OXFORD

Journal of Petrology, 2024, **65**, egae020
https://doi.org/10.1093/petrology/egae020
Advance access publication date 2 March 2024
Original Manuscript

A Machine Learning-Based Thermobarometer for Magmatic Liquids

Gregor Weber[1,2,*] and Jon Blundy[1]

[1]Department of Earth Sciences, University of Oxford, South Parks Road, Oxford, OX1 3AN, United Kingdom
[2]School of Earth Sciences, University of Bristol, Queens Road, Bristol, BS8 1RJ, United Kingdom

*Corresponding author. E-mail: gregor.weber@bristol.ac.uk

Experimentally calibrated models to recover pressures and temperatures of magmas are widely used in igneous petrology. However, large errors, especially in barometry, limit the capacity of these models to resolve the architecture of crustal igneous systems. Here, we apply machine learning to a large experimental database to calibrate new regression models that recover P–T of magmas based on melt composition plus associated phase assemblage. The method is applicable to compositions from basalt to rhyolite, pressures from 0.2 to 15 kbar, and temperatures of 675°C to 1400°C. Testing and optimisation of the model with a filter that removes estimates with standard deviation above the 50th percentile show that pressures can be recovered with root-mean-square-error (RMSE) of 1.1 to 1.3 kbar and errors on temperature estimates of 21°C. Our findings demonstrate that, given constraints on the coexisting mineral assemblage, melt chemistry is a reliable recorder of magmatic variables. This is a consequence of the relatively low thermodynamic variance of natural magma compositions despite their relatively large number of constituent oxide components. We apply our model to two contrasting cases with well-constrained geophysical information: Mount St. Helens volcano (USA), and Askja caldera in Iceland. Dacite whole-rocks from Mount St Helens erupted 1980 to 1986, inferred to represent liquids extracted from cpx–hbl–opx–plag–mt–ilm mush, yield melt extraction source pressures of 5.1 to 6.7 kbar in excellent agreement with geophysical constraints. Melt inclusions and matrix glasses record lower pressures (0.7–3.8 kbar), consistent with magma crystallisation within the upper reaches of the imaged geophysical anomaly and during ascent. Magma reservoir depth estimates for historical eruptions from Askja match the location of seismic wave speed anomalies. V_p/V_s anomalies at 5 to 10 km depth correspond to hot (~990°C) rhyolite source regions, while basaltic magmas (~1120°C) were stored at 7 to 17 km depth under the caldera. These examples illustrate how our model can link petrology and geophysics to better constrain the architecture of volcanic feeding systems. Our model (MagMaTaB) is accessible through a user-friendly web application (https://igdrasil.shinyapps.io/MagmaTaBv4/).

Key words: thermobarometry; machine learning; magmatic plumbing system; magma storage; melt extraction

INTRODUCTION

Advances in igneous petrology and volcanology depend critically on our ability to determine the pressures (P), temperatures (T), and water contents of magmas (Blundy & Cashman, 2008; Putirka, 2017). By understanding these parameters, petrologists can gain insights into a wide range of problems, including the construction and evolution of the crust and mantle (Tibaldi *et al.*, 2013; Jagoutz, 2014; Plank & Forsyth, 2016; Till, 2017; Ducea *et al.*, 2021), and the pre-eruptive storage and transport properties of magmas, which underpin the characterisation of the architecture of igneous plumbing systems (Edmonds *et al.*, 2019; Maclennan, 2019; Nazzareni *et al.*, 2020; Giordano & Caricchi, 2022). P–T conditions are at the core of establishing meaningful links between petrological processes and geophysical monitoring of active volcanoes (Weber & Castro, 2017; Magee *et al.*, 2018; Pritchard *et al.*, 2018; Halldórsson *et al.*, 2022; Dayton *et al.*, 2023) and set fundamental controls on eruption dynamics (Ruprecht & Bachmann, 2010; Andújar & Scaillet, 2012; Cassidy *et al.*, 2018; Popa *et al.*, 2021). In addition, diffusion studies are heavily dependent on temperature information to reconstruct timescales of magmatic processes (Petrone *et al.*, 2016; Weber *et al.*, 2019; Costa *et al.*, 2020; Chakraborty & Dohmen, 2022). A thorough understanding of magmatic variables is, therefore, essential for progress in a variety of petrological fields.

P–T conditions of magmas can be determined by a range of methods that rely on equilibrium exchange reactions between different mineral phases or minerals and melts (e.g. Blundy & Cashman, 2008; Putirka, 2008). Although temperatures can be recovered with fairly small errors (typically better than ±30–60°C; e.g. Holland & Blundy, 1994; Ghiorso & Evans, 2008; Putirka, 2008; Neave & Putirka, 2017), pressure estimates struggle with uncertainties that lie at the very limit of usefulness. For example, the frequently used clinopyroxene–melt (eqn. 32a of Putirka, 2008) and amphibole–melt barometers (Putirka, 2016) show respective errors of at least 3 and 4 kbar, corresponding to a depth uncertainty of approximately 10 to 15 km, which is insufficient to resolve the structure of crustal magmatic systems. This may be attributed to analytical and experimental challenges (Wieser *et al.*, 2023a), low P-dependency of the components used in the regression (Putirka, 2016), or the choice of regression strategy (Higgins *et al.*, 2022).

An alternative approach is to obtain magma storage conditions from the major element chemistry of the melt phase in equilibrium with a particular mineral assemblage. In principle,

Received: April 27, 2023. **Revised:** February 19, 2024. **Accepted:** February 19, 2024

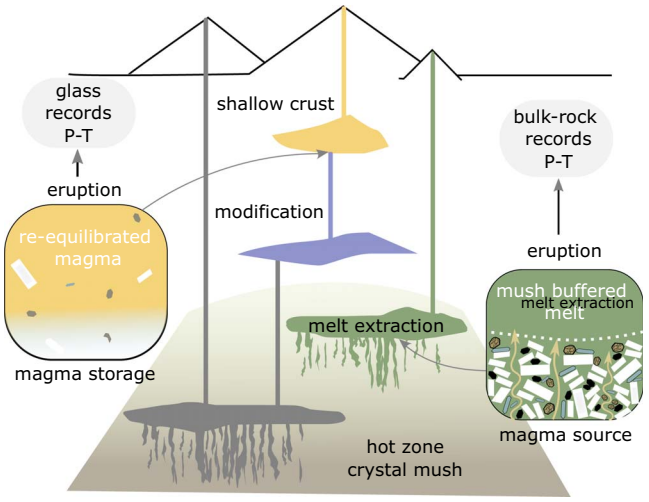

Fig. 1. Conceptual sketch illustrating different types of P–T information stored in volcanic materials. Melts are extracted from crystal mushes (hot zones) and buffered by the volumetrically more dominant mineral assemblage, which will vary as a function of pressure and temperature. If such liquids are erupted without modification, for example through magma mixing or crystal accumulation, then the bulk-rock composition can be used to recover extraction source conditions. Crystallisation conditions of magmas that have been modified and re-equilibrated can be reconstructed from the melt (glass) or melt and mineral phase assemblage pairs. Crystallisation conditions may equate to pre-eruptive storage conditions if additional time information is available.

volcanic glass chemistry can provide P–T conditions through the shift of cotectic lines for quartz- and feldspar-saturated melts (Blundy & Cashman, 2001; Gualda & Ghiorso, 2014), or for mafic liquids saturated with olivine, plagioclase, and clinopyroxene (Yang *et al.*, 1996; Hartley *et al.*, 2013). In crystal-rich mushes, of the type that have been shown geophysically to underlie volcanic areas worldwide, melt chemistry is buffered by the volumetrically dominant mineral assemblage, suggesting the possibility to constrain extraction source conditions in terms of the P and T at which a melt is multiply-saturated at its liquidus (Blundy, 2022). This approach (Fig. 1) requires only that the melt originated in a multi-mineral mush at depth and remained closed, except for loss of volatiles, during subsequent ascent, crystallisation and pre-eruptive storage (Gualda *et al.*, 2019; Blundy, 2022). Blundy (2022) used this method to develop a thermobarometer and hygrometer for volatile-saturated melts in equilibrium with the common mineral assemblage clinopyroxene–hornblende–orthopyroxene–magnetite–plagioclase–ilmenite (CHOMPI) using published phase equilibrium experiments to calibrate the model. As magmas ascend from their source they may degas and cool, crystallising in the process. This modifies the composition of the melt phase to reflect the conditions under which crystallisation and degassing occurred (Fig. 1). These may, for example, equate to immediate pre-eruptive storage conditions or to an integrated polybaric ascent path. P–T information obtained from these melts, for example, as melt inclusions or matrix glasses, should record shallower depths than the original source. Unless crystallisation during ascent leads to physical separation of crystals and melt, the bulk composition of the system, excepting volatiles, will remain unchanged and thus preserve information about the magma source depth.

Although both source extraction and subsequent storage conditions can be successfully determined by melt thermobarometry, the calibration range of current models is relatively limited. The large number of available experiments and recent advances in machine learning make it now possible to test if P–T recovery from melts can be generalised over a wider range of phase stabilities and melt compositions. Here, we build a large compilation of phase equilibrium experiments and calibrate a machine learning model to predict P–T of crustal magmas from the major element chemistry of melts and their associated mineral phase assemblage. We assess model errors based on experimental data not used in the calibration and test different optimisation strategies to obtain the most accurate results. Finally, the model is applied to two case studies of volcanic systems that have well constrained geophysical tomography and independent thermobarometric estimates. Our study introduces an unprecedented methodology, that for the first time, leverages melt chemistry and phase assemblage to estimate P–T conditions, encompassing the full spectrum of igneous compositions, from basaltic to rhyolitic systems.

RATIONALE

Our strategy to use the major element chemistry of the molten phase to serve as a recorder of intensive magmatic variables across a wide range of the compositional space, merits further theoretical reasoning. In this section, we postulate that this approach is consistent with basic thermodynamics (Gibbs Phase Rule), if the number of principal components that vary independently is considered to explain the variance of the system.

The use of mineral–melt assemblages to constrain pressure and temperature, such as that espoused by Blundy (2022) or embedded in the amphibole solidus barometer of Mutch *et al.* (2016), has its roots in the Gibbs Phase Rule, $\Phi + F = C + 2$, where Φ is the number of phases, C the number of components and F the number of degrees of freedom. To limit the latter to the two variables of interest, e.g. pressure and temperature, requires co-existence of a relatively large number of phases. If we consider magmas to be a ten-component mix of major oxide species, SiO_2–Al_2O_3–TiO_2–Fe_2O_3–FeO–MgO–CaO–Na_2O–K_2O–H_2O, then eight minerals need to coexist with melt and fluid to generate a sufficiently low degree of variance, e.g. P and T. If fO_2 is known, then Fe_2O_3/FeO ratio can be constrained, reducing the number of mineral phases to seven. Blundy (2022) settled on six mineral phases (CHOMPI, above), while the amphibole barometer requires seven; amphibole–biotite–plagioclase–alkali feldspar–quartz–magnetite–ilmenite (or titanite). Ensuring that so many minerals are in equilibrium with a given melt is an exacting requirement and, in the case of volcanic rocks that originated at depth in the presence of minerals that may no longer be stable as phenocrysts, difficult to prove.

However, application of the Gibbs Phase Rule to natural magmas overlooks the fact that not all 10 oxide components are independent. In fact, natural terrestrial magmas inhabit only a very small volume of the potential compositional space due to the strong correlations between components in the relatively limited number of common igneous minerals. A glance at any Harker plot of igneous rocks makes this important feature readily apparent. Thus, natural magmas have a much smaller number of *effective* components than their complex compositions would suggest. We can demonstrate this fact by performing Principal Components Analysis (PCA) on a large body of igneous rocks. To calculate the number of independent components necessary to explain the total variation in 8-dimensional major element, anhydrous oxide space (SiO_2–Al_2O_3–TiO_2–FeO–MgO–CaO–Na_2O–K_2O), we carried out a PCA on a global compilation of arc magmas (Fig. 2). PCA reduces the dimensionality of datasets by linearly transforming

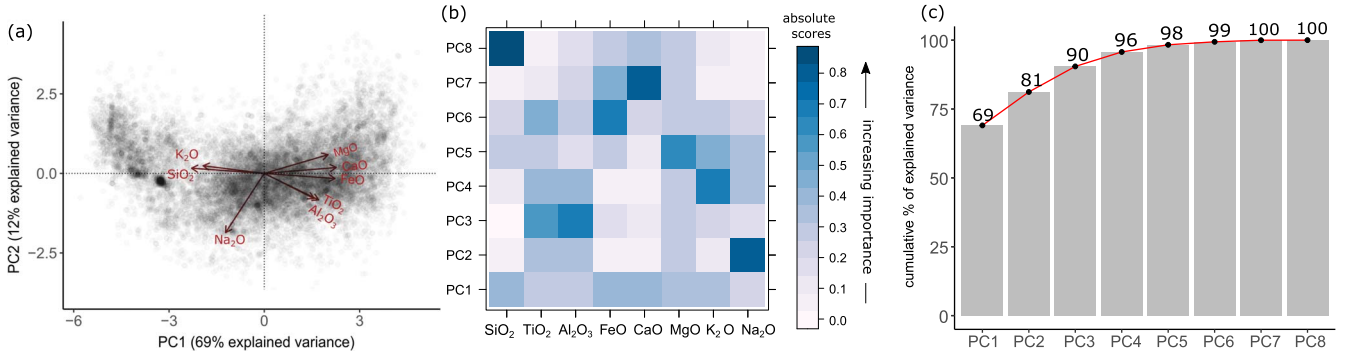

Fig. 2. Principal components analysis of arc magma bulk-rock and glass compositions. a) Cross plot of principal component (PC) 1 versus PC2 with loadings shown as red vectors. b) Matrix of loading scores for different major element oxides and PCs. Darker colours indicate an increasing importance of the particular oxide on the PCs. c) Cumulative percentage of variance explained by the different PCs. 96% of variance is explained by just four components.

the data into a new coordinate system, using Eigenvectors and the covariance matrix (e.g. Aitchison, 1984). As shown in Fig. 2a, the first principal component (PC1) is most strongly impacted by the loadings (eigenvectors ∗ (eigenvalues)$^{1/2}$) of SiO_2, CaO, and FeO, explaining 69% of the data. PC2 is most significantly affected by Na_2O, describing a further 12% of the total variance. PC3 (9%) is mainly influenced by TiO_2 and Al_2O_3, while K_2O provides the strongest contribution to PC4 (6%) (Fig. 2b). A cumulative plot (Fig. 2c) shows that although seven PCs are required to explain 100% of the data, just four PCs encompass 96% of the data. Considering four components and a typical phase assemblage (e.g. clinopyroxene–orthopyroxene–olivine–oxide–melt-fluid), the system would be invariant (F = 0). This finding introduces a novel perspective for thermobarometry. We suggest that common magmas have fewer effective components than previously assumed due to strong covariation between major element oxides, which are not truly independent of each other. This insight not only challenges conventional wisdom but also opens the door to calibrating models using melt chemistry over a much broader range of compositional space than previously considered.

APPLICATION OF RANDOM FOREST MACHINE LEARNING IN THERMOBAROMETRY

Machine learning (ML) is defined as the ability of a computer algorithm to learn from examples, existing in the form of data, and leverage this experience to build statistical models that make predictions on new observations (Zhou, 2021). ML is widely used for pattern or cluster recognition in unlabelled data, as well as for classification and regression tasks on labelled datasets. Among the latter class of models, also called supervised learning, a wide range of algorithms have been developed, with particular strengths and weaknesses for different types of data and applications (Mohammed *et al.*, 2016).

One of the most flexible and widely used supervised learning algorithms is random forests (Breiman, 2001). In principle, this method is based on aggregating re-sampled hierarchical flowcharts (decision trees) to minimise group heterogeneity (Fig. 3). As shown in Fig. 3a, decision trees consist of an initial root node, branches, internal nodes and leaf nodes. Starting from the root node, a tree is built by splitting the data into compartments (internal nodes) based on data features that correspond to the independent variables in the model. Leaf nodes then represent the final outcomes for different possible paths through the tree structure. In a random forest, a large number of decision trees is grown through bootstrapping (sampling with replacement) of the training (calibration) dataset and predictions are made by averaging the outcomes of all trees (Fig. 3b).

This procedure leads to much improved accuracy of random forest models compared to single decision trees (Zhou, 2021). The learning process in random forest models can be tuned through a set of hyperparameters, such as the number of trees in the model or the number of variables that are randomly sampled as candidates for each split. In conventional random forests, split-point values for each feature variable, defining a tree node, can be calculated based on different mathematical criteria. A variation of this method, the extremely randomized trees (ERT) algorithm (Geurts *et al.*, 2006), involves randomizing the choice of variables and split points that define each tree node, which is computationally more efficient and can improve accuracy in some cases.

The best practice for assessing the performance of both ML and conventional regression models involves testing them on a portion of the data, typically around 15% to 20%, which was not utilized during the model training process (as illustrated in Fig. 3c). This assessment usually entails splitting the data into training and testing subsets through a random process, which is repeated multiple times to gauge the impact of various randomized combinations. However, it is essential to acknowledge that this approach can, in certain instances, result in an underestimation of prediction errors. This underestimation may occur due to interdependencies in the data structure that can inadvertently transfer from the training to the testing data (Roberts *et al.*, 2017).

Alternatively, models can be evaluated by dividing the data into training and testing subsets while preserving the integrity of data groups, such as entire experimental studies. This approach can lead to a more robust assessment of prediction errors (Roberts *et al.*, 2017). Nonetheless, it is worth noting that this testing strategy may, in some cases, underestimate a model's predictive capacity, particularly when dealing with large group sizes, such as extensive experimental studies. Consequently, the choice of model testing strategies should be made with careful consideration of the specific context and dataset characteristics.

The accuracy of model predictions is typically assessed using the root-mean-square-error (RMSE), defined as:

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{n}}$$

where $\hat{y}$ is the predicted value, y is the observed value, and n is the number of observations. It is important to note that the

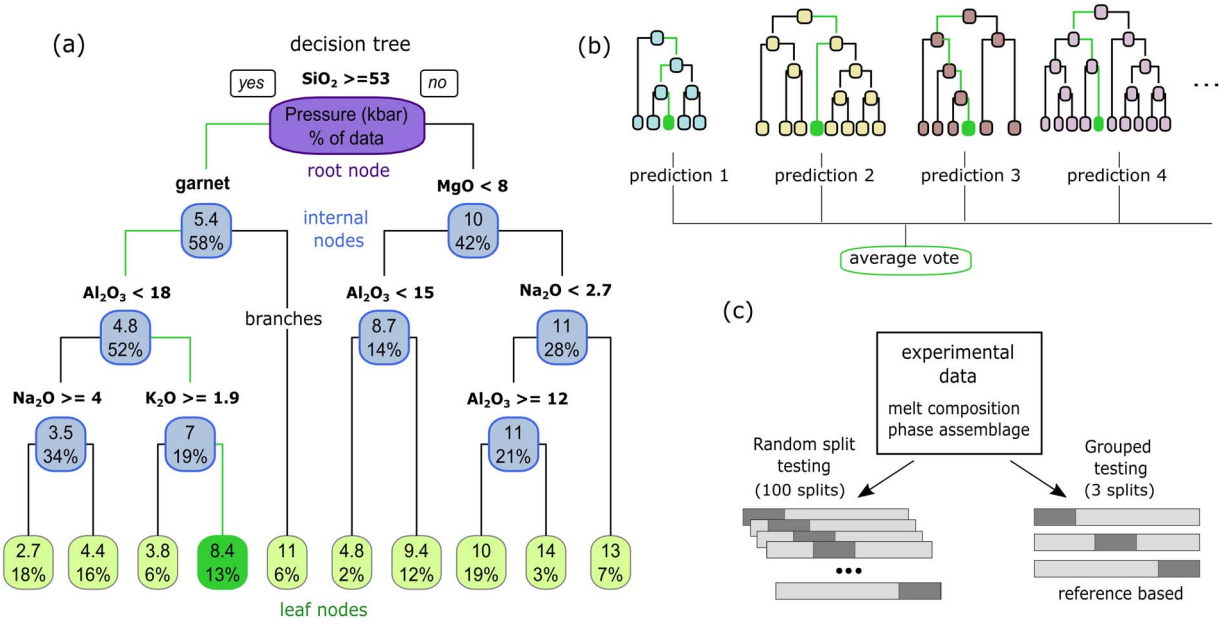

Fig. 3. Methodological summary of the machine learning method (ERT random forest) used for regression. a) Example structure of a decision tree using mineral assemblage and major element oxides to predict pressure. The decision tree is comprised of a root node (purple) that is split based on a series of yes and no questions regarding the regression variables into internal nodes (blue) and finally leaf nodes (green). Left and right branches correspond to yes and no respectively. Each node shows the predicted value (P kbar – top number), and the percentage of observations in the node (bottom number). The green branches illustrate a prediction path and the dark green leaf node the predicted pressure. b) The predictions of a large number of decision trees (typically ~500) are ensembled and averaged without weighting. c) We divided the experimental data into two parts: one for calibrating the model (training), and the other for testing the model's performance (testing). This was carried out in two different ways. First, we randomly split the data into training and testing sets, making sure each split keeps the same proportions of the target variable (P or T). This process was iterated 100 times (here termed: random split testing). Second, we divided the data while making sure that discrete experimental studies stayed together in either the training or testing set (grouped testing). This validation type was carried out with 3 different splits.

RMSE is sensitive to the range of observation (e.g. 0–15 kbar versus 0–30 kbar), which must be considered when comparing the performance of different models. Conversely, classification models with categorical prediction variables are evaluated using a contingency table (or confusion matrix) that compares absolute numbers of predicted and true conditions (i.e. true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN)). Commonly used performance measures for such models include the accuracy ((TP + TN) / (TP + TN + FN + FP)), precision (TP / (TP + FP)), sensitivity, i.e. true positive rate = TP / (TP + FN)), and specificity, i.e. true negative rate = TN/ (TN + FP)).

Several random forest-type models have been calibrated recently to predict pressures and temperatures from mineral chemistry and mineral–melt pairs (Petrelli *et al.*, 2020; Thomson *et al.*, 2021; Higgins *et al.*, 2022; Jorgenson *et al.*, 2022; Li & Zhang, 2022). In the first of these studies, Petrelli et al., (2020) calibrated cpx and cpx–melt thermobarometers using a wide range of different ML algorithms and obtained best performance metrics for an ERT algorithm. This model was subsequently optimised by Jorgenson *et al.* (2022), showing that removing outcomes with high variance in the aggregated trees can improve the accuracy of model predictions. A recent calibration of a random forest amphibole barometer (Higgins *et al.*, 2022) showed that reliable magma equilibration depths can be obtained using the often-maligned hornblende barometer, suggesting that machine learning methods can uncover patterns in data that are not captured by conventional linear regression approaches in thermobarometry. The adaptability of ML methods in this context is further exemplified by the successful calibration of unconventional mineral thermobarometers such as biotite (Li & Zhang, 2022).

METHODOLOGY

Data compilation

To calibrate thermobarometric models based on melt chemistry and associated mineral assemblages, we compiled 2545 phase equilibrium experiments from 127 published studies. The library of experimental phase relations (LEPR; Hirschmann *et al.*, 2008) was used as a starting point and augmented by an extensive literature search. The latter comprises 64% of the final database. Simple synthetic magma compositions, as well as non-terrestrial and alkaline magmas, were not considered in data mining and removed from the LEPR compilation. This choice was made to ensure that the compositions considered lie within the reduced component space of terrestrial magmas shown in Fig. 2. To ensure comparability of different melts, major element oxides (i.e. SiO_2, TiO_2, Al_2O_3, FeO_t, MgO, CaO, Na_2O, K_2O) were normalised to 100% anhydrous. We do not consider P_2O_5 and MnO in the normalisation and calibration of the model given that several experimental studies did not report one or both of these oxides, which would hamper the intercomparison of different melts. There is a high degree of variability in the literature when it comes to reporting stable mineral phase relations, especially with regard to accessory phases and the classification of Fe-Ti oxides. Therefore, we focus on 10 major mineral groups (ol, opx, cpx, plag, amph, bt, ksp, qtz, ox), whose stability is assigned to each experimental melt. In this scheme, oxide (ox) saturation is allocated to a melt if the experiment contained either magnetite, ilmenite or both. Experiments at one atmosphere pressure were not considered in the compilation, as they frequently show rapid disequilibrium crystal growth that may bias melt compositions (Mollo *et al.*, 2010; Ziberna *et al.*, 2017). To assess the impact of analytical uncertainty

on model performance, we further compiled 368 experimental melts that provided errors for each analyte. The full experimental data compilation used in this study is available in the electronic supplementary materials.

Extremely randomised trees—Random Forest algorithm

Random forest models were setup in the open-source programming language R, using the 'ranger' package, which provides a fast and flexible implementation suited for high-dimensional data (Wright & Ziegler, 2017).

To evaluate the model performance and compare the results of different testing strategies, we carried out two approaches:

1. *Random Splitting with Iteration*. In the first approach, we randomly partitioned the compilation of experiments into training and testing subsets, a process repeated 100 times (as depicted in Fig. 3c). To ensure that prediction variables (P–T) were evenly represented in both datasets, we implemented stratified random sampling. This involved dividing each prediction variable into quartiles and resampling them separately.

2. *Grouped Splitting*. The second approach involved dividing the experimental data into training and testing subsets while maintaining the integrity of complete experimental studies. We took special care to ensure that the testing subset covered the entire range of P–T conditions and, when possible, favoured studies with a lower number of experiments. This testing method was repeated three times.

The models were trained and tested using normalized (100% anhydrous) major element oxides (SiO_2, TiO_2, Al_2O_3, FeO_t, MgO, CaO, Na_2O, K_2O) in weight percent and the stability of specific mineral phases (olivine, orthopyroxene, clinopyroxene, plagioclase, amphibole, biotite, potassium feldspar, quartz, oxides) within each melt as input features. The stability of specific mineral phases was encoded as binary variables (1 for stable, 0 for unstable). Training of the model to predict P or T involved recursively splitting the experimental dataset into smaller compartments (nodes) based on randomly selected feature values (as illustrated in Fig. 3a). This process generated decision trees, where each node represented a specific subset of data defined by particular sets of conditions. Random Forests were ensembled from multiple decision trees ($n = 500$), each incorporating unique randomization due to feature selection and data sampling. Predictions of P–T conditions were then computed by passing a composition+mineral assemblage pair through all decision trees. The final P–T prediction is then computed by averaging the predictions of all decision trees for a particular composition+mineral assemblage pair (average vote, Fig. 3b).

After initial performance testing, in accordance with previous work (Petrelli *et al.*, 2020; Li & Zhang, 2022), best results were consistently achieved using an extremely randomised trees approach to random forests (Geurts *et al.*, 2006). We, therefore, use an implementation of this algorithm (ranger: splitrule = 'extraTrees') throughout this study. Sampling was carried out over the entire training or testing dataset without replacement, and nodes were split by choosing fully randomised split-points and variables (Geurts *et al.*, 2006). The number of random splits to consider for each candidate splitting variable was set to 10. Hyperparameter tuning did not significantly impact the results and was discarded, which is in-line with previous studies using igneous geochemical data (Petrelli *et al.*, 2020; Higgins *et al.*, 2022; Jorgenson *et al.*, 2022).

To evaluate the effectiveness of our models in retrieving information on mineral saturation conditions, we further setup three distinct sets of extremely randomized tree models using only melt composition as independent variables. These models serve the following purposes: 1) A classification algorithm designed to forecast the minimum count of saturated mineral phases based on melt composition. 2) A model with the objective of identifying the specific saturated phase assemblage, such as plag-opx-amph-ox. 3) A regression model tasked with estimating the quantity of saturated minerals based on liquid chemistry. These model sets were established to gauge the models' proficiency in handling mineral saturation information.

During evaluation we identified a systematic bias arising from the uneven distribution of training data within the overall P–T space. This bias manifested as overpredictions at lower pressures and underpredictions at higher pressures. To address this issue, we implemented the method proposed by Zhang & Lu (2012), employing a two-step random forest modelling approach using the parameters as previously described. Following the creation of the primary model, we calculated residuals by quantifying the disparities between actual and predicted values on the training dataset. These residuals served as the dependent variable in the subsequent model, utilizing the observed values of the target variable (P or T) and the variables of the initial model as independent predictors. For new or testing datasets we employed the first model to estimate P or T, and the second model to predict the residuals. These predicted residuals were then added to the initial model's output to yield the corrected value. Optimal results were achieved by applying an empirical scaling factor of 2 to the residuals, which was used throughout this study.

The R code used in this study is available in the supplementary materials. A user-friendly web application 'MagMaTaB' was developed using the R package 'shiny' and can be accessed at: https://igdrasil.shinyapps.io/MagmaTaBv4/

COMPILED VARIABLE SPACE

The compiled experimental melts span a wide geochemical spectrum from basalt to rhyolite (Table 1). For the subsequent calibration and application of thermobarometric models, it is crucial to establish whether these melts are representative of natural magma compositions in different tectonic settings. To accomplish this, we examine their variation relative to bulk-rock and glass compositions that have been pre-compiled in the GEOROC database for arc magmas ($n = 24\,593$), East African Rift ($n = 8546$), and Iceland ($n = 7696$). As shown in Fig. 4, the variation of different major element oxides shows that our compilation is consistent with the global arc magma array. Interquartile ranges (IQRs), as well as extreme values (i.e. upper and lower boxplot whiskers, Fig. 4) fully capture the variation. Rare arc magmas with exceptionally high MgO (>25 wt %), K_2O (>7 wt %), CaO (>18 wt %), or low Al_2O_3 (<5 wt %) were not included in the compiled dataset. Icelandic magmas are generally well represented by our compilation but can have higher TiO_2 (IQR, 1.3–2.6 wt %) and FeO contents (IQR, 10.1–14.4 wt %), as well as lower Al_2O_3 (IQR, 13.3–14.8 wt %), and are, therefore, under-represented by the experimental melts (IQR_{TiO2}, 0.4–1.0 wt %; IQR_{FeO}, 2.6–8.8 wt %; IQR_{Al2O3}, 14.9–18.0 wt %). Considerable overlap exists with East African Rift magmas, but the variation in several major element oxides such as Al_2O_3, CaO, Na_2O, and K_2O is greater in natural samples compared to the compiled dataset (Fig. 4). In summary, the compiled major element oxide space of experimental melts overlaps widely with natural rock compositions of different geotectonic association, showing greatest agreement with arc magmas. The dataset captures large parts of the compositional space for rift settings and Iceland, but the compilation and, therefore, the subsequent model calibration

Table 1: Optimization results of the meltassemblage and melt-only models

Statistic	Temperature	Pressure	SiO_2	TiO_2	Al_2O_3	FeO	CaO	MgO	K_2O	Na_2O
	(°C)	(kbar)	wt.%	wt.%	wt.%	wt.%	wt.%	wt.%	wt.%	wt.%
Min	675	0.2	44.6	0.0	5.6	0.3	0.1	0.0	0.0	0.2
10th Percentile	850	1.1	49.0	0.2	13.5	1.4	1.6	0.3	0.1	1.8
1st Quartile	900	2.0	51.1	0.3	14.7	2.5	2.9	0.7	0.4	2.5
Median	1020	4.1	58.7	0.6	16.5	6.2	6.7	3.2	1.4	3.4
Mean	1038	5.6	60.6	0.8	16.4	6.1	6.7	4.3	1.7	3.8
3rd Quartile	1160	10.0	70.7	1.0	18.0	8.8	10.1	6.6	2.7	4.3
90th Percentile	1280	12.0	74.3	1.4	19.3	10.7	12.0	10.1	3.9	4.9
Max	1400	15.0	79.9	5.4	24.1	20.9	17.5	19.5	6.9	7.3

Major element oxides normalised to 100% anhydrous without MnO and P_2O_5 Total number of experiments in the compilation n = 2639

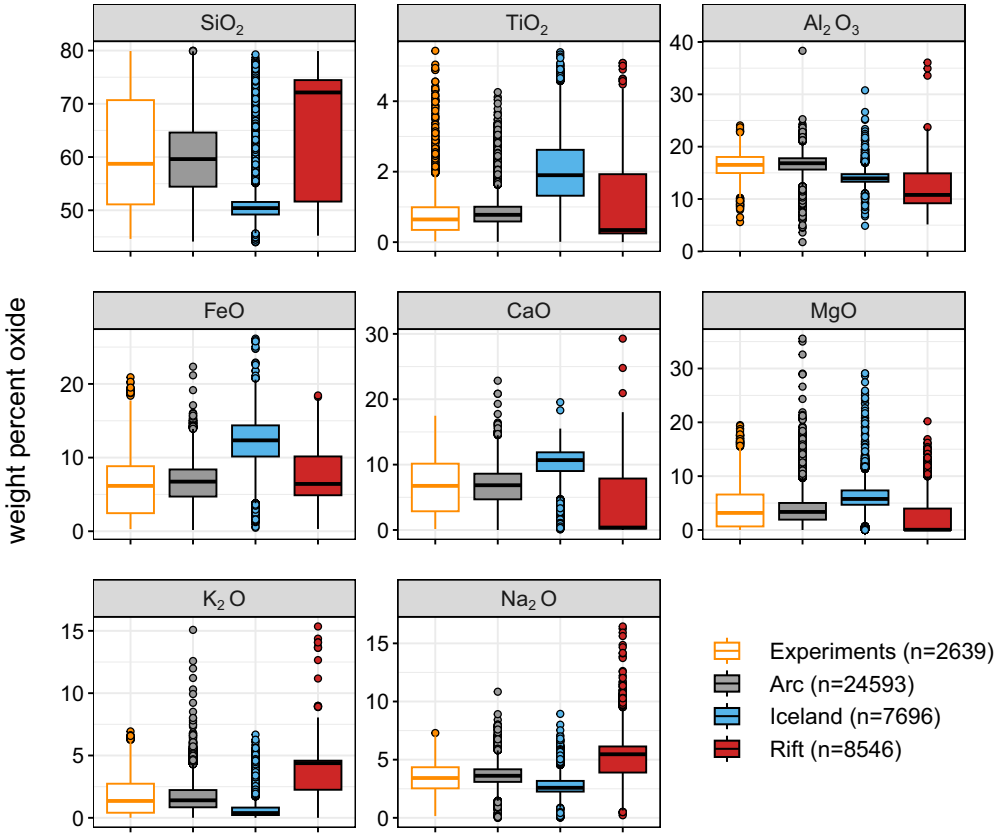

Fig. 4. Boxplots of major element oxides for compiled experimental melts (n = 2243; orange), compared to bulk-rock and glass compositions from subduction zones globally (Arcs; n = 24 593; grey), Iceland (n = 7696; blue), and the East African Rift (n = 8546; red). All natural rock compositions have been taken from precompiled files of the GEOROC database, were filtered between 44–80 wt.% SiO_2 and normalised to 100% anhydrous without MnO and P_2O_5. The Subduction Zones compilation include the: Aegean, Aleutian, Andean, Cascades, Honshu, Izu-Bonin, Kamchatka, Kurile, Lesser Antilles, Mariana, Mexican, New Zealand, and Sunda Arcs.

should not be applied to compositions outside the range (Min-Max) presented in Table 1 (see also Fig. S5).

Focussing on the comparison with arc magmas, Harker diagrams using SiO_2 (wt.%) as an index of magma differentiation, show that the compiled experimental melts are fully consistent with geochemical trends of magma evolution (Fig. 5a–f). The experimental melts span a continuum of compositions, following the curvilinear variation of TiO_2, Al_2O_3, and MgO with SiO_2 (Fig. 5b–d), along with linear behaviour of CaO, FeO, and alkalis (Fig. 5a, e–f). Minor deviations from the predominant trends are observed for intermediate magmas ($SiO_2 \sim 60$ wt %) with MgO >7 wt % and andesites with $Al_2O_3 < 13$ wt %, which are not captured by the experiments (Fig. 5c, d). Evolutionary trends in major element geochemistry for Icelandic magmas show overall equivalent behaviour (Fig. S1). The most striking differences are higher TiO_2 (>3 wt %) in mafic magmas, lower Al_2O_3 (<14 wt %) in intermediate melts, and generally higher FeO contents of Icelandic melts, which are described by a fairly low number density of experimental melts. East African Rift magmas display a much wider variability in geochemical behaviour, such as alkaline differentiation trends, or non-linear SiO_2–CaO variation, that is only partially captured in the experimental dataset (Fig. S2). As shown using colour-coding in Fig. 5, experimental magma evolution trends record a temperature control that, as expected, follows magma differentiation from basalt to rhyolite, but also tracks more subtle, stratified variation of temperature and mafic

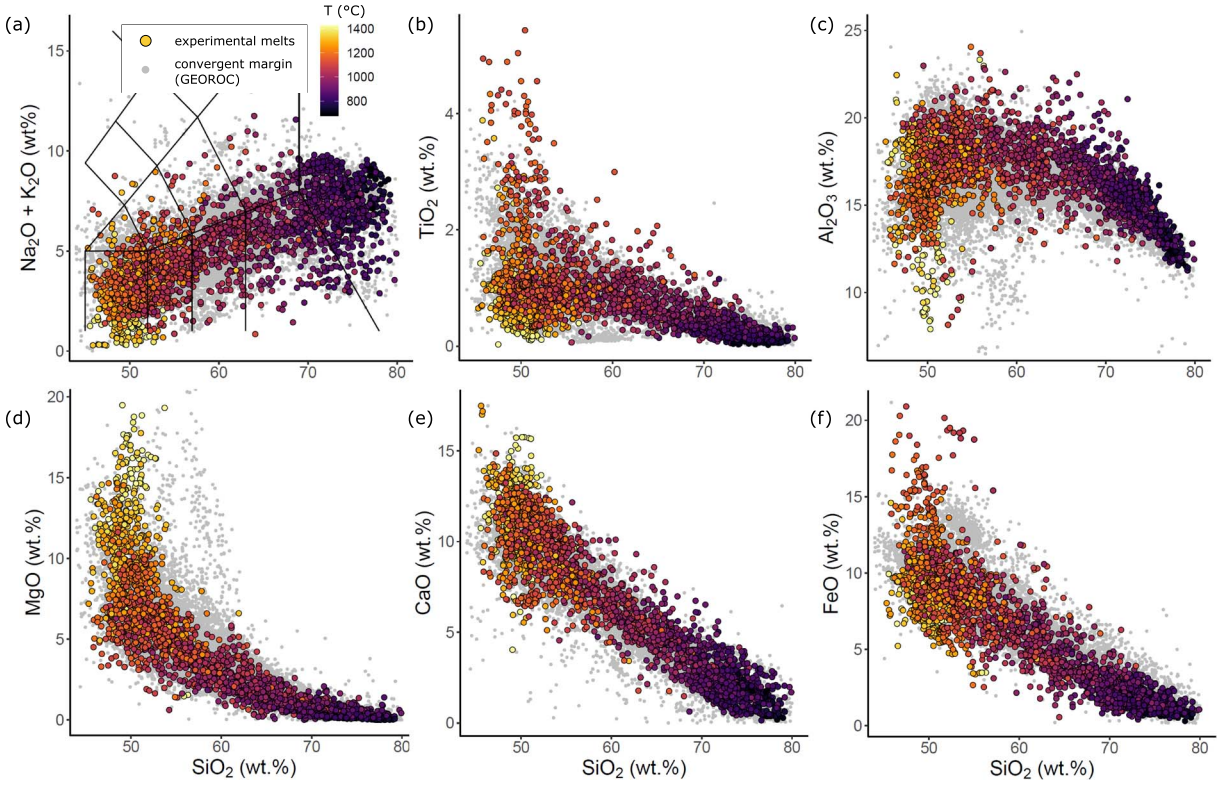

Fig. 5. Major element oxide (wt.%) versus SiO_2 (wt.%) in compiled experimental melts colour-coded for temperature (°C) compared to bulk-rock and glass compositions for global arc magmas (grey dots) from the GEOROC database. a) Total alkalis ($K_2O + Na_2O$) with compositional fields after Le Maitre *et al.* (2002). b) TiO_2, c) Al_2O_3, d) MgO, e) CaO, and f) FeO. All compositions were normalised to 100% anhydrous.

compositions such as low TiO_2 (<0.5 wt %) and high MgO (>10 wt %) for melt temperatures >1200°C.

The compiled experimental melts cover a temperature range of 675°C to 1400°C and pressures between 0.2 and 15 kbar (Fig. 6; Table 1). Temperature is distributed fairly symmetrically within this range, whereas pressures are right (positively) skewed. Experimental pressures below 2.5 kbar, typically the range that can be reached in cold-seal pressure vessels, comprise 41% of the data. Considering bins of 2 kbar, the P range between 2.5 and 10 kbar is typically covered by >200 to 400 experiments. The range from 10 to 15 kbar is covered by 676 experiments. As shown in Fig. 6a, low T data are sparse for high-pressure (>10 kbar) experiments, while high-T melts (>1200°C) are distributed overall uniformly with pressure. Temperature conditions between 800°C and 1200°C are well represented over the entire P-range. Phase relations in the compilation, considering only major minerals (ol, opx, cpx, plag, amph, bt, ksp, qtz, gt, ox), comprise a total number of 124 different assemblages. The most common mineral assemblages are broadly gabbroic: opx–cpx–plag–ox ($n = 208$), ol–cpx–plag ($n = 161$), plag–amph–ox ($n = 109$), and ol–cpx–plag–ox ($n = 104$). Figure 6 gives a good indication of the P–T regions that need to be explored experimentally to provide better calibration coverage. The full list of assemblages and their frequencies of occurrence is provided in Fig. S3.

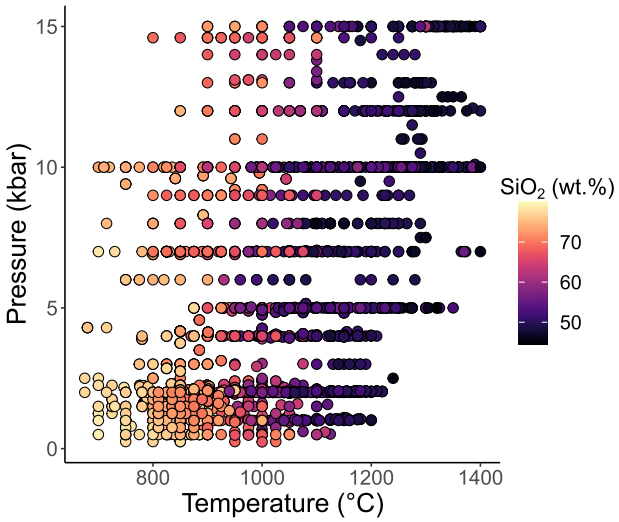

Fig. 6. Pressure (kbar) and temperature (°C) conditions of melts in the compilation of phase equilibrium experiments. Colour coding reflects the SiO_2 (wt.%) of the melt.

RESULTS

Model calibration and validation

The melt-assemblage model recovers experimental pressures with RMSE of 1.7 kbar on random split testing data and 1.9 kbar on grouped testing data (average of three splits) not used in the calibration (Fig. 7a, b). Temperature is predicted with RMSE of 36°C and 42°C on random and grouped testing data, respectively (Fig. 7c, d). Error distributions, quantified as difference between predicted and experimental variable (P or T), show modal values and medians centred at zero in each case, indicating a high accuracy of the predictions. The error distributions are overall symmetrical for T but show a slight tendency towards overprediction at low values and underprediction at high values for P.

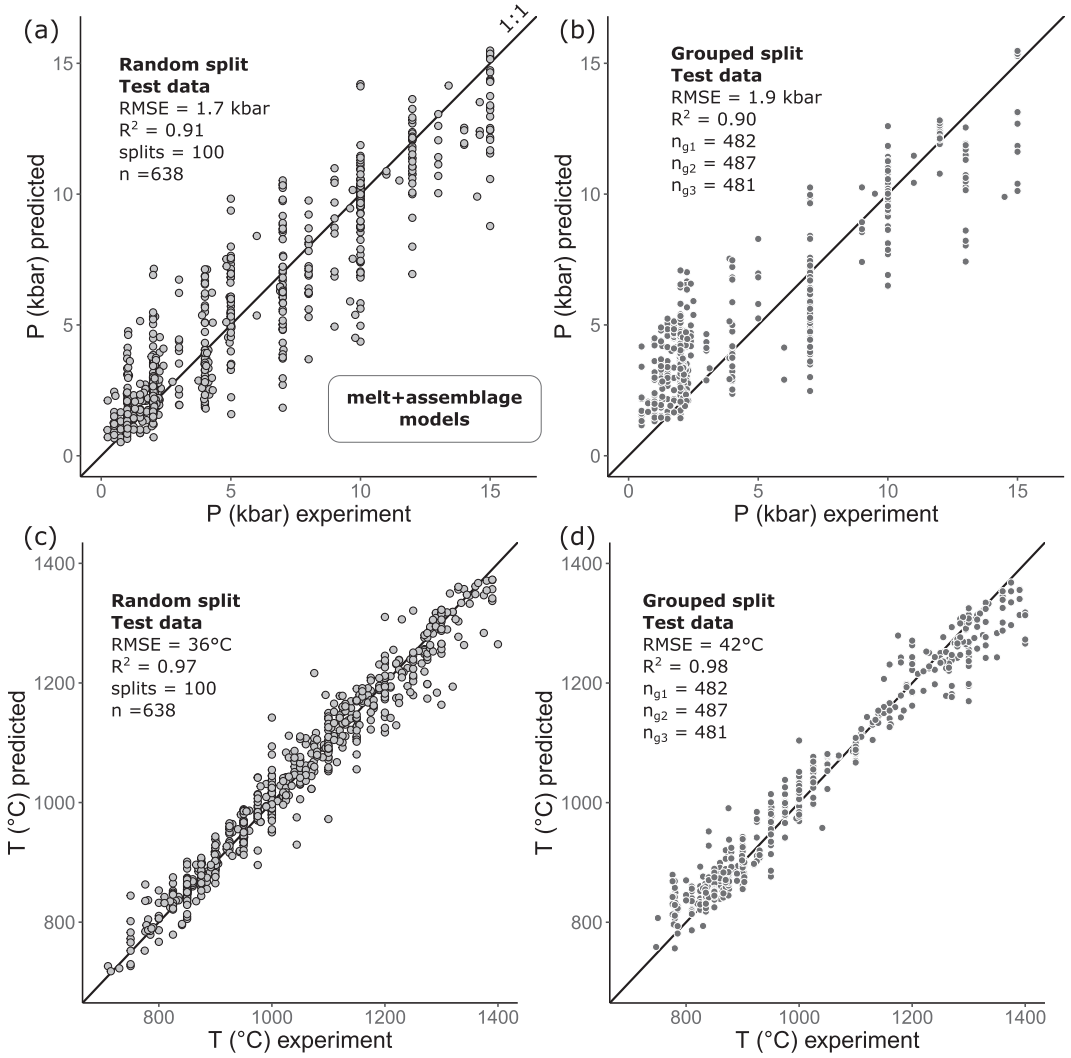

Fig. 7. Testing results of the melt+assemblage model for recovery of P and T. Two different testing approaches are shown: random split test data are presented as grey dots; grouped testing data (where entire experimental studies are kept together) are shown in black. Model performance metrics RMSE and R^2 (Pearson's correlation coefficient) are shown for the testing results, with number of test data denoted by n. Black solid lines illustrate a 1:1 relation. a) Experimental versus predicted pressure (kbar) for random split testing. b) Experimental versus predicted pressure (kbar) for grouped split testing. c) Experimental versus predicted temperature (°C) for random split testing. d) Experimental versus predicted temperature (°C) for grouped split testing.

We conducted an additional analysis using the algorithm, wherein a model was tested to predict P–T only based on melt composition, without specifying the associated mineral assemblage. However, this approach generates consistently higher RMSE values on testing data for each variable (Fig. 8). Pressures were recovered on randomly split and grouped testing data with RMSE of 2.3 and 2.2 kbar (Fig. 8a, b), and temperatures with RMSE of 41°C and 44°C (Fig. 8c, d). The shapes of error distributions are akin to those of the melt-assemblage model (Fig 8b, d, f). Despite, the systematically higher uncertainties of the melt-only model compared to the melt-assemblage model, the disparities are only moderately worse, testifying to the predictive power of the liquid phase composition. To adopt a conservative approach with generally lower prediction uncertainties, we focus solely on the melt+assemblage model for predicting P–T conditions.

In addition to P–T recovery, we set up models to constrain mineral phase saturation conditions based on melt chemistry, given that such information may be useful in model optimisation and for the general theoretical framework of liquid thermobarometry. For example, this method allows for an assessment of whether a given melt is multiply-saturated on its liquidus at the calculated source conditions. As described in the 'Methods' section, three sets of models were developed to recover saturation conditions: 1) A classification algorithm that predicts the minimum number of saturated mineral phases from melt composition. 2) A model aiming to identify the specific saturated phase assemblage (e.g. plag-opx-amph-ox). 3) A regression model that predicts the number of saturated minerals from melt chemistry. Based on melt composition, the algorithm can predict that a liquid is saturated with ≥3 minerals with average accuracy of 83%, precision of 84%, sensitivity (true positive rate) of 93%, and specificity (true negative rate) of 62% (Fig. 9a). While the performance scores for ≥4 saturated minerals show acceptable accuracy (76%), precision (72%), and specificity (87%), the sensitivity is decreased to 58%. Overall, Model 1 is most useful as a test if a specific liquid composition is saturated simultaneously with ≥3 minerals at its liquidus. A model to predict the particular phase assemblage in equilibrium with the melt (Model 2) returned low accuracy scores and is, therefore, not considered further. Lastly, estimating the number of mineral phases (Model 3) results in a symmetrical

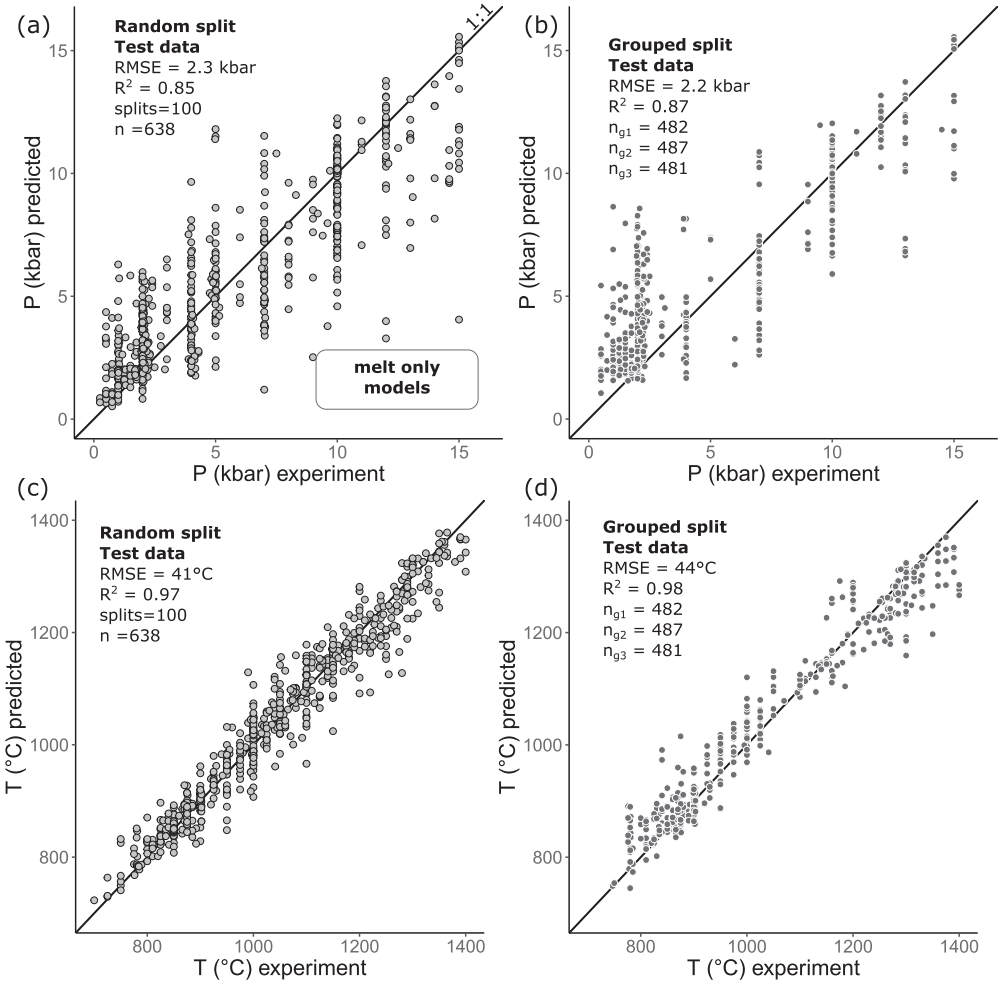

Fig. 8. Testing results of the melt-only model for recovery of P and T. Two different testing approaches are shown: Random split test data are presented as grey dots, while grouped testing data (where entire experimental studies are kept together) are shown in black. Model performance metrics RMSE and R2 (Pearson's correlation coefficient) are shown for the testing results, with number of test data are denoted by 'n'. All shown performance metrics are average of all run simulations. Black solid lines illustrate a 1:1 relation. a) Experimental versus predicted pressure (kbar) for random split testing. b) Experimental versus predicted pressure (kbar) for grouped split testing. c) Experimental versus predicted temperature (°C) for random split testing. d) Experimental versus predicted temperature (°C) for grouped split testing. Note the reduced precision of the melt-only method compared to the melt+assemblage method (Fig. 7).

error distribution and can be achieved with RMSE of 0.9 phases based on testing data not used for calibrating the model (Fig. 9b). In summary, the major element geochemistry of experimental melts records information on the number, but not the identity, of saturated mineral phases.

Optimisation of the modelling approach

From a thermodynamic standpoint, for a given number of chemical components, increasing the number of phases should reduce the degrees of freedom of the system. We, therefore, test if sub-setting the number (but not the identity) of co-saturated minerals can optimise the performance scores of our model. However, it must be noted that the performance is also subject to the effects of processes involving random number generation. Specifically, this involves initial splitting of the full dataset into calibration and testing subsets, as well as building extremely randomised trees. We tested the impact of random number generation on model runs by iterating the splitting operation 100 times, obtaining distributions of RMSE for the testing data (Fig. 10a–c). The results show that random processes induce variability in the RMSE of P–T. As expected, sub-setting the experimental dataset for different minimum numbers of stable mineral phases also impacts the performance of the model. The effect is strongest for T, showing a minimum median RMSE of 32°C for ≥3 minerals compared to a median RMSE of 38°C for the full dataset. Likewise, P-estimates show the lowest median RMSE of 1.6 kbar for saturation with ≥3 minerals (full dataset RMSE: 1.7 and 1.9 kbar on random and grouped split). Sub-setting the database comes, however, at the cost of reducing the number of experiments, leading to overall worse performance in models with ≥4 saturated mineral phases. Sub-setting is, therefore, only recommended to a certain degree (i.e. not more than three mineral phases) to optimise the model in cases with petrographic or circumstantial (e.g. co-genetic plutonic xenoliths) evidence that such conditions prevailed. Further experimental data for multiply-saturated (>4 minerals) melts would help to remedy this situation.

Better performance of random forests may also be achieved through adjusting model-specific parameters (Jorgenson *et al.*, 2022). Large improvements of the model performance can be achieved when a filtering approach based on the variance of the voting structure for individual P–T estimates is applied. Each P or T estimate is ultimately derived by averaging the results of

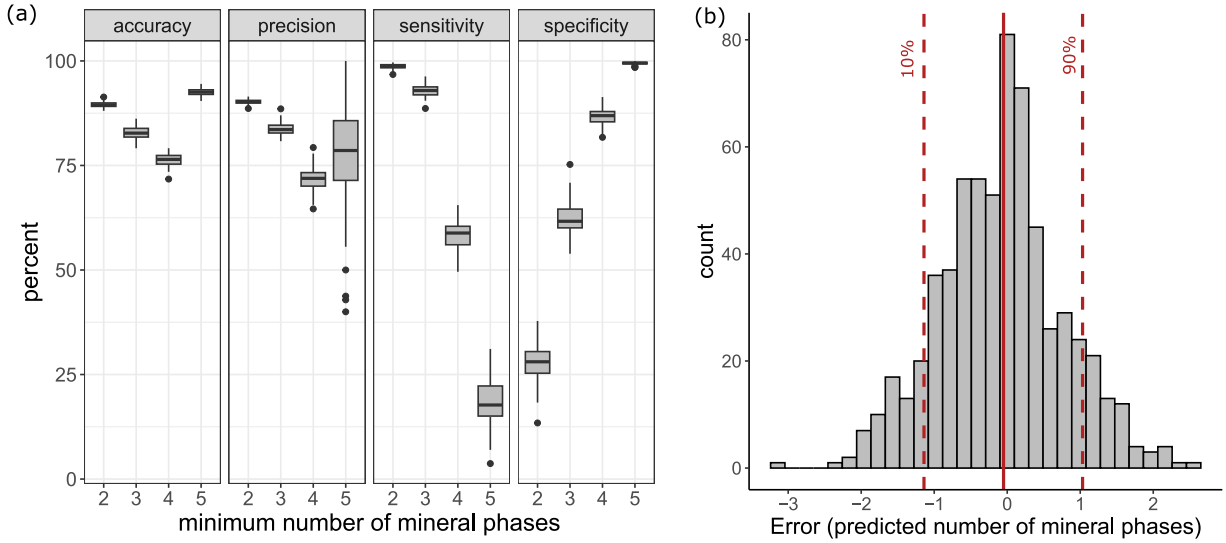

Fig. 9. Classification and regression models to predict mineral phase saturation conditions from experimental melt chemistry. a) Performance measures (accuracy, precision, sensitivity, and specificity) in percent for prediction of the minimum number of saturated mineral phases on random split testing data. See methods section for details on how performance measures are calculated. b) Histogram of error on the predicted number of phases based on random split test data. Red solid line marks the median and dashed lines indicate the 10th and 90th percentile.

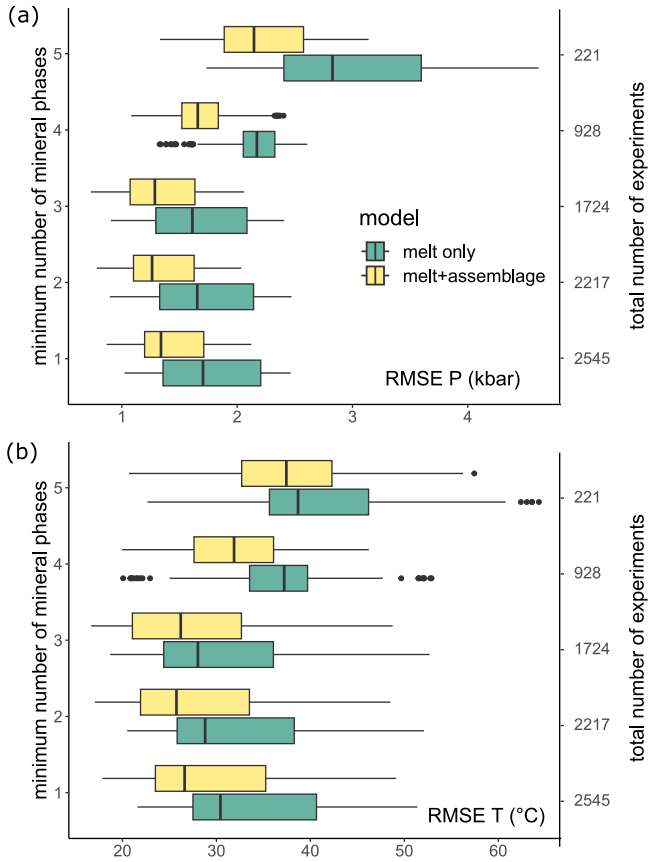

Fig. 10. Boxplots of RMSE calculated on testing data using 100 random splits of training and testing data and three grouped test splits. Results obtained with the melt+assemblage model are shown in yellow and melt-only model in green. Experimental data were subset for different minimum number of saturated mineral phases before splitting into training and testing sets. The number of experimental melts comprising the subset is shown on the right-hand side. a) RMSE of P (kbar), b) RMSE of T (°C).

500 decision trees. This means that for each estimate a standard deviation (SD) can be calculated that quantifies how variable the results of different trees are. As shown in Fig. 11a, errors in P determinations increase systematically with the SD of the voting distribution. We used the SD to quantify the voting variability and removed either 30% or 50% of the highest values for these measures (Table 2). Filtering out the top 30% of estimates with the highest SD reduces RMSE of P determinations down to 1.3 and 1.5 kbar for the random split and grouped melt-assemblage model, respectively. Filtering out 50% of the highest variance tree ensembles can reduce the RMSE further to 1.1 kbar (random split) and 1.3 kbar (grouped split). Estimates of T can be obtained with RMSE of 21°C (50% filtering). Importantly, filtering the estimates based on the variability of the voting distribution does not change the shape of the predicted variable distribution, showing that this approach does not systematically bias the results (Fig. 11b). In summary, as illustrated in Fig. 11c, more robust estimates of P–T can be obtained by the optimisation approach based on SD filtering but are traded off against the potential removal of a number of good estimates.

DISCUSSION

In this section, we discuss the performance and limitations of the model with respect to other thermobarometric calibrations, and the impacts of analytical uncertainties and the application of the model to glass and bulk-rock compositions are critically evaluated. We then apply the model to two volcanic systems in different tectonic environments, which have been chosen based on the availability of geophysical tomography and independent petrological constraints that can be used to cross-validate the results. For each system, we discuss how the model can be applied to map the source regions of magmatic liquids, chart their subsequent ascent paths and provide estimates of pre-eruptive magma storage conditions.

Model performance and limitations

A few factors need to be considered before comparing different thermobarometry models. The comparison should be based on a common measure of the model error that is ideally tested on data not used in the calibration. Given that widely used measures

Table 2: Optimization results of the meltassemblage and melt-only models

Filter	Test data	P (RMSE)				T (RMSE)			
	n	mean	sd	mean	sd	mean	sd	mean	sd
	%	kbar	kbar	kbar	kbar	°C	°C	°C	°C
melt-assemblage model									
Test data		Random[1]		Grouped[2]		Random		Grouped	
non	100	1.74	0.07	1.94	0.01	35.9	2.0	42.2	0.3
SD	70	1.30	0.06	1.58	0.02	25.1	1.5	32.1	0.6
SD	50	1.05	0.08	1.33	0.03	21.4	1.7	27.8	0.7
melt-only model									
Test data		Random		Grouped		Random		Grouped	
non	100	2.26	0.08	2.20	0.01	41.2	1.9	44.5	0.3
SD	70	1.65	0.09	1.86	0.03	29.7	1.6	34.0	0.5
SD	50	1.26	0.10	1.48	0.04	26.0	2.0	31.3	0.8

[1]Average of 100 random splits of training and testing data and decision tree generation
[2]Average of 3 splits of training and testing data and 100 random seeds in decision tree generation

of prediction accuracy, such as the RMSE or the equivalent standard error of estimate (SEE), are impacted by the width of the calibration dataset (scale dependency), it is imperative to establish that models are compared over a similar range in conditions (Hyndman & Koehler, 2006). For these reasons, other melt phase-assemblage thermobarometers, such as ol–plag–cpx(augite)–melt (OPAM; Yang *et al.*, 1996; Kelley & Barton, 2008; Hartley *et al.*, 2018) and melt–cpx–amph(hornblende)–opx–magneite–plag–ilmenite (CHOMPI; Blundy, 2022) are difficult to compare to our model. Sub-setting our calibration dataset to either of these assemblages, we calculate test data RMSEs of 1.5 ± 0.9 kbar (±2SD of the 100 random splits), 36 ± 15°C for OPAM and 2.8 ± 1.1 kbar, 43 ± 23°C for CHOMPI (mean of 100 splits into training and testing data). The high variance of these results reflects the low number of available experiments ($n = 92$ for OPAM and $n = 61$ for CHOMPI) in the compilation, indicating that the training dataset is not sufficient at present to calibrate these models.

We now compare the testing results of our model to errors cited in the literature. As shown in Table 2, our melt+assemblage model, calibrated for P ≤ 15 kbar, returns P–T with RMSE of 1.7 to 1.9 kbar and 36°C to 42°C. The given range reflects the different but overall similar results of random and grouped split testing. These errors can be reduced to 1.1 to 1.3 kbar and < 30°C through SD filtering (Table 2). Calibrated over a similar range (P < 15 kbar, T ~ 750–1300°C), and tested via random splitting, the 'crustal' ML cpx thermobarometer of Jorgenson *et al.* (2022), shows uncertainties on testing data of 2.3 kbar for P and 63°C for T based on cpx compositions only, which can be improved by approximately 0.5 kbar and 30°C if the composition of the melt in equilibrium with cpx is considered. If a wider pressure range is considered in the calibration, this model determines P–T with 3.2 kbar and 48°C (<30 kbar; Jorgenson *et al.* (2022)). The machine learning cpx-liquid thermobarometer of Petrelli *et al.* (2020) recovers P with error of 2.6 kbar and T with 40°C (<40 kbar), similar to the widely used cpx-liquid regression of Putirka (2008) that constrain P with RMSE of 2.6 kbar (Putirka's eqn. 32b) and T with 42°C (Putirka's eqn. 33) over the same P-range. Restricting the P range to <12 kbar and the compositional range to mafic and intermediate magmas, the cpx-liquid barometer of Neave & Putirka (2017) and cpx-only model of Wang *et al.* (2021) yield errors of 1.4 and 1.7 kbar, respectively. Putirka (2016) compared and calibrated different amphibole thermobarometers, concluding that T can be recovered with typical precision of 30°C, and P (<30 kbar) with at best 4 kbar on test data. Subsequently, Higgins *et al.* (2022) showed that the large errors of amph-barometry can be reduced to 1.6 kbar using a ML approach, calibrated for P < 12 kbar. Compared to established and new state-of-the-art mineral–melt and mineral-only thermobarometers, our model represents an enhancement in prediction accuracy for P–T conditions of crustal magmas over a wide range of compositions. Importantly, our approach can be applied effectively and reliably to whole-rocks, if these represent liquid compositions and provided that reasonable assumptions about the source mineralogy can be made (as discussed below). This makes our approach applicable to erupted whole-rock compositions as a probe of melt extraction depths, but also to melt inclusions trapped in a particular phenocryst phase or matrix glasses to gain insights into magma storage conditions. In summary, our approach has greater versatility than most conventional thermobarometers and can be applied to very large compositional datasets with relative ease.

As shown in Fig. 12, the prediction accuracy of the model can be improved based on a SD filtering. However, the accuracy of the model predictions is ultimately limited by the analytical error on major element oxide measurements from the calibrant experiments (e.g. Wieser *et al.*, 2023a). We, therefore, tested the impact of analytical uncertainty on P–T prediction, using a subset of experiments ($n = 368$) for which error estimates were available for each analyte, for example from multiple analyses of the same run product phase. Running the model using the accepted measured values, adding, and subtracting the error bounds, we quantify the deviation on predicted P–T values resulting from analytical uncertainties alone (Fig. 12). These calculations suggest that measurement errors limit P-determinations to ±0.8 kbar, and T-estimates to ±16°C. These are, in effect, estimates of the limit of precision that can be achieved using a melt composition approach calibrated on experimental petrology data. Improvements of the model accuracy of P–T determinations towards these limits should be possible in principle, for example by expanding strategically the experimental database. As illustrated in Fig. 6, the current experimental dataset overrepresents low-P conditions, most likely leading to the systematic over-prediction at low P and under-prediction at high P (Fig. 7, 8), which has been observed in other thermobarometric calibrations (e.g. Putirka, 2008; Neave & Putirka, 2017). Although systematic errors resulting from non-uniform distributions are potentially resolved by re-sampling the experimental data to a regular grid (Higgins *et al.*, 2022), reducing the number of experiments can lead to information loss. Clearly, further experimental work is needed to reinforce P determinations >3 kbar, especially at low T < 850°C, but also to expand the range of melt compositions to alkaline magmas (Figs. 5, 6).

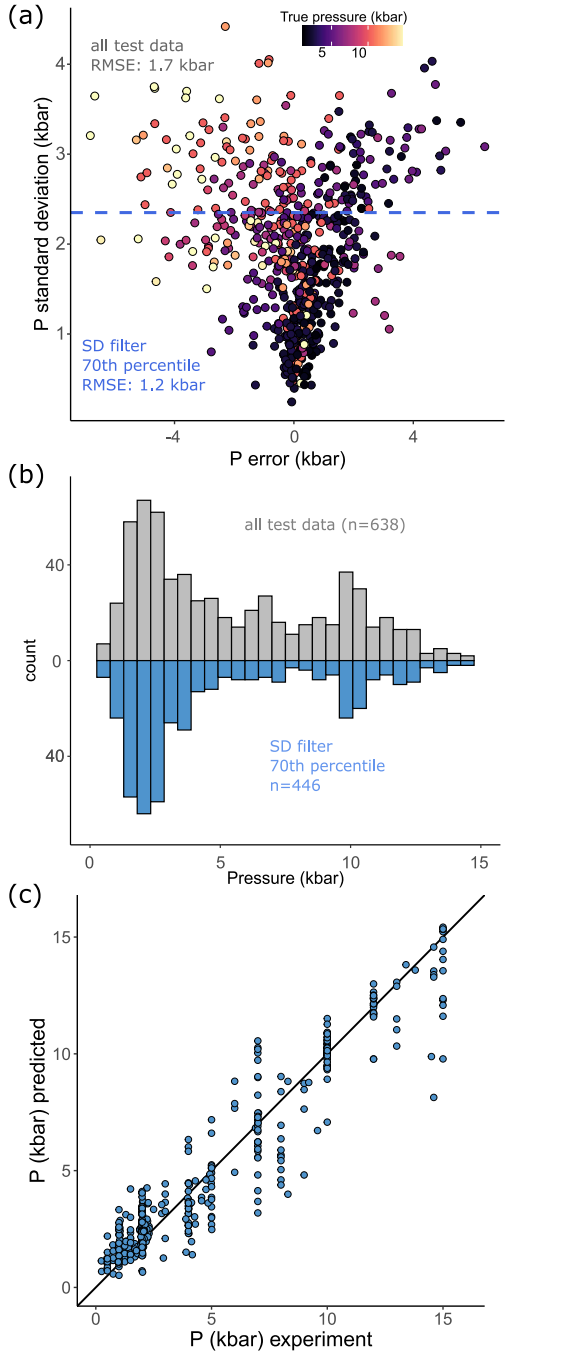

Fig. 11. Filtering approach to optimise model performance. a) P-error (kbar) is plotted against the standard deviation (SD) of individual tree ensembles that produce the average pressure vote for a particular melt composition. The blue dashed lines indicate a threshold value to filter out results with SD above the 70th percentile. Colour coding reflects true experimental pressure in (kbar). b) Histograms of predicted pressure without filtering (grey) and filtered predictions based on the SD threshold (blue). c) Predicted versus experimental pressure (kbar) for SD filtered results.

Moreover, a greater number of precise measurements of H_2O contents in experimental melts would be required to assess the feasibility of our approach as a reliable hygrometer.

Lastly, in accordance with the phase rule discussed above, the most accurate results are obtained when the calibration dataset is confined to liquids that are saturated with ≥3 mineral phases (Fig. 10). Although accurate results (RMSE: P < 2 kbar) can also be obtained when calibrating the model only on melts with ≤2 saturated minerals, we encourage the user to test saturation conditions using the classification and regression approaches shown in Fig. 9. Application of the model to bulk-rocks requires due diligence to establish that the composition represents a true liquid. Further targeted experimental work is likely to boost the performance of these models, given that the accuracy limit dictated by current analytical capabilities is not yet reached.

Constraining melt source conditions from bulk-rock compositions

A promising application for our model is that it can potentially be used to constrain P–T of melt source regions from bulk-rock compositions (Fig. 1; Gualda *et al.*, 2019; Blundy, 2022). This approach differs from constraints on equilibrium magma storage conditions based on melt-assemblage or mineral thermobarometers, in that it is based on mineralogical buffering of a melt phase in a crystal-rich, mushy source region. Field, petrological and geophysical evidence suggests that most magmatic systems are comprised of high crystallinity mushes with distributed interstitial melt (e.g. Bachmann & Bergantz, 2004, 2008; Cashman *et al.*, 2017; Edmonds *et al.*, 2019; Lissenberg *et al.*, 2019; Sparks *et al.*, 2019). As discussed in more detail by Blundy (2022), the volumetric dominance of crystals over melt in mush systems may buffer the composition of the melt via reactive flow, and extraction of the liquid phase with subsequent rapid ascent and quenching upon eruption can retain a chemical memory of P–T of the mush source. However, this approach is under the premise that the bulk-rock composition is representative of the extracted liquid. Processes that could modify the original melt composition include magma mixing and crystal accumulation, as well as fractional crystallisation (i.e. open system removal of solid phases) of the melt at various crustal levels. To constrain the impact of modifications to the original melt composition on P–T estimates, we added 5%, 10%, and 15% of crystals (ol, plag, opx, cpx, amph, ksp, bt) to experimental liquids of different bulk composition (basalt, basaltic andesite, andesite, dacite, rhyolite) in an attempt to simulate the impact of entrained crystal cargoes on recovered P–T. These calculations were performed using only the major element oxide chemistry of the melt composition as independent variable to isolate the effect of melt chemistry changes on P–T estimates. The results show that modification processes may lead to erroneous P–T predictions (up to 5 kbar and 176°C; Fig. S4). It is, therefore, important to verify if a particular bulk composition represents a true liquid composition. We suggest that the following points are useful to consider in establishing this:

1) *Geochemical variation:* Major and trace element variation frequently show evolutionary trends such as curvilinear behaviour in Al_2O_3, MgO, Zr, Ba, or P_2O_5. Magma mixing and crystal accumulation processes can lead to linear trajectories, deviating from this behaviour, which can help to identify such processes (e.g. Sparks & Marshall, 1986; Deering & Bachmann, 2010; Lee & Bachmann, 2014; Masotta *et al.*, 2016).
2) *Textural observations:* The method can be expected to work best for aphyric or crystal-poor magmas, given that these may represent extracted liquids that are unlikely to have experienced significant crystal accumulation.
3) *Mineral zoning:* Reverse zonation textures such as An-Fe-Mg rich plagioclase rims (Ruprecht & Wörner, 2007), Cr-rich rims in pyroxenes (Ubide & Kamber, 2018; Weber *et al.*, 2020)

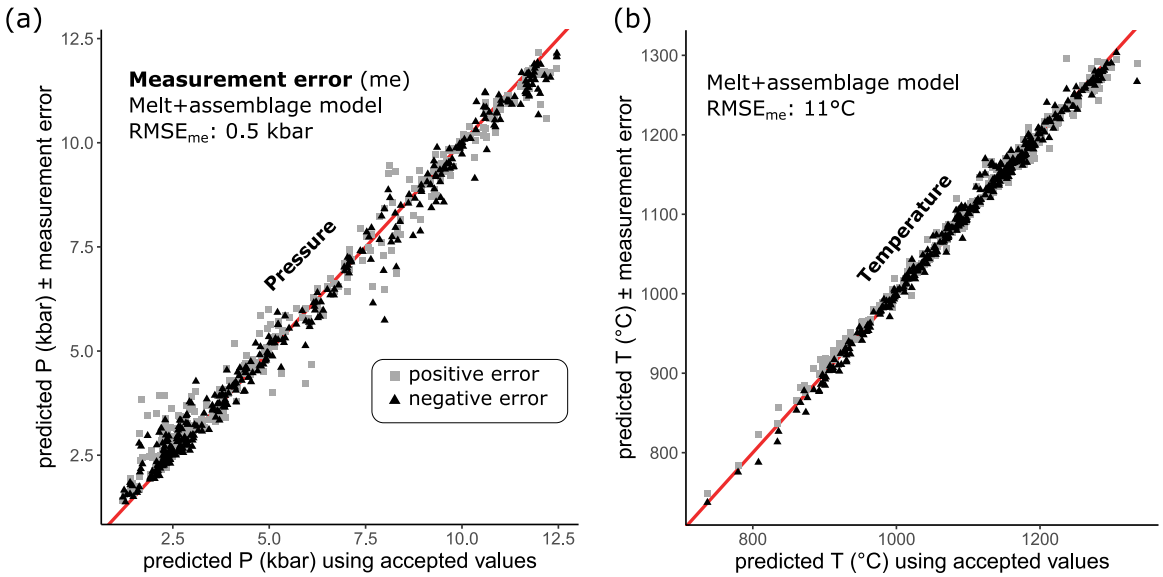

Fig. 12. Impact of analytical errors on results. Models were run using the accepted values and with reported errors for each major element oxide. Positive errors are shown as grey squares and negative errors as black triangles. a) Predicted P (kbar) using accepted value versus predicted P (kbar) ± analytical error, b) Results for temperature.

or Mg-rich rims in olivine (Gordeychik *et al.*, 2018) are all suggestive of magma mixing and re-equilibration processes. Bulk-rocks with crystal cargo dominated by this type of zoning are thus likely not representative of the original melt composition and should be avoided when using the melt-only model.

4) *Crystal–melt equilibrium:* The importance of open-system processes in the history of a bulk-rock may be interrogated by quantifying deviations from the relevant mineral–melt equilibrium constants (Ganne *et al.*, 2018). However, it should be borne in mind than an individual rock may have crystallised over a wide P–T range and this will be reflected in a diversity of mineral compositions that were unable to equilibrate fully with the host melt prior to eruption.
5) *Thermometry:* Mineral thermometry, such as from co-existing oxide or pyroxene pairs, and melt-only (source) thermometry should provide corresponding results within uncertainty, if no cooling has accompanied the pre-eruptive storage since melt extraction from the mush (Blundy, 2022).

Application of the model to bulk-rock composition is theoretically possible by relying solely on the major element chemistry of the melt. However, due to increased errors compared to the melt+assemblage approach and potential challenges in determining whether a bulk-rock composition accurately represents a melt, we adopt a more conservative approach. Our focus is on combining bulk rock compositions with an inferred source mineralogy. Implementing this approach necessitates knowledge about the source mineralogy from which the melt has been derived, information that is typically not known *a priori*. Note that source mineralogy is not necessarily the same as phenocryst mineralogy, as minerals stable in the source may not crystallise from the melt at shallower depths, amphibole being a common case in point (e.g. Davidson *et al.*, 2007). Despite these potential complications, reasonable assumptions about the source mineralogy can be made in many instances based on: 1) The phase assemblages of co-erupted mush fragments or plutonic xenoliths entrained in volcanic bulk-rock samples (e.g. Cooper *et al.*, 2019), 2) Reconstruction of crystallising mineral assemblages from major and/or trace element modelling of geochemical trends in igneous rock suites (e.g. Davidson *et al.*, 2007), and 3) Thermodynamic modelling of igneous phase equilibria through Gibbs Free Energy minimization approaches (e.g. Riel *et al.*, 2022).

Although estimates of source mineralogy require careful consideration on a case-by-case basis, our focus here is to quantify the general impact of misidentifying the source mineralogy. We systematically manipulated the true equilibrium mineral assemblage of experimental melt+assemblage pairs to assess the resulting impact on predicted P–T conditions. These calculations were conducted across various melt compositions, encompassing basalt, basaltic andesite, andesite, dacite, and rhyolite glasses, each in equilibrium with 3 to 5 mineral phases. To alter the true experimental mineral assemblage, we either imposed an entirely randomized set of mineral stabilities to a specific composition or, in a more optimistic scenario, alternated the stability of one of the mineral phases. The findings from this analysis reveals that while P–T estimates significantly degrade using a random mineral assemblage, the degradation is marginal when only one mineral is incorrectly inferred (Fig. 13). Notably, there is an exception for garnet misidentification, which exhibits a pronounced impact due to its high-pressure stability. However, assessing the involvement of garnet during magma evolution can be quite easily accomplished by other geochemical means, for example, through REE systematics (Harrison & Wood, 1980; Davidson *et al.*, 2007). In summary, exercising due diligence in petrological considerations is imperative for the application of our model to reconstructing melt extraction conditions in crustal crystal mushes.

Tracking magma sources and storage conditions

The 1980–86 eruption of Mt. St. Helens volcano

The iconic 1980–86 eruption of Mount St. Helens, a stratovolcano in the Cascades arc (USA) produced porphyritic dacites and silicic andesite (61–65 wt % SiO_2) containing phenocrysts of plagioclase, amphibole, orthopyroxene, magnetite, and ilmenite with scarce clinopyroxene (Rutherford *et al.*, 1985). The eruption initiated with a lateral blast on the morning of May 18th 1980, evolving to a < 19 km high Plinian eruption column and culminating in a

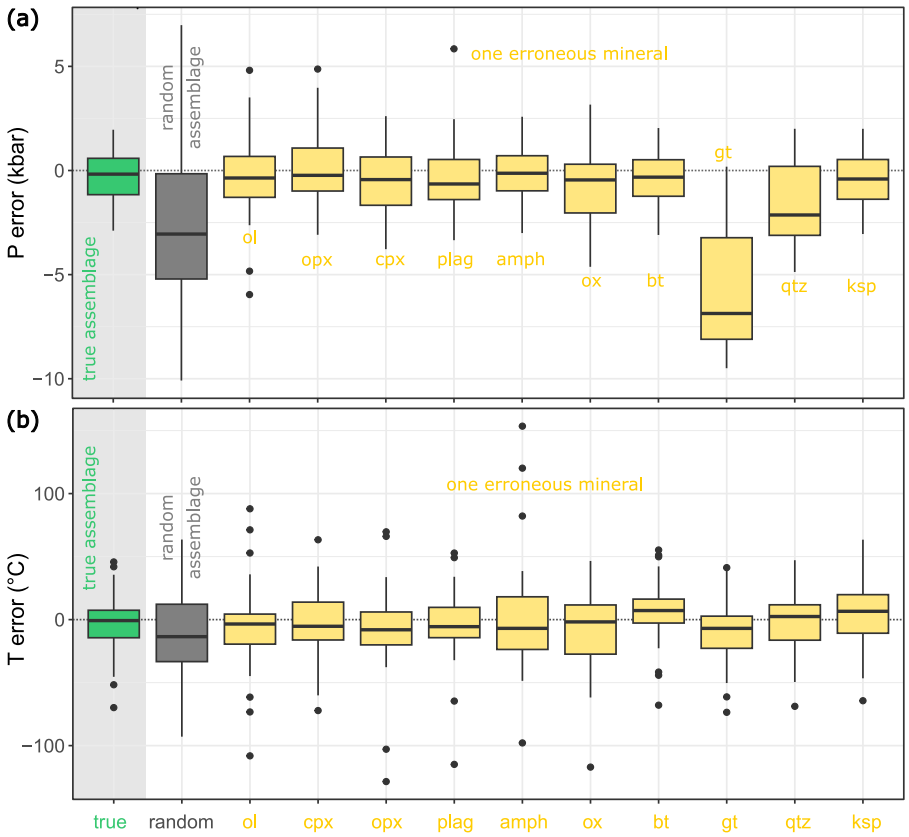

Fig. 13. Errors of P–T reconstructions resulting from misidentification of the mineral phase assemblage. Boxplots in green show errors using the true experimental mineral phase assemblage, grey boxplots show an entirely random mineral phase assemblage, yellow boxplots represent cases in which the stability of one of the mineral phases has been switched. a) Results for pressure in kbar. b) Temperature errors in °C.

series of dome-forming episodes through October 1986. Eruption temperatures, from coexisting Fe-Ti oxides from the entire eruption lie in the range 881 ± 54°C and fO_2 0.2 to 0.8 log units above NNO (Melson, 1983; Rutherford *et al.*, 1985; Blundy *et al.*, 2008). A large number of phenocryst-hosted melt inclusions analysed by SIMS record dissolved H_2O contents up to 6.8 wt % and CO_2 up to 1200 ppm, corresponding to calculated saturation pressures of <3 kbar (Blundy & Cashman, 2005; Blundy *et al.*, 2008; Cashman & Blundy, 2013). The overall range of melt inclusion volatile contents is consistent with volatile-saturated decompression crystallisation of dacite magma at near-constant temperature due to counterbalancing of adiabatic cooling by latent heat of crystallisation (Blundy *et al.*, 2006). Plagioclase phenocryst zoning is consistent with crystallisation over this P–T range (Cashman & Blundy, 2013).

A large body of experimental data aimed at reproducing the phenocryst assemblage and matrix glass compositions in the Plinian phase of the eruption indicate pre-eruptive magma storage pressures of 2.2 kb, with temperatures of 920°C and an fO_2 just above NNO (Merzbacher & Eggler, 1984; Rutherford *et al.*, 1985; Rutherford & Devine, 1988). Cogenetic plutonic xenoliths in 1980–86 eruptive products (Heliker, 1995; Blundy, 2022) are dominated by hornblende gabbronorites with interstitial glass testifying to the presence of such lithologies in the magma source region. Higher pressure (4–9 kb; Blatter et al., 2013) experiments suggest that dacites from a previous (3.5 ka) eruption of Mount St. Helens, similar in composition to 1980–86, were multiply-saturated on their liquidus with the peritectic assemblage hornblende-plagioclase-orthopyroxene-magnetite-ilmenite at 7 to 9 kbar, 925°C and 6 to 7 wt % dissolved H_2O. Blundy (2022) used these observations to justify application of his CHOMPI barometer to dacite bulk-rock compositions and recovered source conditions of 4.3 kbar, 950°C and 5.5 wt % dissolved H_2O. Moreover, there is generally a lack of mixing/mingling textures in the eruptive products, and a well-documented dominance of decompression and thermal effects controlling mineral zonation (Cashman & Blundy, 2013), suggesting that Mt. St Helens dacites whole-rocks can be treated as melt compositions. Geochemistry is consistent with generation of dacite magma by mid- to lower-crustal differentiation of arc basalt or basaltic andesite with minor crustal assimilation (Smith & Leeman, 1987; Wanke *et al.*, 2019).

Mount St. Helens is exceptionally well monitored and studied geophysically. Precisely located earthquake hypocentres extend to almost 20 km below surface and appear to define a high-velocity plug at 6 to 9 km depth overlying an aseismic region thought to be occupied by magma (Lees, 1992). However, the relatively small magnitude of wave-speed anomalies throughout the sub-Mount St. Helens crust is more consistent with a mush region with pockets of higher and lower melt fraction than discrete melt-rich magma chambers (Lees, 1992; Kiser *et al.*, 2016; Ulberg *et al.*, 2020). Magnetotelluric images (Hill *et al.*, 2015; Bedrosian *et al.*, 2018) tell a similar tale of mushy source regions with a well-developed mid-crustal conductive anomaly at depths of ≥15 km connected to the volcanic edifice by elongate conductive tendrils that appear to circumnavigate resistive bodies interpreted as cold, solid precursor intrusions (Bedrosian *et al.*, 2018). Long-period seismic events, consistent with fluid release, occur within the deep conductor at depths of 24 to 40 km below surface (Nichols *et al.*, 2011).

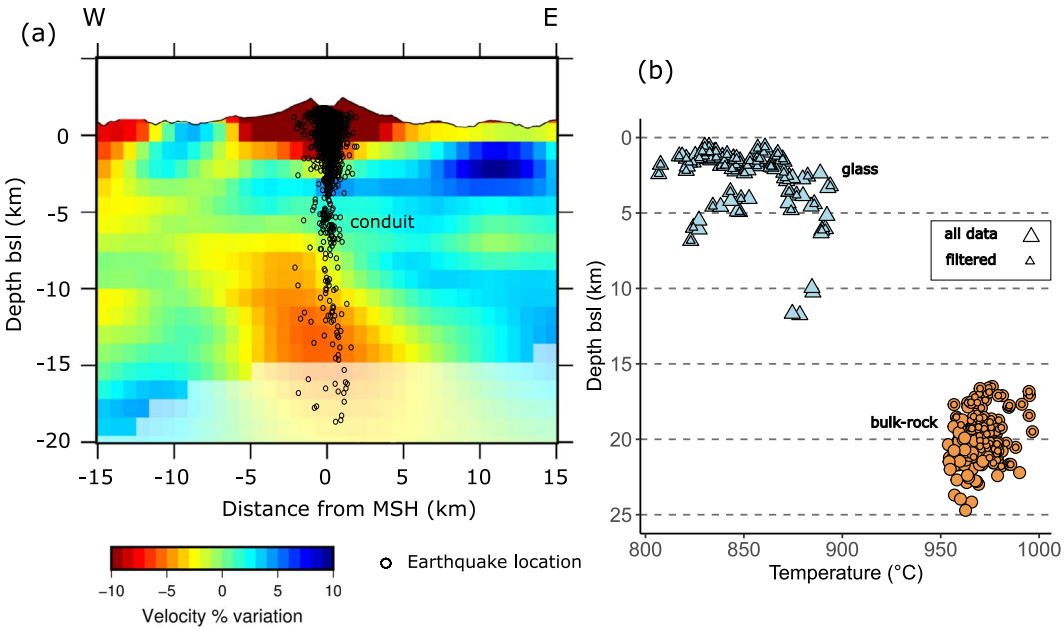

Fig. 14. Thermobarometry for Mount St Helens volcano. a) Comparison of seismic velocity variation (%) and earthquake hypocentre locations with P–T estimates for Mt. St. Helens. The seismic cross-section was modified from Ulberg et al., (2020) and earthquake locations within a 5 km radius of the summit were taken from the Pacific Northwest Seismic Network database (https://pnsn.org/volcanoes/mount-st-helens). Histograms show b) depth below sea level (bsl) versus temperature estimates for calculations involving bulk-rock (orange circles) and glasses (light blue triangles) paired with plag-opx-cpx-amph-ox assemblages. Larger symbols show all calculation results; smaller symbols are results that have been filtered using the SD filtering approach (50th percentile). Pressures were converted to depth by calculating a density profile from Vp velocities within 30 km of St Helens (Ulberg *et al.*, 2020) using the equation of Brocher (2005).

We applied the thermobarometer to a compilation of whole-rock dome lavas and pumices ($n = 42$) from the 1980–86 eruption in apparent equilibrium with the assemblage hornblende–plagioclase–orthopyroxene–magnetite–ilmenite on the basis of abundant hornblende gabbronorite xenoliths (Heliker, 1995; Blundy, 2022). Pressure calculations were carried out either without filtering or with SD filter of 0.5. The minimum number of saturated phases was set to either 1 or 3, showing overall similar results. Whole rock+assemblage pairs yield pressures of 5.1 to 8.8 kbar, which reduces to 5.1 to 6.7 kbar when the filter is applied. Temperatures for unfiltered bulk-rock samples are recovered with 954°C to 997°C (unfiltered) and 957°C to 997°C when the SD filter is applied (Fig. 14). Application of the model to melt inclusions and groundmass glasses ($n = 48$) yields pressures between 0.7 and 3.7 kbar, and temperatures of 807–894°C (mean: 851°C). All melt-assemblage calculations were carried out using opx–cpx–plag–amph–ox. The calculated pressures equate to depths of 16 to 22 km for SD filtered whole-rocks, which we interpret as dacite melt extraction depths and 0.6 to 12 km for melt inclusions and groundmass glasses, which we equate to decompression crystallisation during ascent, including brief periods of pre-eruptive storage. Depth was determined from pressures by utilizing a density profile derived from Vp velocities within a 30-km radius of St. Helens (Ulberg *et al.*, 2020) and employing the equation developed by Brocher (2005). The temperature change from whole-rocks (954–997°C) to melt inclusions (807–894°C) is suggestive of some cooling during ascent and pre-eruptive storage. The depth results are consistent with the geophysical data presented above (Fig. 14).

Overall, our results support a petrogenetic model for Mount St. Helens whereby dacite melts are generated within a laterally extensive, mid-crustal reservoir of electrically conductive mush and ascend to the surface via a narrow conduit region, crystallising, and degassing en route with cooling of up to ~150°C. This scenario is consistent with that proposed by Blundy (2022) for the Cascades arc more generally. Wieser *et al.* (2023b) calculate median pressures for Mount St. Helens dacites of 2 to 3 kbar based on amphibole and cpx barometry, which lie within the range we calculate for melt inclusions and groundmass glasses. Wieser *et al.* (2023b) interpret their pressures in terms of 'magma storage conditions'. However, in the absence of time information, e.g. from diffusion chronometry, these pressures may equally reflect transient points on the magma ascent trajectory where crystallisation of amphibole and cpx occurred. Given that all magmas ascending from a deep source region, as estimated using our methodology, will pass through and crystallise under lower pressures, caution is required in directly ascribing barometry results with magma storage *per se*.

Historic eruptions from the Askja caldera (Iceland)

The Askja volcanic system in Iceland provides an ideal test case to integrate petrological barometry with the results of geophysical tomography. Askja is located in the Icelandic rift zone and was frequently active in the early 20th century, producing at least six small-volume basaltic eruptions in the vicinity of the central volcano (Hartley & Thordarson, 2013). Notably, in 1875 Askja produced a major explosive rhyolite eruption, one of Iceland's largest in the historical record, creating a 4.5-km wide collapse caldera that was subsequently filled by Lake Öskuvatn (Hartley & Thordarson, 2013). Seismic tomography, using local earthquake locations, provides a snapshot of the crustal magmatic feeding system beneath Askja, revealing low-velocity structures at multiple levels beneath the caldera (Greenfield *et al.*, 2016). The volcano is also actively deforming, showing deflation over timescales of years to decades and a strong surface uplift pattern centred at about 3 km depth based on Mogi modelling (Zeeuw-van Dalfsen *et al.*, 2012). It is currently not known how these geophysical features of the volcanic plumbing system relate to the compositional

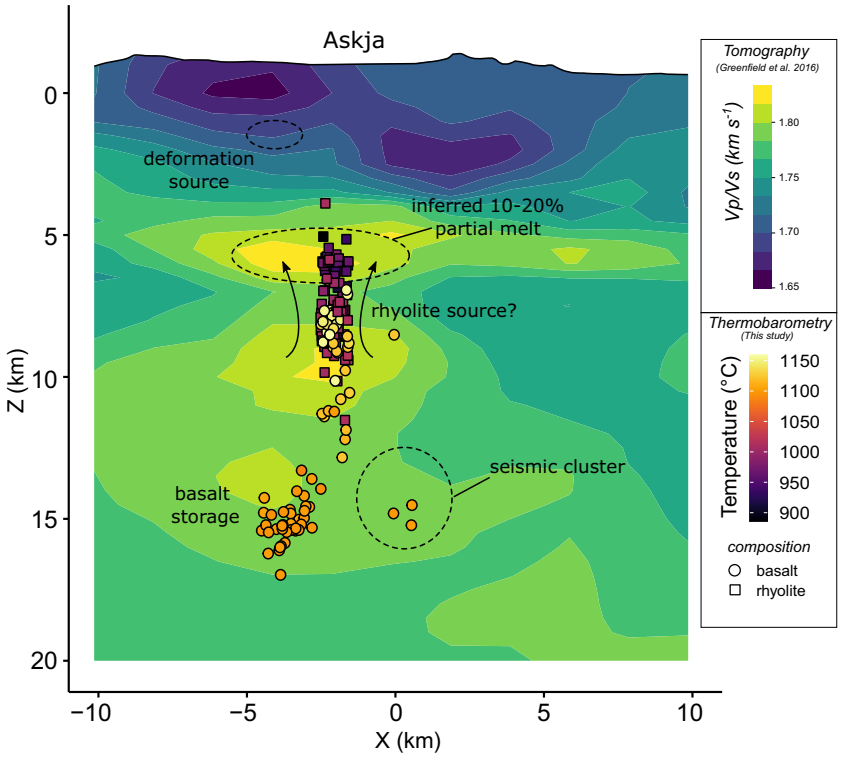

Fig. 15. Comparison of thermobarometric estimates and seismic tomography for the Askja caldera (Iceland). The seismic velocity structure of the crust beneath Askja parallel to the spreading direction was taken from Greenfield *et al.* (2016), with colours reflecting the ratio of P- and S-waves (Vp/Vs) in km s-1. P–T estimates have been orthogonally projected from the vent location onto the 2D crustal section and are shown as circles for basalts and squares for rhyolites. Crustal density for Askja was calculated from seismic data of Greenfield *et al.* (2016) using the relation between wave speeds and density of Brocher (2005).

diversity and storage depths of the bimodal magma associations erupted historically at Askja.

We used the melt-assemblage thermobarometer on glass compositions for 20th century Askja basalts and 1875 rhyolites utilizing data published in Hartley *et al.* (2013) and Schattel *et al.* (2014) (Supplementary tables). The basalts have phase assemblages of ol–cpx–plag–ox (Hartley *et al.*, 2013) and 1875 rhyolites contain opx–cpx–plag–ox (Schattel *et al.*, 2014). Both basalts and rhyolites are nearly aphyric, with phenocrysts in textural equilibrium exhibiting normal or no zoning, and show geochemical trends consistent with magma evolution (Hartley *et al.*, 2013; Schattel *et al.*, 2014; Fig. S1). Minor andesite glasses erupted in 1875 are not considered as these most likely represent mixed compositions, as suggested by SiO_2–Al_2O_3 divergence from the general evolutionary trend (Fig. S1). Although not all Icelandic magmas are covered by our experimental compilation, Askja basalts and rhyolites are captured in most major element oxides, with the exception of TiO_2 and Al_2O_3, which are only marginally represented (Fig. S1). Nevertheless, Askja provides an ideal and contrasting case to test the model on compositions that are not overrepresented experimentally. Temperature reconstructions for the basaltic melts using our model yield averages of 1085°C to 1159°C (mean, 1117°C; SD filter = 0.7). No temperature estimates are currently available for these eruptions, but the calculated temperatures are at the lower end of estimates for older lava flows from the Askja fissure swarm based on cpx-liquid thermometry (Hartley *et al.*, 2013). Our temperature estimates for the 1875 rhyolites (886–1022°C; mean, 987°C) corroborate previous thermometry (940–1008°C; Schattel *et al.*, 2014), showing that these melts were unusually hot. These results support the hypothesis that melting of hydrothermally altered crust, rather than protracted fractional crystallisation, play a key role in the origin of Askja rhyolites (Bindeman *et al.*, 2012).

Depth estimates using our model have been projected orthogonally onto the seismic tomography section of Greenfield *et al.* (2016), showing excellent correspondence with P-/S-wave velocity (Vp/Vs) anomalies (Fig. 15). Glass-assemblage pairs of rhyolites record depth estimates between 4 and 12 km below sea level (bsl). Pressures were converted into depth based on the empirical relationship between wave speeds and density of Brocher (2005), using seismic data from Greenfield *et al.* (2016). These depths correspond to a major zone of reduced wave speeds centred at 6 km and a smaller Vp/Vs anomaly located at 9 km bsl, the upper reaches of which have been inferred to contain about 10% to 20% of partial melt on average (Greenfield *et al.*, 2016). Basaltic glass compositions return storage depths between 7 and 17 km (mean, 12 km bsl), corresponding to the lower part of the Vp/Vs anomaly and a region of active seismicity, which has been interpreted as a zone of melt movement (Greenfield *et al.*, 2016). The storage depth range is narrower compared to previous cpx-liquid barometry on basaltic composition from historical Askja eruptions that indicated a very wide range between 2 and 18 km depth (Hartley *et al.*, 2013). However, while cpx-liquid barometry may suffer from large errors (Wieser *et al.*, 2023a), the majority of these estimates (~11–14 km bsl) overlap with our constrained storage depth. Depth estimates within the range of shallow crustal deformation are absent in our calculations, which may reflect that no eruptions have been sourced from this depth in historical time. Our approach, marked by its remarkable convergence with seismic tomography, offers a unique perspective on the storage depths of magmatic systems, allowing us to assign chemical compositions and temperatures to specific crustal magma bodies beneath the Askja caldera system,

which is critical to our understanding of magmatic processes and assessment of volcanic hazards.

CONCLUSIONS

The ability to quantify magmatic system variables from igneous rocks is invaluable to igneous petrology. However, thermobarometric models are often limited by large errors on pressure determinations and narrow calibration ranges. In this study, we have addressed this need by calibrating a widely applicable ERT random forest machine learning model to reconstruct magmatic pressures, and temperatures with increased predictive accuracy. The model utilizes glass or bulk-rock chemistry together with mineral paragenesis as independent variables, both of which are readily available for most volcanic systems. We conclude that:

1) The major element oxide chemistry of melts (±mineral phase assemblage) is a powerful recorder of P–T for crustal magmas over a wide range of multiple mineral saturation conditions and throughout the compositional spectrum from basalt to rhyolite. This empirical finding can be attributed to a more flexible regression strategy of large experimental datasets but may be theoretically grounded in the limited degrees of freedom of correlated oxide components in natural magmas.
2) Thorough error assessment of the model, using either grouped or 100 random splits of training and testing datasets, indicate that P can be recovered with RMSE of 1.7 kbar on random split and 1.9 kbar on grouped testing data. Comparison of model predictions with independent estimates shows excellent agreement for hot-rhyolites (~1000°C), which are scarcely included in the experimental calibration.
3) Prediction errors can be significantly reduced by filtering out P–T estimates with standard deviation above the 50th percentile in the predictions of decision trees that comprise the random forest (RMSE of 1.1 kbar on random split and 1.3 kbar on grouped test data; Table 2). Better results are also obtained when the calibration data are restricted to a minimum of ≥3 saturated mineral phases, but the effect is small compared to filtering. The model could be further improved through additional experiments in areas that are currently underrepresented in P–T space.
4) Calculations using pairs of glass chemistry and mineral assemblage can be used to reconstruct the last equilibration conditions of crustal magmas. These may represent pre-eruptive storage depths, if additional time information is available, or simply points on the P–T ascent trajectory where crystallisation occurred. The model also provides the opportunity to constrain mush extraction conditions (magma sources) under the premise that bulk-rock or glass compositions are representative of the extracted melt. The latter should be tested based on a set of petrographic and geochemical criteria.
5) Our ERT random forest algorithm allows to better integrate petrology and geophysical imaging of magmatic feeding systems. Application of the model to Mount St. Helens volcano indicates dacite sourcing from a mid-crustal mush followed by decompression crystallisation and degassing accompanied by cooling. Barometric calculations for historical eruptions from the Askja volcanic system in Iceland allow assignment of rhyolite and basaltic compositions to specific seismic velocity anomalies beneath the volcano.

AUTHOR CONTRIBUTIONS

G.W. participated in the conceptualization, investigation, visualization, formal analysis, methodology, software, writing- original draft, funding acquisition. J.B. participated in the conceptualization, investigation, funding acquisition, writing- review and editing.

DATA AND CODE AVAILABILITY

All data and the R code used in this study are available in the supplementary materials to this article. The code can also be accessed as user-friendly web application with full control over model specific parameters at: https://igdrasil.shinyapps.io/MagmaTaBv4/

SUPPLEMENTARY DATA

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

FUNDING

JB acknowledges support from a Royal Society Research Professorship (RP\R1\201048). GW acknowledges support from a SNSF early postdoc mobility fellowship (P2GEP2 195238).

ACKNOWLEDGEMENTS

We thank Madeleine Humphreys for editorial handling, Susan DeBari and an anonymous reviewer for constructive comments that improved the quality of this work. We would like to thank Ollie Higgins for sharing his insights on random forest thermobarometry, especially for valuable suggestions on how to optimise such models. Luca Caricchi is thanked for stimulating discussions on machine learning and bias correction methods. We thank Lena Melekhova for kindly sharing her experimental compilation. Carl Ulberg is thanked for his prompt and generous responses to e-mail requests for data and for sharing an editable version of his MSH seismic section. Penny Wieser is acknowledged for suggesting to test the model on grouped data.

References

Ague, J. J. (1997). Thermodynamic calculation of emplacement pressures for batholithic rocks, California: implications for the aluminum-in-hornblende barometer. *Geology* **25**(6), 563–566. https://doi.org/10.1130/0091-7613(1997)025%3c0563:TCOEPF%3e2.3.CO;2.

Aitchison, J. (1984). Reducing the dimensionality of compositional data sets. *Journal of the International Association for Mathematical Geology* **16**, 617–635. https://doi.org/10.1007/BF01029321.

Andújar, J. & Scaillet, B. (2012). Relationships between pre-eruptive conditions and eruptive styles of phonolite–trachyte magmas. *Lithos* **152**, 122–131. https://doi.org/10.1016/j.lithos.2012.05.009.

Atkins, P. & De Paula, J. (2013) *Elements of physical chemistry*. USA: Oxford University Press.

Bachmann, O. & Bergantz, G. W. (2004). On the origin of crystal-poor rhyolites: extracted from batholithic crystal mushes. *Journal of Petrology* **45**(8), 1565–1582. https://doi.org/10.1093/petrology/egh019.

Bachmann, O. & Bergantz, G. W. (2008). Rhyolites and their source mushes across tectonic settings. *Journal of Petrology* **49**(12), 2277–2285. https://doi.org/10.1093/petrology/egn068.

Bedrosian, P. A., Peacock, J. R., Bowles-Martinez, E., Schultz, A. & Hill, G. J. (2018). Crustal inheritance and a top-down control on arc magmatism at mount St Helens. *Nature Geoscience* **11**(11), 865–870. https://doi.org/10.1038/s41561-018-0217-2.

Bindeman, I., Gurenko, A., Carley, T., Miller, C., Martin, E. & Sigmarsson, O. (2012). Silicic magma petrogenesis in Iceland by remelting of hydrothermally altered crust based on oxygen isotope diversity and disequilibria between zircon and magma with implications for MORB. *Terra Nova* **24**(3), 227–232. https://doi.org/10.1111/j.1365-3121.2012.01058.x.

Blundy, J. (2022). Chemical differentiation by mineralogical buffering in crustal hot zones. *Journal of Petrology* **63**(7), egac054. https://doi.org/10.1093/petrology/egac054.

Blundy, J. & Cashman, K. (2001). Ascent-driven crystallisation of dacite magmas at mount St Helens, 1980–1986. *Contributions to Mineralogy and Petrology* **140**, 631–650. https://doi.org/10.1007/s004100000219.

Blundy, J. & Cashman, K. (2005). Rapid decompression-driven crystallization recorded by melt inclusions from mount St. *Helens volcano. Geology* **33**(10), 793–796.

Blundy, J. & Cashman, K. (2008). Petrologic reconstruction of magmatic system variables and processes. *Reviews in Mineralogy and Geochemistry* **69**(1), 179–239. https://doi.org/10.2138/rmg.2008.69.6.

Blundy, J., Cashman, K. & Humphreys, M. (2006). Magma heating by decompression-driven crystallization beneath andesite volcanoes. *Nature* **443**(7107), 76–80. https://doi.org/10.1038/nature05100.

Blundy, J., Cashman, K. V., Berlo, K., Sherrod, D. R., Scott, W. E. & Stauffer, P. H. (2008). Evolving magma storage conditions beneath Mount St. Helens inferred from chemical variations in melt inclusions from the 1980–1986 and current (2004–2006) eruptions. *US Geological Survey professional paper* **1750**, 755–790.

Brandmeier, M. & Wörner, G. (2016). Compositional variations of ignimbrite magmas in the Central Andes over the past 26 Ma—a multivariate statistical perspective. *Lithos* **262**, 713–728. https://doi.org/10.1016/j.lithos.2016.07.011.

Breiman, L. (2001). Random forests. *Machine learning* **45**, 5–32. https://doi.org/10.1023/A:1010933404324.

Brocher, T. M. (2005). Empirical relations between elastic wavespeeds and density in the Earth's crust. *Bulletin of the Seismological Society of America* **95**(6), 2081–2092. https://doi.org/10.1785/0120050077.

Cashman, K. & Blundy, J. (2013). Petrological cannibalism: the chemical and textural consequences of incremental magma body growth. *Contributions to Mineralogy and Petrology* **166**, 703–729. https://doi.org/10.1007/s00410-013-0895-0.

Cashman, K. V., Sparks, R. S. J. & Blundy, J. D. (2017). Vertically extensive and unstable magmatic systems: a unified view of igneous processes. *Science* **355**(6331), eaag3055. https://doi.org/10.1126/science.aag3055.

Cassidy, M., Manga, M., Cashman, K. & Bachmann, O. (2018). Controls on explosive-effusive volcanic eruption styles. *Nature Communications* **9**(1), 2839. https://doi.org/10.1038/s41467-018-05293-3.

Chakraborty, S. & Dohmen, R. (2022). Diffusion chronometry of volcanic rocks: looking backward and forward. *Bulletin of Volcanology* **84**(6), 57. https://doi.org/10.1007/s00445-022-01565-5.

Cooper, G. F., Blundy, J. D., Macpherson, C. G., Humphreys, M. C. & Davidson, J. P. (2019). Evidence from plutonic xenoliths for magma differentiation, mixing and storage in a volatile-rich crystal mush beneath St. Eustatius, Lesser Antilles. *Contributions to Mineralogy and Petrology* **174**, 1–24.

Costa, F., Shea, T. & Ubide, T. (2020). Diffusion chronometry and the timescales of magmatic processes. *Nature Reviews Earth & Environment* **1**(4), 201–214. https://doi.org/10.1038/s43017-020-0038-x.

Davidson, J., Turner, S., Handley, H., Macpherson, C. & Dosseto, A. (2007). Amphibole "sponge" in arc crust? *Geology* **35**(9), 787–790. https://doi.org/10.1130/G23637A.1.

Dayton, K., Gazel, E., Wieser, P., Troll, V. R., Carracedo, J. C., La Madrid, H. & Perez-Torrado, F. J. (2023). Deep magma storage during the 2021 La Palma eruption. *Science Advances* **9**(6), eade7641. https://doi.org/10.1126/sciadv.ade7641.

Deering, C. D. & Bachmann, O. (2010). Trace element indicators of crystal accumulation in silicic igneous rocks. *Earth and Planetary Science Letters* **297**(1–2), 324–331. https://doi.org/10.1016/j.epsl.2010.06.034.

Ducea, M. N., Chapman, A. D., Bowman, E. & Balica, C. (2021). Arclogites and their role in continental evolution; part 2: relationship to batholiths and volcanoes, density and foundering, remelting and long-term storage in the mantle. *Earth-Science Reviews* **214**, 103476. https://doi.org/10.1016/j.earscirev.2020.103476.

Edmonds, M., Cashman, K. V., Holness, M. & Jackson, M. (2019). Architecture and dynamics of magma reservoirs. *Philosophical Transactions of the Royal Society A* **377**(2139), 20180298. https://doi.org/10.1098/rsta.2018.0298.

Ganne, J., Bachmann, O. & Feng, X. (2018). Deep into magma plumbing systems: interrogating the crystal cargo of volcanic deposits. *Geology* **46**(5), 415–418. https://doi.org/10.1130/G39857.1.

Geurts, P., Ernst, D. & Wehenkel, L. (2006). Extremely randomized trees. *Machine learning* **63**, 3–42. https://doi.org/10.1007/s10994-006-6226-1.

Ghiorso, M. S. & Evans, B. W. (2008). Thermodynamics of rhombohedral oxide solid solutions and a revision of the Fe-Ti two-oxide geothermometer and oxygen-barometer. *American Journal of Science* **308**(9), 957–1039. https://doi.org/10.2475/09.2008.01.

Giordano, G. & Caricchi, L. (2022). Determining the state of activity of transcrustal magmatic systems and their volcanoes. *Annual Review of Earth and Planetary Sciences* **50**, 231–259. https://doi.org/10.1146/annurev-earth-032320-084733.

Glazner, A. F. (2021). Thermal constraints on the longevity, depth, and vertical extent of magmatic systems. *Geochemistry, Geophysics, Geosystems* **22**(4), e2020GC009459. https://doi.org/10.1029/2020GC009459.

Gordeychik, B., Churikova, T., Kronz, A., Sundermeyer, C., Simakin, A. & Wörner, G. (2018). Growth of, and diffusion in, olivine in ultra-fast ascending basalt magmas from Shiveluch volcano. *Scientific Reports* **8**(1), 11775. https://doi.org/10.1038/s41598-018-30133-1.

Greenfield, T., White, R. S. & Roecker, S. (2016). The magmatic plumbing system of the Askja central volcano, Iceland, as imaged by seismic tomography. *Journal of Geophysical Research: Solid Earth* **121**(10), 7211–7229. https://doi.org/10.1002/2016JB013163.

Grocke, S. B., de Silva, S. L., Iriarte, R., Lindsay, J. M. & Cottrell, E. (2017). Catastrophic caldera-forming (CCF) monotonous silicic magma reservoirs: geochemical and petrological constraints on heterogeneity, magma dynamics, and eruption dynamics of the 3· 49 Ma Tara Supereruption, Guacha II caldera, SW Bolivia. *Journal of Petrology* **58**(2), 227–260. https://doi.org/10.1093/petrology/egx012.

Gualda, G. A. & Ghiorso, M. S. (2014). Phase-equilibrium geobarometers for silicic rocks based on rhyolite-MELTS. Part 1: principles, procedures, and evaluation of the method. *Contributions to Mineralogy and Petrology* **168**, 1–17.

Gualda, G. A., Gravley, D. M., Deering, C. D. & Ghiorso, M. S. (2019). Magma extraction pressures and the architecture of volcanic

plumbing systems. *Earth and Planetary Science Letters* **522**, 118–124. https://doi.org/10.1016/j.epsl.2019.06.020.

Halldórsson, S. A., Marshall, E. W., Caracciolo, A., Matthews, S., Bali, E., Rasmussen, M. B. & Stefánsson, A. (2022). Rapid shifting of a deep magmatic source at Fagradalsfjall volcano, Iceland. *Nature* **609**(7927), 529–534. https://doi.org/10.1038/s41586-022-04981-x.

Harrison, W. J. & Wood, B. J. (1980). An experimental investigation of the partitioning of REE between garnet and liquid with reference to the role of defect equilibria. *Contributions to Mineralogy and Petrology* **72**, 145–155. https://doi.org/10.1007/BF00399474.

Hartley, M. E. & Thordarson, T. (2013). The 1874–1876 volcano-tectonic episode at Askja, North Iceland: lateral flow revisited. *Geochemistry, Geophysics, Geosystems* **14**(7), 2286–2309. https://doi.org/10.1002/ggge.20151.

Hartley, M. E., Thordarson, T., Fitton, J. G. & EIMF (2013). Oxygen isotopes in melt inclusions and glasses from the Askja volcanic system, North Iceland. *Geochimica et Cosmochimica Acta* **123**, 55–73. https://doi.org/10.1016/j.gca.2013.09.008.

Hartley, M. E., Bali, E., Maclennan, J., Neave, D. A. & Halldórsson, S. A. (2018). Melt inclusion constraints on petrogenesis of the 2014–2015 Holuhraun eruption, Iceland. *Contributions to Mineralogy and Petrology* **173**, 10–23. https://doi.org/10.1007/s00410-017-1435-0.

Heliker, C. (1995). Inclusions in Mount St. Helens dacite erupted from 1980 through 1983. *Journal of Volcanology and Geothermal Research* **66**(1–4), 115–135. https://doi.org/10.1016/0377-0273(94)00074-Q.

Helz, R. T. & Thornber, C. R. (1987). Geothermometry of Kilauea Iki lava lake. *Bulletin Volcanologique* **49**, 651–668. https://doi.org/10.1007/BF01080357.

Higgins, O., Sheldrake, T. & Caricchi, L. (2022). Machine learning thermobarometry and chemometry using amphibole and clinopyroxene: a window into the roots of an arc volcano (mount Liamuiga, saint Kitts). *Contributions to Mineralogy and Petrology* **177**(1), 10. https://doi.org/10.1007/s00410-021-01874-6.

Hill, G. J., Bibby, H. M., Ogawa, Y., Wallin, E. L., Bennie, S. L., Caldwell, T. G. & Heise, W. (2015). Structure of the Tongariro volcanic system: insights from magnetotelluric imaging. *Earth and Planetary Science Letters* **432**, 115–125. https://doi.org/10.1016/j.epsl.2015.10.003.

Hirschmann, M. M., Ghiorso, M. S., Davis, F. A., Gordon, S. M., Mukherjee, S., Grove, T. L. & 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**(3). https://doi.org/10.1029/2007GC001894.

Holland, T. & Blundy, J. (1994). Non-ideal interactions in calcic amphiboles and their bearing on amphibole-plagioclase thermometry. *Contributions to Mineralogy and Petrology* **116**, 433–447. https://doi.org/10.1007/BF00310910.

Huber, C., Townsend, M., Degruyter, W. & Bachmann, O. (2019). Optimal depth of subvolcanic magma chamber growth controlled by volatiles and crust rheology. *Nature Geoscience* **12**(9), 762–768. https://doi.org/10.1038/s41561-019-0415-6.

Hyndman, R. J. & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting* **22**(4), 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001.

Jackson, M. D., Blundy, J. & Sparks, R. S. J. (2018). Chemical differentiation, cold storage and remobilization of magma in the Earth's crust. *Nature* **564**(7736), 405–409. https://doi.org/10.1038/s41586-018-0746-2.

Jagoutz, O. (2014). Arc crustal differentiation mechanisms. *Earth and Planetary Science Letters* **396**, 267–277. https://doi.org/10.1016/j.epsl.2014.03.060.

Jia, Y. & Ma, J. (2017). What can machine learning do for seismic data processing? *Geophysics* **82**(3), V163–V177. https://doi.org/10.1190/geo2016-0300.1.

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. *Journal of geophysical research: solid. Earth* **127**(4), e2021JB022904.

Kelley, D. F. & Barton, M. (2008). Pressures of crystallization of Icelandic magmas. *Journal of Petrology* **49**(3), 465–492. https://doi.org/10.1093/petrology/egm089.

Kern, J. M., de Silva, S. L., Schmitt, A. K., Kaiser, J. F., Iriarte, A. R. & Economos, R. (2016). Geochronological imaging of an episodically constructed subvolcanic batholith: U-Pb in zircon chronochemistry of the Altiplano-Puna volcanic complex of the Central Andes. *Geosphere* **12**(4), 1054–1077. https://doi.org/10.1130/GES01258.1.

Kiser, E., Palomeras, I., Levander, A., Zelt, C., Harder, S., Schmandt, B. *et al.* (2016). Magma reservoirs from the upper crust to the Moho inferred from high-resolution Vp and vs models beneath Mount St. Helens, Washington state. *USA. Geology* **44**(6), 411–414.

Kuritani, T., Yokoyama, T., Kitagawa, H., Kobayashi, K. & Nakamura, E. (2011). Geochemical evolution of historical lavas from Askja volcano, Iceland: implications for mechanisms and timescales of magmatic differentiation. *Geochimica et Cosmochimica Acta* **75**(2), 570–587. https://doi.org/10.1016/j.gca.2010.10.009.

Le Maitre, R. W., Streckeisen, A., Zanettin, B., Le Bas, M. J., Bonin, B., Bateman, P. & Woolley, A. R. (eds) (2002) Igneous rocks. In: *A Classification and Glossary of Terms: Recommendations of the International Union of Geological Sciences Subcommission on the Systematics of Igneous Rocks*, Vol. **2**. Cambridge: Cambridge University Press.

Lee, C. T. A. & Bachmann, O. (2014). How important is the role of crystal fractionation in making intermediate magmas? Insights from Zr and P systematics. *Earth and Planetary Science Letters* **393**, 266–274. https://doi.org/10.1016/j.epsl.2014.02.044.

Lees, J. M. (1992). The magma system of Mount St. Helens: non-linear high-resolution P-wave tomography. *Journal of Volcanology and Geothermal Research* **53**(1–4), 103–116. https://doi.org/10.1016/0377-0273(92)90077-Q.

Li, X. & Zhang, C. (2022). Machine learning thermobarometry for biotite-bearing magmas. *Journal of geophysical research: solid. Earth* **127**(9), e2022JB024137.

Lindsay, J. M., Schmitt, A. K., Trumbull, R. B., De Silva, S. L., Siebel, W. & Emmermann, R. (2001). Magmatic evolution of the La Pacana caldera system, Central Andes, Chile: compositional variation of two cogenetic, large-volume felsic ignimbrites. *Journal of Petrology* **42**(3), 459–486. https://doi.org/10.1093/petrology/42.3.459.

Lindsay, J. M., De Silva, S., Trumbull, R., Emmermann, R. & Wemmer, K. (2001). La Pacana caldera, N. Chile: a re-evaluation of the stratigraphy and volcanology of one of the world's largest resurgent calderas. *Journal of Volcanology and Geothermal Research* **106**(1–2), 145–173. https://doi.org/10.1016/S0377-0273(00)00270-5.

Lissenberg, C. J., MacLeod, C. J. & Bennett, E. N. (2019). Consequences of a crystal mush-dominated magma plumbing system: a mid-ocean ridge perspective. *Philosophical Transactions of the Royal Society A* **377**(2139), 20180014. https://doi.org/10.1098/rsta.2018.0014.

Maclennan, J. (2019). Mafic tiers and transient mushes: evidence from Iceland. *Philosophical Transactions of the Royal Society A* **377**(2139), 20180021. https://doi.org/10.1098/rsta.2018.0021.

Magee, C., Stevenson, C. T., Ebmeier, S. K., Keir, D., Hammond, J. O., Gottsmann, J. H. & Jackson, M. D. (2018). Magma plumbing systems: a geophysical perspective. *Journal of Petrology* **59**(6), 1217–1251. https://doi.org/10.1093/petrology/egy064.

Masotta, M., Mollo, S., Gaeta, M. & Freda, C. (2016). Melt extraction in mush zones: the case of crystal-rich enclaves at the Sabatini

Volcanic District (Central Italy). *Lithos* **248-251**, 288–292. https://doi.org/10.1016/j.lithos.2016.01.030.

Melson, W. G. (1983). Monitoring the 1980-1982 eruptions of Mount St. Helens: compositions and abundances of glass. *Science* **221**(4618), 1387–1391. https://doi.org/10.1126/science.221.4618.1387.

Merzbacher, C. & Eggler, D. H. (1984). A magmatic geohygrometer: application to mount St. Helens and other dacitic magmas. *Geology* **12**(10), 587–590.

Mohammed, M., Khan, M. B., & Bashier, E. B. M. (2016). *Machine learning: algorithms and applications*. Boca Raton: Crc Press. https://doi.org/10.1201/9781315371658.

Mollo, S., Del Gaudio, P., Ventura, G., Iezzi, G. & Scarlato, P. (2010). Dependence of clinopyroxene composition on cooling rate in basaltic magmas: implications for thermobarometry. *Lithos* **118**(3–4), 302–312. https://doi.org/10.1016/j.lithos.2010.05.006.

Montierth, C., Johnston, A. D. & Cashman, K. V. (1995). An empirical glass-composition-based geothermometer for Mauna Loa lavas. *Mauna Loa Revealed: Structure, Composition, History, and Hazards* **92**, 207–217. https://doi.org/10.1029/GM092p0207.

Mutch, E. J. F., Blundy, J. D., Tattitch, B. C., Cooper, F. J. & Brooker, R. A. (2016). An experimental study of amphibole stability in low-pressure granitic magmas and a revised Al-in-hornblende geobarometer. *Contributions to Mineralogy and Petrology* **171**, 1–27. https://doi.org/10.1007/s00410-016-1298-9.

Nazzareni, S., Rossi, S., Petrelli, M. & Caricchi, L. (2020). Architecture of the magmatic system in the Main Ethiopian rift. *Dynamic Magma Evolution*, 133–151. https://doi.org/10.1002/9781119521143.ch6.

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**(4), 777–794. https://doi.org/10.2138/am-2017-5968.

Nichols, M. L., Malone, S. D., Moran, S. C., Thelen, W. A. & Vidale, J. E. (2011). Deep long-period earthquakes beneath Washington and Oregon volcanoes. *Journal of Volcanology and Geothermal Research* **200**(3–4), 116–128. https://doi.org/10.1016/j.jvolgeores.2010.12.005.

Petrelli, M., Caricchi, L. & Perugini, D. (2020). Machine learning thermo-barometry: application to clinopyroxene-bearing magmas. *Journal of geophysical research: solid. Earth* **125**(9), e2020JB020130.

Petrone, C. M., Bugatti, G., Braschi, E. & Tommasini, S. (2016). Pre-eruptive magmatic processes re-timed using a non-isothermal approach to magma chamber dynamics. *Nature Communications* **7**(1), 12946. https://doi.org/10.1038/ncomms12946.

Plank, T. & Forsyth, D. W. (2016). Thermal structure and melting conditions in the mantle beneath the basin and range province from seismology and petrology. *Geochemistry, Geophysics, Geosystems* **17**(4), 1312–1338. https://doi.org/10.1002/2015GC006205.

Popa, R. G., Bachmann, O. & Huber, C. (2021). Explosive or effusive style of volcanic eruption determined by magma storage conditions. *Nature Geoscience* **14**(10), 781–786. https://doi.org/10.1038/s41561-021-00827-9.

Pritchard, M. E., De Silva, S. L., Michelfelder, G., Zandt, G., McNutt, S. R., Gottsmann, J. & Ward, K. M. (2018). Synthesis: PLUTONS: investigating the relationship between pluton growth and volcanism in the Central Andes. *Geosphere* **14**(3), 954–982. https://doi.org/10.1130/GES01578.1.

Putirka, K. D. (2008). Thermometers and barometers for volcanic systems. *Reviews in Mineralogy and Geochemistry* **69**(1), 61–120. https://doi.org/10.2138/rmg.2008.69.3.

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**(4), 841–858. https://doi.org/10.2138/am-2016-5506.

Putirka, K. D. (2017). Down the crater: where magmas are stored and why they erupt. *Elements* **13**(1), 11–16. https://doi.org/10.2113/gselements.13.1.11.

Riel, N., Kaus, B. J., Green, E. C. R. & Berlie, N. (2022). MAGEMin, an efficient Gibbs energy minimizer: application to igneous systems. *Geochemistry, Geophysics, Geosystems* **23**(7), e2022GC010427. https://doi.org/10.1029/2022GC010427.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G. & Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **40**(8), 913–929. https://doi.org/10.1111/ecog.02881.

Ruprecht, P. & Bachmann, O. (2010). Pre-eruptive reheating during magma mixing at Quizapu volcano and the implications for the explosiveness of silicic arc volcanoes. *Geology* **38**(10), 919–922. https://doi.org/10.1130/G31110.1.

Ruprecht, P. & Wörner, G. (2007). Variable regimes in magma systems documented in plagioclase zoning patterns: El Misti stratovolcano and Andahua monogenetic cones. *Journal of Volcanology and Geothermal Research* **165**(3–4), 142–162. https://doi.org/10.1016/j.jvolgeores.2007.06.002.

Rutherford, M. J. & Devine, J. D. (1988). The may 18, 1980, eruption of Mount St. Helens: 3. Stability and chemistry of amphibole in the magma chamber. *Journal of Geophysical Research: Solid Earth* **93**(B10), 11949–11959. https://doi.org/10.1029/JB093iB10p11949.

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: solid. Earth* **90**(B4), 2929–2947.

Salisbury, M. J., Jicha, B. R., de Silva, S. L., Singer, B. S., Jiménez, N. C. & Ort, M. H. (2011). 40Ar/39Ar chronostratigraphy of Altiplano-Puna volcanic complex ignimbrites reveals the development of a major magmatic province. *Bulletin* **123**(5–6), 821–840. https://doi.org/10.1130/B30280.1.

Schattel, N., Portnyagin, M., Golowin, R., Hoernle, K. & Bindeman, I. (2014). Contrasting conditions of rift and off-rift silicic magma origin on Iceland. *Geophysical Research Letters* **41**(16), 5813–5820. https://doi.org/10.1002/2014GL060780.

Schmitt, A., De Silva, S., Trumbull, R. & Emmermann, R. (2001). Magma evolution in the Purico ignimbrite complex, northern Chile: evidence for zoning of a dacitic magma by injection of rhyolitic melts following mafic recharge. *Contributions to Mineralogy and Petrology* **140**, 680–700. https://doi.org/10.1007/s004100000214.

Seropian, G., Rust, A. C. & Sparks, R. S. J. (2018). The gravitational stability of lenses in magma mushes: confined Rayleigh-Taylor instabilities. *Journal of Geophysical Research: Solid Earth* **123**(5), 3593–3607. https://doi.org/10.1029/2018JB015523.

Smith, D. R. & Leeman, W. P. (1987). Petrogenesis of Mount St. Helens dacitic magmas. *Journal of geophysical research: solid. Earth* **92**(B10), 10313–10334. https://doi.org/10.1029/JB092iB10p10313.

Sparks, R. S. J. & Marshall, L. A. (1986). Thermal and mechanical constraints on mixing between mafic and silicic magmas. *Journal of Volcanology and Geothermal Research* **29**(1–4), 99–124. https://doi.org/10.1016/0377-0273(86)90041-7.

Sparks, R. S. J., Annen, C., Blundy, J. D., Cashman, K. V., Rust, A. C. & Jackson, M. D. (2019). Formation and dynamics of magma reservoirs. *Philosophical Transactions of the Royal Society A* **377**(2139), 20180019. https://doi.org/10.1098/rsta.2018.0019.

Thomson, A. R., Kohn, S. C., Prabhu, A. & Walter, M. J. (2021). Evaluating the formation pressure of diamond-hosted majoritic garnets: a machine learning majorite barometer. *Journal of geophysical research: solid. Earth* **126**(3), e2020JB020604.

Thornber, C. R., Heliker, C., Sherrod, D. R., Kauahikaua, J. P., Miklius, A., Okubo, P. G. & Meeker, G. P. (2003). Kilauea east rift zone magmatism: an episode 54 perspective. *Journal of Petrology* **44**(9), 1525–1559. https://doi.org/10.1093/petrology/egg048.

Tibaldi, A. M., Otamendi, J. E., Cristofolini, E. A., Baliani, I., Walker, B. A., Jr. & Bergantz, G. W. (2013). Reconstruction of the early Ordovician Famatinian arc through thermobarometry in lower and middle crustal exposures, sierra de Valle Fértil, Argentina. *Tectonophysics* **589**, 151–166. https://doi.org/10.1016/j.tecto.2012.12.032.

Till, C. B. (2017). A review and update of mantle thermobarometry for primitive arc magmas. *American Mineralogist* **102**(5), 931–947.

Ubide, T. & Kamber, B. S. (2018). Volcanic crystals as time capsules of eruption history. *Nature Communications* **9**(1), 326. https://doi.org/10.1038/s41467-017-02274-w.

Ulberg, C. W., Creager, K. C., Moran, S. C., Abers, G. A., Thelen, W. A., Levander, A. & Crosson, R. S. (2020). Local source Vp and vs tomography in the Mount St. Helens region with the iMUSH broadband array. *Geochemistry, Geophysics, Geosystems* **21**(3), e2019GC008888. https://doi.org/10.1029/2019GC008888.

Van Zalinge, M. E., Mark, D. F., Sparks, R. S. J., Tremblay, M. M., Keller, C. B., Cooper, F. J. & Rust, A. (2022). Timescales for pluton growth, magma-chamber formation and super-eruptions. *Nature* **608**(7921), 87–92. https://doi.org/10.1038/s41586-022-04921-9.

Wang, X., Hou, T., Wang, M., Zhang, C., Zhang, Z., Pan, R. & Zhang, H. (2021). A new clinopyroxene thermobarometer for mafic to intermediate magmatic systems. *European Journal of Mineralogy* **33**(5), 621–637. https://doi.org/10.5194/ejm-33-621-2021.

Wanke, M., Clynne, M. A., von Quadt, A., Vennemann, T. W. & Bachmann, O. (2019). Geochemical and petrological diversity of mafic magmas from mount St. *Beiträge zur Mineralogie und Petrographie* **174**, 1–25. https://doi.org/10.1007/s00410-018-1544-4.

Ward, K. M., Delph, J. R., Zandt, G., Beck, S. L. & Ducea, M. N. (2017). Magmatic evolution of a cordilleran flare-up and its role in the creation of silicic crust. *Scientific Reports* **7**(1), 9047. https://doi.org/10.1038/s41598-017-09015-5.

Waters, L. E. & Lange, R. A. (2015). An updated calibration of the plagioclase-liquid hygrometer-thermometer applicable to basalts through rhyolites. *American Mineralogist* **100**(10), 2172–2184. https://doi.org/10.2138/am-2015-5232.

Weber, G. & Castro, J. M. (2017). Phase petrology reveals shallow magma storage prior to large explosive silicic eruptions at Hekla volcano, Iceland. *Earth and Planetary Science Letters* **466**, 168–180. https://doi.org/10.1016/j.epsl.2017.03.015.

Weber, G., Arce, J. L., Ulianov, A. & Caricchi, L. (2019). A recurrent magmatic pattern on observable timescales prior to plinian eruptions from Nevado de Toluca (Mexico). *Journal of Geophysical Research: Solid Earth* **124**(11), 10999–11021. https://doi.org/10.1029/2019JB017640.

Weber, G., Caricchi, L. & Arce, J. L. (2020). The long-term life-cycle of Nevado de Toluca Volcano (Mexico): insights into the origin of petrologic modes. *Frontiers in Earth Science* **8**, 563303. https://doi.org/10.3389/feart.2020.563303.

Weber, G., Blundy, J. & Bevan, D. (2023). Mush amalgamation, short residence, and sparse detectability of eruptible magma before Andean super-eruptions. *Geochemistry, Geophysics, Geosystems* **24**. https://doi.org/10.1029/2022GC010732.

Wieser, P. E., Kent, A. J., Till, C. B., Donovan, J., Neave, D. A., Blatter, D. L. & Krawczynski, M. J. (2023a). Barometers behaving badly I: assessing the influence of analytical and experimental uncertainty on Clinopyroxene Thermobarometry calculations at crustal conditions. *Journal of Petrology* **64**(2), egac126. https://doi.org/10.1093/petrology/egac126.

Wieser, P. E., Kent, A. J., Till, C. B. & Abers, G. A. (2023b). Geophysical and geochemical constraints on magma storage depths along the Cascade arc: knowns and unknowns. *Geochemistry, Geophysics, Geosystems* **24**(11), e2023GC011025. https://doi.org/10.1029/2023GC011025.

Wright, M. & Ziegler, A. (2017). Ranger: a fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software* **77**(1), 1–17. https://doi.org/10.18637/jss.v077.i01.

Yang, H. J., Kinzler, R. J. & Grove, T. L. (1996). Experiments and models of anhydrous, basaltic olivine-plagioclase-augite saturated melts from 0.001 to 10 kbar. *Contributions to Mineralogy and Petrology* **124**(1), 1–18. https://doi.org/10.1007/s004100050169.

de Zeeuw-van Dalfsen, E., Pedersen, R., Hooper, A. & Sigmundsson, F. (2012). Subsidence of Askja caldera 2000–2009: modelling of deformation processes at an extensional plate boundary, constrained by time series InSAR analysis. *Journal of Volcanology and Geothermal Research* **213-214**, 72–82. https://doi.org/10.1016/j.jvolgeores.2011.11.004.

Zhang, G. & Lu, Y. (2012). Bias-corrected random forests in regression. *Journal of Applied Statistics* **39**(1), 151–160. https://doi.org/10.1080/02664763.2011.578621.

Zhou, Z. H. (2021). *Machine learning*. Singapore: Springer Nature. https://doi.org/10.1007/978-981-15-1967-3.

Ziberna, L., Green, E. C. & Blundy, J. D. (2017). Multiple-reaction geobarometry for olivine-bearing igneous rocks. *American Mineralogist* **102**(12), 2349–2366. https://doi.org/10.2138/am-2017-6154.