Predicting olivine composition using Raman spectroscopy through band shift and multivariate analyses

LAURA B. BREITENFELD¹,*, M. DARBY DYAR¹, C.J. CAREY², THOMAS J. Tague JR.³, PENG WANG³, TERRY MULLEN⁴, and MARIO PARENTE⁴

¹Department of Astronomy, Mount Holyoke College, South Hadley, Massachusetts 01075, U.S.A.
²College of Information and Computer Sciences, University of Massachusetts Amherst, Amherst, Massachusetts 01003, U.S.A.
³Bruker Optics, Inc., Billerica, Massachusetts 01821, U.S.A.
⁴Department of Electrical and Computer Engineering, University of Massachusetts Amherst, Amherst, Massachusetts 01003, U.S.A.

ABSTRACT

Olivine group minerals are ubiquitous in extrusive igneous rocks and play an important role in constraining equilibria for samples in the upper mantle and above. All Raman spectra of the olivine group minerals in the solid solution between forsterite (Fo, Mg₂SiO₄) and fayalite (Fa, Fe₂SiO₄) have a high-intensity doublet between 800 and 880 cm⁻¹. Previous studies used small sample suites with limited compositional ranges and varying spectrometers to relate energy shifts of these two bands to Mg/Fe contents. In this work, Raman spectra of 93 olivine samples were acquired on either Bruker’s 532 nm (laser wavelength) Senterra or BRAVO (785/852.3 nm) spectrometer. This paper compares the two-peak band shift univariate method with two multivariate methods: partial least squares (PLS) and the least absolute shrinkage operator (Lasso). Data sets from several instruments are also examined to assess the most accurate method for predicting olivine composition from a Raman spectrum.

Our 181-spectra PLS model is recommended for use when determining olivine composition from a Raman spectrum. For Raman spectra of mixed phases where only the olivine doublet can be identified, composition can best be determined using the position of the peak ca. 838–857 cm⁻¹ through use of the regression equation %Fo = –0.179625x² + 310.077x –133 717 (where x = DB2 centroid in units of cm⁻¹).

In situ methods for predicting mineral composition on planetary surfaces are critically important to extraterrestrial exploration going forward; of these, Raman spectroscopy is likely the best, as shown by the impending deployment of several Raman instruments to Mars (ExoMars and Mars 2020). More broadly, application of machine learning methods to spectral data processing have implications to multiple fields that use spectroscopic data.

Keywords: Raman spectroscopy, olivine, forsterite, fayalite, PLS, Lasso

INTRODUCTION

Olivine group minerals control many of the properties of Earth’s upper mantle, affect rheology, and may be diagnostic of crystallization temperature in terrestrial and extraterrestrial rocks. Their solid solution commonly spans the range between forsterite (Fo, or Mg₂SiO₄) and fayalite (Fa, or Fe₂SiO₄) with minor substitutions of alternative cations such as Mn and Ni. Because olivine composition provides an important petrogenetic indicator, development of convenient methods to measure it without microanalytical techniques that require sample preparation are desirable. This paper focuses on the determination of olivine composition using Raman spectroscopy for this purpose. It has the potential to enable compositions to be conveniently determined in the laboratory, in field identifications with portable units, and on other planets such as Mars.

This problem has been extensively studied using conventional regression-based approaches, but a generalization of their results is arguable given their very small (<20) sample suites and coverage of olivine composition. These prior studies (Kuebler et al. 2006; Foster et al. 2007; Gaisler and Kolesov 2007; Mouri and Enami 2008; Yasuzuka et al. 2009; Ishibashi et al. 2011) have related olivine composition to the peak positions of a high-intensity doublet in the range of 800–880 cm⁻¹ (Fig. 1). Peak centroids are regressed against composition to obtain an equation suitable for prediction of composition in unseen samples. Previous workers have used R² values to characterize their prediction algorithms, making their accuracy difficult to quantify and comparisons across models impossible. Moreover, several Raman instruments with varying resolution and laser wavelength and a combination of single crystal and powdered samples were used. It is thus difficult to assess which model to use to estimate olivine composition on unseen data from a different spectrometer than the one used in each study.

In this study, all known publicly accessible olivine Raman data are considered. In addition, new Raman spectra were acquired from a suite of 93 well-characterized synthetic and naturally occurring olivines using Bruker BRAVO and Senterra spectrometers. The accuracies of linear regression (univariate) methods for various data sets are quantified to enable useful