PREDICTING ILMENITE-GEIKIELITE COMPOSITION USING RAMAN SPECTROSCOPY THROUGH MULTIVARIATE ANALYSIS. L. B. Breitenfeld[1], M. D. Dyar[1], Leif Tokle[2], and Kevin Robertson[2], [1]Dept. of Astronomy, Mount Holyoke College, South Hadley, MA 01075, breit22l@mtholyoke.edu, [2]Dept. of Earth, Environmental, and Planetary Sciences, Brown University, 324 Brook St., Providence, RI 02912.

Introduction: Ilmenite is an important mineral group on planetary surfaces, especially including Earth, the Moon [1], and Mars, as evidenced by SNC meteorites [2]. It occurs in all rock types. There is a solid solution among ilmenite ($Fe^{2+}TiO_3$), geikielite ($MgTiO_3$), and pyrophanite ($MnTiO_3$) at high temperatures. Most terrestrial occurrences are Fe-rich except in kimberlites, where Mg substitution occurs [3]. Lunar ilmenite can also be enriched in Mg [4]. Ilmenite on the Moon may be important as a resource because it supplies both Fe and Ti, critical for building *in situ* structures. Characterizing the composition of ilmenite on the Moon and other remote bodies informs its usefulness [5].

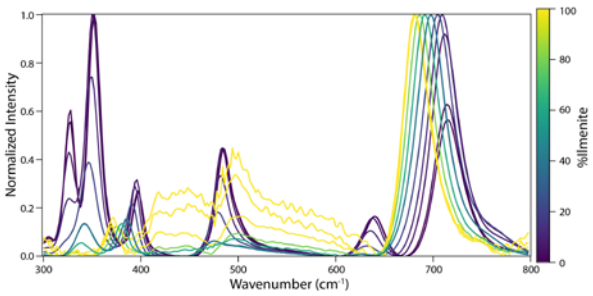

Figure 1. Ilmenite and geikielite normalized Raman spectra acquired on Bruker's BRAVO spectrometer. Spectra are color- coded based on %Ilm content, where pure $Fe^{2+}TiO_3$ is represented with yellow and pure $MgTiO_3$ with purple.

In this study, we explore the use of Raman spectroscopy to measure the composition of synthetic samples covering the solid solution between ilmenite and geikielite. The cation ratio of Fe to Mg (%Ilm= (100 × Fe)/(Mg + Fe)) affects positions of the Raman peaks, allowing compositions to be predicted. Analogous trends have been observed and studied extensively for other mineral solid solutions, such as olivine [6-7]. Here, a multivariate approach is utilized to predict the Fe/Mg for the ilmenite-geikielite solid solution.

Raman Modes: Belonging to the $R\bar{3}$ space group, both ilmenite and geikielite have three acoustic modes and 27 optical modes. Of these, there are 10 major bands in Raman spectra of ilmenite group minerals that occur between 200 and 800 cm^{-1}. These Raman features are caused by the combination of $5A_g$ and $5E_g$ [8, 9].

Methods: All 11 synthetic ilmenite and geikielite powders utilized in this study were synthesized from oxides in a 1 atmosphere $CO:CO_2$ furnace by Leif Tokle. Microprobe analysis shows impurities in all powders are < 1% and chemically homogeneous [10]. Powders were sieved to a grain size fraction of 10 – 20 μm for all samples.

Spectra were acquired on a Bruker BRAVO Raman dual laser spectrometer with five sample scans and an integration time of 10 s, at a spectral resolution of 2.0 cm^{-1}/channel. Spectra were baseline-corrected using adaptive iteratively reweighted penalized least squares (AirPLS) [11] with smoothness set to 100. All spectra were normalized to the intensity of the largest peak.

Results: The spectral range (300-3200 cm^{-1}) of the BRAVO only allowed seven of the ten peaks to be observed. Geikielite spectra (purple) in **Figure 1** have bands at roughly 306 cm^{-1} (E_g), 327 cm^{-1} (A_g), 352 cm^{-1} (A_g), 397 cm^{-1} (E_g), 485 cm^{-1} (E_g), 487 cm^{-1} (merged with 485 cm^{-1}) (A_g), 641 cm^{-1} (E_g) and 714 cm^{-1} (A_g) [8]. In contrast, peaks in spectra of pure ilmenite (yellow) shift to lower energies as Fe is added, although several bands are poorly resolved.

Using multivariate analyses, Fe/Mg was predicted using two multivariate algorithms: partial least squares

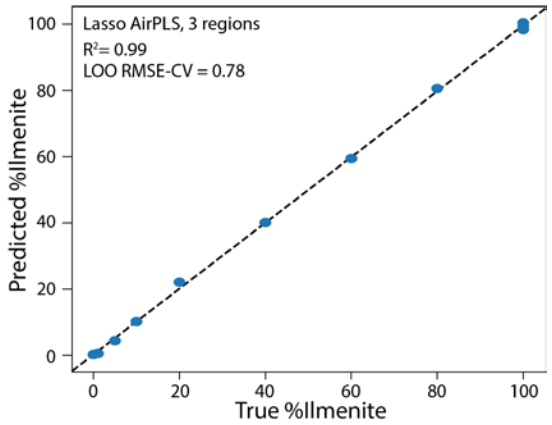

Figure 2. Lasso predictions using wavenumber ranges (310-420 cm^{-1}, 460-530 cm^{-1}, and 600-770 cm^{-1}). All spectra were normalized to maximum peak intensity and baseline-corrected using AirPLS. Error bars are smaller than the symbols for %Ilm by EMPA.

(PLS) and least absolute shrinkage operator (Lasso) [7] (**Figure 2**). These techniques can exploit a broad spectral range, including the multiple diagnostic Raman peaks, and do not depend on any single peak's position.

Three wavenumber ranges were selected for the predictions. The first included the entire spectral range (300-3200 cm^{-1}), the second encompassed the major bands broadly (300-800 cm^{-1}), and finally only the

ranges surrounding known peaks were used for evaluation (310-420, 460-530, and 600-770 cm^{-1}). Prediction accuracy was evaluated with internal R^2 values, internal root mean square error (RMSE) values and leave-one-out (LOO) RMSE cross-validated (CV) values (**Table 1**) as in [7]. LOO RMSE-CV values were calculated by leaving out one spectrum at a time and building a prediction model with the remaining data, then averaging the resultant errors. RMSE values are in units of %Ilm.

Table 1. Error of multivariate ilmenite and geikielite predictions for normalized Raman spectra (units of %Ilm)

	σ	C/α	Internal R^2	Internal RMSE	LOO RMSE-CV
(a)	300-3200	10	1.00	<0.00	5.76
	300-800	10	1.00	<0.00	1.19
	3 regions	9	1.00	0.02	**1.04**
(b)	300-3200	10	1.00	<0.00	5.30
	300-800	9	1.00	0.02	1.07
	3 regions	10	1.00	<0.00	**0.87**
(c)	300-3200	0.1	0.99	0.96	2.59
	300-800	0.1	0.99	0.96	1.87
	3 regions	0.1	0.99	0.96	**0.76**
(d)	300-3200	0.7	0.99	2.69	2.72
	300-800	0.1	0.99	1.06	1.97
	3 regions	0.1	0.99	0.91	**0.78**

Notes: σ = wavenumber range (cm^{-1}), C/α = number of components for PLS models and α for Lasso models. 3 regions uses 310-420 cm^{-1}, 460-530 cm^{-1}, and 600-770 cm^{-1}. Four treatments are used: (a) PLS raw, (b) PLS AirPLS, (c) Lasso raw, and (d) Lasso AirPLS.

The PLS method regresses one response variable (%Ilm) against multiple explanatory variables (intensity at each channel of the spectra), assigning coefficients to every single channel. Because PLS utilizes all available variables (channels) and eliminates multicollinearity, PLS produces small RMSE values.

Lasso uses continuous shrinkage, which allows coefficient values at specific channels to be reduced to zero [9] if they contribute nothing to prediction. The shrinkage parameter *t*, shrinks the residual sum of squares based upon the sum of the absolute value of the coefficients. In other words, this method selects a subset of predictors with the strongest effect on the response variable. Accordingly, the Lasso produces a sparse models with few coefficients (depending on the α parameter), with most channel intensities set to zero.

Discussion: Raman spectroscopy allows Fe/Mg within the ilmenite-geikielite solid solution to be predicted with a high degree of accuracy (Table 1). Error bars on predictions of %Ilm range from ±0.76-1.04 in those units, depending on the algorithm, pre-processing, and energy range used; this is nearly as accurate as electron microscopy. This high-performance result is in keeping with other multivariate analyses predictions observed for Raman [7], LIBS [12, 13] and XAS [14, 15] spectroscopies. It must be caustioned that these errors apply only to analyses of unknown ilmenites under these analytical conditions, and would not directly exptrapolate to unseen data from a completely different instrument. Further work would be needed to quantify prediction errors in such as scenario.

Each prediction model created here used raw data or baseline corrected data (AirPLS). Although baseline removal using AirPLS lowered (improved) prediction errors for the PLS predictions, it increased those for the Lasso models. In general, baseline correction has a minimal effect for this dataset. For broader applications, baseline removal will likely be necessary to mitigate differences among varying instruments [16].

Predictions that isolated only the spectral bands outperformed those that covered larger wavenumber ranges. This trend has previously been observed [7] and is expected because isolating spectral features removes the noise associated with channels that lack predictive value, instead weighting those that account for Raman band shifts associated with cation substitutions.

Finally, Lasso slightly outperformed PLS for this data set. Relative merits of PLS versus Lasso in spectroscopic methods (e.g., [17]) are just beginning to be explored, so there is no consensus yet for which method is better. Their usefulness appears to be highly variable for each dataset and application-dependent.

These models could be strengthened by increasing the size of the sample suite and introducing natural mineral samples. Ongoing research includes generating equivalent models using different Raman spectrometers at different laser wavelengths.

Acknowledgments: We are grateful for support from the RIS^{4}E SSERVI.

References: [1] Heiken G. H., Vaniman, D. T. and French B. M. (Eds.) (1991) *CUP Archive*. [2] Rull F. et al. (2004). *J. Raman Spectros., 35*, 497-503. [3] Shee, S.R. (1984) Kornprobst, 1, 59-73. [4] Papike J. J. et al. (1976) *Geophys. Space Phys.*, 14, 475–540. [5] Heiken G. H. and Vaniman D. T. (1990) *Proc. Lunar Planet Sci. Conf.*, 20, 239-247. [6] Kuebler K. E. et al. (2006) *Geochim. Cosmochim. Acta, 70*, 6201-6222. [7] Breitenfeld L. B. et al. (2017) *Amer. Mineral.*, (in review). [8] Okada T. et al. (2008) *Amer. Mineral., 93*, 39-47. [9] Tibshirani R. (1995) *J. Royal Stat. Soc*., 58, 267-288. [10] Tokle and Robertson (2018) *Lunar Planet Sci. Conf.* [11] Zhang Z.M. et al. (2010) *Analyst*, 135, 1138-1146. [12] Tucker J. et al. (2010) *Chem. Geol., 277*, 137-148. [13] Dyar M. D. et al. (2016) *Spectrochim. Acta B, 123*, 93-104. [14] Dyar M. D. et al. (2012) *Amer. Mineral., 97*, 1726-1740. [15] Dyar M. D. et al. (2016) *Amer. Mineral., 101*, 1171-1189. [16] Carey C. et al. (2015) *J. Raman Spectros.*, DOI: 10.1002/jrs.4757, 894-903. [17] Dyar M.D. et al. (2012) *Spectrochim. Acta B*, 70, 51-67.