Characterizing silicate materials via Raman spectroscopy and machine learning: Implications for novel approaches to studying melt dynamics

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Blake O. LaDouceur
December 2023

DEDICATION

I dedicate this work to my wife, Tiffany N. LaDouceur

ACKNOWLEDGEMENTS

First, I offer my sincerest thanks, admiration, and gratitude to my major advisor, Dr. Molly C McCanta. Her guidance and mentorship throughout my time at the University of Tennessee has been one of the great treasures of my academic career. I am forever grateful for the opportunity to have worked with such a dynamic and tenacious scientist whose intellectual curiosity and expertise helped propel this project beyond what I could have imagined. Thank you, Molly, for everything.

Second, I acknowledge the members of my committee: Drs. Bradley Thomson, Bhavya Sharma, and Annette Engel for their unwavering support in my studies, as well as their key insights into my projects. I also owe a special thanks to several key collaborators: For aiding in the detailed modeling aspect of this work, I thank Dr. M. Darby Dyar and Cai Ytsma. For their help in data collection on various instruments, I recognize Allan Patchen, Dr. Grace Sarabia, Natalie Dunn, and Dr. Elizabeth (Eli) Sklute. Without the contributions from these individuals, this collection of projects would not have been possible.

Lastly, I owe my unwavering appreciation to my family, and my close friends. Tiffany, Maria, Todd, Alec, Cindy, Steve, Spencer, Shelby, Elaine, and Ozzie, you have proved resolute in your familial support of me during the past 6 years of graduate school. It has been a long, fruitful, and daring journey following my professional shift from medicine and healthcare in late 2016. You, my family, have served as a bottomless resource of care and love. I love you all. I offer my gratitude to my closest friends: Chantz, Quinn, Daniel, Brandon, Mina, Anthony, Trevor, Jeffery, Justin, Zach, and Chris. Thank you for your friendship, and for playing major social roles at key inflection points in my life.

ABSTRACT

Silicate melt characteristics impose dramatic influence over igneous processes that operate, or have operated on, differentiated bodies: such as the Earth and Mars. Current understanding of these melt properties, such as composition, primarily comes from investigations on their volcanic byproducts. Therefore, it is imperative to innovate on modalities capable of constraining melt information in environments where a reliance on laboratory methods is severed. Recent investigations have turned to Raman Spectroscopy and amorphous volcanics as a suitable pairing for exploring these ideas. Silicate glasses are a proxy for igneous melts; and Raman spectroscopy is a robust analytical technique capable of operating in-situ. Existing calibrations for retrieving geochemical information from such samples using their Raman data are extremely underdeveloped, with only a handful of approaches available. Here, two supervised machine learning algorithms; Partial Least Squares (PLS) and Least Absolute Shrinkage & Selection Operator (LASSO) are employed with Raman spectroscopy to quantify geochemical information in volcanic glasses and tephra, while also qualifying the underlying atomic mechanics that drive Raman signal variability. This approach establishes a foundation for future explorations into new-age modeling technologies for geoscience experiments. **Chapter I's** PLS geochemical model predicted the concentrations of oxide constituents in synthetic silicate glasses (SiO_2, Na_2O, K_2O, CaO, TiO_2, Al_2O_3, FeO_T, MgO) with increased accuracy and applicability over currently available offerings. The study presents the largest and most diverse sampling suite yet utilized to produce such models. **Chapter II** highlights the limitations to PLS and LASSO based strategies for constraining iron (Fe)-redox information in glasses but uncovers their ability to accurately predict glass structural parameters like polymerization (NBO/T). **Chapter III** yielded accurate predictions of tephra concentrations from various mixed sediment samplings using PLS and LASSO calibrations. Spectra parametrizations highlighted that tephra signatures are unique enough to be readily distinguished from more crystalline profiles using Raman spectroscopy and machine learning. PLS and LASSO technologies are shown to be suitable, yet immature, avenues for unraveling the

geochemical underpinnings of the Raman collections made in this work and help set the stage for future applications to Raman data from planetary missions such as the Perseverance Rover.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

A core aim of modern volcanology and comparative planetology is to unravel the igneous dynamics operating on differentiated silicate bodies, such as the Earth and Mars. Geoscientists are particularly interested in understanding the geochemical and physical properties of silicate melts because they provide records of igneous processes, such as magma generation, transportation, and eruption (Hargraves, 1980; Lange and Carmichael, 1987; Dingwell, 2006; Behrens and Zhang, 2009; Robert et al., 2014). Volcanic glasses are critically important to study because they represent rapidly quenched melt fractions of magma and preserve the pre-eruptive properties of silicate melts that enable extrapolations of the liquid state.

Recently, Raman spectroscopy has been shown to be as a viable technique for analysis of silicate minerals, rocks, and other phases (Mysen et al., 1982; McMillan and Piriou, 1983; McMillan, 1984; White and Minser, 1984; Baert et al., 2011; McMillan and Hofmeister, 1988; Dünnwald and Otto, 1989; de Faria et al., 1997; Di Muro et al., 2009; Prinsloo et al., 2010; Rossano and Mysen, 2012; Le Losq and Neuville, 2014; Beegle et al., 2015 Cochrane and Blacksberg, 2015; Yadav and Singh, 2015; Di Genova et al., 2015, 2016a, 2016, 2017, 2018; Welsch et al., 2017; Giordano et al., 2019, 2020; Pasteris and Beyssac, 2020; González-García et al., 2020; Stabile et al., 2021). Raman spectroscopy is a robust and multifaceted analytical technique that is capable of probing the molecular structure and chemistry of many geologically-relevant phases. The technique is non-destructive, requires little to no sample preparation, can be used over a wide swath of operating wavelengths (i.e., ultraviolet, UV, to near-infrared, NIR), and allows for rapid data collection times at high spatial resolution. Raman spectroscopic analysis is also currently being used for the first time in extraterrestrial environments, as the Perseverance Rover (Mars 2020) operating in Jezero Crater on the surface of Mars, and is collecting data using two separate Raman-capable instruments, the SuperCam and the Scanning Habitable Environments with Raman & Luminescence for Organics & Chemicals, or SHERLOC, instrument (e.g., Beegle et al., 2015; Wiens et al., 2020a, 2020b; Anderson et al., 2022). These applications of Raman spectroscopy techniques in

the geosciences highlight the need to expand the current understanding of Raman-based spectral assessments for geological materials in a general sense, while also refining in the capabilities of Raman spectroscopy of silicate samples suitable for extrapolation to the liquid state.

Although Raman spectroscopic studies of crystalline silicate samples over the last four decades are robust (Prinsloo et al., 2010; Cochrane and Blacksberg, 2015; Chou and Wang, 2017; Giordano et al., 2020), Raman spectroscopic investigations that focus on identifying primary melt properties using synthetic and natural amorphous glasses remain largely underdeveloped. Raman spectra of glasses have predominately been used to determine the relationship between compositional variables and spectral features. But, more recent experiments have shown that melt elements, such as degree of polymerization, chemical composition, oxidation state, volatile content, and even nanocrystal presence, can also be obtained using Raman spectroscopy. Alongside modeling attempts to determine the geochemistry and structure of synthetic or natural glasses, the modality by which glass signals interact in mixed Raman data (i.e., samples that contain both amorphous and crystalline phases) is also poorly understood. Crystalline phases can mask amorphous features in other types of comparable spectroscopies, such as Infrared Spectroscopy (IR). Natural volcanic glass is usually heavily altered or mingled with crystalline material. As such, glass-mineral mixtures are commonly used, and interpreting them is important if natural samples are to be interpreted. Collectively, these studies highlight the abilities of Raman spectroscopy for determining glass composition, but also illustrate current limitations of the field. At the moment, there is no consensus on how to 1) treat, correlate, or interpret collected Raman signals, 2) determine the relationship between envelope intensity ratios and compositional changes, 3) reconcile band feature differences between studies, or 4) advance synthetic-based modeling strategies to natural glass samples. Additionally, most current models are generated using limited composition sample suites and rely on conventional peak parameterization techniques.

Studies using other types of spectroscopic analysis have shown the utility of applying multivariate analytical techniques (MVA) to model volcanic glasses and other

phase geochemistry. MVA is a form of supervised machine learning and has been used to: 1) constrain Fe-oxidation state and elemental concentrations of glasses and igneous rocks using X-ray absorption near edge spectroscopy (XAS-XANES) and laser induced breakdown spectroscopy (LIBS) (Dyar et al., 2012; Lanzirotti et al., 2018); 2) model tephra-phase concentrations in the visible, near, and mid-infrared (VNIR, MIR) (Leight 2020; Leight et al., 2021); and 3) predict trace element concentrations in rock powders with X-ray fluorescence data (XRF) (Lanzirotti et al., 2018). Experiments such as those conducted in Breitenfeld et al. (2018) and Carey et al. (2015) have also shown that machine learning can be successfully paired with Raman spectroscopic data to classify minerals in the RRUFF database (https://rruff.info) and to determine the composition of olivine group minerals.

There are two prominent MVA techniques, Partial Least Squares (PLS) and Least Absolute Shrinkage & Selection Operator (LASSO). These powerful tools rely on modified forms of multiple linear regression and are used for the development of regression models that can reliably predict a host of input variables when paired with a suitable calibration dataset. In the case of predicting geochemical information from spectra profiles, collected Raman spectroscopic data serve as the observational input, with the predicted oxide concentrations representing explanatory variables that are measured as responses to variations in the observed spectra (i.e., quantified by model coefficients). Although such approaches have proven useful on crystalline Raman spectral data, calibrations for amorphous or amorphous-mixed volcanic materials have not yet been developed.

Because of the use of Raman spectroscopy on the *Perseverance* rover, there is now a growing need in the planetary community for Raman spectroscopy-based methodologies that can adequately provide geochemical information on silicate systems via proxy samples. Supervised machine learning methodologies have shown steady progress in modeling such data and are primed for future applications. Therefore, the joining of Raman spectroscopy and PLS/LASSO algorithms to model silicate samples in the following chapters will further refine their quantifying and qualifying capabilities,

while also uncovering the current boundaries for what is possible when pairing these technologies.

Chapter I: This chapter helps to establish a reliable and widely applicable modeling suite and processing calibration for quantifying eight major oxide variables (SiO_2, Na_2O, K_2O, CaO, TiO_2, Al_2O_3, FeO_T, MgO wt%) in a host of 69 silicate glass samples (spanning basaltic – rhyolitic and reduced – oxidized compositions) via their collected Raman spectra using a form of Partial Least Squares (PLS) regression. Alongside the models, the evolution of the Raman spectra as a function of glass chemistry was investigated in an effort to contextualize model accuracy trends. The primary goals of the study were to determine the utility of Raman spectroscopy as a reliable source of geochemical information and probe the use of PLS regression methods to achieve those quantifications. This chapter provides the best modeling package possible using these technologies on a large glass dataset, and the results are a product of iterative testing and experimentation. After establishing the model and Raman spectroscopy parameterizations for the 69 original samples, the approach was validated against a publicly available Raman spectra dataset obtained from 11 mid-ocean ridge basaltic glass samples. This external testing provides context for this new technology pairing approach and outlines the optimization that may still be needed in future iterations of the strategy.

Chapter II: This chapter presents a detailed comparison of the use of PLS or LASSO modeling approaches to predict iron redox variables (%Fe^{3+}, %Fe^{2+}, Fe_2O_3, FeO) and melt polymerization (NBO/T) from a suite of 78 silicate glass samples via their Raman spectra. By quantifying the percent of ferric iron (%Fe^{3+}) in particular, a direct measure of overall sample oxidation state is possible. Building from the successes in modeling oxide concentrations and contextualizing their effect on glassy Raman data in Chapter I, this study expands to constraining additional melt parameters. Redox and melt polymerization are strongly correlated with the melt structure and, therefore, should have extensive effects on Raman spectra. This study highlights the differences between the two machine learning algorithms and serves as a blueprint for application of these methods in future studies.

Chapter III: This chapter functions as a preliminary proving ground for unraveling the characteristics of signal dilution and discrimination for Raman spectra collected from mixed volcaniclastic and sedimentary material, in both a qualitative and numerical fashion. Raman spectroscopic data were collected for 13 pre-mixed tephra:sediment samples, for which a specific amount of volcaniclastic tephra and hemipelagic sediment are mixed in concentrations from 0% - 100%. Band shapes and features were compared and recorded along mixing trends to gain a better understanding if Raman spectroscopy could identify the tephra signals, and the ability of the PLS and LASSO algorithms to predict the amount of tephra material (% tephra) present was evaluated. This research demonstrates the ability of Raman spectroscopy to identify volcaniclastic materials in mixed deposits to tephra concentrations as low as 15%.

BACKGROUND

Silicate melt properties and their influence over igneous processes

Properties of silicate melts are largely dictated by their composition. In a multicomponent silicate magma, silica tetrahedra $(SiO_4)^{4-}$ and assorted cations are linked together through either bridging oxygens (BO) or non-bridging oxygens (NBO). Bridging oxygens are oxygen atoms that link adjacent tetrahedral units together in a network via strong covalent bonds. Tetrahedral units capable of forming BO bonds are comprised of cations like silica (Si^{4+}) and aluminum (Al^{3+}) and are designated as network forming agents. However, other abundant melt cations, such as calcium (Ca^{2+}), sodium (Na^{+}), and magnesium (Mg^{2+}), act as network modifiers. These cations do not form tetrahedra and disrupt the linkages in a silicate network, causing NBOs to form. NBOs are oxygen atoms that do not link with other tetrahedra, but rather create weak linkages between their associated units and the free-floating modifying cation. The chain length (polymerization) and internal friction (viscosity) of a melt is dictated by the number of NBO/BO bonds, and is the primary control over how magmas erupt, move, and crystallize. Polymerization is defined by the ratio NBO/T and provides geoscientists a quantifiable measure of the number of NBO bonds per tetrahedrally coordinated cations (T) in the glass (McMillian, 1984; Richet, 1984). Viscosity is positively correlated with polymerization; that is, the greater the melt polymerization, the greater the melt viscosity.

Oxidation state (f_{O2}) is a measure of variations in multivalent elements as a function of redox conditions. The most common multivalent element in igneous liquids, iron (Fe), can be present as either oxidized ferric (Fe^{3+}), reduced ferrous (Fe^{2+}), or neutral metallic (Fe^{0}) iron (expressed in oxide form as Fe_2O_3 and FeO, respectively), which depends on the prevailing system conditions, specifically temperature, chemical composition and oxygen fugacity (f_{O2}) (Kress and Carmichael, 1991; Mysen, 1991; Ottonello et al., 200; Giuli et al., 2012; Osipov et al., 2015; Borisov et al., 2015). The overall oxygen fugacity of an Fe-bearing silicate system can be directly measured by determining the Fe^{3+}/Fe^{2+} ratio of a melt. The valence state of Fe determines whether it behaves as either a network former and modifier and, therefore, the prevalence of one valence state over the other can greatly affect bonding in the silicate network through changes in NBO and BO species.

Melt oxidation state and general composition control major physical properties by altering polymerization via abundances of tetrahedral linkages. For example, an Fe-bearing, high silica melt equilibrated under oxidizing conditions (more Fe^{3+}) exhibits a greater degree of polymerization through increased T-O bonds. This greater degree of polymerization enhances the length and strength of the network chain, thereby increasing the internal friction capacity of the melt. This leads to a higher overall viscosity profile. This highly viscous liquid strongly resists movement and transport, which allows for more explosive eruption styles. In this manner, the nature and style of igneous processes are controlled by silicate melt geochemical composition (i.e., general and oxidation state) and bonding dynamics, which also dictate melt structural properties like polymerization (Lofgren, 1980; Mysen et al., 1982; Mysen and Piriou, 1983; Richet, 1984; Stebbins, 1995; Liebske et al., 2003; Henderson, 2005; Karki, 2010).

Eruptive products as proxies for melts: volcanic glass and tephra

Volcaniclastic material, including silicate glass and tephra, is extremely abundant in a variety of settings on silicate bodies, such as the Earth, Moon, and Mars (i.e., Rampe et al., 2014; Fu et al., 2017; Stabile et al., 2021). These phases have been identified on Mars through orbital and Earth-based spectroscopic observations, lander and rover operations, geochemical modeling, and meteorite analysis (Singer, 1985; McSween,

1994; Bell et al., 2000) Such phases have similarly been identified on Earth and the Moon through comparable techniques, with the added advantage of direct sampling through field collection.

Glasses can result from primary processes like volcanism and impact events, as well as from secondary processes like radiation damage of crystalline phases, hydrothermal alteration, and low-temperature weathering (e.g., Rampe et al., 2014). Volcanic glasses serve as the closest available proxies for extrapolation to the liquid magmatic state, as they have been rapidly quenched and essentially represent the melts in a frozen state. They are non-crystalline materials that lack long-range structural network order and symmetry. Volcanic glasses can be found as melt inclusions, interstitial rock matrixes, or as pure-form chips.

Volcanic tephra deposits are heterogenous mixtures of mineral fragments, rock debris, glass, and ashy material that are expelled and deposited from explosive eruption events. Similar to volcanic glasses, tephra materials are found on nearly every terrestrial planet in the solar system that harbor evidence of explosive volcanic activity (Hartmann and Berman, 2000). Terrestrial tephra deposits are used as geochronologic markers, due to their instantaneous deposition timelines (Brown et al., 1992). Their large concentrations of glassy phases make tephra samples useful in igneous investigations, as these components can be used to constrain sourcing information in areas containing mixed-source materials (Lowe, 2011; Cassidy et al., 2014), and can be used to determine the composition and physical properties of the silicate melts from which they quenched.

Raman Spectroscopy

Raman spectroscopy is a vibrational spectroscopic technique that relies on the inelastic scattering of incident photons that result from the emission properties of molecular transitions from an excited virtual state back to a particular ground/initial vibrational state. All molecules in a sample, whether in a solid, liquid, or gas phase, contain different arrangements and numbers of chemical bonds. These chemical bonds have specific vibrational state frequencies and energy levels due to their inherent elemental makeup, which can be described by group theory and quantum mechanics. Raman spectroscopy provides information on the concentration of sample molecules, the

molecules themselves, and changes across elements in different states. Molecular bond vibrations that are Raman "active," meaning they can be observed, include vibrations like symmetric stretching and bending deformations (e.g., in-plane and out-of-plane). These atomic bond properties contain a change in polarizability, which is essentially a measure of how easily the electrons (in the electron cloud) can be distorted. A way to visualize such a change can be described by symmetric stretching. In this case, the molecule's bonds "stretch" or vibrate at the same particular interval (i.e., frequency). As the bonds vibrate and allow for a change in the distortion of its electric charge due to the varying bond distances across the molecule in this fashion, they can be exposed to an electric field, such as that from electromagnetic radiation that will highlight this polarizability. If a molecule is Raman "active" in this fashion, it will exhibit the Raman effect or the Raman shift when struck with incident beam photons that occur over a wide range of wavelengths (e.g., 10 nm – 2500 nm).

Most molecules in a sample are initially at an inherent ground vibrational state characteristic of their composition. A small fraction of molecules can also exist in an initially higher ground vibrational state. When incident beam photons of a specific frequency and energy interact with the sample, most of the light is transmitted through the material. However, some of the light is scattered by the molecules. The scattering of light in the sample is due to a transition of the molecules to an excited virtual state, along with a quick relaxation to their ground state or a higher vibrational state. In so doing, the molecular electrons re-emit photons in one of two ways: elastically or inelastically. Most of the time, the re-emitted photon retains the initial incident beam's energy, wavelength, and frequency. This emission is elastically scattered light and is referred to as Rayleigh scattering. There is no change in the light as the molecule relaxes back to the initial ground state. However, in a much rarer case, ~0.0001% of the light will scatter inelastically, in which there is a difference between the re-emitted, and incident photons, in energy, wavelength and frequency. This difference is the Raman shift. In this scenario there are two possibilities. Molecules at the ground vibrational state are excited to the same virtual state as those that expressed Rayleigh scattering but relax back to a higher vibrational level. In this case, the molecule has gained energy from the interaction, and

the scattered photon stream has a higher wavelength and lower frequency. This is called Stokes scattering. In the inverse case, known as Anti-Stokes Raman, a minute portion of the molecules in the sample already at a higher initial vibrational state are excited to a higher virtual state than the other molecules, but relax to the ground state. In so doing, the molecule has lost energy from the interaction with the incident photons and scatters light at a lower wavelength and higher frequency. The Anti-Stokes Raman signal is factors weaker than the Stokes signal and is generally filtered out in the spectrometer system setup.

The shift in wavelength (λ) from the Raman signal is incident-wavelength independent, and the unit of wavenumbers is used to describe the Raman shift in reciprocal centimeters (cm^{-1}), as describing the shift in wavelengths is not practical. The resulting spectrum is a plot of signal intensity vs. the Raman shift in wavenumbers (cm^{-1}). The shift is calculated primarily as the difference between the incident light frequency and the Stokes/Anti-Stokes scattered photon wavelength, so the shift is inherently the same for a molecule excited under different incident beams. The intensity (*e*) of a Raman signal is correlated with incident λ following the efficiency relationship (Eq. 1) that allows shorter incident wavelengths greater energies and a more intense Raman signal than higher λ's. There is also the relationship and influence of signal noise from fluorescence that increases as the incident λ decreases. The use of confocal Raman spectroscopy systems, as in this work, requires constraint estimates on spatial resolution (S.R.) (Eq. 2), and additional information regarding diffraction (D) relationships (Eq. 3)), with incident wavelengths and the numerical aperture size (n.a.).

$$e = \frac{1}{\lambda^4} \tag{1}$$

$$\mathrm{S.R.} = \frac{0.61}{\lambda * \mathrm{n.a.}} \tag{2}$$

$$\mathrm{D} = \frac{1.22}{\lambda * \mathrm{n.a.}} \tag{3}$$

Raman spectroscopy of silicate glasses

The strength of Raman signals from silicate glasses and other amorphous phases are a function of the molecular network structure determined by the geochemical makeup

of the sample. Spectral features like intensity, band area, band/peak position, and overall shape are linked to varying contributions of specific chemical-bonding constituents in the sample glass, and their effects on structural properties like NBO/T or polymerization (De Jong et al., 1981). Due to the amorphous nature of the silicate network, Raman bands of glasses are complicated, and have been shown to vary among non-uniform compositions, which make them difficult to assign, map, and model (Di Genova et al., 2015). Glass Raman bands are typically broad and ill-defined when compared to crystalline phases of similar compositions (McMillan and Piriou, 1983), which necessitates deconvolution procedures to obtain accurate information of Q^n-species contributions. Q^n notation refers to specific tetrahedral units in the glass network with *n* number of BO. It is a measure similar to that of NBO/T but allows for qualitative descriptions of what type of coordination geometries contribute to a band region following Gaussian or Lorentzian deconvolution procedures. The shape of a sample spectrum and its inherent arbitrary intensity are also affected by preprocessing options and collection environments. This makes interpretations related to intensity variations among compositional differences in glasses difficult.

Raman spectral signals between ~200 and 1300 cm^{-1} have been shown to contain the key diagnostic region for molecular bonding information in volatile-free aluminosilicate glasses, and have been the focal point for prior investigations focused on determining glass geochemical data (Di Muro et al., 2009; Baert et al., 2011; Di Genova et al., 2015, 2016a, 2016b, 2017; Welsch et al., 2017; Giordano et al., 2019; González-García et al., 2020) The low-wavenumber (LW) region of the Raman glass spectrum is generally defined between 300/350 – 650 cm^{-1} and is assigned to varied BO vibrational modes (McMillian and Piriou, 1983; McMillian, 1984) with ratioed contributions to three or higher-membered tetrahedra rings (TO_4) in the network. Specifically, the broad signal recorded in this region is comprised of low energy signals that result from tetrahedrally coordinated cations (e.g., T = Si^{4+}, Al^{4+}, Fe^{3+} etc.), bending in T-O-T bridges, and symmetric stretching in O-T-O configurations. Silica cations (Si^{4+}), acting as network formers in Si-O-Si bonding, constitute the major contributors to a single peak that is typically centered between 485 – 530 cm^{-1} (Bell et al., 1968; Mysen et al., 1980; Seifert

et al., 198; McMillan and Wolf, 1995; Neuville and Mysen, 1996; Colomban, 2003; Bykov et al., 2009). Trends in Raman band shape for this region typically involve shifts to 'bimodal' peak arcs in mafic glasses and a single concentrated peak in felsic glasses.

The middle wavenumber (MW) Raman-glass region is ascribed between 650 – 850 cm^{-1}. The information in the MW is attributed predominantly to the abundance of silica oxide and other network formers in the glass network and is less sensitive to other major chemical variations (i.e. Gonźalez-García et al., 2020). The general shape of the region is controlled by contributions of specific Q^n species (generally Q^0) with Si-O stretching modes that represent dominant Si motion in their respective oxygen cages, and O_2 motion in Si-O-Si planes. Network forming aluminum tetrahedral units in Q^1 likewise dominate signals in the MW and can occur near 780 cm^{-1}. These contributions are ascribed to an envelope that can occur between ~680 and 850 cm^{-1}, but can occur farther into the spectrum, sometimes nearing ~ 900 cm^{-1}. Typically, silicate glass Raman spectra show either a broad doublet arc or singular peak in this region (Matson et al., 1983; Mysen, 2003; Kalampounias et al., 2006; Mysen and Toplis, 2007).

The region between 850 - 1300 cm^{-1} is denoted as the high wavenumber (HW) region in previous studies on Raman-glass signals, where antisymmetric Si-O stretching in Q^1 contributes to signals near 880/900 cm^{-1}. Single or doublet peaks occurring between ~ 950 and 1000 cm^{-1} in Fe-bearing glasses are described as "ferric iron" bands. This feature is seemingly induced by a combination of Q^2 Si-O oscillations in the network, and Fe^{3+}– O stretching as a function of overall fugacity of the glass (Mysen et al., 1982; Wang et al., 1995; Magnien et al., 2006). Along with the network activity described by the Fe^{3+} peak between 950 and 1000 cm^{-1} and the oscillations of Si-O linkages, Raman signals up to 1300 cm^{-1} are assigned to both T-NBO and T-BO stretching vibrations (Furkawa et al., 1981). Raman data of volatile-free silicate glasses are typically featureless beyond ~1200/1250 cm^{-1}.

Multivariate Analytical Techniques

Partial Least Squares (PLS) is a popular statistical analysis modality used to investigate the relationships between a dependent variable, and one or more independent variables. PLS is a modified form of multiple linear regression (MLR) (Eq. 4) and has

been a widely utilized algorithm in multiple disciplines. PLS is a class of statistical algorithms of supervised machine learning and a multivariate analytical technique (MVA). It is a subvariant of Least Squares Regression (Eq. 5), in which explanatory variables ($f(X)$) are predicted from a set of observational data (X_j) (Stone and Brooks, 1990).

In the case of studying glass using Raman spectroscopy, the collected spectral channels serve as the observations, and the compositional data (SiO_2 wt%, CaO wt%, etc.) are the explanatory variables being predicted. In least squares, the unput matrix (X), has dimensions equal to the number of spectra (n), and spectral channels (p). Following equation 5, the input variables (X_j) are the collected spectra channels, and take the vector form $X^T = (X_1, X_2, ..., X_p)$, where the length of the input is the number of spectral channels from the collection (p). β_j are the weighted PLS coefficients taking the $\beta = (\beta_1, \beta_2, ..., \beta_p)$, used to predict the compositional variable $f(X)$. These coefficients are estimated by training the PLS algorithm to the spectra and compositional data and are the coefficients that best minimize the sum of the squared residuals. This means, more practically, that a trained PLS model creates coefficients (β) for a predicted variable (SiO_2 wt%) that are correlated to specific spectral channels (X_j) along the profile and places the largest coefficient values on the channels with the highest correlation and variance with the predicted variable. This can be visually transposed on top of a typical X-Y spectral diagram and allows for easy identification of the spectral features that are the most heavily weighted for a particular compositional variable prediction. As an example, if a PLS model is trained on a set of Raman spectra profiles with associated geochemical metadata, then the algorithm will produce a coefficient matrix that shows which regions of the Raman spectrum are most important to predicting a geochemical variable along the entire range of the collected profile.

$$Y = (\beta_0 + \beta_1 X_1 \dots \beta_N X_N) \quad (4)$$

$$f(X) = \beta_0 + \sum_{j=1}^{p} X_j \beta_j \quad (5)$$

The algorithm has multiple extensions and variations, though a popular version of the base technique is PLS-1. PLS-1 is the application of PLS in which the response of only one dependent variable is being measured (predicted) at a time. While multivariate predictions are possible (PLS-2), they are rarely necessary and often lack the accuracy associated with univariate prediction sets (PLS-1) (Dyar et al., 2012). PLS is similar to Principal Component Analysis (PCA), another MVA technique, but has distinct advantages in handling in handling correlated datasets. PLS-1, in the scenarios described below in the following chapters, allows Raman spectral channels to serve as the length of input, assumes they are correlated, and reduces them in the algorithm to create a wieldier expression of prediction when compared to MLR or PCA. This projects the data into the new dimensional vector space X where model complexity is regulated by the number of components (q), effectively removing problems that involve multicollinearity. Components are both a model output and a model parameter. Following the process of cross validation, a specific number of these components is recommended. The components regulate the degree of shrinkage in the final trained regression, while also providing a measure of the regression's ability to model external datasets not used for calibration.

Least Absolute Shrinkage and Selection Operator (LASSO) regression analysis is another supervised machine learning technique useful for multivariate problems that require more robust analyses than simple linear regression can provide (Hastie et al., 2009). Similar to PLS, LASSO regression is an optimized form of MLR (Eq. 4) in which the sum of the squared residuals in the equation is reduced implicitly. However, unlike PLS regression, the LASSO algorithm does not utilize projection to achieve this, reduction and instead seeks to mathematically constrain the values and number of the correlation coefficients (β) in the original MLR matrix. This "shrinkage" of the MLR equation is achieved by 1) calculating β according to (Eq. 6) under the constraints of (Eq. 7), in which "t" is a coefficient boundary limit that dictates the final summed *value* of the coefficients in the expression; and 2) the "α" parameter – a regression value akin to "k" in PLS that determines the final *number* of coefficients, and is determined by cross validation. β_{LASSO} is the estimated coefficient, N = number of coefficients, y_i is the

sample metadata information within each spectra, β_j is the β value for spectral channel *j*. β_0 is the regression coefficient value at the intercept in the matrix. x_{ij} is the intensity at channel *j* for sample/spectrum 'i'. Due to this shrinkage, LASSO regressions produce and use a much smaller number of coefficients in the prediction of a targeted variable. In the case of predicting geochemical variables from Raman spectral data, the smaller number of coefficients are assigned to the channels determined to be most important to that variable's prediction, and all other coefficients are sent to a value of zero. This creates a prediction model that uses "hard" feature selection, rather than the "soft" selection used by PLS, in which every spectral channel is assigned a regression coefficient, even those that may have negative effects on model accuracy. LASSO regression judges which channels are more important to the prediction of a variable (either positively or negatively), and then uses those coefficients to predict the concentration or value of the selected geochemical variable. The shrinkage function in LASSO modeling allows for a more robust prediction model but has also been shown to exclude important data features and lead to poor accuracy in some spectrographic-application scenarios (Dyar et al., 2012).

$$\beta_{LASSO} = arcMin_{\beta} \sum_{i=1}^{N}((y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\,\beta_j))^2 \quad (6)$$

$$\sum_{j=1}^{p}|\beta_j| \leq t \quad (7)$$

Put simply, LASSO models perform a non-projection based, shrunken form of multiple linear regression to predict an assigned variable in which only the most informative Raman spectral channels are used. In this case, LASSO models can predict variables, such as $\%Fe^{3+}$, or SiO_2 wt%, by performing MLR on coefficient values in the original matrix that have been assigned to fall within a set range. The technique has shown tremendous success in applications to solving spectral problems (similar to PLS) and has been utilized in the predictions of chemical variables from many different types of geologic spectra, including rocks in Laser Induced Breakdown Spectroscopy (LIBS), minerals in Raman spectroscopy, and glass in X-Ray Absorption Spectroscopy (Dyar et al., 2012; McCanta et al., 2015; Dyar et al., 2016; Breitenfeld et al., 2018). However,

LASSO prediction models have not yet been tested on the Raman spectra of silicate glasses.

Essentially, both LASSO and PLS. algorithms are constrained versions of MLR that implement least squares principles, offering a more dynamic range of functionality when compared to normal linear regression methodologies. The degree and style of their constraints from MLR are regulated by the cross-validation recommended *q* and *a* parameters whose acceptable values are established in the current literature (Stone and Brooks, 2009). Also, when constructing predictive models with such techniques, the desire is typically to build and test the method on all available data, rather than to hold any data out. However, without "unknown" data points to validate a prediction model's effectiveness, the true power of the calibration is unknown. Therefore, when collected datasets are sufficiently large enough to enable such withholding, it is customary to exclude roughly one-third of the data. The Raman spectroscopic data presented in Chapter's I and II were deemed suitable for such treatment, while the smaller sample suite investigated in Chapter III was not.

Error analysis in MVA techniques

To measure, contextualize, and compare prediction model quality throughout the following chapters, four standardized statistical error metrics are calculated. These metrics help to objectively quantify internal and external predictive accuracy and uncover chemical-related influences in the collected Raman spectra through investigations of their trends. The metrics include the Root-Mean-Square-Error (RMSE) of calibration (RMSE-C), cross-validation (RMSE-CV), prediction (RMSE-P), and normalized prediction (NRMSE-P). RMSE is a useful metric for constraining prediction accuracy because it is generated in the same units as the original measurements. The metric is widely used in reporting supervised machine learning model errors, as it allows for external comparisons to be made for models produced on similar data in other experiments. RMSEs are not directly representative of the error a single prediction has, but rather showcases an average measure of the predictions within a model; meaning an RMSE of 5% does not indicate each prediction in that model has an error of 5%, but that the model varies ~5% when compared to perfect predictions (i.e., RMSE = 0). It is a multifaceted statistic that

helps track model effectiveness across the standard machine learning process, including 1) training a machine learning model algorithm to predict variables within a training dataset, 2) cross-validating the model iteratively on that same data in various combinations, and 3) applying the trained model to unseen testing data. The four metrics are based on the fundamental idea of computing the square root of the mean residuals for each data point in a series (Eq. 8) where N is the number of data points in the series, y_i is the predicted variable value, and x_i is the actual variable value (Hastie et al., 2009).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - x_i)^2}{N}} \quad (8)$$

RMSE-C refers to a first-order prediction error associated with training a model to a set of data. It is only a measure of how well an applied modeling technique predicts information within that dataset and does not denote how well a model performs on "unseen" data, i.e. data that has not been used to train the model. Denoting RMSE-C as a comprehensive measure of model predictive accuracy has been shown to be statistically unsound and can lead to confusion in external dataset comparisons (Dyar and Ytsma, 2021). RMSE-C is only useful in instances where the targeted data cannot be separated into training and test-sets. In this study, RMSE-C is used to measure how well the univariate PLS/LASSO regression predicts the targeted variables (i.e., oxides, Fe-redox variables, or % tephra) from the Raman spectra in a training set (n_{train}) as an initial step towards applying the model on unseen data in the testing set (n_{test}). Test-set Raman spectra are treated as virtual 'unknowns', where they are withheld from the model, and are only used to test the calibration's ability to predict externally.

RMSE-CV, the error following the cross-validation process, is as an effective pseudo-test error that signifies the uncertainty of the calibration parameters (in PLS - *q*, in LASSO - *a*), and contextualizes how well the model would perform on future applications to unseen data. The version of cross validation used in this work is explored in-depth in Chapter I. RMSE-CV is generated by computing the average of the mean squared errors between the actual and predicted values in the internal training model iterations (Hastie et al., 2009).

RMSE-P, the error of prediction, represents the absolute fit of a model to unknown data, where predicted values are compared to true ones following model application to n_{test}. It is a robust measure of accuracy in easy-to-understand units that can be readily compared between studies that choose to report it (i.e. Dyar and Ytsma, 2021).

NRMSE-P, the normalized RMSE-P, is provided to standardize prediction errors between variables containing different value ranges. A variable with a larger or smaller range of possible values can render simple RMSE-P comparisons in a study less valuable. As an example, a model produced RMSE-P of 2.0 wt%, for a compositional variable whose training data spans values of 5.0 – 10.0 wt% is not as accurate when compared to a variable with a range between 1 – 100 wt%. In this case, the 2.0 wt% error occupies a larger portion of possible variance within the smaller training range (only 5 wt%) when compared to the larger 99 wt% in the other variable. Currently, there are no standardized normalization procedures required to produce an NRMSE-P metric, but the one chosen for this study was calculated using (Eq. 9). This normalization procedure places the prediction error within an appropriate scaled range present in the observed training values for a particular variable, where $(y_{max} - y_{\min})_{obsv}$ denotes the difference between the actual maximum and minimum range values. NRMSE-P can be reported as a decimal or a percentage, between values of 0.0 – 1.0, or 0 – 100%. However, reporting NRMSE-P in percentage form can be misleading, as its value is not a direct ratio.

$$\text{NRMSEP} = \left[\frac{RMSEP}{(y_{max} - y_{\min})_{train-obsv}}\right] \quad (9)$$

NRMSE-P is the best recorder of model prediction error but is not itself a hardline uncertainty metric; rather, NRMSE-P expresses the spread of possible variance denoted by the RMSE-P within the range of training values. This means that, if NRMSE-P is = 0.20 for a model, then the predictions within that model can vary up to 20% on average from the actual value. However, this does not mean each prediction will vary up to 20% from the actual value. NRMSE-P is used in this work to identify trends among the models in, and for uncovering model uncertainties. The threshold for statistically sound NRMSE-

P values is ≤ 1.00, as when NRMSE-P approaches or exceeds 1.00. Predictions can vary at values equal to the entire range for that variable, which would lead to less reliable predictions on average.

Prediction coefficients (R^2) have also been recorded to aid in visualizing model accuracy. However, R^2 has been shown to be an erroneous metric in applications of supervised learning methods such as PLS or LASSO (Breitenfeld et al., 2018; Dyar and Ytsma, 2021) because it ignores covariance scaling and depends on the error associated with each individual measurement. R^2 is only a relative measure of fit to a defined regression trend and is not provided in the same units as the response variable. Therefore, the measure is not used in this work.

All MVA analysis referenced in the following chapters was undertaken using the Data Exploration, Visualization and Analysis for Spectroscopy (DEVAS) system (http://nemo.mtholyoke.edu/). DEVAS is a multi-tool web portal platform hosted and operated by researchers at Mt. Holyoke College. The system is an open-source processing and modeling platform that allows users to ingest and transform spectroscopic data, then use that data to train and test PLS and LASSO prediction models for included variables. The system was employed in the following chapters to ingest, visualize, fit, pair, and pre-process the collected Raman spectroscopic data for use in building prediction models. DEVAS allows users to baseline, smooth, truncate, normalize, and squash spectroscopic data while simultaneously observing the processing's effects. Batch and individual spectral peak fitting are also possible using the system. The spectral preprocessing selections discussed in this work, along with the PLS and LASSO modeling parameters and errors, are the final product of extensive rounds of testing using a grid-search iterative approach (explained in detail in Chapter I).

CHAPTER I

QUANTIFYING GEOCHEMICAL OXIDE CONCENTRATIONS IN SILICATE GLASSES USING RAMAN SPECTROSCOPY AND PARTIAL LEAST SQUARES

A version of this chapter has been accepted with edits for peer-review publication in the journal Applied Spectroscopy, and was authored by Blake O. LaDouceur, Molly C. McCanta, M. Darby Dyar, Bhavya Sharma, Grace Sarabia, and Natalie Dunn.

Acceptable Citation: LaDouceur, B. O., McCanta, M. C., Sharma, B., Sarabia, G., Dunn, N., Dyar, M. D. (Accepted July 2023) A machine learning model for predicting geochemistry in silicate glasses via Raman spectroscopy: Implications for PLS applications to amorphous Raman spectra. *Applied Spectroscopy*

This research was originally submitted to Applied Spectroscopy in March of 2023 after undergoing multiple rounds of co-author edits and rewrites. The article was accepted for publication following recommended revisions on July 7th 2023, and is currently being revised based on peer review feedback for resubmission in August 2023. Blake LaDouceur, the first author, is the corresponding author for the submission process and is the primary contact for revisions from co-authors and reviews. Blake was the primary individual responsible for the content of the manuscript, including: data modeling, sample data collection, data processing, writing, and figure/table development. Molly McCanta provided the main source of comments and revisions among the authorship team and was the singular liaison between collecting co-author ideas throughout the entirety of the research effort. Bhavya Sharma, Grace Sarabia, and Natalie Dunn served integral roles in spectroscopic data collection, processing tests, and discipline-related revisions surrounding the Raman data. M. Darby Dyar provided the initial glass sample suite, along with important methodology-related edits.

Abstract

Here, univariate partial least squares (PLS-1) regression is used to develop a modeling suite capable of reliably predicting oxide geochemistry in volcanic glass samples via their Raman spectra. The study yields eight, oxide-specific models that allow predictions of SiO_2, Na_2O, K_2O, CaO, TiO_2, Al_2O_3, FeO_T, and MgO (wt%) in glass samples that span a range of igneous compositions, while also providing PLS coefficient that help highlight which Raman spectral features are related to oxide bonding influences. The regression was trained on 48 of the 69 total glasses for each oxide variable, and then tested against 21 validation samples (i.e., held-out of training). Trends in RMSE-C/CV/P model accuracy metrics are investigated to uncover the efficacy and limitations of utilizing PLS for oxide Raman applications and are contextualized against recently produced strategies. The technique produces a fully generalized calibration in which each model is built on samples spanning the entire compositional range of typical volcanic rock types (basalts – rhyolites) along reduced and oxidized conditions. The technique yields an average root-mean: training error of ~2.4 (wt%), cross-validation error of ~ 2.9 (wt%), prediction error of ~ 2.6 (wt%), and normalized prediction error of ~ 28%. This suite represents the best predictive accuracy possible when balancing considerations for the corrected Raman signal and the precedents of multivariate analysis (MVA) in similar circumstances. This balance is quantified through baseline quality metric investigations using grid-search style testing, and through visual inspection of the effects expressed in the processed Raman signals. Raman band positional shifts are also mapped against underlying chemical variations while holding oxidation state constant; with major influences arising primarily as a function of silica, calcium, iron, and alkali content. A standardized preprocessing scheme for Raman-glass spectra is provided, along with the easily applied PLS models. Additionally, the oxide-model suite is applied to a secondary external Raman dataset of 11 reduced basaltic glasses in order to further refine the use of the technique and provide context for optimization controls. The model's ability to constrain information on the external dataset was seemingly hindered by the lack of targeted sample diversity, yielding an average NRMSE-P of ~61% when applied directly. However, when the models were constrained to include only basaltic samples (along

multiple oxidation states), prediction errors decreased to ~30%. This study represents the first time Raman spectra of amorphous silicates have been paired with PLS algorithms, offering a foundation for future improvements utilizing these systems together.

Introduction

Volcanic glasses represent rapidly quenched fractions of magma and preserve pre-eruptive silicate melt chemical properties and are critically important for volcanologists to study and characterize large-scale processes, like eruption style. Advances in molecular spectroscopy and computer science technologies have revolutionized the ways in which geoscientists can study igneous systems and target volcanic glasses, such as modeling Raman spectroscopic data. Raman spectroscopic systems offer robust analytical capabilities for unravelling chemical compositions in almost any material, and, when paired with suitable modeling techniques, can yield reliable quantifying attributes. The adoption of Raman spectroscopy stems from technological improvements in operating ranges, costs, probing power, and their utility in extreme environments. However, currently available glass modeling methodologies from Raman spectroscopic data have been limited because they rely on restrictive sample compositions, complex spectra, and chemical prediction parameterization techniques (Di Muro et al., 2009; Di Genova et al., 2015; 2016a; 2016b; Giordano et al., 2019; Le Losq et al., 2019; González-García et al., 2020). Improving predictive silicate melt discrimination from glass is important for future scenarios where hands-on field characterization and laboratory work are impossible, for example, on other terrestrial planets whose surfaces record significant volcanic activity.

The main purpose of this study was to pair new synthetic and natural silicate glass experiments (n = 69) from a confocal Raman spectroscopy using a green-laser source (532 nm) with a partial least squares (PLS), supervised machine learning modeling algorithm to produce higher quality glass composition characterization for eight major oxide constituents (SiO_2, Na_2O, K_2O, CaO, TiO_2, Al_2O_3, FeO_T, and MgO). To avoid the limitations of previous calibrations, the modeling predictions were built upon a highly variable glass suite and were internally and externally validated against multiple pairings of uncharacterized spectra. Prediction accuracy was contextualized against Raman spectroscopic band parameterizations along chemical influences, in an effort to uncover strategy shortcomings and advancements. Raman spectra preprocessing options are provided that can be replicated in any collection environment, which should help to move towards a standardized protocol for combining Raman spectroscopy and machine

learning approaches that can solve volcanological problems in both terrestrial and extraterrestrial environments.

Methods

Silicate glass sample suite production

Glass materials were produced by Dyar et al. (2016) after measuring and weighing Alfa Aesar Puratronic carbonate and oxide powders into one of five compositional groups ranging from silica-poor basalts to silica-rich rhyolites (Figure 1.1). Briefly, each of the oxide mixes was ground for one hour under ethanol in an agate mortar and then was decarbonated for two hours at a temperature of 800 °C. To equilibrate each glass-mix at the desired range of redox conditions, the 1-bar wire-loop technique was used in a vertical one-atmosphere gas mixing furnace at Tufts University. Depending on the target fugacity (f_{O2}) for a sample, either a platinum (Pt) or a rhenium (Re) wire was used in the experiment. About 100 mg of sample powder was mixed with polyvinyl alcohol to adhere each sample to either wire loop in the experiments. High silica experiments proceeded with an initial glassing step, and low silica runs utilized the oxide mixes directly. Glass samples were equilibrated at f_{O2} values that ranged from highly oxidizing using pre-determined gas mixtures of CO_2 and air, to reducing quartz-fayalite magnetite (QFM), molybdenum oxide (Mo-MoO_2) and iron-wüstite (IW). Reduced runs at QFM and below were performed on a Re-loop due to the extremely low solubility of iron in these conditions (e.g., Borisov and Jones, 1999). For the remaining f_{O2} values, experiments were run using Pt-loops that had been pre-doped for six hours at a temperature of 1250 °C. Melts were rapidly quenched in water and sectioned to create the final phases. The samples were placed in sealed vials and separated into multiple flakes for electron microprobe, Mössbauer spectroscopy, and X-ray absorption spectroscopy analyses to obtain their geochemical compositions by Dyar et al. (2016).

In total, a suite of 69 subalkaline, anhydrous, Fe-bearing synthetic glasses were generated (Figure 1.1) (Dyar et al., 2016). The set included 21 basalts (BAS), 10 basaltic andesites (BAA), 16 andesites (AND), 15 dacites (DAC), and 7 rhyolites (RHY). Compositions ranged from 5.6 ± 4.72 to 0.29 ± 0.03 wt% total alkali ($Na_2O + K_2O$), 43.45 to 8.02 wt% silica (SiO_2), and 0.08 to 1.0 Fe^{3+}/Fe_{TOT}. The samples were also evenly

split regarding f_{O2}, with 35 reduced and 34 oxidized samples (Table 1.1, Table 1.2). No major crystallization, alteration, or devitrification was recognized in the glasses or on their surfaces (Figure 1.2A–F).

Raman data collection and preprocessing

Raman spectra were collected from the developed glasses using a homebuilt confocal Raman microscope managed and operated by the Sharma Raman Lab (https://sharmalab.utk.edu/) at the University of Tennessee, Knoxville, Department of Chemistry. The Raman microscope consists of a 532 nm optically pumped semiconductor laser (Coherent, Inc.) and a Nikon Ti-U inverted microscope with a 0.45 numerical aperture (n.a.). Raman scattered light was collected by a 20× objective in a 180° backscattering geometry and exited the microscope before being focused onto the 100 μm slit of an ISOPlane SCT 320 spectrometer equipped with a PIXIS 400 charge-coupled device (CCD) detector. Raman spectra were excited at 532 nm with 10 mW of power using an ~5 μm diameter spot at a distance of a few millimeters from the glass samples. Individual glass samples were secured to 1" × 1" reflective aluminum squares with tape, placed on the microscope stage, and optically focused prior to data collection. Raman spectra were collected for each sample in triplicate (t = 40 s per Raman spectrum) for a total 120 s. This resulted in 207 Raman spectra, over 519 channels (p). Data were collected using the LightField® spectroscopy software (Teledyne Princeton Instruments). The collected replicates typically result in identical spectra but allow for identification of random spikes or evidence of photodegradation of samples. The frames can be averaged, or used individually, as required by the study. Here, the three frames were averaged per sample, and the resulting single data stream was used as that sample's Raman spectrum.

Preprocessing non-crystalline Raman spectroscopic data is an absolute necessity, given the inherent broad nature of the features when compared with their crystalline counterparts. The spectra bands are broad and ill-defined and lack narrow peaks and shoulders that would help to produce coherent correction methodologies. There are currently no standard pre-processing protocols for such spectra, which gives the operator free range to determine their own procedural choices for normalizing, truncating, baselining, and smoothing the data. Preprocessing the Raman spectra of silicate glasses

typically considers: (1) spectrometer operating conditions, (2) the geochemistry of the samples, (3) observed raw Raman features, (4) effects on modeling prediction accuracies, and (5) any system collection interferences (McMillan and Piriou, 1983, 1984; Rossano and Mysen, 2012; Schulze et al., 2012; Neuville et al., 2014; Ning et al., 2014; Qian et al., 2017). However, some consistent preprocessing options have appeared in recent experiments, and include options like: the Long correction (Long, 1977), Whitaker-based smoothing, linear, polynomial, or cubic baseline subtractions, maximum-intensity or peak feature normalization schemes, and truncations between ~350 – 1250 cm^{-1} (Dünnwald and Otto, 1989; Baert et al., 2011; Di Genova et al., 2015, 2016a, 2016b, 2017; Welsch et al., 2017; Giordano et al., 2019; Le Losq et al., 2019).

Although truncation, smoothing, and normalization schemes for Raman spectroscopic data are important because they affect final data quality, selecting these schemes relies on which Raman band regions are of interest to the study and what type of isolation is required. Conversely, baseline subtraction choices are singularly critical and can have large effects on the reliability of the final Raman spectra. For recent Raman-glass modeling publications, baselining procedures have been heavily reliant on visual inspection of the correction's effect on spectral profiles. When pursuing batch reproducibility in predictive modeling, selecting a baseline, which can be subjective, can complicate correction applications for other datasets (Carey et al., 2015). In spectroscopic studies that employ MVA modeling techniques however, it is an established practice to choose baseline corrections that result in the lowest errors for a given prediction (Liland et al., 2010; Schulze et al., 2012; Giguere et al., 2017; Carey et al., 2015). Visual inspection is still done in such cases, but the resultant Raman spectra may end up overcorrected in exchange for higher predictive accuracy. This is also an issue. Therefore, baseline selections on silicate glass Raman spectra should strike a suitable balance between visual inspection considerations, and statistical precision.

The 69 glass spectra were initially treated in MATLAB software with a Median Filter (MFilter) noise/temperature/cosmic ray suppression based on a 2:9 Savitzky-Golay smoothing function, resulting in a consistent signal to noise ratio across the collected spectrum of ~400 – 1200 cm^{-1}. The raw anti-stokes Raman data for these glasses

exhibited no signal activity below 400 cm^{-1} or beyond ~1150 cm^{-1} in the profiles. The smoothed Raman spectra were paired with the sample geochemical metadata and input into the Data Exploration, Visualization, and Analysis for Spectroscopy (DEVAS) system (http://nemo.mtholyoke.edu/). Using the system, the 400 – 1150 cm^{-1} region of the Raman profiles were further pre-processed for PLS modeling using an asymmetric least squares (ALS) baseline function (Eq. 1.1) (Eilers and Belens, 2005) with a parametrized smoothness factor (λ) = $5.754e^{+4}$ and asymmetry (p) = 0.002, as well as a maximum feature normalization between 0 and 1 ($Norm_{Max}$) [e.g. Whiteman et al., 1993) (Figure 1.3A). These options were selected based upon an iterative grid-searching methodology, similar to the one described in Liland et al. (2010), in which endmember correction selections whose resulting Raman profiles were visually acceptable were identified initially. After the extremes of preprocessing possibilities were identified, further, smaller refinements to the corrections were tested in an effort to narrow in on the options within those boundaries that produced the lowest prediction errors (RMSE-P's). The ALS baseline algorithm uses a 2nd derivative-constrained weighted regression to automatically select anchor points along a given spectra profile for subtractions in a batch process. The ALS and $Norm_{Max}$ parameters helped eliminate irrelevant signals and Raman band shape disparities in the highly complex signals, while preserving predictive information contained in the glass bands of interest separately. The correction is akin to the baseline provided in similar studies (Di Muro et al., 2009; Giordano et al., 2019; González-García et al., 2020) The technique has a proven record of success when paired with MVA/PLS in many spectroscopic investigations, and more recently, in Raman modeling applications (e.g., Carey et al., 2015; Liland et al., 2010).

$$F = \sum_i w_i (y_i - z_i)^2 + \lambda \sum_i (z_i^{''})^2 \quad (1.1)$$

Building PLS models for oxide predictions

The 69 Raman spectra were randomly separated and selected for modeling into a training set (n_{train} = 48) and validation test set (n_{test} = 21) being sure to have stratified populations of both f_{O2} and wt% SiO_2 variability in the suites. Prediction models were constructed by cross validating, then training, the PLS-1 algorithm to predict a single

oxide variable from the training set spectra (n_{train} = 48), using the "Train Regression Models" functionality in the DEVAS system. Cross-validation is a method of testing the regression on a specified number of iterations (folds) of the training data prior to applying that regression on uncharacterized data. It is recognized as a statistical necessity in predictive modeling studies and helps ensure the optimal set model parameters are selected for a particular variable, helping to avoid subjective choices of parameters that can lead to harsh over, or under, fitting of the dataset (Hastie et al., 2009; Stone and Brooks, 1990). CV errors are used to estimate how accurately a predictive model will perform in practice on external data, and the process of CV produces a model component value that is used to train the model for each variable.

The Leave-One-Out Cross-Validation (LOO-CV) technique was used to iterate several variations of the model within the system automatically and determined the optimal number of PLS components (*q*) that yield the most accurate predictions. Cross validation is dictated by the number of folds (*k*) chosen by the user. *k* is a primary control over how the model is validated against the training data, by regulating the number and style of iterative testing scenarios. This methodology separates the training data suite into *k* number of subsets, and the regression model is tested *k* number of times on those subsets to find the optimal number of *q*, which are used to train the model and regulate algorithm shrinking when compared to the normal M.L.R. formula. In LOO-CV, *k* is set equal to the number of samples in the training set (i.e., *k* = 48). This process was repeated for each oxide variable, producing a unique value of *q* for each model. Following cross-validation, the recommended number of PLS model components (*q*) for each oxide was employed into the regression (Table 1.3), and the resulting final regression model was prepared for application to n_{test}. These steps produce a model for each oxide built on a data suite of Raman spectra that collected from combinations of mafic to felsic and reduced to oxidized compositions (Table 1.1). The strategy then generated a "general" or "universal" modeling suite, in which the regression was trained on a Raman spectroscopic dataset comprised of five major extrusive igneous compositions, along either reducing or oxidizing conditions. This ensured the final oxide models were applicable to nearly any possible encountered melt/glass composition and were capable

of predicting wt% concentrations of SiO_2, Na_2O, K_2O, CaO, TiO_2, Al_2O_3, FeO_T, and MgO.

Results

Evolution of Raman signals as a function of oxide chemistry

Although the shape, intensity, and positions of silicate glass Raman spectra can vary among studies, spectra are usually presented as three primary bands with features in the low-wavenumber (LW: 400 – 650 cm^{-1}), mid-wavenumber (MW: 650 –850 cm^{-1}), and high-wavenumber regions (HW: 850 – 1200 cm^{-1}). The collected Raman spectra in this study also exhibited three major bands that were centered in the LW, MW, and HW envelopes (Figure 1.3B). In the LW region, a low-energy band was centered at ~ 500 cm^{-1} between shoulders at ~ 400 and 550 cm^{-1}. In the MW region, the spectra exhibited a broad arc feature between strong shoulders at ~ 790 and 870 cm^{-1} and was centered at ~830 cm^{-1}. The HW region had a more defined peak feature, whose singular centroid occurred near ~980 cm^{-1} (Figure 1.3B). Yet, the overall feature shapes from glasses used in this study were broader and more ill-defined when compared to chemically similar crystalline samples (i.e., rocks, minerals), in which Raman features are defined by narrow peaks with strong, straighter shoulders.

To explore Raman spectra signal evolution from the 69 glass samples with different overall silica oxide concentration chemistry, as has been done in recent experiments (e.g., Le Losq et al., 2019), this work used the DEVAS system to isolate and map Raman spectra band centroid positional shifts as a function of individual oxide concentrations via batch Gaussian fitting. Although glass structure could not be directly inferred using this approach, the Raman spectra could uncover the effect of individual chemical components on composition and provide context for PLS prediction models. Both reduced and oxidized glass samples exhibited a shift in the LW band centroid towards lower wavenumbers as SiO_2 wt% increased (Figure 1.4). However, notable influences in this band region also resulted from the alkali elements, Na_2O (Figure 1.5D) and K_2O (Figure 1.5E), which shifted towards lower wavenumbers, and other oxides CaO, FeO_T, and MgO (Figure 1.5F, C, G, respectively), which shifted towards higher wavenumbers. The LW band showed no shift as a response to increased Al_2O_3 or TiO_2

concentrations (Figure 1.5B, A, respectively). For the MW band region, there appeared to be competing influences from multiple oxides. Oxidized glass samples had a larger range of centroid band positions compared to reduced samples, which has also been observed by others (e.g., Di Genova et al., 2016a; Giordano and Russel, 2018; González-García et al., 2020). In particular, the center of the MW band exhibited a very slight shift towards lower wavenumbers as reduced glass SiO_2 concentration increased, but there was no clear band shift among glass samples isolated along oxidized compositions (Figure 1.4). Otherwise, the MW band was unaffected by any singular oxide component, with all spectra recording a centroid near ~835 cm^{-1} (Figure 1.5A-G). In the HW region, the Raman band centered near ~975 cm^{-1} shifted towards lower wavenumbers as SiO_2 concentration increased for both reduced and oxidized samples, although oxidized samples had higher variance in their center band positions (Figure 1.4). This HW Raman feature was also sensitive to the alkali elements, which resulted in a negative position shift for reduced glasses (Figure 1.5D, E). With higher FeO_T and CaO concentrations for the reduced glasses, the HW band shifted towards higher wavenumbers (Figure 1.5C, F respectively). Moreover, based on previous Raman spectral collections on Fe-bearing glasses (i.e., Giordano et al., 2020), the ~980 cm^{-1} HW band for the iron-bearing samples reinforced the existence of a "Ferric Iron Band/Peak."

Oxide model prediction accuracies

LOO-CV resulted in PLS component values of $q \leq 6$, on an acceptable possible scale of 2 – 10 and an average recommended value of ~ 3 (Table 1.3). Values of q in PLS exceeding 10 are considered overfitted (Stone and Brooks, 1990), such that models using these values are typically only applicable to the dataset they were trained on and may not be capable of describing uncharacterized data. Therefore, a maximum $q = 6$ was a suitable result. SiO_2, CaO, and TiO_2 exhibited the lowest cross-validated component values with $q = 2$, whereas the remaining five variables had $q \geq 3$, with K_2O at the highest with $q = 6$ (Table 1.3). These values were used to train the final regression calibration for each oxide PLS model, although it is possible to control the selection of q manually in PLS to achieve higher internal predictive accuracy. Manually selecting PLS components guarantees mixed results (Dyar et al., 2012). Therefore, based solely on the

recommendations of cross-validation techniques, the component values of $q = 2$ or 3 for SiO_2, CaO, Al_2O_3, Na_2O, and TiO_2 represent a highly applicable model, but may be prone to higher internal errors. Similarly, the higher q values recommended from LOO-CV for K_2O, FeO_T, and MgO indicated the regression may require slight overfitting to produce lower errors and could be less efficient at predicting oxides from external datasets (i.e. Hastie et al., 2017).

Figure 1.6 shows the predicted vs. actual wt% concentrations for the selected oxides and Table 1.4 provides the predictions, with model vs. microprobe values for each oxide alongside the difference (+ or -) between them. Oxide error metrics in Table 1.3 showed three major trends: 1) errors of SiO_2 concentrations were regularly the highest among the metrics, with RMSE-C = 7.89 (wt%), RMSE-CV = 9.25 (wt%), and RMSE-P = 7.40 (wt%); 2) the entire modeling suite had relatively consistent errors, with comparable RMSE-C/CV/P values for each oxide; and 3) NRMSE-P variance was also extremely consistent, and hovered around ~ 0.26 for all oxides except Al_2O_3 at ~ 0.46. The models had similar predictive accuracies regardless of the concentration range present for any random oxide in that series, which indicated the regression was optimal and did not overfit the response variables. The PLS modeling suite yielded an average NRMSE-P of ~ 0.28, or ~28% error. The model for Al_2O_3 predictions yielded the highest NRMSE-P value at ~ 0.46 for Al_2O_3. TiO_2 prediction errors were the most consistent, with RMSE-C/CV/P errors near ~ 0.65 wt%. The eight PLS models showed affinity for over- or under-estimating any single oxide, with all oxide predictions having a relatively equal spread of such predictions and indicating regression consistency. One of the glass samples, a dacite equilibrated in air (DAC4_Air), was characterized by two erroneous negative predictions for its CaO and TiO_2 concentrations (Figure 1.6). No issues were uncovered for this spectrum or sample, and it was therefore retained in the testing data set. In this use of this general modeling suite, there were no noticeable prediction accuracy trends along glass composition, as the PLS regression was trained and tested against all available compositions.

Correlations between glass spectral bands and oxide components via PLS coefficients

PLS regression-produced coefficients (β) were used to examine the presence of individual oxides at specific Raman channels (*p*) and their relationship to the observed Raman features. In Figure 1.7, the coefficients were plotted at each channel along the collected Raman region 400 – 1200 cm^{-1} at specific wavenumber positions for each PLS model. The magnitudes (i.e., height above or below the coefficient "0" point, either positive or negative) and positions (i.e., wavenumber position) of the PLS coefficients corresponded to the relationship between an oxide and a Raman channel as determined by the algorithm. The sign indicates the positive or negative correlation of an analyte to a specific Raman channel, and the magnitude of the sign from zero expresses the scale of that correlation. Positive component values indicate an oxide is present or exerts network bonding influence at a specific Raman channel, while negative values express the inverse. Values near zero indicate no observable relationship. PLS regression seeks to place the largest coefficient values on the Raman spectral channels that have the highest correlation with the targeted oxide (Stone and Brooks, 1990; James et al., 2013). However, it must be noted that coefficients do not directly infer glass network structural information, as can be done with deconvolution into Q^n species (Breitenfield et al., 2018).

The entire wavenumber range between 400 and 1200 cm^{-1} contained useful information to predict targeted oxides, with only a small number of channels assigned to near-zero coefficient values. The LW Raman spectral band feature centered near ~500 cm^{-1} of these glasses exhibited overly positive correlations for SiO_2, Na_2O, and K_2O variables (Figure 1.7B, C, D) and negative correlations for CaO, TiO_2, Al_2O_3, FeO_T, and MgO (Figure 1.7A, E, G, H). Al_2O_3 coefficients were close to zero near this band (Figure 1.7F). The strongest positive correlations for this region were with SiO_2, with coefficient values reaching up to + 2.0 (Figure 1.7D), while the strongest negative values were associated with FeO_T and MgO variables, reaching ~ -1.0. PLS coefficients within the MW Raman band were highly variable, with swings between positive and negative values occurring along the region for most oxides (i.e. K_2O, Al_2O_3, FeO_T, and MgO). However, CaO, SiO_2, and TiO_2 model coefficients were less scattered, and were near

zero throughout the MW region. The HW Raman spectral region containing the "ferric iron peak" and up to ~1200 cm^{-1} exhibited high magnitude coefficients with positive and negative values (Figure 1.7). The primary channels that mapped heavily to the oxide components included the region immediately before the ferric iron peak center and immediately after (i.e., channels between ~ 900 and 970 cm^{-1}, and 985 and 1000 cm^{-1}). The band center near ~980 cm^{-1} mapped in all oxides to near zero coefficient values. CaO, FeO_T, and MgO variables were assigned strong negative values leading up to the peak, and then strongly positive coefficients after (Figure 1.7A, G, H). Na_2O, K_2O, and SiO_2 variables were described by the inverse relationship. Al_2O_3 and TiO_2 PLS coefficients showed similar relationships to that of CaO, FeO_T, and MgO, but were of smaller magnitudes. Lastly, there were distinct similarities in coefficient trends across the Raman spectra between two groups of variables, SiO_2, Na_2O, + K_2O, and FeO_T + MgO, whereas the remaining oxides (e.g., Al_2O_3, CaO, and TiO_2) exhibited unique trends. Al_2O_3 and K_2O model coefficients exhibited the largest magnitude scattering, which was in good agreement with the uncovered oxide-related Raman band positional shifts.

Discussion

Contextualization of oxide model accuracies and Raman-glass signal evolution

Some progress has been made in producing approaches to obtain geochemical information using the Raman spectra of silicate glasses, and the results of such prior investigations have set valuable precedents for how these approaches perform and can be improved upon in future experiments. The overarching goal of such experiments is to eventually produce a highly generalizable, uniquely applicable, and easy to use model for approximating major geochemistry data in samples using their collected Raman spectra. While recent efforts have been beneficial, they lack sophistication. Below is a comprehensive comparison between the modeling results and Raman signal evolutions provided in this work and that of three other prominent studies in the field. This will help to contextualize the improvements, shortcomings, and distinct advantages that the PLS approach produced here may offer. It must be noted that no currently available study has yet offered a modeling strategy reliant on as large, or as diverse a glass dataset as the one put forth in this work. Similarly, most of these experimental publications do not provide

simple error metrics or accuracy measurements – making comparisons between studies difficult. To make more meaningful model comparisons, RMSE's have been independently calculated from the chosen context studies' training and testing prediction data. This is explained in detail below.

The geochemical prediction model constructed for silicate glass samples initially presented in Di Genova et al. (2015), then later improved and expanded in Di Genova et al. (2016b), produces some of the most accurate measures of glass composition yet published using Raman spectroscopy, with extracted average RMSE-P's of ~1.71 mol% and 2.09 wt% in the 2015 and 2016 studies, respectively (Figure 1.8). Calculations of sample composition are made using a linear ideal mixing formula that constructs a "predicted" Raman spectrum from two predetermined endmember training spectra. The method requires the collection of more than one Raman spectrum from multiple samples to generate a reliable prediction for glasses that fall outside the predetermined chemistries provided. The studies make use of a "Raman fit" parameter (R_p), a metric that helps to produce/predict the composition of the targeted sample and is regulated by the initial training scenarios. Due to the need for an encountered sample to fall between the given training spectra from the glass chemistry (i.e., $SiO_2 + Al_2O_3 < 63$ mol%, and $FeO_T < 17$ wt% if basaltic or rhyolitic, and cannot be an iron rich rhyolite) an investigator must have prior knowledge of the expected compositions, and necessitates two separate collections be made on samples that fall on either side compositionally from the targeted sample's makeup if the glass does not meet these requirements. the authors' modeling strategies assumes any intermediate glass (i.e., andesitic, dacitic etc.) would have Raman spectra that could easily mix linearly between two endmembers. The models were built on 44 (Di Genova et al., 2015) and 186 (Di Genova et al., 2016b) individual Raman collection points along a basaltic to rhyolitic synthetic glass transect and tested externally on seven (Di Genova et al., 2015) and four glasses (Di Genova et al., 2016b). Although the 2016 study was able to produce predictions for all major glass forming oxides individually (SiO_2, Na_2O, K_2O, CaO, TiO_2, Al_2O_3, FeO_T, and MgO wt%), the 2015 strategy could only independently constrain FeO_T, and combined other oxides into composite models where they are predicted together. The calculated RMSE-P values given above are

determined from the available testing prediction data listed within the publication, and, for the 2015 publication, are averaged from the following individual model RMSE-P's: SiO_2 + Al_2O_3, Na_2O + K_2O, CaO + MgO, and FeO_T. As the Di Genova et al. (2015, 2016b) modeling publications do not provide actual vs. predicted matrices for their training data applications, internal RMSE-C metrics and normalized prediction errors cannot be calculated or would otherwise be inappropriate. Di Genova et al. (2015) and Di Genova et al. (2016b) also provide Raman signal evolutionary trends as a function of SiO_2 concentration for reduced glass samples in an effort to constrain the effect of polymerization on model effectiveness. In the collected region between 350 – 1250 cm^{-1}, SiO_2 concentration increases were shown to cause a negative band centroid shift in the LW, and positive shifts in the MW and HW (Figure 1.8).

More recently, González-García et al. (2020) produced a Raman spectroscopy-dependent model for predicting silicate glass geochemistry trained on a diverse calibration sample suite comprised of 31 natural glass samples along oxidized conditions (i.e., 11 basalts, 12 andesites, 5 dacites, and 3 rhyolites). The models are externally tested on a suite of 8 validation glass samples (7 basalts, 1 rhyolite), also along oxidized conditions. The authors use a series of empirical equations to relate the concentration of molar oxide contents to the normalized intensity ratio of the HW and LW Raman band regions (*R*). By utilizing a Spearman correlation coefficient to constrain the influence of independent oxide components on the observed Raman spectral features between ~250 – 1250 cm^{-1}, the authors can create specific relational parameters for each oxide to calculate unknown concentrations in testing spectra via the I_{LW}/I_{HW} ratio. Individual models for SiO_2, Al_2O_3, FeO_T, and MgO are produced, whereas Na_2O, K_2O, and TiO_2 predictions are incorporated into paired calibrations, Na_2O + K_2O and SiO_2 + Al_2O_3 + TiO_2. The authors provide internal RMSE-C errors within the text, but do not report prediction errors or total model errors using the same metrics, complicating and diluting their comparisons. The study does not consider varying redox conditions within the modeling strategy, nor in the Raman signal evolution discussions, further limiting the application to compositions outside of these constraints. The González-García et al. (2020) model yields an average RMSE-C of ~1.23 mol%, based on the individual RMSE-

C's provided in the text for each of the eight targeted oxide variables. RMSE-P's were independently calculated from the actual vs predicted values provided in the supplementary material, and averaged to produce an average RMSE-P of ~ 1.80 mol%. In total, the modeling suite averages a ~21% error when all individual oxide models are considered. % errors are provided by the author individually within the text, and due to the disparity in units and lack of public data from the study, NRMSE-P values were not calculated for comparisons to this study. Total model % error was averaged from the following reported metrics: ~15 – 40% for CaO, Na_2O, K_2O, and MgO; SiO_2 errors up to 10%, FeO_T up to 15%, Al_2O_3 at 20%, and TiO_2 up to 50%.

González-García et al. (2020) report Raman signal evolutionary trends as a function of increasing oxide concentrations for the observed features and is displayed in Figure 1.8. Only oxidized compositions were considered. Increasing concentrations of glass FeO_T, MgO, CaO, and TiO_2 shifted the LW band towards higher wavenumbers, but higher SiO_2, Na_2O, and K_2O concentrations caused the band center to shift lower. In the MW, higher SiO_2 content caused positive shifts, but higher CaO, MgO, FeO_T, and TiO_2 contents caused negative shifts. Increasing SiO_2 content caused the HW band center to shift towards higher wavenumbers, and CaO and FeO_T resulted in negative trends.

Although no modeling study has yet been published whose accuracy metrics are concretely comparable to the standardized quality reports offered by RMSE, the accuracies reported here from the models produced in Di Genova et al. (2015, 2016b) and González-García et al. (2020) show that the PLS-1 oxide modeling suite is only slightly less reliable (Figure 1.8). The average PLS NRMSE-P for this work hovered near ~28% and yielded an average RMSE-P = 2.63 wt%. These metrics were obtained from a modeling suite built and tested on sample suites containing 5 igneous compositions, along 4 distinct oxidation buffers, while the beforementioned studies contained limited compositions, and offered many distinct limitations in their applicability. Raman glass spectral evolutionary trends with oxide concentrations were very similar between the studies, though, again here, neither of the previous authors defined band positional trends along such varied compositions. PLS avoids the complex parametrizations required by previous experiments, and the covariance between the structural influence of oxides in

the glasses as recorded in the Raman features are analyzed and modeled simultaneously. PLS does however, have some notable limitations in handling the complex relationships between oxide-bonding and feature classes in Raman-glass spectra, and further optimization may be needed to help the technique approach its successes in other spectroscopic studies.

External model validation as insight for future PLS applications

Only one attempt has been made to model silicate glass properties via their Raman spectra using machine learning methodologies rather than the more traditional procedures described above. Le Losq et al. (2019) compared conventional peak fitting and intensity ratio methods to both supervised and unsupervised machine learning approaches to determine Fe-redox states for a suite of Mid-Ocean Ridge basaltic glasses using the 850 – 1140 cm^{-1} region of the Raman spectrum. The authors determined that the machine learning techniques tested (Alternative Least Squares Multivariate Curve Resolution (ALS-MCR), Neural Network, Kernal Ridge, and Support Vector regression) produced equivalent results for predictions of Fe^{3+}/Fe_{TOT} in their samples but were preferable due to their ease-of-use and flexibility to the Raman data. In an application using the Neural Network method, the authors also showed in the that estimating oxide concentrations with an average error of < 1 mol % was possible with the calibration scheme. Though their investigation was based on a very limited sample suite and Raman spectral range, the study highlighted the potential opportunity for using machine learning to quantify geochemical information from the Raman spectra of silicate glasses.

To further test the PLS oxide modeling package of this chapter on a preliminary basis, a subset of the Raman spectra publicly provided by Le Losq et al. (2019) was used as a secondary n_{test} (*AM-19-76887*: MORBData). This data is some of the only publicly available Raman-glass spectra that has been used for comparable modeling studies. The subset includes a random selection of 11 (of the 42) natural Mid-Ocean Ridge Basalt (MORB) glass samples presented in the authors' data. The reader is directed to that publication for review of the full dataset. These samples represent a selection of the MORBs from the Department of Mineral Sciences, Smithsonian Institution NMNH catalog (the VG series data, named as "11-XXXXX", n = 9), as well as a set of basalts

from Juan de Fuca and Mid-Atlantic Ridge MAPCO samples (Alv, and CH series respectively, n = 2) (Table 1.6, Table 1.7). A discussion of the glasses, their chemistry, and their Raman spectra can be found in the Le Losq et al. (2019). Two of the randomly selected samples for PLS validation (NMNH No. 115083-41 and 111241) exhibited magnetite crystallization Raman peaks at ~670 cm^{-1} (Figure 1.9A). These samples were still included in the test set as a litmus for how well the PLS modeling suite fits natural glasses with impurities that could be encountered in field scenarios. The model is applied to the Le Losq et al. (2019) spectra using the DEVAS system.

RMSE-P and NRMSE-P results indicate similar trends in oxide prediction accuracy to that of the initial model application (Table 1.5). However, the average model errors are higher in this application with NRMSE-P at ~ 0.61 (61%) and RMSE-P at 4.65 wt%. SiO_2 PLS predictions on the MORB's are significantly more accurate, with a normalized possible variance of only ~ 8%. Na_2O, CaO, and TiO_2 normalized prediction metrics are comparable to the initial validation metrics, 0.16, 0.21, and 0.28 respectively. However, the models show high difficulty in predicting MgO and Al_2O_3 on these Raman data, with similar (though less) strain in K_2O and FeO_T predictions (Table 1.5) (Figure 1.10). The considerably lower RMSE's for silica predictions is likely due to the constrained nature of the test set, where there is very little variability in these concentrations (all MORB's contained ~50% SiO_2). As is the case in all previous Raman-glass modeling publications (Di Muro et al., 2009; Di Genova et al., 2015; 2016a; 2016b; Giordano et al., 2019; González-García et al., 2020; Le Losq et al., 2019), and due to the fact that an average model suite error of > 60% is not suitable, an attempt to provide an 'optimized' or 'constrained' version of the modeling suite is reported on here. To accomplish this constraint, the training data set of 48 glasses is reduced to only include basaltic samples (n=15) (still along all oxidation states). The PLS algorithm is then trained to predict the oxides using the new "BAS Only" constrained Raman dataset, and then applied to the Le Losq et al. (2019) data.

Application of the BAS Only PLS models yielded a dramatic increase in predictive accuracy on average (~ 30%). Average RMSE-P drops by half to 2.08 wt% , while oxide-specific values vary slightly when compared to the general approach but are

generally lower (Figure 1.11): likely due to their constraint along narrower training ranges in the basalt group, as this approach restricts the training data ranges of the other oxides significantly (Table 1.2). The predictions and coefficients of both PLS approaches for major SiO_2 content are displayed against the external MORB spectra in Figure 1.9A, and the predictions are provided in Table 1.6 and Table 1.7. The results here further the notion that in modeling glass chemistry using Raman-based techniques, the use of a generalized approach may not be the most appropriate option in all cases, even though use of a constrained modeling suite would require prior contextual knowledge of possible sample chemistry. Similar to the strategy employed on these glasses in Le Losq et al. (2019) a model calibrated on samples akin to the composition of the intended target samples seems to be preferable.

Conclusions

This chapter demonstrates that PLS-1 calibrations can be used to predict eight major oxides in synthetic silicate glasses by combining multivariate analytical techniques and Raman spectroscopy. Models for each major geochemical constituent are provided, and the effects of chemical variations on the Raman spectra as a function of individual oxide components along constant oxidation states are explored. Associated internal/external accuracy metrics show that PLS is capable of predicting oxide concentrations in this scenario with an average error of ~ 28%. It is asserted that while a "universal" or "general" modeling approach of any type for glass chemistry using Raman is desired, a refined calibration using constrained geologic compositions may be a more appropriate choice when encountering limited glass compositions that can be defined prior to model application. In the case of applying a constrained version of the PLS suite that is built solely on basaltic compositions, then validated against external reduced basaltic glasses, prediction errors are greatly reduced on average when compared to the general approach. Average PLS model errors here are comparable to previous studies in the field, with the added advantage of being widely applicable to any encountered composition. It is hoped this work will serve as a foundation for future refinements of PLS and associated machine learning techniques in these scenarios, while also improving

the utility of Raman spectroscopy in modeling silicate glass properties as a first-order measure where laboratory analyses, or other instruments are unavailable.

References

Anderson, R.B. *et al.* (2022) "Post-landing major element quantification using SuperCam laser induced breakdown spectroscopy," *Spectrochimica Acta Part B: Atomic Spectroscopy*, 188, p. 106347. Available at: https://doi.org/10.1016/j.sab.2021.106347.

Baert, K. *et al.* (2011) "Using Raman spectroscopy as a tool for the detection of iron in glass," *Journal of Raman Spectroscopy*, 42(9), pp. 1789–1795. Available at: https://doi.org/10.1002/jrs.2935.

Beegle, L. *et al.* (2015) "SHERLOC: Scanning habitable environments with Raman & Luminescence for Organics & Chemicals," *2015 IEEE Aerospace Conference*. Available at: https://doi.org/10.1109/aero.2015.7119105.

Behrens, H. and Zhang, Y. (2008) "H_2O diffusion in Peralkaline to peraluminous rhyolitic melts," *Contributions to Mineralogy and Petrology*, 157(6), pp. 765–780. Available at: https://doi.org/10.1007/s00410-008-0363-4.

Bell, J.F. *et al.* (2000) "Mineralogic and compositional properties of martian soil and dust: Results from Mars Pathfinder," *Journal of Geophysical Research: Planets*, 105(E1), pp. 1721–1755. Available at: https://doi.org/10.1029/1999je001060.

Bell, R.J., Bird, N.F. and Dean, P. (1968) "The vibrational spectra of vitreous silica, Germania and beryllium fluoride," *Journal of Physics C: Solid State Physics*, 1(2), pp. 299–303. Available at: https://doi.org/10.1088/0022-3719/1/2/304.

Borisov, A., Behrens, H. and Holtz, F. (2015) "Effects of melt composition on Fe^{3+}/Fe^{2+} in silicate melts: A step to model ferric/ferrous ratio in multicomponent systems," *Contributions to Mineralogy and Petrology*, 169(2). Available at: https://doi.org/10.1007/s00410-015-1119-6.

Botcharnikov, R.E. *et al.* (2005) “The effect of water activity on the oxidation and structural state of Fe in a ferro-basaltic melt,” *Geochimica et Cosmochimica Acta*, 69(21), pp. 5071–5085. Available at: https://doi.org/10.1016/j.gca.2005.04.023.

Breitenfeld, L.B. *et al.* (2018) “Predicting olivine composition using Raman spectroscopy through band shift and multivariate analyses,” *American Mineralogist*, 103(11), pp. 1827–1836. Available at: https://doi.org/10.2138/am-2018-6291.

Brown, F.H., *et al.* (1992) “Correlation of Pliocene and Pleistocene tephra layers between the Turkana Basin of East Africa and the Gulf of Aden,” *Quaternary International*, 13-14, pp. 55–67. Available at: https://doi.org/10.1016/1040-6182(92)90010-Y

Bykov, V.N., Koroleva, O.N. and Osipov, A.A. (2009) “Structure of silicate melts: Raman spectroscopic data and thermodynamic simulation results,” *Geochemistry International*, 47(11), pp. 1067–1074. Available at: https://doi.org/10.1134/s0016702909110020.

Carey, C. *et al.* (2015) “Machine learning tools for mineral recognition and classification from Raman spectroscopy,” *Journal of Raman Spectroscopy*, 46(10), pp. 894–903. Available at: https://doi.org/10.1002/jrs.4757.

Chou, I.-M. and Wang, A. (2017) “Application of Laser Raman micro-analyses to Earth and planetary materials,” *Journal of Asian Earth Sciences*, 145, pp. 309–333. Available at: https://doi.org/10.1016/j.jseaes.2017.06.032.

Cochrane, C.J. and Blacksberg, J. (2015) “A fast classification scheme in Raman spectroscopy for the identification of mineral mixtures using a large database with correlated predictors,” *IEEE Transactions on Geoscience and Remote Sensing*, 53(8), pp. 4259–4274. Available at: https://doi.org/10.1109/tgrs.2015.2394377.

Colomban, P. (2003) “Polymerization degree and Raman identification of ancient glasses used for jewelry, ceramic enamels and mosaics,” *Journal of Non-Crystalline Solids*, 323(1-3), pp. 180–187. Available at: https://doi.org/10.1016/s0022-3093(03)00303-x.

de Faria, D.L., Venâncio Silva, S. and de Oliveira, M.T. (1997) “Raman microspectroscopy of some iron oxides and Oxyhydroxides,” *Journal of Raman Spectroscopy*, 28(11), pp. 873–878. Available at: https://doi.org/10.1002/(sici)1097-4555(199711)28:11<873::aid-jrs177>3.0.co;2-b.

de Jong, B.H.W.S. *et al.* (1981) “Polymerization of silicate and aluminate tetrahedra in glasses, melts, and aqueous solutions—III. local silicon environments and internal nucleation in silicate glasses,” *Geochimica et Cosmochimica Acta*, 45(8), pp. 1291–1308. Available at: https://doi.org/10.1016/0016-7037(81)90223-4.

Di Genova, D. *et al.* (2015) “Approximate chemical analysis of volcanic glasses using Raman spectroscopy,” *Journal of Raman Spectroscopy*, 46(12), pp. 1235–1244. Available at: https://doi.org/10.1002/jrs.4751.

Di Genova, D. *et al.* (2016a) “Models for the estimation of Fe^{3+}/Fe_{tot}ratio in terrestrial and extraterrestrial alkali- and iron-rich silicate glasses using Raman Spectroscopy” *American Mineralogist*, 101(4), pp. 943–952. Available at: https://doi.org/10.2138/am-2016-5534ccbyncnd.

Di Genova, D. *et al.* (2016b) “Raman spectra of Martian glass analogues: A tool to approximate their chemical composition,” *Journal of Geophysical Research: Planets*, 121(5), pp. 740–752. Available at: https://doi.org/10.1002/2016je005010.

Di Genova, D. *et al.* (2017) “Effect of oxygen fugacity on the glass transition, viscosity and structure of silica- and iron-rich magmatic melts,” *Journal of Non-Crystalline*

Solids, 470, pp. 78–85. Available at: https://doi.org/10.1016/j.jnoncrysol.2017.05.013.

Di Genova, D., Caracciolo, A. and Kolzenburg, S. (2018) "Measuring the degree of 'nanotilization' of volcanic glasses: Understanding syn-eruptive processes recorded in melt inclusions," *Lithos*, 318-319, pp. 209–218. Available at: https://doi.org/10.1016/j.lithos.2018.08.011.

Di Muro, A. *et al.* (2009) "Micro-raman determination of iron redox state in dry natural glasses: Application to Peralkaline Rhyolites and Basalts," *Chemical Geology*, 259(1-2), pp. 78–88. Available at: https://doi.org/10.1016/j.chemgeo.2008.08.013.

Dingwell, D.B. (2006) "Transport properties of magmas: Diffusion and rheology," *Elements*, 2(5), pp. 281–286. Available at: https://doi.org/10.2113/gselements.2.5.281.

Dünnwald, J. and Otto, A. (1989) "An investigation of phase transitions in rust layers using Raman spectroscopy," *Corrosion Science*, 29(9), pp. 1167–1176. Available at: https://doi.org/10.1016/0010-938x(89)90052-8.

Dyar, M.D. and Ytsma, C.R. (2021) "Effect of data set size on geochemical quantification accuracy with laser-induced breakdown spectroscopy," *Spectrochimica Acta Part B: Atomic Spectroscopy*, 177, p. 106073. Available at: https://doi.org/10.1016/j.sab.2021.106073.

Dyar, M.D. *et al.* (2012) "Comparison of partial least squares and LASSO regression techniques as applied to laser-induced breakdown spectroscopy of geological samples," *Spectrochimica Acta Part B: Atomic Spectroscopy*, 70, pp. 51–67. Available at: https://doi.org/10.1016/j.sab.2012.04.011.

Dyar, M.D. *et al.* (2016) "Accurate predictions of iron redox state in silicate glasses: A multivariate approach using X-ray absorption spectroscopy," *American*

Mineralogist, 101(3), pp. 744–747. Available at: https://doi.org/10.2138/am-2016-5555.

Eilers, P., and Boelens, H. (2005). "Baseline Correction with Asymmetric Least Squares Smoothing". *Unpubl. Manuscr.* Algorithm at:https://www.wavemetrics.com/code-snippet/asymmetric-least-squares-smoothing

Fu, X. *et al*. (2017) "Characterizing amorphous silicates in extraterrestrial materials: Polymerization effects on Raman and mid-IR spectral features of alkali and alkali earth silicate glasses," *Journal of Geophysical Research: Planets*, 122(5), pp. 839–855. Available at: https://doi.org/10.1002/2016je005241.

Furukawa, T., *et al.* (1981) "Raman spectroscopic investigation of the structure of silicate glasses. III. raman intensities and structural units in sodium silicate glasses," *The Journal of Chemical Physics*, 75(7), pp. 3226–3237. Available at: https://doi.org/10.1063/1.442472.

Giguere, S. *et al.* (2017) "A fully customized Baseline Removal Framework for spectroscopic applications," *Applied Spectroscopy*, 71(7), pp. 1457–1470. Available at: https://doi.org/10.1177/0003702817695624.

Giordano, D. and Russell, J.K. (2018) "Towards a structural model for the viscosity of geological melts," *Earth and Planetary Science Letters*, 501, pp. 202–212. Available at: https://doi.org/10.1016/j.epsl.2018.08.031.

Giordano, D. *et al.* (2019) "A calibrated database of Raman spectra for natural silicate glasses: Implications for modelling melt physical properties," *Journal of Raman Spectroscopy*, 51(9), pp. 1822–1838. Available at: https://doi.org/10.1002/jrs.5675.

Giordano, D. *et al.* (2020) “Raman spectroscopy from laboratory and proximal to remote sensing: A tool for the Volcanological Sciences,” *Remote Sensing*, 12(5), p. 805. Available at: https://doi.org/10.3390/rs12050805.

Giuli, G. *et al.* (2012) “Effect of alkalis on the FE oxidation state and local environment in Peralkaline rhyolitic glasses,” *American Mineralogist*, 97(2-3), pp. 468–475. Available at: https://doi.org/10.2138/am.2012.3888.

González-García, D. *et al.* (2020) “A Raman spectroscopic tool to estimate chemical composition of natural volcanic glasses,” *Chemical Geology*, 556, p. 119819. Available at: https://doi.org/10.1016/j.chemgeo.2020.119819.

Hargraves, R.B. (1980) “Physics of magmatic processes.” Available at: https://doi.org/10.1515/9781400854493.

Hartmann, W. K., and Berman, D.C. (2000) "Elysium Planitia lava flows: Crater count chronology and geological implications," *Journal of Geophysical Research: Planets*, 105, pp. 15011-15025.

Hastie, T., Tibshirani, R. and Friedman, J. (2009) “The elements of Statistical Learning,” *Springer Series in Statistics*. Available at: https://doi.org/10.1007/978-0-387-84858-7.

James, G. *et al.* (2013) “An introduction to statistical learning,” *Springer Texts in Statistics*. Available at: https://doi.org/10.1007/978-1-4614-7138-7.

Kalampounias, A.G., Yannopoulos, S.N. and Papatheodorou, G.N. (2006) “A high-temperature Raman spectroscopic investigation of the potassium tetrasilicate in glassy, supercooled, and Liquid States,” *The Journal of Chemical Physics*, 125(16), p. 164502. Available at: https://doi.org/10.1063/1.2360275.

Kress, V.C. and Carmichael, I.S. (1991) “The compressibility of silicate liquids containing fe2o3 and the effect of composition, temperature, oxygen fugacity and pressure on their redox states,” *Contributions to Mineralogy and Petrology*, 108(1-2), pp. 82–92. Available at: https://doi.org/10.1007/bf00307328.

Lange, R.A. and Carmichael, I.S.E. (1987) “Densities of NA2O-K2O-Cao-MgO-FeO-Fe2O3-Al2O3-Tio2-SIO2 liquids: New measurements and derived partial molar properties,” *Geochimica et Cosmochimica Acta*, 51(11), pp. 2931–2946. Available at: https://doi.org/10.1016/0016-7037(87)90368-1.

Lanzirotti, A. *et al.* (2018) “Accurate predictions of microscale oxygen barometry in basaltic glasses using V K-edge x-ray absorption spectroscopy: A multivariate approach,” *American Mineralogist*, 103(8), pp. 1282–1297. Available at: https://doi.org/10.2138/am-2018-6319.

Le Losq, C. *et al.* (2019) “Determination of the oxidation state of iron in Mid-Ocean Ridge basalt glasses by Raman spectroscopy,” *American Mineralogist*, 104(7), pp. 1032–1042. Available at: https://doi.org/10.2138/am-2019-6887.

Le Losq, C., Moretti, R. and Neuville, D.R. (2014) “Speciation and amphoteric behaviour of water in aluminosilicate melts and glasses: High-temperature Raman spectroscopy and reaction equilibria,” *European Journal of Mineralogy*, 25(5), pp. 777–790. Available at: https://doi.org/10.1127/0935-1221/2013/0025-2322.

Leight, C. (2020) Master’s Thesis: *The University of Tennessee Research and Creative Exchange*

Leight, C. *et al*. (2021) “Identifying volcanic glass concentration and composition with remote sensing using multivariate methods,” *Lunar Planet. Sci.* No. 2548, 2214.

Liland, K.H., Almøy, T. and Mevik, B.-H. (2010) "Optimal choice of baseline correction for multivariate calibration of Spectra," *Applied Spectroscopy*, 64(9), pp. 1007–1016. Available at: https://doi.org/10.1366/000370210792434350.

Long, D.A. (1977) Raman Spectroscopy, pp. 276. McGraw-Hill, Maidenhead, U.K

Lowe, D. J. (2011) "Tephrochronology and its application: a review," *Quaternary Geochronology*, 6 (2), pp. 107–153.

Magnien, V. *et al.* (2006) "Kinetics of iron redox reactions in silicate liquids: A high-temperature X-ray absorption and Raman spectroscopy study," *Journal of Nuclear Materials*, 352(1-3), pp. 190–195. Available at: https://doi.org/10.1016/j.jnucmat.2006.02.053.

Matson, D.W., Sharma, S.K. and Philpotts, J.A. (1983) "The structure of high-silica alkali-silicate glasses. A Raman spectroscopic investigation," *Journal of Non-Crystalline Solids*, 58(2-3), pp. 323–352. Available at: https://doi.org/10.1016/0022-3093(83)90032-7.

McMillan, P. (1984). "Structural Studies of Silicate Glasses and Melts-Applications and Limitations of Raman Spectroscopy", *American Mineralogist*, 69, pp. 622-644.

McMillan, P.F. and Hofmeister, A.M. (1988) "Chapter 4. infrared and Raman spectroscopy," *Spectroscopic Methods in Mineralogy and Geology*, pp. 99–160. Available at: https://doi.org/10.1515/9781501508974-006.

McMillan, P.F. and Piriou, B. (1983) "Raman spectroscopic studies of silicate and related glass structure : A Review," *Bulletin de Minéralogie*, 106(1), pp. 57–75. Available at: https://doi.org/10.3406/bulmi.1983.7668.

McMillan, P.F. and Wolf, G.H. (1995) "Chapter 8. vibrational spectroscopy of silicate liquids," *Structure, Dynamics, and Properties of Silicate Melts*, pp. 247–316. Available at: https://doi.org/10.1515/9781501509384-010.

McSween, H.Y. (1994) “What we have learned about Mars from SNC Meteorites,” *Meteoritics*, 29(6), pp. 757–779. Available at: https://doi.org/10.1111/j.1945-5100.1994.tb01092.x.

Moretti, R. (2009) “Polymerization, basicity, oxidation state and their role in ionic modelling of silicate melts,” *Annals of Geophysics*, 48, pp. 583-608. Available at: https://doi.org/10.4401/ag-3221.

Mysen, B. (2003) “Physics and chemistry of silicate glasses and melts,” *European Journal of Mineralogy*, 15(5), pp. 781–802. Available at: https://doi.org/10.1127/0935-1221/2003/0015-0781.

Mysen, B., Virgo D., and Scarfe C. (1980) “Relations between the anionic structure and viscosity of silicate melts- a Raman spectroscopic study,” *American Mineralogist*, 65 (7-8), pp. 690-710.

Mysen, B.O. (1991) “Relations between structure, redox equilibria of iron, and properties of magmatic liquids,” *Advances in Physical Geochemistry*, pp. 41–98. Available at: https://doi.org/10.1007/978-1-4612-3128-8_2.

Mysen, B.O. and Toplis, M.J. (2007) “Structural behavior of Al^{3+} in Peralkaline, metaluminous, and peraluminous silicate melts and glasses at ambient pressure,” *American Mineralogist*, 92 (5-6), pp. 933–946. Available at: https://doi.org/10.2138/am.2007.2334.

Mysen, B.O., Virgo, D. and Seifert, F.A. (1982) “The structure of silicate melts: Implications for chemical and physical properties of natural magma,” *Reviews of Geophysics*, 20(3), p. 353. Available at: https://doi.org/10.1029/rg020i003p00353.

Neuville, D.R. and Mysen, B.O. (1996) “Role of aluminum in the silicate network: In situ, high-temperature study of glasses and melts on the join SIO2-NaAlO2,”

Geochimica et Cosmochimica Acta, 60(10), pp. 1727–1737. Available at: https://doi.org/10.1016/0016-7037(96)00049.

Neuville, D.R., de Ligny, D. and Henderson, G.S. (2014) "Advances in Raman spectroscopy applied to Earth and material sciences," *Reviews in Mineralogy and Geochemistry*, 78(1), pp. 509–541. Available at: https://doi.org/10.2138/rmg.2013.78.13.

Ning, X., Selesnick, I.W. and Duval, L. (2014) "Chromatogram baseline estimation and denoising using sparsity (beads)," *Chemometrics and Intelligent Laboratory Systems*, 139, pp. 156–167. Available at: https://doi.org/10.1016/j.chemolab.2014.09.014.

Osipov, A., Osipova, L. and Zainullina, R. (2015) "Raman spectroscopy and statistical analysis of the silicate species and group connectivity in cesium silicate glass forming system," *International Journal of Spectroscopy*, 2015, pp. 1–15. Available at: https://doi.org/10.1155/2015/572840.

Ottonello, G. *et al.* (2001) "Oxidation state of iron in silicate glasses and melts: A thermochemical model," *Chemical Geology*, 174(1-3), pp. 157–179. Available at: https://doi.org/10.1016/s0009-2541(00)00314-4.

Pasteris, J.D. and Beyssac, O. (2020) "Welcome to raman spectroscopy: Successes, challenges, and Pitfalls," *Elements*, 16(2), pp. 87–92. Available at: https://doi.org/10.2138/gselements.16.2.87.

Prinsloo, L.C. *et al.* (2010) "A Raman spectroscopic study of the Igneous Rocks on Marion Island: A possible terrestrial analogue for the geology on Mars," *Journal of Raman Spectroscopy*, 42(4), pp. 626–632. Available at: https://doi.org/10.1002/jrs.2756.

Qian, F., Wu, Y. and Hao, P. (2017) “A fully automated algorithm of Baseline Correction based on Wavelet Feature Points and segment interpolation,” *Optics & Laser Technology*, 96, pp. 202–207. Available at: https://doi.org/10.1016/j.optlastec.2017.05.021.

Rampe, E *et al*. (2014) “Amorphous phases on the surface of Mars.” *International Conference on Mars,* Abstract ID- 20140008714. Available at: https://ntrs.nasa.gov/citations/20140008714

Richet, P. (1984) “Viscosity and configurational entropy of silicate melts,” *Geochimica et Cosmochimica Acta,* 48(3), pp. 471–483. Available at: https://doi.org/10.1016/0016-7037(84)90275-8.

Robert, G. *et al.* (2014) “Heat capacity of hydrous basaltic glasses and liquids,” *Journal of Non-Crystalline Solids,* 390, pp. 19–30. Available at: https://doi.org/10.1016/j.jnoncrysol.2014.02.011.

Rossano, S. and Mysen, B. O. (2012). “Raman spectroscopy of silicate glasses and melts in Geologic systems,” *EMU Notes in Mineralogy*, 12, chapter 9, pp. 319-364.

Schulze, H.G. *et al.* (2012) “A small-window moving average-based fully Automated Baseline Estimation Method for Raman Spectra,” *Applied Spectroscopy*, 66(7), pp. 757–764. Available at: https://doi.org/10.1366/11-06550.

Seifert, F.A., Mysen, B.O. and Virgo, D. (1981) “Structural similarity of glasses and melts relevant to Petrological Processes,” *Geochimica et Cosmochimica Acta*, 45(10), pp. 1879–1884. Available at: https://doi.org/10.1016/0016-7037(81)90017-x.

Singer, R.B. (1985) “Spectroscopic observation of Mars,” *Advances in Space Research*, 5(8), pp. 59–68. Available at: https://doi.org/10.1016/0273-1177(85)90242-x.

Stabile, P. *et al.* (2021) “The effect of iron and alkali on the nanocrystal-free viscosity of volcanic melts: A Combined Raman spectroscopy and DSC Study,” *Chemical Geology*, 559, p. 119991. Available at: https://doi.org/10.1016/j.chemgeo.2020.119991.

Stone, M. and Brooks, R.J. (1990) “Continuum regression: Cross-validated sequentially constructed prediction embracing ordinary least squares, partial least squares and principal components regression,” *Journal of the Royal Statistical Society: Series B (Methodological)*, 52(2), pp. 237–258. Available at: https://doi.org/10.1111/j.2517-6161.1990.tb01786.x.

Wang, A., Jolliff, B.L. and Haskin, L.A. (1995) “Raman spectroscopy as a method for mineral identification on lunar robotic exploration missions,” *Journal of Geophysical Research*, 100(E10), p. 21189. Available at: https://doi.org/10.1029/95je02133.

Welsch, A.-M., Knipping, J.L. and Behrens, H. (2017) “Fe-oxidation state in alkali-trisilicate glasses - A Raman spectroscopic study,” *Journal of Non-Crystalline Solids*, 471, pp. 28–38. Available at: https://doi.org/10.1016/j.jnoncrysol.2017.04.033.

White, W.B. and Minser, D.G. (1984) “Raman spectra and structure of natural glasses,” *Journal of Non-Crystalline Solids*, 67(1-3), pp. 45–59. Available at: https://doi.org/10.1016/0022-3093(84)90140-6.

Whiteman, D.N. *et al.* (1993) “Temperature sensitivity of an atmospheric Raman lidar system based on a XEF excimer laser,” *Optics Letters*, 18(3), p. 247. Available at: https://doi.org/10.1364/ol.18.000247.

Wiens, R. *et al.* (2020a) “The SUPERCAM instrument for the Mars 2020 rover,” *2020 IEEE Aerospace Conference*, Available at: https://doi.org/10.1109/aero47225.2020.9172661.

Wiens, R.C. *et al.* (2020b) “The SUPERCAM instrument suite on the NASA mars 2020 rover: Body unit and combined system tests,” *Space Science Reviews*, 217(1). Available at: https://doi.org/10.1007/s11214-020-00777-5.

Yadav, A.K. and Singh, P. (2015) “A review of the structures of oxide glasses by Raman spectroscopy,” *RSC Advances*, 5(83), pp. 67583–67609. Available at: https://doi.org/10.1039/c5ra13043c.

Appendix

Tables

Table 1.1. Experimental conditions and major chemical ranges for the 69 Dyar samples; filtered by geochemical grouping. The sample set was separated into a training set (n_{train} = 48) and a testing set (n_{test} = 21) for building and validating the PLS modeling calibration. Glasses spanned all 5 major extrusive rock types (i.e., basalts – rhyolites) along a $\%Fe^{3+}$ range of ~ 8 – 100 %. This ensured a truly chemically diverse sample suite, with little to no constraining properties

Glass Type	n	SiO_2 Range (wt%)	Fe^{3+}/Fe_{TOT} Range
Basalt	21	48.21 – 51.03	0.09 – 0.90
Basaltic Andesite	10	53.66 – 56.27	0.08 – 0.93
Andesite	16	59.03 – 63.73	0.08 – 1.00
Dacite	15	65.01 – 72.96	0.08 – 1.00
Rhyolite	7	74.29 – 78.02	0.11 – 0.71

Table 1.2. Geochemical compositional data (major oxide concentrations in wt% (SiO_2, Na_2O, K_2O, CaO, TiO_2, Al_2O_3, FeO_T, MgO) for the 69 Dyar glass samples used in the development and validation of PLS oxide models. Samples were named according to their rock type category and mineral redox buffer M.RB. (i.e., BAS, basalt; BAA, basaltic andesite; AND, andesite; DAC, dacite; or RHY, rhyolite) & (Oxidized - Air: air buffer, CO_2: carbon dioxide buffer & Reduced – IW: iron wüstite buffer, QFM: quartz-fayalite-magnetite buffer, Mo-MoO_2: molybdenum oxide buffer). The suite chemical boundaries include compositions with > 48 wt% SiO_2, and > 9% Fe^{3+}

Sample Name[1]	Rock Type	SiO_2[2]	Na_2O	K_2O	CaO	TiO_2	Al_2O_3	FeO_T	MgO
BAS1_Air	Basalt	50.75	3.06	0.29	10.73	1.07	17.72	7.07	9.3
BAS1_IW	Basalt	50.75	3.06	0.29	10.73	1.07	17.72	7.07	9.3
BAS2_Air	Basalt	50.7	2.64	0.15	11.55	1.57	15.34	10.54	7.5
BAS2_QFM	Basalt	50.7	2.64	0.15	11.55	1.57	15.34	10.54	7.5
BAS2_CO_2	Basalt	50.7	2.64	0.15	11.55	1.57	15.34	10.54	7.5
BAS3_CO_2	Basalt	50.28	3.02	1.16	8.76	2.62	16.79	11.66	5.73
BAS3_QFM	Basalt	50.28	3.02	1.16	8.76	2.62	16.79	11.66	5.73
BAS4_Air	Basalt	51.03	2.02	0.29	11.37	0.81	17.8	10.24	6.44
BAS4_CO_2	Basalt	51.03	2.02	0.29	11.37	0.81	17.8	10.24	6.44
BAS4_IW	Basalt	51.03	2.02	0.29	11.37	0.81	17.8	10.24	6.44
BAS4_QFM	Basalt	51.03	2.02	0.29	11.37	0.81	17.8	10.24	6.44
BAS5_CO_2	Basalt	50.47	2.53	0.81	10.68	1.01	17.21	10.14	7.15
BAS5_Air	Basalt	50.47	2.53	0.81	10.68	1.01	17.21	10.14	7.15
BAS5_IW	Basalt	50.47	2.53	0.81	10.68	1.01	17.21	10.14	7.15
BAS5_QFM	Basalt	50.47	2.53	0.81	10.68	1.01	17.21	10.14	7.15
BAS6_CO_2	Basalt	50.61	2.25	0.52	11.57	2.78	13.7	11.24	7.33
BAS6_Air	Basalt	50.61	2.25	0.52	11.57	2.78	13.7	11.24	7.33
BAS6_QFM	Basalt	50.61	2.25	0.52	11.57	2.78	13.7	11.24	7.33
BAS7_Air	Basalt	48.21	1.84	0.03	13.38	0.98	15.58	10.24	9.75
BAS7_CO_2	Basalt	48.21	1.84	0.03	13.38	0.98	15.58	10.24	9.75
BAS7_QFM	Basalt	48.21	1.84	0.03	13.38	0.98	15.58	10.24	9.75
BAA1_Air	Basaltic Andesite	53.66	3.23	0.5	9.33	0.65	17.1	7.14	8.39

Table 1.2. Continued

Sample Name[1]	Rock Type	SiO_2[2]	Na_2O	K_2O	CaO	TiO_2	Al_2O_3	FeO_T	MgO
BAA1_IW	Basaltic Andesite	53.66	3.23	0.5	9.33	0.65	17.1	7.14	8.39
BAA1_QFM	Basaltic Andesite	53.66	3.23	0.5	9.33	0.65	17.1	7.14	8.39
BAA2_CO_2	Basaltic Andesite	53.98	2.84	1.05	9.11	1.77	14.53	11.45	5.27
BAA2_Air	Basaltic Andesite	53.98	2.84	1.05	9.11	1.77	14.53	11.45	5.27
BAA2_IW	Basaltic Andesite	53.98	2.84	1.05	9.11	1.77	14.53	11.45	5.27
BAA2_QFM	Basaltic Andesite	53.98	2.84	1.05	9.11	1.77	14.53	11.45	5.27
BAA3_CO_2	Basaltic Andesite	56.27	3.16	2.01	6.93	2.28	14.09	11.87	3.38
BAA3_Air	Basaltic Andesite	56.27	3.16	2.01	6.93	2.28	14.09	11.87	3.38
BAA3_QFM	Basaltic Andesite	56.27	3.16	2.01	6.93	2.28	14.09	11.87	3.38
AND1_QFM	Andesite	63.73	4.53	1.35	5.08	0.78	17.84	4.75	1.95
AND1_Air	Andesite	63.73	4.53	1.35	5.08	0.78	17.84	4.75	1.95
AND2_Air	Andesite	60.14	2.91	0.77	8.68	0.65	11.72	6.05	9.07
AND2_IW	Andesite	60.14	2.91	0.77	8.68	0.65	11.72	6.05	9.07
AND2_CO_2	Andesite	60.14	2.91	0.77	8.68	0.65	11.72	6.05	9.07
AND2_QFM	Andesite	60.14	2.91	0.77	8.68	0.65	11.72	6.05	9.07
AND3_IW	Andesite	59.03	2.92	0.6	8.13	0.72	17.06	7.93	3.61
AND3_CO_2	Andesite	59.03	2.92	0.6	8.13	0.72	17.06	7.93	3.61
AND3_QFM	Andesite	59.03	2.92	0.6	8.13	0.72	17.06	7.93	3.61
AND4_IW	Andesite	59.73	3.22	1.32	7.16	0.72	17.24	6.87	3.73
AND4_QFM	Andesite	59.73	3.22	1.32	7.16	0.72	17.24	6.87	3.73
AND4_Air	Andesite	59.73	3.22	1.32	7.16	0.72	17.24	6.87	3.73
AND5_Air	Andesite	59.97	3.76	0.87	7.91	0.66	17.78	6.39	2.66
AND5_CO_2	Andesite	59.97	3.76	0.87	7.91	0.66	17.78	6.39	2.66
AND5_IW	Andesite	59.97	3.76	0.87	7.91	0.66	17.78	6.39	2.66
AND5_QFM	Andesite	59.97	3.76	0.87	7.91	0.66	17.78	6.39	2.66
DAC1_Mo-MoO_2	Dacite	67.94	4.59	1.62	3.74	0.58	16.78	3.61	1.14
DAC2_Air	Dacite	65.66	4.7	1.14	5.23	0.43	17.51	3.2	2.14
DAC2_CO_2	Dacite	65.66	4.7	1.14	5.23	0.43	17.51	3.2	2.14

Table 1.2 Continued

Sample Name[1]	Rock Type	SiO_2[2]	Na_2O	K_2O	CaO	TiO_2	Al_2O_3	FeO_T	MgO
DAC2_Mo-MoO_2	Dacite	65.66	4.7	1.14	5.23	0.43	17.51	3.2	2.14
DAC3_Air	Dacite	65.01	4.09	4.09	3.66	0.6	17.13	4.29	1.12
DAC3_CO_2	Dacite	65.01	4.09	4.09	3.66	0.6	17.13	4.29	1.12
DAC3_Mo-MoO_2	Dacite	65.01	4.09	4.09	3.66	0.6	17.13	4.29	1.12
DAC4_Air	Dacite	72.85	4.87	1.54	2.38	0.55	13.51	3.8	0.5
DAC4_CO_2	Dacite	72.85	4.87	1.54	2.38	0.55	13.51	3.8	0.5
DAC4_Mo-MoO_2	Dacite	72.85	4.87	1.54	2.38	0.55	13.51	3.8	0.5
DAC5_Air	Dacite	67.65	3.83	0.92	5.04	0.6	15.12	5.34	1.51
DAC5_Mo-MoO_2	Dacite	67.65	3.83	0.92	5.04	0.6	15.12	5.34	1.51
DAC6_CO_2	Dacite	67.73	4.34	2.11	3.84	0.5	16.34	3.64	1.51
DAC6_Mo-MoO_2	Dacite	67.73	4.34	2.11	3.84	0.5	16.34	3.64	1.51
DAC7_CO_2	Dacite	72.96	4.76	2.81	2.21	0.38	14.03	2.32	0.53
RHY2_Mo-MoO_2	Rhyolite	77.47	4.47	1.44	1.81	0.25	12.2	2.11	0.25
RHY3_CO_2	Rhyolite	78.02	2.89	3.69	0.87	0.38	12.41	1.54	0.19
RHY3_Mo-MoO_2	Rhyolite	78.02	2.89	3.69	0.87	0.38	12.41	1.54	0.19
RHY4_CO_2	Rhyolite	74.29	5.22	1.55	1.87	0.45	13.79	2.46	0.38
RHY4_Mo-MoO_2	Rhyolite	74.29	5.22	1.55	1.87	0.45	13.79	2.46	0.38
RHY6_Mo-MoO_2	Rhyolite	75.06	4.03	1.1	2.83	0.39	13	2.97	0.6
RHY7_Mo-MoO_2	Rhyolite	75.27	4.2	2.71	1.6	0.2	13.52	1.99	0.5

[1]Sample names are in the following format: Sample Type/Number_Mineral Redox Buffer

[2]Oxide concentrations are in wt%

Table 1.3. Calculated RMSE metrics that assess PLS model accuracy and quality. Individual oxide calibration (RMSE-C), cross validation (RMSE-CV), prediction (RMSE-P), and normalized prediction (NRMSE-P) metrics are provided per oxide variable, along with the PLS component (q) selections produced from LOO-CV procedures in the DEVAS system. Errors are averaged to uncover full modeling suite accuracy. The absolute value of the calculated differences between the actual EMPA determined oxide concentrations and the PLS calculated amounts are averaged from Table 1.4

Predicted Oxide	**RMSE-C (wt%)**[1]	**RMSE-CV (wt%)**	**RMSE-P (wt%)**	**NRMSE-P**	**|Avg. Difference| (wt%)**[2]	***q***[3]
SiO_2	7.89	9.25	7.40	0.248	2.16	2
Na_2O	0.78	0.90	0.87	0.257	0.36	3
K_2O	0.53	0.87	1.10	0.271	0.12	6
CaO	3.01	3.44	3.02	0.242	1.11	2
TiO_2	0.63	0.69	0.66	0.257	0.19	2
Al_2O_3	1.68	2.34	2.79	0.456	1.04	3
FeO_T	2.45	3.20	2.62	0.254	0.46	4
MgO	2.12	2.96	2.55	0.267	0.08	5
Average	**2.39**	**2.96**	**2.63**	**0.282**		

[1]RMSE metrics rounded to nearest hundredth.

[2]Absolute value of the averaged difference between actual and predicted oxide wt % per sample.

[3]Number of PLS training model components (*q*) recommended through LOO-CV

Table 1.4. EMPA determined "Actual" oxide concentrations vs PLS model predicted oxide concentration values for the 21 Dyar test samples (n_{test} = 21), following application of the pertinent model (Act. = actual concentration; Pred. = predicted; Diff. = difference between actual and predicted value)

Sample[1]	SiO_2			Na_2O			K_2O			CaO		
	Act.	**Pred.**	**Diff.**	**Act.**	**Pred.**	**Diff.**	**Act.**	**Pred.**	**Diff.**	**Act.**	**Pred.**	**Diff.**
AND1_QFM	63.7	59.2	-4.5	4.5	3.4	-1.1	1.4	1.0	-0.3	5.1	7.6	2.5
AND2_CO_2	60.1	58.4	-1.7	2.9	3.6	0.7	0.8	0.4	-0.3	8.7	7.3	-1.4
AND2_QFM	60.1	67.9	7.8	2.9	3.4	0.5	0.8	1.0	0.2	8.7	4.2	-4.4
AND3_IW	59.0	61.2	2.1	2.9	3.7	0.7	0.6	1.4	0.8	8.1	6.5	-1.7
AND4_Air	59.7	52.0	-7.7	3.2	2.4	-0.9	1.3	1.9	0.5	7.2	9.9	2.7
AND5_Air	60.0	58.8	-1.2	3.8	3.4	-0.4	0.9	0.9	0.0	7.9	7.8	-0.1
BAA1_IW	53.7	58.5	4.8	3.2	3.4	0.2	0.5	0.9	0.4	9.3	8.1	-1.2
BAA2_CO_2	54.0	58.7	4.7	2.8	3.6	0.7	1.1	1.7	0.6	9.1	7.9	-1.2
BAA3_Air	56.3	56.3	0.0	3.2	2.8	-0.4	2.0	0.1	-1.9	6.9	8.9	1.9
BAA3_QFM	56.3	57.2	0.9	3.2	3.3	0.2	2.0	0.9	-1.1	6.9	8.6	1.7
BAS2_Air	50.7	57.5	6.8	2.6	3.1	0.5	0.2	1.4	1.2	11.6	7.5	-4.0
BAS2_QFM	50.7	58.4	7.7	2.6	3.5	0.8	0.2	1.5	1.4	11.6	8.0	-3.5
BAS4_Air	51.0	55.8	4.7	2.0	2.7	0.6	0.3	0.4	0.1	11.4	9.4	-2.0
BAS5_CO_2	50.5	61.3	10.8	2.5	3.5	1.0	0.8	1.3	0.5	10.7	6.8	-3.9
BAS6_CO_2	50.6	57.4	6.8	2.3	3.3	1.1	0.5	1.1	0.6	11.6	8.4	-3.1
BAS7_QFM	48.2	56.1	7.9	1.8	3.2	1.3	0.0	1.2	1.1	13.4	8.9	-4.5
DAC2_Air	65.7	81.5	15.9	4.7	5.3	0.6	1.1	3.3	2.2	5.2	-2.2	-7.4
DAC3_CO_2	65.0	68.8	3.8	4.1	3.9	-0.1	4.1	1.5	-2.6	3.7	3.8	0.1
DAC5_Mo	67.7	63.2	-4.5	3.8	3.4	-0.4	0.9	1.1	0.1	5.0	6.2	1.2
RHY3_Mo	78.0	75.8	-2.2	2.9	5.3	2.4	3.7	2.2	-1.5	0.9	0.4	-0.5
RHY6_Mo	75.1	57.5	17.6	4.0	3.5	-0.5	1.1	1.6	0.5	2.8	8.3	5.5

Table 1.4. Continued

Sample	TiO_2			Al_2O_3			FeO_T			MgO		
	Act.	**Pred.**	**Diff.**	**Act.**	**Pred.**	**Diff.**	**Act.**	**Pred.**	**Diff.**	**Act.**	**Pred.**	**Diff.**
AND1_QFM	0.8	1.0	0.3	17.8	15.1	-2.7	4.8	6.4	1.6	2.0	4.9	3.0
AND2_CO_2	0.7	1.1	0.5	11.7	16.2	4.5	6.1	7.2	1.1	9.1	4.1	-5.0
AND2_QFM	0.7	0.2	-0.5	11.7	20.6	8.9	6.1	11.1	5.1	9.1	8.4	-0.7
AND3_IW	0.7	1.0	0.2	17.1	16.2	-0.9	7.9	6.6	-1.4	3.6	1.4	-2.2
AND4_Air	0.7	1.3	0.5	17.2	18.0	0.7	6.9	9.4	2.5	3.7	3.3	-0.4
AND5_Air	0.7	1.1	0.4	17.8	14.7	-3.0	6.4	6.5	0.1	2.7	4.6	2.0
BAA1_IW	0.7	1.0	0.4	17.1	16.5	-0.6	7.1	8.7	1.6	8.4	6.1	-2.3
BAA2_CO_2	1.8	1.1	-0.7	14.5	15.7	1.2	11.5	6.8	-4.7	5.3	3.7	-1.6
BAA3_Air	2.3	1.1	-1.2	14.1	15.6	1.5	11.9	9.5	-2.4	3.4	4.6	1.2
BAA3_QFM	2.3	1.1	-1.2	14.1	16.2	2.1	11.9	8.6	-3.3	3.4	5.7	2.4
BAS2_Air	1.6	1.0	-0.6	15.3	16.7	1.4	10.5	7.0	-3.6	7.5	13.2	5.7
BAS2_QFM	1.6	1.1	-0.5	15.3	16.1	0.8	10.5	7.3	-3.2	7.5	5.0	-2.5
BAS4_Air	0.8	1.1	0.3	17.8	15.6	-2.2	10.2	9.5	-0.7	6.4	6.6	0.1
BAS5_CO_2	1.0	0.9	-0.1	17.2	16.0	-1.2	10.1	6.9	-3.2	7.2	4.8	-2.3
BAS6_CO_2	2.8	1.1	-1.7	13.7	16.2	2.5	11.2	8.2	-3.0	7.3	5.1	-2.2
BAS7_QFM	1.0	1.1	0.1	15.6	17.5	1.9	10.2	9.6	-0.6	9.8	6.1	-3.7
DAC2_Air	0.4	-0.2	-0.7	17.5	18.1	0.6	3.2	1.9	-1.3	2.1	1.6	-0.6
DAC3_CO_2	0.6	0.3	-0.3	17.1	17.8	0.7	4.3	7.1	2.8	1.1	3.5	2.3
DAC5_Mo	0.6	0.7	0.1	15.1	14.9	-0.2	5.3	5.0	-0.4	1.5	2.9	1.4
RHY3_Mo	0.4	0.2	-0.2	12.4	16.0	3.6	1.5	1.2	-0.3	0.2	0.8	0.7
RHY6_Mo	0.4	1.2	0.8	13.0	15.2	2.2	3.0	6.5	3.5	0.6	3.5	2.9

Table 1.5. Calculated RMSE metrics that assess PLS model accuracy and quality after applying the full "General" model and basaltic-only constrained "BAS" model to the 11 MORB Le Losq spectra (n_{test} = 11). Individual oxide prediction (RMSE-P) and normalized prediction (NRMSE-P) metrics are provided per oxide variables. Errors were averaged to uncover full and constrained modeling suite accuracy. The absolute value of the calculated differences between the actual EMPA determined oxide concentrations and the PLS calculated amounts were averaged from Table 1.6 and Table 1.7

	Predicted Oxide	**RMSE-P (wt%)**	**NRMSE-P**	**|Avg. Difference| (wt%)**
General Model	SiO_2	2.38	0.08	0.44
	Na_2O	0.55	0.16	0.39
	K_2O	2.52	0.62	2.41
	CaO	2.57	0.21	2.32
	TiO_2	0.73	0.28	0.45
	Al_2O_3	10.89	1.78	10.8
	FeO_T	7.59	0.73	7.37
	MgO	9.98	1.04	9.84
	Average	**4.65**	**0.614**	
	Predicted Oxide	**RMSE-P (wt%)**	**NRMSE-P**	**|Avg. Difference| (wt%)**
BAS Model	SiO_2	1.78	0.06	1.75
	Na_2O	1.04	0.31	1.00
	K_2O	0.59	0.15	0.57
	CaO	3.24	0.26	2.80
	TiO_2	1.23	0.48	1.01
	Al_2O_3	3.66	0.60	2.85
	FeO_T	1.84	0.18	0.63
	MgO	3.30	0.35	3.14
	Average	**2.08**	**0.296**	

Table 1.6. Le Losq et al. (2019) report "actual" oxide concentrations vs "general" PLS model predicted oxide concentration values for the 11 Le Losq test samples (n_{test} = 11), following application of the pertinent oxide model (Act. = actual concentration; Pred. = predicted; Diff. = difference between actual and predicted value)

Sample	SiO_2			Na_2O			K_2O			CaO		
	Act.	**Pred.**	**Diff.**	**Act.**	**Pred.**	**Diff.**	**Act.**	**Pred.**	**Diff.**	**Act.**	**Pred.**	**Diff.**
Alv 2262-8	50.53	52.13	1.6	2.43	3.03	0.6	0.14	1.97	1.8	11.9	8.50	-3.4
CH98 DR11	50.32	51.70	1.4	2.6	2.81	0.2	0.07	2.28	2.2	11.01	8.73	-2.3
111241-1	50.93	44.97	-6.0	2.63	3.15	0.5	0.17	3.32	3.2	10.81	10.93	0.1
111237-67	50.18	50.87	0.7	2.72	2.65	-0.1	0.16	2.53	2.4	11.06	9.28	-1.8
111237-37	50.55	50.81	0.3	2.62	2.73	0.1	0.16	1.32	1.2	11.12	9.36	-1.8
111238-1	50.17	50.88	0.7	2.69	2.79	0.1	0.18	2.47	2.3	10.96	9.09	-1.9
115272-19	50.92	52.06	1.1	2.46	2.72	0.3	0.07	1.90	1.8	11.66	8.68	-3.0
115272-37	50.73	48.96	-1.8	1.94	2.94	1.0	0.24	3.57	3.3	13.29	9.49	-3.8
115272-90	50.41	49.22	-1.2	2.05	2.72	0.7	0.28	2.82	2.5	13.35	9.59	-3.8
115089-251	50.39	52.19	1.8	3.01	2.85	-0.2	0.16	2.13	2.0	11.14	8.72	-2.4
115083-41	50.61	47.10	-3.5	2.76	3.77	1.0	0.09	3.92	3.8	11.52	9.87	-1.7

Table 1.6. Continued

Sample	TiO_2			Al_2O_3			FeO_T			MgO		
	Act.	**Pred.**	**Diff.**	**Act.**	**Pred.**	**Diff.**	**Act.**	**Pred.**	**Diff.**	**Act.**	**Pred.**	**Diff.**
Alv 2262-8	1.41	1.11	-0.3	14.57	25.29	10.7	9.97	17.44	7.5	7.65	17.38	9.7
CH98 DR11	1.43	1.04	-0.4	14.55	26.48	11.9	10.18	18.75	8.6	8.43	18.08	9.7
111241-1	2.46	1.74	-0.7	13.36	25.62	12.3	14.21	19.52	5.3	5.88	16.49	10.6
111237-67	2.28	1.09	-1.2	13.62	25.61	12.0	13.51	18.36	4.9	6.12	16.98	10.9
111237-37	2.11	1.09	-1.0	13.82	25.49	11.7	12.8	18.97	6.2	6.34	19.15	12.8
111238-1	2.33	1.13	-1.2	13.48	25.81	12.3	12.99	18.37	5.4	6.21	16.75	10.5
115272-19	1.53	0.99	-0.5	15.53	26.20	10.7	9.44	18.73	9.3	7.99	18.31	10.3
115272-37	0.92	1.34	0.4	15.61	25.81	10.2	7.97	17.06	9.1	9.34	15.42	6.1
115272-90	0.92	1.25	0.3	15.64	26.01	10.4	7.95	18.60	10.6	9.09	17.67	8.6
115089-251	1.73	1.03	-0.7	15.76	25.37	9.6	10.1	17.82	7.7	7.17	17.82	10.6
115083-41	1.44	1.85	0.4	16.56	23.48	6.9	9.69	16.29	6.6	7.97	16.43	8.5

Table 1.7. Le Losq et al. (2019) report "actual" oxide concentrations vs "BAS" PLS model predicted oxide concentration values for the 11 Le Losq test samples (n_{test} = 11), following application of the pertinent oxide model (Act. = actual concentration; Pred. = predicted; Diff. = difference between actual and predicted value)

Sample	SiO_2			Na_2O			K_2O			CaO		
	Act.	**Pred.**	**Diff.**	**Act.**	**Pred.**	**Diff.**	**Act.**	**Pred.**	**Diff.**	**Act.**	**Pred.**	**Diff.**
Alv 2262-8	50.5	52.5	2.0	2.4	3.6	1.2	0.1	-0.4	-0.5	11.9	13.2	1.3
CH98 DR11	50.3	52.4	2.1	2.6	3.5	0.9	0.1	-0.4	-0.4	11.0	14.6	3.6
111241-1	50.9	52.0	1.0	2.6	3.6	0.9	0.2	-0.4	-0.6	10.8	11.4	0.6
111237-67	50.2	52.1	1.9	2.7	3.4	0.7	0.2	-0.4	-0.5	11.1	14.4	3.4
111237-37	50.6	52.4	1.8	2.6	3.5	0.9	0.2	-0.3	-0.5	11.1	12.3	1.2
111238-1	50.2	52.2	2.0	2.7	3.4	0.8	0.2	-0.4	-0.5	11.0	13.7	2.8
115272-19	50.9	52.4	1.5	2.5	3.5	1.1	0.1	-0.3	-0.4	11.7	13.8	2.2
115272-37	50.7	52.0	1.2	1.9	3.5	1.6	0.2	-0.6	-0.8	13.3	19.9	6.6
115272-90	50.4	52.1	1.7	2.1	3.5	1.5	0.3	-0.5	-0.8	13.4	17.8	4.5
115089-251	50.4	52.5	2.1	3.0	3.5	0.5	0.2	-0.3	-0.5	11.1	14.0	2.8
115083-41	50.6	52.4	1.8	2.8	3.7	1.0	0.1	-0.6	-0.7	11.5	13.5	2.0

Table 1.7. Continued

Sample	TiO_2			Al_2O_3			FeO_T			MgO		
	Act.	**Pred.**	**Diff.**	**Act.**	**Pred.**	**Diff.**	**Act.**	**Pred.**	**Diff.**	**Act.**	**Pred.**	**Diff.**
Alv 2262-8	1.4	0.5	-0.9	14.6	18.2	3.6	10.0	11.3	1.4	7.7	10.7	3.0
CH98 DR11	1.4	0.3	-1.1	14.6	19.2	4.6	10.2	11.2	1.0	8.4	10.5	2.1
111241-1	2.5	1.5	-0.9	13.4	15.0	1.6	14.2	12.6	-1.6	5.9	10.7	4.9
111237-67	2.3	0.4	-1.9	13.6	18.3	4.7	13.5	11.8	-1.7	6.1	10.2	4.1
111237-37	2.1	0.4	-1.7	13.8	18.2	4.4	12.8	11.6	-1.2	6.3	10.4	4.0
111238-1	2.3	0.5	-1.9	13.5	18.6	5.1	13.0	11.9	-1.0	6.2	10.3	4.1
115272-19	1.5	0.2	-1.3	15.5	19.2	3.7	9.4	11.6	2.2	8.0	10.4	2.5
115272-37	0.9	0.8	-0.1	15.6	17.6	2.0	8.0	11.1	3.1	9.3	10.9	1.6
115272-90	0.9	0.6	-0.3	15.6	18.1	2.4	8.0	11.2	3.2	9.1	10.8	1.7
115089-251	1.7	0.4	-1.3	15.8	18.4	2.7	10.1	10.8	0.7	7.2	10.4	3.2
115083-41	1.4	1.9	0.5	16.6	13.1	-3.4	9.7	10.5	0.8	8.0	11.3	3.4

Figures

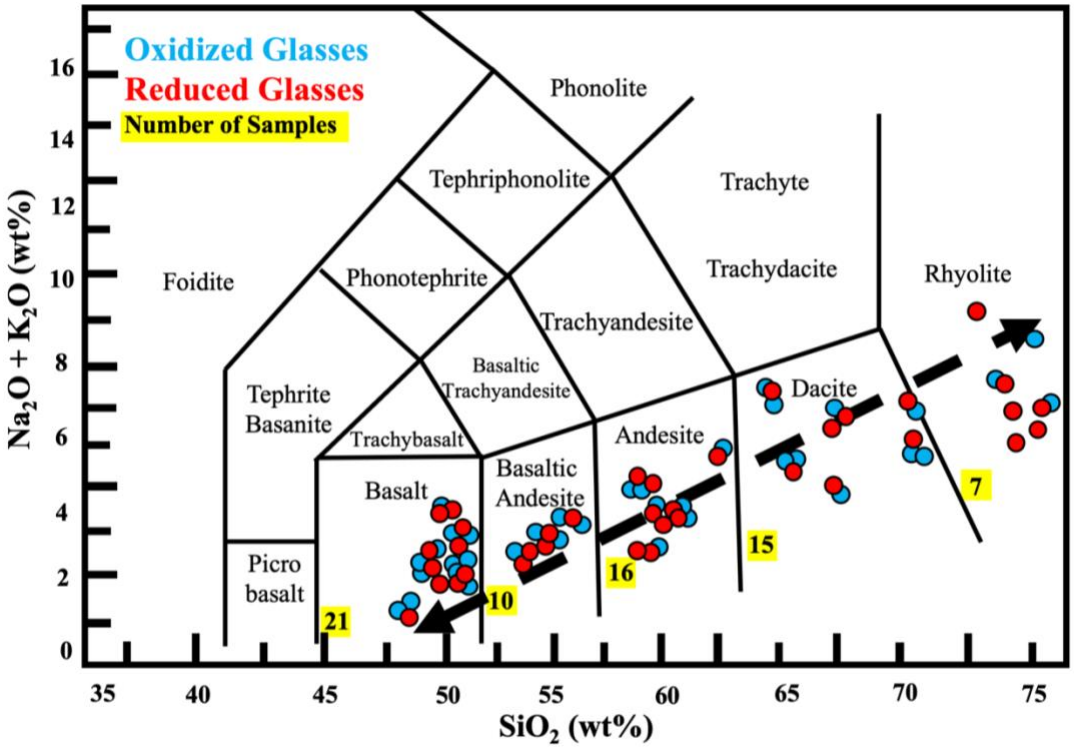

Figure 1.1. Total alkali vs. SiO_2 contents in wt% (T.A.S.) diagram denoting the chemical compositions of the Dyar glass sample suite (n = 69) used to develop and validate the PLS oxide prediction models. Red circles indicate samples equilibrated along reduced buffers: Iron-wüstite (IW), Molybdenum Oxide (Mo-MoO_2), and Quartz-Fayalite-Magnetite (QFM). Blue circles indicate samples equilibrated along oxidizing buffers: CO_2 and Air. Samples spanned the subalkaline–slightly alkaline series of all major extrusive igneous compositions suitable for common volcanic glasses (i.e., basaltic–rhyolite). Symbols slightly overlap in clusters to indicate samples with identical base compositions, but different oxidation states. The number of each compositional group (i.e., basalts, etc.) is indicated by the yellow highlighted number within the T.A.S. division

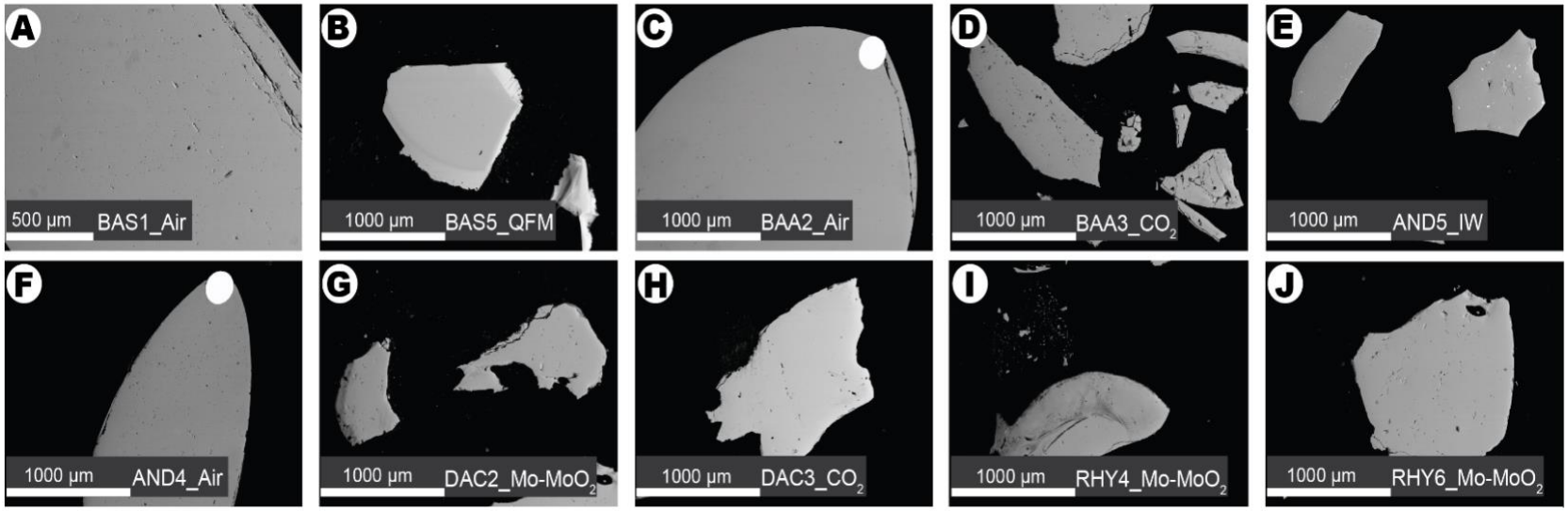

Figure 1.2. Backscattered electron images (B.S.E.) of 10 representative Dyar glass samples entrained within polished epoxy resin mounts BAS1_Air (**A**); BAS5_QFM (**B**); BAA2_Air (**C**); BAA3_CO_2 (**D**); AND5_IW (**E**); AND4_Air (**F**); DAC2_Mo-MoO_2 (**G**); DAC3_CO_2 (**H**); RHY4_Mo-MoO_2 (**I**); RHY6_Mo-MoO_2 (**J**). Shown are pairs of an oxidized Air sample and reduced along QFM example glass for each rock type. Bright white ovals observed at the tips of (**C**) and (**F**) resulted from the Pt wire loop attached to these portions of these imaged glass fragments. Little to no crystalline alteration was observed in these samples. Some images show multiple fragments of the same glass (i.e., **B**, **D**, **E**, **G**)

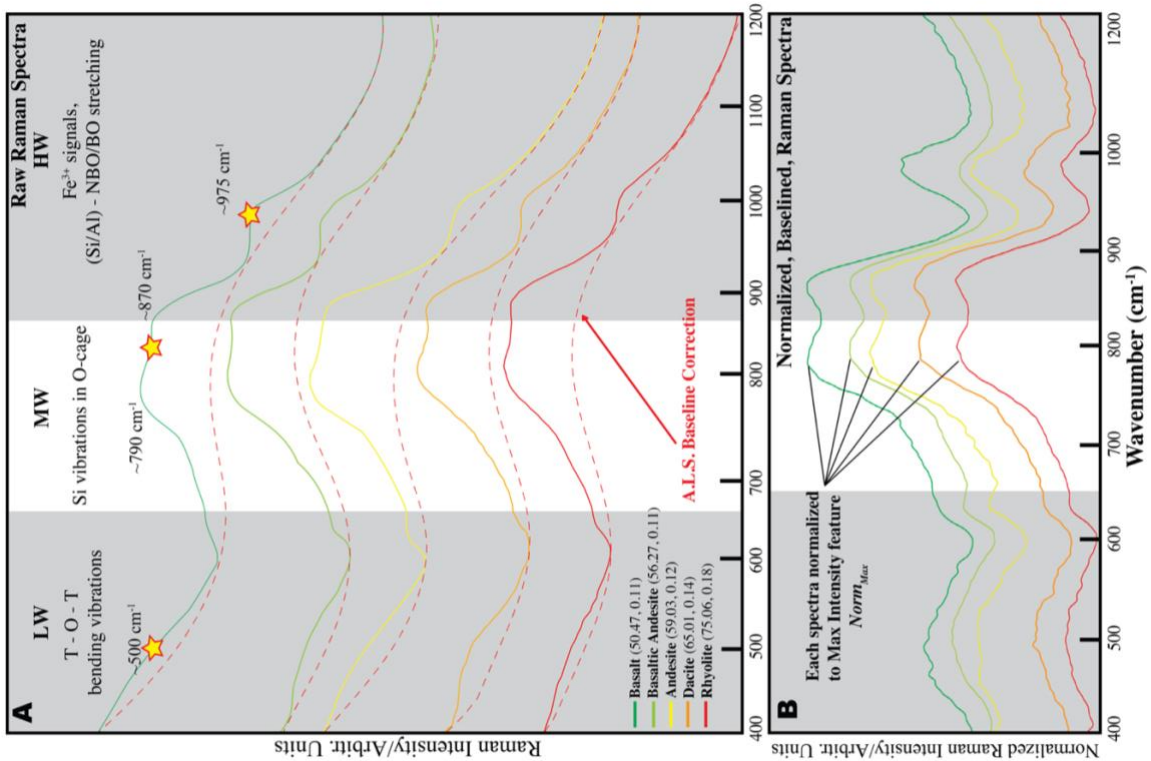

Figure 1.3. A series of five representative Raman spectra from the Dyar glass suite, with examples of reduced samples along each rock type. The spectra are color coded to the legend, where geochemical values are provided according to Sample Rock Type (SiO_2 wt%, Fe^{3+}/Fe_{TOT}). Panel (**A**) exhibits the raw Raman spectra for the samples between 400 – 1200 cm^{-1} and denotes the three primary band regions in the LW, MW, and HWs. Each band region is labeled according to the assigned vibrational signals from Giordano et al. (2020). The A.L.S. baseline correction area is shown as a dotted red line below the glass spectra, and the LW, MW, and HW band feature centroids are denoted with yellow stars. The major features were centered near ~500, 830, and 975 cm^{-1}. Panel (**B**) exhibits the Raman spectra after the A.L.S. baseline correction and normalization (along $Norm_{Max}$). The fully preprocessed spectra shown in (**B**) were utilized for model building and validation. Spectra had a low intensity band in the LW, a broad doublet arc in the MW, and a single narrower peak feature in the HW

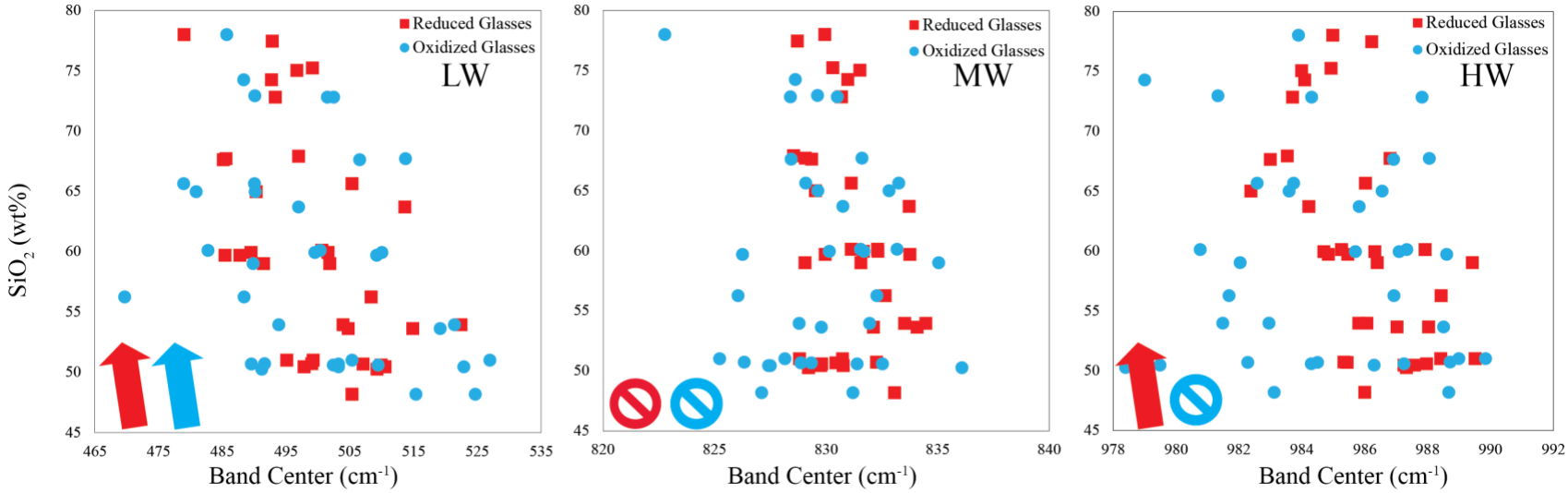

Figure 1.4. Shifts in band centroid positions for the LW, MW, and HW Raman features recorded for the 69 Dyar glass samples as a function of overall SiO2 wt% concentrations. Positional shifts were calculated using batch Gaussian peak fitting techniques within the DEVAS system, and the centers were plotted within major oxidation state groupings: Red squares = reduced glasses along Mo-MoO2, IW, or QFM; Blue circles = oxidized glasses along CO_2 and Air. Arrows indicate observed shifts, and null circles indicate no shift among the color-coded group of samples. The LW and HW bands shifted negatively according to increased SiO_2 wt%, whereas MW bands had less sensitivity to such variations. The oxidized glass samples recorded Raman spectral band centers in all regions across a much wider range, but with less concrete trends

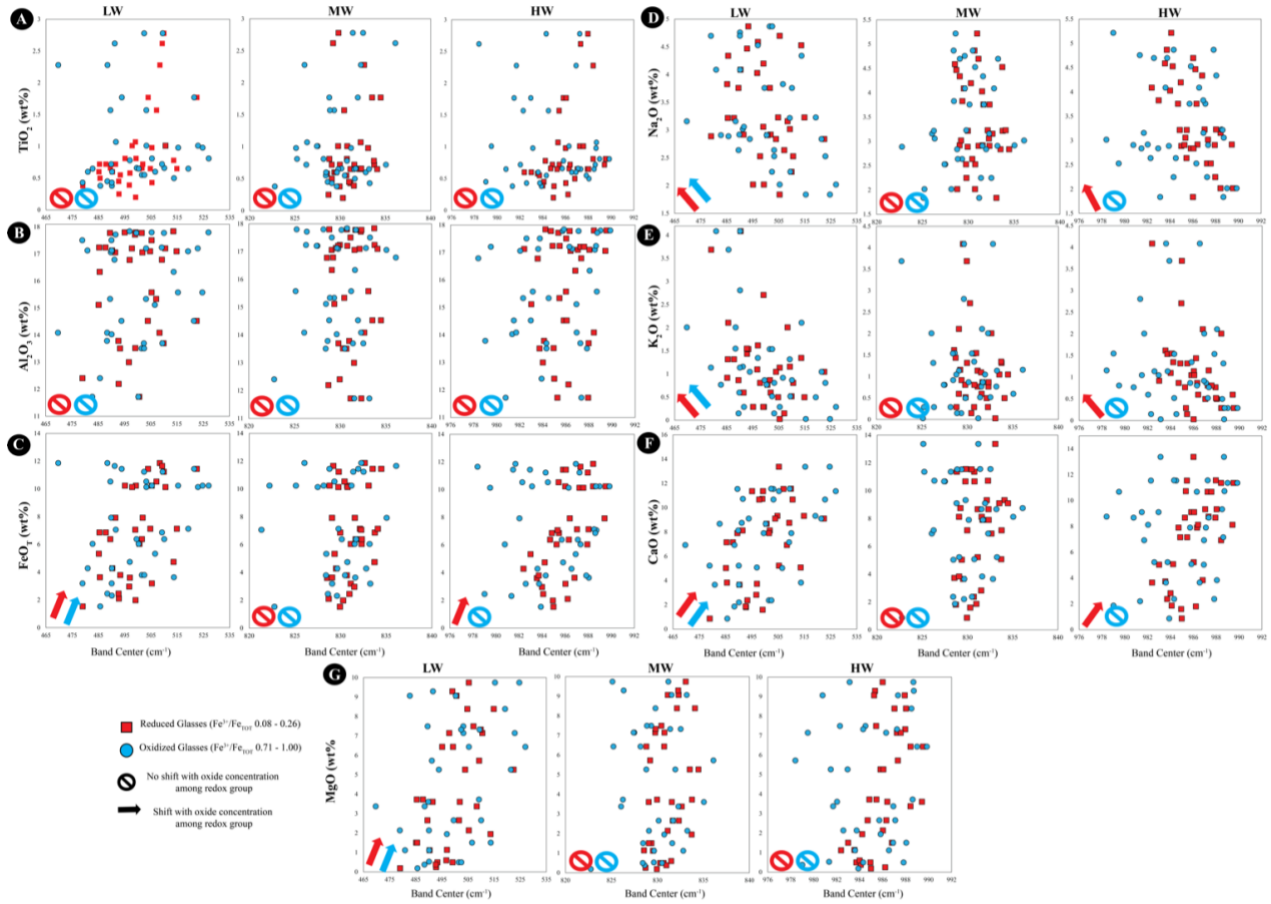

Figure 1.5. Shifts in band centroid positions for the LW, MW, and HW Raman features recorded for the 69 Dyar glass samples as a function of overall oxide concentrations for TiO_2 (A), Al_2O_3 (B), FeO_T (C), Na_2O (D), K_2O (E), CaO (F), and MgO (G). Positional shifts were calculated using batch Gaussian peak fitting techniques within the DEVAS system, and the centers were plotted within major oxidation state groupings: Red squares = reduced glasses along Mo-MoO_2, IW, or QFM; Blue circles = oxidized glasses along CO_2 and Air. Arrows indicate observed shifts, and null circles indicate no shift among the color-coded group of samples

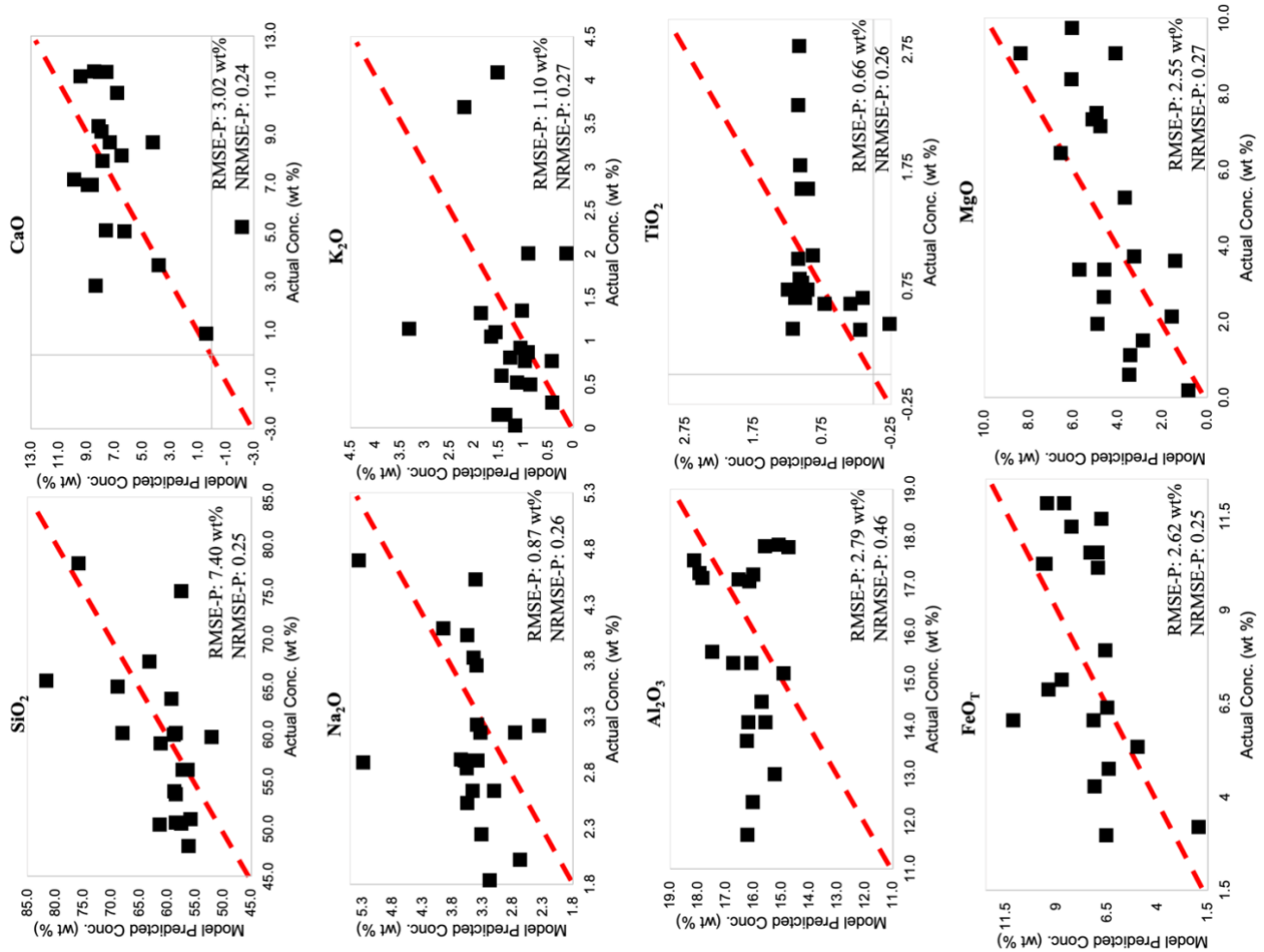

Figure 1.6. Actual vs. predicted oxide concentration output from applying the individual PLS models built and tested on the Dyar glasses. Individual RMSE-P and NRMSE-P metrics are provided. Red trendline is 1:1

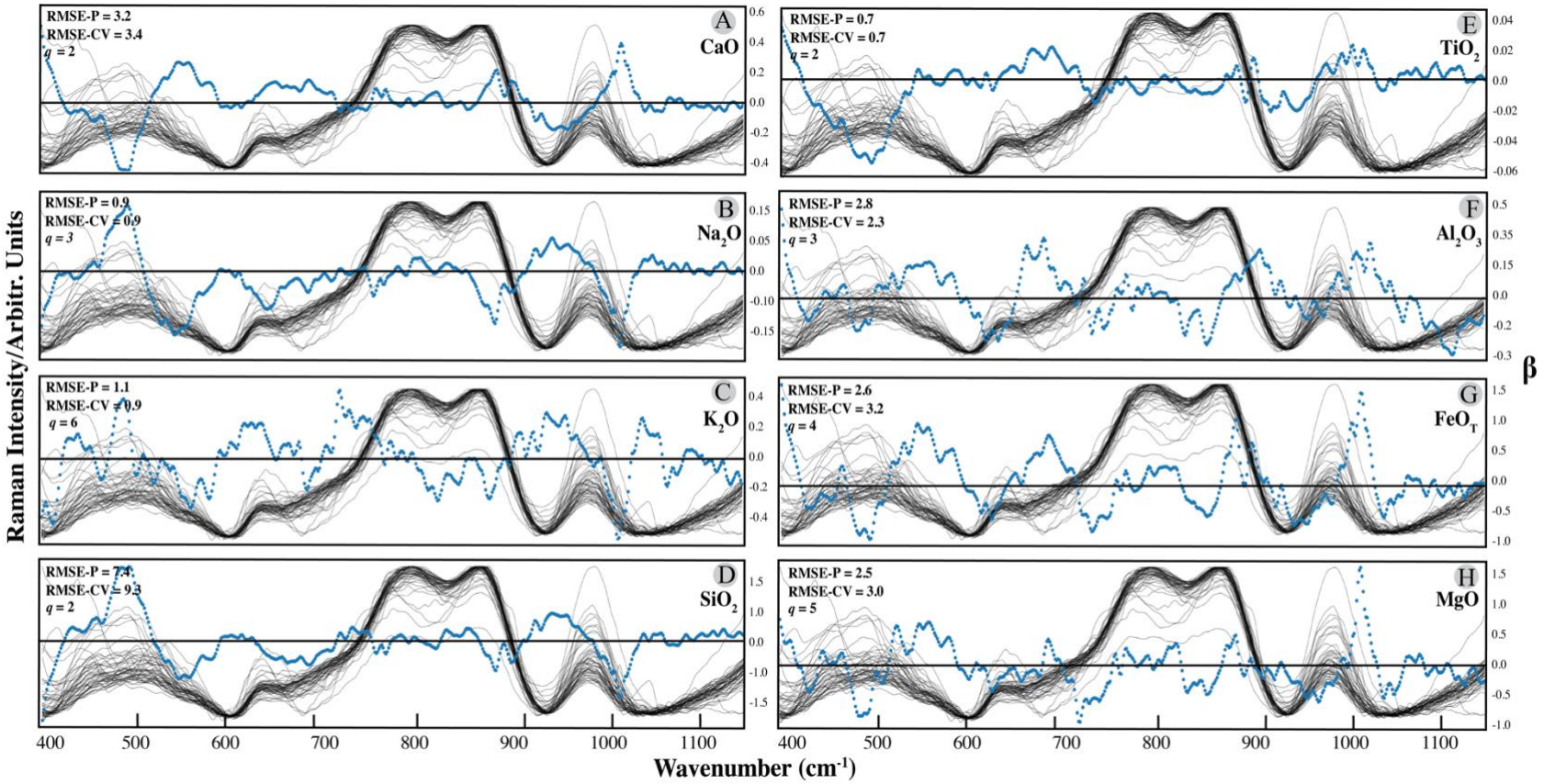

Figure 1.7. PLS model coefficients plotted against the 48 Dyar training set spectra (n_{train} = 48) for each oxide, including CaO (**A**), Na_2O (**B**), K_2O (**C**), SiO_2 (**D**), TiO_2 (**E**), Al_2O_3 (**F**), FeO_T (**G**), and MgO (**H**). Coefficient (***β***) values were unitless and measured the correlation between a selected variable prediction and individual Raman spectra channels (*p*). Coefficient values of non-zero weight and magnitude shed light on positive or negative correlations between a predicted analyte and a Raman feature. RMSE-P and CV accuracy metrics and component values are also provided

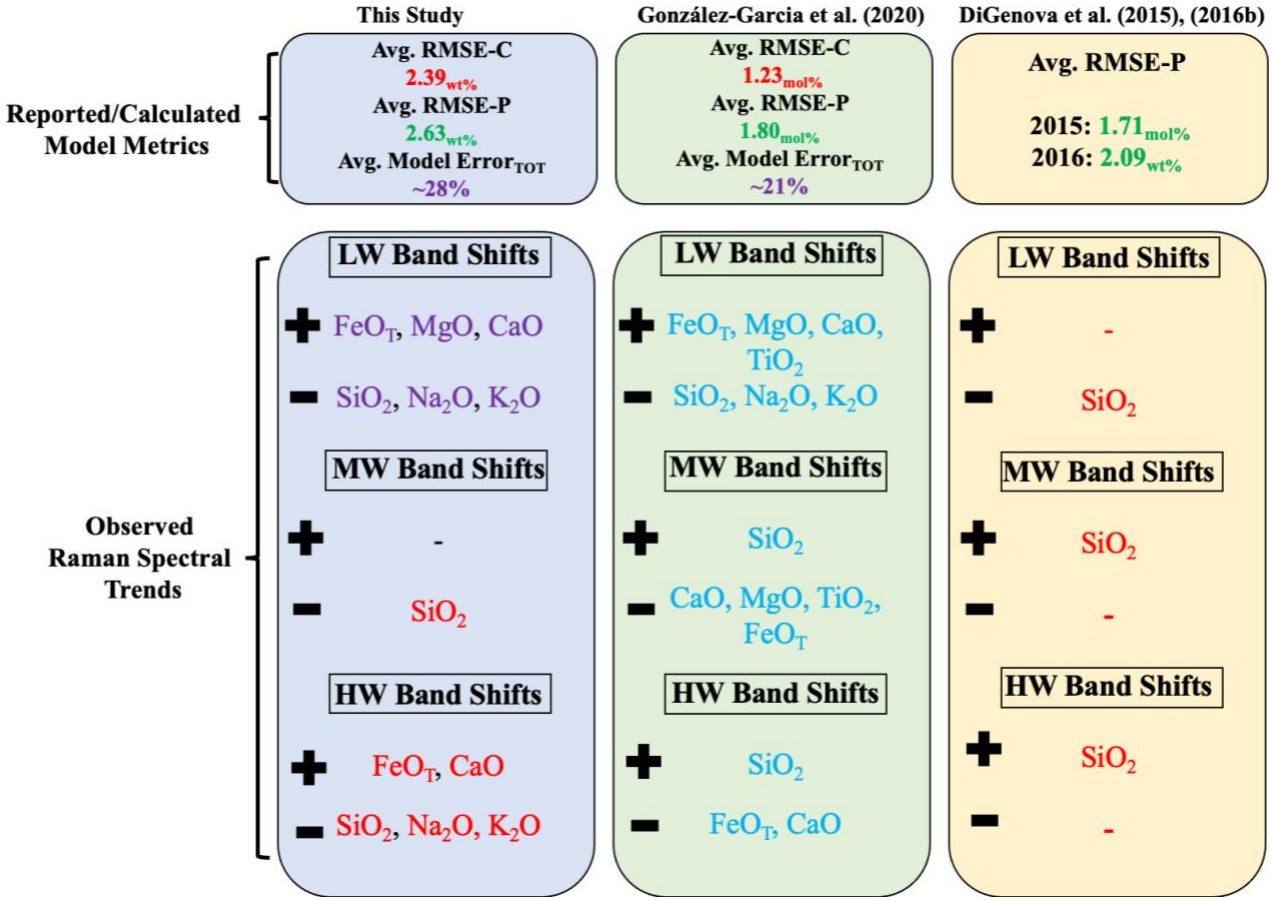

Figure 1.8. Comparisons between the reported/calculated model prediction accuracies and Raman spectral evolutionary trends resulting from this work and the work of González-Garcia et al. (2020) and Di Genova et al. (2015; 2016b). Model average metrics are reported for each study where available and were compared to contextualize the accuracy of each approach. Raman spectral trends are charted according to increasing oxide concentrations, as causing either positive (+) or negative (-) band center shifts in each region (LW, MW, and HW). Oxides labeled in red indicate trends for that variable and found in reduced spectra only. Blue indicates oxidized spectra only. Purple indicates the listed trends recorded for both groups

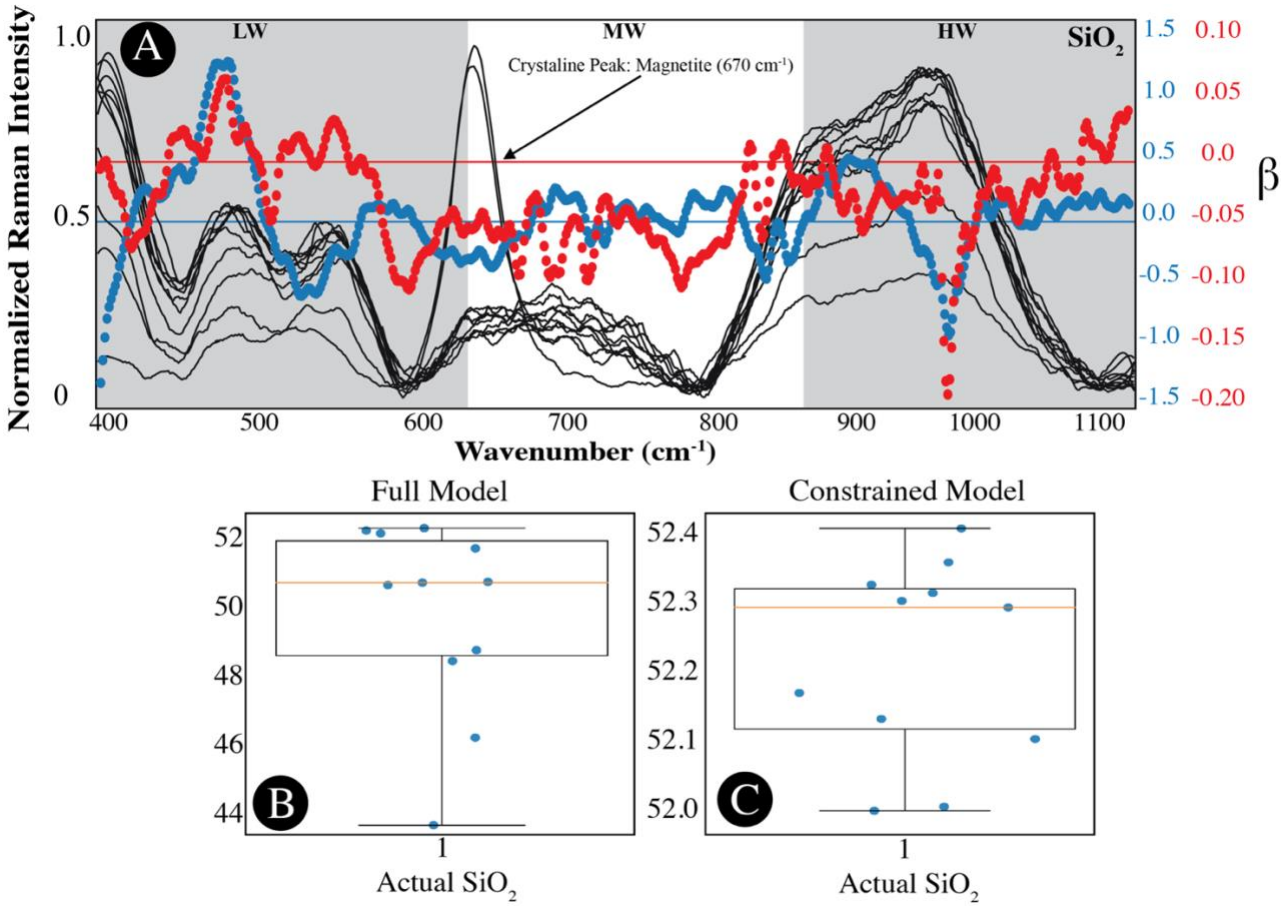

Figure 1.9. PLS model coefficients for SiO_2 predictions from the "general" (blue) and basaltic-only model (red) plotted against the 11 Le Losq MORB sample spectra (A). The general and constrained SiO_2 PLS coefficients illustrated that the models, although trained along different spectra, assigned SiO_2 influence in the exact same channel locations. SiO_2 prediction box plots for the general application (B) and constrained application (C) are also provided

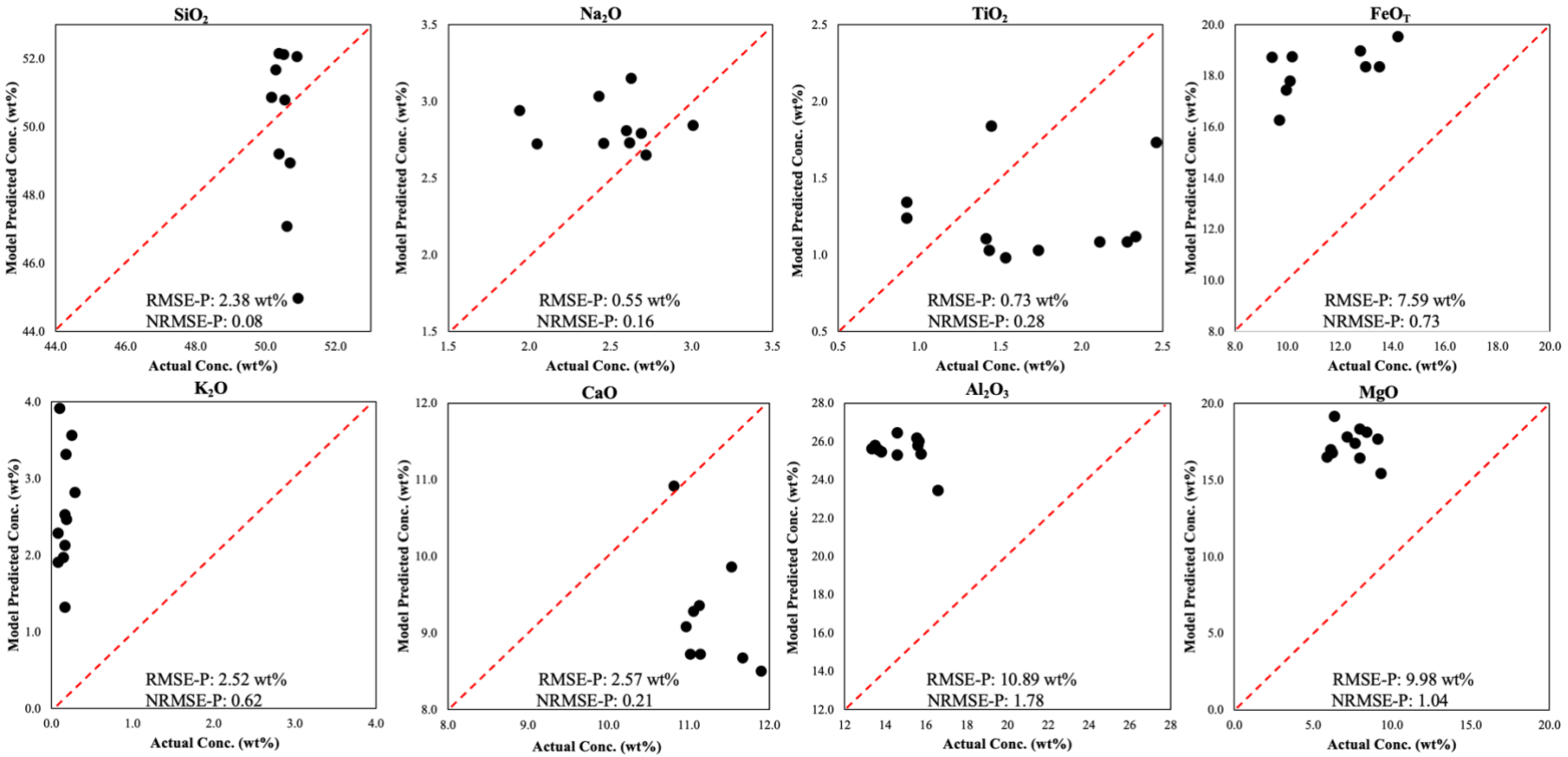

Figure 1.10. Actual vs. predicted oxide concentration output from applying the individual general PLS models tested on the Le Losq MORB glasses. Individual RMSE-P and NRMSE-P metrics are provided. Red trendline is 1:1

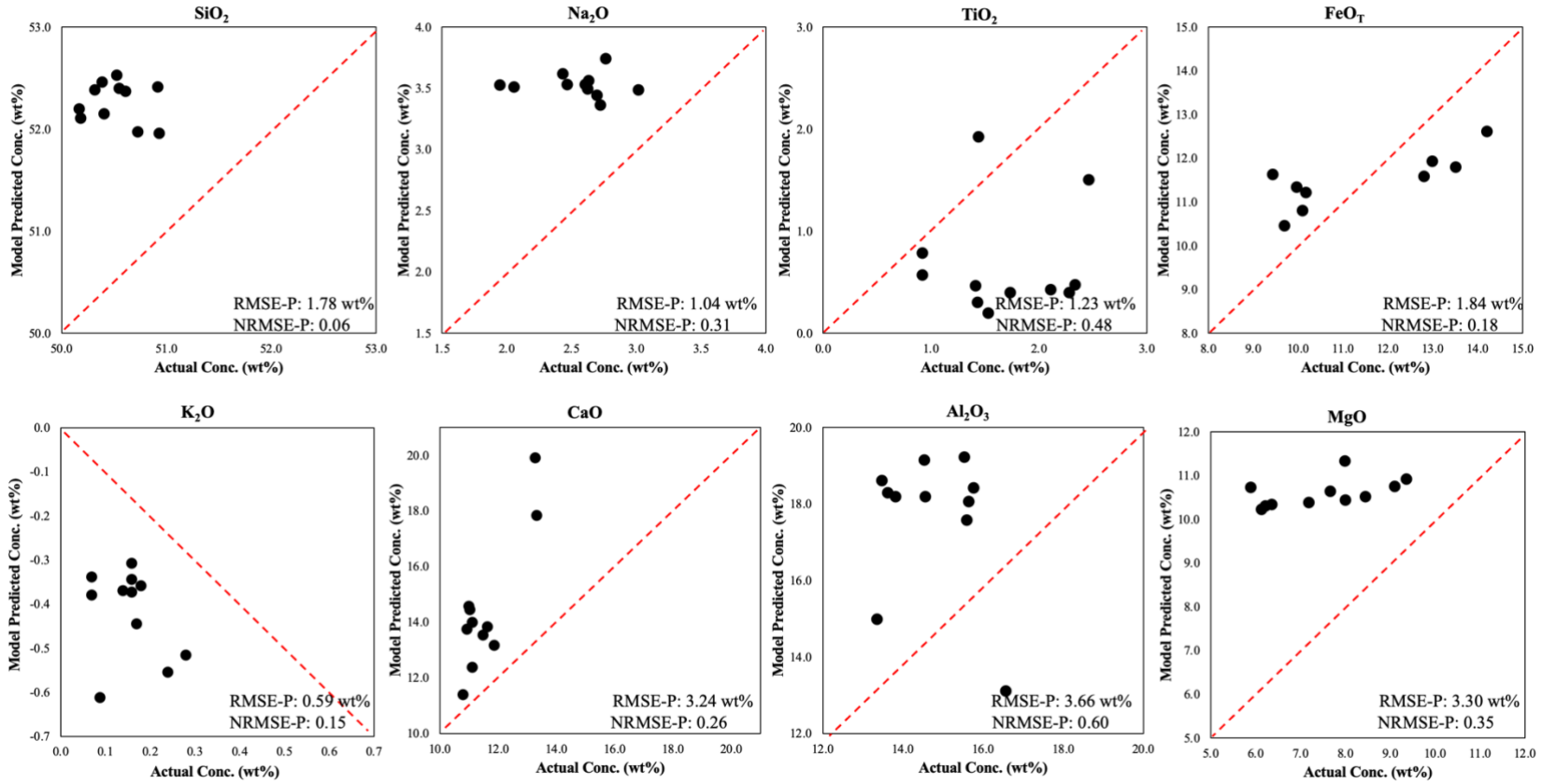

Figure 1.11. Actual vs. predicted oxide concentration output from applying the individual constrained "BAS only" PLS models tested on the Le Losq MORB glasses. Individual RMSE-P and NRMSE-P metrics are provided. Red trendline is 1:1

CHAPTER II

PREDICTING VOLCANIC GLASS OXIDATION STATE USING RAMAN SPECTROSCOPY: A COMPARATIVE ANALYSIS OF PLS AND LASSO REGRESSION METHODOLOGIES

Abstract

The Raman spectra of a suite of 78 silicate glasses are utilized to develop, validate, and test univariate Partial Least Squares (PLS-1) and Least Absolute Shrinkage and Selection Operator (LASSO) methodologies for the prediction of oxygen fugacity (f_{O2}) and polymerization conditions in volcanic glasses. The influence of Fe-redox states on the Raman-glass region between 400 – 1200 cm^{-1} is explored as a function of traditional band parameterization procedures, and through model-derived coefficients. On average, PLS-1 and LASSO regressions seem unable to adequately quantify Fe-redox related variables $\%Fe^{3+}$, $\%Fe^{2+}$, Fe_2O_3 (wt%), or FeO (wt%), but can suitably constrain melt polymerization (defined by the non-bridging oxygen to tetrahedral cation ratio (NBO/T). LASSO models perform worse overall when compared to the PLS suite, with a ~8% increase in predictive errors (RMSE-P and NRMSE-P). NBO/T normalized predictive errors are 0.24 and 0.23 for LASSO and PLS, respectively, while $\%Fe^{3+}$ errors are ~0.35 and 0.32. PLS models perform better than their LASSO equivalents due to the incorporation of additional spectral channels in the regression formula, but also utilize possible erroneous Raman signals that affect model predictive power. LASSO coefficients are limited to < 10 in all models, with most variables utilizing < 3 channels. LASSO coefficients are heavily positive on average, leading to the disregard for important negative correlations between Raman features and predicted variables. The HW Raman band in these glasses is found to shift to lower wavenumbers in response to increasing oxidation state conditions and increases in normalized intensity. PLS and LASSO methodologies are tested for such determinations in Raman signals for the first time and are found to produce much higher errors than previous modeling strategies. It is suggested here that linear based modeling strategies like PLS or LASSO fall short in directly quantifying oxidation state variables from the Raman spectra of silicate glasses but can recover overall measures of polymerization.

Introduction

Sufficient experimentation on modeling geochemical information from silicate melt proxies (i.e., volcanic glass) in the last half-century has shown that such data can be qualified and quantified using Raman spectroscopy (Di Genova et al., 2015, 2016a, 2016b; Giordano et al., 2020; González-García et al., 2020). In these investigations, chemical information such as silica, aluminum, and iron oxide concentrations were successfully obtained from volcanic glass samples using Raman spectra. Similar authors have also expanded on these works to explore the ability of Raman spectroscopy as a reliable modality by which iron oxidation state can be constrained (Di Genova et al., 2016a; Le Losq et al., 2019). These more nuanced investigations have highlighted some of the primary Raman spectral features and subtleties that arise from varying oxidation conditions and have used these attributes to determine $\%Fe^{3+}$ (Di Muro et al., 2009; Di Genova et al., 2016a, Le Losq et al., 2019). A number of publications have also prepared calibrations to directly predict structural metrics from Raman-glass signals that are closely related to the overall oxygen fugacity of the sample, such as NBO/T (polymerization) (Fu et al., 2017; Giordano et al., 2019; González-García et al., 2020). Together, such research encourages further work into new, and more accurate ways to predict Fe-redox and polymerization variables from the Raman spectra of silicate glasses.

In this work, the combined Raman spectra of a large, chemically diverse suite of silicate glasses is used to train, test, and compare quantitative models for Fe-redox parameters and NBO/T. The influence of Fe-oxidation state on the 78 glass samples is investigated as a function of Raman band positions and normalized intensity in an effort to build predictive models. This work builds off previous studies (i.e., Di Muro et al., 2009; Di Genova et al., 2016a; Le Losq et al., 2019) and attempts to expand the use of supervised machine learning (SML) in geologically relevant spectroscopy studies. Two popular machine learning regression algorithms are deployed to achieve these goals: Least Absolute Shrinkage and Selection Operator (LASSO) and Partial Least Squares (PLS). Modeling calibrations are developed and tested using the glass suite Raman data; and the two methodologies are compared to one another in their suitability in predicting the following variables: $\%Fe^{3+}$, $\%Fe^{2+}$, Fe_2O_3 (wt%), FeO (wt%), and NBO/T. This

research provides context to future utilizations of machine learning technologies in Raman-glass applications.

Methods

Silicate glass samples

The synthetic glass sample suite described in Chapter I (the Dyar suite; Dyar et al., 2016) contains 69 anhydrous, subalkaline silicate glasses spanning basaltic to rhyolitic compositions and were equilibrated across four mineral redox buffers. Fe^{3+} concentration, determined by Mössbauer spectroscopy, range between 8 – 100%. Results from additional studies were added to create a larger, more diverse sample suite of 78 glasses in total. The additional samples were incorporated into the Dyar suite to expand the number of glasses with intermediate ferric iron concentrations, due to the lack of robust coverage for those ranges in the initial Dyar suite. Upon inspection of the redox parameters (Table 2.1) of the 69 Dyar glasses, which were equilibrated and mixed according to the process outlined in the first Chapter, there exists a noticeable 'gap' in ferric iron values. The Dyar suite contains $\%Fe^{3+}$ values between 8 - 26%, and 71 - 100% (Figure 2.1), with no 'intermediate' values between ~ 26 – 70%. To fill this gap, four basaltic glasses with $\%Fe^{3+}$ values of 35, 51, 53, and 54% were incorporated from a previous study (Dufresne glasses: Dufresne et al., 2009). Additionally, five glasses from an unpublished suite (Gardner glasses) were added to the suite. The Gardner suite is comprised of two rhyolites, one dacite, and two andesites, with $\%Fe^{3+}$ values of 39, 36, 58, 27, and 70% respectively. Glass geochemical information is provided in Tables 2.1 and 2.2, with Table 2.1 denoting oxygen fugacity information, and Table 2.2 denoting the major silicate network forming oxide concentrations of the newly incorporated glasses. The Dyar suite oxide concentrations can be referenced in Chapter I. Fe-redox concentrations and their associated variables: $\%Fe^{3+}$, $\%Fe^{2+}$, log f_{O2}, ΔNNO, and ΔQFM were obtained from Mössbaur spectroscopic analysis, while total iron information was obtained via electron microprobe (EMP). Non-bridging oxygens per tetrahedrally coordinated cation ratios (NBO/T) were hand calculated according to the formula presented in Lezzi et al. (2008) for reactions between metal oxides and tetrahedrally coordained cation sums (*T*) (Eq. 2.1)

$$\frac{\mathbf{NBO}}{\mathbf{T}} = \frac{(\mathbf{2O - 4T})}{(\mathbf{T})} \qquad \mathbf{where\ NBO\ = (2O - 4T)} \qquad (\mathbf{2.1})$$

The Gardner glasses were prepared from initial composite oxide powders using the Pt/Re wire loop technique. The quenched liquids were synthetized along the manganese oxide ($MnO+Mn_3O_4$) buffer using a vertical gas mixing furnace at an average temperature of ~1050°C for 24 h and rapidly quenched. Following the experiment, glasses were placed into sealed glass vials, and individual chip separates were subjected to Mössbaur and EMP analyses to obtain geochemical information. ΔQFM and NNO metrics were calculated for the samples according to Huebner and Sato (1970) (Table 2.1) (Figure 2.1). These glasses span three extrusive rock types and have SiO_2 concentrations between 60 – 76 wt%, with a minor MnO content. Three of the glass samples contain little to no MgO or CaO, yielding a notable disparity in composition when compared to the remaining sample suites (Table 2.2).

The Dufresne glasses were prepared according to the procedures detailed in Dufresne et al. (2009) from Hawaiian tholeiitic basalt samples (e.g., Holloway and Burnham, 1972). The basaltic rocks were crushed and ground under ethanol into < 30 μm portions and then mixed with polyethylene oxide into a paste suitable for Pt/Re-wire mounting (according to O'Neill and Mavrogenes, 2002). The melts were synthetized from the dehydrated pastes at the desired intermediate fugacity conditions using a 1 atm gas-mixing furnace at ~1400°C for 5 hours, then rapidly quenched in iced water. Small aliquots of the basaltic glasses were mounted in epoxy resin rounds, and major geochemical information was obtained from separates using electron microprobe and Mössbaur analysis (Table 2.1). The samples are labelled according to their log f_{O2} relation to the NNO (nickel-nickel oxide) mineral redox buffer curve (Figure 2.3) in delta notation (log(f_{O2}) ± the NNO curve). The glasses are basaltic (< 51 SiO_2 wt%) and have low alkali contents of < 2.4 wt % ($Na_2O + K_2O$) on average. The glasses also contain minor MnO and P_2O_5 content.

Raman spectra acquisition and preprocessing

Raman spectroscopic data for the glasses were excited with a 532 nm diode laser (Coherent, Inc.) coupled to a confocal Raman microscope (Ti-U Inverted microscope, Nikon, Inc.) in the Sharma lab in the Department of Chemistry at the University of Tennessee. The samples were illuminated and focused with a 20× objective with a 0.45 numerical aperture prior to laser exposure. One < 10 μm spot per sample was exposed to the incident green laser three times over a total exposure time of 120 s, resulting in three unique Raman spectra per sample. Collection of multiple Raman spectra allows operators to review the spectra for any issues with photodegradation. The three replicate Raman spectra are averaged together to create a single Raman spectrum for each sample. The Raman Stokes scattered light was initially collected through the 20× objective in a 180° backscattering geometry, then exited the microscope before being focused onto the 100 μm slit of an ISOPlane SCT 320 spectrometer equipped with a PIXIS 400 charge-coupled device (CCD) camera detector (Teledyne Princeton Instruments). The system allowed for submicron (< 1 μm) spatial resolution over a ~5 μm spot diameter at a distance of a few millimeters from the glass surface. Laser power was locked at 10 mW. The spectrometer collected information for the Raman spectral region between ~ 400 – 1200 cm^{-1}, resulting in 519 total channels (*p*). The spectra were labeled and collected for export using the LightField® spectroscopy software as .csv files and were ingested in this format into the Data Exploration, Visualization, and Analysis for Spectroscopy web system (DEVAS) managed by Mt. Holyoke College (http://nemo.mtholyoke.edu/).

The spectra were treated with a median filter (MFilter) function in MATLAB along a 2:9 Savitzky-Golay smoothing to remove collection excitation noise and temperature, and cosmic ray flux interference signals, yielding a consistent signal-noise ratio for the raw Raman data. Following the considerations for preprocessing Raman-glass data explored in Chapter I, the data were treated with the same Asymmetric Least Squares (ALS) baseline correction using the DEVAS system (Figure 2.2), with a smoothness (λ) and asymmetry (p) factor of $5.754e^{+4}$ and 0.002 respectively (Eilers and Boelen, 2005). The spectra were normalized to their most intense feature ($Norm_{Max}$) resulting in normalized intensity value ranges between 0.0 – 1.0, where a value of 1

represents the most intense feature for that spectrum (Whiteman et al., 1993). These preprocessing options were selected via the same visual and grid-search testing scenarios described in Chapter I, where baseline parameter boundaries between 0.001 - 0.003 (p), and $4.365e^{+4} – 6.607e^{+4}$ (λ) yielded the best predictions and did not overcorrect the Raman data. The final treatment described above produced the best PLS and LASSO predictions, and the following results are the product of these extensive iterative tests.

LASSO and PLS model building

To build regression models for the prediction of the oxygen fugacity variables in this study, the 78 glass Raman spectra were randomly sorted into training and testing datasets along a ~70:30 split, with the training set containing 70% of the data (n_{train}= 54), and the testing set containing 30% (n_{test}= 24). The preprocessed spectra were used to develop (n_{train}) and validate (n_{test}) univariate PLS-1, and LASSO modeling suites for five oxidation state variables (%Fe^{3+}, %Fe^{2+}, NBO/T, FeO & Fe_2O_3 wt%) using the DEVAS system. While measures of %Fe^{3+} in silicate glasses are the primary metric by which system oxidation state is measured, %Fe^{2+} predictions offer further context for total Fe-redox determinations, as these values are directly linked to those of ferric iron. A single variable is predicted at a time using either machine learning algorithm, resulting in a PLS and LASSO modeling suite with five models each, with one model per variable.

Prediction models were initially constructed by cross validating and training PLS/LASSO algorithms with the preprocessed Raman spectra in the training set using the DEVAS system. Cross validation procedures help to ensure statistically sound modeling practices and remove user subjectivity in model development. Cross validation is a necessary step in the type of statistical modeling carried out in this work and helps to determine the optimal set of PLS/LASSO parameters that yield the best predictions. CV procedures are conducted first using the DEVAS "Regression Training" functionality, before the model is fully trained onto the training data. Here, the leave-one-out cross validation (LOO-CV) technique is used (Hastie et al., 2009). LOO-CV works by internally separating and testing the Raman data using '*k*' fold methodologies where the training data is internally split into a set *k* number of iterations, the regression trained on these iterations, then tested on *k*-1. In datasets that are large enough to split (as is the case

here where the number of spectral channel inputs help to serve as the length of the regression, i.e., 519), LOO-CV requires the number of folds (*k*) be set equal to the number of samples in the training set (i.e., $k = 54$). LOO-CV yields the PLS/LASSO model parameter value (components '*q*' in PLS, and alpha 'α' in LASSO) that yields the lowest internal prediction errors for the training suite. These parameters are then trained to obtain the final model file.

LASSO and PLS regression formulas are selected in the DEVAS system, their pertinent *q* and α parameters determined via LOO-CV, and the resulting optimized algorithm is trained to predict the targeted variable from the Raman spectra in n_{train}. Each variable yields a slightly different parameter value following these applications here, and those values are recorded in Table 2.3 and Table 2.5. following the successful training of PLS and LASSO algorithms for each target variable, the fully assembled model file is downloaded and exported for use in predicting Fe-redox from the spectra in n_{test}.

Results

Evolution of glass-based Raman signals as a function of oxidation state

Three major Raman spectral features were observed in the collection made on the 78 glass samples, and occur in the low, mid, and high wavenumber (LW, MW, HW) envelopes typically ascribed to such spectra (i.e., Rossano and Mysen, 2012). These envelopes occur between 400 – 650 cm^{-1} (LW), 650 – 850 cm^{-1} (MW), and 850 – 1200 cm^{-1} (HW) (Figure 2.1). The observed raw features include a low intensity, asymmetric band centered near ~ 500 cm^{-1}, with a well-defined leading shoulder at ~ 600 cm^{-1}, a broad doublet arc with a primary centroid near ~832 cm^{-1} spanning almost the entirety of the MW region, and a more defined narrow peak feature centered at ~979 cm^{-1}. Band centroid positions were determined via batch Gaussian peak fitting procedures within the DEVAS system. To determine and isolate the materialized influence of underlying glass oxidation state on the Raman spectra, and contextualize PLS and LASSO modeling coefficients, the positions and normalized intensities of these band features were measured as a function of overall ferric iron concentrations ($\%Fe^{3+}$). Sample spectra were stratified along SiO_2 concentration (≤ 61 SiO_2 wt% or ≥ 61 SiO_2 wt%), as glass SiO_2 concentration has been shown to affect Raman profiles (e.g., Giordano et al., 2019).

Overall, the centroid reference positions of LW and HW Raman band features show some measurable variations as glass samples become more oxidized (i.e., more $\%Fe^{3+}$) (Figure 2.3). In these regions Raman bands tend to shift towards lower wavenumbers as Fe-oxidation state increases, but these shifts were minor, exhibiting only ~10 – 20 cm^{-1} of movement among the higher SiO_2 glasses. More mafic glasses recorded Raman band shifts of a similar nature, though their positions were more scattered, leading to less defined trends. There exists a noticeable increase in the spread of band center positions between glasses containing less than ~25% Fe^{3+} and those with > 70% Fe^{3+} regardless of overall SiO_2 content (Figure 2.3). More oxidized glasses recorded Raman band centers with larger degrees of variability than did the more reduced samples, while intermediate compositions (the Dufresne and Gardner suites) helped define the slight negative shifts in overall positions. Separately, the more felsic glasses, on average, exhibit Raman band centers in more constrained regions as $\%Fe^{3+}$ increases, while the more mafic samples showed a higher degree of variance.

The LW and MW Raman bands displayed no discernable trends as a function of iron redox. However, a singular peak feature near ~980 cm^{-1} in the HW envelope recorded distinctive fluctuations with changing glass chemistry. This feature has been recorded in previous investigations of Fe-bearing silicate glass studies utilizing Raman spectroscopy and is referred to as the 'ferric iron peak' in the HW band (Di Genova et al., 2016a). To isolate trends in intensity, ALS baseline removal procedures were removed from the region between 900 – 1180 cm^{-1}, and the sample spectra stratified along Fe-redox variables ($\%Fe^{3+}$, $\%Fe^{2+}$, log f_{O2}). Removal of the baseline helped to ensure that normalized intensity differences among the spectra were the primary influence being measured (i.e., Liland et al., 2010). Figure 2.4 plots the "ferric iron peak" according to the processes described. The overwhelming trend exhibited in these Raman glass spectra is the distinctive increase in normalized intensity with increasing oxidation state (Figure 2.4A) (i.e., increasing $\%Fe^{3+}$ and log f_{O2} and decreasing $\%Fe^{2+}$). Similarly, the inverse trend is observed as ferrous iron ($\%Fe^{2+}$) increases (Figure 2.4B)

The relationship between key chemical constituents (SiO_2, Fe^{3+}/Fe^{2+}, and $Na_2O + K_2O$) and overall glass polymerization (NBO/T) was also investigated (Figure 2.5).

While all chemical constituents in a glass network contribute to overall polymerization, the above three metrics play key roles in increasing or decreasing bridging oxygen linkages. Therefore, their charted relationships with NBO/T help highlight Raman sensitivities to their influence. NBO/T (between 0 – 1.20) is observed to be negatively correlated with increasing SiO_2, Na_2O, and K_2O (Figure 2.5A, C). NBO/T is positively correlated with FeO_T up to a maximum near ~1.03, then negatively to ~ 0.51 (Figure 2.5B).

LASSO modeling results and coefficient matrices

The LASSO modeling suite for Fe-redox variables (%Fe^{3+}, %Fe^{2+}, Fe_2O_3, FeO) yields large errors (Table 2.3, Table 2.4). Internal root mean training errors and cross validation scores (RMSE-C, RMSE-CV) show no significant variations between internal and external testing power. The major measure of external predictive power for %Fe^{3+}, NRMSE-P, was reported at ~ 0.35 (± 35% possible variance with a range between 8 – 100 %: Table 2.3). Individual predictions of Fe_2O_3 and FeO oxides also have large NRMSE-P values of ~0.35 and 0.30 (~ ± 35%, ±30% variance), respectively.

NBO/T predictions, when normalized, were more accurate, with RMSE-C = 0.24, RMSE-CV = 0.30, RMSE-P = 0.23, and NRMSE-P = ~0.24 (± 24 % possible variance). While still high when compared to the range of NBO/T's in the sample suites (0.04 - 1.03) (Table 2.4), these predictions were still useable in a standalone environment. Figure 2.6 highlights the final predictive strength of each LASSO model relative to a perfect observation trend (1:1). Only NBO/T predictions were acceptable and were not consistently over or underestimated. Predictions of ferrous or ferric oxide were consistently near values between ~1.5 – 4.0 wt %. %Fe^{3+} predictions were largely underestimated for oxidized samples and overestimated for reduced samples on average.

The LASSO model coefficients shown in Figure 2.7 were mapped against the preprocessed training spectra and highlight the relationships between particular Raman spectral features (channels) and the predicted variable. LASSO assigns two strongly positive coefficients to channels in the HW envelope, near 850 and 980 cm^{-1} for predicting %Fe^{3+}, and two in the MW near ~790 cm^{-1} for predicting %Fe^{2+}, indicating these channels were considered to have strong positive correlations to these variables.

The LASSO models for Fe_2O_3 and FeO wt% use two and one channel, respectively and were similarly positioned to the ones assigned in the prediction of $\%Fe^{3+}$ and Fe^{2+}. The coefficients were also strongly positive for these variables. The NBO/T LASSO model utilizes a larger number of coefficients, assigning 10 to Raman channels across the entire range between 400 - ~1000 cm^{-1} (Figure 2.7). The NBO/T coefficients were largely congregated in the LW and HW bands between 400 – 600 cm^{-1}, and 850 – 1150 cm^{-1}. A single channel was utilized by the regression from the MW were near ~888 cm^{-1}. NBO/T coefficients were spread at both positive and negative values, at low magnitudes.

PLS modeling results and coefficient matrices

The PLS model for predicting $\%Fe^{3+}$ yields an NRMSE-P of ~ 0.32 (or ~32% possible variance) (Table 2.5). $\%Fe^{2+}$ normalized errors were identical to those of $\%Fe^{3+}$, while those of Fe_2O_3 and FeO predictions were slightly more variable, yet close (0.31 and 0.37, respectively). The predictive power of the PLS approach for all variables except NBO/T, where the NRMSE-P = 0.23, was too low to be usable in real-world scenarios (Table 2.6). LOO-CV on the samples yield PLS component values near ~3, with only the NBO/T model calling for $q > 3$. While within the theoretically acceptable 2 – 10 range of PLS regression components, these values do not yield accurate predictions. Typically, a regression trained with a lower number of components (down to 2) yields the best and most reliable predictions possible for a model (Stone and Brooks, 19900; Hastie et al., 2017) Errors hover above ~30% for all variables except NBO/T. Actual vs. predicted value charts in Figure 2.8 show how widespread model determined values were when compared to a 1:1 trend. The models yield negative predictions in a small subset of samples for all variables except $\%Fe^{3+}$, indicating possible underfitting. FeO wt% predictions lie along consistently predicted values between ~1.0 – 6.0. NBO/T, $\%Fe^{3+}$, and Fe_2O_3 predictions were less consistent, but lie closer to the 1:1 fit line.

PLS uses all available Raman spectral channels ($p = 519$) in the prediction of a target metric and assigns coefficients to these channels based on the same positive or negative correlation principles described for LASSO. In Figure 2.9, the PLS coefficients for all of the variables show unique trends across the training spectra – and exhibit little to no scatter. $\%Fe^{3+}$ and $\%Fe^{2+}$ PLS coefficients both exhibit inverse correlations for any

given channel (Figure 2.9). These coefficients show a mix of positive and negative correlations in the LW and MW bands (between 400 – 600 cm^{-1}, and 600 – 850 cm^{-1}) but show definitive shifts from positive to negative (%Fe^{3+}) or vice versa (%Fe^{2+}) for the HW band. Specifically, the model suggests strong positive correlations near the beginning of the "ferric iron band' peak to %Fe^{3+} (Di Muro et al., 2009), but switches to negative correlations as the channels increase past the peak center near ~980 cm^{-1}. The inverse trend was exhibited for the %Fe^{2+} coefficients. A similar relationship was observed for the ferric and ferrous oxide coefficients in the MW's, but not for the LW and HW. Fe_2O_3 and FeO PLS coefficients show strong negative correlations for the LW band peak near 500 cm^{-1}, mixed trends in the MW, and positive values for the HW peak. NBO/T coefficients show moderate positive and negative relationship along the spectrum, with only the channels past ~990 cm^{-1} weighted highly positive.

Discussion

Interpretations of glass-based Raman signal evolution as a function of oxidation state

The LW Raman band observed for these 78 glass samples shows only minor shifts towards lower wavenumbers in the more silicic samples as a function of increasing ferric concentration. Intermediate samples (those with %Fe^{3+} values between 26 – 70%) help highlight this shift but is in contrast to heavily reduced or oxidized samples whose high spread of positions masks any deeper insights (Figure 2.3). The MW Raman band shows no shift in average position based on %Fe^{3+}, with all centroids near ~ 830 cm^{-1}. The HW band, however, exhibits a more noticeable shift in peak position as Fe-oxidation state increases, but still is less consistent than required for accurate oxidation assessment. These shifts are more noticeable in the silicic samples, while the mafic glass positions are more scattered (Figure 2.3). It should also be noted that based on the peak shift investigations in Chapter I, the influence of %Fe^{3+} is not the only recorded influence on Raman spectra as a function chemistry but is accompanied by various oxide contributions causing shifts to higher (CaO, FeO_T) and lower (SiO_2, Na_2O, K_2O) wavenumbers. While accounting for the more minor oxide components in such isolations is not possible, SiO_2 content can be used as a discriminator. The increase in the HW Raman band intensity

with increasing Fe-oxidation state was the only intensity influence found to occur on the samples, over other chemical components in the network (Figure 2.4). These slight negative shifts of the LW and HW band centroid positions, and the increase in intensities with increasing ferric iron concentrations, are in good agreement with previous investigations; though no previous experiment has mapped such trends along such diverse glass suites, with currently published relying heavily on a single rock type, oxidation state, or endmember grouping (Di Muro et al., 2009; Di Genova et al., 2016a; 2016b).

Comparing these influences on the PLS and LASSO coefficient maps provided in Figures 2.7 and 2.9, it is obvious the HW band contains most of Fe-redox information, while the LW and MW bands are less informative. Not much can be gained from inspection of the LASSO coefficients, due to their scarcity, and therefore, the PLS coefficients provide the majority of this information. In the HW region, the PLS coefficients show a strong positive correlation to $\%Fe^{3+}$ in the lagging shoulder of the feature, and peak centroid near ~985 cm^{-1}, trending to distinctly negative values past the center and into the leading shoulder. The inverse trend is shown for the $\%Fe^{2+}$ coefficients. When compared to the ferric and ferrous oxide coefficients, there is a striking similarity between the ratioed metrics and their oxide counterparts, where Fe_2O_3 and FeO coefficients are in good agreement with $\%Fe^{3+}$ and $\%Fe^{2+}$ coefficients in the HW. Based on these findings, the PLS model coefficients in this region are seemingly influenced by the 'switch' in overall network affects arising from ferrous and ferric iron (Figure 2.5). As the peak in this band is also shown to shift to lower wavenumbers with increasing $\%Fe^{3+}$ (i.e., Figure 2.3), the positive coefficient weights for the channels leading up to this peak are sensible. The same is true for the inverse case of $\%Fe^{2+}$ coefficients. NBO/T coefficients are somewhat unique, but also display affinity for overall iron content in the band. The PLS model seemingly attributes erroneous weight to the less important spectral channels in the LW and MW bands for Fe-redox variables, as their influence is notably absent based on peak fitting procedures and prior investigations – leading to worse predictions overall for the model when compared to similar studies.

Fe-redox plays a major role in controlling overall glass polymerization, meaning the relationship between NBO/T and Fe^{3+}/Fe^{2+} ratios should be further investigated. The distinct increase then decrease trends between NBO/T measures and FeO_T in Figure 2.5 is significant. While negative correlations between networking forming cations like Si^{4+} and NBO/T or lower concentration components (e.g., Na_2O, K_2O) make sense, the relationship with FeO and Fe_2O_3 is mixed. Acting as a network former, Fe^{3+} should increase the overall polymerization of the glass network, and therefore decrease NBO/T. Fe^{2+} however, acting as a network modifier, should increase NBO/T. These trends are more diluted according to the calculation plotted in Figure 2.5. Here, NBO/T increases alongside total Fe content to a maximum near ~10.5 wt%, then suddenly decreases. This unique inflection trend has been observed in silicate glass Raman studies previously (e.g., Xuan et al., 2020) and is explained by the simultaneous opposing effects of ferric and ferrous iron in the glass network (Table 2.1). The NBO/T relationship is caused by the increasing presence of Fe^{3+} and Fe^{2+} as FeO_T increases. Initially, at lower FeO_T concentrations, the depolymerizing effect of ferrous iron is stronger than the polymerizing effect of the oxidized Fe^{3+}, causing NBO/T values to increase. As total iron content increases, the average abundance of Fe^{3+} in these samples increases, eventually overtaking the effect of Fe^{2+} and causing NBO/T values to decrease. This is also a likely consequence of other network forming or weak oxides competing in the glass network, which is why NBO/T decreases consistently as both SiO_2 and the alkaline content (Na_2O + K_2O) of the glasses increases.

Comparisons of the LASSO and PLS modeling suites

Overall, the PLS modeling suite exhibits stronger predictive power for the major oxidation state variables targeted in this study (%Fe^{3+}, %Fe^{2+}, Fe_2O_3, FeO, and NBO/T) than does the LASSO calibration. However, while this increase in predictive accuracy is ~22% on average, the PLS model does not produce reliable predictions for variables other than NBO/T. Neither of the regression methodologies predicts Fe-redox variables from the Raman spectra of silicate glasses with errors less than ~30% (variance). LASSO model predictions also showed distinct trends in both over and underpredicting these variables, whereas the PLS models showed no affinity for either in any variable. The PLS

suite also exhibited a larger number of negative predictions, leading to less reliability in the error metrics.

Although both methods were unreliable, the PLS suite performed slightly better than the LASSO technique. As PLS utilizes every spectral channel available in the prediction of an oxidation variable, it is possible that the incorporation of more channels across the entire spectrum is critical in these determinations for Raman signal in glass. However, as the PLS approach is still unusable from a statistical standpoint, it is likely that oxidation-related information is contained in more of the Raman spectrum than is incorporated by LASSO, but using every available channel introduces erroneous information into the PLS regression predictions. The higher amount of non-zero coefficients in the PLS regression is possibly too much for the regression to handle, leading to poor predictive accuracy. Similarly, the LASSO technique highlights the notion that utilizing too few channels with large positive coefficient weights can lead to even worse predictions. This indicates that the use of too few channels is also leading to lower accuracies. The PLS coefficients are also shown to vary between positive and negative values (Figure 2.9), whereas the LASSO values remain strongly positive for $\%Fe^{3+}$, $\%Fe^{2+}$, Fe_2O_3, and FeO (Figure 2.7); indicating there may be important negatively correlated Raman features ignored by the LASSO regression. When LASSO incorporates negative correlations, as is the case in the prediction of NBO/T, predictive power increases (Table 2.3). Figure 2.10 compares the LASSO and PLS regressions against one another visually for the predictions of $\%Fe^{3+}$ and NBO/T. When predicting $\%Fe^{3+}$, neither of the techniques show best-fit trend lines near the 1:1 line, both show a near-horizontal regression function with $R^2 < 0.2$. When predicting NBO/T, both models yield acceptable results with only slight variations between the methods.

Contextualization of the LASSO and PLS modeling suites

Studies focused on the determination of Fe-redox (Di Muro et al., 2009; Di Genova et al., 2016a) and structural parameters such as NBO/T (González-García et al., 2020) in silicate glasses using Raman spectroscopy combined with multivariate methods are limited, but powerful. Few previous works have utilized machine learning technologies such as that in this research. A notable exception is the study of Le Losq et

al. (2019). To contextualize the PLS and LASSO modeling results for redox variables and NBO/T presented in this work, comparisons are drawn between the dominant modeling methodologies currently available.

Previous work on Fe-redox chemical calibrations in Raman-glass signals utilizes band deconvolution procedures reliant on end-member compositions (peralkaline rhyolite, iron-rich basalt) to measure band position and height ratios of the HW envelope (~1000 cm^{-1}) in the collected spectrum to establish determinations of Fe valence ratios and is highly compositionally dependent (Di Muro et al., 2009). As the HW envelope was found to be strongly sensitive to iron redox state, the ratios of heights and positions of this band in between separate rock types was exploited to great effect, generating predicted line fits of <1% in disparity when compared to calibration spectra, with errors near ~5 – 8%.

Di Genova et al. (2016a) similarly deployed oxide prediction models as described in Di Genova et al. (2015), in which $\%Fe^{3+}$ was determined by parameterizing the HW band between ~970 – 1040 cm^{-1} using an empirical formula and fit parameter (R_P) to relate the Raman spectra of oxidized and reduced endmembers to a new spectrum. The authors were able to predict Fe^{3+}/Fe_{tot} metrics with an accuracy of < 5% on average. However, the methodology relied on the need for knowledge of an endmember reference sample (spectrum) to produce a prediction using the mixing model. The modeling strategy is also highly optimized for constrained compositions, limiting its overall generalizability.

Fe-redox modeling of mid-ocean ridge basaltic (MORB) glass samples in Le Losq et al. (2019) were the first of their kind to utilize machine learning. Though no major increase in overall model accuracy was found for the proposed methods employed, the study averaged high precision, with average predictive root mean square deviations (RMSD) < 0.04 for all six of the methodologies tested to determine $\%Fe^{3+}$, translating to extremely small differences between actual and predicted values.

When compared to these previous investigations, the PLS and LASSO Fe-redox predictions are inadequate and contain extremely high errors, while the determinations of NBO/T values begin to approach more comparable metrics. It is suggested that because

PLS and LASSO rely on modified forms of multiple linear regression, in which there are assumed linear relationships between the dependent variables and the independent variable being predicted, and the relationship between redox variables in Raman-glass signals is more complex, accurate predictions of such variables is not possible using these techniques in these adaptations.

Conclusions

The Raman spectra of 78 synthetic silicate glass samples with large variations in compositions and oxidation state are used to develop and test PLS and LASSO-based regression models for the prediction of $\%Fe^{3+}$, $\%Fe^{2+}$, Fe_2O_3, FeO, and NBO/T. It is determined that the predictive power of both the PLS-1 and LASSO techniques is lacking for all targeted variables except NBO/T, with average NRMSE-P metrics near ~30 – 35% for $\%Fe^{3+}$, $\%Fe^{2+}$, Fe_2O_3, and FeO. NBO/T predictions average NRMSE-P of ~24%, closer to a more useable metric when directly compared to similar studies but is still high.

The influence of Fe-redox variables in the Raman spectra of silicate glasses is explored as a function of both peak position, normalized intensity, and polymerization. The findings that the intensity of the 'ferric iron band' near ~980 cm^{-1} increases with increasing ferric iron concentration, and shifts to lower wavenumbers along the same trend, is in good agreement with previous Raman-glass investigations and the determined model coefficients. The peak shifting effect is found to be more prominent in the higher SiO_2 glass samples (dacites and rhyolites). No significant shifts or intensity differences were found in the LW and MW regions of the spectrum as a function of oxidation state. The relationship between polymerization (NBO/T) and different chemical components in the glass samples showed that polymerization increases with increasing SiO_2, and Na_2O+K_2O concentrations. The relationship between NBO/T and glass Fe^{3+}/Fe^{2+} is unique, with NBO/T increasing as FeO_T increases up to a boundary where the effects of Fe^{3+} become more dominant due to the higher proportion of Fe_2O_3 in the system.

While the coefficients from the PLS and LASSO suites seem to accurately identify the influences of the underlying glass oxidation chemistry in the Raman data, their predictions are unusable in a practical sense, suggesting there is a central point of failure between the methodologies in applications to constraining Fe-redox variables as a

function of Raman-glass signals. As other modeling techniques have shown great success and accuracy in such applications, the point of failure must lie with the PLS and LASSO methodologies themselves.

References

Berry, A.J. and O'Neill, H.S.C. (2021) "Oxygen content, oxygen fugacity, the oxidation state of iron, and mid-ocean Ridge Basalts," *Magma Redox Geochemistry*, pp. 155–163. Available at: https://doi.org/10.1002/9781119473206.ch8.

Breitenfeld, L.B. *et al.* (2018) "Predicting olivine composition using Raman spectroscopy through band shift and multivariate analyses," *American Mineralogist*, 103(11), pp. 1827–1836. Available at: https://doi.org/10.2138/am-2018-6291.

Cukierman, M. and Uhlmann, D.R. (1974) "Effects of iron oxidation state on viscosity, lunar composition 15555," *Journal of Geophysical Research*, 79(11), pp. 1594–1598. Available at: https://doi.org/10.1029/jb079i011p01594.

Di Genova, D. *et al.* (2015) "Approximate chemical analysis of volcanic glasses using Raman spectroscopy," *Journal of Raman Spectroscopy*, 46(12), pp. 1235–1244. Available at: https://doi.org/10.1002/jrs.4751.

Di Genova, D. *et al.* (2016a) "Models for the estimation of Fe^{3+}/Fe_{tot}ratio in terrestrial and extraterrestrial alkali- and iron-rich silicate glasses using Raman Spectroscopy" *American Mineralogist*, 101(4), pp. 943–952. Available at: https://doi.org/10.2138/am-2016-5534ccbyncnd.

Di Genova, D. *et al.* (2016b) "Raman spectra of Martian glass analogues: A tool to approximate their chemical composition," *Journal of Geophysical Research: Planets*, 121(5), pp. 740–752. Available at: https://doi.org/10.1002/2016je005010.

Di Muro, A. *et al.* (2009) "Micro-raman determination of iron redox state in dry natural glasses: Application to Peralkaline Rhyolites and Basalts," *Chemical Geology*, 259(1-2), pp. 78–88. Available at: https://doi.org/10.1016/j.chemgeo.2008.08.013.

Dingwell, D.B. (1991) “Redox viscometry of some Fe-bearing silicate melts,” *American Mineralogist*, 76, 1560–1562.

Dufresne, C.D.M. *et al.* (2009) “Effect of SIO2, total FeO, Fe^{3+}/Fe^{2+}, and alkali elements in basaltic glasses on mid-infrared Spectra,” *American Mineralogist*, 94(11-12), pp. 1580–1590. Available at: https://doi.org/10.2138/am.2009.3113.

Dyar, M.D. and Ytsma, C.R. (2021) “Effect of data set size on geochemical quantification accuracy with laser-induced breakdown spectroscopy,” *Spectrochimica Acta Part B: Atomic Spectroscopy*, 177, p. 106073. Available at: https://doi.org/10.1016/j.sab.2021.106073.

Dyar, M.D. *et al.* (2012) “Comparison of partial least squares and LASSO regression techniques as applied to laser-induced breakdown spectroscopy of geological samples,” *Spectrochimica Acta Part B: Atomic Spectroscopy*, 70, pp. 51–67. Available at: https://doi.org/10.1016/j.sab.2012.04.011.

Dyar, M.D. *et al.* (2016) “Accurate predictions of iron redox state in silicate glasses: A multivariate approach using X-ray absorption spectroscopy,” *American Mineralogist*, 101(3), pp. 744–747. Available at: https://doi.org/10.2138/am-2016-5555.

Eilers, P., and Boelens, H. (2005). “Baseline Correction with Asymmetric Least Squares Smoothing”. *Unpubl. Manuscr.* Algorithm at:https://www.wavemetrics.com/code-snippet/asymmetric-least-squares-smoothing

Frost, B.R. (1991) “Chapter 1.Introduction to oxygen fugacity and its petrologic importance,” *Oxide Minerals*, pp. 1–10. Available at: https://doi.org/10.1515/9781501508684-004.

Fu, X., Wang, A. and Krawczynski, M.J. (2017) “Characterizing amorphous silicates in extraterrestrial materials: Polymerization effects on Raman and mid-IR spectral features of alkali and alkali earth silicate glasses,” *Journal of Geophysical Research: Planets*, 122(5), pp. 839–855. Available at: https://doi.org/10.1002/2016je005241.

Giordano, D. *et al.* (2019) “A calibrated database of Raman spectra for natural silicate glasses: Implications for modelling melt physical properties,” *Journal of Raman Spectroscopy*, 51(9), pp. 1822–1838. Available at: https://doi.org/10.1002/jrs.5675.

Giordano, D. *et al.* (2020) “Raman spectroscopy from laboratory and proximal to remote sensing: A tool for the Volcanological Sciences,” *Remote Sensing*, 12(5), p. 805. Available at: https://doi.org/10.3390/rs12050805.

González-García, D. *et al.* (2020) “A Raman spectroscopic tool to estimate chemical composition of natural volcanic glasses,” *Chemical Geology*, 556, p. 119819. Available at: https://doi.org/10.1016/j.chemgeo.2020.119819.

Hastie, T., Robert, T., and Friedman, J. (2017) “The elements of statistical learning: data mining, inference, and prediction,” *Springer Science & Business Media*

Hastie, T., Tibshirani, R. and Friedman, J. (2009) “The elements of Statistical Learning,” *Springer Series in Statistics*. Available at: https://doi.org/10.1007/978-0-387-84858-7.

Henderson, G.S. (2005) “The structure of silicate melts: A glass perspective,” *The Canadian Mineralogist*, 43(6), pp. 1921–1958. Available at: https://doi.org/10.2113/gscanmin.43.6.1921.

Holloway, J.R. and Burnham, C.W. (1972) “Melting relations of basalt with equilibrium water pressure less than total pressure,” *Journal of Petrology*, 13(1), pp. 1–29. Available at: https://doi.org/10.1093/petrology/13.1.1.

Karki, B.B. (2010) “First-principles molecular dynamics simulations of silicate melts: Structural and dynamical properties,” *Reviews in Mineralogy and Geochemistry*, 71(1), pp. 355–389. Available at: https://doi.org/10.2138/rmg.2010.71.17.

Le Losq, C. *et al.* (2019) “Determination of the oxidation state of iron in Mid-Ocean Ridge basalt glasses by Raman spectroscopy,” *American Mineralogist*, 104(7), pp. 1032–1042. Available at: https://doi.org/10.2138/am-2019-6887.

Lezzi, G *et al.* (2008) “Solidification behaviour of natural silicate melts and volcanological implications,” *In Volcanoes: Formation, Eruptions and Modelling – Chp. 6*, ISBN: 978-1-60692-916-2. Available at: https://www.earth-prints.org/handle/2122/3990

Liebske, C. *et al.* (2003) “The influence of pressure and composition on the viscosity of andesitic melts,” *Geochimica et Cosmochimica Acta*, 67(3), pp. 473–485. Available at: https://doi.org/10.1016/s0016-7037(02)01139-0.

Lofgren, G. (1980) “Chapter 11. experimental studies on the dynamic crystallization of silicate melts,” *Physics of Magmatic Processes*, pp. 487–552. Available at: https://doi.org/10.1515/9781400854493.487.

McCanta, M.C. *et al.* (2015) “Identifying cryptotephra units using Correlated Rapid, nondestructive methods: VSWIR spectroscopy, x-ray fluorescence, and magnetic susceptibility,” *Geochemistry, Geophysics, Geosystems*, 16(12), pp. 4029–4056. Available at: https://doi.org/10.1002/2015gc005913.

Mysen, B. O., Virgo, D., and Seifert, F. (1984) “Redox equilibria of iron in alkaline earth silicate melts: relationships between melt structure, oxygen fugacity, temperature

and properties of iron-bearing silicate liquids," *American Mineralogist,* 69 (1984) pp. 834–847.

Mysen, B.O., Virgo, D. and Seifert, F.A. (1982) "The structure of silicate melts: Implications for chemical and physical properties of natural magma," *Reviews of Geophysics,* 20(3), p. 353. Available at: https://doi.org/10.1029/rg020i003p00353.

Naslund, H.R. (1983) "The effect of oxygen fugacity on liquid immiscibility in iron-bearing silicate melts," *American Journal of Science,* 283(10), pp. 1034–1059. Available at: https://doi.org/10.2475/ajs.283.10.1034.

O'Neill H.S.C., and Mavrogenes, J. A. (2002) "The sulfide capacity and the sulfur content at sulfide saturation of silicate melts at 1400degreesc and 1 bar," *Journal of Petrology,* 43(6), pp. 1049–1087. Available at: https://doi.org/10.1093/petrology/43.6.1049.

Richet, P. (1984) "Viscosity and configurational entropy of silicate melts," *Geochimica et Cosmochimica Acta,* 48(3), pp. 471–483. Available at: https://doi.org/10.1016/0016-7037(84)90275-8.

Stebbins, J.F. (1995) "Structure, dynamics, and properties of silicate melts." Available at: https://doi.org/10.1515/9781501509384.

Stone, M. and Brooks, R.J. (1990) "Continuum regression: Cross-validated sequentially constructed prediction embracing ordinary least squares, partial least squares and principal components regression," *Journal of the Royal Statistical Society: Series B (Methodological),* 52(2), pp. 237–258. Available at: https://doi.org/10.1111/j.2517-6161.1990.tb01786.x.

Whiteman, D.N. *et al.* (1993) "Temperature sensitivity of an atmospheric Raman lidar system based on a XEF excimer laser," *Optics Letters,* 18(3), p. 247. Available at: https://doi.org/10.1364/ol.18.000247.

Xuan, W. *et al.* (2020) “Different roles of iron species in the network structure and viscosity of silicate melts,” *Energy & Fuels*, 34(11), pp. 13698–13706. Available at: https://doi.org/10.1021/acs.energyfuels.0c02425.

Appendix

Tables

Table 2.1. Oxidation state related geochemical data for the 78 glass samples included in this study, specifically Dyar (BAS – RHY series), Dufresne (Du series), and Gardner (G series). Major oxygen fugacity (f_{O2}), Fe valence state ratios (%Fe^{3+}, %Fe^{2+}), log-value relationships to the quart-fayalite-magnetite (QFM) and nickel-nickel-oxide (NNO) mineral redox buffers, and polymerization (NBO/T) metrics were recorded for each sample, following equilibration based on methodology described in the text. Mo: Mo-MoO_2 buffer; IW: iron-wüstite buffer; CO_2: carbon-dioxide buffer; Air: ambient conditions

Sample Name	%Fe^{3+}	%Fe^{2+}	Log(f_{o2})	f_{o2} ΔQFM	f_{O2} ΔNNO	NBO/T
BAS7_Air	90	10	-2.29	5.3	4.65	0.97
BAS7_CO_2	85	15	-2.29	5.3	4.65	1.00
BAS7_QFM	20	80	-7.6	0	-0.65	0.92
Du_+3d	51	49	-2.95	3.18	3	1.03
Du_+5	54	46	-0.68	5.45	5	0.99
Du_+3	53	47	-2.95	3.18	3	0.99
BAS3_CO_2	81	19	-2.29	5.3	4.65	0.71
BAS3_QFM	22	78	-7.6	0	-0.65	0.63
BAS5_Air	90	10	-2.29	5.3	4.65	0.76
BAS5_IW	26	74	-11.28	-3.5	-4.15	0.75
BAS5_QFM	11	89	-7.6	0	-0.65	0.69
BAS5_CO_2	78	22	-2.29	5.3	4.65	0.67
BAS6_Air	90	10	-2.29	5.3	4.65	0.91
BAS6_QFM	17	83	-7.6	0	-0.65	0.90
BAS6_CO_2	79	21	-2.29	5.3	4.65	0.81
BAS2_CO_2	80	20	-2.29	5.3	4.65	0.86
BAS2_Air	90	10	-2.29	5.3	4.65	0.84
BAS2_QFM	12	88	-7.6	0	-0.65	0.76

Table 2.1. Continued

Sample Name	%Fe^{3+}	%Fe^{2+}	Log(f_{O2})	f_{O2} ΔQFM	f_{O2} ΔNNO	NBO/T
BAS1_Air	90	10	-2.29	5.3	4.65	0.74
BAS1_IW	13	87	-11.28	-3.5	-4.15	0.67
Du_+1	35	65	-4.55	1.58	1	0.92
BAS4_CO_2	75	25	-2.29	5.3	4.65	0.71
BAS4_IW	19	81	-11.28	-3.5	-4.15	0.69
BAS4_QFM	9	91	-7.6	0	-0.65	0.63
BAS4_Air	89	11	-2.29	5.3	4.65	0.62
BAA1_Air	93	7	-2.29	5.3	4.65	0.66
BAA1_QFM	8	92	-7.6	0	-0.65	0.60
BAA1_IW	13	87	-11.28	-3.5	-4.15	0.59
BAA2_Air	89	11	-2.29	5.3	4.65	0.73
BAA2_IW	17	83	-11.28	-3.5	-4.15	0.71
BAA2_QFM	14	86	-7.6	0	-0.65	0.63
BAA2_CO2	79	21	-2.29	5.3	4.65	0.63
BAA3_CO_2	81	19	-2.29	5.3	4.65	0.61
BAA3_Air	86	14	-2.29	5.3	4.65	0.60
BAA3_QFM	11	89	-7.6	0	-0.65	0.51
AND3_CO_2	78	22	-2.29	5.3	4.65	0.41
AND3_QFM	9	91	-7.6	0	-0.65	0.36
AND3_IW	12	88	-11.28	-3.5	-4.15	0.36
AND4_IW	10	90	-11.28	-3.5	-4.15	0.39
AND4_QFM	8	92	-7.6	0	-0.65	0.32
AND4_Air	100	0	-2.29	5.3	4.65	0.32

Table 2.1. Continued

Sample Name	%Fe^{3+}	%Fe^{2+}	Log(f_{O2})	f_{O2} ΔQFM	f_{O2} ΔNNO	NBO/T
AND5_CO_2	77	23	-2.29	5.3	4.65	0.34
AND5_IW	15	85	-11.28	-3.5	-4.15	0.33
AND5_QFM	8	92	-7.6	0	-0.65	0.29
AND5_Air	88	12	-2.29	5.3	4.65	0.29
AND2_Air	93	7	-2.29	5.3	4.65	0.72
AND2_IW	17	83	-11.28	-3.5	-4.15	0.71
AND2_CO_2	83	17	-2.29	5.3	4.65	0.66
AND2_QFM	13	87	-7.6	0	-0.65	0.66
G_1013B	70	30	-4.66	4.23	3.49	0.33
G_924	27	73	-8.78	0.64	0	0.30
AND1_Air	89	11	-2.29	5.3	4.65	0.21
AND1_QFM	18	82	-7.6	0	-0.65	0.17
DAC3_Air	100	0	-2.29	5.3	4.65	0.17
DAC3_Mo	14	86	-11.28	-3.5	-4.15	0.17
DAC3_CO_2	82	18	-2.29	5.3	4.65	0.14
DAC2_CO_2	85	15	-2.29	5.3	4.65	0.18
DAC2_Mo	11	89	-11.28	-3.5	-4.15	0.17
DAC2_Air	100	0	-2.29	5.3	4.65	0.15
G_1009	58	42	-6.03	4.19	3.45	0.04
DAC5_Air	90	10	-2.29	5.3	4.65	0.22
DAC5_Mo	8	92	-11.28	-3.5	-4.15	0.18
DAC6_CO_2	81	19	-2.29	5.3	4.65	0.15
DAC6_Mo	18	82	-11.28	-3.5	-4.15	0.13
DAC1_Mo	9	91	-11.28	-3.5	-4.15	0.10

Table 2.1. Continued

Sample Name	%Fe^{3+}	%Fe^{2+}	Log(f_{o2})	f_{o2} ΔQFM	f_{O2} ΔNNO	NBO/T
G_920	39	61	-10.22	0.64	0	0.18
DAC4_Air	100	0	-2.29	5.3	4.65	0.13
DAC4_CO_2	84	16	-2.29	5.3	4.65	0.12
DAC4_Mo	12	88	-11.28	-3.5	-4.15	0.10
DAC7_CO2	81	19	-2.29	5.3	4.65	0.08
RHY4_CO_2	71	29	-2.29	5.3	4.65	0.07
RHY4_Mo	12	88	-11.28	-3.5	-4.15	0.06
RHY6_Mo	18	82	-11.28	-3.5	-4.15	0.08
RHY7_Mo	16	84	-11.28	-3.5	-4.15	0.05
G_923	36	64	-10.22	0.64	0	0.07
RHY2_Mo	11	89	-11.28	-3.5	-4.15	0.05
RHY3_CO_2	71	29	-2.29	5.3	4.65	0.10
RHY3_Mo	16	84	-11.28	-3.5	-4.15	0.09

Table 2.2. Geochemical concentration data for 10 glass-forming oxide components for the Gardner and Dufresne glass samples (n = 9)

Sample	SiO_2[1]	TiO_2	Al_2O_3	FeO_T	MgO	CaO	Na_2O	K_2O	MnO	P_2O_5
Du_+3d	49.7	2.59	11.95	12.65	9.59	10.24	1.89	0.45	0.16	0.18
Du_+5	50.19	2.65	11.69	10.97	9.74	10.39	1.86	0.49	0.17	0.22
Du_+3	50.25	2.6	11.68	10.78	9.68	10.49	1.94	0.48	0.15	0.22
Du_+1	50.86	2.71	11.85	9.67	9.9	10.6	1.8	0.48	0.17	0.21
G_1013B	60.8	1.02	16.02	6.62	2.14	4.72	5.47	2.52	0.12	0
G_924	60.8	1.02	16.02	6.62	2.14	4.72	5.47	2.52	0.12	0
G_1009	67.25	0.26	16.59	2.32	0.13	0.24	7.98	4.56	0.08	0
G_920	71.42	0.4	8.47	7.93	0	0.34	6.7	4.36	0.34	0
G_923	76.14	0.29	10.24	4.35	0	0.29	5.19	3.43	0.07	0

[1]Oxides are reported in wt%

Table 2.3. LASSO model error metric and quality statistics for targeted Fe-redox and polymerization variables. RMSE metrics are reported in the same units as the variable, $\%Fe^{3+}$ and $\%Fe^{2+}$ in %, and Fe_2O_3 and FeO in wt%. NBO/T is unitless. The alpha (α) shrinkage parameter trained for the model, following LOO-CV recommendations, is listed for each variable. A difference metric is also provided and was calculated by averaging the combined differences found between the actual Mössbaur (Fe-redox variables) or hand-calculated (NBO/T) values, and the LASSO model predicted values

Variable	RMSE-C	RMSE-CV	RMSE-P	NRMSE-P	|Avg. Difference|	α
$\%Fe^{3+}$	32.7	36.9	34.2	0.35	32.74	1.47
$\%Fe2^{+}$	32.7	36.0	33.0	0.34	31.89	1.21
Fe_2O_3 (wt%)	3.4	3.7	3.8	0.35	3.20	0.14
FeO (wt%)	2.9	3.2	2.9	0.30	2.46	0.11
NBO/T	0.24	0.30	0.23	0.24	0.19	0.003

Table 2.4 Mössbaur/Calculated actual geochemical values vs LASSO model predicted values for the 24 test samples (n_{test} = 24), following application of the pertinent model (Act. = actual value; Pred. = model predicted value; Diff. = difference between actual and predicted values (+ or -)

Sample	$\%Fe^{3+}$			$\%Fe^{2+}$			Fe_2O_3 (wt%)			FeO (wt%)			NBO/T		
	Act.	**Pred.**	**Diff.**	**Act.**	**Pred.**	**Diff.**	**Act.**	**Pred.**	**Diff.**	**Act.**	**Pred.**	**Diff.**	**Act.**	**Pred.**	**Diff.**
Du_+3d	51	70	19	49	38	-11	7.2	4.7	-2	6.2	2.6	-4	1.03	1.0	0.0
G_1009	58	29	-29	42	23	-19	1.5	2.0	1	1.0	1.7	1	0.04	-0.3	-0.3
G_924	27	44	17	73	54	-19	2.0	3.6	2	4.8	3.7	-1	0.30	0.5	0.2
AND1_QFM	18	47	29	82	47	-35	1.0	3.8	3	3.9	3.2	-1	0.17	0.5	0.3
AND2_CO_2	83	66	-17	17	52	35	5.6	4.7	-1	1.0	3.6	3	0.66	0.5	-0.1
AND2_QFM	13	63	50	87	54	-33	0.9	4.7	4	5.3	3.8	-1	0.66	0.9	0.2
AND3_IW	12	58	46	88	51	-37	1.1	4.1	3	7.0	3.4	-4	0.36	0.4	0.1
AND4_Air	100	69	-31	0	39	39	7.6	4.5	-3	0.0	2.8	3	0.32	0.7	0.4
AND5_Air	88	46	-42	12	48	36	6.3	3.7	-3	0.8	3.4	3	0.29	0.5	0.2
BAA1_IW	13	42	29	87	53	-34	1.0	3.5	2	6.2	3.7	-3	0.59	0.5	-0.1
BAA2_CO_2	79	41	-38	21	52	31	10.1	3.5	-7	2.4	3.6	1	0.63	0.4	-0.2
BAA3_Air	86	60	-26	14	52	38	11.4	4.4	-7	1.7	3.6	2	0.60	0.6	0.0
BAA3_QFM	11	43	32	89	54	-35	1.5	3.6	2	10.6	3.7	-7	0.51	0.5	0.0
BAS2_Air	90	44	-46	10	26	16	10.5	3.1	-7	1.1	1.8	1	0.84	0.5	-0.4
BAS2_QFM	12	42	30	88	51	-37	1.4	3.6	2	9.3	3.6	-6	0.76	0.5	-0.3
BAS4_Air	89	48	-41	11	50	39	10.1	3.8	-6	1.1	3.5	2	0.62	0.6	0.0
BAS5_CO_2	78	44	-34	22	46	24	8.8	3.7	-5	2.2	3.2	1	0.67	0.4	-0.3
BAS6_CO_2	79	45	-34	21	53	32	9.9	3.7	-6	2.4	3.7	1	0.81	0.5	-0.3
BAS7_QFM	20	48	28	80	51	-29	2.3	3.9	2	8.2	3.6	-5	0.92	0.6	-0.3
DAC2_Air	100	46	-54	0	48	48	3.6	3.4	0	0.0	3.4	3	0.15	0.0	-0.1
DAC3_CO_2	82	49	-33	18	54	36	3.9	3.8	0	0.8	3.8	3	0.14	0.3	0.2

Table 2.4. Continued

Sample	$\%Fe^{3+}$			$\%Fe^{2+}$			Fe_2O_3 (wt%)			FeO (wt%)			NBO/T		
	Act.	Pred.	Diff.	Act.	Pred.	Diff.	Act.	Pred.	Diff.	Act.	Pred.	Diff.	Act.	Pred.	Diff.
DAC5_Mo	8	43	35	92	49	-43	0.5	3.5	3	4.9	3.4	-1	0.18	0.3	0.1
RHY3_Mo	16	39	23	84	54	-30	0.3	3.1	3	1.3	3.8	2	0.09	-0.1	-0.2
RHY6_Mo	18	41	23	82	52	-30	0.6	3.4	3	2.4	3.6	1	0.08	0.4	0.3

Table 2.5. PLS model error metric and quality statistics for targeted Fe-redox and polymerization variables. RMSE metrics are reported in the same units as the variable, $\%Fe^{3+}$ and $\%Fe^{2+}$ in %, and Fe_2O_3 and FeO in wt%. NBO/T is unitless. The PLS component (*p*) shrinkage parameter trained for the model, following LOO-CV recommendations, is listed for each model variable. A difference metric is also provided and was calculated by averaging the combined differences found between the actual Mössbaur (Fe-redox variables) or hand-calculated (NBO/T) values, and the PLS model predicted values

Variable	**RMSE-C**	**RMSE-CV**	**RMSE-P**	**NRMSE-P**	**|Avg. Difference|**	***q***
$\%Fe^{3+}$	27.9	33.9	31.6	0.32	28.98	3
$\%Fe2^{+}$	27.9	33.9	31.6	0.32	28.98	3
Fe_2O_3 (wt%)	3.1	3.5	3.4	0.31	2.88	2
FeO (wt%)	2.3	2.9	3.6	0.37	3.08	3
NBO/T	0.21	0.30	0.22	0.23	0.19	5

Table 2.6. Mössbaur/Calculated actual geochemical values vs PLS model predicted values for the 24 test samples (n_{test} = 24), following application of the pertinent model (Act. = actual value; Pred. = model predicted value; Diff. = difference between actual and predicted values (+ or -)

Sample	$\%Fe^{3+}$			$\%Fe^{2+}$			Fe_2O_3 (wt%)			FeO (wt%)			NBO/T		
	Act.	Pred.	Diff.	Act.	Pred.	Diff.	Act.	Pred.	Diff.	Act.	Pred.	Diff.	Act.	Pred.	Diff.
Du_+3d	51	64	13	49	36	-13	7.2	7.2	0.0	6.2	2.6	-3.6	1.03	0.79	-0.2
G_1009	58	90	32	42	10	-32	1.5	0.9	-0.6	1.0	-5.5	-6.5	0.04	0.12	0.1
G_924	27	31	4	73	69	-4	2.0	4.0	2.0	4.8	4.3	-0.5	0.30	0.46	0.2
AND1_QFM	18	69	51	82	31	-51	1.0	4.5	3.6	3.9	2.1	-1.8	0.17	0.48	0.3
AND2_CO_2	83	43	-40	17	57	40	5.6	4.2	-1.3	1.0	5.5	4.5	0.66	0.34	-0.3
AND2_QFM	13	28	15	87	72	-15	0.9	1.5	0.6	5.3	12.4	7.1	0.66	0.89	0.2
AND3_IW	12	44	32	88	56	-32	1.1	4.0	2.9	7.0	4.2	-2.8	0.36	0.28	-0.1
AND4_Air	100	40	-60	0	60	60	7.6	8.0	0.4	0.0	2.8	2.8	0.32	0.76	0.4
AND5_Air	88	57	-31	12	43	31	6.3	4.4	-1.9	0.8	3.0	2.2	0.29	0.54	0.3
BAA1_IW	13	39	26	87	61	-26	1.0	3.2	2.1	6.2	5.4	-0.8	0.59	0.58	0.0
BAA2_CO_2	79	44	-35	21	56	35	10.1	3.6	-6.4	2.4	3.1	0.7	0.63	0.43	-0.2
BAA3_Air	86	55	-31	14	45	31	11.4	5.3	-6.1	1.7	5.0	3.4	0.60	0.68	0.1
BAA3_QFM	11	38	27	89	62	-27	1.5	3.5	2.0	10.6	5.1	-5.5	0.51	0.56	0.0
BAS2_Air	90	115	25	10	-15	-25	10.5	7.7	-2.8	1.1	-4.1	-5.1	0.84	0.90	0.1
BAS2_QFM	12	41	29	88	59	-29	1.4	3.7	2.3	9.3	3.9	-5.3	0.76	0.53	-0.2

Table 2.6. Continued

Sample	%Fe^{3+}			%Fe^{2+}			Fe_2O_3 (wt%)			FeO (wt%)			NBO/T		
	Act.	Pred.	Diff.	Act.	Pred.		Act.	Pred.	Diff.	Act.	Pred.		Act.	Pred.	Diff.
BAS4_Air	89	58	-31	11	42	31	10.1	4.9	-5.2	1.1	4.2	3.0	0.62	0.67	0.0
BAS5_CO_2	78	58	-20	22	42	20	8.8	3.8	-5.0	2.2	2.8	0.6	0.67	0.49	-0.2
BAS6_CO_2	79	44	-35	21	56	35	9.9	3.7	-6.2	2.4	4.5	2.2	0.81	0.54	-0.3
BAS7_QFM	20	41	21	80	59	-21	2.3	4.1	1.9	8.2	5.1	-3.1	0.92	0.60	-0.3
DAC2_Air	100	75	-25	0	25	25	3.6	-1.6	-5.2	0.0	2.5	2.5	0.15	-0.04	-0.2
DAC3_CO_2	82	56	-26	18	44	26	3.9	0.9	-3.0	0.8	5.3	4.5	0.14	0.28	0.1
DAC5_Mo	8	60	52	92	40	-52	0.5	3.6	3.1	4.9	1.4	-3.5	0.18	0.32	0.1
RHY3_Mo	16	29	13	84	71	-13	0.3	-0.8	-1.1	1.3	2.9	1.6	0.09	0.03	-0.1
RHY6_Mo	18	39	21	82	61	-21	0.6	4.0	3.4	2.4	2.8	0.3	0.08	0.43	0.4

Figures

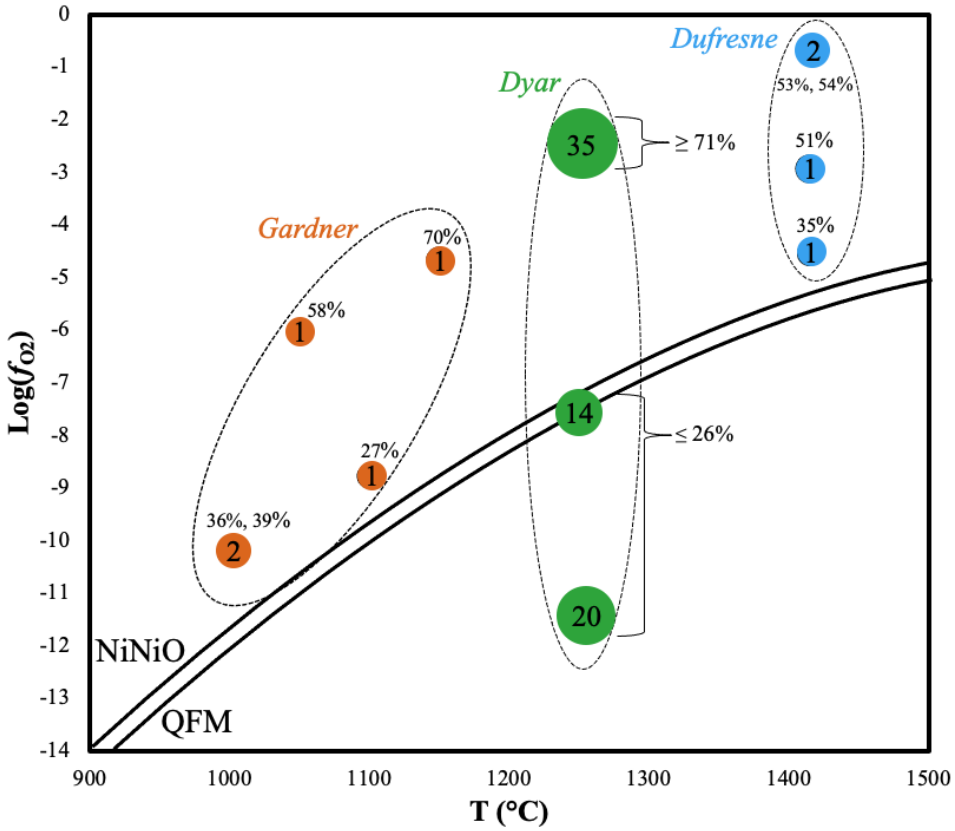

Figure 2.1. Log(f_{O2}) vs temperature diagram of 78 synthetic glass samples relative to one another as a function of their oxidation state and their positions to the QFM and NiNiO MRBs. Colored circles denote relative positions of the glasses in fugacity space, and are sized according to the number of glasses present at that plotting location (i.e., larger circles = more glasses). The number of glasses at each locality is provided inside the colored circles. Dyar glass samples (blue circles) spanned the largest fugacity range and the Gardner samples (multicolored circles) occupied a range in temperature and log(f_{O2}). The Dufresne samples (orange circles) had a smaller chemical spread. The 78 samples formed three distinct groupings (dashed ovals) according to temperature relationships. However, the spread of synthetizing conditions logged for the Gardner samples was wider. The Dufresne and Gardner glasses contained no samples below the QFM or NiNiO buffer lines

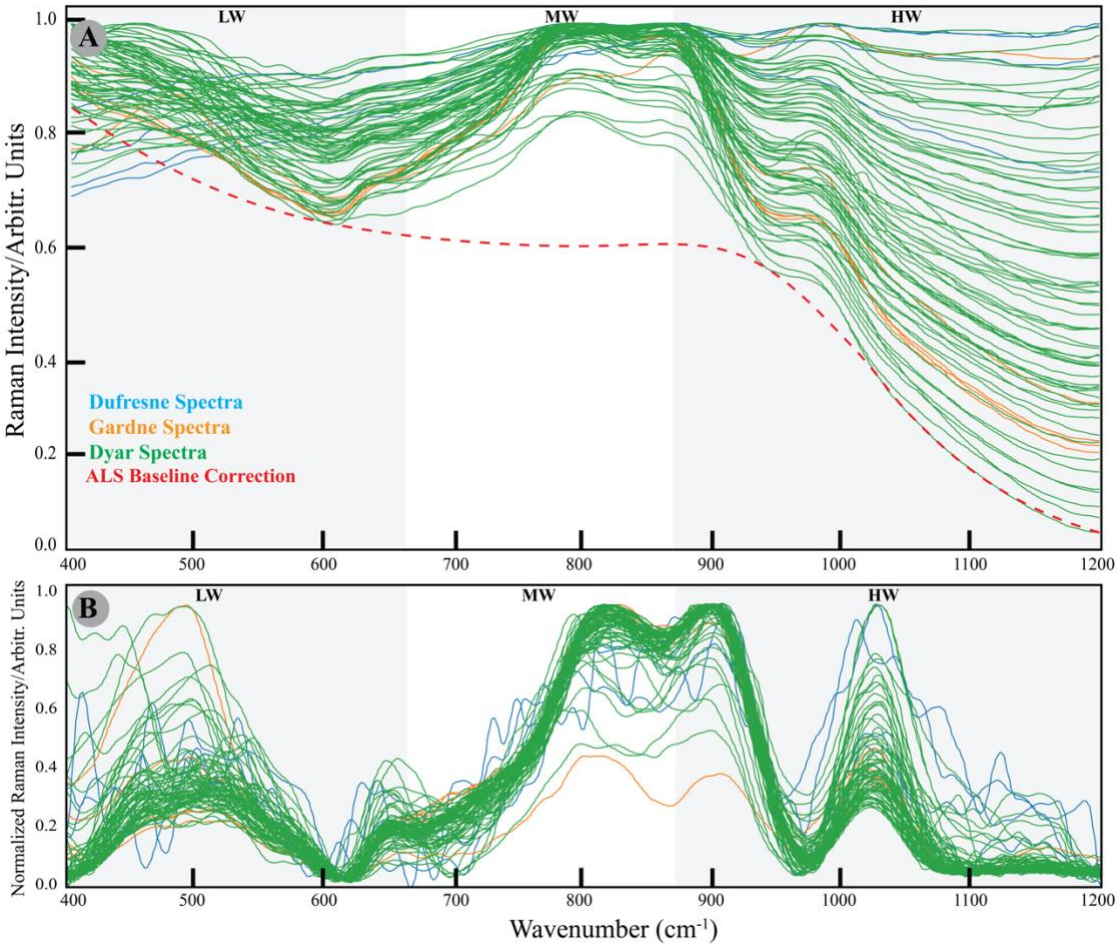

Figure 2.2. Raw (**A**) and processed (**B**) Raman spectra of the 78 synthetic silicate glass samples including the 69 Dyar (green), 4 Dufresne (blue), and 5 Gardner (orange) suites. Raw profiles were treated with a smoothing function and normalized to their maximum feature between 0–1.0 and baselined with an Asymmetric Least Squares (ALS) function with $\lambda=5.754e^{+4}$ and $p = 0.002$ to increase predictive accuracy and retain signal integrity. The processed spectra (**B**) exhibited the three major Raman-glass bands discussed in previous literature and occurring in LW, MW, and HW bands between ~400 – 1200 cm^{-1}

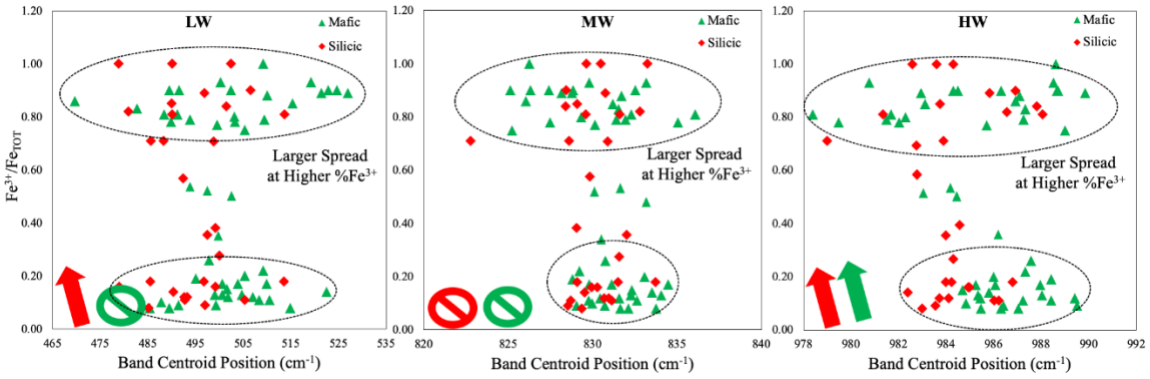

Figure 2.3. Gaussian fitted glass-band centroid positions for the LW, MW, and HW bands as a function of $\%Fe^{3+}$ along constrained rock types (mafic, green triangles; silicic, red diamonds). The LW and MW bands exhibited little to no shift in average wavenumber position, but the HW band showed decreasing position with increasing oxidation state

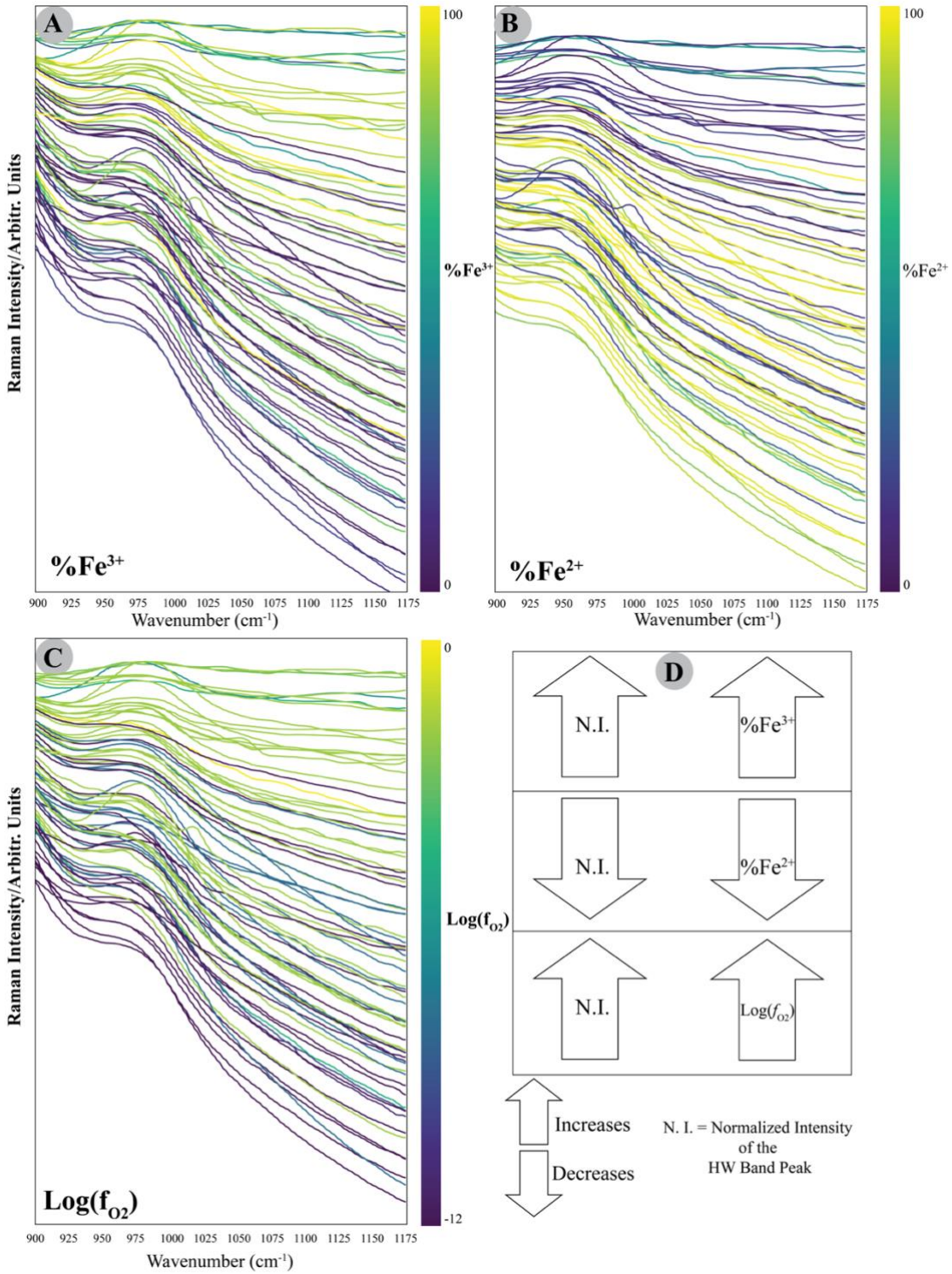

Figure 2.4. Normalized intensity variations of the "ferric-iron peak" near ~980 cm^{-1} as a function of Fe-redox and oxygen fugacity metrics, specifically $\%Fe^{3+}$(A) $\%Fe^{2+}$(B) and Log(f_{O2}) (C). With increasing oxidation state, peak intensity increased

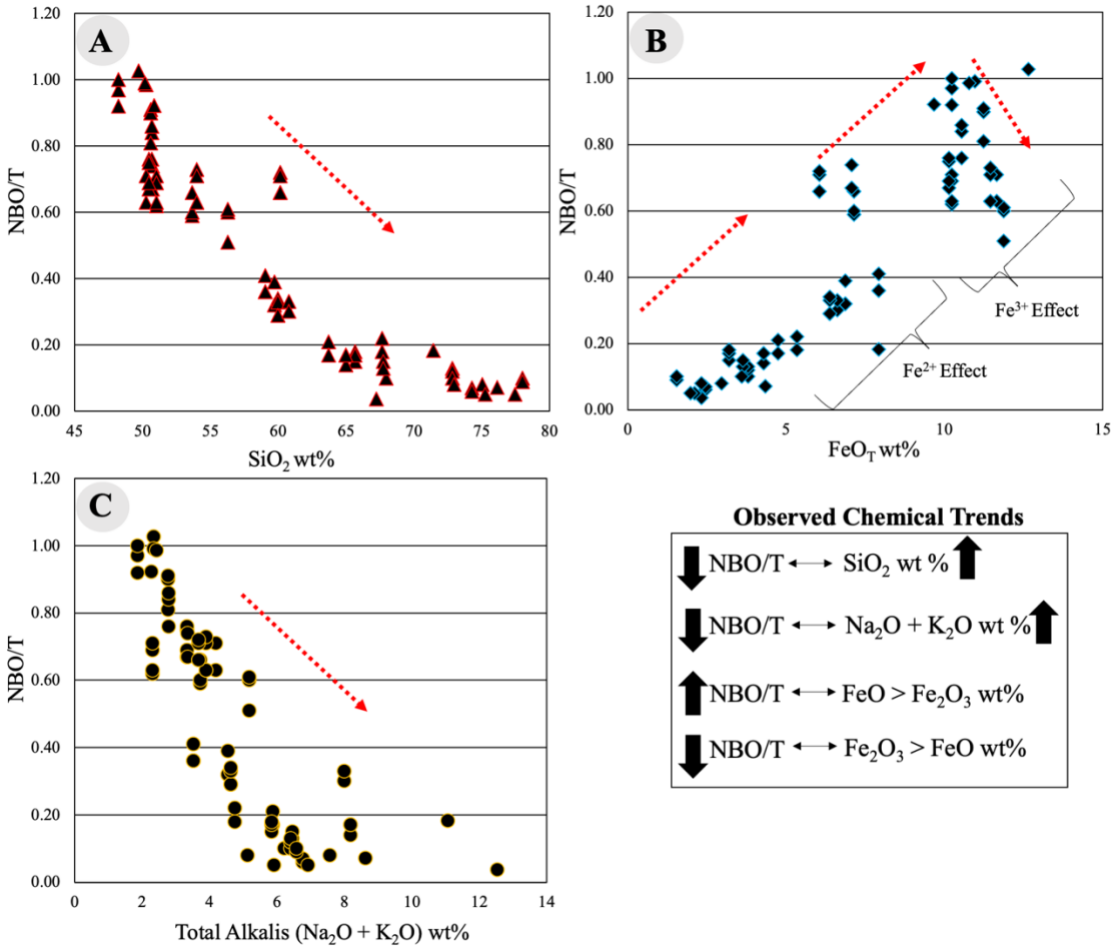

Figure 2.5. Change of calculated NBO/T parameter with glass chemistry: SiO_2 wt%; FeO_T wt%; total alkali content ($Na_2O + K_2O$) wt%

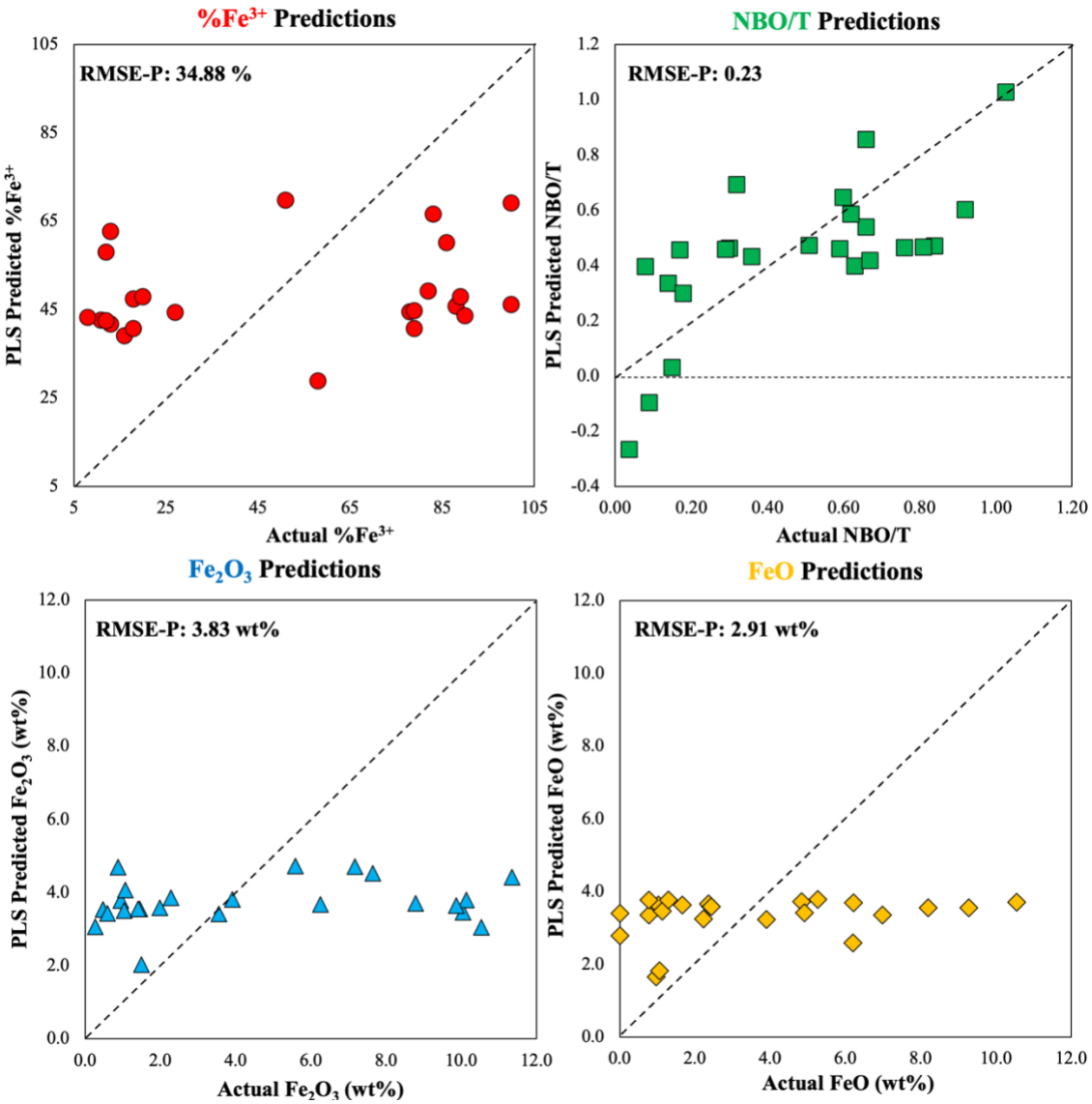

Figure 2.6. Predicted vs actual charts denoting Fe-redox variable values following the LASSO regression model application to n_{test}. RMSE-P errors are reported in comparable units. A 1:1 trend line is also provided (i.e., dashed black line)

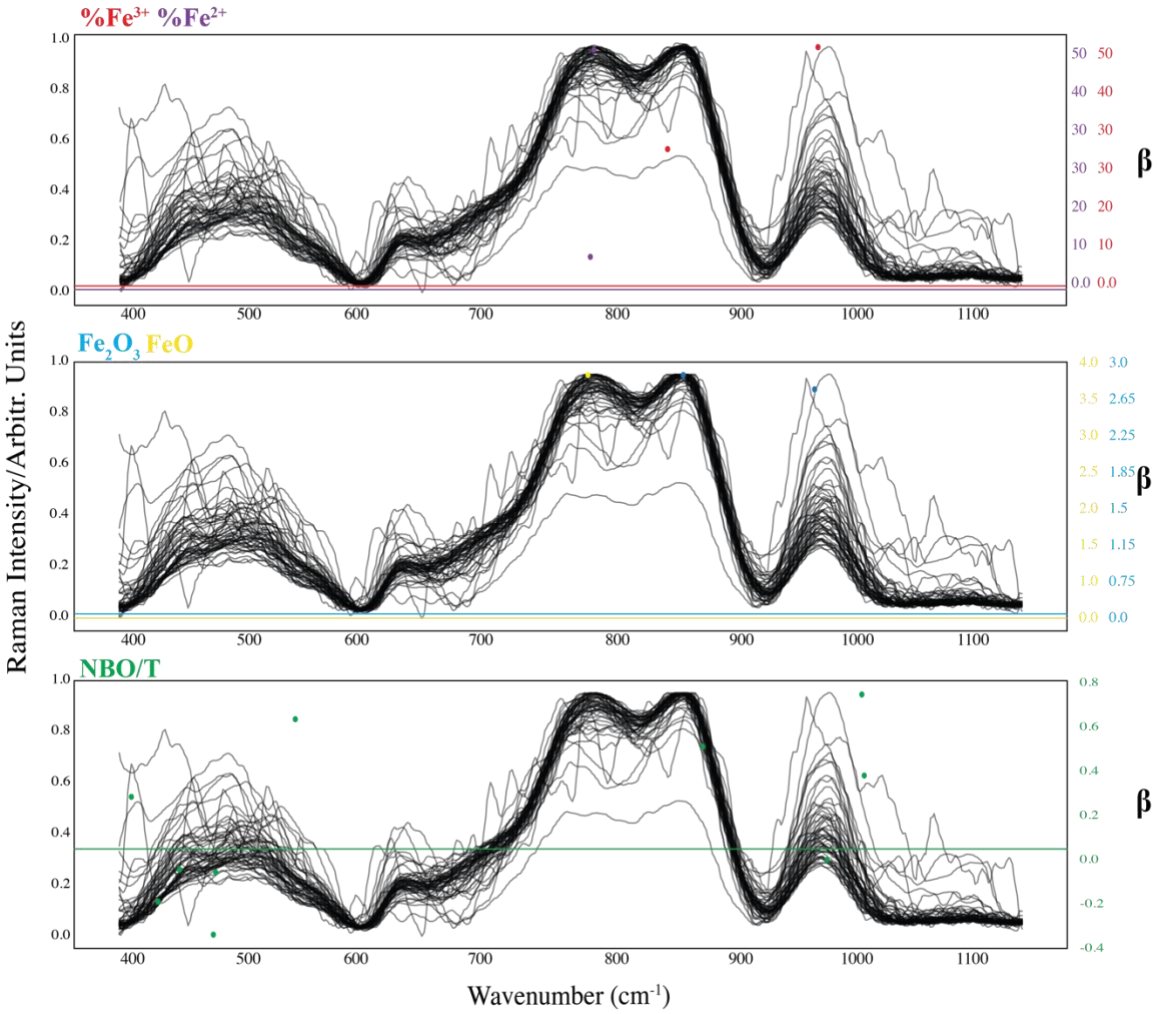

Figure 2.7. LASSO model coefficient maps for each of the five predicted Fe-redox variables. Coefficients (β) values are plotted against the training spectra in reference to a zero point (horizontal line). Nonzero values (above or below the line) indicated a positive or negative relationship between the predicted variable and that Raman spectral channel. Ten channels were selected for the prediction of NBO/T, which was a much higher number than used by the other variable regressions

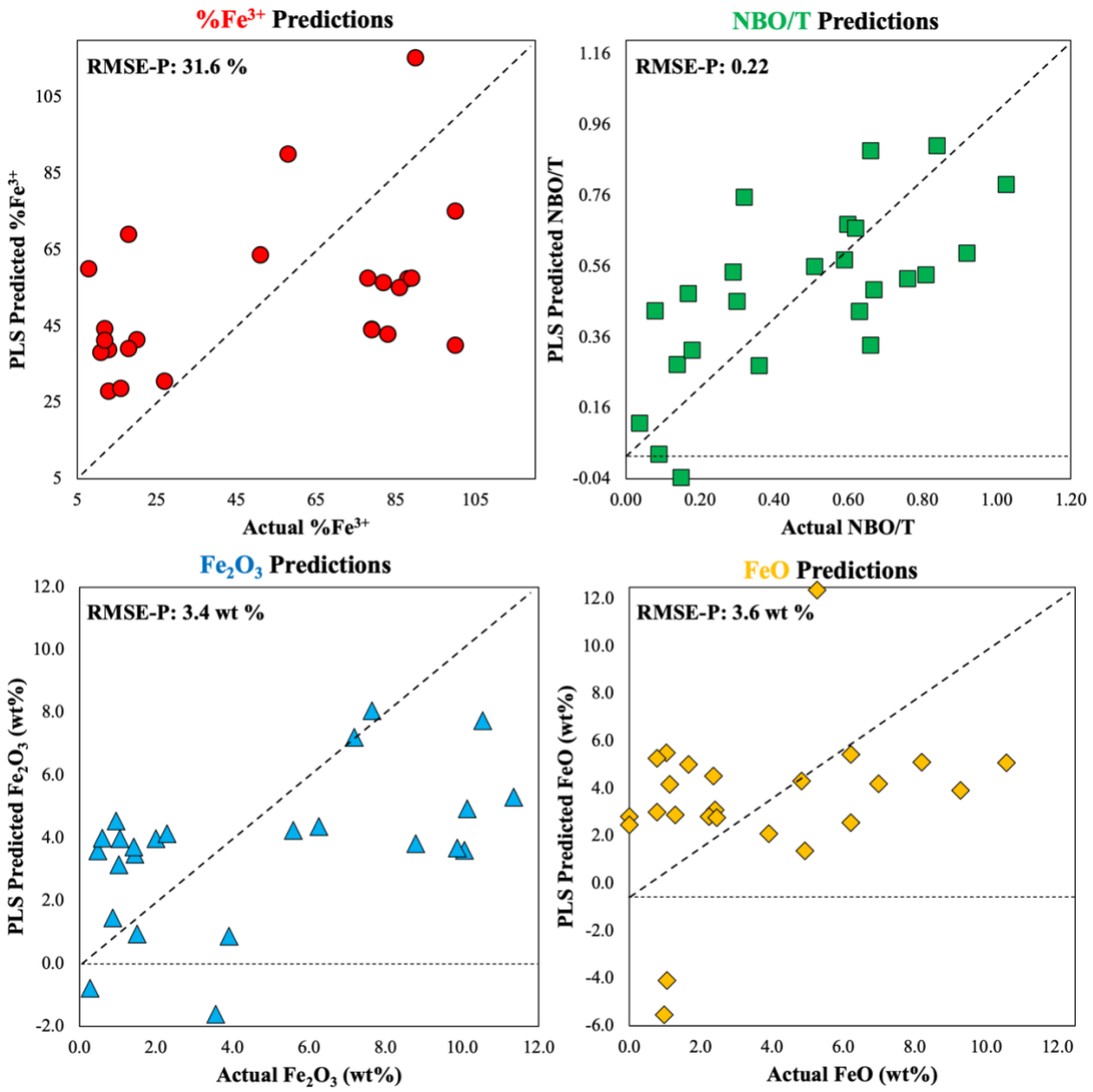

Figure 2.8. Predicted vs actual charts denoting Fe-redox variable values following the PLS regression model application to n_{test}. RMSE-P errors are reported in comparable units. A 1:1 trend line is also provided (i.e., dashed black line)

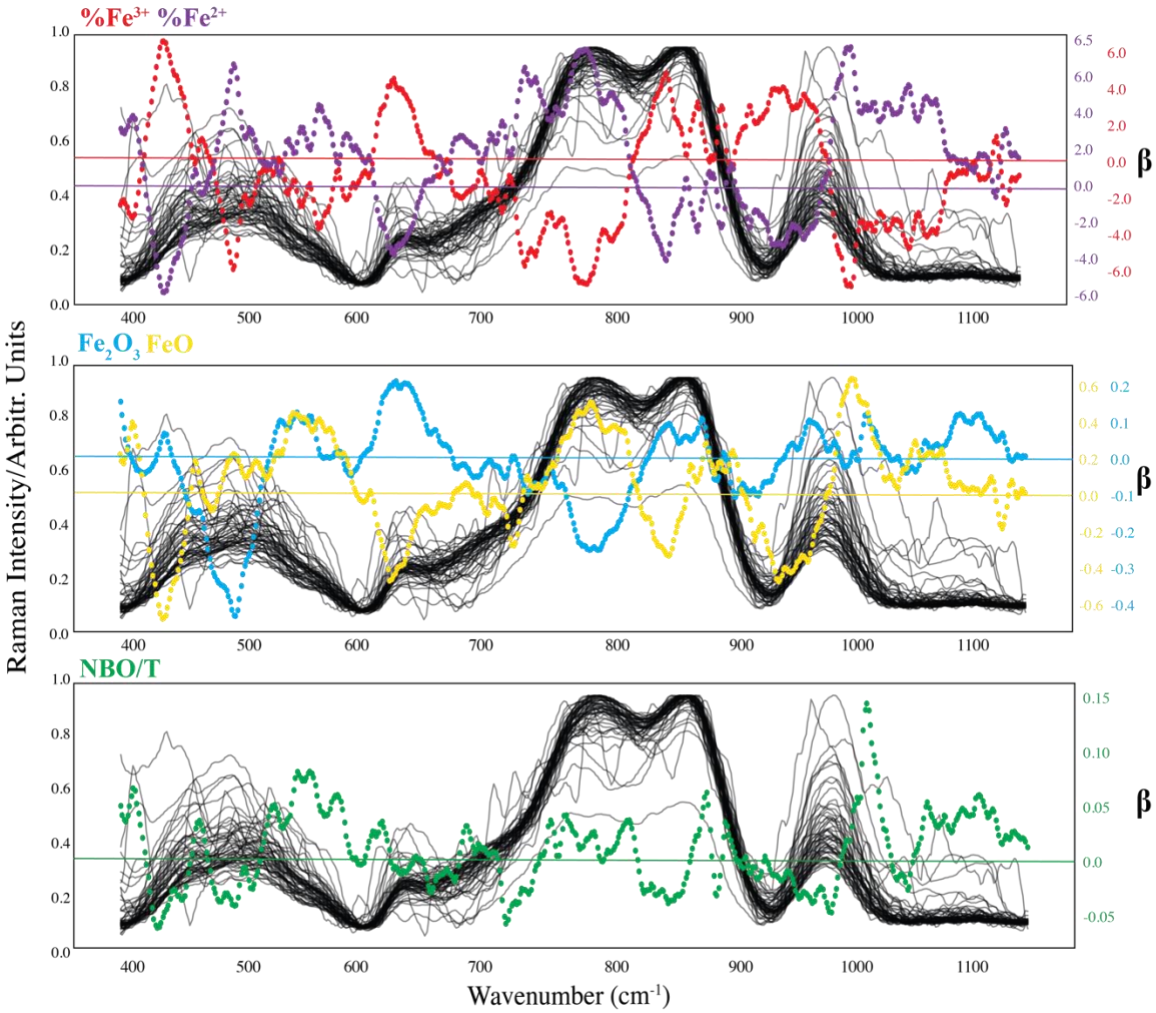

Figure 2.9. PLS-1 model coefficient maps for each of the five predicted Fe-redox variables. Coefficients (β) values are plotted against the training spectra in reference to a zero point (horizontal line). Nonzero values (above or below the line) indicated a positive or negative relationship between the predicted variable and that Raman spectral channel. PLS utilized every channel available and assigned a coefficient to each. This incorporated signals that were not necessarily important to the prediction of the targeted oxide, thereby noting the need for spectral parameterization. The $\%Fe^{3+}$ and $\%Fe^{2+}$ coefficients mirrored each other in the first frame, and indicated the influence of each valence state in the glass network, as picked by the regression function

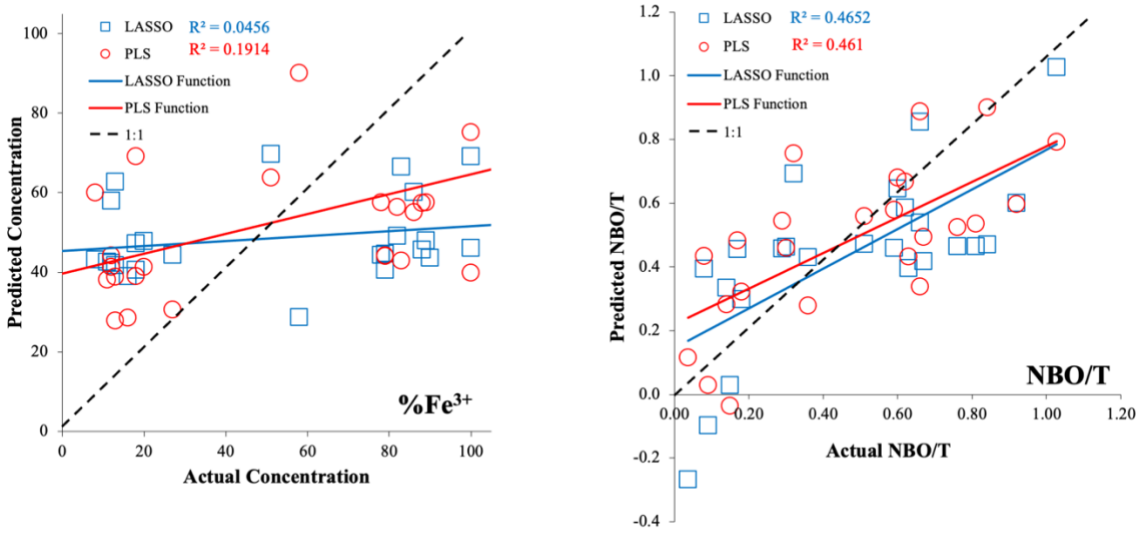

Figure 2.10. Predicted vs actual charts of $\%Fe^{3+}$ and NBO/T values for the 24 test glasses from the PLS-1 (red circles) and LASSO (blue squares) regressions. A 1:1 line is shown in black dashes to provide reference for the best-fit regression lines (red PLS) (blue LASSO)

CHAPTER III

RAMAN SPECTROSCOPY AND SUPERVISED MACHINE LEARNING STRATEGY FOR TEPHRA DISCRIMINATION IN MIXED GEOMATERIALS: A SUITABLE PAIRING FOR EXAMINING EXPLOSIVE ERUPTIVE PRODUCTS

Abstract

Here, volcaniclastic and hemiplegic marine sediment material from the region surrounding the Caribbean island of Montserrat, are combined in varying ratios to obtain 13 mixed-phase, sample targets for dual-beam Raman spectroscopy. Raman data is collected for each sample mixture in an effort to expand the use and capability of Raman techniques in identifying tephra in mixed natural samples. This work highlights the use of Raman spectroscopy to accurately identify tephra concentration as a function of Raman signal intensity in the spectral region between 300 – 1200 cm^{-1} and constitutes a preliminary spectral library of such samples for future mixing investigations. Tephra concentrations can be reliably predicted using a PLS regression strategy with a root-mean-square error of calibration (RMSE-C) of ~8% and an error of cross validation (RMSE-CV) of ~15%. The LASSO calibration cannot reliably predict such variables. Optimal modeling parameters and coefficients were obtained using *k*-fold cross validation, leading to internal R^2 metrics of >0.90 for both techniques. PLS coefficients across all channels map Raman spectral features present in a purely tephra sample (i.e., 100% tephra), while signals inherent to the pure sediment sample are ignored by the regression. Individual mineral species such as calcite can be identified in these mixed samples, and amorphous silicates (volcanic and biogenic) also contribute to the bulk Raman signal. Definitive band shape shifts are observed for the spectra at distinct mixing ratio boundaries and indicate dampening influence of amorphous phases over their crystalline counterparts down to mixes containing only 15% tephra.

Introduction

Pure silicate glass is rare in most natural eruptive products. Most volcanic materials are mixtures of glass and crystalline components. In addition, unless collected or analyzed in the immediate aftermath of an eruptive event, the overwhelming majority of such volcanics are heavily altered (McCloy, 2019). Tephra is a complicated mixture of rock debris, glass, ash, and mineral fragments, and is found on nearly every terrestrial planet (Wilhelms, 1987; Greeley and Crown, 1990; Robinson et al., 1993; Wilson and Head, 1994; Greeley et al., 2000; Rampe et al., 2014; Byrne et al., 2018). As the primary byproduct of explosive eruptions, tephra is the main target for primary melt characterization in many explosive environments (Carter et al., 1995; Lowe, 2011; Cassidy et al., 2014).

Over the last half century, the identification and discrimination of tephra has been achieved with great success using a host of spectroscopic techniques, both *in-situ* and from orbit on Earth and Mars; with infrared (IR) methodologies dominating: (NIR, VSWIR, VNIR, and MIR) (Bell et al., 1993; Bishop et al., 1998; McCanta et al., 2015; Hook et al., 1999). These IR-based approaches have recently been paired with machine learning methodologies (such as Partial Least Squares regression: PLS), opening a new avenue of investigation for tephra characterization (Leight, 2020). However, Raman spectroscopic based methods for such discriminations is limited and has mainly been used to identify mineral components in tephra (White, 2009; Barletta, 2012; Bathgate et al., 2015; Surtees et al., 2016). While such Raman spectroscopic experiments have been promising in identifying individual or grouped mineral signals from a known tephra, the ability to identify the presence of tephra in a mixed tephra-sediment sample has not been demonstrated. With Raman spectrometer technologies currently deployed on Mars, where mixed origin materials are likely dominant, the ability to identify tephra and determine their compositions is of high importance.

In this work, we determined the effect of sediment dilution of tephra samples on the resulting Raman signature. Raman spectra were collected on both pure-phase samples (i.e., tephra only or sediment only) and on intermediary mixed samples, with pre-defined mixing ratios (see Table 3.1). The Raman data was then used to build and internally train

quantitative machine learning models using PLS and LASSO regressions, to determine at which concentration end member discrimination became impossible.

Methods

Starting materials

A collection of terrestrial volcaniclastic (tephra) material was originally collected from a coherent tephra unit in 2012 approximately 5 km downwind from the Soufrière Hills stratovolcano, located on the Caribbean Island of Montserrat (McCanta et al., 2015). Hemipelagic sedimentary material was extracted from a marine drill core collected during the Integrated Ocean Drilling Program (IODP) expedition 340 in March 2012, off the western coast of Montserrat. The marine sediment was taken from a visibly pristine layer of drill core U1396-C (section U1396-C-3H-4A), collected at 786 m water depth ~ 36 km southwest of the volcanic complex. Core U1396-C (16°30.47'N, 62°37.09'W) represented the northwestern most site drilled during the expedition, was drilled ~15 – 20 km closer to the island of Montserrat and Soufrière Hills than previous expeditions, and was comprised of a series of sediment, tephra, and volcaniclastic sand (Le Friant et al., 2013). The material collected from the expedition has been the subject of a major investigation concerned with the geologic characterization of the regions' eruptive history, particularly as it relates to the most recent volcanic activity from Soufrière Hills (McCanta et al., 2015).

The tephra and hemipelagic sediment materials were both disaggregated with mortar and pestle following oven-drying over a period of 24 h at 100°C. A mesh sieve was used to concentrate <74 μm-sized fractions of both materials, and the resulting sample portions were weighed and mixed following the proportions listed in Table 3.1. This process created 13 individual tephra:sediment mixtures with tephra concentrations containing: 100, 95, 90, 80, 75, 50, 25, 20, 15, 10, 5, 2, and 0% tephra. Samples are labeled as 'MC-#' between 1 and 13: with sample MC-1 containing 500 mg sediment and 0 mg tephra, and MC-13 containing 0 mg sediment and 500 mg tephra (Table 3.1). Sample MC-8 represents an even 50:50 by weight split of tephra:sediment.

Sample suite chemistry and major mineralogy

Mineralogic and chemical investigations of the sample materials indicate the volcaniclastic tephra is of a dark grey color, and primarily andesitic in composition (Stinton et al., 2014). The hemipelagic marine sediment is of a consistent bright white color and largely composed of calcium carbonate ($CaCO_3$) (Le Friant et al., 2013). The andesitic tephra contained dome rock fragments, vesicular material (amorphous volcanic glass), minor Fe-Ti oxides (magnetite & ilmenite), plagioclase, and amphibole (hornblende); while the sediment was primarily pure-calcite and aragonite in various portions, with a minor amount of clay silicates, plagioclase, and amorphous material (glass and biogenic silica) (Stinton et al., 2014). The sediment also contained a portion of volcanic products comprising ~ 35% of the background materials (McCanta et al., 2015) due to extensive volcanism in the near environment. Analyses on other cores containing similar material from the U1396 site agree with these descriptions (La Friant et al., 2013; Cassidy et al., 2014).

Raman spectra acquisition and preprocessing

Sample powder separates (MC-1 – MC-13) were placed into portable glass vials, placed on a collection stage, and subjected to Raman spectroscopic analysis in the Astronomy department at Mt. Holyoke College in South Hadley, MA using a DuoLaser^{TM} BRAVO handheld Raman spectrometer. The spectrometer had an operating spectral resolution of 10 – 12 cm^{-1} and a spectral range of 300 – 3200 cm^{-1} using a dual incident beam system. Raman spectra were collected in this range for the powders individually, from an operating distance of < 1mm using the 785 and 852 nm excitation laser beam settings. The use of the dual system helps increase collection coverage, reduces fluorescence automatically, and provides overlapping data streams for preliminary matching. The Raman region between 300 – 2100 cm^{-1} was obtained via the 785 nm beam, and the region from 1200 – 3200 cm^{-1} was collected using the 852 nm beam. Spectra were collected over 6 channels, with two minor shifts in laser position each, resulting in an individual frame-averaged spectrum for each sample (n = 13). No unique Raman activity was observed beyond ~ 1200 cm^{-1}, therefore the region beyond was truncated. The 13 raw spectra between 300 – 1200 cm^{-1} were corrected for excitation

noise, temperature, and cosmic ray flux using a Median Filter (MFilter) based on a 2:9 Savitzky-Golay smoothing function and were imported into the Data Exploration, Visualization and Analysis for Spectroscopy (DEVAS) system for preprocessing and model building (http://nemo.mtholyoke.edu/).

As was the case with preprocessing silicate glass Raman data, no standardized processing strategy exists for baselining or normalizing tephra, sediment, or tephra:sediment mixes. Selections for an appropriate baseline and normalization schema was determined initially through visual inspection of their effects on the Raman data, to avoid overcorrection. Boundary conditions of acceptable approaches were identified through iterative testing. Then, the best possible correction parameters were established by choosing the set of preprocessing options that yielded the highest predictive accuracies within those boundaries (i.e., Liland et al., 2010). Following these preprocessing testing scenarios in DEVAS, it was determined that an Asymmetric Least Squares (ALS) baseline correction (Eilers and Boelen, 2005) with an asymmetry (p) of 0.004 and a parametrized smoothness factor (λ) of $1.585e^{+3}$ was the optimal subtraction selection for these Raman data. Similarly, a maximum feature normalization scheme ($Norm_{Max}$), where the most intense Raman features are used to stratify the spectrum intensities between 0 – 1.0, was determined to be the best for these samples (Figure 3.1, Figure 3.2).

Building PLS and LASSO models for % tephra predictions

Univariate PLS-1 and LASSO prediction models, in which a single variable is predicted, were constructed using the 13 preprocessed MC Raman spectra within the DEVAS system, with tephra concentration as the target variable. Due to the small sample suite size (n = 13), the data was not suitable for breakdowns into a training suite (n_{train}) and testing suite (n_{test}), and therefore only internal modeling was possible (Clegg et al., 2009; Dyar and Ystma, 2021). Therefore, these constructed models serve as a proving ground for future incorporations of similar data streams, with eventual applications to external validation suites.

PLS and LASSO regression algorithms were trained to predict % tephra from the 13 preprocessed Raman spectra using a *k*-fold cross validation technique, in which the shrinkage constrains placed on the algorithm (i.e., number of PLS components (q) or the

a value in LASSO) were determined by the choosing the value of *k* in the Regression Training segment in the DEVAS system. The system implements cross validation by splitting the dataset into *k* equal-sized groups, with $k - 1$ groups used to internally train the regression, while the remaining groups measure error via internal predictions (Ytsma and Dyar, 2019). Typically, *k* is set equal to the square root of the training sample set (n = 13) in situations where no external tests are performed, so in this case, *k* was set equal to 3 (Hastie et al., 2017; Hastie et al., 2009). The errors from each internal *k*-fold model are averaged together in the algorithms to produce root-mean-squared error of cross validation (RMSE-CV) (Stone and Brooks, 1990; Hastie et al., 2017). RMSE-CV in these circumstances (where data suites are limited in scale), serves as the best measure of model effectiveness to future applications on external datasets (Dyar & Ytsma, 2021). Both PLS and LASSO models were cross validation with $k = 3$ to produce the recommended optimal *q* and α values, and these values were used to train the algorithm to predict % tephra.

Results

Evolution of Tephra Raman signals as a function of sediment mixing ratios

The collected Raman spectra for the 13 samples, including the pure tephra, pure sediment, and 11 intermediate mixes, yield both consistent, and unique features depending on the ratio of tephra to sediment observed. The pure tephra spectrum (100% tephra: MC-13) is comprised of low intensity, asymmetric bands occurring between 300 – 520 cm^{-1}, 550 – 750 cm^{-1}, and 850 – 1100 cm^{-1}. The sample does not contain any singular narrow peaks, but does have band centroids near ~ 330, 410, 510, 660, and 950 cm^{-1} based on batch Gaussian peak fitting (Figure 3.2). These feature descriptions are in stark contrast to the Raman spectrum of the pure sediment sample (i.e. 100% sediment: MC-1). This sample spectra exhibits four well defined, high-intensity peaks, centered across the collection region at ~ 450, 635, 880, and 1120 cm^{-1}. A separate minor peak is observed near ~355 cm^{-1} for this sample spectrum as well (Figure 3.1).

The intermediate mixes (i.e., MC-2 → MC-12) show a mix of bands that fall along discrete mixing ratio boundaries. A noticeable disparity in the overall Raman spectrum shape is observed at the 50:50, and 15:85 mixing boundaries (tephra:sediment),

in which samples with > 50% tephra are characterized by broad asymmetric bands in the low (LW), and high (HW) wavenumber regions: LW: 300 – 750 cm^{-1}, HW: 750 – 1200 cm^{-1}, and those with < 50% tephra exhibit two to four narrow, well-defined peaks across these regions (Figure 3.1). Specifically, in the LW region, this disparity arises at the 50:50 boundary, where the broad triplet band of the tephra-rich spectra switches to become two intense, narrow peaks, centered near ~450 cm^{-1} and 635 cm^{-1}, with a minor feature near ~355 cm^{-1} in the sediment-rich samples (Figure 3.1, Figure 3.2). In the HW region, this transition is observed to occur at much lower concentrations of tephra, in which a single broad band transitions into two strong peaks near 880 and 1120 cm^{-1} at the 15:85 boundary (tephra:sediment). Along with these shifts in overall shape, the persistence of a sharp, low-intensity peak at ~1085 cm^{-1} is observed in all sample spectra between the 95% tephra sample - 15% tephra sample. The peak is not present in the pure tephra sample and is not recognized in the samples containing ≤ 15% tephra. A slight positive shift in band center position for the ~635 cm^{-1} feature is observed in the low tephra spectra (Figure 3.2).

Prediction accuracies of the PLS and LASSO training models for tephra concentration

k-fold cross validation recommended $q = 4$ components for the PLS regression calibration and the prediction model yielded an internal error of RMSE-C = ~ 8 %, a cross validation error of RMSE-CV = ~15.5%, and a $R^2 = 0.96$ (RMSE metrics are recorded in the same units as the target variable, in this case, % tephra) (Table 3.2; Figure 3.4). An α value of ~0.79 was recommended for the LASSO regression following cross validation, with the final internal predictions yielding an RMSE-C = ~10.2 %, RMSE-CV = ~41.4 %, and $R^2 = 0.93$ (Table 3.2; Figure 3.4). As these models represent initial training calibrations, the reporting of determination coefficients (R^2) is somewhat acceptable: though it is not as powerful a quality metric as RMSE (Dyar and Ytsma, 2021). Still, the value helps to highlight the closeness of fit to a 1:1 line for the calibrations (Figure 3.4). The tephra concentration PLS model consistently predicts 27.5% tephra for samples MC-5 – MC-8 (Table 3.3), and LASSO models similarly quantify 22.9 % tephra for the same samples. These predicted values are in close

alignment with the actual concentrations recorded for samples MC-6 (20%) and MC-7 (25%), but do not align well with samples MC-5 (15%) and MC-8 (50%). Predictions of % tephra in sample MC-8 show the highest difference errors, with both methods underpredicting the concentrations by > 50%.

Correlations between tephra:sediment spectral bands and sample components via PLS and LASSO coefficients

PLS and LASSO model coefficients (β) are provided in Figure 3.5. These coefficients highlight the weighted importance of a Raman spectral feature to the prediction of the targeted variable. In this case, tephra concentration. Coefficients are assigned by PLS to every spectral channel (p) along the collection region (in this case p=346), whereas LASSO regression only assigns coefficients to Raman channels deemed to be the most important to predicting tephra concentration. Therefore, the calibrations produces 346 PLS coefficients, and 3 LASSO coefficients (Figure 3.5). Coefficient values > 0 represent a positive correlation and relationship between that Raman channel and the concentration of tephra in the sample, whereas values < 0 represent the inverse relation. Values near 0 indicate no relation. The magnitude of the coefficient's value is also of major importance, as it indicates how strong that positive or negative correlation is.

Here, PLS coefficients indicate relatively consistent relationships between the Raman bands in the regions between ~ 300 – 1000 cm^{-1}, with most values falling slightly above or slightly below zero. The coefficients are in good agreement with the regressions' ability to assign weighted importance to the features present in the higher tephra concentration samples (MC-13 – MC-9), rather than attributing strong positive correlations to those inherent to the sediment-heavy samples. They are not highly scattered and match the peaks and valleys of the pure tephra sample well (Figure 3.5). PLS coefficients exhibit a strong negative correlation to the channel near ~ 1085 cm^{-1}. This Raman feature was ascribed to $CaCO_3$, indicating the model's ability to identify Raman features not present in the pure-tephra sample (Figure 3.2).

LASSO coefficients (n = 3) were assigned to channels occurring at ~ 410, 660, and 880 cm^{-1}. The lower positioned coefficients are weighted slightly positive and

slightly negative respectively (Figure 3.5), while the 880 cm^{-1} channel is assigned a strongly negative value near ~ -35.00. When compared to the PLS coefficients, it is evident the LASSO assignment of ~ -35.00 to this channel is more than likely incorrect – leading to the predictions under this strategy fairing worse than that of PLS.

Discussion

Raman signal component discrimination in mixed tephra:sediment materials

The noticeable disparity in overall spectrum shape (from ill-defined asymmetric bands in the tephra-rich samples, to well-defined coherent peaks in the sediment-rich signals) beginning at the 50:50, and 15:85 tephra:sediment ratios is likely due the influence of the large glassy component in the tephra-rich samples. As observed in Chapter 1, the presence of glassy material results in a dampened Raman spectral signal. This diluting effect from a large amorphous content arises from a decrease in the long-range structural order and bonding geometries in the sample mixture that is measured by the Raman analysis (when compared to the sediment-rich samples) (McMillian, 1984; Rossano and Mysen, 2012). With much lower levels of crystallinity, the tephra-rich sample spectra retain these characterizations down to ~50% in the LW region, and to 15% in the HW. The persistence of this effect to much lower tephra concentrations in the HW region of the spectrum is not well understood but may correlate with the Fe-oxidation state of the glassy components that constitute these samples (such effects in a comparable IR region as would be associated in Raman data has been observed for these samples previously) (McCanta et al., 2015). However, Raman spectroscopy cannot be used to directly infer such information from these data.

The narrow peak near ~1085 cm^{-1} present in samples MC-5 – MC-12 is a direct match to pure $CaCO_3$-mineralogies, based on mineralogical Raman comparisons to the RRUFF database, and previous mineralogical descriptions (Figure 3.3), however distinguishing between the polymorphs of calcite or aragonite is not possible here due to the lack of Raman data below 300 cm^{-1} where such distinctions could be made (RRUFF, Pierre et al., 2014). This indicates that calcium carbonate is detected with Raman spectroscopy when even minor amounts (≤ 5%) of the sedimentary material is present. Though the remaining mineralogy is known for the sedimentary and volcaniclastic

material, and preliminary assignments could be made for their matching RRUFF peaks (Figure 3.3), we cannot definitively identify other individual mineral phases beyond $CaCO_3$ in this work.

Overall, the major spectral features recorded by Raman spectroscopy in the Monserrat sample mixes are dominated by complicated contributions from their constituent crystalline and amorphous phases, with the major Raman band envelopes strongly affected by the amorphous component. A pure tephra spectrum is readily distinguishable in the data from a sample bearing pure sediment. However, there are distinct consistencies between the pure tephra signal and those samples mixed with minor amounts of sedimentary material. Spectra from samples containing 50% tephra still retain the same low-intensity band features in the LW region found in the pure tephra sample profile; and the same consistencies are found in the HW region band features in samples with concentrations down to even 15%. These boundaries affected the ability of PLS and LASSO models to accurately predict tephra concentrations at these cut-offs (Table 1.3). Primary features in the sample spectra likely arise from plagioclase, hornblende, and calcium carbonate, while the glassy component in the mixes diluted signal intensities and peak strength until overtaken by stronger vibrational recordings in higher crystalline mixes.

Contextualization and comparisons of the PLS and LASSO models for tephra concentration

Overall, the PLS-1 model is a stronger predictor of tephra concentrations for the 13 sample mixtures than the one produced using LASSO (Table 3.2). Internal RMSE-C metrics for the two strategies are close, with LASSO records fairing slightly worse when the regression is asked to predict % tephra from the same 13 samples it was trained to. RMSE-CV metrics however, a true proxy for how well the linear regressions would fare on external unseen data, vary greatly between the PLS and LASSO models. The LASSO approach's RMSE-CV is more than double the PLS calibration. Both strategies exhibit interesting consistencies in predicting tephra concentrations in the intermediate mixes, beginning with the 15:85 sample, and ending with the 50:50 sample (Table 3.3). These samples represent the Raman spectrum shape disparity boundaries, where the LW and

HW band regions of the samples transitioned from consistent shapes in tephra-rich mixtures, to having both diluted and narrow peak structures (Figure 3.2). The models are seemingly able to predict % tephra variables from mixed samples whose Raman spectra were more consistent with one another (i.e., MC-13 – MC-9, or MC-4 – MC-1), while struggling to pull such information from those Raman spectra that exhibited a mix of well-defined peak features and more asymmetrical bands (i.e., MC-8 – MC-5). Currently, no such quantitative investigation using Raman-tephra data exists, and this work stands alone in the attempt. Previous work has focused on identifying individual mineral components or other crystalline Raman signals in collections from tephra deposits, as a means to discriminate between the mineralogizes of different tephra deposits rather than identifying the tephra deposits themselves with Raman spectroscopy. Therefore, no adequate comparisons can be made to the quantitative results in this work.

Conclusions

This work expands the use of Raman spectroscopy for discrimination of tephra in natural mixtures. By pairing PLS-1 and LASSO regression strategies to the 300 – 1200 cm^{-1} Raman region for 13 mixed samples collected from the region surrounding the Soufriere Hills Volcano, tephra concentration can be predicted with errors of ~15% (PLS) or 41% (LASSO). The PLS model is significantly more accurate when compared to the LASSO calibration and is the recommended method for such applications to uncharacterized datasets. It is shown here that the quantitative methodologies struggle with Raman spectral data in which distinct disparities in Raman peak shape exist between the major LW and HW envelopes but do well for endmember profiles whose spectra exhibited consistent band definitions. LASSO utilizes only three Raman channels for the prediction of % tephra in these data, likely excluding key features essential to identifying volcaniclastic material, and lowering predictive power (Figure 3.5). PLS coefficients are mapped across every channel and show strong negative correlations to the calcium carbonate peak near ~1080 cm^{-1}. As this feature was not recorded in the pure tephra sample MC-13, it is clear the regression algorithm is operating under the correct chemical considerations supported by mineral matching.

Raman profiles of tephra-heavy samples are easily discerned from the sediment-rich samples up to the point of 50%:50% mixing where peak clarity transitions occur in the LW region, and up to the 15%:85% mixing boundary in the HW's. Tephra rich spectra are composed of broad, low intensity, asymmetric bands while the more crystalline sediment spectra are comprised by high intensity narrow peaks along defined centroids. The major disparities between the materials are likely due to the stronger effect of amorphous material in the volcaniclastic-rich samples – which introduces a dampening in similar collections of Raman spectra. When crystallinity is increased among the mixes, this effect is overcome, leading to more defined Raman peaks. This work represents a first of its kind approach to quantifying tephra in scenarios where collections of Raman spectra are the sole source of information, yielding valuable insights into discriminations from such equipment in future work.

References

Barletta, R.E. (2012) "Raman analysis of Blue Ice Tephra: An approach to tephrachronological dating of Ice Cores," *Antarctic Science*, 24(2), pp. 202–208. Available at: https://doi.org/10.1017/s0954102011000885.

Bathgate, E.J. *et al.* (2015) "Raman, FTIR and XRD Study of Icelandic tephra minerals: Implications for Mars," *Journal of Raman Spectroscopy*, 46(10), pp. 846–855. Available at: https://doi.org/10.1002/jrs.4694.

Bart, J.C. *et al.* (1982) "Nature and distribution of iron in volcanic glasses : Mössbauer and EPR study of Lipari Pumice," *Bulletin de Minéralogie*, 105(1), pp. 43–50. Available at: https://doi.org/10.3406/bulmi.1982.7586.

Bishop, J.L., Fröschl, H. and Mancinelli, R.L. (1998) "Alteration processes in volcanic soils and identification of exobiologically important weathering products on Mars using remote sensing," *Journal of Geophysical Research: Planets*, 103(E13), pp. 31457–31476. Available at: https://doi.org/10.1029/1998je900008.

Byrne, P.K. *et al.* (2018) "The volcanic character of Mercury," *Mercury*, pp. 287–323. Available at: https://doi.org/10.1017/9781316650684.012.

Bell, J.F., Morris, R.V. and Adams, J.B. (1993) "Thermally altered palagonitic tephra: A spectral and process analog to the soil and dust of Mars," *Journal of Geophysical Research: Planets*, 98(E2), pp. 3373–3385. Available at: https://doi.org/10.1029/92je02367.

Cassidy, M. *et al.* (2014) "Construction of volcanic records from marine sediment cores: A review and Case Study (Montserrat, West Indies)," *Earth-Science Reviews*, 138, pp. 137–155. Available at: https://doi.org/10.1016/j.earscirev.2014.08.008.

Clegg, S.M. *et al.* (2009) "Multivariate analysis of remote laser-induced breakdown spectroscopy spectra using partial least squares, principal component analysis, and

related techniques," *Spectrochimica Acta Part B: Atomic Spectroscopy*, 64(1), pp. 79–88. Available at: https://doi.org/10.1016/j.sab.2008.10.045.

Di Genova, D. *et al.* (2017) "Effect of oxygen fugacity on the glass transition, viscosity and structure of silica- and iron-rich magmatic melts," *Journal of Non-Crystalline Solids*, 470, pp. 78–85. Available at: https://doi.org/10.1016/j.jnoncrysol.2017.05.013.

Dyar, M.D. and Ytsma, C.R. (2021) "Effect of data set size on geochemical quantification accuracy with laser-induced breakdown spectroscopy," *Spectrochimica Acta Part B: Atomic Spectroscopy*, 177, p. 106073. Available at: https://doi.org/10.1016/j.sab.2021.106073.

Eilers, P., and Boelens, H. (2005). "Baseline Correction with Asymmetric Least Squares Smoothing". *Unpubl. Manuscr.* Algorithm at:https://www.wavemetrics.com/code-snippet/asymmetric-least-squares-smoothing

Greeley, R. and Crown, D.A. (1990) "Volcanic geology of Tyrrhena Patera, Mars," *Journal of Geophysical Research*, 95(B5), p. 7133. Available at: https://doi.org/10.1029/jb095ib05p07133.

Greeley, R., *et al.*(2000) "Volcanism on the Red Planet: Mars," *In: Zimbelman, J.R., Gregg, T.K.P. (Eds.), Environmental Effects on Volcanic Eruptions: From Deep Oceans to Deep Space*. Kluwer, New York, NY, pp. 75–112.

Hastie, T., Robert, T., and Friedman, J. (2017) "The elements of statistical learning: data mining, inference, and prediction," *Springer Science & Business Media*

Hastie, T., Tibshirani, R. and Friedman, J. (2009) "The elements of Statistical Learning," *Springer Series in Statistics*. Available at: https://doi.org/10.1007/978-0-387-84858-7.

Wilhelms, D.E., McCauley, J.F. and Trask, N.J. (1987) “The geologic history of the Moon,” *Professional Paper*. Available at: https://doi.org/10.3133/pp1348.

Hook, S. J., *et* al (1999) “The use of multispectral thermal infrared data in geologic studies,” *in Manual of Remote Sensing, 3rd ed.*, edited by A. N. Rencz and R. A. Ryerson, pp. 59–110

Klaes, B. *et al.* (2022) “Iron (hydr)oxide formation in Andosols under extreme climate conditions.” Available at: https://doi.org/10.21203/rs.3.rs-1822714/v1.

Koralay, T. (2010) “Petrographic and geochemical characteristics of Upper Miocene tekkedag volcanics (Central Anatolia—Turkey),” *Geochemistry*, 70(4), pp. 335–351. Available at: https://doi.org/10.1016/j.chemer.2010.03.002.

Le Friant, A. *et al*. (2008) “Late Pleistocene tephrochronology of marine sediments adjacent to Montserrat, Lesser Antilles volcanic arc,” *J. Geol. Soc.*, 165, 279–290, Available at: doi:**1**0.1144/0016-76492007-019.

Le Friant, A.**,** and Expedition 340 Scientists (2013) **“**Lesse**r** Antilles Volcanism and landslides: Implications for hazard assessment and long-term magmatic evolution of the arc,**”** *Proceedings of the IODP*, 340, doi:10.2204/iodp.pr.340.2013.

Leight, C. (2020) Master’s Thesis: *The University of Tennessee Research and Creative Exchange*

Marshall, R.R. (1961) “Devitrification of natural glass,” *Bull. Geol. Soc. Am*. 72, 1493–1520. Available at: https://doi.org/10.1130/0016-7606(1961)72[1493:DONG]2.0.CO,2.

McCanta, M.C. *et al.* (2015) “Identifying cryptotephra units using Correlated Rapid, nondestructive methods: VSWIR spectroscopy, x-ray fluorescence, and magnetic susceptibility,” *Geochemistry, Geophysics, Geosystems*, 16(12), pp. 4029–4056. Available at: https://doi.org/10.1002/2015gc005913.

McCloy, J.S. (2019) “Frontiers in natural and un-natural glasses: An interdisciplinary dialogue and Review,” *Journal of Non-Crystalline Solids: X*, 4, p. 100035. Available at: https://doi.org/10.1016/j.nocx.2019.100035.

McMillan, P. (1984). “Structural Studies of Silicate Glasses and Melts-Applications and Limitations of Raman Spectroscopy”, *American Mineralogist,* 69, pp. 622-644.

Mysen, B., Virgo D., and Scarfe C. (1980) “Relations between the anionic structure and viscosity of silicate melts- a Raman spectroscopic study,” *American Mineralogist,* 65 (7-8), pp. 690-710.

Orlando, A. *et al.* (2021) “A comprehensive review on Raman spectroscopy applications,” *Chemosensors,* 9(9), p. 262. Available at: https://doi.org/10.3390/chemosensors9090262.

De La Pierre, M. *et al.* (2014) “The Raman spectrum of Caco3Polymorphs calcite and aragonite: A combined experimental and computational study,” *The Journal of Chemical Physics*, 140(16), p. 164509. Available at: https://doi.org/10.1063/1.4871900.

Rampe, E *et al*. (2014) “Amorphous phases on the surface of Mars.” *International Conference on Mars,* Abstract ID- 20140008714. Available at: https://ntrs.nasa.gov/citations/20140008714

Robinson, M.S. *et al.* (1993) “Chronology, eruption duration, and atmospheric contribution of the martian volcano apollinaris patera,” *Icarus*, 104(2), pp. 301–323. Available at: https://doi.org/10.1006/icar.1993.1103.

Rossano, S. and Mysen, B. O. (2012). “Raman spectroscopy of silicate glasses and melts in Geologic systems,” *EMU Notes in Mineralogy*, 12, chapter 9, pp. 319-364.

Stinton, A.J. *et al.* (2014) “Chapter 6 Dome Growth and valley fill during phase 5 (8 October 2009–11 February 2010) at the Soufrière Hills Volcano, Montserrat,”

Geological Society, London, Memoirs, 39(1), pp. 113–131. Available at: https://doi.org/10.1144/m39.6.

Stone, M. and Brooks, R.J. (1990) "Continuum regression: Cross-validated sequentially constructed prediction embracing ordinary least squares, partial least squares and principal components regression," *Journal of the Royal Statistical Society: Series B (Methodological)*, 52(2), pp. 237–258. Available at: https://doi.org/10.1111/j.2517-6161.1990.tb01786.x.

Surtees, A.P.H. (2015) Ph.D. Dissertation- "Development of geochemical identification and discrimination by Raman spectroscopy. The development of Raman spectroscopic methods for application to whole soil analysis and the separation of volcanic ashes for tephrachronology," *University of Bradford Scholars Repository*, Available at: http://hdl.handle.net/10454/14409

Surtees, A.P.H. *et al.* (2016) "Raman spectroscopy for the discrimination of tephras from the hekla eruptions of AD 1510 and 1947," *The Holocene*, 26(3), pp. 432–438. Available at: https://doi.org/10.1177/0959683615609744.

White, S.N. (2009) "Laser Raman spectroscopy as a technique for identification of seafloor hydrothermal and cold seep minerals," *Chemical Geology*, 259(3-4), pp. 240–252. Available at: https://doi.org/10.1016/j.chemgeo.2008.11.008.

Wilson, L. and Head, J.W. (1994) "Mars: Review and analysis of volcanic eruption theory and relationships to observed landforms," *Reviews of Geophysics*, 32(3), p. 221. Available at: https://doi.org/10.1029/94rg01113.

Ytsma, C.R. and Dyar, M.D. (2019) "Accuracies of lithium, boron, carbon, and sulfur quantification in geological samples with laser-induced breakdown spectroscopy in Mars, earth, and vacuum conditions," *Spectrochimica Acta Part B: Atomic Spectroscopy*, 162, p. 105715. Available at: https://doi.org/10.1016/j.sab.2019.105715.

Appendix

Tables

Table 3.1. Concentration mixing ratios by weight (mg) for the 13 natural tephra and hemipelagic sediment samples prepared by McCanta et al. (2015). Samples MC-1 and MC-13 represent pure sediment and pure tephra concentrations, respectively. Sample MC-8 contained equal concentrations of tephra-sediment, with a 50%:50% mixing ratio

Sample	Sediment Weight (mg)	Tephra Weight (mg)	% Tephra
MC-1	500	0	0
MC-2	490	10	2
MC-3	475	25	5
MC-4	450	50	10
MC-5	425	75	15
MC-6	400	100	20
MC-7	375	125	25
MC-8	250	250	50
MC-9	125	375	75
MC-10	100	400	80
MC-11	50	450	90
MC-12	25	475	95
MC-13	0	500	100

Table 3.2. PLS and LASSO regression model quality metrics. RMSE-C represents the internal predictive error of the calibration for each regression against the 13 training samples. RMSE-CV represents a proxy external predictive error and was produced from internal cross validation procedures. Both RMSEs are provided in the same units as the target variable (i.e., %) PLS component (q), and LASSO alpha (α) parameters, as determined via k-fold cross validation, are provided as a measure of algorithm shrinkage from the M.L.R. equation. q values < 10 were considered optimal for PLS regressions. Coefficient of determinations are listed (R^2) and shown in Figure 3.4

Model Type	**RMSE-C (%)**	**RMSE-CV (%)**	**R^2**	**q/α**	**$\sqrt{k}$**
PLS	7.96	15.46	0.955	4	3
LASSO	10.15	41.35	0.927	0.79	3

Table 3.3. Actual vs. PLS/LASSO model predicted % tephra values, along with the calculated difference between the predictions and actual concentrations. Both modeling strategies predicted consistent concentrations for samples whose Raman spectra exhibited band shape disparities due to crystalline mixing ratios

Sample	% Tephra				
	Actual	**LASSO Predicted**	**LASSO Difference**	**PLS Predicted**	**PLS Difference**
MC-1	0	7.8	7.8	2.9	2.9
MC-2	2	7.5	5.5	2.7	0.7
MC-3	5	7.8	2.8	5.9	0.9
MC-4	10	7.7	-2.3	5.7	-4.3
MC-5	15	22.9	7.9	27.5	12.5
MC-6	20	22.9	2.9	27.5	7.5
MC-7	25	22.9	-2.1	27.5	2.5
MC-8	50	22.9	-27.1	27.5	-22.5
MC-9	75	86.9	11.9	75.4	0.4
MC-10	80	85.0	5.0	85.3	5.3
MC-11	90	85.0	-5.0	85.3	-4.7
MC-12	95	85.5	-9.5	91.4	-3.6
MC-13	100	87.9	-12.1	102.6	2.6

Figures

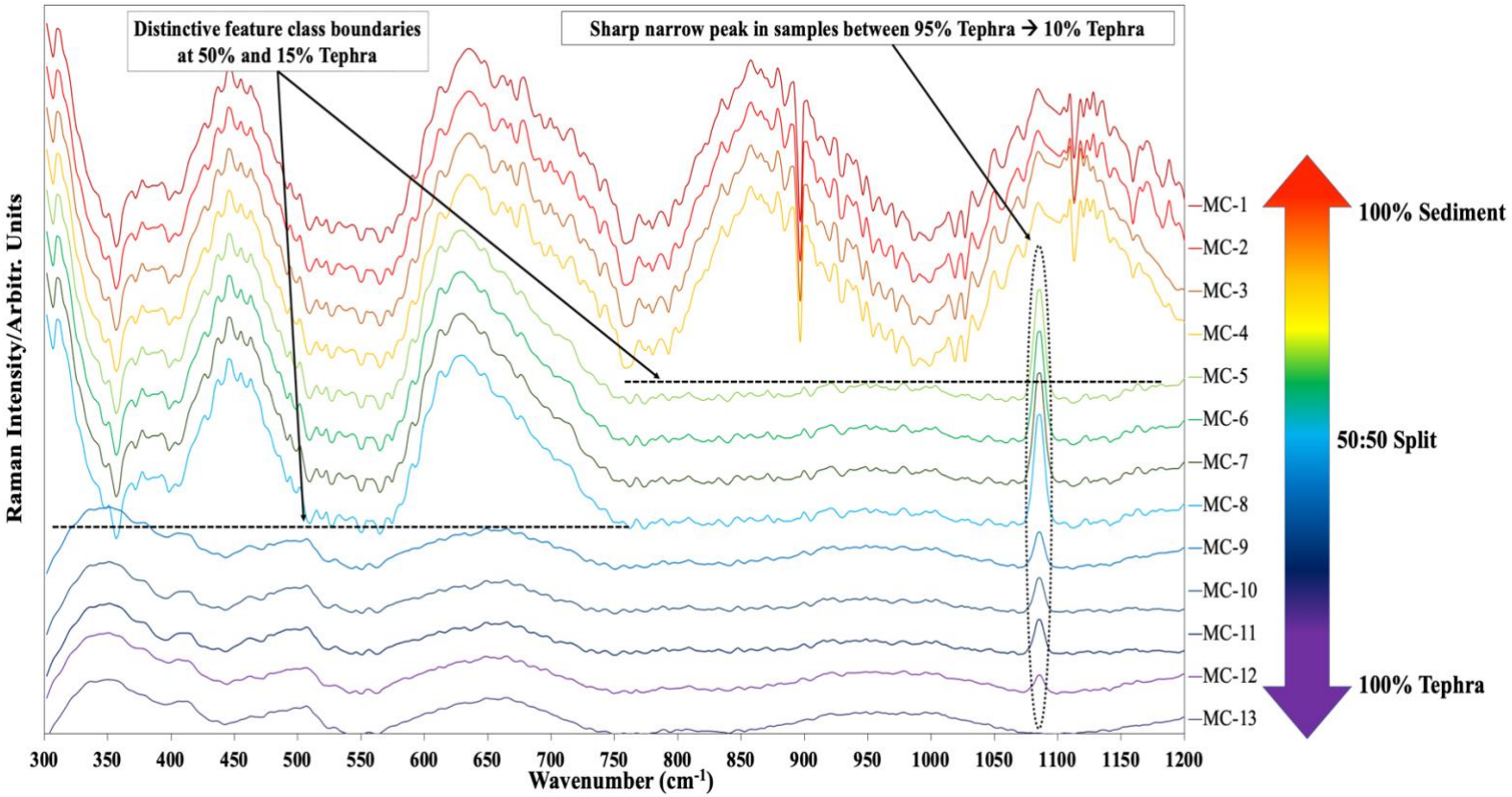

Figure 3.1. Color coordinated Raman spectra of the 13 tephra:sediment mixture samples, plotted with a graphical offset for clarity. Spectra were preprocessed and were mapped according to % tephra (i.e., purple = 100% tephra; red = 0% tephra). General Raman feature trends and descriptions are provided for the distinct shape disparities described in the text, which were observed near the 50%:50% and 15%:85% mixing boundaries of tephra:sediment. A singular peak was exhibited near ~1080 cm^{-1}. Distinct boundaries of signal loss, change, and conversion are detailed

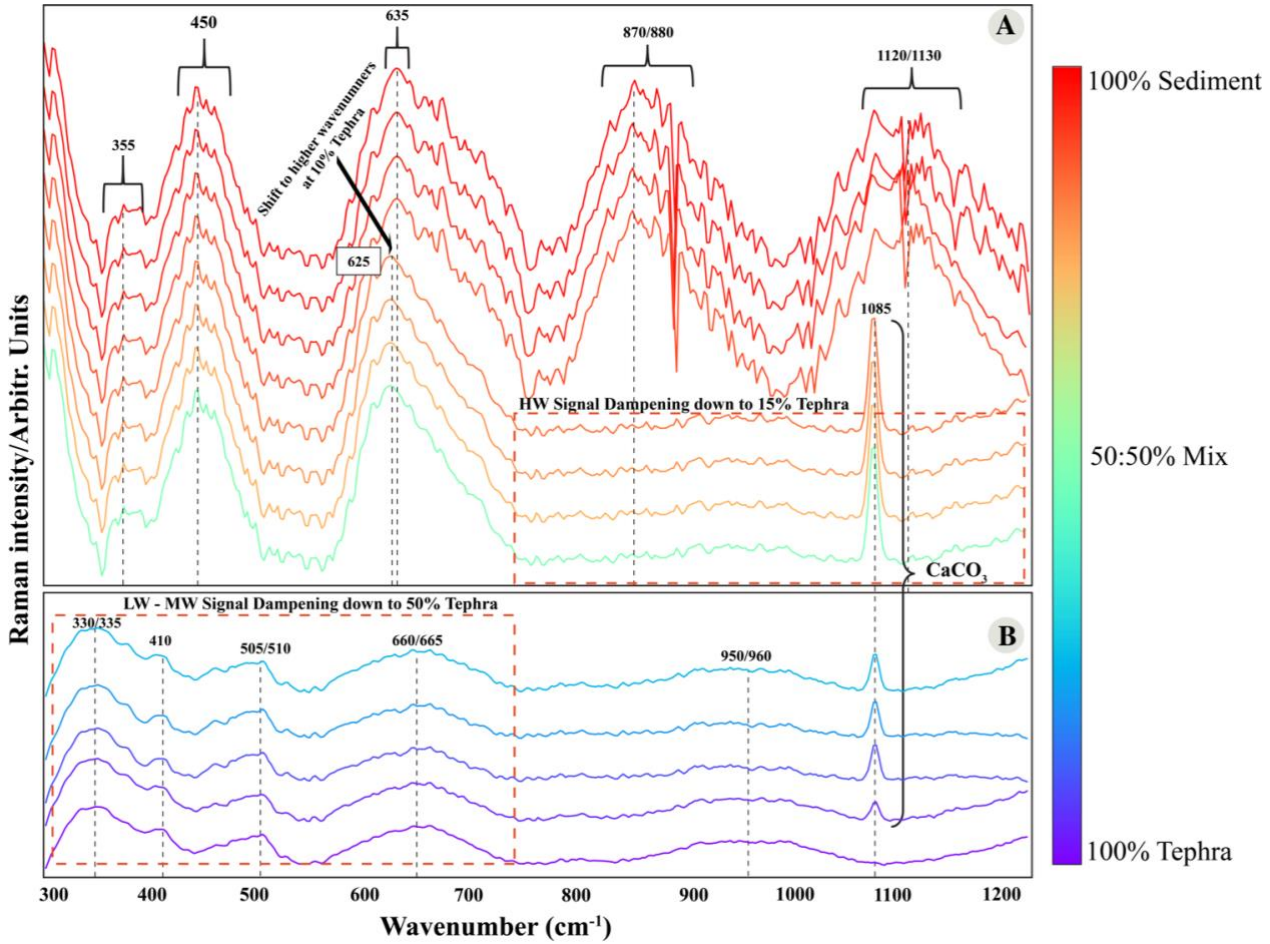

Figure 3.2. Color-coordinated Raman spectra of the 13 tephra:sediment mixtures, denoting major feature assignments based on Gaussian peak fitting procedures. Samples containing less than 50% tephra (**A**) showed disticntve, well-defiend peaks of stong intensity, whereas those samples richer in tephra (**B**) showed the inverse trend. Profiles are plotted with a graphical offset for clairty. Major peaks resulted from complex mixes of indivdual mineral species, and were affected by amorphous content primarily in the tephra-rich samples (**B**). The Raman peak near ~1085 cm^{-1} was assigned to calcite

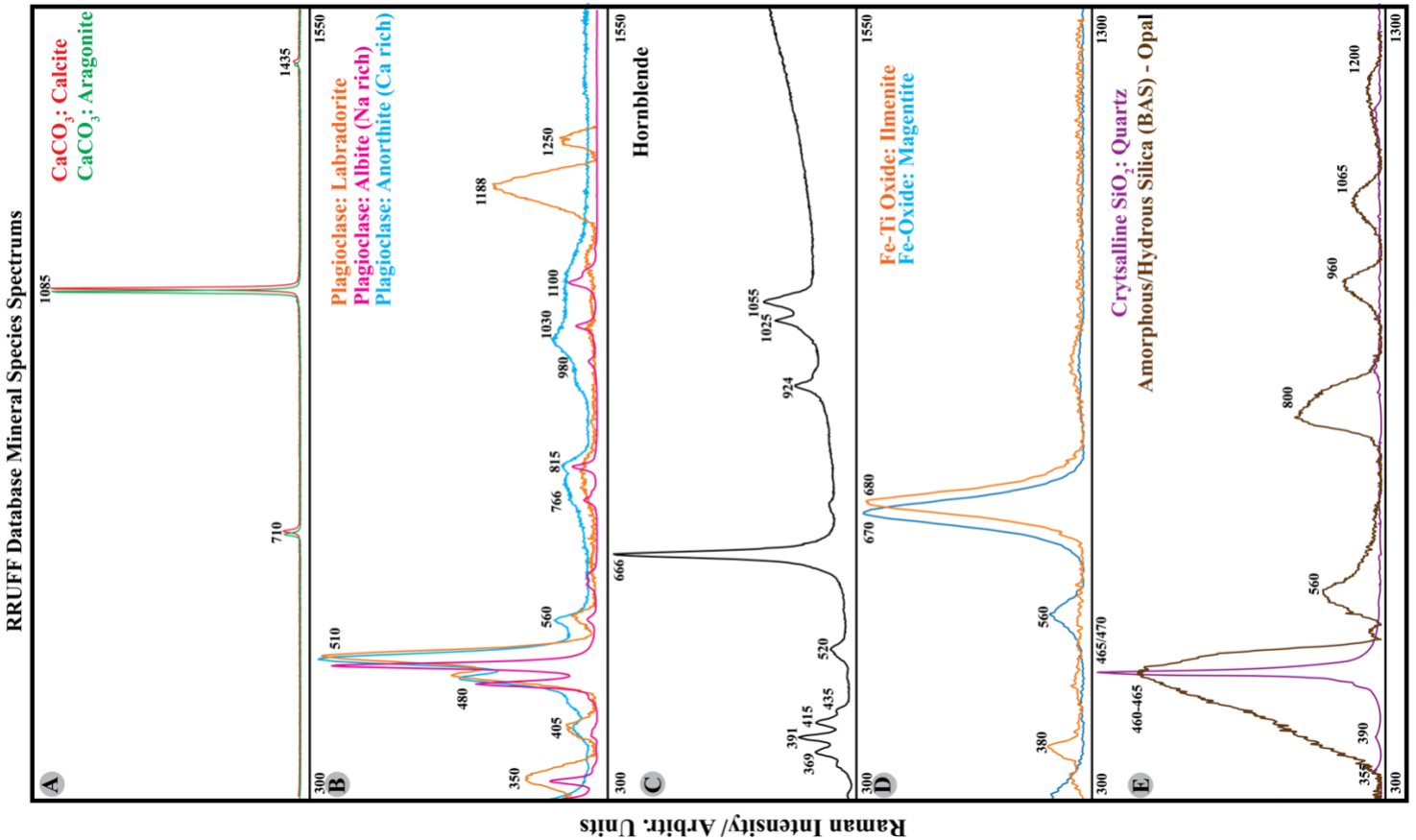

Figure 3.3. RRUFF Raman spectra profiles for mineral standards selected as possible matches for spectral features exhibited in the 13 Montserrat tephra:sediment sample mixes and based on compositonal and mineralogical informaiton from various sources (i.e., McCanta et al., 2015; Stinton et al., 2014; Le Friant et al., 2008, 2013)

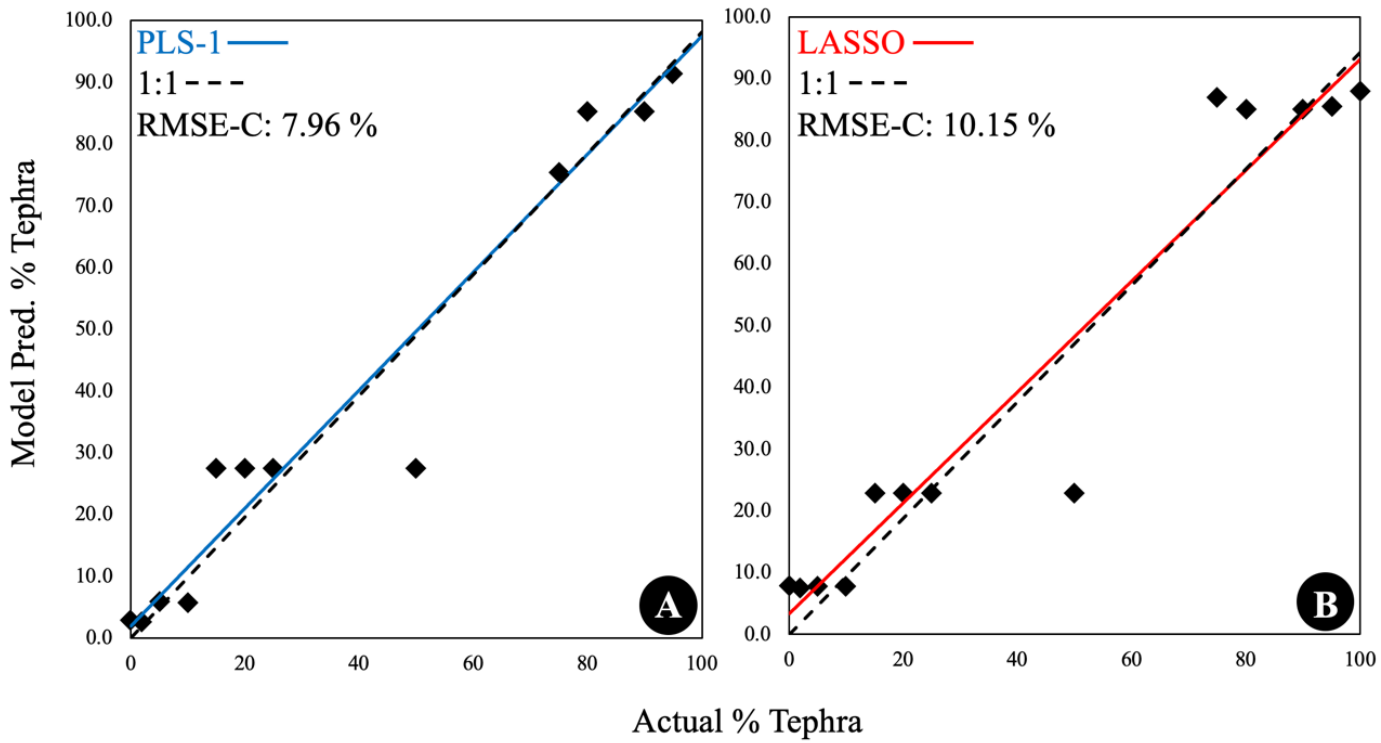

Figure 3.4. Model predicted vs actual tephra concentrations by weight plotted against a perfect 1:1 regression (black line) and their associated trendline from each modeling type using PLS (A) or LASSO (B). RMSE-Cs are also reported. Although both models yielded seemingly acceptable results, their contextualized cross-validation errors indicated a higher probability of less accurate predictions on future external datasets

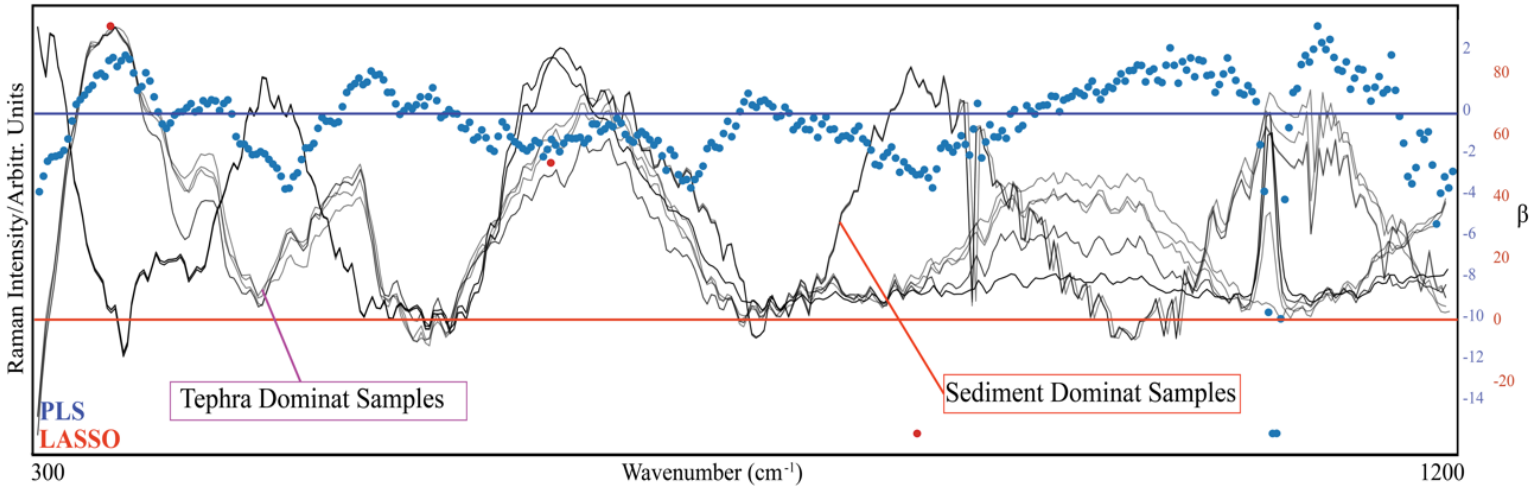

Figure 3.5 PLS and LASSO coefficients mapped against the training data suite of 13 tephra:sediment mixes and plotted without graphical offset for magnitude clarity. PLS coefficients (blue) and LASSO coefficients (red) ($\beta_j \ldots \beta_n$) are unitless, with weight plotted on the y-axis. PLS coefficients showed near-zero correlations to most Raman features in the region between 300–1200 cm^{-1}, except strong negative affinities for the calcite peak near ~1080 cm^{-1}, which was not observed in the pure tephra sample mixture. LASSO assigned only three coefficients along the profile, with a strong negative coefficient near ~880 cm^{-1}

CONCLUSIONS

The research presented in this dissertation strove to develop, test, and contextualize newly adopted, novel modalities for characterizing silicate melts through their eruptive byproducts. With the rise in popularity, technical innovation, and success of Raman spectroscopy-based techniques for such characterizations in the geosciences over the past decade, we decided to focus our efforts on utilizing this method as the primary foundation for developing our unique modeling style. At the same time, the adoption of machine learning technologies like PLS and LASSO regression in the field is also expanding and has shown great success in other spectral techniques solving similar questions. With these simultaneous advancements taking place, it became clear that the merging of the two tools would be inevitable, and only a limited number of such research investigations were publicly underway. Selection of what type of volcanic byproducts to target was an easy choice; and synthetic silicate glass samples were chosen to serve as the best proxies for characterizing silicate melts as an initial step towards solving other Raman-amorphous questions. By combining the robust molecular vibration information contained in Raman-glass spectral data with machine learning models in Chapters I and II, valuable insights were gained on the viability of using these two techniques in solving such questions. Chapter III builds off of those insights and sought to expand the experimental approach to external volcanic questions. The first chapter is that of a successful experimentation; the second – a dive into the nature of a failure, and the third – an encouraging preliminary glance towards the future applications of the tested techniques.

Specifically: Chapter I reports the findings of a successful PLS-1 – Raman modeling calibration, capable of constraining major oxide concentrations (SiO_2, CaO, TiO_2, Al_2O_3, FeO_T, Na_2O, K_2O, and MgO) in a chemically diverse glass suite – that was suitable for extrapolation to liquid state characterizations. Chapter II details the shortcomings of both PLS-1 and LASSO calibrations on the glass suite (with some minor additions) in efforts to predict Fe-redox variables ($\%Fe^{3+}$, $\%Fe^{2+}$, Fe_2O_3, FeO). Chapter III discusses the capability of Raman-MVA approaches for unraveling tephra

concentrations from various sediment mixing scenarios and highlights promising predictive power for both regression methodologies in a preliminary capacity. Each experiment helped to uncover the wild complexity associated with modeling silicate samples – in terms of both the unstandardized parameters for their Raman signals, and the complications with mapping these parameters using linear-based modeling technologies. Key contextual information was obtained from investigating model-derived coefficients as a function of the collected Raman channels, and from correlations in internal glass chemistry to the features present at those channels. Together this information helped shed light on the successes and shortcomings within this works' approaches. A set of these major conclusions is summarized below from each study, and their general interpretations as a function of the three chapters together is also provided.

Chapter I

- PLS-1 is able to effectively quantify general chemistry in silicate glasses via their Raman spectra when compared to similarly constructed models: with a normalized average prediction error near ~28 %. This result is consistent with currently available modeling publications (i.e., González-García et al., 2020 - ~21% error), though no such study has provided the level of variability exhibited in the glass dataset utilized in Chapter I.
- Modeling suite quality metrics detail a consistent treatment among the variables through internal and external testing (cross validation and external validation) – indicating the regression is not unduly overfitting the training data for any of the targeted oxides.
- As the model is built using a sample suite of glasses that span four rock types and large swings in oxidation state (i.e., basalts – rhyolites along reducing and oxidizing MRB's) the model is truly generalizable and does suffer from development constrains like other previous modeling strategies. This model can be used to predict wt% oxides in any type of encountered volcanic glass.
- When held along constant oxidation states, the glass spectra show high sensitivity to glass chemistry, with major positional shifts arising in the low-mid-and high

wavenumber envelopes. Major contributions to these shifts are caused by increasing or decreasing concentrations of SiO_2, Na_2O & K_2O, CaO, and FeO_T.

- PLS model coefficients (β) help contextualize the positive or negative correlations observed by the regression that an oxide variable has with a specific spectral channel (or wavenumber position). All of the oxides are shown to have significant correlation magnitudes with practically every channel but are primarily confined to strong relationships with the LW and HW band centered near ~500 cm^{-1} and 980 cm^{-1}. Based on spectral trends with chemistry, the regression is able to sufficiently map Raman-glass data for use in predictions.

Chapter II

- PLS and LASSO calibrations developed for predicting Fe-redox variables are not appropriately accurate for serious use and exhibit high normalized prediction errors of ~30 – 35% on average (for both techniques). There is no appreciable difference between the predictions of either Fe valance ratios, or their oxide counterparts (i.e., $\%Fe^{3+}$ vs Fe_2O_3 wt%; Or $\%Fe^{2+}$ vs FeO wt%).
- LASSO and PLS calibrations for the structural parameter proxy for polymerization (NBO/T) are suitable for use based on predictive errors (NRMSE-P = ~23 % but are still much higher than those reported in similar investigations.
- Overall, the PLS suite performs slightly better on average than the LASSO suite. This is attributed to the fact that PLS models assign coefficients (the engines of predictions) to all available spectral channels – while the LASSO models use ~ 2 on average for the redox tests. However, it clear that not all of these Raman channels are important in the regression of the variables, and therefore PLS incorporates erroneous information that weighs down predictive accuracy to only a slightly better performance to the LASSO strategy.
- Model coefficients seem to be in good agreement the uncovered effects of Fe-redox in the Raman HW band, and map inverse trends for ferric and ferrous metrics – showing that the regressions (at least PLS) does recognize the correct relationship between the variables and the channels in this region, however, it also maps large trends in the LW and MW bands.

- The glass samples express the expected trends in major chemistry as a function of polymerization (NBO/T), where increasing influences (concentrations) of network forming cations in the glass network cause calculated NBO/T values to decrease. The trend for Fe-redox state is (interestingly) less concrete upon initial inspection but occurs along defined patterns among FeO_T measurements.
- Raman signals for the glass suite in this chapter exhibit strong trends with oxidation state in the HW band, and the observed trends agree with previous experiments. The normalized intensity of the HW band increases with increasing oxidation state (i.e., increasing $\%Fe^{3+}$), and the peak shifts to lower wavenumbers.

Chapter III

- Initial PLS and LASSO training models for predicting % tephra concentrations in the 13 sediment:volcaniclastic mixed samples yield promising results. RMSE-CV's for PLS and LASSO calibrations are ~15 and 41% respectively – with only the LASSO strategy falling short of external usability (though no such tests were possible due to the small sample suite). R^2's for the models were both >0.91.
- Model coefficients for tephra concentration in the PLS regression are not heavily weighted across the spectrum for any particular channel except a $CaCO_3$ feature near 1085 cm^{-1} that is not present in the purely tephra sample. The coefficients distinctly follow the trends observed in the purely tephra sample (MC-13), indicating that the regression is effectively trained to consider only these types of signals in its prediction. LASSO coefficients are sparse, and only place importance on 3 channels in the envelope.
- Distinct disparities in overall band shape and peak structures between tephra-rich and sediment-rich mixes is observed for the region between 300 – 1200 cm^{-1}, at distinct mixing boundaries. Samples with higher tephra concentrations show Raman bands of lower intensity and are definitively broad and ill-defined when compared to those rich in sediment. This signifies that Raman spectroscopy can be used to distinguish between tephra-related signals in scenarios where the material may be mixed down to a ratio of 50%:50% tephra:sediment in the LW

(300 – 750 cm^{-1}) and to 15%:85% in the HW region (750– 1200 cm^{-1}), without the need for direct modeling.

- The Raman signals for the samples can be accurately attributed to individual mineral components when referenced against existing Raman databases and are heavily influenced by the presence of amorphous phases.

Based on these concluding points, it is uncovered that linear-based modeling methodologies are not an extremely ideal fit for directly determining chemical information in Raman-glass signals but do show some major successes that can serve as benchmarks for future advancements and applications when juxtaposed against comparable experiments in the field. Raman data alone can be used to identify key changes in glass chemistry and discern tephra signals in mixed phases – without the need for direct modelling. The lack of pinpoint accuracies for any of the predictions provided through PLS and LASSO in this work is than likely due to one of two items: the relationships between oxide, Fe-redox, and tephra variables in their collected Raman profiles is not directly linear, or it is highly multicollinear, and the modeling strategies are unable to perfectly deal with either scenario. The nature of this problem has yet to be fully established and will require future work. This series of works however, highlight the notion that despite these possible shortcomings, quantitative and qualitative information on volcanic silicates can be adequality obtained using the described methodology pairings.

VITA

Blake O. LaDouceur is originally from a rural county in the Florida panhandle, near Pensacola. After finishing high school with an eye towards studying medicine, he was recruited to play NCAA college football at Johns Hopkins University in Baltimore MD in 2014. While pursuing his pre-med academic track he played offensive tackle for the Blue Jays and was inducted as a member of the Baltimore Adelphic Literary Society (Alpha Delta Phi – JHU Chapter). During the summer of his junior year, he was presented a unique opportunity to combine his studies in epidemiology to the planetary sciences for the first time, in a single internship at the Marshall Space Flight Center in Huntsville, AL. The joint EPA-NASA project involved building prediction models for particulate matter concentrations in atmospheric columns using NASA EOS systems, and projecting the adverse public health impacts such concentrations would have on surrounding populations over time. This project exposed Blake to the wonders of space research, and the geosciences by proxy. Enthralled with this new passion, Blake enrolled in double the allowed college credits in his final undergraduate year to earn a minor in Earth and Planetary Sciences, in an effort to prepare him for a career-trajectory switch. Following his graduation from JHU, Blake was offered the chance to study for a master's degree under Dr. Jennifer Gifford in the Geology and Geological engineering Dept. at the University of Mississippi. His research there focused on piecing together evidence for a new amalgamation model hypothesis for the ancient core of the North American continent Laurentia via U/Pb and Lu/Hf isotopic data reduction techniques. While at UM, Blake gained significant exposure to the foundational scientific principles underlying the geosciences (of which he was asked to teach while being a novice himself), which prepared him well for future endeavors in academic research surrounding such topics. After connecting with student and professor representatives from the Earth and Planetary Sciences Dept. at the University of Tennessee, Knoxville at a national conference in 2018, Blake was offered the opportunity to study under Dr. Molly McCanta on the projects described in this dissertation. Blake and Molly were connected over the phone for the first time in late 2018, while Blake was on the first day of his honeymoon in Jamaica. Under Dr. McCanta's expert direction, Blake researched the viability of newly

adopted spectroscopic techniques in studying volcanic glass properties and looked to implement new-age computing methodologies to perform the necessary experiments. Blake's passion for planetary research runs in tandem with his newfound interest in computer science (spurred from these recent Ph.D. projects). He hopes to continue to combine these interests in future research projects as a professional planetary scientist and offers this dissertation as one the first steps along that journey.