Physics of the Earth and Planetary Interiors 237 (2014) 1–17

Contents lists available at ScienceDirect

Physics of the Earth and Planetary Interiors

journal homepage: www.elsevier.com/locate/pepi

Bayesian inversion of free oscillations for Earth's radial (an)elastic structure

R.W.L. de Wit *, P.J. Käufl, A.P. Valentine, J. Trampert

Department of Earth Sciences, Utrecht University, Budapestlaan 4, 3584 CD Utrecht, The Netherlands

A R T I C L E I N F O

Article history:
Received 2 May 2014
Received in revised form 19 September 2014
Accepted 21 September 2014
Available online 30 September 2014

Keywords:
Radial Earth structure
Normal modes
Body waves
Uncertainty estimation
Inverse theory
Neural networks

A B S T R A C T

We perform a Bayesian inversion of degree-zero spheroidal mode splitting function measurements for radial (1-D) Earth structure, in terms of the Voigt averages of P-wave (V_P) and S-wave (V_S) velocities, density, bulk and shear attenuation, using neural networks. The method is flexible and allows us to assess the robustness of features in existing reference models, such as *PREM*. The Bayesian framework provides a means for quantifying uncertainties in the model parameters and for measuring the information content of the data. The analysis of the information content suggests that the free oscillations constrain most parameters better than body wave travel time data.

Our most important findings can be summarised as follows. The data prefer an inner–outer core boundary (ICB) that lies in the depth range 5154.7–5165.7 km, i.e. deeper than in existing reference models; the effect on the travel time of inner-core-sensitive seismic phases is comparable to the estimated noise in such measurements. The density contrast at the ICB (0.73 g cm^{-3}) is larger than in *PREM* (0.60 g cm^{-3}) and *ak135f* (0.56 g cm^{-3}), but our range including uncertainties (0.52–0.94 g cm^{-3}) encompasses all previous estimates in the literature. The average V_P and V_S in the D″ region are smaller than in *PREM*, whereas the mean density is probably larger. The data cannot uniquely determine whether this density excess is restricted to the D″ region or distributed throughout the lower(most) mantle. The data cannot determine with certainty the presence or absence of a discontinuity at 220 km depth for V_P, V_S and density. If present, the jump in both velocities is likely smaller than in *PREM*. Shear attenuation parameters in the mantle deviate from *PREM* in a similar fashion to results from more recent studies. We find a non-homogeneous shear attenuation in the inner core, reinforcing the hypothesis that a distinct 'innermost inner core' may exist. The bulk attenuation in the mantle and the outer core is stronger than in *PREM*.

We investigate the influence of radial anisotropy on the inversions and analyse possible trade-offs between (anisotropic) parameters. The largest trade-offs are observed in regions that are believed to be anisotropic, such as the D″ region. This illustrates the need to constrain anisotropy in the (deep) mantle.

1. Introduction

Most of our current knowledge of the Earth's internal structure has been inferred from seismological observations made at its surface. The gross features of these measurements can be explained by relatively simple spherically symmetric (1-D) models of wave velocities, density and attenuation, which describe the Earth's average (radial) structure. Such radial earth models are routinely used for the determination of seismic source locations and serve as a starting model for 3-D seismic tomography: see for example Kennett (2006) and references therein. 1-D seismological reference models are also successfully being used in conjunction with mineral physics data and geodynamical modelling to provide constraints on the Earth's thermochemical structure and its dynamics, e.g. Cammarano et al. (2005), Cammarano et al. (2011), Cobden et al. (2008), Cobden et al. (2009).

Existing seismological reference models have been derived using seismic observables with different, yet complementary, sensitivities to the Earth's interior. The tables of Jeffreys and Bullen (1940) summarised the travel times for many different seismic phases in a 1-D earth model. The accumulation of measurements of the Earth's free oscillations made it possible to construct 1-D profiles of compressional (V_P) and shear (V_S) wave velocities and density (*1066A*, *1066B* (Gilbert, 1975)). Subsequently, parametric models were designed to simultaneously explain travel time,

* Corresponding author. Tel.: +31 (0)30 253 5135.
E-mail address: r.w.l.dewit@uu.nl (R.W.L. de Wit).

http://dx.doi.org/10.1016/j.pepi.2014.09.004
0031-9201/

normal mode and regional surface wave dispersion data (*PEM*, Dziewoński et al. (1975)). A similar form of polynomial representation was used by Dziewoński and Anderson (1981) for the Preliminary Reference Earth Model (*PREM*), which was derived from body wave travel times, normal mode frequencies and attenuation measurements, augmented with constraints on the Earth's mass and moment of inertia, and has clearly outlived its 'preliminary' status. The models *iasp91* (Kennett and Engdahl, 1991) and *ak135* (Kennett et al., 1995) were constructed to explain the extensive catalogue of travel times documented by the International Seismological Centre (ISC). More recently, Cammarano et al. (2005) combined seismological and mineral physics data to construct 1-D physical reference models (*PREF*). Kustowski et al. (2008) derived the spherically symmetric model *STW105*, which serves as a basis for a 3-D tomographic mantle model of anisotropic shear wave velocity (*S362ANI*). These models were derived from body wave travel times, long-period waveforms and surface wave phase anomalies.

Besides the elastic structure, the anelastic properties of the Earth have received quite some attention. This interest relates, particularly, to the temperature dependence of attenuation processes in the Earth, through which elastic (seismic) energy is transformed into heat. The Earth's absorption properties can be inferred from the attenuation of free oscillations and surface waves. For instance, *PREM* includes a model for bulk and shear attenuation, represented by their inverses Q_κ and Q_μ, respectively. Montagner and Kennett (1996) aimed to reconcile free oscillation and travel time observations by supplementing *ak135* with density and Q profiles (*ak135f*). Other estimates of the Earth's Q structure include *PAR3C* (Okal and Jo, 1990), *QM1* (Widmer et al., 1991), *QL6* (Durek and Ekström, 1996; Resovsky et al., 2005; Cammarano and Romanowicz, 2008).

Another crucial feature of seismological models concerns anisotropy in elastic parameters. Mantle flow is one mechanism that might introduce anisotropy; therefore, the identification of seismic anisotropy can provide important constraints on the Earth's dynamics. Whereas *PREM* is only transversely isotropic in the uppermost mantle, more recent studies suggest that the deeper parts of the Earth also exhibit radial anisotropy. Anisotropy has been inferred in the inner core, e.g. Morelli et al. (1986), Woodhouse et al. (1986), Beghein and Trampert (2003), Deuss et al. (2010), and the lowermost mantle, e.g. Montagner and Kennett (1996), Panning and Romanowicz (2004), while consensus appears to have been reached that the rest of the lower mantle is devoid of anisotropic structure (see Chang et al. (2014) for a review). Furthermore, the presence of anisotropy is important for estimates of the Earth's seismic structure due to trade-offs between (anisotropic) parameters in earth models. For instance, Beghein et al. (2006) investigated the robustness of radial anisotropy in existing 1-D mantle reference models and found that the strength of anisotropy in V_P trades off with density structure, which was later confirmed by Kustowski et al. (2008).

The aforementioned 1-D seismological reference models correlate with each other to a high degree (Fig. 1), especially in the Earth's deep interior, yet there is disagreement. The fundamental shortcoming of most existing models is the lack of a quantitative assessment of their accuracy, which renders it impossible to determine the significance of the differences between these models. Seismic inverse problems are notoriously non-unique; different

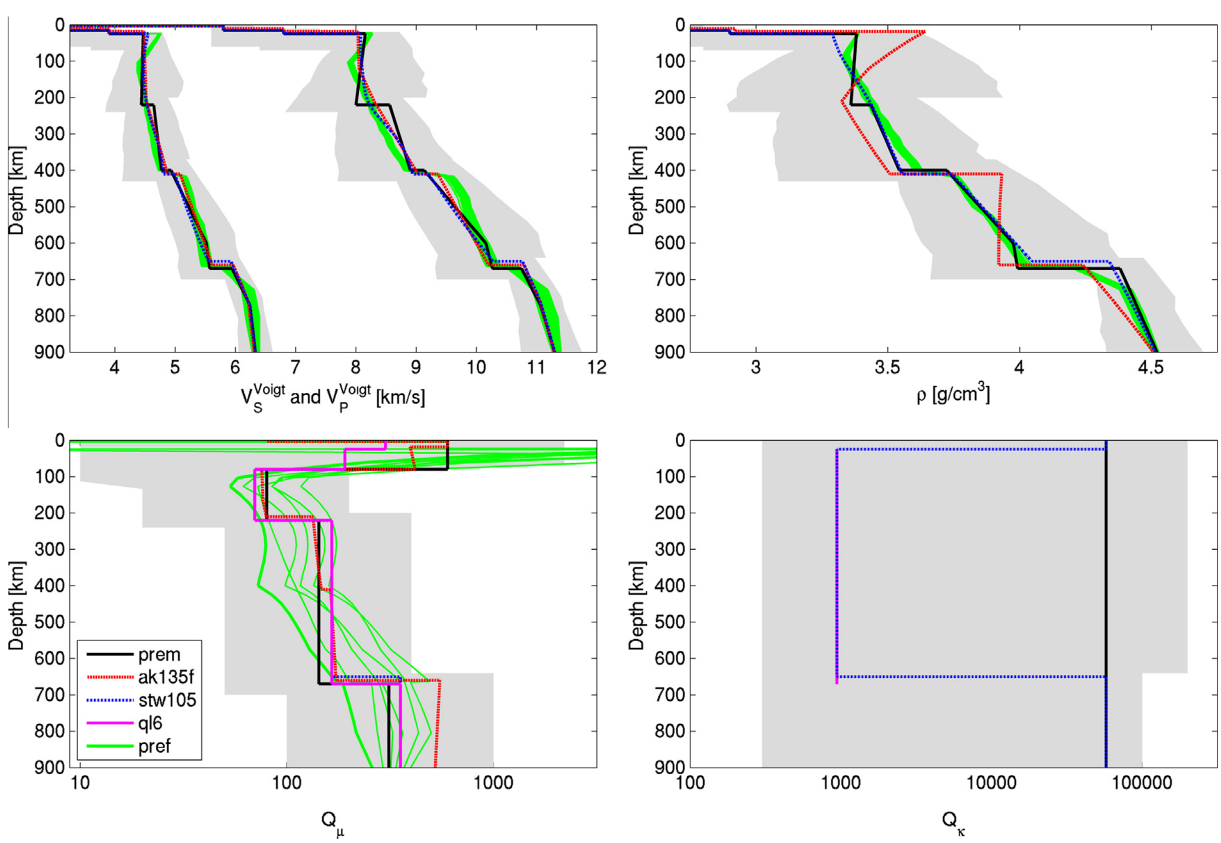

Fig. 1. Radial earth models in the upper mantle. The parameter range spanned by the prior model space is represented by the grey shaded area, along with the 1-D reference models *PREM* (black, solid), *ak135f* (red, dashed), *STW105* (blue, dotted-dashed), *QL6* (magenta, solid) and *PREF* (Cammarano et al., 2005, 99 models, green) for V_S^{Voigt} and V_P^{Voigt} (top-left panel), ρ (top-right panel), Q_μ (bottom-left panel) and Q_κ (bottom-right panel). The horizontal scale for the bottom panels is logarithmic. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

earth models can explain the data equally well, but may lead to incompatible interpretations of the nature of the Earth's interior and dynamics, e.g. Trampert and van (2005), de Wit et al. (2012). Therefore, a quantification of uncertainties in any inferred earth model is essential to assess its quality and the robustness of any subsequent interpretation.

However, quantifying model uncertainties presents a challenge in traditional seismological inverse problems; consequently, most existing techniques are pragmatic and based upon linear approximations. Resolution analyses, for instance using the framework by Backus and Gilbert (1968), Backus and Gilbert (1970), can be employed to determine the robustness of the inferred earth models, e.g. Kennett (1998), Masters and Gubbins (2003). In seismic tomography, resolution and covariance matrices can provide some assessment of model quality, e.g. Aki et al. (1977), Boschi (2003), Vasco et al. (2003), but such measures are usually affected by subjective regularisation criteria. Other examples for the linear case include exploring the model null space, or model non-uniqueness (de Wit et al., 2012), misfit mapping, e.g. in the context of source parameter determination (Valentine and Trampert, 2012) and resolution tests using matrix probing (An, 2012; Trampert et al., 2013). Kennett et al. (1995) adopted a non-linear search procedure to determine the robustness of *ak135*.

An assessment of model uncertainty is natural in a Bayesian framework, in which all inferences are probabilistic. Any inference made about a model is the result of the conjunction of our current (*prior*) knowledge and the ability of the model to explain the observations, e.g. Tarantola and Valette (1982). The *posterior* knowledge on the model, i.e. the knowledge after observing the data, represents the updated degree of belief in the model, expressed by a probability density function (pdf). Ample examples of Bayesian inference exist in the seismological literature; this involves sampling the model space, as is done in Markov Chain Monte Carlo (MCMC) methods via a (guided) random walk, e.g. Mosegaard and Tarantola (1995), Sambridge and Mosegaard (2002), Tarantola (2005).

As an alternative, we propose to employ machine learning techniques to make inferences based on samples of the prior model space. Recent examples in geophysics include Meier et al. (2007), Shahraeeni and Curtis (2011), de Wit et al. (2013) and Käufl et al. (2014), where artificial neural networks and a set of prior samples are used to solve various geophysical inverse problems. Neural networks are very common in pattern recognition problems and can be used to infer an arbitrary non-linear mapping between two parameter spaces, e.g. Bishop (1995), MacKay (2003).

To solve the inverse problem, which in our framework involves obtaining posterior pdfs on earth model parameters, we use a Mixture Density Network (MDN, Bishop, 1995). An MDN takes the seismic data as input, and outputs the parameters describing the posterior marginal pdf for the earth model parameter(s) of interest; for a full description, see e.g. de Wit et al. (2013), Käufl et al. (2014). In the Bayesian paradigm, a 1-D marginal distribution represents our knowledge of (and uncertainties in) a single model parameter, given the variations in all other model parameters. The method is flexible, as we are free to choose the output, or *target*, parameter for the MDN. This allows us to ask specific questions, i.e. test hypotheses, about an arbitrary (combination of) model parameter(s), such as the depth of a seismic discontinuity or the average density in a region. In addition, we can construct 2-D pdfs to investigate the trade-offs between parameters, given the constraint offered by the available data.

We investigate the information on radial Earth structure that is contained in various seismic observations and assess the robustness of features in existing reference models, such as *PREM*. Our aim is threefold. First, we illustrate the flexibility of the method to investigate specific parameters in our earth model. In the process, we can assess the uncertainties in the corresponding estimates and the information content of the data. Second, we exploit the constraint on earth model parameters provided by recently-measured normal mode splitting functions (Deuss et al., 2013; Koelemeijer et al., 2013; Koelemeijer, 2014). We perform a non-linear Bayesian inversion of the available data for radial Earth structure in terms of V_P, V_S, density, bulk and shear attenuation. We focus on specific parameters in the radial distributions rather than presenting a new radial earth model. Third, we investigate potential trade-offs between parameters in the context of radial anisotropy.

This paper is structured as follows. First, we describe the earth model parametrisation and the normal mode data. Second, we train MDNs to construct 1-D marginal posterior pdfs for the parameters in our radial earth model. Third, we address trade-offs between model parameters related to anisotropy. Finally, we discuss the efficacy of a joint inversion of normal mode and travel time data by analysing the information content of the various data sets.

2. Model parametrisation

We base our model parametrisation on that used for *PREM* and parametrise the radial (1-D) structure of the Earth in terms of V_P, V_S, density (ρ), the anisotropic parameter η and bulk and shear attenuation ($1/Q_\kappa$ and $1/Q_\mu$, respectively). The model is parametrised on a discrete set of 185 grid points (as one of the options in the Mineos package (Masters et al., 2011)) and the depths of discontinuities are allowed to vary. These points, or knots, are used by Mineos for a cubic spline interpolation to obtain a continuous representation with depth (between discontinuities). No correlations between physical parameters are imposed, i.e. velocity, density, η and attenuation profiles are constructed independently from each other. Within each profile, except for attenuation, we introduce correlations between adjacent points away from discontinuities to exclude physically implausible, i.e. non-smooth or oscillatory, models and restrict the size of the model space. In addition, we impose constraints on the mass and moment of inertia of the earth models using estimates from Chambat and Valette (2001). The details of the parametrisation are given in A.

We consider two different classes of parametrisation for the anisotropic structure. In a first setup, we allow for radial anisotropy in the uppermost mantle between the Moho and the 220 km discontinuity ("220"), as in *PREM*, which is parametrised by the vertically (V_{PV}, V_{SV}) and horizontally (V_{PH}, V_{SH}) polarised wave velocities and η (Dziewoński and Anderson, 1981). In a second setup, we adopt a similar anisotropic parametrisation in the whole mantle and in the inner core.

We generate 100,000 synthetic models, which are randomly drawn from the prior model distribution. Fig. 1 shows the parameter range spanned by the prior model space and a number of existing 1-D reference models for the upper mantle. Prior ranges for the various parameters in our model are given in Tables A.2–A.4.

3. Methodology

We use artificial neural networks to solve the non-linear Bayesian inverse problem. Neural networks can approximate an arbitrary non-linear function, using a set of examples of corresponding input–output pairs. These examples are presented to a network in a so-called *training* process, during which the free parameters of a network are modified to approximate the function of interest. The particular class of neural network we use here, the MDN, takes seismological observations as input and outputs the parameters governing a conditional probability distribution (B).

We closely follow the methodology outlined in de Wit et al. (2013). We refer the reader to this work and references therein to Bishop (1995) and, for instance, Meier et al. (2007) for details.

Neural network training is sensitive to the random initialisation of the network parameters. Therefore, it is common practice to train several neural networks with different initialisations, and subsequently choose the network which performs best on a given synthetic test data set, e.g. Bishop (1995). de Wit et al. (2013) trained 30 independent networks; the network which performed best on the test set was used to draw inferences from the observed data.

Here, we extend the method by employing *ensembles* of MDNs, as used by—for instance—Cornford et al. (1999) and Käufl et al. (2014). Ensembles, or committees, of networks can result in better generalisation, i.e. achieve a better prediction accuracy on unseen data, e.g. Bishop (1995). The ensemble output is formed by a weighted average of the members, where the individual weights are determined by each network's performance on the same test set (B).

4. Data

We use centre frequencies and mode quality factors derived from 185 self-coupled spheroidal mode splitting functions up to 10 mHz, the majority of which were measured by Deuss et al. (2013). This recent catalogue contains modes sensitive to V_P and inner core structure. We supplement this catalogue with similar measurements for Stoneley modes (Koelemeijer et al., 2013) and fundamental modes ${}_0S_{22}-{}_0S_{30}$ and the mode ${}_2S_{17}$ (Koelemeijer, 2014). Note that both centre frequencies and mode quality factors are only sensitive to radial (1-D) structure and thus only depend on the degree-zero splitting function coefficients c_{00}. Other splitting function coefficients are not used in our analysis, since they relate to lateral variations in Earth structure.

We use the Mineos package (Masters et al., 2011) to calculate exact normal mode frequencies and quality factors for all 100,000 synthetic 1-D earth models. Self-gravitation is taken into account for frequencies below 30 mHz and a reference period of 1 s is used for the attenuative dispersion correction. Since Mineos does not compute the centre frequency and quality factor of mode ${}_0S_4$ for ∼2.5% of the synthetic models due to inherent computational limitations, we exclude this mode from our data set. The synthetic data for the normal modes thus consist of 184 free oscillation centre frequencies and quality factors.

We corrupt the synthetic data by adding Gaussian noise with zero mean and a standard deviation given by the uncertainty estimate accompanying each measurement (Deuss et al., 2013; Koelemeijer et al., 2013; Koelemeijer, 2014). The measurement errors were estimated using a cross-validation approach; it should be noted that this approach may not fully account for any systematic uncertainties. Further, *PREM* was used as a reference model in the iterative damped least-squares inversion of normal-mode spectra for splitting function coefficients, e.g. Deuss et al. (2013). As such, the coefficients could in theory include a bias towards *PREM*. However, Deuss et al. (2013) specify that the centre frequencies and quality factors—or the coefficient c_{00} from which these quantities are derived—are the most robust parameters in their inversion, to which no damping is applied; therefore, we assume that the bias is minimal.

5. Results

5.1. Network configuration

For all results presented in this study, we train MDNs with 40 hidden units and a Gaussian mixture consisting of 15 Gaussian kernels. The number of free parameters in an MDN N_w is given by

$$N_w = (I+1)\cdot J + (J+1)\cdot K, \tag{1}$$

where I, J and K are the number of input, hidden and output units, respectively (Bishop, 1995). For a 1-D target parameter and 15 Gaussian kernels, an MDN has 45 output parameters (the means, the standard deviations and the relative importance of the Gaussian kernels). In combination with a 184-D input and 40 hidden units, such an MDN has 9245 free parameters. Networks are trained using the Scaled Conjugate Gradient (SCG) algorithm (Møller, 1993) for a maximum of 5000 iterations.

As in de Wit et al. (2013), we employ early stopping, which means that network training is halted when the error of a separate validation set reaches a minimum. We use 80% of the 100,000 patterns in the synthetic data set for training, 15% for the validation set and the remaining 5% for the test set. We train an ensemble of 48 networks for each target. For each network realisation, the synthetic data are randomly divided over training, validation and test sets to enhance the generalisation capability of the ensemble.

5.2. Network target parameters

As we pointed out earlier, we can choose any (combination of) parameter(s) in the radial earth model as a target parameter for the MDNs. We aim at asking specific questions, i.e. we test hypotheses, rather than inferring a complete earth model that best fits the data. Our focus lies on discontinuities in the earth model; we investigate their depths and the amplitudes of the corresponding jumps in velocity and density. In particular, we address the ICB depth, the associated density contrast and the existence of the "220" discontinuity in the 1-D earth model. Further, we study the average velocities and density in the D″ region and test the hypothesis of a density excess in the lower(most) mantle. In addition, we infer the mean velocities and density in the upper mantle. In the remainder of this section, we show the results for these parameters in order of decreasing depth. Finally, we investigate the bulk and shear attenuation.

5.2.1. Influence of radial anisotropy

We consider two different classes of radial anisotropy in our earth models (A): (i) radial anisotropy in the uppermost mantle (similar to *PREM*) and (ii) radial anisotropy in the whole mantle and inner core. While spheroidal modes are sensitive to all the radial anisotropy parameters, we would only expect to image them together with toroidal mode data. Therefore, we show results for the parametrisation in which the uppermost mantle is characterised by radial anisotropy (similar to *PREM*). These inferences are conditioned on the assumption that the rest of the mantle and inner core are isotropic. We use the second (fully anisotropic) parametrisation to address possible trade-offs with deeper anisotropic parameters in Section 6.1. Note that the outer core is set to be isotropic in both parametrisations.

5.3. Inferences on radial Earth structure

For networks trained on discontinuity depths, velocities, density and η, we select the centre frequencies as input. MDNs trained on attenuation use the mode quality factors as input. We evaluate the performance of each network ensemble by comparing the target value for 5000 test set samples with the Maximum A Posteriori (MAP) estimate of the ensemble output. The MAP estimate represents the parameter value that is assigned the highest probability in the posterior pdf. As an additional check, we investigate the prediction accuracy of the trained ensemble for *PREM*.

Further, we quantify the information content of the data for each target parameter by computing the Kullback–Leibler divergence D_{KL} in bits (C) between the 1-D marginal posterior and prior pdfs, e.g. MacKay (2003). D_{KL} measures the information gain, or relative entropy, for a particular model parameter upon observing the data, e.g. Meier et al. (2007), Käufl et al. (2014). For reference, consider a 1-D Gaussian distribution with mean μ and standard deviation σ; when comparing this distribution to one with the same mean and standard deviation $\frac{1}{2}\sigma$, the information gain is 1.16 bits. If we do not extract unique information from the data, i.e. information not contained in our prior pdf, the information gain $D_{KL} = 0$. This could occur if the data have no sensitivity to a region within the Earth. In such a case, the MDN output will resemble the prior pdf (de Wit et al., 2013).

5.3.1. Discontinuity depths

Table 1 summarises the posterior statistics for the seven discontinuities by the MAP estimate θ and its asymmetric 2σ error bars, corresponding to $1/e^2$ levels in the unit normalised 1-D marginal posterior pdfs. Our analysis shows that normal mode data constrain the depths of the ICB, the core-mantle boundary (CMB), "660" and Moho, as indicated by the information gain, which is 2.5 bits or more for these parameters (Table 1). The depths of the top of the D″ layer, the "410" and the "220" are not constrained, i.e. $D_{KL} < 1.0$ bits. As an example, Fig. 2 shows the results for the test set, *PREM* and the observed data for the ICB. Prediction accuracy is high for the patterns in the test set and for *PREM*.

The inversion of the normal mode measurements results in clear deviations from the *PREM* reference values. This is to be expected for the "660", which is commonly found to be at a depth of 660 km in (radial) earth models but is fixed to 670 km in *PREM*, e.g. Deuss et al. (2013). For this parameter, $\theta = 663.7$ km, although *PREM* is included at the 2σ level ($\theta \pm 2\ \sigma$ = 650.3–676.2 km). The ICB ($\theta = 5160.1$ km) lies significantly deeper than in *PREM* (5149.5 km) and *ak135* (5153.5 km), which are both outside the $\pm 2\sigma$ range (Fig. 2, Table 1). We discuss the inference on the ICB depth in more detail in Section 6.3. The data indicate a strong preference for a CMB at a shallower depth ($\theta \pm 2\sigma$ = 2886.5–2890.7 km) than in *PREM* and *ak135*. The posterior pdf for the Moho ($\theta = 36.2$ km) is in accord with the continental structure of *ak135* (35.0 km), but also includes the *PREM* value (24.4 km) at the 2σ level.

Table 1
Posterior statistics for the seven discontinuity depths in kilometres, in terms of the MAP estimate θ and asymmetric 2σ model error bars, corresponding to $1/e^2$ levels in the unit normalised 1-D marginal posterior pdfs. The corresponding *PREM* and *ak135* values are given for comparison. The last column shows the information gain D_{KL} in bits (C).

Discontinuity	*PREM*	*ak135*	θ	$\theta - 2\sigma$	$\theta + 2\sigma$	D_{KL} [bits]
ICB	5149.5	5153.5	5160.1	5154.7	5165.7	4.7
CMB	2891.0	2891.5	2888.6	2886.5	2890.7	13.8
D″ (top)	2741.0	2740.0	2722.4	2721.0	2761.0	0.0
"660"	670.0	660.0	663.7	650.3	676.2	2.5
"410"	400.0	410.0	386.2	370.0	413.8	0.7
"220"	220.0	–	200.9	200.0	240.0	0.0
Moho	24.4	35.0	36.2	20.8	47.3	3.4

5.3.2. The density contrast at the ICB

The ICB marks the deepest phase transformation in the Earth. The growth of the inner core forms a source of energy that powers the geodynamo, with the amount of energy delivered through compositional convection strongly dependent on the density contrast across the ICB, e.g. Masters and Gubbins (2003). Therefore, knowledge of the density contrast across this boundary ($\Delta\rho^{\mathrm{ICB}}$) is crucial to our understanding of the generation of the Earth's magnetic field. The density jump can be inferred from various seismic observables, but estimates of this parameter in a spherically symmetric Earth vary significantly in the literature. One approach uses normal modes that are sensitive to the inner core to infer the amplitude of the contrast, e.g. Masters and Gubbins (2003), who find 0.82 ± 0.18 g cm^{-3}. *PREM* is largely based on normal mode data and has $\Delta\rho^{\mathrm{ICB}} = 0.60$ g cm^{-3}, similar to the value in *ak135f* (0.56 g cm^{-3}, Montagner and Kennett (1996)).

A second technique considers the amplitude of the core reflected phase *PKiKP*, and its ratio with respect to the amplitude of *PcP*. Using *PKiKP*/*PcP* amplitude ratios, Cao and Romanowicz (2004) find a range of values 0.6 – 0.9 g cm^{-3} with a preferred value of 0.85 g cm^{-3}. Koper and Dombrovskaya (2005) infer a lower value of 0.52 ± 0.24 g cm^{-3}, as do Shearer and Masters (1990), who find a best fitting value of 0.55 g cm^{-3}. Approximate upper limits to the density jump, as derived from body wave studies, are 1.0 g cm^{-3} (Shearer and Masters, 1990) and 1.1 g cm^{-3} (Tkalčić et al., 2009).

Gubbins et al. (2008) point out that the two data types are sensitive to Earth structure on different length scales, which could explain discrepancies between estimates of the density contrast. Body waves are sensitive on a length scale of several kilometres, whereas the normal modes used in our study have radial wavelengths on the order of hundreds of kilometres. Gubbins et al. (2008) reconcile the different density contrasts from body wave and normal mode studies and estimates of inner core heat flux by introducing a thermochemical boundary layer at the base of the Earth's outer core.

The flexibility of our method allows us to investigate the density contrast across the ICB on different length scales, although we are fundamentally limited by the resolving power (or wavelengths) of the normal modes. We train ensembles of MDNs on two different target parameters: (i) the density contrast at the boundary ($\Delta\rho^{\mathrm{ICB}}$), given by the difference in density between the two points representing the ICB in our model parametrisation, and (ii) the density contrast over a wider region spanning a few hundred kilometres. For all synthetic models and *PREM*, we calculate the average density in two ~200 km wide regions above and

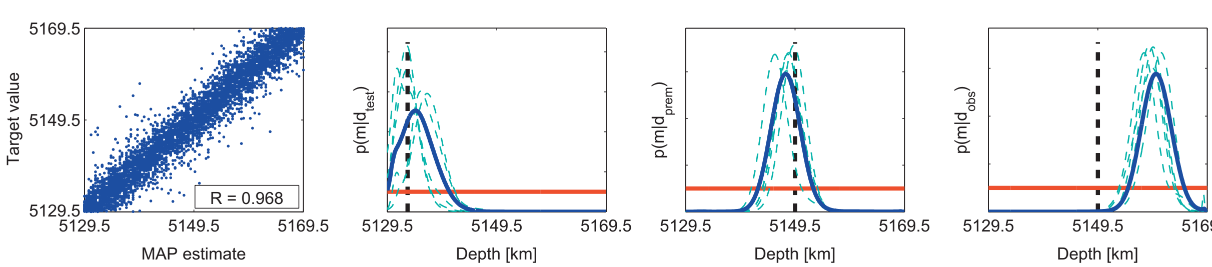

Fig. 2. Ensemble of MDNs trained on the depth of the ICB. The left-hand panel shows the performance on the 5000 test set samples, represented by the correspondence between the MAP estimate and the target value and quantified by the correlation coefficient *R*. The other panels show the 1-D marginal posterior pdf (blue line), the prior pdf (red) and the target value (black, dashed) for a test set pattern (second panel from the left), *PREM* (third panel) and the observed data (right-hand panel). Note that the *PREM* value in the right-hand panel is shown as a reference and does not represent a target for the observed data. The normalised pdfs for five individual networks in the ensemble are shown in cyan. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

below the ICB. The difference between the averages forms the new density jump ($\Delta\rho^{\mathrm{ICB(Wide)}}$).

For the first case, network predictions are accurate, as indicated by the performance on test set samples and *PREM* (Fig. 3). The MAP estimate for the observed data (0.73 g cm^{-3}) is higher than in *PREM* and *ak135f*. The 2σ error levels span a range of 0.52 – 0.94 g cm^{-3} (Table 2). For the second target parameter, network prediction accuracy increases and the width of the posterior pdfs decreases. Again, the observed data assign most probability to values higher than in *PREM* and *ak135f*. On the 2σ level, the density jump lies in the range 0.67 – 0.97 g cm^{-3}, with an MAP estimate of 0.82 g cm^{-3}, which is comparable to the result of Masters and Gubbins (2003) (0.82 $\pm$ 0.18 g cm^{-3}) and in agreement with the upper limits of Shearer and Masters (1990) ($\sim$1.0 g cm^{-3}) and Tkalčić et al. (2009) ($\sim$1.1 g cm^{-3}).

Our uncertainty estimates naturally encompass the discrepancies between the estimates of earlier studies. Thus, we cannot differentiate between these estimates, given the constraint provided by the data we used here; their respective differences relate to the non-uniqueness of the inverse problem and the differences in the data used.

We also investigate the jumps in V_P and V_S across the ICB. The jump in V_P is resolved by the data, while the V_S contrast is poorly constrained ($D_{KL} < 1.0$ bits, Table 2). Whereas the density jump is probably larger than in *PREM* and *ak135f*, the V_P contrast is likely smaller.

5.3.3. Average velocity and density in D″

We train networks for the average V_P and V_S ($\overline{V}_P$ and $\overline{V}_S$, respectively) and density ($\bar{\rho}$) in the D″ layer. The region is 150 km wide in *PREM*, but its thickness varies throughout the training data set due to the change in depth of the two enclosing discontinuities. The data provide a strong constraint on the parameters ($D_{KL} \geqslant 6.0$ bits, Table 3). The MAP estimates for $\overline{V}_P$ and $\overline{V}_S$ are respectively 0.9% and 1.3% lower than in *PREM*, which lies outside the 2σ range for both parameters. The mean density in the D″ region is 1.5% higher than in *PREM*, in agreement with for instance Beghein et al. (2006), who find a mean density excess of $\sim$1.5% with respect to *PREM*, and *ak135f* (Montagner and Kennett, 1996), in which the average density in the D″ layer deviates by 3.7% from *PREM*.

Table 2
Similar to Table 1, but for $\Delta V_P^{\mathrm{ICB}}$, $\Delta V_S^{\mathrm{ICB}}$ and $\Delta\rho^{\mathrm{ICB}}$ at the ICB and the difference between the average V_P, V_S and density in $\sim$200 km wide regions above and below the ICB.

	PREM	*ak135f*	θ	$\theta - 2\sigma$	$\theta + 2\sigma$	D_{KL} [bits]
$\Delta V_P^{\mathrm{ICB}}$	0.67	0.75	0.56	0.39	0.73	3.4
$\Delta V_P^{\mathrm{ICB(Wide)}}$	0.77	0.80	0.66	0.56	0.76	7.0
$\Delta V_S^{\mathrm{ICB}}$	3.50	3.50	3.47	3.43	3.56	0.3
$\Delta V_S^{\mathrm{ICB(Wide)}}$	3.53	3.53	3.48	3.45	3.58	0.5
$\Delta\rho^{\mathrm{ICB}}$	0.60	0.56	0.73	0.52	0.94	2.3
$\Delta\rho^{\mathrm{ICB(Wide)}}$	0.70	0.67	0.82	0.67	0.97	3.9

Table 3
Similar to Table 1, but for $\overline{V}_P$, $\overline{V}_S$ and $\bar{\rho}$ in the D″ layer. Values are given as percentage deviations from *PREM*.

	PREM	θ [%]	$\theta - 2\sigma$ [%]	$\theta + 2\sigma$ [%]	D_{KL} [bits]
$\overline{V}_P$ [%]	13.70 [km/s]	−0.9	−1.7	−0.0	6.9
$\overline{V}_S$ [%]	7.27 [km/s]	−1.3	−2.1	−0.5	8.1
$\bar{\rho}$ [%]	5.53 [g/cm^3]	1.5	0.4	2.5	6.0

5.3.4. Excess density in the lowermost mantle

Kellogg et al. (1999) propose a compositionally distinct layer of $\sim$500 km at the bottom of the mantle and estimate that such a layer would be stable for a density excess of $\sim$1% (with respect to an isochemical mantle). From a linear analysis of normal mode data, Masters and Gubbins (2003) infer a possible excess of 0.4% with respect to a value of 5.45 g cm^{-3} (the median value of the models in their analysis), but they note that this value is within observational uncertainties. Alternatively, the whole lower mantle, ranging from the CMB to the "660", may be slightly more dense, which would imply that the excess density in the lowermost $\sim$500 km of the mantle is less than 0.4%. We investigate whether a density excess in the lower mantle could exist and, if so, whether this density excess is distributed throughout the lower mantle or whether the data can be explained by a strong density excess in the D″ region and a *PREM*-like lower mantle.

First, we consider $\bar{\rho}$ in three layers of variable thickness, ranging from the CMB and upwards: (i) 2891–2376 km, (ii) 2891–1792 km, and (iii) the whole lower mantle (2891–670 km), i.e. ranging up to the "660". The uncertainties in the density estimates are relatively

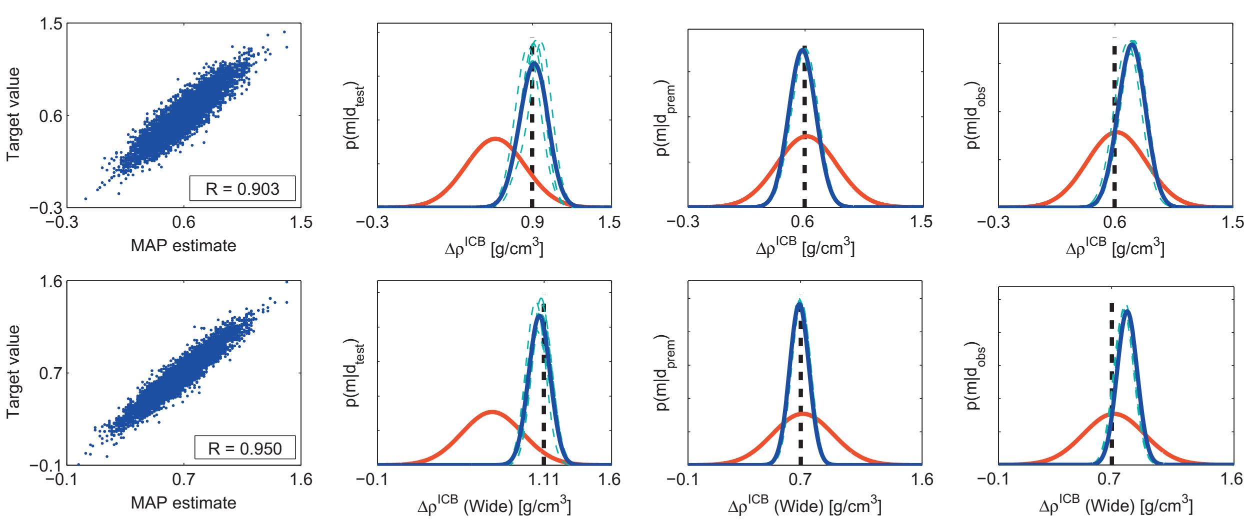

Fig. 3. Ensembles of MDNs trained on the density contrast $\Delta\rho^{\mathrm{ICB}}$ at the ICB (first row) and the difference between the average density in $\sim$200 km wide regions above and below the ICB (second row). The panels are similar to those shown in Fig. 2.

Table 4
Similar to Table 1, but for $\bar{\rho}$ in three layers in the lower mantle, both including and excluding the D″ region (2891–2741 km). Values are given as percentage deviations from *PREM*.

Depth [km]	*PREM* [g/cm^3]	θ [%]	$\theta - 2\sigma$ [%]	$\theta + 2\sigma$ [%]	D_{KL} [bits]
2891–2376	5.44	0.8	0.4	1.1	10.0
2891–1792	5.29	0.6	0.3	0.9	10.7
2891–670	5.00	0.4	0.1	0.7	10.4
2741–2376	5.40	0.2	−0.2	0.6	9.6
2741–1792	5.25	0.2	−0.1	0.6	11.0
2741–670	4.96	0.1	−0.1	0.4	10.8

low for these thick layers and the data prefer a positive density anomaly (Table 4).

Second, we invert for $\bar{\rho}$ in three layers that exclude the D″ region, i.e. layers ranging from the top of the D″ region and upwards: (i) 2741–2376 km, (ii) 2741–1792 km, and (iii) 2741–670 km. For all three layers, the MAP estimates indicate a preference for a weak yet positive density anomaly (θ = 0.1% or 0.2%), but the data do not uniquely constrain the existence of a density anomaly nor its sign on the 2σ level (Table 4).

Thus, with the data used here we cannot uniquely determine whether a density excess with respect to *PREM* is distributed throughout the lower mantle. If we compare the pdfs for the target parameters including and excluding the D″ region, the data clearly prefer a density excess in the D″ layer, as was evident from Table 3.

5.3.5. The upper mantle

We train networks for $\overline{V}_P, \overline{V}_S$ and $\bar{\rho}$ in three layers in the upper mantle: (i) the transition zone (TZ), which is bordered by the "660" and "410" discontinuities, (ii) the region between the "410" and the "220" ("410–220") and (iii) the uppermost mantle between the "220" and the Moho ("220-Moho"). All nine parameters are resolved by the data, although to varying degrees (Table 5). In particular, the data contain much information on density ($D_{KL} \geqslant 7.4$) and the parameters in the TZ ($D_{KL} \geqslant 4.7$).

Table 5
Information gain D_{KL} in bits for $\overline{V}_P, \overline{V}_P$ and $\bar{\rho}$ in the TZ, the "410–220" region and the "220-Moho" region (Fig. 4).

	TZ	"410–220"	"220-Moho"
$\overline{V}_P$	8.2	3.5	5.5
$\overline{V}_S$	4.7	7.3	3.4
$\bar{\rho}$	9.7	7.4	7.5

The TZ is 270 km wide in *PREM*, but its thickness varies throughout the training data set with the change in depth of the two enclosing discontinuities. The MAP estimates for the mean $\overline{V}_P$ and $\overline{V}_S$ deviate respectively 0.5 and −0.7% from *PREM*, but the data do not uniquely determine the existence of a deviation from *PREM* in either velocity on the 2σ level (Fig. 4). For the average TZ density, most of the probability mass in the pdf lies at values lower than in *PREM*, with a most probable estimate of −0.5%, in contrast to a positive density anomaly found by Beghein et al. (2006). However, we note that the posterior pdf inferred by Beghein et al. (2006) is relatively wide, i.e. extends down to −2% with respect to *PREM* on the 2σ level; our uncertainty estimate is narrower and falls within this range. For the "410–220" region, we find a negative deviation from *PREM* for V_P ($\theta \pm 2\sigma$ = −6.0%–−1.4%). A possible deviation from *PREM* is not constrained on the 2σ level for $\overline{V}_S$ and $\bar{\rho}$.

The uppermost mantle, which is enclosed by the "220" and the Moho, is radially anisotropic in our parametrisation. Therefore, we analyse the Voigt averages of V_P and V_S, the density and the anisotropic parameter η. The Voigt averages can be interpreted as the isotropic representation of the wave velocities in an anisotropic medium and are defined as $V_P^{Voigt} = \sqrt{(V_{PV}^2 + 4V_{PH}^2)/5}$ and $V_S^{Voigt} = \sqrt{(2V_{SV}^2 + V_{SH}^2)/3}$ for V_P and V_S, respectively. The region is 195.6 km wide in *PREM*, but its thickness varies throughout the training data set with the change in depth of the two enclosing discontinuities. For the observed data, the MAP estimates for $\overline{V}_P^{Voigt}$ and $\overline{V}_S^{Voigt}$ are respectively 3.1% and 1.6% higher than in *PREM*, which lies outside the $\pm 2\sigma$ range for $\overline{V}_P^{Voigt}$ (Fig. 4). The higher V_P^{Voigt} and V_S^{Voigt} are in agreement with reference models such as *ak135* and *STW105*, which compensate for the absence of a discontinuity at 220 km with velocities in the overlying region that are higher than in *PREM* (Fig. 1). We will investigate the relation with the "220" further in Section 5.3.6. The average density in this region is in agreement with *PREM*. This disagrees with a likely negative density anomaly inferred by Beghein et al. (2006), but again we note that their pdf is more conservative, i.e. represents a larger uncertainty, than the 1-D marginal pdf we obtained here. The average anisotropic parameter $\bar{\eta}$ in the "220–Moho" region is weakly constrained by the data ($D_{KL} = 1.0$ bits), but the $\pm 2\sigma$ error levels span a wide range of 0.91–0.99 ($\bar{\eta} = 0.94$ in *PREM*).

Further, we train ensembles of MDNs on the jumps in V_P, V_S and density across the "660" and "410" discontinuities. For the "660", the posterior pdfs for V_P and V_S are centred on *PREM* (Table 6).

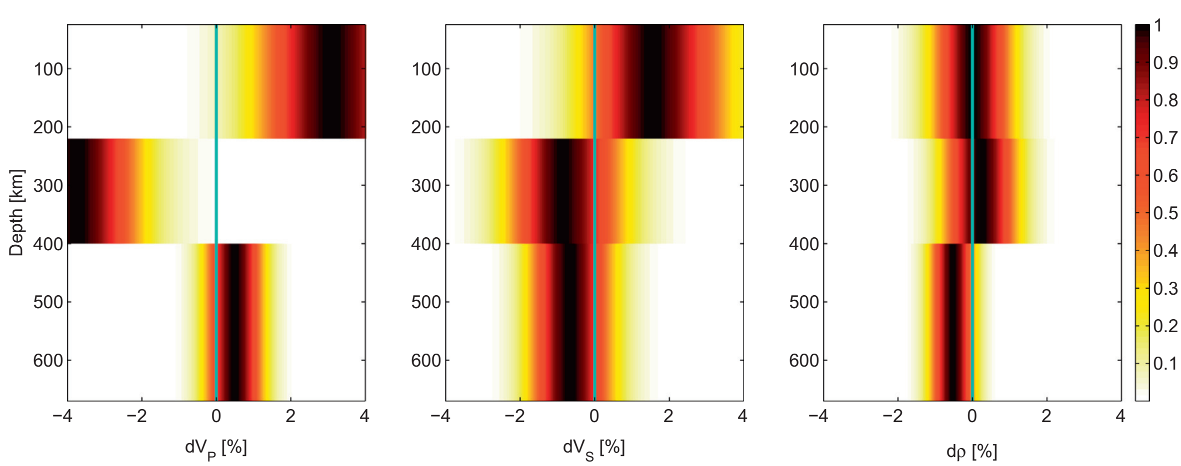

Fig. 4. 1-D marginal posterior pdfs for $\overline{V}_P^{Voigt}$ (left-hand panel), $\overline{V}_S^{Voigt}$ (middle panel) and $\bar{\rho}$ (right-hand panel) in the upper mantle, expressed as percentage deviations with respect to *PREM*. The probability for each 1-D pdf is rescaled so that the maximum equals 1. Asymmetric 1σ and 2σ error bars correspond to the $1/e^{1/2}$ (0.61) and $1/e^2$ (0.14) contours, respectively.

Table 6
Similar to Table 1, but for the jumps in V_P, V_S and ρ across the "660" and "410" discontinuities.

	PREM	*ak135f*	θ	$\theta - 2\sigma$	$\theta + 2\sigma$	D_{KL} [bits]
ΔV_P^{660} [km/s]	0.49	0.59	0.52	0.24	0.79	2.1
ΔV_S^{660} [km/s]	0.37	0.35	0.40	0.24	0.55	2.0
$\Delta\rho^{660}$ [g/cm^3]	0.39	0.31	0.43	0.34	0.51	4.5
ΔV_P^{410} [km/s]	0.23	0.33	0.60	0.23	0.95	2.0
ΔV_S^{410} [km/s]	0.16	0.21	0.17	−0.03	0.38	3.0
$\Delta\rho^{410}$ [g/cm^3]	0.18	0.42	0.14	0.02	0.26	6.1

The density jump is resolved best ($D_{KL} = 4.5$ bits) and tends towards slightly higher values. This is in agreement with our finding that the average density in the TZ may be slightly lower than in *PREM* (Fig. 4). For the "410", the jumps in V_S and density are in accord with *PREM*. The MAP estimate for the V_P contrast is relatively large ($\theta = 0.60$ km/s), although *PREM* (0.23 km/s) lies just within our uncertainty bounds.

5.3.6. Existence of a discontinuity at 220 km

The hotly-debated Lehmann discontinuity at 220 km depth is another good target for our flexible method. The discontinuity has been observed by many workers using several distinct data types (see e.g. Deuss et al. (2013) for a review). Consensus seems to have been reached on its regional existence, but its nature and whether the discontinuity extends globally are still debated. The controversy is illustrated by the presence, e.g. *PREM*, and absence, e.g. *ak135*, of the discontinuity in existing reference earth models.

We construct target parameters similar in fashion to those used to investigate the velocity and density contrasts at the "660" and "410" and invert for the jumps in V_P^{Voigt}, V_S^{Voigt} and density at the "220". Our aim is to investigate the probability that the contrasts at the "220" in these parameters is positive, as they are in *PREM*. We show a test set pattern in Fig. 5 with a near-zero jump, which we interpret as the absence of a discontinuity, to demonstrate the ability of the trained network ensemble to make correct predictions for such an earth model.

Table 7
Similar to Table 1, but for $\Delta V^{220}_{P(Voigt)}, \Delta V^{220}_{S(Voigt)}$ and $\Delta\rho^{220}$.

	PREM	θ	$\theta - 2\sigma$	$\theta + 2\sigma$	D_{KL} [bits]
$\Delta V^{220}_{P(Voigt)}$ [km/s]	0.56	0.10	−0.44	0.61	1.1
$\Delta V^{220}_{S(Voigt)}$ [km/s]	0.20	0.07	−0.17	0.31	1.6
$\Delta\rho^{220}$ [g/cm^3]	0.08	0.09	−0.09	0.27	3.4

We find that the data prefer contrasts in V_P^{Voigt} ($\theta = 0.10$ km/s) and V_S^{Voigt} ($\theta = 0.07$ km/s) at the "220" that are smaller than in *PREM* (Table 7). However, the information gain is relatively low ($D_{KL} \leqslant 1.6$ bits) and the 1-D posterior pdfs include both zero (no jump) and *PREM*-like values (Fig. 5). The smaller jump in V_P^{Voigt} is in agreement with the higher $\overline{V}_P^{Voigt}$ in the "220-Moho" region and the lower $\overline{V}_P$ in the "410–220" region (Fig. 4). Similarly, the smaller contrast in V_S^{Voigt} corresponds to a higher $\overline{V}_S^{Voigt}$ in the "220-Moho" region. By contrast, the density contrast is similar to *PREM*, although the data do not exclude a (near-) zero amplitude on the 2σ level. The probability of a positive jump can be extracted from the marginal pdfs and is 0.63 $\left(\Delta V^{220}_{P(Voigt)}\right)$, 0.72 $\left(\Delta V^{220}_{S(Voigt)}\right)$ and 0.84 $(\Delta\rho^{220})$. Thus, the normal mode data cannot determine with certainty the presence or absence of a discontinuity at 220 km in the radial Earth structure, particularly for V_P^{Voigt} and V_S^{Voigt}. If a discontinuity exists, the jump in density and both velocities is probably small (~0.1 g/cm^3, ~0.1 km/s).

5.3.7. Attenuation

We train MDNs that take the normal mode Q measurements as input and produce bulk and shear attenuation parameters, $1/Q_\kappa$ and $1/Q_\mu$, respectively, as output. The 13 Q_κ and Q_μ parameters are independently drawn from prior distributions that are uniform on a logarithmic scale (A.2). Fig. 6 shows the 1-D posterior pdfs for the 13 Q_μ and Q_κ parameters, as well as the 1-D earth models *PREM*, *ak135f* (Montagner and Kennett, 1996) and *QL6* (Durek and Ekström, 1996). In addition, we show the most probable values obtained from a sampling-based inversion of surface wave and normal mode attenuation measurements (Resovsky et al., 2005).

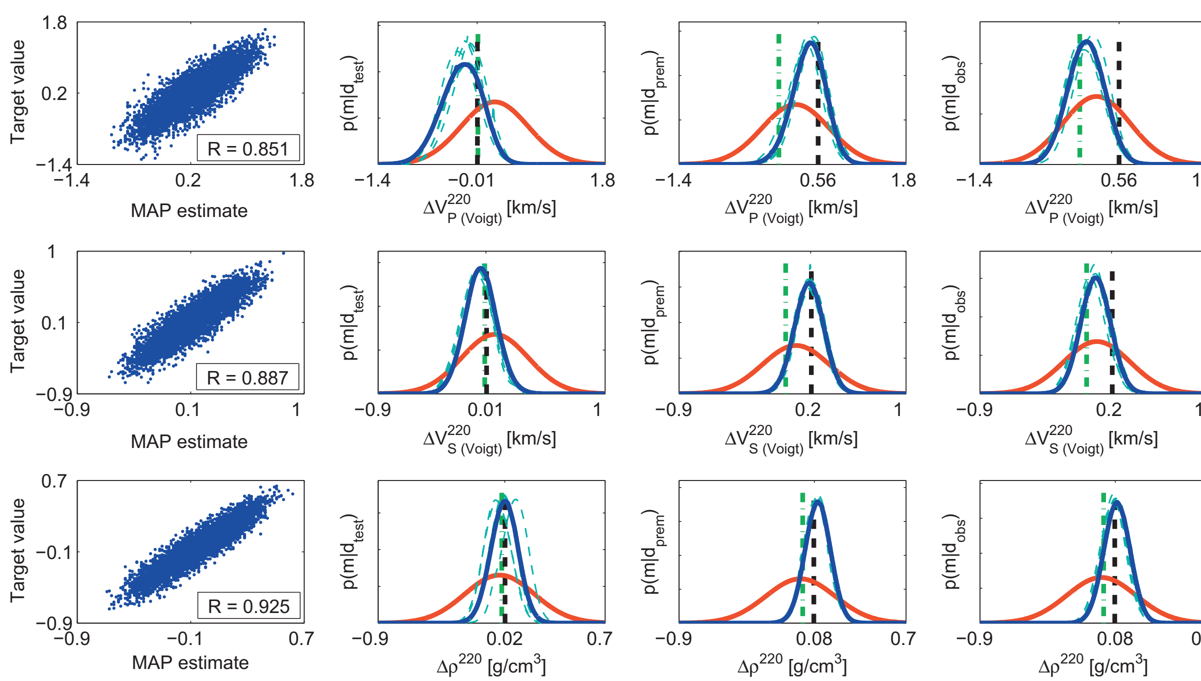

Fig. 5. Ensembles of MDNs trained on $\Delta V^{220}_{P(Voigt)}$ (first row), $\Delta V^{220}_{S(Voigt)}$ (second row) and $\Delta\rho^{220}$ (third row). The green dashed-dotted line indicates no jump, such as in *ak135*. The panels are similar to those shown in Fig. 2. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

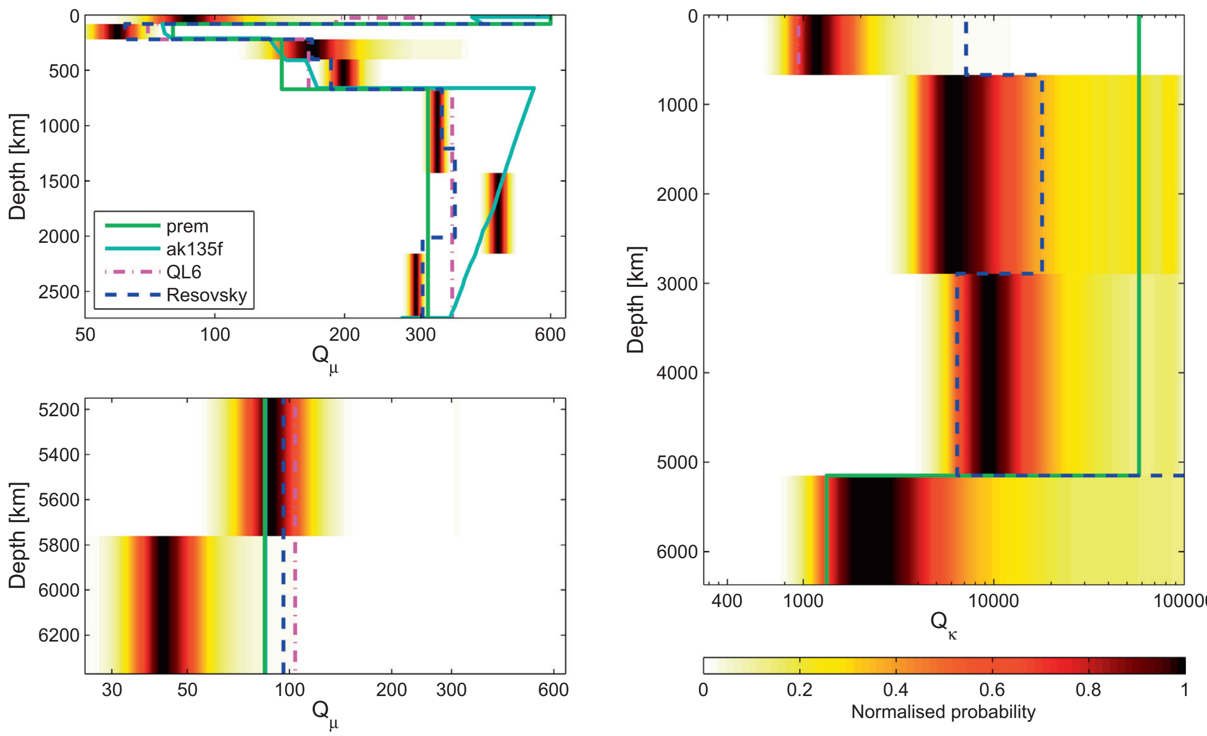

Fig. 6. 1-D marginal posterior pdfs for Q_μ (left-hand panels) and Q_κ (right-hand panel). The probability for each 1-D pdf is rescaled so that the maximum equals 1. 1σ and 2σ error levels correspond to the $1/e^{1/2}$ (0.61) and $1/e^2$ (0.14) contours, respectively. Several 1-D earth models and the most probable values from Resovsky et al. (2005) are added for comparison.

Note that our inferences and the mentioned 1-D earth models are based on the assumption of frequency-independent attenuation, despite the consensus that $Q_\mu \propto \omega^\alpha$ with α between 0.1 and 0.5 (e.g. Romanowicz and Mitchell, 2007), which may result in a bias in estimates of Q_μ (Lekić et al., 2009).

The information content is $D_{KL} \geqslant 3.1$ bits for all Q_μ parameters (Table 8). This indicates that the data are sensitive to the attenuation structure in both inner core and mantle, given the particular model parametrisation we haven chosen. For most Q_μ parameters, we find that our MAP estimates deviate from *PREM* in a similar fashion to the most probable values reported by Resovsky et al. (2005). Notable exceptions are the high-velocity lid and crust, for which we find a lower Q_μ, and the middle layer in the lower mantle, for which the data prefer a higher Q_μ that is similar to the value in *ak135f*.

Table 8
Information gain D_{KL} in bits for the 13 bulk and shear attenuation parameters.

Q_μ	Region	D_{KL} [bits]
	Inner core (IC)	
	6371–5760 km	3.4
	5760–5150 km	8.3
	Outer core (OC)	–
	Lower mantle (LM)	
	2891–2157 km	13.2
	2157–1428 km	12.3
	1428–670 km	13.1
	Transition zone (TZ)	11.5
	410–220	6.3
	Low-velocity zone (LVZ)	9.4
	High-velocity lid + crust	5.1
Q_κ	Region	D_{KL} [bits]
	Inner core (IC)	2.2
	Outer core (OC)	6.4
	Lower mantle (LM)	6.3
	Upper mantle (UM)	2.8

In comparison to the 1-D models shown in Fig. 6, we parametrised the inner core shear attenuation using an additional layer. The shear Q in the upper half of the inner core is in agreement with existing models ($\theta = 89$), but we observe a preference for a lower shear Q in the lower half of the inner core ($\theta = 43$). We estimate the compressional attenuation using these values for Q_μ, our MAP estimate for Q_κ of 2243, $V_P = 11.1$ km s^{-1}, $V_S = 3.6$ km s^{-1} and the relation (Anderson and Hart, 1978)

$$Q_\alpha^{-1} = \frac{4}{3}\left(\frac{V_S}{V_P}\right)^2 Q_\mu^{-1} + \left[1 - \frac{4}{3}\left(\frac{V_S}{V_P}\right)^2\right] Q_\kappa^{-1}. \quad (2)$$

We find $Q_\alpha \approx 510$ and $Q_\alpha \approx 274$ in the upper and lower half of the inner core, respectively. This contrasts with the consensus in the literature that both compressional and shear attenuation decrease with depth in the inner core (see Romanowicz and Mitchell (2007) for a review). Andrews et al. (2006) showed that measurements of inner core Q could be biased by neglecting mode coupling. This could influence the data we use here, which only contains self-coupled modes, and requires further investigation.

The existence of an innermost inner core, with approximate radius between 300 and 600 km, has been suggested previously based on studies of anisotropy, e.g. Ishii and Dziewoński (2002); Beghein and Trampert, 2003, and attenuation, e.g. Li and Cormier (2002); Cormier and Stroujkova, 2005, in the inner core, although Lythgoe et al. (2014) find that an innermost inner core is not required to explain *PKIKP* travel times. A review of studies on the inner core structure and dynamics can be found in e.g. Alboussière and Deguen (2012); Deguen, 2012. Admittedly, we cannot address the structure of the inner core in detail within the limitations of our two-layer parametrisation. However, we verified that the MDNs make accurate predictions for *PREM* and test set models with similar Q_μ values in the two inner-core layers. By contrast, the normal mode data used here strongly prefer a non-homogeneous structure.

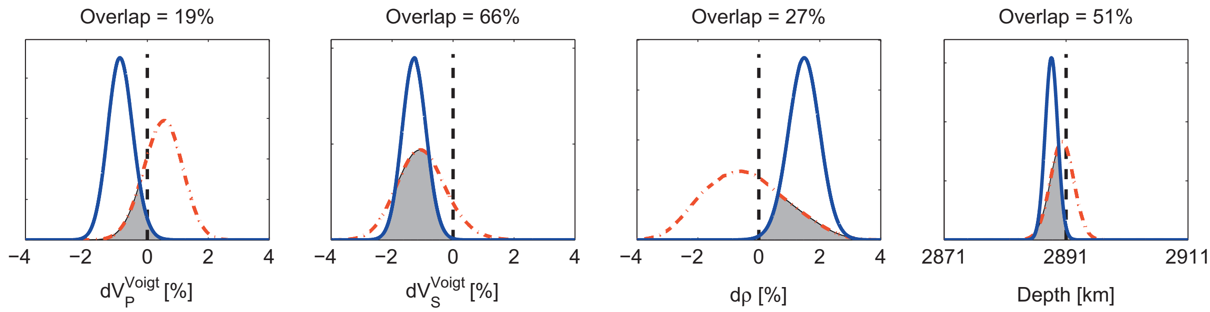

Fig. 7. 1-D marginal posterior pdfs for the partly (blue, solid) and fully (red, dashed-dotted) anisotropic model parametrisation for $\overline{V}_P^{Voigt}$ (left-hand panel), $\overline{V}_S^{Voigt}$ (middle-left panel), $\bar{\rho}$ (middle-right panel) in the D″ region and the CMB depth (right-hand panel). *PREM* is shown as a reference (black, dashed) and the percentage of overlap between the two pdfs is shown above each panel. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The data contain less information on Q_κ, most notably for the bulk attenuation in the inner core and upper mantle (Table 8). Our results indicate that Q_κ in the mantle and the outer core is lower than in *PREM*. In the upper mantle, our MAP estimate for Q_κ ($\theta = 1338$) is similar to model *QL6*. In contrast to Resovsky et al. (2005), we find a relatively low Q_κ in the inner core ($\theta = 2243$) that is similar to *PREM*. However, we note that the posterior pdf is strongly asymmetric and spans several orders of magnitude ($\theta + 2\sigma = 195349$, Fig. 6).

6. Discussion

6.1. Trade-offs with anisotropy

Since spheroidal mode data alone do not fully constrain radial anisotropy, we showed results for the *PREM*-like parametrisation, in which only the uppermost mantle is characterised by radial anisotropy. However, it is instructive to investigate whether trade-offs between (anisotropic) parameters exist, in light of the information contained in the spheroidal modes. To this end, we train MDNs on the same target parameters as in Section 5, but for the second parametrisation with an anisotropic inner core and mantle, and compare the resulting 1-D posterior pdfs with those for the isotropic parametrisation (Figs. 7 and D.2). To quantify their similarity, we compute the percentage of overlap between the pdfs; a low overlap indicates that trade-offs related to potential anisotropy influence the result of the inversion. The discrepancies between the pdfs illustrate where we need to constrain radial anisotropy in the model. As an example, we discuss the average velocities and density in the D″ region and the CMB depth and analyse possible trade-offs between anisotropic parameters.

For the D″ region, we find a strong discrepancy between the pdfs for the two parametrisations (Fig. 7). Additional trade-offs result in larger uncertainties and hence a lower information gain, especially for the density (Table 9). This may not be surprising, as the D″ region is in general believed to be (radially) anisotropic, at least in terms of V_P and the anisotropic parameter η, e.g. Montagner and Kennett (1996); Beghein et al., 2006. The sign of

Table 9
Information gain D_{KL} in bits for $\overline{V}_P, \overline{V}_P$ and $\bar{\rho}$ in the D″ region and the CMB depth for the partly and fully anisotropic model parametrisation (Fig. 7).

Anisotropy	Partly	Fully
$\overline{V}_P^{Voigt}$	6.9	2.3
$\overline{V}_S^{Voigt}$	8.1	1.4
$\bar{\rho}$	6.0	0.3
CMB	13.8	8.7

a possible deviation from *PREM* is not constrained on the 2σ level for both velocities and density (red lines, Fig. 7). Further, we observe that the MAP estimates for $\overline{V}_P^{Voigt}$ and $\bar{\rho}$ are opposite in sign compared to the results for the isotropic parametrisation.

We further investigate whether the weaker constraint is the result of trade-offs related to anisotropic parameters. Therefore, we construct 2-D marginal pdfs for two model parameters m_1 and m_2 using the decomposition, e.g. Tarantola (2005),

$$\sigma(m_1, m_2|\mathbf{d}) = \sigma(m_2|m_1, \mathbf{d})\sigma(m_1|\mathbf{d}), \qquad (3)$$

where the l.h.s. 2-D marginal pdf is given by the product of the 1-D marginal pdf $\sigma(m_1|\mathbf{d})$ and the conditional pdf $\sigma(m_2|m_1, \mathbf{d})$, i.e. the pdf for m_2 conditioned on m_1. All pdfs in Eq. 3 are conditioned on the observed data $\mathbf{d}$. For both pdfs in the r.h.s. product, a separate MDN is trained. Note that it is straightforward to train MDNs on conditional pdfs; this merely requires the conditional model parameter(s) to be appended to the input pattern (de Wit et al., 2013).

Fig. 8 shows an example of 2-D pdfs for $\bar{\eta}$ versus both $\overline{V}_P^{Voigt}$ and $\overline{V}_S^{Voigt}$ in the D″ region. $\bar{\eta}$ is unresolved by the data, as is evident from the 1-D marginal pdf (left-hand panels). In both cases, the conditional and the 2-D marginal pdfs show a clear correlation between $\bar{\eta}$ and the velocities. The trade-offs in the 2-D pdfs indicate that the data do not uniquely constrain the individual parameters. If we impose isotropy in the D″ region *a priori*, i.e. $\eta = 1$, we would infer a different $\overline{V}_P^{Voigt}$ and $\overline{V}_S^{Voigt}$ than if anisotropy is present.

6.1.1. Constraining radial anisotropy

For some of our model parameters, we do not obtain a similar pdf for the two parametrisations (Figs. 7 and D.2), which highlights the need to constrain radial anisotropy in the whole earth model. We can potentially resolve these (anisotropic) parameters better, and thereby reduce trade-offs between parameters, if we augment the spheroidal mode data with complementary measurements, such as toroidal modes and surface waves. We did not use such data here and limited our inversion to spheroidal mode data. However, the input to a neural network can easily be extended by such additional data; the same flexibility applies as we have already highlighted for the target parameters.

Finally, we note that our radially anisotropic parametrisation in the inner core ignores a part of the existing knowledge of this region. This choice was deliberate, since we used centre frequencies and quality factors that are only sensitive to the radial (degree-zero) structure. Inner core anisotropy can be better described by a cylindrical symmetry with the symmetry axis aligned with the Earth's rotation axis, e.g. Morelli et al. (1986), Woodhouse et al. (1986). However, such cylindrical symmetry manifests itself in splitting function coefficients of higher degrees, which we did not use in this work. We checked whether the assumption of radial anisotropy in the inner core biased the results of the second (anisotropic) parametrisation by constructing a new

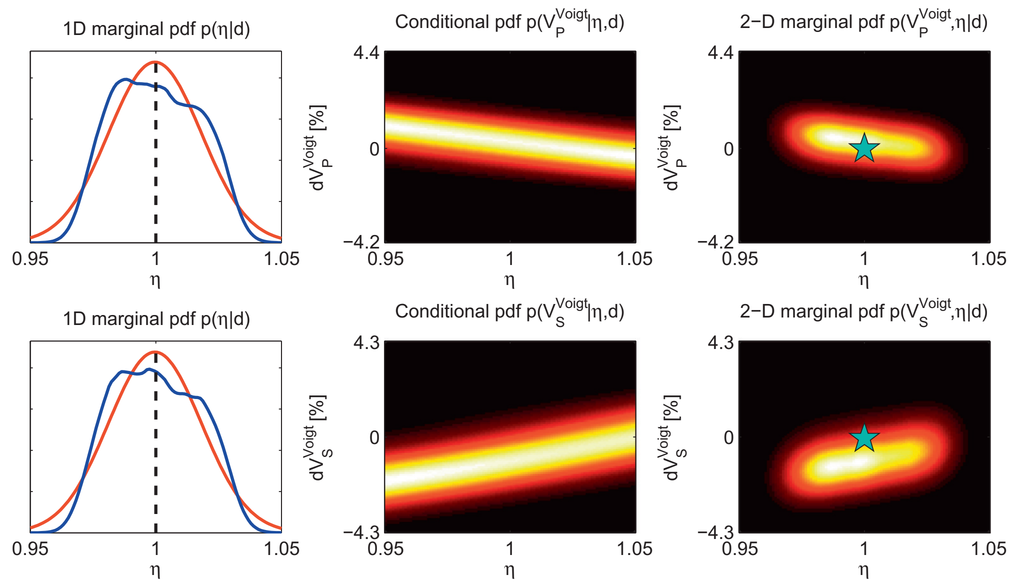

Fig. 8. Construction of 2-D marginal posterior pdfs via Eq. 3 for $\overline{V}_P^{Voigt}$ (top row) and $\overline{V}_S^{Voigt}$ (bottom row) versus $\bar{\eta}$ in the D″ region. The three panels in each row show the 1-D marginal (blue) and prior (red) pdfs (left-hand panel), the conditional pdf of the velocity given $\bar{\eta}$ and the observed data **d** (middle panel) and the 2-D marginal pdf of $\bar{\eta}$ and the average velocity for the observed data **d** (right-hand panel). Lighter colours denote higher probabilities. The corresponding *PREM* values are denoted by the black line (left-hand panel) and the cyan star (right-hand panel). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

training set with a radially anisotropic mantle and an isotropic inner (and outer) core. We trained MDNs on this new training set and found that the resulting pdfs were similar, with overlap $\geqslant 87\%$, to the pdfs for the parametrisation with a radially anisotropic inner core (not shown here). Therefore, we concluded that our analysis of trade-offs, as presented above, was not biased by either an isotropic or a radially anisotropic parametrisation in the inner core.

6.2. Joint inversion with travel time data

All results shown in this study are based on normal mode centre frequencies and quality factors. We also performed joint inversions of normal mode data and body wave travel times (E). We adopted a setup similar to de Wit et al. (2013), who used first-arrival travel time data from the EHB bulletin (Engdahl et al., 1998) for the phases *Pn*, *P*, *PP*, *PKP* for the years 2001 to 2008. We supplemented this data set with data for the *Sn* and *S* phases (Table E.5). For the joint inversion, the input to the network consisted of 184 free oscillations and 186 travel time measurements. The 186-D travel time vector was a concatenation of data for the Pn (8 travel time measurements), P (32), PP (62), PKPab (14), PKPbc (4), PKIKP (29), Sn (9) and S (28) phases (E).

However, we analysed the information content of the data and found that the travel time measurements do not provide additional information on Earth structure to the inversion of the normal mode data for the chosen parametrisation. We computed the information gain D_{KL} for an inversion of the normal mode data, an inversion of the travel time data and a joint inversion of the two data sets. As an example, Fig. 9 shows the information gain for the three different inversions for the depth of four discontinuities. With the exception of the Moho depth, and the velocities near this discontinuity, we found that the travel time data have a low information content compared to the normal mode data for most model parameters. Consequently, the joint inversion did not yield a better constraint on the model parameters than a separate inversion of the normal modes. We performed this analysis for a synthetic data set in which the bulk and shear attenuation structures were fixed to that of *PREM*, as body wave travel times have very little sensitivity to the attenuation parameters.

Based on this analysis, we have focused on results without travel time data. By removing the travel times, we reduced the size of the (input of the) neural network, the required number of training patterns and thus computation time, whilst retaining the information on the model parameters. This does not mean that the information in travel times is never complementary to the information contained in the normal mode data; their relative contributions to our inversions relate to the choices we have made in our setup. A joint inversion of normal modes and travel times may be beneficial if one is able to construct a more complete noise model for the travel times and include additional seismic phases. In addition, more information on Earth structure may be available when using travel time picks of higher quality than the data in the ISC (and EHB) catalogues, as these bulletins contain measurements of varying precision.

6.3. Shift in ICB depth

We evaluated the ICB depth for the fully anisotropic parametrisation (Fig. D.2) and found that the pdfs for the two parametrisation overlap by 69%. The pdf for the anisotropic case is wider and corresponds more to *PREM* (5149.5 km) and *ak135* (5153.5 km), but most of the probability mass is still assigned to greater depths ($\theta = 5157.8$ km and $\theta \pm 2\sigma = 5149.9 - 5165.7$ km). For the isotro-

Table E.5
Epicentral distance range for the seismic phases used in this study and by Kennett et al. (1995) and de Wit et al. (2013).

Distance range [°]	Pn	P	PP	PKPab	PKPbc	PKIKP	Sn	S
This study	3:18	25:88	50:173	147:173	147:153	122:179	2:19	25:80
de Wit et al. (2013)	3:18	25:88	50:173	145:174	145:155	122:179	–	–
Kennett et al. (1995)	–	25:99	53:180	156:178	151:153	118:180	–	25:80

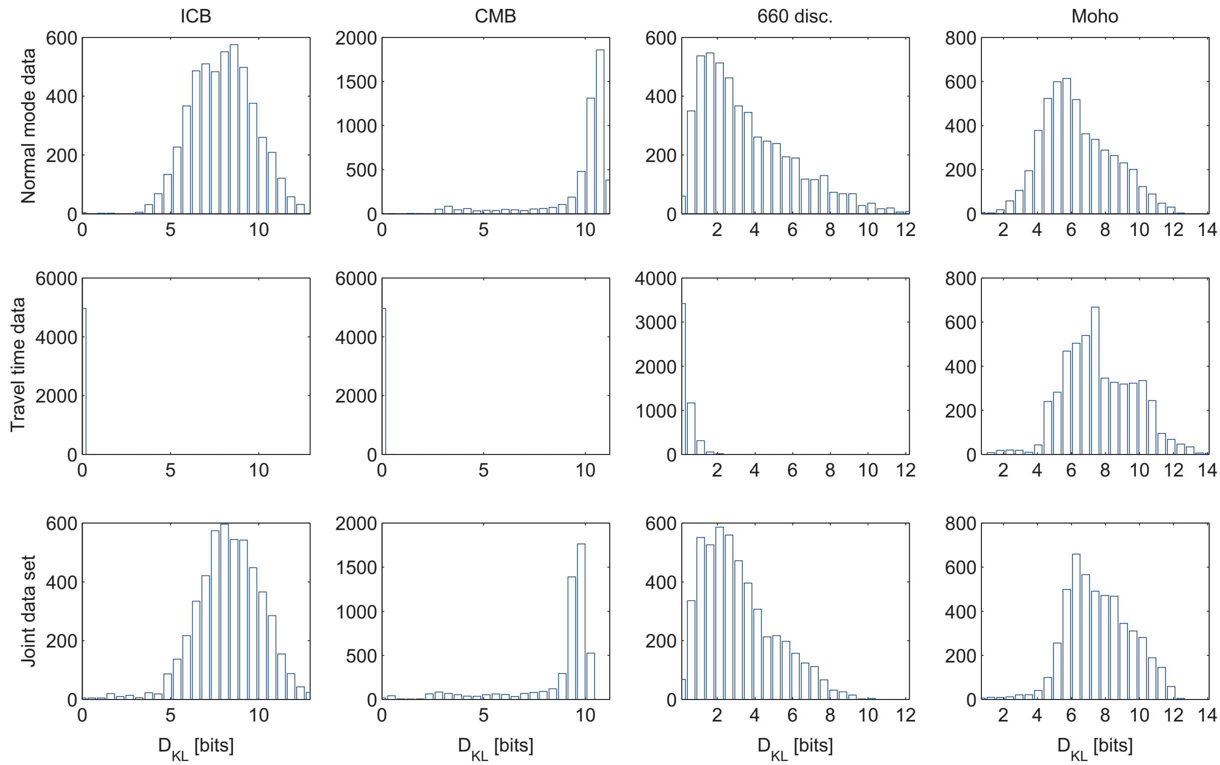

Fig. 9. Information gain, as measured by the Kullback–Leibler divergence D_{KL} (Eq. B.7), for networks trained on the normal mode data (first row), the travel time data (second row) and the joint data set (third row). D_{KL} (in bits) is shown for the depths of the ICB, the CMB, the 660 km discontinuity and the Moho for the ~5000 patterns in the test set.

pic case, our MAP estimate is 5160.1 km compared to the values in *PREM* and *ak135*, or earlier estimates, e.g. 5155.0 ± 1.0 km (Bolt, 1977) and 5153.9 km in *PEM* (Dziewoński et al., 1975).

Could such a shift in ICB depth be detected by body waves sensitive to the inner core? We investigate the effect on the travel times of two inner core sensitive phases *PKIKP* and *PKiKP* for a modified version of *PREM* with the ICB depth at 5160 km, denoted $PREM_{ICB}$. Travel times for *PKIKP* and *PKiKP* are typically studied in an epicentral distance range of 130 to 143 degrees to avoid interference between the two phases and the outer core sensitive *PKPab* and *PKPbc* phases, e.g. Waszek et al. (2011), Waszek and Deuss (2013). We compute the travel time difference between *PREM* and $PREM_{ICB}$ in this distance range at one degree intervals using TauP and a surface source. The average difference is 0.38 s ($\sigma = 0.06$ s) for *PKIKP* and 0.33 ($\sigma = 0.08$ s) for *PKiKP*.

For the *PKIKP* data we considered in this study, the estimated noise, i.e. the sample standard deviation, varies between 0.18 s and 0.43 s in the 130 to 143 degree distance range. Thus, the effect of this shift in ICB depth on the travel times is on average of the same order of magnitude as the assumed data noise, albeit larger for some distances. The ICB depth may be difficult to constrain accurately using these body wave phases, considering that the data are affected by more than just a simple shift in ICB depth, i.e. considering additional trade-offs between other model parameters. Note that we did not use *PKiKP* travel times in this study (Table E.5); we simply consider TauP synthetic data for *PKIKP* and *PKiKP* for the straightforward analysis performed here.

6.4. Gradients

To understand the Earth's thermochemical structure and dynamics requires an accurate knowledge of the sensitivities of velocity and density with respect to temperature and composition, e.g. Deschamps et al. (2007). The gradients in the radial distribution of these elastic parameters can provide important constraints. For instance, several phase transitions are necessary to explain the gradients in the upper mantle of seismological models, e.g. Cammarano et al. (2005), Cobden et al. (2009).

An insightful first-order inference would be to constrain the sign of the gradient in a certain region. We investigated the potential of the normal mode data to constrain the depth gradients in velocity and density in our model. However, we found that in general very little information on gradients is directly available, especially if the gradient is evaluated over relatively narrow regions of ~200–300 kilometres.

6.5. A note on the number of synthetic samples

Compared to model space searches with Monte Carlo methods, e.g. Sambridge (1999), Bodin and Sambridge (2009), which commonly involve ~10^6 models, we have a relatively low number of samples (100,000). However, with the setup presented here, we are still able to draw insightful inferences on the Earth's radial structure for reasons addressed here. First, the success of the training process and accuracy of the trained networks can be verified by testing network performance on the 5000 test set samples and synthetic data for *PREM* (see Figs. 2, 3 and 5). Second, before network training commences the MDN output is initialised to resemble the prior pdf for the target parameter. If no systematic relation between input (seismological data) and output (earth model parameter) is found during training, no information is extracted from the data. Consequently, the final MDN output will resemble the prior, the information gain $D_{KL} = 0$ bits (Appendix C) and two possibilities exist.

First, the data may simply not be sensitive to the given earth model parameter, compared to the assumed data noise. Second, the set of training samples, which we generated from the prior model space, may not include all regions of the model space of non-zero probability, which is a potential issue of any sampling-based method. As an additional check, we trained networks with a larger set of models (200,000 samples) for some of the target parameters investigated in Section 5. We found that for this larger training set, uncertainties decrease slightly for some parameters (the width of the posterior pdf decreases; D_{KL} is a bit higher), but the bulk of probability mass is assigned to similar values as for the set with 100,000 earth models. More importantly, we find that in no case the pdf is wider for the larger training set, which would indeed be undesirable. In that sense, the method is conservative and therefore allows us to make insightful inferences on Earth structure.

7. Conclusions

We used artificial neural networks to obtain 1-D marginal posterior pdfs for parameters in a radial earth model, thereby solving seismological inverse problems. We have illustrated the flexibility of the method, which offers the freedom to invert for an arbitrary combination of model parameters and allows us to test specific hypotheses.

The 1-D distributions can be used to quantify uncertainties in the model parameters. In addition, they provide a basis to measure the information contained in the normal mode splitting function measurements, using the Kullback–Leibler divergence. The information content for the various data sets showed that the free oscillations constrain most parameters better than the travel time data. By removing the travel times, the MDNs had fewer free parameters and thus solving the inverse problem was computationally more efficient, whilst the information on the earth model was retained.

The results of our inversions can be summarised as follows:

1. The spheroidal mode data constrain the depths of the ICB, CMB, "660" and Moho. The data strongly prefer an ICB that lies deeper (5154.7–5165.7 km) than in existing reference models; the result is robust with respect to a radially anisotropic inner core, although less pronounced (5149.9–5165.7 km). The effect on the travel time of inner core-sensitive seismic phases is comparable to the estimated noise in such measurements.
2. The most probable value for the density contrast at the ICB ($0.73\ \mathrm{g\,cm^{-3}}$) is larger than in *PREM* ($0.60\ \mathrm{g\,cm^{-3}}$) and *ak135f* ($0.56\ \mathrm{g\,cm^{-3}}$); the 2σ error levels span a range of 0.52–$0.94\ \mathrm{g\,cm^{-3}}$, which encompasses all previous estimates in the literature. With the data used here, we cannot differentiate between these estimates; their respective differences reflect the non-uniqueness of the inverse problem.
3. We observed a negative deviation with respect to *PREM* for both V_P and V_S in the D″ layer, in contrast to a positive anomaly found for the average density. However, these deviations are not robust with respect to a fully anisotropic parametrisation. The data cannot uniquely determine whether the possible density excess is restricted to the D″ region or whether it is distributed throughout the lower(most) mantle.
4. In the upper mantle, the strongest deviations from *PREM* were observed for V_P^{Voigt} and V_S^{Voigt} in the "220-Moho" region and V_P in the "410–220" layer, although the latter result is not robust with respect to the anisotropic parametrisation.
5. We found that the data cannot uniquely determine the presence or absence of a discontinuity at the "220". If present, the V_P^{Voigt} and V_S^{Voigt} contrasts are likely smaller than in *PREM*, while the density jump is similar to *PREM*.
6. The MAP estimates for most shear attenuation parameters in the mantle deviate from *PREM* in a similar fashion to results from more recent studies. The data strongly prefer a non-homogeneous shear attenuation in the inner core, enforcing the hypothesis that a distinct innermost inner core may exist.
7. The bulk attenuation is stronger than in *PREM*, save for the inner core, for which most probability is assigned to a Q_κ higher than in *PREM*.

We have also addressed the influence of radial anisotropy on the inversions. We compared the results with the posterior pdfs for a parametrisation in which the inner core and mantle were anisotropic. This enabled us to analyse possible trade-offs between (anisotropic) parameters by constructing conditional and 2-D pdfs. The largest discrepancies were observed in regions that are believed to be anisotropic, such as the D″ region. This illustrates the need to constrain anisotropy in the (deep) mantle and suggests the addition of complementary data, such as toroidal modes.

Acknowledgements

We are grateful for the constructive comments of two anonymous reviewers. We thank Paula Koelemeijer for providing additional normal mode measurements. Ralph de Wit, Paul Käufl and Andrew Valentine are funded by the Netherlands Organisation for Scientific Research (NWO) under the grant ALW Top-subsidy 854.10.002. Computational resources for this work were provided by the Netherlands Research Center for Integrated Solid Earth Science (ISES 3.2.5 High End Scientific Computation Resources). We acknowledge the International Seismological Centre for making the EHB bulletin available.

Appendix A. Model parametrisation

A.1. Velocity and density structure

We define V_P, V_S and ρ on a discrete set of 185 grid points (as one of the options in the Mineos package (Masters et al., 2011)), the depths of which range from the Earth's surface to its centre. We parametrise the depths of seven discontinuities in the model: the inner–outer core boundary (ICB) and core-mantle boundary (CMB), the top of the D″ layer, the discontinuities around 660, 410 and 220 km depth ("660", "410" and "220", respectively) and the Moho. The lower mantle (LM) represents the region between the top of the D″ layer and the "660", while the transition zone (TZ) spans the region between the "660" and the "410". For both velocities and density, the core and mantle are defined at 157 points in the earth model (Table A.1). The remaining 28 points in the earth model represent the crust, which is parametrised by two homogeneous layers. No sediment or water layers are present.

We consider two different parametrisations for anisotropy. In a first case, we include radial anisotropy only in the uppermost mantle between the Moho and the "220", as in *PREM* (Dziewoński and Anderson, 1981). A radially anisotropic, or trans-

Table A.1
Number of grid points L between each pair of adjacent discontinuities in the earth model. The full model is defined on a discrete set of 185 grid points.

Region	L
IC	33
OC	33
D″	5
LM	59
TZ	9
"410–220"	9
"220–Moho"	9
LC	11
UC	17

versely isotropic, medium has hexagonal symmetry with a radial symmetry axis and can be described by five independent parameters (Love, 1927). Here, we parametrise the anisotropy by the vertically polarised P- and S-wave velocities (V_{PV}, V_{SV}), the horizontally polarised velocities (V_{PH}, V_{SH}) and the anisotropic parameter η on the nine points in this region. In the rest of the model, the wave velocities are isotropic and $\eta = 1$.

Tables A.2–A.4 define the prior model distribution $p(\mathbf{m})$ for this first case. Discontinuity depths are independently drawn from uniform priors, as are V_{PV}, V_{SV} and ρ directly below the seven discontinuities and V_{PH}, V_{SH} and η below the Moho (Table A.2). The prior distributions are centred on the corresponding values in *PREM*. For the upper mantle and D″, we chose priors similar to those of de Wit et al. (2013). We define narrower priors for the core and lower mantle, taking into account the constraint on these parameters provided by body wave travel times (de Wit et al., 2013). In the "220-Moho" region and directly below the "220" discontinuity, we allowed for the largest variations in our prior, since existing models, such as *PREM* and *ak135f* (Montagner and Kennett, 1996), vary strongly in this region.

To exclude physically implausible models from the prior model space, we introduce correlations between adjacent points (layers) in each region, i.e. between discontinuities. First, we draw the value of the independent point directly below a discontinuity. We then use this value and the local gradient in *PREM* to calculate the value for the underlying point. Subsequently, this value is perturbed, with the amount of perturbation drawn from a uniform prior (Table A.3). This procedure is performed for all the points successively with increasing depth and introduces a correlation between the parameters in each region. In general, the radial velocities and density increase with depth, i.e. the velocity and density gradients are mostly positive. The η profile in the uppermost mantle is constructed in a similar fashion.

Table A.2
Prior information on *independent* model parameters. Prior distributions are uniform over the specified ranges, which are given as percentage perturbations from *PREM*, except for the discontinuity depths and the two crustal layers. V_P, V_S and ρ parameters represent the points located directly below a discontinuity. The tops of the lower mantle (LM) and the transition zone (TZ) are formed by the "660" and "410", respectively.

Discontinuity	Range [km]
ICB	5129.5–5169.5
CMB	2871–291
D″ layer (top)	2721–2761
"660"	640–700
"410"	370–430
"220"	200–240
Moho	20–70
V_P, V_S, ρ, η	Range [%]
Inner core (IC)	±2
Outer core (OC)	±2
D" layer	±3
Lower mantle (LM)	±2
Transition zone (TZ)	±5
"410–220"	
V_P	[−10,+2.5]
V_S	[−10,+5.0]
ρ	[−10,+5.0]
"220–Moho"	±7
Lower crust (LC)	
V_P [km/s]	6.4–7.4
V_S [km/s]	3.6–4.1
ρ [g cm^{-3}]	2.8–3.0
Upper crust (UC)	
V_P [km/s]	5.6–6.3
V_S [km/s]	3.1–3.6
ρ [g cm^{-3}]	2.6–2.8
Mass (10^{24} kg)	5.9733 ± 0.0090
Moment of inertia (10^{37} kg m^2)	8.018 ± 0.012

Table A.3
Prior information on *dependent* model parameters. Prior distributions are uniform over the specified ranges, which are given as percentage perturbations from the updated model value (see text). The corresponding independent parameters are listed in Table A.2.

V_P, V_S, ρ, η	Range [±%]
IC	0.5
OC	0.5
D″	1
LM	0.5
TZ	1
"410–220"	1
"220–Moho"	2

To avoid physically unrealistic 1-D density profiles, we constrain the mass and moment of inertia of the earth models using the error estimates reported by Chambat and Valette (2001). A model is discarded whenever its mass or moment of inertia does not lie within $(5.9733 \pm 0.0090) \cdot 10^{24}$kg or $(8.018 \pm 0.012) \cdot 10^{37}$kg m^2, respectively. Fig. 1 shows the parameter range spanned by the prior model space and a number of existing 1-D reference models for the upper mantle.

In the second case of model parametrisation, we allow radial anisotropy in the whole mantle and in the inner core. The parametrisation is the same as for the uppermost mantle in the first case, i.e. in terms of $V_{PV}, V_{SV}, V_{PH}, V_{SH}$ and η. The outer core is isotropic in both parametrisations.

A.2. Attenuation structure

The bulk and shear attenuation are parametrised by the inverses of Q_κ and Q_μ, respectively. We closely follow the parametrisation of Resovsky et al. (2005) and define the radial bulk and shear attenuation structure by 13 parameters. Q_κ is parametrised as four layers of constant attenuation: the inner core, the outer core, the lower mantle and the upper mantle. The latter two are separated by the "660". Note that the depths of the discontinuities separating these regions are free parameters in the earth model. Q_μ is parametrised by two and three layers of roughly equal thickness in the inner core and the lower mantle, respectively, and is zero in the outer core. We add a second layer to the inner core compared to the parametrisation of Resovsky et al. (2005). The upper mantle consists of four layers, which represent the TZ, the "410–220" region, the low-velocity zone (LVZ) between the "220" and 80 km depth in *PREM* and a layer encompassing both the overlying high-velocity lid and the crust.

The prior distributions are given in Table A.4. No correlations exist between the 13 parameters and all priors are uniform on a base-10 logarithmic scale. Fig. 1 shows the prior model range and existing 1-D attenuation models for the upper mantle.

Appendix B. Ensembles of MDNs

An MDN outputs the parameters of a Gaussian Mixture Model (GMM), which describes a conditional probability density for a set of model parameters $\mathbf{m}\prime$ as

$$\sigma(\mathbf{m}'|\mathbf{d};\mathbf{w}^*) \approx \sum_{j=1}^{M} \alpha_j(\mathbf{d};\mathbf{w}^*)\phi_j(\mathbf{m}'|\mathbf{d};\mathbf{w}^*), \tag{B.1}$$

where the coefficients α_j give the relative importance of the M Gaussian kernels ϕ_j (Bishop, 1995). Note the explicit conditioning on both the observed data $\mathbf{d}$ and the set of optimal network weights

Table A.4
Prior information on the attenuation parameters. Prior distributions are uniform on a base-10 logarithmic scale over the specified ranges.

Q_μ	Region	Range
	Inner core (IC)	
	6371–5760 km	10–300
	5760–5150 km	10–300
	Outer core (OC)	0
	Lower mantle (LM)	
	2891–2157 km	100–1000
	2157–1428 km	100–1000
	1428–670 km	100–1000
	Transition zone (TZ)	50–400
	"410–220"	50–400
	Low-velocity zone (LVZ)	20–200
	High-velocity lid + crust	10–2200
Q_κ	Region	Range
	Inner core (IC)	300–300,000
	Outer core (OC)	300–100,000
	Lower mantle (LM)	300–100,000
	Upper mantle (UM)	300–200,000

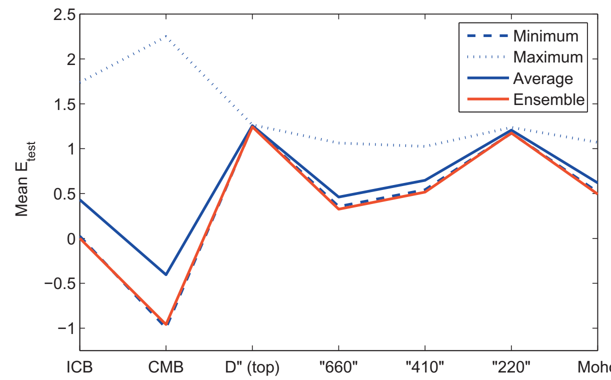

Fig. B.1. Comparison of the ensemble error (solid red, Eq. B.2) with the average (solid blue), minimum (dashed blue) and maximum (dotted blue) error for the ensemble members. All network members are applied to the same test set. The 1-D targets consist of the seven discontinuity depths, which are shown along the horizontal axis. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$\mathbf{w}^*$, i.e. the free parameters that minimise a cost function for a validation data set (de Wit et al., 2013).

Despite measures to maximise generalisation, such as the addition of noise, a neural network will be biased in its performance to the data sets used to train and validate it. Furthermore, a single network is sensitive to the initialisation of its weights. By combining the output of multiple networks into one *ensemble*, we aim to integrate out (marginalise over) the influence of the random initialisation of the network weights. The marginalisation of so-called nuisance parameters plays a central role in the Bayesian framework, e.g. MacKay (2003). Ensembles of networks can achieve better generalisation, i.e. can make more accurate predictions for unseen data, e.g. Bishop (1995).

The output of an ensemble of C MDNs can be constructed from a weighted average of the members (Käufl et al., 2014)

$$\sigma(\mathbf{m}'|\mathbf{d},\mathbf{w}^*_{i\in 1:C}) = \sum_{i=1}^{C} \frac{\omega_i}{\sum_j \omega_j} \sigma(\mathbf{m}'|\mathbf{d};\mathbf{w}^*_i), \tag{B.2}$$

where the individual weights ω_i are determined by each network's performance on the same test set

$$\omega_i = \exp\left\{-\frac{E(D_{test},\mathbf{w}^*_i)}{N}\right\}. \tag{B.3}$$

N is the number of samples in the test set $D_{test} = \{\mathbf{d}_n, \mathbf{m}'_n\}$ and $E(D_{test},\mathbf{w}^*_i)$ is the error for the ith member (Bishop, 1995)

$$E(D_{test},\mathbf{w}^*_i) = -\sum_{n=1}^{N} \ln[\sigma(\mathbf{m}'_n|\mathbf{d}_n;\mathbf{w}^*)]. \tag{B.4}$$

Effectively, the output of an ensemble of C MDNs is a GMM with $C \cdot M$ kernels of relative importance

$$\beta_{M\cdot i+j} = \frac{\omega_i}{\sum_j \omega_j} \cdot (\alpha_j)_i, \tag{B.5}$$

where $(\alpha_j)_i$ is the relative importance of the jth kernel in the ith ensemble member, akin to α_j in Eq. (B.1).

Admittedly, the use of a simple ensemble as proposed here is no replacement of the integration over the full weight space, e.g. Bishop (1995), Käufl et al. (2014). However, Bishop (1995) shows that the upper bound on the ensemble error is given by the average error of the individual networks. Fig. B.1 illustrates the advantage of an ensemble of MDNs. The performance of the ensemble error on a test set is favourable compared to the performance of the individual members; the ensemble error is similar or slightly lower than the error for the best performing individual network in the committee. In addition, the ensemble encompasses a larger volume of the weight space than any individual member.

Appendix C. Kullback–Leibler divergence

The Kullback–Leibler divergence, or relative entropy, e.g. MacKay (2003), measures the difference between two probability distributions A and B

$$D_{KL}(A||B) = H(A,B) - H(A), \tag{B.6}$$

where $H(A,B)$ is the cross-entropy of A and B and $H(A)$ is the entropy of A. The measure can be interpreted as the amount of information lost when using B to approximate A. For a continuous random variable x, D_{KL} can be expressed by the integral

$$D_{KL}(A||B) = \int_{-\infty}^{+\infty} \log_2\left(\frac{A(x)}{B(x)}\right) A(x)dx, \tag{B.7}$$

which measures D_{KL} in units of bits for a logarithm taken to base 2.

For a single model parameter m, we quantify the information gain upon observing the data by calculating the Kullback–Leibler divergence D_{KL} between the 1-D marginal posterior and prior probability distributions. A similar measure was used by for instance Meier et al. (2007), Käufl et al. (2014). In the above integral (Eq. B.7), $A(x)$ is the prior pdf $p(m)$ and $B(x)$ represents the 1-D marginal posterior pdf $p(m|\mathbf{d})$, with the observed data $\mathbf{d}$. If the posterior pdf equals the prior pdf, $D_{KL} = 0$ and our knowledge on the parameter m remains unchanged after observing the data $\mathbf{d}$. For reference, consider a 1-D Gaussian distribution with mean μ and standard deviation σ; the difference with a second distribution with the same mean and standard deviation $\frac{1}{2}\sigma$, as measured by the information gain, is 1.16 bits.

Appendix D. 1-D marginals for (an) isotropic parametrisations

Fig. D.2 shows the 1-D marginal posterior pdfs for the partly (blue) and fully (red) anisotropic model parametrisation for all parameters addressed in the main text, except for parameters related to the D″ region (Fig. 7) and attenuation. Note that the outer core is isotropic in both parametrisations. The differences between the pdfs are assumed to result from trade-offs between anisotropic parameters, which cannot be constrained by the spheroidal mode data alone.

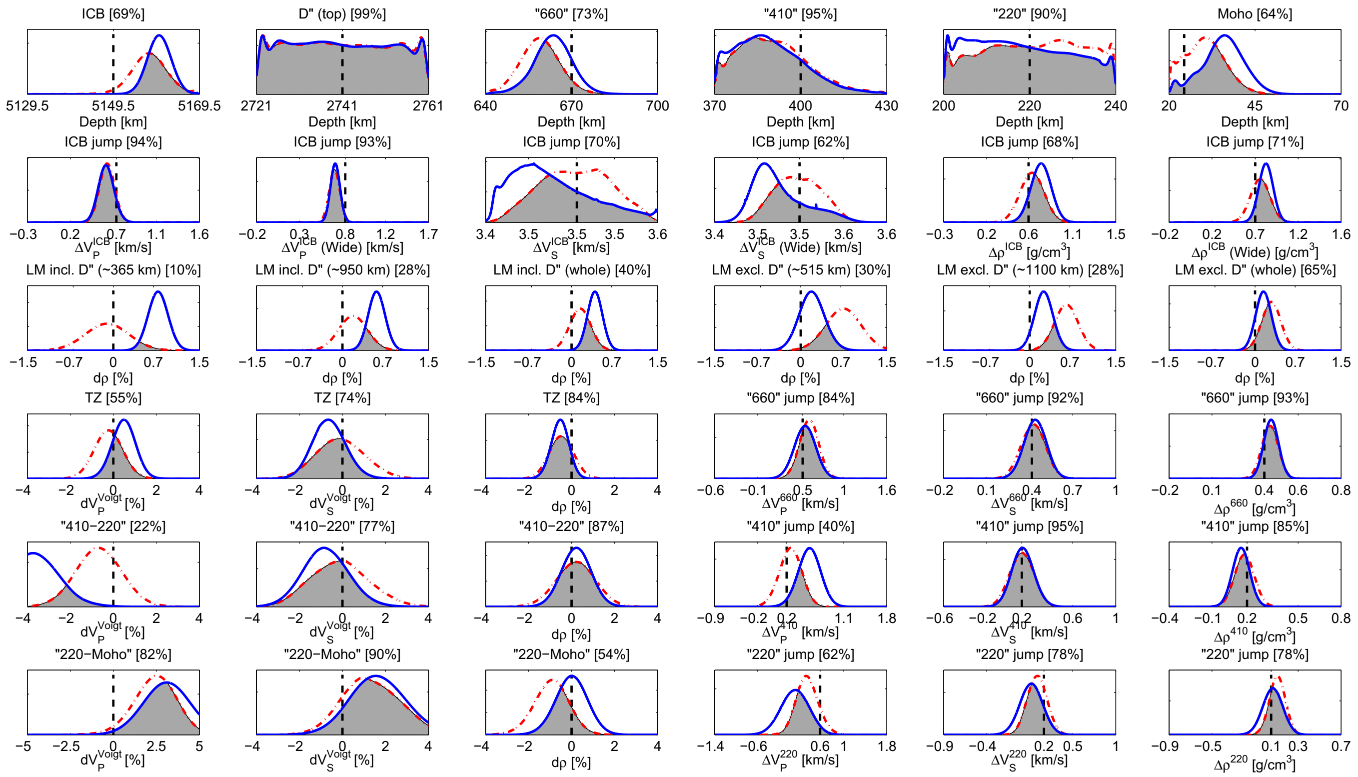

Fig. D.2. 1-D marginal posterior pdfs for the partly (blue, solid) and fully (red, dashed-dotted)) anisotropic model parametrisation for all parameters addressed in the main text, except for parameters related to D″ (Fig. 7). *PREM* is shown as a reference (black, dashed) and the percentage of overlap between the two pdfs is shown above each panel. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Appendix E. Body wave travel times

In addition to free oscillation centre frequencies and quality factors, we used body wave travel times to perform a joint inversion of normal mode and travel time data for radial Earth structure. Similar to de Wit et al. (2013), we used first-arrival travel time data from the EHB bulletin for the years 2001 to 2008, as collected by the International Seismological Centre (ISC) and reprocessed by Engdahl et al. (1998), for the phases *Pn*, *P*, *PP*, *PKPab*, *PKPbc* and *PKIKP* (*PKPdf*). This data set was augmented by measurements from the EHB bulletin (years 2001–2008) for the *Sn* and *S* phases (Table E.5). Synthetic first-arrival travel time curves were computed using the TauP package (Crotwell et al., 1999).

We processed the travel times following the procedure by de Wit et al. (2013). Since the travel time curve for each phase was rather smooth, a large (linear) correlation exists between the travel time at different epicentral distances. Therefore, the travel time curves were sampled at 2° intervals. This reduced the number of free network parameters and thus made network training faster. As in de Wit et al. (2013), we assumed that this downsampling did not result in a significant loss of information on the earth model parameters, given the high correlation between the measurements. The resulting 186-D travel time vector was a concatenation of data for the Pn (8 travel time measurements), P (32), PP (62), PKPab (14), PKPbc (4), PKIKP (29), Sn (9) and S (28) phases.

The measurement errors for the centre frequencies were estimated by Deuss et al. (2013) using a cross-validation approach. This is in contrast to the conservative noise estimates in the travel time data in de Wit et al. (2013), which were based on the scatter in the available measurements in the EHB bulletin, i.e. the maximum difference between the data for each seismic phase and epicentral distance interval. This spread originates from measurement errors, phase misidentifications, uncertainties in the estimated source depth and lateral heterogeneities (3-D structure) in the Earth. To align the two data sets and the associated noise estimates, we defined a new noise model for the travel time data. For a given epicentral distance, the average of the travel time data may be more representative of 1-D Earth structure, as the contribution of (incoherent) 3-D structure to the measurement scatter is averaged out. The uncertainty in this average is given by the sample variance, which we computed for the EHB travel time data for each phase and distance. For most travel time measurements, these new noise estimates are one to two orders of magnitudes smaller than the conservative error levels in de Wit et al. (2013).

References

Aki, K., Christoffersson, A., Husebye, E.S., 1977. Determination of the three-dimensional seismic structure of the lithosphere. J. Geophys. Res. 82, 277–296.
Alboussière, T., Deguen, R., 2012. Asymmetric dynamics of the inner core and impact on the outer core. J. Geodyn. 61, 172–182.
An, M., 2012. A simple method for determining the spatial resolution of a general inverse problem. Geophys. J. Int. 191 (2), 849–864.
Anderson, D.L., Hart, R.S., 1978. Q of the Earth. J. Geophys. Res. 83 (B12), 5869–5882.
Andrews, J., Deuss, A., Woodhouse, J., 2006. Coupled normal-mode sensitivity to inner-core shear velocity and attenuation. Geophys. J. Int. 167 (1), 204–212.
Backus, G.E., Gilbert, F., 1968. The resolving power of gross earth data. Geophys. J. Roy. Astron. Soc. 16, 169–205.
Backus, G.E., Gilbert, F., 1970. Uniqueness in the inversion of inaccurate gross Earth data. Phil. Trans. Roy. Soc. Lond. 266, 123–192.
Beghein, C., Trampert, J., 2003. Robust normal mode constraints on inner-core anisotropy from model space search. Science 299 (5606), 552–555.
Beghein, C., Trampert, J., van Heijst, H.J., 2006. Radial anisotropy in seismic reference models of the mantle. J. Geophys. Res. 111, B02303.
Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Oxford University Press, New York.
Bodin, T., Sambridge, M., 2009. Seismic tomography with the reversible jump algorithm. Geophys. J. Int. 178 (3), 1411–1436.
Bolt, B.A., 1977. The detection of PKIIKP and damping in the inner core. Ann. Geophys. 30.

Boschi, L., 2003. Measures of resolution in global body wave tomography. Geophys. Res. Lett. 30 (3–4).
Cammarano, F., Romanowicz, B., 2008. Radial profiles of seismic attenuation in the upper mantle based on physical models. Geophys. J. Int. 175 (1), 116–134.
Cammarano, F., Deuss, A., Goes, S., Giardini, D., 2005. One-dimensional physical reference models for the upper mantle and transition zone: combining seismic and mineral physics constraints. J. Geophys. Res.: Solid Earth 110 (B1), n/a–n/a.
Cammarano, F., Goes, S., Deuss, A., Giardini, D., 2005. Is a pyrolitic adiabatic mantle compatible with seismic data? Earth Planet. Res. Lett. 232, 227–243.
Cammarano, F., Tackley, P., Boschi, L., 2011. Seismic, petrological and geodynamical constraints on thermal and compositional structure of the upper mantle: global thermo-chemical models. Geophys. J. Int. 187, 1301–1318.
Cao, A., Romanowicz, B., 2004. Constraints on density and shear velocity contrasts at the inner core boundary. Geophys. J. Int. 157 (3), 1146–1151.
Chambat, F., Valette, B., 2001. Mean radius, mass and inertia for reference Earth's models. Phys. Earth Planet. Inter. 124, 237–253.
Chang, S.J., Ferreira, A.M., Ritsema, J., van Heijst, H.J., Woodhouse, J.H., 2014. Global radially anisotropic mantle structure from multiple datasets: a review, current challenges, and outlook. Tectonophysics 617, 1–19.
Cobden, L., Goes, S., Cammarano, F., Connolly, J.A.D., 2008. Thermochemical interpretation of one-dimensional seismic reference models for the upper mantle: evidence for bias due to heterogeneity. Geophys. J. Int. 175 (2), 627–648.
Cobden, L., Goes, S., Ravenna, M., Styles, E., Cammarano, F., Gallagher, K., Connolly, J.A.D., 2009. Thermochemical interpretation of 1-D seismic data for the lower mantle: the significance of nonadiabatic thermal gradients and compositional heterogeneity. J. Geophys. Res.: Solid Earth 114 (B11).
Cormier, V.F., Stroujkova, A., 2005. Waveform search for the innermost inner core. Earth Planet. Sci. Lett. 236, 96–105.
Cornford, D., Nabney, I.T., Bishop, C.M., 1999. Neural network-based wind vector retrieval from satellite scatterometer data. Neural Comput. Appl. 8 (3), 206–217.
Crotwell, H.P., Owens, T.J., Ritsema, J., 1999. The taup toolkit: flexible seismic travel-time and ray-path utilities. Seismol. Res. Lett. 70, 154–160.
Deguen, R., 2012. Structure and dynamics of earth's inner core. Earth Planet. Sci. Lett. 333–334, 211–225.
Deschamps, F., Trampert, J., Tackley, P.J., 2007. Thermo-chemical structure of the lower mantle: seismological evidence and consequences for geodynamics. In: Yuen, D., Maruyama, S., Karato, S.I., Windley, B. (Eds.), Superplumes: Beyond Plate Tectonics. Springer, Netherlands, pp. 293–320.
Deuss, A., Irving, J.C.E., Woodhouse, J.H., 2010. Regional variation of inner core anisotropy from seismic normal mode observations. Science 328 (5981), 1018–1020.
Deuss, A., Andrews, J., Day, E., 2013. Seismic observations of mantle discontinuities and their mineralogical and dynamical interpretation. In: Physics and Chemistry of the Deep Earth. John Wiley & Sons Ltd, pp. 295–323.
Deuss, A., Ritsema, J., van Heijst, H., 2013. A new catalogue of normal-mode splitting function measurements up to 10 mhz. Geophys. J. Int. 193 (2), 920–937.
de Wit, R.W.L., Trampert, J., van der Hilst, R.D., 2012. Toward quantifying uncertainty in travel time tomography using the null-space shuttle. J. Geophys. Res. 117, B03301.
de Wit, R.W.L., Valentine, A.P., Trampert, J., 2013. Bayesian inference of Earth's radial seismic structure from body-wave traveltimes using neural networks. Geophys. J. Int. 195, 408–422.
Durek, J.J., Ekström, G., 1996. A radial model of anelasticity consistent with long-period surface-wave attenuation. Bull. Seismol. Soc. Am. 86, 144–158.
Dziewoński, A.M., Anderson, D.L., 1981. Preliminary reference Earth model. Phys. Earth Planet. Inter. 25, 297–356.
Dziewoński, A.M., Hales, A.L., Lapwood, E.R., 1975. Parametrically simple earth models consistent with geophysical data. Phys. Earth Planet. Inter. 10, 12–48.
Engdahl, E.R., van der Hilst, R.D., Buland, R., 1998. Global teleseismic earthquake relocation with improved travel times and procedures for depth determination. Bull. Seismol. Soc. Am. 88, 722–743.
Gilbert, F.D.A, 1975. An application of normal mode theory to the retrieval of structural parameters and source mechanisms from seismic spectra. Philos. Trans. Roy. Soc. Lond. 278, 187–269.
Gubbins, D., Masters, G., Nimmo, F., 2008. A thermochemical boundary layer at the base of Earth's outer core and independent estimate of core heat flux. Geophys. J. Int. 174 (3), 1007–1018.
Ishii, M., Dziewoński, A.M., 2002. The innermost inner core of the earth: evidence for a change in anisotropic behavior at the radius of about 300 km. Proceedings of the National Academy of Sciences 99 (22), 14026–14030.
Jeffreys, H., Bullen, K.E., 1940. Seismological Tables. British Association for the Advancement of Science, London, UK.
Käufl, P., Valentine, A.P., O'Toole, T.B., Trampert, J., 2014. A framework for fast probabilistic centroid-moment-tensor determination—inversion of regional static displacement measurements. Geophys. J. Int.
Kellogg, L.H., Hager, B.H., van der Hilst, R.D., 1999. Compositional stratification in the deep mantle. Science 283 (5409), 1881–1884.
Kennett, B., 1998. On the density distribution within the Earth. Geophys. J. Int. 132, 374–382.
Kennett, B., 2006. On seismological reference models and the perceived nature of heterogeneity. Phys. Earth Planet. Inter. 159, 129–139.
Kennett, B., Engdahl, E.R., 1991. Travel times for global earthquake location and phase association. Geophys. J. Int. 105, 429–465.
Kennett, B., Engdahl, E.R., Buland, R., 1995. Constraints on seismic velocities in the Earth from travel times. Geophys. J. Int. 122, 108–124.
Koelemeijer, P.J., 2014. Normal Mode Studies of Long Wavelength Structures in Earth's Lowermost Mantle. Ph.D. Thesis. University of Cambridge, Cambridge, UK.
Koelemeijer, P., Deuss, A., Ritsema, J., 2013. Observations of core-mantle boundary Stoneley modes. Geophys. Res. Lett. 40 (11), 2557–2561.
Koper, K., Dombrovskaya, M., 2005. Seismic properties of the inner core boundary from PKiKP/P amplitude ratios. Earth Planet. Res. Lett. 237, 680–694.
Kustowski, B., Ekström, G., Dziewoński, A.M., 2008. Anisotropic shear-wave velocity structure of the Earth's mantle: A global model. J. Geophys. Res.: Solid Earth 113 (B6).
Lekić, V., Matas, J., Panning, M., Romanowicz, B., 2009. Measurement and implications of frequency dependence of attenuation. Earth Planet. Sci. Lett. 282, 285–293.
Li, X., Cormier, V.F., 2002. Frequency dependent attenuation in the inner core: part I. A viscoelastic interpretation. J. Geophys. Res. 107.
Love, A.E.H., 1927. A Treatise on the Mathematical Theory of Elasticity. Cambridge University Press, New York.
Lythgoe, K., Deuss, A., Rudge, J., Neufeld, J., 2014. Earth's inner core: innermost inner core or hemispherical variations? Earth Planet. Sci. Lett. 385, 181–189.
MacKay, D., 2003. Information Theory, Inference, and Learning Algorithms. Cambridge Univ Press.
Masters, G., Gubbins, D., 2003. On the resolution of density within the Earth. Phys. Earth Planet. Inter. 140, 159–167.
Masters, G., Barmine, M., Kientz, S., 2011. Mineos: User Manual Version 1.0.2. Calif. Inst. of Tech., Pasadena, CA.
Meier, U., Curtis, A., Trampert, J., 2007. Fully nonlinear inversion of fundamental mode surface waves for a global crustal model. Geophys. Res. Lett. 34.
Meier, U., Curtis, A., Trampert, J., 2007. Global crustal thickness from neural network inversion of surface wave data. Geophys. J. Int. 169, 706–722.
Mller, M.F., 1993. A scaled conjugate gradient algorithm for fast supervised learning. Neural Networks 6, 525–533.
Montagner, J.P., Kennett, B.L.N., 1996. How to reconcile body-wave and normal-mode reference earth models. Geophys. J. Int. 125, 229–248.
Morelli, A., Dziewonski, A.M., Woodhouse, J.H., 1986. Anisotropy of the inner core inferred from PKIKP travel times. Geophys. Res. Lett. 13 (13), 1545–1548.
Mosegaard, K., Tarantola, A., 1995. Monte Carlo sampling of solutions to inverse problems. J. Geophys. Res. 100, 12431–12447.
Okal, E.A., Jo, B.G., 1990. Q measurements for phase X overtones. Pure Appl. Geophys. 132, 331–362.
Panning, M., Romanowicz, B., 2004. Inference on flow at the base of Earth's mantle based on seismic anisotropy. Science 303, 351–353.
Resovsky, J., Trampert, J., van der Hilst, R.D., 2005. Error bars for the global seismic Q profile. Earth Planet. Res. Lett. 230, 413–423.
Romanowicz, B., Mitchell, B., 2007. Q in the Earth from crust to core. In: Schubert, G. (Ed.), Seismology and the Structure of the Earth, 1. Elsevier, Amsterdam, pp. 731–774.
Sambridge, M., 1999. Geophysical inversion with a neighbourhood algorithm-I. searching a parameter space. Geophys. J. Int. 138, 479–494.
Sambridge, M., Mosegaard, K., 2002. Monte Carlo methods in geophysical inverse problems. Rev. Geophys. 40, 1–29.
Shahraeeni, M.S., Curtis, A., 2011. Fast probabilistic nonlinear petrophysical inversion. Geophysics 76, 45–58.
Shearer, P., Masters, G., 1990. The density and shear velocity contrast at the inner core boundary. Geophys. J. Int. 102 (2), 491–498.
Tarantola, A., 2005. Inverse problem theory and methods for model parameter estimation. SIAM.
Tarantola, A., Valette, B., 1982. Inverse problems = quest for information. J. Geophys. 50, 159–170.
Tkalčić, H., Kennett, B.L.N., Cormier, V.F., 2009. On the inner–outer core density contrast from PKiKP/PcP amplitude ratios and uncertainties caused by seismic noise. Geophys. J. Int. 179 (1), 425–443.
Trampert, J., van der Hilst, R., 2005. Towards a quantitative interpretation of global seismic tomography. In: van der Hilst, R.D., Bass, J.D., Matas, J., Trampert, J. (Eds.), Earth's Deep Mantle: Structure, Composition, and Evolution, 160. AGU, Washington, D.C., USA, pp. 47–62.
Trampert, J., Fichtner, A., Ritsema, J., 2013. Resolution tests revisited: the power of random numbers. Geophys. J. Int. 192 (2), 676–680.
Valentine, A.P., Trampert, J., 2012. Assessing the uncertainties on seismic source parameters: towards realistic error estimates for centroid-moment-tensor determinations. Phys. Earth Planet. Inter. 210–211, 36–49.
Vasco, D.W., Johnson, L.R., Marques, O., 2003. Resolution, uncertainty, and whole Earth tomography. J. Geophys. Res. 108 (B1), 2022.
Waszek, L., Deuss, A., 2013. A low attenuation layer in the Earth's uppermost inner core. Geophys. J. Int.
Waszek, L., Irving, J., Deuss, A., 2011. Reconciling the hemispherical structure of Earth's inner core with its super-rotation. Nat. Geosci. 4, 264–267.
Widmer, R., Masters, G., Gilbert, F., 1991. Spherically symmetric attenuation within the Earth from normal mode data. Geophys. J. Int. 104, 541–553.
Woodhouse, J.H., Giardini, D., Li, X.D., 1986. Evidence for inner core anisotropy from free oscillations. Geophys. Res. Lett. 13 (13), 1549–1552.