

675 **Supplementary materials**

676 **Proposal distributions and acceptance probability**

677 In this section we summarize how new models are generated and accepted within the rj-
 678 MCMC algorithm. For the operations involving updates to the slowness model, we follow the
 679 developments of Bodin & Sambridge (2009) with minor alterations for 3D geometry. For the
 680 hierarchical updates, we follow Bodin et al. (2012a).

681 At each step in the chain, we create a new model by updating one randomly selected
 682 parameter (or hyperparameter) using one of five operations with the following proportions:
 683 Change slowness (0.5), move nucleus (0.15), birth parameter (0.15), kill parameter (0.15),
 684 update hierarchical parameter (0.05). The new value for the selected parameter is chosen
 685 from the proposal distribution, a normal distribution around the current value with a standard
 686 deviation given for each operation.

New models are accepted or rejected based on the Metropolis-Hastings algorithm (Metropo-
 lis et al. 1953). The probability that a change to the model be accepted, $\alpha(m'|m)$, is given by
 the ratio between the posterior probabilities of the new model m' and the old model m :

$$\alpha(m'|m) = \min \left[1, \frac{p(m')p(d|m')}{p(m)p(d|m)} \right] \quad (.1)$$

687 Since we use a uniform prior $p(m')$ equals $p(m)$ if the new model value is within the prior
 688 range and is identically zero otherwise. The acceptance probability for a new model is therefore
 689 based on the ratio of the likelihoods, which are determined by the misfit function $\chi^2(m)$
 690 (dependent on both the travelttime residuals and the data covariance), and the dimensionality
 691 of the models. For the birth and death operations, which change the number of dimensions
 692 in the model, the acceptance probability includes a parsimonious term derived from Bayes'
 693 Theorem (Bodin & Sambridge 2009). The effect of this term is to penalize the addition of
 694 parameters whose existence is not warranted by the data.

695 (i) *Change slowness.* One Voronoi nucleus is randomly chosen and its slowness, s_i , is per-
 696 turbed by a value drawn from a normal distribution with a standard deviation of $\sigma_{\delta s}$.
 697 The new slowness value is then given by $s'_i = s_i + N(0, \sigma_{\delta s})$, where $N(0, \sigma)$ indicates a
 698 normal distribution with a mean of 0 and standard deviation of σ .

The probability that the change of state between the original model m and model m'
 with s'_i is accepted is:

$$\alpha_{\delta s}(m'|m) = \begin{cases} \min \left[1, \exp \left(-\frac{\chi^2(m') - \chi^2(m)}{2} \right) \right] & \text{if } s'_i \in J \\ 0 & \text{otherwise,} \end{cases} \quad (.2)$$

where J is the prior range for slowness.

- (ii) *Birth parameter.* A new Voronoi nucleus (and, therefore, volume) is created in a location \mathbf{x}_{N+1} , drawn from a uniform distribution within the model domain. The slowness within the volume, s'_{N+1} , is drawn from a normal distribution with a standard deviation of σ_{bd} around the current slowness, s_i at the nucleus location (i.e. the slowness value of the nearest neighbor nucleus, i). The slowness within the new volume is $s'_{N+1} = s_i + N(0, \sigma_{bd})$.

The probability that model m' containing the new nucleus $N + 1$ will be accepted is given by:

$$\alpha_b(m'|m) = \begin{cases} \min \left[1, \frac{\sigma_{bd}\sqrt{2\pi}}{\Delta s} \exp \left(\frac{(s'_{N+1} - s_i)^2}{2\sigma_{bd}^2} - \frac{\chi^2(m') - \chi^2(m)}{2} \right) \right] & \text{if } (N + 1) \in I \text{ and } s'_{N+1} \in J \\ 0 & \text{otherwise,} \end{cases} \quad (.3)$$

where Δs is the total range of the slowness prior and I is the prior range for the number of nuclei. The additional terms come from the change in likelihood due to the addition of a new parameter (see Bodin & Sambridge 2009), and have the affect of penalizing the creation of a new volume with slowness close to the existing value.

- (iii) *Kill parameter.* A Voronoi nucleus, i , is randomly selected and removed. The space occupied by the removed volume is then split between adjacent volumes. For the purpose of calculating the acceptance probability, the slowness of the removed nucleus, s_i , is then replaced with the slowness, s'_j , of its nearest neighbor, j .

The probability of accepting model m' without nucleus i is given by:

$$\alpha_d(m'|m) = \begin{cases} \min \left[1, \frac{\Delta s}{\sigma_{bd}\sqrt{2\pi}} \exp \left(-\frac{(s'_j - s_i)^2}{2\sigma_{bd}^2} - \frac{\chi^2(m') - \chi^2(m)}{2} \right) \right] & \text{if } (N - 1) \in I \\ 0 & \text{otherwise.} \end{cases} \quad (.4)$$

The addition terms containing Δs and σ_{bd}^2 are the inverse of those for the *birth parameter* operation. They have the affect of increasing the probability of accepting a model with fewer dimensions even if it slightly increases the misfit.

- (iv) *Move nucleus.* One Voronoi nucleus located at \mathbf{x}_i is chosen at random and moved in a random Cartesian direction by a distance drawn from a three dimensional normal proposal distribution with a standard deviation of σ_x . The new location for the nucleus is then given by $\mathbf{x}'_i = \mathbf{x}_i + \mathbf{N}(0, \sigma_x)$. We choose the most straightforward implementation of the proposal distribution for moving a nucleus, but other options exist. For instance, including separate proposal distributions for radial movement and lateral movement

within the Earth would yield an alternate sampling scheme. The efficiency of such a scheme will be explored in future work.

The probability of accepting the new model m' is

$$\alpha_x(m'|m) = \begin{cases} \min \left[1, \exp \left(-\frac{\chi^2(m') - \chi^2(m)}{2} \right) \right] & \text{if } \mathbf{x}'_i \in K \\ 0 & \text{otherwise,} \end{cases} \quad (.5)$$

where K is the model domain.

(v) *Update noise hyperparameter.* One noise hyperparameter, i , which corresponds to a subset of the data, is chosen at random and is perturbed by a value drawn from a proposal distribution with a standard deviation of σ_h . The new value for the noise hyperparameter is then given by $h'_i = h_i + N(0, \sigma_h)$.

Due to the normalization factor that ensures the likelihood function integrates to one, there are data covariance terms both inside and outside of the exponent. In Equation 3, the normalization factor C is

$$C = \frac{1}{\prod_{j=1}^M \sqrt{2\pi}\sigma_j}, \quad (.6)$$

for M data, where σ is parametrized by the noise hyperparameters, $\{h\}$. Thus, when calculating the acceptance probability under a change in one noise hyperparameter, the normalization factors no longer cancel out completely. Terms not dependent on the selected h_i continue to cancel out, and the ratio of the normalization factors is

$$\frac{k'}{k} = \frac{\prod_{i=1}^{M_h} h_i}{\prod_{i=1}^{M_h} h'_i}, \quad (.7)$$

where M_h is the length of the data covariance vector parametrized by h_i (e.g. the number of data in the i^{th} subset of data). For numerical stability, it is useful to bring the multiplicative terms into the exponent by using $\prod_{i=1}^{M_h} h_i = \exp(M_h \log(h_i))$.

The acceptance probability for a change in noise hyperparameter can be formulated as:

$$\alpha_h(m'|m) = \begin{cases} \min \left[1, \exp \left(-M_h \log(h'_i) + M_h \log(h_i) - \frac{\chi^2(m') - \chi^2(m)}{2} \right) \right] & \text{if } h'_i \in H \\ 0 & \text{otherwise,} \end{cases} \quad (.8)$$

where H is the prior range for noise hyperparameters.

Efficient generation of new models

Previous algorithms for transdimensional tomography generate Voronoi parameterizations of arbitrary complexity and either find the intersection between their boundaries and the

precomputed rays (Bodin et al. 2012a) or trace rays in each model (Piana Agostinetti et al. 2015). For the teleseismic body-wave tomography on a continental scale, these calculations remain too computationally intensive for use in the rj-MCMC algorithm. We therefore retain the underlying grid for model m_j and sensitivity matrix G_{ij} , and merge their entries according to the Voronoi volumes as a way to efficiently sample model space.

We have a Voronoi 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.

In order to produce a viable rj-MCMC algorithm for the 3D tomography problem, it is vital to update the model and sensitivity matrix and to evaluate the misfit function as quickly as possible. For each model operation in the search, we exploit efficiencies to reduce computational time by an order of magnitude. An overview of the algorithm can be found in Figure 2.

We begin by selecting a randomized set of Voronoi nuclei. We compute the Delaunay triangulation between the nuclei, which yields a connectivity list—essentially a list of which volumes border each other—and allows us to take advantage of algorithms to efficiently update the model and sensitivity matrix when changing dimensions. We then begin generating new models, choosing at each step a random operation. For the *update slowness*, *kill parameter*, and *move nucleus* operations, we choose a random model parameter, n (or for *update noise hyperparameter*, a random hyperparameter).

- (i) *Update slowness*. This operation is the least intensive since it requires no update to G_{ik}^{vor} , but the time for calculating the data and error function can be improved. We consider only data with non-zero entries in the n^{th} column of G_{ik}^{vor} for update. We calculate the difference in each of these data due to the update in the model value, and then calculate the change in the misfit function from only these updated data.
- (ii) *Kill parameter*. For the following three operations, G_{ik}^{vor} must be updated, but we can avoid starting again from G_{ij} by only addressing the cells affected by the change. When a Voronoi cell is removed, the regular grid volumes that it contained can only be split amongst the cells previously adjacent to it. Therefore, we need only to find the nearest neighbors to these volumes among the nuclei previously connected to the removed nucleus to find their new index and update G_{ik}^{vor} .
- (iii) *Birth parameter*. When adding a new cell, we recompute the Delaunay triangulation

771 and connectivity list to find the existing cells that it borders. Then, we calculate the
772 nearest neighbor nuclei to all volumes within these cells to determine whether or not
773 they belong in the new cell.

774 (iv) *Move nucleus.* Depending on the distance, nuclei n may border new nuclei before and
775 after the move. Therefore, we record the indices of connected nuclei prior to the move
776 and then recompute the triangulation and connectivity. Nearest neighbors are found for
777 all grid volumes within cell n and connected cells before and after.

778 (v) *Update noise hyperparameter.* For a diagonal representation of the data covariance,
779 minor efficiencies can be found by updating the misfit function only for the affected
780 data.

781 **Supplementary tables and figures**

Dataset	Phase	Total rays	Rays in domain
USArray	P	1.29 mil	1.27 mil
EHB (West)	P	1.87 mil	0.57 mil
EHB (East) + ABCE	P	2.1 mil	0.38 mil
EHB + ABCE	pP	0.44 mil	0.125 mil
EHB + ABCE	Pn	1.1 mil	0.035 mil

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.

	Proposal σ	Prior Range
Change slowness	$0.0025 \frac{dV_P}{V_P}$	$\pm 0.05 \frac{dV_P}{V_P}$
Move nucleus (lateral)	100 km	Model domain
Move nucleus (vertical)	10 km	Model domain
Birth nucleus	$0.00375 \frac{dV_P}{V_P}$	$\pm 0.05 \frac{dV_P}{V_P}$
Noise hyperparameter	0.1 s	0–10 s
Nuclei count	–	3–10,000

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

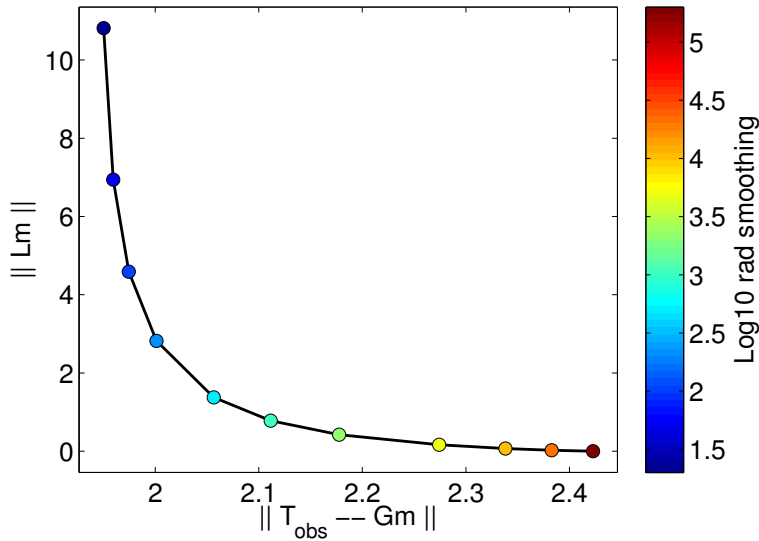


Figure S1. Results of the L-curve test used to determine damping parameters. L represents the regularization operator. Colors give the regularization parameter for smoothing in the radial direction. From this test, we choose a radial smoothing parameter of 1000. Lateral smoothing scales as twice this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

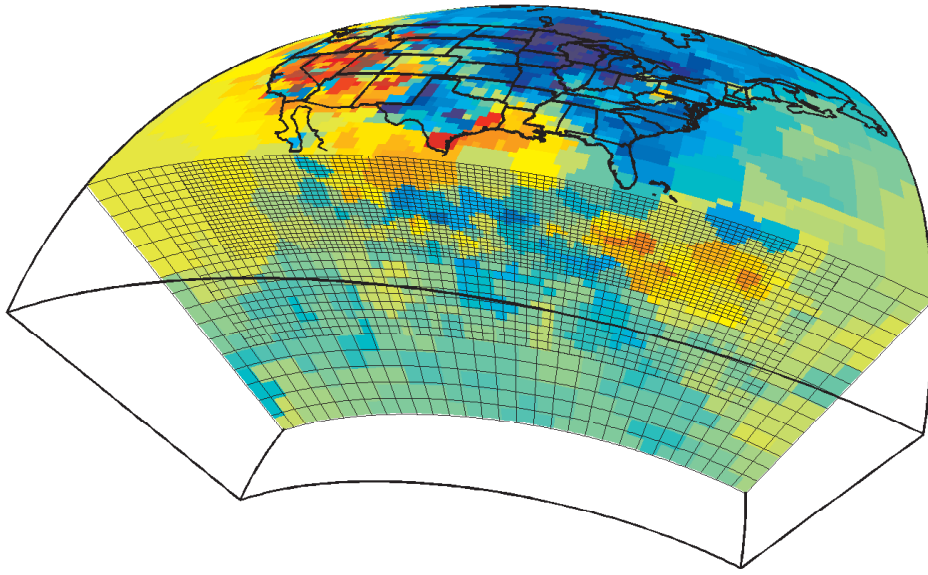


Figure S2. Example starting model for THB tomography inversion. The starting models consist of randomized Voronoi volumes with velocity values derived from a global adaptive grid projected onto a base grid. The base grid is nested set of regular grids with widths of 0.7° , 1.4° , and 2.8° in latitude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