

PyBaselines: An Open-Source MCMC Algorithm for Fitting Baselines to FTIR Spectra of Basaltic-Rhyolitic Glasses

Sarah Shi^{1,4}, Henry Towbin¹, Terry Plank¹, Anna Barth^{1,2}, Dan Rasmussen^{1,3}

¹Lamont-Doherty Earth Observatory, Columbia University, New York, NY USA

²University of California, Berkeley, Berkeley, CA USA

³National Museum of Natural History, Smithsonian Institution, Washington, DC USA

⁴University of Cambridge

Key Points:

- 1
- 2

Abstract

1 Introduction

Degassing of volatiles in magmas modulate magma phase equilibria (Sisson & Grove, 1993), crystallinity (Blundy & Cashman, 2001), density (Ochs III & Lange, 1999), and viscosity (Hess & Dingwell, 1996) from generation to eruption. The most abundant of these volatiles are H₂O and CO₂, both of which dissolve in silicate melts with strong pressure dependence. Analyses of H₂O and CO₂ in melt inclusions trapped in mineral crystals can thus be utilized to estimate the conditions of magma storage and evolution, whereas those in experimental charges inform understandings of solubility and phase equilibria. H₂O and CO₂ in volcanic glasses are commonly measured with the two analytical techniques of Fourier Transform Infrared Spectroscopy (FTIR) and Secondary-Ion Mass Spectrometry (SIMS) to characterize natural volcanic systems and experimental recreations of high-pressure and high-temperature magmatic conditions. *necessary to talk a bit about benefits and drawbacks here? or later?*

FTIR spectroscopy excites materials with visible light through wavelengths in the infrared region to determine transitions in rotational, vibrational, and electronic energies. Volatiles in silicate glasses absorb light proportional to concentration in certain wavelength regions, allowing for recognition and quantification of species with the Beer-Lambert Law:

$$c = \frac{AM}{\epsilon l \rho} \quad (1)$$

where c is concentration, A is absorbance, M is the molar mass of the absorbing volatile species (g·mol⁻¹), ϵ is the absorptivity of the species (L·mol⁻¹·cm), l is the optical path length or thickness (cm), and ρ is density (kg·m⁻³). **Provide image of spectrum, define IR near IR regions, describe a bit.**

Linear relationships between absorbance and concentration dominate, but non-linear relationships can be introduced when increased sample thicknesses result in insufficient light transmission to the detector, pushing peaks to become saturated as with the H₂O_{t,3550} peak (McIntosh et al., 2017; von Aulock et al., 2014). The H₂O_{t,3550} peak can be well quantified with a linear to near-linear baseline between the two absorbance minima on either side of peak when the peak is unsaturated with a raw absorbance of <2 (Dixon & Stolper, 1995; von Aulock et al., 2014). When ϵ H₂O_{t,3550} is saturated, the combination of H₂O_{m,5200} or the H₂O_{m,1635} and the OH₄₅₀₀⁻ must be used instead.

Volatile concentrations in silicate glasses require the measurement of peak heights or areas, by measuring the difference in absorbance or integrated absorbance between peak and baseline. Variability and subjectivity in baseline definition constitutes a significant uncertainty in studies determining volatile concentrations, from inception to present. The H₂O_{m,5200}, OH₄₅₀₀⁻, and ϵ H₂O_{t,3550} peaks in the near-infrared (12800-4000 cm⁻¹) to mid-infrared (4000-200 cm⁻¹) region can be well quantified with linear to near-linear baselines between the two absorbance minima on either side of the peak. The H₂O_{m,1635} peak and CO₃²⁻ doublet peaks pose greater challenges due to the steep and sharp increase of baseline absorbance at wavenumbers lower than 1430 cm⁻¹. Furthermore, the convolution of the tails of the H₂O_{m,1635} and the CO₃²⁻ 1515 doublet peak in melt inclusion spectra further complicates the appropriate baselines in the near infrared region. Baselines beneath the H₂O_{m,1635} and CO₃²⁻ doublet peaks have been approximated with the spectra of matrix- and chemistry-matched devolatilized glasses (Dixon et al., 1988; Newman et al., 2000), splines and flexicurve (Dixon & Stolper, 1995), and other curve functions (Dixon & Clague, 2001). Devolatilized glass baselines are the gold standard but are not possible in melt inclusion studies, where large pieces of glass are not readily available. Spline, flexicurve, and curve function baselines generated without open-source code are not replicable and not associated with appropriate uncertainties.

We examine each parameter within the Beer-Lambert Law to develop PyBaselines, a Python implementation of a Bayesian-informed Multi-Core Markov Chain Monte Carlo algorithm for fitting baselines and peaks for all H₂O and CO₂ species. We assess the fundamental shape and variability of the baseline using principal component analysis (PCA) applied to a database of absorbance spectra for volatile-free and volatile-poor melt inclusions from the Aleutian arc (Rasmussen, 2019) and MORB glasses (Newman et al., 2000). This dataset of naturally devolatilized spectra is relevant for basaltic to rhyolitic silicate melts. The application of PCA to a training dataset of devolatilized spectra allows for the assessment of the fundamental shape of the baseline and the spectral features contributing to variability (Carvajal et al., 2016). We refit the relationships between molar absorptivity and composition with implicit iterative inversion with updated datasets from the last parameterization of Mandeville et al. (2002) and to derive molar absorptivity with uncertainty. We develop automated implementations for calculating thicknesses from reflectance spectra by the reflectance FTIR interference fringe method. PyBaselines allows for the reproducible quantification of volatile concentrations with meaningful uncertainties.

2 Background

2.1 Fourier Transform Infrared Spectroscopy (FTIR) Background

I currently have everything incorporated into computational methodology – both background and our solutions. This is a lot of the structural trouble I was thinking about. I don't really know to split this well. How to split????????????????????

Nonlinearities, linear Species dependent dissolution, etc. Peak appearances

3 Computational Methodology

3.1 Assessing Variability in Baselines and Peaks

The shape and variability of the baseline and the H₂O_{m,1635} and CO₃²⁻ peaks were assessed with a database of ~90 absorbance FTIR spectra for Aleutian melt inclusions and MORB glasses *need Henry's numbers for each type of spectrum, statistics on composition, data to upload to Github repository – Henry is collaborator so can maybe post a new directory about all this with code*. The spectra variably lack CO₃²⁻ (decarbonated), H₂O_{m,1635} (dehydrated), or both CO₃²⁻ and H₂O_{m,1635} (devolatilized) and span a wide compositional range (SiO₂ = 43-60 wt.%, MgO = 2.1-6.4 wt.%, Na₂O = 2.2-4.3 wt.%, K₂O = xy value).

Definition of baseline and peak shapes was iterative. Completely devolatilized absorbance spectra were scaled and shifted to ensure equal absorbances in the fitting range from 2400 to 1275 cm⁻¹. We first generate the initial mean baseline ($\overline{\text{Baseline}}_i$) and described variability within the baseline with principal component analysis (PCA). Four principal components describe $\geq 95\%$ of the variance within the baseline.

Decarbonated spectra containing only H₂O_{m,1635} peaks were fit with the $\overline{\text{Baseline}}_i$ and two principal components to isolate the shape of the H₂O_{m,1635} peak, which is neither Gaussian nor Lorentzian peaks. We thus identify the mean H₂O_{m,1635} ($\overline{\text{H}_2\text{O}_{m,1635}}$) peak shape and describe variability within the peak with PCA (Figure 1). The downturn of the H₂O_{m,1635} peak occurs more rapidly than that of a Gaussian or Lorentzian peak *Maybe need a figure from Henry to show our proof of this*. Principal components (PC) 1 and 2 allow for slight lateral deviations of the H₂O_{m,1635} central peak location, generating better fits when shifted away from the standard peak location at 1635 cm⁻¹. Two principal components describe $\geq 95\%$ of the variance within the baseline. Identification of the H₂O_{m,1635} peak shape further allows for H₂O_{m,1635} peak subtrac-

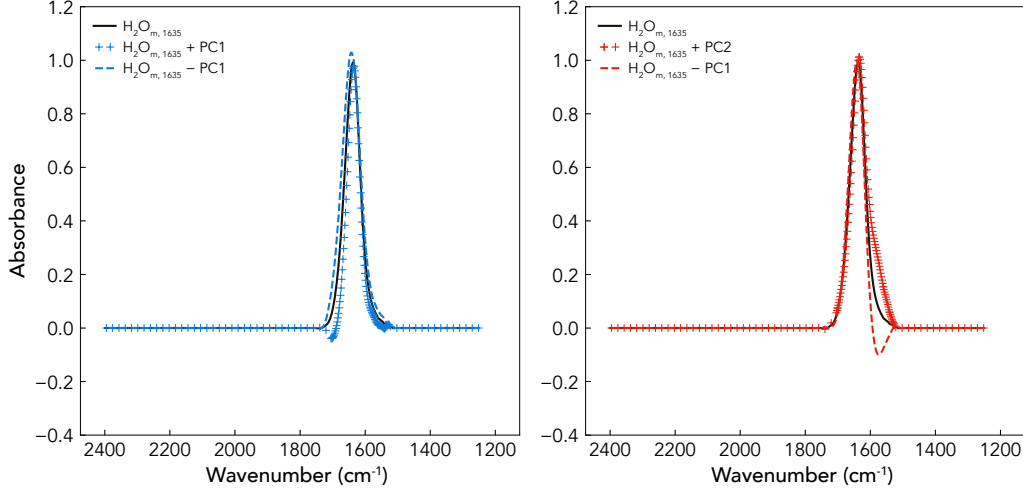


Figure 1: Mean $\text{H}_2\text{O}_{m,1635}$ peak with PC components describing variability

tion from the decarbonated spectra with least squares, generating additional devolatilized spectra.

Dehydrated spectra with CO_3^{2-} were similarly utilized to describe the CO_3^{2-} peaks, which are simple Gaussian peaks *Maybe need a figure from Henry to show our proof of this*. CO_3^{2-} peak subtraction from dehydrated spectra with least squares generate more devolatilized spectra.

Decarbonated and dehydrated spectra were thus converted to entirely devolatilized spectra. We iteratively identify and iteratively define the mean devolatilized baseline (Baseline) with associated variability (Figure 2). Variability in the baseline can be described with the addition or subtraction of these PC components, highlighting and describing the different baselines observed within the devolatilized spectrum database. We further disentangle the convoluted tails of the overlapping $\text{H}_2\text{O}_{m,1635}$ and the $\text{CO}_{3,1515}^{2-}$ peaks by identifying $\text{H}_2\text{O}_{m,1635}$ with associated principal components and the Gaussian nature of the CO_3^{2-} peaks. Given the sharp upturn in absorbance while approaching 1275 cm^{-1} , we provide additional accommodation for shifting and tilting the $\text{H}_2\text{O}_{m,1635}$ and $\text{CO}_{3,1515}^{2-}$ peaks with a linear shift. We can therefore credibly describe the range of possible devolatilized baselines while simultaneously fitting peaks appropriate for basaltic-rhyolitic glasses with the net equation:

$$\hat{A} = \hat{B} + \text{H}_2\text{O}_{m,1635}^{\hat{}} + \text{CO}_3^{2-}^{\hat{}} + \hat{L}, \quad (2)$$

where \hat{A} is the best-fit absorbance spectrum. \hat{B} or the best-fit baseline is:

$$\hat{B} = x_0 \bar{B} + x_1 \bar{B}_{PC1} + x_2 \bar{B}_{PC2} + x_3 \bar{B}_{PC3} + x_4 \bar{B}_{PC4}, \quad (3)$$

where the $x_{\text{subscript}}$ terms are best-fit scaling parameters, \bar{B} is the mean baseline, and \bar{B}_{PC} is the n-th principal component of the baseline. $\text{H}_2\text{O}_{m,1635}^{\hat{}}$ or the best-fit $\text{H}_2\text{O}_{m,1635}$ peak is:

$$\text{H}_2\text{O}_{m,1635}^{\hat{}} = y_0 \overline{\text{H}_2\text{O}_{m,1635}} + y_1 \overline{\text{H}_2\text{O}_{m,1635}}_{PC1} + y_2 \overline{\text{H}_2\text{O}_{m,1635}}_{PC2}, \quad (4)$$

where the $y_{\text{subscript}}$ terms are best-fit scaling parameters, $\overline{\text{H}_2\text{O}_{m,1635}}$ is the mean $\text{H}_2\text{O}_{m,1635}$ peak, and $\overline{\text{H}_2\text{O}_{m,1635}}_{PC}$ is the n-th principal component of the $\text{H}_2\text{O}_{m,1635}$ peak. Peak amplitude is fit with $\overline{\text{H}_2\text{O}_{m,1635}}$ and the principal components accommodate variations in

133 peak location. $\hat{CO}_{3,1515/1430}^{2-}$ or the best-fit CO_3^{2-} peak is:

$$CO_{3,1515/1430}^{2-} = \frac{a}{2\sigma^2} \cdot e^{-(v^- - \mu)^2}, \quad (5)$$

134 where a is the amplitude, σ is the peak half-width, v^- is the wavenumber range of 2400
135 to 1275 cm^{-1} , and μ is the central wavenumber of the peak. The μ parameter accounts
136 for slight deviations in peak location. \hat{L} or the best-fit line accomodating for shifting and
137 tilting of peaks is:

$$\hat{L} = mv + b \quad (6)$$

138 where m is the slope, v is the wavenumber range of interest, and b is the intercept.

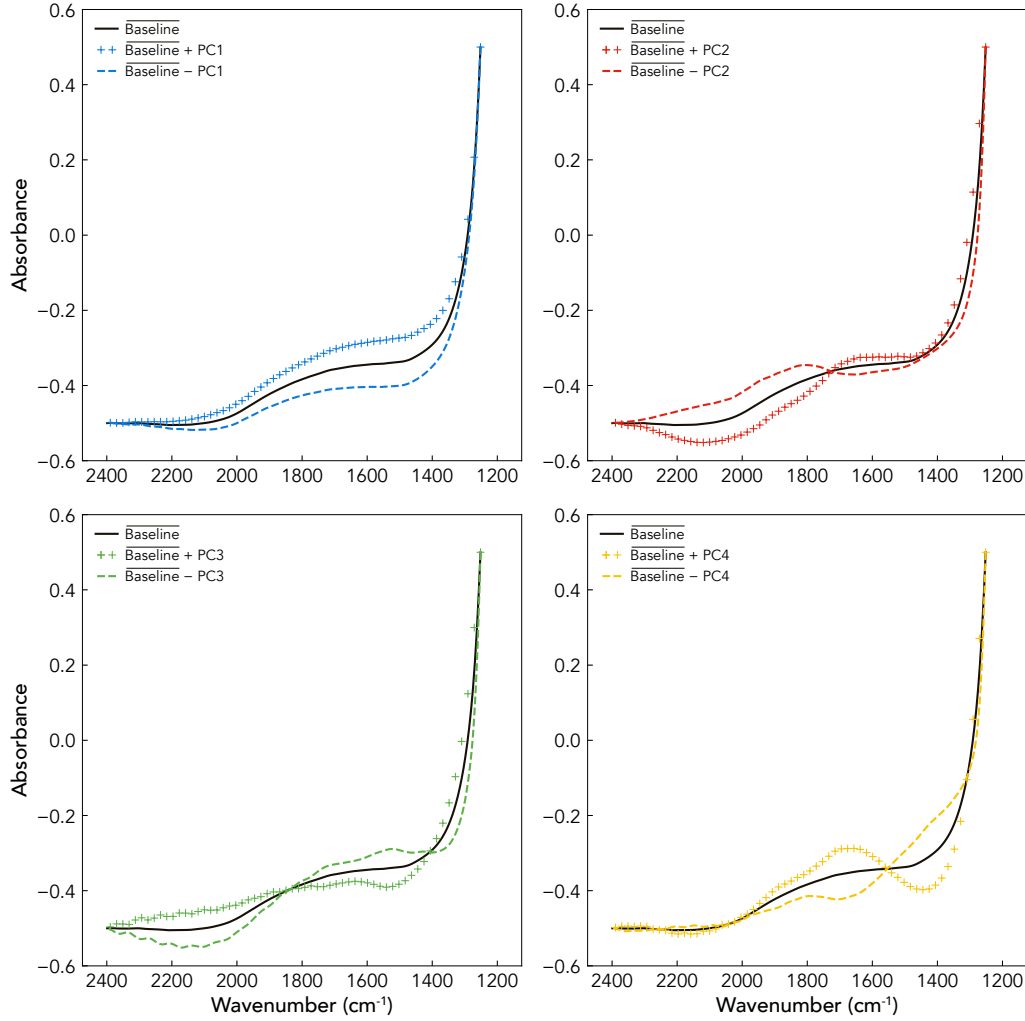


Figure 2: Mean devolatilized baseline (—) with PC components describing potential variability within the baseline. A.

139 3.2 Calibrating Molar Absorptivity (ϵ)

140 Molar absorptivities or extinction coefficients (ϵ) determine the intensity at which
141 light attenuates when passing through a material and relates absorbance to concentra-
142 tion, calibrated by pairing analyses of volatile content by transmission FTIR and by an

additional independent analytical technique. ϵ is dependent on glass composition, related to both the tetrahedral cation fraction of $\tau = (\text{Si}^{4+} + \text{Al}^{3+}) / (\text{total cations})$ and the cation fraction of $\text{Na}^+ / (\text{Na}^+ + \text{Ca}^{2+})$. The ϵ data are limited to studies with independent methods of volatile measurement – such as SIMS, Karl Fischer titration, experimental mass balance, vacuum heating, or manometry – and with $\text{H}_2\text{O} > 0.5$ wt.% for peak resolution. In the full dataset, τ spans the range of 0.50-0.90 and $\text{Na}^+ / (\text{Na}^+ + \text{Ca}^{2+})$ from 0.23-0.84. We examine whether these compositional parameters remain the best descriptors of extinction coefficient. We refit the linear relationships between ϵCO_3^{2-} , $\epsilon\text{OH}_{4500}^-$, $\epsilon\text{H}_2\text{O}_{\text{t},3550}$, $\epsilon\text{H}_2\text{O}_{\text{m},1635}$, and ϵCO_3^{2-} and the corresponding compositional parameters values to include more recent data and broader compositional ranges since the last compilation of extinction coefficients by Mandeville et al. (2002), with an implicit inversion.

The extinction coefficients related to H_2O species increase with tetrahedral cation fraction, τ , reflecting increased polymerization of the glass (Stolper, 1982). Tetrahedral cations have also been proposed to compete with non-tetrahedral cations ($\text{M} = \text{Mg}^{2+}$, Ca^{2+} , Na^+) to bond with free hydroxyl groups in melt (Mercier et al., 2010; Pandya et al., 1992). More depolymerized melts will have higher proportions of free hydroxyl, which translates to more $\text{M}(\text{OH})_x$ bonding and hydrogen bonding (Mercier et al., 2010; Xue, 2009). Dixon et al. (1995) and Mandeville et al. (2002) demonstrate the dependence of the $\epsilon\text{H}_2\text{O}_{\text{m},5200}$, $\epsilon\text{OH}_{4500}^-$, and $\epsilon\text{H}_2\text{O}_{\text{m},1635}$ on glass tetrahedral cation fraction. Tetrahedral cation fraction is similarly observed to be positively correlated with $\epsilon\text{H}_2\text{O}_{\text{t},3550}$, but this relationship is less straightforward (Mercier et al., 2010). Quench temperatures and quench rates drive the formation of variable proportions of $\text{H}_2\text{O}_{\text{m}}$ and OH^- (Silver & Stolper, 1989; Stolper, 1989). $\epsilon\text{H}_2\text{O}_{\text{t},3550}$ is thus considered as a combination of two endmember absorption coefficients, $\epsilon\text{OH}_{3550}^-$ and $\epsilon\text{H}_2\text{O}_{\text{m},3550}$, multiplied by the relative proportion of each respective species (McIntosh et al., 2017; Newman et al., 1986; Okumura et al., 2003). Studies quantifying $\epsilon\text{OH}_{3550}^-$ and $\epsilon\text{H}_2\text{O}_{\text{m},3550}$ are the gold standard, but are currently limited in compositional range; McIntosh et al. (2017) considered the range of $\tau = 0.746$ -0.800, spanning silicic compositions from rhyolitic to albitic glasses. More experiments are needed to properly quantify the species-dependent nature of the $\text{H}_2\text{O}_{\text{t},3550}$ peak. We find a positive relationship between $\epsilon\text{H}_2\text{O}_{\text{t},3550}$ and τ , despite compilations presented by speciation. While endmember data for $\text{H}_2\text{O}_{\text{t},3550}$ remain limited, $\epsilon\text{H}_2\text{O}_{\text{t},3550}$ is best described by τ . The relationship observed with NBO/T , calculated by methods outlined in Mysen and Richet (2018), is not observed in this dataset, despite the observation in Mercier et al. (2010) (Supplemental Figure xx).

Dixon and Pan (1995) first demonstrate the dependence of the CO_3^{2-} doublet absorption coefficients on the molar or cationic ratio of $\text{Na}^+ / (\text{Na}^+ + \text{Ca}^{2+})$. The proportion of Na^+ to Ca^{2+} modulates the dissolution of C as molecular CO_2 and/or CO_3^{2-} . As the proportion of Ca^{2+} increases, ϵCO_3^{2-} decreases. C dissolves as both molecular CO_2 and CO_3^{2-} in NaAl-rich silicate glasses (albitic, jadeitic, nephelinitic compositions) (Dixon & Pan, 1995; Fine & Stolper, 1985; Mysen & Virgo, 1980b, 1980a). C solely dissolves as CO_3^{2-} in Ca and CaMg-rich silicate glasses (diopsidic, sodamelilitic, and akermanitic compositions) (Dixon & Pan, 1995; Fine & Stolper, 1986). ϵCO_3^{2-} data are often limited to either the $\epsilon\text{CO}_{3,1515}^{2-}$ or $\epsilon\text{CO}_{3,1430}^{2-}$ peak, which are also assumed to be approximately equal (Dixon & Pan, 1995). A cumulative ϵCO_3^{2-} is fit with data from $\epsilon\text{CO}_{3,1515}^{2-}$ or $\epsilon\text{CO}_{3,1430}^{2-}$, where available, given the variability in the baseline fitting method.

Uncertainties exist within both the compositional parameter and the quantification of ϵ . Variability in compositional parameters may exist in older experimental studies where microprobe measurements quantify limited oxides. Variability in ϵ can be attributed to variability in the baseline fitting method. $\text{H}_2\text{O}_{\text{m},5200}$ and $\epsilon\text{OH}_{4500}^-$ have been quantified with both linear and Gaussian baselines, resulting in likely differences in the determined absorption coefficient. The inversion applies Newton’s method for determining the minimum of a function in a non-linear inverse problem. The Newtonian inversion derives strength from the ability to incorporate experimental and compositional un-

certainties into the calibration and to utilize information on the shape of the error for a trial solution to successively derive improved fits for model parameters which minimize error. We apply a variant of the inversion described in Chapter 9 of Menke (2018), with code available on GitHub. The implicit theory or model is formulated as:

$$f(x) = -\varepsilon + m_0 + m_1 p = 0, \quad (7)$$

where m is the vector of coefficients to be solved, p is the compositional parameter of interest, and x is matrix:

$$\varepsilon \quad (8)$$

The least squares covariance matrix accounts for uncertainty in ε but not in compositional parameter. We thus move to successively solve for m and c with an implicit Newtonian inversion to account for these uncertainties. Uncertainties in ε and compositional parameter can be applied with a diagonal covariance matrix:

$$c = \quad (9)$$

We successively solve for the m vector, and thus the x matrix with the implicit Newtonian algorithm:

$$x_n = x_i + MO * (F * (x_{n-1} - x)) \quad (10)$$

where x_n is the matrix at iteration n , x_{n-1} is the matrix at the previous iteration, x_i is the initial matrix, f_{n-1} is the model at the previous iteration, and F is the diagonal gradient matrix (matrix of derivatives of implicit theory f with respect to each component of matrix x):

$$F = \quad (11)$$

and MO is the Lagrange multiplier:

$$MO... \quad (12)$$

The x_i column matrix is defined by both the measured data and a simple least squares model, with the simplest least squares solution to the implicit model defined as:

$$m_{ls} = (G^T * G)^{-1} (G * \varepsilon), \quad (13)$$

and the covariance matrix described by:

$$c_{ls} = ((G^T * \sigma_\varepsilon)^{-2} * G)^{-1}, \quad (14)$$

where G is the data matrix from experiment 1 to N :

$$G = something. \quad (15)$$

Two forms of error—the error of calibration and the error of a single application of the thermometer to melt—can be quantified. The 95% confidence interval of the error of calibration is two times the posterior covariance in ε , between the predicted ε of the model (varying within error) and the experimental ε . The 95% confidence interval of a single application of the model incorporates the error of calibration as well as analytical uncertainty of compositional parameter derivation, applied with:

$$c_T = Z^T c_m Z + m^T c_z m, \quad (16)$$

where Z is a horizontal matrix of the measured compositional parameter, c_m is the diagonal posterior covariance on the model parameter coefficients, m is the horizontal matrix of the posterior model parameter coefficients, and c_z is a diagonal matrix of the uncertainty of measured compositional parameter.

We assume an uncertainty of 10% is applied to $\varepsilon\text{CO}_3^{2-}$, $\varepsilon\text{H}_2\text{O}_{t,3550}$, $\varepsilon\text{H}_2\text{O}_{m,1635}$, and $\varepsilon\text{CO}_3^{2-}$ peaks. An uncertainty of 20% is applied to the fits with $\varepsilon\text{OH}_{4500}^-$ given the increased uncertainty on the contrasting linear and Gaussian baseline fits to the peak.

Compositional parameters were assigned a 2.5% uncertainty, based on the propagated uncertainties in calculating cation fractions from the oldest study with reported microprobe uncertainties. We derive best-fit parameters and covariance matrices **add table with parameters here**, evolving from the initial least squares fit to reach a minimum on the error surface 3. The problem is sufficiently overdetermined such that the solution does not depend significantly on the starting model and converges within 10 iterations. We thus are able to derive ϵ with uncertainty, accounting for both model and compositional uncertainty, propagated forward to ultimately derive concentration.

3.3 Calculating Thickness from Reflectance FTIR Spectra

We develop an automated implementation for calculating thicknesses from reflectance spectra by the interference fringe method. Wavelengths of interference fringes are proportional to the thickness and refractive index of the sample for thin silica films (Nishikida et al., 1996; Tamic et al., 2001; Wysoczanski & Tani, 2006; Sun et al., 2007). Nishikida et al. (1996) show the relationship applicable to glasses and olivine:

$$l = \frac{m}{2n(v_1 - v_2)} \quad (17)$$

where l is the thickness of area analyzed, m is the number of fringes in the wavenumber range, n is the refractive index of the material, and v_1 and v_2 are the highest and lowest wavenumbers in the interval. Interference fringes are produced by the interactions between reflected light and reflections internal to the sample, with fringe wavelengths inversely proportional to thickness.

Signal-to-noise ratios associated with interference fringes can be low. Reflectance spectra are preprocessed with median filtering to remove potential single spike noise and successively processed with Savitzky-Golay filtering, fitting low-degree polynomials to data with linear least squares (Savitzky & Golay, 1964). Local extrema associated with each interference fringe wave are located by identifying regions of greatest absorbance change. Maxima are surrounded by points that are lower in absorbance and minima are surrounded by points that are greater in absorbance. Local extrema identification by differentiation can fail with noisier spectra. Thicknesses are calculated from the wavenumber differences between each sequential grouping of local maxima or minima within the inputted wavenumber range.

Mean thicknesses and standard deviations from local maxima and minima are determined and compared against those from digital micrometer to assess the uncertainty associated with this method. Two to eight interference fringes in the wavenumber range of 1700 to 2850 cm^{-1} for Volcán de Fuego olivines. A refractive index of 1.709 was determined with the average Fo_{72} composition for spots adjacent to melt inclusions with the linear, orientation-dependent parameterization of olivine refractive index from Howie et al. (1992)

Increasing the wavenumber range compared to that of Nichols and Wysoczanski (2007) increases the number of high-quality interference fringe waves while reproducing digital micrometer measurements. Average thickness uncertainties associated with five measurements is similar to previously reported values at less than 3 μm (Allison et al., 2019; Nichols & Wysoczanski, 2007; von Aulock et al., 2014). Thickness by the reflectance fringe method was uniformly greater than that by the digital micrometer, likely suggesting that the olivine refractive index calculated from Howie et al. (1992) is too low. Uncertainties in the refractive index may exist, given the lack of parameterization of composition dependence and the lack of recent analyses. **Add stuff on Fuego wafers as proof of concept.**

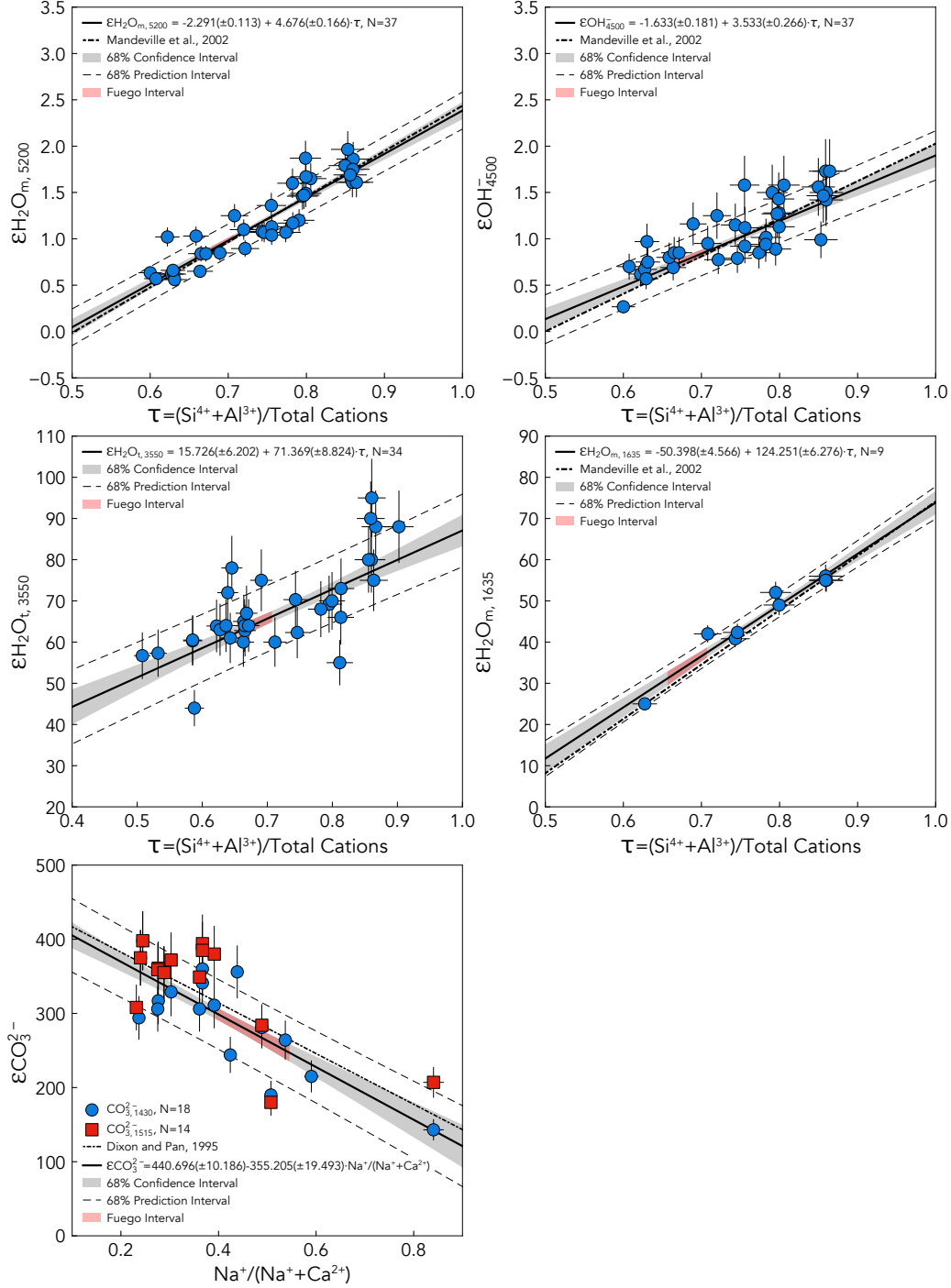


Figure 3: Epsilon regressions

3.4 Calculating Density

Melt density is a function of glass composition, and is calculated from the gram formula weight and partial molar volumes (at ambient temperature and pressure of the analytical conditions) from Lesher and Spera (2015). The large molar volume of H_2O

significantly impacts density and requires the implementation of an iterative solver. Density is initially assumed to have no contribution from H_2O but is iteratively update with the calculated amount of $\text{H}_2\text{O}_{\text{t},3550}$ or $\text{H}_2\text{O}_{\text{m},1635} + \text{OH}_{4500}^-$ if the sample is saturated. The $\epsilon\text{H}_2\text{O}_{\text{t},3550}$ is analyzed to determine saturation, returning an error message when raw absorbance is >2 . The $\text{H}_2\text{O}_{\text{m},1635}$ and OH_{4500}^- peaks are used as total H_2O in determining density when saturated. The density calculation is repeated to ensure stability and converges within 10 iterations.

3.5 Asymmetric Least Squares for $\text{H}_2\text{O}_{\text{m},5200}$, OH_{4500}^- , and $\text{H}_2\text{O}_{\text{t},3550}$ Peak Heights

The $\text{H}_2\text{O}_{\text{t},3550}$, $\text{H}_2\text{O}_{\text{m},5200}$, and OH_{4500}^- baselines are fit with the asymmetric least squares (ALS) method, iteratively solving for an interpolated fitting which balances smoothness and asymmetry (Eilers, 2004; Lee et al., 2017; Peng et al., 2010). **Look up ALS a bit more, define more properly.** Data are preprocessed by first applying a median filter to remove single spike noise. Three repeat baselines, varying the regions of interest surrounding each peak, are fit with ALS to assess uncertainty within peak height and subtracted from each peak. The three baseline-subtracted peaks are kriged to reduce noise and to obtain maximum peak height. **Look up kriging a bit more, define more properly.** The mean peak absorbance and standard deviation in absorbance are determined from the repeat baseline-subtracted peaks. $\text{H}_2\text{O}_{\text{t},3550}$ saturation prompts the return of an error message and suggestion of use of the $\text{H}_2\text{O}_{\text{m},5200}$ and OH_{4500}^- peaks for determining $\text{H}_2\text{O}_{\text{t}}$

3.6 Markov Chain Monte Carlo Algorithm for $\text{H}_2\text{O}_{\text{m},1635}$ and CO_3^{2-} Doublet

The net equation describing the absorbance spectrum of a glass or melt inclusion, with ν^- defined as the wavenumber range from 2400 to 1275 cm^{-1} , is thus expressed in the following form:

$$\begin{aligned}\hat{A} = & x_0\bar{B} + x_1\bar{B}_{PC1} + x_2\bar{B}_{PC2} + x_3\bar{B}_{PC3} + x_4\bar{B}_{PC4} \\ & + y_0\overline{\text{H}_2\text{O}_{\text{m},1635}} + y_1\overline{\text{H}_2\text{O}_{\text{m},1635}_{PC1}} + y_2\overline{\text{H}_2\text{O}_{\text{m},1635}_{PC2}} \\ & + \frac{a_{1515}}{2\sigma_{1515}^2} \cdot e^{-(\nu^- - \mu_{1515})^2} + \frac{a_{1430}}{2\sigma_{1430}^2} \cdot e^{-(\nu^- - \mu_{1430})^2} + m\nu + b\end{aligned}$$

where $x_0, x_1, x_2, x_3, x_4, y_0, y_1, y_2, a_{1515}, a_{1430}, \sigma_{1515}, \sigma_{1430}, \mu_{1515}, \mu_{1430}, m$, and b are the best fit parameters. These fit parameters are first determined by ordinary least squares. The model parameter space of the modeled absorbance of the melt inclusion spectrum is sampled and explored by leveraging a Monte-Carlo Markov Chain algorithm in a Bayesian parametric framework, to account for uncertainty. In this framework of Bayesian inferences, the posterior joint probability distribution of model parameters is quantified as a function of the prior probability of model parameters and a likelihood function. Bayes' Theorem dictates that the posterior joint probability distribution is updated as more information becomes available, following Cubillos et al. (2017):

$$P(\theta|y, M) \propto P(\theta|M)P(y|\theta, M), \quad (18)$$

where y denotes data, θ is the set of parameters, $P(\theta|y, M)$ is the posterior probability distribution of parameters, $P(\theta|M)$ is the prior probability distribution (not incorporating new information), and $P(y|\theta, M)$ is the likelihood. The prior probability distribution of parameters is defined by the expected range and distributions of these parameters. The likelihood serves as the probability density function of the modeled data given an array of parameters. A Markov-Chain Monte Carlo (MCMC) algorithm was utilized to generate random samples from the parameter space to generate and assess the parameter space, where the random samples generated by the algorithm hold a probability density distribution that is proportional to the posterior probability distribution. The open-

source Python package, Multi-Core Markov-Chain Monte Carlo, or MC³, integrates these statistical methods (Cubillos et al., 2017) and was implemented for this study.

Fixed absorbance spectra provide the information required to generate improved posterior probability distributions of model parameters. Prior probability distributions of the parameters generating the baseline, peaks, and lines require different functions – either uniform or Gaussian – to accurately represent this space. Parameters related to the baselines, H_{2,m}O peak, CO₃²⁻ doublet peak heights, and linear adjustment are sampled uniformly, given the greater variation within these parameters that is not defined around one value. The exploration space provided these uniformly sampled priors is guided by the initial model generation process. The CO₃²⁻ doublet peak locations and standard deviations are defined with a Gaussian prior, given the well-defined nature of these parameters in basaltic melt inclusion absorbance spectra. The initial prior parameters are optimized with the least-squares Trust Region Reflective algorithm, which is suitable for well-constrained, large problems (Branch et al., 1999).

Random MCMC sampling is performed with the Snooker Updater Differential Evolution Markov Chain method (DEMC), which involves the parallel evolution of multiple chains and computation of appropriate scales and orientations for the jumping distribution (ter Braak, 2006; ter Braak & Vrugt, 2008). The scale and direction of the proposed jump is computed by the parallel computation of several chains, which allow for the differences between one chain and two additional randomly selected chains to evolve forwards. The chains will converge towards a posterior distribution and will align the evolution along the proper orientation, with good scaling. The Snooker Updater DEMC optimizes acceptance rates and efficiency, allowing for one absorbance spectrum to be run within 15 seconds.

The Bayesian and MCMC framework underlying this model allows for numerous parameters to be sampled and fit concurrently, while accounting for covariance and quantifying uncertainties. The peak heights and standard deviation of the H₂O_{m,1635} and CO₃²⁻ doublet peaks are exported from the algorithm and used to determine volatile concentrations.

3.7 Markov Chain Monte Carlo for Calculating Concentration

The final density obtained following the iteration and appropriate absorption coefficients are applied to the Beer Lambert Law, with the ALS and MC³ determined peak heights, to determine the concentration of total H₂O, molecular H₂O, and CO₂ of each doublet peak.

The final density, the appropriate absorption coefficient, and the ALS and MC³ determined peak heights of absorbance are applied through the Beer Lambert Law to determine the concentration of H₂O_t, H₂O_m, OH⁻, and CO₂ of each doublet peak. A standard deviation on each concentration is taken by running a simple Monte Carlo error assessment in which all parameters (except molar mass) are allowed to vary with a normal distribution for 5x10⁵ samples.

4 Analytical Methodology

4.1 Transmission Fourier Transform Infrared Spectroscopy (FTIR)

Natural glass standards and olivine-hosted melt inclusions (MIs) were analyzed with the Thermo Scientific Nicolet iN10 MX Fourier Transform Infrared (FTIR) Spectrometer at the Lamont-Doherty Earth Observatory. Glass standards and MIs were placed on a CaF₂ plate within the sample holder for measurement of IR absorbance during transmission in the detector spectral range of 8000-450 wavenumbers (cm⁻¹). The FTIR was purged for twenty minutes to decrease the signal of atmospheric H₂O and CO₂. Glasses

and MIs were mapped for absorbance to determine potential heterogeneity in volatile concentrations and to determine the boundaries of measurement — particularly for MI to ensure double intersection. Absorbance maps were generated at a point spacing of $10 \times 10 \mu\text{m}$, with 16 scans taken at each point at a resolution of 16 cm^{-1} . The relatively few scans and low resolution were balanced to yield manageable mapping times of approximately 15 minutes for each MI. Maps guided the selection of optimal regions for volatile analyses. Three repeat measurements were collected for each glass standard or MI, with 256 collection scans at 4 cm^{-1} spectral resolution. Background scans were collected under the same conditions through the CaF_2 plate. Aperture sizes ($30\text{--}200 \mu\text{m}$ for glass standards and $15\text{--}50 \mu\text{m}$ for MIs) were selected to maximize analytical area and to ensure that light propagated solely through the MI without including the host olivine. The Happ-Genzel apodization function within the Thermo-Nicolet OMNIC Picta software was applied to each spectrum to maintain resolution and reduce noise. Three repeat measurements were taken for each MI. Internal standards of submarine arc lava GH88-1-D1010 analyzed by Newman et al. (2000) with FTIR and basaltic MIs CN_C.OL1', CN92C.OL2, and ETF46 analyzed by Barth et al. (2019) and Barth (2021) with SIMS were measured at the beginning and end of each analytical session as internal check standards.

4.2 Reflectance Fourier Transform Infrared Spectroscopy (FTIR)

Three repeat thickness measurements of olivine-hosted melt inclusions were acquired using a Mitutoyo 543-783B Digimatic Indicator, as well as with the reflectance method described by Nichols and Wysoczanski (2007). Two reflectance spectra were taken adjacent to olivine-hosted melt inclusions with 256 collection scans at 4 cm^{-1} spectral resolution, with aperture sizes of $50 \times 50 \mu\text{m}$. Background scans were collected under the same conditions on a highly reflective gold plate with a reflectance coefficient of unity (Nichols & Wysoczanski, 2007). The basaltic glass refractive index is quantified to be 1.546 (Kumagai & Kaneoka, 2003). The rhyolitic glass refractive index is not as well quantified, showing additional variability and spanning the range of $1.48 < n < 1.51$ (Tröger et al., 1959; Tamic et al., 2001). Refractive indices for glasses of variable composition can be calibrated with the digital micrometer (Duncan & Dasgupta, 2015; Tamic et al., 2001). The olivine mean refractive index, averaged across the three crystallographic axes, is described by forsterite content with linear relationships described in Howie et al. (1992).

4.3 Electron Probe Microanalyzer (EPMA)

Major and trace elements in glass and olivine were analyzed by wavelength dispersive X-ray spectroscopy with the Cameca SXFive-TACTIS electron microprobe (EPMA) at the American Museum of Natural History. Calibration was completed with natural and synthetic mineral and oxide standards. Glasses were measured with an accelerating voltage of 15 keV, variable working current of 4 nA, 10 nA, and 40 nA, and spot size of $10 \mu\text{m}$. Duplicate MI analyses were collected proximal to the center of the inclusion. Na was measured with a working current of 4 nA with 2 s counting on peak and 20 s on background to minimize Na loss. Other major and trace elements were measured with a working current of 10 nA with 30 s counting on peak and 15 s on background. Fe was analyzed with 20 s counting on peak and 15 s on background. Volatiles S and Cl were measured with an increased working current of 40 nA. Olivines were measured with an accelerating voltage of 15 keV, variable working current of 10 nA and 300 nA, and spot size of $1 \mu\text{m}$. Duplicate olivine-host analyses were collected approximately $15 \mu\text{m}$ from the MI. Mg, Si, and Fe were analyzed with a working current of 10 nA, and all remaining minor and trace elements were analyzed with a working current of 300 nA. **I can't figure out all the counting times from the files, need a bit of help with this for appendix.** Replicate analyses of the check standard MR:ND-70-01 and secondary standards San Carlos Olivine and VG-2 were performed every 10 analyses. Column con-

ditions and additional assessments of Instrumental drift, precision, and accuracy from analysis of standard analyses are provided in Appendix A.

5 Validation

6 Application

6.1 PyBaselines Package and Structure

PyBaselines can be locally installed with Python versions ≥ 3.6 using the command line:

```
1 pip install PyBaselines
```

and installed with Google Colab and Binder, two cloud-based Python development environments, where installation occurs within a code cell:

```
1 !pip install PyBaselines
```

The PyBaselines package must then be loaded and initialized with:

```
1 import PyBaselines as bl
```

Functions from PyBaselines are then called with the abbreviated package name, followed by a dot, followed by the function name. The function loading all spectrum files is called as follows:

```
1 bl.Load_SampleCSV(args)
```

and the function processing each spectrum file is called as follows:

```
1 bl.Run_All_Spectra(args)
```

6.2 Volcán de Fuego Melt Inclusions

Appendix A Supplement: EPMA Analytical Conditions, Precision, and Accuracy

Analytical precision is quantified as the relative standard deviation (standard deviation of repeat analysis / mean composition of standard during session) of repeated analyses of a secondary standard during an analytical session. Analytical accuracy is quantified as the mean composition of the secondary standard during an analytical session divided by the laboratory value for the standard. Specific acquisition parameters and assessments of analytical precision and accuracy are presented in the following tables.

Acknowledgments

Enter acknowledgments, including your data availability statement, here.

References

- Allison, C. M., Roggensack, K., & Clarke, A. B. (2019). H₂O–CO₂ solubility in alkali-rich mafic magmas: new experiments at mid-crustal pressures. *Contributions to Mineralogy and Petrology*, 174(7), 1–24.
- Barth, A. (2021). *Ascent rates and volatiles of explosive basaltic volcanism* (Unpublished doctoral dissertation). Columbia University in the City of New York.
- Barth, A., Newcombe, M., Plank, T., Gonnermann, H., Hajimirza, S., Soto, G. J., ... Hauri, E. (2019). Magma decompression rate correlates with explosivity

Table A1: Assessment of EPMA analytical conditions, precision, and accuracy for analyzed phases including glass and olivine

Glass							
Element	Spectrometer	Crystal	Peak Location	Counting Time (s)	Calibration Material	Precision (%)	Accuracy (%)
Na	Sp1	LTAP	46415	10	Jadeite	2.7	
Si	Sp4	TAP	27741	10	Diopside	0.5	
K	Sp2	PET	42761	10	KSpK	13.1	
Ca	Sp2	PET	38389	20	Diopside	0.9	
Ti	Sp2	PET	31416	40	Rutile	2.1	
Mg	Sp4	TAP	38515	20	Olivine	1.5	
Al	Sp1	LTAP	32473	30	Corundum	0.6	
P	Sp1	LTAP	23949	30	Apatite	20.0	
Fe	Sp5	LIF	48084	60	Fayalite	0.6	
Mn	Sp3	LLIF	52199	40	MnMnSp	6.9	
Cr	Sp3	LLIF	56860	20	CrCrSp3	94	
Olivine							
Element	Spectrometer	Crystal	Peak Location	Counting Time (s)	Calibration Material	Precision (%)	Accuracy (%)
C1							
Si	Sp4	TAP	27741	10	Diopside	0.4	
Fe	Sp3	LLIF	48084	10	Fayalite	1.0	
Mg	Sp1	LTAP	38530	10	Olivine	0.2	
C2							
Ca	Sp2	PET	38389	60	Diopside	4.4	
Ti	Sp2	PET	31416	90	Rutile	36.5	
Cr	Sp5	LIF	56866	60	CrCrSp3	10.6	
Mn	Sp5	LIF	52205	90	MnMnSp3	3.1	
Ni	Sp3	LLIF	41167	150	NiONiSp3	0.4	
Al	Sp1	LTAP	32473	150	Corundum	4.2	
P	Sp4	TAP	23958	150	Apatite	148	

- at basaltic volcanoes—Constraints from water diffusion in olivine. *Journal of Volcanology and Geothermal Research*, 387, 106664.
- Blundy, J., & Cashman, K. (2001). Ascent-driven crystallisation of dacite magmas at Mount St Helens, 1980–1986. *Contributions to Mineralogy and Petrology*, 140(6), 631–650.
- Branch, M. A., Coleman, T. F., & Li, Y. (1999). A Subspace, Interior, and Conjugate Gradient Method for Large-Scale Bound-Constrained Minimization Problems. *SIAM Journal on Scientific Computing*, 21(1), 1–23.
- Carvajal, R. C., Arias, L. E., Garces, H. O., & Sbarbaro, D. G. (2016). Comparative analysis of a principal component analysis-based and an artificial neural network-based method for baseline removal. *Applied Spectroscopy*, 70(4), 604–617.
- Cubillos, P., Harrington, J., Lored, T. J., Lust, N. B., Blecic, J., & Stemm, M. (2017). On Correlated-noise Analyses Applied to Exoplanet Light Curves. *The Astronomical Journal*, 153(1), 3.
- Dixon, J. E., & Clague, D. A. (2001). Volatiles in basaltic glasses from loihi seamount, hawaii: Evidence for a relatively dry plume component. *Journal of Petrology*, 42(3), 627–654.
- Dixon, J. E., & Pan, V. (1995). Determination of the molar absorptivity of dissolved carbonate in basaltic glass. *American Mineralogist*, 80(11-12), 1339–1342.
- Dixon, J. E., Stolper, E., & Delaney, J. R. (1988). Infrared spectroscopic measurements of CO₂ and H₂O in Juan de Fuca Ridge basaltic glasses. *Earth and Planetary Science Letters*, 90(1), 87–104.
- Dixon, J. E., & Stolper, E. M. (1995). An experimental study of water and carbon dioxide solubilities in mid-ocean ridge basaltic liquids. Part II: applications to degassing. *Journal of Petrology*, 36(6), 1633–1646.
- Dixon, J. E., Stolper, E. M., & Holloway, J. R. (1995). An experimental study of water and carbon dioxide solubilities in mid-ocean ridge basaltic liquids. Part I: calibration and solubility models. *Journal of Petrology*, 36(6), 1607–1631.
- Duncan, M. S., & Dasgupta, R. (2015). Pressure and temperature dependence of CO₂ solubility in hydrous rhyolitic melt: implications for carbon transfer to mantle source of volcanic arcs via partial melt of subducting crustal lithologies. *Contributions to Mineralogy and Petrology*, 169(6), 1–19.
- Eilers, P. H. (2004). Parametric time warping. *Analytical Chemistry*, 76(2), 404–411.
- Fine, G., & Stolper, E. (1985). The speciation of carbon dioxide in sodium aluminosilicate glasses. *Contributions to Mineralogy and Petrology*, 91(2), 105–121.
- Fine, G., & Stolper, E. (1986). Dissolved carbon dioxide in basaltic glasses: concentrations and speciation. *Earth and Planetary Science Letters*, 76(3-4), 263–278.
- Hess, K., & Dingwell, D. B. (1996). Viscosities of hydrous leucogranitic melts: A non-Arrhenian model. *American Mineralogist: Journal of Earth and Planetary Materials*, 81(9-10), 1297–1300.
- Howie, R., Zussman, J., & Deer, W. (1992). *An introduction to the rock-forming minerals*. Longman London, UK.
- Kumagai, H., & Kaneoka, I. (2003). Relationship between submarine MORB glass textures and atmospheric component of MORBs. *Chemical Geology*, 200(1-2), 1–24.
- Lee, L. C., Liong, C., Khairul, O., & Jemain, A. A. (2017). Effects of baseline correction algorithms on forensic classification of paper based on ATR-FTIR spectrum and principal component analysis (PCA). *Pertanika Journal of Science and Technology*, 25(3), 767–774.
- Leshner, C. E., & Spera, F. J. (2015). Thermodynamic and transport properties of silicate melts and magma. In *The encyclopedia of volcanoes* (pp. 113–141). Elsevier.

- Mandeville, C. W., Webster, J. D., Rutherford, M. J., Taylor, B. E., Timbal, A., & Faure, K. (2002). Determination of molar absorptivities for infrared absorption bands of H₂O in andesitic glasses. *American Mineralogist*, 87(7), 813–821.
- McIntosh, I. M., Nichols, A. R., Tani, K., & Llewellyn, E. W. (2017). Accounting for the species-dependence of the 3500 cm⁻¹ H₂O_i infrared molar absorptivity coefficient: Implications for hydrated volcanic glasses. *American Mineralogist: Journal of Earth and Planetary Materials*, 102(8), 1677–1689.
- Menke, W. (2018). *Geophysical data analysis: Discrete inverse theory*. Academic Press.
- Mercier, M., Di Muro, A., Métrich, N., Giordano, D., Belhadj, O., & Mandeville, C. W. (2010). Spectroscopic analysis (FTIR, Raman) of water in mafic and intermediate glasses and glass inclusions. *Geochimica et Cosmochimica Acta*, 74(19), 5641–5656.
- Mysen, B. O., & Richet, P. (2018). *Silicate glasses and melts*. Elsevier.
- Mysen, B. O., & Virgo, D. (1980a). The solubility behavior of CO₂ in melts on the join NaAlSi₃O₈-CaAl₂Si₂O₈-CO₂ at high pressures and temperatures: a Raman spectroscopic study. *American Mineralogist*, 65(11-12), 1166–1175.
- Mysen, B. O., & Virgo, D. (1980b). Solubility mechanisms of carbon dioxide in silicate melts: a Raman spectroscopic study. *American Mineralogist*, 65(9-10), 885–899.
- Newman, S., Stolper, E., & Stern, R. (2000). H₂O and CO₂ in magmas from the Mariana arc and back arc systems. *Geochemistry, Geophysics, Geosystems*, 1(5).
- Newman, S., Stolper, E. M., & Epstein, S. (1986). Measurement of water in rhyolitic glasses; calibration of an infrared spectroscopic technique. *American Mineralogist*, 71(11-12), 1527–1541.
- Nichols, A. R., & Wysoczanski, R. (2007). Using micro-FTIR spectroscopy to measure volatile contents in small and unexposed inclusions hosted in olivine crystals. *Chemical Geology*, 242(3-4), 371–384.
- Nishikida, K., Nishio, E., & Hannah, R. W. (1996). *Selected applications of modern FTIR techniques*. CRC Press.
- Ochs III, F. A., & Lange, R. A. (1999). The density of hydrous magmatic liquids. *Science*, 283(5406), 1314–1317.
- Okumura, S., Nakamura, M., & Nakashima, S. (2003). Determination of molar absorptivity of IR fundamental OH-stretching vibration in rhyolitic glasses. *American Mineralogist*, 88(11-12), 1657–1662.
- Pandya, N., Muenow, D. W., & Sharma, S. K. (1992). The effect of bulk composition on the speciation of water in submarine volcanic glasses. *Geochimica et Cosmochimica Acta*, 56(5), 1875–1883.
- Peng, J., Peng, S., Jiang, A., Wei, J., Li, C., & Tan, J. (2010). Asymmetric least squares for multiple spectra baseline correction. *Analytica Chimica Acta*, 683(1), 63–68.
- Rasmussen, D. J. (2019). *Vapor bubble growth in olivine-hosted melt inclusions* (Unpublished doctoral dissertation). Columbia University in the City of New York.
- Savitzky, A., & Golay, M. J. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8), 1627–1639.
- Silver, L., & Stolper, E. (1989). Water in albitic glasses. *Journal of Petrology*, 30(3), 667–709.
- Sisson, T., & Grove, T. (1993). Experimental investigations of the role of H₂O in calc-alkaline differentiation and subduction zone magmatism. *Contributions to mineralogy and petrology*, 113(2), 143–166.
- Stolper, E. (1982). Water in silicate glasses: an infrared spectroscopic study. *Contributions to Mineralogy and Petrology*, 81(1), 1–17.

- Stolper, E. (1989). Temperature dependence of the speciation of water in rhyolitic melts and glasses. *American Mineralogist*, 74(11-12), 1247–1257.
- Sun, W., Binns, R., Fan, A., Kamenetsky, V. S., Wysoczanski, R., Wei, G., ... Arculus, R. (2007). Chlorine in submarine volcanic glasses from the eastern Manus basin. *Geochimica et Cosmochimica Acta*, 71(6), 1542–1552.
- Tamic, N., Behrens, H., & Holtz, F. (2001). The solubility of H₂O and CO₂ in rhyolitic melts in equilibrium with a mixed CO₂–H₂O fluid phase. *Chemical Geology*, 174(1-3), 333–347.
- ter Braak, C. J. (2006). A Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution: easy Bayesian computing for real parameter spaces. *Statistics and Computing*, 16, 239–249.
- ter Braak, C. J., & Vrugt, J. A. (2008). Differential Evolution Markov Chain with snooker updater and fewer chains. *Statistics and Computing*, 18(435–446).
- Tröger, W. E., Bambauer, H. U., Taborszky, F., & Trochim, H.-D. (1959). *Optische bestimmung der gesteinsbildenden minerale-teil i.: Bestimmungstabellen*. Schweizerbart'sche Verlagsbuchhandlung.
- von Aulock, F. W., Kennedy, B. M., Schipper, C. I., Castro, J., Martin, D., Oze, C., ... others (2014). Advances in Fourier transform infrared spectroscopy of natural glasses: From sample preparation to data analysis. *Lithos*, 206, 52–64.
- Wysoczanski, R., & Tani, K. (2006). Spectroscopic FTIR imaging of water species in silicic volcanic glasses and melt inclusions: An example from the Izu-Bonin arc. *Journal of Volcanology and Geothermal Research*, 156(3-4), 302–314.
- Xue, X. (2009). Water speciation in hydrous silicate and aluminosilicate glasses: Direct evidence from ²⁹Si-1H and ²⁷Al-1H double-resonance NMR. *American Mineralogist*, 94(2-3), 395–398.