

Velocity variations and uncertainty from transdimensional *P*-wave tomography of North America

Scott Burdick and Vedran Lekić

Department of Geology, University of Maryland, College Park, MD 20742, USA. E-mail: scott.a.burdick@gmail.com

Accepted 2017 March 1. Received 2017 February 13; in original form 2016 August 20

SUMMARY

High-resolution models of seismic velocity variations constructed using body-wave tomography inform the study of the origin, fate and thermochemical state of mantle domains. In order to reliably relate these variations to material properties including temperature, composition and volatile content, we must accurately retrieve both the patterns and amplitudes of variations and quantify the uncertainty associated with the estimates of each. For these reasons, we image the mantle beneath North America with *P*-wave traveltimes from USArray using a novel method for 3-D probabilistic body-wave tomography. The method uses a Transdimensional Hierarchical Bayesian framework with a reversible-jump Markov Chain Monte Carlo algorithm in order to generate an ensemble of possible velocity models. We analyse this ensemble solution to obtain the posterior probability distribution of velocities, thereby yielding error bars and enabling rigorous hypothesis testing. Overall, we determine that the average uncertainty (1σ) of compressional wave velocity estimates beneath North America is ~ 0.25 per cent dV_P/V_P , increasing with proximity to complex structure and decreasing with depth. The addition of USArray data reduces the uncertainty beneath the Eastern US by over 50 per cent in the upper mantle and 25–40 per cent below the transition zone and ~ 30 per cent throughout the mantle beneath the Western US. In the absence of damping and smoothing, we recover amplitudes of variations 10–80 per cent higher than a standard inversion approach. Accounting for differences in data coverage, we infer that the length scale of heterogeneity is ~ 50 per cent longer at shallow depths beneath the continental platform than beneath tectonically active regions. We illustrate the model trade-off analysis for the Cascadia slab and the New Madrid Seismic Zone, where we find that smearing due to the limitations of the illumination is relatively minor.

Key words: Composition and structure of the mantle; Inverse theory; Body waves; Seismic tomography.

1 INTRODUCTION

Body-wave tomography is a powerful tool for understanding the present-day structure of the Earth's mantle. By minimizing the misfit between observed and predicted seismic data (e.g. traveltimes, waveforms), models of 3-D variations in seismic velocity, anisotropy, and attenuation can be created. Although analysis of these seismic models can be interesting in itself, ultimately we wish to use the models as evidence to test hypotheses about Earth processes and structure. The geometry, volume, and continuity of features observed in tomographic models can help test hypotheses concerning the character and timing of past episodes of subduction, the preservation of chemical heterogeneity, and the viscosity structure of the mantle. Relative variations in seismic velocity can be related to effects from temperature (e.g. Cammarano *et al.* 2003), grain size (e.g. Faul & Jackson 2005), lattice preferred orientation (e.g. Karato *et al.* 2008), partial melting (e.g. Hammond & Humphreys 2000) and volatile content (e.g. Jacobsen & Smyth 2013). In turn, the

variation in these properties informs our understanding of mantle convection, thermal evolution, and hydration.

Recordings from USArray, the seismological component of EarthScope, have revealed the mantle beneath North America at unprecedented resolution, motivating the development of hypotheses concerning the processes at work within. Numerous body wave studies have been performed with USArray data on regional, continental, and global scales, with schemes ranging from ray tracing with first arrivals (e.g. Burdick *et al.* 2008) to finite-frequency traveltime kernels (e.g. Sigloch *et al.* 2008) and waveform inversion (e.g. Schaeffer & Lebedev 2014). These tomographic models have been used to formulate hypotheses concerning, for example, the gap in seismicity along the Cascadia subduction zone (Roth *et al.* 2008; Burdick *et al.* 2009), the magmatic system beneath the Yellowstone hotspot (Huang *et al.* 2015), and its connection with the deep mantle (Schmandt *et al.* 2012; Tian & Zhao 2012). Models incorporating USArray data have tracked relict slabs in the deep mantle beneath the eastern continental margin (Pavlis *et al.* 2012; Sigloch &

Mihalynuk 2013) and, combined with geodynamic modelling, helped constrain the geometry and timing of past subduction.

As USArray moved into the eastern half of the United States where data were previously even more sparse, studies based on new models addressed hypotheses concerning mechanisms that control the length scale of structural variations (Lou & van der Lee 2014), the connection between rift structure and deeper processes at the Midcontinent Rift (Burdick *et al.* 2014) and the Reelfoot Rift (Chen *et al.* 2014), and the thermochemical origins of anomalous volcanism in Virginia (Schmandt & Lin 2014; Biryol *et al.* 2016) and at the Great Meteor hotspot track (Villemare *et al.* 2012).

Relating tomographic body-wave models to Earth properties is made difficult by a number of limitations arising from standard tomographic inversion. First, the ability to determine the size and continuity of features such as deep slabs and plume conduits is complicated by regularization (e.g. damping and smoothing) which can artificially broaden or smear structures. Second, in order to infer properties like temperature and partial melt, it is necessary to recover the true amplitude of variations in seismic velocity. Damping velocity values towards a background model can lead to an underestimation of the amplitude of the variations and obscure the strength of the mechanisms causing them. Third, the scale of variations and spatial gradients of seismic heterogeneity offer insight into viscosity structure and tectonic development, but their recovery is affected by uneven data coverage and choices in model parametrization. The accuracy of these observations is further affected by spatial smoothing.

Hypothesis-testing of Earth structure and processes requires that a probability be assigned to a given velocity model. However, standard tomographic methods yield a single velocity model that best fits the observed data, along with an approximated estimate of uncertainty about that model. Therefore, in order to assign uncertainty to inferences on the continuity of features, petrological properties, and scale of structure that we draw from seismic models, it is imperative that we understand the full range of velocities that explain the data. Yet, rigorous examination of model uncertainty is exceedingly difficult, and therefore also rare, in the field of body-wave tomography. Uncertainty is commonly assessed via ‘checkerboard’ resolution tests, wherein synthetic data are generated in a trial model and inverted to estimate where structure can and cannot be recovered. These tests are useful for giving a qualitative picture of where the Earth is resolved by the data, but they have two important shortcomings: (1) since the synthetic data are often generated using the same forward modelling used in the tomographic inversion, they do not account for error in physical approximation; (2) they must be repeated for each scale of target heterogeneity (Leveque *et al.* 1993). Formal model covariance can also be estimated, but it is commonly assumed that error can be represented by a generalized Gaussian centred on the best model. Therefore, the uncertainty estimates are usually only valid in the immediate vicinity of the best fit model, and more complicated aspects of tomographic model uncertainty—including multimodal distributions—are often ignored.

Efforts to overcome these limitations have tended to follow probabilistic approaches. For example, probabilistic tomography methods were developed to infer the distribution of chemical and temperature heterogeneity (Deschamps & Trampert 2003; Trampert & Yuen 2004). Recently, much work has been done in applying transdimensional inversion methods implemented with reversible-jump Markov Chain Monte Carlo (rj-MCMC; Green 1995; Sambridge *et al.* 2006) to estimate the posterior probability distributions of geophysical models (e.g. Bodin & Sambridge 2009; Agostinetti & Malinverno 2010; Dettmer *et al.* 2010). Development has pro-

ceeded in two major directions. First, through the use of transdimensional parametrizations, which allow for models of arbitrary scale and sharpness, the numbers of model parameters are allowed to vary based on the demands of the data coverage and actual structural variations (Malinverno 2002). Second, hierarchical parameters have been introduced that enable the variance of the data errors to be estimated simultaneously with the velocity structure, resulting in probability distributions that account for the total uncertainty in the data (Malinverno & Briggs 2004). When applied to joint inversions of multiple data types, this approach yields optimal weights for those data types (Bodin *et al.* 2012b). A recent overview of Bayesian developments in geophysics can be found in Sambridge *et al.* (2013).

In this paper, we follow the work of Bodin *et al.* (2012a) on surface waves and apply Transdimensional Hierarchical Bayesian (THB) tomography to the body wave problem. The algorithm does not require damping and smoothing of the model in order to find a solution. Instead, possible values for the model are explored through repeated forward modelling of the data. Therefore, the output of the algorithm is not simply the best-fit model, but a large ensemble of possible models, which enables a more thorough understanding of the range of velocity models that can fit the traveltimes, including their covariance and non-uniqueness.

Previous applications of transdimensional Bayesian tomography have primarily been limited to surface wave tomography (e.g. Bodin & Sambridge 2009; Bodin *et al.* 2012a) due in part to the large number of parameters and the computational expense of forward modelling in 3-D. 3-D surface wave models have been constructed using the two-step process—first performing 2-D transdimensional inversions for phase velocity then inverting dispersion for a series of 1-D velocity models (e.g. Galetti *et al.* 2016). Applications with body waves have focused on recovering 2-D layers (Young *et al.* 2013; Tkalčić *et al.* 2015) or small regions using local seismicity (Piana Agostinetti *et al.* 2015). Here, we present an application of THB methodology to teleseismic body waves on a continental scale. In order to viably create an ensemble of 3-D velocity models using large volumes of data available from USArray, we linearize the forward problem and perform calculations on an invariant underlying grid. Transdimensional parameters are then used to efficiently project models onto the grid. In this way, we trade the ability to create models of arbitrarily fine scale for increased computational speed. The resulting method is fast and flexible, and will be easily tailored for a wide variety of 3-D applications.

This article consists of three major parts. In Section 2, we develop a general method for probabilistic body-wave tomography and describe an algorithm for the fast construction of a chain of transdimensional models. In Section 3, we apply the method to teleseismic *P*-wave traveltimes from USArray Transportable Array and other catalogues in order to create an ensemble of models of the mantle beneath North America. Finally, in Section 4, we use the ensemble of models to produce a mean model, PTHB_NA16, investigate uncertainty in mantle structure beneath North America, assess the improvement of velocity estimation due to USArray, and compare the scale and amplitude of velocity variation with a model created using a more standard, Adaptive Grid (AG) method (Li *et al.* 2008).

2 METHOD

We wish to estimate the probability that a particular velocity model explains a set of observed traveltimes. Bayes’ Theorem

states that the posterior probability of a model is proportional to two quantities. First, if the model is correct, how likely is it that we would observe the data? Second, how well does the model agree with any prior knowledge about the values that a model can take?

To estimate the posterior, we adapt the THB surface wave tomography method of Bodin *et al.* (2012a) and sample model space using rj-MCMC. An overview of the algorithm is shown in Fig. 2. At each step in the Markov chain, a new model is proposed by changing one parameter of the velocity model. The proposed model is accepted or rejected based on the change in posterior probability according to the Metropolis–Hastings algorithm (Metropolis *et al.* 1953). At regular intervals, the current model is saved to an ensemble, and after many iterations the frequency with which a model appears in the ensemble will be proportional to the probability that it fits the data.

In order to efficiently sample the model space with rj-MCMC, we reduce the effective dimensionality of the model by using a transdimensional approach that allows the number, shape, and position of parameters to vary based on the demands of the data. For flexibility, we use 3-D Voronoi volumes (Okabe *et al.* 1992), whose shapes and positions are defined by a set of Voronoi nuclei, though transdimensional sampling over tree structures (Hawkins & Sambridge 2015) provides another way forward. Previous tomographic algorithms using transdimensional Voronoi parametrizations performed forward calculations either by ray tracing in each model (Galetti *et al.* 2015; Piana Agostinetti *et al.* 2015) or by finding the intersections between the volumes and precomputed rays (Bodin & Sambridge 2009; Bodin *et al.* 2012a). Continental scale tomographic models (e.g. Burdick *et al.* 2008, 2014) are constructed using millions of seismic traveltimes and typically comprise hundreds of thousands of model parameters, making these approaches computationally impractical. In order to develop a viable method for the body wave problem, we take a two part approach to parametrizing the problem. First, we set up a linearized forward problem with the velocity model and sensitivity matrix defined on an underlying ‘regular grid’. Second, we merge elements of this regular grid according to the nearest Voronoi nucleus, yielding a model and sensitivity matrix based on the Voronoi parametrization.

Although a solution and uncertainty estimate of the linearized tomography problem can be found with matrix inverse methods, rj-MCMC has three major advantages. First, since it does not require model regularization, it more accurately retrieves the amplitude of variations and allows for sharp velocity gradients in the model. Second, it can yield both Gaussian and non-Gaussian posterior distributions on model parameters, allowing us to analyse non-uniqueness in the tomography problem. In contrast, most standard least-squares methods only recover the Gaussian approximation to the true uncertainty in the vicinity of the solution, which can underestimate true uncertainty if multiple minima are present. Third, we need not assume a normal prior on model values, which is implicit in inverse methods using quadratic misfits. In this study we will begin from a uniform prior, weakening the assumption that the model values are distributed about zero.

2.1 Tomographic problem

We set up a linearized forward problem to model body wave traveltimes from seismic slowness. We first determine the ray path between each earthquake source and seismic station through a back-

ground model using ray theory. The data are defined as a set of traveltime residuals, $d_i = T_i - T_i^{\text{obs}}$, $i = 1, M$, that is, the difference between the observed traveltimes and traveltimes calculated in the background model. The Earth is divided up into N regular volumes, and we construct the sensitivity matrix G_{ij} where each entry is the length of the i th ray in the j th grid volume. We can then model each traveltime residual as

$$d_i = \sum_{j=1}^M G_{ij} m_j, \quad (1)$$

where model parameter m_j , $j = 1, N$ is the difference in seismic slowness, (e.g. $s = 1/V_p$, where V_p is seismic velocity) from the background slowness in the j th volume.

We seek models of seismic slowness that explain our traveltime data, so we choose a function to evaluate the misfit between observed and modelled traveltimes. If the residuals are independent and normally distributed about a mean of zero, the misfit will follow the χ^2 function:

$$\chi^2 = \sum_{i=1}^N \left(\frac{T_i - T_i^{\text{obs}}}{\sigma_i} \right)^2, \quad (2)$$

where σ_i is the variance of the data errors, also referred to here as the data variance.

The likelihood function is then given by:

$$p(\{d_i\}|\{m_j\}, I) = C \exp\left(\frac{-\chi^2}{2}\right), \quad (3)$$

where I is relevant background information. The normalization factor $C = (2\pi)^{-N/2} \prod_{i=1}^N \sigma_i^{-1}$ ensures that the likelihood integrates to one. Note that the likelihood thus depends not only upon the traveltime residuals, but also upon the data variance, σ_i . The trade-off between σ_i terms in C and in the exponent will determine the choice of data variance in this method (see Supporting Information Section S1).

Finally, we relate the likelihood to the posterior according to Bayes’ Theorem. The posterior probability, $p(\{m_j\}|\{d_i\}, I)$, for model $\{m_j\}$ and observed data $\{d_i\}$, is commonly written as a proportionality:

$$p(\{m_j\}|\{d_i\}, I) \propto p(\{d_i\}|\{m_j\}, I)p(\{m_j\}|I), \quad (4)$$

where $p(\{m_j\}|I)$ is the prior, or the probability of the model given all that we know about it beforehand. The choice of prior has implications for the character of models in THB methods in poorly sampled regions. I contains assumptions about the physics, parametrization and data corrections. The background assumptions for this study are discussed below.

We reduce the effective number of parameters involved in the forward calculation by using transdimensional sampling. The underlying regular grid parameters and sensitivity matrix are merged according to a set of 3-D Voronoi volumes determined within the inversion (see Fig. 1). For a set of N^{vor} nuclei, we have a model m_k^{vor} , $k = 1, N^{\text{vor}}$ and an $M \times N^{\text{vor}}$ sensitivity matrix, G_{ik}^{vor} . Each underlying grid volume j lies within the Voronoi volume corresponding to the nearest nucleus, and m_j will take the value of m^{vor} within that volume. Similarly, we compute G_{ik}^{vor} by summing the columns of G_{ij} corresponding to elements of the underlying grid within Voronoi volume k , i.e. by finding the total lengths of rays within each Voronoi volume. Thus we can equivalently model the data as

$$d_i = \sum_{k=1}^M G_{ik}^{\text{vor}} m_k^{\text{vor}}. \quad (5)$$

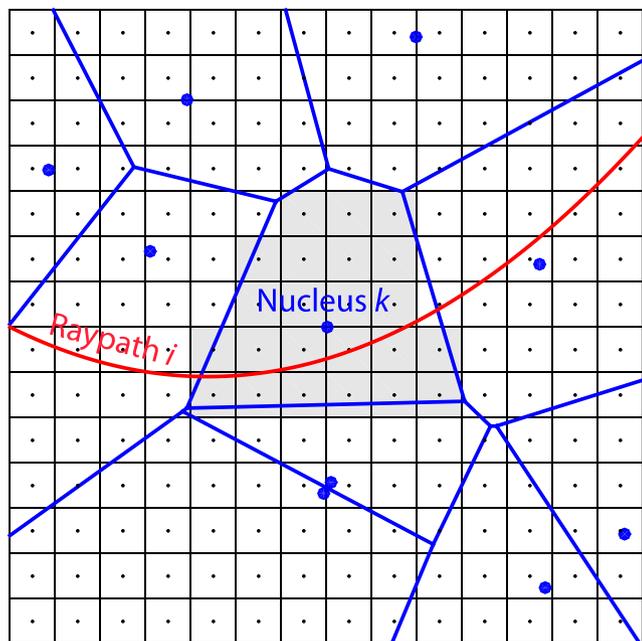


Figure 1. Schematic of model parametrization. Voronoi cells (blue lines) defined by a set of nuclei (blue dots) are projected onto a regular grid (black lines). Regular grid volumes whose centre points (black dots) lie within the cell defined by Nucleus k are coloured grey. An example ray path i is shown in red. G_{ik}^{vor} is then defined by the total length of the ray within the grey region.

This formulation will all for efficient transitions between parametrizations in rj-MCMC.

2.2 THB tomography with rj-MCMC

To find the posterior probability distribution, we use rj-MCMC to explore the model space by varying the model one parameter at a time. The subsequent models are then accepted or rejected based on their affect on the posterior probability according to the Metropolis-Hastings algorithm (Metropolis *et al.* 1953). After an initial ‘burn-in’ period, the frequency with which models appear will be proportional to their probability. By evaluating the ensemble of models thus derived, we estimate the posterior probability and compute the statistical properties such as the model mean, standard deviation, and covariance.

To assess the effects of data and modelling uncertainty on the uncertainty of the retrieved velocity models, we follow a hierarchical approach that estimates the variance of data errors simultaneously with the seismic velocity. The data variance vector is parametrized according to a set of noise hyperparameters that specify the standard deviation associated with the traveltime picks for each subset of data (i.e. different P phases, different data sources). As the velocity model varies, the noise hyperparameters vary in response to the changing misfit, keeping the data variance similar in magnitude to the residuals. As a result, uncertainty in the data and forward-modelling scheme is correctly mapped into the posterior probability. Although in this study we assume that the noise is independent and normally distributed, it is possible to improve estimates of data uncertainty by using more accurate parametrizations of the data covariance (e.g. Dettmer *et al.* 2012; Kolb & Lekić 2014).

Following Bodin *et al.* (2012a), new models are generated by randomly varying one parameter according to normal ‘proposal’

distributions with a standard deviation defined for each of five operations:

- (i) *Update slowness.* Change slowness value in one Voronoi volume.
- (ii) *Birth parameter.* Add new Voronoi nucleus with randomly selected location and slowness value drawn from proposal distribution.
- (iii) *Kill parameter.* Randomly remove one Voronoi nucleus.
- (iv) *Move nucleus.* Move one Voronoi nucleus in a random direction with Cartesian distance drawn from move proposal distribution.
- (v) *Update noise hyperparameter.* Change one noise hyperparameter in order to update the data variance.

Parameter values are constrained to be within the uniform prior. The prior ranges and the proposal distributions used in this study can be found in Supporting Information Table S2.

In order to produce a viable algorithm for the 3-D tomography problem, we seek to minimize the computational cost by updating only the parts of m^{vor} and G^{vor} affected by each model operation. We find that by taking this approach we are able to reduce the compute time by an order of magnitude. The acceptance criteria for new models and efficient algorithms for updating m^{vor} and G^{vor} for each model operation are described in depth in the supplementary material.

3 APPLICATION TO THE NORTH AMERICAN MANTLE

3.1 Data and AG inversion

As a starting point and reference for comparison, we perform a global AG inversion in addition to the THB inversion. We use the data set described by Burdick *et al.* (2014) with two additional years of data from USArray Transportable Array (TA). The updated USArray TA data set includes 500 000 new P -wave onset arrival times hand-picked by the Array Network Facility, available at <http://anf.ucsd.edu/tools/events/>. The new data were observed at 800 stations, primarily in the Eastern US, from 4000 earthquakes occurring between January 2013 and May 2015.

In order to reduce the effects of uneven sampling due to the irregular distribution of global seismicity and seismic stations, summary rays are created for multiple co-located sources and receivers. In both the AG and THB inversions, the data variance term in the misfit is weighted by the square root of the number of data included in each summary ray. For the THB inversion, we divide the summary data into bins by data set and phase, and assign to them different noise hyperparameters in order to investigate their relative data uncertainty.

In addition to the USArray data, we create two roughly equally sized bins from the global data sets. The first comprises data from the EHB catalogue (Engdahl *et al.* 1998), largely from stations in the western hemisphere, and the second mainly comprises EHB data from the eastern hemisphere plus data from the Annual Bulletin of Chinese Earthquakes (see Li *et al.* 2008). Overall, the AG inversion uses ~ 4 million global summary rays compared to ~ 1 million for the THB inversion, while both inversions include ~ 1.3 million USArray summary rays. A breakdown of the data can be found in Supporting Information Table S1.

We perform an AG inversion according to the methodology of Li *et al.* (2008). Rays are traced through the mantle using the 1-D reference model *ak135* Kennett *et al.* (1995). Starting from a regular

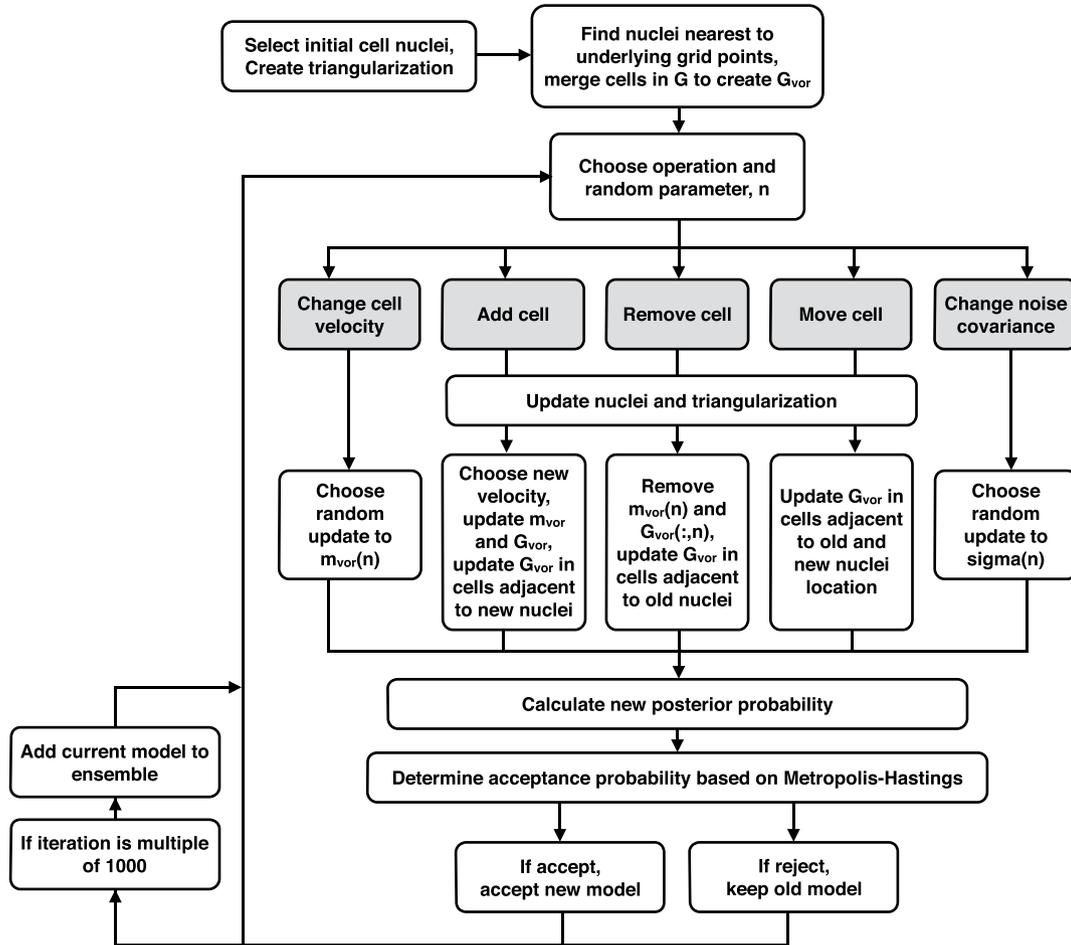


Figure 2. Flow chart of 3-D Transdimensional Hierarchical Bayesian tomography method.

global grid of $0.7^\circ \times 0.7^\circ \times 45$ km, grid volumes with inadequate ray-path density are iteratively merged. We include correction terms for source mislocation. Finally, we regularize the inversion via norm and gradient damping, the parameters for which are determined by L-curve test (Supporting Information Fig. S1). The tomographic problem is solved via LSQR (Paige & Saunders 1982). The resulting model is shown in Fig. 3 given as per cent difference from *ak135*.

3.2 THB inversion

For the THB inversion, we limit our analysis to a region surrounding North America. The underlying grid consists of a nested set of regular grids. The finest inner grid is $0.7^\circ \times 0.7^\circ$ from $25\text{--}50^\circ\text{N}$ and $235\text{--}295^\circ\text{W}$, the middle grid is $1.4^\circ \times 1.4^\circ$ from $20\text{--}55^\circ\text{N}$ and $230\text{--}300^\circ\text{W}$ and outer grid is $2.7^\circ \times 2.7^\circ$ from $12\text{--}60^\circ\text{N}$ and $225\text{--}308^\circ\text{W}$. The depth spacing is 45, 90 and 180 km down to a depth of 1100 km, 2000 km and the core–mantle boundary, respectively.

Because the majority of the ray paths travel outside of the THB model domain, we correct the data for variation outside of the domain based on the AG model. Underestimation of heterogeneity by the damped inversion likely contributes to the data uncertainty inferred by THB. Furthermore, the large number of source coordinates used makes it computationally difficult to include the mislocation terms in the THB inversion, although it is theoretically possible (Myers *et al.* 2007; Piana Agostinetti *et al.* 2015). We in-

stead include a data correction for source mislocation from the AG result.

Within the model domain, we choose an initial set of Voronoi nuclei. In practice, we find that the rj-MCMC algorithm achieves convergence more quickly when starting from an overly complex model and discarding unnecessary parameters, rather than starting from a constant model and adding nuclei. Olugboji *et al.* (2015) show that the results of THB inversions of surface wave dispersion do not depend on the starting model, and we expect the same to hold true for body-wave inversions. We therefore choose as our initial Voronoi nuclei a random selection of one eighth of the adaptive grid centre-points within the domain (~ 3900 nuclei for the model without USArray and ~ 6800 for the model with it). Their respective slownesses are retained as the initial model values. One such starting parametrization is shown in Supporting Information Fig. S2. This choice has the effect of giving our starting models complexity based entirely on ray-path density.

We build an ensemble of models from 100 rj-MCMC chains run in parallel on the DeepThought2 Cluster at the University of Maryland. Each chain is run for 1.5 ± 0.1 million iterations. The chains initially run for a burn-in period (approximately 400 thousand iterations) before the number of dimensions and the noise hyperparameters stabilize and the algorithm begins to search around a mean value or values for slowness within each background grid volume. Following the burn-in period, we save every 1000th model to the ensemble, approximately 100,000 models in total. The mean and 1σ error bar for the model ensemble are displayed in Fig. 4.

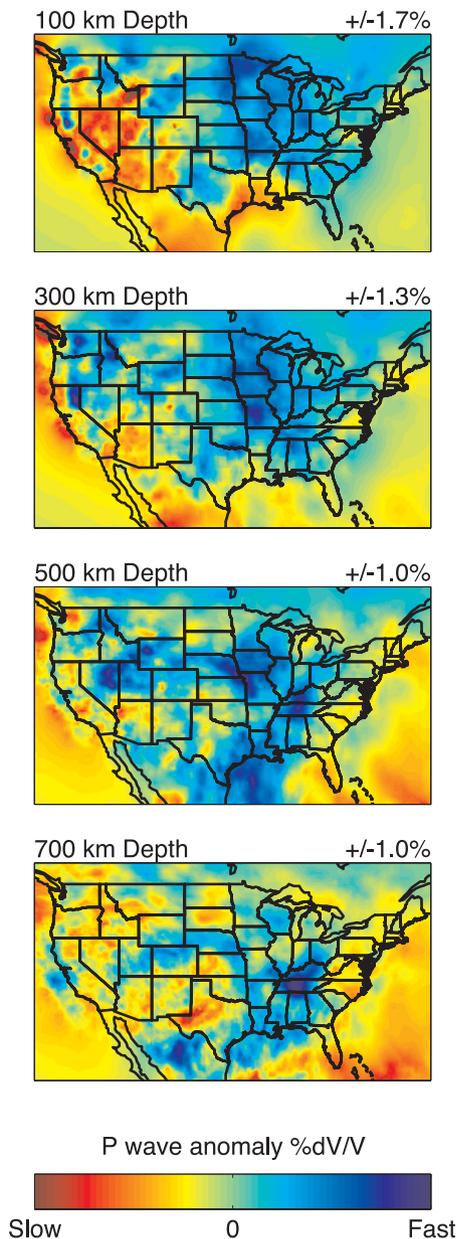


Figure 3. Adaptive Grid model selected depths.

4 RESULTS AND DISCUSSION

4.1 Background assumptions

The model probability distribution presented in this study and the statistical values we derive from it are not universal—that is, they do not apply to all velocity models of North America or those constructed with comparable data sets—but rather are dependent both on the data and the particular background assumptions used in this study. These assumptions include:

- (i) The arrival times of teleseismic P -waves can be adequately modelled by ray theory.
- (ii) The traveltime residuals can be expressed as perturbations around rays traced through a 1D spherical model.
- (iii) The distribution of data error is Gaussian and uncorrelated.
- (iv) The underlying grid is sufficiently fine for expressing the target heterogeneity.

(v) Out-of-domain traveltime delays and source mislocation can be accounted for by a correction from the AG model.

These background assumptions represent common approximations for P -wave traveltime tomography. Nevertheless, by refining the assumptions (i.e. using more accurate forward modelling and finer parametrization) it may be possible to decrease the uncertainty in the model. Additional hierarchical inversions could help isolate the relative effects of individual assumptions. For example, a comparison between posterior distributions for models created using rays traced in 1-D and 3-D background models will illuminate how much the 1-D assumption contributes to uncertainty in seismic slowness. Such investigations will provide the basis for future work.

4.2 Characteristics of the ensemble solution

In spite of the coarse Voronoi parametrization of models in the ensemble, we find that each model yields an equivalent or better fit to the data than the AG solution. This is mainly due to the fact that the AG inversion minimizes model regularization terms in addition to the data residuals. Note that different choices of regularization could lead to improved data fit. The mean properties of the model ensembles further improve the data fit. The misfit in the mean model for each chain is ~ 1.2 per cent lower than the average misfit for individual models, and the ensemble mean outperforms the individual chains by ~ 0.1 per cent.

The hierarchical approach allows us to build an ensemble of values for the data variance, and thereby compare the relative uncertainties associated with data subsets. Fig. 5 (left) shows the posterior distributions for the noise hyperparameters in the ensemble. Each peak shows the range of values the variance had during inversion, indicating how well each data set was fit by models in the ensemble. The mean data variance for the USArray P is significantly smaller than the other P data sets. This is likely because of the consistent instrumentation and traveltime picking, and because the THB model domain encompasses a larger number of summary data from USArray. The pP and Pn data sets are relatively poorly fit in the inversion.

In addition to enabling the analysis of data uncertainty, the THB approach allows us to analyse the number of Voronoi volumes and their spatial distribution. Fig. 5 (right) shows the distribution of the Voronoi cell counts in the model ensemble, which peaks at ~ 2700 for the model without USArray data and ~ 4600 with it. The difference in the number of model parameters obtained with the two data sets can be attributed to the added information contained in the data from USArray.

Because we start the rj-MCMC at an overly complex model, most regions in the model decrease in Voronoi density during the burn-in period; however, this decrease is not equal everywhere. This is because while the positions of the initial nuclei are based solely on ray-path coverage, the final positions are based on a combination of ray-path coverage and structural complexity. Additionally, in the absence of any ray paths, the nucleus density is determined by choices in prior and proposal distribution (Gao *et al.* 2015). To quantify the final parameter distribution and change from the starting model, we compute the mean distance between nuclei. For each underlying grid cell, the left column of Fig. 6 displays the Cartesian distance between the nearest nucleus and its adjacent nuclei, averaged over the final model in each chain. The nuclei have the closest spacing (< 200 km) at shallow depths beneath the Western US, while in the east the distances range from 200 to 350 km. The nuclei typically get further apart with increasing depth

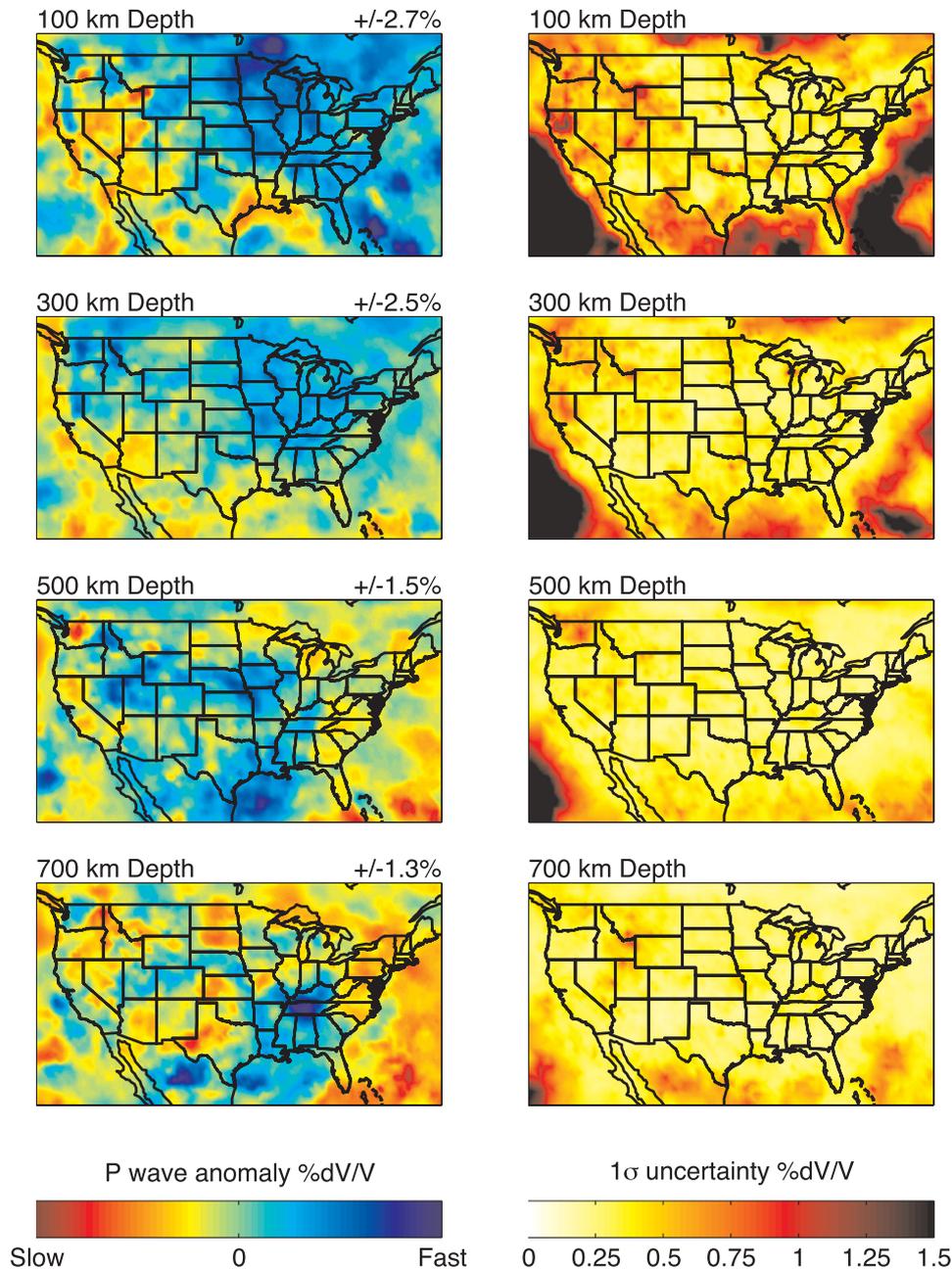


Figure 4. Average model and model uncertainty for selected depths. Areas of high uncertainty within the continental US are generally indicative of multimodal velocity distributions in regions with sharp boundaries.

and the spacing increases dramatically in areas with little or no ray path coverage, such as in the upper mantle beneath the oceans.

The effect of structural scale can be further investigated by comparing the final nuclei spacing to the beginning spacing. The right column of Fig. 6 shows the change in nuclei distance, with a positive change indicating that fewer parameters were necessary to explain the structure in the region. The first notable feature is that the spacing decreased in areas without ray sensitivity. This is controlled by the search algorithm, which determines the optimal number of parameters to efficiently search the model space. At shallow depths, continental regions west of the Rocky Mountains experienced a slightly smaller increase in spacing than regions to the east, indicating that the tectonically active region requires greater

model complexity. The greatest increase in nuclei spacing in the upper mantle occurs beneath the Great Plains and Central Lowlands (~ 40 per cent at 100 km depth), provinces commonly imaged as seismically uniform. The relative nuclei spacing generally increases with increasing depth, corroborating the widely observed decrease in heterogeneity in the mid-to-lower mantle.

4.3 Mean model and model uncertainty

Computing the mean and covariance of the velocity values at each background grid point provides us with a basic view of the model ensemble and its uncertainty. The mean THB model (left-hand column of Fig. 4) images large scale structures consistent with those

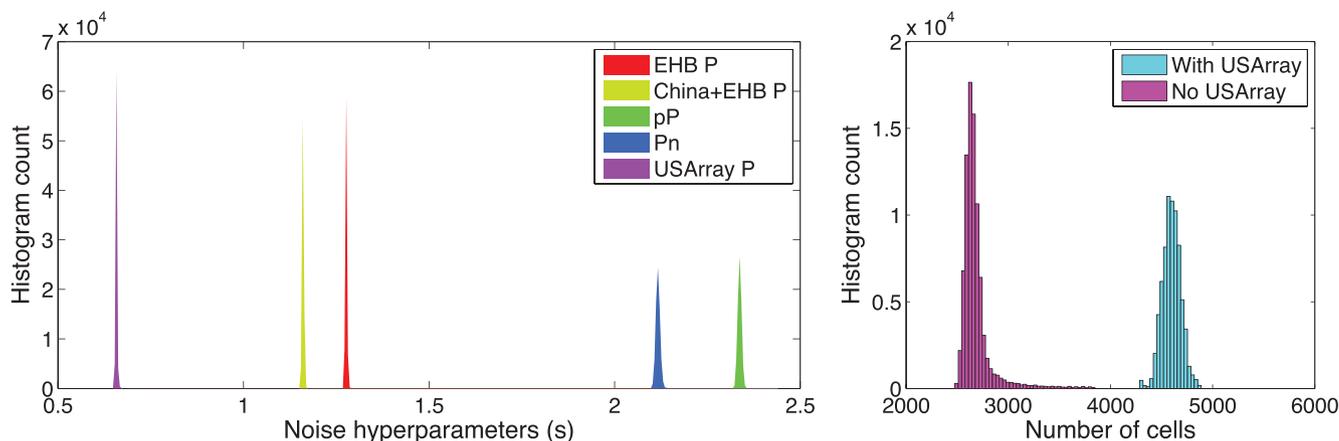


Figure 5. Left: posterior distribution for noise hyperparameters for different data subsets in the model constructed with all data. P picks show the lowest residual data variance. Picks from USArray were better fit than those from the EHB catalogue and Chinese data sets. P_n and pP phases had relatively higher data variance. Right: posterior distribution for number of Voronoi volumes in models with and without the USArray data set.

seen in the AG model (Fig. 3). However, the THB model features larger amplitude variations and slightly less structure at short wavelengths. This is, in part, a feature of the inherent parsimony of THB (Malinverno 2002) which prefers the model with fewer Voronoi nuclei, data residuals being equal. Given the relatively large traveltime residuals remaining in the ‘best fit’ model, a relatively small number of nuclei are necessary to fit the data to within its uncertainty. Improvements to the residuals (e.g. by retracing of rays or using finite-frequency kernels) would yield a model with finer structure (Bodin *et al.* 2012a).

With the addition of two years’ worth of USArray P traveltime picks beyond Burdick *et al.* (2014), both the THB and AG models reveal new features along the eastern margin of the US. We recover a slow velocity structure beneath central Virginia related to Cenozoic volcanism similar to the structures imaged by Chu *et al.* (2013) and Biryol *et al.* (2016), among others. In northern New England, we image a slow structure that coincides with the Great Meteor hotspot track (Villemare *et al.* 2012).

Of note is the anomalous high-amplitude heterogeneity in regions with sparse ray coverage like at shallow depths beneath the oceans. In the AG inversion, the model regularization outweighs any data misfit in these regions, effectively muting these variations. In parts of the model without any sensitivity to the data, the THB algorithm should recover the (uniform) prior (see Fig. 7). In this case, however, posterior velocity distributions are biased by a small number of poorly fit measurements, confirmed by the fact that similar mean heterogeneity is recovered in each model chain. The resulting posteriors for these locations remain relatively uniform, and their large uncertainty prevents over-interpretation. The issue could be avoided through the use of a non-uniform prior—for instance a Gaussian prior centred on zero slowness perturbation. However, for this study we prefer a uniform prior since it avoids the assumption that the slowness perturbations should be distributed around zero.

The right-hand column of Fig. 4 gives the 1σ error bar for the model at the specified depths. Generally, the error bar is the smallest where the ray-path density is adequate—that is, within the continent in the upper mantle and spreading outwards with depth. However, there are areas of the model, particularly in the tectonically active Western US, where there the error bar is puzzlingly large. We find that these large uncertainties typically indicate regions where the model posterior is bimodal or otherwise non-Gaussian due to sharp structural boundaries, as discussed further in Section 4.4.1.

The average uncertainty for regions within the US is summarized in Fig. 8. Overall, the 1σ error bar on the velocity model including USArray can be taken as ~ 0.4 – 0.55 per cent dV_P/V_P for upper-mantle depths beneath the Western US and ~ 0.35 per cent below the east. In the transition zone, the error bar ranges from 0.25 per cent to 0.35 per cent on average, and at depths of 700–1700 km, it is ~ 0.15 – 0.2 per cent, before increasing below due to lack of ray coverage. For most depths and regions, the maximum amplitude of heterogeneity is 3σ – 4σ away from zero. These simplified error bars are inflated by the non-Gaussianity of many of the posterior distributions, as discussed below.

In addition to error bars, the covariance between model values at different locations can also be calculated from the ensemble. Fig. 9 gives examples of model covariance at three geological points of interest. In these figures, blue represent positive covariance, usually indicating that the points tend to fall within the same Voronoi volume and giving a sense of the limits of resolvable structure, while red represents negative covariance, or a trade-off between parameters at the two locations. Figs 9(a) and (b) show the covariance for a point within the inferred slab window beneath the High Lava Plains in Oregon. Resolution testing by Burdick *et al.* (2009) found that the weakening of the slab signal was unlikely due to smearing perpendicular to the slab of low velocity bodies to the west, and the covariance in this study confirms that any smearing is more likely along the strike of the slab. Figs 9(c) and (d) beneath the New Madrid Seismic Zone demonstrate that the low velocities in the lithosphere do not smear downwards past 200 km depth. Figs 9(e) and (f) show that the resolution of our model is unable give depth constraints on the slow anomaly related to the Midcontinental Rift.

4.4 The effect of USArray data

The improvement to multi-scale P -wave velocity models due to the inclusion of USArray data has been a major focus of a number of studies, including Burdick *et al.* (2008, 2014). Previous studies have focused on comparisons of ‘checkerboard’ resolution tests between models created with and without USArray data. These tests are useful inasmuch as they provide a qualitative picture of where the additional data are useful for resolving structure. With the THB method, the benefit of USArray for mantle tomography can now be quantified. The results of the hierarchical inversion for variance

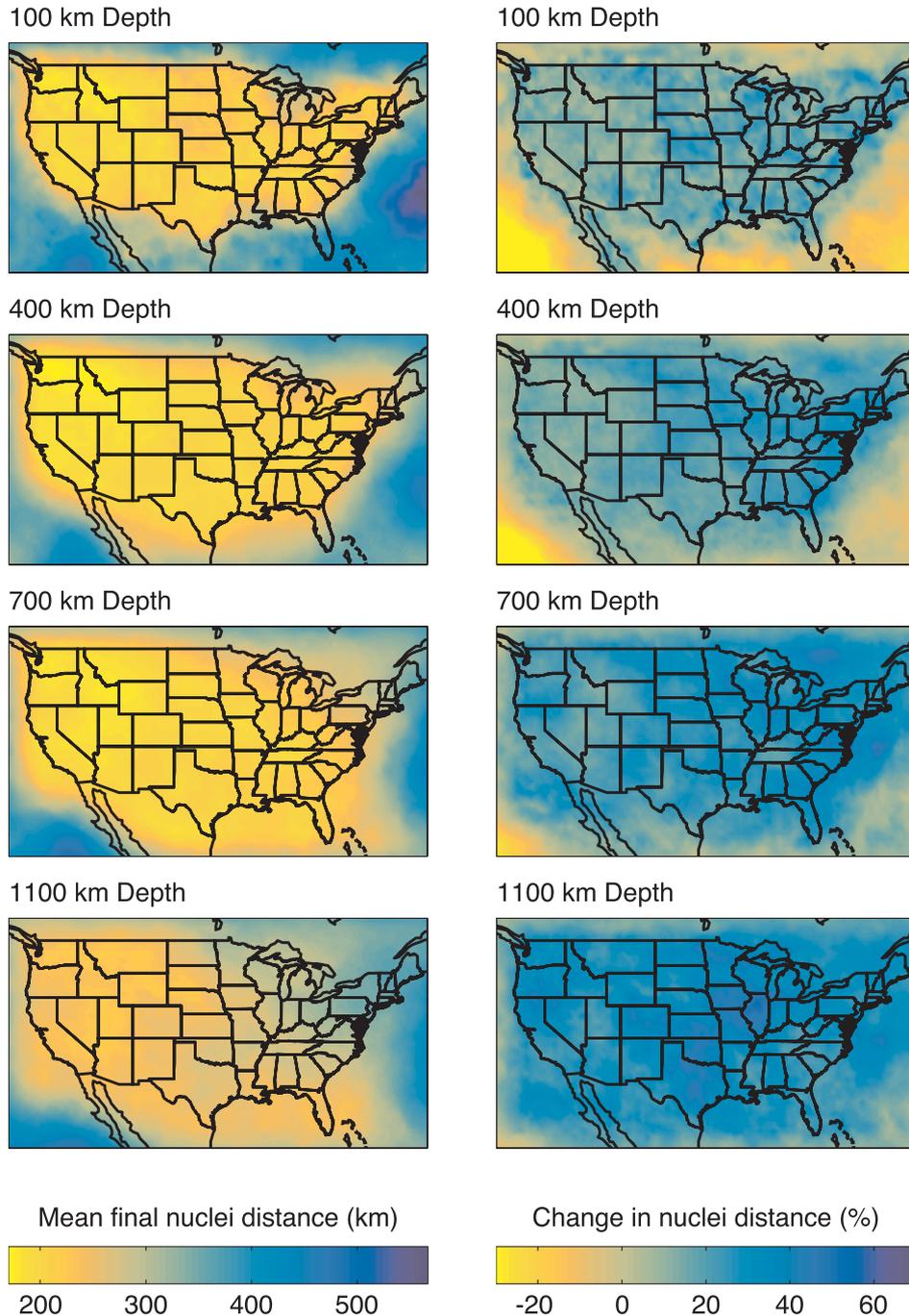


Figure 6. Left: the mean distance between nuclei of adjacent Voronoi volumes for the final model in each chain at selected depths. Right: The per cent difference in nuclei distance for adjacent volumes between the initial model (determined by ray-path density) and final model of each chain. Locations with reduced nuclei distance indicate regions where additional parameters were needed to explain the scale of structure.

of data error (Fig. 5) show that USArray provides a data set with lower data error than other available data sets. Together with the increased ray-path coverage it provides, it is no surprise that the addition of USArray lowers the uncertainty across every region of the Continental US; however, we find that the benefits are not evenly distributed.

To investigate the regional effect of USArray on the model uncertainty, Fig. 8 shows three large provinces for comparison: Central Lowland and Coastal Plain, which were sparsely instrumented prior to USArray, and the Basin and Range, which is well sampled in

the global data sets. Beneath the Basin and Range, the improvement due to USArray is 40–50 per cent for all depths, while in the eastern provinces, the improvement is more substantial from the surface to transition zone (55–70 per cent lower uncertainty) and weaker below the transition zone (35–50 per cent lower). This can be understood in terms of the ray geometries of events recorded at the new stations. Teleseismic *P*-waves arriving in the Eastern US from earthquakes around the Pacific pass beneath the Western US through the lower mantle, so the addition of stations in the east is beneficial at those depths. Meanwhile, those stations add few

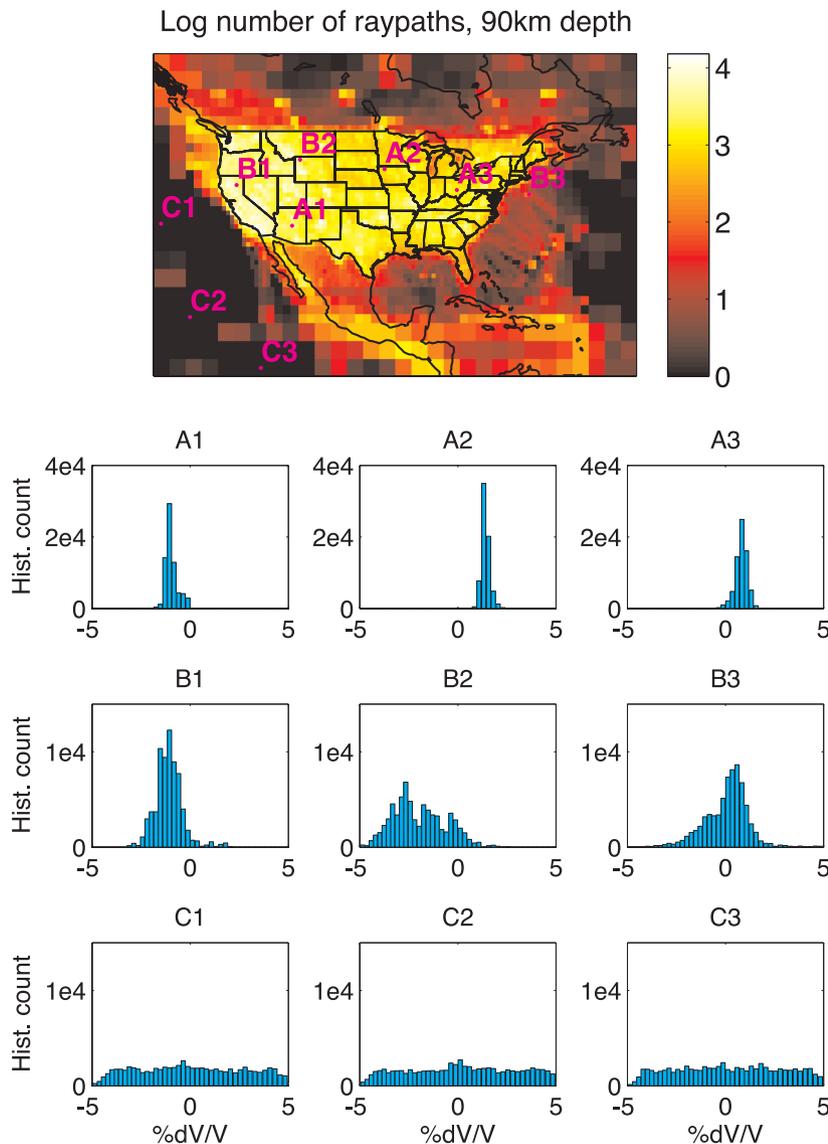


Figure 7. Ray coverage at 90 km depth with selected velocity values distributions. Points A1–A3 are within the continent in the region of good ray coverage, and their distributions are approximately normal. Points C1–C3 are in within underlying cells with low or no sensitivity to the data and their distributions are allowed to trace out the uniform prior. Points B1–B3 are in areas on the margin of coverage. With weak sensitivity to the data, their distributions are have large variance and are skewed and/or bimodal.

additional ray paths in the lower mantle beneath the Eastern US, and the ray paths contributed by new stations in the west makes a relatively small difference compared with existing coverage.

Setting aside the complications of non-Gaussian posteriors, the standard deviations of the distributions within the upper mantle are typically below 0.3 per cent dV_p/V_p for the model with USArray and up to 0.9 per cent dV_p/V_p without. If the velocity variation is attributed entirely to temperature variation according to Cammarano *et al.* (2003) (for 1300 K adiabat), the uncertainty in the USArray-derived velocity propagates into uncertainty of 25–45 °C in the upper 200 km, rising to ~ 80 °C in the transition zone. Without USArray, the uncertainty in temperature exceeds 100 °C throughout, and is in excess of 200 °C in the transition zone beneath the Eastern US. This difference has implications for how well we can resolve hypotheses about buoyancy-driven mantle convection and the longevity of ancient subducted slabs. It is vital to further note that the conversion between velocity perturbation and tem-

perature is subject to its own additional uncertainty and potential trade-offs with compositional effects.

4.4.1 Non-Gaussian model posteriors

Although it is possible to estimate model uncertainty in the least-squares formalism by computing the posterior model covariance matrix (e.g. Tarantola 2005), these estimates only pertain to a multi-dimensional Gaussian distribution fit to the immediate vicinity of the best model. Furthermore, use of model regularization itself affects the estimates of uncertainty. Since we create the model without regularization and allow non-Gaussian posterior distributions, the description of the posterior probability in terms of a mean, standard deviation, and covariance is inadequate in many places. We find that the posterior can be bimodal, multi-modal, or uniform depending on the complexity of the velocity structure and ray coverage. In places without ray sensitivity the algorithm traces out the prior velocity

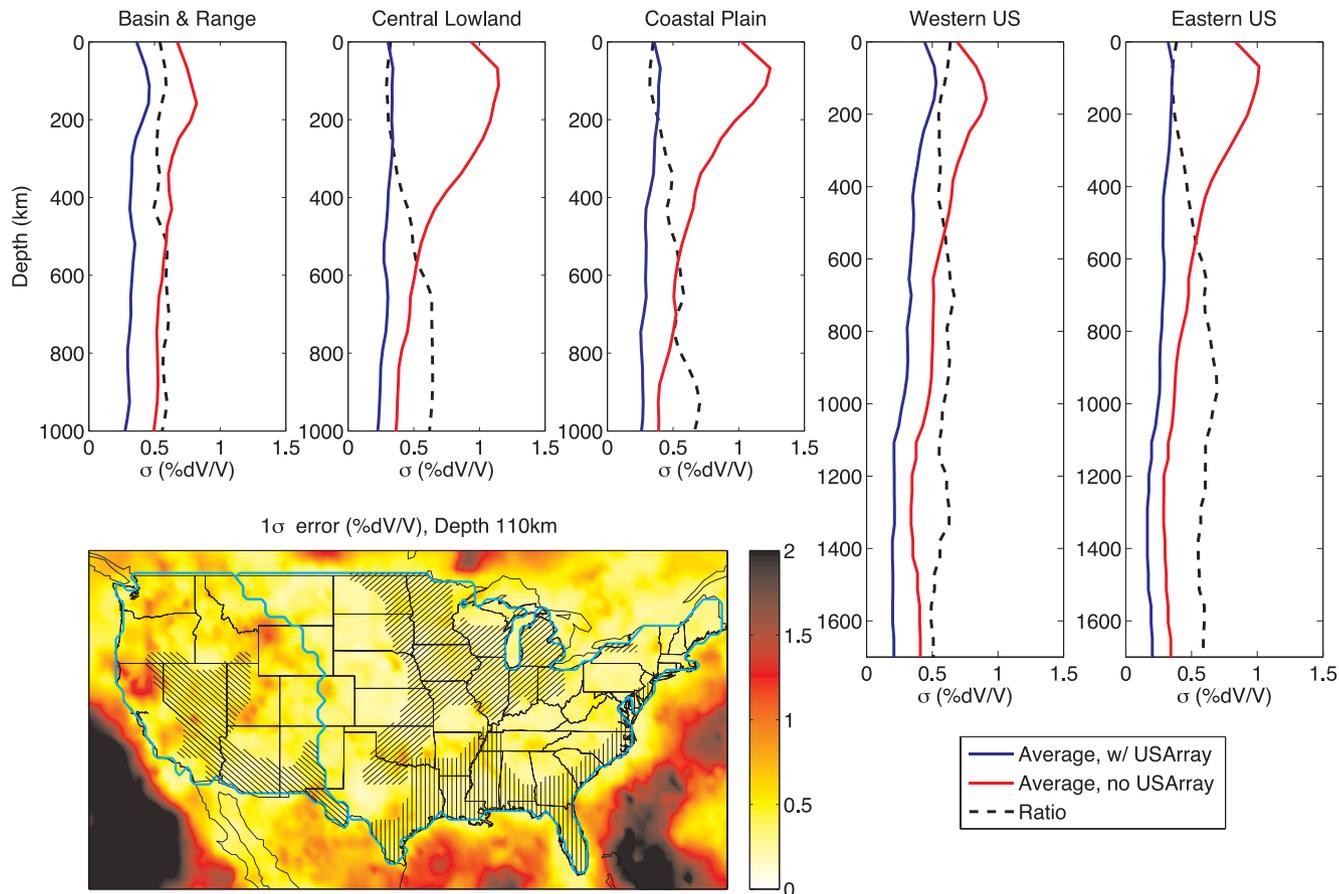


Figure 8. Depth profiles of uncertainty for models with and without USArray data. Black dotted lines represent the ratio of uncertainty with USArray to uncertainty without. Location of selected physiological provinces are indicated on map of 1σ error, with the cyan line running along the Rocky Mts denoting the boundary between Western and Eastern US. In general, the uncertainty in the model is greater at depths in the upper mantle where the absolute strength of the heterogeneity is greater. The model with USArray shows less uncertainty for all regions and depths, but particularly makes improvements in the upper mantle and transition zone beneath the Eastern US and lower-mantle depths beneath the Western US.

range, resulting in a uniform distribution between the minimum and maximum velocity allowed.

At sharp boundaries in the Earth, particularly where the boundary cannot be captured by the underlying grid or its location cannot be uniquely determined from the data coverage, bimodal distributions arise since the data are equally well explained by moving the interface between Voronoi volumes with respect to the true structural boundary. Galetti *et al.* (2015) observe similar multi-modal distributions in transdimensional surface wave tomography and posit that they are due to the non-linear effect of waves bending around low velocity bodies. We do not believe these effects are the cause of multi-modality in this study, as we do not allow the ray paths to vary within the algorithm.

Fig. 10 shows one such location where multi-modality occurs: the Eastern Slab Pile, which is a fast eastward sloping feature extending into the lower mantle that has been variously interpreted as a remnant of the Farallon plate (Pavlis *et al.* 2012) or another slab resulting from intra-oceanic closure (Sigloch & Mihalynuk 2013). The marked locations on the velocity and uncertainty cross-sections correspond to the five posterior probability plots above. At locations 1 (below the slab) and 5 (above the slab) the most likely velocity perturbation is slightly negative, while within the slab (location 3), the most likely value is 0.7 per cent dV_p/V_p . At locations 2 and 4, where the slab meets the mantle, the probability distributions are

broad or bimodal, indicating that at those locations the models are split between the two end-members. On the cross-section, bimodal distributions appear as regions of higher uncertainty, although both of their peaks have roughly the same standard deviations as the single-mode distributions surrounding them. This phenomenon is responsible for much of the high uncertainty in the Western US, where tectonic activity gives rise to many fine-scale structures and sharp gradients in velocity.

We verify that the high uncertainty in these locations is not attributable to measurement error by carrying out a test in which we increase all noise hyperparameters by a multiplicative factor. We find that this increase in data variance leads to a corresponding increase in model variance in well-resolved regions not associated with sharp gradients in structure. However, the model variance is not substantially affected in regions with multi-modal or uniform posterior distributions. We surmise that this is because, in these regions, the majority of posterior uncertainty results from inadequate data sampling or structural complexity.

The behaviour seen in Fig. 10 illustrates that by parametrizing models with Voronoi volumes which allow for discrete jumps in velocity, THB tomography places sharp boundaries between the slab and mantle and maps out how its location is distributed across the ensemble. With further analysis, these distributions have the potential to estimate the likelihood that the fast anomaly extends

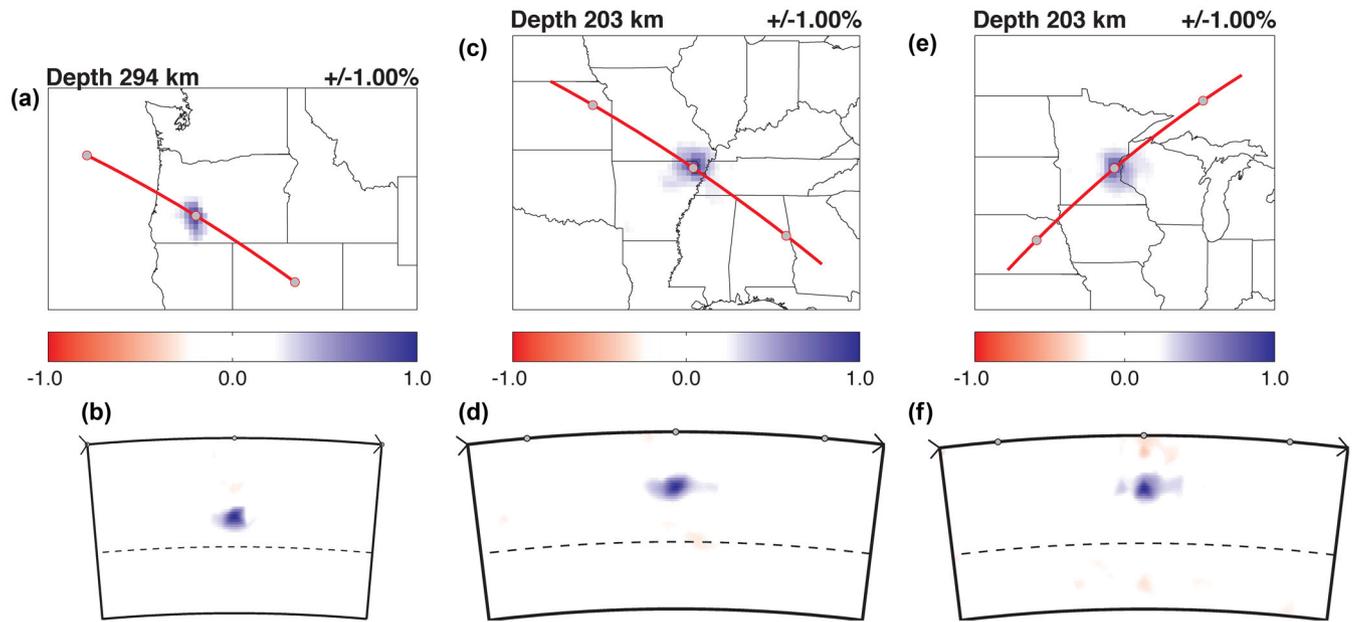


Figure 9. Covariance maps and cross-sections for three different points. Stars indicate the location of points where we calculate the covariance with all other model parameters and red lines indicate location of cross-sections. Blue represents points where there is a high covariance and red represents points where there is a negative covariance with starred location. Left: slab window beneath the High Lava Plains. Middle: New Madrid Seismic Zone at 203 km depth. Right: mid-continental Rift at 203 km depth.

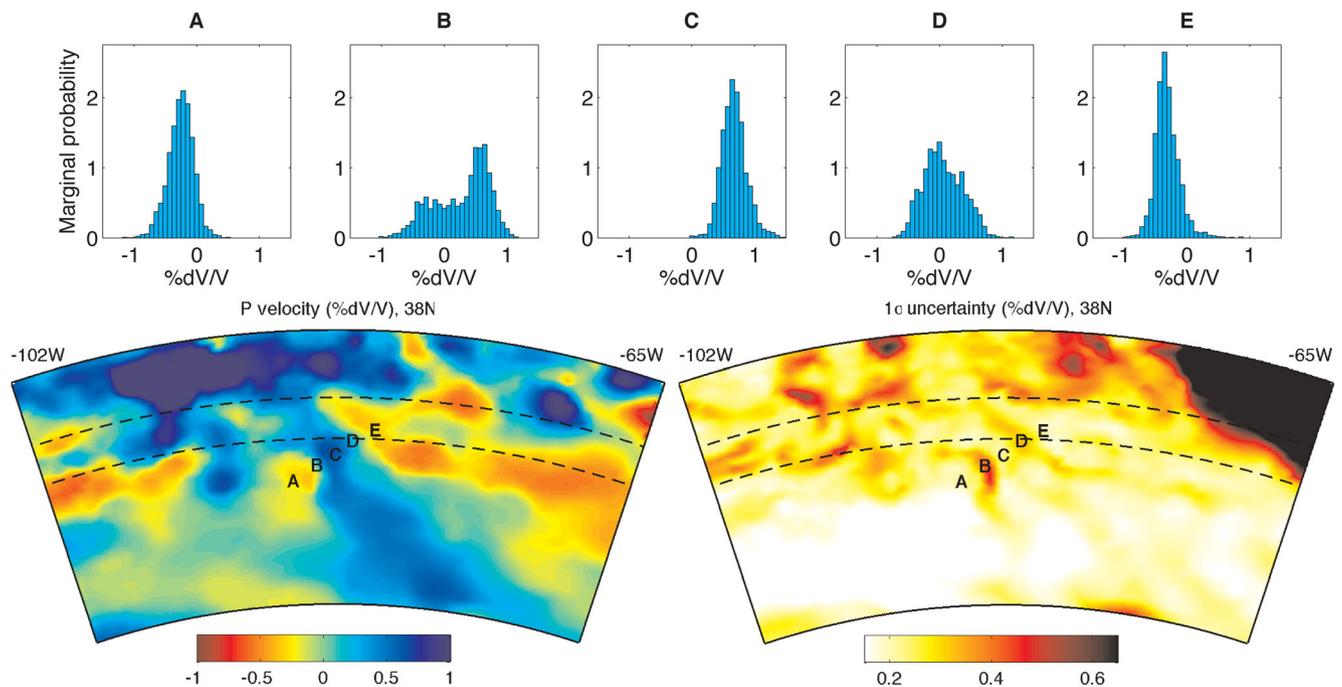


Figure 10. The effect of sharp boundaries on the velocity distributions across the Eastern Slab Pile. Left bottom: cross-section through mean model at 38°N latitude from -102°W to -65°W longitude. Right bottom: corresponding section for 1σ uncertainty. Top: histograms of distributions indicated by letters on cross-sections. The distribution of velocities within the slab has a peak at 0.7 per cent dV_p/V_p , while the outlying areas have a slightly negative peak perturbation. For regular grid volumes at the boundary between the slab and surrounding mantle, some models in the ensemble place points within the slab and some place it without, causing the posterior probability to have broad or bimodal distributions.

to a given point, allowing us to put error bars on slab volume. Estimating the volume of slab material in the lower mantle is key for distinguishing between these two scenarios (Liu 2014), but typical smoothed approaches make it difficult to estimate the location of the boundary between the slab and outlying mantle.

4.5 Amplitude comparison with AG model

In order to accurately relate tomographic models to variations in physical properties like temperature, partial melt, and volatile content, it is vital to estimate the absolute strength of velocity

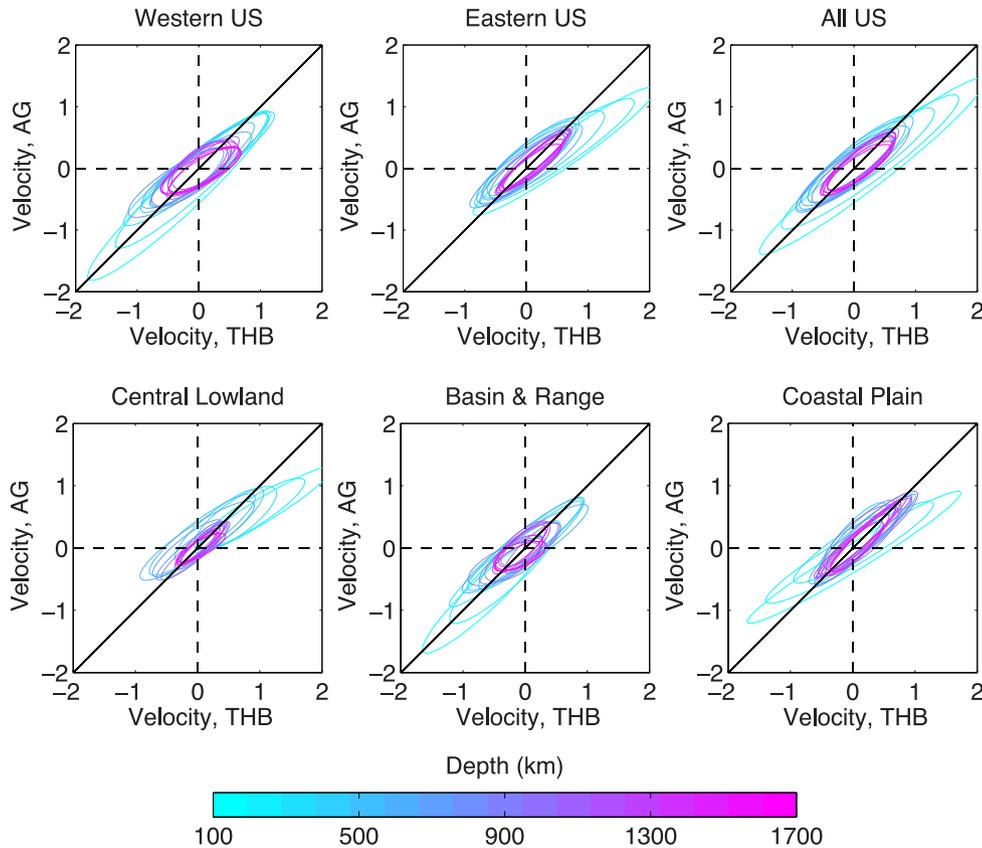


Figure 11. Confidence ellipses for correlation in velocity value (per cent dV_p/V_p) between adaptive grid and probabilistic models at different depths. Regions are indicated in Fig. 8. Ellipses represent 95 per cent confidence region (2.45σ). The tilt of the ellipses shows that velocity variations in the probabilistic model have higher amplitude in general, particularly at shallow depths and in the Eastern US, where there is a factor of ≈ 1.5 between them. Note that the tilt of the ellipses is dependent on regularization choices in the adaptive grid inversion.

heterogeneity. Sampling the model space using the *rj*-MCMC algorithm allows us to create models where the amplitude of the heterogeneity is not affected by arbitrary choices of model damping and smoothing. Fig. 11 shows a comparison between the velocity values from the AG inversion and the mean THB model. Each ellipse represents the 95 per cent confidence region determined by the covariance between the two models, with cyan and magenta representing shallow and deep depths, respectively. The width of the ellipses (minor axes) shows the extent to which the models are linearly related and deviation of the slope from the diagonal shows the scaling factor between them.

Over broad regions within the continent, there is a strong correlation between the two models, but with scaling in amplitude. The THB model consistently has higher overall amplitudes than the AG model, revealing heterogeneity 10–80 per cent greater at shallow depths beneath some eastern provinces. This difference in amplitude recovery is comparable to that attributed to the use of finite-frequency kernels (~ 30 – 50 per cent Montelli *et al.* 2004) and accurate forward modelling techniques (~ 50 per cent Lekic & Romanowicz 2011).

At upper-mantle depths, the amplitudes are more consistent between the models in the Western US compared to the east. We attribute this consistency to the fact that regularization in the AG inversion has a lesser effect when data coverage is dense. In the Eastern US, and other regions where there are fewer data, regularization plays a larger role and the amplitudes are underestimated. In the transition zone and deeper depths where ray-path density is higher, amplitudes between the AG and THB models become once

again more consistent. Over some large regions, such as the Coastal Plain above 200 km, the THB model has higher velocities overall compared to the AG model. This is because in the absence of damping, the THB model values are not constrained to be close to $ak135$, allowing us to recover regions where the velocity varies strongly from any 1-D global average.

The higher amplitudes recovered by the THB inversion indicate larger variations in temperature than would be inferred from the AG inversion. As a worst case scenario, the AG model could cause the temperature difference between the subcratonic and average mantle to be underestimated by ~ 100 °C. The THB model gives evidence of a wider degree of variation in temperature, and thus in buoyancy and viscosity for mantle structures, the implications of which we hope will motivate study by geodynamicists and geochemists.

5 CONCLUSIONS

We present a first implementation of the THB method for a 3-D, continental scale tomographic problem. The implementation is made computationally efficient by considering an underlying ‘base’ parametrization and seeking efficient ways to merge these base parameters according to a set of Voronoi volumes. Although we apply the method to teleseismic P -wave tomography of North America, the approach is flexible enough to enable straightforward extension to other 2-D and 3-D problems with large, relatively sparse

sensitivity matrices. By analysing the properties of the tomographic models in the ensemble solution, we:

(i) Quantify the uncertainty in seismic velocity beneath North America, including non-uniqueness and covariance between parameters. Overall, velocities have a 1σ error bar of 0.25 per cent dV_P/V_P , decreasing towards the Eastern US and with greater depth. The posterior probability frequently features multi-modal and non-Gaussian distributions, which renders standard analyses (e.g. mean and standard deviation) insufficient. Bimodal distributions give useful insight into sharp structural gradients and allow us to quantify the extent and geometry of tectonic domains and slab fragments.

(ii) More accurately recover the absolute amplitudes and sharp velocity gradients of V_P variations in the absence of model regularization. Overall, the THB mean model has amplitudes 10–80 per cent higher than the comparable AG model, introducing up to $\sim 100^\circ\text{C}$ error in inferences of relative temperature variations.

(iii) Quantify the relative variance of traveltimes from different data subsets. Our results show that traveltimes from USArray, picked at the Array Network Facility, have the lower variance than traveltimes from global catalogues. This is due in part to consistent instrumentation and traveltime picking methodology and to the geographical boundaries of the model region.

(iv) Quantify the decrease in uncertainty due to the addition of traveltimes from USArray. The model with USArray has smaller error bars across all regions and depths down to the lower mantle. However, the improvement varies between regions, with the greatest improvement coming above 700 km beneath the Eastern US. Due to ray coverage prior to EarthScope, data from USArray have more effect in the mid-mantle beneath the Western US than the Eastern US. Little significant improvement in V_P retrieval is seen at depths deeper than 2000 km.

ACKNOWLEDGEMENTS

Thanks to Frank Vernon, Tolulope Olugboji, Pritwiraj Moulik, Lauren Waszek and Chao Gao for helpful discussions. SB acknowledges support from an NSF Postdoctoral Fellowship EAR-1349771. VL acknowledges support from the National Science Foundation (CA-REER grant EAR1352214) and from the Packard Fellowship for Science and Technology.

REFERENCES

- Agostinetti, N.P. & Malinverno, A., 2010. Receiver function inversion by trans-dimensional Monte Carlo sampling, *Geophys. J. Int.*, **181**(2), 858–872.
- Biryol, C.B., Wagner, L.S., Fischer, K.M. & Hawman, R.B., 2016. Relationship between observed upper mantle structures and recent tectonic activity across the Southeastern United States, *J. geophys. Res.*, **121**, 3393–3414.
- 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., 2012a. Transdimensional tomography with unknown data noise, *Geophys. J. Int.*, **189**(3), 1536–1556.
- Bodin, T., Sambridge, M., Tkalčić, H., Arroucau, P., Gallagher, K. & Rawlinson, N., 2012b. Transdimensional inversion of receiver functions and surface wave dispersion, *J. geophys. Res.*, **117**, doi:10.1029/2011JB008560.
- Burdick, S. *et al.*, 2008. Upper mantle heterogeneity beneath North America from travel time tomography with global and USArray Transportable Array data, *Seismol. Res. Lett.*, **79**, 384–390.
- Burdick, S. *et al.*, 2009. Model update December 2008: upper mantle heterogeneity beneath North America from travel time tomography with global and USArray Transportable Array data., *Seismol. Res. Lett.*, **80**, 638–645.
- Burdick, S. *et al.*, 2014. Model update January 2013: upper mantle heterogeneity beneath North America from travel-time tomography with global and USArray Transportable Array data, *Seism. Res. Lett.*, **85**(1), 77–81.
- Cammarano, F., Goes, S., Vacher, P. & Giardini, D., 2003. Inferring upper-mantle temperatures from seismic velocities, *Phys. Earth planet. Inter.*, **138**(3–4), 197–222.
- Chen, C., Zhao, D. & Wu, S., 2014. Crust and upper mantle structure of the New Madrid Seismic Zone: insight into intraplate earthquakes, *Phys. Earth planet. Inter.*, **230**, 1–14.
- Chu, R., Leng, W., Helmberger, D.V. & Gurnis, M., 2013. Hidden hotspot track beneath the eastern United States, *Nat. Geosci.*, **6**(11), 963–966.
- Deschamps, F. & Trampert, J., 2003. Mantle tomography and its relation to temperature and composition, *Phys. Earth Planet. Inter.*, **140**, 277–291.
- Dettmer, J., Dosso, S.E. & Holland, C.W., 2010. Trans-dimensional geoaoustic inversion, *J. acoust. Soc. Am.*, **128**(6), 3393–3405.
- Dettmer, J., Molnar, S., Steininger, G., Dosso, S.E. & Cassidy, J.F., 2012. Trans-dimensional inversion of microtremor array dispersion data with hierarchical autoregressive error models, *Geophys. J. Int.*, **188**(2), 719–734.
- Engdahl, E.R., van der Hilst, R. & Buland, R., 1998. Global teleseismic earthquake relocation with improved travel times and procedures for depth determination, *Bull. seism. Soc. Am.*, **88**(3), 722–743.
- Faul, U.H. & Jackson, I., 2005. The seismological signature of temperature and grain size variations in the upper mantle, *Earth planet. Sci. Lett.*, **234**(1–2), 119–134.
- Galetti, E., Curtis, A., Meles, G.A. & Baptie, B., 2015. Uncertainty loops in travel-time tomography from nonlinear wave physics, *Phys. Rev. Lett.*, **114**, doi:10.1103/PhysRevLett.114.148501.
- Galetti, E., Curtis, A., Baptie, B., Jenkins, D. & Nicolson, H., 2016. Transdimensional Love-wave tomography of the British Isles and shear-velocity structure of the East Irish Sea Basin from ambient-noise interferometry, *Geophys. J. Int.*, **208**(1), 36–58.
- Gao, C., Lekic, V. & Olugboji, T., 2015. Constraining anisotropy in the US continental lithosphere using a joint inversion of receiver function and ambient noise data, in *AGU Fall Meeting*, no. S14A-05, San Francisco, CA.
- Green, P.J., 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika*, **82**(4), 711–732.
- Hammond, W.C. & Humphreys, E.D., 2000. Upper mantle seismic wave velocity: Effects of realistic partial melt geometries, *J. geophys. Res.*, **105**(B5), 10 975–10 986.
- Hawkins, R. & Sambridge, M., 2015. Geophysical imaging using trans-dimensional trees, *Geophys. J. Int.*, **203**(2), 972–1000.
- Huang, H.-H., Lin, F.-C., Schmandt, B., Farrell, J., Smith, R.B. & Tsai, V.C., 2015. The Yellowstone magmatic system from the mantle plume to the upper crust, *Science*, **348**(6236), 773–776.
- Jacobsen, S.D. & Smyth, J.R., 2013. *Effect of Water on the Sound Velocities of Ringwoodite in the Transition Zone*, vol. 168 of *Geophysical Monograph Series*, pp. 131–145, American Geophysical Union.
- Karato, S., Jung, H., Katayama, I. & Skemer, P., 2008. Geodynamic significance of seismic anisotropy of the upper mantle: new insights from laboratory studies, *Annu. Rev. Earth Planet. Sci.*, **36**, 59–95.
- Kennett, B.L.N., Engdahl, E.R. & Buland, R., 1995. Constraints on seismic velocities in the Earth from travel-times, *Geophys. J. Int.*, **122**(1), 108–124.
- Kolb, J.M. & Lekic, V., 2014. Receiver function deconvolution using trans-dimensional hierarchical Bayesian inference, *Geophys. J. Int.*, **197**(3), 1719–1735.
- Lekic, V. & Romanowicz, B., 2011. Tectonic regionalization without a priori information: a cluster analysis of upper mantle tomography, *Earth planet. Sci. Lett.*, **308**(1), 151–160.
- Leveque, J.-J., Rivera, L. & Wittlinger, G., 1993. On the use of the checkerboard test to assess the resolution of tomographic inversions, *Geophys. J. Int.*, **115**(1), 313–318.
- Li, C., van der Hilst, R.D., Engdahl, E.R. & Burdick, S., 2008. A new global model for P wave speed variations in Earth's mantle, *Geochem. Geophys. Geosyst.*, **9**, Q05018, doi:10.1029/2007GC001806.

- Liu, L., 2014. Constraining Cretaceous subduction polarity in eastern Pacific from seismic tomography and geodynamic modeling, *Geophys. Res. Lett.*, **41**(22), 8029–8036.
- Lou, X. & van der Lee, S., 2014. Observed and predicted North American teleseismic delay times, *Earth planet. Sci. Lett.*, **402**, 6–15.
- Malinverno, A., 2002. Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem, *Geophys. J. Int.*, **151**(3), 675–688.
- Malinverno, A. & Briggs, V.A., 2004. Expanded uncertainty quantification in inverse problems: Hierarchical Bayes and empirical Bayes, *Geophysics*, **69**(4), 1005–1016.
- Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. & Teller, E., 1953. Equation of state calculations by fast computing machines, *J. Chem. Phys.*, **21**(6), 1087–1092.
- Montelli, R., Nolet, G., Masters, G., Dahlen, F. & Hung, S.-H., 2004. Global P and PP traveltimes tomography: rays versus waves, *Geophys. J. Int.*, **158**(2), 637–654.
- Myers, S.C., Johannesson, G. & Hanley, W., 2007. A Bayesian hierarchical method for multiple-event seismic location, *Geophys. J.*, **171**(3), 1049–1063.
- Okabe, A., Boots, B. & Sugihara, K., 1992. *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*, John Wiley & Sons Inc.
- Olugboji, T., Lekic, V., Gao, C. & McDonough, W., 2015. Evaluating models of the US continental crust using ambient noise datasets: a transdimensional approach, in *AGU Fall Meeting*, no. T11A-2860, San Francisco, CA.
- Paige, C.C. & Saunders, M.A., 1982. LSQR: An algorithm for sparse linear equations and sparse least squares, *ACM Trans. Math. Softw.*, **8**, 43–71.
- Pavlis, G.L., Sigloch, K., Burdick, S., Fouch, M.J. & Vernon, F.L., 2012. Unraveling the geometry of the Farallon plate: Synthesis of three-dimensional imaging results from USArray, *Tectonophysics*, **532**, 82–102.
- Piana Agostinetti, N., Giacomuzzi, G. & Malinverno, A., 2015. Local three-dimensional earthquake tomography by trans-dimensional Monte Carlo sampling, *Geophys. J. Int.*, **201**(3), 1598–1617.
- Roth, J.B., Fouch, M.J., James, D.E. & Carlson, R.W., 2008. Three-dimensional seismic velocity structure of the northwestern United States, *Geophys. Res. Lett.*, **35**, L15304, doi:10.1029/2008GL034669.
- Sambridge, M., Gallagher, K., Jackson, A. & Rickwood, P., 2006. Trans-dimensional inverse problems, model comparison and the evidence, *Geophys. J. Int.*, **167**(2), 528–542.
- Sambridge, M., Bodin, T., Gallagher, K. & Tkalčić, H., 2013. Transdimensional inference in the geosciences, *Phil. Trans. R. Soc. A*, **371**(1984), doi:10.1098/rsta.2011.0547.
- Schaeffer, A.J. & Lebedev, S., 2014. Imaging the North American continent using waveform inversion of global and USArray data, *Earth planet. Sci. Lett.*, **402**, 26–41.
- Schmandt, B. & Lin, F.-C., 2014. P and S wave tomography of the mantle beneath the United States, *Geophys. Res. Lett.*, **41**(18), 6342–6349.
- Schmandt, B., Dueker, K., Humphreys, E. & Hansen, S., 2012. Hot mantle upwelling across the 660 beneath Yellowstone, *Earth planet. Sci. Lett.*, **331**, 224–236.
- Sigloch, K. & Mihalynuk, M.G., 2013. Intra-oceanic subduction shaped the assembly of Cordilleran North America, *Nature*, **496**(7443), 50–56.
- Sigloch, K., McQuarrie, N. & Nolet, G., 2008. Two-stage subduction history under North America inferred from multiple-frequency tomography, *Nat. Geosci.*, **1**(7), 458–462.
- Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM.
- Tian, Y. & Zhao, D., 2012. P-wave tomography of the western United States: Insight into the Yellowstone hotspot and the Juan de Fuca slab, *Phys. Earth planet. Inter.*, **200**, 72–84.
- Tkalčić, H., Young, M., Muir, J.B., Davies, D.R. & Mattesini, M., 2015. Strong, multi-scale heterogeneity in Earth's lowermost mantle, *Scientific Reports*, **5**, doi:10.1038/srep18416.
- Trampert, J.D. & F.R.J. & Yuen, D., 2004. Probabilistic tomography maps chemical heterogeneities throughout the mantle, *Science*, **306**, 853–856.
- Villemaire, M., Darbyshire, F.A. & Bastow, I.D., 2012. P-wave tomography of eastern North America: evidence for mantle evolution from Archean to Phanerozoic, and modification during subsequent hot spot tectonism, *J. geophys. Res.*, **117**(B12), doi:10.1029/2012JB009639.
- 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 [GJIRAS](https://doi.org/10.1002/gjras) online.

Table S1. Number of summary rays used in inversions. The entire set of summary rays was used for the global adaptive grid inversion, while the THB inversion was limited to rays within the domain.

Table S2. Values for proposal distribution and prior range used in this study.

Figure S1. Results of the L-curve test used to determine damping parameters. L represents the regularization operator. Colours give the regularization parameter for smoothing in the radial direction. From this test, we choose a radial smoothing parameter of 1000. Lateral smoothing scales as twice this value.

Figure S2. Example starting model for THB tomography inversion. The starting models consist of randomized Voronoi volumes with velocity values derived from a global adaptive grid projected onto a base grid. The base grid is nested set of regular grids with widths of 0.7°, 1.4°, and 2.8° in latitude and longitude and 45 km, 90 km, and 180 km in depth.

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.