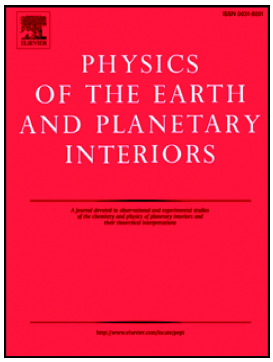
Journal Pre-proof



Inferring material properties of the lower mantle minerals using Mixture Density Networks

Ashim Rijal, Laura Cobden, Jeannot Trampert, Jennifer M. Jackson, Andrew Valentine

PII: S0031-9201(21)00142-4

DOI: https://doi.org/10.1016/j.pepi.2021.106784

Reference: PEPI 106784

To appear in: *Physics of the Earth and Planetary Interiors*

Please cite this article as: A. Rijal, L. Cobden, J. Trampert, et al., Inferring material properties of the lower mantle minerals using Mixture Density Networks, *Physics of the Earth and Planetary Interiors* (2021), https://doi.org/10.1016/j.pepi.2021.106784

Highlights

- Equation of state (EOS) modelling is a powerful tool to estimate mineral properties at conditions not accessed by high pressure and temperature experiments.
- Experimental errors, both random and systematic (e.g. pressure scale, functional forms), data consistency and sparsity all contribute to the uncertainties in mineral seismic properties.
- Conventional explicit EOSs which are assumed to follow certain form provide a priori information by fixing their functional form or pressure scale, thereby providing a biased estimate of uncertainties.
- Neural networks based approach can implicitly capture full uncertainties together with highlighting data gaps and identifying data inconsistencies.

Inferring material properties of the lower mantle minerals using Mixture Density Networks

Ashim Rijal[a,*], Laura Cobden[a], Jeannot Trampert[a], Jennifer M. Jackson[b], Andrew Valentine[c]

[a] *Department of Earth Sciences, Utrecht University, Princetonlaan 8a, 3584 CB Utrecht, The Netherlands.*

[b] *Seismological Laboratory, California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125, USA.*

[c] *Department of Earth Sciences, Durham University, South Road, Durham, DH1 3LE, UK.*

Abstract

Interpretation of information available from seismic data in terms of temperature and composition requires an understanding of the physical properties of minerals, in particular, the elastic properties of candidate Earth minerals at the relevant (here, lower mantle) pressure and temperature. A common practise for the bulk elastic properties is to measure volume at a range of pressures and temperatures using experiments or computational methods. These datasets are then typically fit to a pre-determined functional form, or equation of state to allow computation of elastic properties at any other pressure or temperature. However, errors, both random and systematic, limitations in the number of data and choice of pressure marker and scale, as well as different functional forms of equations of state, all contribute to the uncertainties in mineral seismic properties. In an attempt to present a more comprehensive view of these uncertainties, we use neural-network based techniques to infer the relationship among: pressure, temperature, volume, bulk modulus, and thermal expansivity of MgO. We illustrate our approach on experimental data, but an extension to ab initio data is straightforward. The type of neural network used is called a Mixture Density Network (MDN) which is a combination of a conventional feed-forward neural network and a mixture

*Corresponding author
Email address: a.rijal@uu.nl (Ashim Rijal)

model that consists of Gaussian functions. MDNs are capable of approximating arbitrary probability density functions, which allows us to compute the uncertainties in the predicted equations of state. Since the networks interpolate locally between input samples, pressure-volume-temperature relations are implicitly learned from data without imposing any explicit thermodynamic assumptions or ad-hoc relationships. We use the partial derivatives of the mapping between inputs (pressure and temperature) and output (volume) to compute the isothermal bulk modulus and thermal expansivity. Flexibility of the MDNs allows us to investigate the uncertainty due to certain data in one region of pressure-temperature space without influencing the posterior probability density everywhere. In general, we find that the elastic properties of MgO are well-constrained by experimental data. However, our study highlights regions in which sparse or inconsistent data lead to poorly constrained elastic properties, namely: at low pressure and high temperature (<25GPa and >1500 K), and temperatures above 2700 K. While the former conditions are likely not important for the Earth's lower mantle, they are relevant in other planetary bodies such as the Moon and Mars. Comparison with conventional equation of state forms shows that assuming a certain functional form of the pressure-volume-temperature relationship leads to potential bias in uncertainty quantification, because the uncertainties are then specific to the underlying form. In combination with data sets of other lower mantle minerals, this technique should improve uncertainty quantification in interpretations of seismic data.

Keywords: equations of state; lower mantle; neural networks; periclase; MgO

1. Introduction

Information such as variation of wave speeds (e.g. Dziewonski and Anderson 1981, Kennett et al. 1995), obtained by studying seismic data is crucial for understanding the internal structure of the Earth. Various studies have reported the presence of seismically distinct structures at multiple scales in the Earth's mantle (e.g. Garnero and Helmberger 1998, Ritsema et al. 1999, Romanowicz 2008, Hernlund and Houser 2008, Deschamps et al. 2012,

Garnero et al. 2016). In order to relate those observed seismic structures to appropriate temperature and composition, constraints from mineral physics on the sensitivity of seismic wave speeds to these parameters are required (e.g. Jackson 1998, Trampert et al. 2001). The sensitivities have been used to infer the probable existence of chemical heterogeneities within the mantle (e.g. Trampert et al. 2004, Dobrosavljevic et al. 2019, Jackson and Thomas 2021). Other studies have tried to constrain the (average) mantle geotherm and composition by combining seismic data and mineral seismic properties (e.g. Cammarano et al. 2003, 2005a,b, Deschamps and Trampert 2004, Stixrude and Lithgow-Bertelloni 2005, Matas et al. 2007, Cobden et al. 2008, 2009, Simmons et al. 2010, Khan et al. 2009, 2011, 2013). Mantle convection simulations (e.g. Nakagawa et al. 2009, 2010, 2012, Schuberth et al. 2009, 2012) have also incorporated mineral properties to illustrate the importance of joint geodynamical-mineralogical approaches to explain the seismic anomalies in the mantle. Mineral properties can be derived from experimental or theoretical methods. In particular, information on the density (or volume V), incompressibility and rigidity are required to obtain the seismic wave speeds in a material. Since it is not practical or feasible yet to perform experiments at each pressure (P) and temperature (T) that may exist within the Earth, the convention is to use equations of state (EOSs) to define the relationship among the thermodynamic variables P, V and T (e.g. Duffy and Wang, 1998), and hence be able to estimate mineral properties at the conditions not accessed by experiments.

However, a number of uncertainties are associated with this procedure. Experimental measurements contain random and systematic errors. The choice of pressure scale as well as different functional forms of the EOS (e.g. Vinet EOS, third/fourth order finite strain equations, also called Birch-Murnaghan EOSs, as well as the choice of Grüneisen models) all contribute to the uncertainties in mineral seismic properties. As a result, it becomes challenging to determine realistic uncertainties for the interpretations which relate seismic observations to temperature and composition.

In this study, we present an Artificial Neural Network (ANN) based approach to infer the

pressure-volume-temperature (P-V-T) relationship of MgO, with a view to extend the application to other major lower mantle minerals. We collate experimental P-V-T data for MgO together with reported uncertainties, regardless of pressure scale or functional form used. By applying ANN techniques, P-V-T relationships are implicitly learned from data without any prior assumption on the functional form (or thermodynamic model) of the relationship. Specifically, we use Mixture Density Networks to infer material properties and assess their uncertainties. We compute the partial derivatives of inferred volume with respect to pressure and temperature to extract the bulk modulus and thermal expansivity, respectively. In order to test the feasibility of this approach, we train the networks only on experimental data, although a combination of theoretical and experimental data is also possible and straightforward.

2. Equations of state: Uncertainties

Experimental approaches (e.g. Vassiliou and Ahrens 1981, Yoneda 1990, Utsumi et al. 1998, Duffy and Ahrens 1995, Fei 1999, Sinogeikin and Bass 2000, Sinogeikin et al. 2000, Dewaele et al. 2000, Speziale et al. 2001, Li et al. 2006, Dorogokupets and Dewaele 2007, Hirose et al. 2008, Murakami et al. 2009, Kono et al. 2010, Dorfman et al. 2012, Ye et al. 2017) have been used to establish the P-V-T relationship of MgO. Experiments using a diamond anvil cell (DAC), a multi-anvil press (MAP) and shock compression have provided a huge number of data covering a wide range of pressure and temperature. Laboratory measurements of volume are done at a discrete set of pressure and temperature points. To cover the entire pressure and temperature range of lower mantle requires pressure extrapolation and/or interpolation of the measurements using a thermal equation of state. The most common procedure (e.g. Matas et al. 2007, Cobden et al. 2009) is to use an isothermal equation of state with a Mie-Grüneisen model for thermal pressure. In this approach, the total pressure is considered to be the sum of a static pressure and a quasiharmonic thermal pressure. The static pressure term describes the pressure-volume relationship at a reference temperature

(usually 300 K). Different functional forms, such as third/fourth order finite strain and Vinet, have been widely used to model isothermal compression curves often leading to different estimates of fitting parameters or ambient mineral properties such as volume (V_0), bulk modulus (K_{0T}) and pressure derivative of bulk modulus (K'_{0T}) at 0 GPa pressure (e.g. Speziale et al. 2001, Dorogokupets and Dewaele 2007, Tange et al. 2009). To compute temperature effects (more precisely, thermal pressure) this framework uses a Grüneisen parameter whose volume dependence is uncertain (Ye et al. 2017). Although anharmonic effects are very small compared to the harmonic contribution to thermal pressure, some authors (e.g. Dorogokupets and Dewaele 2007) use models to account for this term as well.

Additionally, the exact determination of pressure using a reliable pressure scale in static high pressure and temperature experiments is still a challenging task. The ruby pressure scale of Forman et al. 1972 used in DAC experiments has been largely calibrated (Liu and Bi 2016) using both static and dynamic compression data, but still suffers from large experimental uncertainties. Dynamic shock compression experiments provide an absolute pressure scale. But the correction for thermal effects can be very uncertain (e.g. Dorfman et al. 2012, Duffy and Wang 1998), especially at high shock temperatures because the corresponding thermal contribution also increases. Other widely used pressure scales are gold, platinum and MgO scales. A recent study by Ye et al. 2017 shows the inter-comparison of those scales up to 140 GPa and 2500 K. They report ± 1 to 4 GPa (sometimes systematic) differences in pressure among those pressure scales. Although their study optimized different Au, Pt and MgO pressure scales to make them agree within ± 1 GPa, it concludes that the most preferred form of EOS (and the pressure standard itself) remains uncertain.

Measurement errors, lack of an absolute pressure scale, and a variety of functional forms of EOSs all contribute to the uncertainties in mineral seismic properties. Assuming one particular EOS or pressure scale has the potential to produce biased uncertainty estimates that are specific to the underlying functional form. In this study we train neural networks to learn the implicit relation between pressure and temperature (as inputs) and volume,

bulk modulus and thermal expansivity (as outputs). The results are entirely data-driven without a priori selection of experiments or a functional form to explain the data. In this way, we can infer the relative contributions of data sparsity versus prior conditioning to the uncertainties. We can also map the level of certainty of the elastic parameters in pressure-temperature space, which can be propagated into seismic interpretation.

3. The Mixture Density Network (MDN)

3.1. Background

Conventional neural networks (Hornik et al. 1989) are general function approximators, which can be used to infer an (arbitrary nonlinear) relationship (Cybenko 1989) between inputs and targets/outputs. However, the conditional average (i.e. the mean value of output conditioned on input data) given by such networks only provides limited information about that relationship (Bishop 1994). Since experimental P-V-T data contain measurement errors, and inferring P-V-T relationship using those data is an inverse problem which can have multiple solutions, naturally we seek to treat the problem in a probabilistic framework. Hence, instead of having only the average volume output, we want to find the posterior probability density function (pdf) for volume. The pdf for volume at a given pressure and temperature can be denoted as

$$\sigma(V|P,T). \tag{1}$$

We can represent a general pdf by combining a conventional feed-forward neural network with a Gaussian Mixture Model (GMM), which is then called a Mixture Density Network (MDN) (Bishop 1994 and Bishop 1995). The architecture of the MDN used in this study is shown in Figure 1, and consists of a two layer feed-forward neural network and a GMM. The GMM contains a mixture of a finite number of Gaussian kernels which are then weighted to give the posterior pdf. The mean, standard deviation and weight of each Gaussian kernel are parameterized by weights and biases of the feed-forward neural network, also known as

network parameters ($\boldsymbol{\alpha}$).

Application of MDNs in Earth Sciences ranges from inversion of surface wave data for global crustal thickness (Meier et al. 2007a,b), temperature and water content variations within the transition zone (Meier et al. 2009), inference of Earth's radial seismic structure (de Wit et al. 2013), inversion of free oscillations (de Wit et al. 2014), constraints on lower mantle anisotropy (de Wit and Trampert 2015), nonlinear petrophysical inversion (Shahraeeni and Curtis 2011), source inversion of strong-motion data (Käufl et al. 2016b), inferring parameters governing mantle convection (Atkins et al. 2016) to travel-time tomography (Earp and Curtis 2020). In our case, based on some experimental P-V-T data, we seek to approximate the true posterior pdf (Equation 1) by a parameterized posterior

$$p(V|P,T;\boldsymbol{\alpha}) \approx \sigma(V|P,T). \tag{2}$$

In other words, for a given pressure and temperature, the posterior probability density for volume is given by the pdf in expression 2 which is parameterized by the weights and biases ($\boldsymbol{\alpha}$) of the feed-forward neural network. These parameters are learned during the network training process (see Sub-section 3.2). The posterior pdf (Equation 2) can be expressed as a linear combination of a fixed number of Gaussian kernels (also see Figure 1) as

$$p(V|P,T;\boldsymbol{\alpha}) = \sum_{n=1}^{M} \pi_n(P,T;\boldsymbol{\alpha})\phi_n(V|P,T;\boldsymbol{\alpha}) \tag{3}$$

where M denotes the number of kernels used, and π_n are mixing coefficients which satisfy

$$\sum_{n=1}^{M} \pi_n(P,T;\boldsymbol{\alpha}) = 1. \tag{4}$$

If the number of Gaussian kernels is M, then the total number of outputs from the feed-forward network is K= 3M because each kernel is parameterized by its weight (π_n), mean (μ_n) and standard deviation (σ_n). Equation 4 ensures that the posterior integrates to 1

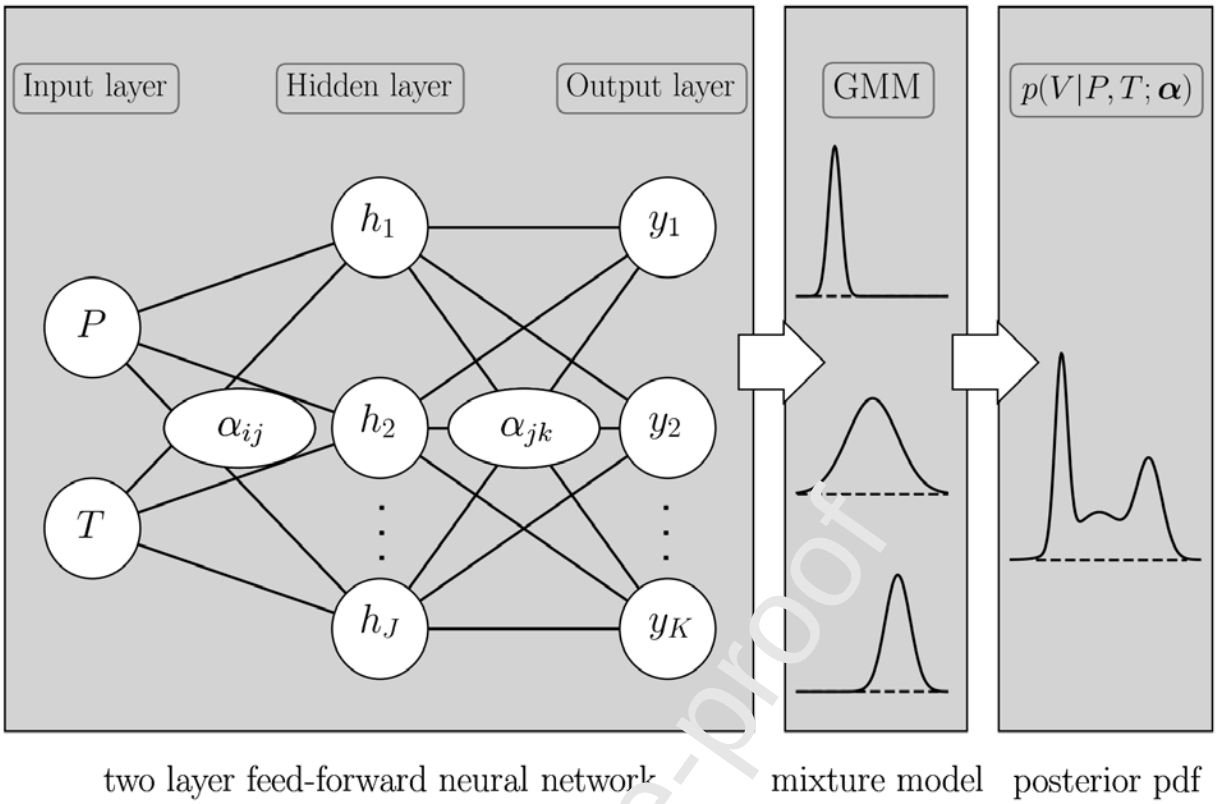


Figure 1: Architecture of the Mixture Density Network (MDN). A two layer feed-forward neural network (left) is combined with a GMM (centre) to get the posterior pdf (right). P & T denote the network inputs, h_j are the hidden nodes, and y_k are the outputs of the feed-forward network. Indices J and K represent the number of hidden and output nodes, respectively. Except for the input nodes, each circle represents a computational node. Hidden layer nodes take a weighted sum (with weights α_{ij}, where $i \neq 0$) of input data (P & T) plus a bias term (α_{0j}) as inputs and apply a sigmoid activation function. The output layer nodes take a weighted sum (weighted by α_{jk}, where $j \neq 0$) of the outputs from the hidden layer plus a bias (α_{0k}) and apply a linear activation function to give the outputs y_k. These outputs are related to the mean, standard deviation and weight of each Gaussian in the GMM (see Appendix A for details). Each Gaussian in the GMM is then weighted to give the final posterior pdf.

making it a valid probability density. ϕ_n in equation 3 are Gaussian kernels of the form

$$\phi_n(V|P,T;\boldsymbol{\alpha}) = \frac{1}{\sqrt{2\pi}\sigma_n(P,T;\boldsymbol{\alpha})} exp\left\{ -\frac{(V-\mu_n(P,T;\boldsymbol{\alpha}))^2}{2\sigma_n(P,T;\boldsymbol{\alpha})^2} \right\} \tag{5}$$

where μ_n and σ_n are the mean and standard deviation of Gaussian kernels in the GMM. These parameters of the GMM are related to the outputs (y_k) of the feed-forward network (see details in Appendix A).

3.2. MDN initialization and training

In order to find the appropriate weights and biases of the feed-forward neural network, we train the MDN using a sub-set of the experimental P-V-T data. In fact, the total experimental P-V-T data, shown in Figure 2 (Fei 1999, Jacobsen et al. 2008, Fei et al. 2004a, Fei et al. 2004b, Dewaele et al. 2000, Speziale et al. 2001, Utsumi et al. 1998, Fiquet et al. 1999, Ye et al. 2017, Kono et al. 2010, Dorfman et al. 2012, Zhang 2000, Fiquet et al. 1996, Dubrovinsky and Saxena 1997, Hirose et al. 2008, Litasov et al. 2005, Murakami et al. 2012, Sinogeikin and Bass 2000, Li et al. 2006 and Fan et al. 2019), is divided into three sets: training (70%), monitoring (20%) and test (10%) sets. During training, the MDN takes pressure and temperature from the training data and outputs a pdf for volume according to Equation 3. However, we need to decide on the initial values of the network parameters of the feed-forward neural network to compute the first output. We randomly draw the input layer and hidden layer weights (Bishop 1995) according to Gaussian distributions (see Appendix B for details). Once the MDN is initialized and training has started, the difference between the output and the target can be computed according to an error function defined in Appendix B. This function is also called the loss function which is minimized iteratively using the ADAM optimization method (see detailed algorithm in Kingma and Ba 2014). We use TensorFlow (1.13.1) (Abadi et al. 2015) to construct, train and evaluate the MDN.

Overfitting is a general property of the maximum likelihood technique (Bishop 1995). We use a separate monitoring data set to monitor the error decay during training. We evaluate the monitoring set error at the end of each iteration; if the monitoring error starts to increase (i.e. the network starts to over-fit the training data) then we stop the training procedure and save the last best trained model. This technique is also called the early-stopping technique.

It is known that the inverse problem can have multiple solutions (i.e. a range of network parameters can possibly provide equally likely solutions). We train a number of independent MDNs, and combine them by a weighted sum (e.g. Käufl et al. 2016a). The weight of each network is based on how well it performs on the test data which is not used during training.

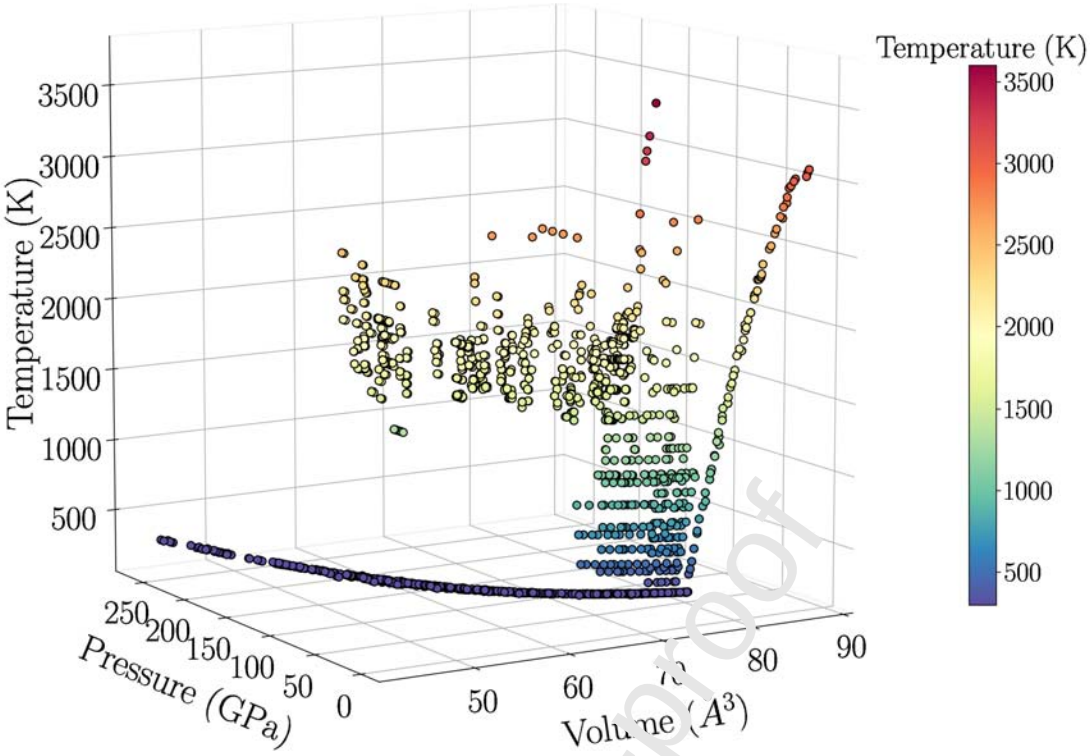


Figure 2: Experimental P-V-T data for MgO used in this study (Fei 1999, Jacobsen et al. 2008, Fei et al. 2004a, Fei et al. 2004b, Dewaele et al. 2000, Speziale et al. 2001, Utsumi et al. 1998, Fiquet et al. 1999, Ye et al. 2017, Kono et al. 2010, Dorfman et al. 2012, Zhang 2000, Fiquet et al. 1996, Dubrovinsky and Saxena 1997, Hirose et al. 2008, Litasov et al. 2005, Murakami et al. 2012, Sinogeikin and Bass 2000, Li et al. 2006 and Fan et al. 2019) to train the MDNs. Data with uncertainties from X-ray diffraction experiments (in static high P-T, Brillouin spectroscopy and ultrasonic interferometry) are collected for the analysis. Note: uncertainties in collected experimental data are not plotted because the scale would be inappropriate to visualize them.

The performance is measured by the same error function that we use to calculate training and monitoring errors (for details see Appendix B). In this way, the explicit dependence of the posterior on the network parameters can be avoided. The choice of the number of MDNs depends on the problem at hand. A rough estimate for a relatively simple problem (e.g. a few inputs and a target/output) may lie in the range 10-20 (Käufl et al. 2016a). However, in order to compute the uncertainties in bulk modulus and thermal expansivity (details in Section 5) we train a large number of MDNs (10^3). The number of hidden nodes to use in each MDN are randomly selected from a pre-defined range which is 16-32. We conducted a separate test (not shown here) to find the range that provides the lowest errors for the test set. Similarly, we propagate the uncertainties in experimental data through the MDNs by randomly perturbing the thermodynamic variables within the reported uncertainty range.

3.3. Network performance

We use the test data set to examine how well the trained MDNs perform when a new datum is presented. Since the test data are not used in network training, we can use them to predict the output and subsequently compare with target data. In Figure 3 (top panel) the predicted volume is compared with the target data. The MDNs predict pdfs for volume, and for this comparison we compute the conditional mean volume (conditioned on inputs P & T), instead of using the full posterior pdfs on volume, as

$$< V|P,T;\boldsymbol{\alpha} >= \sum_{n=1}^{M} \pi_n(P,T;\boldsymbol{\alpha})\mu_n(P,T;\boldsymbol{\alpha}). \tag{6}$$

This special case of MDN corresponds to the standard neural network output (Bishop 1994), i.e. only the feed-forward network with one volume output. Equation 6 shows the mean volume output for one MDN, and we calculate the weighted sum (weights are chosen according to the test set error as mentioned previously) of mean volumes from all MDNs. One alternative to the conditional posterior mean could be the posterior mode. However, the posterior mode may be biased towards certain pressure scales which contain relatively more data in the training set compared to other scales.

In the region of high temperatures and low pressures (Figure 3, top panel) the trained MDNs show lower resolving capacity, providing more uncertain volume predictions. We found that this discrepancy in network predictions comes from the inclusion of specific training data points (high temperature data of Fiquet et al. 1996) in those ranges. We note that Fiquet et al. 1996 did not include a thermal pressure term in their experiments and so it is likely that the total pressure is underestimated. Moreover, the reported temperatures are likely overestimated by about 20 to 50%. We trained another network excluding these data in our training set and access the prediction performance (Figure 3, bottom panel). In doing so, MgO volumes are resolved within the prior range of experimental data, also in the region of low pressure and high temperature. This shows the networks' ability to capture the

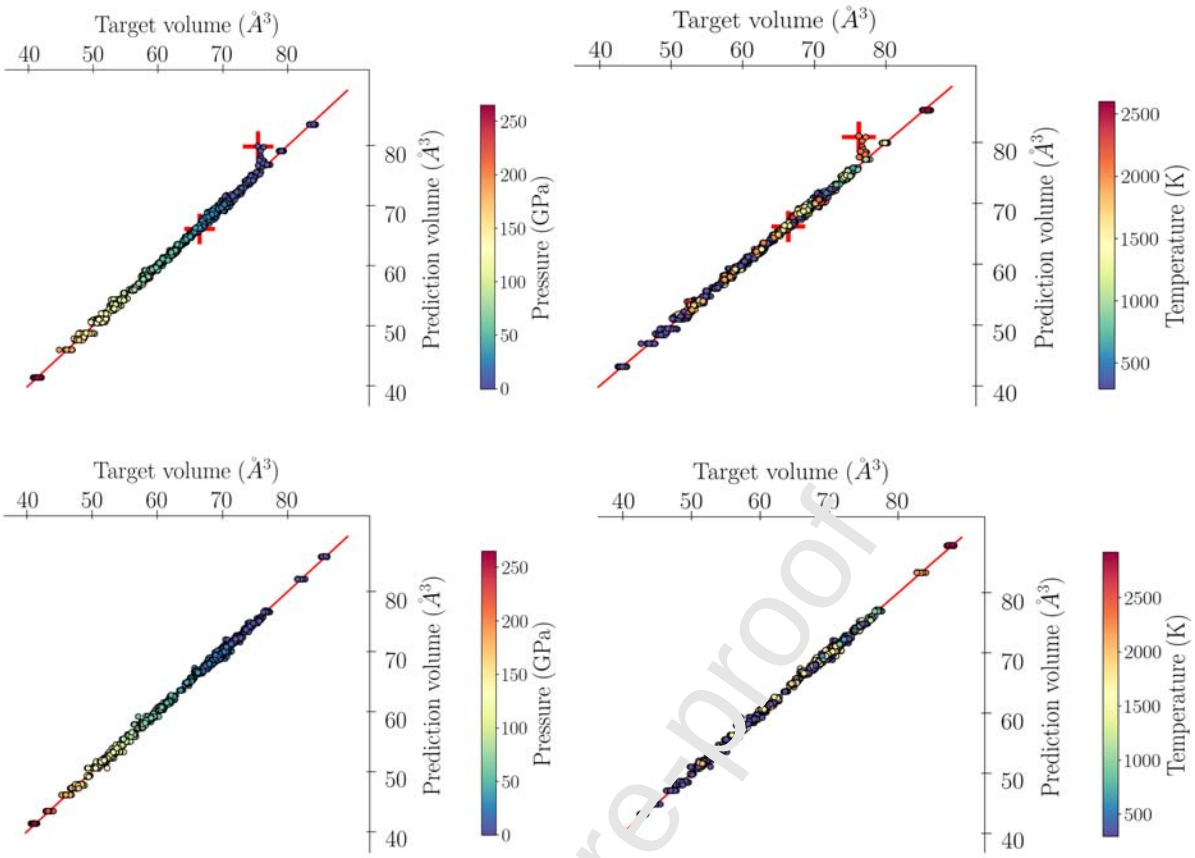


Figure 3: Performance of MDNs. Target volumes from the test data set are compared with mean volumes (Equation 6) predicted by the MDNs. Top panel shows mean volumes predicted by the MDNs trained with all experimental data while bottom shows results with high temperature data of Fiquet et al. 1996 and Murakami et al. 2012 excluded (also see Subsection 4.2). The pressure (left) and temperature (right) range of the test data set is shown by colourbars on both panels. We note that the solid red line in the Figure refers to a perfectly resolved network prediction. Points located near this line are well resolved and those located away represent more uncertain volume predictions. The MDNs best predict the volumes in low temperature regions and at simultaneous high temperature and pressure. However, including high temperature data of Fiquet et al. 1996 into training provides more uncertain volume predictions in the low pressure, high temperature region. For two data points marked with "+" in both left and right plots in the top panel, we plot posterior pdfs for volume in Figure 4. One datum is located in the low pressure, high temperature region where the effect of high temperature data from Fiquet et al. 1996 is significant and another away from it.

underlying data consistency.

Low pressure data (approximately less than 30 GPa) are relatively dense up to about 1400 K compared to higher temperatures. Similarly, most of the high pressure data, i.e. extending to the lower mantle environment, come either from approximately between 1500 K to 2700 K or from ambient temperature measurements. Besides that, the experimental data doesn't cover simultaneous high temperature and high pressure regions, for example temperatures greater

than ∼2700 K at pressures expected near the bottom of lower mantle. Hence, we expect wider posterior probability density functions for volume in regions of sparse experimental data coverage.

So far we have only shown the mean of the posterior pdf for volume. To illustrate more clearly the effect of the high temperature data of Fiquet et al. 1996 on the posterior pdf at low pressure, high temperature, we take two data points from the test set (denoted by '+' in Figure 3, top panel). Both points are drawn at low pressures, but one is at high temperature and located away from the solid line and another at low temperature is close to it. In Figure 4 posterior pdfs at those points are shown. They show a more uncertain prediction for the high temperature, low pressure input. Once we remove Fiquet et al. 1996 data from training (see Sub-section 4.2), the network predicts narrow posterior pdfs showing less uncertainty (cf. including those in training) in volume. Although excluding Fiquet et al. 1996 provides less uncertain volume predictions, due to limited availability of experimental data at high temperature and low pressure (approximately >1500 K and < 25 GPa) the predicted posterior pdfs are still slightly wider than at similar temperatures and high pressures (also see Sub-section 4.2 and Appendix C.1).

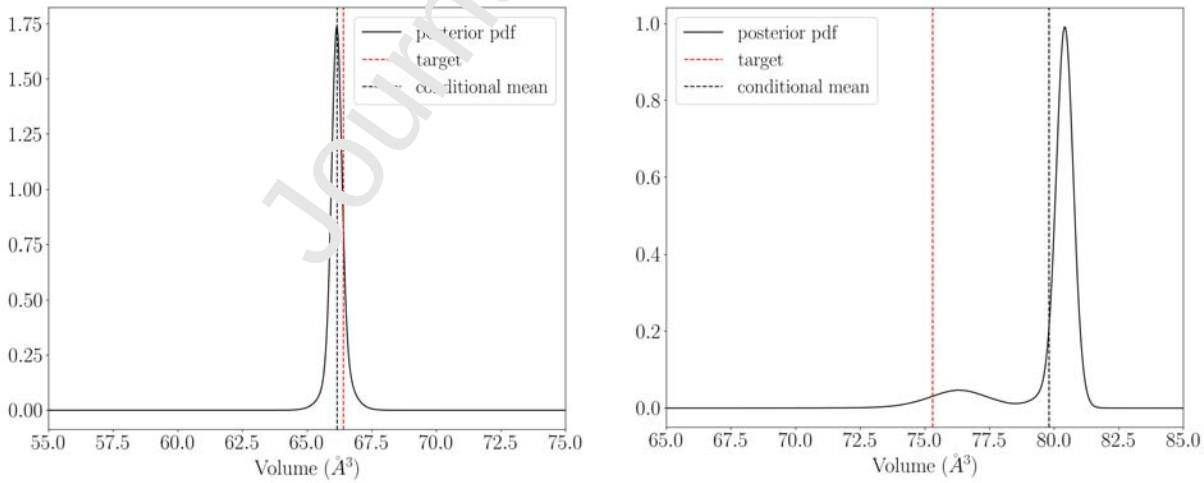


Figure 4: Posterior pdfs for MgO volume (solid curves) for two data points from top panel of Figure 3 together with their target values (red dashed line) and conditional mean volume (black dashed line). Left: inputs are 24.86 GPa and 300.19 K. The posterior pdf is narrow and uni-modal with the posterior mode located close to the target value. Right: inputs are 1.36 GPa and 2116.03 K. The posterior pdf is broad and multi-modal with target volume located away from the posterior modes. The smaller peak is the due to experimental P-V-T data of Fiquet et al. 1996.

4. MDN predicted material properties

4.1. P-V relationship at 300 K

The predicted pdfs for volume along a 300 K isotherm are presented in Figure 5. A subset of the training data (i.e. only around 300 K temperature) is also shown along with the MDN predictions. The uncertainty in volume increases with pressure as shown by the increasing width of pdfs. This is expected as the training data (around 300 K) are more consistent with each other at lower pressures.

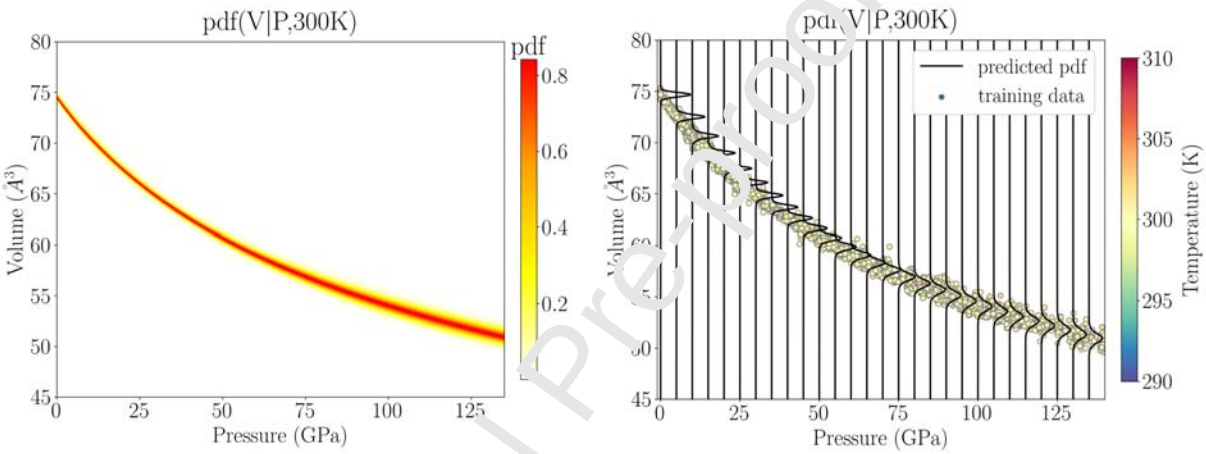


Figure 5: The predicted pdf by the MDNs for volume of MgO along a 300 K isotherm. Left: pdf for volume up to lower mantle pressures is shown as a continuous function of pressure. The colour scale shows the value of the probability density function. Right: pdfs on volume are shown at 5 GPa pressure intervals together with training data around 300 K (shown as circles in the background). The training data show less variation at low pressures which results in narrower pdfs compared to high pressures.

In Figure 6 we compare pdfs for the volume of MgO along a 300 K isotherm with EOSs of Tange et al. 2009, Speziale et al. 2001, Stixrude and Lithgow-Bertelloni 2005, 2011 and Dorogokupets and Dewaele 2007 (denoted as T09, S01, SLB0511 and DD07, respectively). In this study, we use MINUTI (Sturhahn 2020) to compute volume, bulk modulus and thermal expansivity as a function of pressure (and temperature) from these EOSs. For ambient temperature comparisons, static equations (i.e. third-order finite strain or Vinet) together with respective fitting parameters (V_0, K_{0T} and K'_{0T}) as reported in the literature are used. We show the pdfs for volume (Figure 6, left panel) at every 5 GPa. The EOSs diverge as the pressure increases. At 135 GPa, the difference in volume between the equations of state

of Stixrude and Lithgow-Bertelloni 2005, 2011 and Tange et al. 2009 is ∼0.68 $\AA^3$, whereas one standard deviation predicted by the neural networks is ±0.54 $\AA^3$. Moreover, the slope of each individual EOS differs. This can best be visualized by computing $\frac{\partial P}{\partial V}$ for all EOSs (see Figure 6, right panel). Although Speziale et al. 2001 and Stixrude and Lithgow-Bertelloni 2005, 2011 are based on third order Birch-Murnaghan EOSs, their fitting parameters are different. Comparisons between different EOSs and their fitting parameters are given by other studies (e.g. Dorogokupets and Dewaele 2007, Tange et al. 2009, Ye et al. 2017, etc.). The mean slope predicted by the neural network shows a slightly stiffer EOS compared to the "standard" EOSs from the literature. This may be due to the fact that our training data include experiments which make use of different pressure standards (e.g. Ruby, NaCl, Pt, Au) than the EOSs considered for comparison (which are based on MgO). Nevertheless, such a difference in slope together with the volume difference will inevitably lead to a significant divergence in the inferred compressibility and thermal expansivity (see Section 5).

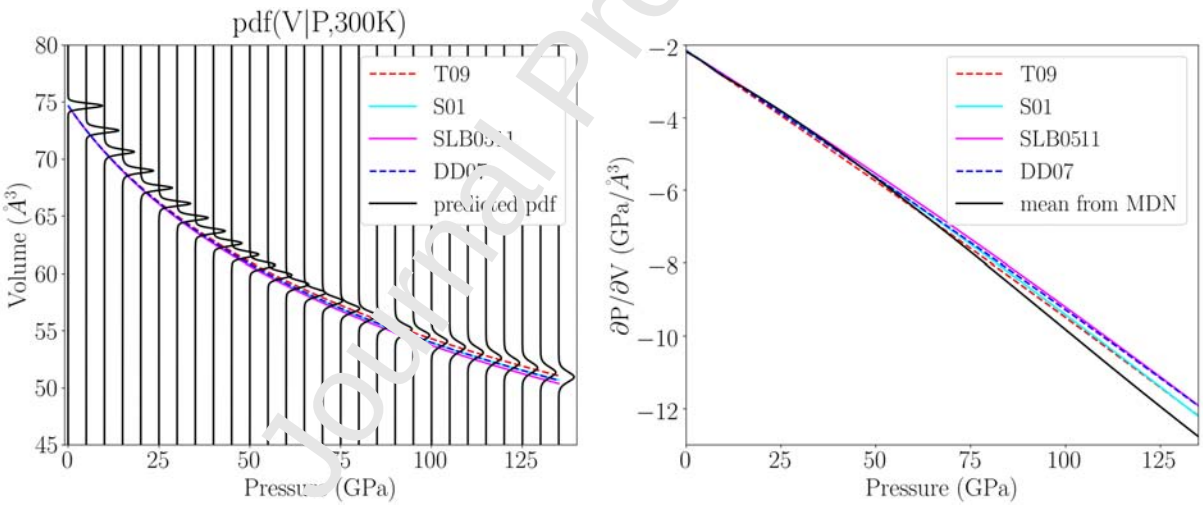


Figure 6: Left: our predicted pdfs for volume of MgO along a 300 K isotherm (black lines) compared with previously published EOSs (Tange et al. 2009, Speziale et al. 2001, Stixrude and Lithgow-Bertelloni 2005, 2011 and Dorogokupets and Dewaele 2007) (coloured lines). Pdfs for volume are shown at 5 GPa pressure intervals. Right: $\frac{\partial P}{\partial V}$ of MgO EOSs from the left panel. For this computation, we take the mean (Equation 6) of the output posterior on volume at every 0.1 GPa interval. The divergence between different EOSs increases with pressure.

4.2. High temperature P-V-T relationships

We use the trained MDNs to predict volumes of MgO at different temperatures. As an example, we plot the predicted pdfs for volume along a 2500 K isotherm in Figure 7, left panel (other isotherms are provided in Appendix C.1). Similar to the ambient temperature (Subsection 4.1), the 2500 K isotherm shows a well-constrained volume prediction at lower mantle pressures. However, the high temperature pdfs show more uncertain volume predictions at low pressures (except at 0 GPa). For example, at 5 GPa the pdf is relatively wide and bimodal compared to that at high pressures (e.g. 100 GPa) which is unimodal. As discussed earlier in Section 3.3, high temperature experimental data of Fiquet et al. 1996 do not include a thermal pressure term, and it is likely the total pressure is underestimated. This can be visualised in Figure 7, left panel, where training data points located approximately between 5-15 GPa have a smaller volume compared to data around 20 GPa and ∼2500 K. We train another network without the high temperature data of Fiquet et al. 1996 and plot the results on the right panel of Figure 7. The posterior pdf for volume at 5 GPa now shows a unimodal peak and the width is decreased by approximately a factor of 2 (cf. left panel at 5 GPa). Although removing Fiquet et al. 1996 reduces the uncertainties in volume, the posterior pdf is still wider than at high pressures for the same temperature. This region of low pressure, high temperature is known to be dominated by anharmonic effects. Although these effects are implicitly represented in our volume pdfs, there are limited experimental data in this region (temperature >1500 K and pressure <25 GPa) to further constrain them.

We compare the MDN predicted pdfs along a 2500 K isotherm (Figure 7) with some conventional EOSs (Tange et al. 2009, Speziale et al. 2001, Stixrude and Lithgow-Bertelloni 2005, 2011 and Dorogokupets and Dewaele 2007). The variation in volume between these EOSs at high pressures is similar to that observed at 300 K. It has been noted in earlier studies (e.g. Ye et al. 2017) that the discrepancies in high temperature EOSs are partly due to persistence of the disagreement between them at 300 K (reference isotherm). Furthermore, at low pressure (<25 GPa) Speziale et al. 2001 diverges from other EOSs. This deviation is

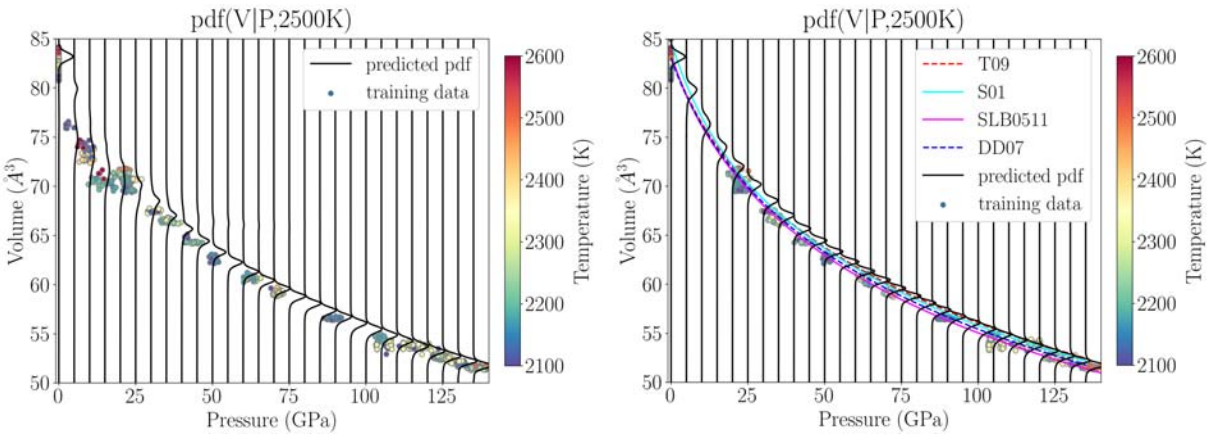


Figure 7: Left: pdfs for volume of MgO along a 2500 K isotherm predicted by MDNs trained with all data. Right: same as left but Fiquet et al. 1996 and Murakami et al. 2012 data are excluded. For comparison, volumes along the high temperature isotherm for some previously published EOSs (Tange et al. 2009, Speziale et al. 2001, Stixrude and Lithgow-Bertelloni 2005, 2011 and Dorogokupets and Dewaele 2007) are computed using MINUTI (Sturhahn 2020). On both panels we plot a sub-set of the total training data, namely those data at temperatures between 2100 and 2600 K. Excluding Fiquet et al. 1996 data from neural network training significantly reduces the width of the pdfs at high temperature and low pressure.

likely due to different values of fitting parameters together with distinct Grüneisen models to compute the thermal behavior. For example, Speziale et al. 2001 do not consider anharmonic effects, and their ambient Grüneisen parameters are also different than other studies (see e.g. Ye et al. 2017, Dorogokupets and Dewaele 2007). Besides that, as with the case of the

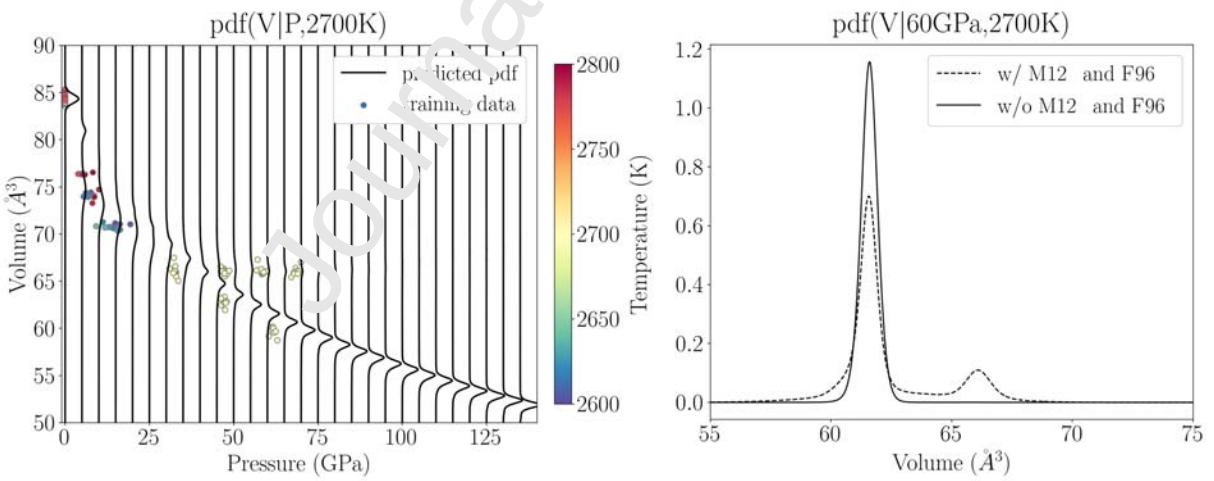


Figure 8: Left: pdfs for volume of MgO along a 2700 K isotherm predicted by the MDNs trained with all data. We also plot a sub-set of the training data, namely those whose temperatures lie between 2600 and 2800 K. Note: the large uncertainty in volume in the low pressure region (approximately below 25 GPa) is due to inclusion of data from Fiquet et al. 1996 as discussed in the text. Right: Comparison of posterior pdfs for volume predicted by MDNs trained with and without Murakami et al. 2012 (M12) and Fiquet et al. 1996 (F96) data at 2700 K and 60 GPa. The small peak at around 66 $\AA^3$ is due to Murakami et al. 2012 data.

300 K isotherm, all explicit EOSs lie within the uncertainty range predicted by our MDNs, which is expected because some training data come from the MgO pressure scales described by these EOSs.

At 2700 K, the MDN predicted pdfs (Figure 8) show bimodal volumes in the pressure range of approximately 45-90 GPa. Once we plot the associated training data on top, it becomes clear that the smaller peaks in the pdfs are the representation of experimental data points of Murakami et al. 2012. Surprisingly, for the same reported volume and temperature they report pressures which are different from each other by about 36 GPa. However, their reported densities appear to be physically reasonable. Nevertheless, we train another network to discriminate how much uncertainty is coming from those specific data points. In doing so, the posterior becomes unimodal. At 60 GPa, including Murakami et al. 2012 data leads to a factor of approximately 3.5 wider pdfs for volume (Figure 8, right panel) compared to results without those data. However, the effect of those data points seems to be local in P-V-T space and their influence decreases for example, at higher pressures. This is because MDNs interpolate locally in between samples, and data in one region of P-T space doesn't influence uncertainties everywhere.

5. Bulk modulus and thermal expansivity

Since the training data do not contain explicit values for the volume derivatives with respect to the inputs (P and T), getting constraints on bulk modulus ($-V\frac{\partial P}{\partial V}$) and thermal expansivity ($\frac{1}{V}\frac{\partial V}{\partial T}$) is less straightforward than constraining the volumes. Hence, we follow a slightly different approach compared to volume. We calculate the mean volume using Equation 6 for any given P and T from each earlier obtained MDN. Then we perturb pressure ($P+\delta P$) while keeping the temperature fixed and compute the mean volume ($< V(P+\delta P, T) >$) for that pressure from the same MDN. This way, we can compute the mean isothermal bulk modulus (K) as shown in Equation 7. Similarly, we evaluate mean volumes for two slightly different temperatures but at a fixed pressure, and use that to compute the thermal expansivity, α

(Equation 8). For numerical differentiation, we use $\delta P = 0.1$ GPa and $\delta T = 1$ K. Using a different value for ∂P or ∂T provides similar results.

$$< K|P,T;\boldsymbol{\alpha} > \quad = \quad < -V(P,T) > \frac{\delta P}{< V(P+\delta P,T) > - < V(P,T) >} \tag{7}$$

$$< \alpha|P,T;\boldsymbol{\alpha} > \quad = \quad \frac{1}{< V(P,T) >} \frac{< V(P,T+\delta T) > - < V(P,T) >}{\delta T} \tag{8}$$

Hence, in this approach, we take the derivatives of the P-V (or T-V) curve defined by the mean of the posterior pdfs from each neural network rather than fitting P-V-T data to a predefined EOS to get fitting parameters (such as K_{0T} and K'_{0T}). Since we have trained a large number of MDNs (10^3) to predict the posterior pdf for volume, we get the same number of mean isothermal bulk modulus and thermal expansivity values. This way, each neural network approximates a slightly different mapping and its derivatives, and the distribution on the mean bulk modulus and thermal expansivity can approximate the uncertainties on them. Moreover, we use the same networks to compute the pdfs for volume and the mean volumes; the volume that goes into the calculation of bulk modulus and thermal expansivity is therefore consistent.

As an example, Figure 9 shows bulk modulus as a function of pressure along two selected isotherms (refer to Appendix C.2 for other isotherms). The bulk modulus predicted by neural networks shows a higher value at high pressure along the 300 K isotherm compared to conventional EOSs. As mentioned earlier, this is likely due to the fact that the training data come from experiments which make use of different EOSs and pressure standards than those (MgO based) EOSs considered for comparison. Moreover, the fitting parameters (V_0, K_{0T} and K'_{0T}) are different for different EOSs. Hence, although these EOSs predict volume within the uncertainty range predicted by MDNs (Figure 6, left panel), their derivatives (Figure 6, right panel) differ significantly from each other and also from the MDN prediction, leading to different values of bulk modulus.

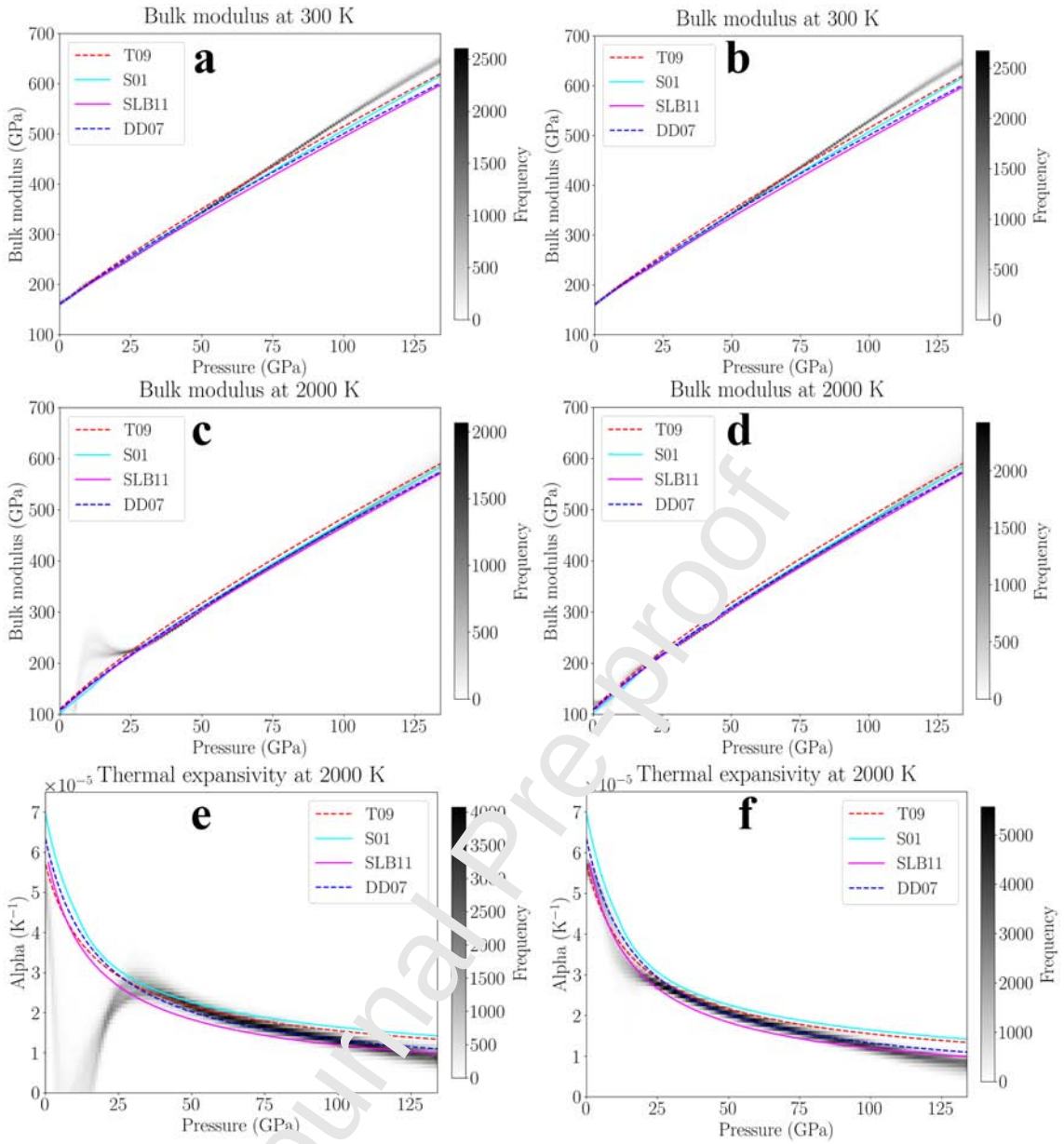


Figure 9: Comparison of the mean bulk modulus (a, b, c and d) and thermal expansivity (e and f) predicted by the neural networks with previously published equations of state for MgO (Tange et al. 2009, Speziale et al. 2001, Stixrude and Lithgow-Bertelloni 2011 and Dorogokupets and Dewaele 2007) as a function of pressure. The output from the neural networks is shown with greyscale- the darker the region of the plot, the greater the number of MDNs which predict the bulk modulus (or thermal expansivity) has that value. Frequency counts for output from the MDNs are at intervals of 1 GPa for pressure and bulk modulus, and 10^{-7} K^{-1} for thermal expansivity. For (a), (c) and (e) neural networks are trained with all collected data, whereas for (b), (d) and (f) data from Fiquet et al. 1996 and Murakami et al. 2012 have been excluded. Due to the inclusion of Fiquet et al. 1996 data we obtain large uncertainties in bulk modulus and thermal expansivity in low pressure, high temperature regions. Note: the overlapping of different EOSs makes the background histogram difficult to visualise.

One high temperature (2000 K) comparison between the neural network predicted mineral properties and other studies is shown in Figure 9- c, d, e and f. In general, bulk modulus

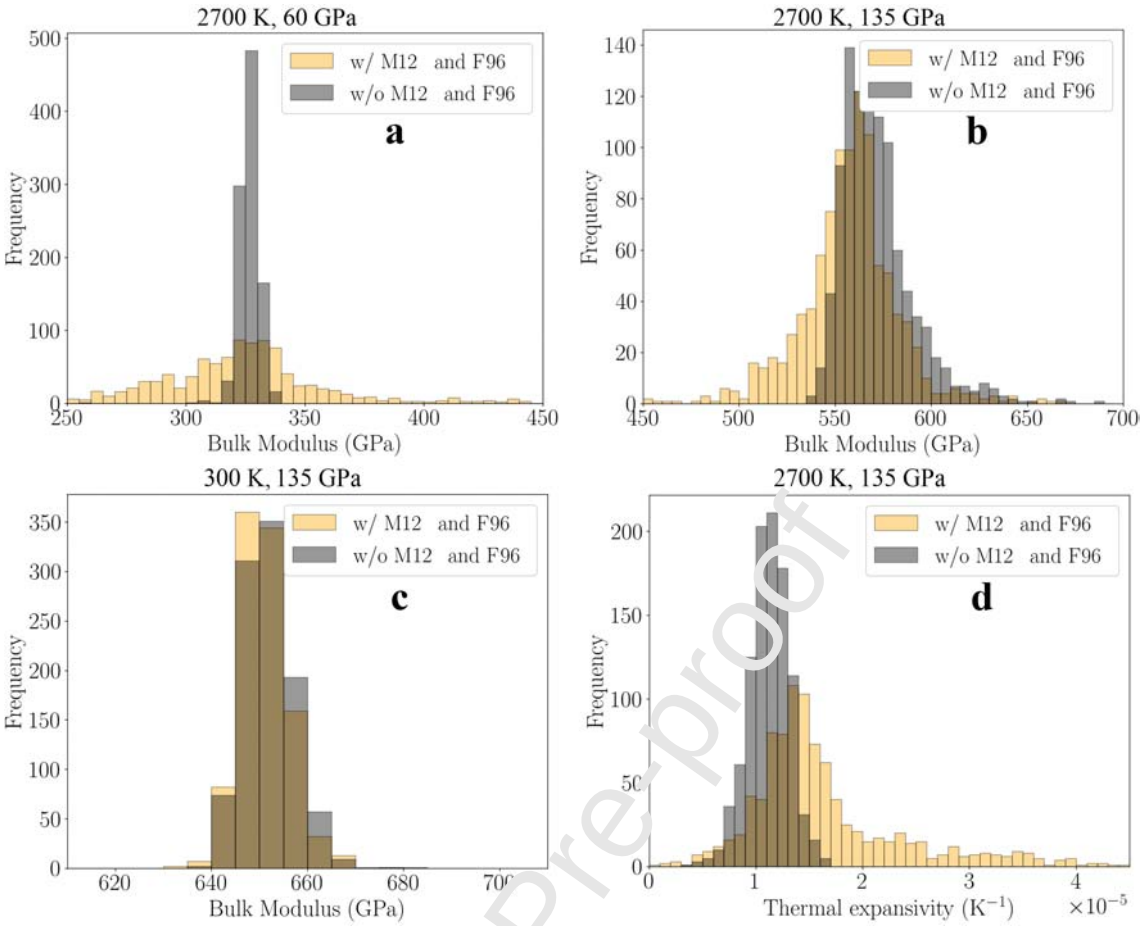


Figure 10: Comparison of the MDN predicted mean bulk modulus at (a) 2700 K, 60 GPa, (b) 2700 K, 135 GPa, (c) 300 K, 135 GPa and (d) thermal expansivity at 2700 K, 135 GPa of MgO trained with and without Murakami et al. 2012 (M12) and Fiquet et al. 1996 (F96). The effect of Murakami et al. 2012 data on bulk modulus and thermal expansivity is mainly around 2700 K, and it gradually reduces as pressure decreases or increases outside the interval approximately 45-90 GPa.

values predicted by the neural networks agree well with explicit EOSs, although Tange et al. 2009 shows slightly higher values at moderate pressures (e.g. 60 GPa). The mean bulk modulus predicted by the neural networks shows a large uncertainty at low pressures (below $\sim$ 25 GPa) when high temperature data by Fiquet et al. 1996 are included. In Figure 9- d, we show the bulk modulus predicted by the neural network trained without Fiquet et al. 1996 (and Murakami et al. 2012). Here, the uncertainties at low pressure are significantly decreased. Similarly, neural networks trained without those two data sets predict physically reasonable thermal expansivities (Figure 9- f) compared to those trained with all data sets (Figure 9- e). At high temperatures, we still see a sharp bend around 20 GPa (also see Appendix C) which we suggest may be related to anharmonic effects. As the

experimental data is relatively sparse in this region, one would need additional measurements (or theoretical studies) to confirm this. Furthermore, the thermal expansivity of Speziale et al. 2001 deviates from other EOSs. As mentioned in earlier studies (e.g. Dorogokupets and Dewaele 2007), this may be improved by including anharmonic terms in the EOS. In equation of state formalisms, one can add an anharmonic term to the total free energy. This additional term has a T^2 dependence, rather than simply a linear temperature term. The effect of adding this term is most significant at low pressures, and can potentially capture more accurately the volume dependence at high temperatures compared with the standard thermal models without anharmonicity (for temperatures less than or equal to 2700 K in this meta dataset).

Besides low pressure, including Murakami et al. 2012 data during network training provides mean bulk modulus uncertainties that are more than 4 times larger (Figure 10- a) than excluding them together with Fiquet et al. 1996, and this discrepancy reduces at higher pressures (Figure 10- b). Moreover, as expected, neither Fiquet et al. 1996 nor Murakami et al. 2012 data influence bulk modulus at low temperatures, as shown in Figures 10- c and 9- a, b.

6. Discussion

Fitting parameters (such as K_{0T} and K'_{0T}) are inherent to explicit global EOSs, and a correlation between them tells us how one parameter changes with another providing optimal global fit. We do not estimate the uncertainties on fit parameters of EOSs which are specific to the underlying global functional form. Instead, we directly provide the uncertainties on volumes which are local in P-T space. The MDN is a kernel based method where we fit (a mixture of Gaussian) kernels to the experimental data and get an arbitrary probability density function on volume at any given P and T. The neural networks are flexible and interpolate locally; the uncertainties in one region of P-T space don't impact the posterior pdf everywhere. For example, Figure 7 shows no change in high pressure pdfs while removing

Fiquet et al. 1996 data in the region of low pressures. Our approach is also very powerful at identifying data inconsistencies when using different data sources.

The posterior pdfs given by the MDNs represent uncertainties in volume due to experimental errors, data gaps and data inconsistencies from different studies. Together with the uncertainties in mean isothermal bulk modulus and thermal expansivity, these results can be used by, for example, seismologists working on thermochemical interpretation of seismic data. Although uncertainties in volume, bulk modulus and thermal expansivity vary locally depending on sparsity and consistency of the experimental data, using these outputs from MDNs, one can directly compute bulk wave speed ($\phi^2 = K_S/\rho$) and density (ρ) at any given pressure and temperature. However, in order to compute bulk wave speeds at temperatures applicable to the lower mantle, we need the adiabatic bulk modulus ($K_S = K_T(1 + \alpha\gamma T)$), where γ is Grüneisen parameter and α is the thermal expansivity. Nevertheless, assuming that the difference between isothermal (K_T) and adiabatic (K_S) bulk moduli, at 300 K is roughly within ±1.0% (Marquardt et al. 2018), the bulk wave speed of MgO is 11.14±0.07 km/s at 135 GPa. At the same condition, the relative uncertainty (one standard deviation around mean) in density predicted by the MDNs is about ±1.0%. This is larger than or comparable to the relative density variations in lower mantle estimated by previous studies (e.g. Ishii and Tromp 1999, Trampert et al. 2004, Koelemeijer et al. 2017). Although the Grüneisen parameter varies as a function of volume that ultimately depends on pressure (and temperature), we assume it to be approximately 1.1±0.3 (e.g. Stixrude and Lithgow-Bertelloni 2011, Ye et al. 2017) at 2700 K and 135 GPa to give an estimate of uncertainties in bulk wave speed. In doing so, the relative uncertainty in bulk wave speed is about ±1.77% which is larger than the reported bulk sound speed variation in the lower mantle (e.g. Trampert et al. 2004).

Estimation of mineral properties beyond the range of experimental data requires extrapolation. The standard EOSs can easily be used for extrapolation provided that the assumptions of the functional form hold in the region of no data. In general, it has been observed that

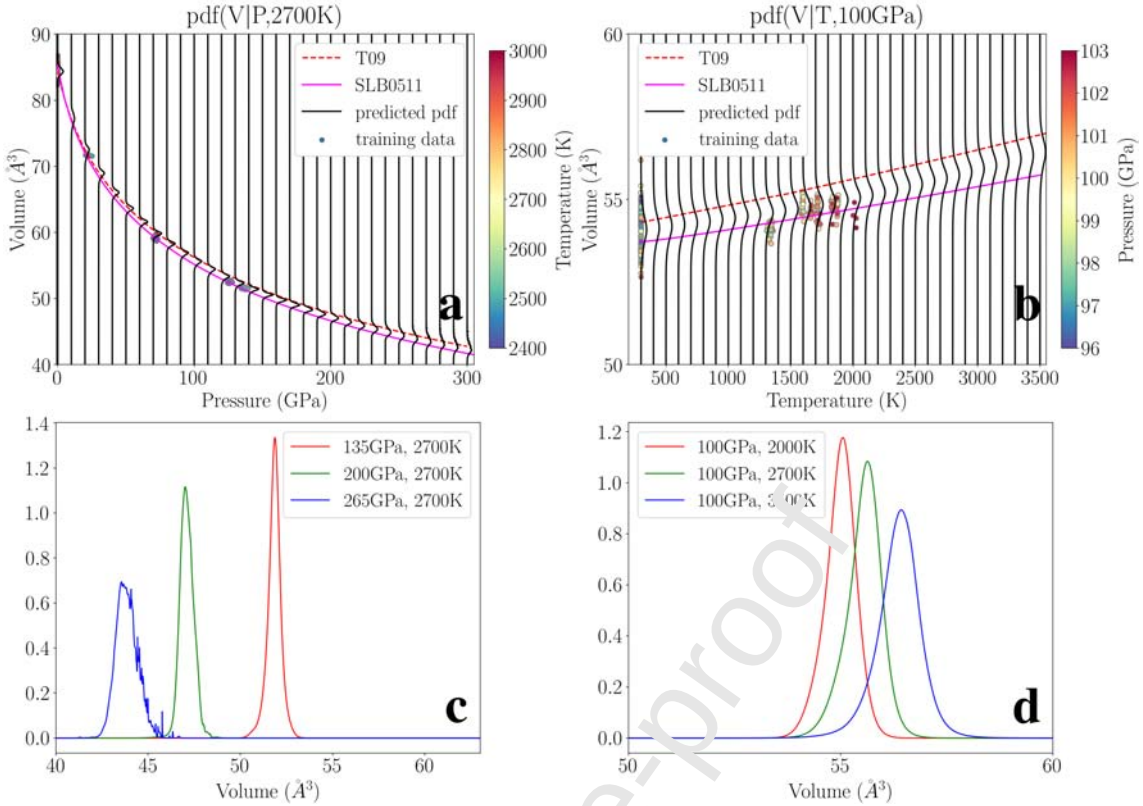


Figure 11: Probability density function for volume of MgO along a 2700 K isotherm (a) and 100 GPa isobar (b). Training data belonging to temperature between 2400 and 3000 K (a), and pressure range from 96 to 103 GPa (b) are also shown. Magenta (SLB0511) and red (T09) curves are Stixrude and Lithgow-Bertelloni 2005, 2011 and Tange et al. 2009 EOS, respectively. They follow the volume trend predicted by the network. In the region outside the prior data, the trained MDNs provide wider pdfs as they are forced to extrapolate the volume. To illustrate this more clearly, volume pdfs at a fixed temperature (and pressure) and three different pressure (and temperature) are also shown in c (and d).

MDNs provide a wider estimate of uncertainties in the region of little to no training data (Käufl et al. 2016a). Here too, as shown by the wider pdfs in Figure 11, the uncertainty in predicted mineral properties increases when the network has to extrapolate from distant training data. We note that EOSs of Stixrude and Lithgow-Bertelloni 2005, 2011 and Tange et al. 2009 closely follow the pdf predicted by the network indicating that it learns a functional form present in the data, but errs on the cautious side by returning larger uncertainties. From a Bayesian perspective, we would advise against extrapolation as this covers a region outside the prior. Figure 11, however, demonstrates some capability of neural networks to extrapolate beyond the ranges of the data, although we would need to establish how far this is related to the precise network architecture.

The shear modulus is required to calculate compressional and shear wave speeds. There is no thermodynamic expression for the shear modulus, but functional forms are often assumed, for example third order finite-strain and shear counterpart of the Keane EOSs (Keane 1954) by Kennett 2017, to compute the shear modulus which are based on the bulk modulus calculation. One can also use the linear relationship among shear modulus, adiabatic bulk modulus and pressure given by Stacey 1995. However, the uncertainties in shear modulus would then be dependent on those in bulk modulus, and the assumption that shear properties can be constrained from the bulk properties. An alternative is to use data from experiments such as Brillouin Spectroscopy that provide shear wave speed information. Together with unit-cell volume, as measured by X-ray diffraction on the same sample (e.g. Murakami et al. 2012, Kurnosov et al. 2017) and known sample composition, the density and thus shear moduli can be determined. However, these data sets do not cover simultaneous high pressure and temperature regions that are expected in the Earth's lowermost mantle. For example, the highest temperature and pressure data for MgO reported in Murakami et al. 2012 are six measurements at 2700 K and between 32.5-68.4 GPa. Nevertheless, a combination of wave speed data from ultrasonic techniques and Brillouin Spectroscopy together with high P-V-T data from x-ray diffraction techniques has the potential to exhaustively sample the lower mantle geotherm in the near future (Marquardt and Thomson 2020).

We note that, in principle, a combination of experimental data and theoretical calculations (e.g. Karki et al. 1999, Oganov and Dorogokupets 2003, Wu et al. 2008) is possible. This may provide additional constraints on the predicted mineral properties covering a wider range of pressure and temperature. Since our approach implicitly identifies the consistency between different data sources, a proper rationale can be developed to mix data and uncertainties from theory with experiments. Furthermore, the MDN based approach can easily be extended to the upper mantle and the core. Since MDNs are flexible, they can be employed to model multi-mode targets/outputs. This would be helpful to model for example volume anomalies induced by the iron spin transition (e.g. Marquardt et al. 2009, Speziale et al. 2007, Lin et al. 2006, Crowhurst et al. 2008, Solomatova et al. 2016). A natural progression of this

work is to extend it for solid solution. It is straightforward to include composition, e.g. the Mg/Fe ratio, by including it as an extra dimension in the input data (i.e. P, T and mol% Fe in ferropericlase) provided there is enough training data.

7. Conclusions

This study demonstrates the feasibility of a neural network based approach to infer the material properties of lower mantle minerals. In our approach, we learn the underlying P-V-T relationship providing a reasonable approximation of the P-V-T data of MgO. This allows us to compute the uncertainties in density, thermal expansivity and bulk modulus without prescribing an explicit EOS. Once the networks are trained, it is a simple function that can be evaluated at any given pressure and temperature to get volume, mean bulk modulus and thermal expansivity with uncertainties. In order to train the networks, we collect data from high P-V-T experiments without prior selection of data (e.g. based on pressure scale or functional form used). Hence, our uncertainties are not biased towards a subjective selection of experimental data. Furthermore, our approach identifies inconsistencies between data from different sources. The assumption that an EOS follows a particular form provides a priori information by fixing their form (or thermodynamic model) and/or pressure scale. It remains to be determined which EOS form best describes the thermodynamic behaviour of MgO at wide range of pressures and temperatures. In this study, we compare a few "standard" EOSs with the material properties inferred from neural networks and show that choosing one particular explicit form provides a biased estimate of uncertainties.

Based on the prediction performance of the MDNs and comparison with conventional EOSs (such as Figures 3, 7, 9, and Appendix C), we can be most confident about physical interpretation of seismic data in the lower mantle within the prior range of experimental data (Figure 2). In the regions where there exists little evidence about how the P-V-T relationship behaves, such as at low pressure, high temperature (<25 GPa, >1500 K), and temperatures approximately >2700 K at pressures expected towards the core-mantle boundary, neural

networks show increasingly uncertain predictions. Although for the Earth's lower mantle, low pressure and high temperature environments may not be relevant, they are expected in other planetary bodies such as the Moon and Mars (e.g. Khan et al. 2014, 2018). With currently available data, it likely provides meaningful uncertainties that could be used by seismologists within certain ranges of pressure and temperature, while highlighting the P, T regions in which more experimental (or theoretical) data is needed before we can draw robust conclusions on temperature and composition.

Acknowledgements

We would like to thank an anonymous reviewer for helpful comments which improved the manuscript. AR and LC received funding from the Dutch Research Council (NWO) on grant number 016.Vidi.171.022. JMJ is thankful for discussions with Wolfgang Sturhahn and for support of this research by the National Science Foundation's Collaborative Studies of the Earth's Deep Interior (EAR-1161046 and EAR-2009735).

Appendices

Appendix A. Generalised theory of the MDN

Let, $\boldsymbol{x} = \{x_1, x_2, ..., x_I\}$ be the input data to the feed-forward part of the MDN. Please note, to generalise this section, we write inputs as $\boldsymbol{x}$ and targets as m_k instead of P & T and V, respectively. The feed-forward network outputs y_k are computed as a weighted sum of the outputs from the hidden nodes plus a bias

$$y_k = f_2\left(\sum_{j=1}^{J} \alpha_{jk} h_j + \alpha_{0k}\right) \tag{A.1}$$

where the function f_2 is an identity function such that $f_2(p) = p$, α_{jk} is the hidden layer weight matrix and α_{0k} represents a bias term of each output node. Now, the hidden node

outputs h_j are computed as

$$h_j = f_1\left(\sum_{i=1}^{I} \alpha_{ij} x_i + \alpha_{0j}\right) \tag{A.2}$$

where the function f_1 is a logistic sigmoid function $f_1(p) = \frac{1}{1+exp(-p)}$, α_{ij} is the input layer weight matrix, α_{0j} are the biases of hidden nodes and x_i are input data. y_k are related to the parameters, namely weights (π_n), means (μ_n) and standard deviations (σ_n) of Gaussians in the Gaussian Mixture Model (GMM) by the following relationship (for details see e.g. Bishop 1994, de Wit et al. 2013)

$$\pi_n(\boldsymbol{x};\boldsymbol{\alpha}) = \frac{exp\left(y_k^{(\pi)}(\boldsymbol{x};\boldsymbol{\alpha})\right)}{\sum_{n=1}^{M} exp\left(y_k^{(\pi)}(\boldsymbol{x};\boldsymbol{\alpha})\right)}, \tag{A.3}$$

$$\mu_n(\boldsymbol{x};\boldsymbol{\alpha}) = y_k^{(\mu)}(\boldsymbol{x};\boldsymbol{\alpha}) \text{ and} \tag{A.4}$$

$$\sigma_n(\boldsymbol{x};\boldsymbol{\alpha}) = exp\left(y_k^{(\sigma)}(\boldsymbol{x};\boldsymbol{\alpha})\right). \tag{A.5}$$

Appendix B. MDN initialization and training details

The total data ($\boldsymbol{x}$) is divided into three sets- training (70%), monitoring (20%) and test (10%) sets such that

$$\boldsymbol{x}^{train} \subset \boldsymbol{x},\ \boldsymbol{x}^{monitor} \subset \boldsymbol{x}\ and\ \boldsymbol{x}^{test} \subset \boldsymbol{x} \tag{B.1}$$

with $\boldsymbol{x}^{train} \cap \boldsymbol{x}^{monitor} = \emptyset$, $\boldsymbol{x}^{train} \cap \boldsymbol{x}^{test} = \emptyset$ and $\boldsymbol{x}^{monitor} \cap \boldsymbol{x}^{test} = \emptyset$. Using the training data ($\boldsymbol{x}^{train}$) we train the MDN. However, before we train the MDN we need to decide on initial values of the network parameters. We randomly draw the input layer and hidden layer weights (Bishop 1995) according to the following Gaussian distributions

$$\alpha_{ij} \sim \mathcal{N}\left(0, \frac{1}{I+1}\right) \tag{B.2}$$

and

$$\alpha_{jk} \sim \mathcal{N}\Big(0, \frac{1}{J+1}\Big), \tag{B.3}$$

respectively. Where I and J are number of input and hidden nodes, respectively. Similarly, the output layer biases are initialized by a K-means clustering algorithm (i.e. fitting a GMM to the training data set). Once the initialization is done and the training begins, the difference between the output and the target can be computed according to the error function

$$E^{train} = \sum_{train} -\ln\Big(p(m_k|\boldsymbol{x}^{train};\boldsymbol{\alpha})\Big) \tag{B.4}$$

which is summed over all training data providing the average error. This function is also called the loss function which is minimized iteratively using the ADAM optimization method (see detailed algorithm in Kingma and Ba 2014). The explicit dependence of output posterior on the network parameters (see Käufl et al. 2016a and references therein) can be avoided by using multiple MDNs and combining them by weighted sum. The weight of each MDN is determined by the test set error as

$$w_i = \exp\Big(-\frac{E^{test}(\boldsymbol{x}^{test}, \boldsymbol{\alpha}_i)}{N}\Big) \tag{B.5}$$

where index i denotes the i-th MDN (C MDNs in total) and N is the size of the test data set, and the MDNs are combined according to

$$p(m_k|\boldsymbol{x};\boldsymbol{\alpha}) = \sum_{i=1}^{C} \frac{w_i}{\sum_j w_j} p_i(m_k|\boldsymbol{x};\boldsymbol{\alpha}_i). \tag{B.6}$$

Appendix C. Mineral properties

Appendix C.1. P-V-T EOS

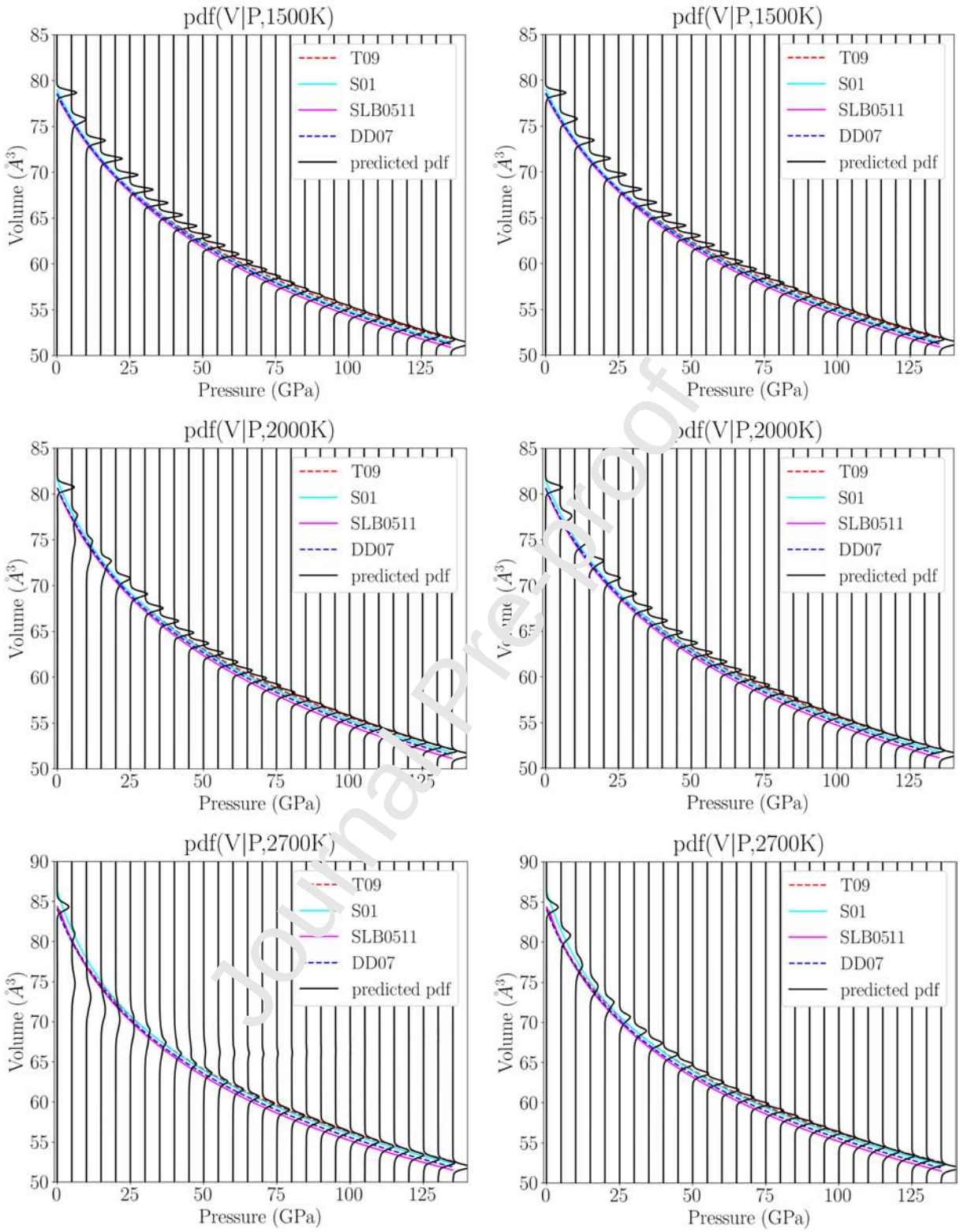


Figure Appendix C.1: P-V relationship of MgO predicted by MDNs trained with (left) all data and (right) excluding Murakami et al. 2012 and Fiquet et al. 1996. Comparison with previously published EOSs (Tange et al. 2009, Speziale et al. 2001, Stixrude and Lithgow-Bertelloni 2005, 2011 and Dorogokupets and Dewaele 2007) along 1500 K (top), 2000 K (middle) and 2700 K (bottom) isotherms also shown.

Appendix C.2. Bulk modulus

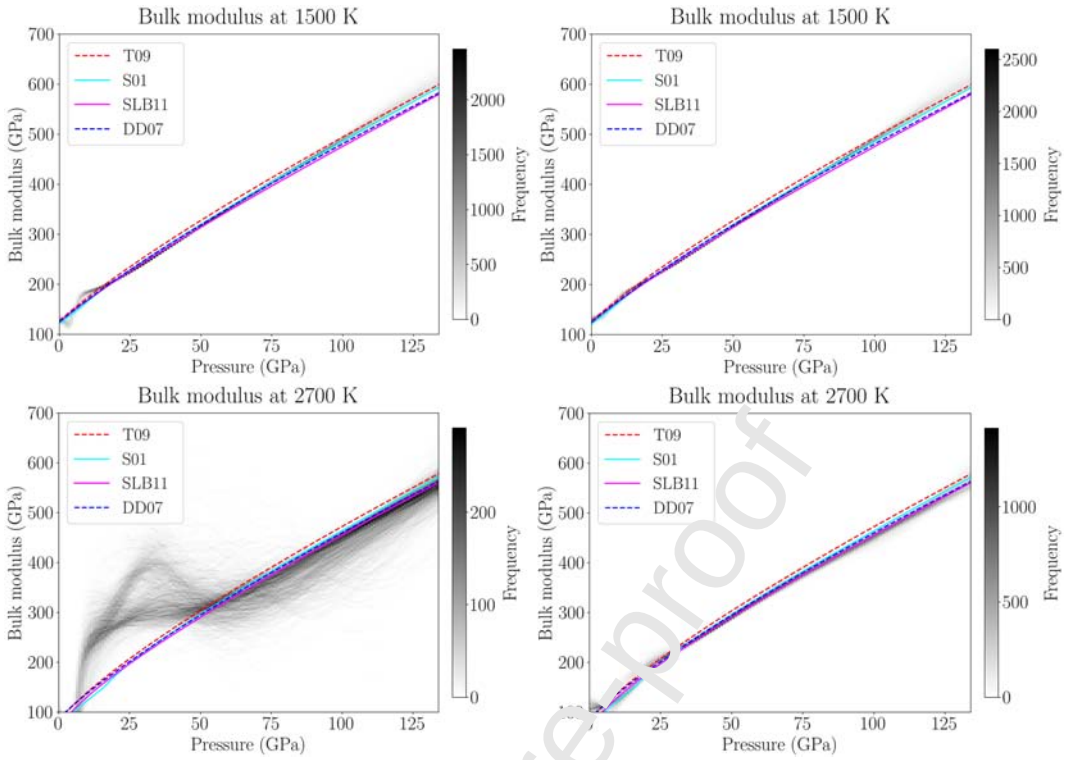


Figure Appendix C.2: Comparison of the bulk modulus of MgO predicted by the neural network along 1500 K (top) and 2700 K (bottom) isotherms with other studies (Tange et al. 2009, Speziale et al. 2001, Stixrude and Lithgow-Bertelloni 2011 and Dorogokupets and Dewaele 2007) as a function of pressure. Left panel shows results from MDNs trained with all data and the right panel shows results from MDNs excluding Murakami et al. 2012 and Fiquet et al. 1990 data in training.

Appendix C.3. Thermal expansivity

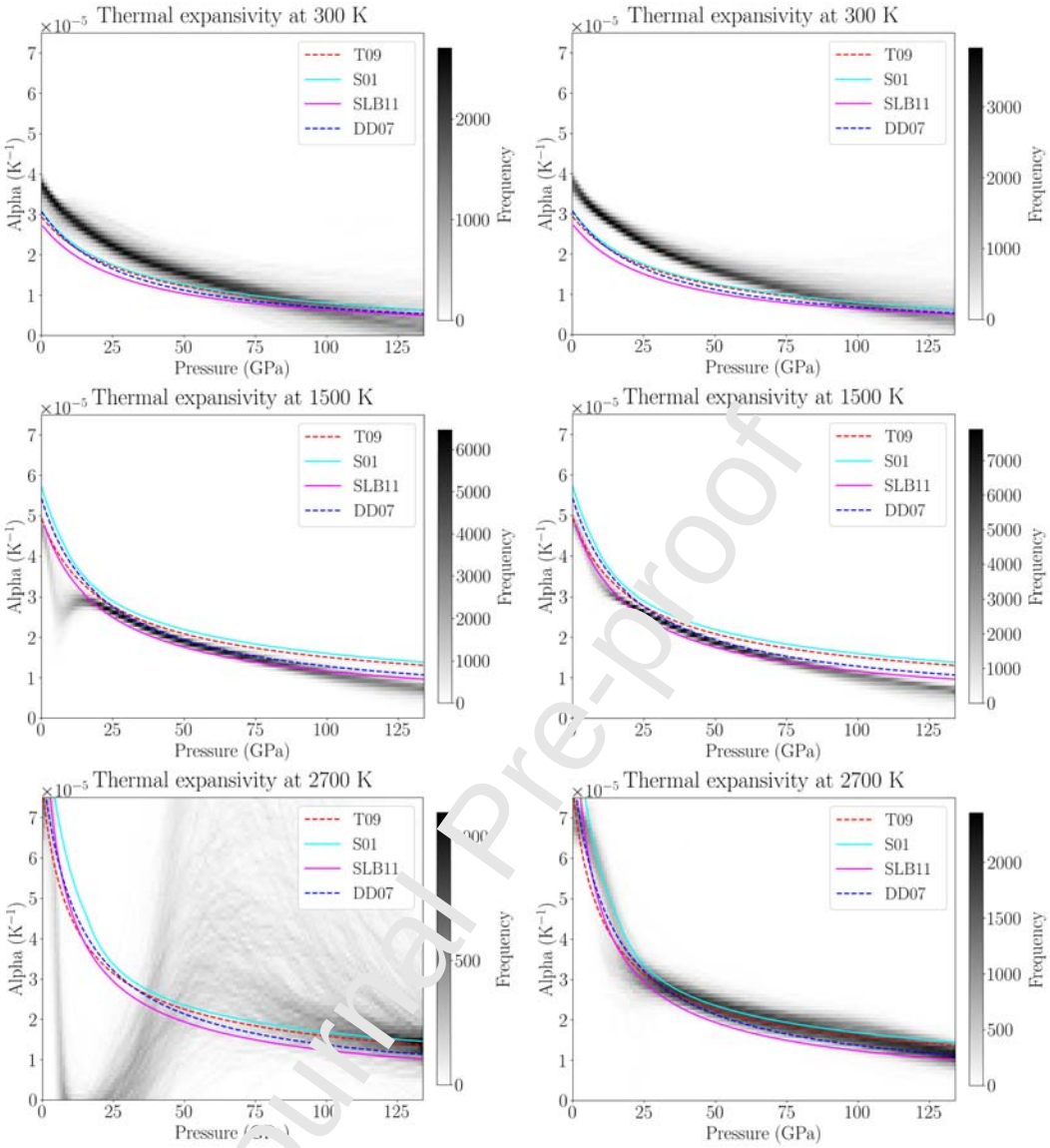


Figure Appendix C.3: Comparison of the thermal expansivity of MgO predicted by neural networks with Tange et al. 2009, Stixrude and Lithgow-Bertelloni 2011 and Dorogokupets and Dewaele 2007 along 300 K (top), 1500 K (middle) and 2700 K (bottom) isotherms as a function of pressure. Left panel: MDNs trained with all data. Right: MDNs trained without Murakami et al. 2012 and Fiquet et al. 1996 data.

References

M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL `https://www.tensorflow.org/`. Software available from tensorflow.org.

S. Atkins, A. P. Valentine, P. J. Tackley, and J. Trampert. Using pattern recognition to infer parameters governing mantle convection. *Physics of the Earth and Planetary Interiors*, 257:171 – 186, 2016. ISSN 0031-9201. doi: https://doi.org/10.1016/j.pepi.2016.05.016. URL `http://www.sciencedirect.com/science/article/pii/S0031920116300863`.

C. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, January 1995. URL `https://www.microsoft.com/en-us/research/publication/neural-networks-pattern-recognition-2/`.

C. M. Bishop. Mixture density networks. Technical report, Aston University, Birmingham, 1994.

F. Cammarano, S. Goes, P. Vacher, and D. Giardini. Inferring upper-mantle temperatures from seismic velocities. *Physics of the Earth and Planetary Interiors*, 138(3):197 – 222, 2003. ISSN 0031-9201. doi: https://doi.org/10.1016/S0031-9201(03)00156-0. URL `http://www.sciencedirect.com/science/article/pii/S0031920103001560`.

F. Cammarano, A. Deuss, S. Goes, and D. Giardini. One-dimensional physical reference models for the upper mantle and transition zone: Combining seismic and mineral physics constraints. *Journal of Geophysical Research: Solid Earth*, 110(B1), 2005a. doi: 10.1029/2004JB003272. URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2004JB003272`.

F. Cammarano, S. Goes, A. Deuss, and D. Giardini. Is a pyrolitic adiabatic mantle compatible with seismic data? *Earth and Planetary Science Letters*, 232(3):227 – 243, 2005b. ISSN 0012-821X. doi: https://doi.org/10.1016/j.epsl.2005.01.031. URL http://www.sciencedirect.com/science/article/pii/S0012821X05000804.

L. Cobden, S. Goes, F. Cammarano, and J. A. D. Connolly. Thermochemical interpretation of one-dimensional seismic reference models for the upper mantle: Evidence for bias due to heterogeneity. *Geophysical Journal International*, 175(2):627–648, 11 2008. ISSN 0956-540X. doi: 10.1111/j.1365-246X.2008.03903.x. URL https://doi.org/10.1111/j.1365-246X.2008.03903.x.

L. Cobden, S. Goes, M. Ravenna, E. Styles, F. Cammarano, K. Gallagher, and J. A. D. Connolly. Thermochemical interpretation of 1-D seismic data for the lower mantle: The significance of nonadiabatic thermal gradients and compositional heterogeneity. *Journal of Geophysical Research: Solid Earth*, 114(B12), 2009. doi: 10.1029/2008JB006262. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2008JB006262.

J. C. Crowhurst, J. M. Brown, A. F. Goncharov, and S. D. Jacobsen. Elasticity of (mg,fe)o through the spin transition of iron in the lower mantle. *Science*, 319(5862):451–453, 2008. ISSN 0036-8075. doi: 10.1126/science.1149606. URL https://science.sciencemag.org/content/319/5862/451.

G. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2(4):303–314, Dec 1989. doi: 10.1007/BF02551274. URL https://doi.org/10.1007/BF02551274.

R. de Wit and J. Trampert. Robust constraints on average radial lower mantle anisotropy and consequences for composition and texture. *Earth and Planetary Science Letters*, 429: 101 – 109, 2015. ISSN 0012-821X. doi: https://doi.org/10.1016/j.epsl.2015.07.057. URL http://www.sciencedirect.com/science/article/pii/S0012821X15005002.

R. de Wit, P. Käufl, A. Valentine, and J. Trampert. Bayesian inversion of free oscillations

for earth's radial (an)elastic structure. *Physics of the Earth and Planetary Interiors*, 237: 1 – 17, 2014. ISSN 0031-9201. doi: https://doi.org/10.1016/j.pepi.2014.09.004. URL http://www.sciencedirect.com/science/article/pii/S0031920114002039.

R. W. L. de Wit, A. P. Valentine, and J. Trampert. Bayesian inference of Earth's radial seismic structure from body-wave traveltimes using neural networks. *Geophysical Journal International*, 195(1):408–422, 06 2013. ISSN 0956-540X. doi: 10.1093/gji/ggt220. URL https://doi.org/10.1093/gji/ggt220.

F. Deschamps and J. Trampert. Towards a lower mantle reference temperature and composition. *Earth and Planetary Science Letters*, 222(1):161 – 175, 2004. ISSN 0012-821X. doi: https://doi.org/10.1016/j.epsl.2004.02.024. URL http://www.sciencedirect.com/science/article/pii/S0012821X04001426.

F. Deschamps, L. Cobden, and P. J. Tackley. The primitive nature of large low shear-wave velocity provinces. *Earth and Planetary Science Letters*, 349-350:198 – 208, 2012. ISSN 0012-821X. doi: https://doi.org/10.1016/j.epsl.2012.07.012. URL http://www.sciencedirect.com/science/article/pii/S0012821X12003718.

A. Dewaele, G. Fiquet, D. Andrault, and D. Hausermann. P-V-T equation of state of periclase from synchrotron radiation measurements. *Journal of Geophysical Research: Solid Earth*, 105(B2):2869–2877, 2000. doi: 10.1029/1999JB900364.

V. V. Dobrosavljevic, W. Sturhahn, and J. M. Jackson. Evaluating the role of iron-rich (Mg,Fe)O in ultralow velocity zones. *Minerals*, 9(12), 2019. doi: 10.3390/min9120762. URL https://www.mdpi.com/2075-163X/9/12/762.

S. M. Dorfman, V. B. Prakapenka, Y. Meng, and T. S. Duffy. Intercomparison of pressure standards (Au, Pt, Mo, MgO, NaCl and Ne) to 2.5 Mbar. *Journal of Geophysical Research: Solid Earth*, 117(B8), 2012. doi: 10.1029/2012JB009292.

P. I. Dorogokupets and A. Dewaele. Equations of state of MgO, Au, Pt, NaCl-B1, and NaCl-B2: Internally consistent high-temperature pressure scales. *High Pressure Research*, 27

(4):431–446, 2007. doi: 10.1080/08957950701659700. URL https://doi.org/10.1080/08957950701659700.

L. Dubrovinsky and S. Saxena. Thermal expansion of periclase (MgO) and tungsten (W) to melting temperatures. *Physics and Chemistry of Minerals*, 24(8):547–550, 1997.

T. Duffy and Y. Wang. Pressure-volume-temperature equation of state. *Reviews in Mineralogy*, 37, 1998.

T. S. Duffy and T. J. Ahrens. Compressional sound velocity, equation of state, and constitutive response of shock-compressed magnesium oxide. *Journal of Geophysical Research: Solid Earth*, 100(B1):529–542, 1995. doi: 10.1029/94JB02065.

A. M. Dziewonski and D. L. Anderson. Preliminary reference earth model. *Physics of the Earth and Planetary Interiors*, 25(4):297 – 356, 1981. ISSN 0031-9201. doi: https://doi.org/10.1016/0031-9201(81)90046-7. URL http://www.sciencedirect.com/science/article/pii/0031920181900467.

S. Earp and A. Curtis. Probabilistic neural network-based 2D travel-time tomography. *Neural Computing and Applications*, 32:17077–17095, 2020. doi: 10.1007/s00521-020-04921-8. URL https://doi.org/10.1007/s00521-020-04921-8.

D. Fan, S. Fu, J. Yang, S. Tkachev, V. Prakapenka, and J.-F. Lin. Elasticity of single-crystal periclase at high pressure and temperature: The effect of iron on the elasticity and seismic parameters of ferropericlase in the lower mantle. *American Mineralogist*, 104:262–275, 02 2019. doi: 10.2138/am-2019-6656.

Y. Fei. Effects of temperature and composition on the bulk modulus of (Mg,Fe)O. *American Mineralogist*, 84:272–276, 1999.

Y. Fei, J. Li, K. Hirose, W. Minarik, J. V. Orman, C. Sanloup, W. van Westrenen, T. Komabayashi, and K. ichi Funakoshi. A critical evaluation of pressure scales at high

temperatures by in situ x-ray diffraction measurements. *Physics of the Earth and Planetary Interiors*, 143-144:515 – 526, 2004a. ISSN 0031-9201. doi: https://doi.org/10.1016/j.pepi.2003.09.018. URL http://www.sciencedirect.com/science/article/pii/S0031920104000858. New Developments in High-Pressure Mineral Physics and Applications to the Earth's Interior.

Y. Fei, J. Van Orman, J. Li, W. van Westrenen, C. Sanloup, W. Minarik, K. Hirose, T. Komabayashi, M. Walter, and K. Funakoshi. Experimentally determined postspinel transformation boundary in Mg_2SiO_4 using MgO as an internal pressure standard and its geophysical implications. *Journal of Geophysical Research: Solid Earth*, 109(B2), 2004b. doi: 10.1029/2003JB002562. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2003JB002562.

G. Fiquet, D. Andrault, J. Itié, P. Gillet, and P. Richet. X-ray diffraction of periclase in a laser-heated diamond-anvil cell. *Physics of the Earth and Planetary Interiors*, 95(1):1 – 17, 1996. ISSN 0031-9201. doi: https://doi.org/10.1016/0031-9201(95)03109-X. URL http://www.sciencedirect.com/science/article/pii/003192019503109X.

G. Fiquet, P. Richet, and G. Montagnac. High-temperature thermal expansion of lime, periclase, corundum and spinel. *Physics and Chemistry of Minerals*, 27:103–111, 12 1999. doi: 10.1007/s002690050246. URL https://doi.org/10.1007/s002690050246.

R. A. Forman, G. J. Piermarini, J. D. Barnett, and S. Block. Pressure measurement made by the utilization of ruby sharp-line luminescence. *Science*, 176(4032):284–285, 1972. doi: 10.1126/science.176.4032.284. URL https://science.sciencemag.org/content/176/4032/284.

E. J. Garnero and D. V. Helmberger. Further structural constraints and uncertainties of a thin laterally varying ultralow-velocity layer at the base of the mantle. *Journal of Geophysical Research: Solid Earth*, 103(B6):12495–12509, 1998. doi: https://doi.org/

10.1029/98JB00700. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/98JB00700.

E. J. Garnero, A. K. McNamara, and S.-H. Shim. Continent-sized anomalous zones with low seismic velocity at the base of earth's mantle. *Nature Geoscience*, 9:481–489, 2016. doi: 10.1038/ngeo2733. URL https://doi.org/10.1038/ngeo2733.

J. W. Hernlund and C. Houser. On the statistical distribution of seismic velocities in earth's deep mantle. *Earth and Planetary Science Letters*, 265(3):423 – 437, 2008. ISSN 0012-821X. doi: https://doi.org/10.1016/j.epsl.2007.10.042. URL http://www.sciencedirect.com/science/article/pii/S0012821X07006784.

K. Hirose, N. Sata, T. Komabayashi, and Y. Ohishi. Simultaneous volume measurements of Au and MgO to 140GPa and thermal equation of state of Au based on the MgO pressure scale. *Physics of the Earth and Planetary Interiors*, 167(3):149 – 154, 2008. ISSN 0031-9201. doi: https://doi.org/10.1016/j.pepi.2008.03.002. URL http://www.sciencedirect.com/science/article/pii/S0031920108000460.

K. Hornik, M. Stinchcombe, and H. White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359 – 366, 1989. ISSN 0893-6080. doi: https://doi.org/10.1016/0893-6080(89)90020-8. URL http://www.sciencedirect.com/science/article/pii/0893608089900208.

M. Ishii and J. Tromp. Normal-mode and free-air gravity constraints on lateral variations in velocity and density of earth's mantle. *Science*, 285(5431):1231–1236, 1999. ISSN 0036-8075. doi: 10.1126/science.285.5431.1231. URL https://science.sciencemag.org/content/285/5431/1231.

I. Jackson. Elasticity, composition and temperature of the Earth's lower mantle: a reappraisal. *Geophysical Journal International*, 134(1):291–311, 07 1998. ISSN 0956-540X. doi: 10.1046/j.1365-246x.1998.00560.x. URL https://doi.org/10.1046/j.1365-246x.1998.00560.x.

J. M. Jackson and C. Thomas. *Seismic and Mineral Physics Constraints on the D″ Layer*, chapter 8, pages 193–227. American Geophysical Union (AGU), 2021. ISBN 9781119528609. doi: https://doi.org/10.1002/9781119528609.ch8. URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/9781119528609.ch8`.

S. Jacobsen, C. Holl, K. Adams, R. Fischer, E. Martin, C. Bina, J.-F. Lin, V. Prakapenka, A. Kubo, and P. Dera. Compression of single-crystal magnesium oxide to 118 gpa and a ruby pressure gauge for helium pressure media. *American Mineralogist*, 93:1823–1828, 11 2008. doi: 10.2138/am.2008.2988.

B. B. Karki, R. M. Wentzcovitch, S. de Gironcoli, and S. Baroni. First-principles determination of elastic anisotropy and wave velocities of MgO at lower mantle conditions. *Science*, 286(5445):1705–1707, 1999. ISSN 0036-8075. doi: 10.1126/science.286.5445.1705.

P. Käufl, A. Valentine, R. Wit, and J. Trampert. Solving probabilistic inverse problems rapidly with prior samples. *Geophysical Journal International*, 205:ggw108, 03 2016a. doi: 10.1093/gji/ggw108.

P. Käufl, A. P. Valentine, and J. Trampert. Probabilistic point source inversion of strong-motion data in 3-d media using pattern recognition: A case study for the 2008 mw 5.4 chino hills earthquake. *Geophysical Research Letters*, 43(16):8492–8498, 2016b. doi: 10.1002/2016GL069887. URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016GL069887`.

A. Keane. An investigation of finite strain in an isotropic material subjected to hydrostatic pressure and its seismological applications. *Australian Journal of Physics*, 7:322 – 333, 1954.

B. Kennett. Towards constitutive equations for the deep earth. *Physics of the Earth and Planetary Interiors*, 270:40 – 45, 2017. ISSN 0031-9201. doi: https://doi.org/10.1016/j.pepi.2017.06.012. URL `http://www.sciencedirect.com/science/article/pii/S003192011730153X`.

B. L. N. Kennett, E. R. Engdahl, and R. Buland. Constraints on seismic velocities in the earth from traveltimes. *Geophysical Journal International*, 122(1):108–124, 1995. doi: 10.1111/j.1365-246X.1995.tb03540.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-246X.1995.tb03540.x.

A. Khan, L. Boschi, and J. A. D. Connolly. On mantle chemical and thermal heterogeneities and anisotropy as mapped by inversion of global surface wave data. *Journal of Geophysical Research: Solid Earth*, 114(B9), 2009. doi: 10.1029/2009JB006399. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2009JB006399.

A. Khan, A. Zunino, and F. Deschamps. The thermo-chemical and physical structure beneath the north american continent from bayesian inversion of surface-wave phase velocities. *Journal of Geophysical Research: Solid Earth*, 116(B9), 2011. doi: 10.1029/2011JB008380. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011JB008380.

A. Khan, A. Zunino, and F. Deschamps. Upper mantle compositional variations and discontinuity topography imaged beneath australia from bayesian inversion of surface-wave phase velocities and thermochemical modeling. *Journal of Geophysical Research: Solid Earth*, 118(10):5285–5306, 2013. doi: 10.1002/jgrb.50304. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/jgrb.50304.

A. Khan, J. A. D. Connolly, A. Pommier, and J. Noir. Geophysical evidence for melt in the deep lunar interior and implications for lunar evolution. *Journal of Geophysical Research: Planets*, 119(10):2197–2221, 2014. doi: 10.1002/2014JE004661. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014JE004661.

A. Khan, C. Liebske, A. Rozel, A. Rivoldini, F. Nimmo, J. A. D. Connolly, A.-C. Plesa, and D. Giardini. A geophysical perspective on the bulk composition of mars. *Journal of Geophysical Research: Planets*, 123(2):575–611, 2018. doi: 10.1002/2017JE005371. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017JE005371.

D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

P. Koelemeijer, A. Deuss, and J. Ritsema. Density structure of earth's lowermost mantle from stoneley mode splitting observations. *Nature Communications*, 8:15241, 2017. doi: 10.1038/ncomms15241. URL `https://doi.org/10.1038/ncomms15241`.

Y. Kono, T. Irifune, Y. Higo, T. Inoue, and A. Barnhoorn. P-V-T relation of MgO derived by simultaneous elastic wave velocity and in situ x-ray measurements: A new pressure scale for the mantle transition region. *Physics of the Earth and Planetary Interiors*, 183 (1-2):196–211, 2010. doi: 10.1016/j.pepi.2010.03.010.

A. Kurnosov, H. Marquardt, D. J. Frost, T. B. Ballaran, and L. Ziberna. Evidence for a fe3+-rich pyrolitic lower mantle from (al,fe)-bearing bridgmanite elasticity data. *Nature*, 543:543–546, 2017. doi: 10.1038/nature21390.

B. Li, K. Woody, and J. Kung. Elasticity of MgO to 11 GPa with an independent absolute pressure scale: Implications for pressure calibration. *Journal of Geophysical Research: Solid Earth*, 111(B11), 2006. doi: 10.1029/2005JB004251.

J.-F. Lin, S. D. Jacobsen, W. Sturhahn, J. M. Jackson, J. Zhao, and C.-S. Yoo. Sound velocities of ferropericlase in the earth's lower mantle. *Geophysical Research Letters*, 33 (22), 2006. doi: 10.1029/2006GL028099. URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2006GL028099`.

K. Litasov, E. Ohtani, A. Sano, A. Suzuki, and K. Funakoshi. In situ x-ray diffraction study of post-spinel transformation in a peridotite mantle: Implication for the 660-km discontinuity. *Earth and Planetary Science Letters*, 238(3):311 – 328, 2005. ISSN 0012-821X. doi: https://doi.org/10.1016/j.epsl.2005.08.001. URL `http://www.sciencedirect.com/science/article/pii/S0012821X05005236`.

L. Liu and Y. Bi. How far away are accurate equations of state determinations? some issues on pressure scales and non-hydrostaticity in diamond anvil cells. *Matter and Ra-*

diation at Extremes, 1(4):224 – 236, 2016. ISSN 2468-080X. doi: https://doi.org/10.1016/j.mre.2016.06.002. URL `http://www.sciencedirect.com/science/article/pii/S2468080X16300309`.

H. Marquardt and A. R. Thomson. Experimental elasticity of earth's deep mantle. *Nature Reviews Earth & Environment*, 1:455–469, 2020. doi: 10.1038/s43017-020-0077-3. URL `https://doi.org/10.1038/s43017-020-0077-3`.

H. Marquardt, S. Speziale, H. J. Reichmann, D. J. Frost, and F. R. Schilling. Single-crystal elasticity of $(Mg_{0.9}Fe_{0.1})O$ to 81 GPa. *Earth and Planetary Science Letters*, 287(3):345 – 352, 2009. ISSN 0012-821X. doi: https://doi.org/10.1016/j.epsl.2009.08.017. URL `http://www.sciencedirect.com/science/article/pii/S0012821X09004889`.

H. Marquardt, J. Buchen, A. S. J. Mendez, A. Kurnosov, M. Wendt, A. Rothkirch, D. Pennicard, and H.-P. Liermann. Elastic softening of $(Mg_{0.8}Fe_{0.2})O$ ferropericlase across the iron spin crossover measured at seismic frequencies. *Geophysical Research Letters*, 45(14):6862–6868, 2018. doi: https://doi.org/10.1029/2018GL077982. URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018GL077982`.

J. Matas, J. Bass, Y. Ricard, E. Mattern, and M. S. T. Bukowinski. On the bulk composition of the lower mantle: predictions and limitations from generalized inversion of radial seismic profiles. *Geophysical Journal International*, 170(2):764–780, 08 2007. ISSN 0956-540X. doi: 10.1111/j.1365-246X.2007.03454.x. URL `https://doi.org/10.1111/j.1365-246X.2007.03454.x`.

U. Meier, A. Curtis, and J. Trampert. Global crustal thickness from neural network inversion of surface wave data. *Geophysical Journal International*, 169(2):706–722, 2007a. doi: https://doi.org/10.1111/j.1365-246X.2007.03373.x. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-246X.2007.03373.x`.

U. Meier, A. Curtis, and J. Trampert. Fully nonlinear inversion of fundamental mode surface waves for a global crustal model. *Geophysical Research Letters*, 34(16), 2007b. doi: https:

//doi.org/10.1029/2007GL030989. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2007GL030989.

U. Meier, J. Trampert, and A. Curtis. Global variations of temperature and water content in the mantle transition zone from higher mode surface waves. *Earth and Planetary Science Letters*, 282(1):91 – 101, 2009. ISSN 0012-821X. doi: https://doi.org/10.1016/j.epsl.2009.03.004. URL http://www.sciencedirect.com/science/article/pii/S0012821X09001460.

M. Murakami, Y. Ohishi, N. Hirao, and K. Hirose. Elasticity of MgO to 130 GPa: Implications for lower mantle mineralogy. *Earth and Planetary Science Letters*, pages 123–129, 2009. doi: 10.1016/j.epsl.2008.10.010.

M. Murakami, Y. Ohishi, N. Hirao, and K. Hirose. A perovskitic lower mantle inferred from high-pressure, high-temperature sound velocity data. *Nature*, 485:90–94, 2012. doi: 10.1038/nature11004.

T. Nakagawa, P. J. Tackley, F. Deschamps, and J. A. D. Connolly. Incorporating self-consistently calculated mineral physics into thermochemical mantle convection simulations in a 3-d spherical shell and its influence on seismic anomalies in earth's mantle. *Geochemistry, Geophysics, Geosystems*, 10(3), 2009. doi: 10.1029/2008GC002280. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2008GC002280.

T. Nakagawa, P. J. Tackley, F. Deschamps, and J. A. Connolly. The influence of morb and harzburgite composition on thermo-chemical mantle convection in a 3-d spherical shell with self-consistently calculated mineral physics. *Earth and Planetary Science Letters*, 296(3):403 – 412, 2010. ISSN 0012-821X. doi: https://doi.org/10.1016/j.epsl.2010.05.026. URL http://www.sciencedirect.com/science/article/pii/S0012821X10003444.

T. Nakagawa, P. J. Tackley, F. Deschamps, and J. A. D. Connolly. Radial 1-d seismic structures in the deep mantle in mantle convection simulations with self-consistently calculated mineralogy. *Geochemistry, Geophysics, Geosystems*, 13(11), 2012. doi:

10.1029/2012GC004325. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2012GC004325.

A. Oganov and P. I. Dorogokupets. All-electron and pseudopotential study of MgO: Equation of state, anharmonicity, and stability. *Phys. Rev. B*, 67:224110, 06 2003. doi: 10.1103/PhysRevB.67.224110.

J. Ritsema, H. J. v. Heijst, and J. H. Woodhouse. Complex shear wave velocity structure imaged beneath africa and iceland. *Science*, 286(5446):1925–1928, 1999. ISSN 0036-8075. doi: 10.1126/science.286.5446.1925. URL https://science.sciencemag.org/content/286/5446/1925.

B. Romanowicz. Using seismic waves to image earth's internal structure. *Nature*, 451:266–268, 2008. doi: 10.1038/nature06583. URL https://doi.org/10.1038/nature06583.

B. S. A. Schuberth, H.-P. Bunge, G. Steinle-Neumann, C. Moder, and J. Oeser. Thermal versus elastic heterogeneity in high resolution mantle circulation models with pyrolite composition: High plume excess temperatures in the lowermost mantle. *Geochemistry, Geophysics, Geosystems*, 10(1), 2009. doi: 10.1029/2008GC002235. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2008GC002235.

B. S. A. Schuberth, C. Zaroli, and G. Nolet. Synthetic seismograms for a synthetic Earth: long-period P- and S-wave traveltime variations can be explained by temperature alone. *Geophysical Journal International*, 188(3):1393–1412, 03 2012. ISSN 0956-540X. doi: 10.1111/j.1365-246X.2011.05333.x. URL https://doi.org/10.1111/j.1365-246X.2011.05333.x.

M. S. Shahraeeni and A. Curtis. Fast probabilistic nonlinear petrophysical inversion. *GEOPHYSICS*, 76(2):E45–E58, 2011. doi: 10.1190/1.3540628. URL https://doi.org/10.1190/1.3540628.

N. A. Simmons, A. M. Forte, L. Boschi, and S. P. Grand. Gypsum: A joint tomographic model of mantle density and seismic wave speeds. *Journal of Geophysical Re-*

search: Solid Earth, 115(B12), 2010. doi: https://doi.org/10.1029/2010JB007631. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010JB007631.

S. V. Sinogeikin and J. D. Bass. Single-crystal elasticity of pyrope and MgO to 20 GPa by brillouin scattering in the diamond cell. *Physics of the Earth and Planetary Interiors*, 120 (1):43 – 62, 2000. ISSN 0031-9201. doi: https://doi.org/10.1016/S0031-9201(00)00143-6.

S. V. Sinogeikin, J. M. Jackson, B. O'Neill, J. W. Palko, and J. D. Bass. Compact high-temperature cell for brillouin scattering measurements. *Review of Scientific Instruments*, 71(1):201–206, 2000. doi: 10.1063/1.1150183. URL https://doi.org/10.1063/1.1150183.

N. Solomatova, J. Jackson, W. Sturhahn, J. Wicks, J. Zhao, T. Toellner, B. Kalkan, and W. Steinhardt. Equation of state and spin crossover of (Mg,Fe)O at high pressure, with implications for explaining topographic relief at the core-mantle boundary. *American Mineralogist*, 101:1084–1093, 05 2016. doi: 10.2138/am-2016-5510.

S. Speziale, C.-S. Zha, T. S. Duffy, R. J. Hemley, and H.-k. Mao. Quasi-hydrostatic compression of magnesium oxide to 52 GPa: Implications for the pressure-volume-temperature equation of state. *Journal of Geophysical Research: Solid Earth*, 106(B1):515–528, 2001. doi: 10.1029/2000JB900318.

S. Speziale, V. E. Lee, S. M. Clark, J. F. Lin, M. P. Pasternak, and R. Jeanloz. Effects of fe spin transition on the elasticity of (Mg, Fe)O magnesiowüstites and implications for the seismological properties of the earth's lower mantle. *Journal of Geophysical Research: Solid Earth*, 112(B10), 2007. doi: 10.1029/2006JB004730. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2006JB004730.

F. D. Stacey. Theory of thermal and elastic properties of the lower mantle and core. *Physics of the Earth and Planetary Interiors*, 89(3):219 – 245, 1995. ISSN 0031-9201. doi: https://doi.org/10.1016/0031-9201(94)03005-4. URL http://www.sciencedirect.com/science/article/pii/0031920194030054.

L. Stixrude and C. Lithgow-Bertelloni. Thermodynamics of mantle minerals – I. physical properties. *Geophysical Journal International*, 162(2):610–632, 2005. doi: 10.1111/j.1365-246X.2005.02642.x.

L. Stixrude and C. Lithgow-Bertelloni. Thermodynamics of mantle minerals- II. phase equilibria. *Geophysical Journal International*, 184(3):1180–1213, 2011. doi: 10.1111/j.1365-246X.2010.04890.x. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-246X.2010.04890.x`.

W. Sturhahn. MINeral physics UTIlity (MINUTI) open source software package, 2020. URL `www.nrixs.com`.

Y. Tange, Y. Nishihara, and T. Tsuchiya. Unified analyses for P-V-T equation of state of MgO: A solution for pressure-scale problems in high P-T experiments. *Journal of Geophysical Research: Solid Earth*, 114(B3), 2009. doi: 10.1029/2008JB005813. URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2008JB005813`.

J. Trampert, P. Vacher, and N. Vlaar. Sensitivities of seismic velocities to temperature, pressure and composition in the lower mantle. *Physics of the Earth and Planetary Interiors*, 124(3):255 – 267, 2001. ISSN 0031-9201. doi: https://doi.org/10.1016/S0031-9201(01)00201-1. URL `http://www.sciencedirect.com/science/article/pii/S0031920101002011`.

J. Trampert, F. Deschamps, J. Resovsky, and D. Yuen. Probabilistic tomography maps chemical heterogeneities throughout the lower mantle. *Science*, 306(5697):853–856, 2004. ISSN 0036-8075. doi: 10.1126/science.1101996. URL `https://science.sciencemag.org/content/306/5697/853`.

W. Utsumi, D. J. Weidner, and R. C. Liebermann. Volume measurement of MgO at high pressures and high temperatures. *Washington DC American Geophysical Union Geophysical Monograph Series*, 101:327–333, 1998. doi: 10.1029/GM101p0327.

M. S. Vassiliou and T. J. Ahrens. Hugoniot equation of state of periclase to 200 GPa. *Geophysical Research Letters*, 8(7):729–732, 1981. doi: 10.1029/GL008i007p00729.

Z. Wu, R. M. Wentzcovitch, K. Umemoto, B. Li, K. Hirose, and J.-C. Zheng. Pressure-volume-temperature relations in MgO: An ultrahigh pressure-temperature scale for planetary sciences applications. *Journal of Geophysical Research: Solid Earth*, 113(B6), 2008. doi: 10.1029/2007JB005275.

Y. Ye, V. Prakapenka, Y. Meng, and S.-H. Shim. Intercomparison of the gold, platinum, and MgO pressure scales up to 140 GPa and 2500 K. *Journal of Geophysical Research: Solid Earth*, 122(5):3450–3464, 2017.

A. Yoneda. Pressure derivatives of elastic constants of single crystal MgO and $MgAl_2O_4$. *Journal of Physics of the Earth*, 38(1):19–55, 1990. doi: 10.4294/jpe1952.38.19.

J. Zhang. Effect of pressure on the thermal expansion of MgO up to 8.2 gpa. *Physics and Chemistry of Minerals*, 27:145–148, 02 2000. doi: 10.1007/s002690050001.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Author statement

Manuscript title: Inferring material properties of the lower mantle minerals using Mixture Density Networks.

Ashim Rijal: methodology, software, validation, formal analysis, investigation, visualisation, writing - original draft

Laura Cobden: conceptualisation, supervision, funding acquisition, writing - review and editing

Jeannot Trampert: supervision, writing - review and editing

Jennifer M. Jackson: validation, formal analysis, writing - review and editing

Andrew Valentine: methodology, writing - review and editing