RocMLMs: Predicting Rock Properties through Machine Learning Models

Buchanan Kerswell [1] **Nestor Cerpa** [1] **Andréa Tommasi** [1] **Marguerite Godard** [1] **José Alberto Padrón-Navarta** [2]

[1]Geosciences Montpellier, University of Montpellier, CNRS, University of Antilles, Place Eugène Bataillon, 34095 Montpellier, France

[2]Instituto Andaluz de Ciencias de la Tierra (IACT), CSIC, Avda. de las Palmeras, 4, 18100 Armilla (Granada), Spain

Key Points:

- RocMLMs predict rock properties up to 10^1–10^3 faster than commonly used methods
- RocMLMs trained with Neural Networks are more efficient compared to other regression algorithms
- RocMLM training data show good agreement with PREM and STW105 for an average mantle geotherm

Corresponding author: Buchanan Kerswell, buchanan.kerswell@umontpellier.fr

Abstract

Mineral phase transformations significantly alter the bulk density and elastic properties of mantle rocks and consequently have profound effects on mantle dynamics and seismic wave propagation. These changes in the physical properties of mantle rocks result from evolution in the equilibrium mineralogical composition, which can be predicted by the minimization of the Gibbs Free Energy with respect to pressure (P), temperature (T), and chemical composition (X). Thus, numerical models that simulate mantle convection and/or probe the elastic structure of the Earth's mantle must account for varying mineralogical compositions to be self-consistent. Yet coupling Gibbs Free Energy minimization (GFEM) approaches with numerical geodynamic models is currently intractable for high-resolution simulations because execution speeds of widely-used GFEM programs (10^0–10^2 ms) are impractical in many cases. As an alternative, this study introduces machine learning models (RocMLMs) that have been trained to predict thermodynamically self-consistent rock properties at arbitrary PTX conditions between 1–28 GPa, 773–2273 K, and mantle compositions ranging from fertile (lherzolitic) to refractory (harzburgitic) end-members defined with a large dataset of published mantle compositions. RocMLMs are 10^1–10^3 times faster than GFEM calculations or GFEM-based look-up table approaches with equivalent accuracy. Depth profiles of RocMLMs predictions are nearly indistinguishable from reference models PREM and STW105, demonstrating good agreement between thermodynamic-based predictions of density, Vp, and Vs and geophysical observations. RocMLMs are therefore capable, for the first time, of emulating dynamic evolution of density, Vp, and Vs in high-resolution numerical geodynamic models.

Plain language summary

The mineralogical makeup of rocks within Earth's mantle largely determines how the mantle flows over geologic time, and how it responds to seismic waves triggered by earthquakes, because mineral assemblages control important rock properties such as density and stiffness (elasticity). The mineralogy of mantle rocks is not constant, however. It changes depending on three factors: pressure, temperature, and the chemical composition of the rock. Thus, it is important for computer simulations of mantle convection to account for the evolution of rock mineralogy. Computer programs that can predict rock properties based on thermodynamic calculations are available, but are generally too slow to be used in high-resolution simulations. As an alternative approach, this study intro-

duces machine learning models (RocMLMs) that have "learned" how to predict rock properties (density and elasticity) by "training" on a large dataset of thermodynamic calculations. We demonstrate that RocMLMs can then predict rock properties up to 10^1–10^3 times faster than state-of-the-art methods. We tested RocMLM predictions against reference mantle models based on observations of seismic waves and found good agreement. RocMLMs are therefore capable of fast and highly-accurate predictions of changes in rock properties and can be implemented in high-resolution computer simulations of mantle convection.

1 Introduction

The dominant mineral phases in Earth's mantle are olivine, pyroxene, garnet, wadsleyite, ringwoodite, bridgmanite, ferropericlase, calcium silicate perovskite, and $MgSiO_3$ post-perovskite (e.g., Stixrude and Lithgow-Bertelloni, 2012). Mantle mineralogy evolves with depth by a series of relatively discontinuous phase transformations that define sharp transitions in the physical properties of mantle rocks (Ringwood, 1991). The most important phase transformations occur at depths between 410 km and 670 km beneath Earth's surface, defining the transition from the upper to the lower mantle (Equation (1)). This mantle transition zone (MTZ) is characterized by sharp variations in density and elastic properties that strongly impact mantle convection (Christensen, 1995; Fukao et al., 2001; Jenkins et al., 2016; Karato et al., 2001; Kuritani et al., 2019; Nakagawa and Buffett, 2005; Ringwood, 1991; Schubert et al., 1975; Tackley et al., 1994; Wang et al., 2015), and the propagation of teleseismic waves (Dziewoński and Anderson, 1981; Ita and Stixrude, 1992; Ringwood, 1991). The MTZ is therefore an essential feature for modeling mantle structure and dynamics. With respect to a simple FeO-MgO-SiO_2 chemical system, the most important MTZ reactions can be written as:

$$\begin{aligned} &\text{olivine} \xrightarrow{410\text{ km}} \text{wadsleyite} \rightarrow \text{ringwoodite} \xrightarrow{670\text{ km}} \text{bridgmanite} + \text{ferropericlase} \\ &(\text{Mg,Fe})_2\text{SiO}_4 \rightarrow (\text{Mg,Fe})_2\text{SiO}_4 \rightarrow (\text{Mg,Fe})_2\text{SiO}_4 \rightarrow (\text{Mg,Fe})\text{SiO}_3 + (\text{Mg,Fe})\text{O} \end{aligned} \tag{1}$$

These phase changes (e.g., Equation (1)) are often parameterized in numerical geodynamic simulations with simple pressure-temperature (PT)-dependent reaction boundaries based on high-pressure experiments (e.g., Agrusta et al., 2017; Ballmer et al., 2015;

Christensen, 1995; Čížková and Bina, 2013; Kerswell et al., 2021; Liu et al., 1991; Nakagawa and Buffett, 2005; Tackley et al., 1994; Torii and Yoshioka, 2007). Alternatively, some numerical geodynamic experiments (e.g., Li et al., 2019; Yang and Faccenda, 2020) use Gibbs Free Energy minimization (GFEM) programs (e.g., Connolly, 2009; Riel et al., 2022) to precompute Lookup Tables of rock properties, which are subsequently referenced to adjust material properties as the numerical experiments evolve. These implementations usually consider fixed ideal mantle compositions, such as pyrolite, and/or approximate phase transitions with simple functions. These approaches neglect the PT dependency of mineral transitions on natural variations of mantle composition (X) such as variations of Fe-Mg and Al-Ca that may be either primordial or result from melt extraction or reactions during melt transport. Despite these simplifications, these models have corroborated that the MTZ is a critical feature impacting subduction dynamics, mantle plume dynamics, and water cycling in the deep Earth.

More self-consistent numerical models of mantle convection would track changes in physical properties of mantle rocks by computing GFEM as a function of the evolution of PTX conditions. However, this is currently intractable for high-resolution geodynamic models because GFEM programs remain too slow ($\geq$ 4–228 ms per PTX point) to be applied recursively during a geodynamic simulation (see Supporting Information). Parallelization of GFEM programs can increase efficiency by scaling the number of parallel processes (Riel et al., 2022), but continuously computing phase relations during geodynamic simulations would require GFEM efficiency on the order of $\leq 10^{0}$–10^{-1} ms to be feasible (see Supporting Information), which may be difficult to achieve solely by parallelisation and/or direct improvements to the current GFEM paradigm.

Here, we propose an alternative approach to predicting rock properties based on the use of machine learning models (referred to as RocMLMs) that have been "trained" on a multidimensional dataset of precomputed rock properties using classical (k-Neighbors, Decision Trees) and deep (Neural Network) regression algorithms. These later regression algorithms compress large amounts of thermodynamic information into highly efficient nonlinear functions, allowing RocMLMs to infer (predict) rock properties across arbitrary PTX conditions faster than any current GFEM algorithm. We demonstrate that RocMLMs are thus highly efficient emulators of GFEM programs and are well-suited for predicting bulk rock properties in numerical geodynamic models.

This article begins by detailing our method for building, training, and evaluating RocMLMs. We then demonstrate that RocMLMs can predict densities and seismic velocities in a dry upper mantle and transition zone up to 10^1–10^3 times faster than commonly used GFEM programs with equivalent accuracies. Finally, we compare RocMLM predictions with reference models derived from seismological datasets (Dziewoński and Anderson, 1981; Kustowski et al., 2008) and discuss the accuracy and performance of RocMLMs with respect to their future implementation in numerical geodynamic models.

2 Methods

The following sections describe the methodologies employed in constructing, training, and assessing RocMLMs, with a focus on four primary objectives. First, define the size and scope of RocMLM training data to ensure widespread applicability of RocMLMs to the upper mantle and transition zone (Section 2.1). Second, define a generalized approach for generating RocMLM training data to ensure applicability to any GFEM program (e.g., MAGEMin, Perple_X, and others, Section 2.2). Third, train RocMLMs on a set of input features that can be routinely computed during geodynamic simulations to ensure widespread applicability of RocMLMs to various geodynamic codes (Section 2.3). Fourth, rank the overall performance of RocMLMs in terms of accuracy and efficiency (Section 2.4).

2.1 RocMLM Training Dataset Design

2.1.1 Pressure-Temperature Conditions

High-pressure experiments constrain the reaction olivine → wadsleyite between 14.0 ± 1.0 GPa and 1600 ± 400 K with Clapeyron slopes between 2.4x10^{-3} ± 1.4x10^{-3} GPa/K (Akaogi et al., 1989; Katsura and Ito, 1989; Li et al., 2019; Morishima et al., 1994). Likewise, the reaction ringwoodite → bridgmanite + ferropericlase is constrained between 24.0 ± 1.5 GPa and 1600 ± 400 K with negative Clapeyron slopes between -2.0x10^{-3} ± 1.6x10^{-3} GPa/K (Akaogi et al., 2007; Bina and Helffrich, 1994; Hirose, 2002; Ishii et al., 2018; Ito, 1982; Ito et al., 1990; Ito and Katsura, 1989; Ito and Takahashi, 1989; Katsura et al., 2003; Litasov et al., 2005). We therefore compute RocMLM training data within a rectangular PT region bounded between 1–28 GPa and 773–2273 K to encompass ex-

pected conditions for the entire upper mantle and MTZ—from approximately 30 km to 865 km depth (Figure 1).

Figure 1 shows that our training dataset PT range includes PT conditions that are not expected to exist in neither the Earth's mantle, nor geodynamic simulations (e.g., very cold conditions with thermal gradients $\leq$ 5 K/km, Cerpa et al., 2022; Maruyama et al., 1996; Syracuse et al., 2010). Such a large rectangular PT range might be considered impractical with respect to training efficiency (unnecessary amounts of training data) and accuracy (outside the bounds of calibrated thermodynamic data) compared to an irregular PT range bounded between arbitrary geotherms. However, initial sensitivity tests showed comparable RocMLM performance irrespective of the range of PT conditions used to generate RocMLM training data. Thus, we adopted a regular rectangular training dataset design because it is computationally convenient and does not deteriorate RocMLM accuracy.

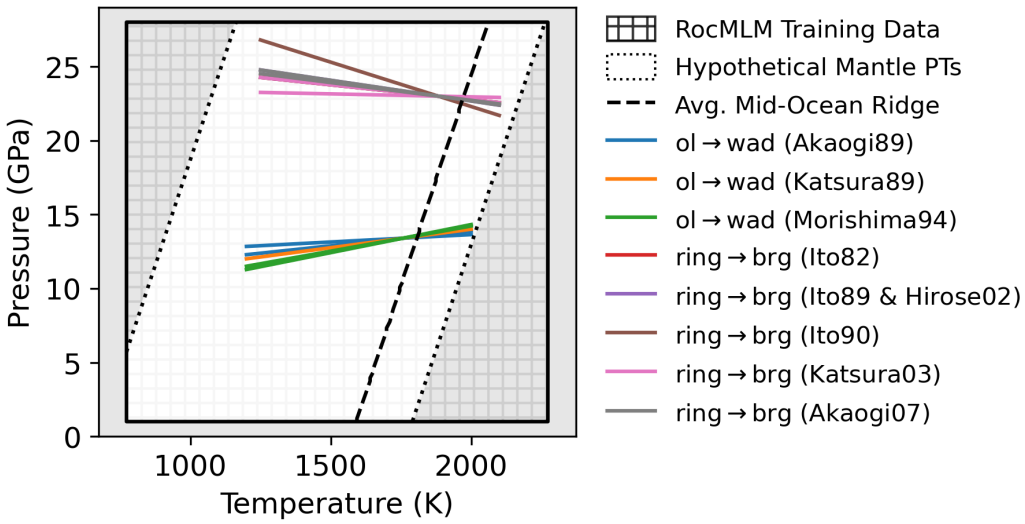

Figure 1: PT diagram showing the range of conditions considered for generating RocMLM training data (hatched region) compared to a range of possible upper mantle conditions (inner white region). The dotted black lines are geotherms with arbitrary mantle potential temperatures of 673 K and 1773 K and a constant adiabatic gradient of 0.5 K/km, representing hypothetical lower and upper bounds for mantle PT conditions (including hypothetical cold lithospheric slabs). The dashed black line is an average geotherm for a mid-ocean ridge (1573 K adiabat). Phase boundaries for the 410 km and 670 km discontinuities (colored lines) are from a compilation by Li et al. (2019).

2.1.2 Bulk Mantle Compositions

We derived an array of synthetic bulk mantle compositions with the aim of encompassing the widest range of chemical variability in Earth's mantle. For this, we applied a statistical analysis to publicly-available geochemical data from thousands of natural peridotite samples. The procedure was as follows.

Bulk chemical analyses of peridotite samples were downloaded using the Earthchem.org Search Portal with a single search criterion: *"set sample type > igneous rocks > names from Earthchem categories > igneous-plutonic-ultramafic"*. The search queried 19791 samples with rock type classifications that we did not modify from their original labels. Samples lacking analyses for SiO_2, MgO, Al_2O_3, or CaO were excluded from the dataset. All samples classified as "unknown", chromitite, limburgite, wehrlite, undifferentiated peridotite, dunite, or pyroxenite were also excluded from the dataset to focus on samples that are most likely mantellic, that is, residues of partial melting modified (or not) by refertilization, rather than products of fractional crystallization (Bowen, 1915). The data were grouped according to the remaining rock types (lherzolite and harzburgite) and outliers were removed from each group using a 1.5 interquartile range threshold applied to each chemical component. Cr and Ni measured as minor elements (ppm) were converted to Cr_2O_3 and NiO (wt.%) and all Fe oxides were converted to Fe_2O_3T. Total oxides were then checked against H_2O, CO_2, and LOI to determine if chemical analyses were performed before or after ignition. Analyses with total oxides summing to $\leq 97\%$ or $\geq 103\%$ were considered erroneous, or otherwise low-quality, and excluded from the dataset. All analyses were then normalized to a volatile-free basis before converting Fe_2O_3T to FeOT. After normalization, the final compositional space investigated includes the components Na_2O-CaO-FeO-MgO-Al_2O_3-SiO_2-TiO_2 (NCFMAST system). The final dataset contains 3111 chemical analyses of classified peridotite samples (Table 1).

We applied Principal Component Analysis (PCA) to the standardized peridotite dataset to reduce its dimensionality from the original 7-oxides space. PCA requires complete data, so samples were first arranged by decreasing MgO and increasing SiO_2 content and a k-Neighbors algorithm was applied to impute missing oxide analyses, which were mainly the Na_2O component (see Table 1 for missing analyses counts). Following common practice, a "z-score normalization" was applied to all oxide components before running PCA. The first two principal components (PC1 and PC2) explain 78% of the

variance of the dataset, which we considered to be sufficient for modeling a broad range of peridotitic mantle compositions. PC1 separates samples by their TiO_2, Al_2O_3, MgO, CaO, and Na_2O contents, while PC2 separates samples by SiO_2 and FeO (Figure 2).

In this PC space, we drew a mixing line connecting the lherzolite and harzburgite group centroids (i.e., the median values for PC1 and PC2 for each group). The lherzolite-harzburgite mixing line was then extended until reaching the approximate location of the most fertile (Al_2O_3-CaO-TiO_2-rich) and most refractory (MgO-rich, SiO_2-poor) peridotite samples, hereafter referred to as Primitive Synthetic Upper Mantle (PSUM) and Depleted Synthetic Upper Mantle (DSUM, Figure 2b), respectively. The mixing line approximates the widest array of mantle compositions derived from the natural rock record and may be interpreted as representing the first order composition variation in response to melt extraction (depletion) or addition (refertilization) in the mantle. The mixing line therefore provides a basis for sampling synthetic bulk mantle compositions directly from PC space, which were then used to generate RocMLM training data.

Table 1: Summary of the filtered and standardized peridotite dataset from Earthchem.org. Columns with an asterisk are in wt.%. Std = standard deviation, IQR = interquartile range.

Oxide	Measured	Missing	Min*	Max*	Mean*	Median*	Std*	IQR*
SiO_2	3111	0	36.7	52	44.1	44.1	1.16	1.24
TiO_2	2835	276	0	0.268	0.051	0.03	0.05	0.068
Al_2O_3	3111	0	0.023	4.95	1.65	1.31	1.14	1.82
FeOT	3111	0	5.98	15.3	8.05	8.01	0.675	0.569
MgO	3111	0	31.8	50.8	43	43.6	2.96	4.38
CaO	3111	0	0.01	5.2	1.46	1.17	1.04	1.66
Na_2O	2008	1103	0	0.525	0.127	0.098	0.11	0.171

2.1.3 Reducing Bulk Mantle Compositions to a Single Fertility Index Value

Training RocMLMs with either 7 oxide components or two PCs as inputs is possible. However, our targeted application (e.g., implementing RocMLMs in geodynamic codes) discourages the use of the two options because in either case it would require track-

ing the oxides in numerical geodynamic codes, which is currently impractical. Thus, we aimed to reduce the dimensionality of the training dataset from nine dimensions (7 oxide components + PT) to three dimensions (1 compositional dimension + PT) by estimating the amount of melt extraction (depletion) that might have produced the synthetic bulk mantle compositions in the training dataset. Assuming that all synthetic samples were derived from a PSUM source, we adopt a simple modal fractional melting model (after Shaw, 1970):

$$\frac{C^s_{\mathrm{TiO_2}}}{C^0_{\mathrm{TiO_2}}} = R = (1-F)^{\frac{1}{D_0}-1} \tag{2}$$

where R is the ratio of the TiO_2 concentration of the sample to the initial PSUM source (Table 2), F is the melt fraction, and $D_0 = 0.05$ is the bulk distribution coefficient for TiO_2 in peridotite (after Brown and Lesher, 2016). Note that unlike the dataset of natural peridotite samples, synthetic samples were drawn directly from PC space and their TiO_2 concentrations (and other oxide components) change monotonically with PC1 from the initial PSUM source (Figure 2b,c). Synthetic samples therefore represent a smooth and idealized variability from fertile (PSUM) to depleted (DSUM) mantle compositions that captures the average variation in natural peridotite samples.

A Fertility Index (ξ) is calculated by rearranging Equation (2) for F and subtracting F from 1:

$$\xi = 1 - F = R^{\frac{1}{(\frac{1}{D_0})-1}} \tag{3}$$

Training RocMLMs on ξ instead of seven oxide components is beneficial for two reasons: 1) it greatly increases RocMLM efficiency and 2) unlike oxide components or PCs, melt fraction is routinely implemented in numerical geodynamic simulations (e.g., Cerpa et al., 2019; Gerya and Yuen, 2003; Kelley et al., 2010; Li et al., 2019; Sizova et al., 2010; Yang and Faccenda, 2020). Likewise, tracking the depletion/fertility of the mantle in geodynamics models with Lagrangian tracers and/or compositional fields is more conceivable (Agrusta et al., 2015; Cagnioncle et al., 2007; Gerya and Meilick, 2011; Tackley and Xie, 2003). Although we chose ξ for RocMLM training, ξ and F represent opposite reference frames for the same time-integrated melting process, and are therefore

interchangeable. This approach offers a generalized solution for coupling RocMLMs to geodynamic codes.

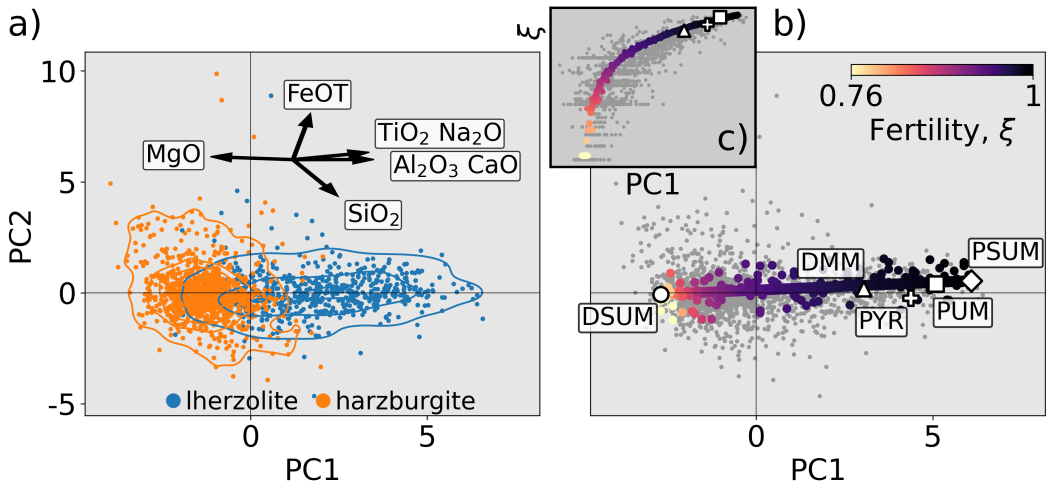

Figure 2: PC1-PC2 diagrams showing the standardized geochemical dataset of natural peridotite samples (a) and a mixing array between hypothetical end-member mantle compositions Primitive Synthetic Upper Mantle (PSUM) and Depleted Synthetic Upper Mantle (DSUM, b). Black arrows in (a) indicate PCA loading vectors. Colored data points in (b) are the synthetic mantle compositions used to train RocMLMs, which were sampled independently from the natural peridotite samples (gray data points). The inset (c) shows how the Fertility Index (ξ) changes nonlinearly with PC1. DMM, PUM, and PYR are from Table 2.

The melting model in Equation (2) is oversimplified since it assumes: 1) melt is instantaneously removed from the source region, 2) D_0 is constant, and 3) minerals melt in the same proportions that they exist in the source rock. It nevertheless provides an efficient parameterization of the variation in mantle composition as a function of melt extraction and addition. Equation (2) predicts that a Depleted MORB Mantle (DMM) composition is produced through a time-integrated 2.2% melt extraction from a Primitive Upper Mantle (PUM) source (Table 2). This result is consistent with the degree of depletion inferred from trace element patterns and mass balance constraints (2-3% melt removal from PUM, Workman and Hart, 2005). We therefore consider ξ an adequate first-order proxy for describing the variations in bulk mantle composition used in our RocMLM training dataset. However, given that TiO_2 concentrations are strongly affected by reactive melt transport (e.g., Le Roux et al., 2007), ξ may only be estimated for the average compositional trend as expressed in PC1-PC2 space, rather than on individual peridotite samples.

Table 2: Hypothetical upper mantle end-member compositions. Columns with an asterisk are in wt.%. Depleted MORB Mantle (DMM) is from Workman and Hart (2005), Primitive Upper Mantle (PUM) is from Sun and McDonough (1989), and Pyrolite (PYR) is from Green (1979). Primitive Synthetic Upper Mantle (PSUM) and Depleted Synthetic Upper Mantle (DSUM), are end-member compositions derived in this study.

Sample	SiO_2^*	TiO_2^*	$Al_2O_3^*$	$FeOT^*$	MgO^*	CaO^*	Na_2O^*	ξ
DSUM	44.1	0.0012	0.261	7.96	47.4	0.22	0.042	0.764
DMM	44.7	0.13	3.98	8.18	38.7	3.17	0.13	0.974
PYR	45	0.16	4.4	7.6	38.8	3.4	0.34	0.984
PUM	44.9	0.2	4.44	8.03	37.7	3.54	0.36	0.996
PSUM	46.2	0.216	4.88	8.88	35.2	4.34	0.33	1

2.2 Generating RocMLM Training Data

We used the GFEM program Perple_X (version 7.0.9, Connolly, 2009) to generate RocMLM training data across PT conditions as described in Section 2.1.1 and synthetic bulk mantle compositions as described in Section 2.1.2. The Perple_X calculations were constrained to the Na_2O-CaO-FeO-MgO-Al_2O_3-SiO_2 (NCFMAS) chemical system to comply with the thermodynamic data and solution models of Stixrude and Lithgow-Bertelloni (2022). The Stixrude and Lithgow-Bertelloni (2022) dataset (stx21ver.dat) was used because our initial tests with alternative thermodynamic datasets (hp02ver.dat and hp633ver.dat, Connolly and Kerrick, 2002; Holland et al., 2018; Holland and Powell, 2001) failed to reproduce the seismic wave velocities of geophysical reference models (PREM and STW105, Dziewoński and Anderson, 1981; Kustowski et al., 2008) with sufficient accuracy because these datasets lack a parametrization of the shear modulii of the minerals phases. Note that our Perple_X calculations ignored TiO_2, which was initially included to define ξ and derive synthetic bulk mantle compositions. Despite being measured as a major oxide component, the average TiO_2 content of our standardized samples is 0.05 ± 0.1 wt.% (2σ, Table 1). Such small concentrations of TiO_2 may safely be ignored in phase relation calculations with negligible effects on the RocMLM training dataset.

The Perple_X models used to generate the present RocMLM training database included equations of state for solution phases: olivine, plagioclase, spinel, clinopyroxene,

wadsleyite, ringwoodite, perovskite, ferropericlase, high-pressure C2/c pyroxene, orthopyroxene, akimotoite, post-perovskite, Ca-ferrite, garnet, and Na-Al phase. Melt was not considered due to the absence of melt models in the Stixrude and Lithgow-Bertelloni (2022) dataset, but may be considered in future versions of training datasets if the elastic parameters in hp02ver.dat are corrected. Once configured, Perple_X generated RocMLM training data (density, as well as P- and S-wave seismic velocities) by minimizing the total Gibbs Free Energy of a multicomponent multiphase thermodynamic system at fixed PTX conditions (Gibbs, 1878; Spear, 1993). The reader is referred to Connolly (2009) and Riel et al. (2022) for a complete description of the GFEM problem.

In principle, applying identical sets of solution phase models, thermodynamic data, and bulk compositions will define identical Gibbs Free Energy hyperplanes. This implies that any GFEM algorithm should converge on identical phase relations. Thus, although this study uses Perple_X exclusively, an identical set of training data can be generated by applying the procedures outlined above to other GFEM programs. Note that RocMLM capabilities and performance are primarily dependent on the size and the range of PTX conditions of the training dataset, not on the choice of GFEM algorithm.

2.3 Training RocMLMs

RocMLM training data were preprocessed using the following procedure. First, two-dimensional grids of rock properties ("pseudosections") calculated by Perple_X were stacked into a three-dimensional array, $Z = (z_{1,1,1}, \ldots, z_{n,w,w})$, where $w = 128$ is the resolution of the PT grid and $n = 128$ is the number of random synthetic bulk mantle compositions represented by a ξ value. Z was flattened into arrays of training features (PT and ξ), $X = (x_{1,1,1}, \ldots, x_{v,v,v})$, and training targets (density, Vp, and Vs), $y = (y_{1,1,1}, \ldots, y_{v,v,v})$, where $v = n \cdot w^2 = 128^3$ is the total number of training examples. Following common practice, X and y were scaled using "z-score normalization" before training.

The preprocessed training data were then fit with three different nonlinear regression algorithms (Decision Tree: DT, k-Neighbors: KN, and Neural Networks: NN) from the scikit-learn python library (Pedregosa et al., 2011). Each regression algorithm was tuned with a grid search approach, where a performance score (RMSE) was evaluated over all hyperparameter combinations relevant to the particular regression algorithm (Ta-

ble 3). The set of hyperparameters that produced the best score (lowest RMSE) was used to train the RocMLM.

Table 3: RocMLM configuration. Hyperparameter values in parentheses are tested sequentially by a cross-validation grid search algorithm and the best set of hyperparameters is chosen by the lowest RMSE. Hyperparameters that are not shown use default values (see regression model documentation on scikit-learn.org).

Model	Hyperparameter	Value	Tuned
DT	splitter	(best, random)	tuned
	max features	(1, 2, 3)	tuned
	min samples leaf	(1, 2, 3)	tuned
	min samples split	(2, 4, 6)	tuned
KN	n neighbors	(2, 4, 8)	tuned
	weights	(uniform, distance)	tuned
NN1	hidden layer sizes	(8, 16, 32)	tuned
NN2	hidden layer sizes	([16, 16], [32, 16], [32, 32])	tuned
NN3	hidden layer sizes	([32, 16, 16], [32, 32, 16], [32, 32, 32])	tuned
NN(all)	learning rate	(0.001, 0.005, 0.001)	tuned
	batch size	20%	fixed
	max epochs	100	fixed

2.4 Evaluating RocMLM Accuracy and Performance

Connolly and Khan (2016) estimated the uncertainties of Vp and Vs to be on the order of 3–5% within the same thermodynamic framework used to generate RocMLM training data (Stixrude and Lithgow-Bertelloni, 2005). We can therefore consider the base-uncertainty of RocMLM predictions to be 3–5%. RocMLM predictions must also account for additional uncertainties that are introduced during RocMLM training (i.e., the variance of residuals between RocMLM predictions and targets), which are about 2% for NN1 and $< 1\%$ for DT, KN, and NN3. Assuming the lowest-uncertainty models (DT, KN, NN3) would be preferred for geodynamic applications, we ignore the small variances introduced during training ($< 1\%$) and evaluate the total RocMLM prediction uncertain-

ties to be on the same order as the base GFEM uncertainty (3–5%) after Connolly and Khan (2016).

RocMLM accuracy (in terms of RMSE) was evaluated by: 1) testing RocMLMs on a separate validation dataset to determine the generalization capacity of RocMLMs to unseen mantle conditions (internal accuracy), and 2) comparing RocMLMs predictions with geophysical reference models PREM and STW105 (external accuracy). The first test evaluates the degree to which RocMLMs can reproduce GFEM predictions. The second test evaluates the degree to which the "true data" used for RocMLM training reproduces the phase transitions actually observed in Earth's upper mantle, which depend on the thermodynamic data, GFEM algorithm, and parameterization used to describe the composition of mantle rocks (i.e., ξ).

The validation dataset was generated by Perple_X in the same manner as the training dataset, but shifted by one-half step (in the positive PT directions) so that RocMLM predictions could be evaluated at completely independent PT conditions. RocMLM performance was evaluated by: 1) measuring single-point prediction times (execution speed), and 2) scaling execution speed by RocMLM file size (disk space) to account for information compression (model efficiency).

The number of PT points and synthetic bulk mantle compositions used for generating training data were varied from 8 to 128 (2^{11}–2^{21} total training examples) to test the sensitivity of RocMLM accuracy and performance with respect to the size ("capacity") and composition of the training dataset. The same sets of training data were also used to evaluate single-point execution speed using a common Lookup Table approach, where a cubic spline interpolation was applied to the training dataset and rock properties were evaluated at arbitrary PTX conditions. Prediction accuracy and performance were measured in a consistent manner so that direct comparisons could be made between RocMLMs, Lookup Tables, and GFEM programs.

3 Results

3.1 RocMLM Accuracy

The following examples of Decision Tree (DT, Figure 3), single-layer Neural Network (NN1, Figure 4), and three-layer Neural Network (NN3, Figure 5) models demon-

strate how different regression algorithms ultimately influence the accuracy of RocMLM predictions (see Supplementary Information for all regression algorithms).

DT predictions are practically indistinguishable from that of Perple_X, indicating a nearly-perfect mapping of the validation dataset by the DT algorithm (RMSE for density: 0.01 g/cm^3, Vp and Vs: 0.02 km/s, Figure 3). Absolute differences between Perple_X and DT predictions (residuals) are broadly dispersed and approach zero in most regions of PT space. Some concentrations of residuals exist near phase transitions, but are subtle and discontinuous (Figure 3g–i).

In contrast, NN1 predictions are notably smoother than Perple_X (Figure 4), with higher errors (RMSE for density: 0.02 g/cm^3, Vp: 0.06 km/s, Vs: 0.05 km/s) that indicate an inability to resolve sharp gradients in physical properties when using a single-layer Neural Network with a small to moderate amount of neurons. This is evident by the NN1 residuals, which are systematically concentrated near phase transitions (Figure 4g–i). NN1 profiles display relatively weak discontinuities with gradual changes in physical properties across the olivine $\rightarrow$ wadsleyite and ringwoodite $\rightarrow$ bridgmanite + ferropericlase transitions (Figure 4j–l), and phase transformations within the MTZ are virtually absent compared to DT and NN3 profiles. While NN1 predictions do not reproduce the validation dataset or geophysical profiles with the highest accuracy, deeper (and/or wider) NN architectures with more hidden-layers (e.g., NN3) are more capable (Figure 5). NN3 predictions fit the validation dataset and resolve discontinuities in geophysical profiles with nearly equivalent accuracy as DT and KN algorithms (compare profiles in Supplementary Information).

Comparing density, Vp, and Vs depth profiles predicted by RocMLMs (for an average mid-ocean ridge-like geotherm with a mantle potential temperature of 1573 K) with PREM and STW105 reveals relatively low errors (density: $\leq$ 0.08 g/cm^3, Vp: $\leq$ 0.26 km/s, Vs: $\leq$ 0.14 km/s) and high correlations ($R^2 \geq 0.94$) that indicate good agreement between seismically-derived profiles and thermodynamic predictions, irrespective of regression algorithm (compare profiles in the Supplementary Information). The largest deviations between RocMLM profiles, PREM, and STW105 fall within two regions: 1) between 1–8 GPa, and 2) at the base of the MTZ (Figures 3–5j–l). At pressures lower than 5 GPa, the divergence between RocMLM profiles and seismically-derived profiles may be explained by the low resolution of the 1D geophysical profiles relative to the extreme

spatial variability in composition and geotherms on Earth. Tests using an average continental geotherm to calculate RocMLM profiles results in less divergence between RocMLM profiles and PREM at < 5 GPa compared to the mid-ocean ridge-like geotherms used to build the profiles presented in Figures 3–5. At pressures between 5–8 GPa, the two geophysical models show a discrepancy: PREM contains a discontinuity, especially in seismic velocities, while STW105 has a gradual and continuous increase. RocMLM profiles between 5–8 GPa are more similar to STW105, which does not map any discontinuities until the olivine → wadsleyite transition at 410 km depth (Figures 3–5j–l).

Within the MTZ, DT and NN3 profiles predict intermediate discontinuities, while PREM and STW105 are gradual and continuous (Figures 3,5g–i). As expected, comparing RocMLM profiles for different geotherms shows that the choice of a mantle potential temperature leads to contrasting predictions of: 1) the overall evolution of rock properties with depth, and 2) the depths, magnitudes, and sharpness of phase transitions within the MTZ (Figures 3–5g–i). RocMLM profiles show, similarly to those directly derived from the Perple_X calculation, temperature-sensitive discontinuities at the olivine → wadsleyite and wadsleyite → ringwoodite transitions, but a rather temperature insensitive ringwoodite → bridgmanite + ferropericlase transition (Figures 3–5g–i). This can be explained by differences in Clapeyron slopes modeled by the Stixrude and Lithgow-Bertelloni (2022) dataset.

3.2 RocMLM Performance

We now compare RocMLM performance to two other tools classically used to predict the variations of physical properties of mantle rocks in geodynamic models: GFEM programs and Lookup Tables. Note that RocMLM, GFEM, and Lookup Table performance is platform specific. Running analogous implementations with other programming languages and/or on alternative computer hardware will differ from the results presented here. All computations in this study were made using CPUs of a Macbook Pro (2022; M2 chip) with macOS 13.4 and using Python 3.11.4. All performance metrics were evaluated with a single CPU core.

Figure 6 shows how execution speed, efficiency, and accuracy scale with the capacity of Lookup Tables and RocMLMs. Here, "capacity" refers to the number of scalar values stored by Lookup Tables, or alternatively, the number of pseudosection PTX points

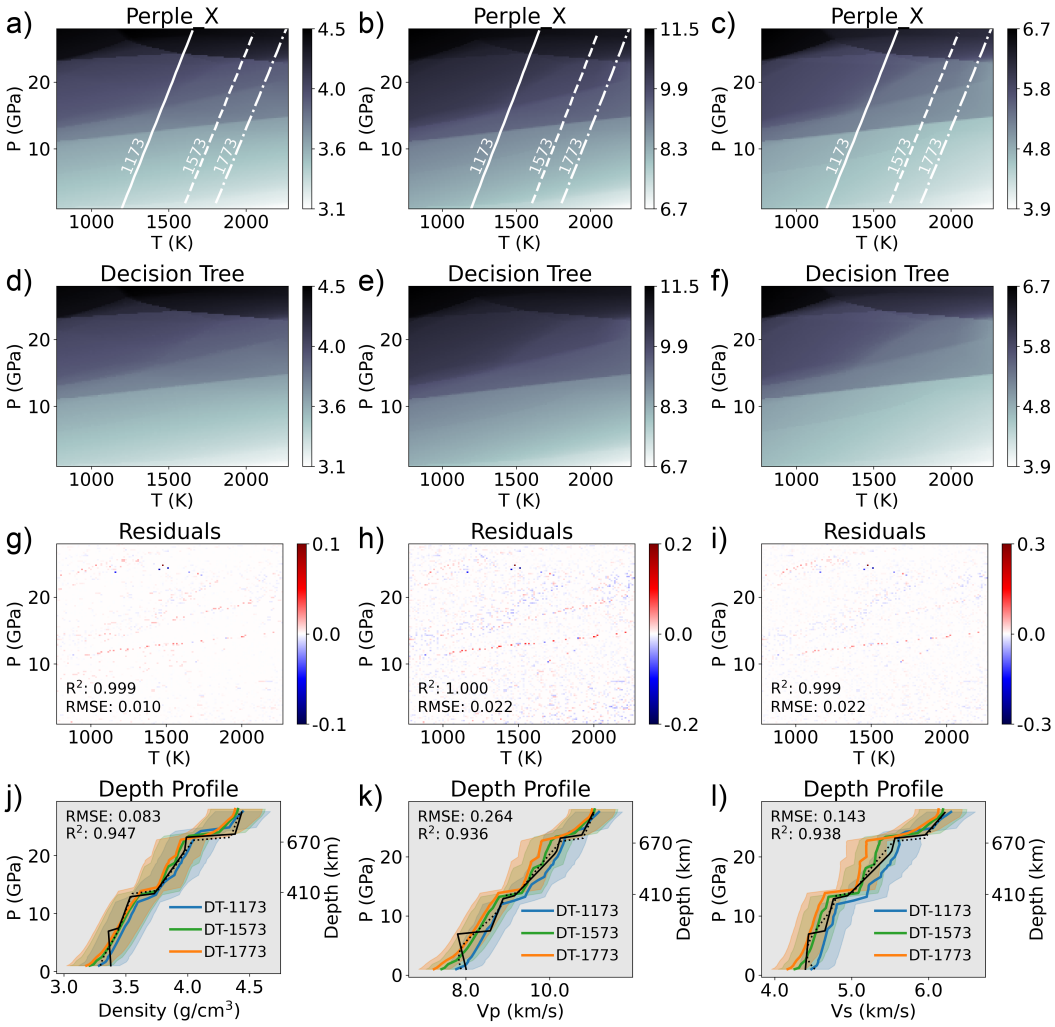

Figure 3: PT diagrams showing density (left column, g/cm^3), Vp (middle column, km/s), and Vs (right column, km/s) predictions from a Perple_X model with a PUM bulk composition (a–c), a Decision Tree RocMLM (d–f), and absolute differences between Perple_X and DT (g–i) measured on the validation dataset. Depth profiles (j–l) compare Perple_X and DT predictions extracted along a 0.5 K/km adiabat with different mantle potential temperatures (white lines) with reference models PREM (solid black line, Dziewoński and Anderson, 1981) and STW105 (dotted black line, Kustowski et al., 2008). The RMSE in (j–l) indicates the measured differences between DT-1573 and PREM. Colored ribbons indicate 5% uncertainty in RocMLM predictions.

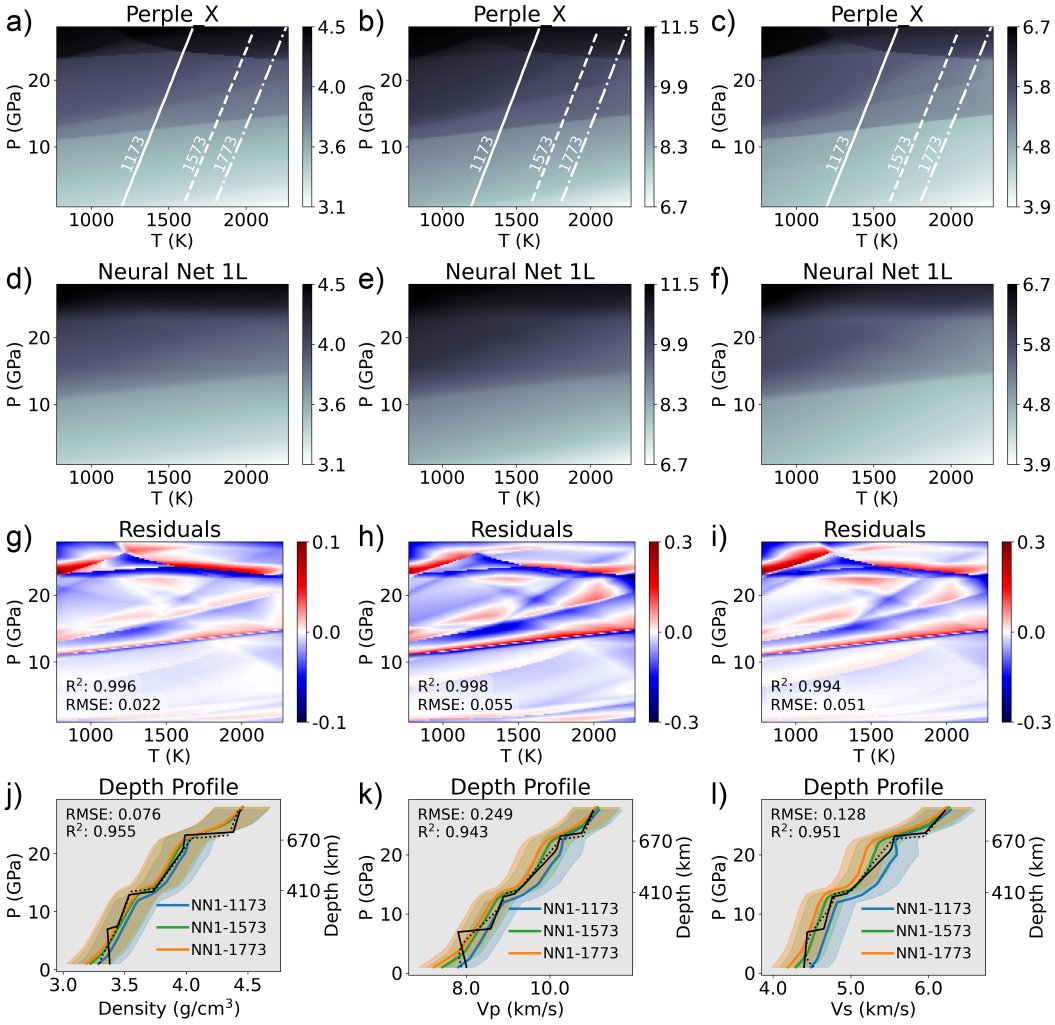

Figure 4: PT diagrams showing density (left column, g/cm^3), Vp (middle column, km/s), and Vs (right column, km/s) predictions from a Perple_X model with a PUM bulk composition (a–c), a single-layer Neural Network RocMLM (d–f), and absolute differences between Perple_X and NN1 (g–i) measured on the validation dataset. Other legend details are the same as in Figure 3.

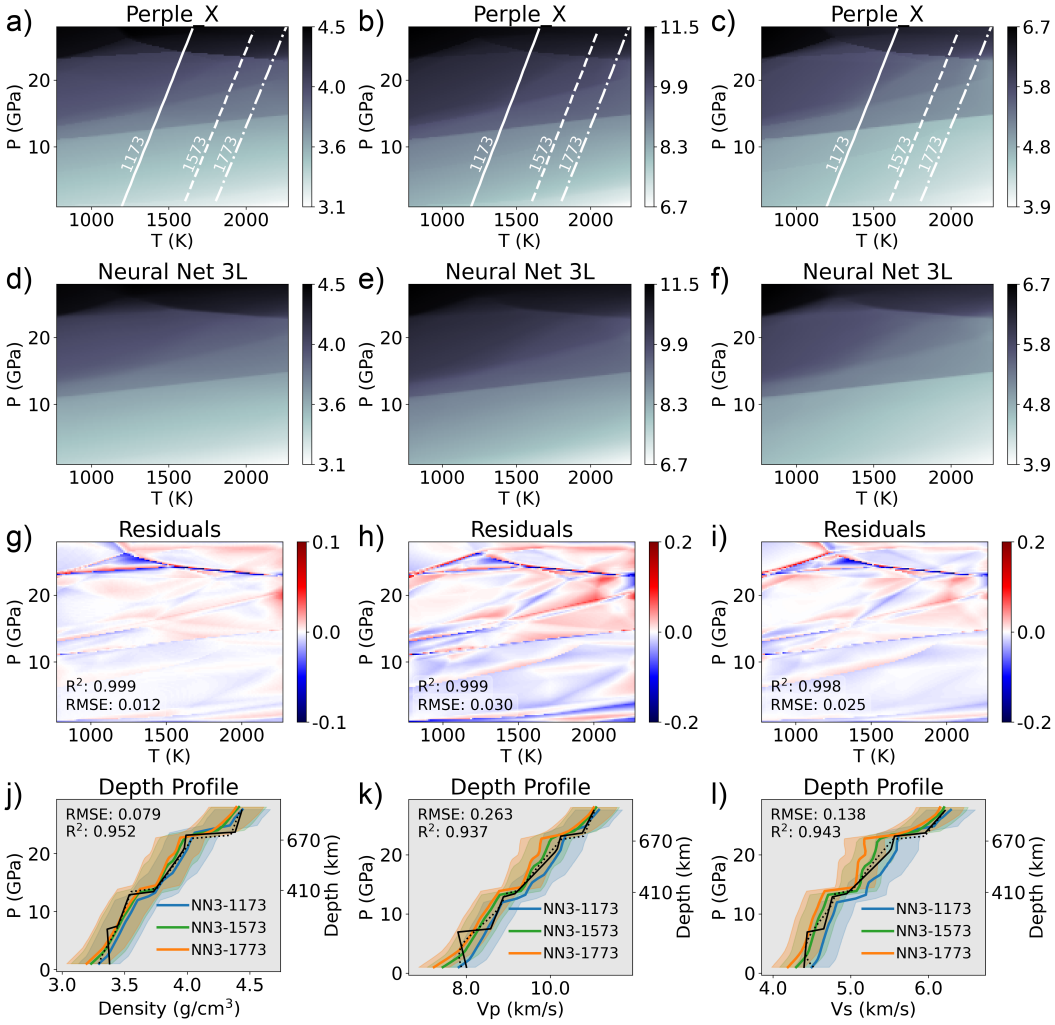

Figure 5: PT diagrams showing density (left column, g/cm^3), Vp (middle column, km/s), and Vs (right column, km/s) predictions from a Perple_X model with a PUM bulk composition (a–c), a three-layer Neural Network RocMLM (d–f), and absolute differences between Perple_X and NN3 (g–i) measured on the validation dataset. Other legend details are the same as in Figure 3.

"learned" by RocMLMs. Thus, "capacity" is intended to convey and compare the breadth of petrological "knowledge", or predictive capabilities, of Lookup Tables and RocMLMs. Within the same context, the notion of "capacity" is irrelevant for GFEM programs. Rather, GFEM performance primarily scales with the number of chemical components, phase solutions, and size of the compositional space defined by the user, as well as automatic grid refinement settings and other user-defined configuration options.

GFEM performance is reported using the range of average execution speeds (4–228 ms) and efficiencies (60–3138 ms·Mb) that we measured while generating our RocMLM training datasets as described in Section 2.2. To demonstrate the sensitivity of GFEM performance to alternative Perple_X configurations, we also show GFEM execution speed and efficiency for similar calculations using the thermodynamic data and phase solutions of Holland et al. (2018). Note that none of the Perple_X calculations using the Holland et al. (2018) configuration were used to train RocMLMs due to inaccurate seismic velocity predictions, and their performance metrics are only shown for illustrative purposes.

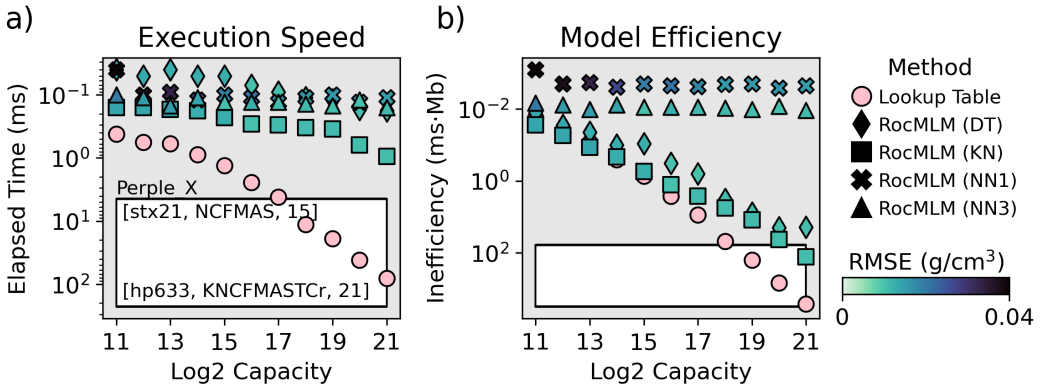

Figure 6: Computational efficiency of various approaches in terms of execution speed (a) and model efficiency (b). "Capacity" (x-axis) is a proxy for the petrological "knowledge", or predictive capabilities, of Lookup Tables and RocMLMs. White regions indicate GFEM efficiencies for different Perple_X configurations (thermodynamic dataset, chemical system, and number of solution phases are indicated in square brackets). stx21: Stixrude and Lithgow-Bertelloni (2022), hp633: Holland and Powell (2011) updated in Holland et al. (2018). Perple_X was run without multilevel grid refinement. RMSE is measured between density predictions and the validation dataset.

For Lookup Tables, execution speed and efficiency both scale roughly linearly with capacity on a logarithmic scale—indicating an inverse power-law relationship between Lookup Table capacity and performance (Figure 6). RocMLM performance, in contrast,

scales differently depending on the performance metric and regression algorithm. For example, RocMLM execution speed remains roughly constant, or increasing slightly with capacity, and shows relatively small variance among all regression algorithms (0.14 ± 0.26 ms, 2σ, Figure 6a). Yet RocMLM efficiency is markedly different for DT and KN algorithms compared to NN algorithms (Figure 6b). Despite the fast execution times of KN and DT algorithms (Figure 6a), their efficiency scales roughly linearly with capacity on a logarithmic scale—indicating an inverse power-law relationship between efficiency and capacity similar to Lookup Tables (Figure 6b). NN algorithms, on the other hand, show roughly constant efficiencies that indicate a high degree of information compression without sacrificing execution speed (Figure 6b). We note that training times for NN algorithms are many orders of magnitude larger than DT and KN algorithms (Supplementary Information). However, training times are neither limiting nor critical for geodynamic applications as training is independent from, and precedes numerical simulations.

Since accuracy is measured relative to the rock properties generated by GFEM programs, GFEM programs have perfect accuracy by definition. With respect to RocMLMs, validation accuracies (RMSE) are observed to be roughly constant for regression algorithms that apply binary decisions or local distance-based weights (DT and KN), while algorithms that apply global activation-based weights (NNs) show a positive correlation between accuracy and capacity (Figure 6). In addition to improving accuracy with increasing amounts of training examples, NN accuracy also increases with the number of hidden-layers (Figure 6) because deeper networks are more capable of fitting sharp gradients in the training data (see Supplementary Information for examples of NN1, NN2, and NN3 RocMLMs). We also tested the effects of NN width (changing the number of nodes within each hidden layer), but this had a negligible impact on NN performance and accuracy compared to increasing NN depth.

4 Discussion

4.1 RocMLM Performance Tradeoffs

RocMLM performance and accuracy are both critical for geodynamic applications and crucial for determining if RocMLMs are an improvement over methods commonly used for predicting rock properties in numerical geodynamic simulations. In terms of pure execution speed, our testing demonstrates that RocMLMs can make predictions between

10^1–10^3 times faster than GFEM programs and Lookup Tables (Figure 6), depending on the GFEM program configuration. The difference in execution speed between Lookup Tables and RocMLMs is small for low-resolution models (Figure 6) that are limited to $\leq$ 16 mantle compositions and large PT intervals ($\geq$ 1.7 GPa and 100 K PT step sizes). However, such low-resolution models are not an obvious improvement over simple polynomial approximations of a selective number of important phase transformations. At higher resolutions, RocMLMs can accurately resolve the physical properties of all thermodynamically-stable mineral assemblages in fine detail (at PT intervals of $\leq$ 0.2 GPa and 12 K) and for a wide variety of bulk mantle compositions (Figure 2). In addition to their broad predictive capabilities, high-resolution RocMLMs make predictions at speeds (approximately 0.1–1 ms, Figure 6) that allow computation of physical properties at the node-scale during geodynamic simulations. We therefore argue that high-resolution RocMLMs overcome all practical limitations for implementing thermodynamically self-consistent density evolution in numerical geodynamic models.

With respect to ranking the practicality of different RocMLM for geodynamic applications, execution speeds and accuracies alone suggest that high-resolution RocMLMs will perform with roughly equivalent outcomes regardless of the regression algorithm (Figure 6a). However, our testing reveals an obvious tradeoff between RocMLM performance and accuracy when accounting for compression ratio (i.e., the amount of "learned" information relative to the RocMLM file size). Figure 6b shows DT and KN algorithms becoming rapidly inefficient compared to NNs as the capacity of the training dataset increases. This is because NN algorithms require relatively little information to make predictions after training (weights and biases for each neuron) compared to DT (tree structure: nodes, splits, and predictions) and KN (entire training dataset with distance weights) algorithms. Moreover, accuracy tends to improve monotonically with dataset capacity for NN, but not for DT or KN. We therefore argue that deep NN RocMLMs are the most practical choice for geodynamic applications for three reasons: 1) modeling more rock types only requires adding more training data, 2) adding more training data improves prediction accuracy without diminishing performance, and 3) further improvements and adaptations to different geodynamic applications are possible by exploring different architectures than the simple NN models we have tested thus far.

The main limitations of NN RocMLMs are twofold: 1) training is computationally expensive compared to other regression algorithms (Supplementary Information) and 2)

shallow NN architectures imply smoother gradients in rock properties than GFEM calculations. We do not consider these limitations critical because training time is independent from RocMLM performance and even if deeper NN architectures are needed to fit discontinuities in rock properties with high accuracy, the number of layers and neurons in each layer remains small (Table 3). We note that our testing has been limited to the prediction of three properties that are mostly P-dependent and are relatively continuous despite a few large discontinuities. In principle, RocMLMs can be trained on any thermodynamic variable output by GFEM programs. However, we have not yet trained RocMLMs on more discrete, discontinuous, and/or highly T-dependent variables, such as modal proportions of minerals, volatile contents, or melt fraction, which will be treated in future developments of RocMLMs.

4.2 Geophysical and Thermodynamic Estimates of Elastic Properties

The amount of overlap between RocMLM profiles and PREM (Figures 3–3) suggests good agreement between thermodynamic and geophysical estimates of the elastic properties of mantle rocks within the limits of our training dataset and Perple_X configuration (see Sections 2.1 and 2.2). Discrepancies between thermodynamic profiles and PREM can be explained by chemical heterogeneity and/or differences in mantle geotherms that modify phase relations (Goes et al., 2022; Karki and Stixrude, 1999; Karki et al., 2001; Stixrude and Lithgow-Bertelloni, 2012; Waszek et al., 2021; Xu et al., 2008). Because the RocMLM training dataset spans a wide array of synthetic bulk mantle compositions, we can directly test the sensitivity of thermodynamic estimates to changes in bulk FeO–MgO contents (Figure 7).

As Fertility Index (ξ) increases by refertilization and/or lack of melt extraction and the bulk mantle composition becomes more Fe-rich (and more dense), Vp and Vs respond (both positively and negatively) according to the equations of state described in Stixrude and Lithgow-Bertelloni (2005). RocMLM training data show that density is the least sensitive parameter to ξ overall with only modest variations across a broad range of mantle rocks from fertile to highly depleted ($\xi = 0.76$, Figure 7a). The largest density variations occur at pressures below the olivine $\rightarrow$ wadsleyite transition (< 410 km), yet are still small enough (approximately 3–5 %) to imply that spontaneous mantle convection requires strong thermal gradients and/or hydration by metamorphic fluids in addition to melt extraction.

In contrast to density, Vp and Vs are more sensitive to ξ overall, especially at pressures above the olivine → wadsleyite transition (> 410 km). RocMLM training data suggests that an "optimal" Vp/Vs profile requires a more depleted mantle between 410–670 km and a more fertile mantle at < 410 km (Figure 7b,c). Forming this compositional layering pattern is counterintuitive, however, as partial melting is expected to be more pervasive at lower pressures. Moreover, density profiles are incongruent with this pattern, suggesting instead that a depleted mantle at < 410 km and more fertile mantle at > 410 km are required for an optimal fit with PREM and STW105 (Figure 7a).

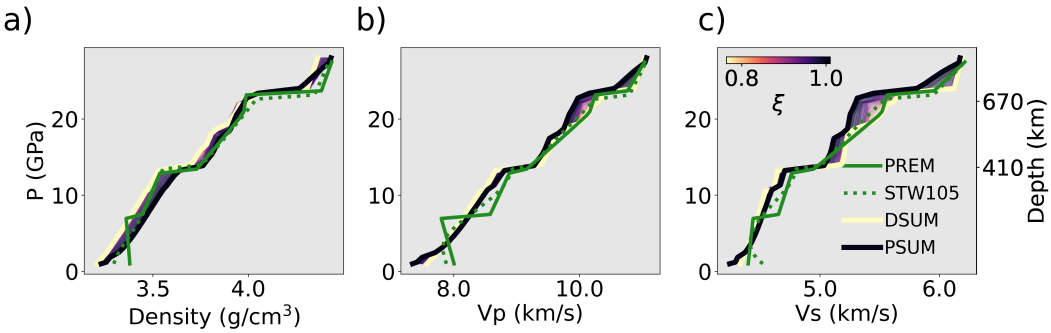

Figure 7: Depth profiles of RocMLM training data along a 1573 K mantle adiabat showing the sensitivities of thermodynamic estimates of density (a), Vp (b), and Vs (c) to changes in bulk mantle composition (as represented by the Fertility Index, ξ). Geophysical profiles PREM and STW105 (green lines) and the profiles of synthetic mantle end-member compositions PSUM and DSUM (thick colored lines) are shown for reference. Thin colored lines show profiles for the entire range of RocMLM training data.

5 Conclusions

The dynamics of Earth's upper mantle is largely driven by density contrasts stemming from changes in PT conditions, which lead to phase transformations in mantle rocks. These phase transformations also modify the elastic properties of mantle rocks. Therefore phase changes must be considered when inverting present-day mantle structure from seismic data. Likewise, numerical geodynamic simulations of mantle convection must account for thermodynamics, but are typically implemented with simple PT-dependent parameterizations of rock properties and phase boundaries that do not explicitly account for changes in Gibbs Free Energy resulting from changes in PT and in bulk composition. Here, we introduce RocMLMs as an alternative to GFEM programs and we evaluate RocMLM performance and accuracy. We also show how the RocMLM predictions compare to PREM and STW105. Our main findings are as follows:

1. RocMLMs predict density and elastic properties with high accuracy and are up to 101–103 faster than commonly used methods. This improvement in prediction speed makes thermodynamically self-consistent mantle convection within high-resolution numerical geodynamic models practical for the first time.
2. RocMLMs trained with moderately deep (3 hidden layers) NNs are more robust and efficient compared to RocMLMs trained with other regression algorithms, and are therefore the most practical models for coupling with numerical geodynamic codes.
3. RocMLM training data are sensitive to bulk mantle composition and geothermal gradients, yet show good agreement with PREM and STW105 for an average mantle geotherm.

Based on our results, we argue that moderately deep NN RocMLMs can be exceptional emulators of GFEM programs in geodynamic simulations that require computationally efficient predictions of rock properties. We have demonstrated that RocMLMs perform remarkably well for dry mantle rocks with compositions ranging from very fertile lherzolites to strongly depleted harzburgites and PT conditions ranging from 1–28 GPa and 773–2273 K.

Moreover, the RocMLM approach can be used with any GFEM program and thermodynamic dataset. Any improvement to the underlying thermodynamic data should further increase the accuracy of RocMLM predictions. Testing RocMLMs predictions on other thermodynamic variables of interest, including modal proportions of minerals, volatile contents, and melt fractions will be the focus of future studies. Likewise, in future works, we will extend the training data to include hydrous systems and additional end-member mantle compositions (e.g., pyroxenites and dunites).

6 Acknowledgements

This work was supported by the Tremplin-ERC grant LEARNING awarded to Nestor Cerpa by the I-SITE excellence program at the Université de Montpellier. We thank Maurine Montagnat, Fernando Carazo, Nicolas Berlie, and many researchers and students at Géosciences Montpellier for their thoughtful feedback during the development of this work. We gratefully acknowledge additional support from the European Research Coun-

cil (ERC) under the European Union Horizon 2020 Research and Innovation program grant agreement No. 882450 (ERC RhEoVOLUTION) awarded to Andréa Tommasi.

7 Open Research

All data, code, and relevant information for reproducing this work can be found at `https://github.com/buchanankerswell/kerswell_et_al_rocmlm`, and at `https://doi.org/10.17605/OSF.IO/K23TB`, the official Open Science Framework data repository (Kerswell et al., 2024). All code is MIT Licensed and free for use and distribution (see license details). Reference models PREM and STW105 are freely available from the Incorporated Research Institutions for Seismology Earth Model Collaboration (IRIS EMC, doi: 10.17611/DP/EMC.1, Trabant et al., 2012). All computations were made using CPUs of a Macbook Pro (2022; M2 chip) with macOS 13.4 and using Python 3.11.4.

8 References

References

Agrusta, R., Goes, S., and van Hunen, J. (2017). Subducting-slab transition-zone interaction: Stagnation, penetration and mode switches. *Earth and Planetary Science Letters*, 464:10–23.

Agrusta, R., Tommasi, A., Arcay, D., Gonzalez, A., and Gerya, T. (2015). How partial melting affects small-scale convection in a plume-fed sublithospheric layer beneath fast-moving plates. *Geochemistry, Geophysics, Geosystems*, 16(11):3924–3945.

Akaogi, M., Ito, E., and Navrotsky, A. (1989). Olivine-modified spinel-spinel transitions in the system mg2sio4-fe2sio4: Calorimetric measurements, thermochemical calculation, and geophysical application. *Journal of Geophysical Research: Solid Earth*, 94(B11):15671–15685.

Akaogi, M., Takayama, H., Kojitani, H., Kawaji, H., and Atake, T. (2007). Low-temperature heat capacities, entropies and enthalpies of mg 2 sio 4 polymorphs, and α- β- γ and post-spinel phase relations at high pressure. *Physics and Chemistry of Minerals*, 34:169–183.

Ballmer, M., Schmerr, N., Nakagawa, T., and Ritsema, J. (2015). Compositional mantle layering revealed by slab stagnation at˜ 1000-km depth. *Science advances*, 1(11):e1500815.

Bina, C. and Helffrich, G. (1994). Phase transition clapeyron slopes and transition zone seismic discontinuity topography. *Journal of Geophysical Research: Solid Earth*, 99(B8):15853–15860.

Brown, E. and Lesher, C. (2016). Reebox pro: a forward model simulating melting of thermally and lithologically variable upwelling mantle. *Geochemistry, Geophysics, Geosystems*, 17(10):3929–3968.

Cagnioncle, A., Parmentier, E., and Elkins-Tanton, L. (2007). Effect of solid flow above a subducting slab on water distribution and melting at convergent plate boundaries. *Journal of Geophysical Research: Solid Earth*, 112(B9).

Cerpa, N., Arcay, D., and Padrón-Navarta, J. (2022). Sea-level stability over geological time owing to limited deep subduction of hydrated mantle. *Nature Geoscience*, 15(5):423–428.

Cerpa, N., Wada, I., and Wilson, C. (2019). Effects of fluid influx, fluid viscosity, and fluid density on fluid migration in the mantle wedge and their implications

for hydrous melting. *Geosphere*, 15(1):1–23.

Christensen, U. (1995). Effects of phase transitions on mantle convection. *Annual Review of Earth and Planetary Sciences*, 23(1):65–87.

Čížková, H. and Bina, C. (2013). Effects of mantle and subduction-interface rheologies on slab stagnation and trench rollback. *Earth and Planetary Science Letters*, 379:95–103.

Connolly, J. (2009). The geodynamic equation of state: what and how. *Geochemistry, geophysics, geosystems*, 10(10).

Connolly, J. and Kerrick, D. (2002). Metamorphic controls on seismic velocity of subducted oceanic crust at 100–250 km depth. *Earth and Planetary Science Letters*, 204(1-2):61–74.

Connolly, J. and Khan, A. (2016). Uncertainty of mantle geophysical properties computed from phase equilibrium models. *Geophysical Research Letters*, 43(10):5026–5034.

Dziewoński, A. and Anderson, D. (1981). Preliminary reference earth model. *Physics of the earth and planetary interiors*, 25(4):297–356.

Fukao, Y., Widiyantoro, S., and Obayashi, M. (2001). Stagnant slabs in the upper and lower mantle transition region. *Reviews of Geophysics*, 39(3):291–323.

Gerya, T. and Meilick, F. (2011). Geodynamic regimes of subduction under an active margin: effects of rheological weakening by fluids and melts. *Journal of Metamorphic Geology*, 29(1):7–31.

Gerya, T. and Yuen, D. (2003). Rayleigh–taylor instabilities from hydration and melting propel 'cold plumes' at subduction zones. *Earth and Planetary Science Letters*, 212(1-2):47–62.

Gibbs, J. (1878). On the equilibrium of heterogeneous substances. *American Journal of Science*, 3(96):441–458.

Goes, S., Yu, C., Ballmer, M., Yan, J., and van der Hilst, R. (2022). Compositional heterogeneity in the mantle transition zone. *Nature Reviews Earth & Environment*, 3(8):533–550.

Green, D. (1979). Petrogenesis of mid ocean ridge basalts. *The Earth: its origin, structure and evolution*, pages 200–299.

Hirose, K. (2002). Phase transitions in pyrolitic mantle around 670-km depth: Implications for upwelling of plumes from the lower mantle. *Journal of Geophysical*

Research: Solid Earth, 107(B4):ECV–3.

Holland, T., Green, E., and Powell, R. (2018). Melting of peridotites through to granites: a simple thermodynamic model in the system kncfmashtocr. *Journal of Petrology*, 59(5):881–900.

Holland, T. and Powell, R. (2001). Calculation of phase relations involving haplogranitic melts using an internally consistent thermodynamic dataset. *Journal of Petrology*, 42(4):673–683.

Holland, T. and Powell, R. (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(3):333–383.

Ishii, T., Huang, R., Fei, H., Koemets, I., Liu, Z., Maeda, F., Yuan, L., Wang, L., Druzhbin, D., Yamamoto, T., et al. (2018). Complete agreement of the post-spinel transition with the 660-km seismic discontinuity. *Scientific reports*, 8(1):6358.

Ita, J. and Stixrude, L. (1992). Petrology, elasticity, and composition of the mantle transition zone. *Journal of Geophysical Research: Solid Earth*, 97(B5):6849–6866.

Ito, E. (1982). Stability relations of silicate spinels, ilmenites, and perovskite. *High pressure research in geophysics*, pages 405–419.

Ito, E., Akaogi, M., Topor, L., and Navrotsky, A. (1990). Negative pressure-temperature slopes for reactions forming mgsio3 perovskite from calorimetry. *Science*, 249(4974):1275–1278.

Ito, E. and Katsura, T. (1989). A temperature profile of the mantle transition zone. *Geophysical Research Letters*, 16(5):425–428.

Ito, E. and Takahashi, E. (1989). Postspinel transformations in the system mg2sio4-fe2sio4 and some geophysical implications. *Journal of Geophysical Research: Solid Earth*, 94(B8):10637–10646.

Jenkins, J., Cottaar, S., White, R., and Deuss, A. (2016). Depressed mantle discontinuities beneath iceland: Evidence of a garnet controlled 660 km discontinuity? *Earth and Planetary Science Letters*, 433:159–168.

Karato, S., Riedel, M., and Yuen, D. (2001). Rheological structure and deformation of subducted slabs in the mantle transition zone: implications for mantle circulation and deep earthquakes. *Physics of the Earth and Planetary Interiors*,

127(1-4):83–108.

Karki, B. and Stixrude, L. (1999). Seismic velocities of major silicate and oxide phases of the lower mantle. *Journal of Geophysical Research: Solid Earth*, 104(B6):13025–13033.

Karki, B., Stixrude, L., and Wentzcovitch, R. (2001). High-pressure elastic properties of major materials of earth's mantle from first principles. *Reviews of Geophysics*, 39(4):507–534.

Katsura, T. and Ito, E. (1989). The system mg2sio4-fe2sio4 at high pressures and temperatures: Precise determination of stabilities of olivine, modified spinel, and spinel. *Journal of Geophysical Research: Solid Earth*, 94(B11):15663–15670.

Katsura, T., Yamada, H., Shinmei, T., Kubo, A., Ono, S., Kanzaki, M., Yoneda, A., Walter, M., Ito, E., Urakawa, S., et al. (2003). Post-spinel transition in mg2sio4 determined by high p–t in situ x-ray diffractometry. *Physics of the Earth and Planetary Interiors*, 136(1-2):11–24.

Kelley, K., Plank, T., Newman, S., Stolper, E., Grove, T., Parman, S., and Hauri, E. (2010). Mantle melting as a function of water content beneath the mariana arc. *Journal of Petrology*, 51(8):1711–1738.

Kerswell, B., Cerpa, N., Tommasi, A., Godard, M., and Padrón-Navarta, J. (2024). RocMLMs: Predicting rock properties through machine learning models [dataset repository].

Kerswell, B., Kohn, M., and Gerya, T. (2021). Backarc lithospheric thickness and serpentine stability control slab-mantle coupling depths in subduction zones. *Geochemistry, Geophysics, Geosystems*, 22(6):e2020GC009304.

Kuritani, T., Xia, Q., Kimura, J., Liu, J., Shimizu, K., Ushikubo, T., Zhao, D., Nakagawa, M., and Yoshimura, S. (2019). Buoyant hydrous mantle plume from the mantle transition zone. *Scientific Reports*, 9(1):6549.

Kustowski, B., Ekström, G., and Dziewoński, A. (2008). Anisotropic shear-wave velocity structure of the earth's mantle: A global model. *Journal of Geophysical Research: Solid Earth*, 113(B6).

Le Roux, V., Bodinier, J., Tommasi, A., Alard, O., Dautria, J., Vauchez, A., and Riches, A. (2007). The lherz spinel lherzolite: refertilized rather than pristine mantle. *Earth and Planetary Science Letters*, 259(3-4):599–612.

Li, Z., Gerya, T., and Connolly, J. (2019). Variability of subducting slab morphologies in the mantle transition zone: Insight from petrological-thermomechanical modeling. *Earth-Science Reviews*, 196:102874.

Litasov, K., Ohtani, E., Sano, A., Suzuki, A., and Funakoshi, K. (2005). Wet subduction versus cold subduction. *Geophysical Research Letters*, 32(13).

Liu, M., Yuen, D., Zhao, W., and Honda, S. (1991). Development of diapiric structures in the upper mantle due to phase transitions. *Science*, 252(5014):1836–1839.

Maruyama, S., Liou, J., and Terabayashi, M. (1996). Blueschists and eclogites of the world and their exhumation. *International geology review*, 38(6):485–594.

Morishima, H., Kato, T., Suto, M., Ohtani, E., Urakawa, S., Utsumi, W., Shimomura, O., and Kikegawa, T. (1994). The phase boundary between α-and β-mg2sio4 determined by in situ x-ray observation. *Science*, 265(5176):1202–1203.

Nakagawa, T. and Buffett, B. (2005). Mass transport mechanism between the upper and lower mantle in numerical simulations of thermochemical mantle convection with multicomponent phase changes. *Earth and Planetary Science Letters*, 230(1-2):11–27.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Riel, N., Kaus, B., Green, E., and Berlie, N. (2022). Magemin, an efficient gibbs energy minimizer: application to igneous systems. *Geochemistry, Geophysics, Geosystems*, 23(7):e2022GC010427.

Ringwood, A. (1991). Phase transformations and their bearing on the constitution and dynamics of the mantle. *Geochimica et Cosmochimica Acta*, 55(8):2083–2110.

Schubert, G., Yuen, D., and Turcotte, D. (1975). Role of phase transitions in a dynamic mantle. *Geophysical Journal International*, 42(2):705–735.

Shaw, D. (1970). Trace element fractionation during anatexis. *Geochimica et Cosmochimica Acta*, 34(2):237–243.

737 Sizova, E., Gerya, T., Brown, M., and Perchuk, L. (2010). Subduction styles in the
738 precambrian: insight from numerical experiments. *Lithos*, 116(3-4):209–229.
739 Spear, F. (1993). Metamorphic phase equilibria and pressure-temperature-time
740 paths. *Mineralogical Society of America Monograph*, 799.
741 Stixrude, L. and Lithgow-Bertelloni, C. (2005). Thermodynamics of mantle miner-
742 als—i. physical properties. *Geophysical Journal International*, 162(2):610–632.
743 Stixrude, L. and Lithgow-Bertelloni, C. (2012). Geophysics of chemical heterogeneity
744 in the mantle. *Annual Review of Earth and Planetary Sciences*, 40:569–595.
745 Stixrude, L. and Lithgow-Bertelloni, C. (2022). Thermal expansivity, heat ca-
746 pacity and bulk modulus of the mantle. *Geophysical Journal International*,
747 228(2):1119–1149.
748 Sun, S. and McDonough, W. (1989). Chemical and isotopic systematics of oceanic
749 basalts: implications for mantle composition and processes. *Geological Society,*
750 *London, Special Publications*, 42(1):313–345.
751 Syracuse, E., van Keken, P., and Abers, G. (2010). The global range of subduction
752 zone thermal models. *Physics of the Earth and Planetary Interiors*, 183(1-
753 2):73–90.
754 Tackley, P., Stevenson, D., Glatzmaier, G., and Schubert, G. (1994). Effects of mul-
755 tiple phase transitions in a three-dimensional spherical model of convection in
756 earth's mantle. *Journal of Geophysical Research: Solid Earth*, 99(B8):15877–
757 15901.
758 Tackley, P. and Xie, S. (2003). Stag3d: a code for modeling thermo-chemical multi-
759 phase convection in earth's mantle. In *Computational Fluid and Solid Mechan-*
760 *ics 2003*, pages 1524–1527. Elsevier.
761 Torii, Y. and Yoshioka, S. (2007). Physical conditions producing slab stagnation:
762 Constraints of the clapeyron slope, mantle viscosity, trench retreat, and dip
763 angles. *Tectonophysics*, 445(3-4):200–209.
764 Trabant, C., Hutko, A., Bahavar, M., Karstens, R., Ahern, T., and Aster, R. (2012).
765 Data products at the iris dmc: Stepping stones for research and other applica-
766 tions. *Seismological Research Letters*, 83(5):846–854.
767 Wang, X., Wilde, S., Li, Q., and Yang, Y. (2015). Continental flood basalts derived
768 from the hydrous mantle transition zone. *Nature Communications*, 6(1):7700.

769 Waszek, L., Tauzin, B., Schmerr, N., Ballmer, M., and Afonso, J. (2021). A poorly
770 mixed mantle transition zone and its thermal state inferred from seismic
771 waves. *Nature Geoscience*, 14(12):949–955.
772 Workman, R. and Hart, S. (2005). Major and trace element composition of the de-
773 pleted morb mantle (dmm). *Earth and Planetary Science Letters*, 231(1-2):53–
774 72.
775 Xu, W., Lithgow-Bertelloni, C., Stixrude, L., and Ritsema, J. (2008). The effect of
776 bulk composition and temperature on mantle seismic structure. *Earth and*
777 *Planetary Science Letters*, 275(1-2):70–79.
778 Yang, J. and Faccenda, M. (2020). Intraplate volcanism originating from upwelling
779 hydrous mantle transition zone. *Nature*, 579(7797):88–91.