Why is This Important?

- Olivine composition varies over a solid solution between forsterite (Mg$_2$SiO$_4$) and fayalite (Fe$_2$SiO$_4$).
- Fe/Mg is diagnostic of crystallization temperature (Figure 1).
- Fe/Mg also affects phase relations. For example, forsterite does not occur with quartz, and is stable to great depths, while fayalite occurs only rarely in granites and high-Si rocks like rhyolites. Olivine minerals control many of the properties of Earth’s upper mantle, affect tectonics, and may store H in the mantle.

Strategy:

- Use 93 well-characterized synthetic and naturally occurring olivines.
- Test existing univariate methods for prediction of Fe/Mg from Raman data that build on effects of Raman shift with composition (Figure 2).
- Develop improved machine learning algorithms with quantitatively determined accuracies for measuring Mg and Fe in olivine with Raman.

Diagnostic olivine modes:

There are 36 Raman active modes including: (1) A$_1$ + (11 B$_{1g}$ + 7 B$_{2g}$ + 7 B$_{3g}$) [5]. The olivine doublet that is key to this study occurs at: ~815-825 cm$^{-1}$ (DB1) and ~838-857 cm$^{-1}$ (DB2) [3]. Five vibrational modes form this doublet: (2 A$_g$ + 2 B$_{1g}$ + 2 B$_{3g}$) [1,3,6,7]

Methods:

- EMPA at Brown University to determine composition (Figure 3).
- Mössbauer spectroscopy at Mount Holyoke for %Fe$^{2+}$.
- Raman spectra (Figure 4) acquired on two Bruker instruments: 1) E-RAVO dual laser system (758/852 nm): Powdered samples with 3 sample scans and 10s integration.
- 2) Senterra 532 nm laser: Single crystals using 10mW laser, 2 sample scans and 10s integration.
- Baseline correction using AirPLS (smoothness = 100).
- Normalization to account for arbitrary intensity differences.
- Multivariate (MVA) models used partial least-squares (PLS), least absolute shrinkage and selection operator (lasso), or least angle regression (LARS). See nemo.umass.cs.edu:54321 for tools.

Univariate (2-peak) Results:

- Fo-Fa ratio can be estimated [3] using central peak positions of the olivine peaks (DB1 and DB2) as in the literature.
- Linear, 2nd, and 3rd-order polynomial fits relating peak position to composition were performed (Figure 5).
- Equations for fits are given in Table 1 with the $R^2$ of the fit and the RMSE-CV in units of %Fo. The latter is the error bar obtained using leave-one-out cross validation.

Multivariate (800-880 cm$^{-1}$) Results:

- Machine learning methods using MVA significantly outperform peak centroid-based regression models, as is also observed in LIBS [16,17] and XAS [19,20] spectroscopies.
- PLS or Lasso are preferred techniques for determining Fo contents when optimum prediction accuracy is desired.

Figure 1. Phase diagram showing the relationship between olivine composition and temperature.

Figure 2. Normalized Raman spectra of lauhunite (blue), forsterite (orange) and fayalite (green). The high intensity doublet (DB1 and DB2) is present in all samples between 815 and 857 cm$^{-1}$.

Figure 3. Histogram of compositions of 93 samples used in this study. Typical natural olivines form with $\approx$90% Fo, resulting in an imbalanced distribution on the Fo-Fa series.

Figure 4. Raman spectra of the olivine doublets (DB1 and DB2) in this study from both Bruker spectrometers. Spectra were baseline-corrected using Air-PLS and normalized.

Figure 5. Relationships among peak centroids and forsterite content. RRUFF data acquired from http://rruff.info. Kuebler et al. (2006) fits here match those published. Other literature (purple) refers to values from [1, 4-9, 18-15].

Figure 6. Prediction errors for data sets (Senterra, BRAVO, and RRUFF) when the trained models use spectra from the same or different instrument(s).

Effects of Instrument:

- To examine the generalizability of our results, we predicted each data set independently and with aggregated data sets (Figure 6).
- Predictions from each data set (BRAVO, Senterra and RRUFF) predict their own Fo contents most accurately, despite differences in detection sensitivities, resolutions and number of spectra within the model. These factors must be detangled.
- Further work comparing data on the same set of samples on multiple instruments is in progress.

Table 1. Comparison of regression vs. multivariate methods for %Fo predictions. Multivariate models use wavenumber range of 800-880 cm$^{-1}$. Only BRAVO spectrometer data.

<table>
<thead>
<tr>
<th>Model</th>
<th>Components/alpha</th>
<th>$R^2$</th>
<th>RMSE-CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lasso</td>
<td>0.91</td>
<td>2.33</td>
<td></td>
</tr>
<tr>
<td>Lars</td>
<td>2</td>
<td>0.93</td>
<td>7.32</td>
</tr>
</tbody>
</table>

Conclusions and Ongoing Work:

- Multivariate models produce far superior prediction accuracy using data from wavenumbers covering from 800-880 cm$^{-1}$.
- We are working to develop a penultimate general model for predicting Fo (Mg vs. Fe) composition from Raman data.

Acknowledgements: We are grateful for support from the Massachusetts Space Grant Consortium and the NASA RISE (SSERVI) award.

References: