials. Cong Pan and A. Deanne Rogers, Stony Brook University, Department of Geosciences, 255 Earth and Space Science Building, Stony Brook, New York, 11794-2100, panconggeosciences@gmail.com.

Introduction: Characterizing the thermal infrared (TIR) spectral mixing behavior of compacted fine-grained mineral assemblages is necessary for facilitating quantitative mineralogy of sedimentary surfaces from spectral measurements. Previous researchers have demonstrated that TIR spectra from igneous and metamorphic rocks as well as coarse-grained (>60 micron) sand mixtures combine in proportion to their volume abundance [1][2]. However, the spectral mixing behavior of compacted, fine-grained (<10 μm) mineral mixtures that would be characteristic of sedimentary depositional environments has received little attention. In spectral regions where the absorption coefficient is large, surface scattering should be reduced for packed grains relative to loose powders, and thus spectral shapes should resemble that of coarse particles in those regions. However, where the absorption coefficient is small, transmission through small grains (<5μm) is likely to occur [e.g. 3, 4]. Thus we would expect some differences between pressed powders and coarse grains in some spectral regions. Our previous work [5] has shown that many pressed pellet samples of <10 μm mineral mixtures have contributions from volume scattering over portions of the mixed spectrum, leading to non-linear spectral combinations and reduced accuracy of mineral abundance retrievals using the linear least squares minimization (NNLS) [1,6] technique that has been traditionally applied to remote TIR data sets. Here we assess the applicability of partial least squares (PLS) analysis [7] to these mixture spectra, to determine whether model accuracy can be improved. PLS is a statistical method that generalizes and combines features from principal component analysis (PCA) and multiple regression. It has been widely used in chemometrics; for example, it is one of the primary data reduction techniques employed by the Mars Science Laboratory Rover Curiosity ChemCam team [8,9].

Data and Methods: Major primary and secondary minerals found on the Martian surface including feldspar, pyroxene, smectite, sulfate and carbonate, were crushed with an agate mortar and pestle and centrifuged to obtain < 10 μm size. Pure phases and mixtures of two, three and four components were made in varying proportions by volume. All of the samples were pressed into pellets at 15000PSI.

Thermal emission spectra of pellets were measured at Stony Brook University from ~225 to 2000 cm^{-1}. In order to avoid dehydration of sulfate during measurement, sulfate-bearing mixtures were cooled to >30° below the detector temperature and then measured [10].

Non-negative linear least squares minimization (NNLS) using spectra of pellets and powders of pure minerals was used to derive mineral abundance of mixtures over the 400-1400cm^{-1} spectral range. PLS analysis [7] was then employed to generate a calibration model from which unknown mineral abundance of mixture spectra (testing data set) can be predicted using known spectra and mineral abundance (training data set). Here, we used feldspar and sulfate mixtures as the testing data set and all the pure phases and mixtures except feldspar and sulfate mixtures as the training data set (a total of 74 mixtures). Both NNLS and PLS involve generating regression models to solve the linear problem: Y=BX, where Y is one or several dependent variables, X is the independent variable or predictor variable and B is the matrix of regression coefficients. For NNLS, Y is the mixture spectrum and X is the spectral library (usually consisting of pure mineral sample spectra [1,6]). For PLS, Y is known mineral abundance and X is the spectra of pure and mixture samples with known abundances.

PLS analysis determines the statistical linear correlation between the known mineral abundance and observed spectra. Unlike NNLS, PLS can deal with data with strongly correlated, noisy and numerous X-variables. Also, similar to PCA, PLS weights X variables, which reflects the covariance between X and Y. PLS may be suited to TIR spectra because for any given mineral, the emissivity in many channels will vary together. And, it can weight the major and minor mineral absorption features unequally, which is fit to derive abundance of mixtures here.

Results: Here, two component mixtures of varying proportions of feldspar and sulfate are shown as an example. Modeled spectra from NNLS show close similarity to measured spectra (**Figure 1**). However, the predicted abundances are further from the known abundance compared to the PLS result (**Table 1**). **Figure 2** shows the PLS-modeled abundance and known abundance of feldspar for the training data set. Except for a few points, most of the points fall close to the y=x line, indicating the model is working well to predict feldspar abundance in the training data set. The resulting regression analysis generates correlation coefficients between feldspar abundance and emissivity for each spectral channel (**Figure 3**). It identifies the correlation with feldspar thermal infrared spectral featurse at

~450, 500,600, and 1000 cm^{-1} with two shifted absorption features near ~1100 cm^{-1}. Applying the training data to the test data set (a two-component feldspar and sulfate mixture), the predicted abundances fall within+/-15% of known abundance and show an improvement over NNLS-modeled values.

Discussion and Future Work: Our preliminary results suggest that PLS is a promising technique for retrieving mineral abundance from TIR spectra of fine-grained mineral mixtures. Though the spectral components combine non-linearly across the full spectral range, portions of the spectral range (generally, where absorption coefficients are high) do vary in a predictable manner (**Figure 1**), which likely contributes to the success of PLS. Future work will evaluate the performance of the PLS model for all minerals in our controlled mixtures, and for a variety of expanded training data sets being built from previous and ongoing, complementary studies.

References: [1] Ramsey, M. S., and Christensen, P. R., JGR, 1998, 103,577-596.[2] Feely, K. C., and Christensen, P. R., JGR, 1999, 104, 24195-24210.[3] Salisbury, J. W., et al. JGR, 1987,92, 702-710.[4] Cooper, C. D., and Mustard, J. F, Icarus, 2002, 158, 42-55.[5]Pan, C. and Rogers, A. D., AGU 2012, Abstract #P13A-1906. [6] Lawson and Hanson, *Solving Least-Squares Problems*, Prentice-Hall, 1974 [7] Wold, H 1982, North-Holland, Amsterdam. [8]Clegg S., M., et al, spectrochimica Acta PartB, 2009, 79-88. [9] Dyar et al., *Spechemica. Acta. B, 70,* 51-67 (2012) [10] Baldridge, A. M., and Christensen, P. R., Applied spectroscopy, 2009, 63, 678-688.
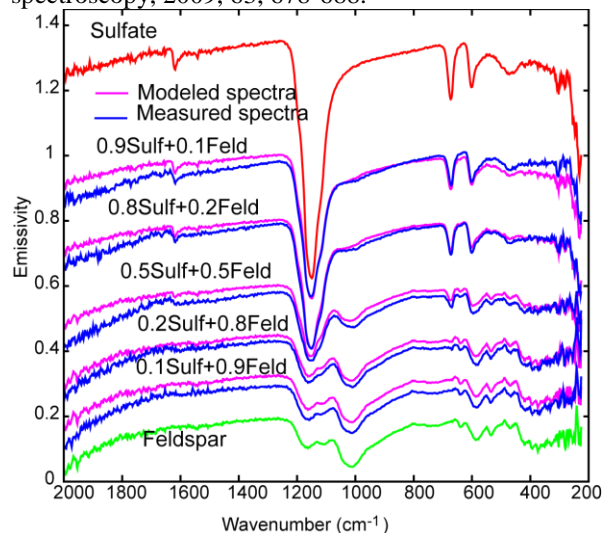

Figure 1 Example two-component series using feldspar and sulfate and the NNLS model. Spectra from pressed pellets are included for the end-members. The remaining spectra are measured and modeled spectra of the controlled mixtures. The modeling range used was 400 to 1400 cm^{-1}.

Table 1 Known abundance v.s. predicted abundance of feldspar and sulfate mixtures of NNLS and partial least squares models

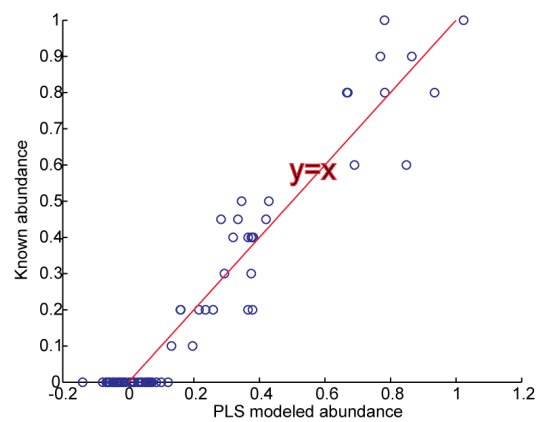| known feldspar abundance	NNLS predicted abundance	PLS predicted abundance
0.1	0.244	0.1702
0.2	0.3409	0.2752
0.5	0.7245	0.6548
0.8	0.9335	0.9252
0.9	0.9724	0.9571

Figure 2 Known abundance and partial least square analysis modeled abundance of feldspar from training data set. The line is y=x.
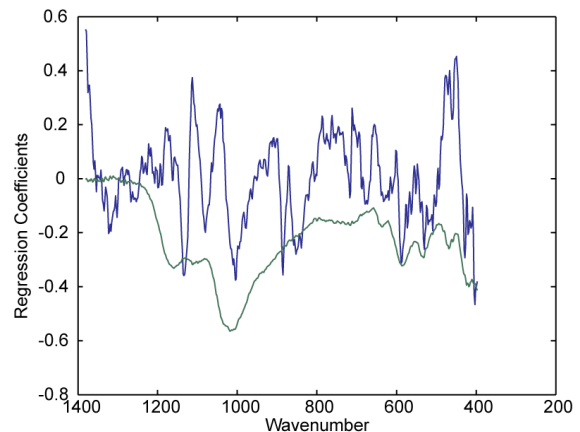

Figure 3 Regression coefficients (blue) of feldspar generate from partial least squares analysis and thermal infrared spectrum of feldspar (green, the y-axis of spectrum is emissivity).The model identifies feldspar absorpstion features in ~450,500,600,1000 cm^{-1} wavenumbers.