

Probabilistic lowermost mantle *P*-wave tomography from hierarchical Hamiltonian Monte Carlo and model parametrization cross-validation

Jack B. Muir^{1,2} and Hrvoje Tkalčić²

¹Seismological Laboratory, California Institute of Technology, 1200 E. California Blvd, Pasadena, CA 91125, USA. E-mail: jmuir@caltech.edu

²Research School of Earth Sciences, The Australian National University, 142 Mills Rd, Acton, ACT 0200, Australia

Accepted 2020 August 19. Received 2020 January 2; in original form 2020 August 10

SUMMARY

Bayesian methods, powered by Markov Chain Monte Carlo estimates of posterior densities, have become a cornerstone of geophysical inverse theory. These methods have special relevance to the deep Earth, where data are sparse and uncertainties are large. We present a strategy for efficiently solving hierarchical Bayesian geophysical inverse problems for fixed parametrizations using Hamiltonian Monte Carlo sampling, and highlight an effective methodology for determining optimal parametrizations from a set of candidates by using efficient approximations to leave-one-out cross-validation for model complexity. To illustrate these methods, we use a case study of differential traveltime tomography of the lowermost mantle, using short period *P*-wave data carefully selected to minimize the contributions of the upper mantle and inner core. The resulting tomographic image of the lowermost mantle has a relatively weak degree 2—instead there is substantial heterogeneity at all low spherical harmonic degrees less than 15. This result further reinforces the dichotomy in the lowermost mantle between relatively simple degree 2 dominated long-period *S*-wave tomographic models, and more complex short-period *P*-wave tomographic models.

Key words: Structure of the Earth; Inverse theory; Tomography; Body waves.

1 INTRODUCTION

The core–mantle boundary (CMB) is the most pronounced internal boundary of the Earth, separating the turbulently convecting, well mixed, iron alloy outer core (Stevenson 1981) from the highly heterogeneous silicate lower mantle (Helffrich & Wood 2001). The lowermost few hundred kilometres of the mantle exhibit particularly pronounced, high amplitude heterogeneity in the spectra of high resolution *S*-wave tomographic images (Dziewonski *et al.* 2010; Meschede & Romanowicz 2015), and is also found in the spectra of geodynamic modeling (Becker & Boschi 2002). In particular, *S*-wave tomographic images give strongly concordant images of large slow velocity provinces (LLSVPs) underneath Africa and the Pacific Ocean (Li & Romanowicz 1996; Ritsema *et al.* 2011; Lekić *et al.* 2012; Garnero *et al.* 2016; McNamara 2018), and faster regions correlated with subduction zones (Garnero & McNamara 2008). Evidence from both *P* and *S* upper and mid-mantle tomography appear to show subducting slabs penetrating through the transition zone into the lower mantle (despite generally poor agreement in the mid-mantle; van der Hilst *et al.* (1997), Ritsema *et al.* (2011), Simmons *et al.* (2012), Fukao & Obayashi (2013), Tesoniero *et al.* (2015), Cottaar & Lekić (2016), Moulík & Ekström (2016) and Shephard *et al.* (2017)), suggesting that these fast zones

are relatively colder remnants of lithospheric material (Tan *et al.* 2002). Additional evidence for recycled material contributing to a strongly heterogeneous lowermost mantle comes from imaged interfaces that constitute the upper surface of the *D*' layer (Lay & Helmberger 1983; van der Hilst *et al.* 2007; Shang *et al.* 2014) that may be formed by the Bridgmanite—Post-Perovskite phase transition within colder regions (Bower *et al.* 2013) and the presence of high frequency scatterers at small length scales further suggesting recycled crustal material (Frost *et al.* 2017). Consequently, the lowermost mantle forms a highly inhomogeneous boundary condition for outer core heat transfer, chemical transfer and mechanical coupling.

Within the outer core, the pattern of overturning convection is thought to be influenced by this boundary condition. In particular, thermal anomalies of the lowermost mantle and radial perturbations of the CMB act on outer core flow at the CMB (an isothermal boundary) by changing the heat flux through the CMB and deflecting the gravitational equipotential from isotherms. Order of magnitude calculations by Bloxham & Gubbins (1987) indicated that relatively small perturbations to heat flux could act to force persistent patterns in outer core flow, explaining static features in the inferred pattern of reconstructed CMB magnetic flux, especially the Siberian and

Canadian flux patches (Jackson *et al.* 2000). Further quantitative development (e.g. Gibbons & Gubbins (2000); Gubbins *et al.* (2007)) has indicated that candidate geodynamo solutions forced by inferred heat fluxes computed by *S*-wave tomography produce realistic inferred CMB magnetic flux maps. Analyses of this type have further been extended to include the effect of the post-Perovskite phase transition and heterogeneous chemical compositions (Nakagawa & Tackley 2008; Amit & Choblet 2009, 2012). Amit *et al.* (2015) have used tomographic images that have been further inverted for thermal and chemical heterogeneity using mineral physics data (Mosca *et al.* 2012) to show that decoupling thermal perturbations from seismic observations, rather than using a simple linear relationship between them, may have a large effect on the inferred CMB heat flux, further emphasizing the strong control that CMB and lowermost mantle properties have on the perturbation of the geodynamo away from a dipole. The inner core is in turn thermally coupled to the CMB through the rapidly convecting outer core, which may result in different patterns of crystallization (Aubert *et al.* 2008) or even patches of melting on the inner core boundary (Gubbins *et al.* 2011). Observational studies have long shown very large scale hemispherical heterogeneity in inner core seismic behaviour (e.g. Tanaka & Hamaguchi 1997). More recent studies of the inner core show that the inner core is laterally heterogeneous with substantial complexity at more regional scales, both in anisotropic seismic velocity (Attanayake *et al.* 2014; Yee *et al.* 2014; Irving & Deuss 2015; Burdick *et al.* 2019), and in seismic attenuation (Pejić *et al.* 2017, 2019), with the general trend being towards an increasing recognition of the complex structure of upper layers of the inner core, which is heavily controlled by the CMB heat flux boundary condition.

Towards providing better constraints on the effect of lowermost mantle structures on the dynamics of both the outer core and the mantle, our study performs an inversion of lowermost mantle V_p perturbations using handpicked body wave differential traveltimes. Past body wave inversions have typically relied on absolute traveltimes for *P* phases. Because of the potentially significant accumulated perturbations to absolute traveltimes, large residuals in absolute data are not uncommon. Use of differential traveltimes provides better specificity to the area of interest, and typically smaller starting residuals, at the expense of reduced spatial coverage (e.g. Cormier & Choy (1986); Tkalčić *et al.* (2002); Tanaka (2010); Young *et al.* (2013)). We use a combination of two data sets (PcP–P, PKPab–PKPbc) chosen to allow relatively high sensitivity to the lowermost mantle whilst reducing the effect of the crust, upper mantle and core (under the assumption that the seismic velocity of the upper core is laterally homogeneous at long length scales). Because the relative quality of the data sets, as well as the appropriate scales of the velocity and radius perturbations, are *a priori* unknown, we use a hierarchical Bayesian formulation (Malinverno & Briggs 2004; Bodin *et al.* 2012) to quantify these scales within our inverse framework. We use the Markov Chain Monte Carlo (MCMC) method to sample the posterior distribution of our inverse problem. The hierarchical structure of our inverse problem makes the posterior non-linear; consequently we use a highly efficient Hamiltonian Monte Carlo (HMC) sampler, as well as a non-centred hierarchical parametrization, to allow for full exploration of the posterior. We furthermore investigate the meta-problem of determining optimal model complexity from a data-driven perspective by utilizing an efficient approximate leave-one-out cross-validation (LOO-CV) approach (Vehtari *et al.* 2017) for maximum degree of spherical harmonic expansion utilized in the model.

2 DATA

The data for this study consist of two sets of differential *P*-wave traveltime measurements with strong sensitivity to the lowermost mantle. The phase pairs used were PcP–P and PKPab–PKPbc. Initial residuals were calculated with respect to the ak135 1-D velocity model (Kennett *et al.* 1995). Following Young *et al.* (2013), residuals from all data sets were then corrected for mantle structure above the study region using the *P*-wave velocity model of Della Mora *et al.* (2011), and for the Earth's ellipticity (Dziewonski & Gilbert 1976). For the remainder of this paper, the resulting corrected residual differential traveltimes will be denoted as Δt_{A-B} where *A* and *B* are the two phases used in the measurement. A global view of the phases used in this study are shown in Fig. 1.

2.1 PcP–P

The PcP–P data set was taken from Tkalčić *et al.* (2002) and Young *et al.* (2013). It consists of 680 differential traveltime measurements, with an epicentral distance of 55° to 70°. PcP–P data were constructed via hand-picked waveform matching of PcP and *P* phases to maximize the accuracy of measured differential times. Data were picked using either the raw waveform, or waveforms bandpass filtered at 1–3 Hz. PcP–P measurements for distances less than 55° were not used as the increased separation between the ray paths in the crust and bulk mantle reduces the relative sensitivity to the lower mantle, and because restriction to wider epicentral distances reduces the potential effect of CMB topography on the PcP reflection. Even with this distance restriction in place, the PcP–P data set is the most affected by heterogeneity in the upper mantle. The PcP–P data set provides the best global coverage of the lowermost mantle in our data set, and does not suffer from a trade-off between the contributions of velocity and CMB topography perturbations on the source versus receiver side of the mantle as is the case in the PKPab–PKPbc data set.

2.2 PKPab–PKPbc

The PKPab–PKPbc data set was created by combining the PKPab–PKPdf and PKPbc–PKPdf data sets of Tkalčić *et al.* (2002), Leykam *et al.* (2010), and Young *et al.* (2013). These data sets consist of meticulously hand picked phases. Both data sets pick the unfiltered waveforms; in the case of PKPab–PKPdf, a Hilbert transform is performed to the PKPab phase to correct the waveform phase shift associated with this path. In order to generate a PKPab–PKPbc data set, pairs of PKPab–PKPdf and PKPbc–PKPdf measurements were identified that corresponded to the same source/receiver pair. The PKPab–PKPbc differential traveltime was then calculated by subtracting the PKPbc–PKPdf from the PKPab–PKPdf, to cancel the PKPdf component. The PKPbc phase only exists at an epicentral distance of 145° to 155°, and PKPab and PKPbc phases are not always both identifiable on for a particular event even within this epicentral distance range. Hence only a subset of the original data sets could be used to generate PKPab–PKPbc differential traveltime measurements; specifically, from 1871 PKPab–PKPdf and 1292 PKPbc–PKPdf measurements, 385 PKPab–PKPbc measurements could be made. This represents a loss of measurements of 88 per cent. Nevertheless, in our opinion it is important to consider PKPab–PKPbc measurements only, as the PKPdf phase samples the inner core, which is expected to affect traveltimes both through possible isotropic and anisotropic velocity heterogeneity.

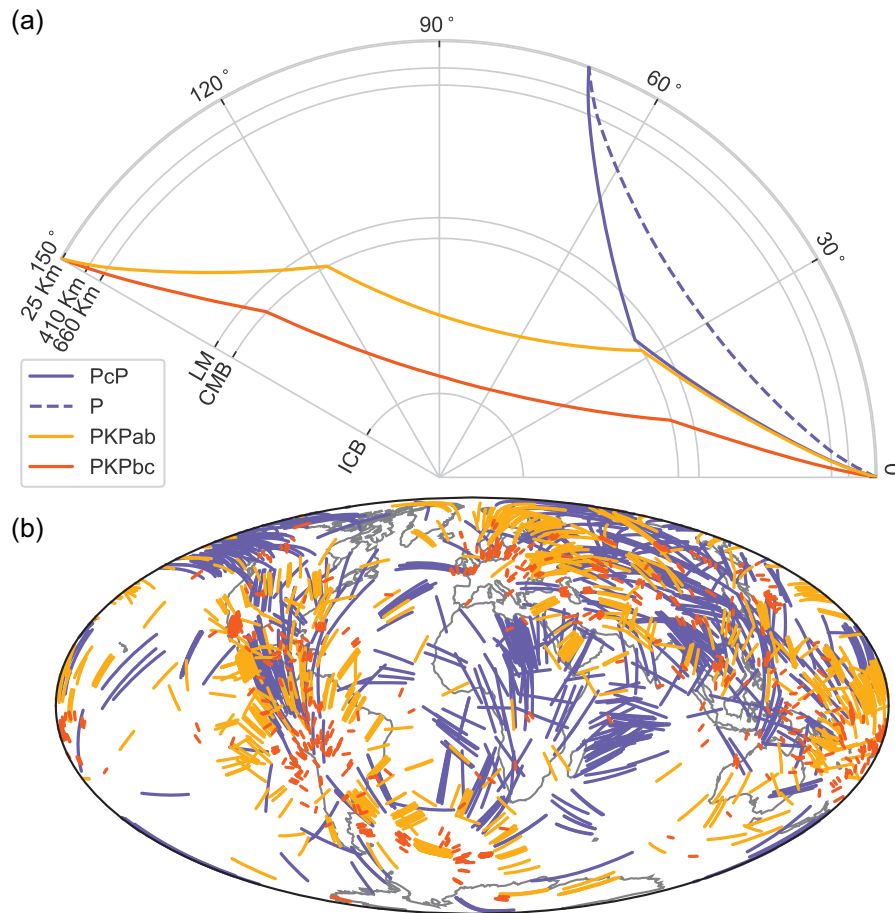


Figure 1. (a) Seismic phases used in this study. Depth LM represents the top of the lowermost mantle inverted in this study. (b) Ray paths within the lowermost mantle used in this study.

3 THE INVERSE PROBLEM DEFINITION

In this work we use the Bayesian, or probabilistic, formulation of the inverse problem (Tarantola 2005). In this formulation the solution is defined by a posterior probability density function

$$p(q, \sigma | d) = p(d | q, \sigma) p(q | \sigma) p(\sigma), \quad (1)$$

where d is the measured data, q are some model parameters of interest and σ are hyperparameters that are not necessarily of direct interest (such as the unknown level of data noise) but which nevertheless impact the structure of the posterior. To better understand high-dimensional posterior distributions, we are normally interested in computing expectation values in respect to the posterior, such as the mean over domain Ω

$$\mathbb{E}[q] = \int_{\Omega} q p(q, \sigma | d) d\Omega. \quad (2)$$

Other typical expectation values studied are the variance or standard deviation, histograms and quantiles of interesting parameters, etc. Computation of expectation values becomes very difficult in high dimensions when analytical results are not available, as is typically the case in geophysical applications. Therefore, we use MCMC methods to sample from the posterior to estimate these expectation values. In particular, in this work we will use the HMC method that is particularly efficient in high dimensions (Neal 2011). Note that we have chosen to not use bold-font ‘vector’ notation in describing our Bayesian model, to emphasize that the final models are the product of many independent scalar distributions on the data and

model parameters, due to the assumption of diagonal covariance matrices for these quantities.

3.1 Forward model

We modeled perturbations to the slowness relative to the AK135 (Kennett *et al.* 1995) reference model using a spherical harmonic expansion up to maximum degree l . While there has been recent success in using adaptive transdimensional Bayesian parametrizations for global tomography from the upper mantle to the inner core (Young *et al.* 2013; Burdick & Lekić 2017; Burdick *et al.* 2019; Pejić *et al.* 2019), the geodynamic community typically engages with tomographic results expressed in spherical harmonics, so we are motivated to develop robust tomographic methods that operate by directly using that parametrization. We calculated linear kernels relating these perturbations to the differential traveltimes using rays drawn using the AK135 reference model by the Obspy `taup` module (Beyreuther *et al.* 2010). Similarly to Young *et al.* (2013), we perturbed slowness in the lowermost 300 km of the mantle only, and assumed that slowness varied only laterally. Motivated by the statistical analysis of Garcia *et al.* (2009), which found the highly heterogeneous layer of the lowermost mantle to be approximately 350 km thick, Young *et al.* (2013) performed inversions using a variety of different layer thicknesses, settling on 300 km as the optimal parametrization, which we have maintained for consistency.

This parametrization results in a linear forward model

$$\begin{bmatrix} \Delta t_{PcP-P} \\ \Delta t_{PKPab-PKPbc} \end{bmatrix} = \begin{bmatrix} G_{PcP-P}^{\delta\tau} \\ G_{PKPab-PKPbc}^{\delta\tau} \end{bmatrix} q_{\delta\tau}, \quad (3)$$

where the $G^{\delta\tau}$ and $q_{\delta\tau}$ are the kernels and model parameter vectors for slowness perturbations, respectively, and Δt_{PcP-P} and $\Delta t_{PKPab-PKPbc}$ are the residual differential traveltime data used for this study. This linearized forwards model is simple but appropriate for the low volume of data and resolution target used in this study. To better scale the problem, we use units of $s(10^6\text{m})^{-1}$ for slowness. Using a linear forward model is essentially free in comparison to the MCMC sampling process as the G s are precomputed. However, this linear formulation does introduce potential theory errors, of which the most significant are the lack of ray tracing updates through a potentially highly heterogeneous model, inconsistencies between the background mantle model used for corrections and the model used for inversion, lack of correction for CMB topography and assumption of a single layer for the lowermost mantle. The use of differential data that contains both reflected and transmitted phases is designed to mitigate the theory errors caused of background model and CMB topography effects. Furthermore, our inversion utilizes relatively coarse spherical harmonics, which ensure that at the finest resolutions tested the effective ‘voxel’ aspect ratio at the CMB induced by our assumption of a 300-km-thick layer is of order one, so that we are not inverting for an unrealistic vertically elongated parametrization. Current software developments are increasingly allowing for more complex, large scale non-linear Monte Carlo inversions (Gebraad *et al.* 2020), however robust, well-tested implementations are currently limited to linear models such as the one considered here for performance reasons.

3.2 Data likelihood model

If the residual differential traveltime data after inversion are expected to be identically independently distributed (i.i.d.) as a Gaussian around zero, it is appropriate to use a Gaussian noise error model for the likelihood of the data; for example, indexing the N_{PcP-P} differential PcP–P ray paths by i , we have

$$p(\Delta t_{PcP-P} | q_{\delta\tau}, \sigma_{PcP-P}) = \prod_{i=1}^{N_{PcP-P}} \frac{1}{\sqrt{2\pi\sigma_{PcP-P}^2}} \times \exp\left[-\frac{(\Delta t_{PcP-P} - G_{PcP-P}^{\delta\tau} q_{\delta\tau})_i^2}{2\sigma_{PcP-P}^2}\right] \quad (4)$$

and similarly for PKPab–PKPbc. In this case, assuming that the data variance parameters σ_{PcP-P} , and $\sigma_{PKPab-PKPbc}$ are known, the result for the posterior distribution of $q_{\delta\tau}$ is analytic and given by a suitably multivariate Gaussian (Tarantola 2005), from which the mean and covariance of the posterior are immediately apparent. When the data standard deviations σ_{PcP-P} and $\sigma_{PKPab-PKPbc}$ are not known *a priori*, we may treat them as hyperparameters, including them in the inversion. Formulating the model in this way typically results in a posterior for which MCMC sampling methods are required for the computation of expectations (such as the mean and covariance of the posterior). When using hierarchical parametrizations, we are therefore motivated to use a highly efficient sampler, such as HMC, to draw samples from the posterior and from them to compute the expectations of interest.

3.3 Prior distributions for model parameters

The choice of prior is an essential component of any Bayesian approach to geophysical inverse problems. In our case, we have a clear motivation for the prior for model variables $q_{\delta\tau}$. As a result of many decades of extensive study, we have high quality 1-D models of the Earth (e.g. Dziewonski & Anderson 1981; Kennett *et al.* 1995). We have used the ak135 reference model for this study; consequently we are strongly motivated to assume perturbations away from these reference models are small. The appropriate prior for this assumption is a normal distribution, centred on zero, for $q_{\delta\tau}$. As we do not know the appropriate scales of these distributions *a priori*, we introduce another hyperparameter $\sigma_{\delta\tau}$, for a total of three hyperparameters. The choice of prior for hyperparameters is difficult. A common approach has been to use the non-informative, or Jeffries, prior, an improper prior that takes the form $1/\sigma$ for scale parameters. The intuition behind this prior is that all potential magnitudes of scales are equally likely, or equivalently that the distribution of scales is log-uniform. The Jeffries prior has extremely long tails, and blows up near $\sigma = 0$. Current best practice in the statistical community suggests non-informative priors, such as the Jeffries prior, are in fact poor choices for hyperparameters as they tend to bias hyperparameters to extreme values, and that more restrictive priors such as the half-Cauchy or even half-Normal are better choices (Gelman 2006). We use a half-Normal prior for our scale parameters as we have a strong intuition as to a maximum unacceptable scale for velocity perturbations. Additionally, the lightly tailed half-Normal is significantly more efficient to sample from than the heavier tailed half-Cauchy. The choice of scale of these hyperpriors is set to an arbitrary, large value, such that all likely scales are contained within the area of high prior density to avoid unduly biasing the posterior. Concretely, the prior is therefore set to be

$$p(q_{\delta\tau} | \sigma_{\delta\tau}) = \prod_{l=0}^{l'} \prod_{m=-l}^l \frac{1}{\sqrt{2\pi\sigma_{\delta\tau}^2}} \exp\left[-\frac{q_{\delta\tau_{lm}}^2}{2\sigma_{\delta\tau}^2}\right] \quad (5)$$

$$p(\sigma_{\delta\tau}) = \frac{\sqrt{2}}{\sqrt{\pi^2(v_1)^2}} \exp\left[-\frac{\sigma_{\delta\tau}^2}{2(v_1)^2}\right], \quad \sigma_{\delta\tau} > 0 \quad (6)$$

$$p(\sigma_{PcP-P}) = \frac{\sqrt{2}}{\sqrt{\pi^2(v_2)^2}} \exp\left[-\frac{\sigma_{PcP-P}^2}{2(v_2)^2}\right], \quad \sigma_{PcP-P} > 0 \quad (7)$$

$$p(\sigma_{PKPab-PKPbc}) = \frac{\sqrt{2}}{\sqrt{\pi^2(v_2)^2}} \exp\left[-\frac{\sigma_{PKPab-PKPbc}^2}{2(v_2)^2}\right], \quad \sigma_{PKPab-PKPbc} > 0 \quad (8)$$

with $v_1 = 5$ and $v_2 = 1$ setting the scales of the hyperpriors.

3.4 Centred versus non-centred hierarchical parametrizations

A well known pathology in MCMC sampling in which scale hyperparameters σ are used is the ‘funnel’ effect (Neal 2003), in which the curvature of the posterior increases steeply as the σ become small. The funnel pathology has the tendency to drastically increase rejection rates for all common MCMC samplers in the vicinity of low σ areas of the posterior distribution. In turn, lack of samples with low σ has the effect of biasing any estimators of expectation values computed for MCMC chains of finite length, even if asymptotic convergence to the true posterior is theoretically guaranteed (Betancourt & Girolami 2015). Advanced posterior-curvature aware samplers such as Riemannian HMC methods may ameliorate

these issues (Girolami & Calderhead 2011), however these methods are computationally expensive and robust implementations are not readily available. In certain cases, reparametrization of the model can also remove the issue. The formulation of the prior in eq. (5) supposes that the $q_{\delta\tau}$ are *a priori* normally distributed with standard deviation σ_{τ} . It is apparent from this formulation that the funnel effect occurs when these σ become small, so that the range of q acceptable under the prior is limited and difficult to search using MCMC. This form of parametrization is typically known as the centred form. Alternatively, we may use the mathematically equivalent non-centred form

$$q_{\delta\tau} = \bar{q}_{\delta\tau} + \sigma_{\delta\tau} q'_{\delta\tau}, \quad (9)$$

where the variable $\bar{q}_{\delta\tau}$ represent the mean bias for $q_{\delta\tau}$, respectively and the auxiliary variables $q'_{\delta\tau}$ have prior standard (zero mean, unit standard deviation) normal distributions. As previously indicated, we believe that our 1-D reference model is good enough that we expect the mean bias terms to be zero, so we discount them. Global static corrections may still be included by the Y_{00} component of our forward model. Whilst the centred and non-centred forms give rise to the same posterior, the funnel effect is transferred from the prior (in the centred form) to the likelihood (in the non-centred form). If the data noise is relatively large, as is typically the case in global geophysical applications, the funnel is smeared out in the likelihood and correlations between the non-centred parameters are reduced, allowing MCMC to efficiently sample the full posterior in non-centred coordinates (Betancourt & Girolami 2015).

3.5 Selection of model complexity

The family of models proposed in this study are characterized by their maximum degree of spherical harmonic expansion l . Higher l allows the model to be more expressive, at the expense of potentially overfitting the data. This trade-off induces a metaproblem of model selection into the inverse scheme (Claeskens 2016). The model selection problem is ubiquitous in geophysical inverse theory, and consequently many methods have been developed to choose the ‘optimal’ model when there is not a good geophysical justification for preferring a particular one. Broadly speaking, these can be categorized into two classes; first adaptive methods, such as split-block models (e.g. Tkalčić *et al.* 2002; Auer *et al.* 2014), transdimensional Bayesian sampling (e.g. Bodin & Sambridge 2009; Dettmer *et al.* 2011; Sambridge *et al.* 2013) and random or directed model subspace probing methods (e.g. Fang *et al.* 2020), in which the model parametrization itself changes before or during the inversion to reflect data distributions. The second class consists of *post hoc* selection methods in which multiple models are fitted, and then compared using some scoring function to determine the optimal one. These methods include information criteria applied to the *maximum a posteriori* point such as the Bayesian and Akaike information criterion (Pachhai *et al.* 2014; Muir & Tkalčić 2015), information criterion applied to MCMC chains such as the deviance information criterion (Kowsari *et al.* 2019), and cross-validation methods. This class also conceptually includes the class of highly underdetermined problems for which selection of damping and smoothing parameters via the l-curve, cross-validation or discrepancy principle methods (Aster *et al.* 2018) or correction of the model towards a preferred form via the null-space shuttle (Deal & Nolet 1996; Fichtner & Zunino 2019) acts as a form of effective *post hoc* selection of model complexity, whilst the explicit number of parameters does not change. Whilst

these *post hoc* methods typically require inversion of multiple models, robustly fitting adaptive parametrizations may be extremely challenging and costly. Additionally, global basis functions such as spherical harmonics do not naturally fit into adaptive schemes. Consequently in this study we have used a particular *post hoc* selection method to determine the optimal degree of expansion l . In particular, we use LOO-CV to evaluate models with different l . We chose LOO-CV as its estimates are relatively robust and produce unbiased estimates of the predictive power of the model, compared to information criteria methods when the number of data is low (Vehtari & Ojanen 2012). In classical LOO-CV, each individual data point is removed from the data set, the inverse problem is solved and then the posterior predictive distribution for the left out data point is computed (Vehtari *et al.* 2017). The final LOO-CV score is the sum of the estimated loss relative to the predictive distribution for every data point; that is for data y

$$\text{LOO-CV}(y) = \sum_i \log p(y_i | y_{j \neq i}) \quad (10)$$

$$p(y_i | y_{j \neq i}) = \int p(y_i | q) p(q | y_{j \neq i}) dq. \quad (11)$$

In practice, explicit computation of the posterior predictive $p(y_i | y_{j \neq i})$ for every data point is prohibitively expensive, so we use Pareto-Smoothed Importance Sampling (PSIS) using MCMC results for *all* data to estimate the LOO-CV scores (Vehtari *et al.* 2017). For particularly influential data points, the PSIS-LOO estimate may be poorly constrained, and for those data points we perform explicit LOO computations as suggested by Vehtari *et al.* (2017).

3.6 Hamiltonian Monte Carlo

The essential idea of HMC is to exploit currently existing high performance algorithms of classical mechanics to improve the *per sample* efficiency of Monte Carlo over traditional Metropolis–Hastings or Gibbs methods. HMC variants exploit geometry of the posterior likelihood to propose MCMC samples that have a higher acceptance ratio and lower integrated trace autocorrelation length than traditional MCMC samplers such as random-walk Metropolis–Hastings (Neal 2011; Betancourt *et al.* 2017). Through the proliferation of domain-specific languages such as STAN (Carpenter *et al.* 2017) that define and assemble statistical models, compute log-posterior gradients and tune sampling parameters automatically, HMC is becoming increasingly popular as a general-purpose MCMC sampler. In geophysical contexts, HMC has been applied to uncertainty quantification in source inversion problems (Fichtner & Simutè 2018) and 1D acoustic impedance inversion (Sen & Biswas 2017), data exploration (Muir & Tsai 2020), and some tomography problems (Biswas & Sen 2017; Fichtner *et al.* 2019), but overall utilization of the method is limited. An essential requirement for HMC is that the log-posterior be differentiable; this limits its applicability to popular geophysical MCMC sampling techniques such as reverse-jump MCMC in which the number of parameters changes (Bodin & Sambridge 2009) to the parts of the model that are differentiable, with non-differentiable components handled using the normal random-walk Metropolis–Hastings step method (Sen & Biswas 2017). In a linearized tomography problem such as the one discussed here, the data likelihood and priors are almost certainly differentiable, making the problem very well suited to the HMC method.

The Hamiltonian of a system with coordinates q and momenta p is

$$H(q, p) = U(q) + K(p), \quad (12)$$

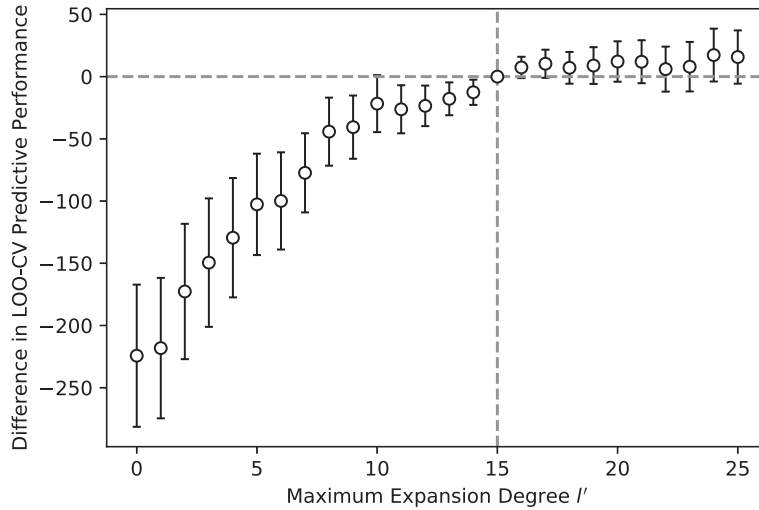


Figure 2. Estimated difference in Leave-one-out cross-validation (LOO-CV) predictive performance relative to a maximum expansion degree $l' = 15$ —higher is better. Error bars are shown at the 2σ level.

where U is the potential energy of the system and K is the kinetic energy. The equations of motion for the system are

$$\frac{\partial q}{\partial t} = \frac{\partial H}{\partial p} \quad (13)$$

$$\frac{\partial p}{\partial t} = -\frac{\partial H}{\partial q}. \quad (14)$$

Setting

$$U(q) = -\log[p(q|d)] = -\log[p(d|q)p(q)] \quad (15)$$

results in a potential that reflects the desired posterior distribution. The kinetic energy is set using arbitrary symmetric positive-definite mass matrix M as

$$K(p) = \frac{1}{2}p^T M^{-1}p. \quad (16)$$

In the context of the geophysical inverse problem, the conjugate momenta p have no physical significance, they merely serve to facilitate the HMC algorithm; similarly the choice of mass matrix affects the efficiency of sampling but not the asymptotic behaviour of the Markov chain, that is guaranteed to converge to the posterior (Neal 2011). The HMC step algorithm begins with the current vector of parameter coordinates q . A momentum vector p is drawn randomly with a multinomial distribution with identity covariance matrix and zero mean; $p \sim N(0, I)$. The equations of motion are then integrated with starting conditions (q, p) for some time period to produce new coordinates (q', p') . The momentum reversed coordinates $(q', -p')$ are then accepted as the next element of the MCMC chain with probability

$$\min[1, \exp(H(q, p) - H(q', -p'))]. \quad (17)$$

The results of Bayesian geophysical inverse problems are almost always given as various expectation values (e.g. mean, variance, marginal probability distributions), which are integrals over the full posterior. For a particular model, the computation of these integrals is dominated by the *typical set*, which counterintuitively is often far from the maximum *a posteriori* (MAP) point. The volume of the typical set decreases relative to that of all model space drastically as the dimensionality of the problem increases, so samplers that do not account for its structure tend to become increasingly inefficient as a function of dimension—this is the problem of *concentration*

of measure. HMC guides the sampling to draw efficiently from the typical set, allowing it to scale well with dimension (Betancourt 2017). The HMC algorithm is conceptually intuitive, and a naive implementation of the algorithm is easily produced if gradients of the log posterior are available. In practice, however, tuning the step size and mass matrix for efficient sampling of the problem is difficult; as such we use the well tested STAN sampler, which automatically handles tuning, to perform HMC sampling for this study.

4 RESULTS

We determined the optimal maximum degree of spherical harmonic expansion l' from the range $0 \leq l' \leq 25$ using the PSIS-LOO method applied to the posterior predictive derived from MCMC sampling. We explicitly computed the posterior predictive distribution using HMC for the data points for which the PSIS estimates failed. In both these cases, 1000 burn in samples were used to initialize the chain, followed by 1000 saved samples. Six chains were run in parallel and combined to produce 6000 final samples, as well as to check that the chains were well mixed and that the HMC convergence diagnostics provided by STAN were satisfied. As with any model selection method with noisy data, comparisons between the model scores are inherently noisy. Applying Occam's razor to model selection suggests that we should choose the simplest model for which adequately describes the data—that is, the improvement from more complex models is not statistically significant. Using a 2σ level of significance, $l' = 15$ is the smallest l' for which no other models performed significantly better, as shown in Fig. 2.

Once a maximum degree of $l' = 15$ was selected, a final HMC run with 6 chains of 2500 burn in samples and 2500 saved samples each was performed, for a final total of 15 000 samples. Residuals from the inversion with $l' = 15$, shown in Fig. 3 show a significant improvement compared to the data, with a variance reduction of 28.5 per cent for PcP-P and 29.4 per cent for PKPab-PKpbc. Fig. 4 shows the spatial distribution of mean residual differential traveltimes, and indicate that there is no significant spatial coherence in the remaining residuals, as projected into the ray bottoming points. Posterior distributions of the scale parameters are shown in Fig. 5, with resulting posterior means of $0.12 \text{ s}(10^6\text{m})^{-1}$ for

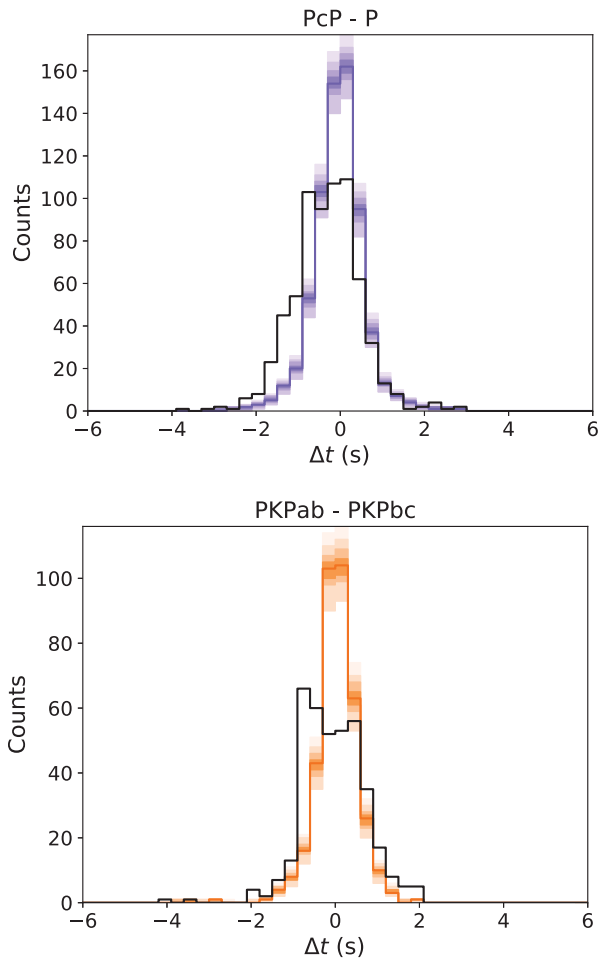


Figure 3. Traveltime residuals for the two data sets used in this study. Black histograms show the initial distribution of residuals, and purple/orange histograms show the quantiles of remaining residual distributions after inversion with maximum $l = 15$.

$\sigma_{d\tau}$, 0.61 s for σ_{PcP-P} and 0.54 s for $\sigma_{PKPab-PKPbc}$. Fig. 6(a) shows a mean model that is comparable to similar short-period V_P inversions of the lowermost mantle, in particular Houser *et al.* (2008) and Young *et al.* (2013). Many features are also similarly placed to the geodynamically coupled joint CMB topography/lowermost mantle P -wave inversion of Soldati *et al.* (2012). Similarity to Young *et al.* is unsurprising as the data set is a subset of Young *et al.*'s, reprocessed to avoid contamination by the inner core, although the model parametrization is very different. We find an RMS velocity perturbation of 0.71 per cent, significantly smaller than Young *et al.*'s value of 0.87 per cent, which may be a byproduct of our global basis functions and assumed priors inducing greater regularization of the inverse problem—noting that a spherical harmonic expansion of $l = 15$ induces a minimum half-wavelength resolution scale of approximately 12° . The most consistent feature across short-period V_P inversions of this depth is the large positive perturbation underneath Asia, coupled with a decrease in velocity mainly underneath the Pacific and Africa in regions roughly concordant with the large low shear wave velocity provinces found in longer period V_S inversions (Houser *et al.* 2008; Soldati *et al.* 2012, 2014; Obayashi *et al.* 2013; Young *et al.* 2013; Tesoniero *et al.* 2015; Moulik & Ekström 2016; Hosseini *et al.* 2020). The lateral extent of P -wave velocity reductions in the Pacific and Africa is relatively inconsistent when compared to LLSVPs revealed by comparable S -wave studies. In

particular, a significantly more complicated pattern of perturbations seems to be required by inversions utilizing short period data, such as that used in this study (e.g. Tkalčić *et al.* (2015)).

Fig. 6(b) shows that due to lack of ray path coverage, the model main in the Southeastern Pacific, Southern Indian and Southern oceans are generally unreliable, but that there is low uncertainty underneath most continental regions (except Antarctica) as well as in the West/Northwestern Pacific and Atlantic oceans. The marginal distributions for spectral power in each degree show that the power is spread relatively evenly between degrees 1–15 with degree 2 not a dominant contributor to the inversion. The strong degree 0 component can be explained as a byproduct of assuming a constant reference velocity throughout the lowermost 300 km of the mantle that is different to the true ray path averaged velocity as encountered by the rays in our data set.

5 RESOLUTION TESTS

In order to confirm that our model selection criterion was performing adequately, we posed a synthetic resolution test using the same phases and source/receiver locations as in our real data set. We constructed a model by initially constructing a binary velocity perturbation image (shown in Fig. S4), and then filtering it at a maximum harmonic degree $l = 10$. The purpose of this model is to explore the algorithm's ability to successfully recover structures at various spatial scales, starting from a simple degree 2 and including higher harmonic degrees superimposed onto it. In order to avoid an 'inversion crime', we used separate ray tracing and perturbation integration codes for creating the synthetic data and for generating the data kernel matrix G . After constructing synthetic differential traveltime data, we imposed Gaussian noise on the differentials with a standard deviation of 0.5 s, which represents a 'realistic' noise contribution for depth-range restricted global tomography equal to 65 per cent of the true differentials and which is similar to the remaining residuals for our real data inversion, as seen in Fig. 5.

Once the synthetic data were created, we performed model selection using the same methodology as in our real data results, with the LOO-CV curve shown in Fig. 7, which shows that a maximum harmonic degree of $l = 9$ is preferred by the data, with the true $l = 10$ within 2σ standard error in the difference. Given the large proportion of noise in the data, it is to be expected that model selection may favour a simpler model than the true model, especially where large parts of the model domain are poorly sampled by rays, as is the case with our data set. The resultant tomographic test at $l = 9$ is shown in Fig. 8. These results indicate that the PSIS-LOO method can obtain a good approximation of the true structure for our problem under realistic amplitudes of noise that are a substantial proportion of the overall initially measured residuals.

As a further test of the resolution of the two data sets we used in this study, we performed a synthetic inversion utilizing the DETOX-P3 synthetic model (Hosseini *et al.* 2020), evaluated at a depth of 2741 km using the SubMachine webtool (Hosseini *et al.* 2018). The DETOX-P2/3 models represent the current state-of-the-art in global P -wave tomography, with the full DETOX-P3 model utilizing ~ 10.7 million traveltimes in multiple frequency bands, notably including the lesser used Pdiff and PP phases. The inclusion of large amounts of Pdiff data, in particular, allows for heightened sensitivity to the lowermost mantle, although we note that the diffraction process inherently smooths Pdiff wavefields, effectively limiting the temporal resolution of Pdiff picks deep into the shadow zone (Hosseini & Sigloch 2015). For our resolution test, we again assumed

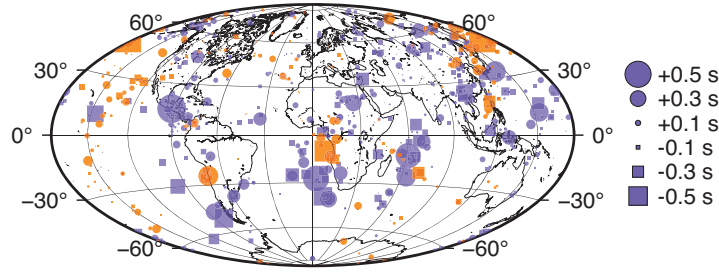


Figure 4. Posterior mean residual differential traveltimes for PcP–P/PKpab–PKPbc in purple/orange, respectively. The posterior mean residual is $\text{mean}(\Delta t_{\text{obs}} - \Delta t_{\text{pred}})$. Symbols are plotted at the bottoming points of the PcP and PKPbc rays.

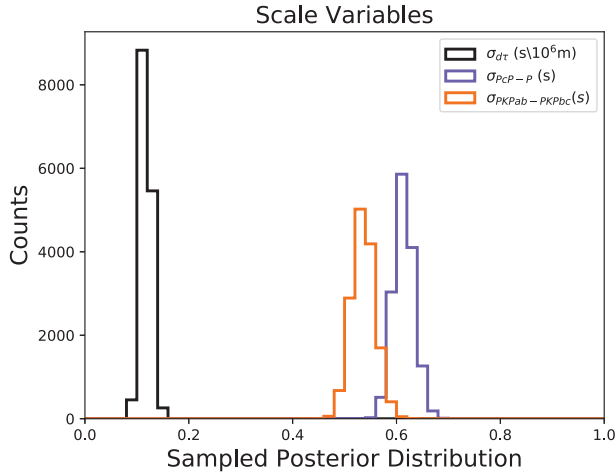


Figure 5. Posterior distributions of scale variables $\sigma_{d\tau}$, σ_{PcP-P} and $\sigma_{PKpab-PKPbc}$ for maximum $l' = 15$. Prior distributions were half-normal for all three scale variables, with prior standard deviation $\nu_1 = 5$ for $\sigma_{d\tau}$ and $\nu_2 = 2$ for σ_{PcP-P} and $\sigma_{PKpab-PKPbc}$.

that perturbations were restricted to the 300 km depth slice used in this study, which was assumed to have constant radial velocity. Differential traveltimes were calculated and then contaminated with 0.2 s of additive Gaussian noise, which is the same ratio of noise to velocity perturbation RMS as our final model. We then performed a hierarchical HMC run with $l' = 15$.

The results of the inversion for a DETOX-P3 synthetic data set are shown in Fig. 9, and show that nearly all major features of the DETOX-P3 model are recovered, albeit smoothed by our lower resolution inversion. The primary exception is the interior of the Pacific low velocity province (LVP), although we note that this is both flagged by a high uncertainty region by our method, and that the boundaries of the province are recovered. Fig. S7 shows that the input noise of 0.2 s is recovered by the hierarchical inversion, while Fig. 9(d) and Fig. S5(d) show the low-degree weighted power spectrum of DETOX-P3 is recovered, albeit with a reduced degree 2 component compared to the input model; we note that the recovered power spectrum from the synthetics is substantially redder than that recovered from our real data despite this, and our conclusions would not change even if the degree 2 component of our real inversion was doubled. The results of this inversion test show that our data set is in principle capable of recovering the LVPs typically found in lowermost mantle tomography derived from long period data sets, suggesting that the flatter power spectrum recovered by our model is a reflection of the details in our data, rather than a product of insufficient ray coverage.

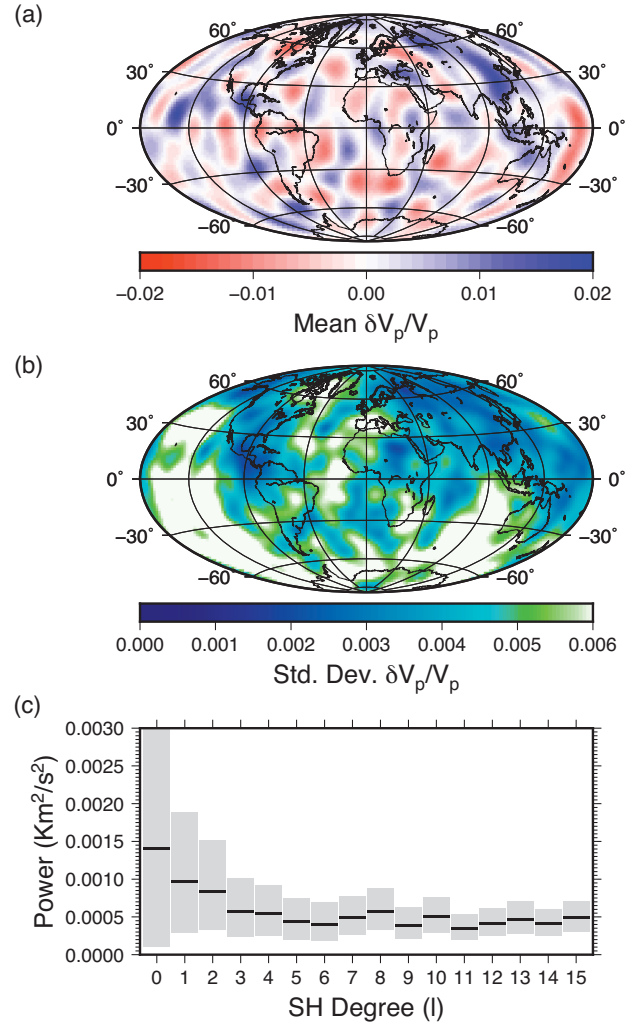


Figure 6. Summary statistics for V_p from combined HMC chains. (a) Mean, (b) standard deviation and (c) power in each harmonic degree (black line shows the median and grey bars show the 95 per cent credible interval).

6 DISCUSSION AND CONCLUSIONS

We have presented a joint inversion of lowermost mantle velocity using a high quality, handpicked data set of PcP–P and PKpab–PKPbc differential traveltimes, chosen to minimize the effects the crust and upper mantle using coincident ray paths and tomographic corrections, and independent of the inner core. Our inversion uses a hierarchical Bayesian formulation to marginalize out the effect

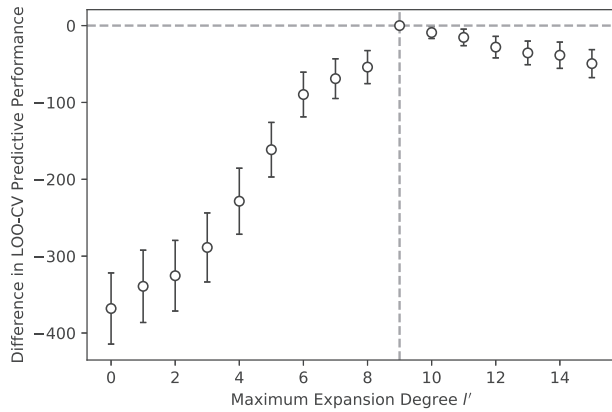


Figure 7. Estimated difference in leave-one-out cross-validation (LOO-CV) predictive performance for the synthetic resolution test relative to a maximum expansion degree $l' = 9$ —higher is better. Error bars are shown at the 2σ level.

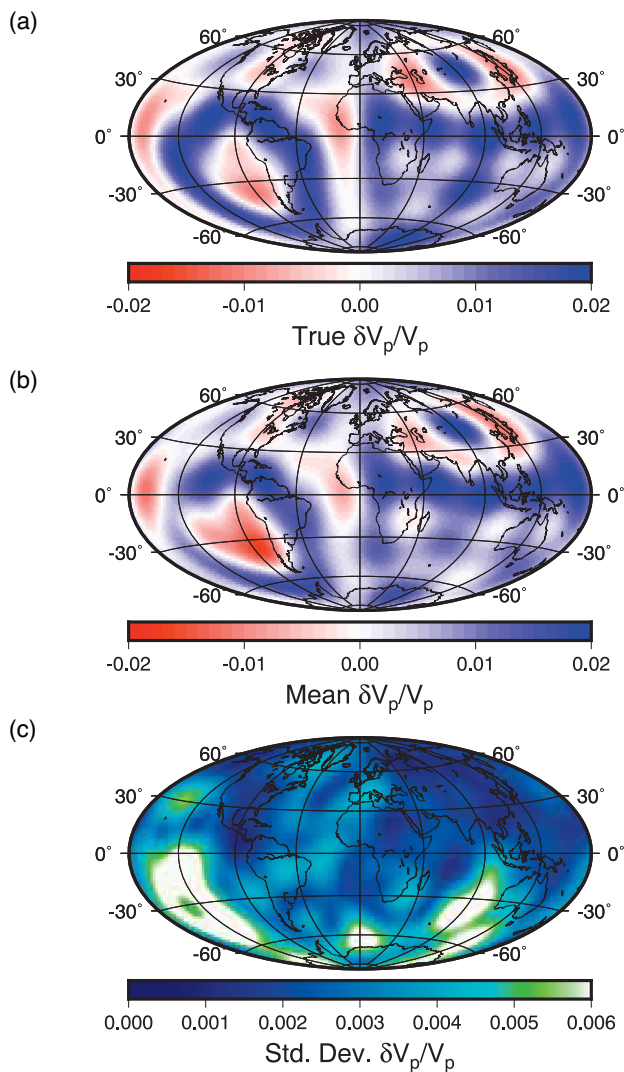


Figure 8. Summary statistics for V_p from combined HMC chains for the synthetic test on the geometric model. (a) True model, (b) mean of recovered modelled and (c) standard deviation of recovered model.

of unknown data noise and uncertain perturbation scales from the inverse problem. To handle the high-curvature non-linear posterior distribution, we used the HMC method, specifically the highly optimized STAN sampler, which uses posterior gradient information to dramatically improve sampling in high dimensions. Furthermore, we used a non-centred parametrization to avoid the ‘funnel’ effect that is a common pathology in hierarchical Bayesian models. Our results show relatively strong RMS velocity perturbations of 0.71 per cent with the pattern of CMB velocity perturbations conforming to the results of Houser *et al.* (2008) and Young *et al.* (2013). The behaviour of the LOO-CV curve in Fig. 2, with values slowly increasing after our selected degree 15 model, suggests there may be additional complexity that is not adequately captured byproduct by our forward model, noting that in contrast the synthetic LOO-CV begins decaying immediately after the true model degree. The two prime candidates for this complexity are contribution to the residuals by the topography of the CMB, considering the phases that we have used, and the use of a global P -wave tomographic model to correct for the relatively homogeneous upper and mid-mantle, which despite the use of differential traveltimes may integrate to substantial values over the long ray paths used in this study. We expect the crustal corrections to be minor given the use of differential traveltimes with paths that are very close in the crust.

Our results support a relatively complex lowermost mantle V_p model, with significant spectral power at spherical harmonic degrees up to $l = 15$, although the V_p perturbations are smaller than Young *et al.* (2013) from which a large part of our data is derived. Large perturbations in Bayesian inversions have often been ascribed to the lack of added damping to the inverse problem. We have taken an alternative perspective in this study, which is that given the existence of pre-existing high quality reference models, a good *a priori* constraint on the data is that the perturbations should be close to zero—the priors we have chosen to reflect this are equivalent to standard $L2$ Tikhonov regularization (i.e. damping) when the three scale parameters σ are fixed. The crucial difference is that we have allowed these parameters to be set parsimoniously by the data by using a hierarchical inversion (Malinverno & Briggs 2004). The final reported spatial maps and power spectra are marginalized over all values of these parameters that are compatible with the data. The combination of explicit inclusion of a normal prior on model parameters, along with the global low-degree spherical harmonic expansion which induces implicit regularization by basis truncation, means that our inversion is relatively conservative compared transdimensional Bayesian inversions such as Young *et al.* (2013). Despite this, large, spatially complex perturbations are required to account for the short-period residuals used in this study. Given the intrinsically low resolution of lowermost mantle seismic tomography, there is continued uncertainty as to whether velocity heterogeneities may be described by a purely thermal model containing a phase transition from bridgmanite to post-perovskite, or whether additional chemical heterogeneity is required (Davies *et al.* 2012; Koelemeijer *et al.* 2018).

Following the analysis of Tkalčić *et al.* (2015), the velocity perturbation model derived in this study is only weakly dominated by large scale degree 1–3 structure, which indicates compatibility with the shorter wavelength lowermost mantle structures found in thermally driven geodynamic models, and in previous tomographic studies utilizing PKP and PcP data (Tkalčić & Romanowicz 2002; Soldati *et al.* 2012). Recent large scale deterministic P -wave tomographic results with significant sensitivity to the lowermost mantle

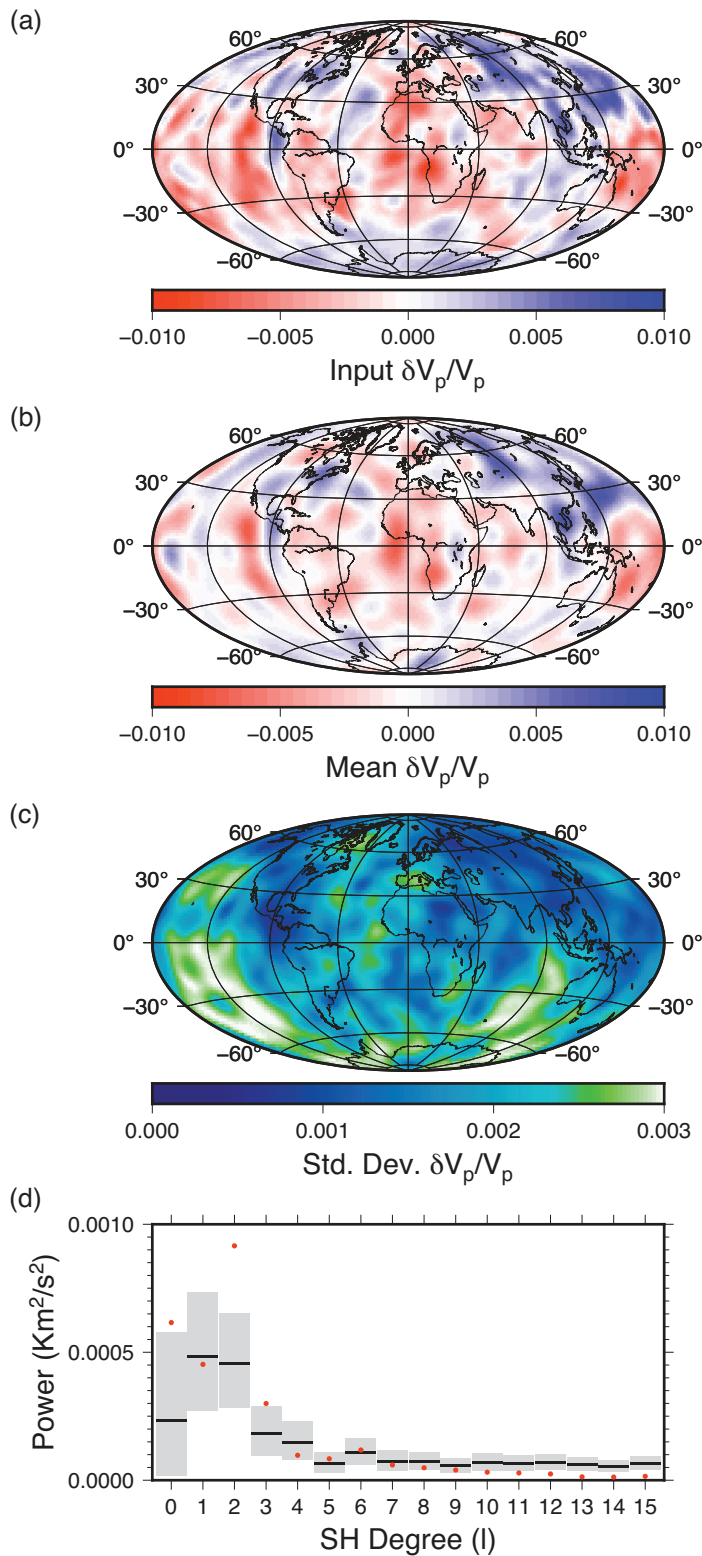


Figure 9. Summary statistics for V_p from combined HMC chains for the synthetic test on the DETOX-P3 model. (a) True model, (b) mean of recovered model, (c) standard deviation of recovered model and (d) power in each harmonic degree (black line shows the median and grey bars show the 95 per cent credible interval, red dots show power of the input model).

obtained by Pdiff data sets, reported by Hosseini *et al.* (2020), have also indicated substantial structural complexity with the most predominant features being a fast patch under East Asia and a wide, quasi-hemispherical slow band that includes the locations of the

African and Pacific LVPs. Pdiff data sets provide the most comprehensive coverage of lowermost mantle V_p structure (Hosseini & Sigloch 2015; Euler & Wysession 2017), however their maximum frequency content is limited by the diffraction process and so our

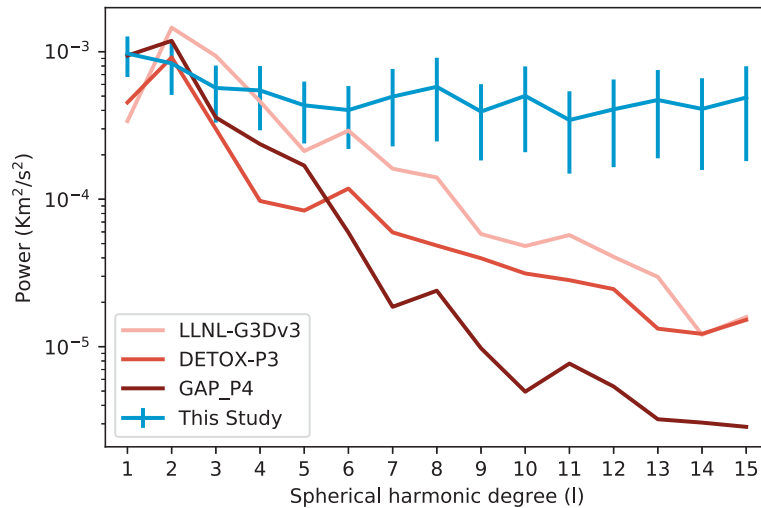


Figure 10. Comparison of power per degree between this study and three global P -wave tomography models evaluated at a depth of 2741 km: LLNL-G3Dv3, DETOX-P3 and GAP_P4, showing that the results of this study support a less red spectrum for V_P perturbations in the lowermost mantle than those models. Vertical lines indicate the 95 per cent credible interval for the results of our model.

high-frequency PcP-P and PKPab-PKPbc provide useful additional constraints on the finer scale structure of the lowermost mantle region. Quantifying these observations, Fig. 10 shows a comparison in power per degree between this study and three global V_P tomography methods—GAP_P4 (Obayashi *et al.* 2013) and LLNL-G3Dv3 (Simmons *et al.* 2012), which are well established 3-D models based principally on direct P -wave data in the lowermost mantle, and DETOX-P3 (Hosseini *et al.* 2020), which as earlier stated utilizes Pdiff to enhance sensitivity to the lowermost mantle. We find that the preferred power spectrum of our model is significantly less red than these global V_P models, supporting the findings of significant multiscale structure of Tkalčić *et al.* (2015) albeit now using a data set that is free from any contribution from the inner core.

The most prominent similarity of our model with previous P -wave tomography models of the lowermost mantle is also this high speed feature obtained under the entire Asian continent. That faster patch in our model, however, continues from Asia across Indonesia and extends to Australia. This is different from previous models (e.g. Tkalčić *et al.* 2002; Houser *et al.* 2008; Soldati *et al.* 2012; Young *et al.* 2013) which all had moderate to low speeds in the area between Asia and Australia. We attribute the difference between heterogeneity maps obtained by Tkalčić *et al.* (2002) and Young *et al.* (2013) and our model in that area to inner core structure, which could be anomalously slow and thus affect PKPbc traveltimes that were included in those studies, but are omitted from the inversion here by considering PKPab-PKPbc traveltimes only. Aubert *et al.* (2008) argued for the existence of a large downward motion of chemically depleted liquid in the outer core beneath the southeast Asia, which is compatible with the highest-rate growth of the inner core. The fast growth would induce a small grain texture at the top of the inner core, which could be responsible for anomalously low speeds.

We note another significant change, beneath Americas, with respect to the previous models. In the models that included the PKPbc traveltimes data, both North and South America, including the southern parts of the Pacific Ocean had pronounced low speeds. We do not observe this low speed area in the current model. At least part of these anomalies in the previous models could be attributed either to low speed of the inner core on the source side of the paths from South America to east Asia and on the receiver side from the southern Indian Ocean to the west coast of the United States (see fig. 8

in Tkalčić *et al.* (2002) and animation 3 in Tkalčić (2017) for more details on those ray paths). Alternatively, a cylindrical anisotropy in the inner core with a slow axis for intermediate angles w.r.t. the Earth's spin axis could be responsible for mapping into the previous models of the lowermost mantle. Such model of anisotropy had been proposed before for the innermost part of the inner core (Ishii & Dziewonski 2002).

The further interpretation of these features is contingent on future joint tomographic imaging of the lowermost mantle and CMB topography, ideally including phases with a high sensitivity to CMB topography, such as P4KP, as well as better ray coverage, which will give greater constraints on the relative contributions of isostasy and dynamic topography at the CMB and hence our understanding of the density and viscosity profiles of the deep mantle.

7 SUPPLEMENTARY MATERIAL

We also show a Pacific-centred version of Fig. 6 in Fig. S1. The supplementary material also contains the results of an inversion restricted to maximum degree $l = 6$ in Figs S2 and S3, which show long wavelength structures including dominant slow regions underneath Africa and the southwest Pacific. Note that this model is strongly unpreferred by PSIS-LOO analysis compared to our presented model at maximum degree $l = 15$, which indicates strong regional-scale heterogeneity is required to account for the short-period data used in this study. The apparent uncertainty of these restricted l models is low—this is a reflection of the inexpressiveness of the low degree parametrization being unable to correctly model data, leading to high residuals and hence a strongly restricted likelihood function. Fig. S4 shows the binary model used to generate the filtered degree 10 model used in our geometric synthetic inversion to test PSIS-LOO. Figs S5–S8 show additional results pertaining to the synthetic test performed on the DETOX-P3 model.

ACKNOWLEDGEMENTS

The authors would like to thank associate editor Frederik Simons and editorial assistants Fern Storey and Anna Evripidou for handling the publication process, and two anonymous reviewers for substantially improving the manuscript. JBM would like to thank the

General Sir John Monash Foundation and the Origin Energy Foundation for financial support. In addition, JBM would like to thank Michael Betancourt for providing an intensive STAN workshop gratis at Caltech, that helped to reinvigorate this study. The newly analysed PKPab-PKPbc dataset can be constructed from the raw PKPab-PKPbc and PKPbc-PKPbc data found at http://rses.anu.edu.au/~hrvoje/+INNER_CORE_SUPPLEMENTS/IC_suppl.html; the PcP-P data were previously reported.

REFERENCES

- Amit, H. & Choblet, G., 2009. Mantle-driven geodynamo features—effects of post-Perovskite phase transition, *Earth, Planets Space*, **61**(11), 1255–1268.
- Amit, H. & Choblet, G., 2012. Mantle-driven geodynamo features—effects of compositional and narrow D anomalies, *Phys. Earth planet. Inter.*, **190–191**, 34–43.
- Amit, H., Choblet, G., Olson, P., Monteux, J., Deschamps, F., Langlais, B. & Tobie, G., 2015. Towards more realistic core-mantle boundary heat flux patterns: a source of diversity in planetary dynamos, *Prog. Earth planet. Sci.*, **2**(1), doi:10.1186/s40645-015-0056-3.
- Aster, R.C., Borchers, B. & Thurber, C.H., 2018. *Parameter Estimation and Inverse Problems*, Elsevier.
- Attanayake, J., Cormier, V.F. & de Silva, S.M., 2014. Uppermost inner core seismic structure—new insights from body waveform inversion, *Earth planet. Sci. Lett.*, **385**, 49–58.
- Aubert, J., Amit, H., Hulot, G. & Olson, P., 2008. Thermochemical flows couple the Earth's inner core growth to mantle heterogeneity, *Nature*, **454**(7205), 758–761.
- Auer, L., Boschi, L., Becker, T.W., Nissen-Meyer, T. & Giardini, D., 2014. Savani: a variable resolution whole-mantle model of anisotropic shear velocity variations based on multiple data sets, *J. geophys. Res.: Solid Earth*, **119**(4), 3006–3034.
- Becker, T.W. & Boschi, L., 2002. A comparison of tomographic and geodynamic mantle models, *Geochem. Geophys. Geosyst.*, **3**(1), doi:10.1029/2001GC000168.
- Betancourt, M., 2017. A conceptual introduction to Hamiltonian Monte Carlo, arXiv:1701.02434 [stat].
- Betancourt, M.J. & Girolami, M., 2015. Hamiltonian Monte Carlo for hierarchical models, *Current trends in Bayesian methodology with applications*, 79–101.
- Betancourt, M., Byrne, S., Livingstone, S. & Girolami, M., 2017. The geometric foundations of Hamiltonian Monte Carlo, *Bernoulli*, **23**(4A), 2257–2298.
- Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y. & Wassermann, J., 2010. ObsPy: a Python toolbox for seismology, *Seismol. Res. Lett.*, **81**(3), 530–533.
- Biswas, R. & Sen, M., 2017. 2D full-waveform inversion and uncertainty estimation using the reversible jump Hamiltonian Monte Carlo, in *SEG Technical Program Expanded Abstracts 2017*, pp. 1280–1285, Society of Exploration Geophysicists.
- Bloxham, J. & Gubbins, D., 1987. Thermal core–mantle interactions, *Nature*, **325**(6104), 511.
- Bodin, T. & Sambridge, M., 2009. Seismic tomography with the reversible jump algorithm, *Geophys. J. Int.*, **178**(3), 1411–1436.
- Bodin, T., Sambridge, M., Rawlinson, N. & Arroucau, P., 2012. Transdimensional tomography with unknown data noise: transdimensional tomography, *Geophys. J. Int.*, **189**(3), 1536–1556.
- Bower, D.J., Gurnis, M. & Sun, D., 2013. Dynamic origins of seismic wavespeed variation in D", *Phys. Earth planet. Inter.*, **214**, 74–86.
- Burdick, S. & Lekić, V., 2017. Velocity variations and uncertainty from transdimensional P-wave tomography of North America, *Geophys. J. Int.*, **209**(2), 1337–1351.
- Burdick, S., Waszek, L. & Lekić, V., 2019. Seismic tomography of the uppermost inner core, *Earth planet. Sci. Lett.*, **528**, 115789, doi:10.1016/j.epsl.2019.115789.
- Carpenter, B., et al., 2017. Stan: a probabilistic programming language, *J. Stat. Software*, **76**(1), doi:10.18637/jss.v076.i01.
- Claeskens, G., 2016. Statistical model choice, *Ann. Rev. Stat. Appl.*, **3**(1), 233–256.
- Cormier, V.F. & Choy, G.L., 1986. A search for lateral heterogeneity in the inner core from differential traveltimes near PKP-D and PKP-C, *Geophys. Res. Lett.*, **13**(13), 1553–1556.
- Cottaar, S. & Lekić, V., 2016. Morphology of seismically slow lower-mantle structures, *Geophys. J. Int.*, **207**(2), 1122–1136.
- Davies, D.R., Goes, S., Davies, J., Schubert, B., Bunge, H.-P. & Ritsema, J., 2012. Reconciling dynamic and seismic models of Earth's lower mantle: the dominant role of thermal heterogeneity, *Earth planet. Sci. Lett.*, **353–354**, 253–269.
- Deal, M.M. & Nolet, G., 1996. Nullspace shuttles, *Geophys. J. Int.*, **124**(2), 372–380.
- Della Mora, S., Boschi, L., Tackley, P.J., Nakagawa, T. & Giardini, D., 2011. Low seismic resolution cannot explain S/P decorrelation in the lower mantle, *Geophys. Res. Lett.*, **38**(12), doi:10.1029/2011GL047559.
- Dettmer, J., Dosso, S.E. & Holland, C.W., 2011. Sequential trans-dimensional Monte Carlo for range-dependent geoaoustic inversion, *J. acoust. Soc. Am.*, **129**(4), 1794–1806.
- Dziewonski, A.M. & Anderson, D.L., 1981. Preliminary reference Earth model, *Phys. Earth planet. Inter.*, **25**(4), 297–356.
- Dziewonski, A.M. & Gilbert, F., 1976. The effect of small, aspherical perturbations on travel times and a re-examination of the corrections for ellipticity, *Geophys. J. R. astr. Soc.*, **44**(1), 7–17.
- Dziewonski, A.M., Lekic, V. & Romanowicz, B.A., 2010. Mantle anchor structure: an argument for bottom up tectonics, *Earth planet. Sci. Lett.*, **299**(1–2), 69–79.
- Euler, G.G. & Wyssession, M.E., 2017. Geographic variations in lowermost mantle structure from the ray parameters and decay constants of core-diffracted waves: CMBR structure from P_{diff} and S_{diff} , *J. geophys. Res.: Solid Earth*, **122**(7), 5369–5394.
- Fang, H., van der Hilst, R.D., de Hoop, M.V., Kothari, K., Gupta, S. & Dokmanić, I., 2020. Parsimonious seismic tomography with Poisson Voronoi projections: methodology and validation, *Seismol. Res. Lett.*, **91**(1), 343–355.
- Fichtner, A. & Simutė, S., 2018. Hamiltonian Monte Carlo inversion of seismic sources in complex media, *J. geophys. Res.*, **123**(4), 2984–2999.
- Fichtner, A. & Zunino, A., 2019. Hamiltonian nullspace shuttles, *Geophys. Res. Lett.*, **46**(2), 644–651.
- Fichtner, A., Zunino, A. & Gebraad, L., 2019. Hamiltonian Monte Carlo solution of tomographic inverse problems, *Geophys. J. Int.*, **216**(2), 1344–1363.
- Frost, D.A., Rost, S., Garnero, E.J. & Li, M., 2017. Seismic evidence for Earth's crusty deep mantle, *Earth planet. Sci. Lett.*, **470**, 54–63.
- Fukao, Y. & Obayashi, M., 2013. Subducted slabs stagnant above, penetrating through, and trapped below the 660 km discontinuity, *J. geophys. Res.*, **118**(11), 5920–5938.
- Garcia, R.F., Chevrot, S. & Calvet, M., 2009. Statistical study of seismic heterogeneities at the base of the mantle from PKP differential traveltimes, *Geophys. J. Int.*, **179**(3), 1607–1616.
- Garnero, E.J. & McNamara, A.K., 2008. Structure and dynamics of Earth's lower mantle, *Science*, **320**(5876), 626.
- Garnero, E.J., McNamara, A.K. & Shim, S.-H., 2016. Continent-sized anomalous zones with low seismic velocity at the base of Earth's mantle, *Nat. Geosci.*, **9**(7), 481–489.
- Gebraad, L., Boehm, C. & Fichtner, A., 2020. Bayesian elastic full-waveform inversion using Hamiltonian Monte Carlo, *J. geophys. Res.*, **125**(3), doi:10.1029/2019JB018428.
- Gelman, A., 2006. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper), *Bayesian Anal.*, **1**(3), 515–534.
- Gibbons, S.J. & Gubbins, D., 2000. Convection in the Earth's core driven by lateral variations in the core-mantle boundary heat flux, *Geophys. J. Int.*, **142**, 631–642.
- Girolami, M. & Calderhead, B., 2011. Riemann manifold Langevin and Hamiltonian Monte Carlo methods: Riemann Manifold Langevin

- and Hamiltonian Monte Carlo Methods, *J. R. Stat. Soc., B*, **73**(2), 123–214.
- Gubbins, D., Willis, A.P. & Sreenivasan, B., 2007. Correlation of Earth's magnetic field with lower mantle thermal and seismic structure, *Phys. Earth planet. Inter.*, **162**(3–4), 256–260.
- Gubbins, D., Sreenivasan, B., Mound, J. & Rost, S., 2011. Melting of the Earth's inner core, *Nature*, **473**(7347), 361–363.
- Helfrich, G.R. & Wood, B.J., 2001. The Earth's mantle, *Nature*, **412**, 501.
- Hosseini, K. & Sigloch, K., 2015. Multifrequency measurements of core-diffracted P waves (Pdiff) for global waveform tomography, *Geophys. J. Int.*, **203**(1), 506–521.
- Hosseini, K., Matthews, K.J., Sigloch, K., Shephard, G.E., Domeier, M. & Tsekhmistrenko, M., 2018. SubMachine: web-based tools for exploring seismic tomography and other models of Earth's deep interior, *Geochem. Geophys. Geosyst.*, **19**(5), 1464–1483.
- Hosseini, K., Sigloch, K., Tsekhmistrenko, M., Zaheri, A., Nissen-Meyer, T. & Igel, H., 2020. Global mantle structure from multifrequency tomography using P, PP and P-diffracted waves, *Geophys. J. Int.*, **220**(1), 96–141.
- Houser, C., Masters, G., Shearer, P. & Laske, G., 2008. Shear and compressional velocity models of the mantle from cluster analysis of long-period waveforms, *Geophys. J. Int.*, **174**(1), 195–212.
- Irving, J. & Deuss, A., 2015. Regional seismic variations in the inner core under the North Pacific, *Geophys. J. Int.*, **203**(3), 2189–2199.
- Ishii, M. & Dziewonski, A.M., 2002. The innermost inner core of the earth: Evidence for a change in anisotropic behavior at the radius of about 300 km, *Proc. Natl. Acad. Sci.*, **99**(22), 14026–14030.
- Jackson, A., Jonkers, A.R.T. & Walker, M.R., 2000. Four centuries of geomagnetic secular variation from historical records, *Phil. Trans. R. Soc. Lond., A*, **358**(1768), 957–990.
- Kennett, B.L.N., Engdahl, E.R. & Buland, R., 1995. Constraints on seismic velocities in the Earth from traveltimes, *Geophys. J. Int.*, **122**(1), 108–124.
- Koelemeijer, P., Schubert, B., Davies, D., Deuss, A. & Ritsema, J., 2018. Constraints on the presence of post-perovskite in Earth's lowermost mantle from tomographic-geodynamic model comparisons, *Earth planet. Sci. Lett.*, **494**, 226–238.
- Kowsari, M., Halldorsson, B., Hrafnkelsson, B. & Jónsson, S., 2019. Selection of earthquake ground motion models using the deviance information criterion, *Soil Dyn. Earthq. Eng.*, **117**, 288–299.
- Lay, T. & Helmberger, D.V., 1983. A lower mantle S-wave triplication and the shear velocity structure of D", *Geophys. J. R. astr. Soc.*, **75**(3), 799–837.
- Lekić, V., Cottaar, S., Dziewonski, A. & Romanowicz, B., 2012. Cluster analysis of global lower mantle tomography: a new class of structure and implications for chemical heterogeneity, *Earth planet. Sci. Lett.*, **357–358**, 68–77.
- Leykam, D., Tkalčić, H. & Reading, A.M., 2010. Core structure re-examined using new teleseismic data recorded in Antarctica: evidence for, at most, weak cylindrical seismic anisotropy in the inner core, *Geophys. J. Int.*, **180**(3), 1329–1343.
- Li, X.-D. & Romanowicz, B., 1996. Global mantle shear velocity model developed using nonlinear asymptotic coupling theory, *J. geophys. Res.*, **101**(B10), 22 245–22 272.
- Malinverno, A. & Briggs, V.A., 2004. Expanded uncertainty quantification in inverse problems: hierarchical Bayes and empirical Bayes, *Geophysics*, **69**(4), 1005–1016.
- McNamara, A.K., 2018. A review of large low shear velocity provinces and ultra low velocity zones, *Tectonophysics*, **760**, 199–220.
- Meschede, M. & Romanowicz, B., 2015. Lateral heterogeneity scales in regional and global upper mantle shear velocity models, *Geophys. J. Int.*, **200**(2), 1078–1095.
- Mosca, I., Cobden, L., Deuss, A., Ritsema, J. & Trampert, J., 2012. Seismic and mineralogical structures of the lower mantle from probabilistic tomography, *J. geophys. Res.*, **117**(B6), doi:10.1029/2011JB008851.
- Moulik, P. & Ekström, G., 2016. The relationships between large-scale variations in shear velocity, density, and compressional velocity in the Earth's mantle, *J. geophys. Res.*, **121**(4), 2737–2771.
- Muir, J.B. & Tkalčić, H., 2015. A method of spherical harmonic analysis in the geosciences via hierarchical Bayesian inference, *Geophys. J. Int.*, **203**(2), 1164–1171.
- Muir, J.B. & Tsai, V.C., 2020. Did Oldham discover the core after all? Handling imprecise historical data with hierarchical Bayesian model selection methods, *Seismol. Res. Lett.*, **91**(3), 1377–1383.
- Nakagawa, T. & Tackley, P.J., 2008. Lateral variations in CMB heat flux and deep mantle seismic velocity caused by a thermal–chemical-phase boundary layer in 3D spherical convection, *Earth planet. Sci. Lett.*, **271**(1–4), 348–358.
- Neal, R.M., 2003. Slice sampling, *Ann. Stat.*, **31**(3), 705–767.
- Neal, R.M., 2011. MCMC using Hamiltonian dynamics, in *Handbook of Markov Chain Monte Carlo*, eds S. Brooks, A. Gelman, G. L. Jones, X.-L. Meng, Chapman & Hall/CRC Press.
- Obayashi, M., Yoshimitsu, J., Nolet, G., Fukao, Y., Shiobara, H., Sugioka, H., Miyamachi, H. & Gao, Y., 2013. Finite frequency whole mantle P wave tomography: improvement of subducted slab images, *Geophys. Res. Lett.*, **40**(21), 5652–5657.
- Pachhai, S., Tkalčić, H. & Dettmer, J., 2014. Bayesian inference for ultralow velocity zones in the Earth's lowermost mantle: Complex ULVZ beneath the east of the Philippines: Bayesian Inference for ULVZs, *J. geophys. Res.*, **119**(11), 8346–8365.
- Pejić, T., Tkalčić, H., Sambridge, M., Cormier, V.F. & Benavente, R., 2017. Attenuation tomography of the upper inner core, *J. geophys. Res.*, **122**(4), 3008–3032.
- Pejić, T., Hawkins, R., Sambridge, M. & Tkalčić, H., 2019. Transdimensional Bayesian attenuation tomography of the upper inner core, *J. geophys. Res.*, **124**(2), 1929–1943.
- Ritsema, J., Deuss, A., van Heijst, H.J. & Woodhouse, J.H., 2011. S40RTS: a degree-40 shear-velocity model for the mantle from new Rayleigh wave dispersion, teleseismic traveltimes and normal-mode splitting function measurements: S40RTS, *Geophys. J. Int.*, **184**(3), 1223–1236.
- Sambridge, M., Bodin, T., Gallagher, K. & Tkalčić, H., 2013. Transdimensional inference in the geosciences, *Phil. Trans. R. Soc., A*, **371**(1984), 20110547, doi:10.1098/rsta.2011.0547.
- Sen, M.K. & Biswas, R., 2017. Transdimensional seismic inversion using the reversibly jump Hamiltonian Monte Carlo algorithm, *Geophysics*, **82**(3), R119–R134.
- Shang, X., Shim, S.-H., de Hoop, M. & van der Hilst, R., 2014. Multiple seismic reflectors in Earth's lowermost mantle, *Proc. Natl. Acad. Sci.*, **111**(7), 2442–2446.
- Shephard, G.E., Matthews, K.J., Hosseini, K. & Domeier, M., 2017. On the consistency of seismically imaged lower mantle slabs, *Sci. Rep.*, **7**(1), doi:10.1038/s41598-017-11039-w.
- Simmons, N.A., Myers, S.C., Johannesson, G. & Matzel, E., 2012. LLNL-G3Dv3: global P wave tomography model for improved regional and teleseismic travel time prediction, *J. geophys. Res.*, **117**(B10), doi:10.1029/2012JB009525.
- Soldati, G., Boschi, L. & Forte, A.M., 2012. Tomography of core-mantle boundary and lowermost mantle coupled by geodynamics: tomography of core-mantle boundary, *Geophys. J. Int.*, **189**(2), 730–746.
- Soldati, G., Boschi, L., Mora, S.D. & Forte, A.M., 2014. Tomography of core-mantle boundary and lowermost mantle coupled by geodynamics: joint models of shear and compressional velocity, *Ann. Geophys.*, **57**(6), S0652.
- Stevenson, D.J., 1981. Models of the Earth's core, *Science*, **214**(4521), 611–619.
- Tan, E., Gurnis, M. & Han, L., 2002. Slabs in the lower mantle and their modulation of plume formation, *Geochem. Geophys. Geosyst.*, **3**(11), 1–24.
- Tanaka, S., 2010. Constraints on the core-mantle boundary topography from P 4 KP PLBIBITALIC–PcP differential travel times, *J. geophys. Res.*, **115**(B4), doi:10.1029/2009JB006563.
- Tanaka, S. & Hamaguchi, H., 1997. Degree one heterogeneity and hemispherical variation of anisotropy in the inner core from PKP (BC)–PKP (DF) times, *J. geophys. Res.*, **102**(B2), 2925–2938.
- Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM.
- Tesoniero, A., Auer, L., Boschi, L. & Cammarano, F., 2015. Hydration of marginal basins and compositional variations within the continental lithospheric mantle inferred from a new global model of shear and compressional velocity, *J. geophys. Res.*, **120**(11), 7789–7813.

- Tkalčić, H., 2017. *The Earth's Inner Core: Revealed by Observational Seismology*, Cambridge Univ. Press.
- Tkalčić, H. & Romanowicz, B., 2002. Short scale heterogeneity in the lowermost mantle: insights from PcP-P and ScS-S data, *Earth planet. Sci. Lett.*, **201**(1), 57–68.
- Tkalčić, H., Romanowicz, B. & Houy, N., 2002. Constraints on D'' structure using PKP(AB-DF), PKP(BC-DF) and PcP-P traveltime data from broadband records: constraints on D'' structure, *Geophys. J. Int.*, **149**(3), 599–616.
- Tkalčić, H., Young, M., Muir, J.B., Davies, D.R. & Mattesini, M., 2015. Strong, multi-scale heterogeneity in Earth's lowermost mantle, *Sci. Rep.*, **5**, 18416.
- van der Hilst, R.D., Widiyantoro, S. & Engdahl, E.R., 1997. Evidence for deep mantle circulation from global tomography, *Nature*, **386**(6625), 578–584.
- van der Hilst, R.D., de Hoop, M.V., Wang, P., Shim, S.-H., Ma, P. & Tenorio, L., 2007. Seismostratigraphy and thermal structure of Earth's core-mantle boundary region, *Science*, **315**(5820), 1813–1817.
- Vehtari, A. & Ojanen, J., 2012. A survey of Bayesian predictive methods for model assessment, selection and comparison, *Stat. Surv.*, **6**(0), 142–228.
- Vehtari, A., Gelman, A. & Gabry, J., 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC, *Stat. Comput.*, **27**(5), 1413–1432.
- Yee, T.-G., Rhie, J. & Tkalčić, H., 2014. Regionally heterogeneous uppermost inner core observed with Hi-net array: variations of isotropic V_p in the UIC, *J. geophys. Res.*, **119**(10), 7823–7845.
- Young, M.K., Tkalčić, H., Bodin, T. & Sambridge, M., 2013. Global P wave tomography of Earth's lowermost mantle from partition modeling, *J. geophys. Res.*, **118**(10), 5467–5486.

SUPPORTING INFORMATION

Supplementary data are available at [GJI](https://doi.org/10.1093/gji/ggaa000) online.

Figure S1. Summary statistics for V_p from combined HMC chains, Pacific centred view. (a) Mean, (b) standard deviation and (c) power in each harmonic degree (black line shows the median and grey bars show the 95 per cent credible interval).

Figure S2. Summary statistics for V_p from combined HMC chains for an inversion restricted to $l' = 6$. (A) Mean, (b) standard deviation and (c) power in each harmonic degree (black line shows the median and grey bars show the 95 per cent credible interval).

Figure S3. Summary statistics for V_p from combined HMC chains for an inversion restricted to $l' = 6$, Pacific centred view. (a) Mean, (b) standard deviation and (c) power in each harmonic degree (black line shows the median and grey bars show the 95 per cent credible interval).

Figure S4. Unfiltered binary synthetic model used for the geometric synthetic test.

Figure S5. Summary statistics for V_p from combined HMC chains for the synthetic test on the DETOX-P3 model, including power per degree, Pacific centred view. (A) Synthetic model, (b) mean recovered model, (c) standard deviation and (d) power in each harmonic degree (black line shows the median and grey bars show the 95 per cent credible interval, red dots show power of the input model).

Figure S6. Traveltime residuals for the 2 data sets used in this study using the DETOX-P3 as a synthetic input model and 0.2 s applied Gaussian noise. Black histograms show the initial distribution of residuals, and purple/orange histograms show the quantiles of remaining residual distributions after inversion with maximum $l' = 15$.

Figure S7. Posterior distributions of scale variables σ_{dt} , σ_{PcP-P} and $\sigma_{PKPab-PKPbc}$ for maximum $l' = 15$ using the DETOX-P3 as a synthetic input model and 0.2s applied Gaussian noise. Prior distributions were half-normal for all three scale variables, with prior standard deviation $\nu_1 = 5$ for σ_{dt} and $\nu_2 = 2$ for σ_{PcP-P} and $\sigma_{PKPab-PKPbc}$.

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the paper.