

ABSTRACT

Olivine minerals control many of the properties of Earth's upper mantle. Additionally, olivines affect rheology, they may store hydrogen, and they are diagnostic of crystallization temperature. Olivine's presence in meteorites and terrestrial bodies makes the study of this mineral group crucial to the planetary science community. Olivine composition can be characterized using Raman spectroscopy. This *in situ* technique will be used to characterize Martian minerals in upcoming missions such as *ExoMars* and *Mars 2020*.

Raman spectra of 93 olivines were acquired on Bruker's 532 nm Senterra spectrometer. Of these samples, 25 were also run on Bruker's BRAVO and Senterra (785 nm) spectrometers. Raman spectra of the olivine group minerals in the solid solution between forsterite (Mg_2SiO_4) and fayalite (Fe_2SiO_4) have a high intensity doublet between 800 and 880 cm^{-1} . Historically, the band shift of these two peaks was utilized to predict the Mg/Fe contents (Kuebler et al., 2006; Foster et al., 2007; Gaisler and Kolesov, 2007; Mouri and Enami, 2008; Yasuzuka et al., 2009; Ishibashi et al., 2011), though these studies used different instruments and only limited data sets. This thesis compares the band shift method for understanding olivine composition with a more novel method using partial least squares (PLS), the least absolute shrinkage and selection operator (lasso), and least angle regression (LARS).

To evaluate the accuracy of each univariate model, both the R^2 value of each fit (on a plot of composition versus peak centroid) and the cross-validated root mean squared error (RMSE-CV) of each model were used. Internal cross-validation of each data set was the most accurate %Fo predictor. For example, the model for which all data were acquired on the BRAVO spectrometer (under identical operating conditions) predicts %Fo content best for other data acquired on the BRAVO. For this reason, previous studies may appear deceptively accurate because they did not

use cross-validation nor evaluate the accuracy of predictions on “unseen data” (data not included within the model).

Aggregated data sets from multiple instruments show excellent performance and can be generalized to other instruments for which calibrations are not available, such as the upcoming Raman Mars instruments. The most accurate %Fo predictions that avoid instrument bias result from an aggregated model produced through PLS or lasso multivariate analyses. A model that isolates the olivine doublet is also suggested instead of utilizing the entirety of the spectrum.

Overall, this thesis demonstrates that multivariate analyses are superior to univariate methods for prediction of olivine composition from Raman spectroscopy. Multivariate analyses that use multiple instrument data avoid instrument bias and leverage multiple aspects of the spectra. Recommended PLS and lasso models with the smallest errors are listed in the appendix of this thesis for the use of future workers.

Predicting Olivine Composition Using Raman Spectroscopy

Through Band Shift and Multivariate Analyses

Laura Brucia Breitenfeld

Submitted to the Department of Astronomy at

Mount Holyoke College with the purpose of honors consideration

Thesis Advisor: M. Darby Dyar, Ph.D.

May, 2017

For my Mom, Dad, Sarah and Paul

We are grateful for the support of the Massachusetts Space Grant Consortium and the NASA
RIS⁴E SSERVI node.

Tom Tague, Peng Wang, and Bruker Optics Inc. have been essential in generously providing
support for this research.

Thanks to the many Dyar lab students for supporting my work in Raman spectroscopy over the
last four years.

And most importantly, thank you Darby for being my inspiration and mentor.

TABLE OF CONTENTS

LIST OF TABLES	vi
LIST OF FIGURES	vii
Chapter 1: INTRODUCTION.....	1
Chapter 2: BACKGROUND.....	5
A. RAMAN SPECTROSCOPY	5
B. OLIVINE STRUCTURE AND CHEMISTRY	9
C. OLIVINE RAMAN SPECTRA	10
D. RATIONALE FOR THIS STUDY.....	16
Chapter 3: METHODS	18
A. SAMPLE PROVENANCE	18
B. SAMPLE CHARACTERIZATION	24
C. RAMAN MEASUREMENTS	25
D. DATA ANALYSIS	27
E. MODEL COMPARISONS	33
Chapter 4: RESULTS	36
A. UNIVARIATE (TWO-PEAK) RESULTS	36
B. MULTIVARIATE (MACHINE LEARNING) RESULTS	59
C. CONTRASTING UNIVARIATE AND MULTIVARIATE RESULTS	71
Chapter 5: DISCUSSION AND CONCLUSIONS.....	74
A. EFFECTS OF PRE-PROCESSING.....	74
B. IMPORTANCE OF TESTING UNSEEN DATA.....	76
C. USEFULNESS OF MULTIVARIATE ANALYSIS.....	77
D. CHOICE OF WAVENUMBER RANGE.....	78
E. EFFECTS OF SPECTRAL RESOLUTION.....	79
F. CONCLUSION.....	79
APPENDIX.....	82
REFERENCES CITED	83

LIST OF TABLES

Table 2.1. Olivine group minerals	10
Table 2.2. Summary of olivine doublet Raman modes.....	12
Table 2.3. Summary of olivine bands in literature.....	14
Table 3.1. Natural samples (this study)	19
Table 3.2. Synthetic samples (this study)	23
Table 3.3. Average composition of natural samples by EMPA	28
Table 3.4. Compositions of natural samples in formula units	30
Table 4.1. DB1 centroids (this study)	39
Table 4.2. DB2 centroids (this study)	42
Table 4.3. RRUFF centroids	45
Table 4.4. Other literature centroids	46
Table 4.5. DB1 univariate regressions with RMSE-CV	55
Table 4.6. DB2 univariate regressions with RMSE-CV	56
Table 4.7. Internal multivariate analyses with 3.0 cm ⁻¹ /channel resolution	63
Table 4.8. RMSE-CV tests for multivariate models with 3.0 resolution.....	68
Table 4.9. Internal multivariate analyses with as acquired resolution (matching spectral data) ..	69
Table 4.10. RMSE-CV tests for multivariate models with actual resolution	69
Table 4.11. Multivariate analyses that isolate the laser wavelength effect.....	71
Table 4.12. Summary of univariate and multivariate analyses	73

LIST OF FIGURES

Figure 1.1. Natural forsterite olivine sample	1
Figure 1.2. Olivine phase diagram.....	2
Figure 1.3. Forsterite and fayalite Raman spectra	3
Figure 2.1. Raman scattering (anti-stokes and stokes)	6
Figure 2.2. Rayleigh scattering	7
Figure 2.3. Visualization of Raman activity	8
Figure 2.4. Fayalite crystal orientations	11
Figure 2.5. Olivine doublet (DB1 and DB2).....	15
Figure 3.1. Histogram of synthetic and natural samples on the Fo-Fa series	21
Figure 3.2. Synthetic sample synthesis tube (Skulte, 2006)	22
Figure 3.3. %Fo prediction modeling (Kuebler et al., 2006).....	35
Figure 4.1. Univariate DB1 predictions (this study).....	48
Figure 4.2. Univariate DB2 predictions (this study).....	49
Figure 4.3. Univariate predictions on BRAVO and Senterra spectrometers (this study).....	51
Figure 4.4. Univariate predictions of published literature	53
Figure 4.5. Data predictions using univariate models.....	58
Figure 4.6. Summary of univariate RMSE-test values	59
Figure 4.7. Coefficients based on lasso alpha values.....	62
Figure 4.8. Wavenumber effect on multivariate predictions	66
Figure 4.9. Spectrometer effect on multivariate predictions.....	67
Figure 4.10. Summary of multivariate RSME-tests.....	70
Figure 4.11. Laser wavelength effect on centroid position.....	72



Figure 1.1. Natural forsteritic olivine sample.

Chapter 1

INTRODUCTION

The content of this thesis lies at the intersection of multiple disciplines: planetary science, mineralogy, chemistry, and computer science. This research primarily has geologic applications because its goal is to develop a method to measure the chemistry of olivine. The field of Raman spectroscopy is also central to this research. Specifically, spectral data processing and spectral multivariate analyses are discussed in detail and have implications to the greater spectroscopy community. Finally, this research is relevant to numerous planetary bodies within our solar system because olivine is a common mineral constituent of basalts, which are ubiquitous on terrestrial bodies.

Olivine group minerals are typically solid solutions between forsterite (Fo₁₀₀: Mg₂SiO₄) and fayalite (Fa: Fe₂SiO₄) (**Figure 1.1**). Limited substitutions of alternative cations may exist, such as Mn and Ni. Although the olivine group encompasses these minerals, this thesis focuses

primarily on forsterite and fayalite because these are by far the most commonly occurring olivine group minerals.

Because Fe/Mg is diagnostic of crystallization temperature (**Figure 1.2**) (Dyar et al., 2008), quantifying the ratio of forsterite to fayalite is critical. 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 such as rhyolites. Olivine is well-studied because of its presence in meteorites and on terrestrial surfaces (i.e., Earth and Mars) as a typical component of Mg and Fe rich igneous rocks. Olivine minerals control many of the properties of Earth's upper mantle, they affect rheology, and they may store hydrogen, therefore the study of olivine composition is essential to the planetary science community for understanding surface processes.

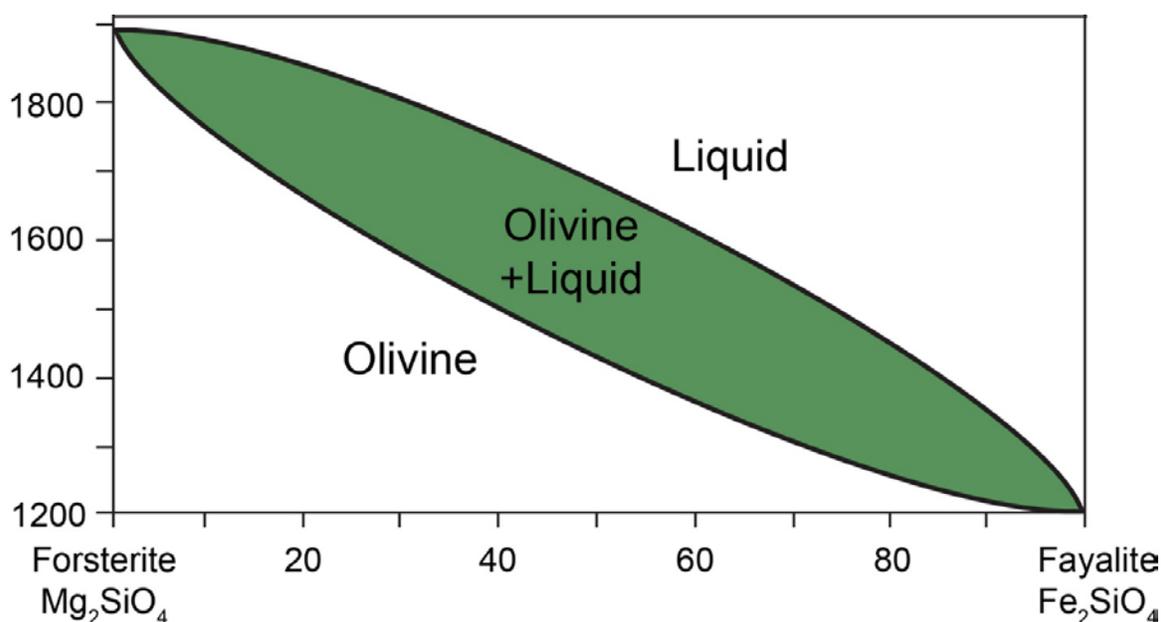


Figure 1.2. Phase diagram of the relationship between olivine composition and temperature in degrees Celsius.

Brown (1980) provides a comprehensive review of research performed on synthetic and natural olivines that span the Fo-Fa solid solution series. Ideally, olivine composition is

measured in a laboratory using sophisticated electron microprobe equipment, but such technology is impractical on a planetary surface. Thus it has become more imperative to be able to predict olivine composition using methods suitable for *in situ* uses, and of these, Raman spectroscopy is likely the best, as evidenced by the impending deployment of several Raman instruments to Mars (*ExoMars* and *Mars 2020*).

Raman spectroscopy is an impressive tool for analyzing various materials. Here the application is mineralogy. Determining the Fe/Mg composition of olivine depends on the band energies of olivine peaks in a Raman spectrum. Peak positions of a high intensity doublet in the range of 800-880 cm^{-1} can be used to identify the relative amounts of Mg (forsterite) and Fe (fayalite) in the sample (**Figure 1.3**).

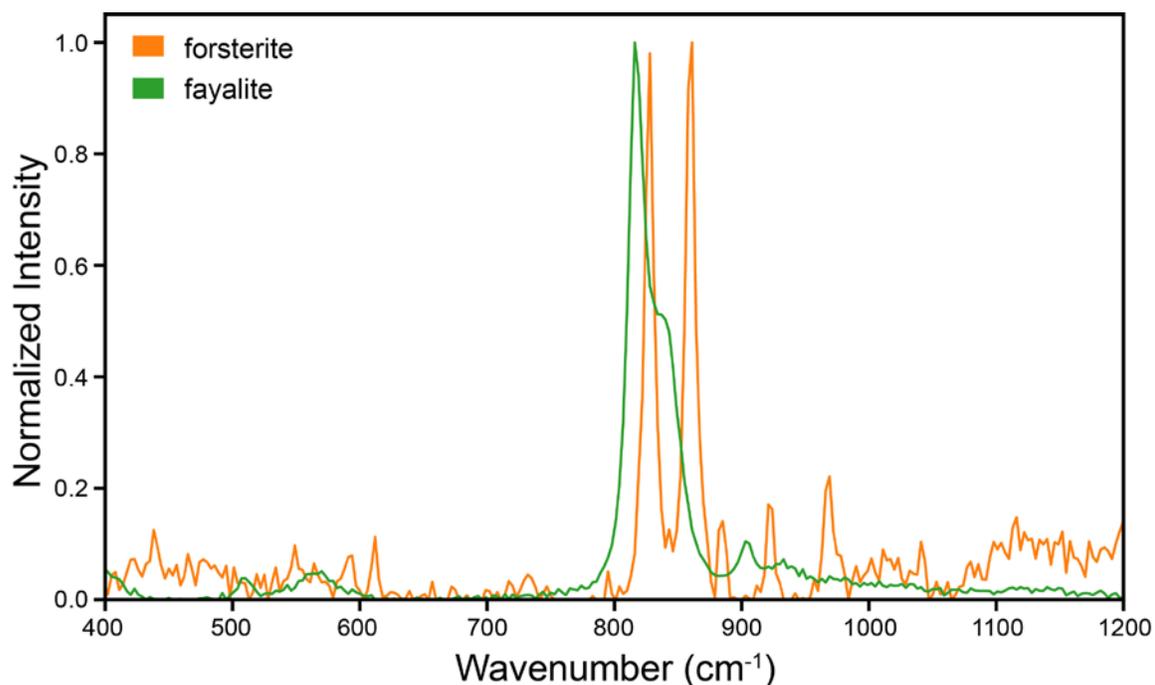


Figure 1.3. Normalized Raman spectra of forsterite (orange) and fayalite (green). The high intensity doublet (DB1 and DB2) is present in all samples between 815 and 857 cm^{-1} .

The energy of the forsterite doublet is greater than the fayalite doublet. Thus, as %Fo (i.e., $(100 \times \text{Mg}) / (\text{Mg} + \text{Fe})$ expressed as a percentage) content increases, the doublet energy increases. Univariate analyses for %Fo predictions have been made using the centroid positions of the two olivine peaks as seen in **Figure 1.3** (Kuebler et al., 2006; Foster et al., 2007; Gaisler and Kolesov, 2007; Mouri and Enami, 2008; Yasuzuka et al., 2009; Ishibashi et al., 2011). Alternatively, %Fo can be predicted using multivariate analyses, which use many channels within the spectra. There are advantages to this method because characteristics of the spectra (i.e., a band shoulder) are considered, instead of solely the centroid.

In this thesis, all known publicly-accessible olivine Raman data are considered. In addition, Raman spectra were acquired from a suite of 93 well-characterized synthetic and naturally-occurring olivines from the Dyar research lab using Bruker's BRAVO and Senterra spectrometers. These data sets are compared and contrasted to develop robust algorithms with quantitatively determined accuracies for measuring Mg and Fe in olivine through Raman spectroscopy. Conventional univariate approaches that relate peak position to composition are compared with a multivariate analysis approach that leverages a much broader wavenumber range and produces more accurate predictions. Results of this thesis will be immediately useful to the science teams of *Mars 2020* and *ExoMars* for use in accurately predicting olivine composition from Martian data.

Chapter 2

BACKGROUND

A. RAMAN SPECTROSCOPY

The Raman Effect, named after Chandrasekhara Venkata Raman, was first observed in 1928, although Smekal theoretically predicted this phenomenon in 1923 (Nasdala et al., 2004). Raman shared the following inspiring words regarding the capabilities of Raman spectroscopy in his acceptance speech for the Nobel Prize in Physics.

The frequency differences determined from the spectra, the width and character of the lines appearing in them, and the intensity and state of polarization of the scattered radiations enable us to obtain an insight into the ultimate structure of the scattering substance ... It follows that the new field of spectroscopy has practically unrestricted scope in the study of problems related to the structure of matter.

In Raman spectroscopy, band wavenumber position (expressed as energy in units of cm^{-1} along the x -axis) is linked to the chemistry of the material. Wavenumber is defined as the inverse of wavelength (λ) or Stokes ($\nu = \nu_0 - \nu^1$) divided by the speed of light (c) (**Equation 2.1**). Using the energies and shapes of Raman bands, identification of the primary material can generally be made. In the geosciences, this allows mineralogy to be determined for an Earth hand sample and for surface materials using *in situ* rover analyses on planetary bodies.

$$\tilde{\nu} = \frac{\nu}{c} = \frac{1}{\lambda}$$

Equation 2.1. Definition of wavenumber (cm^{-1}).

Through the interaction between the incident light and molecular vibrations, the wavelength of light is altered. This rare inelastic interaction occurs when the material energy rises or falls to a higher or lower state than the starting vibrational state (**Figure 2.1**) (Nasdala et al., 2004). The exciting photon energy is either increased or decreased compared to the phonon (vibrational)

energy. It follows that the change in these energies is equal to the difference between the vibrational levels of the given material.

Gain in vibrational energy and loss of the scattered photon energy is denoted as Stokes Raman scattering. Loss of vibrational energy and gain in scattered photon energy is referred to as anti-Stokes Raman scattering. Anti-Stokes scattering has a negative Raman shift and Stokes

Raman Scattering (inelastic)

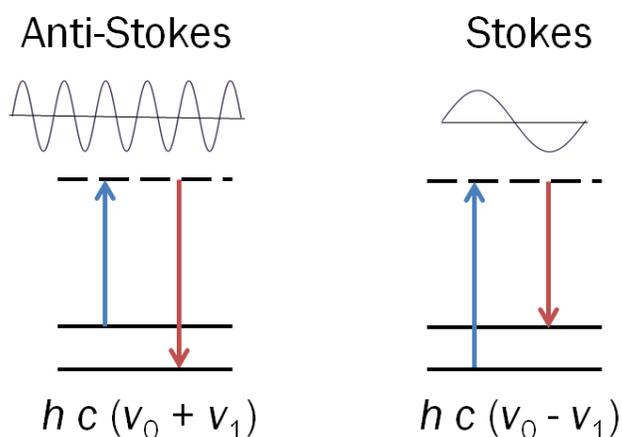


Figure 2.1. The anti-Stokes and Stokes processes of Raman scattering. Solid lines represent vibrational states while the dashed lines represent non-vibrational energy states.

scattering is recorded on the positive portion of the Raman spectrum. For context, infrared absorption occurs when light is absorbed and the photon energy corresponds to the exact energy difference between two vibrational levels.

The Rayleigh line is positioned at zero wavenumbers and therefore coincides with the y-axis. This line is named for the position at which Rayleigh scattering occurs. This phenomenon is the release of vibrational energy because of elastic scattering. Similar to Stokes scattering, the photon energy surpasses a higher vibrational state. Yet, in the case of Rayleigh scattering the original vibrational state is revisited (**Figure 2.2**). The ratio of Rayleigh to Raman scattering is 100,000:1, demonstrating the rarity of the Raman Effect (Nasdala et al., 2004).

Rayleigh Scattering (elastic)

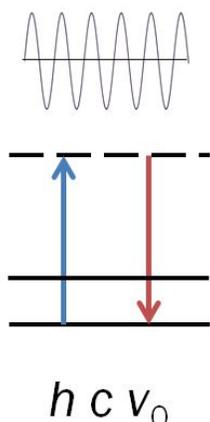


Figure 2.2. The process of Rayleigh scattering. Solid lines represent vibrational states while the dashed lines represent non-vibrational energy states.

Typically, wavenumber or Raman shift (the shift from the Rayleigh line) is used in Raman spectroscopy to denote the position of a band. The intensities of Raman peaks are arbitrary and depend on factors including, but not limited to, Raman cross section (the relative activity of the Raman signal for a given material), crystal orientation, laser power and integration time. Therefore, peak intensity should be analyzed comparatively by normalizing the spectra.

Raman spectroscopy can be utilized as an identification tool because each Raman band within a spectrum corresponds to a vibration of the nuclei which depends on mass, size and valences of the atomic material. Other factors such as bond forces and crystal symmetry can also affect the bands of the spectrum (Nasdala et al., 2004).

Modes within Raman spectra dictate the existence of bands within each spectrum, and not all modes are necessarily Raman active. Raman spectroscopy provides information on covalency of molecular bonds because the energies and intensities of Raman peaks reflect changes in bond polarization. When there is a change in polarizability, the mode will be Raman active. For

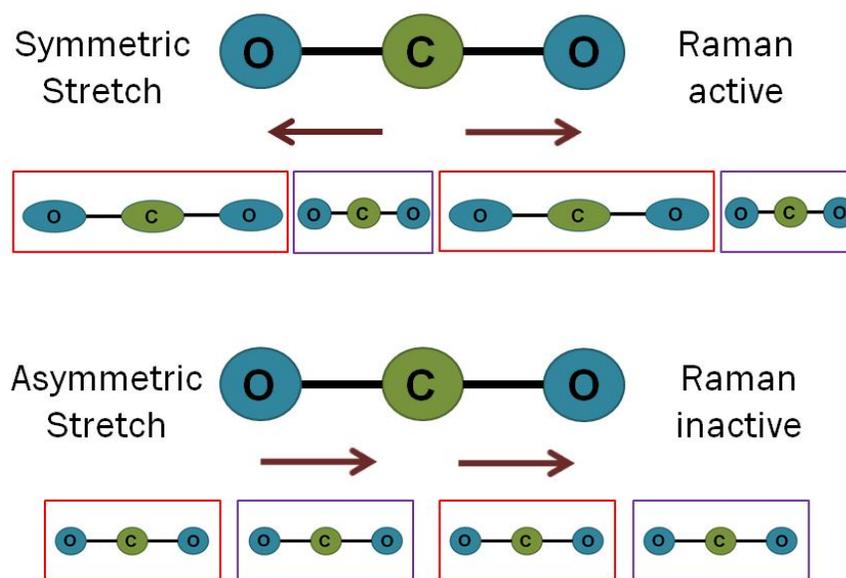


Figure 2.3. Visualization of Raman activity and inactivity for CO₂. The mode is active as polarizability changes.

example, in the case of a linear molecule such as CO₂, symmetric stretching would be Raman active yet, asymmetric stretching is Raman inactive (**Figure 2.3**).

Nomenclature for Raman vibrational modes and symmetries is extensive and terms are often used inconsistently within literature. In this thesis, symmetry terminology is employed, which consists of a capital letter with a subscript (i.e., A_g, B_{1g}). The capital letter specifies the degeneracy of the mode. “A” denotes single degenerate symmetry and “B” corresponds to single degenerate anti-symmetry in relation to the main symmetry axis. The lower case letter “g” (gerade) defines the mode as symmetric to the symmetry center. The number is listed to distinguish between vibrations of the same type within the system. Symmetric stretch (ν_1), symmetric bend (out of plane) (ν_2), antisymmetric stretch (ν_3) and antisymmetric bend (in plane bend) (ν_4) describe the nomenclature for mode assignments.

This nomenclature is used in this thesis to discuss relevant modes and their assignments to specific coordination types in the structure of olivine group minerals. As will be explained

further, chemical substitutions in the olivine group cause the site symmetries to change, and thus affect the energies of the Raman modes, allowing composition to be inferred through Raman.

B. OLIVINE STRUCTURE AND CHEMISTRY

Olivine is a mineral group that encompasses several silicate minerals (**Table 2.1**). Within the field of mineralogy, the Dana classification system is used to organize minerals by their physical characteristics. It is a hierarchical classification system similar to that of taxonomy within biology. There are four levels of Dana classification: class (composition and structure), type (ratio of cations to anions), group (similar structures) and species (unique structure and composition). By definition, all the mineral species in the olivine group (**Table 2.1**) have the same crystal structures, with minor variations in bond lengths and symmetries. Similarities among these species' compositions and structures result in shared characteristics within their Raman spectra (**Figure 1.3**). This thesis focuses on the Raman features that result from the substitutions of Fe^{2+} and Mg because they are by far the most common in naturally occurring rock types. Raman features may also be diagnostic of other cation substitutions (Ni, Mn, Ca, Fe^{3+} etc.), but these minerals are so rare that an appropriate number of samples could not be obtained to characterize them.

Olivine species are rarely pure, and solid solutions between divalent cations are common. However, the name used for any given specie is defined on the basis of which cation predominates in the octahedral site. For example, in forsterite or fayalite, the cation component of olivine is Mg^{2+} or Fe^{2+} , respectively, and the species' name is for whichever cation is >50% of the divalent cations.

Table 2.1. Mineral species within the olivine group with formulae and Dana numbers.

Species	Formula	Dana Number
fayalite	Fe ₂ SiO ₄	51.03.01.01
forsterite	Mg ₂ SiO ₄	51.03.01.02
liebenbergite	(Ni,Mg) ₂ SiO ₄	51.03.01.03
tephroite	Mn ₂ SiO ₄	51.03.01.04
laihunite	Fe ²⁺ Fe ³⁺ ₂ SiO ₄	51.03.01.05
monticellite	CaMgSiO ₄	51.03.02.01
kirschsteinite	CaFe ²⁺ SiO ₄	51.03.02.02
glaucochroite	CaMn ²⁺ SiO ₄	51.03.02.03

The structure of fayalite shown in **Figure 2.4** is representative of all the olivine group minerals; it is useful in understanding the bands within the Raman spectra of olivine. In mineralogical nomenclature, olivine is termed to be a nesosilicate. The structure is based on isolated tetrahedra composed of an Si⁴⁺ cation bonded to four O²⁻ anions, with divalent cations between them in octahedral sites. In its structure, each oxygen ligand is bonded to only one silicon atom, creating discrete SiO₄ tetrahedra. These non-bridging oxygen atoms form two nonequivalent octahedrally coordinated sites in the olivine structure, designated M1 and M2 (M is for metal), into which divalent cations can fit (**Figure 2.4**). The M1 site is smaller than the M2 site by 0.03-0.05Å, and is the more distorted of the two sites (Lumpkin and Ribbe, 1983). The M1 site is roughly distorted into tetragonally elongated octahedron where the M2 site is approximately resembles a trigonally elongated octahedron (Burns, 1993).

C. OLIVINE RAMAN SPECTRA

The assignments of Raman active forsterite and fayalite modes have evolved over time (**Table 2.2**). Generally, forsterite Raman bands above 500 cm⁻¹ can be classified as internal movement within the (SiO₄)⁴⁻ tetrahedra. Below this threshold energy, peaks are caused by rotation and translation of the tetrahedra as well as divalent cation motion (Mckeown et al., 2010). Forsterite and fayalite have 84 vibrational modes; only 36 are Raman active (11A_g + 11 B_{1g} + 7B_{2g} + 7B_{3g}) (Mckeown et al., 2010).

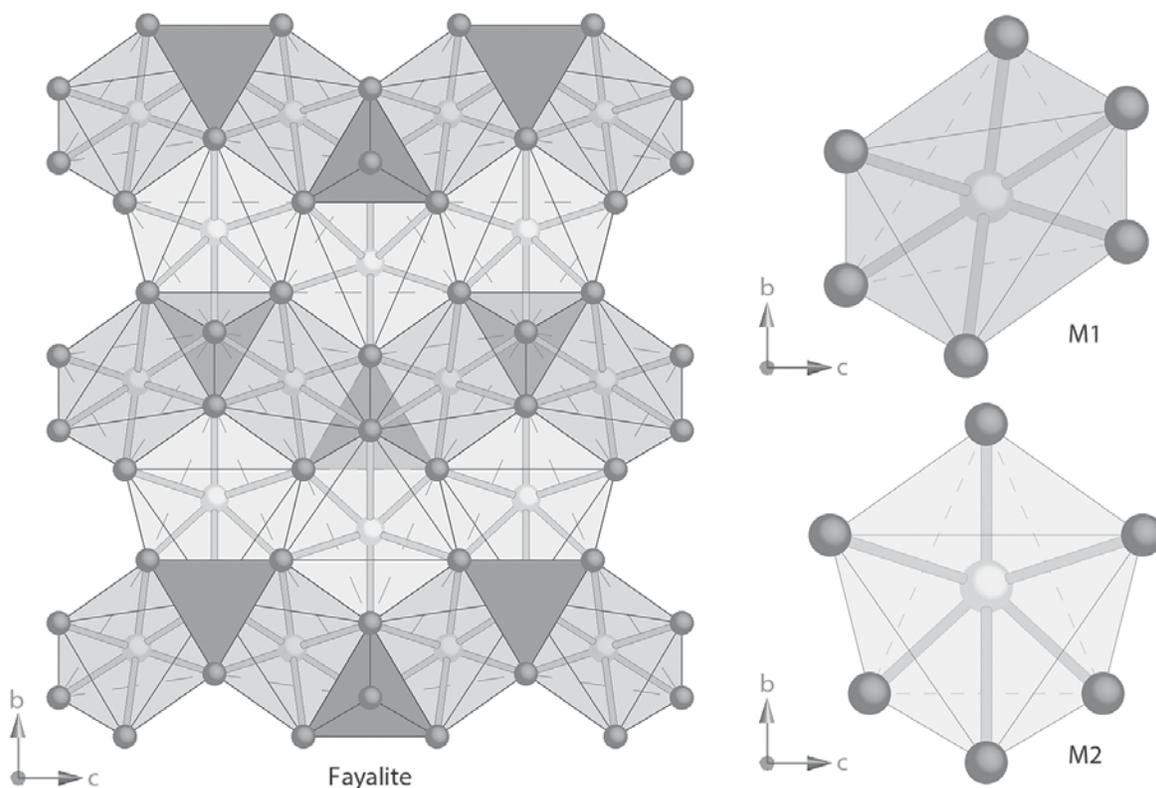


Figure 2.4. The crystal structure of fayalite, Fe_2SiO_4 , which is representative of all the olivine group minerals. The dark gray polyhedra denote the isolated Si^{4+} tetrahedra. The M1 and M2 octahedra form edge-sharing chains, which are parallel to the c axis. In these sites, Fe is the central atom, which is bonded to O. The M1 and M2 Fe sites are shown as dark and light tones. On the right of the diagram, the M1 and M2 octahedra have been emphasized to display their different sizes and bond lengths (Dyar et al., 2008).

Key to this study are the two principle Raman bands that form a doublet composed of five vibrational modes ($2A_g + 2B_{1g} + B_{2g}$) (**Table 2.2**). Peaks in this doublet occur between ~ 815 - 825 cm^{-1} (DB1) and ~ 838 - 857 cm^{-1} (DB2) (Kuebler et al., 2006). This doublet (DB1 and DB2) is primarily attributed to A_g (single degenerate symmetric stretch) though B_{1g} and B_{2g} (single degenerated antisymmetric stretch) also affects the shape of the spectrum (**Table 2.2**). The energy shift of the A_g stretch from the SiO_4 tetrahedra is caused by changes in site geometry due to cation substitutions in adjacent sites. Cation substitutions between forsterite (Mg^{2+}) and fayalite (Fe^{2+}) thus result in band shifts (Kuebler et al., 2006). The B_{1g} and B_{2g} modes also have low intensity contributions to the doublet. This doublet has thus been utilized for %Fo

(100×Mg)/(Mg+Fe) predictions of olivine composition using Raman spectroscopy (**Table 2.3**).

This predication method is tested within this thesis, yet it is also expanded upon with multivariate analyses.

Other peaks within olivine spectra have been utilized rarely for prediction of composition, such as ~200-230 cm⁻¹, ~290-310 cm⁻¹, ~410-440 cm⁻¹, ~540-553 cm⁻¹, ~881-883 cm⁻¹, ~914-920 cm⁻¹, 950-966 cm⁻¹ (**Table 2.3**). However, these features are relatively low in intensity compared to the DB1 and DB2 doublet. This makes peak centroid fitting difficult and less accurate and therefore, these regions are not used within this thesis. Additionally, Raman bands

Table 2.2. Summary of Raman olivine doublet modes with corresponding band positions (cm⁻¹).

~Band (cm ⁻¹)	Symmetry	Assignment	Literature
815-825 (DB1)	A _g	n.a.	Servoin et al. (1972)
		v ₁ +v ₃	Paques-Ledent and Tarte (1973)
		v ₁	Iishi (1978)
		v ₁ +v ₃	Piriou (1983)
		v ₁ +v ₃	Chopelas et al. (1991)
		SiO ₄ ²⁻ stretching	Kolesov and Tanskaya (1996)
		Si-O stretch, v ₃	McKeown et al. (2010)
838	B _{1g}	v ₁	Iishi (1978)
		v ₁ (+v ₃)	Chopelas et al. (1991)
		v ₁ (+v ₃)	Kolesov and Geiger (2004a)
		v ₁	McKeown et al. (2010)
837-857 (DB2)	A _g	n.a.	Servoin et al. (1972)
		v ₁ +v ₃	Paques-Ledent and Tarte (1973)
		v ₃	Iishi (1978)
		v ₁ (+v ₃)	Piriou (1983)
		v ₃	Price et al. (1987)
		v ₁ +v ₃	Chopelas et al. (1991)
		SiO ₄ ²⁻ stretching	Kolesov and Tanskaya (1996)
Si-O stretch; SiO ₄ breathing v ₃	McKeown et al. (2010)		
866	B _{1g}	v ₃	Iishi (1978)
		v ₃	Price et al. (1987)
		v ₃ (+v ₁)	Chopelas et al. (1991)
		v ₃ (+v ₁)	Kolesov and Geiger (2004a)
882	B _{2g}	v ₃	McKeown et al. (2010)
		v ₃	Paques-Ledent and Tarte (1973)
		v ₃	Iishi (1978)
		v ₃	Chopelas et al. (1991)
		v ₃	Kolesov and Geiger (2004a)
		v ₃	McKeown et al. (2010)

caused by different vibrational modes are not affected by octahedral substitutions. For example, features between 400 and 700 cm^{-1} have been attributed to the internal bending modes of the anion, which have minimal centroid shifts (Chopelas, 1991; Kuebler et al., 2006).

Surprisingly, previous studies used only very small sample suites, and thus only a limited number of olivine samples with confirmed compositions have been published with corresponding Raman peak centroids (**Table 2.3** and RRUFF database <http://rruff.info/>). Although several of these studies did not focus on Raman composition predictions, their data were nonetheless appropriate for use within this thesis for prediction models.

In the case of the DB1/DB2 doublet, Raman bands shifts (compared to the Rayleigh line) occur because of variation in Mg and Fe contents (**Figure 2.5**). The shape and relative intensity of DB1 and DB2 also change because of composition. Many previous workers (**Table 2.3**) have used the peak centroids of the DB1 and DB2 doublet peaks to derive olivine composition. However, this practice does not allow other information in the spectra to be utilized, such as shifts arising from minor modes that affect the shape of the primary doublet and give rise to other, more subtle features elsewhere in the wavenumber range.

There are other studies of Raman spectra of olivine that focus on different aspects of its paragenesis. For example, the ubiquity of olivine in Earth's mantle has inspired numerous Raman studies focusing on changes in spectra because of the influence of pressure and temperature (Besson et al., 1982; Heymann and Cellucci, 1988; Durben et al., 1993; Liu, 1993; Wang et al., 1993; Gillet, et al., 1997; Kolesov and Geiger, 2004a; Kolesov and Geiger, 2004b; Farrell-Turner et al., 2005; Rouquette et al., 2008; Manghnani et al., 2013; Weber et al., 2014). This research amply demonstrates that pressure and temperature affect the Raman signature in shape and peak position. As pressure increases, the wavenumber of the peak centroid increases.

Table 2.3. Summary of %Fo prediction models using band shift method within the literature.

Paper	# samples*	~Bands (cm⁻¹)	Main conclusions
Iishi (1978)	1 (3)	All bands	Comprehensive assignment of olivine modes and report of electronic polarizability.
Guyot et al. (1986)	4 (4)	815-825, 837-857, 881-883, 914-920, 950-966	Raman studies of 4 olivines with centroid positions. Early observation of centroid movement with compositional differences. Detection of band shift with replacement of Si with Ge in olivines.
Chopelas et al. (1991)	1 (4)	All bands	The forsterite sample was utilized within this thesis. The fayalite, monticellite and tephroite samples were not used because of composition impurities. Examination of single crystal lattice modes.
Mohanan et al. (1993)	1 (8)	All bands from 200-1000	Study includes 8 samples on the forsterite- monticellite solid solution. Only Fo ₁₀₀ is considered here. All 8 samples have reported peak centroids for DB1 and DB2. Band broadening based on composition is discussed.
Kolesov and Tanskaya, (1996)	2 (14)	All bands from 200-1000	Centroid positions are reported for 2 of the 14 samples. Confirms DB1 and DB2 are ideal bands for %Fo predictions. Examines cation distribution of the olivines.
Wang et al. (2004)	0 (2)	815-825, 837-857	First to quantify pyroxene and olivine composition through Raman band shifts of two Martian meteorites.
Kuebler et al. (2006)	10 (10)	815-825, 837-857	Develop comprehensive models to quantify change in olivine doublet due to composition. 10 samples used to build DB1 and DB2 models. Proposed peak centroid Fo-Fa prediction models.
Foster et al. (2007)	2 (6)	815-825, 837-857	Reports peak centroid shift with change in compositions using 6 samples; centroid positions listed for 2. Focuses on hypervelocity impacts.
Gaisler and Kolesov (2007)	0 (11)	200-230, 290-310, 410-440, 815-825, 837-857	Reports peak centroid shift with change in compositions using 11 samples; centroid positions listed for 0 samples. Conclusions made regarding the percolation threshold and spin-vibration interaction of olivines.
Mouri and Enami (2008)	0 (25)	815-825, 837-857	Reports centroid shift with change in compositions using 25 samples; centroid positions listed for 0 samples. Applications include fluid inclusions and pressure phases of diamond and zircon.
Ishibashi et al. (2008)	1 (1)	815-825, 837-857	Examined relationship between Raman spectrum and crystallographic orientation (acquired through EBSD).
Yasuzuka et al. (2009)	10 (10)	540-553, 815-825, 837-857	Reports centroid shift with change in pressure using 10 samples; centroid positions listed for all samples. As pressure increases, the wavenumber of the centroid increases.
McKeown et al. (2010)	1 (6)	All bands	Peak centroids reported for 1 of 6 samples. Forsterite single crystal examination and theoretical determination of modes.
Abdu et al. (2011)	3 (3)	815-825, 837-857	All 5 samples have reported peak centroids for DB1 and DB2. Raman, FTIR and Mössbauer spectroscopy applications to D'Orbigny meteorite.
Ishibashi et al. (2011)	15 (15)	815-825, 837-857, 881-883, 914-920, 950-966	Reports centroid shift with change in pressure using 15 samples; centroid positions listed for 15 samples. New model equation with accuracy for olivine inclusions within other minerals.
Weber et al. (2014)	5 (5)	815-825, 837-857	All 5 samples have reported peak centroids for DB1 and DB2. Band shifting from temperature changes are explored for Martian application.

*Number of samples with reported centroid positions and compositions and total samples in parentheses.

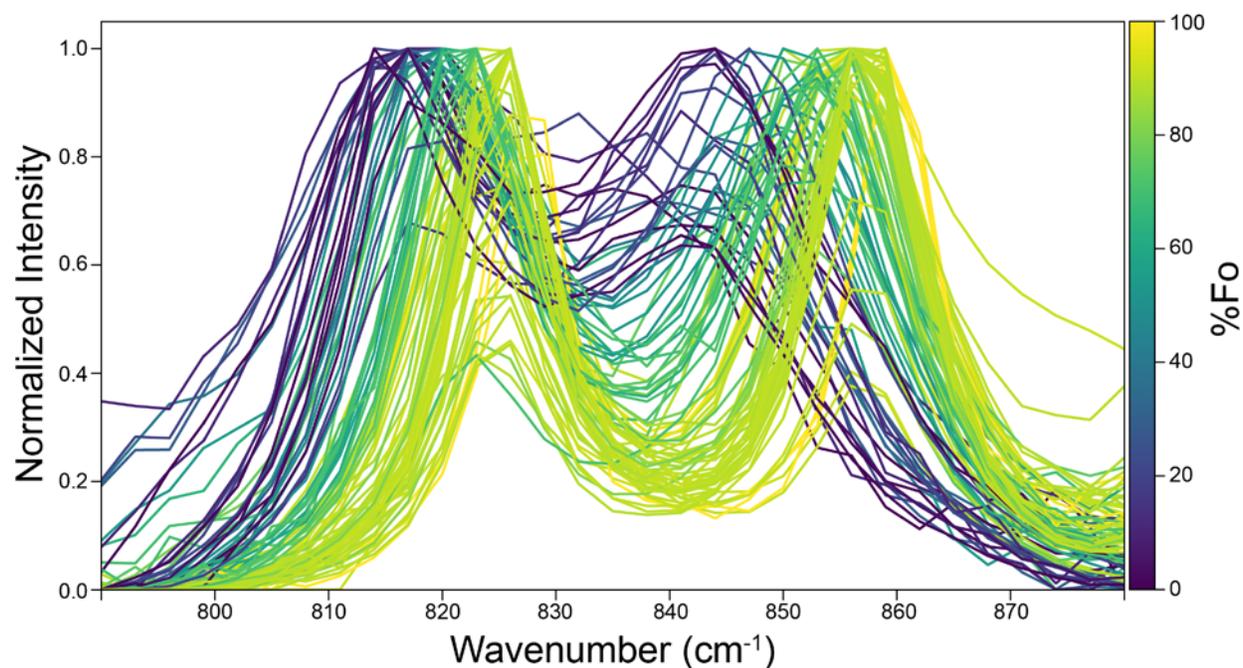


Figure 2.5. Raman spectra of olivine doublet (DB1 and DB2) of 93 samples acquired on Bruker's Senterra and BRAVO spectrometers. All spectra were baseline removed using Air-PLS and normalized to a maximum intensity of 1. Spectra are color-coded based on Fo content, where forsterite is represented with yellow, fayalite with purple, and intermediate compositions in between.

Raman spectra from these studies were generally not included within the models in this thesis because of the pressure and temperature operating conditions used.

Fluid inclusions within olivines (Pasteris and Wanamaker, 1988; Miura et al., 2011; Bolfan-Casanova et al., 2014), Cr-doped olivines (an active substance for solid-state lasers) (Demos and Alfano, 1995; Calistru et al., 1995a; Calistru et al., 1995b; Calistru et al., 1996; Golubovic et al., 2009), and shocked meteorite olivines (Heymann, 1990; Miyamoto, 1995; Nagy et al., 2008) have all been studied using Raman spectroscopy. While these studies have yielded many olivine Raman spectra, spectra from these studies were not useful for this thesis because of sample compositions or data acquisition information were not published.

This thesis evaluates how well the compositions of 93 well-characterized synthetic and naturally occurring olivine group minerals can be predicted using Raman spectroscopy. Using

the suite of samples from this thesis as well as utilizing published peak centroids, univariate methods for the prediction of Fe/Mg were tested. Additionally, the quantification of olivine composition through multivariate analyses was performed and tested with machine learning algorithms (PLS, lasso and LARS). Improving prediction accuracies is important for the exciting arrival of Raman instruments on the *ExoMars* and *Mars 2020* missions (Rull et al., 2014; Beegle et al., 2014; Maurice et al., 2015).

D. RATIONALE FOR THIS STUDY

In an Earth-bound laboratory, there are many other different ways to measure the Mg^{2+} and Fe^{2+} content of olivine, either as a single crystal (electron probe microanalysis or scanning electron microscopy) or a powder (x-ray diffraction, x-ray fluorescence). However, these techniques are cumbersome or impossible for remote analyses. Thus it is no accident that *in situ* use of Raman spectroscopy on Mars is planned for three different instruments in the next decade. The *ExoMars* mission carries an RLS instrument inside the analytical laboratory of the rover (Rull, 2014). It will use a 532 nm laser to probe powdered samples obtained by a drill. *Mars 2020*'s SHERLOC will utilize deep ultraviolet (DUV) resonance to scan habitable environments with both Raman and luminescence for organics and chemicals (Beegle et al., 2014). It is an arm-mounted instrument with a 248.6 nm DUV laser. Finally, Super-Cam on *Mars 2020* will probe Mars surface materials at distance up to 12 m (Maurice et al., 2015) using a pulsed doubled Nd:YAG laser (532 nm, 10Hz, 10 mJ/pulse) to produce Raman emission at energies up to $4,200 \text{ cm}^{-1}$ (Clegg et al., 2015) and footprints of 0.67 mrad (1.3 mm at 2 m distance) (Maurice et al., 2015). Although implementations differ, all these Raman instruments share the potential to identify surface minerals and organics and inform Martian geology and geochemistry. Their success depends on the availability of appropriate databases and software for phase

identification. This study provides models that can be used by these instruments to estimate olivine composition (%Fo).

Chapter 3

METHODS

A. SAMPLE PROVENANCE

Natural samples for this study came from collections of the Mineral Spectroscopy Lab at Mount Holyoke, the National Museum of Natural History (NMNH, Smithsonian), and from S.A. Morse (University of Massachusetts Amherst) (Morse, 2001). Roughly one-third of the natural samples came from previous studies on olivines from mantle xenoliths (McGuire et al., 1991; Dyar et al., 1989; Dyar et al., 1992;) or Fe³⁺-bearing samples studied by Schaefer (1983), Banfield et al. (1992) and Dyar et al. (1998). Another third of the samples were provided by S.A. Morse of the University of Massachusetts Amherst. These samples come from the Kiglapait layered mafic body, which is a large 1.3 Ga layered intrusion on the coast of Labrador, Canada (Morse, 1996; Morse, 2001). As the original melt crystalized, the Fe/Mg ratio of the remaining liquid changed, so a range of olivine compositions were produced. Lower Mg and higher Fe contents occurred successively higher within the intrusion. Finally, roughly one-third of the samples came from the NMNH.

The natural olivines studied in this thesis are listed in **Table 3.1** with species, locality, composition source, MB source and the Raman instrument indicated for each sample. This is the largest suite of naturally occurring olivine samples studied by Raman (or any other type of) spectroscopy.

Table 3.1. Natural Samples Studied.

Sample Name	Species	Locality	Composition source	MB source	Raman Instrument
Ba-1-61	Fo	Dish Hill, CA	UTK	[12]	Sen 532nm
Ba-1-74	Fo	Dish Hill, CA	UTK	[12]	Sen 532nm
Ba-2-1 WR1	Fo	Dish Hill, CA	Brown	[6]	Sen 532nm
Ba-2-1 WR2	Fo	Dish Hill, CA	Brown	[6]	Sen 532nm
Ba-2-1 WR3	Fo	Dish Hill, CA	Brown	[6]	Sen 532nm
Ba-2-1 WR4	Fo	Dish Hill, CA	Brown	[6]	Sen 532nm
Ba-2-1 D-1	Fo	Dish Hill, CA	[1]	[6]	Sen 532nm
Ci-1-183	Fo	Dish Hill, CA	[12]	[12]	Sen 532nm
Ci-1-25	Fo	Dish Hill, CA	[12]	[12]	Sen 532nm
DH101-B	Fo	Dish Hill, CA	Brown	[6]	Sen 532nm
DH101-C	Fo	Dish Hill, CA	Brown	[6]	Sen 532nm
DH101-D	Fo	Dish Hill, CA	Brown	[6]	Sen 532nm
DH101-E	Fo	Dish Hill, CA	Brown	[6]	Sen 532nm
Dyar 89-190	Fo	unknown	Brown	[12]	BR, Sen 532nm and 785nm
Dyar 89-12	Fo	unknown	Brown	[12]	BR, Sen 532nm and 785nm
Dyar 89-187	Fo	unknown	Brown	[12]	BR, Sen 532nm and 785nm
Dyar 89-194	Fo	unknown	Brown	[12]	BR, Sen 532nm and 785nm
Ep-1-13	Fo	Potrillo maar, NM	[12]	[7]	BR, Sen 532nm and 785nm
Ep-3-139-C	Fo	Kilbourne Hole, NM	Brown	[8]	Sen 532nm
Ep-3-139-D	Fo	Kilbourne Hole, NM	Brown	[8]	Sen 532nm
Ep-3-44	Fo	Kilbourne Hole, NM	UTK	[12]	Sen 532nm
Ep-3-46	Fo	Kilbourne Hole, NM	UTK	[12]	Sen 532nm
Ep-3-72	Fo	Kilbourne Hole, NM	UTK	[12]	Sen 532nm
Ep-3-7A	Fo	Kilbourne Hole, NM	Univ. Houston	[12]	Sen 532nm
KI-3003	Fa	Kiglapait Formation	Brown	[12]	Sen 532nm
KI-3373	Fa	Kiglapait Formation	Brown	[12]	Sen 532nm
NMNH 112085	Fa	Red Rock Ridge	UTK, Brown	[12]	Sen 532nm
NMNH 1210672	Fa	Germany Greifensteiner Kalk	UTK, Brown	[12]	Sen 532nm
NMNH 135841	Fa	Sweden Nykopig Tunaberg	Brown	[12]	Sen 532nm
NMNH 85539	Fa	unknown	UTK, Brown	[12]	Sen 532nm
Rockport	Fa	Rockport	Brown	[9]	Sen 532nm
Globe	Fo	Globe, AZ	[1]	[12]	BR, Sen 532nm and 785nm
H279-12	Fo	Harrat al Kishb, Saudi Arabia	[12]	[12]	Sen 532nm
H30-82-8	Fo	Harrat al Kishb, Saudi Arabia	UTK	[12]	Sen 532nm
H30-B1	Fo	Harrat al Kishb, Saudi Arabia	Brown	[10]	BR, Sen 532nm and 785nm
H30-B2	Fo	Harrat al Kishb, Saudi Arabia	[12]	[7]	BR, Sen 532nm and 785nm
H30-B3	Fo	Harrat al Kishb, Saudi Arabia	UTK	[12]	Sen 532nm
H30-B4	Fo	Harrat al Kishb, Saudi Arabia	UTK, Brown	[12]	Sen 532nm
H30-B5	Fo	Harrat al Kishb, Saudi Arabia	UTK	[12]	Sen 532nm
H312-1	Fo	Harrat Uwayrid, Saudi Arabia	[12]	[12]	Sen 532nm
H366-28	Fo	Harrat Hutaymah, Saudi Arabia	[12]	[12]	BR, Sen 532nm and 785nm
H366-30	Fo	Harrat Hutaymah, Saudi Arabia	[12]	[12]	Sen 532nm
NMNH 9140	Fa	Orange Co. NY	UTK, Brown	[12]	Sen 532nm
KBH-94-23-B	Fo	Kilbourne Hole, NM	UTK	[12]	Sen 532nm
KBH-94-23-E	Fo	Kilbourne Hole, NM	UTK	[12]	Sen 532nm

Forsterite (Fo), fayalite (Fa), Senterra spectrometer (Sen), and BRAVO spectrometer (BR). Sources are abbreviated as follows: [1] Byrne et al. (2015), [2] Floran et al. (1978), [3] McSween and Jarosewich (1983), [4] McCanta et al. (2008), [5] Mikouchi & Kurihara (2008), [6] McGuire et al. (1991), [7] Dyar et al. (1989), [8] Dyar et al. (1992), [9] Schaefer (1983), [10] McGuire et al. (1992), [11] Dyar (2003), [12] this study.

Table 3.1 (continued). **Natural Samples Studied.**

Sample Name	Species	Locality	Composition source	MB source	Raman Instrument
KBH-94-23-E	Fo	Kilbourne Hole, NM	UTK	[12]	Sen 532nm
KI-3005	Fa	Kiglapait Formation	Brown	[12]	Sen 532nm
KI-3289	Fa	Kiglapait Formation	Brown	[12]	Sen 532nm
KI-3362	Fo	Kiglapait Formation	Brown	[12]	Sen 532nm
KI-3648	Fo	Kiglapait Formation	UTK	[12]	Sen 532nm
KI-4030	Fo	Kiglapait Formation	Brown	[12]	Sen 532nm
Ki-5-16	Fo	Cima volcanic field, CA	[12]	[12]	Sen 532nm
Ki-5-235	Fo	Cima volcanic field, CA	UTK	[12]	Sen 532nm
Ki-5-35	Fo	Cima volcanic field, CA	UTK	[12]	Sen 532nm
Ki-5-62	Fo	Cima volcanic field, CA	UTK	[12]	BR, Sen 532nm and 785nm
Pakistan	Fo	Pakistan Sapatime Kohistan District	Brown	[12]	BR, Sen 532nm and 785nm
San Carlos AZ	Fo	San Carlos AZ	[1]	[12]	BR, Sen 532nm and 785nm
ALHA 77005	Fo	Mars	UTK	[11]	Sen 532nm
ALHA-77005-193	Fo	Mars	UTK	[11]	Sen 532nm
Chassigny USNM E24	Fo	Mars	[2]	[11]	Sen 532nm
EETA-79001 60B	Fo	Mars	[3]	[11]	Sen 532nm
EETA-79001-A	Fo	Mars	[3]	[11]	Sen 532nm
LAP-0484016	Fo	Mars	[4]	[12]	Sen 532nm
NWA2737	Fo	Mars	Brown	[12]	Sen 532nm
Y000097 86	Fo	Mars	[5]	[12]	Sen 532nm

Forsterite (Fo), fayalite (Fa), Senterra spectrometer (Sen), and BRAVO spectrometer (BR). Sources are abbreviated as follows: [1] Byrne et al. (2015), [2] Floran et al. (1978), [3] McSween and Jarosewich (1983), [4] McCanta et al. (2008), [5] Mikouchi & Kurihara (2008), [6] McGuire et al. (1991), [7] Dyar et al. (1989), [8] Dyar et al. (1992), [9] Schaefer (1983), [10] McGuire et al. (1992), [11] Dyar (2003), [12] this study.

Naturally occurring olivine typically has high Fo (Mg) content of roughly 89.5%. For a solid %Fo prediction model, wide representation of the Fo-Fa (Fe) continuum is needed. Therefore, synthetic samples were added to our collection of naturally formed olivines to represent %Fo from 0 to 100 (**Figure 3.1**).

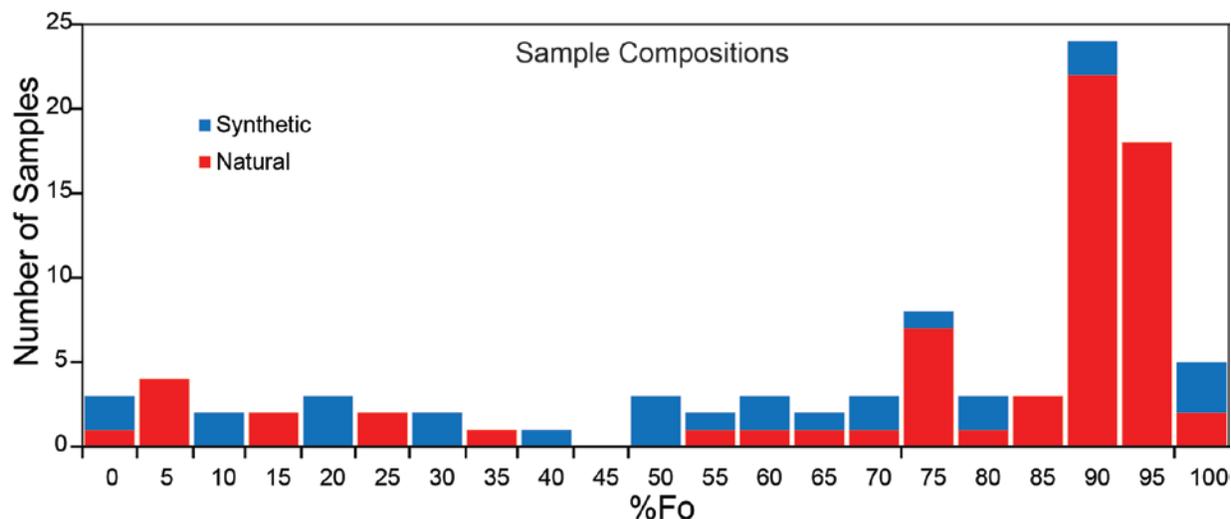


Figure 3.1. Histogram of 93 synthetic (blue) and natural (red) samples on the Fo-Fa series. Natural olivines typically form with a %Fo of ~89.5 resulting in an unbalanced distribution on the Fo-Fa series.

All synthetic samples were synthesized using the same methods by Donald Lindsley in his laboratory at SUNY Stony Brook. A mixture of hematite and silicon was ground for 1-2 hours under ethanol. An iron sponge was then added and grinding continued for less than 1 hour. The product was wrapped in silver foil and placed in a silicon glass capsule. One end was sealed and the middle of the capsule was drawn out into a capillary, leaving the sample by the sealed end. An Fe getter was placed next to the open end of the capsule (**Figure 3.2**). The capsule was put into a vertical tube furnace at ~800°C (the Fe getter remained at ~600°C) for 10-20 minutes. Next, the capsule was taken out of the furnace and sealed across the capillary (**Figure 3.2**). The completely sealed capsule section containing the sample was then placed in a horizontal tube furnace at ~920-940°C and cooked for 10 days (Sklute, 2006).

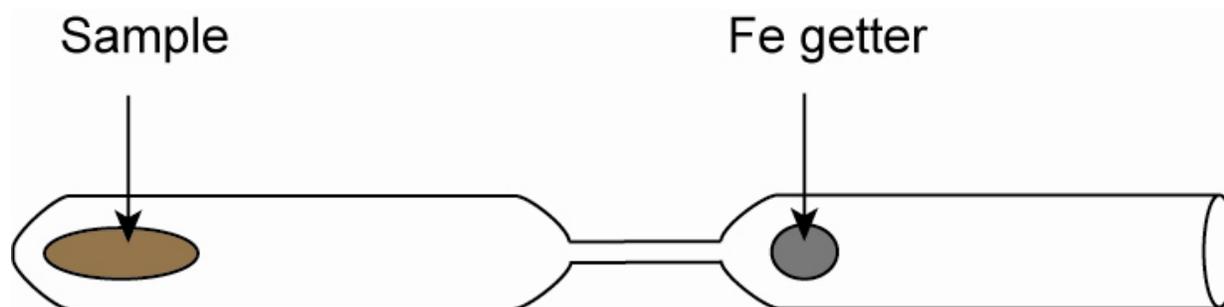


Figure 3.2. Schematic of synthesis tube (Sklute 2006).

Synthetic olivine samples of 10-30 mg were ground using an Fe-free diamonite mortar and pestle under acetone to prevent oxidation. Synthetic olivines used within this thesis are listed in **Table 3.2** with sample composition and the instrument on which each Raman spectrum was acquired.

Table 3.2. Synthetic Samples Studied.

Sample Name	Raman Instrument
Fa0Fo100a	Sen 532nm
Fa0Fo100b	Sen 532nm
Fa0Fo100c	Sen 532nm
Fa10.5 Fo89.5	BR, Sen 532nm and 785nm
Fa100Fo0a	BR, Sen 532nm and 785nm
Fa100Fo0b	BR, Sen 532nm and 785nm
Fa10Fo90 Menzies	BR, Sen 532nm and 785nm
Fa20Fo80a	BR, Sen 532nm and 785nm
Fa20Fo80b	BR, Sen 532nm and 785nm
Fa25 Fo75	BR, Sen 532nm and 785nm
Fa30Fo70a	BR, Sen 532nm and 785nm
Fa30Fo70b	BR, Sen 532nm and 785nm
Fa35 Fo65	BR, Sen 532nm and 785nm
Fa40Fo60a	Sen 532nm
Fa40Fo60b	Sen 532nm
Fa45 Fo55	Sen 532nm
Fa50Fo50a	Sen 532nm
Fa50Fo50b	Sen 532nm
Fa50Fo50c	Sen 532nm
Fa60Fo40	BR, Sen 532nm and 785nm
Fa70Fo30a	BR, Sen 532nm and 785nm
Fa70Fo30b	Sen 532nm
Fa80Fo20a	BR, Sen 532nm and 785nm
Fa80Fo20b	Sen 532nm
Fa80Fo20c	Sen 532nm
Fa90Fo10a	BR, Sen 532nm and 785nm
Fa90Fo10b	Sen 532nm

Forsterite (Fo), fayalite (Fa), Senterra spectrometer (Sen), and BRAVO spectrometer (BR). All samples described in Sklute (2006).

The 93 olivine samples examined were either a single crystal or in powdered form. To produce a powdered sample, first each sample was visually inspected and handpicked for purity. Then each grain was treated using oxalic acid (2 tsp. in 2 gal. of water) for 1 hour to remove surface weathering, followed by three cycles of washing and rinsing with clean water. As needed, samples were then crushed in a tungsten shatterbox or ground by hand in a diamonite mortar. We chose to study both single crystals and powders because crystal orientation affects the Raman spectrum of a given sample (Price et al., 1987) and we sought to evaluate the magnitude of the difference. When the quantity of the sample was too small to produce sufficient powder, then only a single crystal was analyzed.

B. SAMPLE CHARACTERIZATION

Many samples have published compositions (**Table 3.1**) and Mössbauer work had already been done to determine Fe^{3+} contents. For samples with unknown compositions, electron microprobe analyses of 10 spots on various grains of each sample were acquired at Brown University using standard operation conditions. The %Fo was calculated for each sample by normalizing the contents to contain only Mg and Fe as commonly done with the formula $\% \text{Fo} = (100 \times \text{Mg}) / (\text{Mg} + \text{Fe})$. Here, %Fo represents the total Fe contents ($\Sigma \text{Fe}^{2+} + \text{Fe}^{3+}$). All samples were analyzed by Mössbauer spectroscopy using standard methods (Sklute, 2006) and the $\% \text{Fe}^{3+}$ in most samples was zero. Samples with such impurities were eliminated from analyses as will be discussed in further detail.

The process for evaluating Fe^{3+} through Mössbauer spectroscopy is as follows. Rh was used on a WEB Research Co. model W100 spectrometer equipped with a Janus closed-cycle He refrigerator. Run times ranged from 2-12 hours; results were calibrated against α -Fe foil. Typical count rates were between 500,000 and 900,000 non-resonant counts/hour.

Paramagnetic spectra of antiferromagnetic minerals (unsplit spectra) were fit with Lorentzian line shapes using the method of Wivel and Mørup (1981). The program used was Mexfieldd, a component of a suite of programs created by Eddy De Grave and Toon van Alboom (Gent, Belgium). Mexfieldd uses Lorentzian line shapes to fit doublets with a fixed area ratio of 1:1 for the peaks. It solves the full Hamiltonian to determine single quadrupole splitting values (this is distinguished from other programs where distributions are found for one or more parameters). Other variables are isomer shift and width. Best fits are determined by minimizing the chi squared (χ^2) value. See Grant (1995) for a full description of how this is done. Comparison fits for some of our spectra were performed using Disd3e_dd, a program that uses velocity approximations instead of solving full Hamiltonians to obtain values for isomer shift and

quadrupole splitting. It searches for a distribution of quadrupole splitting values, rather than the single value sought in Mexdisdd. Quadrupole splitting distributions provide a non-Lorentzian lineshape (the lineshape is a sum of Voigt lines) that has proved more correct for fitting certain types of Mössbauer spectra (Rancourt 1994). Magnetically split spectra were fit using Mexdisdd, which solves the full Hamiltonian to obtain values for quadrupole splitting, isomer shift, and magnetic field, but provides a distribution of values for the magnetic field parameter, similar to the quadrupole splitting distributions mentioned above.

C. RAMAN MEASUREMENTS

Spectra were acquired on two Bruker instruments: the BRAVO dual laser system and the Senterra 532 or 785 nm laser. The 532 and 785 nm lasers within the Senterra are independent whereas the BRAVO dual laser system consists of two lasers that operate simultaneously. Handheld samples can be placed adjacent to the aperture of the BRAVO for acquisition of a spectrum. Although this method limits the operator's control, the large sample spot (area of the sample hit by the incoming laser beam) allows for multiple grains to be analyzed collectively. This can be helpful for analyses of mixtures of several minerals (Berlanga et al., 2017). Additionally, the BRAVO spectrometer is handheld, weighing only several pounds, making it an ideal instrument for *in situ* analyses. The Senterra differs from the BRAVO in its overall size and sampling aperture. This instrument has a microscope attachment resulting in a smaller sampling spot.

The quantity of each olivine sample in this thesis varied drastically, resulting in the examination of both powdered and single crystal samples. Single crystals could not be analyzed on the BRAVO because of its large aperture. The twenty-five samples run on both spectrometers allows for a comparison to be made between these micro and macro systems. For example,

examining a powder with a large aperture spectrometer (i.e., BRAVO) multiple crystal orientations can be observed within the sample. Therefore, this spectrum will encompass all the expected bands for the given material because all modes are accounted for. This demonstrates the advantage of acquiring spectra on the BRAVO spectrometer even if the sample is not a mixture of multiple minerals.

Single crystals and powders were analyzed on the Senterra using 10 mW laser power for two sample scans and integrated for 10s. All samples analyzed on the Senterra were analyzed through a 20x objective. Several resolution settings are available on the Senterra. The highest resolution available of $0.5 \text{ cm}^{-1}/\text{channel}$ was utilized. The powdered olivine samples were also analyzed on the BRAVO with three sample scans and an integration time of 10s. The BRAVO spectrometer has a $2.0 \text{ cm}^{-1}/\text{channel}$ spectral resolution.

D. DATA ANALYSIS

The Raman spectra of 93 olivine samples were utilized to produce several models within this thesis. Univariate models rely on a peak centroid value for every sample considered. Therefore, some univariate peak centroid models have fewer samples due to anomalies within the spectra. For example, in rare cases the DB2 ($837\text{-}857 \text{ cm}^{-1}$) olivine peak may not be present, yet the DB1 ($815\text{-}825 \text{ cm}^{-1}$) peak can still be distinguished. This is another difficulty associated with univariate peak centroid models that is not present with multivariate analyses. Average compositional data of all 93 olivines were evaluated through EMPA are reported in **Table 3.3**. Using the wt.% oxides from **Table 3.3**, olivine compositions expressed as formula units and calculated %Fo are listed in **Table 3.4**.

When a portion of the Raman signal is not necessary to the features of interest, a baseline or continuum removal can be a helpful pre-processing step. Manual baseline removal is unprincipled and may produce variable results. Therefore, all spectra were baseline removed using AirPLS with smoothness set to 100 and normalized to account for arbitrary intensity differences between the two spectrometers. The DB1 and DB2 bands of the doublet were peak fitted for each spectrum using a Gaussian and Lorentzian method. Additionally, the multivariate models used included partial least-squares (PLS), least absolute shrinkage and selection operator (lasso), or least angle regression (LARS). Pre-processing and predictions were produced using tools on the superman website nemo.umass.cs.edu:54321 (Carey et al., 2017).

Multivariate analyses provide an alternative method for Fo-Fa predictions. The partial least squares (PLS) method used here regresses one response variable (%Fo) against multiple explanatory variables (intensity at each channel of the spectra). PLS predictions utilize every channel of the spectral range, assigning coefficients to every single channel. Because PLS

Table 3.3. Average compositions of natural samples by EMPA. All olivine samples were evaluated for ferric content using Mössbauer spectroscopy.

Sample Name	SiO ₂	Al ₂ O ₃	TiO ₂	FeO	Fe ₂ O ₃	MgO	MnO	Cr ₂ O ₃	CaO	Na ₂ O	NiO	3+/tot	FeO
Ba-1-61	40.7	0.0	0.0	10.0	0.3	48.9	0.1	0.0	0.1	0.0	0.4	2.2	10.2
Ba-1-74	40.6	0.0	0.0	10.6	0.0	48.6	0.2	0.0	0.1	0.0	0.4	0.0	10.6
Ba-2-1 WR1	39.9	0.0	0.0	12.9	0.2	45.8	0.3	0.0	0.1	0.1	0.3	1.1	13.0
Ba-2-1 WR2	40.5	0.0	0.0	11.2	0.4	47.5	0.2	0.0	0.1	0.0	0.4	2.8	11.5
Ba-2-1 WR3	40.6	0.0	0.0	9.5	0.4	48.3	0.2	0.0	0.1	0.0	0.4	3.8	9.9
Ba-2-1 WR4	40.5	0.0	0.0	10.0	0.1	48.5	0.2	0.0	0.1	0.0	0.4	1.3	10.1
Ba-2-1 D-1	39.8	0.0	0.0	10.2	0.4	48.8	0.2	0.2	0.0	0.0	0.2	3.2	10.5
Ci-1-183	40.1	0.0	0.0	11.9	0.0	47.7	0.2	0.1	0.0	0.0	0.3	0.0	11.9
Ci-1-25	40.6	0.0	0.0	8.3	0.0	50.0	0.1	0.1	0.0	0.0	0.4	0.0	8.3
DH101-B	40.7	0.0	0.0	10.5	0.5	47.6	0.2	0.0	0.1	0.0	0.4	3.9	10.9
DH101-C	40.9	0.0	0.0	8.1	0.4	49.1	0.1	0.0	0.1	0.0	0.4	4.7	8.4
DH101-D	40.7	0.0	0.0	8.3	0.4	49.4	0.1	0.0	0.1	0.0	0.4	4.5	8.7
DH101-E	41.8	0.0	0.0	8.1	0.6	49.0	0.2	0.0	0.1	0.0	0.4	6.0	8.6
Dyar 89-190	38.4	0.0	0.0	22.5	0.0	39.0	0.3	0.0	0.2	0.0	n.a.	0.0	22.5
Dyar 89-12	40.7	0.0	0.0	8.8	0.2	49.5	0.1	0.0	0.1	0.0	n.a.	2.0	9.0
Dyar 89-187	41.7	0.0	0.0	4.2	0.1	53.4	0.2	0.0	0.0	0.0	n.a.	2.0	4.3
Dyar 89-194	41.3	0.0	0.0	7.6	0.0	50.6	0.2	0.0	0.0	0.0	n.a.	0.0	7.6
Ep-1-13	40.5	0.0	0.0	10.3	0.0	47.9	0.1	0.1	0.0	0.0	0.4	0.0	10.3
Ep-3-139-C	40.4	0.0	0.0	9.5	0.0	48.7	0.2	0.0	0.1	0.0	0.0	0.0	9.5
Ep-3-139-D	41.2	0.0	0.0	9.0	0.0	50.1	0.2	0.0	0.0	0.0	0.0	0.0	9.0
Ep-3-44	40.7	0.0	n.a.	10.4	0.0	48.7	0.1	n.a.	0.1	n.a.	0.4	0.0	10.4
Ep-3-46	40.8	0.0	n.a.	9.6	0.0	49.4	0.1	n.a.	0.1	n.a.	0.4	0.0	9.6
Ep-3-72	40.7	0.0	n.a.	9.9	0.0	49.3	0.1	n.a.	0.1	n.a.	0.4	0.0	9.9
Ep-3-7A	38.9	0.0	0.0	16.9	0.0	40.9	0.2	0.2	0.0	0.0	0.2	0.0	16.9
KI-3003	30.7	0.0	0.0	62.3	0.7	5.9	1.4	0.0	0.2	0.0	0.0	1.0	62.9
KI-3373	32.6	0.0	0.0	52.2	0.0	15.3	0.9	0.0	0.1	0.0	0.0	0.0	52.2
NMNH 112085	29.6	0.0	0.0	57.4	2.0	1.2	9.4	0.0	0.2	0.0	0.0	3.0	59.2
NMNH 1210672	29.1	0.2	0.0	69.6	0.0	0.4	1.0	0.0	0.1	0.0	0.0	0.0	69.6
NMNH 135841	29.9	0.0	0.0	62.7	0.0	1.8	5.0	0.0	0.0	0.0	0.0	0.0	62.7
NMNH 85539	29.1	0.0	0.0	64.5	0.0	0.2	5.1	0.0	0.1	0.0	0.0	0.0	64.5
Rockport	28.9	0.0	0.0	67.7	0.0	0.1	2.7	0.0	0.0	0.0	0.0	0.0	67.7
Globe	41.2	0.0	0.0	8.7	0.0	50.4	0.1	0.0	0.1	0.0	0.3	0.0	8.7
H279-12	40.4	0.0	0.0	9.2	0.0	49.6	0.1	0.0	0.0	0.0	0.3	0.0	9.2
H30-82-8	39.2	0.1	n.a.	17.6	0.0	43.3	0.2	n.a.	0.1	n.a.	0.3	0.0	17.6
H30-B1	41.2	0.0	0.0	8.7	0.3	50.1	0.1	0.0	0.0	0.0	0.0	2.5	9.0
H30-B2	40.9	0.0	0.0	8.2	0.0	49.9	0.1	0.0	0.0	0.0	0.4	0.0	8.2
H30-B3	40.9	0.0	0.0	8.7	0.5	49.9	0.1	0.0	0.0	0.0	0.4	5.1	9.2

Table 3.3 (continued). Average compositions of natural samples by EMPA. All olivine samples were evaluated for ferric content using Mössbauer spectroscopy.

Sample Name	SiO ₂	Al ₂ O ₃	TiO ₂	FeO	Fe ₂ O ₃	MgO	MnO	Cr ₂ O ₃	CaO	Na ₂ O	NiO	3+/tot	FeO
H30-B4	41.0	0.0	0.0	8.4	0.2	50.0	0.1	0.0	0.0	0.0	0.0	1.6	8.5
H30-B5	40.8	0.0	0.0	10.2	0.0	49.2	0.1	0.0	0.1	0.0	0.4	0.0	10.2
H312-1	40.0	0.0	0.0	9.1	0.0	51.3	0.1	0.0	0.0	2.0	0.5	0.0	9.1
H366-28	40.8	0.0	0.0	10.0	0.4	49.0	0.1	0.1	0.0	0.0	0.4	3.1	10.3
H366-30	40.4	0.0	0.0	10.9	0.0	48.4	0.1	0.1	0.0	0.0	0.4	0.0	10.9
NMNH 9140	31.1	0.0	0.0	56.5	1.3	8.4	3.6	0.0	0.1	0.0	0.0	2.0	57.6
KBH-94-23-B	39.4	0.0	n.a.	17.4	0.0	43.6	0.2	n.a.	0.1	n.a.	0.2	0.0	17.4
KBH-94-23-E	39.9	0.0	n.a.	13.7	0.0	46.1	0.2	n.a.	0.1	n.a.	0.3	0.0	13.7
KI-3005	30.3	0.0	0.0	60.6	2.8	4.2	1.5	0.0	0.1	0.0	0.0	4.0	63.1
KI-3289	31.6	0.0	0.0	56.8	0.0	9.7	1.2	0.0	0.2	0.0	0.0	0.0	56.8
KI-3362	36.0	0.0	0.0	35.0	0.0	28.6	0.5	0.0	0.0	0.0	0.1	0.0	35.0
KI-3648	37.8	0.0	0.0	26.2	0.0	36.5	0.3	0.0	0.0	0.0	0.1	0.0	26.2
KI-4030	34.6	0.0	0.0	40.2	0.0	24.3	0.5	0.0	0.0	0.0	0.0	0.0	40.2
Ki-5-16	40.9	n.a.	n.a.	8.2	0.0	50.0	0.1	0.1	n.a.	n.a.	0.4	0.0	8.2
Ki-5-235	40.5	0.0	n.a.	10.3	0.0	48.6	0.1	n.a.	0.1	n.a.	0.4	0.0	10.3
Ki-5-35	40.5	0.0	n.a.	10.5	0.0	48.6	0.2	n.a.	0.1	n.a.	0.4	0.0	10.5
Ki-5-62	42.4	0.0	0.0	2.0	0.0	55.5	0.1	0.1	n.a.	0.0	0.3	0.0	2.0
Pakistan Kohistan District	41.4	0.0	0.0	8.0	0.1	50.5	0.2	0.0	0.0	0.0	0.4	1.3	8.1
San Carlos AZ	41.1	0.0	0.0	9.0	0.3	50.0	0.1	0.0	0.1	0.0	0.4	3.0	9.2
ALHA 77005	37.3	0.0	0.0	25.6	0.7	35.7	0.5	0.0	0.2	0.0	0.1	2.3	26.2
ALHA-77005-193	37.3	0.0	0.0	25.6	0.7	35.7	0.5	0.0	0.2	0.0	0.1	2.3	26.2
Chassigny USNM E24	37.1	0.1	0.0	27.6	0.0	34.0	0.5	0.0	0.2	n.a.	n.a.	0.0	27.6
EETA-79001 60B	38.3	0.0	0.0	22.1	2.1	36.8	0.6	0.1	0.3	0.0	n.a.	7.9	24.0
EETA-79001-A	38.3	0.0	0.0	22.1	2.1	36.8	0.6	0.1	0.3	0.0	n.a.	7.9	24.0
LAP-0484016	36.0	0.0	0.0	31.3	1.8	29.6	0.4	0.0	0.0	0.0	0.4	5.0	33.0
NWA2737	37.4	0.0	0.0	25.5	0.6	35.9	0.5	0.0	0.0	0.0	0.0	2.0	26.0
Y000097 86	38.0	0.0	0.0	25.9	0.0	35.7	0.6	0.0	0.2	n.a.	0.0	n.a.	25.9

Table 3.4. Olivine compositions expressed as formula units, calculated from wt.% oxides (Table 3.4).

Sample Name	Si	Al	Ti	Fe ²⁺	Fe ³⁺	Mg	Mn	Cr	Ca	Na	Ni	Fo ²⁺	Fa ²⁺	Fo ²⁺³⁺	Fa ²⁺³⁺	Fo vs. MgFe ²⁺³⁺	Fe ²⁺ vs. MgFe ²⁺³⁺	Fe ³⁺ vs. MgFe ²⁺³⁺
Ba-1-61	1	0	0	0.2	0	1.8	0	0	0	0	0	90	10	89	11	89	10	0
Ba-1-74	1	0	0	0.22	0	1.8	0	0	0	0	0	89	11	89	11	89	11	0
Ba-2-1 WR1	1	0	0	0.27	0	1.7	0	0	0	0	0	86	14	86	14	86	14	0
Ba-2-1 WR2	1	0	0	0.23	0.01	1.7	0	0	0	0	0	88	12	88	12	88	12	0
Ba-2-1 WR3	1	0	0	0.2	0.01	1.8	0	0	0	0	0	90	10	90	10	90	10	0
Ba-2-1 WR4	1	0	0	0.21	0	1.8	0	0	0	0	0	90	10	90	10	90	10	0
Ba-2-1 D-1	1	0	0	0.21	0.01	1.8	0	0	0	0	0	90	10	89	11	89	10	0
Ci-1-183	1	0	0	0.25	0	1.8	0	0	0	0	0	88	12	88	12	88	12	0
Ci-1-25	1	0	0	0.17	0	1.8	0	0	0	0	0	91	9	91	9	91	9	0
DH101-B	1	0	0	0.22	0.01	1.8	0	0	0	0	0	89	11	89	11	89	11	0
DH101-C	1	0	0	0.17	0.01	1.8	0	0	0	0	0	92	8	91	9	91	8	0
DH101-D	1	0	0	0.17	0.01	1.8	0	0	0	0	0	91	9	91	9	91	9	0
DH101-E	1	0	0	0.16	0.01	1.8	0	0	0	0	0	92	8	91	9	91	8	1
Dyar 89-190	1	0	0	0.49	0	1.5	0	0	0	0	0	76	24	76	24	76	24	0
Dyar 89-12	1	0	0	0.18	0	1.8	0	0	0	0	0	91	9	91	9	91	9	0
Dyar 89-187	1	0	0	0.08	0	1.9	0	0	0	0	0	96	4	96	4	96	4	0
Dyar 89-194	1	0	0	0.15	0	1.8	0	0	0	0	0	92	8	92	8	92	8	0
Ep-1-13	1	0	0	0.21	0	1.8	0	0	0	0	0	89	11	89	11	89	11	0
Ep-3-139-C	1	0	0	0.2	0	1.8	0	0	0	0	0	90	10	90	10	90	10	0
Ep-3-139-D	1	0	0	0.18	0	1.8	0	0	0	0	0	91	9	91	9	91	9	0
Ep-3-44	1	0	0	0.21	0	1.8	0	0	0	0	0	89	11	89	11	89	11	0
Ep-3-46	1	0	0	0.2	0	1.8	0	0	0	0	0	90	10	90	10	90	10	0
Ep-3-72	1	0	0	0.2	0	1.8	0	0	0	0	0	90	10	90	10	90	10	0
Ep-3-7A	1	0	0	0.37	0	1.6	0	0	0	0	0	81	19	81	19	81	19	0
KI-3003	1	0	0	1.67	0.02	0.3	0	0	0	0	0	14	86	14	86	14	85	1
KI-3373	1	0	0	1.32	0	0.7	0	0	0	0	0	34	66	34	66	34	66	0
NMNH 112085	1	0	0	1.61	0.05	0.1	0.3	0	0	0	0	4	96	3	97	3	94	3
NMNH 1210672	1	0	0	1.97	0	0	0	0	0	0	0	1	99	1	99	1	99	0
NMNH 135841	1	0	0	1.76	0	0.1	0.1	0	0	0	0	5	95	5	95	5	95	0
NMNH 85539	1	0	0	1.85	0	0	0.2	0	0	0	0	1	99	1	99	1	99	0
Rockport	1	0	0	1.94	0	0	0.1	0	0	0	0	0	100	0	100	0	100	0
Globe	1	0	0	0.18	0	1.8	0	0	0	0	0	91	9	91	9	91	9	0
H279-12	1	0	0	0.19	0	1.8	0	0	0	0	0	91	9	91	9	91	9	0
H30-82-8	1	0	0	0.37	0	1.6	0	0	0	0	0	81	19	81	19	81	19	0
H30-B1	1	0	0	0.18	0	1.8	0	0	0	0	0	91	9	91	9	91	9	0
H30-B2	1	0	0	0.17	0	1.8	0	0	0	0	0	92	8	92	8	92	8	0
H30-B3	1	0	0	0.18	0.01	1.8	0	0	0	0	0	91	9	91	9	91	9	0

Table 3.4 (continued). Olivine compositions expressed as formula units, calculated from wt.% oxides (Table 3.4).

Sample Name	Si	Al	Ti	Fe ²⁺	Fe ³⁺	Mg	Mn	Cr	Ca	Na	Ni	Fo ²⁺	Fa ²⁺	Fo ²⁺³⁺	Fa ²⁺³⁺	Fo vs. MgFe ²⁺³⁺	Fe ²⁺ vs. MgFe ²⁺³⁺	Fe ³⁺ vs. MgFe ²⁺³⁺
H30-B4	1	0	0	0.17	0	1.8	0	0	0	0	0	91	9	91	9	91	9	0
H30-B5	1	0	0	0.21	0	1.8	0	0	0	0	0	90	10	90	10	90	10	0
H312-1	1	0	0	0.18	0	1.8	0	0	0	0.1	0	91	9	91	9	91	9	0
H366-28	1	0	0	0.2	0.01	1.8	0	0	0	0	0	90	10	89	11	89	10	0
H366-30	1	0	0	0.22	0	1.8	0	0	0	0	0	89	11	89	11	89	11	0
NMNH 9140 Orange Co. NY	1	0	0	1.49	0.03	0.4	0.1	0	0	0	0	21	79	21	79	21	78	2
KBH-94-23-B	1	0	0	0.37	0	1.6	0	0	0	0	0	82	18	82	18	82	18	0
KBH-94-23-E	1	0	0	0.28	0	1.7	0	0	0	0	0	86	14	86	14	86	14	0
KI-3005	1	0	0	1.66	0.07	0.2	0	0	0	0	0	11	89	11	89	11	86	4
KI-3289	1	0	0	1.5	0	0.5	0	0	0	0	0	23	77	23	77	23	77	0
KI-3362	1	0	0	0.81	0	1.2	0	0	0	0	0	59	41	59	41	59	41	0
KI-3648	1	0	0	0.58	0	1.4	0	0	0	0	0	71	29	71	29	71	29	0
KI-4030	1	0	0	0.96	0	1	0	0	0	0	0	52	48	52	48	52	48	0
Ki-5-16	1	0	0	0.17	0	1.8	0	0	0	0	0	92	8	92	8	92	8	0
Ki-5-235	1	0	0	0.21	0	1.8	0	0	0	0	0	89	11	89	11	89	11	0
Ki-5-35	1	0	0	0.22	0	1.8	0	0	0	0	0	89	11	89	11	89	11	0
Ki-5-62	1	0	0	0.04	0	2	0	0	0	0	0	98	2	98	2	98	2	0
Pakistan Kohistan District	1	0	0	0.16	0	1.8	0	0	0	0	0	92	8	92	8	92	8	0
San Carlos AZ	1	0	0	0.18	0.01	1.8	0	0	0	0	0	91	9	91	9	91	9	0
ALHA 77005	1	0	0	0.57	0.01	1.4	0	0	0	0	0	71	29	71	29	71	28	1
ALHA-77005-193	1	0	0	0.57	0.01	1.4	0	0	0	0	0	71	29	71	29	71	28	1
Chassigny USNM E24	1	0	0	0.62	0	1.4	0	0	0	0	0	69	31	69	31	69	31	0
EETA-79001 60B	1	0	0	0.48	0.04	1.4	0	0	0	0	0	75	25	73	27	73	25	2
EETA-79001-A	1	0	0	0.48	0.04	1.4	0	0	0	0	0	75	25	73	27	73	25	2
LAP-0484016	1	0	0	0.72	0.04	1.2	0	0	0	0	0	63	37	62	38	62	37	2
NWA2737	1	0	0	0.56	0.01	1.4	0	0	0	0	0	72	28	71	29	71	28	1
Y000097 86	1	0	0	0.57	0	1.4	0	0	0	0	0	71	29	71	29	71	29	0

utilizes all available variables (channels) and eliminates multicollinearity (peaks whose intensities are dependent, as is the case for the doublet in the Raman spectra of olivine). This algorithm was created for the analysis of data with high collinear explanatory (p) variables, which are significantly greater in number compared to the observations (N). Therefore, $p \gg N$ (Butler and Denham, 2000). For the given predictions, the number of components was chosen at which the mean-square error (MSE) was smallest. PLS can predict multiple dimensional data sets and has been utilized for the specific application of spectroscopy (Wold et al., 1983). This thesis utilizes PLS2 (hereafter referred to as PLS) rather than alternative versions.

The least absolute shrinkage and selection operator (lasso) is an additional multivariate analysis method utilized within this study. The lasso is a continuous shrinkage, which allows for the production of coefficient values to be reduced even to as small as zero (Tibshirani, 1995). This shrinkage is in agreement with the shrinkage parameter t , by shrinking 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. Unlike PLS, lasso does not utilize every channel available. Therefore, PLS provides a stable model yet, lasso can be utilized when less data are available. For the lasso predictions within this thesis, alpha values were chosen at which the MSE was smallest.

The least angle regression (LARS) differs from PLS and lasso in that the algorithm begins with all coefficients assigned to zero. Next, steps are taken toward the most correlated predictor. When a new predictor relates to the residual more accurately, the equiangular direction between the two predictors is utilized. When a third predictor becomes favorable, the direction used is that of the least angle direction between the three predictors (Efron et al., 2004). Similar to PLS,

the number of channels examined with the LARS algorithm was assigned to produce the smallest MSE.

E. MODEL COMPARISONS

Whether a multivariate or univariate prediction is made, it is essential to evaluate the accuracy of the given prediction. The R^2 value and cross-validated root mean square error (RMSE-CV) of each prediction was reported within this thesis. Each models was also utilized to predict the %Fo of every sample within a different data set. This allows the evaluation of the accuracy of predicting “unseen data”. The errors for these secondary predictions are referred to as RMSE-test values within this thesis.

The R^2 value is a statistic that describes how accurately the data points within a given model matches the fit of the data set (difference between the predicted and true values). An R^2 value of 1.0 would indicate that every data point falls directly on the fit. More specifically, the R^2 value is the square of the Pearson product moment correlation coefficient.

RMSE is typically calculated by finding the difference between the predicted and actual values within the data set, squaring these new values, summing the squares and finally taking the square root. This method was utilized for calculating RMSE-CV. The “CV” of RMSE-CV refers to cross-validation, which means that the error was obtained using leave-one-out cross-validation. In other words, for a data set of 93 samples, one sample at a time was held out, and its composition was predicted using a regression equation developed from the other 92 samples. Next, the 93 predictions would be used to calculate RMSE, which represents the error bar on the predicted composition in units of %Fo. It is important to cross-validate the RMSE values, especially for small data sets, in order for the prediction not to be biased towards its own data. Cross-validation typically produces larger error than the standard RMSE, yet it tells the more

honest story of the accuracy of the prediction. The calculations of the RMSE-test values are simply performed like a standard RMSE, because the prediction has already been cross-validated within the first treatment of the data set.

Univariate predictions of %Fo have been made using Raman spectra (**Table 2.3**). These predictions have been made by modeling the peak centroids of the olivine doublet (DB1 and DB2). For example, **Figure 3.3** depicts %Fo second order models using the DB1 and DB2 peak centroids (Kuebler et al., 2006).

When examining fits within **Figure 3.3**, the model appears highly successful with large R^2 values. The R^2 values are listed for the DB1 and DB2 peak centroids as 0.918 and 0.981 respectively. Yet, when these data are used to calculate RMSE-CV, a different story emerges. The RMSE-CV of the **Figure 3.3** models are ± 4.33 and ± 4.57 %Fo units. And the inaccuracy of this model becomes even more apparent when it is used to predict a different data set. The RMSE-test values are ± 16.17 and ± 14.66 %Fo for the prediction of the combined BRAVO and Senterra (532 nm) data sets. Calculating the R^2 value, RMSE-CV and RMSE-test value for each model will assist in evaluating the accuracy of both univariate and multivariate analyses. Through these statistics, the best prediction method is recommended within this thesis.

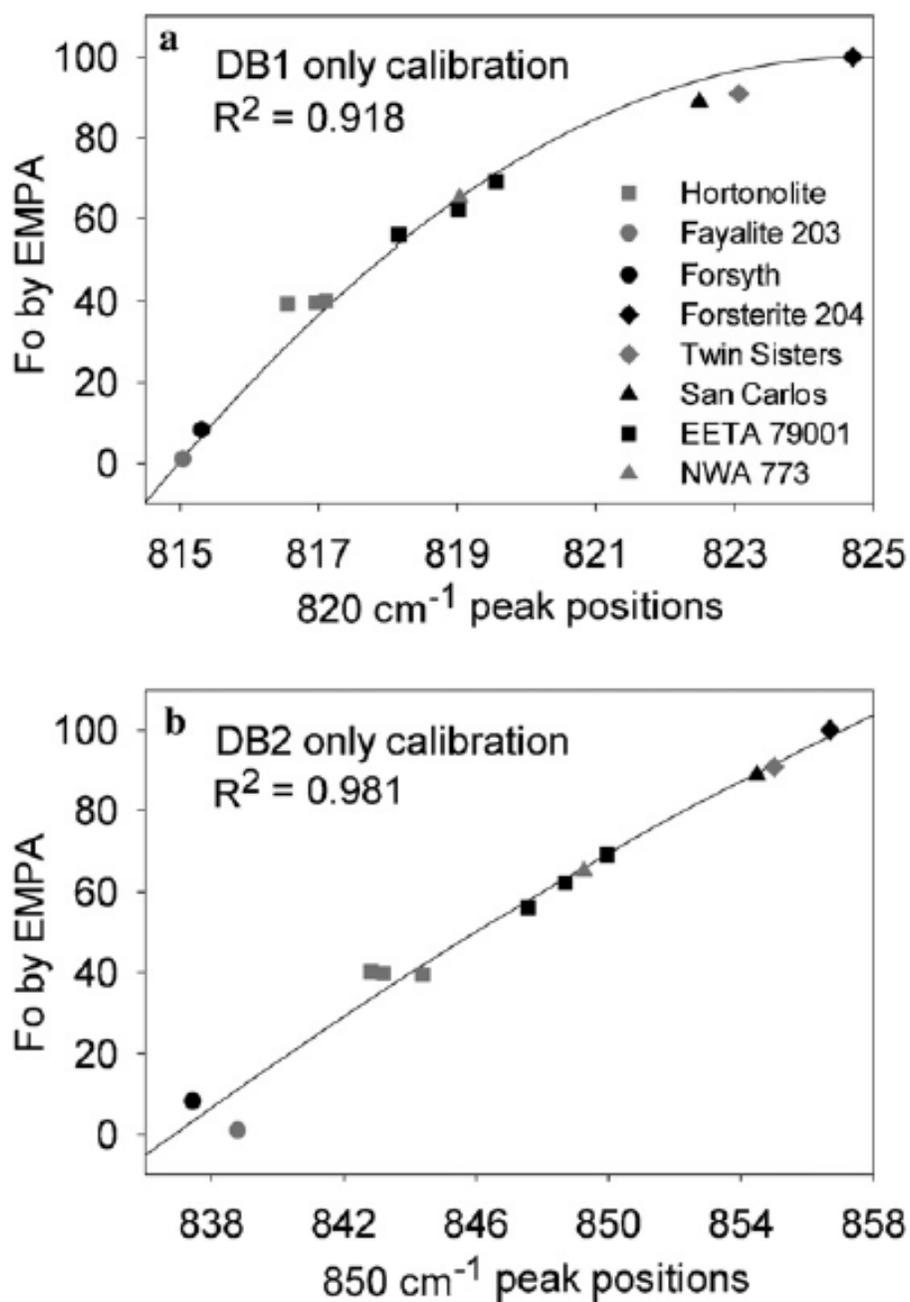


Figure 3.3. First calibration models for predicting %Fo using Raman spectroscopy. Here peak centroid positions and actual %Fo content of olivines are used to build a univariate second order fit. The R^2 values are reported yet, RMSE-CVs are not reported (Kuebler et al., 2006).

Chapter 4

RESULTS

The results of this thesis can be divided into two categories: univariate and multivariate. Univariate predictions refer to modeling of the relationship between %Fo and a single variable (peak centroid wavenumber position) using linear or polynomial regression. These predictions utilize the position of either the DB1 (815-825 cm^{-1}) or DB2 (837-857 cm^{-1}) bands of the olivine doublet. First-, second-, and third-degree polynomial fits were applied to the %Fo versus peak centroid graphs. This thesis compares the band shift univariate results with an innovative alternative method employing multivariate analysis. These techniques utilize many channels of a spectrum instead of a single peak centroid using a range of wavenumbers to predict the composition of the sample. There are many multivariate techniques suited for this purpose (Boucher et al., 2015), but for this thesis, predictions used partial least squares (PLS), the least absolute shrinkage and selection operator (lasso), and least angle regression (LARS) algorithms. Because multivariate analyses examine a section of spectra instead of a single data point, spectral information such as shape, intensity and secondary features are preserved and exploited in the regression, which ought to lead to improved prediction accuracy.

A. UNIVARIATE (TWO-PEAK) RESULTS

Numerous permutations of univariate analysis were used to predict %Fo from peak centroids. Plots of composition (%Fo or $(100 \times \text{Mg}) / (\text{Mg} + \text{Fe})$) versus DB1 or DB2 peak positions replicate the modeling of **Figure 3.3** (Kuebler et al., 2006). Spectra of the 93 well-characterized synthetic and naturally-occurring olivines acquired on either the BRAVO or Senterra 532 nm spectrometers (**Tables 3.1** and **3.2**) were considered for the univariate predictions. Additionally,

all known publicly accessible olivine Raman data were considered, mined from the many publications on this topic as described in Chapter 2.

The normalized height, centroid, full width at half maximum intensity (FWHM) and area of the DB1 and DB2 peaks of all samples studied for this thesis are given in **Tables 4.1** and **4.2**. Peak fitting was performed using both Gaussian and Lorentzian methods. All spectra were normalized over the entire wavenumber range prior to peak fitting.

Peaks were also fitted to olivine data from the RRUFF data set posted at <http://rruff.info/>, also using both Gaussian and Lorentzian methods. Centroid energies are reported in **Table 4.3** for those data, which were acquired on randomly-oriented single crystals using a 532 nm spectrometer. **Table 4.4** includes peak centroids of olivine spectra reported within surveyed literature as described in Chapter 2. The normalized height, FWHM and area of the bands are rarely reported in previous studies, and therefore are not listed in **Table 4.4**. Previous Raman studies of olivine that did not report peak centroids could not be included in this thesis, such as those in Wang et al. (2004), Gaisler and Kolesov (2007), and Mouri and Enami (2008). The Fo (Mg) versus Fa (Fe^{2+}) compositions of all RRUFF and literature samples were obtained from either <http://rruff.info/> or from the papers themselves.

Samples selected for use in this thesis, either for our own measurements or when taken from the literature, were specifically chosen to have limited amounts of other minor elements substituting into the octahedral sites; a criterion of <5 wt.% of those cations was applied, and samples with >5% of other cations were excluded from this work.

The peak centroids listed in **Tables 4.1- 4.4** demonstrate shifting of the peaks as Fo/Fa changes as seen in previous studies that used far fewer samples (Kuebler et al., 2006; Foster et al., 2007; Gaisler and Kolesov, 2007; Mouri and Enami, 2008; Yasuzuka et al., 2009; Ishibashi

et al., 2011). The details of these papers are summarized in **Table 2.3**. Linear-, second-, and third-order polynomial fits relating peak centroid position to composition were created for the BRAVO and Senterra data. The second order polynomial fits are shown in **Figures 4.1** and **4.2**. These models replicate univariate %Fo prediction methods developed by Kuebler et al. (2006) (**Figure 3.3**) but show the effect of using an order of magnitude more samples and varied instrumentation.

Table 4.1. Raman spectra characteristics of DB1 (peak at 815-825 cm^{-1}). Height, centroid, FWHM and area of DB1 band with Fo-Fa content based on EMPA. There is an observable trend that as Fo content increases the peak centroids increase in their wavenumber (cm^{-1}) value.

Sample ID	%Fo	DB1 Lorentzian				DB1 Gaussian			
		Height	Centroid	FWHM	Area	Height	Centroid	FWHM	Area
Fa100Fo0a	0	0.40	815.94±0.05	37.63±0.5	23.65±0.31	0.40	816.31±0.06	33.5±0.41	14.28±0.18
Fa100Fo0b	0	1.00	816.12±0.01	17.37±0.07	27.32±0.1	1.00	816.01±0.01	14.86±0.05	15.87±0.05
Rockport	0	1.00	817.21±0.01	17.26±0.07	27.14±0.11	1.00	817.22±0.01	15.12±0.06	16.13±0.06
NMNH 85539	1	1.00	814.5±0.01	16.4±0.05	25.75±0.07	1.00	814.45±0.01	15.63±0.06	16.55±0.06
NMNH 1210672	1	0.78	818.15±0.01	18.04±0.09	22.08±0.1	0.77	817.95±0.02	18.29±0.13	15.05±0.1
NMNH 112085	4	1.00	816.43±0.01	16.61±0.08	26.13±0.12	1.00	816.48±0.01	14.17±0.06	15.16±0.06
NMNH 135841	5	1.00	816.45±0.01	14.64±0.04	23.01±0.06	0.99	816.37±0.01	14.31±0.06	15.15±0.06
Fa90Fo10a	10	0.48	817.47±0.03	30.56±0.24	22.96±0.18	0.48	816.83±0.05	32.53±0.32	16.5±0.16
Fa90Fo10b	10	0.27	815.18±0.01	12.63±0.07	5.3±0.03	0.26	815.25±0.02	14.29±0.11	3.99±0.03
KI-3005	11	1.00	815.05±0.01	16.37±0.08	25.63±0.12	1.00	814.83±0.02	15.66±0.1	16.6±0.1
KI-3003	14	0.69	817.86±0.01	16.44±0.09	17.83±0.1	0.69	817.93±0.01	14.67±0.08	10.78±0.06
Fa80Fo20a	20	0.55	817.48±0.01	31.43±0.13	27.01±0.11	0.55	817.53±0.02	28.39±0.14	16.53±0.08
Fa80Fo20b	20	1.00	816.42±0.01	20.81±0.08	32.68±0.12	1.00	816.39±0.01	18.45±0.07	19.62±0.07
Fa80Fo20c	20	1.00	815.55±0.01	21.67±0.12	33.96±0.18	1.00	815.63±0.01	17.57±0.08	18.78±0.08
NMNH 9140	21	0.87	818.72±0.01	15.79±0.05	21.49±0.07	0.87	818.79±0.01	14.40±0.05	13.27±0.04
KI-3289	23	1.00	817.17±0.01	17.79±0.04	27.96±0.07	1.00	817.26±0.01	16.06±0.04	17.1±0.05
Fa70Fo30a	30	0.54	816.16±0.01	32.52±0.1	27.66±0.09	0.54	816.04±0.02	29.8±0.15	17.15±0.09
Fa70Fo30b	30	1.00	818.09±0.01	17.03±0.05	26.71±0.08	1.00	818.16±0.01	15.38±0.05	16.33±0.05
KI-3373	34	0.36	818.15±0.00	11.23±0.02	6.3±0.01	0.36	818.16±0	10.45±0.02	3.96±0.01
Fa60Fo40	40	0.65	817.02±0.02	22.76±0.14	23.23±0.14	0.65	816.96±0.03	21.08±0.16	14.56±0.11
Fa50Fo50a	50	1.00	819.93±0.00	16.82±0.04	26.4±0.06	1.00	819.94±0.01	15.3±0.04	16.25±0.05
Fa50Fo50b	50	1.00	818.89±0.01	17.59±0.04	27.63±0.07	1.00	818.93±0.01	15.92±0.05	16.93±0.05
Fa50Fo50c	50	1.00	819.09±0.01	18.97±0.06	29.67±0.09	0.99	819.03±0.01	17.15±0.06	18.16±0.06
KI-4030	52	1.00	819.63±0.00	10.74±0.02	16.91±0.03	1.00	819.6±0	9.85±0.01	10.49±0.01
Fa45 Fo55	55	0.40	817.38±0.02	21.78±0.13	13.51±0.08	0.40	817.11±0.03	20.41±0.16	8.58±0.07
KI-3362	59	0.98	820.91±0.01	13.56±0.03	20.83±0.05	0.97	820.98±0.01	13.16±0.03	13.62±0.03
Fa40Fo60a	60	1.00	821.64±0.00	10.34±0.01	16.23±0.02	1.00	821.6±0	9.64±0.02	10.22±0.02
Fa40Fo60b	60	1.00	820.88±0.01	8.7±0.04	13.93±0.06	1.00	820.98±0.01	8.17±0.02	8.84±0.03
SNC LAP-0484016	63	1.00	820.7±0.00	13.05±0.02	20.55±0.03	1.00	820.79±0.01	12.05±0.03	12.84±0.03
Fa35 Fo65	65	0.81	818.05±0.02	24.89±0.21	31.5±0.26	0.81	818.4±0.03	21.47±0.16	18.48±0.13
Chassigny USNM	69	1.00	821.48±0.00	8.84±0.01	13.83±0.02	0.99	821.44±0	8.43±0.02	8.88±0.02
Fa30Fo70a	70	0.43	819.15±0.01	22.54±0.11	15.38±0.08	0.43	819.48±0.02	23.03±0.17	10.58±0.08
Fa30Fo70b	70	0.49	820.41±0.02	25.53±0.2	19.76±0.15	0.49	820.42±0.02	22.3±0.14	11.69±0.08

Table 4.1 (continued). Raman spectra characteristics of DB1 (peak at 815-825 cm^{-1}). Height, centroid, FWHM and area of DB1 band with Fo-Fa content based on EMPA. There is an observable trend that as Fo content increases the peak centroids increase in their wavenumber (cm^{-1}) value.

Sample ID	%Fo	DB1 Lorentzian				DB1 Gaussian			
		Height	Centroid	FWHM	Area	Height	Centroid	FWHM	Area
KI-3648	71	0.40	822.58±0.004	12.15±0.03	7.66±0.02	0.40	822.61±0.01	11.18±0.03	4.77±0.01
SNC Y000097 86	71	1.00	821.6±0.00	14.81±0.03	23.25±0.04	1.00	821.58±0.01	13.59±0.04	14.43±0.04
SNC ALHA 77005	71	1.00	821.92±0.00	10.66±0.01	16.77±0.02	1.00	821.92±0	9.83±0.01	10.44±0.01
SNC ALHA-77005-193	71	1.00	822.01±0.00	14.25±0.03	22.45±0.04	1.00	822.03±0	12.95±0.02	13.79±0.02
SNC NWA2737	72	0.99	823.38±0.01	12.93±0.04	20.2±0.07	1.00	823.24±0.01	11.6±0.05	12.29±0.05
SNC EETA-79001 60B	75	1.00	822.41±0.00	11.78±0.01	18.52±0.01	1.00	822.4±0	10.83±0.02	11.5±0.02
SNC EETA-79001-A	75	0.93	821.41±0.00	10.62±0.02	15.53±0.02	0.93	821.41±0	9.85±0.01	9.72±0.01
Fa25 Fo75	75	0.44	820.75±0.01	26.78±0.1	18.53±0.07	0.44	820.73±0.01	23.13±0.11	10.86±0.05
Dyar 89-190	76	0.89	821.13±0.01	19.62±0.08	27.29±0.11	0.89	821.2±0.01	17.61±0.08	16.6±0.07
Fa20Fo80a	80	0.68	820.31±0.01	19.56±0.08	20.76±0.08	0.67	820.12±0.01	19.08±0.1	13.64±0.07
Fa20Fo80b	80	1.00	822.85±0.00	9.89±0.02	15.53±0.02	0.99	822.75±0.01	9.65±0.02	10.19±0.03
Ep-3-7A	81	0.83	824.05±0.00	11.11±0.02	14.44±0.02	0.82	823.98±0.01	10.68±0.03	9.35±0.02
H30-82-8	81	1.00	823.61±0.00	9.53±0.01	14.98±0.02	1.00	823.6±0	8.65±0.01	9.2±0.01
KBH-94-23-B	82	0.47	824.07±0.00	11.77±0.02	8.68±0.02	0.47	824.02±0.01	11.13±0.03	5.54±0.02
Ba-2-1 WR1	86	0.53	824.35±0.00	8.97±0.02	7.45±0.02	0.53	824.39±0	8.27±0.01	4.64±0.00412
KBH-94-23-E	86	0.99	824.36±0.00	10.91±0.01	17.01±0.02	0.99	824.28±0.01	10.36±0.02	10.9±0.02
Ba-2-1 WR2	88	1.00	824.52±0.00	9.53±0.01	14.99±0.02	1.00	824.54±0	8.8±0.01	9.36±0.01
Ci-1-183	88	0.55	825.41±0.00	8.86±0.03	7.66±0.02	0.55	825.41±0	8.54±0.02	4.95±0.01
Ba-1-74	89	0.89	824.88±0.00	7.93±0.01	11.1±0.02	0.89	824.94±0	7.51±0.00385	7.09±0.00346
DH101-B	89	1.00	824.62±0.00	9.53±0.01	15.02±0.02	1.00	824.63±0	8.94±0.01	9.5±0.01
Ep-1-13	89	0.39	822.98±0.01	17.09±0.11	10.5±0.07	0.39	822.98±0.01	16.01±0.11	6.63±0.04
Ep-3-44	89	1.00	825±0.00	9.91±0.01	15.59±0.01	1.00	824.98±0	9.11±0.01	9.69±0.01
H366-30	89	1.00	824.84±0.00	8.46±0.01	13.33±0.02	1.00	824.86±0	7.94±0.01	8.44±0.01
Ki-5-235	89	0.45	824.97±0.00	10.08±0.01	7.19±0.01	0.45	824.87±0.01	9.69±0.03	4.65±0.01
Ki-5-35	89	0.84	824.99±0.00	10.35±0.01	13.66±0.01	0.84	825.01±0	9.58±0.02	8.54±0.02
Ki-5-35	89	0.76	825.47±0.00	10.38±0.01	12.4±0.01	0.76	825.5±0	9.63±0.02	7.77±0.01
Ba-1-61	90	0.76	824.49±0.00	8.28±0.02	9.95±0.02	0.76	824.57±0	8.02±0.01	6.47±0.01
Ba-2-1 WR3	90	0.66	824.96±0.00	9.24±0.01	9.54±0.01	0.66	824.98±0	8.56±0.01	5.97±0.01
Ba-2-1 WR4	90	0.54	825.17±0.00	8.89±0.01	7.58±0.01	0.54	825.2±0	8.44±0.01	4.85±0.01
Ba-2-1 D-1	90	0.35	825.4±0.00	8.84±0.03	4.86±0.01	0.35	825.45±0	8.01±0.01	2.98±0.00389
Ep-3-139-C	90	0.98	825.52±0.00	10.14±0.01	15.67±0.01	0.98	825.54±0	9.4±0.01	9.82±0.01
Ep-3-46	90	0.51	825±0.00	8.84±0.01	7.11±0.01	0.51	825.02±0	8.43±0.01	4.57±0.00493
Ep-3-72	90	0.95	825.09±0.00	8.95±0.01	13.31±0.01	0.94	825.06±0	8.43±0.01	8.45±0.01

Table 4.1 (continued). Raman spectra characteristics of DB1 (peak at 815-825 cm^{-1}). Height, centroid, FWHM and area of DB1 band with Fo-Fa content based on EMPA. There is an observable trend that as Fo content increases the peak centroids increase in their wavenumber (cm^{-1}) value.

Sample ID	%Fo	DB1 Lorentzian				DB1 Gaussian			
		Height	Centroid	FWHM	Area	Height	Centroid	FWHM	Area
Fa10.5 Fo89.5	90	0.56	823.41±0.01	19.02±0.11	16.85±0.09	0.56	823.45±0.02	18.4±0.1	10.97±0.06
Fa10Fo90 Menzies	90	0.84	822.8±0.01	18.98±0.06	25.05±0.08	0.84	822.82±0.01	17.15±0.06	15.32±0.05
H30-B5	90	0.98	825.16±0.00	10.26±0.01	15.72±0.01	0.97	825.18±0	9.5±0.01	9.83±0.01
H366-28	90	0.79	822.92±0.01	16.78±0.05	20.73±0.06	0.78	822.92±0.01	15.62±0.05	13.03±0.04
San Carlos AZ	91	0.83	822.81±0.01	18.44±0.06	24.11±0.08	0.83	822.84±0.01	16.68±0.06	14.76±0.05
Ep-3-139-D	91	0.74	825.44±0.00	9.94±0.01	11.61±0.01	0.74	825.45±0	9.27±0.02	7.31±0.01
Ci-1-25	91	0.86	825.14±0.00	9.17±0.01	12.38±0.02	0.85	825.16±0	8.7±0.01	7.91±0.01
DH101-D	91	1.00	825.32±0.00	9.2±0.01	14.5±0.01	1.00	825.32±0	8.49±0.01	9.04±0.01
Dyar 89-12	91	0.83	823.33±0.01	18.29±0.06	23.9±0.08	0.83	823.39±0.01	16.87±0.06	14.89±0.05
Globe	91	0.77	823.18±0.01	19.93±0.09	24.23±0.11	0.77	823.17±0.01	17.54±0.07	14.45±0.06
H279-12	91	1.00	825.16±0.00	9.11±0.01	14.36±0.02	1.00	825.16±0	8.55±0.01	9.1±0.01
H30-B3	91	1.00	825.09±0.00	9.28±0.01	14.58±0.01	1.00	825.05±0	8.8±0.01	9.33±0.01
H30-B4	91	0.94	825.43±0.00	9.95±0.01	14.62±0.01	0.93	825.45±0	9.21±0.01	9.14±0.01
H312-1	91	0.98	825.37±0.00	7.86±0.01	12.11±0.02	0.98	825.38±0	7.34±0.01	7.63±0.01
Ki-5-31	91	0.47	825.45±0.00	9.16±0.01	6.8±0.01	0.47	825.42±0	8.14±0.01	4.11±0.01
H30-B1	91	0.75	823.6±0.01	17.95±0.06	21.1±0.07	0.75	823.66±0.01	16.35±0.06	13±0.05
DH101-C	92	0.32	825.02±0.00	9.65±0.03	4.9±0.01	0.32	825.02±0	8.77±0.01	3.01±0.00474
Pakistan Kohistan	92	0.68	823.78±0.01	17.28±0.05	18.32±0.05	0.67	823.76±0.01	15.6±0.05	11.2±0.04
DH101-E	92	1.00	824.98±0.00	9.24±0.01	14.59±0.02	1.00	825.04±0	8.82±0.01	9.38±0.01
Dyar 89-194	92	0.81	823.81±0.01	18.42±0.06	23.34±0.07	0.81	823.82±0.01	16.46±0.05	14.12±0.04
H30-B2	92	0.28	824.27±0.01	15.77±0.07	7.04±0.03	0.28	824.34±0.01	14.35±0.05	4.34±0.02
Ki-5-16	92	1.00	825.4±0.00	8.96±0.01	14.11±0.01	1.00	825.43±0	8.39±0.01	8.91±0.01
Dyar 89-187	96	0.79	824.4±0.01	17.67±0.05	21.97±0.06	0.79	824.42±0.01	15.93±0.05	13.41±0.04
Ki-5-62	98	0.86	823.16±0.01	17.41±0.05	23.41±0.07	0.85	823.19±0.01	16.05±0.05	14.58±0.05
Fa0Fo100a	100	0.93	827.51±0.00	7.72±0.01	11.3±0.01	0.93	827.58±0	7.45±0.01	7.33±0.01
Fa0Fo100b	100	0.70	827.42±0.00	8.63±0.02	9.53±0.02	0.70	827.44±0	7.82±0.01	5.84±0.01
Fa0Fo100c	100	0.93	827.31±0.00	8.61±0.02	12.56±0.02	0.93	827.39±0	7.76±0.01	7.69±0.01

Table 4.2. Raman spectra characteristics of DB2 (peak at 837-857 cm^{-1}). Height, centroid, FWHM and area of DB2 band with Fo-Fa content based on EMPA. There is an observable trend that as Fo content increases the peak centroids increase in their wavenumber (cm^{-1}) value.

Sample ID	%Fo	DB2 Lorentzian				DB2 Gaussian			
		Height	Centroid	FWHM	Area	Height	Centroid	FWHM	Area
Fa100Fo0a	0	0.30	833.01±0.26	102.11±5.09	48.59±2.42	0.30	832.78±0.34	97.31±5.39	31.31±1.73
Fa100Fo0b	0	0.51	842.56±0.03	23.59±0.25	18.79±0.2	0.50	841.03±0.04	27.32±0.28	14.6±0.15
Rockport	0	0.65	841.18±0.02	28.2±0.2	28.96±0.21	0.65	841.22±0.03	26.09±0.21	18.13±0.14
NMNH 85539	1	0.64	841.13±0.02	27.36±0.15	27.64±0.15	0.64	841.22±0.02	26.61±0.19	18.16±0.13
NMNH 1210672	1	0.83	842.68±0.02	27.31±0.15	35.67±0.19	0.82	841.96±0.02	31.66±0.21	27.72±0.18
NMNH 112085	4	0.92	842.69±0.01	30.66±0.14	44.28±0.21	0.92	842.68±0.02	27.47±0.13	26.87±0.13
NMNH 135841	5	0.62	841.98±0.02	29.18±0.2	28.47±0.19	0.62	841.98±0.03	27.74±0.21	18.3±0.14
Fa90Fo10a	10	0.38	834.3±0.07	38.32±0.78	23.02±0.47	0.38	833.29±0.19	58.32±1.87	23.47±0.75
Fa90Fo10b	10	0.24	845.38±0.01	9.15±0.04	3.46±0.01	0.21	841.74±0.03	11.65±0.15	2.63±0.03
KI-3005	11	0.96	842.79±0.01	24.05±0.09	36.19±0.14	0.96	842.83±0.01	21.62±0.1	22.04±0.1
KI-3003	14	1.00	844±0.01	22.59±0.05	35.54±0.08	1.00	844.03±0.01	19.95±0.04	21.26±0.05
Fa80Fo20a	20	0.41	839.48±0.03	21.5±0.24	13.91±0.16	0.41	839.29±0.05	20.72±0.27	9.08±0.12
Fa80Fo20b	20	0.92	842.84±0.01	21.67±0.1	31.45±0.15	0.92	842.62±0.02	22.18±0.12	21.68±0.12
Fa80Fo20c	20	0.96	843.74±0.01	23.87±0.08	35.87±0.13	0.96	843.71±0.01	21.9±0.09	22.27±0.09
NMNH 9140	21	1.00	846.96±0.01	21.58±0.06	33.92±0.09	1.00	846.94±0.01	19.91±0.06	21.15±0.06
KI-3289	23	0.81	844.02±0.01	23.26±0.09	29.54±0.11	0.81	843.88±0.01	21.44±0.09	18.44±0.08
Fa70Fo30a	30	0.39	838.98±0.08	57.2±1.25	34.84±0.76	0.39	838.9±0.10	53.84±1.22	22.21±0.5
Fa70Fo30b	30	0.81	844.89±0.01	25.87±0.09	32.83±0.11	0.81	844.81±0.01	23.32±0.1	20.05±0.08
KI-3373	34	0.16	845.61±0.01	11.31±0.05	2.82±0.01	0.16	845.61±0.01	10.21±0.03	1.72±0.01
Fa60Fo40	40	0.46	844.86±0.03	30.44±0.37	22.01±0.27	0.46	844.78±0.04	28.04±0.36	13.73±0.17
Fa50Fo50a	50	0.86	850.42±0.01	25.52±0.1	34.49±0.14	0.86	850.43±0.01	22.66±0.09	20.75±0.08
Fa50Fo50b	50	0.96	848.14±0.01	22.1±0.06	33.21±0.09	0.96	848.15±0.01	19.72±0.06	20.08±0.06
Fa50Fo50c	50	0.92	848.84±0.01	21.32±0.06	30.65±0.08	0.91	848.89±0.01	19.49±0.07	18.97±0.07
KI-4030	52	0.70	849.98±0.01	13.67±0.05	15.08±0.05	0.70	849.83±0.01	12±0.03	8.95±0.02
Fa45 Fo55	55	0.29	845.23±0.02	19.34±0.16	8.93±0.07	0.29	845.23±0.03	17.88±0.16	5.59±0.05
KI-3362	59	1.00	851.11±0.00	19.07±0.03	29.97±0.05	1.00	851.1±0.01	17.16±0.04	18.27±0.04
Fa40Fo60a	60	0.83	853.1±0.01	16.94±0.05	22±0.07	0.83	853.16±0.01	14.97±0.03	13.19±0.03
Fa40Fo60b	60	0.84	852±0.01	11.51±0.06	15.25±0.08	0.85	852.14±0.01	9.78±0.04	8.84±0.04
SNC LAP-0484016	63	0.75	851.52±0.01	18.77±0.05	22.05±0.06	0.75	851.5±0.01	16.86±0.05	13.41±0.04
Fa35 Fo65	65	0.67	846.54±0.04	42.97±0.46	44.91±0.48	0.67	846.35±0.04	37.47±0.39	26.56±0.27
SNC Chassigny USNM E24	69	0.52	851.56±0.02	25.32±0.58	20.5±0.47	0.52	851.67±0.02	20.76±0.38	11.39±0.21
Fa30Fo70a	70	0.38	851.76±0.03	20±0.17	12.01±0.1	0.38	851.19±0.02	18.26±0.13	7.44±0.05
Fa30Fo70b	70	0.53	851.11±0.01	21.4±0.07	17.86±0.06	0.53	851.15±0.01	19.45±0.08	10.99±0.05
KI-3648	71	1.00	853.25±0.00	14.98±0.02	23.52±0.03	1.00	853.22±0.00	13.95±0.02	14.8±0.02
Y000097 86	71	0.90	851.63±0.01	21.69±0.09	30.72±0.12	0.90	851.58±0.01	20.04±0.08	19.19±0.08

Table 4.2 (continued). Raman spectra characteristics of DB2 (peak at 837-857 cm^{-1}). Height, centroid, FWHM and area of DB2 band with Fo-Fa content based on EMPA. There is an observable trend that as Fo content increases the peak centroids increase in their wavenumber (cm^{-1}) value.

Sample ID	%Fo	DB2 Lorentzian				DB2 Gaussian			
		Height	Centroid	FWHM	Area	Height	Centroid	FWHM	Area
SNC ALHA 77005	71	0.87	853.41±0.01	14.69±0.05	20.18±0.06	0.87	853.54±0.00	13.28±0.02	12.36±0.02
SNC ALHA-77005-193	71	0.92	853.04±0.00	16.05±0.03	23.15±0.05	0.92	852.88±0.01	15.13±0.05	14.76±0.05
SNC NWA2737	72	0.93	855.09±0.02	18.37±0.12	26.93±0.17	0.94	854.98±0.01	13.97±0.08	14.03±0.08
SNC EETA-79001 60B	75	0.56	853.8±0.00	13.55±0.03	12.01±0.03	0.56	853.73±0.01	12.25±0.03	7.35±0.02
SNC EETA-79001-A	75	1.00	852.81±0.00	15.97±0.03	25.15±0.05	1.00	852.85±0.00	14.41±0.01	15.36±0.02
Fa25 Fo75	75	0.36	851.43±0.01	19.37±0.1	11.08±0.06	0.36	851.77±0.03	18.11±0.14	7.01±0.05
Dyar 89-190	76	1.00	851.89±0.01	22.56±0.05	35.43±0.08	1.00	851.83±0.01	20.36±0.08	21.67±0.08
Fa20Fo80a	80	0.67	851.1±0.02	23.18±0.18	24.34±0.18	0.67	851.23±0.02	21.15±0.15	15.02±0.1
Fa20Fo80b	80	0.95	855.32±0.00	12.7±0.02	19.04±0.03	0.95	855.33±0.00	11.72±0.02	11.88±0.02
Ep-3-7A	81	1.00	855.87±0.00	14.2±0.01	22.32±0.02	1.00	855.89±0.00	12.92±0.02	13.74±0.02
H30-82-8	81	0.91	855.72±0.00	11.38±0.01	16.18±0.02	0.90	855.7±0.00	10.48±0.02	10.08±0.02
KBH-94-23-B	82	1.00	855.68±0.00	14.1±0.01	22.17±0.01	1.00	855.69±0.00	12.85±0.02	13.67±0.02
Ba-2-1 WR1	86	1.00	856.36±0.00	10.4±0.02	16.39±0.03	1.00	856.37±0.00	9.5±0.00	10.12±0.00
KBH-94-23-E	86	1.00	856.27±0.00	13.21±0.01	20.77±0.02	1.00	856.29±0.00	12.08±0.02	12.85±0.02
Ba-2-1 WR2	88	0.87	856.85±0.00	11.34±0.02	15.53±0.03	0.87	856.86±0.00	10.22±0.01	9.47±0.01
Ci-1-183	88	1.00	857.36±0.00	10.38±0.03	16.3±0.04	1.00	857.36±0.00	9.25±0.01	9.84±0.01
Ba-1-74	89	0.79	856.94±0.00	9.36±0.02	11.69±0.02	0.79	856.95±0.00	8.54±0.01	7.21±0.01
DH101-B	89	0.83	856.62±0.00	10.17±0.01	13.21±0.02	0.82	856.62±0.00	9.47±0.02	8.3±0.01
Ep-1-13	89	0.89	850.93±0.01	18.91±0.05	26.37±0.07	0.88	850.88±0.01	18.08±0.07	16.99±0.06
Ep-3-44	89	0.43	856.9±0.00	11.23±0.02	7.5±0.02	0.42	856.92±0.00	10.28±0.02	4.64±0.01
H366-30	89	0.87	856.96±0.00	10.05±0.03	13.79±0.04	0.87	856.99±0.00	9.11±0.01	8.46±0.01
Ki-5-235	89	1.00	856.9±0.00	11.71±0.00462	18.4±0.01	1.00	856.92±0.00	10.76±0.02	11.44±0.02
Ki-5-35	89	1.00	857±0.00	12.09±0.01	19±0.01	1.00	857.02±0.00	11.12±0.02	11.82±0.02
Ki-5-35	89	1.00	857.62±0.00	11.76±0.01	18.5±0.01	1.00	857.67±0.00	11.11±0.02	11.79±0.02
Ba-1-61	90	0.95	856.97±0.00	9.45±0.02	14.15±0.03	0.95	857.01±0.00	8.68±0.00	8.78±0.00
Ba-2-1 WR3	90	1.00	857.08±0.00	10.58±0.01	16.64±0.02	1.00	857.06±0.00	9.8±0.01	10.41±0.01
Ba-2-1 WR4	90	1.00	857.22±0.00	10.56±0.01	16.63±0.02	1.00	857.23±0.00	9.69±0.01	10.31±0.01
Ba-2-1 D-1	90	0.47	856.85±0.00	8.4±0.02	6.26±0.01	0.47	856.9±0.00	7.69±0.01	3.88±0.01
Ep-3-139-C	90	1.00	857.66±0.00	12.01±0.01	18.92±0.02	1.00	857.68±0.00	10.98±0.01	11.69±0.01
Ep-3-46	90	1.00	857.16±0.00	10.59±0.01	16.67±0.02	1.00	857.14±0.00	9.83±0.01	10.45±0.01
Ep-3-72	90	1.00	857.41±0.00	10.39±0.01	16.35±0.02	1.00	857.39±0.00	9.52±0.01	10.13±0.01
Fa10.5 Fo89.5	90	0.69	855.56±0.01	20.82±0.05	22.49±0.05	0.69	855.63±0.01	18.94±0.05	13.84±0.04
Fa10Fo90 Menzies	90	1.00	854.66±0.00	20.64±0.04	32.71±0.07	1.00	854.67±0.01	18.67±0.06	20.03±0.06
30-B5	90	1.00	857.35±0.00	11.53±0.01	18.12±0.01	1.00	857.38±0.00	10.67±0.02	11.33±0.02

Table 4.2 (continued). Raman spectra characteristics of DB2 (peak at 837-857 cm^{-1}). Height, centroid, FWHM and area of DB2 band with Fo-Fa content based on EMPA. There is an observable trend that as Fo content increases the peak centroids increase in their wavenumber (cm^{-1}) value.

Sample ID	%Fo	DB2 Lorentzian				DB2 Gaussian			
		Height	Centroid	FWHM	Area	Height	Centroid	FWHM	Area
H366-28	90	0.95	854.9±0.00	18.33±0.04	27.25±0.05	0.94	854.89±0.01	17.2±0.04	17.26±0.04
San Carlos AZ	91	1.00	854.73±0.00	19.83±0.04	31.38±0.06	1.00	854.71±0.01	17.85±0.05	19.13±0.05
Ep-3-139-D	91	1.00	857.48±0.00	11.78±0.01	18.51±0.01	1.00	857.52±0.00	10.83±0.02	11.5±0.02
Ci-1-25	91	1.00	857.5±0.00	10.02±0.01	15.76±0.02	1.00	857.51±0.00	9.21±0.01	9.8±0.01
DH101-D	91	0.32	857.3±0.00	10.08±0.03	5.09±0.01	0.32	857.34±0.00	9.25±0.01	3.16±0.00
Dyar 89-12	91	1.00	855.06±0.00	20.45±0.05	32.43±0.08	1.00	855.12±0.01	18.96±0.05	20.3±0.05
Globe	91	1.00	855.44±0.00	19.33±0.03	30.47±0.05	1.00	855.42±0.01	17.45±0.04	18.62±0.05
H279-12	91	0.69	857.36±0.00	10.02±0.01	10.88±0.01	0.69	857.34±0.00	9.25±0.01	6.79±0.01
H30-B3	91	0.89	857.3±0.00	10.87±0.01	15.25±0.02	0.89	857.31±0.00	9.97±0.01	9.46±0.01
H30-B4	91	1.00	857.59±0.00	11.2±0.01	17.61±0.01	1.00	857.63±0.00	10.34±0.02	10.99±0.02
H312-1	91	1.00	857.96±0.00	8.67±0.01	13.62±0.02	0.99	858±0.00	8.32±0.02	8.79±0.02
Ki-5-31	91	1.00	857.34±0.00	10.27±0.01	16.17±0.02	1.00	857.32±0.00	9.59±0.01	10.19±0.01
H30-B1	91	1.00	855.6±0.00	18.84±0.03	29.65±0.05	1.00	855.65±0.01	17.02±0.04	18.14±0.04
DH101-C	92	0.94	857.36±0.00	10.24±0.01	15.17±0.02	0.94	857.36±0.00	9.41±0.01	9.42±0.01
Pakistan Kohistan	92	1.00	855.72±0.00	16.75±0.03	26.35±0.04	1.00	855.76±0.01	15.12±0.03	16.1±0.03
DH101-E	92	0.33	857.07±0.00	9.32±0.02	4.86±0.01	0.33	857.08±0.00	8.57±0.01	3.02±0.00
Dyar 89-194	92	1.00	855.69±0.01	19.98±0.05	31.44±0.08	1.00	855.77±0.01	18.04±0.04	19.21±0.05
H30-B2	92	0.46	856.81±0.01	21.11±0.08	15.4±0.06	0.46	856.77±0.01	19.62±0.1	9.67±0.05
Ki-5-16	92	0.51	857.49±0.00	9.74±0.02	7.85±0.01	0.51	857.45±0.00	9.39±0.02	5.1±0.01
Dyar 89-187	96	1.00	856.57±0.00	19.09±0.03	30.11±0.05	1.00	856.56±0.01	17.17±0.04	18.33±0.04
Ki-5-62	98	1.00	855.19±0.00	18.99±0.04	30.05±0.06	1.00	855.19±0.01	17.23±0.04	18.45±0.05
Fa0Fo100a	100	1.00	859.8±0.00	7.63±0.01	12.03±0.01	1.00	859.82±0.00	7.18±0.01	7.64±0.01
Fa0Fo100b	100	1.00	859.6±0.00	7.96±0.01	12.53±0.01	1.00	859.61±0.00	7.42±0.01	7.88±0.01
Fa0Fo100c	100	1.00	859.94±0.00	7.45±0.01	11.67±0.01	0.99	859.98±0.00	6.97±0.01	7.37±0.01

Table 4.3. Raman spectra centroids, height, FWHM and area of DB1 and DB2 bands from the RRUFF data set. Spectra were acquired through <http://rruff.info/> on a 532nm unoriented spectrometer. All samples have <5 wt% of cations beyond Mg and Fe²⁺ as confirmed through EMPA.

DB1		Lorentzian				Gaussian			
Sample ID	%Fo	Height	Centroid	FWHM	Area	Height	Centroid	FWHM	Area
R100103	2	1.00	815.31±0.01	20.15±0.1	31.65 ± 0.15	1.00	815.44±0.02	17.69±0.09	18.87±0.09
R100102	45	0.63	818.23±0.02	22.31±0.14	22.00 ± 0.14	0.63	818.2±0.02	20.78±0.15	13.86±0.1
R060539	84	0.83	821.92±0.00	13.06±0.02	17.06 ± 0.02	0.83	822±0.01	12.11±0.03	10.71±0.03
R040057	84	0.73	821.86±0.00	13.46±0.02	15.42 ± 0.02	0.73	821.83±0.01	12.29±0.03	9.53±0.02
R100101	91	0.52	823.8±0.00	12.79±0.03	10.36 ± 0.02	0.52	823.71±0.01	11.59±0.04	6.37±0.02
R040018	91	1.00	822.49±0.00	11.94±0.02	18.78 ± 0.03	1.00	822.45±0.00	10.82±0.02	11.52±0.02
R050117	91	0.48	823.27±0.00	13.14±0.03	9.82 ± 0.02	0.48	823.22±0.01	11.84±0.04	6±0.02
R060535	91	1.00	823.42±0.00	12.17±0.02	19.10 ± 0.02	1.00	823.35±0.01	11.05±0.02	11.75±0.03
R060551	91	1.00	823.32±0.00	11.88±0.01	18.67 ± 0.02	1.00	823.33±0.00	10.97±0.02	11.65±0.02
R100100	92	1.00	823.66±0.00	10.59±0.01	16.63 ± 0.01	1.00	823.61±0.00	9.72±0.02	10.33±0.02
R040052	100	0.45	824.49±0.00	10.17±0.02	7.16 ± 0.01	0.45	824.43±0.01	9.45±0.02	4.5±0.01
R100099	100	0.37	825.38±0.00	11.12±0.03	6.54 ± 0.02	0.37	825.38±0.01	10.38±0.03	4.12±0.01
DB2		Lorentzian				Gaussian			
Sample ID	%Fo	Height	Centroid	FWHM	Area	Height	Centroid	FWHM	Area
R100103	2	0.73	836.55±0.03	43±0.41	49.60 ± 0.47	0.73	836.76±0.05	40.93±0.45	31.96±0.35
R100102	45	1.00	845.76±0.01	24.91±0.05	39.07 ± 0.09	1.00	845.82±0.01	22.46±0.07	23.87±0.07
R060539	84	1.00	853.46±0.00	15.7±0.02	24.68 ± 0.03	1.00	853.4±0.01	14.11±0.03	15.03±0.03
R040057	84	1.00	853.49±0.00	15.58±0.02	24.49 ± 0.03	1.00	853.42±0.00	14.06±0.03	14.96±0.03
R100101	91	1.00	855.52±0.00	14.14±0.02	22.22 ± 0.04	1.00	855.51±0.00	12.98±0.03	13.8±0.03
R040018	91	0.66	854.55±0.00	13.3±0.03	13.73 ± 0.03	0.66	854.51±0.00	11.82±0.02	8.27±0.01
R050117	91	1.00	855.09±0.00	14.06±0.02	22.11 ± 0.03	1.00	855.07±0.00	12.86±0.03	13.68±0.03
R060535	91	0.99	855.43±0.00	13.36±0.02	20.81 ± 0.03	0.99	855.42±0.01	12.33±0.03	12.97±0.03
R060551	91	0.79	855.35± 0.00	13.41±0.02	16.63 ± 0.03	0.79	855.36±0.01	12.51±0.03	10.47±0.03
R100100	92	0.88	855.86±0.00	11.68±0.01	16.21 ± 0.02	0.88	855.86±0.00	10.74±0.02	10.08±0.02
R040052	100	1.00	856.55±0.00	10.42±0.03	16.35 ± 0.04	1.00	856.54±0.01	9.69±0.03	10.27±0.03
R100099	100	1.00	857.56±0.00	10.69±0.02	16.83 ± 0.04	1.00	857.55±0.01	10.17±0.03	10.79±0.03

Table 4.4. Raman spectra centroids of DB1 and DB2 bands from literature. All samples have <5 wt% of cations beyond Mg and Fe²⁺ as confirmed through EMPA.

Sample Set	Sample ID	%Fo	DB1		DB2	
			Centroid	S.D.	Centroid	S.D.
Foster et al. (2007)	Fo0	0	812.00	n.a.	837.00	n.a.
Kolesov and Tanskaya, (1996)	fayalite	0	815.00	n.a.	839.00	n.a.
Kuebler et al. (2006)	fayalite 203 synthetic	1	815.00	0.09	838.80	0.52
Weber et al. (2014)	63	2	818.00	n.a.	832.00	n.a.
Kuebler et al. (2006)	Forsyth Iron Mine	8	815.30	0.07	837.40	1.18
Guyot et al. (1986)	Fo20	20	815.00	n.a.	841.00	n.a.
Kuebler et al. (2006)	LAP 02224,24	34	816.20	0.29	840.70	1.17
Kuebler et al. (2006)	Hortonolite high	39	816.50	0.03	844.40	0.03
Kuebler et al. (2006)	Hortonolite av.	40	817.00	0.26	843.20	0.54
Kuebler et al. (2006)	Hortonolite low	40	817.10	0.00	842.80	0.00
Guyot et al. (1986)	Fo41	41	818.00	n.a.	845.00	n.a.
Kuebler et al. (2006)	rim EETA 79001,530	56	818.20	0.05	847.60	0.03
Kuebler et al. (2006)	av. EETA 79001,530	62	819.00	0.54	848.70	0.90
Ishibashi et al. (2011)	15	63	819.81	0.05	849.36	0.09
Abdu et al. (2011)	honey olivine (Hol)	64	820.00	n.a.	850.00	n.a.
Kuebler et al. (2006)	NWA 773	65	819.00	0.16	849.20	0.26
Ishibashi et al. (2011)	14	65	820.10	0.04	849.66	0.04
Ishibashi et al. (2011)	13	67	820.50	0.02	850.43	0.01
Kuebler et al. (2006)	core EETA 79001,530	69	819.60	0.06	849.90	0.01
Ishibashi et al. (2011)	12	71	820.41	0.04	851.22	0.01
Yasuzuka et al. (2009)	Fo70.7	71	819.46	0.06	850.38	0.04
Ishibashi et al. (2011)	11	72	820.93	0.01	851.81	0.02
Yasuzuka et al. (2009)	Fo73.7	74	819.64	0.10	851.25	0.04
Guyot et al. (1986)	Fo74	74	820.00	n.a.	851.00	n.a.
Ishibashi et al. (2011)	10	75	821.29	0.03	852.17	0.03
Yasuzuka et al. (2009)	Fo76.7	77	820.36	0.07	852.00	0.05
Ishibashi et al. (2011)	9	77	821.67	0.02	852.53	0.02
Ishibashi et al. (2011)	8	81	822.28	0.01	853.33	0.03
Yasuzuka et al. (2009)	Fo82.3	82	821.80	0.01	853.09	0.01
Ishibashi et al. (2011)	7	83	822.12	0.06	853.47	0.04
Ishibashi et al. (2011)	6	85	822.98	0.01	854.47	0.01
Yasuzuka et al. (2009)	Fo84.6	85	821.28	0.02	852.70	0.02
Abdu et al. (2011)	green olivine (Gol)	85	823.00	n.a.	854.00	n.a.
Ishibashi et al. (2011)	5	86	823.06	0.05	854.67	0.02
Ishibashi et al. (2011)	4a	87	823.56	0.02	855.16	0.02
Yasuzuka et al. (2009)	Fo87.5	88	822.56	0.05	854.22	0.05
Chopelas et al. (1991)	olivine, Fo88	88	822.00	n.a.	854.00	n.a.
Ishibashi et al. (2008)	n.a.	89	822.00	n.a.	854.00	n.a.
Kuebler et al. (2006)	San Carlos	89	822.50	0.01	854.50	0.05
Yasuzuka et al. (2009)	Fo89.5	90	822.50	0.04	854.39	0.03
Abdu et al. (2011)	olivinite (Pol)	90	824.00	n.a.	856.00	n.a.
Ishibashi et al. (2011)	3a	90	823.91	0.01	856.02	0.04
Yasuzuka et al. (2009)	Fo90.1	90	822.93	0.02	855.27	0.03
Weber et al. (2014)	61	90	824.00	n.a.	858.00	n.a.
Weber et al. (2014)	21	91	825.00	n.a.	859.00	n.a.
Weber et al. (2014)	36	91	824.00	n.a.	858.00	n.a.
Kuebler et al. (2006)	Twin Sisters	91	823.10	0.00	855.00	0.00
Weber et al. (2014)	62	91	824.00	n.a.	858.00	n.a.
Ishibashi et al. (2011)	2	93	824.52	0.02	856.38	0.02
Yasuzuka et al. (2009)	Fo92.7	93	823.40	0.04	855.58	0.05
Guyot et al. (1986)	Fo94	94	824.00	n.a.	855.00	n.a.

Table 4.4 (continued). Raman spectra centroids of DB1 and DB2 bands from literature. The Fo-Fa content of all samples was known through EMPA. All samples <5 wt% of cations beyond Mg and Fe²⁺.

Sample Set	Sample ID	%Fo	DB1		DB2	
			Centroid	S.D.	Centroid	S.D.
Kolesov and Tanskaya, (1996)	forsterite	99	824.00	n.a.	856.00	n.a.
Ishibashi et al. (2011)	1	100	825.77	0.02	857.99	0.03
Foster et al. (2007)	Fo100	100	826.00	n.a.	858.00	n.a.
Iishi,(1978)	Fo100	100	826.00	n.a.	856.00	n.a.
McKeown et al. (2010)	Fo100	100	826.00	n.a.	856.00	n.a.
Mohanan et al. (1993)	Fo100	100	824.00	n.a.	855.00	n.a.
Yasuzuka et al. (2009)	Fo100	100	824.71	0.02	857.00	0.02
Kuebler et al. (2006)	forsterite 204 synthetic	100	824.70	0.03	856.70	0.02
Chopelas et al. (1991)	synthetic forsterite	100	824.00	n.a.	856.00	n.a.

S.D. denoting standard deviation.

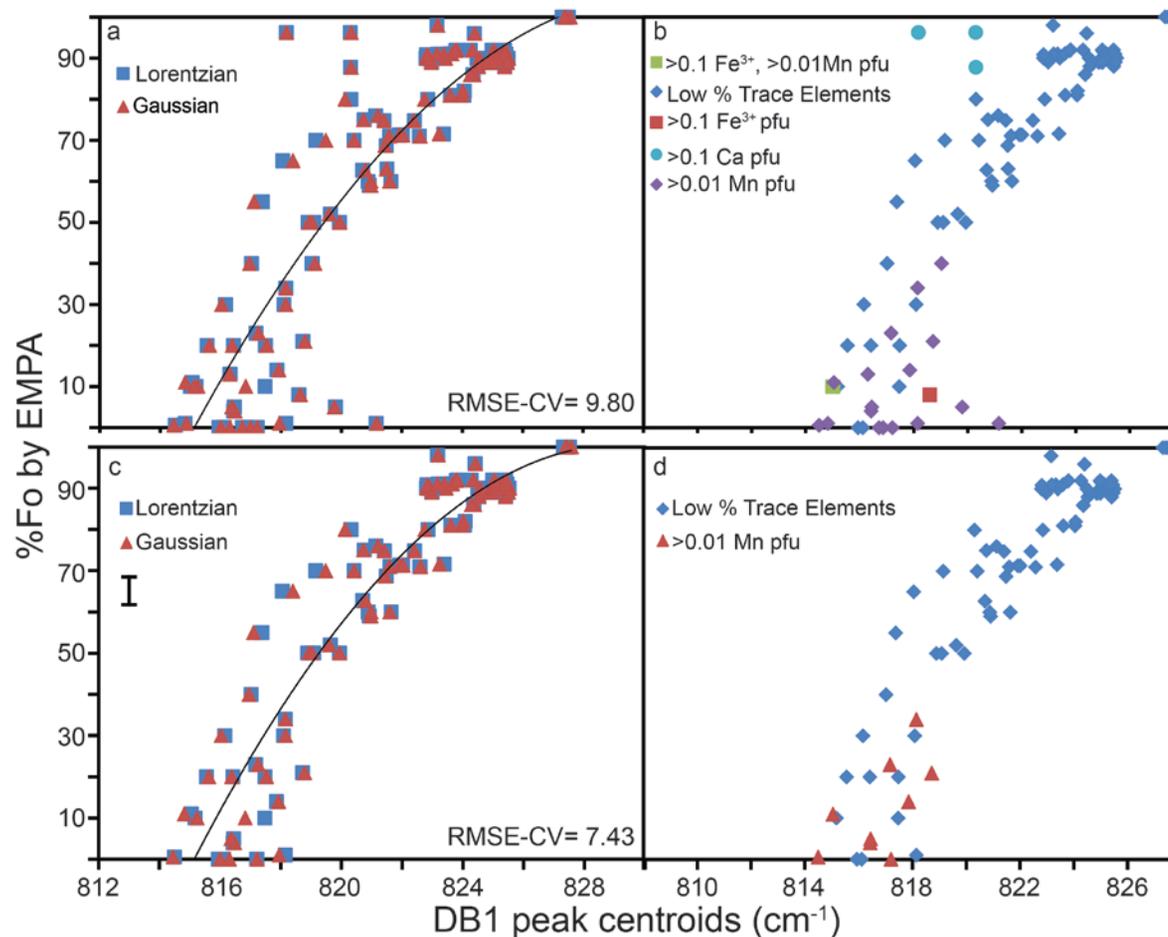


Figure 4.1 (a) %Fo versus peak centroids position with a second order polynomial fit. All olivine samples with a distinguishable DB1 peak are included in this model (106 samples). Peak centroids are estimated in a Lorentzian (blue) and Gaussian (red) methods. Part (b) replicates (a) of figure while showing sample compositions. Samples are color coded (>0.1 Fe³⁺ pfu and >0.01 Mn pfu (green), >0.1 Fe³⁺ pfu (red), >0.1 Ca pfu (light blue), >0.01 Mn pfu (purple) and remaining samples with low % of trace elements (dark blue)). Part (c) replicated part (a) of figure yet, with a reduced data (93 samples) set of samples with low % of trace elements as shown in part (d). In part (d) Samples are color coded with >0.01 Mn pfu in red and all remaining samples with low % of trace elements in dark blue). Error bars are smaller than the symbols for %Fo by EMPA and centroid errors are given in **Table 4.1**.

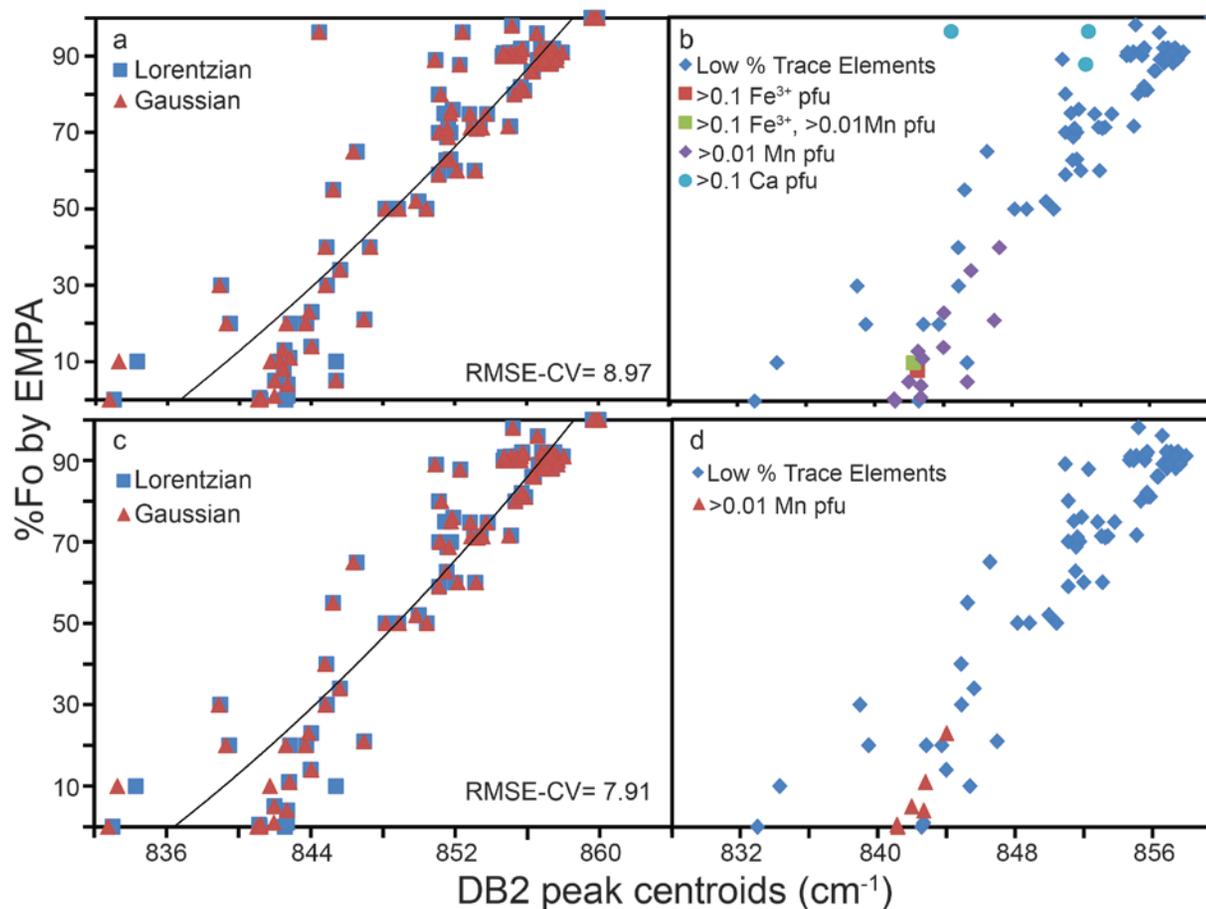


Figure 4.2. (a) %Fo versus peak centroids position with a second order polynomial fit. All olivine samples with a distinguishable DB2 peak are included in this model (102 centroids). Peak centroids are estimated in a Lorentzian (blue) and Gaussian (red) methods. Part (b) replicates (a) of figure while showing sample compositions. Samples are color coded (>0.1 Fe³⁺ pfu and >0.01Mn pfu (green), >0.1 Fe³⁺ pfu (red), >0.1 Ca pfu (light blue), >0.01 Mn pfu (purple) and remaining samples with low % of trace elements (dark blue)). Part (c) replicated part (a) of figure yet, with a reduced data (93 samples) set of samples with low % of trace elements as shown in part (d). In part (d) Samples are color coded with >0.01 Mn pfu in red and all remaining samples with low % of trace elements in dark blue). Error bars are smaller than the symbols for %Fo by EMPA and centroid errors are given in **Table 4.2**.

To understand the effect of elemental substitution from cations other than Mg^{2+} and Fe^{2+} , **Figures 4.1a** and **4.2a** plot the %Fo against the DB1 and DB2 peak centroids for all samples using an extended data set that includes laihunites (Fe^{3+} substitution), monticellites (Ca^{2+} substitution) and other species (**Table 2.1**). **Figures 4.1b** and **4.2b** color code each sample based on its composition. As previously discussed, the samples with odd substitutions are rare and we could not obtain a sufficient quantity of samples to model their properties effectively.

Based on the compositions of the samples in **Figures 4.1b** and **4.2b**, the data set was modified in **Figures 4.1c** and **4.2c** to include only samples with low percentages of Fe^{3+} , Ca, and Mn, leaving the 93 samples that form the focus of this study. The extended data set has RMSE-CV values for accuracy of ± 9.80 (DB1) and ± 8.97 (DB2) yet, the reduced data set values are ± 7.43 (DB1) and ± 7.91 (DB2) in units of %Fo. Reducing the data sets to include only samples with low percentages of trace elements (**Figures 4.1d** and **4.2d**) improves their prediction accuracies for most general geological scenarios (including those on planetary surfaces) where these odd compositions would not be expected.

Separating the spectral data acquired on the BRAVO and Senterra spectrometers also improves the accuracy of the %Fo prediction (**Figure 4.3**). In **Figure 4.3a**, it is apparent that the DB1 BRAVO data do not follow a linear trend, but lie on a polynomial curve. Results from the Senterra for DB1 (**Figure 4.3b**) and DB2 (both spectrometers) produce trends that are closer to linear. It is apparent that DB2, with a centroid that covers a much wider energy shift with changing composition (compare x -axis limits in **Figures 4.3a,b**), produces better prediction accuracy: RMSE-CV = 7.69 and 4.94 %Fo for DB1 versus 4.35 and 3.35 %Fo for DB2 with the BRAVO and Senterra instruments, respectively. As previously discussed within Chapter 3, the BRAVO and Senterra have differing resolutions and detector sensitivities, which likely impact

the consistency of these univariate predictions. Additionally, these two data sets were acquired using different laser wavelengths for excitation, which may also impact the peak centroid positions, as discussed in the following chapter.

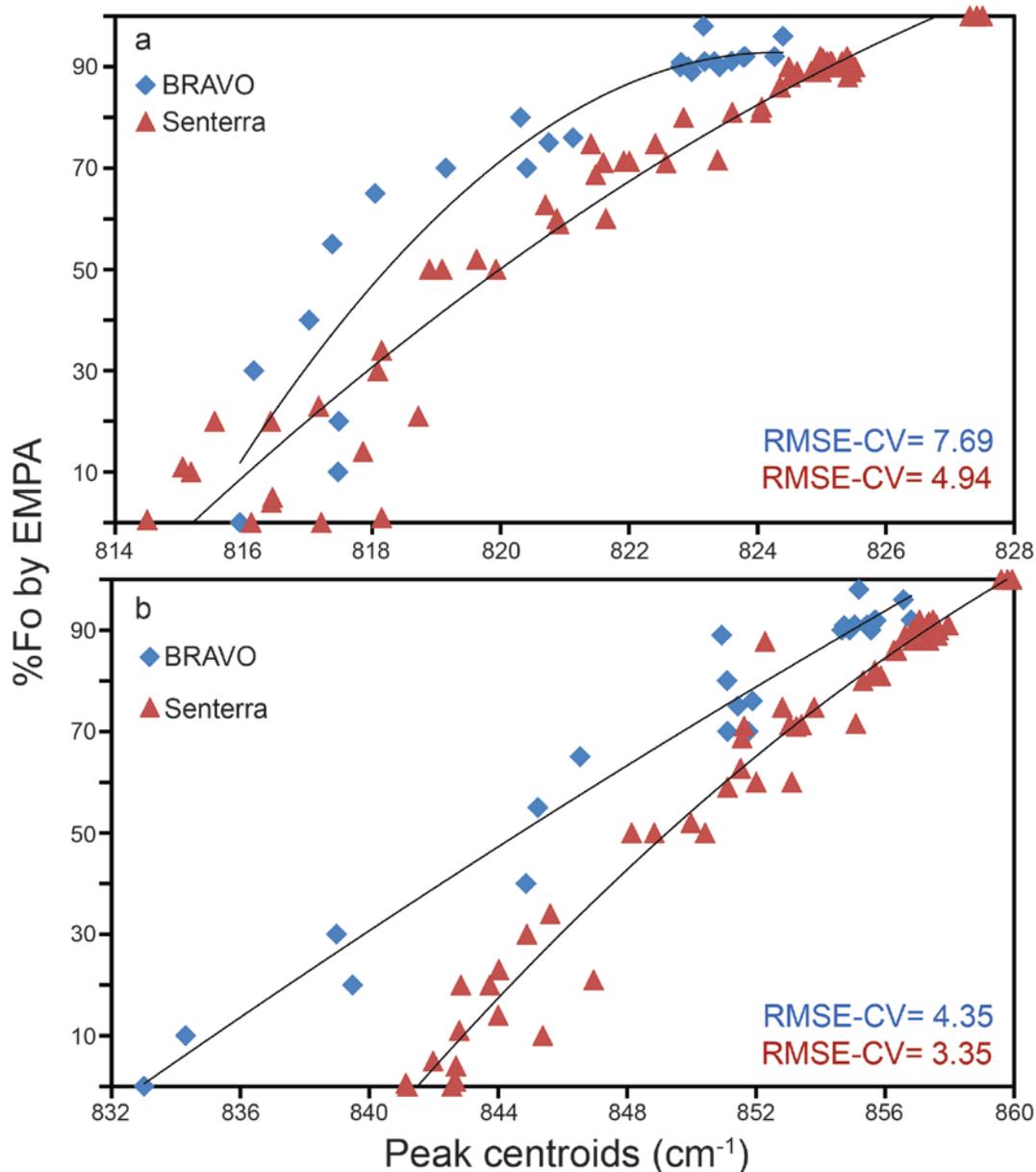


Figure 4.3. %Fo by EMPA versus peak centroids of DB1 (a) and DB2 (b). Second order polynomial fits and RSME-CV values are included for the data acquired on Bruker's BRAVO ($n=25$) and Senterra ($n=68$) spectrometers. Error bars are smaller than the symbols for %Fo by EMPA and centroid errors are given in **Tables 4.1 and 4.2**.

Figure 4.4 merges the BRAVO and Senterra results from **Figure 4.3** with data data from additional sources. **Figure 4.4** thus compares data on %Fo content versus peak centroid energy for five data sets: BRAVO (blue diamonds) and Senterra (red triangles) data from this study (**Tables 4.1 and 4.2**), olivine spectra from the RRUFF data set (**Table 4.3**, light blue squares), data from Kuebler et al.'s (2006) paper (green triangles), and other data from the literature (**Table 4.4**, purple circles).

The univariate %Fo prediction, which uses the DB1 peak, has RMSE-CVs ranging from ± 2.09 to ± 7.69 . The DB2 prediction has an RMSE-CV range of ± 3.33 to ± 5.39 (**Figure 4.4**). Of the five data sets, the RRUFF data produce the most accurate estimates of composition when using DB1, while RRUFF and Senterra data sets produce comparable accuracies when using DB2. However, great caution must be used when comparing the accuracy of these predictions because each data set is a different size. For example, the RRUFF data set has the smallest error for both the DB1 and DB2 second order polynomial fits but the data set consists of only 12 centroids.

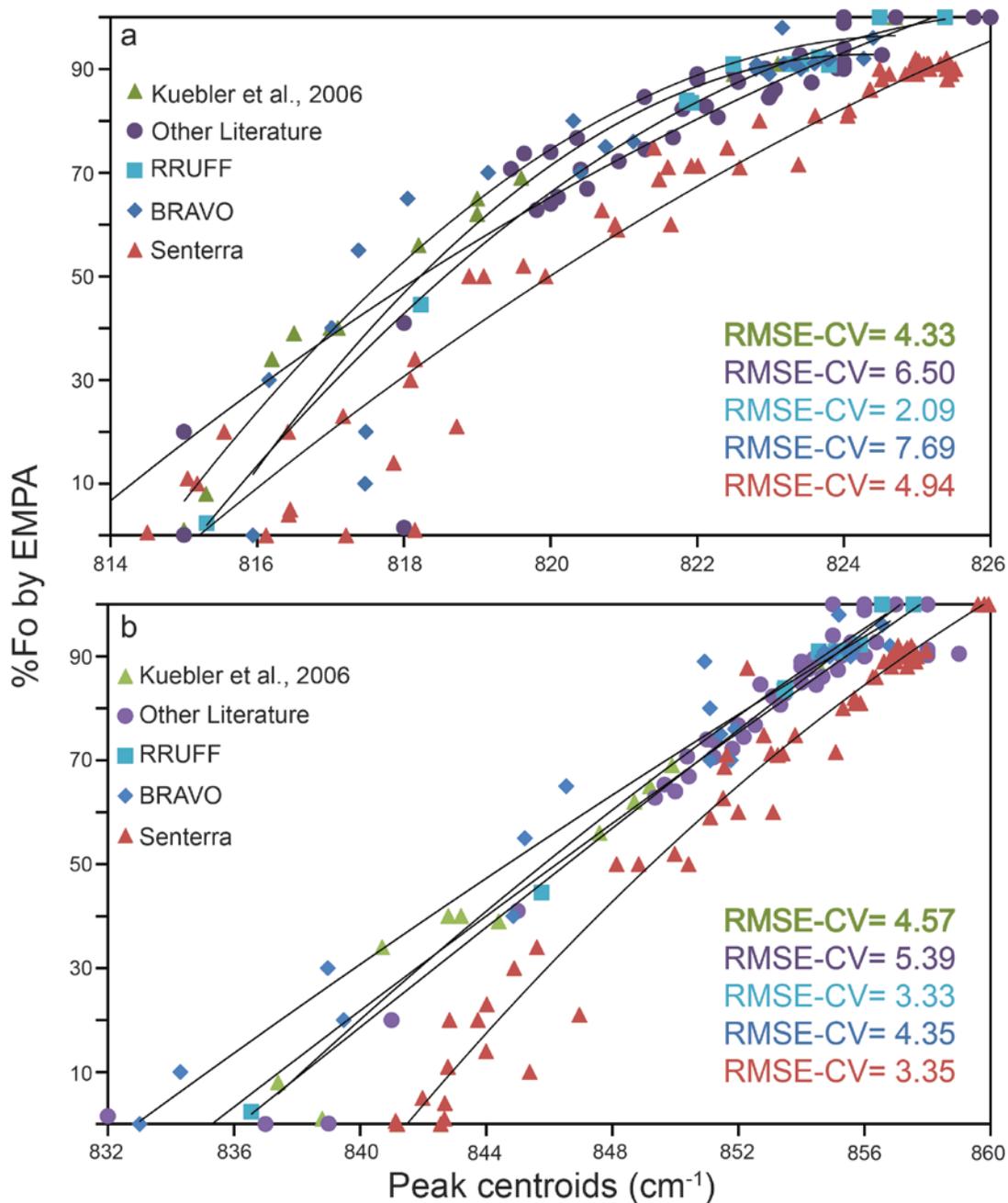


Figure 4.4. %Fo by EMPA versus peak centroids positions of (a) DB1 and (b) DB2. Lorentzian peak centroids of BRAVO (dark blue), the Senterra (red) and the RRUFF data set (light blue) are shown. The RRUFF data was acquired from <http://rruff.info/>. Kuebler et al. (2006) samples are shown in green. Other literature (purple) refers to peak values from **Table 4.4**. Error bars are smaller than the symbols for %Fo by EMPA and centroid errors are given in **Tables 4.1- 4.4**.

The equations of the first-, second-, and third-order polynomial fits that correspond to **Figures 4.1- 4.4** are reported for the DB1 (**Tables 4.5**) and DB2 (**Figure 4.5**) bands. Additionally, R^2 values and the RMSE-CVs are listed. RMSE-CV is the error bar obtained using leave-one-out cross validation. This method for estimating prediction accuracy uses results from 93 different regressions; in each case, one sample is held out and the remaining 92 samples are used to build a regression model and predict the 93rd sample's composition. The 93 %Fo predictions are then compared to the known %Fo for each sample and then used to calculate the RMSE-CV (cross-validated root mean square error of prediction), which represents the error bar on the composition predictions in units of %Fo. Generally, the second-order fits produce the best %Fo prediction models; however, in several cases the third-order polynomial RMSE-CV is smaller as is expected because higher order fits are more flexible and generally produce better fits.

Table 4.5. DB1 regressions for predicting %Fo. First-, second- and third-order fits are provided for each model.

Data	Polynomial order	Equation (y=)	R ²	RMSE-CV
Senterra & BRAVO (all)	1st	8.52x-6934.43	0.78±0	11±11.8
Senterra & BRAVO (all)	2nd	-0.43x ² +720.56x-299186	0.8±0	9.8±12.12
Senterra & BRAVO (all)	3rd	0x ³ +0.9x ² -373.55x-1.37	0.8±0	9.79±12.12
Senterra & BRAVO (reduced)	1st	8.34x-6791.37	0.86±0	8.88±7.73
Senterra & BRAVO (reduced)	2nd	-0.52x ² +855.54x-354569	0.89±0	7.43±7.56
Senterra & BRAVO (reduced)	3rd	0x ³ +1.06x ² -440.62x-1.61	0.89±0	7.43±7.56
BRAVO	1st	9.25x-7525.74	0.83±0	9.46±8.92
BRAVO	2nd	-1.16x ² +1912.64x-788200	0.88±0	7.69±8.6
BRAVO	3rd	0x ³ +2.35x ² -969.95x-3.55	0.88±0	7.69±8.59
Senterra	1st	8.57x-6981.74	0.94±0	5.64±5.97
Senterra	2nd	-0.27x ² +448.51x-187594	0.94±0	4.94±6.11
Senterra	3rd	0x ³ +0.57x ² -237.93x-0.87	0.94±0	4.93±6.11
Kuebler et al. 2006	1st	9.5x-7723.5	0.92±0	8.46±5.92
Kuebler et al. 2006	2nd	-0.92x ² +1514.98x-624752	0.98±0	4.33±3.47
Kuebler et al. 2006	3rd	0x ³ +1.87x ² -771.02x-2.82	0.98±0	4.35±3.49
RRUFF	1st	9.82x-7998.02	0.95±0.01	6.77±6.97
RRUFF	2nd	-0.74x ² +1229.27x-508120	0.99±0	2.09±1.32
RRUFF	3rd	0x ³ +1.52x ² -628.36x-2.3	0.99±0	2.08±1.3
Other literature	1st	8.02x-6512.56	0.85±0	6.85±7.78
Other literature	2nd	-0.28x ² +462.66x-192946	0.87±0.01	6.5±9.79
Other literature	3rd	0x ³ +0.59x ² -244.84x-0.9	0.87±0.01	6.49±9.76
Combined data	1st	8.25x-6709.76	0.84±0.00	9.70±7.04773
Combined data	2nd	-0.48x ² +804.08x-333316	0.87±0.00	7.92±7.50397
Combined data	3rd	0x ³ +1x ² -415.7x-1.52	0.87±0.00	7.91±7.49783

Table 4.6. DB2 regressions for predicting %Fo. First-, second- and third-order fits are provided for each model.

Data	Polynomial Order	Equation (y=)	R ²	RMSE-CV
Senterra & BRAVO (all)	1st	4.74x-3968.83	0.84±0	8.71±9.73
Senterra & BRAVO (all)	2nd	0.04x ² -57.9x+22619	0.84±0	8.97±9.72
Senterra & BRAVO (all)	3rd	-0.02x ³ +39.63x ² -33596.5x+9492950	0.87±0	7.62±9.49
Senterra & BRAVO (reduced)	1st	4.71x-3949.93	0.88±0	7.59±7.63
Senterra & BRAVO (reduced)	2nd	0.05x ² -73.41x+29204.6	0.89±0	7.91±7.41
Senterra & BRAVO (reduced)	3rd	-0.02x ³ +40.48x ² -34327.6x+9701590	0.92±0	6.67±6.77
BRAVO	1st	4.01x-3338.1	0.96±0	4.26±3.77
BRAVO	2nd	-0.02x ² +31.78x-15083.9	0.97±0	4.35±4.08
BRAVO	3rd	0x ³ +10.62x ² -8963.42x+2519900	0.97±0	4.84±4.11
Senterra	1st	5.49x-4617.2	0.97±0	3.75±4.46
Senterra	2nd	-0.09x ² +164.92x-72410.4	0.97±0	3.35±4.33
Senterra	3rd	0x ³ +8.89x ² -7479.88x+2095250	0.97±0	3.41±4.4
Kuebler et al. 2006	1st	4.78x-3994.52	0.97±0	3.9±5.11
Kuebler et al. 2006	2nd	-0.04x ² +76.51x-34377.4	0.97±0	4.57±4.86
Kuebler et al. 2006	3rd	0.01x ³ -25.46x ² +21607.6x-6113160	0.97±0.01	5.17±5.9
RRUFF	1st	4.76x-3980.84	1±0	1.69±0.87
RRUFF	2nd	0x ² +12.31x-7178.25	0.99±0.01	3.33±4.42
RRUFF	3rd	-0.01x ³ +12.94x ² -10954x+3090010	0.99±0.01	3.06±4.45
Other literature	1st	4.44x-3705.35	0.94±0	4.85±5.1
Other literature	2nd	-0.01x ² +22.68x-11433	0.94±0.01	5.39±7.6
Other literature	3rd	-0.01x ³ +37.6x ² -31830.5x+8979790	0.98±0	3.56±2.44
Combined data	1st	4.63x-3879.14	0.89±0.00	7.98± 6.29
Combined data	2nd	0.03x ² -37.87x+14151.6	0.89±0.00	8.18± 6.04
Combined data	3rd	-0.02x ³ +39.51x ² -33485.2x+9457710	0.92±0.00	6.55± 5.87

Each model listed in **Tables 4.5** and **4.6** evaluates the accuracy of that given model but does not necessarily inform how well it will predict “unseen data”. This issue is evaluated within this thesis with RMSE-test values, which will be discussed further for these models. Ignoring the predictability of a model on “unseen data” is an issue not accounted for in previous studies. Because there are many factors that affect a spectrum (i.e., laser wavelength, detector sensitivity, resolution), the accuracy of predicting “unseen data” with a model built under different conditions (instrument, resolution, or sample suite) needs to be evaluated.

Because the goal of this thesis project was to develop a generalizable expression to predict %Fo from Raman spectra on planetary surfaces from a diverse set of instruments, the issue of

generalizability was examined. The %Fo of the combined BRAVO and Senterra (532 nm) data sets were predicted using second-order polynomial fits from **Tables 4.5** and **4.6**. Four models were used for these predictions: Kuebler et al. (2006), RRUFF, other literature (**Table 2.3**) and an aggregated data set of Senterra (532 nm), Kuebler et al. (2006), BRAVO, RRUFF and other literature data. Next, a predicted %Fo versus %Fo by EMPA was graphed in **Figure 4.5** for DB1 and DB2. Finally, the four corresponding RMSE-test values were calculated.

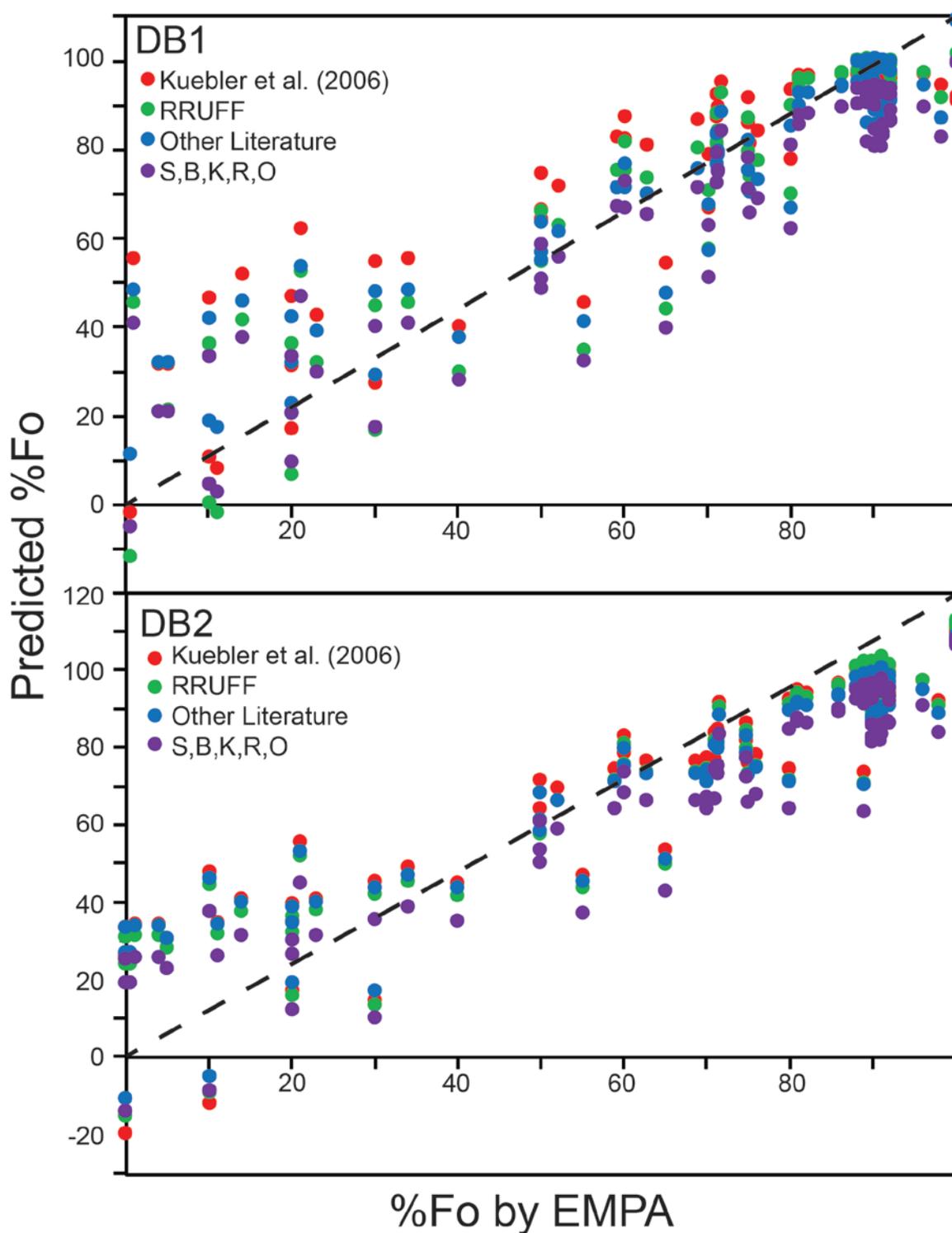


Figure 4.5. Predicted %Fo versus %Fo by EMPA using DB1 (top) and DB2 (bottom). The four prediction models used include the Kuebler et al. (2006) (dark green), RRUFF (light blue), other literature (Table 2.3) (dark blue) and an aggregated data set of Senterra (532 nm), BRAVO, Kuebler et al. (2006), RRUFF and the other literature data (listed as S,B,K,R,O) (purple). These four models were used to predict %Fo of combined Senterra (532 nm) and BRAVO data. Error bars for %Fo by EMPA are smaller than the size of the symbols.

The RMSE-test values were calculated from the data in **Figure 4.5** and are summarized with the histogram of **Figure 4.6**. The univariate RMSE-test values range from ± 10.4 to ± 16.2 (in %Fo units). The RMSE-test values are comparable for the DB1 and DB2 bands for each prediction model (comparing red and blue bars). A slightly smaller RMSE-test value is associated with the aggregated data set compared to the individual instrument models. Aggregated model superiority will be discussed in further detail for univariate and multivariate predictions.

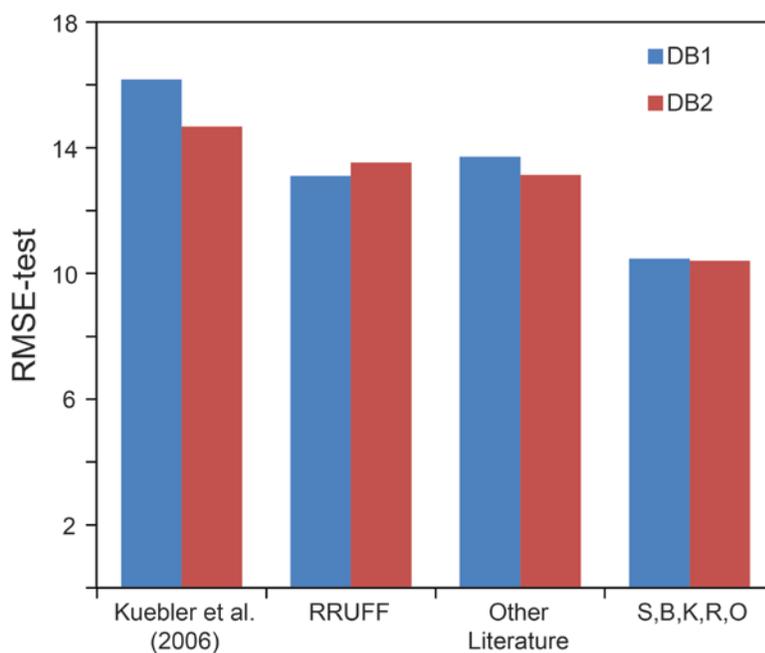


Figure 4.6. Summary of the univariate RMSE-test values. The four prediction models used include the Kuebler et al. (2006), RRUFF, other literature (**Table 2.3**), and an aggregated data set of Senterra (532 nm), BRAVO, Kuebler et al. (2006), RRUFF and the other literature data (listed as S,B,K,R,O). These four models were used to predict %Fo of combined Senterra 532nm and BRAVO data.

B. MULTIVARIATE (MACHINE LEARNING) RESULTS

Raman spectra acquired on the BRAVO and Senterra (532 and 785 nm) were utilized for multivariate predictions and modeling along with RRUFF data for which the full spectra are published online. No other publicly accessible olivine Raman spectra are published.

The three multivariate algorithms used within this research were PLS, lasso and LARS. For PLS, the number of components in each model is chosen using the lowest (best) prediction accuracy; generally this values was 4-7 components. Higher numbers of components produce more accurate predictions of %Fo but they reduce the generalizability of the model. This is comparable to the concept of using very large order polynomials to predict data – they can be quite accurate but not applicable to any other data sets.

For lasso models, an alpha value is chosen for each prediction to train the model depending on the desired “sparseness” of the model – i.e., how few channels the user desires to employ to predict %Fo. Alpha can be trained to various values, which changes the number of channels used by the model. As the alpha value increases, fewer channels are examined in the multivariate analysis prediction (**Figure 4.7**). As the number of channels in a model increases, the RMSE-CV decreases, showing the importance of models with a large number of channels (small alpha lasso models). However, use of large numbers of channels may over tune the model and reduce its generalizability.

The number of channels chosen for LARS modeling seeks to find the solution with the lowest error. As the number of channels increases for any given model, more channels are examined. However, because this model type performs least angle regressions, very few channels are examined. As an example, consider the Senterra data set, resampled from 800-880 cm^{-1} and baseline-removed (Air-PLS), normalized, and square root squashed (see next chapter for more information). If LARS is performed on:

1 channel → RMSE-CV is 30.35 %Fo

2 channels → RMSE-CV is 17.73 %Fo

3 channels → RMSE-CV is 17.67 %Fo

4 channels → RMSE-CV is 19.11 %Fo

5 channels → RMSE-CV is 20.27 %Fo

These results imply that the best predictions possible with LARS use only three channels of the spectrum, but it is also clear that the accuracy of that prediction is poor, especially when compared to the other multivariate methods.

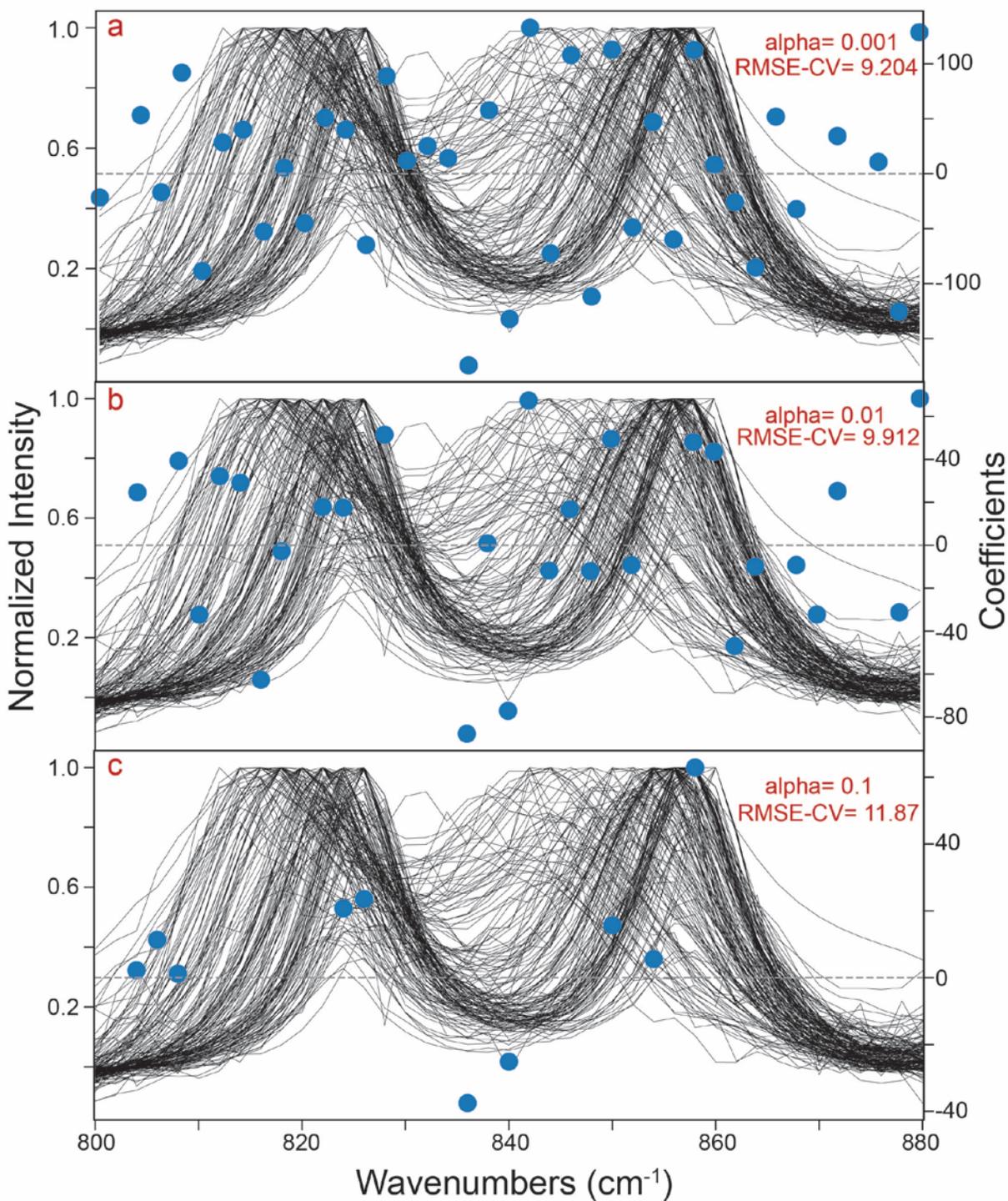


Figure 4.7. Lasso model coefficients. As the alpha value increases, fewer channels are examined: (a) 38 channels for $\alpha = 0.001$, (b) 27 channels for $\alpha = 0.01$, and (c) 10 channels for $\alpha = 0.1$. As the number of channels examined is decreased (fewer coefficients within the model), the RMSE-CV of the models increases in value (gets worse). This demonstrates the value of a models that examines a high number of channels, which is achieved in a small alpha value lasso model or PLS models.

Table 4.7. Multivariate predictions for internally cross-validated models using different combinations of available data sets, using a 3.0 cm^{-1} /channel resolution for all data for consistency.

Data	Model	Wavenumber range (cm^{-1})	Components/alpha	R^2	RMSE-CV
Senterra (532 nm), BRAVO	PLS	300-1500	5	0.92	8.48
Senterra (532 nm), BRAVO	Lasso	300-1500	0.1	0.92	8.64
Senterra (532 nm), BRAVO	LARS	300-1500	5	0.74	15.55
Senterra (532 nm), BRAVO	PLS	800-880	6	0.96	6.44
Senterra (532 nm), BRAVO	Lasso	800-880	0.01	0.93	8.02
Senterra (532 nm), BRAVO	LARS	800-880	3	0.72	16.15
BRAVO	PLS	800-880	7	0.99	2.56
BRAVO	Lasso	800-880	0.01	0.99	2.33
BRAVO	LARS	800-880	2	0.93	7.32
Senterra (532 nm)	PLS	800-880	7	0.97	5.52
Senterra (532 nm)	Lasso	800-880	0.01	0.93	8.52
Senterra (532 nm)	LARS	800-880	4	0.65	18.55
RRUFF	PLS	800-880	2	0.96	5.77
RRUFF	Lasso	800-880	0.004	0.99	2.44
RRUFF	LARS	800-880	3	0.84	10.79
Senterra (532 nm), BRAVO, RRUFF	PLS	800-880	6	0.95	6.86
Senterra (532 nm), BRAVO, RRUFF	Lasso	800-880	0.02	0.92	8.785
Senterra (532 nm), BRAVO, RRUFF	LARS	800-880	3	0.72	16.28

PLS and lasso generally outperform univariate methods for predictions in spectroscopic applications. LIBS (Tucker et al., 2010; Dyar et al., 2016a) and XAS (Dyar et al., 2012; Dyar et al., 2016b) spectroscopies follow this trend. PLS or lasso are the preferred techniques compared to the LARS method. LARS performs least angle regressions that typically rely on few channels within the spectra producing poor results. Second- and third-order polynomial fits have been demonstrated as preferable to linear regressions for modeling %Fo versus wavenumber. Therefore, as is the case for univariate models, relying on a linear regression for prediction produces poor results. For this reason, LARS results will be omitted from subsequent figures and analysis in this thesis.

Multivariate models can be set up to use all or any subset of the spectral data/wavenumber range. Two different approaches were tested in this thesis: using the whole spectrum (WS) from $300\text{-}1500 \text{ cm}^{-1}$ and doublet examination (DE) from $800\text{-}880 \text{ cm}^{-1}$. Note that the former method may produce better results when only pure olivines are being studied. However, in practice (as

when deployed on a planetary surface), it is far more likely that the olivine will be mixed in with other phases such as glass and other minerals. Those other phases would be expected to have bands somewhere in the whole spectral range, and hopefully not overlapping with the olivine doublet. Thus it is desirable to develop a method for predicting %Fo that isolates the olivine part of the spectra.

The first six lines of **Table 4.7** show the internal RMSE-CV results (i.e., leave one out cross-validation for a single group of data) for the combined BRAVO and Senterra data sets merged together, comparing the DE (800-880 cm^{-1}) to the whole spectrum (WS) methods. The DE method consistently performs better than the WS method. For example, the PLS and lasso RMSE-CV values are ± 8.48 and ± 8.64 (in units of %Fo) for whole spectrum range (300-1500 cm^{-1}), while the PLS and lasso RMSE-CV values for the DE model are ± 6.44 to ± 8.02 %Fo (**Table 4.7** and **Figure 4.8**). Reducing the width of the wavenumber range to examine solely the olivine doublet improves the accuracy of the %Fo predictions compared to utilizing a larger spectral range (**Figure 4.8**), likely because of instrument noise within the spectrum. Moreover, other types of features within the entire spectral range may degrade olivine predictions. Some other unwanted features within the spectra include sample heating, fluorescence and, cosmic spikes. These types of noise are not consistent for each spectrum and they do not relate to Raman features resulting from compositional variations. They thus result in small differences that detract from the capability of the model to accurately predict %Fo.

The bottom 15 lines of **Table 4.7** also show the relative accuracies of the BRAVO, Senterra, and RRUFF predictions individually and collectively for the DE models (800-880 cm^{-1}) only. Each of the three instruments' data sets predicts best internally, i.e., BRAVO model predicting BRAVO data. This is to be expected because the instrument and operating conditions are

identical. This same trend was observed in univariate predictions (**Figure 4.3**). For example, the combined BRAVO and Senterra data sets had PLS and lasso RMSE-CV values of ± 6.44 and ± 8.02 respectively for the DE range (**Figure 4.8**), while their individual, independently, produced PLS and lasso RMSE-CVs are ± 2.56 and ± 2.33 (BRAVO) and ± 5.52 and ± 8.52 (Senterra) (**Figure 4.9**) (here the 8.52 value for the Senterra is an outlier). These data sets were acquired on different instruments and therefore have different detector sensitivities, resolutions and laser wavelengths. Additionally, the BRAVO data set ($n=25$) is smaller than the Senterra data set ($n=68$). Although it is possible that the BRAVO data provide a better prediction model than the Senterra data set examined in **Figure 4.9**, other factors may impact this comparison and will be discussed further in the following chapter.

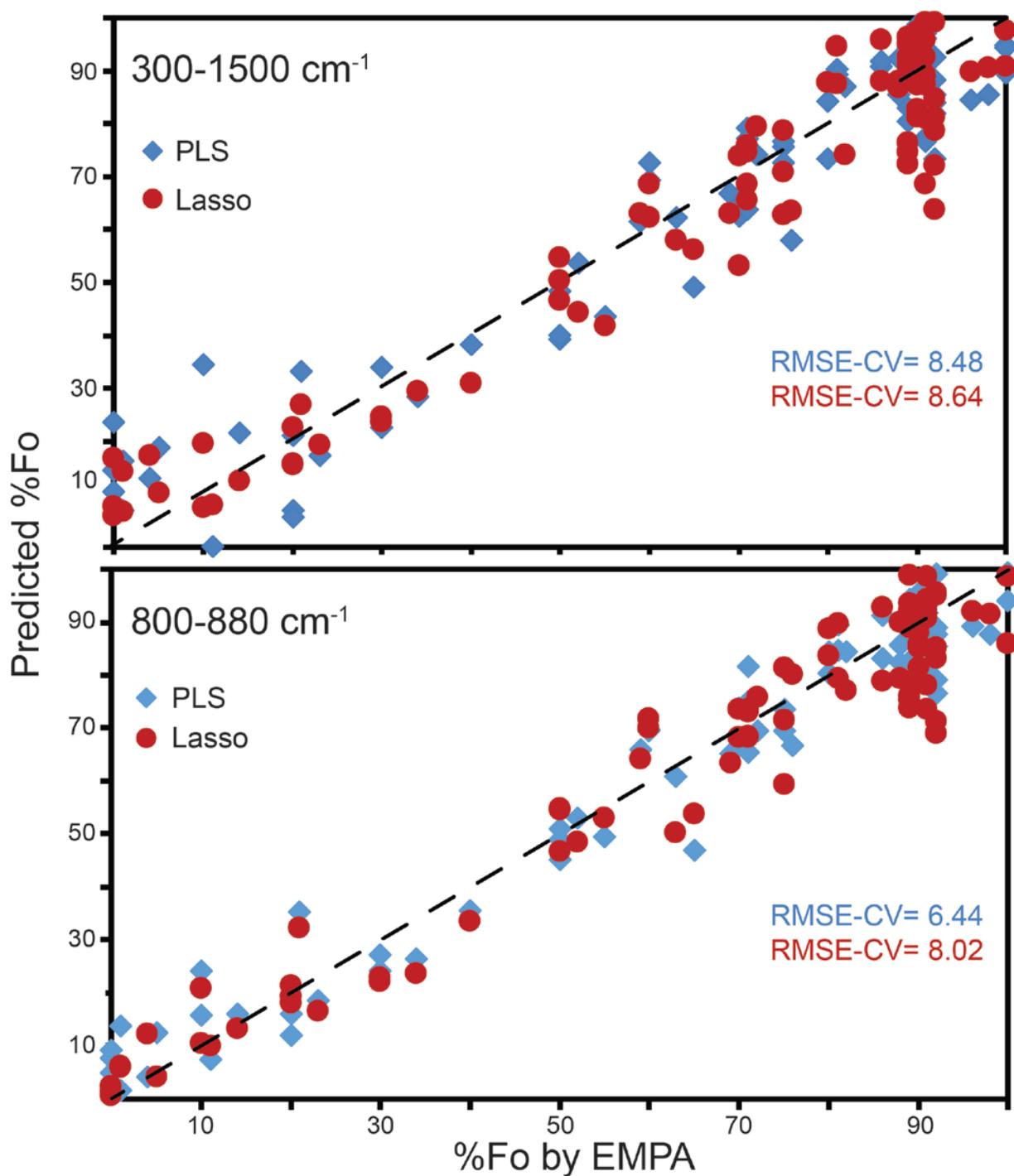


Figure 4.8. PLS and lasso predictions of Fo content using wavenumber ranges (top) 300-1500 cm⁻¹ and (bottom) 800-880 cm⁻¹. Spectra acquired on Bruker's BRAVO and Senterra (532 nm) were utilized for these predictions ($n = 93$). Error bars on EMPA are smaller than the symbols; error bars on %Fo predictions are given in red and blue text.

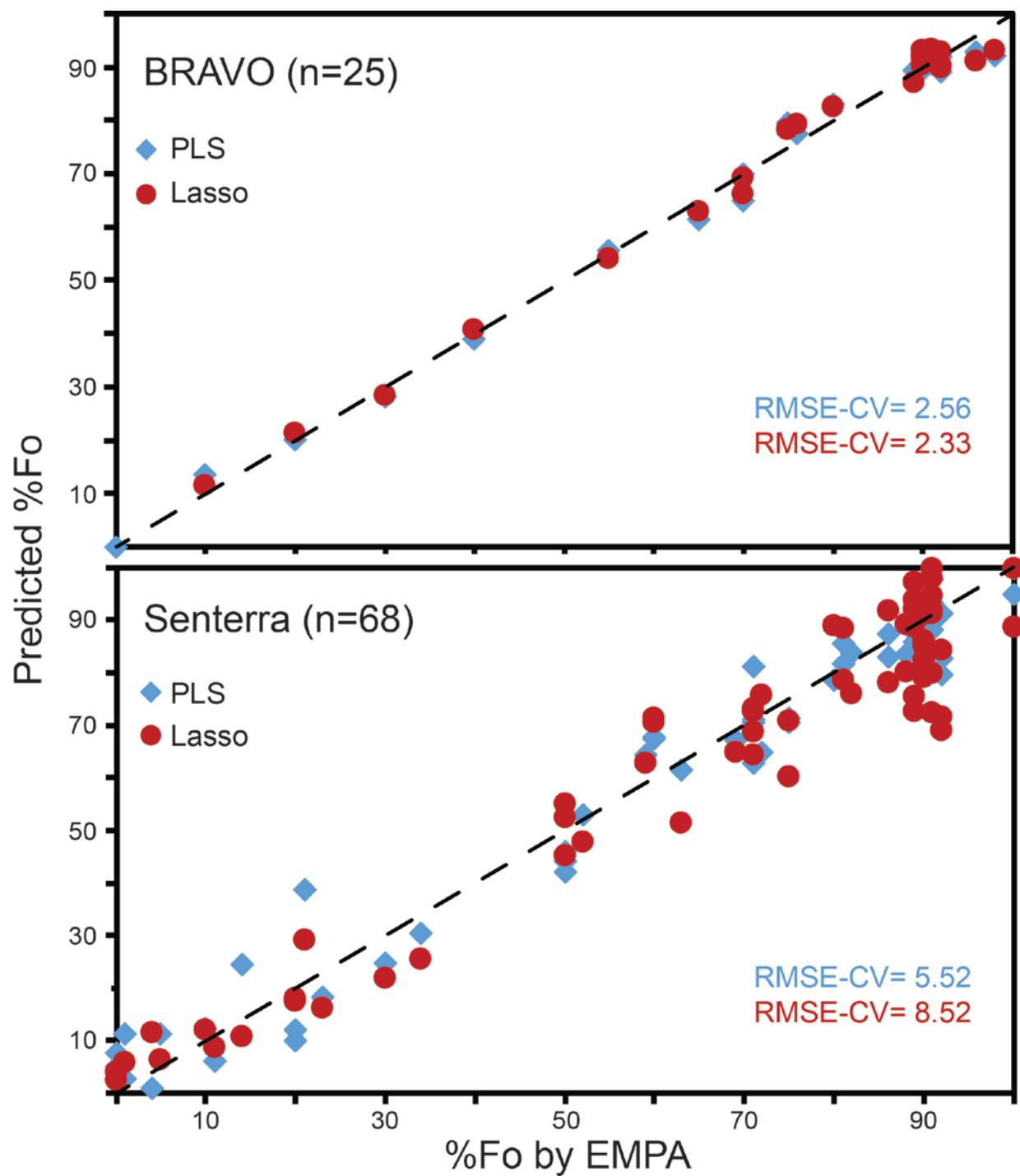


Figure 4.9. PLS and lasso predictions of Fo for samples run on (top) the BRAVO ($n=25$) and (bottom) the Senterra (532 nm) ($n=68$). A wavenumber range of 800-880 cm^{-1} was used in all predictions. Error bars on EMPA are smaller than the symbols; error bars on %Fo predictions are given in red and blue text.

Table 4.8 takes the models from **Table 4.7** (BRAVO only, Senterra only, RRUFF only, Senterra+BRAVO, and all) and uses them to predict one another. The models and data utilized here were explicitly set to have the same spectral resolution of 3.0 cm^{-1} /channel to examine the predictions with a constant resolution. **Table 4.8** again demonstrates that PLS and lasso have good predictive performance. Within this table, the internally cross-validated prediction models (shaded gray) consistently produce the smallest RMSE-test values for each data set with only one exception. It follows that the predictions of data acquired on a different instrument are generally less successful, which is not surprising.

Table 4.8. RMSE-test values compared with internally-cross-validated values (in shaded cells) acquired with a wavenumber range of $800\text{-}880 \text{ cm}^{-1}$ and resolution of 3.0 cm^{-1} /channel. Here, each of five dataset is used to predict %Fo in the other four.

PLS Model:					
Predict:	Senterra (532 nm)	BRAVO	RRUFF	Senterra (532 nm), BRAVO	All
Senterra (532 nm)	5.52	11.91	11.28	6.47	6.57
BRAVO	13.01	2.56	11.44	6.37	7.03
RRUFF	3.61	10.12	5.77	10.25	8.04
Senterra (532 nm), BRAVO	8.23	10.27	11.32	6.44	6.69
All	8.295	10.68	10.83	7.006	6.86
Lasso Model:					
Predict:	Senterra (532 nm)	BRAVO	RRUFF	Senterra (532 nm), BRAVO	All
Senterra (532 nm)	8.52	15.15	39.83	8.86	9.36
BRAVO	8.45	2.33	33.93	5.02	4.79
RRUFF	15.3	14.22	2.44	16.14	11.47
Senterra (532 nm), BRAVO	9.23	13.02	38.33	8.02	8.38
All	10.11	13.17	36.09	9.308	8.79

Tables 4.9 and **4.10** replicate what was in **Tables 4.7** and **4.8** except the source data used were in their as acquired resolutions, which are slightly variable. The BRAVO, Senterra (532 nm) and RRUFF data sets have spectral resolutions of 2.0, 0.5 and 0.48 cm^{-1} /channel respectively. When evaluating RMSE-test values within **Table 4.10**, only Senterra (532 nm), BRAVO, and RRUFF models were used because aggregated data sets would have various spectral resolutions, and the resolution must be uniform for the multivariate analyses. **Table 4.10** again shows that predicting data from one instrument with a model trained on a different spectrometer produces less accurate results. These results are also shown graphically in **Figure 4.10**.

Table 4.9. Multivariate predictions for internally cross-validated models using a wavenumber range of 800-880 cm^{-1} with the highest spectral resolution per spectrometer. The BRAVO, Senterra and RRUFF spectra have resolutions of 2.0, 0.5 and 0.48 cm^{-1} /channel.

Data	Model	Components/ alpha	R ²	RMSE-CV
BRAVO	PLS	8	0.992	2.611
BRAVO	Lasso	0.005	0.996	1.883
Senterra (532 nm)	PLS	7	0.965	5.845
Senterra (532 nm)	Lasso	0.008	0.945	7.371
RRUFF	PLS	4	0.991	2.617
RRUFF	Lasso	0.01	0.916	7.903

Table 4.10. RMSE-test values for models compared with internally-cross-validated values (in shaded cells) with a wavenumber range of 800-880 cm^{-1} and as acquired resolutions.

PLS models:			
Predict:	Senterra (532 nm)	BRAVO	RRUFF
Senterra (532 nm)	5.845	13.77	15.07
BRAVO	14.18	2.611	18.22
RRUFF	12.14	9.629	2.617
Lasso models:			
Predict:	Senterra	BRAVO	RRUFF
Senterra (532 nm)	7.371	17.68	18.05
BRAVO	17.05	1.883	9.327
RRUFF	15.89	18.31	7.903

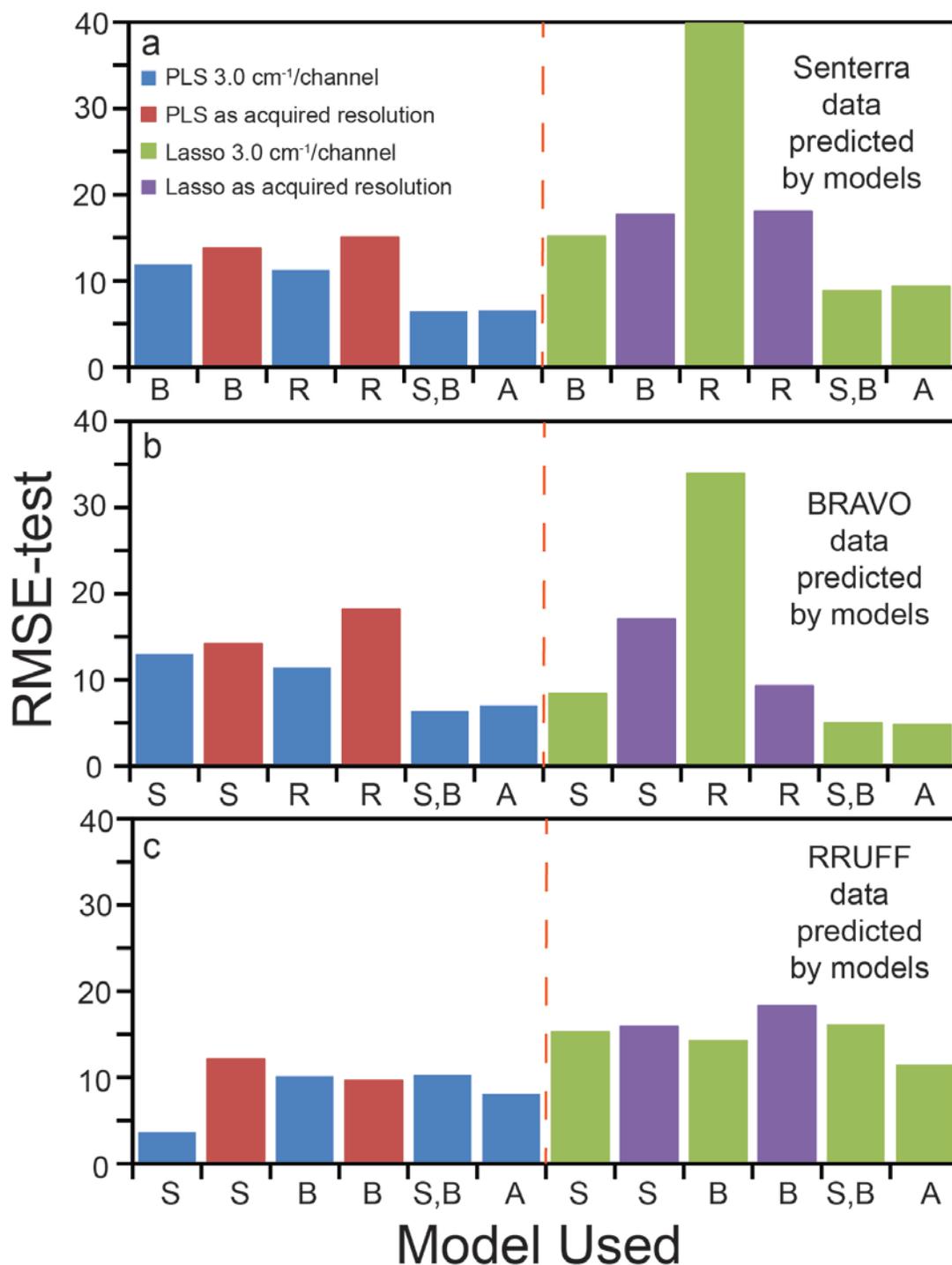


Figure 4.10. RMSE-test values from Tables 4.8 and 4.10 that predict %Fo in each of the data sets using (a) Senterra data as the held-out test set, (b) BRAVO data as the held-out test set and (c) RRUFF data as the held-out test set predicted through PLS and lasso. The models included Senterra (S), BRAVO (B), RRUFF (R), Senterra and BRAVO (S,B) and all three data sets together (A). Either a $3.0 \text{ cm}^{-1}/\text{channel}$ resolution was utilized or the as acquired resolution was used where the BRAVO, Senterra and RRUFF spectra have resolutions of 2.0 , 0.5 and $0.48 \text{ cm}^{-1}/\text{channel}$ respectively.

C. CONTRASTING UNIVARIATE AND MULTIVARIATE RESULTS

Detector sensitivity varies among instruments. To evaluate this effect, 25 samples (that happen to be the same 25 samples run on the BRAVO) were run on the Senterra using the 532 nm and 785 nm lasers (**Table 3.1** and **3.2**). We then predicted the %Fo content of these samples (**Table 4.11**), deliberately keeping resolution constant.

Table 4.11. Multivariate predictions for the spectral wavenumber range of 800-880 cm^{-1} and resolution of 0.5 $\text{cm}^{-1}/\text{channel}$. A re-examination of 25 samples (previously referenced as BRAVO samples).

Data	Model	Components/ alpha	R ²	RMSE-CV
Senterra 532nm	PLS	6	0.99	2.56
Senterra 532nm	Lasso	0.001	1.00	0.49
Senterra 785nm	PLS	6	0.99	2.63
Senterra 785nm	Lasso	0.001	0.99	2.86

For all these data, the sensitivity of the detector was identical; only the laser wavelength changed. Even the specific sampling spot remained constant. This is important because if the two spectra were acquired in two different places on the sample, slight compositional differences may be present.

The spectra that correspond with these models are shown in **Figure 4.11**. Peak centroid position was considerably different for these two data sets (Senterra 532 and 785 nm) yet, as **Table 4.11** depicts, there are minimal differences between the multivariate models. This again demonstrates that multivariate predictions would be successful with this data set whereas univariate predictions would be poor.

The centroids of the Senterra 785 nm data are lower in wavenumber (cm^{-1}) value compared to the Senterra 532 nm data. Laser wavelength calibration (how the instrument does its calibrations at the two wavelengths) may be affecting the peak centroid positions within these samples, demonstrating instrument biases, but these are not specifically related to detector sensitivity. Although RMSE-CV multivariate values differ between the Senterra 532 nm and 785 nm data

sets, the values are comparable and small in every case except for the 532 nm lasso prediction (**Table 4.11**).

The improvement in prediction accuracy from the lasso model for the Senterra 532 nm data may imply that the lower wavelength laser energy does a better job of identifying forsterite composition. This follows the suggestion of Bartholomew et al. (2015) who note that Raman scattering intensity varies with the fourth power of laser wavelength. In other words, Raman peak intensity is roughly $5\times$ better for the 532 nm data, and this increase in signal to noise likely improves the peak resolution of the olivine features in the spectra. For this reason, it seems likely that the increased accuracy resulting from use of the 532 nm laser is real, though this conclusion could bear further testing.

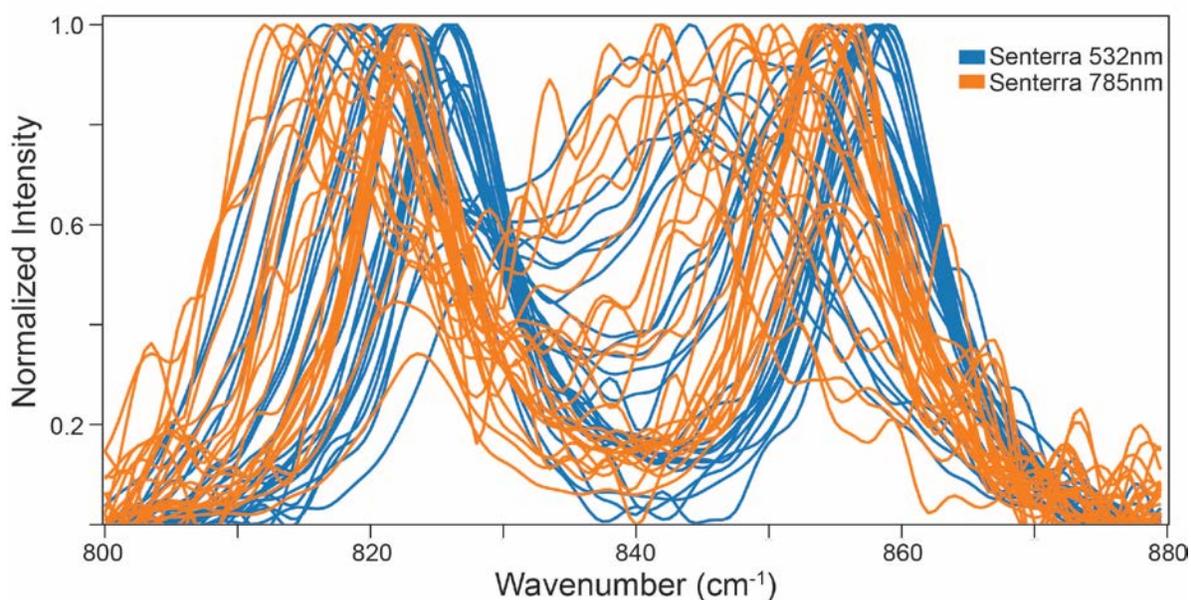


Figure 4.11. Doublet of 25 sample spectra acquired on both the Senterra 532 (blue) and 785nm (orange). It is apparent that the Senterra 785 nm data are shifted to slightly lower wavenumbers due to instrument calibration.

In summary, **Table 4.12** reports RMSE-test values for univariate and multivariate analyses previously discussed to allow comparisons to be made between these methods. For univariate predictions, second order polynomial fits from **Tables 4.5** and **4.6** of RRUFF, BRAVO, Senterra

(532 nm), and aggregated data were used to predict the combined Senterra (532 nm) and BRAVO data sets. Univariate predictions made using models built with published centroid data are also reported. Finally, PLS and lasso models were utilized for all possible multivariate predictions.

Table 4.12. Summary of univariate and multivariate RMSE values of the predictions from individual models (columns) using the combined BRAVO and Senterra (532nm) data sets for testing. Univariate predictions utilized second order polynomial fits from **Tables 4.5** and **4.6**. Multivariate predictions used models with a wavenumber rang of 800-880 cm^{-1} and 2.0 cm^{-1} /channel spectra resolution.

Model	RRUFF	Kuebler et al. (2006)	BRAVO	Other literature (Table 4.4)	Senterra (532nm)	*Combine d data	*All data
# Samples in model	12	13	25	47	68	105	155
# Samples in model & data	0	0	25	0	68	93	93
Univariate DB1	13.10	16.17	12.41	13.71	11.29	10.46	n.a.
Univariate DB2	13.52	14.67	15.98	13.14	14.28	10.40	n.a.
Multivariate PLS	11.14	n.a.	12.11	n.a.	7.99	6.08	8.65
Multivariate lasso	20.73	n.a.	13.59	n.a.	8.13	7.19	8.97

*Combined data references models that includes the BRAVO, RRUFF and Senterra 532nm data. All data references models that include BRAVO, RRUFF, Senterra 532 nm and Senterra 785 nm data.

PLS and lasso multivariate predictions are generally superior to univariate DB1 and DB2 predictions because they have smaller RMSE-test values, especially in data sets with more than a handful of samples. Generally, as the data sets are predicted with larger models (going from left to right within **Table 4.12**), the RMSE values decrease in size. This trend is present for both univariate and multivariate RMSE indicating prediction success with aggregated data set models. The gray shaded cells within **Table 4.12** correspond to recommended models listed within the appendix of this thesis. The rationale for the recommendation of the models shaded in gray of **Table 4.12** will be discussed in greater detail.

Chapter 5

DISCUSSION AND CONCLUSIONS

The goal of this thesis is to develop a generalizable method for measuring the composition of olivine, a common rock-forming mineral, to serve the needs of upcoming planetary exploration using Raman spectroscopy. Results from the previous chapter demonstrate that both the conventional univariate methods based on the positions of the most prominent olivine doublet and multivariate analyses produce reliable robust predictions of olivine composition. But the latter, particularly the lasso, produces optimum accuracy.

Within this chapter, elements that effect univariate and multivariate analyses are discussed. These include pre-processing effects such as baseline removal, normalization, squashing and smoothing. A discussion on the effects of the wavenumber range and spectral resolution of the multivariate method is also presented. Additionally, the importance of predicting “unseen data” is discussed. An argument for using multivariate analyses over univariate analyses is explored. A step by step process for creating multivariate models is presented within the conclusion section of this thesis. Finally, the major conclusions of this thesis are summarized.

A. EFFECTS OF PRE-PROCESSING

As discussed by Carey et al. (2015), pre-processing steps for Raman spectra are important for building prediction models. Within this paper, the effects of several types of spectral pre-processing (performed prior to a prediction) were tested for prediction accuracy; these included baseline removal, normalization, squashing, and smoothing.

Baseline removal is used to remove the signal within the spectra that is not important to the characteristic bands to show only useful information for predictions. Carey et al. (2015) and subsequent papers (Dyar et al., 2016c; Giguere et al., 2017) have experimented with automated

baseline removal techniques such as the Dietrich methods (Dietrich et al., 1991), asymmetric least squares (ALS; Eilers and Boelens, 2005), fast and automatic baseline correction (FABC; Cobas et al., 2006), wavelet transforms (Kajfosz and Kwiatek, 1987), and adaptive iteratively reweighted quantile regression (airPLS; Zhang et al., 2010). Carey et al. (2015) found that AirPLS worked best for predicting the class, type, group and species of a mineral, and so this technique was used in this thesis.

Normalization to total spectral intensity was also used consistently in this thesis, but other methods for normalization have been proposed and bear further testing. It is known that intensity varies with laser wavelength, sample composition, differences in sample crystal orientation, focus, and other instrumental parameters. So in future work, we plan to test the effects of other methods for normalization, which will include normalizing to the maximum value (L^∞ norm), the sum of absolute values ($L1$ norm), the sum of squared values ($L2$ norm), and scaling to intensity at a specific energy.

Squashing and smoothing were used by Carey et al. (2015) to apply different types of nonlinear, monotonic pre-processing before Raman spectral analysis. “Squashing” uses a transformation function f that is applied to each wavelength of a spectrum independently to produce a new spectrum with compressed/reduced distances between strong and weak spectral features. Square root squashing uses $f(x) = \sqrt{x}$, while sigmoid squashing uses $f(x) = 1 - \cos(\pi x)$. Each squashing or normalizing preprocessing step preserves the relative ordering of intensity values while mitigating the effect of peak intensity differences. Smoothing by techniques such as the Savitsky-Golay (Savitsky and Golay, 1964) filter are also useful to reduce noise and enhance spectral peaks.

Time did not permit testing of these pre-processing techniques on our own data sets, though this is obviously an area ripe for research. So in future work, we plan to test the effects of the baseline removal methods noted above, other methods for normalization, and squashing and smoothing techniques. We are hopeful that when these steps are optimized, our accuracy for prediction of %Fo from Raman spectra of olivine will be improved.

B. IMPORTANCE OF TESTING ON UNSEEN DATA

Univariate %Fo predictions made within this thesis have comparable RMSE-CV values to published models (**Figure 4.4**), all of which also use univariate methods. However, previous workers did not test their univariate prediction models on “unseen data,” so their claims of accuracy are unsubstantiated for application to other data sets such as those on Mars. In this thesis, all models were both internally cross-validated and tested on “unseen data”. The error on these secondary predictions was denoted as RMSE-test values. The univariate RMSE-test values are summarized in **Figure 4.6**. For univariate predictions, RMSE-test values are greater (i.e., the errors are larger, indicating they are less accurate) compared to the RMSE-CVs of the original model. However, only accuracies tested on “unseen data” can evaluate the accuracy of a prediction with data acquired under different operating conditions such as those on Mars, with a different instrument design, or with a non-identical laser wavelengths. Thus the accuracies provided in this thesis for “unseen data” are (for now) the best predictions for accuracy in application on Mars (i.e., SHERLOC).

As demonstrated by **Figure 4.11**, multivariate analysis is necessary for %Fo predictions across data sets because peak centroids are affected by laser wavelength. **Tables 4.8** and **4.10** demonstrate that each model is predicted best by itself. In other words, the prediction model predicts its own data best. This is expected seeing as detector sensitivity, resolution and

wavelength of the laser would be constant within a given data set. However, as previously discussed, prediction of “unseen data” is important to predict data acquired on alternative spectrometers. Given this need, using aggregated data sets is recommended (**Figure 4.10**).

Table 4.12 provides a summary of RMSE-test values for both univariate and multivariate predictions. The prediction with the smallest error in **Table 4.12** is the PLS combined data model (BRAVO, Senterra 532nm and RRUFF) with an RMSE-test value of ± 6.08 . The models that have all spectral data (BRAVO, RRUFF, Senterra 532 nm and 785 nm) have greater RMSE-test values ± 8.65 (PLS) and ± 8.97 (lasso) compared to the combined data set. However, it is important to note that the combined data model consists of only twelve spectra that are not also present within the prediction data set in question. Yet again, the issue of “unseen data” arises. Because this is the first development of %Fo Raman multivariate models, it is difficult to evaluate the success of models on “unseen data”. The models highlighted gray in **Table 4.12** are the recommended models for future workers. The coefficients of these models are included in the appendix section of this thesis. Despite the fact that these models have slightly larger RMSE-test values than the combined data set models, their prediction ability is more transferable to other instruments because they include a more diverse group of spectra.

C. USEFULNESS OF MULTIVARIATE ANALYSIS

It is quickly apparent that the multivariate methods outperform the univariate regression models (**Table 4.12**); this result is also observed in LIBS (Tucker et al., 2010; Dyar et al., 2016a) and XAS (Dyar et al., 2012; Dyar et al., 2016b) spectroscopies, among many others. As previously discussed, the olivine doublet examined within this study is the result of five vibrational modes ($2A_g + 2B_{1g} + B_{2g}$) (**Table 2.2**). Using solely the peak centroid to model olivine composition does not rely on other characteristics within the spectrum such as band

shape, intensity, FWHM, and area. Multivariate analyses account for these features resulting from the five doublet vibrational modes that are lost with univariate methods. The univariate prediction method is therefore not recommended. However, if future workers use this method, it is recommended that an aggregated data set be utilized (**Figure 4.6**).

PLS and lasso examine multiple channels within the spectra to build a %Fo prediction model. The number of coefficients per model is based on the assigned number of components based on the alpha value. For lasso predictions, fewer channels are examined in multivariate analysis prediction as the alpha value increases (**Figure 4.7**). As the number of channels in a lasso model increases, RMSE-CV decreases, showing the importance of models with a large number of channels (small alpha lasso models). However, there is a trade-off between the generalizability of the model that is optimized by smaller numbers of channels versus improved accuracy from using larger numbers of channels.

D. CHOICE OF WAVENUMBER RANGE

All multivariate predictions are conducted over a range of channels in each spectrum. In the case of Raman spectroscopy, a wavenumber range must be selected. The effectiveness of %Fo predictions was evaluated using WS (300-1500 cm^{-1}) and DE (800-880 cm^{-1}) methods. **Figure 4.8** examined this spectral range issue. The DE method produced smaller RMSE-CVs compared to the WS method (**Table 4.7**). It is recommended that future models utilize a reduced spectral range that examines solely the doublet (800-880 cm^{-1}). Another reason that the DE method should be utilized for building prediction models is because for rocks (multiple species), the WS method is impractical. This is because each individual component will give rise to several bands within the Raman spectrum that are difficult to quantitatively detangle.

E. EFFECTS OF SPECTRAL RESOLUTION

A comparison of spectral resolution was conducted within this thesis to evaluate its effect on the accuracy of prediction models. Spectral data and models were evaluated under 3.0 $\text{cm}^{-1}/\text{channel}$ resolution and the resolution the spectra were acquired under. The BRAVO, Senterra (532 nm) and RRUFF data sets have resolutions of 2.0, 0.5 and 0.48 $\text{cm}^{-1}/\text{channel}$ respectively. **Figure 4.10** visually depicts the RMSE-test values of these two resolution methods. There are differences between these two groups yet, the trend is difficult to describe. Depending on the data and model each method is potentially favorable. Often the difference between the two resolution groups are insignificant, with several lasso exceptions, allowing us to conclude that resolution has a minimal affect for these groups.

It is also recommended for future workers to publish and add Raman spectral data to these recommended models. Increasing the number of spectra within the models on additional instruments will likely improve the accuracy of prediction results. Because multivariate analyses rely on significant portions of spectra, providing raw data is assistive to the scientific community. All Raman spectra within this thesis are posted on the website nemo.umass.cs.edu:54321 (Carey et al., 2017). Additionally, beyond spectra within this thesis, over 4,000 Raman spectra have been catalogued here.

F. CONCLUSION

It has been demonstrated that multivariate analyses are superior to univariate band shift predictions. The recommended process to follow for the most accurate %Fo prediction using Raman spectroscopy is as follows.

- (1) *Acquire Raman spectra*: Acquire Raman spectra under the highest resolution possible.

This may mean that the wavenumber range will be reduced. If the wavenumber range

extends to roughly 1500 cm^{-1} , olivine and the majority of common rock forming mineral bands will be accounted for.

- (2) *Pre-processing*: Perform baseline removal (AirPLS) and normalization of all spectral data prior to constructing a %Fo model.
- (3) *Examine the spectrum*: Prior to making a %Fo prediction, examine the Raman spectrum to identify three or more diagnostic olivine bands. Samples often have closely associated mineral phases. It is important to identify the mineral prior to the Fo-Fa prediction. Additionally, evaluate whether two mineral phases are present within the same spectra (i.e., olivine and pyroxene). If multiple phases are present, the error on the Fo-Fa prediction may be higher than reported within this thesis.
- (4) *Select spectral range of the model*: In the case of predicting Fo-Fa content, use the DE method to isolate the greatest intensity bands DB1 and DB2 that are encompassed within the wavenumber range $800\text{ to }880\text{ cm}^{-1}$.
- (5) *Assign spectral resolution to the model*: If %Fo predictions were acquired under the highest resolution available, spectral resolution can be reduced to match the resolution of the model.
- (6) *Select model type*: Use a multivariate PLS or lasso aggregated prediction model. A model that incorporates more spectra acquired under different conditions (i.e., laser wavelength) will improve the accuracy of all predictions.
- (7) *Training the model*: Generally, the greater number of channels examined within a multivariate prediction results in smaller error. It is recommended to train the model to examine as many channels as possible while maintaining small MSE.

(8) *Improve the model*: Publish and incorporate olivine Raman spectra (with secondary composition confirmation i.e., EMPA) into multivariate models to improve the accuracy of future predictions.

In conclusion, four major accomplishments have been made within this thesis.

- (1) 155 new spectra have been made publicly available on nemo.umass.cs.edu:54321 (Carey et al., 2017) of 93 olivine samples.
- (2) Error was evaluated for univariate and multivariate predictions on “unseen data”, which is an issue previous workers have ignored. Whether univariate or multivariate analyses are deemed useful, **Table 4.12** gives a quantitative estimate of resultant errors on %Fo predictions.
- (3) It was determined that multivariate predictions typically produce smaller RMSE-CVs and RMSE-test values when compared to univariate predictions.
- (4) The coefficients of %Fo multivariate PLS and lasso prediction models are reported within the appendix of this thesis for future workers.

Olivine is an important rock-forming mineral group. Yet, there are other mineral groups that will need to be evaluated and quantitatively described through Raman spectroscopy. Creating equivalent multivariate Raman models for mineral groups such as pyroxene and feldspar is a direction for future research. Additionally, bringing these mineral group analyses together to evaluate mineral mixtures through Raman spectroscopy is a difficult issue that will need to be addressed in future work.

APPENDIX

Recommended PLS and lasso model coefficients of an aggregated data set (RRUFF, BRAVO, Senterra 532nm and 785nm) with the wavenumber range of 800-880 cm^{-1} . Spectral resolution of 2.0 cm^{-1} /channel was used.

PLS		Lasso	
Component	4	alpha	0.001
R ²	0.92	R ²	0.91
RMSE-CV	±8.23	RMSE-CV	±9.204
Wavenumber (cm^{-1})	%Fo	Wavenumber (cm^{-1})	%Fo
800	7.60476	858	112.467
802	7.35613	824	40.1726
804	4.88258	840	-132.066
806	4.65443	836	-174.37
808	4.03649	850	112.831
810	-0.138954	804	53.2794
812	0.00553941	808	91.9573
814	-2.35866	828	88.7236
816	-5.78988	816	-52.619
818	-1.68219	860	8.38199
820	-1.85993	812	28.4459
822	-0.165762	880	128.72
824	5.14083	842	132.832
826	4.90716	864	-85.3615
828	2.33167	862	-25.9238
830	-2.56251	814	40.0891
832	-10.5647	846	107.946
834	-15.6866	872	34.319
836	-20.9845	878	-125.442
838	-21.7018	822	50.7902
840	-21.9789	810	-88.5651
842	-14.8889	868	-32.2702
844	-12.1839	844	-72.5025
846	-0.468039	852	-48.7609
848	6.94607	848	-111.78
850	9.4293	838	57.7587
852	7.25251	826	-64.7224
854	3.92042	854	47.0273
856	-0.431912	830	11.774
858	3.19703	856	-59.904
860	7.17296	832	25.2158
862	3.85628	800	-21.6578
864	2.11808	866	51.8513
866	3.11679	820	-45.069
868	1.78301	876	10.7031
870	-2.42815	834	14.1753
872	2.30592	806	-16.9617
874	1.7329	818	5.48771
876	-0.428335		
878	-3.16827		
880	-2.03983		

REFERENCES CITED

- Abdu, Y.A., Varela, M.E., and Hawthorne, F.C. (2011) Raman, FTIR, and Mössbauer spectroscopy of olivines from the D'Orbigny meteorite. Annual Meteoritical Society Meeting, 74, Abstract no.5112.
- Banfield, J.M., Dyar, M.D., and McGuire, A.V. (1992) The defect microstructure of oxidized mantle olivine from Dish Hill, California. *American Mineralogist*, 77, 959-975.
- Bartholomew, P.R., Dyar, M.D., and Brady, J.B. (2015) The role of intensity and instrument sensitivity in Raman mineral identification. *Journal of Raman Spectroscopy*, 46, 889-893.
- Beegle, L.W., Bhartia, R., DeFlores L., Darrach, M., Kidd, R.D., Abbey, W., Asher, S., Burton, A., Clegg, S., Conrad, P.G., Edgett, K., Ehlmann, B., Langenhorst, F., Fries, M., Hug, W., Nealson, K., Popp, J., Sobron, P., Steele, A., Wiens, R., and Williford, K. (2014) SHERLOC: Scanning Habitable Environments with Raman & Luminescence for Organics & Chemicals, an investigation for 2020. International GeoRaman Conference, 11, Abstract no.5101.
- Besson, J.M., Pinceaux, J.P., and Anastopoulos, C. (1982) Raman spectra of olivine up to 65 kilobars. *Journal of Geophysical Research*, 87, 10,773-10,775.
- Berlanga, G., Dyar, M.D., Breitenfeld, L.B., Wagnor, C., Hanlon, A., Bartholomew, P., Sharma, S.K., and Misra, A.K. (2017) Detection limits for silicates in Raman spectra of mixtures with volcanic glass. Lunar and Planetary Science Conference, 48, Abstract no.2255.
- Bolfan-Casanova, N., Montagnac, G., and Reynard, B. (2014) Measurement of water content in olivine using Raman spectroscopy. *American Mineralogist*, 99, 149-156.
- Boucher, T.F., Ozanne, M.V., Carmosino, M.L., Dyar, M.D., Mahadevan, S., Breves, E.A., Lepore, K.H., and Clegg, S.M. (2015) Nine machine learning regression methods for major elemental analysis of rocks using laser-induced breakdown spectroscopy. *Spectrochimica Acta Part B*, 107, 1-10.
- Breitenfeld, L.B., Dyar, M.D., Carey, C.J., Bartholomew, P., Tague, T.J., Wang, P., Mertzmann, S., Byrne, S.A., Crowley, M.C., Leight, C., and Watts, E. (2016) Quantifying mineral abundances in mixtures using Raman spectroscopy: calculating Raman coefficients using a diamond reference. Lunar and Planetary Science Conference, 47, Abstract no.2186.
- Brown, G.E. (1980) Olivines and silicate spinels. *Reviews in Mineralogy and Geochemistry*, 5, 275-381.
- Burns, R.G. (1993) *Mineralogical applications of crystal field theory*, 551 p. Cambridge University Press, New York.
- Butler, N.A., and Denham, M.C. (2000) The peculiar shrinkage properties of partial least squares regression. *Journal of the Royal Statistical Society*, 62, 585-593.

- Byrne, S.A., Dyar, M.D., Bessette, E.E., Breitenfeld, L.B., Crowley, M.C., Hoff, C.M., Marchand, G.J., Ketley, M.N., Roberts, A.L., Sklute, E.C., and Parente, M. (2015) Pure mineral separates for mixing experiments to simulate planetary surfaces. Lunar and Planetary Science Conference, 46, Abstract no.1499.
- Calistru, D.M., Demos, S.G., and Alfano, R.R. (1995a) Anharmonic effects and second neighbor interactions of local modes in Cr⁴⁺ doped forsterite probed by higher-order resonance Raman scattering. *Physical Review B*, 52, 15253-15260.
- Calistru, D.M., Demos, S.G., and Alfano, R.R. (1996) Direct observation of second neighbor interactions for Cr⁴⁺ doped forsterite by resonance Raman scattering. *Applied Physics Letters*, 68, 2207-2209.
- Calistru, D.M., Wang, W.B., Petricevic, V., and Alfano, R.R. (1995b) Resonance Raman scattering in Cr⁴⁺ doped forsterite. *Physical Review B*, 51, 14980-14986.
- Carey, C., Dyar, M.D., Boucher, T., and Giguere, S. (2017) Web-based software for preprocessing, matching, fitting, prediction, and visualization of spectroscopic data: the data exploration, visualization, and analysis of spectra (Devas) website. Lunar and Planetary Science Conference, 48, Abstract no.1097.
- Carey, C., Dyar, M.D., Boucher, T., Giguere, S., Hoff, C.M., Breitenfeld, L.B., Parente, M., Tague, T.J., Wang, P., and Mahadevan, S. (2015) Baseline removal in Raman spectroscopy: optimization techniques. Lunar and Planetary Science Conference, 46, Abstract no.2464.
- Chopelas, A. (1991) Single crystal Raman spectra of forsterite, fayalite, and monticellite. *American Mineralogist*, 76, 1101-1109.
- Clegg, S.M., Wiens, R.C., Maurice, S., Gasnault, O., Sharma, S.K., Misra, A.K., Newell, R., Forni, O., Lasue, J., Anderson, R.B., Nowak-Lovato, K.L., Fouchet, T., Angel, S.M., Rull, F., Johnson J.R., and the SuperCam Science Team (2015) Remote geochemical and mineralogical analysis with SuperCam for the Mars 2020 rover. Lunar and Planetary Science Conference, 46, Abstract no.2781.
- Cobas J.C., Bernstein, M.A., Martín-Pastor, M., and Tahoces, P.G. (2006) A new general-purpose fully automatic baseline-correction procedure for 1D and 2D NMR data. *Journal of Magnetic Resonance*, 183,145-151.
- Demos, S.G., and Alfano, R.R. (1995) Subpicosecond time-resolved Raman investigation of optical phonon modes in Cr-doped forsterite. *Physical Review B*, 52, 987-996.
- Dietrich, W., Rüdél, C., and Neumann, M. (1991) Fast and precise automatic baseline correction of one- and two-dimensional nmr spectra. *Journal of Magnetic Resonance* 91, 1-11.
- Durben, D.J., McMillan, P.F., and Wolf, G.H. (1993) Raman study of the high-pressure behavior of forsterite (Mg₂SiO₄) crystal and glass. *American Mineralogist*, 78, 1143-1148.

- Dyar, M.D. (2003) Ferric iron in SNC meteorites as determined by Mössbauer spectroscopy: Implications for Martian landers and Martian oxygen fugacity. *Meteoritics & Planetary Science*, 38, 1733-1752.
- Dyar, M.D., McGuire, A.V., and Ziegler, R.D. (1989) Redox equilibria and crystal chemistry of coexisting minerals from spinel lherzolite mantle xenoliths. *American Mineralogist*, 74, 969-980.
- Dyar, M.D., McGuire, A.V., and Harrell, M.D. (1992) Crystal chemistry of iron in two styles of metasomatism in the upper mantle. *Geochimica et cosmochimica acta*, 56, 2579-2586.
- Dyar, M.D., Gunter, M.E., and Tasa, D. (2008) *Mineralogy and optical mineralogy*, 544-634 p. Mineralogical Society of America, Chantilly.
- Dyar, M.D., Delaney, J.S., Sutton, S.R., and Schaefer, M.W. (1998) Fe³⁺ distribution in oxidized olivine: A synchrotron micro-XANES study. *American Mineralogist*, 83, 1361-1365.
- Dyar, M.D., Giguere, S., Carey, C., and Boucher, T. (2016c) Comparison of baseline removal methods for laser-induced breakdown spectroscopy of geological samples.
- Dyar, M.D., Fassett, C.I., Giguere, S., Lepore, K., Byrne, S., Boucher, T., Carey, C., and Mahadevan, S. (2016a) Comparison of univariate and multivariate models for prediction of major and minor elements from laser-induced breakdown spectra with and without making. *Spectrochimica Acta B*, 123, 93-104.
- Dyar, M.D., Breves, E.A., Gunter, M.E., Lanzirotti, A., Tucker, J.M., Carey, C., Peel, S.E., Brown, E.B., Oberti, R., Lerotic, M., and Delaney, J.S. (2016b) Use of multivariate analysis for synchrotron micro-XANES analysis of iron valence state in amphiboles. *American Mineralogist*, 101, 1171-1189.
- Dyar, M.D., Sklute, E.C., Menzies, O.N., Bland, P.A., Lindsley, D., Glotch, T., Lane, M.D., Wopenka, B., Klima, R., Bishop, J.L., Hiroi, T., Pieters, C.M., and Sunshine, J. (2009) Spectroscopic characteristics of synthetic olivines, with an emphasis on fayalite: An integrated multi-wavelength approach. *American Mineralogist*, 94, 883-898.
- Dyar, M.D., Breves, E.A., Emerson, E., Bell, S.W., Nelms, M., Ozannes, M.V., Peel, S.E., Carmosino, M.L., Tucker, J.M., Gunter, M.E., Delaney, J.S., Lanzirotti, A., and Woodland, A.B. (2012) Accurate determination of ferric iron in garnets by bulk Mössbauer spectroscopy and synchrotron micro-XANES. *American Mineralogist*, 97, 1726-1740.
- Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004) Least angle regression. *The Annals of Statistics*, 32, 407-499.
- Eilers, P. H. (2004) Parametric time warping. *Analytical Chemistry*, 76, 404-411.
- Farrell-Turner, S., Reimold, W.U., Nieuwoudt, M., and Erasmus, R.M. (2005) Raman

- spectroscopy of olivine in dunite experimentally shocked to pressures between 5 and 59 GPa. *Meteoritics and Planetary Science*, 40, 1311-1327.
- Floran, R.J., Prinz, M., Hlava, P.F., Keil, K., Nehru, C.E., and Hinthorne, J.R. (1978) The Chassigny meteorite: A cumulate dunite with hydrous amphibole-bearing melt inclusions. *Geochimica et Cosmochimica Acta*, 42, 1213-1229.
- Foster, N.J., Wozniakiewicz, P.J., Burchell, M.J., Kearsley, A.T., Creighton, A.J., and Cole, M.J. (2007) Identification by Raman spectroscopy of processing effects in forsterite- fayalite samples during hypervelocity impacts on foils and capture in aerogel. *Annual Meteoritical Society Meeting*, 70, Abstract no.5186.
- Foster, N.F., Wozniakiewicz, P.J., Price, M.C., Kearsley, A.T., and Burchell, M.J. (2013) Identification by Raman spectroscopy of Mg-Fe content of olivine samples after impact at 6 km s^{-1} onto aluminium foil and aerogel: In the laboratory and in Wild-2 cometary samples. *Geochimica et Cosmochimica Acta*, 121, 1-14.
- Gaisler, S.V., and Kolesov, B.A. (2007) Raman spectra of olivine solid solution ($\text{Fe}_x\text{Mg}_{1-x}$) $_2\text{SiO}_4$ and spin-vibration interaction. *Journal of Structural Chemistry*, 48, 61-65.
- Giguere, S., Boucher, T., Carey, C., Mahadevan, S., and Dyar, M.D. (2017) A fully customized baseline removal framework for spectroscopic applications. *Applied Spectroscopy*, in press.
- Gillet, P., Daniel, I., and Guyot, F. (1997) Anharmonic properties of Mg_2SiO_4 -forsterite measured from the volume dependence of the Raman spectrum. *European Journal of Mineralogy*, 9, 255-262.
- Golubovic, A., Gajic, R., and Scepanovic, M. (2009) Raman and IR Spectra of Pure and Doped Forsterite Single Crystals. *Science of Sintering*, 41, 43-47.
- Guyot, F., Boyer, H., Madon, M., Velde, B., and Poirier, J.P. (1986) Comparison of the Raman microprobe spectra of $(\text{Mg}, \text{Fe})_2\text{SiO}_4$ and Mg_2GeO_4 with olivine and spinel structures. *Physics and Chemistry of Minerals*, 13, 91-95.
- Grant, C.A. (1995) Sources of Experimental and Analytical Error in Measurements of the Mössbauer Effect in Amphibole, 213p. PhD thesis, University of Oregon, Eugene.
- Haskin, L.A., Wang, A., Rockow, K.M., Jolliff, B.L., Korotev, R.L., and Viskupic, K.M. (1997) Raman spectroscopy for mineral identification and quantification for in situ planetary surface analysis: a point count method. *Journal of Geophysical Research*, 102, 19293-19306.
- Heymann, D. (1990) Raman study of olivines in 37 heavily and moderately shocked ordinary chondrites. *Geochimica et Cosmochimica Acta*, 54, 2507-2510.
- Heymann, D., and Cellucci, T.A. (1988) Raman spectra of shocked minerals 1: olivine. *Meteoritics*, 23, 353-357.

- Iishi, K. (1978) Lattice dynamics of forsterite. *American Mineralogist*, 63, 1198-1208.
- Isaacson, P.J., Klima, R.L., Sunshine, J.M., Pieters, C.M., Hiroi, T., and Dyar, M.D. (2014) Visible to near-infrared reflectance spectroscopy of pure synthetic olivine across the olivine solid solution. *American Mineralogist*, 99, 467-478.
- Ishibashi, H., Arakawa, M., Yamamoto, J., and Kagi, H. (2011) Precise determination of Mg/Fe ratios applicable to terrestrial olivine samples using Raman spectroscopy. *Journal of Raman Spectroscopy*, 43, 331-337.
- Ishibashi, H., Arakawa, M., Ohi, S., Yamamoto, J., Miyake, A., and Kagi, H. (2008) Relationship between Raman spectral pattern and crystallographic orientation of a rock-forming mineral: a case study of $\text{Fo}_{89}\text{Fa}_{11}$ olivine. *Journal of Raman Spectroscopy*, 39, 1653-1659.
- Kajfosz, J., and Kwiatek, W.M. (1987) Nonpolynomial approximation of background in X-ray spectra. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 22, 78-81.
- Kolesov, B.A., and Geiger, C.A. (2004a) A Raman spectroscopic study of Fe–Mg olivines. *Physics and Chemistry of Minerals*, 31, 142–154.
- Kolesov, B.A., and Geiger, C.A. (2004b) A temperature-dependent single-crystal Raman spectroscopic study of fayalite: evidence for phonon-magnetic excitation coupling. *Physics and Chemistry of Minerals*, 31, 155-161.
- Kolesov, B.A., and Tanskaya, J.V. (1996) Raman spectra and cation distribution in the lattice of olivines. *Materials Research Bulletin*, 31, 1035-1044.
- Kuebler, K.E., Jolliff, B.L., Wang, A., and Haskin, L.A. (2005) Extracting olivine (Fo-Fa) compositions from Raman spectral peak positions. *Lunar and Planetary Science Conference*, 36, Abstract no.2086.
- Kuebler, K.E., Jolliff, B.L., Wang, A., and Haskin, L.A. (2006) Extracting olivine (Fo-Fa) compositions from Raman spectral peak positions. *Geochimica et Cosmochimica Acta*, 70, 6201-6222.
- Lane, M.D., Glotch, T.D., Dyar, M. D., Pieters, C.M., Klima, R., Hiroi, T., Bishop, J.L., and Sunshine, J. (2011) Mid-infrared spectroscopy of synthetic olivines: Thermal emission, attenuated total reflectance, and spectral and diffuse reflectance studies of forsterite to fayalite. *Journal of Geophysical Research: Planets*, 116, 1-20.
- Liu, L. (1993) Raman spectra of forsterite and fayalite at high temperatures and room temperature. *High Pressure Research*, 5, 241-256.

- Lumpkin, G.R., and Ribbe, P.H. (1983). Composition, order-disorder and lattice parameters of olivines: relationships in silicate, germate, beryllate, phosphate and borate olivines. *American Mineralogist*, 68, 164-76.
- Manghnani, M.H., Hushur, A., Smyth, J.R., Nestola, F., Dera, P., Sekar, M., Amulele, G., and Frost, D.J. (2013) Compressibility and structural stability of two hydrated olivine samples ($\text{Fo}_{97}\text{Fa}_3$) to 34 GPa by X-ray diffraction and Raman Spectroscopy. *American Mineralogist*, 98, 1972-1979.
- Maurice, S., Wien, R.C., Anderson, R., Beyssac, O., Bonal, L., Clegg, S., DeFlores, L., Dromart, G., Fischer, W., Forni, O., Gasnault, O., Grotzinger, J., Johnson, J., Martinez-Frias, J., Mangold, N., McLennan, S., Montmessin, F., Rull, F., Sharma, S., Fouchet, T., Poulet, F., and the SuperCam team (2015) Science objectives of the SuperCam instrument for the Mars2020 rover. *Lunar and Planetary Science Conference*, 46, Abstract no.2818.
- McCanta, M.C., Treiman, A.H., Dyar, M.D., Alexander, C.O.D., Rumble, D., and Essene, E. J. (2008) The LaPaz Icefield 04840 meteorite: Mineralogy, metamorphism, and origin of an amphibole-and biotite-bearing R chondrite. *Geochimica et Cosmochimica Acta*, 72, 5757-5780.
- McGuire, A.V., Dyar, M.D., and Nielson, J.E. (1991) Metasomatic oxidation of upper mantle periodotite. *Contributions to Mineralogy and Petrology*, 109, 252-264.
- McGuire, A.V., Francis, C.A., and Dyar, M.D. (1992). Mineral standards for electron microprobe analysis of oxygen. *American Mineralogist*, 77, 1087-1087.
- McKeown, D.A., Bell, M.I., and Caracas, R. (2010) Theoretical determination of the Raman spectra of single-crystal forsterite (Mg_2SiO_4). *American Mineralogist*, 95, 980-986.
- McSween, H.Y., and Jarosewich, E. (1983) Petrogenesis of the Elephant Moraine A79001 meteorite: Multiple magma pulses on the shergottite parent body. *Geochimica et Cosmochimica Acta*, 47, 1501-1513.
- Miura, M., Arai, S., and Mizukami, T. (2011) Raman spectroscopy of hydrous inclusions in olivine and orthopyroxene in ophiolitic harzburgite: implications for elementary processes in serpentinization. *Journal of Mineralogical and Petrological Science*, 106, 91-96.
- Mikouchi, T., and Kurihara, T. (2008) Mineralogy and petrology of paired lherzolithic shergottites Yamato 000027, Yamato 000047, and Yamato 000097: Another fragment from a Martian "lherzolite" block. *Polar Science*, 2, 175-194.
- Miyake, M., Nakamura, H., Kojima, H., and Marumo, F. (1987) Cation ordering in Co-Mg olivine solid-solution series. *American Mineralogist*, 72, 594-598.
- Miyamoto, M. (1995) Micro Raman spectroscopy of olivines in L6 chondrites: evaluation of the degree of shock. *Geophysical Research Letters*, 22, 437-440.

- Mohanan, K., Sharma, S.K., and Bishop F.C. (1993) A Raman spectral study of forsterite-monticellite solid solutions. *American Mineralogist*, 78, 42-48.
- Morse, S.A. (1996) Kiglapait mineralogy .3. Olivine compositions and Rayleigh fractionation models. *Journal of Petrology*, 37, 1037-1061.
- Morse, S.A. (2001) Augite-olivine equilibria in the Kiglapait intrusion, Labrador, Canada. *Canadian Mineralogist*, 39, 267-274.
- Mouri, T., and Enami, M. (2008) Raman spectroscopic study of olivine-group minerals. *Journal of Mineralogical and Petrological Sciences*, 103, 100-104.
- Nagy, S., Gucsik, A., Ninagawa, K., Józsa1, S., Bérczi, S., Nishido, H., Okumura, T., Veres, M., Kereszturi1, Á., and Hargitai, H. (2008) Micro-Raman spectroscopy of the shocked olivine in the Martian meteorite ALH77005. *Meteoritics & Planetary Science*, A187-A187.
- Nasdala, L., Smith, C., Kaindl, R., and Ziemann, M.A. (2004) Raman spectroscopy: Analytical perspectives in mineralogical research. *EMU Notes in Mineralogy*, 6, 1-63.
- Parques-Ledent, M.T., and Tarte, P. (1973) Vibrational studies of olivine-type compound-I. The i.r. and Raman spectra of the isotopic species of Mg_2SiO_4 . *Spectrochimica Acta*, 20A, 1007-1016.
- Pasteris, J.D., and Wanamaker, B.J. (1988) Laser Raman microprobe analysis of experimentally re-equilibrated fluid inclusions in olivine: some implications for mantle fluids. *American Mineralogist*, 73, 1074-1088.
- Pirou, B. (1983) The high-frequency vibrational spectra of vitreous and crystalline orthosilicates. *American Mineralogist*, 68, 426-443.
- Price, G.D., Parker, S.C., and Leslie, M., (1987) The lattice dynamics of forsterite. *Mineralogical Magazine* 51, 157-170.
- Rancourt, D. G. (1994). Mössbauer spectroscopy of minerals I. Inadequacy of Lorentzian-line doublets in fitting spectra arising from quadrupole splitting distributions. *Physics and Chemistry of Minerals*, 21, 244-249.
- Rouquette, J., Kantor, I., McCammon, C.A., Dmitriev, V., and Dubrovinsky, L.S. (2008) High-pressure studies of $(Mg_{0.9}Fe_{0.1})_2SiO_4$ olivine using Raman spectroscopy, X-ray diffraction, and Mössbauer spectroscopy. *Inorganic Chemistry*, 47, 2668-2673.
- Rull, F. (2014) The laser Raman instrument (RLS) for ExoMars 2018 mission and the scientific operation on Mars. *International Conference of Mars*, 8, Abstract no.1277.
- Savitsky, A. and Golay, J.E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36, 1627-1639.

- Schaefer, M.W. (1983) Crystal chemistry of ferric-rich fayalites, 38-56 p. Ph.D. thesis, MIT, Cambridge.
- Servoin, J., Piriou, B., and Alain, P. (1972) Diffusion Raman de la forsterite sythetique pure, Mg_2SiO_4 . *Comptes Rendus de l'Académie des Sciences Paris*, 274, 135-137.
- Sklute, E.C. (2006) Mössbauer spectroscopy of synthetic olivine across the Fe-Mg solid solution, 1-38 p. B.A. thesis, Mount Holyoke College, South Hadley.
- Smyth, J.R. (1975) High temperature crystal chemistry of fayalite. *American Mineralogist*, 60, 1092-1097.
- Tibshirani, R. (1995) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*, 58, 267-288.
- Tucker, J.M., Dyar, M.D., Schaefer, M.W., Clegg, S.M., and Wiens, R.C. (2010) Optimization of laser-induced breakdown spectroscopy for rapid geochemical analysis. *Chemical Geology*, 277, 137-148.
- Wang, S.Y., Sharma, S.K., and Cooney, T.F. (1993). Micro-Raman and infrared spectral study of forsterite under high pressure. *American Mineralogist*, 78, 469-476.
- Wang, A., Kuebler, K., Jolliff, B., and Haskin, L.A. (2004) Mineralogy of a Martian meteorite as determined by Raman spectroscopy. *Journal of Raman Spectroscopy*, 35, 504-514.
- Weber, I., Bottger, U., Pavlov, S.G., Jessberger, E.K., and Hubers, H.W. (2014) Mineralogical and Raman spectroscopy studies of natural olivines exposed to different planetary environments. *Planetary and Space Science*, 104, 163-172.
- Wivel, C., and Mørup, S. (1981). Improved computational procedure for evaluation of overlapping hyperfine parameter distributions in Mossbauer spectra. *Journal of Physics E: Scientific Instruments*, 14, 605-610.
- Wold, S., Martens, H., and Wold, H. (1983) The multivariate calibration problem in chemistry solved by the PLS method. *Lecture Notes in Mathematics*, 973, 286-293.
- Yasuzuka, T., Ishibashi, H., Arakawa, M., Yamamoto, J., and Kagi, H. (2009) Simultaneous determination of Mg# and residual pressure in olivine using micro-Raman spectroscopy. *Journal of Mineralogical and Petrological Sciences*, 104, 395-400.
- Zhang, Z.M., Chen, S., and Liang, Y.Z. (2010) Baseline correction using adaptive iteratively reweighted penalized least squares. *Analyst*, 135, 1138-1146.