Introduction: Pyroxene is an important liquidus phase in basalt and it also occurs in many other rock types throughout the solar system including Earth, Mars and many asteroid families. Both remote and in situ spectroscopic methods for identification of pyroxene are thus needed. Although interpretations of remote sensing of pyroxene are well-understood [1,2], in situ mineral identification is less well studied. Such work is needed to interpret data from upcoming Mars rovers equipped with Raman spectrometers: ExoMars RLS, Mars 2020’s SHERLOC and Super-Cam [3-5].

Raman spectra of pyroxene group minerals have many different modes, of which the high intensity bands at ~670 and ~1010 cm$^{-1}$ may be most diagnostic of composition. Here, we use well-characterized synthetic [6] and naturally occurring pyroxenes with compositions covering the bottom half of the Wo (Ca$_2$Si$_2$O$_6$), En (Mg$_3$Si$_2$O$_6$) and Fs (Fe$^{2+}$Si$_2$O$_6$) quadrilateral to develop unmixing multivariate models to predict composition using Raman spectra. This work seeks to improve upon previous univariate methods that predict composition based on peak shifts [7] by testing other types of models using the largest sample suite ever studied [6].

Background: Raman spectra of pyroxene group minerals have many modes that contribute to the complex Raman signature. Dominant features at ~670 and ~1010 cm$^{-1}$ are attributed to Si-O-Si symmetric stretching and Si-O symmetric stretching vibrations [6,8], respectively. These bands may appear as singlets (e.g., augite) or doublets (e.g., enstatite) depending on composition and crystal structure. Additionally, peak positions shift in wavenumber with changing composition.

Spectral differences are apparent in Figure 1 where clinopyroxenes (CPX) and orthopyroxenes (OPX) samples are shown separately. These Raman spectra were acquired by Aaron Celestian at Western Kentucky University (WKU) and illustrate how crystal system (monoclinic versus orthorhombic) affects the spectra.

Methods: Raman spectra of 233 pyroxene samples (107 synthetic and 126 natural) have been collected to date. Synthetic samples were made by Donald Lindsley at Stony Brook University (SBU) and are fine particulates (~45 µm). Natural samples vary in size and the amount of sample available but are mostly granular.

Figure 2. Ternary diagram of synthetic and natural pyroxene samples utilized in this study.

Compositions of all natural samples were confirmed with electron microprobe. Analyses of 10 spots on each sample were acquired either by Molly McCanta at the University of Tennessee in Knoxville or by Joseph Boesenberg at Brown University. Compositions of the 233 pyroxenes are summarized in Figure 2.

Raman spectra of each pyroxene were acquired at WKU, SBU, and/or Mount Holyoke College on four different types of Raman instruments. Although several samples were run on multiple systems, some samples have only been examined with one instrument. This work is ongoing.

Univariate Results: Spectra from Figure 1 show the ~670 cm$^{-1}$ feature (Figure 3) shifting to higher wave-
numbers with lower %Fs content. But this trend is complex and dependent on Mg, Fe, and Ca contents. These data show that an individual peak position can only roughly be used to infer composition. In contrast, multivariate analysis techniques can utilize the entire Raman spectral energy range (or subsets thereof) to predict all three compositional variables.

**Machine Learning Results:** Partial least squares (PLS) and least absolute shrinkage operator (Lasso) are used in this study for compositional predictions. PLS utilizes all channels of the spectra while Lasso reduces the number of channels within the model [9,10]. PLS-1 is used to predict individual variables while PLS-2 predicts multiple variables simultaneously, leveraging information about their co Dependence. Both techniques have been applied to compositional predictions using several spectroscopic techniques [11-13].

Preliminary results from PLS and Lasso predictions (Table 1) show prediction errors of the three elements separately and collectively (assuming Ca+Mg+Fe=1) for one dataset. These predictions rely on 95 AirPLS baseline-corrected and normalized Raman spectra from WKU. Model accuracy is reported as leave-one-out cross-validated root mean square error (LOO RMSE-CV) values. This error is calculated by removing one sample at a time using a regression model based on the other N-1 samples to predict the Nth sample. LOO RMSE-CV gives a robust estimate of model performance on unseen data and is expressed in the same units as the measurements being predicted.

The most accurate predictions are obtained from models that separate clinopyroxene from orthopyroxene spectra, though prediction of Wo is useful for all samples. For these data, PLS-1 outperforms Lasso models consistently. Additionally, PLS-2 models that consider Wo, En and Fs together outperform individual predictions (PLS-1).

**Discussion and Future Work:** Separating CPX and OPX predictions and using multivariate PLS-2 models reduce prediction error. These two conclusions will be applied to expanded future pyroxene datasets. Work planned for this project includes four steps:

1. Analyze the entire sample suite with multiple instruments.
2. Compare prediction accuracy among instruments to evaluate the effects of resolution, energy range, and laser excitation wavelength.
3. Experiment further with pre-processing techniques to determine if they affect results (baseline removal, squashing, normalization, etc.).
4. Develop mathematical expressions that consider the centroids of multiple peaks to provide best estimates of composition in mixtures.

These results will be invaluable for analyses of Martian rocks, minerals, and soils by Raman instruments on upcoming missions.