

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

Alphabetical Ordering for Now: Anna Barth<sup>1,2</sup>, Terry Plank<sup>1</sup>, Dan  
Rasmussen<sup>1,3</sup>, Sarah Shi<sup>1,4</sup>, Henry Towbin<sup>1</sup>

<sup>1</sup>Lamont-Doherty Earth Observatory, Columbia University, New York, NY USA

<sup>2</sup>University of California, Berkeley, Berkeley, CA USA

<sup>3</sup>National Museum of Natural History, Smithsonian Institution, Washington, DC USA

<sup>4</sup>University of Cambridge ? Necessary, I don't know.

## Key Points:

- 1
- 2

---

Corresponding author: Sarah Shi, [scs2202@columbia.edu](mailto:scs2202@columbia.edu)

Corresponding author: Henry Towbin, [wht2106@columbia.edu](mailto:wht2106@columbia.edu)

## Abstract

## 1 Introduction

The variability and subjectivity in Fourier Transform Infrared Spectroscopy (FTIR) baseline fitting techniques constitutes a major uncertainty in the determination of volatile concentrations. We develop quantifiable and reproducible approaches to defining baselines beneath each  $\text{H}_2\text{O}$  and  $\text{CO}_2$  species peak to determine concentrations with meaningful uncertainties, unlike most FTIR studies, where baselines are fit by eye, spline, instrument software, or unspecified means.

The  $\epsilon_{\text{H}_2\text{O}_{\text{t},3550}}$  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  $\epsilon_{\text{H}_2\text{O}_{\text{t},3550}}$  is saturated, the combination of  $\text{H}_2\text{O}_{\text{m},5200}$  or the  $\text{H}_2\text{O}_{\text{m},1635}$  and the  $\text{OH}_{4500}^-$  must be used instead. The  $\text{H}_2\text{O}_{\text{m},5200}$  and  $\text{OH}_{4500}^-$  peaks can be fit with a linear to near-linear baseline, or a baseline consisting of two Gaussians (Ohlhorst et al., 2001; Stabile et al., 2020; Stolper, 1982; Withers & Behrens, 1999). The linear to near-linear baseline is advantageous given the reproducibility of peak heights but can underestimate the peak height or area given the presence of iron and water-related bands ( $\sim 5700$  and  $\sim 4000$   $\text{cm}^{-1}$ , respectively) in dacitic to basaltic compositions (Ohlhorst et al., 2001). The species concentration of  $\text{H}_2\text{O}_{\text{m},5200}$  and  $\text{OH}_{4500}^-$  by the two Gaussian method agrees well with NMR spectroscopy but is highly sensitive to raw absorbance and can be difficult to reproduce across studies (Ohlhorst et al., 2001). The  $\epsilon_{\text{H}_2\text{O}_{\text{t},3550}}$ ,  $\text{H}_2\text{O}_{\text{m},5200}$ , and  $\text{OH}_{4500}^-$  baselines are thus 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).

The  $\text{H}_2\text{O}_{\text{m},1635}$  peak and  $\text{CO}_3^{2-}$  doublet peaks pose greater challenges due to the steep and sharp increase of baseline absorbance at wavenumbers lower than  $1430$   $\text{cm}^{-1}$ . Furthermore, the convolution of the tails of the  $\text{H}_2\text{O}_{\text{m},1635}$  and the  $\text{CO}_3^{2-}$  1515 doublet peak in melt inclusion spectra further complicates the appropriate baselines in the near infrared region. Baselines beneath the  $\text{H}_2\text{O}_{\text{m},1635}$  and  $\text{CO}_3^{2-}$  doublet peaks have been approximated with the spectra of devolatilized samples with similar chemical compositions (Dixon et al., 1988; Newman et al., 2000), splines and flexicures (Dixon & Stolper, 1995), and other curve functions (Dixon & Clague, 2001). We assess the fundamental shape and variability of the baseline using principal component analysis (PCA) applied to a database of absorbance spectra for 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 to a wide range of arc and MORB compositions. 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 develop and present a method for fitting baselines and peaks with uncertainties for the  $\text{H}_2\text{OT}$ ,  $\text{H}_2\text{O}_{\text{m},5200}$ , and  $\text{OH}_{4500}^-$  peaks with repeat asymmetric least squares baselines and the  $\text{H}_2\text{O}_{\text{m},1635}$  peak and the  $\text{CO}_3^{2-}$  doublet peaks with Bayesian statistics and a Python Multi-Core Markov Chain Monte Carlo algorithm.

## 2 Analytical Methodology

### 2.1 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. Transmission FTIR spectroscopy mea-

sures the amount of light absorbed at each wavelength, allowing for the recognition and quantification of volatile 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 ( $\text{g}\cdot\text{mol}^{-1}$ ),  $\epsilon$  is the absorptivity of the species ( $\text{L}\cdot\text{mol}^{-1}\cdot\text{cm}$ ),  $l$  is the optical path length or thickness ( $\text{cm}$ ), and  $\rho$  is density ( $\text{kg}\cdot\text{m}^{-3}$ ). Linear relationships between absorbance and concentration dominate. 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  $\text{H}_2\text{O}_{\text{t},3550}$  (McIntosh et al., 2017; von Aulock et al., 2014).

Glass standards and MIs were placed on a  $\text{CaF}_2$  plate within the sample holder for measurement of IR absorbance during transmission in the detector spectral range of 8000-450 wavenumbers ( $\text{cm}^{-1}$ ). The FTIR was purged for twenty minutes to decrease the signal of atmospheric  $\text{H}_2\text{O}$  and  $\text{CO}_2$ . 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  $16\ \text{cm}^{-1}$ . 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  $4\ \text{cm}^{-1}$  spectral resolution. Background scans were collected under the same conditions through the  $\text{CaF}_2$  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.

## 2.2 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). 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:

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

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. Two reflectance spectra were taken adjacent to olivine-hosted melt inclusions with 256 collection scans at  $4\ \text{cm}^{-1}$  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 (Tamic et al., 2001). The olivine mean refractive index, averaged across the three crystallographic axes, was determined with the average Fo<sub>72</sub> composition for spots adjacent to melt inclusions using the linear relationships described in Howie et al. (1992).

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

## 3 PyBaselines Structure and Computational Methodology

### 3.1 PyBaselines Structure

PyBaselines can be locally installed with Python versions  $\geq 3.6$  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)
```

### 3.2 Calculating Thickness from Reflectance FTIR Spectra

We develop an automated implementation for calculating thicknesses from reflectance spectra by the interference fringe method with the function:

```
bl.ThicknessProcessing(dfs_dict, n, wn_high, wn_low, peak_height_min_delta, peak_search_width)
```

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  $\text{cm}^{-1}$  for Volcán de Fuego olivines. 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  $\mu\text{m}$  (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.

### 3.3 Calibrating Molar Absorptivity ( $\epsilon$ )

Molar absorptivities or extinction coefficients ( $\epsilon$ ) determine the intensity at which light attenuates when passing through a material and relates absorbance to concentration, calibrated by pairing analyses of volatile content by transmission FTIR and by an additional independent analytical technique.  $\epsilon$  is dependent on the composition of glass, related to both tetrahedral cation fraction of  $\tau = (\text{Si}^{4+} + \text{Al}^{3+}) / (\text{total cations})$  and cation fraction of  $\text{Na}^+ / (\text{Na}^+ + \text{Ca}^{2+})$ . We refit these extinction coefficients to incorporate all quantifications of extinction coefficients since the last compilation of extinction coefficients by Mandeville et al. (2002).

The extinction coefficients related to  $\text{H}_2\text{O}$  species increase with tetrahedral cation fraction,  $\tau$ , reflecting increased polymerization of the glass (Stolper, 1982). Tetrahedral cations have also been proposed to compete with non-tetrahedral cations ( $\text{M} = \text{Mg}^{2+}$ ,  $\text{Ca}^{2+}$ ,  $\text{Na}^+$ ) 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 rates will drive the formation of variable proportions of  $\text{H}_2\text{O}_{\text{m}}$  and  $\text{OH}^-$  (Silver & Stolper, 1989; Stolper, 1989).  $\epsilon_{\text{H}_2\text{O}_{\text{t},3550}}$  has thus been considered as a combination of two end-member 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  $\tau = 0.746\text{--}0.800$ , 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.

Dixon and Pan (1995) first demonstrate the dependence of the  $\text{CO}_3^{2-}$  doublet absorption coefficients on the molar or cationic ratio of  $\text{Na}^+ / (\text{Na}^+ + \text{Ca}^{2+})$ . The proportion of  $\text{Na}^+$  to  $\text{Ca}^{2+}$  modulates the dissolution of C as molecular  $\text{CO}_2$  and/or  $\text{CO}_3^{2-}$ . As the proportion of  $\text{Ca}^{2+}$  increases,  $\epsilon_{\text{CO}_3^{2-}}$  decreases. C dissolves as both molecular  $\text{CO}_2$  and  $\text{CO}_3^{2-}$  in NaAl-rich silicate glasses (albitic, jadeitic, nephelinitic compositions) (Dixon & Pan, 1995; Fine & Stolper, 1985; Mysen & Virgo, 1980b, 1980a). C solely dissolves as  $\text{CO}_3^{2-}$  in Ca and CaMg-rich silicate glasses (diopsidic, sodamelilitic, and akermanitic compositions) (Dixon & Pan, 1995; Fine & Stolper, 1986).

We refit the linear relationships between  $\epsilon_{\text{CO}_3^{2-}}$ ,  $\epsilon_{\text{OH}_{4500}^-}$ ,  $\epsilon_{\text{H}_2\text{O}_{\text{t},3550}}$ ,  $\epsilon_{\text{H}_2\text{O}_{\text{m},1635}}$ , and  $\epsilon_{\text{CO}_3^{2-}}$  and the corresponding compositional parameters of  $\tau$  or  $\text{Na}^+ / (\text{Na}^+ + \text{Ca}^{2+})$  values to include more recent data and broader compositional ranges, with an implicit inverse method based on that described in Chapter 9 of Menke (2018).

Uncertainties in  $\epsilon$ , due to variability in independent volatile measurements, and compositional parameters, related to older microprobe measurements with incomplete oxide measurements, can thus be accounted for within the model. The  $\epsilon$  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,  $\tau$  spans the range of 0.50–0.90 and  $\text{Na}^+ / (\text{Na}^+ + \text{Ca}^{2+})$  from 0.23–0.84, encompassing that for the Fuego 2018 MIs ( $\tau = 0.66\text{--}0.71$  and  $\text{Na}^+ / (\text{Na}^+ + \text{Ca}^{2+}) = 0.39\text{--}0.56$ ). If  $\epsilon_{\text{CO}_3^{2-}}$  or  $\epsilon_{\text{OH}_{4500}^-}$  are quantified with both a linear and two Gaussian baseline, we average the two estimates. Although not typically quantified, we find a positive relationship between  $\epsilon_{\text{H}_2\text{O}_{\text{t},3550}}$  and  $\tau$ , despite complications presented by the speciation (Supplemental Figure xx). 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).

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 fit with  $\tau$ .  $\epsilon_{\text{CO}_3^{2-}}$  data are often limited to either the  $\epsilon_{\text{CO}_3^{2-},1515}$  or  $\epsilon_{\text{CO}_3^{2-},1430}$  peak, which are also assumed to be approximately equal (Dixon & Pan, 1995). A cumulative  $\epsilon_{\text{CO}_3^{2-}}$  is fit with data from  $\epsilon_{\text{CO}_3^{2-},1515}$  or  $\epsilon_{\text{CO}_3^{2-},1430}$ , where available, given the variability in the baseline fitting method. This regression is appropriate for the compositional range of calibration ( $\text{Na}^+ / (\text{Na}^+ + \text{Ca}^{2+}) = 0.23\text{--}0.84$ ). The behavior of the absorption coefficient at the  $\text{Na}^+$  and  $\text{Ca}^{2+}$  endmembers is complex and these compositions are not considered in this study. An uncertainty of 10% is applied to  $\epsilon_{\text{CO}_3^{2-}}$ ,  $\epsilon_{\text{H}_2\text{O}_{\text{t},3550}}$ ,  $\epsilon_{\text{H}_2\text{O}_{\text{m},1635}}$ , and  $\epsilon_{\text{CO}_3^{2-}}$  peaks. An uncertainty of 20% is applied to the fits with  $\epsilon_{\text{OH}_{4500}^-}$  given the increased uncertainty on the contrasting linear and Gaussian baseline fits to the peak. Uncertainties in the extinction coefficient fit, in combination with those for the microprobe session, are propagated forward to the posterior covariance on model parameter coefficients.

### 3.4 Calculating Density

Melt density is a function of the composition of the melt inclusion, 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  $\text{H}_2\text{O}$  in arc basaltic melt inclusions strongly impacts density and requires the implementation of an iterative solver. The initial solution for density is assumed to have no contribution from  $\text{H}_2\text{O}$  but is updated 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 density converges within 10 it-

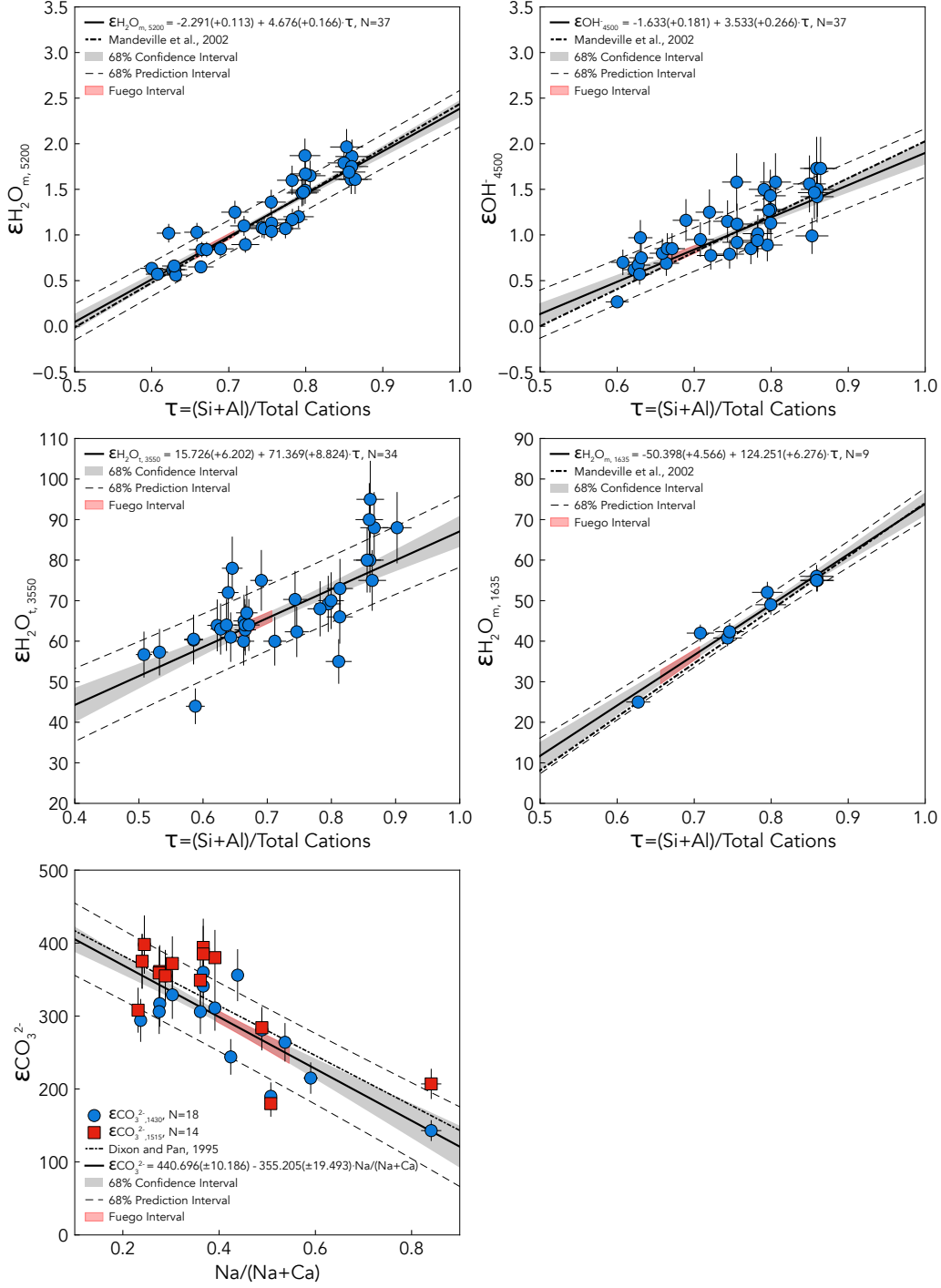


Figure 1: Epsilon regressions

erations. The final density, the appropriate absorption coefficient, and the ALS and MC<sup>3</sup> determined peak heights of absorbance are applied through the Beer Lambert Law to determine the concentration of H<sub>2</sub>O<sub>t</sub>, H<sub>2</sub>O<sub>m</sub>, OH<sup>-</sup>, and CO<sub>2</sub> of each doublet peak. A standard deviation on each concentration is taken by running a simple Monte Carlo er-



ror assessment in which all parameters (except molar mass) are allowed to vary with a normal distribution for  $5 \times 10^5$  samples.

### 3.5 Assessing Baseline Variability

#### Henry's section

### 3.6 Monte Carlo Framework

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

where  $x_1, x_2, x_3, y_1, y_2, a_{1515}, a_{1430}, 1515, 1430, 1515, 1430, m$ , and  $b$  are the best fit parameters. These optimum 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):

$$P(\theta|y, M) \propto P(\theta|M)P(y|\theta, M), \quad (3)$$

where  $y$  denotes data,  $\theta$  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 MC<sup>3</sup>, 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, molecular  $\text{H}_2\text{O}$  peak,  $\text{CO}_3^{2-}$  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  $\text{CO}_3^{2-}$  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 melt inclusion 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 credibly quantifying uncertainties.

## Acknowledgments

Enter acknowledgments, including your data availability statement, here.

## 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.

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	

## References

- Allison, C. M., Roggensack, K., & Clarke, A. B. (2019). H<sub>2</sub>O–CO<sub>2</sub> 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 at basaltic volcanoes—Constraints from water diffusion in olivine. *Journal of Volcanology and Geothermal Research*, 387, 106664.
- 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 CO<sub>2</sub> and H<sub>2</sub>O 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.
- 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.
- 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 H<sub>2</sub>O 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<sup>-1</sup> H<sub>2</sub>O<sub>t</sub> 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 CO<sub>2</sub> in melts on the join NaAlSi<sub>3</sub>O<sub>8</sub>-CaAl<sub>2</sub>Si<sub>2</sub>O<sub>8</sub>-CO<sub>2</sub> 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). H<sub>2</sub>O and CO<sub>2</sub> 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.
- Ohlhorst, S., Behrens, H., & Holtz, F. (2001). Compositional dependence of molar absorptivities of near-infrared OH-and H<sub>2</sub>O bands in rhyolitic to basaltic glasses. *Chemical Geology*, 174(1-3), 5–20.
- 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.
- Stabile, P., Appiah, E., Bello, M., Giuli, G., Paris, E., & Carroll, M. R. (2020). New IR spectroscopic data for determination of water abundances in hydrous pantelleritic glasses. *American Mineralogist: Journal of Earth and Planetary Materials*, 105(7), 1060–1068.
- Stolper, E. (1982). Water in silicate glasses: an infrared spectroscopic study. *Contributions to Mineralogy and Petrology*, 81(1), 1–17.

- 426 Stolper, E. (1989). Temperature dependence of the speciation of water in rhyolitic  
427 melts and glasses. *American Mineralogist*, 74(11-12), 1247–1257.
- 428 Sun, W., Binns, R., Fan, A., Kamenetsky, V. S., Wysoczanski, R., Wei, G., ... Ar-  
429 culus, R. (2007). Chlorine in submarine volcanic glasses from the eastern  
430 Manus basin. *Geochimica et Cosmochimica Acta*, 71(6), 1542–1552.
- 431 Tamic, N., Behrens, H., & Holtz, F. (2001). The solubility of H<sub>2</sub>O and CO<sub>2</sub> in rhy-  
432 olitic melts in equilibrium with a mixed CO<sub>2</sub>–H<sub>2</sub>O fluid phase. *Chemical Geol-*  
433 *ogy*, 174(1-3), 333–347.
- 434 ter Braak, C. J. (2006). A Markov Chain Monte Carlo version of the genetic al-  
435 gorithm Differential Evolution: easy Bayesian computing for real parameter  
436 spaces. *Statistics and Computing*, 16, 239-249.
- 437 ter Braak, C. J., & Vrugt, J. A. (2008). Differential Evolution Markov Chain with  
438 snooker updater and fewer chains. *Statistics and Computing*, 18(435-446).
- 439 Tröger, W. E., Bambauer, H. U., Taborszky, F., & Trochim, H.-D. (1959). *Optis-*  
440 *che bestimmung der gesteinsbildenden minerale-teil i.: Bestimmungstabellen.*  
441 Schweizerbart'sche Verlagsbuchhandlung.
- 442 von Aulock, F. W., Kennedy, B. M., Schipper, C. I., Castro, J., Martin, D., Oze, C.,  
443 ... others (2014). Advances in Fourier transform infrared spectroscopy of  
444 natural glasses: From sample preparation to data analysisadvances in Fourier  
445 transform infrared spectroscopy of natural glasses: From sample preparation to  
446 data analysis. *Lithos*, 206, 52–64.
- 447 Withers, A. C., & Behrens, H. (1999). Temperature-induced changes in the NIR  
448 spectra of hydrous albitic and rhyolitic glasses between 300 and 100 K. *Physics*  
449 *and Chemistry of Minerals*, 27(2), 119–132.
- 450 Wysoczanski, R., & Tani, K. (2006). Spectroscopic FTIR imaging of water species  
451 in silicic volcanic glasses and melt inclusions: An example from the Izu-Bonin  
452 arc. *Journal of Volcanology and Geothermal Research*, 156(3-4), 302–314.
- 453 Xue, X. (2009). Water speciation in hydrous silicate and aluminosilicate glasses:  
454 Direct evidence from 29Si-1H and 27Al-1H double-resonance NMR. *American*  
455 *Mineralogist*, 94(2-3), 395–398.