OXFORD

Journal of Petrology, 2023, **64**, 1–25
https://doi.org/10.1093/petrology/egad050
Advance access publication date 4 July 2023
Original Manuscript

Barometers Behaving Badly II: a Critical Evaluation of Cpx-Only and Cpx-Liq Thermobarometry in Variably-Hydrous Arc Magmas

Penny E. Wieser[1,2,*], Adam J.R. Kent[2] and Christy B. Till[3]

[1]Department of Earth and Planetary Sciences, McCone Hall, UC Berkeley, Berkeley 94720, USA
[2]College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis 97331, USA
[3]School of Earth and Space Exploration, Arizona State University, Tempe, AZ 85281, USA

*Corresponding author. E-mail: Penny_wieser@berkeley.edu.

Abstract

The chemistry of erupted clinopyroxene crystals (±equilibrium liquids) have been widely used to deduce the pressures and temperatures of magma storage in volcanic arcs. However, the large number of different equations parameterizing the relationship between mineral and melt compositions and intensive variables such as pressure and temperature yield vastly different results, with implications for our interpretation of magma storage conditions. We use a new test dataset composed of the average Clinopyroxene-Liquid (Cpx-Liq) compositions from N = 543 variably hydrous experiments at crustal conditions (1 bar to 17 kbar) to assess the performance of different thermobarometers and identify the most accurate and precise expressions for application to subduction zone magmas. First, we assess different equilibrium tests, finding that comparing the measured and predicted Enstatite-Ferrosillite and K_D (using Fe_t in both phases) are the most useful tests in arc magmas, whereas CaTs, CaTi and Jd tests have limited utility. We then apply further quality filters based on cation sums (3.95–4.05), number of analyses (N > 5) and the presence of reported H_2O data in the quenched experimental glass (hereafter 'liquid') to obtain a filtered dataset (N = 214). We use this filtered dataset to compare calculated versus experimental pressures and temperatures for different combinations of thermobarometers. A number of Cpx-Liq thermometers perform very well when liquid H_2O contents are known, although the Cpx composition contributes little to the calculated temperature relative to the liquid composition. Most Cpx-only thermometers perform very badly, greatly overestimating temperatures for hydrous experiments. These two findings demonstrate that the Cpx chemistry alone holds very little temperature information in hydrous systems.

Most Cpx-Liq and Cpx-only barometers show similar performance to one another (mostly yielding root mean square errors [RMSEs] of 2–3.5 kbar), although the best Cpx-only barometers currently outperform the best Cpx-Liq barometers. We also assess the sensitivity of different equations to melt H_2O contents, which are poorly constrained in many natural systems. Overall, this work demonstrates that Cpx-based barometry on individual Cpx only provides sufficient resolution to distinguish broad storage regions in continental arcs (e.g. upper, mid, lower crust). Significant averaging of Cpx compositions from experiments reported at similar pressures can reduce RMSEs to ~1.3–1.9 kbar. We hope our findings motivate the substantial amount of experimental and analytical work that is required to obtain precise and accurate estimates of magma storage depths from Cpx ± Liq equilibrium in volcanic arcs.

Keywords: magma storage conditions, arc magmatism, subduction zones, thermobarometry, clinopyroxene

INTRODUCTION

The compositions of erupted clinopyroxene (Cpx) and Cpx-liquid (Liq) pairs are commonly used to calculate pressures (P) and temperatures (T) in a variety of igneous systems. Cpx is a stable phase over a very wide range of pressures, temperatures, melt compositions and oxygen fugacities (e.g. Costa, 2004; Nandedkar *et al.*, 2014; Ulmer *et al.*, 2018; Brugman & Till, 2019). This means that Cpx-based thermobarometry has broad utility, and so has been applied in a wide variety of tectonic settings. However, recent work has also shown that there are limitations in our current state of knowledge about Cpx-based barometry and thermometry (Wieser *et al.*, 2022a). Specifically, low count times and/or beam currents used for analysis of elements with concentrations <0.5 wt % yield highly imprecise measurements (e.g. 1σ errors of 10–40% for Na_2O). Combined with a small number of measurements per experimental charge, this obscures the true composition of experimental Cpx. This low analytical precision causes large errors (±2–3 kbar) in Cpx barometry when tested using global experimental datasets (Wieser *et al.*, 2023).

In this contribution, we specifically focus on evaluating sources of uncertainty associated with the application of Cpx-based barometry to arc magmas, where Cpx is a common mineral phase in both mafic and more evolved lavas (so has been used by a number of different studies determining storage conditions, e.g. Auer *et al.*, 2013; Belousov *et al.*, 2021; Cassidy *et al.*, 2015; Caulfield *et al.*, 2012; Cigolini *et al.*, 2018; Dahren *et al.*, 2012; Deegan *et al.*, 2016; Freundt & Kutterolf, 2019; Geiger *et al.*, 2018; Hollyday *et al.*, 2020; Jeffery *et al.*, 2013; Lai *et al.*, 2018; Lormand *et al.*, 2021; Moussallam *et al.*, 2021, 2019; Namur *et al.*, 2020; Preece *et al.*, 2014; Romero *et al.*, 2022;

Received: December 12, 2022. **Revised:** June 21, 2023. **Accepted:** June 27, 2023

Table 1: Compilation of different thermometers and barometers used in the literature since 2008 to place constraints on magma storage conditions at arc volcanoes

Clinopyroxene-Liquid barometry	
P = Putirka (2008) eq31, T = Putirka (2008) eq33	**P = Neave & Putirka (2017), T = Putirka (2008) eq33**
• Mt Baker and Glacier Peak, Cascades - Sas *et al.* (2017)	• Ebeko Volcano, Kurile arc - Belousov *et al.* (2021)
• Whangaeuhu Gorge, New Zealand - Auer *et al.* (2013)	• Lassen Peak, Cascades - Hollyday *et al.* (2020)
	• Lassen Peak, Cascades - *Scruggs & Putirka (2018)*
	• Taupo Volcanic Zone - Lormand *et al.* (2021)
	• Calbuco Volcano - Namur *et al.* (2020)
P = Putirka (2008) eq30, T = Putirka (2008) eq33	**P = Putirka *et al.* (2003), T = Putirka *et al.* (2003)**
• Agung and Batur, Indonesia - Geiger *et al.* (2018)	• Agung and Batur, Indonesia - Geiger *et al.* (2018)
• Ambae, Vanuatu - Moussallam *et al.* (2019)	• Ambrym, Vanuatu - Sheehan & Barclay (2016)
• Ambrym, Vanuatu - Moussallam *et al.* (2021)	• Soufrière Hills, Monseratt - Cassidy *et al.* (2015)
• Ambrym, Vanuatu - Sheehan & Barclay (2016)	• Krakatau, Indonesia - Dahren *et al.* (2012)
• Villarrica, Chile – Romero *et al.* (2022)	• Tofua Volcano, Tonga - Caulfield *et al.* (2012)
P = Putirka (2008) eq32c, T = . . .	
T = Putirka *et al.* (1996) eqT2 (spreadsheet default):	**T = Putirka (2008) eq33:**
• Mayon Volcano, Phillipines - Ruth & Costa (2021)	Chiltepe, Nicaragua - Freundt & Kutterolf (2019)
Thermometer not stated in paper:	**T = Putirka *et al.* (2003)**
• Miravalles-Guayabo Caldera, Costa Rica - Cigolini *et al.* (2018)	• Agung and Batur, Indonesia - Geiger *et al.* (2018)
• Mt Baker and Glacier Peak, Cascades - Sas *et al.* (2017)	• Merapi, Indonesia - Preece *et al.* (2014)
	• Krakatau, Indonesia - Dahren *et al.* (2012)
Cpx-only Barometry	
P = Putirka *et al.* (2003) eq32b, T = . . .	**P = Putirka (2008) eq32a, T = . . .**
Thermometer not stated in paper:	**Thermometer not stated in paper:**
• Agung and Batur, Indonesia - Geiger *et al.* (2018)	• Taupo Volcanic Zone - Beier *et al.* (2017)
• Merapi Volcano, Indonesia - Deegan *et al.* (2016)	**T from Putirka *et al.* (2003):**
• Ambae, Vanuatu - Moussallam *et al.* (2019)	• Volcan Melimoyu, Andes. Geoffroy *et al.* (2018)
• Merapi, Indonesia - Preece *et al.* (2014)	**T from Putirka (2008) eq32d:**
T = Putirka (2008) eq32d:	• Ambrym, Vanuatu - Sheehan & Barclay (2016)
• Kelut Volcano, Indonesia - Jeffery *et al.* (2013)	• Ambrym, Vanuatu - Moussallam *et al.* (2021)
T = Putirka (2008) eq33:	**T from Putirka (2008) eq33:**
Chiltepe Volcanic Complex, Nicaragua - *Freundt & Kutterolf (2019)*	• Mariana trough back-arc basin - Lai *et al.* (2018)
T from Putirka *et al.* (1996):	• Okinawa Trough - Chen *et al.* (2021)
• Etna, Italy - Ubide & Kamber (2018)	

Ruth & Costa, 2021; Sas *et al.*, 2017; Scruggs & Putirka, 2018; Sheehan & Barclay, 2016).

Existing expressions relating P and/or T to Cpx(±Liq) compositions are generally calibrated on the products of experiments conducted at known conditions, using a wide variety of equations based on multilinear regressions (e.g. Putirka, 2008, Neave & Putirka, 2017) or machine learning techniques using decision trees (e.g. Petrelli *et al.*, 2020, Jorgenson *et al.*, 2022). A number of different Cpx-Liq and Cpx-only parameterizations exist (Nimis, 1999; Putirka, 1999, 2008; Neave & Putirka, 2017; Petrelli *et al.*, 2020; Wang *et al.*, 2021; Jorgenson *et al.*, 2022). This diversity of choice means that it is not clear which equation is best to use, or how much this choice affects the geological interpretation. This is particularly true in the variably hydrous magma compositions found in volcanic arcs because there has been no detailed evaluation of which thermobarometers behave best in this setting. This is in contrast to extensive work evaluating thermobarometers in more H_2O-poor tectonic settings (e.g. Iceland, Neave *et al.*, 2019; Neave & Putirka, 2017) and alkaline volcanic systems (Mollo *et al.*, 2013; Masotta *et al.*, 2016). An additional problem in more hydrous settings is that many existing equations do not contain a term for the H_2O content of the liquid, and even if such a term is included, the underlying calibration datasets contain a significant number of experiments where H_2O contents in experimental glasses are either not measured or are not reported (convoluting the calibration of such terms, Wieser *et al.*, 2023).

The lack of consensus as to which thermobarometers are best is demonstrated by the breadth of equations selected by studies performing Cpx ± Liq thermobarometry in arc magmas after the publication of the seminal thermobarometry review of Putirka (2008, Table 1). Because many barometers contain a term for temperature, in natural systems where neither pressure not temperature is known, studies tend to iteratively calculate pressures and temperature using a thermometer and a barometer. Iteration of two equations greatly increases the number of possible equation combinations (to $N_{barometers}$ X $N_{thermometers}$).

Iteration of P and T from Putirka *et al.* (2003, hereafter P2003) has remained a popular choice even in recent years, despite the fact that more up-to-date recalibrations of these equations were provided in Putirka (2008, hereafter P2008). Eq32c from P2008 is another popular barometer and has been iterated with a wide variety of different thermometers (Table 1). Another common choice is the iteration of P2008 eq30 or eq31 for pressure with eq33 for temperature. Alternatively, P2008 eq33 (T) has been iterated with the Neave & Putirka (2017, hereafter NP17) barometer. This choice is particularly interesting given that Neave & Putirka (2017) caution that their barometer may not be applicable to the more hydrous and oxidizing conditions found in volcanic arcs.

Cpx-only thermobarometry has also been used for arc systems, but has been less popular than Cpx-Liq. Most studies have used the two Cpx-only barometers from Putirka (2008, eq32a for

H_2O-independent, eq32b for H_2O-dependent) iterated with a wide variety of different temperature estimates from Cpx-only and Cpx-Liq thermometers (Table 1). Three new Cpx-only thermobarometers have recently been published (Petrelli *et al.*, 2020; Wang *et al.*, 2021; Jorgenson *et al.*, 2022), which will likely increase the use of Cpx-only methods in a wide variety of tectonic settings, including volcanic arcs. Thus, it is important to evaluate their performance.

Comparison of existing thermobarometers

The diversity of published equations being used in the literature for Cpx-based thermobarometry is concerning because these equations can give different results for individual Cpx and Cpx-Liq pairs. To demonstrate the magnitude of these differences, we calculate pressures for 10 experiments from Blatter *et al.* (2017) performed at 7 kbar (Fig. 1a) and four experiments from Blatter *et al.* (2013) performed at 4 kbar (Fig. 1b) using the different combinations of equations highlighted in Table 1 along with other recently published thermobarometers (see also Supporting Fig. 1). We show error bars with the published root mean square error (RMSE) for each thermobarometer centred around the mean calculated P and T.

Iteration of Neave & Putirka (2017) with P2008 eq33 (blue squares) and iteration of the PT expressions from Putirka (2003, orange squares) would suggest storage at ∼ −2–12 km, whereas iteration of P2008 eq32c with Putirka *et al.* (1996) eqT2 (burgundy stars, Fig. 1a), would indicate storage at ∼35–60 km. There is also no overlap for calculated pressures using the Cpx-only thermobarometers of Jorgenson *et al.* (2022, white circles) and P2008 eq32b–32d (purple circles). Similar discrepancies also exist for the experiments conducted at 4 kbar from Blatter *et al.* (2013), Fig. 1b). The Cpx-Liq thermobarometers of Petrelli *et al.* (2020, cyan circles) and Jorgenson *et al.* (2022, white circles) suggest crystallization at ∼10–20 km (mid crust), whereas iteration of P2008 eq32c for pressure with P2003 for temperature yields pressures in the lower crust (∼30–40 km, brown diamonds), and the default iteration of eq32c in the P2008 spreadsheet indicates storage at >40-km depth (brown stars, Fig. 1a). To assess differences in calculated temperatures, we select an experiment performed at ∼995 °C and 4 kbar (Fig. 1c). Calculated Cpx-only temperatures from different equations span a very wide range (∼200 °C), whereas most Cpx-Liq temperatures lie within ∼50–100 °C of each other and the experimental temperature (Fig. 1c).

Despite these large discrepancies in calculated P and T using different equations, and obvious implications for geological interpretation, only a small proportion of studies applying Cpx-based barometers to natural systems have performed calculations using more than one thermobarometer (e.g.Erdmann *et al.*, 2016; Sheehan & Barclay, 2016; Sas *et al.*, 2017; Geiger *et al.*, 2018). There is also a general lack of justification in the literature for why a specific equation was chosen. In many cases, the stated error statistics from the paper presenting the thermobarometers are quoted as the rationale. For example, some studies seem to select their thermobarometers based on a smaller quoted RMSE (or standard error estimate, SEE) from the original publication (e.g. Dahren *et al.*, 2012; Preece *et al.*, 2014). However, the way in which RMSE is calculated for these different equations is highly variable, so these statistics are not directly comparable. For example, Putirka *et al.* (2003) state an RMSE of ±1.7 kbar in their abstract based on the model fit to the calibration dataset (four studies, N = 77 experiments). Similarly, Putirka (2008) states an RMSE of ±1.5 kbar for eq32c based on the calibration dataset (four studies, N = 99 experiments). These are the RMSEs quoted by Dahren *et al.* (2012) and Preece *et al.* (2014) to justify their use of these barometers. However, when Putirka (2008) applied these equations to all available experimental data (n = 1303), eq32c had an RMSE of ±5 kbar and Putirka *et al.* (2003) has an RMSE of ±5 kbar for n = 324 hydrous experiments (and ±4.8 kbar for 848 anhydrous experiments). Similarly, the RMSE = ±1.4 kbar commonly quoted by studies using the Neave & Putirka (2017) barometer reflects the fit to the calibration dataset (n = 113), whereas the error on a global regression is ±3.6–3.8 kbar.

Assessing uncertainty using only the calibration data can greatly underestimate the true error because the model has been tuned to those experiments. It is far more statistically robust to assess error using experiments that were not used during thermobarometer calibration (often termed a 'test dataset'), especially when such datasets share important compositional features with target natural systems. Studies that state the more realistic errors associated with a test dataset in their abstract (e.g. Petrelli *et al.*, 2020) may be less widely used simply because they have quoted a larger error, even though this error is more realistic.

Another issue associated with comparing published statistics and taking these as representative of the true error in natural systems is the variable pressure range of calibration and test datasets. For example, Petrelli *et al.* (2020) compute statistics for their test dataset using experiments conducted at 0–30 kbar, Putirka *et al.* (2003) from 0–35 kbar and Neave & Putirka (2017) from 0–20 kbar. However, it is uncommon that Cpx from the higher end of these pressure ranges are encountered when examining products from arc volcanoes. Using the compilation of crustal thicknesses from Profeta *et al.* (2016), all arc segments have Moho depths <45 km (∼12 kbar), apart from the Northern and Central Volcanic Zone in Chile which have Moho depths >50 km (∼14–17 kbar). For the test dataset provided with the Cpx-only barometer of Petrelli *et al.* (2020), if experiments are restricted to those performed at 0–15 kbar, the R^2 value drops from 0.92 to 0.59.

Another factor that can affect the statistics presented for thermobarometers is that most thermometers have a pressure term, and most barometers have a temperature term. When assessing equation performance, most papers input the experimental temperature when calculating pressure, or the experimental pressure when calculating temperature. However, in natural systems, it is most common that neither pressure nor temperature is known, so a thermometer and a barometer must be selected and iteratively solved (Table 1). This will increase the error compared with comparisons using experimentally constrained pressures and temperatures (see Neave & Putirka, 2017).

Additional uncertainties when thermobarometers are applied to natural systems stem from the fact that many equations have a term for the melt H_2O content. Statistics calculated using iterative P–T calculations on experiments with known H_2O contents will underestimate the uncertainty associated with application to natural systems with uncertain H_2O contents. In this study, we investigate the sensitivity of different equations to H_2O to better constrain this often-neglected source of uncertainty.

To summarize—to get a realistic estimate of the errors associated with thermobarometry when applied to natural systems, we should be assessing performance using experimental datasets that were not used during calibration, restricting comparisons to the pressure range of interest, iterating P&T, and propagating uncertainty in melt H_2O content. Comparing calculated P and T from different equations for the samples of interest is also vital to assess systematic errors associated with the choice of

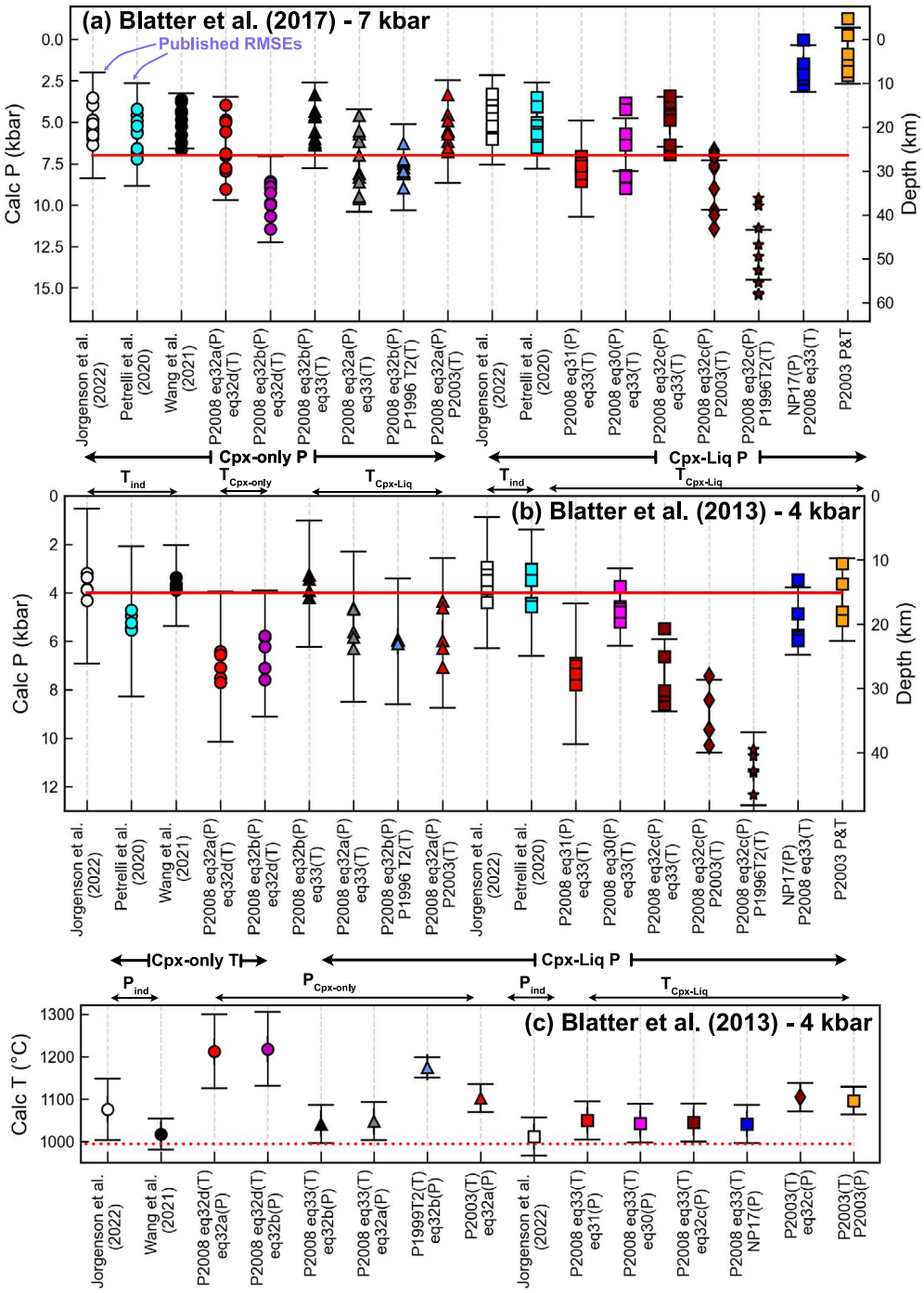

Fig. 1. Comparison of calculated P and T using the various combinations of thermometers and barometers summarized in Table 1. a) Pressure calculations performed for the 10 experiments of Blatter *et al.* (2017) performed at 7 kbar. b) Pressure calculations for the 4 experiments of Blatter *et al.* (2013) performed at 4 kbar. c) Temperatures calculated for experiment no. 2358 conducted at 995 °C, 5.5 wt % H2O, 9 kbar from Blatter *et al.* (2013). Error bars are plotted at the average calculated P, and calculated T, showing the quoted RMSE from each equation. See Supporting Fig. 1 for an additional comparison.

thermobarometry equation(s). Unless one calibration can be robustly selected as the 'best' for a given system, the range of P and T from different calibrations may be representative of the true uncertainty in calculated PT conditions. The best equation for a given system may be identified by compiling experiments with similar compositions to the system of interest and assessing which thermobarometry equations best reproduce the experimental values (e.g. Hammer *et al.*, 2016; Neave & Putirka, 2017). Alternatively, if no suitable experimental data exist, insight may be gained by comparing the dataset used to calibrate each thermobarometry equation against the natural compositions of interest to evaluate the degree of extrapolation required (e.g. Wieser *et al.*, 2022b; Wieser *et al.*, 2022c).

This discussion demonstrates that there is a clear need for igneous petrologists to be able to directly compare the suitability, accuracy and precision of different thermobarometric expressions in arc magmas (and other tectonic settings). To address this, we compile a new experimental dataset of variably hydrous,

tholeiitic to calc-alkaline compositions ranging from basalts to rhyolites. We make sure that none of the data we used to test each model were used during its calibration.

METHODS

ArcPL: A new test dataset for variably hydrous arc magmas

Our new test dataset mostly comprises experiments published since 2008, when the Library of Experimental Phase Relationship (LEPR) dataset used to calibrate most existing thermobarometers was formally compiled and made available to the community (Hirschmann *et al.*, 2008). We also compile a handful of experiments that were conducted before 2008 but not included in the original LEPR compilation (Berndt, 2004; Sisson *et al.*, 2005). The full list of studies is as follows: Almeev *et al.*, 2013; Andújar *et al.*, 2015; Berndt, 2004; Blatter *et al.*, 2023, 2017, 2013; Bogaerts *et al.*, 2006; Cadoux *et al.*, 2014; Costa, 2004; Erdmann & Koepke, 2016; Erdmann *et al.*, 2016; Feig *et al.*, 2010; Firth *et al.*, 2019; Hamada & Fujii, 2008; Husen *et al.*, 2016; Koepke *et al.*, 2018; Krawczynski *et al.*, 2012; Mandler *et al.*, 2014; Marxer *et al.*, 2022; Melekhova *et al.*, 2015; Nakatani *et al.*, 2022; Nandedkar *et al.*, 2014; Neave *et al.*, 2019, Parat *et al.*, 2014; Parman *et al.*, 2011; Pichavant & Macdonald, 2007; Rader & Larsen, 2013; Riker *et al.*, 2015; Rutherford *et al.*, 1985; Sisson *et al.*, 2005; Solaro *et al.*, 2019; Ulmer *et al.*, 2018; Waters *et al.*, 2021. The liquid compositions in these studies have been normalized in different ways; many studies analyzing glasses with high H_2O contents have reported oxides renormalized to 100% on an anhydrous basis, whereas others have reported analysed totals. For consistency, we normalize all glass analyses to have an anhydrous total of 100%.

For other parts of the discussion (e.g. assessing values of equilibrium tests), we also consider experiments conducted on compositions relevant to arc magmas that are present in LEPR (Baker & Eggler, 1987; Bartels *et al.*, 1991; Draper & Johnston, 1992; Kawamoto, 1996; Gaetani & Grove, 1998; Moore & Carmichael, 1998; Martel *et al.*, 1999; Berndt *et al.*, 2001; Blatter & Carmichael, 2001; Grove *et al.*, 2003, 1997, 1982; Hesse & Grove, 2003; Barclay, 2004; Di Carlo, 2006; Feig *et al.*, 2006; Mercer & Johnston, 2008). We refer to these experiments as ArcLEPR and to our newly compiled dataset as ArcPL (post-LEPR).

Our ArcPL dataset contains 543 Cpx-Liq pairs. There is a small amount of overlap with the training datasets of the most recent models (Wang *et al.*, 2021; Jorgenson *et al.*, 2022). These overlapping experiments are not used to test these specific equations. We restrict comparisons with experiments conducted at 0–17 kbar based on the crustal thickness compilation of Profeta *et al.* (2016), assuming a crustal density of 2700 kg/m^3). The compositional, pressure (P) and temperature (T) range of this dataset is shown in Fig. 2. Although 66% of Cpx-containing experiments in LEPR do not have compiled H_2O contents for experimental glasses, we endeavour to compile as much glass H_2O data as possible for our new dataset. In cases where glass H_2O data were not reported but the experiment was said to be volatile saturated, we calculate dissolved H_2O using the quoted experimental P, T and the fluid composition if given (X_{H2O}) using the solubility model MagmaSat (Ghiorso & Gualda, 2015; implemented in VESIcal; Iacovino *et al.*, 2021). MagmaSat has been shown to provide the best fit to volatile solubility experiments relevant to arc magmas (Wieser *et al.*, 2022c). Overall, only 22% of our ArcPL dataset is missing H_2O data; these experiments are not considered when assessing thermobarometers (see Section 2.2).

Equilibrium tests

One of the main issues when performing Cpx-Liq thermobarometry in natural systems is selecting which measured Cpx compositions to pair with which liquid compositions. In arc (and other tectonic settings), it is difficult to identify and sample erupted liquids that were in equilibrium with a given Cpx. A number of different literature studies have taken the approach of compiling all available whole-rock and glass data from the volcanic system and considering all possible matches between each Cpx and Liq composition, discarding pairs that fall outside preferred ranges of several equilibrium tests (Scruggs & Putirka, 2018; Neave *et al.*, 2019; Gleeson *et al.*, 2021). This approach, although popular, is strongly reliant on having reliable tests by which to assess Cpx-Liq equilibrium. Equilibrium filters have also been applied to experimental datasets when calibrating thermobarometers (Neave & Putirka, 2017). However, a variety of equilibrium tests and cut-off values have been proposed, and it is not always clear what criteria should be used. Thus, we start by using the ArcPL dataset to evaluate commonly used filters.

The most widely used equilibrium test assesses partitioning of Fe-Mg between Cpx and liquid ($K_{D,Fe-Mg}^{Cpx-Liq}$ abbreviated as K_D). Putirka (2008) calibrate an expression (eq35) using LEPR experiments to calculate K_D solely as a function of temperature:

$$K_D = e^{-0.107 - \frac{1719}{T\,(K)}}.$$

There is ambiguity in the thermobarometry literature as to the best way to compute the measured value of K_D from Electron probe microanalyzer (EPMA) measurements of Fe-Mg in the Cpx and Liq. Some studies perform the calculation using only Fe^{2+} in the liquid and Fe_T in the Cpx (e.g. Neave & Putirka, 2017, Gleeson *et al.*, 2021), whereas others use the total amount of Fe (Fe_T) in the liquid (Putirka, 2016). It is important to work out which approach works better before discarding specific Cpx-Liq pairs, particularly in experiments and natural samples from arcs, which are generally quite rich in Fe^{3+} (Carmichael, 1991; Kelley & Cottrell, 2009).

We calculate the proportion of Fe^{2+} in each experiment using the experimental fO_2, with the equations of Kress & Carmichael (1988) implemented in Thermobar (an open-source Python-based thermobarometry tool, Wieser *et al.*, 2022b, Fig. 3). If fO_2 is not given, we calculate it from the quoted buffer position, experimental pressure and temperature. For completeness of our assessments, we also calculate Fe^{2+} in the Cpx using the method of Lindsley (1983), acknowledging that stoichiometric methods estimating Fe^{2+} in minerals are associated with large errors.

K_D values are significantly closer to predicted values from Putirka (2008) when Fe_T in both the liquid and Cpx are used (Fig. 3a). When using Fe^{2+} in the liquid and Fe_T in the Cpx (red dots, Fig. 3b), many more experiments lie outside the ±0.08 error window around the predicted value for Putirka (2008), particularly for experiments with >30% Fe^{3+} (Fig. 3d). Using Fe^{2+} in the Liq and Cpx (white squares, Fig. 3b) also results in more experiments failing the equilibrium test. The superior performance using Fe_T is perhaps unsurprising, given that eq35 of Putirka (2008) was calibrated this way. Thus, we suggest that until equilibrium tests are recalibrated on a substantial volume of experimental data with well-constrained Fe^{2+} proportions in liquid (and Cpx), it is best to use Fe_T in both phases for consistency with the calibration of eq35. Using Fe^{2+} in the liquid could lead to more oxidized experiments (or natural samples) being discarded incorrectly (Fig. 3d). It is also interesting that the ±0.03 filter used by many authors

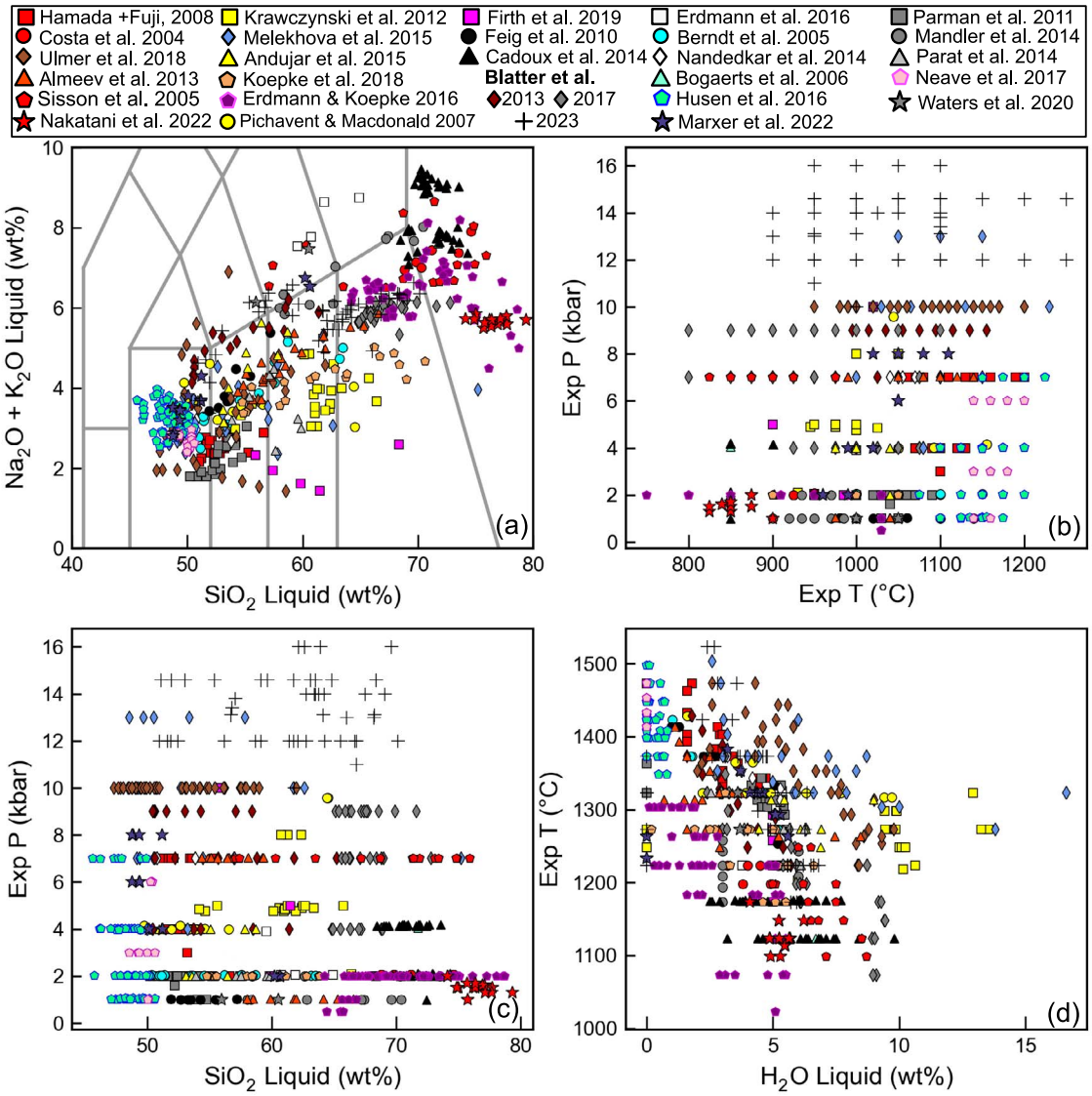

Fig. 2. Compositional and P–T spread for the ArcPL dataset before the application of filters for Cpx-Liq equilibrium, cation sums, number of analyses and melt H_2O contents.

(Scruggs & Putirka, 2018; Neave *et al.*, 2019) would exclude a large amount of experimental data (66%), whereas the ±0.08 value from Putirka (2008) only results in 23% of data being discarded (±0.03 marked by dotted lines, ±0.08 by the grey box on Fig. 3). Although deviation from the predicted equilibrium values may represent true disequilibrium in experiments, to retain a reasonably sized dataset, we proceed with the following comparisons using only experiments with measured K_D values that are within ±0.08 of predicted values calculated from eq35 of Putirka (2008).

Wood & Blundy (1997) also propose an expression (their eq 35):

$$K_D = 0.109 + \frac{0.186}{Mg\#Cpx}$$

We find that this performs very poorly, with the offset between calculated and predicted K_D correlating with Cpx Mg# (Supporting Fig. 2).

Comparing the predicted and measured value of the Calcium-Tschermak's (CaTs) component is another popular equilibrium test, with most studies calculating the predicted CaTs value using the expression of Putirka (1999 *eq3.4*, e.g., Gleeson *et al.*, 2021; Neave *et al.*, 2019). In the ArcPL dataset, there is a very poor correspondence between predicted and measured CaTs values (Fig. 4a). This is also true for experiments from LEPR conducted on compositions relevant to arc magmas (ArcLEPR, red crosses, Fig. 4a). The discrepancy between predicted and measured values is most apparent at higher measured values of CaTs (Fig. 4a), and correlates most strongly with the Al^{VI} content of the Cpx (Fig. 4b). Al^{VI} in Cpx is one of the key parameters used to calculate the CaTs component (along with the anhydrous cation fraction of Na, X_{Na}):

$$CaTs = Al^{VI} - X_{Na}.$$

This means that it is not possible to resolve this offset simply by adding in a term for Al^{VI} in the expression for predicting this component from the liquid component (because this would make it a useless equilibrium test). Clearly, the terms for liquid components, pressure and temperature used in Putirka (1999) eq3.4 to predict the CaTs component are insufficient to account

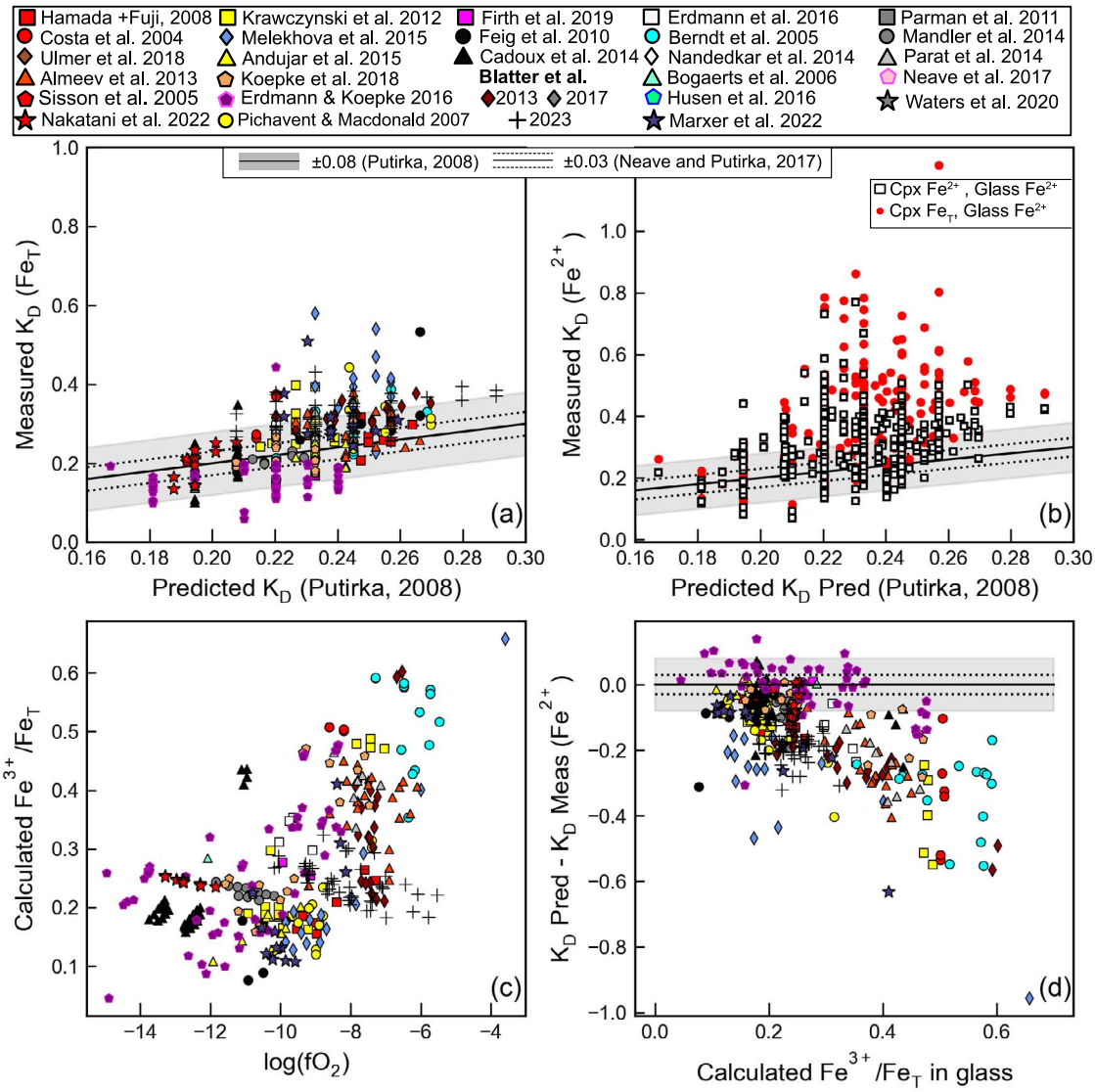

Fig. 3. Assessment of Fe-Mg partitioning between Cpx and Liq (K_D) in the ArcPL dataset. a–b) Predicted values calculated using Putirka (2008) eq35 versus measured values. In a), Fe_T is used in Cpx and Glass to calculate the measured value. In b) the red dots show calculations using Fe^{2+} in the liquid calculated using Kress & Carmichael (1988) from the quoted experimental fO_2 or redox buffer (see part c), and Fe_T in the Cpx. Black squares show Fe^{2+} in the liquid and Fe^{2+} in the Cpx calculated using Lindsley (1983). d) The discrepancy between calculated and predicted KD values using Fe^{2+} in the liquid and Fe_T in the Cpx increases with increasing proportion of Fe^{3+}. Dashed lines show the ±0.03 value used for equilibrium tests by Neave *et al.* (2019), whereas the grey field shows the ±0.08 value suggested by Putirka (2008).

for variation in CaTs in experimental Cpx in hydrous experiments. Although the RMSE = ±0.07 value of Putirka (1999) would result in most experimental pairs being 'in equilibrium', use of the ±0.03 filter of Neave *et al.* (2019) would cause a significant number of experimental pairs to be discarded. Given the poor correlation between predicted and measured values, and the correlation of the discrepancy with Al^{VI}, we suggest that CaTs in its current state is not a useful equilibrium test when working with arc magmas.

The measured Enstatite-Ferrosillite (EnFs) component shows good agreement with the predicted component using eq6 of Mollo *et al.* (2013) for ArcLEPR (Fig. 5a) and ArcPL (Fig. 5b). Relatively few values lie outside the ±0.05 error window. In contrast, measured Diopside-Hedenbergite (DiHd) values show poor agreement with predicted values using eq7 of Mollo *et al.* (2013) for lower temperature experiments, particularly for the ArcPL dataset (Fig. 5c–d). The discrepancy is even worse if DiHd is predicted using Putirka (1999, eq3.1a). The overprediction of calculated DiHd values at low temperatures seems to result from the high T-sensitivity of the Mollo *et al.* (2013) equation at these temperatures. To demonstrate this, we calculate the predicted DiHd values for temperatures of 750–1400 °C using 20 randomly-selected Cpx-Liq pairs. Below ~1000 °C, the predicted value rapidly kicks up to higher values (Fig. 5e, red lines). When the measured values for these 20 pairs are subtracted from the predicted value, the resulting curves recreate the trend to higher values seen in the whole dataset, indicating that this strong temperature dependency is the cause of the discrepancy (Fig. 5f). The equilibrium tests of Mollo *et al.* (2013) were calibrated using LEPR, which contains relatively few experiments at these low temperatures (white squares, Fig. 5f). The lower temperatures of our dataset of arc magmas relative to their calibration range likely result from the higher H_2O contents. These results suggest that care should be taken when applying a DiHd filter to Cpx-Liq pairs in arcs that may have crystallized <900–1000 °C, and that this expression likely needs recalibrating with a dataset of lower temperature experiments.

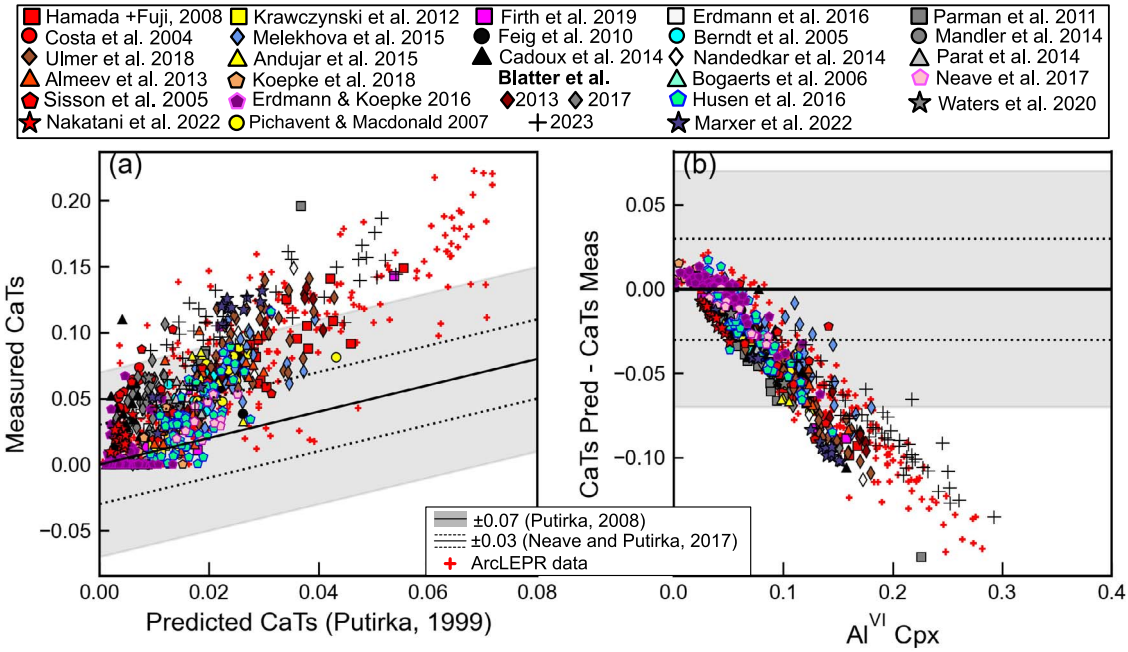

Fig. 4. a) Comparison of predicted and measured values of CaTs. The grey colored box shows the ±0.07 error window suggested by Putirka (1999), and the dashed lines show the ±0.03 value used by Neave *et al.* (2019) to filter natural Cpx-Liq pairs. b) There is a strong correlation between the discrepancy and the Al^{VI} component of Cpx

Finally, lesser used equilibrium tests are also of questionable utility. Only two Cpx-Liq pairs in ArcPL have a measured CaTi component outside the 1σ range of Putirka (1999 eq3.6, Fig. 6a), but there is a poor correlation between predicted and measured values (indicating that it is not a useful test). There is a similarly poor correspondence between predicted and calculated Jd components (Putirka, 1999, eq3.5), with the discrepancy correlating as a function of the Na_2O content of the Cpx (the main component used to calculate Jd, Fig. 6b–c). As for CaTs, this means that recalibration without including terms for the Cpx composition is unlikely to be successful.

Overall, these comparisons demonstrate that EnFs, DiHd (for Cpx crystallized at >1000 °C) and K_D calculated using Fe_T with a filter of ±0.08 rather than ±0.03 are currently the most robust tests of equilibrium when assessing possible Cpx-Liq pairs in arc magmas. Because of the wide range of temperatures in our compiled dataset, we filter our ArcPL dataset to include only experiments within ±0.08 of the predicted value for K_D (using Fe_T) and within ±0.05 for EnFs. We also only include Cpxs with [Ca/(Ca + Mg + Fe) atomic] between 0.2 and 0.5 (excluding pigeonites), and cation sums between 3.95 and 4.05. To help alleviate random scatter associated with analytical imprecision, we only use experiments that measured at least 5 Cpx in each experimental charge (see Wieser *et al.*, 2022a). Because many of the thermobarometers assessed here contain a term for H_2O, we also only consider experiments with some form of reported H_2O contents (e.g. SIMS, FTIR or Raman measurements, volatiles-by-difference or enough information to calculate H_2O using a solubility model). Of the compiled N = 543 new experimental charges, N = 123 fail the K_D filter, N = 71 fail the EnFs filter, N = 20 fail the cation sums filter, N = 156 fail based on having <5 Cpx analyses and N = 53 are discarded based on having no reported H_2O data (see Supporting Fig. 3). Obviously, some experiments fail multiple criteria. Overall, we are left with N = 214 experimental charges. We use these experiments to assess the best performing thermobarometers in arc magmas.

Statistical metrics used in this paper

To assess thermobarometer performance, we calculate the statistics for a linear regression between experimental P or T (x) and calculated P or T (y). We use five statistical metrics associated with this regression to assess the performance of the equation: the correlation coefficient (R^2), the gradient and intercept of the regression, the RMSE and the mean absolute error (MAE), where:

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

$$MAE = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)$$

The MAE doesn't have a squared term like the RMSE, such that it can more easily identify systematic uncertainty. The gradient and intercept of the regression also helps to identify systematic uncertainty.

DISCUSSION

Assessing Cpx-Liq thermobarometers

When estimating temperature from Cpx-Liq equilibrium, iteration of the temperatures calculated using eq33 of P2008 with a variety of different barometers (P2008 eq30, P2008 eq32c and NP17) do a good job of reproducing experimental temperatures in the ArcPL dataset (Fig. 7a-c). The best fit is obtained from iteration of P2008 eq33 with Neave & Putirka (2017), Fig. 7a), returning an R^2 value of 0.92, a gradient close to 1 (0.94), and an intercept of ~93 °C. This iteration also has the lowest RMSE (31.8 °C) and MAE (13.6 °C).

Iteration of the thermometer and barometer of Putirka *et al.* (2003) substantially overestimates temperature (Fig. 7d, MAE = 79 °C, RMSE = 90.5 °C), and the discrepancy is correlated with the melt H_2O content (Supporting Fig. 4a). This is unsurprising given

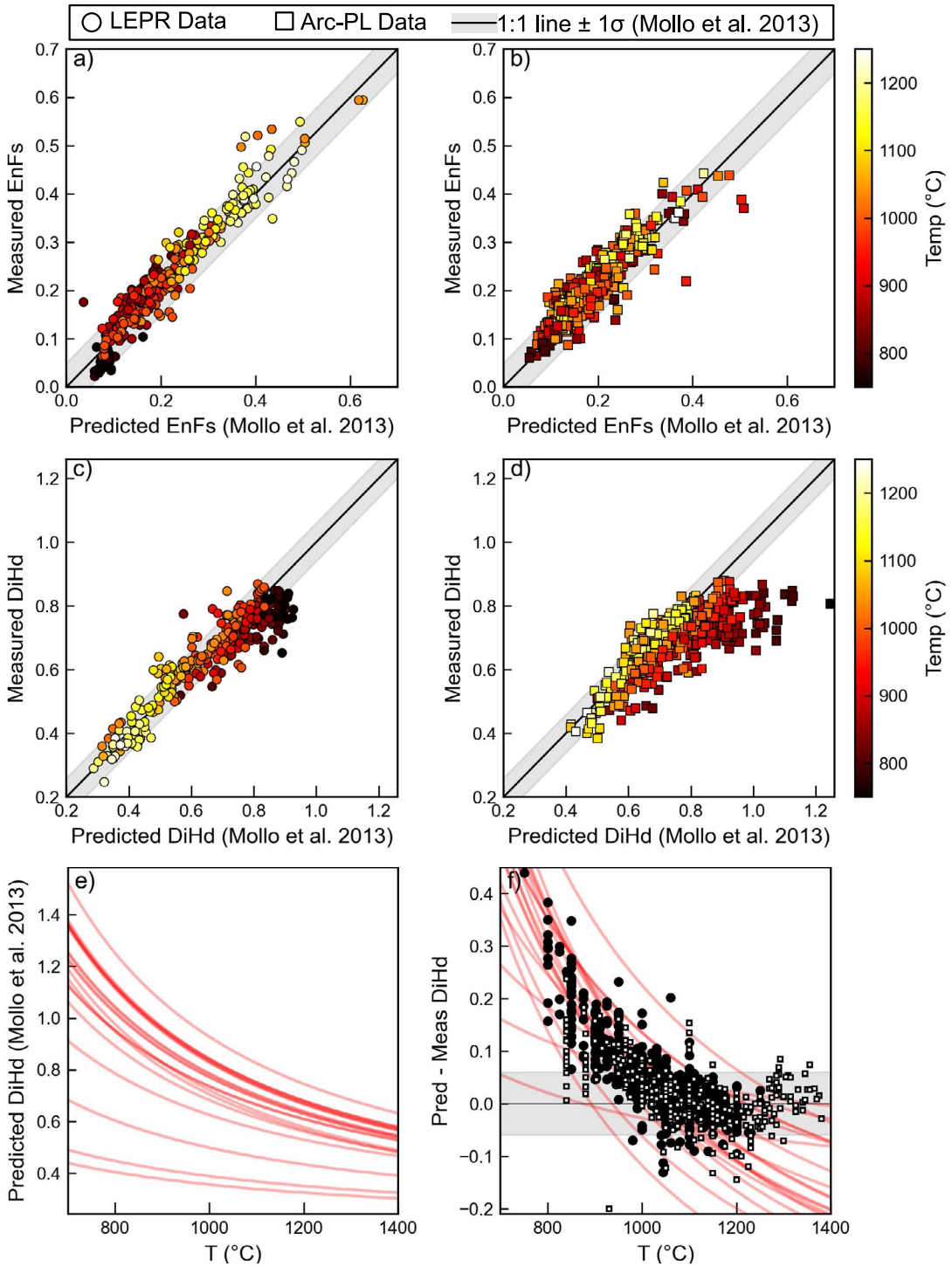

Fig. 5. Comparison of measured values of EnFs and DiHd with those predicted from the equations 6 and 7 of Mollo *et al.* (2013). In a), the grey bar shows ±0.05, whereas in b), the grey bar shows ±0.06 (both cut-offs from Mollo *et al.*, 2013). Symbols are coloured based on the experimental temperature. e) Predicted values of DiHd as a function of temperature using the expression of Mollo *et al.* (2013) for 20 randomly selected Cpx-Liq pairs. e) The discrepancy between predicted and measured DiHd contents for these 20 Cpx (red lines), with experimental data from Arc-PL (black dots) and LEPR (white squares) overlain.

that this equation does not contain a term for H_2O in the liquid (Putirka, 2008). However, this offset is concerning because this equation has been widely applied to hydrous arc magmas (Table 1).

The machine learning P-independent thermometer of Petrelli *et al.* (2020) yields worse statistics than P2008 eq33 for the ArcPL test dataset, largely because it overpredicts temperature at <1000 °C (Fig. 7f). Overprediction at low values and underprediction at high values is common for regression tree methods because these algorithms will not return a value outside the calibration range of the training dataset. Because some of the experiments in ArcPL were used to calibrate the machine learning thermobarometer of Jorgenson *et al.* (2022), we exclude these data when assessing this equation (Fig. 7e). For the

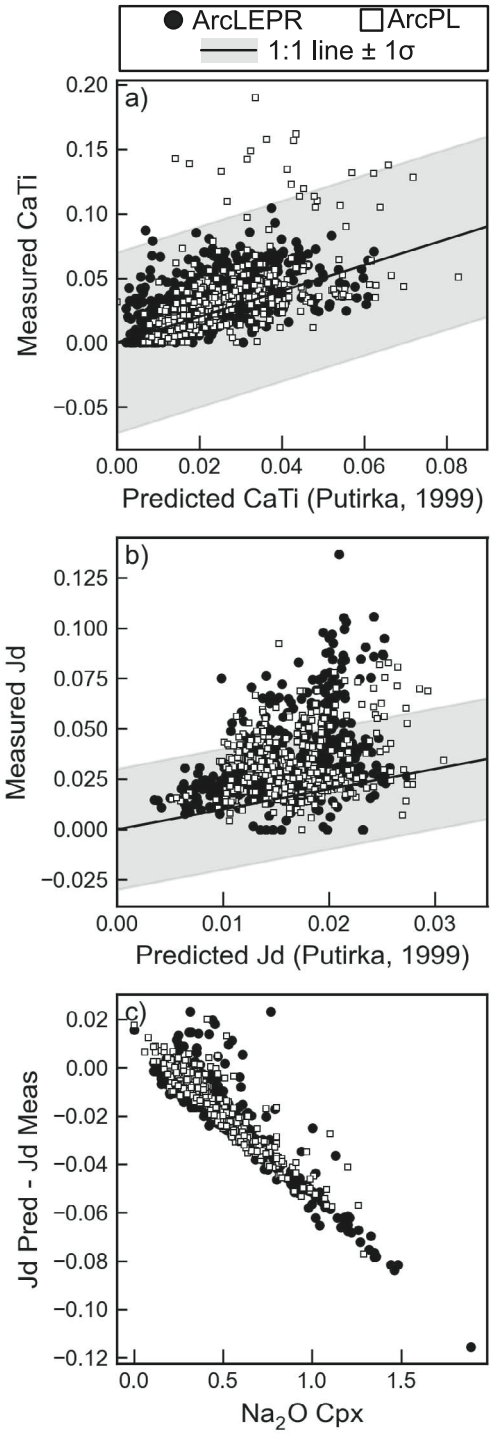

Fig. 6. The discrepancy between calculated and measured a) CaTi and b) Jd components. The discrepancy increases as a function of Na_2O content of the Cpx.

Jorgenson *et al.* (2022) thermobarometers, the authors recommend using the median value calculated across all trees, rather than the mean as used by Petrelli *et al.* (2020). For the ArcPL dataset, the median and mean show very similar statistics for temperature (Supporting Fig. 5c–d), with the median having a slightly better RMSE but slightly worse R^2. We proceed using the mean value because this results in slightly less scatter (and less visibly 'boxy' results, see Supporting Figs 5–7).

It is commendable how well the Jorgenson *et al.* (2022) thermometer performs given that unlike P2008 eq33 or Petrelli *et al.* (2020), it does not contain a term for H_2O in the liquid (a parameter that is often poorly constrained in natural systems). As for calculations using Putirka *et al.* (2003), there is a correlation between the discrepancy between calculated and experimental temperature and H_2O, but the R^2 value and gradient of this correlation is smaller ($R^2 = 0.12$ for Jorgenson *et al.*, 2022 vs 0.33 for Putirka *et al.*, 2003, and similarly, Grad = −7.33 vs −10.47°C/1 wt % H_2O, Supporting Fig. 4b).

Interestingly, the Cpx saturation thermometer of P2008 (eq34), which only uses the liquid composition, also performs well when iterated with Neave & Putirka (2017), having a slightly higher RMSE and MAE than eq33, but a gradient closer to 1 (0.99), and a very low intercept (29.9 °C, Fig. 8b). The similar performance of eq33 and eq34 raises an interesting question as to how much the temperature calculated with a Cpx-Liq thermometer is sensitive to the Cpx composition versus the liquid composition. Petrelli *et al.* (2020) examine the relative feature importance of each oxide in their machine learning model, showing that for Cpx-Liq temperatures, the three dominant features are MgO, CaO and H_2O in the liquid. We examine the relative importance of the Cpx versus Liq term for eq33 of P2008 by pairing each experimental liquid with each of the N = 214 Cpxs in our filtered dataset. For each liquid, we compare the temperatures obtained from these 214 Cpxs to the temperature obtained from the true experimental Cpx. Although the range of experimental temperatures varies by 350 °C, the temperature only changes by ∼±50 °C based on the Cpx composition (Supporting Fig. 8). Thus, users should be aware when performing Cpx-Liq thermometry that the thermometer is mostly tracking information held within the liquid composition, not the Cpx composition (see also Fig. 2d of Till *et al.*, 2012). The lack of temperature information help by Cpx is also apparent from the poor performance of Cpx-only thermometers (see Section 3.2). The dominance of the liquid composition when calculating Cpx-Liq temperatures means that it is vital to have reliable equilibrium tests to be able to determine which liquids were actually in equilibrium with which Cpx compositions.

Cpx-Liq barometers yield substantially worse statistics than thermometers (Fig. 8d-f, Fig. 9). All barometers return calculated pressures that form a flatter array than the 1:1 line (i.e. gradients <1). The machine learning barometers of Jorgenson *et al.* (2022) shows the best performance, closely followed by Petrelli *et al.* (2020), although these barometers overpredict for low P experiments, as expected for regression tree algorithms. It should be noted that the dataset being used to test the Jorgenson *et al.* (2022) barometer is slightly different (to avoid overlap with the model calibration dataset). Using the median tree value as suggested by the authors (Supporting Fig. 5a–b) rather than the mean tree results in a substantially lower R^2 value (0.66 vs 0.74), a higher RMSE (2.5 vs 2.1 kbar) but a slightly better gradient (0.74 vs 0.69) and intercept (0.8 vs 1.7 kbar, Fig. 5). Notably, at the very lowest pressures, the median tree does not experience the overestimation issues to the same degree as the mean tree.

Jorgenson *et al.* (2022) also suggest that the interquartile range (IQR) of the values returned by all trees for a given Cpx could be used to help filter out poor results in machine learning- based thermobarometers, which may help improve the performance of their thermobarometers further. They suggest filtering out analyses where the IQR is more than twice the stated RMSE on the thermobarometer. Unfortunately, we find that there is no significant correlation between the discrepancy between experiment and calculated pressure (or temperature) and the IQR of trees (Supporting Fig. 7). This

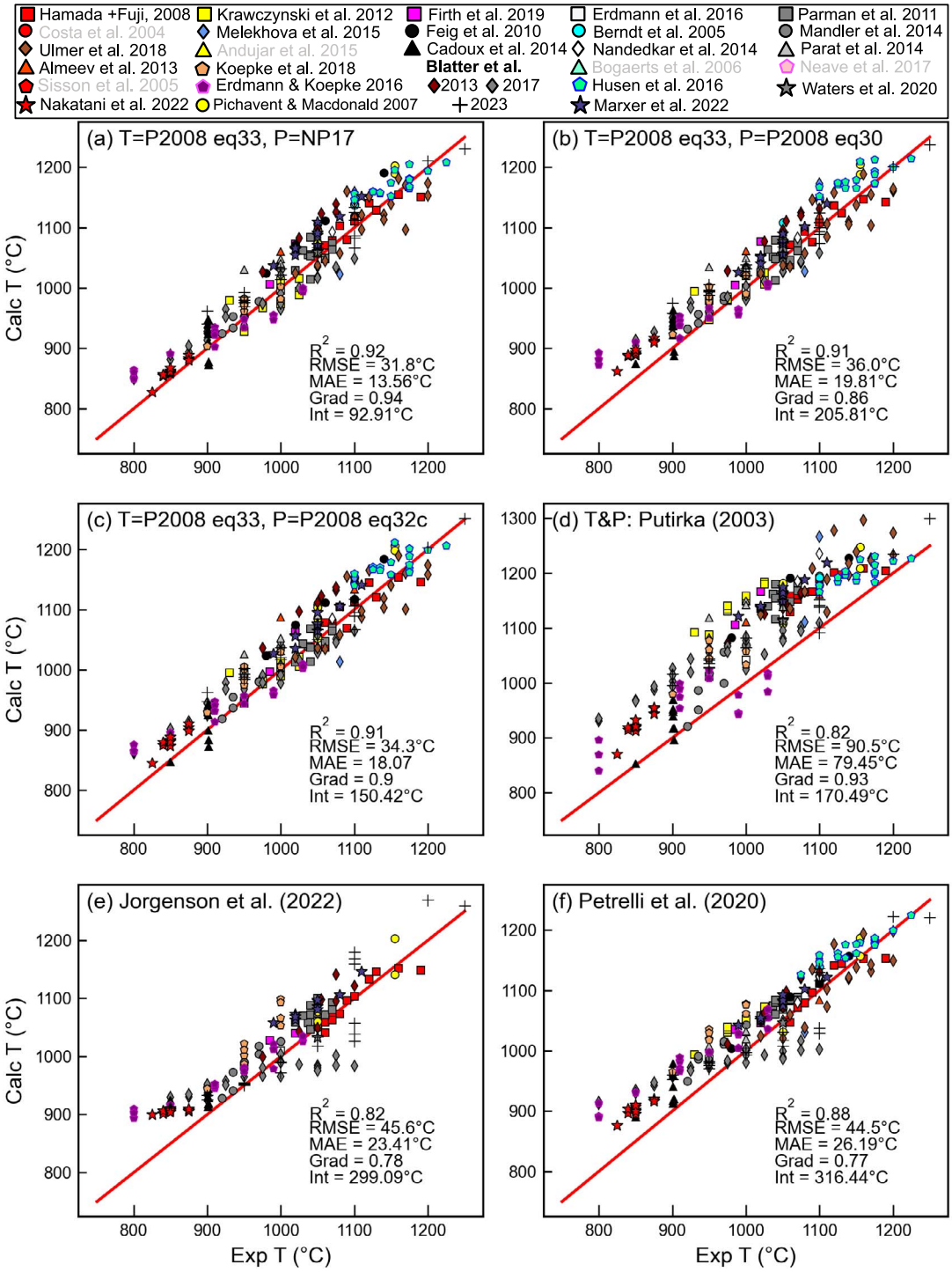

Fig. 7. Evaluation of various Cpx-Liq thermometers (iterating P and T) for the filtered ArcPL dataset. Experimental studies shown in Fig. 2 where all charges failed equilibrium or quality filters are greyed out in the legend. The best thermometer for this dataset is Putirka (2008) eq33 iterated with Neave & Putirka (2017). When testing the Jorgenson *et al.* (2022) expression, experiments in their calibration dataset are excluded, resulting in fewer symbols being shown on this panel than others. A red 1:1 line is shown on each plot.

figure also shows there is no significant correlation between the offset and IQR for the Petrelli *et al.* (2020) thermobarometers. Thus, at present, it does not seem that applying such a filter is useful (we also find that the IQR filter doesn't improve statistics for the Cpx-only thermobarometers discussed below, Supporting Fig. 14).

The iterated barometer of Neave & Putirka (2017) with P2008 eq33 thermometer underpredicts P for most experiments (Fig. 9a), shown by the strongly negative MAE (−2.12 kbar) and large RMSE (3.9 kbar). P2008 eq31 and eq30 iterated with eq33 show similar performance to one another, with both having relatively low gradients and high intercepts (eq30, Grad = 0.52, Int = 2.11, Fig. 9b, eq31, Grad = 0.58, Int = 3.39, Fig. 9c). Thus, they will both overestimate the pressures of low P Cpx and underestimate for high P Cpx. Putirka *et al.* (2003, Fig. 8d) forms a very scattered cloud, with similar pressures returned for experiments conducted at 2

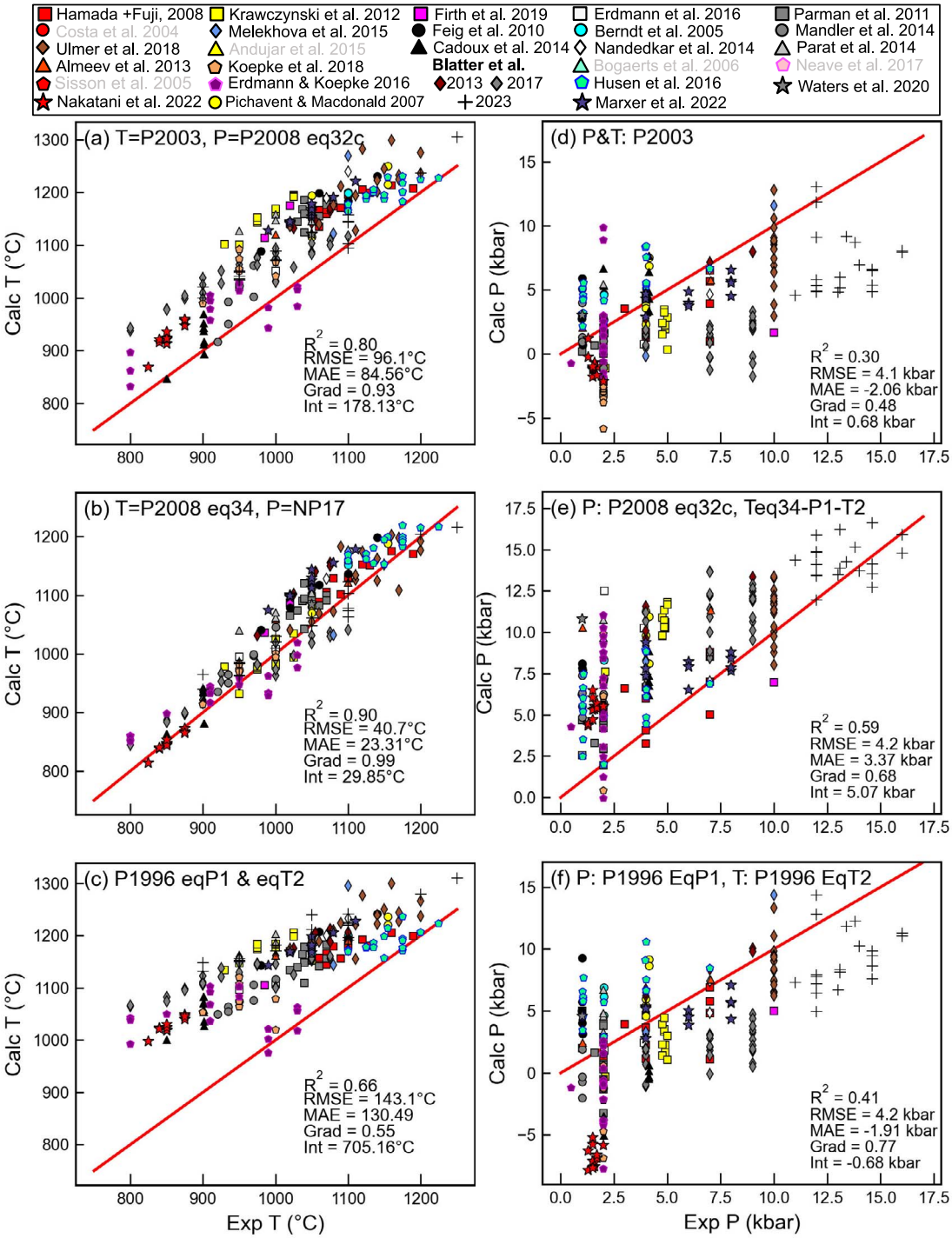

Fig. 8. Evaluation of various Cpx-Liq thermobarometers (iterating P and T) using experimental water contents for the filtered ArcPL dataset. Calculations shown for temperature (left columns, parts a–c) and pressure (right columns, parts d–f). Panel e) shows the default iteration for eq32c in the supporting spreadsheet of Putirka (2008).

and 17 kbar ($R^2 = 0.3$, Grad = 0.48, RMSE = 4.1 kbar). P2008 eq32c using various T estimates (Fig. 8e, Fig. 9d) and P1996 EqP1 and EqT2 (Fig. 8f) also show poor performance.

Sensitivity of Cpx-Liq thermobarometry to H_2O

For the comparisons shown in Figs 7–9, experimental H_2O contents were used when calculations required it. However, in most natural systems, the H_2O content of the melt is highly uncertain, particularly in volcanic systems with no melt inclusion data (e.g. because of a paucity of rapidly quenched tephra, or at understudied volcanoes). Indeed, most Cpx-Liq thermobarometry in arcs has been done using X-ray fluorescence (XRF) analyses of whole-rock samples. Thus, we investigate how much changing H_2O influences the calculated temperature (Fig. 10) and pressure (Fig. 11) to gain insight into the additional sources of uncertainty affecting calculations in variably hydrous arc systems.

We randomly select 41 Cpx-Liq pairs from ArcPL. For each of these pairs, we perform calculations at the experimental

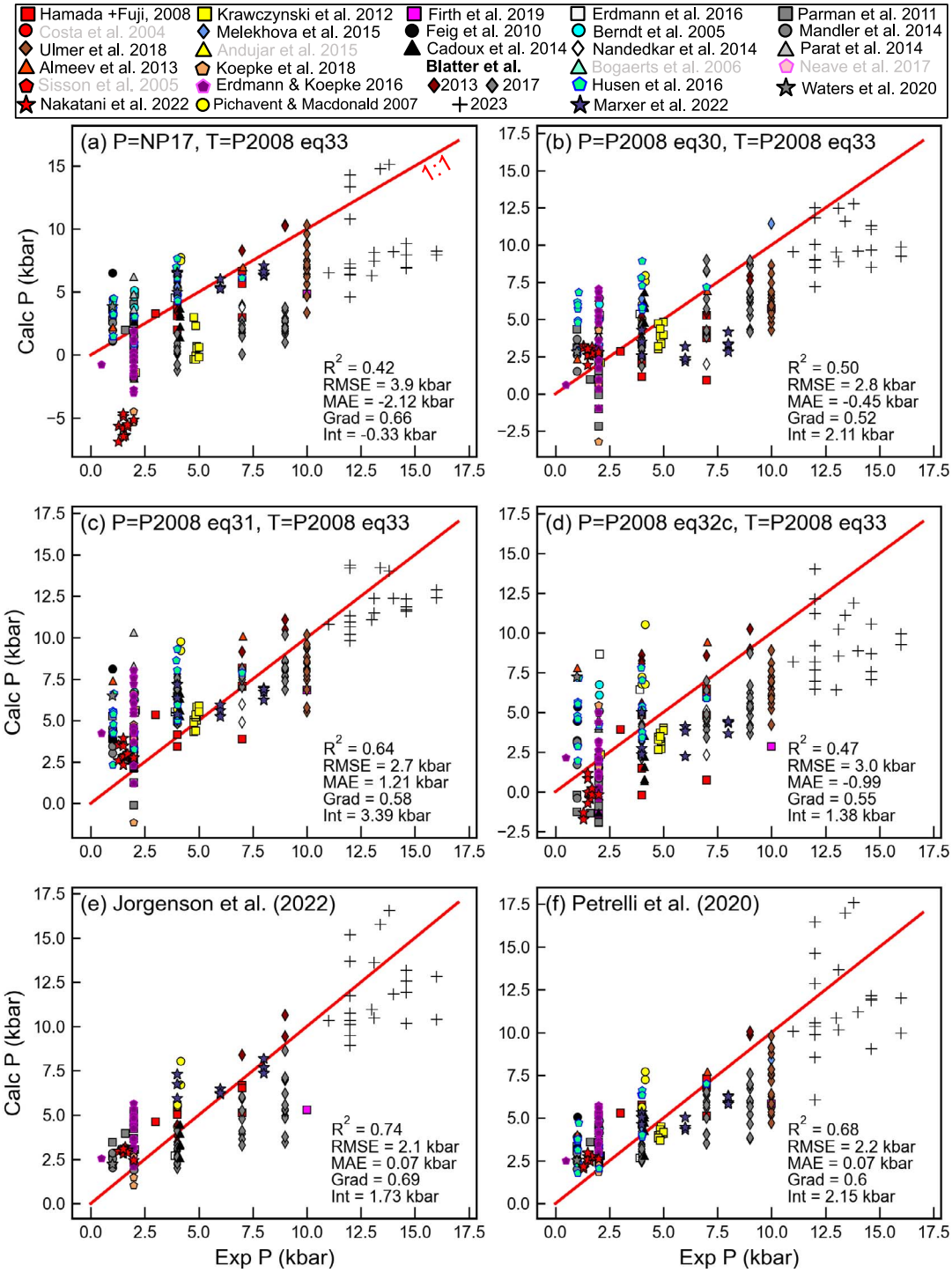

Fig. 9. Evaluation of Cpx-Liq pressures for the filtered ArcPL dataset

H_2O and then perturb H_2O by ±3 wt %, which represents a reasonable uncertainty on the water content of arc systems (where melt inclusion measurements of H_2O generally vary between ~0–6 wt % H_2O; Plank *et al.*, 2013). For each discrete H_2O content, we take the calculated temperature and pressure and subtract the value calculated using the experimental H_2O content. We do not show calculations performed using negative water contents. The variation in H_2O for each Cpx-Liq pair is shown as a single line, stretching either side of a black circle showing the experimental H_2O content (where the difference between the perturbed and experimental calculation is 0, Figs 10–11).

Different calibration approaches show different sensitivity to H_2O perturbations. The regression tree nature of the Cpx-Liq thermometer of Petrelli *et al.* (2020) means that it exhibits a complex non-linear sensitivity to H_2O (blue lines, Fig. 10a). At lower H_2O contents, it is extremely sensitive to H_2O, with temperatures decreasing by as much as 70 °C for a ~2 wt % increase in H_2O. At higher H_2O contents, calculated temperature changes very little, and in some cases, actually increases with increasing H_2O.

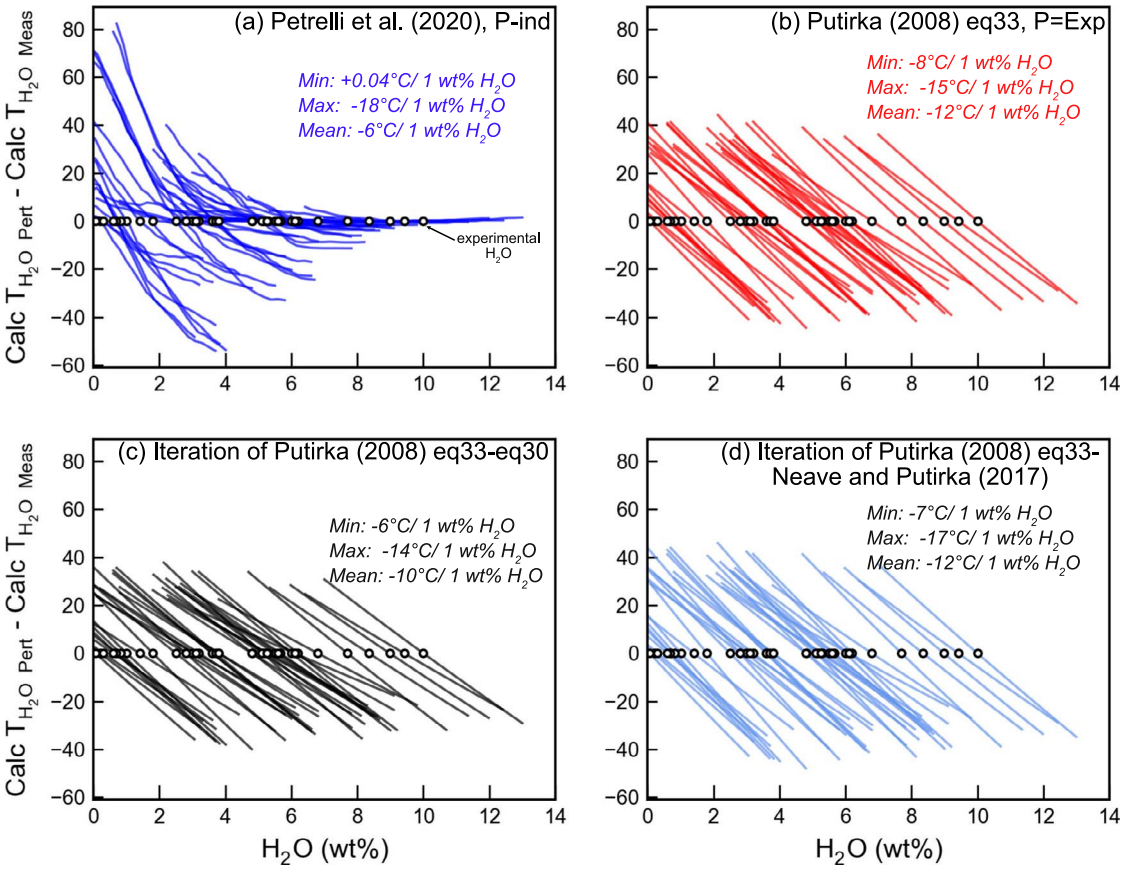

Fig. 10. Sensitivity of calculated temperature to melt H_2O content for 41 randomly selected Cpx-Liq pairs. For each Cpx-Liq pair, we add a linearly spaced array with values ranging from −3 to +3 to the experimental H_2O content, and calculations are performed for each discrete H_2O value (e.g. for $H_2O = 4$ wt %, calculations are performed from 1–7 wt %). The calculated temperature for the measured H_2O content are subtracted from the calculation for the perturbed H_2O content. This change in temperature for each Cpx-Liq pair is displayed as a colored line, passing through the H_2O content of the experiment (where the T discrepancy is 0). We calculate the max and min change in temperature, and the mean change, for all 41 selected pairs.

P2008 eq33 (using experimental pressures) shows a clear decline in calculated temperature with increasing H_2O, with much more similar trends between different samples than for Petrelli *et al.* (2020), Fig. 10b). When P2008 eq33 is iterated with P from eq30 instead of using experimental pressures, the temperature still drops (Fig. 10c), but there is a smaller decrease per unit increase in H_2O than in Fig. 10b. There is also a reasonably similar drop with increasing H_2O for eq33 iterated with NP17 (Fig. 10d).

Performing the same exercise for Cpx-Liq barometers, we find that Petrelli *et al.* (2020) shows non-linear behaviour, with calculated pressure decreasing with increasing H_2O until ~6 wt %, then increasing again (Fig. 11a). However, the change for all samples is relatively small (<1 kbar). When using experimental temperatures, eq30, eq31 and eq32c show an increase in calculated pressure with increasing H_2O, and all samples show the same gradient (because the H_2O term is multiplied by a constant in each of these equations, Fig. 11b). In contrast, these three barometers show very different behaviour when iterated with eq33, reflecting the fact that temperature and pressure are both affected by H_2O, and they are being iteratively solved (Fig. 11c). In all cases, the change in calculated pressure with changing H_2O for iterative calculations is smaller than when using experimental temperatures. This is because increasing H_2O decreases the temperature, which decreases the pressure, counteracting the effect of increasing H_2O increasing the pressure. The effect of changing temperature is so dominant for eq31 that iterative calculations see a decrease in pressure with increasing H_2O (although the effect is relatively subtle). Iteration of eq30 and eq33 is the most sensitive to H_2O, with uncertainty of just 1 wt % in H_2O resulting in an uncertainty in pressure of 0.26–0.63 kbar. The NP17 barometer has no H_2O term, but iterative calculations using this barometer will be H_2O-sensitive if a H_2O-sensitive thermometer is used (because of the T term in the barometer). Iteration with eq33 results in a relatively small H_2O effect (0.09 kbar per 1 wt % H_2O, Fig. 11d).

Given that temperature and pressure sensitivity are highly dependent on the choice of equations to iterate, we suggest that in systems where H_2O is not very well constrained, uncertainties should be propagated using methods similar to those shown here to assess the possible systematic uncertainty introduced by H_2O terms in equations.

Assessing suitable Cpx-only thermobarometers

The poor behaviour of many Cpx-Liq equilibrium tests and the difficulty identifying liquid compositions in arc magmas means that it would be advantageous to be able to use Cpx-only thermobarometers to deduce magma storage conditions. We assess the performance of the Cpx-only thermobarometers from Putirka (2008), Petrelli *et al.* (2020), Jorgenson *et al.* (2022) and Wang *et al.* (2021). None of the experiments in our ArcPL test dataset are present in the calibration datasets of Putirka (2008) and Petrelli *et al.* (2020). Wang *et al.* (2021) include the experiments of Berndt (2004) and Husen *et al.* (2016). The calibration dataset

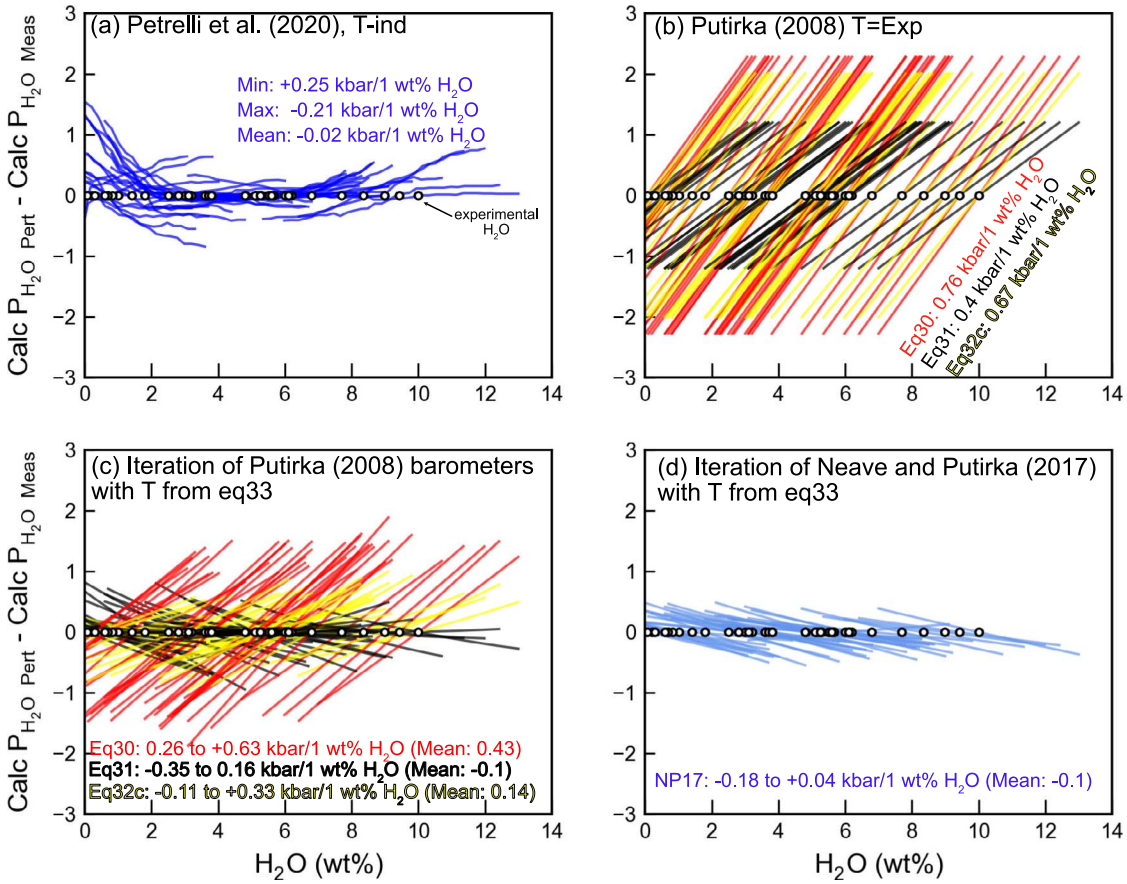

Fig. 11. Using the same method described in Fig. 10, we investigate the sensitivity of calculated pressure to H_2O. Eq32c shown in yellow colors and yellow font with a black outline.

of Jorgenson *et al.* (2022) has substantial overlap with our test dataset (Berndt, 2004; Feig *et al.*, 2010; Krawczynski *et al.*, 2012; Almeev *et al.*, 2013; Nandedkar *et al.*, 2014; Parat *et al.*, 2014; Melekhova *et al.*, 2015; Husen *et al.*, 2016; Ulmer *et al.*, 2018). To obtain the largest possible test dataset here, we exclude these overlapping experiments only when testing each thermobarometer in Figs 12–13. For fair comparisons between the best barometers in Fig. 15, we only use data that are not present in the calibration datasets of any of the models being compared.

Putirka (2008) presents a number of Cpx-only thermobarometers. P2008 eq32d is a P-sensitive, H_2O-independent Cpx-only thermometer. There is also a subsolidus version of eq32d. P2008 eq32a is a T-sensitive barometer that uses only uses the composition of the Cpx, whereas eq32b also requires users to specify the H_2O content of the liquid. Petrelli *et al.* (2020) present a Cpx-only barometer calibrated using an extra trees regression (with no H_2O term), but do not present a Cpx-only thermometer. Jorgenson *et al.* (2022) present a Cpx-only thermometer and barometer, neither of which include a H_2O term. Finally, Wang *et al.* (2021) present a thermometer (eq2) that has a H_2O term but is P-independent and a barometer (eq1) that is independent of T and H_2O.

When P2008 eq32d (T) is iterated with eq32a or eq32b (P), very similar temperatures are returned regardless of the experimental temperature (Fig. 12a–b, $R^2 = 0.1$–0.27, Grad = 0.14–0.24). For completeness we also test the subsolidus version of eq32d. This performs slightly better (lower RMSE and MAE), but it greatly underestimates higher temperature experiments (Fig. 12c–d). Putirka (2008) note that eq32d underestimates temperatures in hydrous systems, and indeed, we find a correlation between the discrepancy (Exp-Calc T) and H_2O in the liquid (eq32d–32b, $R^2 = 0.47$, grad = ~−26 °C/1 wt %, eq32d–32a: $R^2 = 0.35$, grad = ~−20 °C/1 wt %, Supporting Fig. 10). Thus, we do not recommend using either of these Cpx-only thermometers in hydrous arc magmas.

The Jorgenson *et al.* (2022) Cpx-only thermometer also overpredicts (Fig. 12e) for lower temperature experiments, and the discrepancy correlates with H_2O ($R^2 = 0.49$, grad = ~−28 °C/1 wt %, Supporting Fig. 10). The median and mean of trees show similarly poor performance (Supporting Fig. 11–12).

The Wang *et al.* (2021) thermometer performs the best (Fig. 12f), which is perhaps unsurprising given that this is the only Cpx-only thermometer that contains a H_2O term. However, it is worth noting that this equation was only calibrated using liquids with SiO_2 contents <60 wt % (e.g. basalts and basaltic-andesites). We find that the discrepancy between the experimental and calculated temperatures increases greatly at higher SiO_2 contents (overpredicting by 200–300 °C for the most silicic compositions in our test dataset, Supporting Fig. 13). When only experimental Cpx crystallized in liquids with $SiO_2 < 60$ wt % are considered, this thermometer performs much better (Supporting Fig. 13), with an $R^2 = 0.57$ and RMSE = 41.6 °C. The main problem is that it is difficult to identify from Cpx compositions alone whether a given crystal formed from a liquid with $SiO_2 > 60$ wt %. We do not find any robust correlations between Cpx composition and the calculated temperature discrepancy that could be used to apply this filter in natural

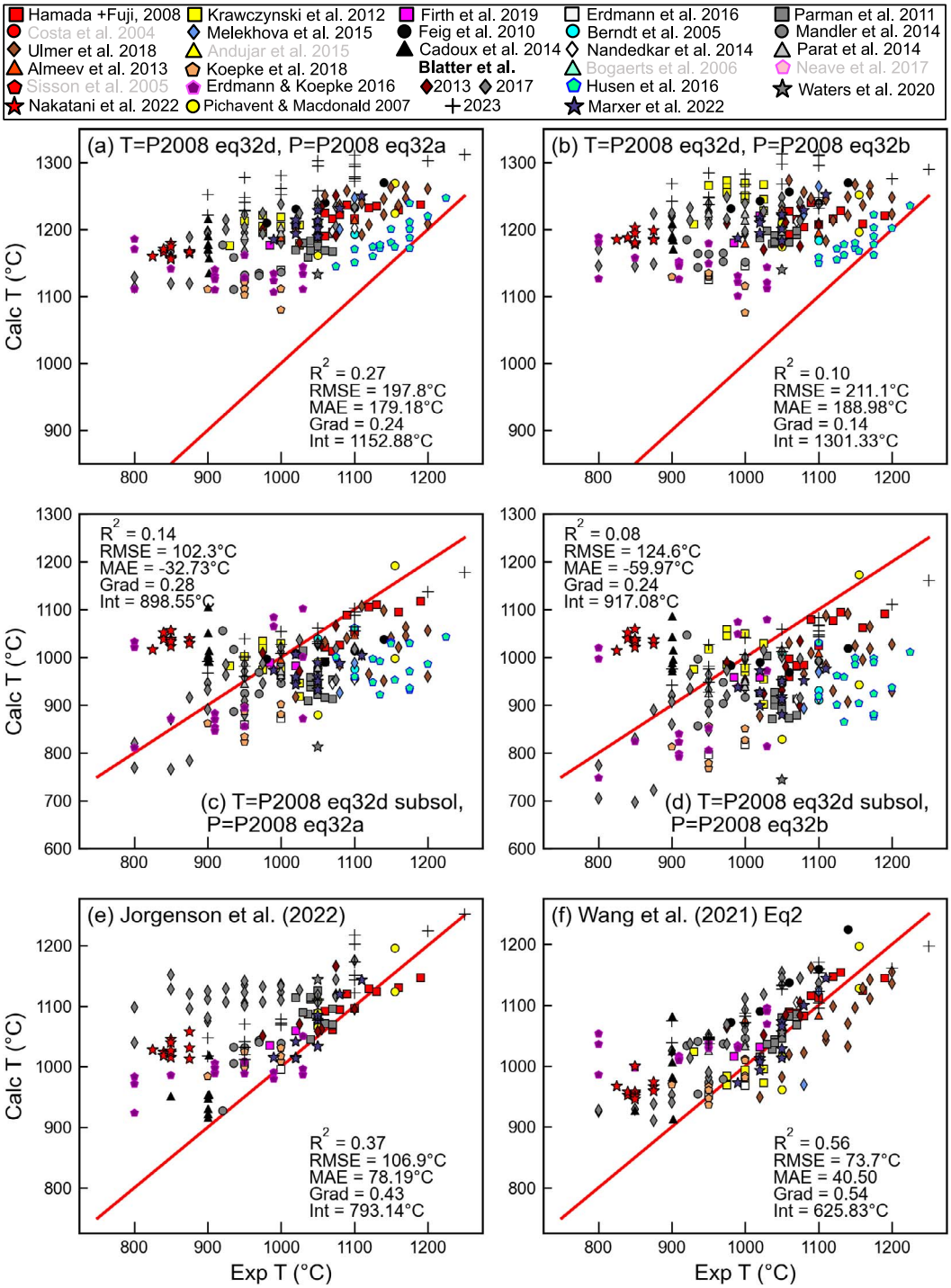

Fig. 12. Comparison of calculated and experimental temperatures for different Cpx-only thermobarometry combinations. For Jorgenson *et al.* (2022) and Wang *et al.* (2021), experiments in their calibration dataset are excluded.

systems. Thus, the Wang *et al.* (2021) thermometer needs to be used with extreme care in systems where Cpx may have crystallized from higher SiO_2 liquids. Overall, it is clear from this comparison that Cpx compositions grown from arc magmas do not hold sufficient temperature information without an independent estimate on the melt H_2O content from which they grew. Even when H_2O is included in the regression, Cpx compositions do not result in a very precise or accurate thermometer.

As for Cpx-Liq barometers, all Cpx-only barometers have intercepts >0 and gradients <1 (Fig. 13). These statistics indicate that all equations overpredict pressure for the lowest pressure experiments (e.g. intercept of 2.4 kbar for eq32d-eq32a, 4 kbar for eq32d-eq32b and 2 kbar for Petrelli *et al.*, 2020). The Jorgenson *et al.* (2022) barometer performs the best, with a high gradient (0.8) and a reasonably low intercept (1.1 kbar) and RMSE (1.9 kbar). As for Cpx-Liq, if the median tree is used rather than the mean, the R^2 and RMSE value is worse, but the gradient and intercept slightly

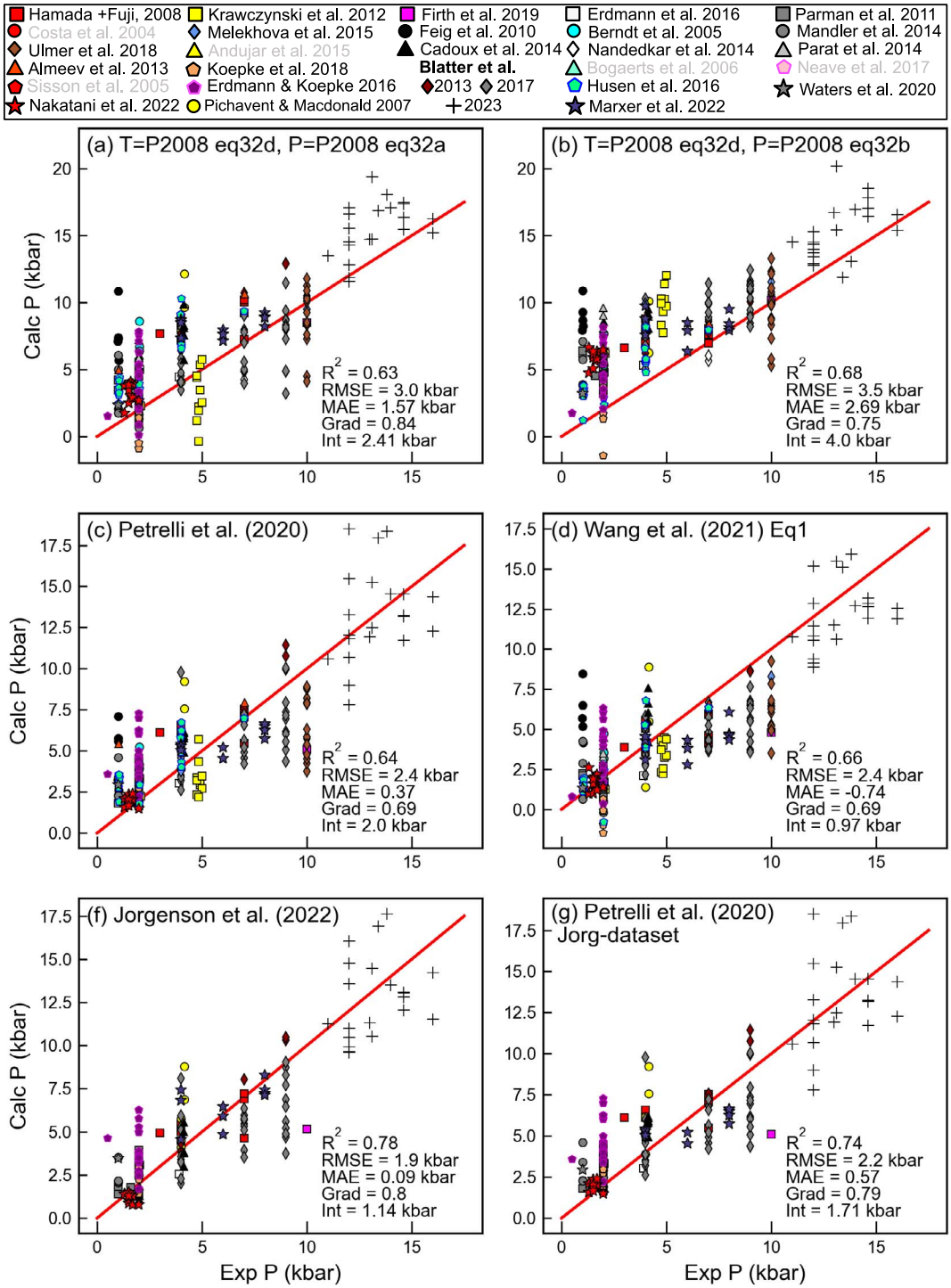

Fig. 13. Assessment of Cpx-only barometers. For Jorgenson *et al.* (2022) and Wang *et al.* (2021), experiments in their calibration dataset are excluded.

better (largely because the median is zero for many low-pressure experiments, Supporting Fig. 11). When Petrelli *et al.* (2020) is applied to the same small dataset used to assess Jorgenson *et al.* (2022), it is clear Jorgenson *et al.* (2022) performs slightly better (Fig. 13f vs g). This is likely because Jorgenson *et al.* (2022) have more hydrous arc-like magma compositions in their calibration dataset.

Sensitivity of Cpx-only thermobarometry to H_2O

It is unlikely that H_2O contents will be well constrained when applying Cpx-only thermobarometry to natural systems (unless melt inclusions in Cpx are analysed). As for Cpx-Liq, we assess the sensitivity of different equation combinations to the melt H_2O contents to give insights into the additional uncertainties when applying these equations to natural systems. The Wang *et al.* (2021) 'Cpx-only' thermometer (eq2) is extremely sensitive to H_2O, with calculated temperature decreasing by 23.4 °C per 1 wt % H_2O added (Fig. 14a). Although P2008 eq32d does not have a H_2O term itself, P2008 eq32b does, meaning that when these are iterated, calculated temperature actually increases with increasing H_2O (because H_2O changes pressure, which changes temperature). Fortunately, this seemingly spurious effect arising

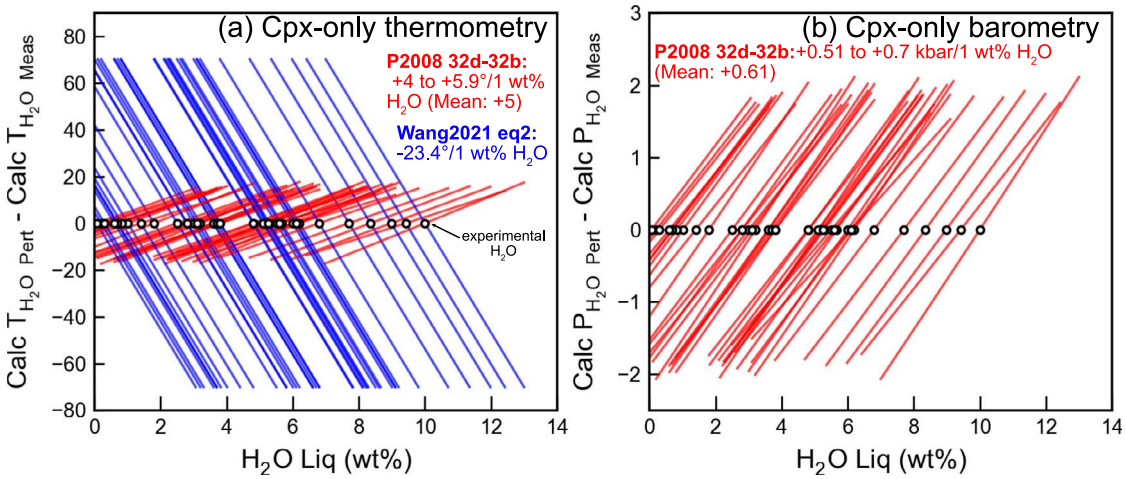

Fig. 14. As for Fig. 10, investigating the sensitivity of Cpx-only pressures and temperatures to H_2O content in the melt.

from iteration is quite subtle, with temperature only increasing by ~4–5.6 °C per 1 wt % H_2O added (Fig. 14a).

Of the Cpx-only barometers discussed here, only eq32b contains a H_2O term. Calculated pressures increase quite dramatically with added H_2O (mean increase of +0.61 kbar per 1 wt % H_2O, Fig. 14b). This represents an additional source of error when applying this equation in natural systems and should be propagated to obtain an uncertainty estimate. The strong sensitivity to H_2O for some of these equations raises a semantic point of whether equations with a H_2O term can truly be considered Cpx-only thermobarometers. However, most studies deploying Cpx-Liq barometry in arc magmas are using whole-rock XRF analyses, which do not provide any information on pre-eruptive H_2O, in place of measured glass compositions. Thus, it is normally necessary for studies to independently estimate the liquid H_2O content to perform calculations (e.g. from melt inclusion analyses in the system of interest, Scruggs & Putirka, 2018), regardless of whether they are using Cpx-Liq or Cpx-only expressions.

What is the resolution of Cpx-based thermobarometry in natural systems?

Our new test dataset shows that if melt H_2O contents are well constrained, Cpx-Liq thermometers can achieve RMSE errors of ~30–40 °C, although it should be noted that this is mostly attributed to the temperature information held in the liquid (shown by the similar performance of Liq-only and Cpx-Liq thermometers and poor performance of Cpx-only thermometers). Figure 10 also shows the strong sensitivity of the best performing thermometers to melt H_2O contents; uncertainty of just 1 wt % for H_2O contents leads to a systematic uncertainty in temperature of ~8–15 °C for the most successful thermometer (eq33, varies with sample and selected barometer). Cpx-only thermometers are inaccurate and imprecise when applied to arc magmas (Fig. 12f) even if melt H_2O is taken into account. The discrepancy between calculated and experimental temperatures is particularly large for low-temperature Cpx forming from more silicic melt compositions ($\Delta T > 300$ °C).

Many popular Cpx-Liq and Cpx-only barometers are associated with large, systematic and random errors. Systematic error is the reason why many equations have gradients and intercepts substantially different from the 1:1 line when plotted in experimental P versus Calc P space. These uncertainties may arise from the fact that hydrous experiments on arc-like magma compositions are relatively poorly represented in the calibration dataset of many barometers, as well as the regression strategy in the case of tree-based regressions. Sources of random error account in part for large RMSEs, and low R^2 values are discussed in detail in part I of this series (Wieser *et al.*, 2022a). Briefly, Wieser *et al.* (2022a) suggests that a substantial amount of random error is introduced because of low analytical precision during analyses of minor components such as Na_2O in experimental Cpx. We have attempted to mitigate the effect of this by only using experiments that averaged >5 Cpx measurements, but barometers still show very scattered performance when applied to individual experimental charges. Thus, we investigate whether averaging multiple different experiments conducted at similar pressures can help improve barometer performance by averaging out sources of random analytical and experimental error (following Putirka *et al.* 1996).

We show average calculated pressures for the best behaving Cpx-Liq barometer and Cpx-only barometer (Jorgenson *et al.*, 2022), as well as the second best Cpx-only barometer of Wang *et al.* (2021, eq1). Given their popularity in papers performing thermobarometry in arc magmas, we also show iteration of P2008 eq33 with Neave & Putirka (2017), iteration of the P and T equations from Putirka *et al.* (2003) and iteration of P2008 eq33–30 (Table 1, Fig. 15). To compare these thermobarometers, we only use experiments that do not appear in the calibration datasets of any of these six thermobarometers. We round experimental pressures to the nearest 0.2 kbar. For each unique rounded pressure in the dataset, we calculate the total number of Cpx compositions within that pressure 'bin'. For example, at 1 kbar (experiments performed between 0.8 and 1.2 kbar), there are seven different experimental charges, each with an averaged Cpx composition. In total, 65 individual Cpx analyses were performed across these seven charges. For these seven experimental charges, we calculate the mean and $\pm 1\sigma$ of the calculated pressures and experimental pressures. If more than one experimental charge was present in this pressure bin, we display the average pressure as a red triangle with an error bar (Fig. 15). We calculate statistics for the regression between calculated and predicted pressures in each bin. If only one experimental charge was present in that pressure window, we show a blue triangle (and exclude this data point when calculating statistics). A comparable figure using all bin averages to calculate statistics is shown in Supporting Fig. 15.

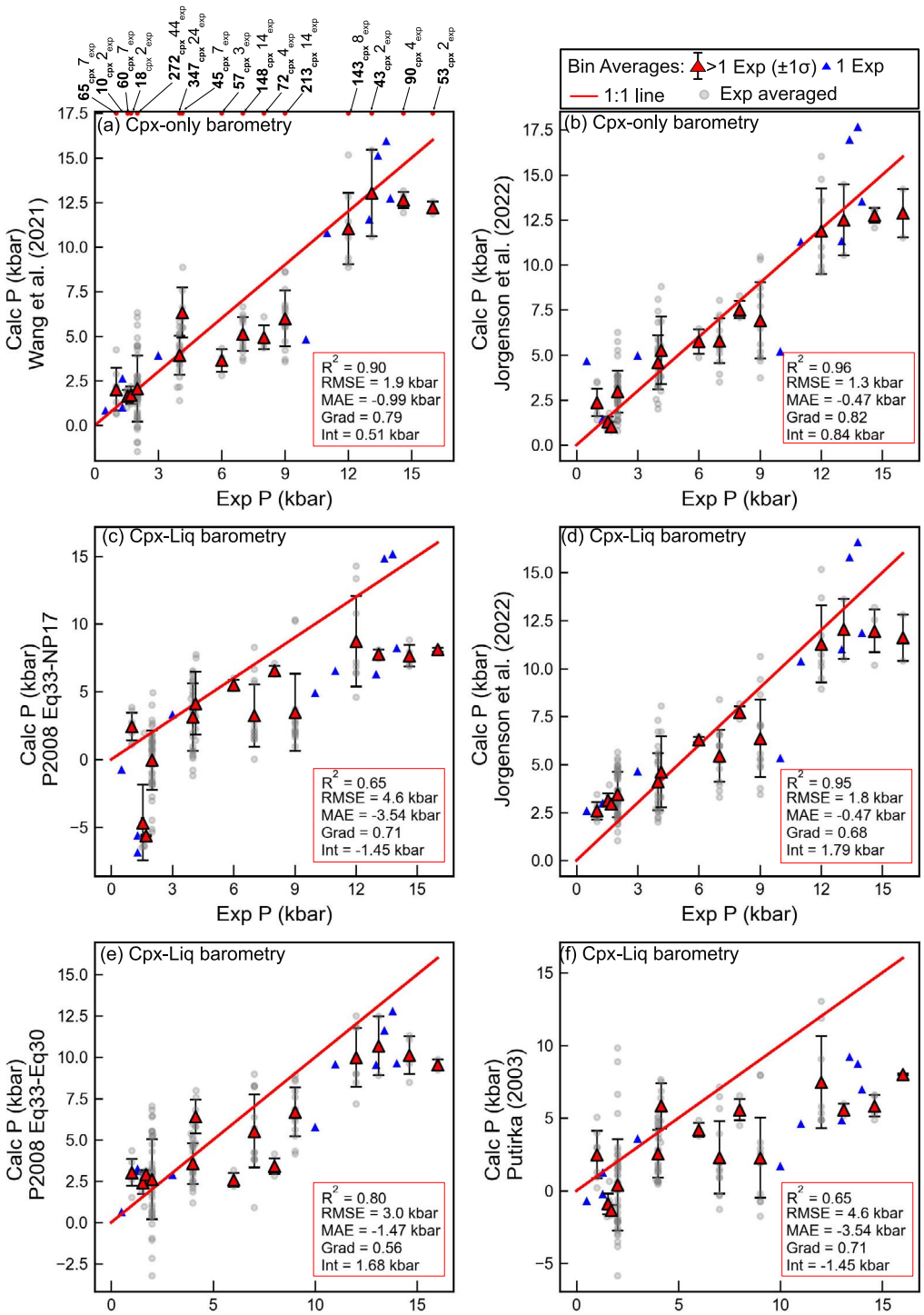

Fig. 15. Assessment of effect of averaging on Cpx-only (a–b) and Cpx-Liq barometry (c–f). After rounding experimental pressures to the nearest 0.2 kbar, we average calculated pressures for experiments in the same pressure bin. We show bin averages with >1 experimental charge in red as diamonds, with statistics within the red box. Bins with only 1 experimental charge are shown as blue triangles and not used to calculate statistics. A figure showing statistics for all bins is shown in the supporting information. Experiments averaged for the red triangles are shown as transparent grey symbols. The number of Cpx and number of experimental charges for each bin is indicated in part a).

Averaging multiple experiments yields greatly improved statistics compared with calculations using the reported average composition of each experimental charge (compare Fig 15a vs Fig. 13d, Fig 15b vs Fig. 13f, Fig. 15c vs Fig. 9a, Fig. 15d vs Fig. 9e; see also Putirka, 1999). For example, using individual experiments, the Cpx-only barometer of Jorgenson *et al.* (2022) has an R^2 of 0.78 using individual charges versus 0.96 when experiments with similar pressures are averaged, and the RMSE reduces from 1.9 to 1.3 kbar. This indicates that significant improvements to Cpx-based barometers could be made if analytical and experimental sources of random uncertainty are mitigated.

Although averaging greatly improves the performance of the Cpx-only and Cpx-Liq barometers of Jorgenson *et al.* (2022), and to a lesser extent Wang *et al.* (2021), it does not have the same

effect on the Cpx-Liq barometers of Neave & Putirka (2017) and Putirka *et al.* (2003). This indicates that systematic uncertainties are at play in these barometers (which are not improved by averaging). Overall, we suggest that extreme caution should be taken when interpreting published results from the iteration of Neave & Putirka (2017) with eq33 from Putirka *et al.* (2003) in volcanic arcs, given that it substantially underestimates pressures for the dataset tested here.

The improvement after averaging also emphasizes the point of Wieser *et al.* (2022a) and Putirka *et al.* (1996) that pressures calculated from individual Cpx compositions are very hard to interpret given the influence of analytical (and/or experimental) uncertainty, but averages of pressures calculated from large numbers of Cpx compositions may delineate the approximate region of the crust where the crystals grew. However, averaging can be problematic in natural systems, where crystals may have formed at a range of depths, and averaging eliminates true variations. Thus, ultimately it is preferable to obtain higher quality data for individual Cpx analyses in experiments and natural samples than have to rely on averaging, which could smear out true variations.

It is interesting that Cpx-only barometers behave just as well, if not slightly better, than Cpx-Liq barometers in the experiments examined here. This suggests that the additional uncertainty and effort associated with identifying equilibrium liquids in natural systems is likely not justified to obtain pressures. However, liquid compositions are certainly required to obtain reliable temperature information, given the poor performance of Cpx-only thermometers.

Using quoted RMSE estimates from thermobarometer calibrations as estimates of uncertainty

Overall, the statistics calculated here using the ArcPL dataset demonstrate that many of the commonly quoted RMSE estimates for Cpx-based thermobarometers are extremely optimistic and not representative of the errors associated with the application of these methods in natural systems (where P and T must be iteratively solved, compositions are not the exact ones used to calibrate the model and H_2O contents are not well known). For example, using P and T calculated from individual experimental charges, we calculated an RMSE of 4.1 kbar versus the often-quoted 1.3 kbar for Putirka *et al.* (2003), Fig. 8d), and 3.9 vs 1.4 kbar for Neave & Putirka (2017). RMSEs of 3–4 kbar (equivalent to ~11–15 km for 2700 kg/m^3) mean that barometers are not able to reliably distinguish storage in the upper, middle and lowermost crust in many volcanic arcs. For example, a 3 kbar RMSE value indicates that depth can only be constrained to within a 22 km wide region with 67% confidence. Thus, when these methods are applied, it should be appreciated that they can only pinpoint broad areas of crustal storage. Only after very extensive averaging do the best Cpx-based barometers (Fig. 15, Jorgenson *et al.*, 2022, and Wang *et al.*, 2021) yield RMSEs (1.3 and 1.8 kbar) that permit storage depths to be identified within a region spanning 10–15 km with 67% confidence. In addition, all these statistics were calculated using known H_2O contents. When applied in nature, the additional uncertainty introduced by using H_2O-sensitive Cpx-only and Cpx-Liq barometers must be propagated and will result in even larger (and systematic) uncertainties.

FUTURE DIRECTIONS

The large systematic errors exhibited by many popular Cpx-only thermobarometers and Cpx-Liq barometers in volcanic arcs are disappointing (e.g. Neave & Putirka, 2017, P2008 eq32c, P2008 eq32d–32b, Putirka, 2003) and have implications for published interpretations of magma storage based on these equations. However, given that thermobarometry is often one of the only available petrological tools for investigating magma plumbing system geometries because many arc volcanoes have no rapidly quenched tephra for melt inclusion work and Amp-only barometry is equally problematic (Erdmann *et al.*, 2014), it is thus critical to find pathways forward. The great improvement in calculated statistics that we observe through averaging of multiple experiments conducted at similar pressures suggests that recalibration of Cpx-based barometry based on higher quality experimental data using longer count times for minor elements such as Na, and more analyses per experiment, may help to provide a new dataset on which to recalibrate the next generation of thermobarometers (see Wieser *et al.*, 2022a). This could be based on re-analysis of existing experiments or higher quality analysis of new experiments. Higher quality analyses of Cpx in natural and experimental samples will mean that far less averaging is required to reduce scatter in calculated pressures. However, without access to such a high-quality dataset, it is difficult to determine how much improvement this may yield, and whether barometer performance will always be restricted by the relatively weak thermodynamic relationship between mineral components in Cpx and pressure (Putirka, 2008). If the main limitation is thermodynamic, extensive averaging of different experiments (and natural Cpx) may still be required.

It is also worth considering that Cpx-based thermobarometry may be fundamentally limited by the fact we are currently calculating the pressure and temperature-sensitive stoichiometric components (e.g. Jd, DiHd) using EPMA analyses of minerals and melts. Tommasini *et al.* (2022) examine natural Cpx crystals from Popocat'epetl Volcano and show that mineral components (e.g. Jd) calculated from XRD-informed site assignments differ greatly from the routines used by Neave & Putirka (2017) and P2008 using EPMA data alone. It is very plausible that if the Cpx components could be calculated more precisely and accurately (Tommasini *et al.*, 2022), the performance of thermobarometers would greatly increase.

In addition, it has been suggested that the presence of Fe^{3+} in Cpx from more oxidized melts stabilizes an aegirine component ($NaFe^{3+}SiO_6$), which convolutes the relationship between pressure and the Cpx Jd component (see Blundy *et al.*, 1995; Neave *et al.*, 2019; Neave & Putirka, 2017 for further discussion). For tholeiitic magmas, Neave *et al.* (2019) conclude that the aegirine component is not a significant issue and that perhaps Fe^{3+} is incorporated as a Ca-Al–bearing CaFe Tschermak's component. Our dataset, spanning a wider range of fO_2 that of Neave & Putirka (2017) and Neave *et al.* (2019), shows no clear correlation between the discrepancy between calculated and predicted pressure and the calculated Fe^{3+} proportion in the liquid (from the experimental fO_2) or the proportion of Fe^{3+} predicted in the Cpx using the parameterization in the spreadsheet of Putirka (2008) after Lindsley (1983). However, stoichiometric techniques to calculate Fe^{3+} in Cpx are 'misleading, inconsistent and inaccurate' (Dyar *et al.*, 1989) and extremely sensitive to propagated uncertainties from the measurement of other cations (Sobolev *et al.*, 1999; McCanta *et al.*, 2018). Although Mössbauer spectroscopy offers high-precision detection of Fe^{3+}, it is a bulk analysis method requiring >100 mg of sample, so it cannot be applied to

most experimental products (Rudra, 2021). XANES measurements are challenging and must take crystal orientation into account because of the anisotropy of Cpx to x-ray absorption (McCanta *et al.*, 2018; Rudra, 2021). Extensive work determining Fe^{3+} proportions for Cpx in different experimental charges at different redox conditions by XANES, combined with more accurate determination of mineral components, is likely needed to investigate why barometers seem to perform more poorly in arc magmas than tholeiitic magmas (e.g. Neave *et al.*, 2019).

Calibrating or recalibrating models using this dataset

It would certainly be tempting to recalibrate existing models or develop new models using the ArcPL dataset. However, doing so would mean we no longer have a truly independent test dataset to assess model quality and quantify errors. As a broad generalization, machine learning researchers encourage a test–train split of ~80:20 or 70:30 (Nguyen *et al.*, 2021), with the default in the popular Python-based machine learning package Sklearn being 75:25. Jorgenson *et al.* (2022) use their entire dataset of N = 2080 Cpx-Liq pairs to calibrate the final model. A new independent dataset to test this would thus require N = 520 unique experiments to achieve the Sklearn default ratio. Assessing Cpx-based barometers after recalibration using the ArcPL dataset combined with previous datasets would perhaps require waiting another 10–15 years for enough new experiments to be published to reassess how the newly calibrated models are performing. Thus, here we choose not to recalibrate models, and instead reiterate the importance of keeping any test dataset truly isolated during model training and validation. This means that the final 'publishable' model is developed in a way that avoids overconfidence in the model, which can occur when the test dataset has been used at any point during model tuning (Wujek *et al.*, 2016; Tampu *et al.*, 2022).

CONCLUSION

Our evaluation of a new dataset of hydrous experiments filtered for K_D and EnFs disequilibrium, cation sums and number of analyses per experiment provides new insights into the best thermobarometry calibrations to use when investigating pressures and temperatures in hydrous arc magmas using Cpx-liquid and Cpx alone. We show that the Cpx-Liq thermometers from Petrelli *et al.* (2020), Jorgenson *et al.* (2022) and P2008 eq33 all perform well and can give important insights into magma storage temperatures. However, this work also reveals that most temperature information is stored in the liquid, rather than the Cpx. In contrast, Cpx-only thermometers that do not have a term for H_2O in the liquid perform very poorly indeed, substantially overestimating temperatures for hydrous magmas (e.g. P2008 eq32d, Jorgenson *et al.* (2022)). Only the thermometer of Wang *et al.* (2021), which includes a H_2O term, shows a reasonable correspondence between experimental and calculated temperatures, and this equation still performs poorly for Cpx grown in liquids with $SiO_2 > 60$ wt %.

Cpx-Liq barometers all behave relatively poorly when applied to individual experimental charges, with large random and systematic errors (all RMSE >2.1, $R^2 < 0.74$, gradient <0.77). Cpx-only barometers are slightly better, but still show relatively large RMSEs (>1.9 kbar), and overpredict at low pressures. Importantly, these observed RMSE are all substantially larger than the RMSE reported for many of the equations, which are often quoted as an estimate of uncertainty in studies of natural magmas. Although random uncertainties can be addressed by averaging large numbers of Cpx (Fig. 14), even then, the best performing barometers can only just distinguish between storage zones ~10-15 km apart. Some commonly used barometers behave extremely poorly, overpredicting pressures by ~4 kbar (Fig. 8e). After averaging, systematic offsets are still very prominent for many barometers. We suggest that additional experimental and analytical work are required to obtain precise (or even accurate) pressures from Cpx compositions in volcanic arcs, to have a large enough high-quality dataset for model calibration and testing without having to perform such extensive averaging (which is hard to translate into natural systems). Given the importance of determining magma storage depths in arcs (Hilley *et al.*, 2022), this should be a key focus of the experimental and petrological community moving forward.

Data Availability Statement

The Excel file containing the new experimental dataset (ArcPL), along with the Jupyter Notebooks used to make every figure can be found on Penny Wieser's GitHub https://github.com/PennyWieser/BarometersBehavingBadly_PartII. This repository is also archived on Zenodo. 10.5281/zenodo.8067330

Supplementary Data

Supplementary data are available at *Journal of Petrology* online.

Conflict of interest

The authors decline no conflicts of interest.

Acknowledgements

P.E.W. thanks helpful conversations with Matt Gleeson and Keith Putirka, and to Dawnika Blatter for sharing her at-the-time unpublished experimental data and for helpful thoughts regarding normalization protocols. We are grateful for helpful comments from Luca Ziberna, one anonymous reviewer and editorial handling from Madeleine Humphries. This contribution was supported by funding from National Science Foundation grants 1948862 and 1949173 to A.J.R.K. and C.B.T., and startup funds to P.E.W. from UC Berkeley.

References

Almeev, R. R., Holtz, F., Ariskin, A. A. & Kimura, J.-I. (2013). Storage conditions of Bezymianny volcano parental magmas: results of phase equilibria experiments at 100 and 700 MPa. *Contributions to Mineralogy and Petrology* **166**, 1389–1414. https://doi.org/10.1007/s00410-013-0934-x.

Andújar, J., Scaillet, B., Pichavant, M. & Druitt, T. H. (2015). Differentiation conditions of a basaltic magma from Santorini, and its bearing on the production of andesite in arc settings. *Journal of Petrology* **56**, 765–794. https://doi.org/10.1093/petrology/egv016.

Auer, A., White, J., Nakagawa, M. & Rosenberg, M. (2013). Petrological record from young Ruapehu eruptions in the 4.5 ka Kiwikiwi formation, Whangaehu Gorge, New Zealand. *New Zealand Journal of Geology and Geophysics* **56**, 121–133. https://doi.org/10.1080/00288306.2013.796998.

Baker, D. R. & Eggler, D. H. (1987). Compositions of anhydrous and hydrous melts coexisting with plagioclase, augite, and olivine or low-ca pyrxene from 1 atm to 8 kbar: application to the Aleutian volcanic center of Atka. *American Mineralogist* **72**.

Barclay, J. (2004). A hornblende basalt from Western Mexico: water-saturated phase relations constrain a pressure-temperature window of eruptibility. *Journal of Petrology* **45**, 485–506. https://doi.org/10.1093/petrology/egg091.

Bartels, K. S., Kinzler, R. J. & Grove, T. L. (1991). High pressure phase relations of primitive high-alumina basalts from Medicine Lake volcano, northern California. *Contr. Mineral. and Petrol.* **108**, 253–270. https://doi.org/10.1007/BF00285935.

Beier, C., Haase, K. M., Brandl, P. A. & Krumm, S. H. (2017). Primitive andesites from the Taupo Volcanic Zone formed by magma mixing. *Contributions to Mineralogy and Petrology* **172**, 1–14. https://doi.org/10.1007/s00410-017-1354-0.

Belousov, A., Belousova, M., Auer, A., Walter, T. R. & Kotenko, T. (2021). Mechanism of the historical and the ongoing vulcanian eruptions of Ebeko volcano, northern Kuriles. *Bulletin of Volcanology* **83**, 4. https://doi.org/10.1007/s00445-020-01426-z.

Berndt, J. (2004). An experimental investigation of the influence of water and oxygen fugacity on differentiation of MORB at 200 MPa. *Journal of Petrology* **46**, 135–167. https://doi.org/10.1093/petrology/egh066.

Berndt, J., Holtz, F. & Koepke, J. (2001). Experimental constraints on storage conditions in the chemically zoned phonolitic magma chamber of the Laacher See volcano. *Contributions to Mineralogy and Petrology* **140**, 469–486. https://doi.org/10.1007/PL00007674.

Blatter, D. L. & Carmichael, I. S. E. (2001). Hydrous phase equilibria of a Mexican high-silica andesite: a candidate for a mantle origin? *Geochimica et Cosmochimica Acta* **65**, 4043–4065. https://doi.org/10.1016/S0016-7037(01)00708-6.

Blatter, D. L., Sisson, T. W. & Hankins, W. B. (2013). Crystallization of oxidized, moderately hydrous arc basalt at mid- to lower-crustal pressures: implications for andesite genesis. *Contributions to Mineralogy and Petrology* **166**, 861–886. https://doi.org/10.1007/s00410-013-0920-3.

Blatter, D. L., Sisson, T. W. & Hankins, W. B. (2017). Voluminous arc dacites as amphibole reaction-boundary liquids. *Contributions to Mineralogy and Petrology* **172**, 27. https://doi.org/10.1007/s00410-017-1340-6.

Blatter, D. L., Sisson, T. W. & Hankins, W. B. (2023). Garnet stability in arc basalt, andesite, and dacite – an experimental study. *Contributions to Mineralogy and Petrology.* **178**. https://doi.org/10.1007/s00410-023-02008-w.

Blundy, J. D., Falloon, T. J., Wood, B. J. & Dalton, J. A. (1995). Sodium partitioning between clinopyroxene and silicate melts. *Journal of Geophysical Research* **100**, 15501–15515. https://doi.org/10.1029/95JB00954.

Bogaerts, M., Scaillet, B. & Auwera, J. V. (2006). Phase equilibria of the Lyngdal granodiorite (Norway): implications for the origin of metaluminous ferroan granitoids. *Journal of Petrology* **47**, 2405–2431. https://doi.org/10.1093/petrology/egl049.

Brugman, K. K. & Till, C. B. (2019). A low-aluminum clinopyroxene-liquid geothermometer for high-silica magmatic systems. *American Mineralogist* **104**, 996–1004. https://doi.org/10.2138/am-2019-6842.

Cadoux, A., Scaillet, B., Druitt, T. H. & Deloule, E. (2014). Magma storage conditions of large Plinian eruptions of Santorini volcano (Greece). *Journal of Petrology* **55**, 1129–1171. https://doi.org/10.1093/petrology/egu021.

Carmichael, I. S. E. (1991). The redox states of basic and silicic magmas: a reflection of their source regions? *Contr. Mineral. and Petrol.* **106**, 129–141. https://doi.org/10.1007/BF00306429.

Cassidy, M., Watt, S. F. L., Talling, P. J., Palmer, M. R., Edmonds, M., Jutzeler, M., Wall-Palmer, D., Manga, M., Coussens, M., Gernon, T., Taylor, R. N., Michalik, A., Inglis, E., Breitkreuz, C., Le Friant, A., Ishizuka, O., Boudon, G., McCanta, M. C., Adachi, T., Hornbach, M. J., Colas, S. L., Endo, D., Fujinawa, A., Kataoka, K. S., Maeno, F., Tamura, Y. & Wang, F. (2015). Rapid onset of mafic magmatism facilitated by volcanic edifice collapse. *Geophysical Research Letters* **42**, 4778–4785. https://doi.org/10.1002/2015GL064519.

Caulfield, J. T., Turner, S. P., Smith, I. E. M., Cooper, L. B. & Jenner, G. A. (2012). Magma evolution in the primitive, intra-oceanic Tonga arc: petrogenesis of basaltic andesites at Tofua volcano. *Journal of Petrology* **53**, 1197–1230. https://doi.org/10.1093/petrology/egs013.

Chen, Z., Zeng, Z., Tamehe, L. S., Wang, X., Chen, K., Yin, X., Yang, W. & Qi, H. (2021). Magmatic sulfide saturation and dissolution in the basaltic andesitic magma from the Yaeyama Central Graben, southern Okinawa Trough. *Lithos* **388**, 106082. https://doi.org/10.1016/j.lithos.2021.106082.

Cigolini, C., Taticchi, T., Alvarado, G. E., Laiolo, M. & Coppola, D. (2018). Geological, petrological and geochemical framework of Miravalles-Guayabo caldera and related lavas, NW Costa Rica. *Journal of Volcanology and Geothermal Research* **358**, 207–227. https://doi.org/10.1016/j.jvolgeores.2018.05.013.

Costa, F. (2004). Petrological and experimental constraints on the pre-eruption conditions of Holocene dacite from Volcan San Pedro (36 S, Chilean Andes) and the importance of sulphur in silicic subduction-related magmas. *Journal of Petrology* **45**, 855–881. https://doi.org/10.1093/petrology/egg114.

Dahren, B., Troll, V. R., Andersson, U. B., Chadwick, J. P., Gardner, M. F., Jaxybulatov, K. & Koulakov, I. (2012). Magma plumbing beneath Anak Krakatau volcano, Indonesia: evidence for multiple magma storage regions. *Contributions to Mineralogy and Petrology* **163**, 631–651. https://doi.org/10.1007/s00410-011-0690-8.

Deegan, F. M., Whitehouse, M. J., Troll, V. R., Budd, D. A., Harris, C., Geiger, H. & Hålenius, U. (2016). Pyroxene standards for SIMS oxygen isotope analysis and their application to Merapi volcano, Sunda arc, Indonesia. *Chemical Geology* **447**, 1–10. https://doi.org/10.1016/j.chemgeo.2016.10.018.

Di Carlo, I. (2006). Experimental crystallization of a high-K arc basalt: the golden pumice, Stromboli Volcano (Italy). *Journal of Petrology* **47**, 1317–1343. https://doi.org/10.1093/petrology/egl011.

Draper, D. S. & Johnston, A. D. (1992). Anhydrous PT phase relations of an Aleutian high-MgO basalt: an investigation of the role of olivine-liquid reaction in the generation of arc high-alumina basalts. *Contr. Mineral. and Petrol.* **112**, 501–519. https://doi.org/10.1007/BF00310781.

Dyar, M. D., McGuire, J. & Ziegler, R. (1989). Redox equilibria and crystal chemistry of coexisting minerals from spinel lherzolite mantle xenoliths. *American Mineralogist* **74**, 969–980.

Erdmann, M. & Koepke, J. (2016). Silica-rich lavas in the oceanic crust: experimental evidence for fractional crystallization under low water activity. *Contributions to Mineralogy and Petrology* **171**, 83. https://doi.org/10.1007/s00410-016-1294-0.

Erdmann, S., Martel, C., Pichavant, M. & Kushnir, A. (2014). Amphibole as an archivist of magmatic crystallization conditions: problems, potential, and implications for inferring magma storage prior to the paroxysmal 2010 eruption of Mount Merapi, Indonesia. *Contributions to Mineralogy and Petrology* **167**, 1016. https://doi.org/10.1007/s00410-014-1016-4.

Erdmann, S., Martel, C., Pichavant, M., Bourdier, J.-L., Champallier, R., Komorowski, J.-C. & Cholik, N. (2016). Constraints from phase equilibrium experiments on pre-eruptive storage conditions in mixed magma systems: a case study on crystal-rich basaltic andesites from Mount Merapi, Indonesia. *Journal of Petrology* **57**, 535–560. https://doi.org/10.1093/petrology/egw019.

Feig, S. T., Koepke, J. & Snow, J. E. (2006). Effect of water on tholeiitic basalt phase equilibria: an experimental study under oxidizing conditions. *Contributions to Mineralogy and Petrology* **152**, 611–638. https://doi.org/10.1007/s00410-006-0123-2.

Feig, S. T., Koepke, J. & Snow, J. E. (2010). Effect of oxygen fugacity and water on phase equilibria of a hydrous tholeiitic basalt. *Contributions to Mineralogy and Petrology* **160**, 551–568. https://doi.org/10.1007/s00410-010-0493-3.

Firth, C., Adam, J., Turner, S., Rushmer, T., Brens, R., Green, T. H., Erdmann, S. & O'Neill, H. (2019). Experimental constraints on the differentiation of low-alkali magmas beneath the Tonga arc: implications for the origin of arc tholeiites. *Lithos* **344-345**, 440–451. https://doi.org/10.1016/j.lithos.2019.07.008.

Freundt, A. & Kutterolf, S. (2019). The long-lived Chiltepe volcanic complex, Nicaragua: magmatic evolution at an arc offset. *Bulletin of Volcanology* **81**, 60. https://doi.org/10.1007/s00445-019-1321-x.

Gaetani, G. A. & Grove, T. L. (1998). The influence of water on melting of mantle peridotite. *Contributions to Mineralogy and Petrology* **131**, 323–346. https://doi.org/10.1007/s004100050396.

Geiger, H., Troll, V. R., Jolis, E. M., Deegan, F. M., Harris, C., Hilton, D. R. & Freda, C. (2018). Multi-level magma plumbing at Agung and Batur volcanoes increases risk of hazardous eruptions. *Scientific Reports* **8**, 10547. https://doi.org/10.1038/s41598-018-28125-2.

Geoffroy, C. A., Alloway, B. V., Amigo, À., Parada, M. A., Gutierrez, F., Castruccio, A., Pearce, N. J. G., Morgado, E. & Moreno, P. I. (2016). A widespread compositionally bimodal tephra sourced from Volcán Melimoyu (44° S, Northern Patagonian Andes): Insights into magmatic reservoir processes and opportunities for regional correlation. *Quaternary Science Reviews* **200**, 141–159. https://doi.org/10.1016/j.quascirev.2018.09.034.

Ghiorso, M. S. & Gualda, G. A. R. (2015). An H2O–CO2 mixed fluid saturation model compatible with rhyolite-MELTS. *Contributions to Mineralogy and Petrology* **169**, 53. https://doi.org/10.1007/s00410-015-1141-8.

Gleeson, M. L. M., Gibson, S. A. & Stock, M. J. (2021). Upper mantle mush zones beneath low melt flux ocean island volcanoes: insights from Isla Floreana, Galápagos. *Journal of Petrology* **61**, egaa094. https://doi.org/10.1093/petrology/egaa094.

Grove, T. L., Gerlach, D. C. & Sando, T. W. (1982). Origin of calc-alkaline series lavas at Medicine Lake volcano by fractionation, assimilation and mixing. *Contr. Mineral. and Petrol.* **80**, 160–182. https://doi.org/10.1007/BF00374893.

Grove, T. L., Donnelly-Nolan, J. M. & Housh, T. (1997). Magmatic processes that generated the rhyolite of Glass Mountain, Medicine Lake volcano, N. California. *Contributions to Mineralogy and Petrology* **127**, 205–223. https://doi.org/10.1007/s004100050276.

Grove, T. L., Elkins-Tanton, L. T., Parman, S. W., Chatterjee, N., Muntener, O. & Gaetani, G. A. (2003). Fractional crystallization and mantle-melting controls on calc-alkaline differentiation trends. *Contributions to Mineralogy and Petrology* **145**, 515–533. https://doi.org/10.1007/s00410-003-0448-z.

Hamada, M. & Fujii, T. (2008). Experimental constraints on the effects of pressure and H2O on the fractional crystallization of high-mg island arc basalt. *Contributions to Mineralogy and Petrology* **155**, 767–790. https://doi.org/10.1007/s00410-007-0269-6.

Hammer, J., Jacob, S., Welsch, B., Hellebrand, E. & Sinton, J. (2016). Clinopyroxene in postshield Haleakala ankaramite: 1. Efficacy of thermobarometry. *Contributions to Mineralogy and Petrology* **171**, 7. https://doi.org/10.1007/s00410-015-1212-x.

Hesse, M. & Grove, T. L. (2003). Absarokites from the western Mexican volcanic belt: constraints on mantle wedge conditions. *Contributions to Mineralogy and Petrology* **146**, 10–27. https://doi.org/10.1007/s00410-003-0489-3.

Hilley, G., Brodsky, E., Roman, D., Shillington, D. & Tobin, H. (2022). SZ4D implementation plan. *Stanford Digital Repository* . https://doi.org/10.25740/HY589FC7561.

Hirschmann, M. M., Ghiorso, M. S., Davis, F. A., Gordon, S. M., Mukherjee, S., Grove, T. L., Krawczynski, M., Medard, E. & Till, C. B. (2008). Library of experimental phase relations (LEPR): a database and Web portal for experimental magmatic phase equilibria data. *Geochemistry, Geophysics, Geosystems* **9**, n/a-n/a. https://doi.org/10.1029/2007GC001894.

Hollyday, A. E., Leiter, S. H. & Walowski, K. J. (2020). Pre-eruptive storage, evolution, and ascent timescales of a high-mg basaltic andesite in the southern Cascade arc. *Contributions to Mineralogy and Petrology* **175**, 88. https://doi.org/10.1007/s00410-020-01730-z.

Husen, A., Almeev, R. R. & Holtz, F. (2016). The effect of H2O and pressure on multiple saturation and liquid lines of descent in basalt from the Shatsky rise. *Journal of Petrology* **57**, 309–344. https://doi.org/10.1093/petrology/egw008.

Iacovino, K., Matthews, S., Wieser, P. E., Moore, G. & Begue, F. (2021). VESIcal part I: an open-source thermodynamic model engine for mixed volatile (H2O-CO2) solubility in silicate melt. *Earth and Space Science*. **8**. https://doi.org/10.1029/2020EA001584.

Jeffery, A. J., Gertisser, R., Troll, V. R., Jolis, E. M., Dahren, B., Harris, C., Tindle, A. G., Preece, K., O'Driscoll, B., Humaida, H. & Chadwick, J. P. (2013). The pre-eruptive magma plumbing system of the 2007–2008 dome-forming eruption of Kelut volcano, East Java, Indonesia. *Contributions to Mineralogy and Petrology* **166**, 275–308. https://doi.org/10.1007/s00410-013-0875-4.

Jorgenson, C., Higgins, O., Petrelli, M., Bégué, F. & Caricchi, L. (2022). A machine learning-based approach to clinopyroxene thermobarometry: model optimization and distribution for use in earth sciences. *JGR Solid Earth* **127**, e2021JB022904. https://doi.org/10.1029/2021JB022904.

Kawamoto, T. (1996). Experimental constraints on differentiation and H2O abundance of calc-alkaline magmas. *Earth and Planetary Science Letters* **144**, 577–589. https://doi.org/10.1016/S0012-821X(96)00182-3.

Kelley, K. A. & Cottrell, E. (2009). Water and the oxidation state of subduction zone magmas. *Science* **325**, 605–607. https://doi.org/10.1126/science.1174156.

Koepke, J., Botcharnikov, R. E. & Natland, J. H. (2018). Crystallization of late-stage MORB under varying water activities and redox conditions: implications for the formation of highly evolved lavas and oxide gabbro in the ocean crust. *Lithos* **323**, 58–77. https://doi.org/10.1016/j.lithos.2018.10.001.

Krawczynski, M. J., Grove, T. L. & Behrens, H. (2012). Amphibole stability in primitive arc magmas: effects of temperature, H2O content, and oxygen fugacity. *Contributions to Mineralogy and Petrology* **164**, 317–339. https://doi.org/10.1007/s00410-012-0740-x.

Kress, V. C. & Carmichael, I. S. E. (1988). Stoichiometry of the iron oxidation reaction in silicate melts. *American Mineralogist* .

Lai, Z., Zhao, G., Han, Z., Huang, B., Li, M., Tian, L., Liu, B. & Bu, X. (2018). The magma plumbing system in the Mariana trough back-arc basin at 18° N. *Journal of Marine Systems* **180**, 132–139. https://doi.org/10.1016/j.jmarsys.2016.11.008.

Lindsley, D. H. (1983). Pyroxene thermometry. *American Mineralogist* **68**(5–6), 477–493.

Lormand, C., Zellmer, G. F., Kilgour, G. N., Németh, K., Palmer, A. S., Sakamoto, N., Yurimoto, H., Kuritani, T., Iizuka, Y. & Moebis, A. (2021). Slow ascent of unusually hot intermediate magmas triggering Strombolian to sub-Plinian eruptions. *Journal of Petrology* **61**, egaa077. https://doi.org/10.1093/petrology/egaa077.

Mandler, B. E., Donnelly-Nolan, J. M. & Grove, T. L. (2014). Straddling the tholeiitic/calc-alkaline transition: the effects of modest amounts of water on magmatic differentiation at Newberry volcano, Oregon. *Contributions to Mineralogy and Petrology* **168**, 1066. https://doi.org/10.1007/s00410-014-1066-7.

Martel, C., Pichavant, M., Holtz, F., Scaillet, B., Bourdier, J.-L. & Traineau, H. (1999). Effects of f_{O2} and H_2O on andesite phase relations between 2 and 4 kbar. *Journal of Geophysical Research* **104**, 29453–29470. https://doi.org/10.1029/1999JB900191.

Marxer, F., Ulmer, P. & Müntener, O. (2022). Polybaric fractional crystallisation of arc magmas: an experimental study simulating trans-crustal magmatic systems. *Contributions to Mineralogy and Petrology* **177**, 3. https://doi.org/10.1007/s00410-021-01856-8.

Masotta, M., Keppler, H. & Chaudhari, A. (2016). Fluid-melt partitioning of sulfur in differentiated arc magmas and the sulfur yield of explosive volcanic eruptions. *Geochimica et Cosmochimica Acta* **176**, 26–43. https://doi.org/10.1016/j.gca.2015.12.014.

McCanta, M. C., Dyar, M. D., Steven, C., Gunter, M. & Lanzirotti, A. (2018) In Situ Measurements Of Fe3+ In Pyroxene Using X-Ray Absorption Spectroscopy. *In: Using An Oriented Crystal Calibration To Refine Geothermobarometric Calculations 2*.

Melekhova, E., Blundy, J., Robertson, R. & Humphreys, M. C. S. (2015). Experimental evidence for Polybaric differentiation of primitive arc basalt beneath St. Vincent, Lesser Antilles. *Journal of Petrology* **56**, 161–192. https://doi.org/10.1093/petrology/egu074.

Mercer, C. N. & Johnston, A. D. (2008). Experimental studies of the P–T–H2O near-liquidus phase relations of basaltic andesite from North Sister Volcano, high Oregon Cascades: constraints on lower-crustal mineral assemblages. *Contributions to Mineralogy and Petrology* **155**, 571–592. https://doi.org/10.1007/s00410-007-0259-8.

Mollo, S., Putirka, K., Misiti, V., Soligo, M. & Scarlato, P. (2013). A new test for equilibrium based on clinopyroxene–melt pairs: clues on the solidification temperatures of Etnean alkaline melts at post-eruptive conditions. *Chemical Geology* **352**, 92–100. https://doi.org/10.1016/j.chemgeo.2013.05.026.

Moore, G. & Carmichael, I. S. E. (1998). The hydrous phase equilibria (to 3 kbar) of an andesite and basaltic andesite from western Mexico: constraints on water content and conditions of phenocryst growth. *Contributions to Mineralogy and Petrology* **130**, 304–319. https://doi.org/10.1007/s004100050367.

Moussallam, Y., Rose-Koga, E. F., Koga, K. T., Médard, E., Bani, P., Devidal, J.-L. & Tari, D. (2019). Fast ascent rate during the 2017–2018 Plinian eruption of Ambae (Aoba) volcano: a petrological investigation. *Contributions to Mineralogy and Petrology* **174**, 90. https://doi.org/10.1007/s00410-019-1625-z.

Moussallam, Y., Médard, E., Georgeais, G., Rose-Koga, E. F., Koga, K. T., Pelletier, B., Bani, P., Shreve, T. L., Grandin, R., Boichu, M., Tari, D. & Peters, N. (2021). How to turn off a lava lake? A petrological investigation of the 2018 intra-caldera and submarine eruptions of Ambrym volcano. *Bulletin of Volcanology* **83**, 36. https://doi.org/10.1007/s00445-021-01455-2.

Nakatani, T., Kudo, T. & Suzuki, T. (2022). Experimental constraints on magma storage conditions of two caldera-forming eruptions at Towada volcano, Japan. *JGR Solid Earth* **127**. https://doi.org/10.1029/2021JB023665.

Namur, O., Montalbano, S., Bolle, O. & Vander Auwera, J. (2020). Petrology of the April 2015 eruption of Calbuco volcano, southern Chile. *Journal of Petrology* **61**, egaa084. https://doi.org/10.1093/petrology/egaa084.

Nandedkar, R. H., Ulmer, P. & Müntener, O. (2014). Fractional crystallization of primitive, hydrous arc magmas: an experimental study at 0.7 GPa. *Contributions to Mineralogy and Petrology* **167**, 1015. https://doi.org/10.1007/s00410-014-1015-5.

Neave, D. A. & Putirka, K. D. (2017). A new clinopyroxene-liquid barometer, and implications for magma storage pressures under Icelandic rift zones. *American Mineralogist* **102**, 777–794. https://doi.org/10.2138/am-2017-5968.

Neave, D. A., Bali, E., Guðfinnsson, G. H., Halldórsson, S. A., Kahl, M., Schmidt, A.-S. & Holtz, F. (2019). Clinopyroxene–liquid equilibria and geothermobarometry in natural and experimental tholeiites: the 2014–2015 Holuhraun eruption, Iceland. *Journal of Petrology* **60**, 1653–1680. https://doi.org/10.1093/petrology/egz042.

Nguyen, Q. H., Ly, H.-B., Ho, L. S., Al-Ansari, N., Le, H. V., Tran, V. Q., Prakash, I. & Pham, B. T. (2021). Influence of data splitting on performance of machine learning models in prediction of shear strength of soil. *Mathematical Problems in Engineering* **2021**, 1–15. https://doi.org/10.1155/2021/4832864.

Nimis, P. (1999). Clinopyroxene geobarometry of magmatic rocks. Part 2. Structural geobarometers for basic to acid, tholeiitic and mildly alkaline magmatic systems. *Contributions to Mineralogy and Petrology* **135**, 62–74. https://doi.org/10.1007/s004100050498.

Parat, F., Streck, M., Holtz, F. & Almeev, R. R. (2014). Experimental study into the petrogenesis of crystal-rich basaltic to andesitic magmas at Arenal volcano. *Contributions to Mineralogy and Petrology.* **168**, 1040. https://doi.org/10.1007/s00410-014-1040-4.

Parman, S. W., Grove, T. L., Kelley, K. A. & Plank, T. (2011). Along-arc variations in the pre-eruptive H2O contents of Mariana arc magmas inferred from fractionation paths. *Journal of Petrology* **52**, 257–278. https://doi.org/10.1093/petrology/egq079.

Petrelli, M., Caricchi, L. & Perugini, D. (2020). Machine learning thermo-barometry: application to clinopyroxene-bearing magmas. *Journal of Geophysical Research - Solid Earth* **125**. https://doi.org/10.1029/2020JB020130.

Pichavant, M. & Macdonald, R. (2007). Crystallization of primitive basaltic magmas at crustal pressures and genesis of the calc-alkaline igneous suite: experimental evidence from St Vincent, Lesser Antilles arc. *Contributions to Mineralogy and Petrology* **154**, 535–558. https://doi.org/10.1007/s00410-007-0208-6.

Plank, T., Kelley, K. A., Zimmer, M. M., Hauri, E. H. & Wallace, P. J. (2013). Why do mafic arc magmas contain ∼4 wt % water on average? *Earth and Planetary Science Letters* **364**, 168–179. https://doi.org/10.1016/j.epsl.2012.11.044.

Preece, K., Gertisser, R., Barclay, J., Berlo, K. & Herd, R. A. (2014). Pre- and syn-eruptive degassing and crystallisation processes of the 2010 and 2006 eruptions of Merapi volcano, Indonesia. *Contributions to Mineralogy and Petrology* **168**, 1061. https://doi.org/10.1007/s00410-014-1061-z.

Profeta, L., Ducea, M. N., Chapman, J. B., Paterson, S. R., Gonzales, S. M. H., Kirsch, M., Petrescu, L. & DeCelles, P. G. (2016). Quantifying crustal thickness over time in magmatic arcs. *Scientific Reports* **5**, 17786. https://doi.org/10.1038/srep17786.

Putirka, K. (1999). Clinopyroxene + liquid equilibria to 100 kbar and 2450 K. *Contributions to Mineralogy and Petrology* **135**, 151–163. https://doi.org/10.1007/s004100050503.

Putirka, K. (2016). Amphibole thermometers and barometers for igneous systems and some implications for eruption mechanisms of felsic magmas at arc volcanoes. *American Mineralogist* **101**, 841–858. https://doi.org/10.2138/am-2016-5506.

Putirka, K. D. (2008). Thermometers and barometers for volcanic systems. *Reviews in Mineralogy and Geochemistry* **69**, 61–120. https://doi.org/10.2138/rmg.2008.69.3.

Putirka, K., Johnson, M., Kinzler, R., Longhi, J. & Walker, D. (1996). Thermobarometry of mafic igneous rocks based on

clinopyroxene-liquid equilibria, 0–30 kbar. *Contributions to Mineralogy and Petrology* **123**, 92–108.

Putirka, K. D., Mikaelian, H., Ryerson, F. & Shaw, H. (2003). New clinopyroxene-liquid thermobarometers for mafic, evolved, and volatile-bearing lava compositions, with applications to lavas from Tibet and the Snake River plain, Idaho. *American Mineralogist* **88**, 1542–1554. https://doi.org/10.2138/am-2003-1017.

Rader, E. L. & Larsen, J. F. (2013). Experimental phase relations of a low MgO Aleutian basaltic andesite at XH2O = 0.7–1. *Contributions to Mineralogy and Petrology* **166**, 1593–1611. https://doi.org/10.1007/s00410-013-0944-8.

Riker, J. M., Blundy, J. D., Rust, A. C., Botcharnikov, R. E. & Humphreys, M. C. S. (2015). Experimental phase equilibria of a Mount St. Helens rhyodacite: a framework for interpreting crystallization paths in degassing silicic magmas. *Contributions to Mineralogy and Petrology* **170**, 6. https://doi.org/10.1007/s00410-015-1160-5.

Romero, J. E., Morgado, E., Pisello, A., Boschetty, F., Petrelli, M., Cáceres, F., Alam, M. A., Polacci, M., Palma, J. L., Arzilli, F., Vera, F., Gutiérrez, R. & Morgavi, D. (2023). Pre-eruptive conditions of the 3 march 2015 lava fountain of Villarrica volcano (southern Andes). *Bulletin of Volcanology* **85**, 2. https://doi.org/10.1007/s00445-022-01621-0.

Rudra, A. (2021) Ferric iron partitioning between pyroxene and melt: experiments, microbeam analysis, and consequences for mantle redox. PhD thesis. University of Minnesota. Retrieved from the University of Minnesota Digital Conservancy, https://hdl.handle.net/11299/226367

Ruth, D. C. S. & Costa, F. (2021). A petrological and conceptual model of Mayon volcano (Philippines) as an example of an open-vent volcano. *Bulletin of Volcanology* **83**, 62. https://doi.org/10.1007/s00445-021-01486-9.

Rutherford, M. J., Sigurdsson, H., Carey, S. & Davis, A. (1985). The May 18, 1980, eruption of Mount St. Helens: 1. Melt composition and experimental phase equilibria. *Journal of Geophysical Research* **90**, 2929. https://doi.org/10.1029/JB090iB04p02929.

Sas, M., DeBari, S., Clynne, M. & Rusk, B. (2017). Using mineral geochemistry to decipher slab, mantle, and crustal input in the generation of high-mg andesites and basaltic andesites from the northern Cascade arc. *msam*. https://doi.org/10.2138/am-2017-5756.

Scruggs, M. A. & Putirka, K. D. (2018). Eruption triggering by partial crystallization of mafic enclaves at Chaos Crags, Lassen volcanic center, California. *American Mineralogist* **103**, 1575–1590. https://doi.org/10.2138/am-2018-6058.

Sheehan, F. & Barclay, J. (2016). Staged storage and magma convection at Ambrym volcano, Vanuatu. *Journal of Volcanology and Geothermal Research* **322**, 144–157. https://doi.org/10.1016/j.jvolgeores.2016.02.024.

Sisson, T. W., Ratajeski, K., Hankins, W. B. & Glazner, A. F. (2005). Voluminous granitic magmas from common basaltic sources. *Contributions to Mineralogy and Petrology* **148**, 635–661. https://doi.org/10.1007/s00410-004-0632-9.

Sobolev, V. N., McCammon, C. A., Taylor, L. A., Snyder, G. A. & Sobolev, N. V. (1999). Precise Moessbauer milliprobe determination of ferric iron in rock-forming minerals and limitations of electron microprobe analysis. *American Mineralogist* **84**, 78–85. https://doi.org/10.2138/am-1999-1-208.

Solaro, C., Martel, C., Champallier, R., Boudon, G., Balcone-Boissard, H. & Pichavant, M. (2019). Petrological and experimental constraints on magma storage for large pumiceous eruptions in Dominica island (Lesser Antilles). *Bulletin of Volcanology* **81**, 55. https://doi.org/10.1007/s00445-019-1313-x.

Tampu, I. E., Eklund, A. & Haj-Hosseini, N. (2022). Inflation of test accuracy due to data leakage in deep learning-based classification of OCT images. *Sci Data* **9**, 580. https://doi.org/10.1038/s41597-022-01618-6.

Till, C. B., Grove, T. L. & Krawczynski, M. J. (2012). A melting model for variably depleted and enriched lherzolite in the plagioclase and spinel stability fields: a MODEL FOR MELTING MODIFIED MANTLE. *Journal of Geophysical Research* **117**, n/a-n/a. https://doi.org/10.1029/2011JB009044.

Tommasini, S., Bindi, L., Savia, L., Mangler, M. F., Orlando, A. & Petrone, C. M. (2022). Critical assessment of pressure estimates in volcanic plumbing systems: the case study of Popocatépetl volcano, Mexico. *Lithos* **408-409**, 106540. https://doi.org/10.1016/j.lithos.2021.106540.

Ubide, T. & Kamber, B. S. (2018). Volcanic crystals as time capsules of eruption history. *Nature Communications* **9**, 326. https://doi.org/10.1038/s41467-017-02274-w.

Ulmer, P., Kaegi, R. & Müntener, O. (2018). Experimentally derived intermediate to silica-rich arc magmas by fractional and equilibrium crystallization at 1·0 GPa: an evaluation of phase relationships, compositions, liquid lines of descent and oxygen fugacity. *Journal of Petrology* **59**, 11–58. https://doi.org/10.1093/petrology/egy017.

Wang, X., Hou, T., Wang, M., Zhang, C., Zhang, Z., Pan, R., Marxer, F. & Zhang, H. (2021). A new clinopyroxene thermobarometer for mafic to intermediate magmatic systems. *European Journal of Mineralogy* **33**, 621–637. https://doi.org/10.5194/ejm-33-621-2021.

Waters, L. E., Cottrell, E., Coombs, M. L. & Kelley, K. A. (2021). Generation of Calc-alkaline magmas during crystallization at high oxygen fugacity: an experimental and petrologic study of tephras from Buldir volcano, Western Aleutian arc, Alaska, USA. *Journal of Petrology* **62**, egaa104. https://doi.org/10.1093/petrology/egaa104.

Wieser, P. E., Iacovino, K., Matthews, S., Moore, G. & Allison, C. M. (2022a). VESIcal: 2. A critical approach to volatile solubility modeling using an open-source Python3 engine. *Earth and Space Science* **9**. https://doi.org/10.1029/2021EA001932.

Wieser, P., Petrelli, M., Lubbers, J., Wieser, E., Kent, A. & Till, C. (2022b). Thermobar: an open-source thermobarometry hygrometry and chemometer Python 3 tool. *Preprint submitted to EarthArxiv.* . https://doi.org/10.31223/X5FD0K.

Wieser, P. E., Petrelli, M., Lubbers, J., Wieser, E., Ozaydin, S., Kent, A. & Till, C. (2022c). Thermobar: an open-source Python3 tool for thermobarometry and hygrometry. *Volcanica* **5**, 349–384. https://doi.org/10.30909/vol.05.02.349384.

Wieser, P. E., Kent, A., Till, C., Donovan, J., Neave, D., Blatter, D. & Mike Krawczynski, M. (2023). Barometers behaving badly: assessing the influence of analytical and experimental uncertainty on clinopyroxene thermobarometry calculations at crustal conditions (preprint). *Earth Science* . doi.org/10.31223/X5JT0N.

Wood, B. J. & Blundy, J. (1997). Predictive model for rare earth element partitioning between clinopyroxene and anhydrous silicate melt. *Contributions to Mineralogy and Petrology.* **129**, 166–181. https://doi.org/10.1007/s004100050330.

Wujek, B., Hall, P. & Günes, F. (2016) *Best Practices for Machine Learning Applications*. San Fransisco: SAS Institute Inc.