Journal of Metamorphic Geology

WILEY

Journal of METAMORPHIC GEOLOGY

ORIGINAL ARTICLE OPEN ACCESS

Calibration, Validation and Evaluation of Machine Learning Thermobarometers in Metamorphic Petrology: An Application to Biotite and Outlook for Future Strategy

Philip Hartmeier[1] | Jacob B. Forshaw[2] | Pierre Lanari[1,3]

[1]Institute of Earth Sciences, University of Lausanne, Lausanne, Switzerland | [2]Mineral Deposit Research Unit, University of British Columbia, Vancouver, Canada | [3]Institute of Geological Sciences, University of Bern, Bern, Switzerland

Correspondence: Philip Hartmeier (philip.hartmeier@unil.ch)

Received: 24 March 2025 | **Revised:** 26 June 2025 | **Accepted:** 30 June 2025

Funding: This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant 850530).

Keywords: biotite | machine learning | metapelite | mineral assemblage sequences | thermobarometry

ABSTRACT

Geothermobarometry provides crucial constraints on the physical conditions of metamorphism, offering insights into petrogenetic processes and providing key information on thermal regimes and metamorphic depths to other geological disciplines. However, calibrating a thermobarometer from the natural record is challenging because independent pressure (P) and temperature (T) estimates are required, and the compositional variation of minerals—governed by multiple metamorphic reactions—must be captured in a complex function. This work calibrates a machine learning thermobarometer for biotite using relative P–T estimates based on mineral assemblage sequences. A neural network is used as a flexible model to fit a high-dimensional thermobarometric regression curve. To address the challenge of sparse training data, a transfer learning strategy is employed, where the model is primarily trained on a large dataset generated with phase equilibrium modelling before refinement with natural data. A general framework for calibrating machine learning thermobarometers is outlined using a neural network thermobarometer for biotite as an example. Selection of the best-performing model is guided by k-fold cross-validation alongside complementary accuracy checks using metamorphic sequences and precision assessments via Monte Carlo error propagation. Evaluation on an independent test dataset, compiled from the literature, indicates that the model is a potential biotite single-crystal thermometer with a root mean square error of ±45°C, consistent with the estimated uncertainty of Ti-in-Bt thermometry applied to the same data. A potential barometer is affected by systematic underestimation of pressures above 0.6 GPa due to regression to the mean of the natural database, which is biased towards low-pressure metamorphism. This limits its applicability in higher-pressure regimes. This study highlights the potential of using neural networks with transfer learning in petrological applications since they are often constrained by limited natural data.

1 | Introduction

Geothermobarometry, the determination of the pressure (P) and temperature (T) conditions of metamorphism, is a fundamental petrological problem. These conditions provide a reference for the depth and temperature regime that operated during the metamorphism of a rock unit and therefore provide important constraints on the geodynamic setting (Ernst 1971; Thompson and England 1984), mountain-building processes (Kohn 2008), the modelling of geo-mechanical processes (Hertgen et al. 2017) and cycling of elements (Hermann and Lakey 2021; Yardley and Cleverley 2013). The P–T conditions

of metamorphism in the Earth's crust and mantle can never be probed directly. Instead, they must be determined indirectly using geothermobarometry, which involves inverting from the mineral assemblage and mineral compositions to recover *P–T* conditions from frozen equilibria (Fyfe et al. 1958). Various approaches to quantitative geothermobarometry have been developed (see detailed reviews by Lanari and Duesterhoeft 2019; Powell and Holland 2008; Spear et al. 2016). For this work, the techniques relying on (semi-)empirical calibration of thermobarometers are relevant and considered below. Experimental calibrations of exchange (e.g., Ellis and Green 1979; Ferry and Spear 1978) and net-transfer reactions (e.g., Koziol and Newton 1988; Liu and Bohlen 1995) can provide thermobarometric relations in simplified systems. For minor or trace elements, their *P–T*-dependent saturation in minerals can be calibrated as geothermobarometer if their incorporation is buffered by accessory phases stable over a broad *P–T* range, for example, the incorporation of Ti into various phases (Ferry and Watson 2007; Henry et al. 2005; Osborne et al. 2022). High-*T* batch equilibration experiments can directly provide high-dimensional phase compositions for multiple *P–T* points and have successfully been used in magmatic petrology to calibrate geothermobarometers. The experiments presumably capture the complexity of natural geochemical equilibration through multiple reactions (Jorgenson et al. 2022; Li and Zhang 2022; Weber and Blundy 2024).

A regression to calibrate a geothermobarometer can either be done by fitting a purely empirical thermo-, baro- or thermobarometric function (e.g., Cathelineau and Nieva 1985; Chicchi et al. 2023; Henry et al. 2005; Jorgenson et al. 2022; Weber and Blundy 2024) or a physically based expression, often a formulation of the Gibbs free energy difference of a reaction (ΔG), where either *P*, *T* or both are variables (e.g., Ellis and Green 1979; Ferry and Spear 1978; Lanari et al. 2014). Studies using linear least-square-fitting for the calibration (Ferry and Spear 1978; Ferry and Watson 2007) rely on reformulating a suitable thermobarometer as a linear expression, hence limiting the complexity of the model. Some studies have used higher-order functions, that can be fitted by using numerical optimisation, to expand the complexity of calibrated expressions (Henry et al. 2005; Osborne et al. 2022). In recent years, petrology has seen an increase in the successful application of modern machine learning algorithms, which no longer require an a priori choice of expression, and therefore provide an additional increase in flexibility to directly fit potentially complex and undiscovered thermobarometric relations in the data (Jorgenson et al. 2022; Li and Zhang 2022; Weber and Blundy 2024). The reader is referred to Petrelli (2024) for a compilation of recent advances in the application of machine learning algorithms to thermobarometric regression problems.

To fit a thermobarometer, an estimate of *P–T*, independent of the composition (*X*), is required for each data point. For experimental data, the laboratory-measured *P–T* conditions can be used directly (e.g., Chicchi et al. 2023; Ferry and Spear 1978; Jorgenson et al. 2022; Weber and Blundy 2024) while considering the degree to which equilibrium was achieved (Pattison 1994). For natural samples, there is no precise numeric *P–T* readily available. Two different strategies have been used to overcome this limitation in the calibration of thermobarometers from natural data. One approach is to use an existing thermobarometer to estimate the *P–T* conditions of each analysis, and then use these in the calibration of the new thermobarometer; for example, the Graham and Powell (1984) calibration of the Grt-Hbl thermometer which used the Grt-Cpx thermometer of Ellis and Green (1979) or the Wu and Chen (2015) calibration of the Ti-in-Biotite thermometer which used the Grt-Bt thermometer and Grt-Al_2SiO_5-Qz-Pl barometer of Holdaway (2000, 2001). A different strategy relies on metamorphic field gradients that, if anchored in *P–T,* can provide constraints for points within an observable sequence. Henry et al. (2005), following Henry and Guidotti (2002), tied the *P–T* conditions of metapelitic isograds in the Grt-St-Sil-Kfs sequence in west-central Maine to the petrogenetic grid of Spear and Cheney (1989). This grid is based on the thermodynamic data of Berman and Brown (1985) and Berman (1988). Assuming these isograds correspond to isotherms, they interpolated the temperature in between, obtaining precise constraints for each sample. Mineral abbreviations are after Warr (2021).

A metamorphic mineral well-suited for the calibration of a geothermobarometer is biotite. Biotite is a ubiquitous mineral in metasiliciclastic rocks, which are globally abundant throughout the Earth's history (Ronov 1972), and is stable from greenschist to granulite facies conditions (Spear 1993). In metapelites, the Fe–Mg exchange between biotite and garnet systematically increases $K_{D\ Mg\text{–}Fe}$ with increasing *T* (Hietanen 1969; Lyons and Morse 1970; Saxena 1969; Sen and Chakraborty 1968). In addition, Ti saturation in biotite has also been shown to increase with increasing *T* of metamorphism (Guidotti et al. 1977; Kwak 1968). Building on these observations, Henry and Guidotti (2002), and later Henry et al. (2005), calibrated an empirical Ti-in-Bt thermometer. Experimental studies performed at high *P* conditions (>2.0 GPa) suggest a Tschermak-type Si–Al substitution in biotite as a function of pressure (Hermann 2003). While there is experimental data for high-temperature biotite (>625°C, Li and Zhang 2022), sluggish experimental kinetics at low temperatures prevent the acquisition of such data at greenschist to amphibolite facies conditions.

To address the limitations outlined above, this study investigates the process of training a neural network (NN) for coupled biotite thermobarometry using natural mineral compositions from metapelitic sequences compiled by Pattison and Forshaw (2025). A systematic approach to the calibration, validation and performance evaluation of a machine learning thermobarometer is presented and used for the calibration of a NN biotite thermobarometer. The strategy and specific performance evaluations outlined below offer a reproducible framework to fit any machine-learning-based thermobarometers, applicable to future studies using different ML algorithms and calibrating thermobarometers for minerals other than biotite.

Calibrating any thermobarometer based on natural data is challenging since first suitable analyses in metamorphic petrology are often sparse and second, compared to experiments, an additional independent constraint for *P–T* is required. In this work, *P–T* conditions are estimated based on the relational information of mineral assemblage sequences (see Carmichael 1978; Hietanen 1967; Miyashiro 1961; Pattison and Forshaw 2025; Pattison and Tracy 1991). This approach was successful for a single metamorphic sequence,

as demonstrated by Henry and Guidotti (2002) and Henry et al. (2005) in the calibration of their Ti-in-Bt thermometer. To address the challenge of limited data availability, phase equilibrium modelling was used to generate large datasets of synthetic biotite-bearing metapelites. The use of NNs as a machine learning model allows leveraging of a technique called transfer-learning (see Weiss et al. 2016 for a review), where a functional relation in sparse data is fitted, building upon a prior model obtained from a larger dataset. This study demonstrates the potential of transfer learning using phase equilibrium modelling to fit machine learning models to sparsely available petrological data, both collections of natural data and experimental data points.

2 | Methods

2.1 | Dataset of Natural Pelitic Biotite Compositions

Fitting an empirical thermobarometer on natural biotite compositions from metapelites requires a large database of constraints, consisting of pressure–temperature–biotite composition (P–T–$X_{biotite}$) points. This study uses a database of biotite compositions from metapelitic mineral assemblage sequences (MAS), derived from the global compilation of the natural metapelite record by Pattison and Forshaw (2025). The data cover a significant part of the domain of typical crustal metamorphic conditions spanning approximately 400°C to 850°C, and 0.15 GPa to 0.95 GPa (Figure 1a). The biotite dataset shows variation in the major elements which is thought to capture potential P–T-dependent substitutions (Figure 1b,c). Volatile elements, including H, F and Cl, were excluded either due to analytical limitations and because they are not routinely measured by electron microprobe. Biotite stoichiometry is therefore calculated based on 11-anionic charges with a constant H_2O of 1.0 per formula unit.

A further limitation is that biotite contains both ferrous (Fe^{2+}) and ferric (Fe^{3+}) iron, yet the electron microprobe cannot differentiate between these oxidation states. To assess if considering Fe^{2+} and Fe^{3+} separately would have a large effect on the calibration of a biotite thermobarometer, this was tested during initial feature selection by using a rough estimate of $X_{Fe3+} = Fe^{3+}/(Fe^{2+}+Fe^{3+})$ in biotite, obtained based on the Fe-oxide present in the stable assemblage (Forshaw and Pattison 2021; Guidotti and Dyar 1991). In trioctahedral micas, the Ca content is typically negligible, and the Ca-rich clintonite does not form a solid solution with biotite (Tischendorf et al. 2018). Consequently, CaO is not considered in the compositional data of biotite used in the training dataset. In contrast to white mica, the Na content in metamorphic biotite is typically low, and the interlayer is dominated by K (Dubacq and Forshaw 2024). To check for a potential P–T dependence of the Na–K substitution in the interlayer site, a second set in the MnO–Na_2O–K_2O–FeO–MgO–Al_2O_3–SiO_2–TiO_2 (MnNKFMAST) system was assessed, in addition to the MnO–FeO–MgO–Al_2O_3–SiO_2–TiO_2 (MnFMAST) system that assumes K is fixed at 1.0 per formula unit.

The metamorphic zones within each MAS provide a relative framework for obtaining an independent estimate of P–T (Figure 1a). Zones of characteristic assemblages in metamorphic field gradients are the result of increasing temperature, while the sequence of mineral assemblages varies systematically with changing pressure (see Pattison and Forshaw 2025, and references therein). The supplementary data of Pattison and Forshaw (2025) outlines the estimated P–T ranges of the zones within each MAS. Each analysis of biotite composition was then assigned a unique P–T by random sampling within the given P–T zones. Pressures were sampled from a uniform probability distribution within each zone. This results in an uncertainty of ±0.1 GPa in the data's P estimate. Temperatures were sampled from a uniform probability distribution across a metamorphic zone and then assigned to each sample by ordering them according to the temperature predicted by the Ti-in-Bt

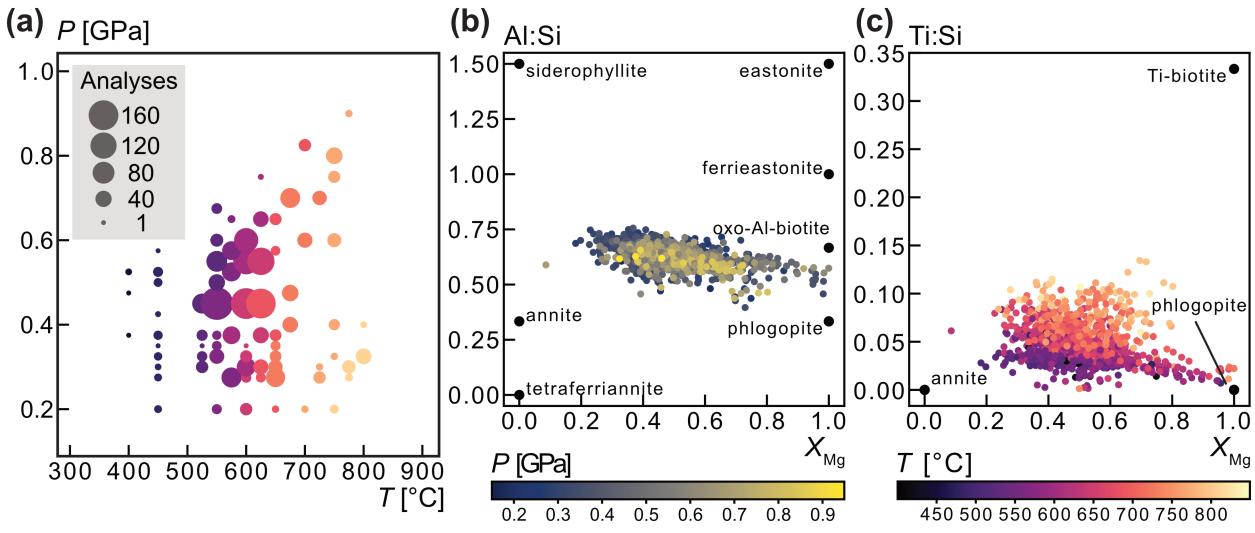

FIGURE 1 | Dataset of P–T–$X_{biotite}$ based on the database of Pattison and Forshaw (2025). (a) Number of biotite analyses in the metamorphic zones of mineral assemblage sequences plotted in P–T space; see Supplementary Material S4 for the inferred P–T of the zones. (b) Compositional variance of biotite in X_{Mg} and Al:Si. The colour-coding after pressure indicates a rough trend of decreasing Al:Si with increasing pressure. (c) Compositional variance of biotite in X_{Mg} and Ti:Si. The colour-coding after temperature shows a clear trend of increasing Ti with increasing temperature. Compositional endmembers of biotite are those proposed by Dubacq and Forshaw (2024).

thermometer of Henry et al. (2005). This procedure hard-codes the well-established correlation between T and Ti in biotite (Guidotti et al. 1977; Henry and Guidotti 2002; Kwak 1968) into the dataset, artificially increasing the data resolution in T, below the ±50°C to 100°C of the metamorphic zones. By not directly taking the prediction of the Ti-in-Bt thermometry but only ordering the randomly generated temperatures for each zone after these predictions, we aim to minimise the potential bias towards the calibration of Henry et al. (2005). See Supplementary Material S1 for a discussion of the effect of different methods to assign a P–T estimate to each analysis on the performance of a NN thermobarometer during training. In Supplementary Material S4, a complete record of the P–T ranges used in this study is provided. Minor deviations from those given in Pattison and Forshaw (2025) are discussed there. A training dataset of 2148 biotite compositions was extracted for the MnFMAST system and 2092 for the MnNKFMAST system. The smaller number in the MnNKFMAST dataset results from incomplete analyses of Na in biotite, which were filtered out. The exact P–T–X_{biotite} datasets, based on the database of Pattison and Forshaw (2025), are provided alongside the code in an open-source repository to ensure reproducibility of the results presented here.

2.2 | Validation and Test Dataset

When calibrating a machine learning model, data from the training dataset can be used in two different ways. The training subset is used during the numerical optimisation of the model parameters, also known as training, to find the best set of parameters. A validation subset is used to evaluate the model performance during training and adapt well-performing architectures, feature sets and training strategies. During the calibration procedure, the model is optimised and, therefore, a priori biased towards both the training and validation data. Towards the training data the model is biased directly by optimising its parameters during training. It is important to emphasise that there is also a potential bias towards the validation data by subsequently selecting input features, hyperparameters and training strategies that perform well on the validation data. To avoid this optimisation leading to overfitting, it is crucial to maintain a separate independent test dataset throughout the calibration process, which is never used in any decision affecting the model performance and which allows the performance of a final model to be assessed.

During the calibration of the models, randomly selected validation data from the training dataset are complemented with 10 additional metamorphic sequences, for which P–T can be estimated based on the mineral assemblage zones (Pattison and Forshaw 2025). While sequential data does not provide precise P–T for each individual sample, it can be used to test if a thermobarometer accurately predicts the increase in metamorphic grade across a sequence. Additionally, a set of biotite analyses from sample 28HF18 of Yogi et al. (2024) are used to investigate the propagation of compositional uncertainty found in nature at the thin section scale.

To test the accuracy of the NN-thermobarometer an independent test set of samples not used during training and validation was compiled. The test dataset comprises 48 biotite-bearing samples with published P–T estimates obtained by the combination of multiple methods, preferentially including a combination of phase equilibrium modelling (Bingo-Antidote, Duesterhoeft and Lanari 2020; LinaForma, Mackay-Champion and Cawood 2024), multi-equilibrium thermobarometry (TWQ, Berman 1991; AvPT, Powell and Holland 1994; ChlMicaEqui, Lanari 2012), empirical thermobarometry (Ti-in-Bt, Henry et al. 2005; Grt-Bt thermometry, Ferry and Spear 1978, Grt-Bt-Ms-Pl barometry, Wu 2015; Grt-Bt-Al_2SiO_5-Qz barometry, Wu 2017) and non-conventional thermobaromery (e.g., Ti-in-quartz thermometry, Ferry and Watson 2007; quartz-in-garnet barometry [QuiG], Kohn 2014; Raman spectra of carbonaceous material thermometry [RSCM], Beyssac et al. 2002).

2.3 | Machine Learning Methodology

To fit the thermobarometer, feed-forward NNs were used (Figure 2a, Table 1). A NN is a machine learning algorithm that is made up of multiple layers, through which input data is consecutively processed (see in Petrelli 2024 for a review). The input layer takes in the compositional vector of biotite ($\vec{C}$), and the last layer consists of a two-dimensional P–T vector as output. In between sit l hidden layers with each a certain number of neurons (k). In each layer, after the input, two mathematical operations take place. First, the previous layer's output vector ($\vec{x}$), of size $n\times1$, is multiplied with a $k\times n$ matrix of weights ($\mathbb{W}$) and a $k\times1$ bias vector ($\vec{\mathbb{b}}$) is added, resulting in a $k\times1$ vector ($\vec{y}$):

$$\vec{y} = \mathbb{W}\vec{x} + \vec{\mathbb{b}} \tag{1}$$

This linear transformation is then followed by the non-linear activation function, a rectified linear unit (ReLU):

$$\mathrm{ReLU}(y_i) = \begin{cases} y_i, & y_i > 0 \\ 0, & y_i \le 0 \end{cases} \tag{2}$$

The full NN is a concatenation of $l+1$ multi-linear transformations, each followed by a ReLU. In the last layer, the output, an identity function is used as the activation instead of the ReLU. During training of the NN, the parameters of all the weight-matrices and bias vectors are adapted using a numerical optimiser to minimise the value of a loss function. This loss is calculated as mean squared error (MSE) between the true (Y_i) and predicted ($\hat{Y}_i$) for values P and T.

$$\mathrm{MSE} = f\left(\mathbb{W}_{i,\dots,j}, \vec{\mathbb{b}}_{u,\dots,w}\right) = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 \tag{3}$$

To evaluate the performance of a trained NN thermobarometer and estimate the uncertainty of its predictions, the root-mean-square error (RMSE) is calculated:

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}} \tag{4}$$

As an optimiser the Adam algorithm (Kingma and Ba 2014) was used with an inverse-time learning rate decay scheme. The compositional input data was normalised, while the P–T data were scaled using minimum–maximum scaling in the range [0.15, 1.0] GPa and [400, 900]°C.

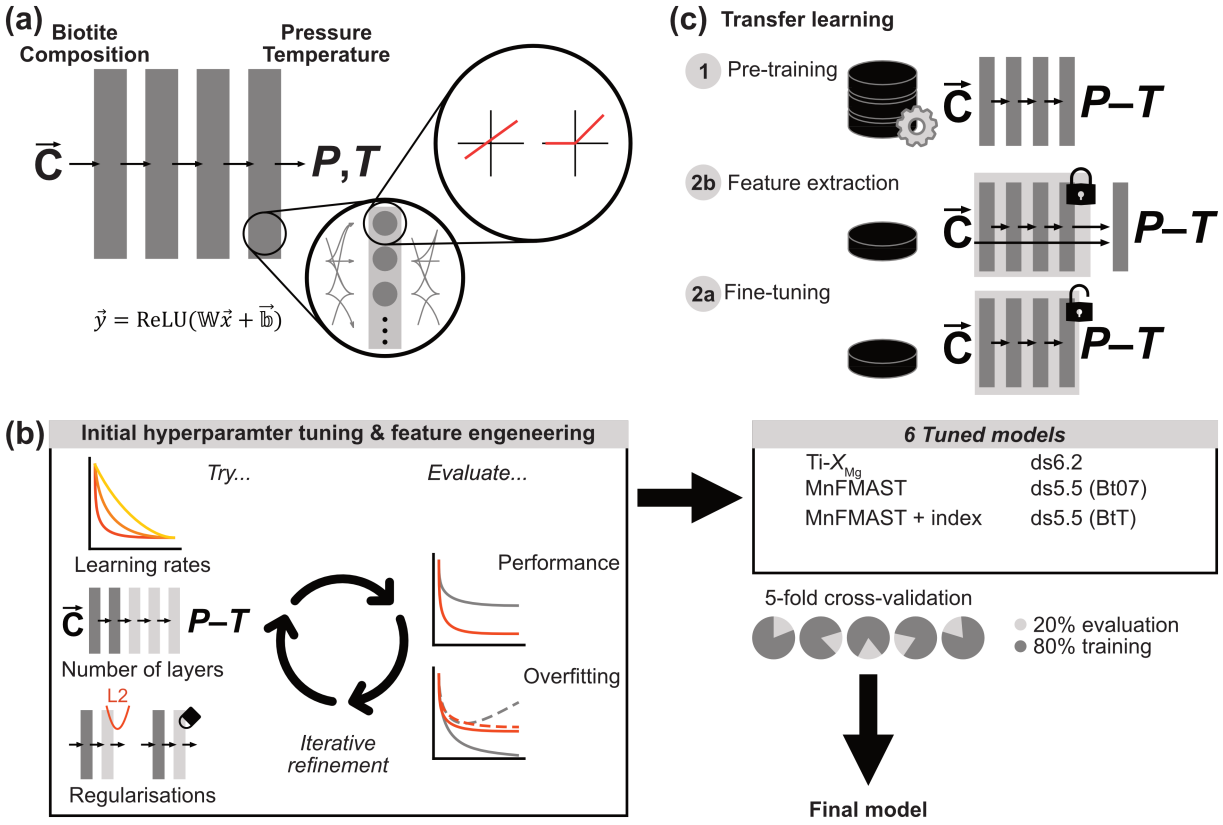

FIGURE 2 | Machine learning methodology of this study. (a) A neural network thermobarometer consisting of an input compositional vector of biotite, four hidden layers and a two-dimensional output with the predicted pressure and temperature. Each layer represents a linear followed by a ReLU transformation. (b) Hyperparameters and feature sets are refined iteratively in experiments by optimising the performance on a validation subset. Afterwards, the best performing models are evaluated using 5-fold cross-validation to select the features, hyperparameter and training strategy for the final model. (c) Transfer learning is a training strategy used on problems with sparse constraints. First, a prior model is fitted on a large dataset of a similar domain, for example, data simulated using phase equilibrium modelling. Afterwards, the smaller dataset containing the relation to fit is used in transfer learning. This is either done by feature extraction, where the prior's prediction is used as an additional input feature helping the regression, or by fine-tuning, where the prior model's parameter is finely updated, assuming they provide an initial parametrisation close to the global optimum.

TABLE 1 | NN models for thermobarometry.

Name	Input features	Input dimension[a]	Model architecture[b]	Trainable parameters	Regularisation	Transfer learning	Number of constraints[c]
M1	Ti-X_{Mg}	(2)	(16)	65	—	No	(1718)
M2a	MnFMAST	(6)	(16)	146	—	No	(1718)
M2b	MnFMAST + index minerals[d]	(14)	(64, 64, 64)	9410	Dropout layers	No	(1718)
M3a	MnFMAST	(6)	(64, 64, 64, 64)	13,058	Dropout layers	Yes (ds62)	(1718/37763)[e]
M3b	MnFMAST	(6)	(64, 64, 64, 64)	13,058	Dropout layers	Yes (ds55Bt07)	(1718/44205)[e]
M3c	MnFMAST	(6)	(64, 64, 64, 64)	13,058	Dropout layers	Yes (ds55BtT)	(1718/39473)[e]

[a]Input dimension is the number of neurons in the input vector, corresponding to the number of features used in a given model.
[b]The model architecture is given in number of neurons in each hidden layer of the neural network.
[c]The number of constraints given here are 80% of the training database of P–T-$X_{biotite}$ triplets, because during calibration 20% of the data is used for validation.
[d]Index minerals considered are: Chl, Grt, Crd, And, St, Ky, Sil and Kfs; mineral abbreviations after Warr (2021).
[e]For the transfer learning models both the constrains of the training and pre-training are given (# training/# pre-training).

2.3.1 | Hyperparameter Tuning and Feature Engineering

In addition to the trainable parameters that are optimised during training, other parameters that are set manually before training can influence a NN's performance. These hyperparameters are optimised in a semi-quantitative way (Figure 2b). Multiple NNs are trained, varying one hyperparameter at a time, and then evaluated. For each hyperparameter, the value is chosen so that the NN best performs on 20% of the data, which has been randomly selected and held out of training as validation data. For the data simulated using phase equilibrium modelling, the fraction is reduced to 10% as this dataset is much larger. Special emphasis is placed on minimising overfitting, identified by discrepancies in performance between validation and training data. The following hyperparameters were systematically tested (Figure 2b): (i) different combinations of the number of layers (l) and neurons (k) per layer in the NN; (ii) various initial learning rates for the numerical optimiser; and (iii) regularisation methods to reduce overfitting. Additional regularisation techniques such as larger weight penalties (L2 regularisation), dropout layers, normalisation layers, or their combinations can help mitigate overfitting while improving overall performance, an alternative to simply constraining model capacity with a smaller NN (Chollet 2021).

Similarly, different features put into the NN can also influence its performance in predicting P–T. As with the hyperparameters different sets of features are evaluated in a semi-quantitative way by comparing the NN performance on a randomly selected validation subset (20%) of the training data. A complete list of all hyperparameters and feature sets tested can be found in the Supplementary Materials S1 and S2.

2.3.2 | Transfer Learning

Another regularisation methodology to train a NN on sparse data and reduce overfitting is transfer learning (Figure 2c). For this, a prior NN is first trained on an analogous regression task with a significantly larger dataset containing similar information as the sparse data on which the final task should be trained (Weiss et al. 2016). In this study, phase equilibrium modelling was used to generate large datasets for training a prior NN (see below). Transfer learning is then performed using one of two strategies (Figure 2c). In the feature extraction strategy, the prior model prediction is used as an additional input combined with the biotite composition in a successive NN. In this case, only the parameters of the added NN layers are trained, while the prior model remains unchanged. In the fine-tuning strategy, the prior model is retrained on the new dataset. The parameterisation of the prior model is used as an initialisation that is assumed to be closer to an optimal solution to the regression problem, and the parameters are then updated with a small learning rate.

Hyperparameters of the prior NNs are optimised analogous to the procedure described above. The semi-quantitative tests were conducted using the data simulated with ds62 (see below) to systematically assess the effect of (i) model capacity, the number of trainable parameters, with different architectures, (ii) the effect of different activation functions and (iii) regularisation by parameter-regularisation or regularisation layers.

2.3.3 | Cross-Validation for Model Selection

Initial hyperparameter tuning and exploration of different sets of input features resulted in six possible NN thermobarometers (see Table 1): a minimal model M1 considering the Ti content and $X_{Mg} = Mg/Mg + Fe_{Tot}$ of biotite, a model M2a trained on the MnFMAST system, a model M2b trained on the MnFMAST dataset with the additional input of a categorical variable denoting which index minerals (Chl, Grt, Crd, And, St, Ky, Sil and Kfs) are part of the inferred stable peak mineral assemblage, and three MnFMAST models, M3a, M3b and M3c, trained using transfer learning. For the selection of the final model architecture and training setup, k-fold cross-validation has been used. The training dataset is split 5-fold into random splits of 80% training data used to calibrate the model parameters and 20% validation data used to evaluate the model performance. This strategy allows a more reproducible performance estimate to be obtained on small training datasets, as the effect of randomly selecting a well-performing validation subset is visible through outliers in the calculated performance metric, which is calculated five times. The randomly selected k validation sets are complemented with the additional independent validation data described above.

2.4 | Simulation of Biotite Data

Phase equilibrium modelling is used to generate a large database of P–T–$X_{biotite}$ data for training a prior NN in transfer learning. This forward modelling method allows the stable mineral assemblage, modes and compositions to be simulated at different P–T conditions for metapelitic bulk rock compositions using Gibbs free energy minimisation. The minimisations are performed in the $MnO–Na_2O–CaO–K_2O–FeO–MgO–Al_2O_3–SiO_2–H_2O–TiO_2–O_2$ (MnNCKFMASHTO) system using Theriak (de Capitani and Brown 1987) via pytheriak (Hartmeier and Lanari 2024). Modelled biotite compositions are then extracted from the minimisation in the MnFMAST system assuming all Fe as Fe^{2+} analogous to the natural data. Bulk rock compositions were randomly sampled from 5729 pelitic whole-rock analyses in the database of Forshaw and Pattison (2023b). The bulk rock compositions were reduced to the MnNCKFMASHTO system by projection from apatite while ensuring a minimum CaO content of 0.01 mol% in each sample. For subsolidus conditions, the system is modelled assuming H_2O saturation. For suprasolidus conditions, the H_2O content is incrementally halved starting from 8 mol (≈ 2 wt%), until the melt fraction is below 7 vol% or a minimum amount of 1 mol H_2O (≈ 0.3 wt%) has been reached. The X_{Fe3+} is set to a fixed value of 0.10, which is lower than the average X_{Fe3+} of 0.23 reported by Forshaw and Pattison (2023b). This choice was made to account for the overestimation of Fe^{3+} in pelites by wet chemical titrations (Fitton and Gill 1970; Forshaw and Pattison 2023a; Hillebrand 1908; Reay 1981) and to be consistent with the expected X_{Fe3+} from mass balance of Fe^{3+}-bearing minerals in metapelites (Forshaw and Pattison 2021).

The P and T are sampled from two uniform distributions in the intervals of [0.15, 1.0] GPa and [400, 900]°C.

Three different thermodynamic databases, each of them consisting of a set of standard state properties and a set of solution models (as defined in Lanari and Duesterhoeft 2019), are used to generate three P–T–$X_{biotite}$ datasets. The first database is the internally consistent thermodynamic dataset ds62 (Holland and Powell 2011) with solution models from White et al. (2014a) and White et al. (2014b) for garnet, biotite, white mica, orthopyroxene, cordierite, staurolite, chlorite, chloritoid and ilmenite, and from Holland et al. (2022) for feldspar. The second database is the internally consistent thermodynamic dataset ds55 (Holland and Powell 1998) extended to include ideal Mn-endmembers for cordierite, staurolite and chlorite, with solution models from White et al. (2007) expanded to include ideal Mn-endmembers for garnet and biotite, from Coggon and Holland (2002) for white mica, from White et al. (2002) for orthopyroxene, from Holland and Powell (2003) for feldspar, from White et al. (2000) for spinel and from White et al. (2005) for ilmenite. The third database is identical to the second database except that it uses the biotite model from Tajčmanová et al. (2009) expanded to include an ideal Mn-endmember, instead of the model from White et al. (2007).

For each thermodynamic database, 100,000 random P–T–$X_{bulk\text{-}rock}$ triplets are used to simulate synthetic metapelitic systems. Since the stability of biotite varies depending on the database and the P–T–$X_{bulk\text{-}rock}$ conditions, this results in 41,959 P–T–$X_{biotite}$ data points for ds62 with the White et al. (2014a) and White et al. (2014b) biotite model (hereafter referred to as ds62), 49,117 for ds55 with the White et al. (2007) biotite model (hereafter referred to as ds55Bt07) and 43,859 for ds55 with the Tajčmanová et al. (2009) biotite model (hereafter referred to as ds55BtT).

3 | Results—Biotite Datasets in Compositional Space

To extract any thermobarometric function from the biotite composition in metapelitic sequences or simulated datasets generated with phase equilibrium modelling, systematic variation of the elements in biotite with P and T must be present. This section explores such biotite compositional variations from a model calibration point of view. The reader is referred to Pattison and Forshaw (2025) for a further discussion of potential petrological implications that such variation might bring.

3.1 | Biotite Compositions from Natural Samples

The MnFMAST training dataset consists of 2′148 biotite analyses. Most biotite compositions in the dataset lie close to the annite–phlogopite solid solution line, with a subordinate contribution from the more Al-rich endmembers siderophyllite and eastonite (Figure 1b). There is a weak negative correlation between the Al:Si ratio in biotite and the assigned P of equilibration (Figure 1b), Pearson's correlation coefficient $r = -0.28$ (95% C.I.: −0.31 to −0.24). This indicates a potential P dependence of a Tschermak-type exchange reaction in biotite,

$$Al^{IV} + Al^{VI} \rightleftarrows Si^{IV} + (Mg, Fe)^{VI} \quad (R1)$$

where with increasing P, the Tschermak's component decreases in favour of the smaller Si-ion on the tetrahedral site. Between the assigned T of equilibration and the Ti:Si ratio, there is a clear positive correlation (Figure 1c, $r = 0.71$, 95% C.I.: 0.69–0.73). These correlations agree with any of the proposed incorporation mechanisms of Ti into biotite: deprotonation, Ti-Tschermak's exchange, or Ti-vacancy exchange (Cesare et al. 2003; Waters and Charnley 2002). As T and X_{Mg} only correlate very weakly or not at all (Figure 1c, $r = 0.11$, 0.07–0.16 [95% C.I.]), temperature-dependent Fe–Mg exchange reactions, including biotite (e.g., Ferry and Spear 1978) do not exert first-order control on biotite composition. We interpret this as a masking of the temperature dependence by varying which stable phase is partitioning Fe–Mg with biotite, as different limiting assemblages can control the biotite's X_{Mg} depending on variations in the X_{Mg} of different bulk rock composition (Thompson 1976).

Examining biotite compositions in atoms per formula unit (apfu) for a 11-anionic charges basis over P–T space, the following observations can be made (Figure 3). Ti increases with T over the whole range of P, with the highest values of > 0.2 apfu Ti in biotite occurring above 750°C (Figure 3f). The Al content in biotite shows at first order a slight decrease with increasing P, and additionally a decrease below 1.5 apfu for the $T > 750$°C (Figure 3d). For the Si content, an increase towards low T–high P (LT–HP) conditions is visible. In the LT–HP region of P–T space Si and Al trends show an anticorrelation indicative of a P-sensitive Tschermak exchange. Both Mn and Fe_{tot} show negative compositional gradients with increasing P–T, whereas Mg shows the opposite trend (Figure 3a–c). A comparison of the compositional gradients suggests that these trends mimic the topology of the garnet-in reaction in metapelites (compare Figure 3a-c with figure 3 in Pattison and Forshaw 2025). This behaviour is well explained by the partitioning of Mg–Fe between garnet and biotite, with Fe showing a stronger affinity for garnet than Mg (Hietanen 1969; Lyons and Morse 1970; Sen and Chakraborty 1968). For Mn, the effect is thought to be even more strongly pronounced, as garnet typically shows bell-shaped zoning in Mn, effectively fractionating Mn from the reactive part of the rock (Hollister 1966; Tracy 1982). The second-order anticorrelation of Fe_{tot} and Mg is likely due to an increase in garnet mode with increasing P–T (Baxter et al. 2017) or a decrease in the Fe–Mg partitioning coefficient with increasing T (Ferry and Spear 1978).

3.2 | Biotite Compositions from Phase Equilibrium Modelling

For the comparison with the natural data, the simulated biotite composition is reported as cations in apfu in a 20 × 20 grid across P–T space (Figure 4). For each grid cell, the biotite composition is an average of biotite generated using a randomly sampled metapelitic whole-rock compositions. This approach considers the effect of the natural geochemical variability of metapelites and its influence on the phase equilibrium modelling of biotite.

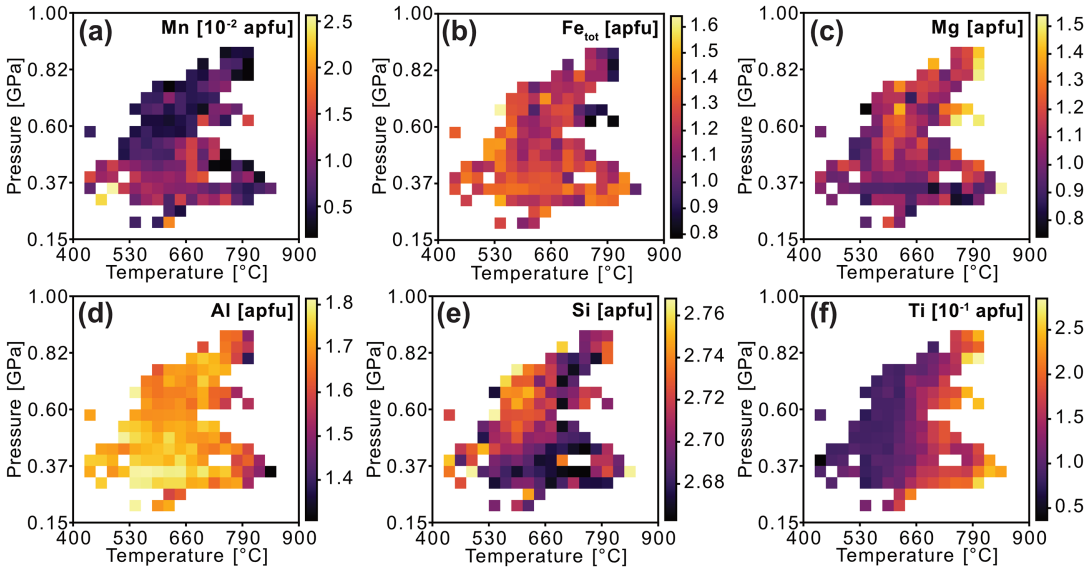

FIGURE 3 | Compositional variation of natural biotite from metamorphic sequences in *P–T* space for the elements: (a) Mn, (b) Fe_{tot}, (c) Mg, (d) Al, (e) Si and (f) Ti, all in atoms per formula unit (apfu).

3.2.1 | Compositional Variation of Simulated Biotite and Comparison to Natural Data

The Mn in biotite generated with each of the three different thermodynamic databases show patterns that mimic the topology of the garnet-in reaction line (cf. Figure 4a–c with figures 3 and 20 in Pattison and Forshaw 2025), as discussed above for the natural data. In the ds62 data, Mn reaches values above 0.025 apfu at low *P–T* conditions where garnet is most likely absent. These values are significantly higher than the natural data, which are generally below 0.015 apfu. For both ds55 datasets, the topology and the modelled absolute Mn content are consistent with the natural record.

Both Fe_{tot} and Mg show a strong anti-correlation with each other. Fe_{tot} is high and Mg is low in the *P–T* region outside garnet stability and vice versa within (Figure 4d–i). The dent of high Mg relative to Fe_{tot} at ~550°C best matches the natural data in extent and relative amplitude for the ds62 dataset. The absolute values for Fe_{tot} are slightly overestimated in ds62 but generally match well for both elements between simulated and natural data.

A clear anti-correlation can be observed for Al and Si in all three datasets, with Si decreasing from L*T*-H*P* conditions to H*T* conditions before increasing again at L*P* suprasolidus conditions (Figure 4j–o). The compositional gradient for both Si and Al is significantly steeper than the one observed in the natural data (Figure 3d,e). There are large differences in the absolute values and resulting ranges between the minimum and maximum Si and Al contents of biotite between the different thermodynamic databases. Both ds55 databases have overpronounced extrema for Al and Si, whereas ds62 is more consistent with natural biotite compositions.

The compositional gradient of Ti in all three datasets shows a pronounced increase towards higher temperatures while being close to zero with respect to pressure (Figure 4p–r). Compared to the natural data where Ti ranges from 0.1 to 0.25 apfu between 450°C and 800°C (Figure 3f), the data generated with the ds55Bt07 database, and to a lesser extent with ds62, systematically underestimate the Ti content with predicted values ranging from 0.05 to 0.15 apfu. The ds55BtT data are the most consistent with the natural data.

3.2.2 | Interpretation of Compositional Data

In summary, for all three datasets generated using phase equilibrium modelling, the topology of the compositional pattern of biotite in *P–T* space matches that observed in the natural data (Figure 3) to a first-order approximation. Minor topological changes are observed in the slope of the Si and Al surfaces. Differences emerge when looking at the absolute content of certain elements in the data, the most striking being the overestimation of Mn in ds62, and the overestimation of Al and corresponding underestimation of Si in both ds55 datasets. Based on these observations, there is no clear recommendation for or against using a specific thermodynamic database to train a prior model for transfer learning.

4 | Results—Model Calibration

In this section the results of the model selection experiments using 5-fold cross-validation are presented. The detailed results of the hyperparameter tuning and feature selection, including the training curves, of all experiments can be found in the Supplementary Materials S1 and S2.

4.1 | Cross-Validation for Model Selection

To select the final NN with the optimal set of hyperparameters, features and training procedure, the performance of potential candidates was evaluated using 5-fold cross-validation. The training dataset of natural biotite composition was split into five

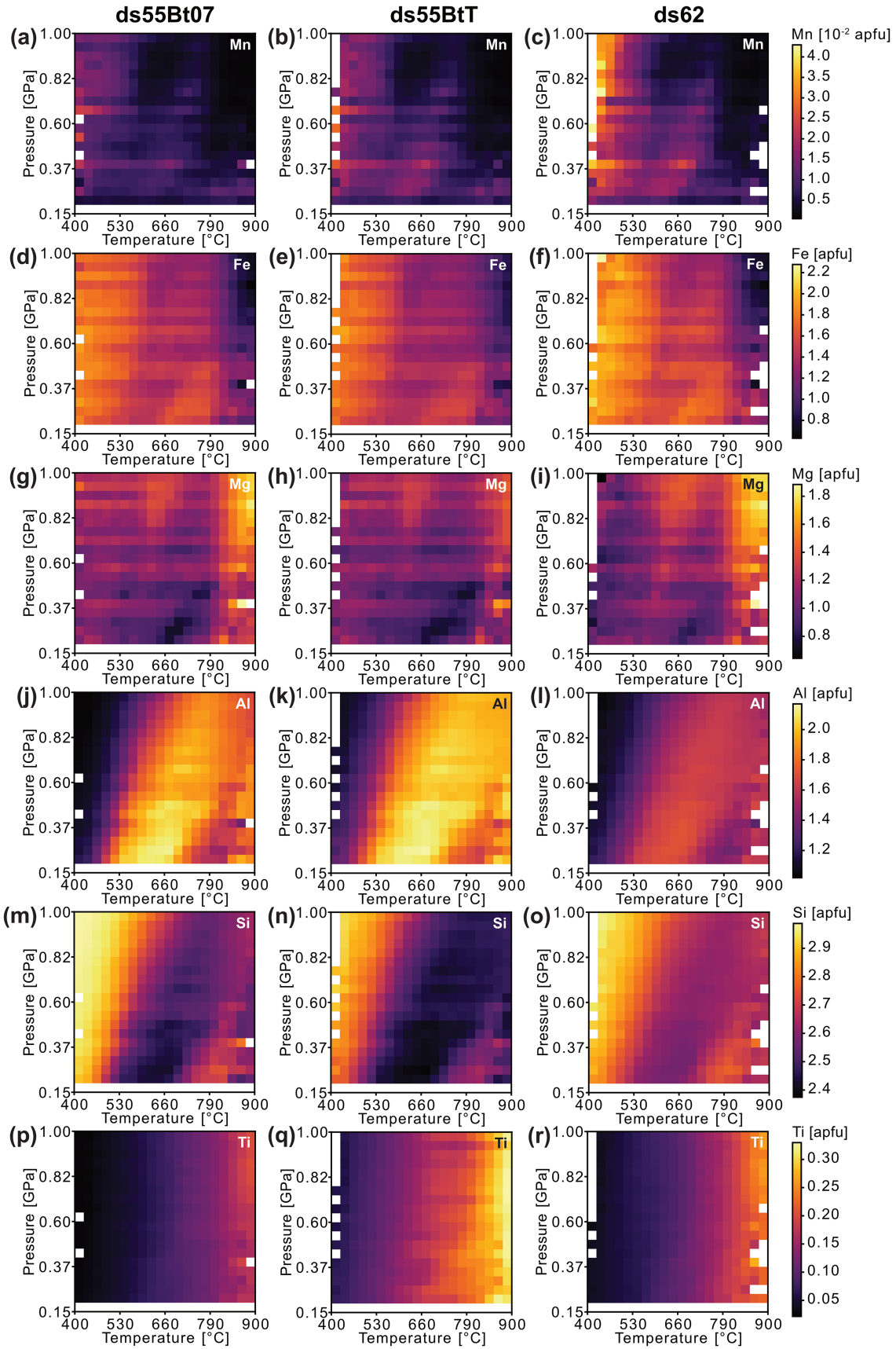

FIGURE 4 | Biotite composition for the three datasets generated with phase equilibrium modelling. (a, d, g, j, m, p) Biotite generated using the thermodynamic database ds62. (b, e, h, k, n, q) Biotite generated using the thermodynamic database ds55Bt07. (c, f, i, l, o, r) Biotite generated using the thermodynamic database ds55BtT.

80:20 training–validation splits, which are used in the training and evaluation of six models selected as a result of the initial hyperparameter tuning and feature selection (Table 1).

The RMSE for T is significantly lower for M2a and M3a/b/c (MnFMAST data) than for M1 (Ti-X_{Mg}) or M2b (MnFMAST + index minerals) (Figure 5a). Model M2a and the transfer learning models M3a/b/c perform similarly, with RMSE T in the range of 43°C–51°C for M2a and 41°C–46°C for M3a/b/c. A very similar systematic is found in the RMSE P (Figure 5b) with RMSE P of 0.12 GPa–0.13 GPa for M2a and 0.11 GPa–0.12 GPa for M3a/b/c. M2b shows significantly higher estimated uncertainties with a large spread for the five models evaluated in the cross-validation.

On the same validation sets the prior models, trained on simulated data, show orders of magnitude larger RMSE compared to models calibrated using natural data. The mean RMSE for T and P are 260°C and 0.33 GPa for the model trained on ds6.2, 530°C and 2.19 GPa for the model trained on ds55Bt07, and 88°C and 0.272 GPa for the model trained on ds55BtT (full results can be found in Supplemetary Material S3 and Figure S3.1).

A systematic relationship has been found between the RMSE for P and T and the range in P–T where the model is applied (Figure 6). The RMSE T, evaluated in 100°C wide zones, shows a u-shaped trend with the lowest uncertainties in the two zones between 500°C and 700°C. At higher temperatures, the RMSE increases, while the spread of RMSE between the 5-fold validation sets evaluated also increases systematically. Similarly, an even more pronounced increase in RMSE T is observed in the lowest temperature zones (400°C–500°C). A similar pattern is observed for the RMSE P calculated in 0.17 GPa width zones (Figure 6). The RMSE P shows a stronger increase and higher RMSE values at high pressure. The RMSE P of model M2b shows that for some of the five models evaluated this increase in uncertainty towards higher P is avoided (Figure 6c). However, the cross-validation shows a strong increase in the dispersion of the RMSE P towards this zone.

4.1.1 | Evaluation on Metamorphic Sequences

As a complementary assessment of the model's accuracy, an independent set of biotite analyses from metamorphic sequences not in the training dataset were used (Table 2). This evaluation is exemplified on the sequence from the Niagara Creek region, Quesnel Lake (Engi 1984) in Figure 7. The predicted P–T for each biotite analysis in these independent test sequences can be compared to the relative position of the metamorphic zone of the sample in the corresponding MAS in P–T space. The mean and standard deviation of the predicted P–T are calculated over all samples from each mapped zone in a MAS, resulting in a simplified single P–T estimate with associated uncertainty for each zone (Figure 7b).

To test the hypothesis of increasing metamorphic grade across any sequence, we examine whether there is a systematic ranking of P–T estimates from different zones within any MAS. In contrast to assessing absolute P–T, this does not require a precise and independent reference P–T. The rank test is conducted using Kendall's τ rank correlation coefficient between the predicted P–T and the authors' zone in the MAS. Kendall's τ is calculated with the number of concordant (C), and discordant (D) pairs of ranks in both sets (Kendall 1938):

$$\tau = (C - D) / (C + D) \tag{5}$$

For each model, τ is calculated for all test sequences and then averaged, weighted by the number of zones in each sequence, to obtain a single metric τ_{av} that captures whether the predicted P–T monotonous increase ($\tau \rightarrow 1.0$) or decrease ($\tau \rightarrow -1.0$) across zones in a MAS.

For each model tested (Table 1), the five individual models trained in the cross-validation process are used to evaluate τ_{av}, which accounts for the effect of performance variation due to training data selection and provides a reproducible estimate of the model's performance. For the temperature predictions, all evaluated models show a positive τ_{av}. All τ_{av} values of the

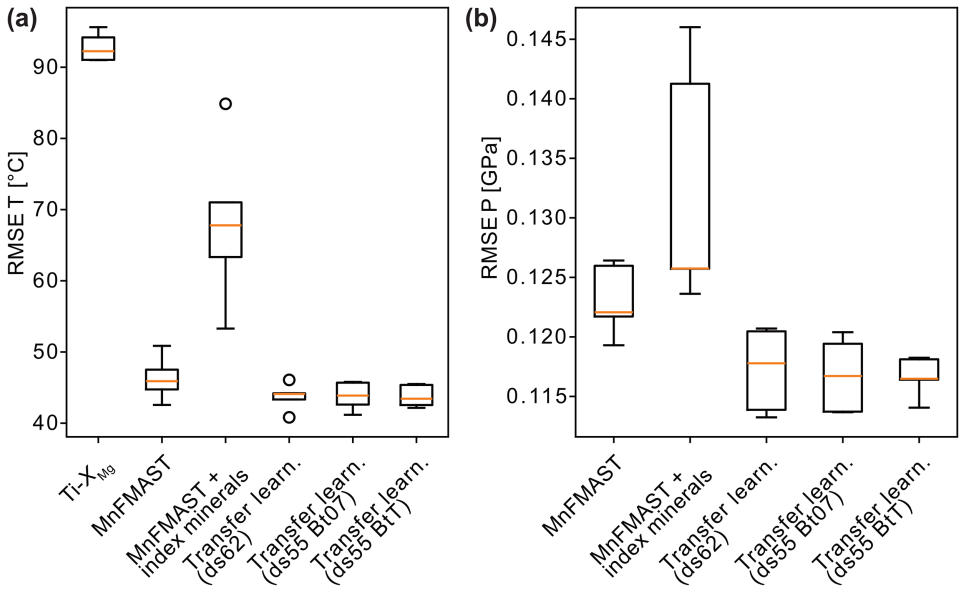

FIGURE 5 | Root-mean-square error (RMSE) evaluated using 5-fold cross-evaluation for different models. (a) The RMSE T and (b) RMSE P for all models evaluated. The models that were calibrated either directly or via transfer learning using the MnFMAST data perform best.

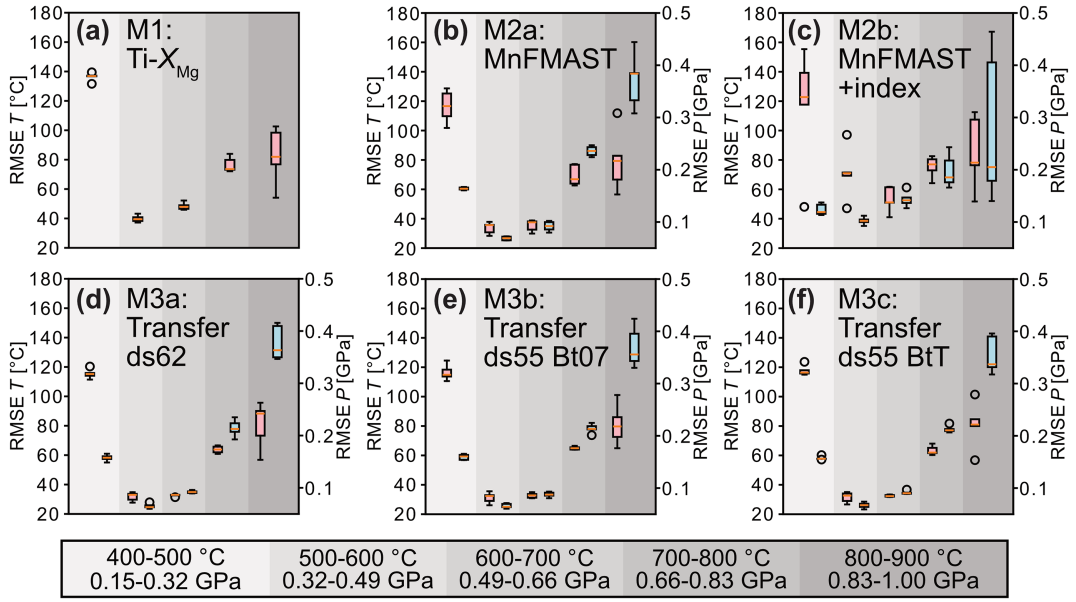

FIGURE 6 | *T*- and *P*-resolved RMSE *P* and *T* for all models evaluated. Data are binned separately (red markers, left boxplots) in *T* zones to evaluate the RMSE *T* and in P zones for the RMSE *P* (blue markers, right boxplots). (a) M1: Ti-X_{Mg} model, only RMSE *T* is evaluated, since the model is only trained as a thermometer. (b) M2a: MnFMAST model (c) M2b: MnFMAST + index model (d) M3a: transfer ds62 (e) M3b: transfer ds55 White07 (f) M3c: transfer ds55 BtT. The models perform best in zones where the most training data is available, between 500°C–700°C and 0.23–0.66 GPa, respectively.

different models lie within the range of 0.33–0.66, with no significant differences between them (Figure 8a). For the pressure, the τ_{av} values range from −0.28 to 0.08 for M2a and are weakly negative with a spread from −0.46 to −0.06 for M3a/b/c (Figure 8b). When only MAS with more than two zones are considered, the τ for *T* increases and all models fall within the range of 0.56–0.79. The τ for *P* range increases from −0.38 to −0.01.

4.1.2 | Investigation of the Propagation of Measurement Uncertainty

The propagation of measurement uncertainty is assessed below using two strategies. A first test was carried out using 27 biotite analyses from sample 28HF18 of Yogi et al. (2024), covering a range close to the natural variance in this sample. Predictions for *P*–*T* are made with the different models (M2a, M3a–c) and the results are shown in Figure 9. The second test involves a Monte Carlo simulation of the analytical uncertainty and aims to isolate the analytical uncertainty from other variables that could influence the observed variance in the model predictions. Analytical errors are assumed to be Gaussian distributed and to be primarily controlled by the counting statistics. A normal distribution with a standard deviation of ± 1% was used to simulate the uncertainty of an EPMA analysis. For Mn, which is typically very low in biotite, a larger uncertainty of ± 10% was used to account for the higher uncertainty associated with low counts. The 1000 simulated biotite measurements are then used to assess the variability of *P*–*T* predictions resulting from the propagation of the analytical uncertainty through the model (Figure 10).

The distributions of predicted pressures on sample 28HF18 from are narrower for the predictions of the models M3a–c (Figure 9b–d) than for the model M2a (Figure 9a). A similar trend, although less pronounced, was observed for distributions of predicted temperatures (Figure 9e–h). For both *P* and *T*, smaller differences between the five individual predicted distributions are observed for the transfer learning models M3a–c (Figure 9). This indicates a more reproducible fit as the predicted *P*–*T* is less affected by the choice of the training data during the cross-validation. A similar result is observed for the Monte Carlo simulated measurement uncertainty, where the models M3a–c (Figure 10b–d/f–h) show narrower distributions for both pressure and temperature predictions.

5 | Discussion of Model Calibration

The evaluation of the prediction uncertainty, using the RMSE for *P* and *T* and 5-fold cross-validation, shows that the models trained on the MnFMAST dataset (M2a and M3a-c) perform best (Figure 5). The comparison with the simple system model M1 analogous to the thermometer of Henry et al. (2005) confirms that the incorporation of additional information in the form of the MnFMAST system benefits model accuracy. This is consistent with the observation of systematic changes in these elements in the *P*–*T* space (Figure 3) and the fact that this system covers almost all proposed exchange reactions in biotite (Dubacq and Forshaw 2024). A potential K–Na exchange in the interlayer site and exchanges including Fe^{3+} explicitly are not considered, because the MnNKFMAST and $MnFe^{2+}Fe^{3+}$MAST systems have been discarded during the initial feature selection. Because $X_{Fe^{3+}}$ in metapelitic biotite is generally low (mean $\pm\sigma = 0.11 \pm 0.08$) and does not vary systematically with metamorphic grade (Forshaw and Pattison 2021), any bias introduced into the *P*–*T* predictions by simplifying assumptions about Fe oxidation state is considered negligible. This is further supported by the lack of performance improvement when Fe^{2+} and Fe^{3+} were treated as separate input features in a feature selection experiment.

TABLE 2 | Test dataset of metamorphic sequences.

Locality	References	MAS[a]	Zone[a]	Sample names of original authors
El Oro	Riel et al. (2013)	2a	Crd-And	AV0819a, AV0813
			Crd/And	PU0804
			Sil-(Crd-And)	VI0803, VI0804, AV0815
			Kfs-Sil-Crd-(And)	VI0806, AV0827a, PU0810
			Kfs-Crd ± Grt	PU0806, AV0828d, AV0832a, AV0833
N Bavaria (Moldanubian)	Blümel and Schreyer (1977)	2a	Kfs-Sil-Crd-(And)	1, 2, 3, 15
			Kfs-Crd ± Grt	6, 7, 8, 9, 10, 11, 12, 13, 14, 16, 17
Harpswell Neck	Forshaw (unpublished)	2c	Grt	PSM-50, PSM-52, PSM-65, PSM-68, PSM-70, JFM-24, H-153
			St	PSM-7, JFM-11, H-7, H-71, H-77
			St-And	PSM-11, PSM-12, PSM-13, PSM-83, PSM-148, JFM-25, H-91,H-172, H-180, H-184
			Sil-(St-And)	PSM-108, PSM-118, PSM-120, PSM-121,H-24, PSM-18, PSM-112, JFM-42, H-10, H-15-2, H-102, H-126
Leech River	Geen and Canil (2023)	2c	St-And	AG042, AG067, AG047, AG1161, AG049, AG0742
			Sil-(St-And)	AG080, AG041
Pichilemu	Hyppolito et al. (2015)	2c	Bt	DTH-1C
			Grt	DTH-1F
			St-And	DTH-68B, 154, DTH-70A, DTH-69A
Kluane	McKenzie (2023)	2c	Grt	19WM262
			Sil	19WM120
			Kfs-Crd	19WM116, 19WM118, 19WM123
Quesnel Lake	Engi (1984)	4a	Grt	154:2, 202:1, 205:1, 208:1
			St-Ky	26:1, 33:1, 40:1, 60:1, 100:1
			Sil	27:1, 30:2, 61:1, 63:1, 73:2, 101:1, 104:1, 110:2
Sierra de Guadarrama	López Ruiz et al. (1978)	4a/4b/5	Grt	36.329
			St	37.152, 36.328, 35.263, 35.325, 33.313
			St-Ky	35.326, 37.074
			Sil-(St-Ky)	37.043, 37.083, 8.481, 31.970, 47.172
			Kfs-Sil ± Crd	32.501, 16.966, 36.764, 22.035, 39.832
Clachncuydainn	Crowley et al. (1996)	4a/4b/5	Grt	135, 63
			St	13b, 13d
Saglek	Mengel and Rivers (1994)	4a/4b/5	Bt	F84-211
			St-Ky	F83-22, F83-79, F83-76, BR277
			Sil-(St-Ky)	BR-169, F83-131

[a]Nomenclature follows Pattison and Forshaw (2025).

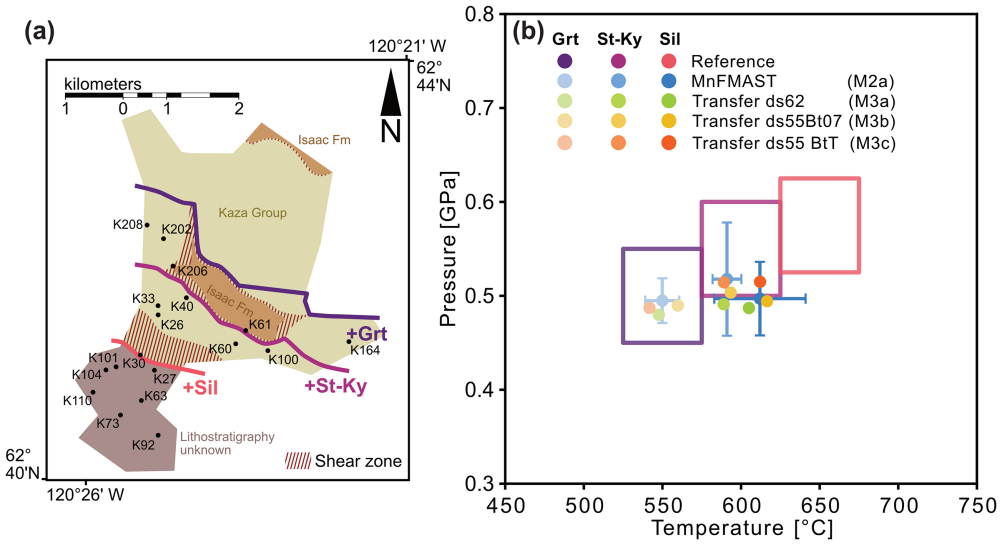

FIGURE 7 | Model evaluation using the Quesnel Lake metamorphic sequence. (a) Map of metamorphic zones and sample location at Quesnel Lake modified after Engi (1984). Samples from these zones can be used as test data even in the absence of a precise estimate of the exact *P–T* conditions, as it can be tested whether their relative order within the sequence is correctly predicted. (b) *P–T* predictions for biotite analyses reported in Engi (1984) from the Grt, St-Ky and Sil zones obtained with different models (M2a and M3a–c). The dots represent the mean *P–T* calculated from multiple analyses within each zone. The error bars (representing the standard deviation) are only shown for the predictions made with M2a for better readability but are of similar amplitude for the other models. A reference *P–T* field is given based on the *P–T* conditions for each zone in Supplementary Material S4.

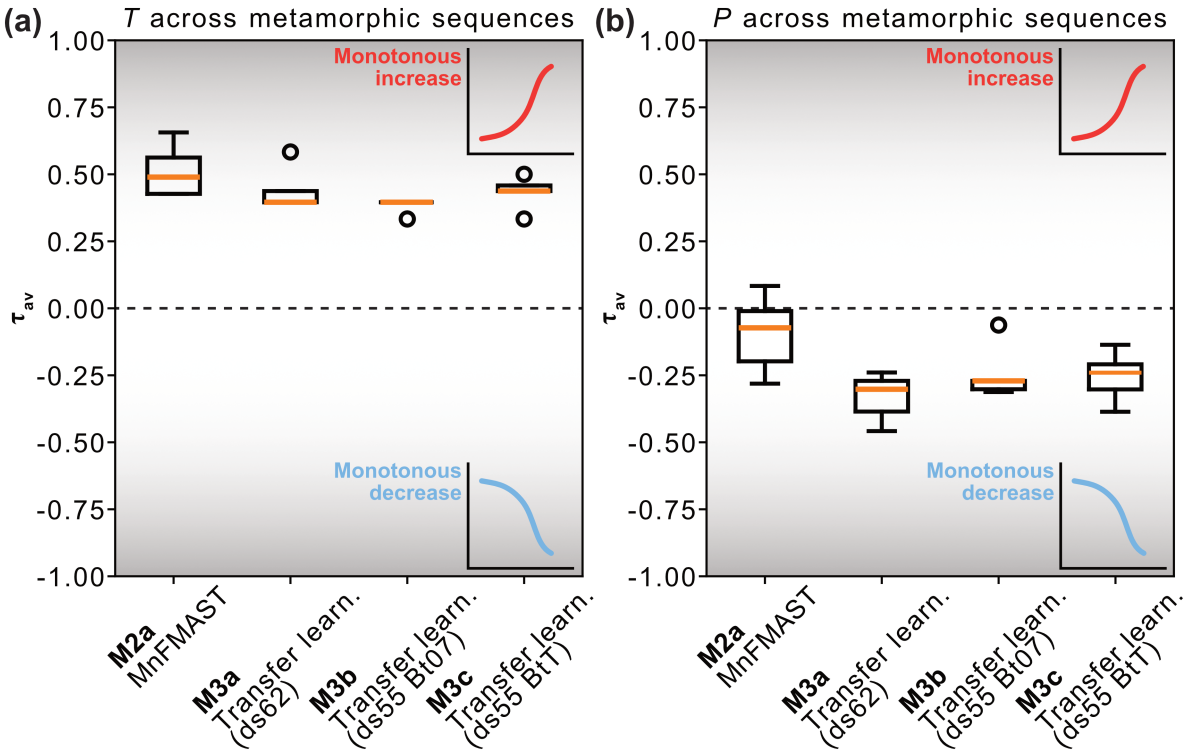

FIGURE 8 | Boxplots showing the rank correlation coefficient τ_{av}, to test if a model is predicting the systematics of *P–T* changes across metamorphic sequences. A weighted average τ_{av} of τ from 10 test sequences is calculated considering the length and the number of zones for each sequence. For each model, the evaluation is done five times using the five models trained for cross-validation to obtain a reproducible metric. (a) τ_{av} for the temperature across metamorphic sequences. The positive τ_{av} indicates that the models tend to correctly predict increasing temperatures across the zones of a MAS. (b) τ_{av} for the pressure across metamorphic sequences.

However, in highly oxidised systems—where most Fe may be present as Fe^{3+}—different exchange mechanisms could control Fe incorporation in biotite, and the current calibration is unlikely to be valid under such conditions.

The RMSE of ±~95°C of model M1 trained on Ti and X_{Mg} is substantially higher compared to the uncertainty reported for the original calibration of the Ti-in-Bt thermometer of ±25°C (Henry et al. 2005). We suggest that this difference is mainly

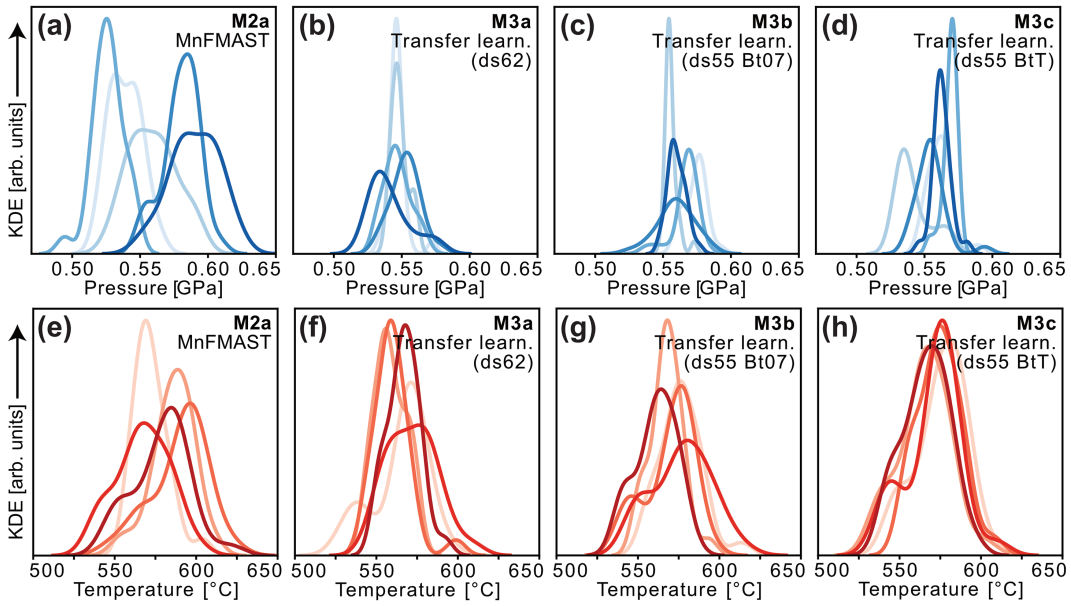

FIGURE 9 | Distribution of predicted *P* (top, blue) and *T* (bottom, red) for 27 chemical analyses of biotite from sample 28HF18 of Yogi et al. (2024). Predicted pressures for five models trained during cross-evaluation for M2a (a), M3a (b), M3b (c) and M3c (d). Predicted temperatures for five models trained during cross-evaluation for M2a (e), M3a (f), M3b (g) and M3c (h). The pressure and temperature distributions are shaded differently as the same prediction is repeated five times for each of the five models calibrated in the cross-validation process (see text). Narrower and more overlapping distributions for the models using transfer learning (b–d, f–h) indicate a more precise and reproducible fit.

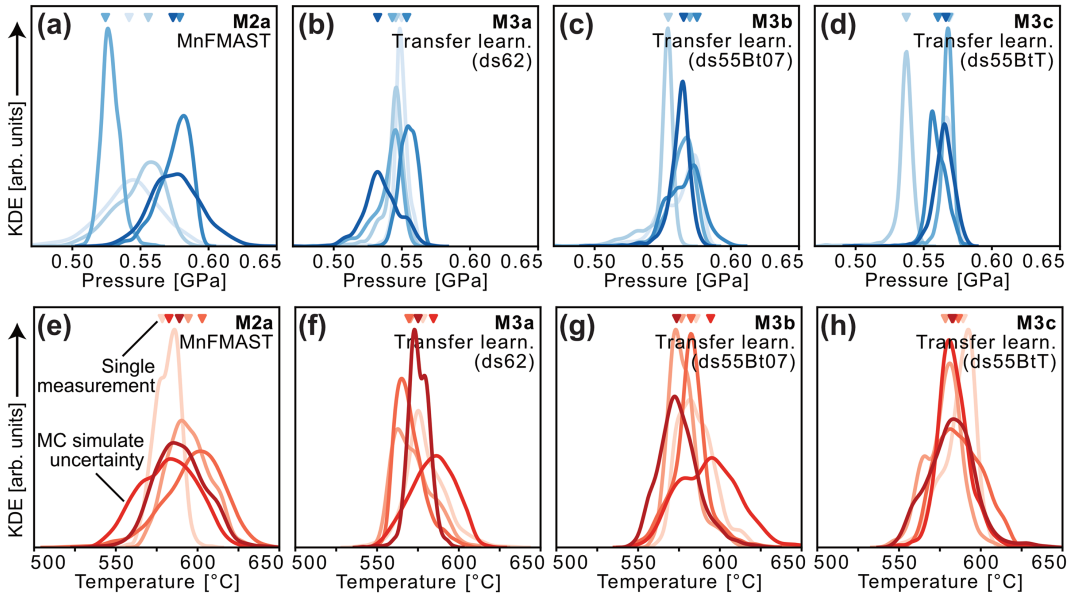

FIGURE 10 | Predicted *P* (top, blue) and *T* (bottom, red) on a single analysis picked from sample 28HF18 of Yogi et al. (2024) shown as a down-facing triangle. Distribution of predicted *P* and *T* for multiple compositional analyses from that single biotite analysis augmented using Monte Carlo simulation to reproduce a theoretical measurement uncertainty (see text). Predicted pressures for five models trained during cross-evaluation for M2a (a), M3a (b), M3b (c) and M3c (d). Predicted temperatures for five models trained during cross-evaluation for M2a (e), M3a (f), M3b (g) and M3c (h). Narrower and more overlapping distributions for the models using transfer learning (b–d, f–h) indicate a more precise and reproducible fit.

due to the method used by Henry et al. (2005) to obtain very precise *T* estimates for their constraints. These results imply that there is great potential for future calibration of an accurate biotite thermometer by combining their *T* allocation method with a model using biotite compositional vectors in the MnFMAST system as input. The large spread in both RMSE *T* and RMSE *P* for model M2b is a strong indication of overfitting (Figure 5). This emphasises the importance of thorough model evaluation, as only cross-validation revealed the tendency to overfit. In the initial feature selection tests, the model appeared to perform better on a randomly selected validation set when including index minerals (Figure S2.2).

The significantly higher RMSE of the prior models (Figure S3.1) is interpreted as resulting from the differences in biotite compositional patterns across *P*–*T* in the natural record compared

to phase equilibrium modelling, as described in Section 3.2.2. Although these differences have been found to be relatively minor, they represent shifts in the distribution between natural and simulated data, resulting in relevant systematic changes in *P–T* prediction.

Resolving the RMSE in *P* for zones of different *P* and the RMSE *T* for zones in *T,* respectively, is inconclusive for the model selection (Figure 6). However, this test shows that all the models evaluated are susceptible to regression to the mean. All models show an excellent fit with the lowest RMSE values for pressures between 0.3 and 0.6 GPa and temperatures between 500°C and 700°C (Figure 6). A comparison with the distribution of the training data in *P–T* (Figure 1a) shows that this is the region in *P–T* space where most data points fall.

Metamorphic sequences provide a record of frozen equilibrium peak metamorphic conditions preserved as metamorphic field gradients (Pattison and Forshaw 2025; Pattison and Tracy 1991). They are a promising complementary approach to systematically assess the accuracy of a thermobarometer without the need for precise numerical *P–T* estimates for individual test samples. Instead of comparing absolute differences between a *P–T* estimate and a reference, sequences allow an assessment of the accuracy by focusing on the relation between metamorphic zones within a MAS leveraging the sequential information. Samples from many low *P* sequences are interpreted to have equilibrated along isobaric *P–T* paths developed in contact aureoles (Pattison and Tracy 1991). While at elevated *P,* field gradients are typically established by a succession of increasing thermal maxima from *P–T* loops of regional metamorphism (e.g., Copley and Weller 2022). In the second case, the P_{peak} and T_{peak} do not coincide, and the pressure recorded may not be P_{peak}. This complicates the interpretation of tests assessing whether *P–T* predictions across different MAS zones increase monotonously. As the temperature can be safely assumed to increase along sequences of all *P*, testing for a monotonous increase across the sequential data provides a reproducible test for a thermometer's accuracy. For the pressure, no clear relation is expected to be found as the low-pressure MAS are most likely very close to isobaric and in the highest-pressure MAS, the possibility of (partial) re-equilibration during exhumation at the T_{max} prevents the prediction of a simple systematic increase in the recorded *P*. This makes MAS less suited to test barometers. Therefore, we focus our discussion on the thermometer. The τ_{av} values for *T* of all models show positive values around 0.5, confirming that the predictions of the NN thermobarometer reproduce a monotonous increase in *T* across sequences. A possible explanation for the only weak correlation, besides the limited predictive accuracy, may be the scattered distribution of predicted *P–T* for different biotite analyses from the same zone (see the error bars in Figure 7b). This distribution is interpreted to result from the partial persistence of metastable biotite or localised partial retrogression in the samples as shown by studies reporting compositional maps (e.g., Blackburn 1968; Lanari and Duesterhoeft 2019; Lanari and Hermann 2021).

The increase of the correlation metric for *T* when reducing the test dataset to sequences of two or more recorded zones highlights another complicating factor for the tests relying on MAS. Some metamorphic sequences might not record a straight-forward metamorphic field gradient but a more complex geological history that is especially hard to identify with only a few zones of the MAS realised in the rock record. This can result in misinterpreting the underlying *P–T* path and, therefore, inferred *P–T*.

The precision of the biotite thermobarometer is sensitive to the measurement uncertainty of the input biotite analyses. For a precise thermobarometer, it is desirable that the analytical uncertainty of the EPMA does not itself result in a large spread of the predicted *P–T*. To some extent, small compositional perturbations can naturally cause large differences in the predicted *P–T*, when exchange reactions involving an element *i* have a very steep slope of the concentration c_i with respect to *P* or *T*, large $\frac{\partial c_i}{\partial P}$ or $\frac{\partial c_i}{\partial T}$. One element that shows this is Ti in biotite, which was the property exploited in the calibration of the Ti-in-Bt thermometer (Henry et al. 2005; Wu and Chen 2015). Furthermore, analytical uncertainties typically result in normally distributed uncertainties for the biotite composition. Consequently, the predicted *P–T* should also be normally distributed so that no systematic shifts in distribution are introduced by the transformation of the model. The tests conducted to assess the propagation of compositional uncertainty to the *P–T* predictions show that the models using transfer learning (M3a–c) tend to result in more congregated predictions for both *P* and *T* for a biotite composition with associated uncertainty (Figures 9 and 10). Furthermore, this test suggests that a minimum uncertainty on the precision of the order of ±25°C and ±0.02 GPa is to be expected solely from the propagation of typical analytical errors in a biotite composition measurement.

5.1 | Model Selection

In conclusion, the models using the transfer learning strategy systematically perform best in the evaluation of RMSE, the test for monotonous increase across metamorphic sequences, and the qualitative assessment of propagating analytical uncertainty in the biotite composition. All models using transfer learning perform similarly, regardless of which dataset was used to train the prior model. However, M3c using the prior model trained with the dataset ds55BtT was selected, as this model showed narrow and almost perfectly normal distributed *P–T* predictions in the propagation of Monte Carlo simulated measurement uncertainty with a minimal scatter of the individual distributions for the five models from cross-validation.

The prior models used for transfer learning with a fine-tuning approach represent different starting points for converging to a minimum of the loss function. One might argue that the prior model trained on ds55BtT is closer to an optimal solution for the natural dataset, given its lower RMSE compared to the other priors when evaluated on the natural data. However, it is important to note that a better-performing starting point does not necessarily lie closer to the optimal solution. It simply provides a more favourable starting position for convergence on a high-dimensional loss surface, which may be flat in places and contain multiple local minima. This should not be interpreted as an assessment of how well a particular thermodynamic database reproduces the natural record of metamorphic biotite.

5.1.1 | Training of the NN-Thermobarometer

A NN with four hidden layers consisting of 64 neurons each was trained as the final model. The model requires an input biotite composition in the MnFMAST system, and outputs a *P–T* vector. The model's parameters were pre-trained using the generated dataset ds55BtT to train a prior model. For the final training, 90% of the training database was used, while the validation fraction was reduced to 10%. This change was made to maximise the available training data. It was facilitated because the previous tests had shown that overfitting was not an issue for the chosen model setup. The input data were normalised to a range of [−1, 1] and the *P–T* data in the training set were scaled to the range of [0.15, 1.0] GPa and [400, 900]°C, respectively, using a minimum–maximum scaler. Training was performed in batches of 50 data points using the Adam algorithm (Kingma and Ba 2014) as the optimiser with inverse time learning rate decay starting with a small initial learning rate of 0.0005. During training, additional regularisation in the form of dropout layers was applied with a rate of 20% after each of the first three hidden layers. The training was stopped by a validation loss minimum after 109 epochs at a validation RMSE of ± 0.0106 GPa and ± 40.8°C.

6 | Application of the Thermobarometer

The final NN biotite thermobarometer is applied to biotite of a metapelite from the Central Alps. This metapelite, sample MA9330, has been previously studied by Todd and Engi (1997) and Lanari and Hermann (2021). The inferred peak mineral assemblage consists of garnet porphyroblasts in a matrix of kyanite, white mica, plagioclase, biotite, quartz and accessory rutile (Lanari and Hermann 2021; Todd and Engi 1997). Staurolite, chlorite and ilmenite are interpreted as retrograde phases locally overgrowing kyanite, garnet, biotite and rutile (Lanari and Hermann 2021). A first estimate for the peak metamorphic conditions of 617°C ± 11°C and 0.618 ± 0.050 GPa was proposed by Todd and Engi (1997) who used multi-equilibrium calculations (TWQ software, Berman (1991)). A possible *P–T* path was refined by Lanari and Hermann (2021) using iterative thermodynamic modelling (Bingo-Antidote, Duesterhoeft and Lanari 2020). These authors proposed at least three distinct metamorphic stages recording prograde relics of the maximum burial at 1.1 GPa–560°C, the equilibration at the peak temperatures at 0.9 GPa–620°C, and a partial retrogression at 0.65 GPa–580°C.

The NN thermobarometer is applied to the compositional map 'Area 1' of Lanari and Hermann (2021). The newly applied thermobarometer predicts a pressure of 0.45 GPa (interquartile range: 0.41–0.51 GPa) and temperature of 606°C (interquartile range: 587°C–627°C) for the crystallisation of matrix biotite (Figure 11).

The predicted temperature overlaps with both determinations of the peak metamorphic conditions by multi-equilibrium calculation (Todd and Engi 1997) or iterative thermodynamic modelling (Lanari and Hermann 2021). The temperature for the retrograde stage proposed by Lanari and Hermann (2021) lies just below the lower end of the interquartile range of temperatures predicted by the NN thermobarometer. Partial re-equilibration of biotite resulting in a mixed signal of peak and retrograde equilibration temperature could possibly explain the predicted temperature distribution (Figure 11c). However, Lanari and Hermann (2021) report textural evidence of such retrogression in domains close to garnet porphyroblasts and not in the matrix domain analysed here. Therefore, we interpret the temperature variation as the result of propagating the analytical uncertainty from the compositional map. This is further justified by the excellent agreement between the Ti-in-Bt thermometers of Henry et al. (2005) and Wu and Chen (2015), and the predicted temperature of the NN thermobarometer for biotite (Figure 11c).

For the pressure predictions, a large discrepancy of 0.2–0.5 GPa is observed between the NN thermobarometer's predictions and the reference *P–T* determined by multi-equilibrium and iterative thermodynamic modelling (Figure 11). The predicted pressure of 0.45 GPa is well below the pressures proposed by Lanari and Hermann (2021) for peak metamorphic conditions of 0.9 GPa, or even 1.1 GPa at the maximum burial depth. The 0.62 GPa determined by Todd and Engi (1997), which is in good agreement with the metamorphic stage of the proposed partial retrogression, suggests that biotite may record a pressure closer to the NN predictions. However, it cannot be excluded that this is coincidental, and that the underestimation of pressure is an inherent problem of the NN thermobarometer, since systematic analysis of the NN performance shows a tendency for a strong regression towards the mean in the predicted pressures (Figures 6f and 12b) as discussed below.

7 | Performance Evaluation

To test the validity of the NN as a thermobarometer applicable to metamorphic biotite and to assess the uncertainty associated with the predicted *P–T*, the final model was evaluated on a test dataset compiled from the literature (Table 3 and Figure 12). For the thermometric predictions the uncertainty is estimated as a RMSE of $\Delta T = \pm 45$°C (Figure 12a). The predictions on the test samples with the uppermost and lowermost reference temperature indicate a minor regression to the mean phenomena. Although only weakly pronounced, this should be considered when applying the model to other samples and interpreting the predicted temperature. Looking at the different types of test data individually, the predictions on the 'single analysis' tend to overestimate the temperature given by the original authors to these samples and the errors are not centred around zero. This is explained by the regression to the mean because these 'single analysis' data points are all in the lower half of the temperature range tested. No such systematic overestimation can be observed for 'multi analyses' and 'EPMA maps'. The thermometer uncertainty ΔT, of ± 45°C, is within the expected uncertainty arising from the *P–T* assignment method. The metamorphic zones in MAS used to assign *P–T* to the biotite analyses in the training dataset are of a width of 50°C–100°C (see in Section 2.1). Therefore, the model performance is most likely limited by the uncertainty of the temperature in the constraints used and convergence to a generally valid optimal solution was successful.

The barometric predictions on the test set show an RMSE of $\Delta P = \pm 0.37$ GPa (Figure 12b). However, the pressure predictions

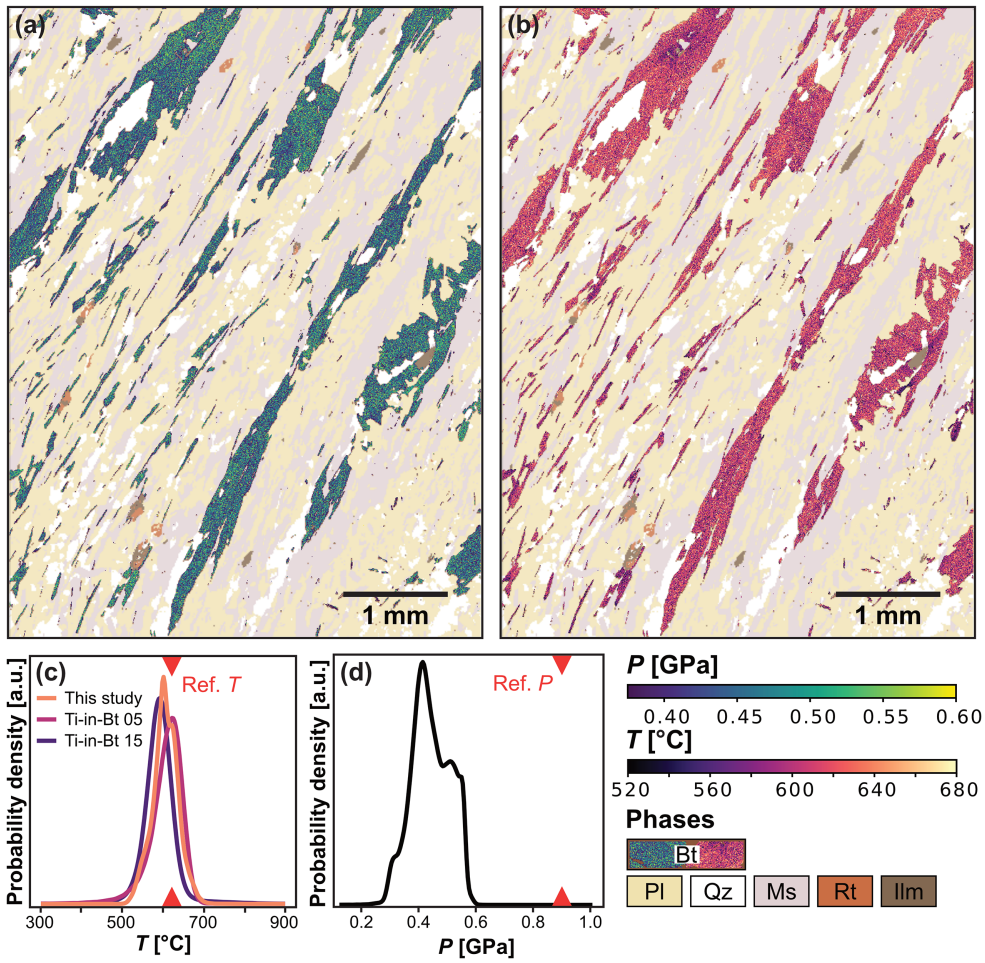

FIGURE 11 | An application example of the final model applied to sample MA9330, a metapelite from the Central Alps. The compositional map obtained by Lanari and Hermann (2021) with predicted pressure (a) and temperature (b) for biotite. (c) The predicted temperatures of biotite equilibration are compared with the results of Ti-in-Bt thermometry using the calibrations of Henry et al. (2005) and Wu and Chen (2015). The reference temperature was determined by Lanari and Hermann (2021) using iterative thermodynamic modelling (see text). (d) Comparison between the NN predicted pressures and the reference pressure for the equilibration at the thermal peak determined by iterative thermodynamic modelling.

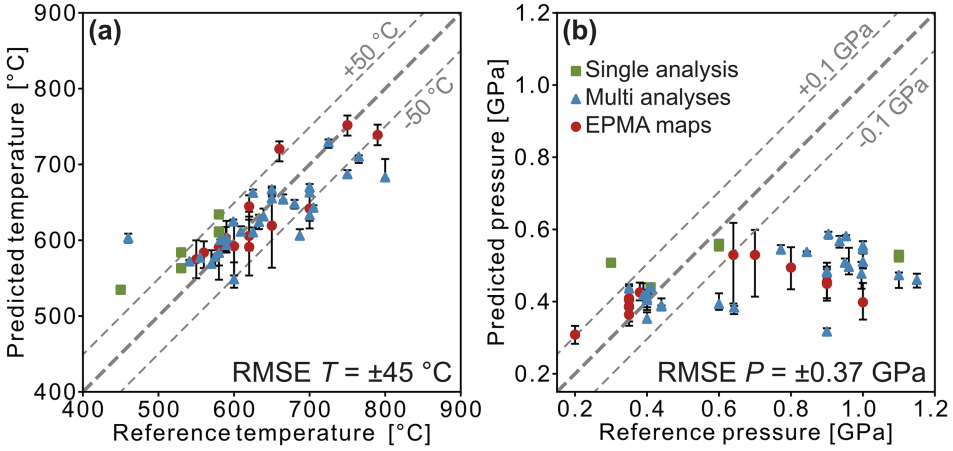

FIGURE 12 | Evaluation of the final model on the test dataset compiled from the literature. The dataset consists of single analyses selected by the original authors, multiple analyses of the biotite composition, or compositional maps of biotite for which *P*–*T* were determined using a combination of methods, most including phase equilibrium modelling, to obtain an estimate of the metamorphic conditions. (a) The predicted *T* by the model plotted against the reference *T*. (b) The predicted *P* plotted against the reference *P*.

TABLE 3 | Test dataset.

Locality	References	Reference *P* (GPa)	Reference *T* (°C)	Thermobarometry[a]	Sample	Biotite analysis
Vedrette di Ries	Tajčmanová et al. (2009)	0.410	580	PEM	VR515	Single analysis
Everest Region	Cottle et al. (2011)	0.300	450	PEM, MS	Ka-71	Single analysis
Longmen Shan	Airaghi et al. (2017)	1.100	530	PEM, TiBt, RSCM, GBT, ChlMicaEqui	To13-4/7 Bt core	Single analysis
Longmen Shan	Airaghi et al. (2017)	0.600	580	PEM, TiBt, RSCM, GBT, ChlMicaEqui	To13-4/7 Bt rim	Single analysis
Longmen Shan	Airaghi et al. (2017)	1.050	580	PEM, TiBt, RSCM, GBT, ChlMicaEqui	w15-5	Single analysis
Seve Nappe Complex	Jeanneret et al. (2022)	1.150	600	QuiG, TiQ, PEM, GBMP	516-2	Multiple analyses
Kluane	McKenzie (2023)	0.400	650	PEM, AvPT, GBT	19wm116	Multiple analyses
Kluane	McKenzie (2023)	0.400	650	PEM, AvPT, GBT	19wm118	Multiple analyses
Kluane	McKenzie (2023)	0.400	590	PEM, AvPT, GBT	19wm120	Multiple analyses
Kluane	McKenzie (2023)	0.410	725	PEM, AvPT, GBT	19wm123	Multiple analyses
Kluane	McKenzie (2023)	0.350	460	PEM, AvPT, GBT	19wm262	Multiple analyses
Gneiss Dome Belt	Hillenbrand et al. (2023)	0.900	800	TWQ, TiBt	ga-92-4-6	Multiple analyses
Gneiss Dome Belt	Hillenbrand et al. (2023)	1.100	765	TWQ, TiBt	gg-92-5-1	Multiple analyses
Gneiss Dome Belt	Hillenbrand et al. (2023)	0.900	680	TWQ, TiBt	IWH19-220	Multiple analyses
Gneiss Dome Belt	Hillenbrand et al. (2023)	1.000	625	TWQ, TiBt	IWH19-290	Multiple analyses
Gneiss Dome Belt	Hillenbrand et al. (2023)	1.400	687	TWQ, TiBt	IWH19-299	Multiple analyses
Gneiss Dome Belt	Hillenbrand et al. (2023)	0.950	700	TWQ, TiBt	IWH19-i5	Multiple analyses
Gneiss Dome Belt	Hillenbrand et al. (2023)	1.000	700	TWQ, TiBt	IWH19-i21	Multiple analyses
Gneiss Dome Belt	Hillenbrand et al. (2023)	1.000	625	TWQ, TiBt	IWH19-312	Multiple analyses
Gneiss Dome Belt	Hillenbrand et al. (2023)	0.900	750	TWQ, TiBt	gsg92-1-2	Multiple analyses
Leech River	Geen and Canil (2023)	0.390	590	RSCM, GBT, GBAQ	AG080	Multiple analyses
Leech River	Geen and Canil (2023)	0.440	570	RSCM, GBT, GBAQ	AG042	Multiple analyses
Leech River	Geen and Canil (2023)	0.440	555	RSCM, GBT, GBAQ	AG067	Multiple analyses
Leech River	Geen and Canil (2023)	0.440	580	RSCM, GBT, GBAQ	AG1161	Multiple analyses

(Continues)

TABLE 3 | (Continued)

Locality	References	Reference *P* (GPa)	Reference *T* (°C)	Thermobarometry[a]	Sample	Biotite analysis
Suru Valley	Cawood (2024)	0.772	575	LinaForma	ICSV86	Multiple analyses
Suru Valley	Cawood (2024)	0.844	541	LinaForma	ICSV130	Multiple analyses
Suru Valley	Cawood (2024)	0.904	599	LinaForma	ICSV109	Multiple analyses
Suru Valley	Cawood (2024)	0.953	609	LinaForma	ICSV113	Multiple analyses
Suru Valley	Cawood (2024)	0.961	582	LinaForma	ICSV13	Multiple analyses
Suru Valley	Cawood (2024)	0.935	633	LinaForma	ICSV12	Multiple analyses
Suru Valley	Cawood (2024)	1.215	665	LinaForma	ICSV91	Multiple analyses
Suru Valley	Cawood (2024)	0.996	639	LinaForma	ICSV81	Multiple analyses
Suru Valley	Cawood (2024)	0.600	700	LinaForma	ICSV145	Multiple analyses
Suru Valley	Cawood (2024)	0.642	705	LinaForma	ICSV150	Multiple analyses
Nelson Aureole	Diwa (2023)	0.350	550	MAS, TiBt	08-CW-7.5A	Compositional map
Nelson Aureole	Diwa (2023)	0.350	560	MAS, TiBt	19-CW-4A	Compositional map
Nelson Aureole	Diwa (2023)	0.350	580	MAS, TiBt	19-CW-5	Compositional map
Nelson Aureole	Forshaw (unpublished)	0.350	600	MAS	19-CW-12A	Compositional map
Nelson Aureole	Forshaw (unpublished)	0.350	650	MAS	93-CW-19A	Compositional map
Seridó Belt	Cioffi et al. (2021)	0.380	589	PEM, Bingo-Antidote	3A2	Compositional map
Southern Central Alps	Lanari et al. (2021)	0.800	700	Bingo-Antidote	AA19	Compositional map
Croveo, Central Alps	Lanari and Hermann (2021)	0.900	620	PEM, Bingo-Antidote, TiBt	MA9330	Compositional map
Croveo, Central Alps	Lanari and Hermann (2021)	0.900	620	PEM, Bingo-Antidote, TiBt	MA9330	Compositional map
Sikkim, Himalaya	Lanari and Duesterhoeft (2019)	0.640	790	PEM, Bingo-Antidote	TG8C-03	Compositional map
Grenville Province	Duesterhoeft and Lanari (2020)	0.700	660	PEM, Bingo-Antidote	HAL104	Compositional map
Cima Lunga	Piccoli et al. (2022)	1.000	620	PEM, Bingo-Antidote	CdG	Compositional map
El Oro	Lanari (pers. comm.)	0.200	750	PEM, Bingo-Antidote	Xenolith	Compositional map

Note: See in Section 2.2 in the main text for the references of the thermobarometers mentioned in this table.

[a]Phase equilibrium modelling (PEM), microstructures (MS), Ti-in-Bt (TiBt), garnet-biotite thermometry (GBT), metamorphic sequence (MAS) chlorite-mica equilibria (ChlMicaEqui), quartz in garnet (QuiG), Grt-Bt-Ms-Plg barometer (GBMP), average pressure–temperature estimate (avPT), Raman spectra of carbonaceous material thermometry (RSCM); garnet-biotite-aluminosilicate-quartz (GBAQ).

seem to be systematically biased towards the lower pressures. As the training dataset is dominated by biotite coming from low-pressure ($P_{\text{last equilibrated}} < 0.6$ GPa) rocks (Figure 1a), this is interpreted as a strongly pronounced asymmetric regression to the mean phenomena. This strong systematic bias limits the use of the barometer to this pressure range. The uncertainty estimated using the RMSE P should not be taken as the true model uncertainty, as this metric does not account for the systematic bias towards lower pressures. When applied to low-pressure rocks the uncertainty could be significantly lower. However, it will be drastically underestimated at pressures > 0.6 GPa.

The test dataset also reveals two important limitations of investigating metapelitic phase equilibria using MAS. The MAS are limited towards a relatively low P_{max}, recorded by the Grenvillian-type sequence (MAS 5) with a reported pressure maximum of ~0.8 GPa (Pattison 2001; Rivers 1983). Only six localities in the worldwide compilation show this MAS, further skewing the database towards lower-pressure sequences (Pattison and Forshaw 2025). Further, as the sequences are laid out by the equilibration of a metamorphic rock during the T_{max} of its P–T path (Pattison and Forshaw 2025), they are inherently biased towards the pressure at T_{max} and not P_{max}. Evidence of equilibration at higher pressures during orogenesis, sometimes only present as relics like cores in zoned minerals or metastable phases at the thermal maximum, can be found in metapelites from the Alps (Lanari and Hermann 2021; Piccoli et al. 2022), the Scandinavian Caledonides (Jeanneret et al. 2022), the Appalachians (Hillenbrand et al. 2023) and the Himalayas (Airaghi et al. 2017; Cawood 2024; Lanari and Duesterhoeft 2019).

The absolute uncertainties for empirical thermobarometers are generally on the order of ±0.1 GPa and ±50°C (Essene 1989). Therefore, the machine learning calibration presented here has resulted in a model that can predict the temperature of biotite equilibration with an uncertainty comparable to other methods. The calibration of a barometer has failed due to a systematic bias towards the low P dominated training data that prevents the model from being applied. Potential sources of uncertainty in the training data that hinder a more accurate calibration are as follows:

1. A lack of accuracy in the constraints, when using the MAS with individual metamorphic zones spanning over 50°C–100°C to determine a P–T for each biotite analysis. While the method used with ordering the samples after the temperature predicted by the Ti-in-Bt thermometer of Henry et al. (2005) proved itself effective, it seems to be limited to the uncertainty achieved here and further is potentially biasing the model towards that older calibration of Henry et al. (2005).
2. The persistence of metastable biotite or localised partial retrogression. If the measured biotite composition does not represent the equilibrium composition that corresponds to the P–T determined by a peak assemblage characteristic for a specific zone in a MAS, it introduces additional uncertainty in this P–T–X_{biotite} triplet.
3. The analytical precision of the EPMA analysis and the treatment of compositional data.

Focusing on the thermometer in the NN model, the calibration of this study is compared with existing single-crystal thermometers for biotite from Henry et al. (2005) and Wu and Chen (2015), both of which are based on the temperature dependence of Ti incorporation in biotite. All three models are (re-)evaluated on the test dataset to ensure comparability. They all perform similarly on the test data (Figure 13). The calibration of Henry et al. (2005) shows the lowest RMSE and a distribution of absolute deviations that is symmetrical around zero (Figure 13c). The calibration presented in this work and the Ti-in-Bt thermometer of Wu and Chen (2015) show similar RMSE. The distribution of absolute errors for the Wu and Chen (2015) model is biased towards negative absolute deviations suggesting a possible systematic underestimation of temperatures in the calibration (Figure 13d), while

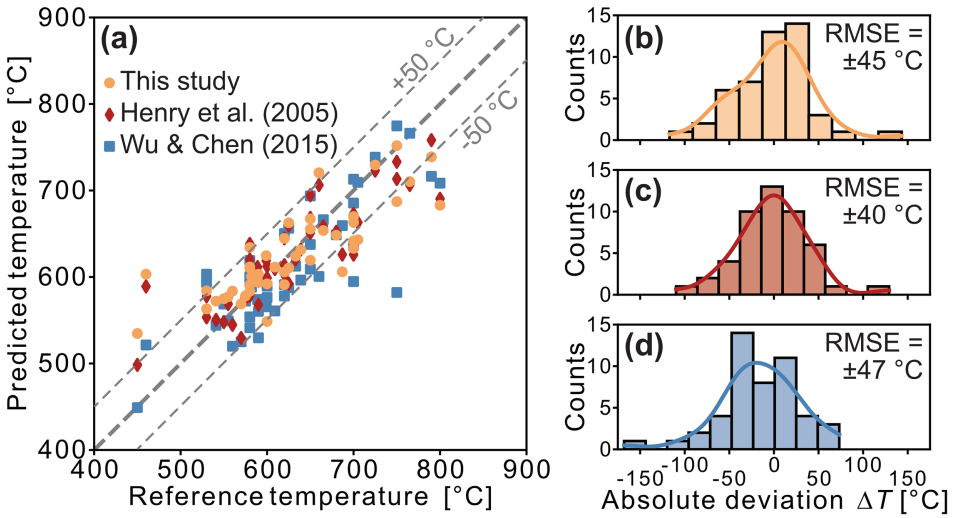

FIGURE 13 | Comparison of the single-crystal biotite temperature predictions by the neural network obtain in this work with Ti-in-Bt thermometry. (a) Temperature predictions on the test dataset plotted against the reference temperature for the neural network, the thermometer of Henry et al. (2005), and Wu and Chen (2015). Distribution of the absolute errors for the neural network (b), Henry's thermometer (c) and the thermometer of Wu and Chen (d).

the NN thermometer shows a similar but less pronounced bias towards positive absolute deviations (Figure 13b). It is quite possible that this test underestimates the uncertainty of the Henry et al. (2005) Ti-in-Bt model, because it is a well-established thermometer and was used along with other methods to constrain the peak *P–T* in 18 of the 48 samples included in the test dataset (Table 3).

8 | Physical Significance of the Empirical Biotite Thermometer

The NN thermobarometer is a purely empirical fit to the natural data of biotite from metamorphic sequences. The underlying metamorphic processes controlling a biotite composition to crystallise at a certain *P–T*, typically thought of as a combination of thermodynamic equilibrium and reaction kinetics, are not explicitly modelled but must be implicitly considered by the model, being the causal link causing the observable trends captured with the NN.

A NN is a curve-fitting algorithm, which is fundamentally different from the Random Forest based ensemble models that have recently been successfully used to calibrate machine learning geothermobarometers (Jorgenson et al. 2022; Petrelli et al. 2020; Weber and Blundy 2024). The fitted curve is a differentiable mathematical function that can be examined using calculus. By using partial derivatives, the elements that cause the largest changes in the *P–T* predictions can be identified. The partial derivatives of the barometer $\frac{\partial P}{\partial x_i}$ and thermometer $\frac{\partial T}{\partial x_i}$ with respect to each input element x_i is calculated at all 2148 points in the training dataset. The amplitude of compositional variation in biotite differs between elements. By multiplying a vector $\vec{u}$, 10% of the interquartile range for each element as entries, and the derivatives each partial derivative is normalised:

$$\frac{\partial P}{\partial x_i} \cdot \vec{u} \text{ or } \frac{\partial T}{\partial x_i} \cdot \vec{u} \tag{6}$$

The normalised partial derivatives (PDVs) of the barometer with respect to Si, Ti, Fe_{tot} and Mg are symmetrical around zero (Figure 14a). The negative normalised partial derivatives with respect to Al would be consistent with a pressure-dependent Tschermak substitution observed in synthetic biotite from piston-cylinder experiments at higher pressures (Hermann 2003). The mostly negative PDVs with respect to Mn are interpreted to result from the scavenging of Mn by garnet within the garnet stability field at elevated pressures (Baxter et al. 2017; Tracy 1982). Since it is not possible to distinguish artefacts from features of the function with the poorly performing calibration, the analysis of the barometer is not pursued further.

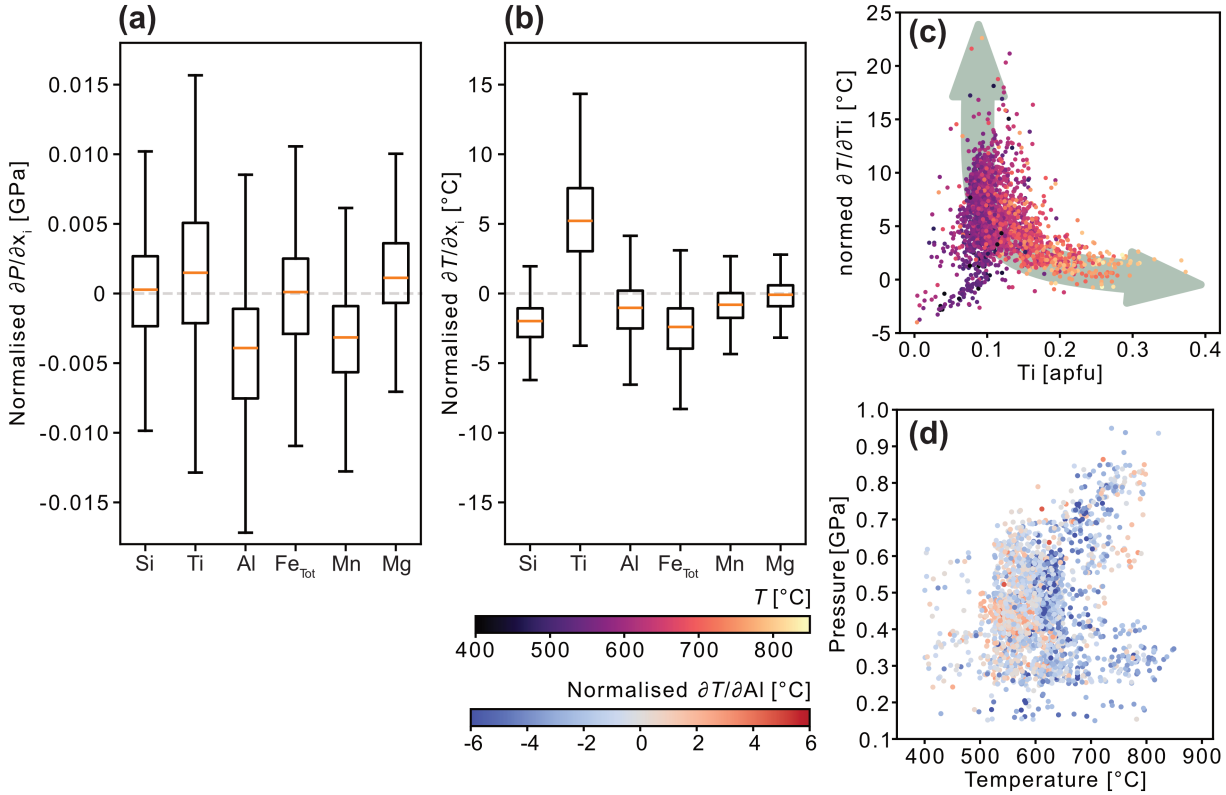

FIGURE 14 | Analysis of the curve fitted as a thermobarometric function by the machine learning model. (a) Partial derivative (PDV) of the barometer with respect to each input element normalise by a 10% change of the expected natural variation for each element. (b) Partial derivative of the thermometer with respect to each input element normalised by a 10% change of the expected natural variation for each element. (c) Normalised PDVs of the thermometer with respect to Ti plotted against the Ti content of the biotite evaluated. The trend shows that the increase in predicted temperatures with increasing Ti diminishes towards biotite with a high Ti content. (d) Normalised PDVs of the thermometer with respect to Al in *P–T* space indicate a switch of the effect of Al incorporation on the predicted temperatures. For temperatures below ~620°C the incorporation of Al has little effect on the predicted temperature, while above predicted temperatures are lowered when Al increases.

The normalised PDVs of the thermometer are largely negative with respect to Si, Al, Fe_{tot} and Mn (Figure 14b). Only for Ti are the PDVs predominantly positive. The normalised PDVs with respect to Mg are small and centred around zero. As the biotite composition is put in the NN as cations per formula unit, the sum of cations is fixed, and therefore, its variables are dependent. A dependency of Mg potentially explains why it is not used in the thermometer by the NN. The PDVs of T with respect to Ti show an increase of the predicted temperature between 3.0°C (25th percentile of PDVs) and 7.6°C (75th percentile PDVs) for an increase in the Ti content corresponding to 10% of the natural variation. When the Ti content in biotite is already high, the same increase of Ti results in a much smaller increase in predicted T compared to biotite with low Ti (Figure 14c). This agrees with the proposed mathematical descriptions of the Ti-saturation in biotite using a power-law (Henry et al. 2005) and indicates the dominant incorporation mechanism of Ti might change towards higher Ti content in biotite. Further evidence of such a change can be seen by examining the normalised PDVs of the thermometer with respect to Al (Figure 14d). For temperatures below ~620°C, these PDVs are small (±2°C) and scattered around zero. Above ~620°C, negative PDVs dominate, extending as low as −6°C per normed change in Al. This is not compatible with a Tschermak-type Ti substitution, $^{(VI)}Mg^{2+} + 2^{(IV)}Si^{4+} \leftrightarrow {}^{(VI)}Ti^{4+} + 2^{(IV)}Al^{3+}$, where Al and Ti would strongly correlate. Instead, the decoupling of the effect on T by Al and Ti favours Ti deprotonation, $^{(VI)}Mg^{2+} + 2(OH)^{-} \leftrightarrow {}^{(VI)}Ti^{4+} + 2O^{2-}$, as the dominant process. Ti incorporation by deprotonation has been shown to most likely be the dominant process resulting in Ti enrichment in metamorphic biotite (Dubacq and Forshaw 2024). Henry et al. (2005) note that the dominant Ti-incorporation mechanism may change with increasing metamorphic grade. In their data, deprotonation best fits biotite above the staurolite zone, while a Tschermak-type substitution best explains the observed compositional variation below. Likewise, the analysis of the NN thermometer suggests that the Ti-deprotonation might only become relevant at temperatures above ~620°C. While the study of the atom-scale geochemical processes is beyond the scope of this study, this analysis using PDVs demonstrates the potential feedback of purely empirical machine learning into physical models of metamorphic processes.

9 | Conclusion

To study metamorphic processes, the ability to accurately predict P–T for the natural record of metamorphic rocks and minerals is crucial. However, calibrating thermobarometers on natural data is challenging since suitable constraints with precise P–T estimates, that are difficult to obtain without biasing the model towards an existing calibration, are often sparse. In this work, we addressed this by training a NN on the natural record of biotite from mineral assemblage sequences. The model is pre-trained on a large dataset simulated with phase equilibrium modelling before fitting the natural data. Utilising this transfer-learning approach allows the fitting of a high-capacity machine learning model to sparse natural data without overfitting. We outline a calibration strategy that provides petrologists with a systematic and reproducible approach for calibrating machine learning models as thermobarometers and provide complementary tests to evaluate the performance of a model. We conclude the following:

1. Metamorphic assemblage sequences provide relative natural constraints for the calibration of a thermobarometer. The sequential information of metamorphic zones can provide a P–T estimate independent of the measured mineral composition. This approach reduces potential bias towards a specific calibration by not relying on existing geothermobarometers. This is because it is based on mineral assemblages and not mineral composition. Therefore, any interpretation of a metamorphic field gradient, ultimately providing the P–T, emerges from the examination of evidence from multiple observations across a sequence.
2. The available natural constraints are relatively sparse and are strongly clustered around intermediate-pressure amphibolite-facies conditions. This sparsity and clustering pose critical limitations to the performance of any calibrated model due to pronounced regression to the mean phenomena, which is a lasting challenge when applying machine learning to petrological problems. In the case presented here, this hinders the successful calibration of a barometer using the MAS data.
3. Phase equilibrium modelling using three different thermodynamic databases can provide large datasets that cover the complete P–T space of biotite stability, filling data gaps in the sparse natural record. This synthetic data provides a suitable prior to employ transfer learning in the calibration of NNs as thermometers. All three of the generated datasets resemble the natural record of biotite composition, with only minor deviations in the topology of compositional gradients or absolute concentrations of elements.
4. The calibration strategy presented for machine learning thermobarometers provides a workflow for systematically selecting input features, tuning hyperparameters (including learning rates, model architectures and regularisation methods) and benchmarking potential models. The combined use of K-fold cross-validation, evaluation of prediction accuracy along metamorphic sequence P–T trends, and precision assessment via Monte Carlo propagation of measurement uncertainties is well suited to machine learning problems with limited data. Importantly, this approach avoids the need to withhold additional data solely for model validation.
5. The final comparison with test data compiled from published estimates using a combination of multiple thermobarometric methods indicates that the temperature can be predicted with an RMSE of ±45°C. A reliable prediction of pressure is hindered by a systematic underestimation of pressures >0.6 GPa, which is interpreted as a result of scarce natural data above these conditions. The estimated uncertainty of the thermometer is in good agreement with Ti-in-Bt thermometry evaluated on the same test dataset.
6. The comparison of the calibration presented here with other single-crystal biotite thermometers shows a very similar performance of all three calibrations. The uncertainty

of predicted T on natural test data spanning the P–T of crustal metamorphism is around ± 50°C for all thermometers. The original calibration of Henry et al. (2005) appears to be the most precise; although calibrated only for P between 0.4 and 0.6 GPa in graphitic ilmenite- or rutile-bearing assemblages, its simple formulation has wide general validity.

Acknowledgements

We thank Jörg Hermann for the insightful discussions on natural and experimental phase equilibria, which greatly enriched this work. Hugo Dominguez is thanked for his help, many stimulating discussions and shared passion for Python programming. Dave Pattison is thanked for his probing comments and general interest in this work. Maurizio Petrelli and a anonymous reviewer are thanked for their constructive comments, and Richard White for his editorial handling. Open access publishing facilitated by Universite de Lausanne, as part of the Wiley - Universite de Lausanne agreement via the Consortium Of Swiss Academic Libraries.

Data Availability Statement

The code for the training of the neural network thermobarometer underlying this article is available in the GitHub repository (https://github.com/Philipsite/ml-biotite-thermobarometry). The code for using the single-crystal biotite thermometer presented in this article is available in the GitHub repository (https://github.com/Philipsite/ml-bt-thermometer).

References

Airaghi, L., P. Lanari, J. de Sigoyer, and S. Guillot. 2017. "Microstructural vs Compositional Preservation and Pseudomorphic Replacement of Muscovite in Deformed Metapelites From the Longmen Shan (Sichuan, China)." *Lithos* 282-283: 262–280. https://doi.org/10.1016/j.lithos.2017.03.013.

Baxter, E. F., M. J. Caddick, and B. Dragovic. 2017. "Garnet: A Rock-Forming Mineral Petrochronometer." *Reviews in Mineralogy and Geochemistry* 83, no. 1: 469–533. https://doi.org/10.2138/rmg.2017.83.15.

Berman, R. G. 1988. "Internally-Consistent Thermodynamic Data for Minerals in the System Na_2O-K_2O-CaO-MgO-FeO-Fe_2O_3-Al_2O_3-SiO_2-TiO_2-H_2O-CO_2." *Journal of Petrology* 29, no. 2: 445–522. https://doi.org/10.1093/petrology/29.2.445.

Berman, R. G. 1991. "Thermobarometry Using Multi-Equilibrium Calculations; A New Technique, With Petrological Applications." *Canadian Mineralogist* 29, no. 4: 833–855.

Berman, R. G., and T. H. Brown. 1985. "Heat Capacity of Minerals in the System Na_2O-K_2O-CaO-MgO-FeO-Fe_2O_3-Al_2O_3-SiO_2-TiO_2-H_2O-CO_2: Representation, Estimation, and High Temperature Extrapolation." *Contributions to Mineralogy and Petrology* 89, no. 2: 168–183. https://doi.org/10.1007/BF00379451.

Beyssac, O., B. Goffé, C. Chopin, and J. N. Rouzaud. 2002. "Raman Spectra of Carbonaceous Material in Metasediments: A New Geothermometer." *Journal of Metamorphic Geology* 20, no. 9: 859–871. https://doi.org/10.1046/j.1525-1314.2002.00408.x.

Blackburn, W. H. 1968. "The Spatial Extent of Chemical Equilibrium in Some High-Grade Metamorphic Rocks From the Grenville of Southeastern Ontario." *Contributions to Mineralogy and Petrology* 19, no. 1: 72–92. https://doi.org/10.1007/BF00371730.

Blümel, P., and W. Schreyer. 1977. "Phase Relations in Pelitic and Psammitic Gneisses of the Sillimanite—Potash Feldspar and Cordierite-Potash Feldspar Zones in the Moldanubicum of the lam—Bodenmais Area, Bavaria." *Journal of Petrology* 18, no. 3: 431–459. https://doi.org/10.1093/petrology/18.3.431.

Carmichael, D. M. 1978. "Metamorphic Bathozones and Bathograds: A Measure of the Depth of Post-Metamorphic Uplift and Erosion on the Regional Scale." *American Journal of Science* 278: 769–797. https://doi.org/10.2475/ajs.278.6.769.

Cathelineau, M., and D. Nieva. 1985. "A Chlorite Solid Solution Geothermometer the Los Azufres (Mexico) Geothermal System." *Contributions to Mineralogy and Petrology* 91, no. 3: 235–244. https://doi.org/10.1007/BF00413350.

Cawood, I. P. 2024. "*Structural and Metamorphic Evolution of the Zanskar Himalaya, Suru Valley Region, NW India*." Doctoral thesis, University of Oxford. Oxford, United Kingdom.

Cesare, B., G. Cruciani, and U. Russo. 2003. "Hydrogen Deficiency in Ti-Rich Biotite From Anatectic Metapelites (El Joyazo, SE Spain): Crystal-Chemical Aspects and Implications for High-Temperature Petrogenesis." *American Mineralogist* 88, no. 4: 583–595. https://doi.org/10.2138/am-2003-0412.

Chicchi, L., L. Bindi, D. Fanelli, and S. Tommasini. 2023. "Frontiers of Thermobarometry: GAIA, a Novel Deep Learning-Based Tool for Volcano Plumbing Systems." *Earth and Planetary Science Letters* 620: 118352. https://doi.org/10.1016/j.epsl.2023.118352.

Chollet, F. 2021. *Deep Learning With Python*. 2nd ed. Manning Publications Co.

Cioffi, C. R., V. T. Meira, R. I. F. Trindade, P. Lanari, C. E. Ganade, and A. Gerdes. 2021. "Long-Lived Intracontinental Deformation Associated With High Geothermal Gradients in the Seridó Belt (Borborema Province, Brazil)." *Precambrian Research* 358: 106141. https://doi.org/10.1016/j.precamres.2021.106141.

Coggon, R., and T. J. B. Holland. 2002. "Mixing Properties of Phengitic Micas and Revised Garnet-Phengite Thermobarometers." *Journal of Metamorphic Geology* 20, no. 7: 683–696. https://doi.org/10.1046/j.1525-1314.2002.00395.x.

Copley, A., and O. Weller. 2022. "The Controls on the Thermal Evolution of Continental Mountain Ranges." *Journal of Metamorphic Geology* 40, no. 7: 1235–1270. https://doi.org/10.1111/jmg.12664.

Cottle, J. M., D. J. Waters, D. Riley, O. Beyssac, and M. J. Jessup. 2011. "Metamorphic History of the South Tibetan Detachment System, Mt. Everest Region, Revealed by RSCM Thermometry and Phase Equilibria Modelling." *Journal of Metamorphic Geology* 29, no. 5: 561–582. https://doi.org/10.1111/j.1525-1314.2011.00930.x.

Crowley, J. L., E. D. Ghent, and R. L. Brown. 1996. "Metamorphism in the Clachnacudainn Terrane and Implications for Tectonic Setting in the Southern Omineca Belt, Canadian Cordillera." *Canadian Journal of Earth Sciences* 33, no. 11: 1570–1582. https://doi.org/10.1139/e96-119.

de Capitani, C., and T. H. Brown. 1987. "The Computation of Chemical Equilibrium in Complex Systems Containing Non-Ideal Solutions." *Geochimica et Cosmochimica Acta* 51, no. 10: 2639–2652. https://doi.org/10.1016/0016-7037(87)90145-1.

Diwa, A. J. 2023. "Vergleiche von Metapelite in Einer Progradierenden Sequenz, an der Kontaktaureole am Nelson Batholit." In English: "A Comparison of Metapelites within the Prograding Sequence of the Contact Aureole of the Nelson Batholith." Bachelor's thesis, Universität Bern. Bern, Switzerland.

Dubacq, B., and J. B. Forshaw. 2024. "The Composition of Metapelitic Biotite, White Mica, and Chlorite: A Review With Implications for Solid-Solution Models." *European Journal of Mineralogy* 36, no. 4: 657–685. https://doi.org/10.5194/ejm-36-657-2024.

Duesterhoeft, E., and P. Lanari. 2020. "Iterative Thermodynamic Modelling—Part 1: A Theoretical Scoring Technique and a Computer Program (Bingo-Antidote)." *Journal of Metamorphic Geology* 38, no. 5: 527–551. https://doi.org/10.1111/jmg.12538.

Ellis, D. J., and D. H. Green. 1979. "An Experimental Study of the Effect of Ca Upon Garnet-Clinopyroxene Fe-Mg Exchange Equilibria." *Contributions to Mineralogy and Petrology* 71, no. 1: 13–22. https://doi.org/10.1007/BF00371878.

Engi, J. E. 1984. "Structure and Metamorphism North of Quesnel Lake and East of Niagara Creek, Cariboo Mountains, British Columbia." Master's thesis, University of British Columbia. Vancouver, Canada.

Ernst, W. G. 1971. "Do Mineral Parageneses Reflect Unusually High-Pressure Conditions of Franciscan Metamorphism?" *American Journal of Science* 270, no. 2: 81–108.

Essene, E. J. 1989. "The Current Status of Thermobarometry in Metamorphic Rocks." *Geological Society, London, Special Publications* 43, no. 1: 1–44. https://doi.org/10.1144/GSL.SP.1989.043.01.02.

Ferry, J. M., and F. S. Spear. 1978. "Experimental Calibration of the Partitioning of Fe and Mg Between Biotite and Garnet." *Contributions to Mineralogy and Petrology* 66, no. 2: 113–117. https://doi.org/10.1007/BF00372150.

Ferry, J. M., and E. B. Watson. 2007. "New Thermodynamic Models and Revised Calibrations for the Ti-in-Zircon and Zr-in-Rutile Thermometers." *Contributions to Mineralogy and Petrology* 154, no. 4: 429–437. https://doi.org/10.1007/s00410-007-0201-0.

Fitton, J., and R. Gill. 1970. "The Oxidation of Ferrous Iron in Rocks During Mechanical Grinding." *Geochimica et Cosmochimica Acta* 34, no. 4: 518–524. https://doi.org/10.1016/0016-7037(70)90143-2.

Forshaw, J. B., and D. R. M. Pattison. 2021. "Ferrous/Ferric (Fe^{2+}/Fe^{3+}) Partitioning Among Silicates in Metapelites." *Contributions to Mineralogy and Petrology* 176, no. 9: 63. https://doi.org/10.1007/s00410-021-01814-4.

Forshaw, J. B., and D. R. M. Pattison. 2023a. "Bulk Compositional Influence on Diverse Metapelitic Mineral Assemblages in the Whetstone Lake Area, Ontario." *Journal of Petrology* 64, no. 10: egad071. https://doi.org/10.1093/petrology/egad071.

Forshaw, J. B., and D. R. M. Pattison. 2023b. "Major-Element Geochemistry of Pelites." *Geology* 51, no. 1: 39–43. https://doi.org/10.1130/g50542.1.

Fyfe, W. S., F. J. Turner, and J. Verhoogen. 1958. "Metamorphic Reactions and Metamorphic Facies." In *Metamorphic Reactions and Metamorphic Facies*, vol. 73, 251. Geological Society of America. https://doi.org/10.1130/MEM73-p1.

Geen, A. C., and D. Canil. 2023. "Pattern and Source of Unusually High-Temperature Metamorphism in an Eocene Forearc Recorded by the Pacific Rim Terrane, British Columbia, Canada." *Journal of Metamorphic Geology* 41, no. 4: 583–602. https://doi.org/10.1111/jmg.12709.

Graham, C. M., and R. Powell. 1984. "A Garnet–Hornblende Geothermometer: Calibration, Testing, and Application to the Pelona Schist, Southern California." *Journal of Metamorphic Geology* 2, no. 1: 13–31. https://doi.org/10.1111/j.1525-1314.1984.tb00282.x.

Guidotti, C. V., J. T. Cheney, and S. Guggenheim. 1977. "Distribution of Titanium Between Coexisting Muscovite and Biotite in Pelitic Schists From Northwestern Maine." *American Mineralogist* 62, no. 5–6: 438–448.

Guidotti, C. V., and M. D. Dyar. 1991. "Ferric Iron in Metamorphic Biotite and Its Petrologic and Crystallochemical Implications." *American Mineralogist* 76, no. 1–2: 161–175.

Hartmeier, P., and P. Lanari. 2024. "Pytheriak." In (Version v1.1.1) Zenodo. https://doi.org/10.5281/zenodo.7945481.

Henry, D. J., and C. V. Guidotti. 2002. "Titanium in Biotite From Metapelitic Rocks: Temperature Effects, Crystal-Chemical Controls, and Petrologic Applications." *American Mineralogist* 87, no. 4: 375–382. https://doi.org/10.2138/am-2002-0401.

Henry, D. J., C. V. Guidotti, and J. A. Thomson. 2005. "The Ti-Saturation Surface for Low-to-Medium Pressure Metapelitic Biotites: Implications for Geothermometry and Ti-Substitution Mechanisms." *American Mineralogist* 90, no. 2–3: 316–328. https://doi.org/10.2138/am.2005.1498.

Hermann, J. 2003. "Experimental Evidence for Diamond-Facies Metamorphism in the Dora-Maira Massif." *Lithos* 70, no. 3–4: 163–182. https://doi.org/10.1016/s0024-4937(03)00097-5.

Hermann, J., and S. Lakey. 2021. "Water Transfer to the Deep Mantle Through Hydrous, Al-Rich Silicates in Subduction Zones." *Geology* 49, no. 8: 911–915. https://doi.org/10.1130/g48658.1.

Hertgen, S., P. Yamato, L. F. G. Morales, and S. Angiboust. 2017. "Evidence for Brittle Deformation Events at Eclogite-Facies P-T Conditions (Example of the Mt. Emilius Klippe, Western Alps)." *Tectonophysics* 706-707: 1–13. https://doi.org/10.1016/j.tecto.2017.03.028.

Hietanen, A. 1967. "On the Facies Series in Various Types of Metamorphism." *Journal of Geology* 75, no. 2: 187–214. https://doi.org/10.1086/627246.

Hietanen, A. 1969. "Distribution of Fe and Mg Between Garnet, Staurolite, and Biotite in Aluminum-Rich Schist in Various Metamorphic Zones North of the Idaho Batholith." *American Journal of Science* 267, no. 3: 422–456. https://doi.org/10.2475/ajs.267.3.422.

Hillebrand, W. F. 1908. "The Influence of Fine Grinding on the Water and Ferrous-Iron Content of Minerals and Rocks." *Journal of the American Chemical Society* 30, no. 7: 1120–1131. https://doi.org/10.1021/ja01949a010.

Hillenbrand, I. W., M. L. Williams, E. M. Peterman, M. J. Jercinovic, and C. W. Dietsch. 2023. "Petrochronologic Constraints on Inverted Metamorphism, Terrane Accretion, Thrust Stacking, and Ductile Flow in the Gneiss Dome Belt, Northern Appalachian Orogen." *Journal of Metamorphic Geology* 41, no. 9: 1197–1235. https://doi.org/10.1111/jmg.12741.

Holdaway, M. J. 2000. "Application of New Experimental and Garnet Margules Data to the Garnet-Biotite Geothermometer." *American Mineralogist* 85, no. 7–8: 881–892. https://doi.org/10.2138/am-2000-0701.

Holdaway, M. J. 2001. "Recalibration of the GASP Geobarometer in Light of Recent Garnet and Plagioclase Activity Models and Versions of the Garnet-Biotite Geothermometer." *American Mineralogist* 86, no. 10: 1117–1129. https://doi.org/10.2138/am-2001-1001.

Holland, T., and R. Powell. 2003. "Activity-Composition Relations for Phases in Petrological Calculations: An Asymmetric Multicomponent Formulation." *Contributions to Mineralogy and Petrology* 145, no. 4: 492–501. https://doi.org/10.1007/s00410-003-0464-z.

Holland, T. J. B., E. C. R. Green, and R. Powell. 2022. "A Thermodynamic Model for Feldspars in KAlSi3O8–NaAlSi3O8–CaAl2Si2O8 for Mineral Equilibrium Calculations." *Journal of Metamorphic Geology* 40, no. 4: 587–600. https://doi.org/10.1111/jmg.12639.

Holland, T. J. B., and R. Powell. 1998. "An Internally Consistent Thermodynamic Data Set for Phases of Petrological Interest." *Journal of Metamorphic Geology* 16, no. 3: 309–343. https://doi.org/10.1111/j.1525-1314.1998.00140.x.

Holland, T. J. B., and R. Powell. 2011. "An Improved and Extended Internally Consistent Thermodynamic Dataset for Phases of Petrological Interest, Involving a New Equation of State for Solids." *Journal of Metamorphic Geology* 29, no. 3: 333–383. https://doi.org/10.1111/j.1525-1314.2010.00923.x.

Hollister, L. S. 1966. "Garnet Zoning: An Interpretation Based on the Rayleigh Fractionation Model." *Science* 154, no. 3757: 1647–1651. https://doi.org/10.1126/science.154.3757.1647.

Hyppolito, T., C. Juliani, A. Garcia-Casco, V. Meira, A. Bustamante, and C. Hall. 2015. "LP/HTmetamorphism as a Temporal Marker of Change

of Deformation Style Within the Late Palaeozoic Accretionary Wedge of Central Chile." *Journal of Metamorphic Geology* 33, no. 9: 1003–1024. https://doi.org/10.1111/jmg.12166.

Jeanneret, P., I. Klonowska, C. Barnes, et al. 2022. "Deciphering the Tectonometamorphic History of Subducted Metapelites Using Quartz-in-Garnet and Ti-in-Quartz (QuiG–TiQ) Geothermobarometry—A Key for Understanding Burial in the Scandinavian Caledonides." *Journal of Metamorphic Geology* 41, no. 2: 235–270. https://doi.org/10.1111/jmg.12693.

Jorgenson, C., O. Higgins, M. Petrelli, F. Begue, and L. Caricchi. 2022. "A Machine Learning-Based Approach to Clinopyroxene Thermobarometry: Model Optimization and Distribution for Use in Earth Sciences." *J Geophys res Solid Earth* 127, no. 4: e2021JB022904. https://doi.org/10.1029/2021JB022904.

Kendall, M. G. 1938. "A New Measure of Rank Correlation." *Biometrika* 30, no. 1–2: 81–93.

Kingma, D. P., and J. Ba. 2014. "Adam: A Method for Stochastic Optimization." preprint, arXiv. arXiv:1412.6980.

Kohn, M. J. 2008. "P-T-T Data From Central Nepal Support Critical Taper and Repudiate Large-Scale Channel Flow of the Greater Himalayan Sequence." *Geological Society of America Bulletin* 120, no. 3–4: 259–273. https://doi.org/10.1130/b26252.1.

Kohn, M. J. 2014. ""Thermoba-Raman-Try": Calibration of Spectroscopic Barometers and Thermometers for Mineral Inclusions." *Earth and Planetary Science Letters* 388: 187–196. https://doi.org/10.1016/j.epsl.2013.11.054.

Koziol, A. M., and R. C. Newton. 1988. "Redetermination of the Anorthite Breakdown Reaction and Improvement of the Plagioclase-Garnet-Al2SiO5-Quartz Geobarometer." *American Mineralogist* 73, no. 3–4: 216–223.

Kwak, T. A. 1968. "Ti in Biotite and Muscovite as an Indication of Metamorphic Grade in Almandine Amphibolite Facies Rocks From Sudbury, Ontario." *Geochimica Et Cosmochimica Acta* 32, no. 11: 1222–1229. https://doi.org/10.1016/0016-7037(68)90124-5.

Lanari P. 2012. "Micro–Cartographie P–T–ε dans Les Roches Métamorphiques. Applications aux Alpes à l'Himalaya." Doctoral thesis, University of Grenoble. Grenoble, France.

Lanari, P., J. Dietrich, and J. Hermann. 2021. "Adding the Third Dimension to the Textural and Chemical Analysis of Metamorphic Rocks and Implications for Petrological Models Goldschmidt2021 Abstracts".

Lanari, P., and E. Duesterhoeft. 2019. "Modeling Metamorphic Rocks Using Equilibrium Thermodynamics and Internally Consistent Databases: Past Achievements, Problems and Perspectives." *Journal of Petrology* 60, no. 1: 19–56. https://doi.org/10.1093/petrology/egy105.

Lanari, P., and J. Hermann. 2021. "Iterative Thermodynamic Modelling—Part 2: Tracing Equilibrium Relationships Between Minerals in Metamorphic Rocks." *Journal of Metamorphic Geology* 39, no. 6: 651–674. https://doi.org/10.1111/jmg.12575.

Lanari, P., T. Wagner, and O. Vidal. 2014. "A Thermodynamic Model for Di-Trioctahedral Chlorite From Experimental and Natural Data in the System MgO–FeO–Al_2O_3–SiO_2–H_2O: Applications to P–T Sections and Geothermometry." *Contributions to Mineralogy and Petrology* 167, no. 2: 986. https://doi.org/10.1007/s00410-014-0968-8.

Li, X., and C. Zhang. 2022. "Machine Learning Thermobarometry for Biotite-Bearing Magmas." *Journal of Geophysical Research: Solid Earth* 127, no. 9: e2022JB024137. https://doi.org/10.1029/2022jb024137.

Liu, J., and S. R. Bohlen. 1995. "Mixing Properties and Stability of Jadeite-Acmite Pyroxene in the Presence of Albite and Quartz." *Contributions to Mineralogy and Petrology* 119, no. 4: 433–440. https://doi.org/10.1007/BF00286940.

López Ruiz, J., A. Aparicio, and L. García Cacho. 1978. "Chemical Variations in Biotites During Prograde Metamorphism, Sierra de Guadarrama, Sistema Central, Spain." *Chemical Geology* 21, no. 1: 113–129. https://doi.org/10.1016/0009-2541(78)90009-8.

Lyons, J. B., and S. A. Morse. 1970. "Mg/Fe Partitioning in Garnet and Biotite From Some Granitic, Pelitic, and Calcic Rocks." *American Mineralogist* 55, no. 1–2: 231–245.

Mackay-Champion, T., and I. Cawood. 2024. "LinaForma." In (Version v1.0). 10.5281/zenodo.11110440.

McKenzie, W. F. 2023. "The Significance of the Cretaceous Kluane Schist in the Mesozoic Assembly of the North American Cordillera Simon Fraser University." Burnaby, Canada.

Mengel, F., and T. Rivers. 1994. "Metamorphism of Pelitic Rocks in the Paleoproterozoic Ramah Group, Saglek Area, Northern Labrador; Mineral Reactions, P-T Conditions and Influence of Bulk Composition." *Canadian Mineralogist* 32, no. 4: 781–801.

Miyashiro, A. 1961. "Evolution of Metamorphic Belts." *Journal of Petrology* 2, no. 3: 277–311. https://doi.org/10.1093/petrology/2.3.277.

Osborne, Z. R., J. B. Thomas, W. O. Nachlas, R. J. Angel, C. M. Hoff, and E. B. Watson. 2022. "TitaniQ Revisited: Expanded and Improved Ti-In-Quartz Solubility Model for Thermobarometry." *Contributions to Mineralogy and Petrology* 177, no. 3: 31. https://doi.org/10.1007/s00410-022-01896-8.

Pattison, D. R. M. 1994. "Are Reversed Fe-Mg Exchange and Solid Solution Experiments Really Reversed?" *American Mineralogist* 79, no. 9–10: 938–950.

Pattison, D. R. M. 2001. "Instability of Al_2SiO_5 "Triple-Point" Assemblages in Muscovite+Biotite+Quartz-Bearing Metapelites, With Implications." *American Mineralogist* 86, no. 11–12: 1414–1422. https://doi.org/10.2138/am-2001-11-1210.

Pattison, D. R. M., and J. B. Forshaw. 2025. "Contact-Metamorphosed to Regionally-Metamorphosed Pelites: The Natural Record." *Journal of Petrology* 66, no. 7: egaf39. https://doi.org/10.1093/petrology/egaf039.

Pattison, D. R. M., and R. J. Tracy. 1991. "Phase Equilibria and Thermobarometry of Metapelites." *Reviews in Mineralogy and Geochemistry* 26, no. 1: 105–206.https://doi.org/10.1515/9781501509612-007.

Petrelli, M. 2024. "Machine Learning in Petrology: State-of-the-Art and Future Perspectives." *Journal of Petrology* 65, no. 5: egae36. https://doi.org/10.1093/petrology/egae036.

Petrelli, M., L. Caricchi, and D. Perugini. 2020. "Machine Learning Thermo-Barometry: Application to Clinopyroxene-Bearing Magmas." *Journal of Geophysical Research: Solid Earth* 125, no. 9: e2020JB020130. https://doi.org/10.1029/2020jb020130.

Piccoli, F., P. Lanari, J. Hermann, and T. Pettke. 2022. "Deep Subduction, Melting, and Fast Cooling of Metapelites From the Cima Lunga Unit, Central Alps." *Journal of Metamorphic Geology* 40, no. 1: 121–143. https://doi.org/10.1111/jmg.12621.

Powell, R., and T. Holland. 1994. "Optimal Geothermometry and Geobarometry." *American Mineralogist* 79, no. 1–2: 120–133.

Powell, R., and T. J. B. Holland. 2008. "On Thermobarometry." *Journal of Metamorphic Geology* 26, no. 2: 155–179. https://doi.org/10.1111/j.1525-1314.2007.00756.x.

Reay, A. 1981. "The Effect of Disc Mill Grinding on Some Rock-Forming Minerals." *Mineralogical Magazine* 44, no. 334: 179–182. https://doi.org/10.1180/minmag.1981.044.334.10.

Riel, N., S. Guillot, E. Jaillard, et al. 2013. "Metamorphic and Geochronogical Study of the Triassic El Oro Metamorphic Complex, Ecuador: Implications for High-Temperature Metamorphism in a Forearc Zone." *Lithos* 156-159: 41–68. https://doi.org/10.1016/j.lithos.2012.10.005.

Rivers, T. 1983. "Progressive Metamorphism of Pelitic and Quartzofeldspathic Rocks in the Grenville Province of Western Labrador - Tectonic Implications of Bathozone 6 Assemblages." *Canadian Journal of Earth Sciences* 20, no. 12: 1791–1804. https://doi.org/10.1139/e83-171.

Ronov, A. B. 1972. "Evolution of Rock Composition and Geochemical Processes in the Sedimentary Shell of the Earth." *Sedimentology* 19, no. 3–4: 157–172. https://doi.org/10.1111/j.1365-3091.1972.tb00019.x.

Saxena, S. K. 1969. "Silicate Solid Solutions and Geothermometry." *Contributions to Mineralogy and Petrology* 22, no. 4: 259–267. https://doi.org/10.1007/BF00400124.

Sen, S., and K. Chakraborty. 1968. "Magnesium-Iron Exchange Equilibrium in Garnet-Biotite and Metamorphic Grade." *Neues Jahrbuch Für Mineralogie - Abhandlungen* 108: 181–207.

Spear, F. S. 1993. *Metamorphic Phase Equilibria and Pressure-Temperature-Time Paths*. 2nd ed. Mineralogical Society of America.

Spear, F. S., and J. T. Cheney. 1989. "A Petrogenetic Grid for Pelitic Schists in the System SiO_2-Al_2O_3-FeO-MgO-K_2O-H_2O." *Contributions to Mineralogy and Petrology* 101, no. 2: 149–164. https://doi.org/10.1007/BF00375302.

Spear, F. S., D. R. M. Pattison, and J. T. Cheney. 2016. *The Metamorphosis of Metamorphic Petrology*, Vol. 523, 31–74. Geological Society of America Special Papers. https://doi.org/10.1130/2016.2523(02).

Tajčmanová, L., J. A. D. Connolly, and B. Cesare. 2009. "A Thermodynamic Model for Titanium and Ferric Iron Solution in Biotite." *Journal of Metamorphic Geology* 27, no. 2: 153–165. https://doi.org/10.1111/j.1525-1314.2009.00812.x.

Thompson, A. B. 1976. "Mineral Reactions in Pelitic Rocks: I. Prediction of P-T-X (Fe-Mg) Phase Relations." *American Journal of Science* 276, no. 4: 401–424. https://doi.org/10.2475/ajs.276.4.401.

Thompson, A. B., and P. C. England. 1984. "Pressure—Temperature—Time Paths of Regional Metamorphism II. Their Inference and Interpretation Using Mineral Assemblages in Metamorphic Rocks." *Journal of Petrology* 25, no. 4: 929–955. https://doi.org/10.1093/petrology/25.4.929.

Tischendorf, G., H. J. Förster, B. Gottesmann, and M. Rieder. 2018. "True and Brittle Micas: Composition and Solid-Solution Series." *Mineralogical Magazine* 71, no. 3: 285–320. https://doi.org/10.1180/minmag.2007.071.3.285.

Todd, C. S., and M. Engi. 1997. "Metamorphic Field Gradients in the Central Alps." *Journal of Metamorphic Geology* 15, no. 4: 513–530. https://doi.org/10.1111/j.1525-1314.1997.00038.x.

Tracy, R. J. 1982. "Compositional Zoning and Inclusions in Metamorphic Minerals." *Reviews in Mineralogy and Geochemistry* 10, no. 1: 355–397. https://doi.org/10.1515/9781501508172-013.

Warr, L. N. 2021. "IMA–CNMNC Approved Mineral Symbols." *Mineralogical Magazine* 85, no. 3: 291–320. https://doi.org/10.1180/mgm.2021.43.

Waters, D. J., and N. R. Charnley. 2002. "Local Equilibrium in Polymetamorphic Gneiss and the Titanium Substitution in Biotite." *American Mineralogist* 87, no. 4: 383–396. https://doi.org/10.2138/am-2002-0402.

Weber, G., and J. Blundy. 2024. "A Machine Learning-Based Thermobarometer for Magmatic Liquids." *Journal of Petrology* 65, no. 4: egae20. https://doi.org/10.1093/petrology/egae020.

Weiss, K., T. M. Khoshgoftaar, and D. Wang. 2016. "A Survey of Transfer Learning." *Journal of big Data* 3, no. 1: 9. https://doi.org/10.1186/s40537-016-0043-6.

White, R. W., N. E. Pomroy, and R. Powell. 2005. "An In Situ Metatexite–Diatexite Transition in Upper Amphibolite Facies Rocks From Broken Hill, Australia." *Journal of Metamorphic Geology* 23, no. 7: 579–602. https://doi.org/10.1111/j.1525-1314.2005.00597.x.

White, R. W., R. Powell, and G. L. Clarke. 2002. "The Interpretation of Reaction Textures in Fe-Rich Metapelitic Granulites of the Musgrave Block, Central Australia: Constraints From Mineral Equilibria Calculations in the System K_2O–FeO–MgO–Al_2O_3–SiO_2–H_2O–TiO_2–Fe_2O_3." *Journal of Metamorphic Geology* 20, no. 1: 41–55. https://doi.org/10.1046/j.0263-4929.2001.00349.x.

White, R. W., R. Powell, and T. J. B. Holland. 2007. "Progress Relating to Calculation of Partial Melting Equilibria for Metapelites." *Journal of Metamorphic Geology* 25, no. 5: 511–527. https://doi.org/10.1111/j.1525-1314.2007.00711.x.

White, R. W., R. Powell, T. J. B. Holland, T. E. Johnson, and E. C. R. Green. 2014a. "New Mineral Activity–Composition Relations for Thermodynamic Calculations in Metapelitic Systems." *Journal of Metamorphic Geology* 32, no. 3: 261–286. https://doi.org/10.1111/jmg.12071.

White, R. W., R. Powell, T. J. B. Holland, and B. A. Worley. 2000. "The Effect of TiO_2 and Fe_2O_3 on Metapelitic Assemblages at Greenschist and Amphibolite Facies Conditions: Mineral Equilibria Calculations in the System K_2O–FeO–MgO–Al_2O_3–SiO_2–H_2O–TiO_2–Fe_2O_3." *Journal of Metamorphic Geology* 18, no. 5: 497–511. https://doi.org/10.1046/j.1525-1314.2000.00269.x.

White, R. W., R. Powell, and T. E. Johnson. 2014b. "The Effect of Mn on Mineral Stability in Metapelites Revisited: New a–x Relations for Manganese-Bearing Minerals." *Journal of Metamorphic Geology* 32, no. 8: 809–828. https://doi.org/10.1111/jmg.12095.

Wu, C. M. 2015. "Revised Empirical Garnet–Biotite–Muscovite–Plagioclase Geobarometer in Metapelites." *Journal of Metamorphic Geology* 33, no. 2: 167–176. https://doi.org/10.1111/jmg.12115.

Wu, C. M. 2017. "Calibration of the Garnet–Biotite–Al_2SiO_5–Quartz Geobarometer for Metapelites." *Journal of Metamorphic Geology* 35, no. 9: 983–998. https://doi.org/10.1111/jmg.12264.

Wu, C.-M., and H.-X. Chen. 2015. "Revised Ti-in-Biotite Geothermometer for Ilmenite- or Rutile-Bearing Crustal Metapelites." *Science Bulletin* 60, no. 1: 116–121. https://doi.org/10.1007/s11434-014-0674-y.

Yardley, B. W. D., and J. S. Cleverley. 2013. "The Role of Metamorphic Fluids in the Formation of Ore Deposits." *Geological Society, London, Special Publications* 393, no. 1: 117–134. https://doi.org/10.1144/sp393.5.

Yogi, M. T. A. G., F. Gaidies, O. K. A. Heldwein, and A. H. N. Rice. 2024. "Mechanisms and Durations of Metamorphic Garnet Crystallization in the Lower Nappes of the Caledonian Kalak Nappe Complex, Arctic Norway." *Journal of Metamorphic Geology* 42, no. 5: 637–664. https://doi.org/10.1111/jmg.12766.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.