submitted to *Geophys. J. Int.*

3D Monte Carlo Surface Wave Tomography

Xin Zhang[1], Andrew Curtis[1,2], Erica Galetti[1], Sjoerd de Ridder[3]

[1] *School of GeoSciences, University of Edinburgh, Edinburgh,* EH9 3FE*, United Kingdom*

[2] *Department of Earth Sciences, ETH Zürich, Switzerland*

[3] *School of Mathematics, University of Edinburgh, Edinburgh,* EH9 3FD*, United Kingdom*

20 August 2018

SUMMARY

Seismic surface wave tomography is a tried and tested method to reveal the subsurface structure of the Earth. However, the conventional 2-step scheme of inverting first for two-dimensional (2D) maps of surface wave phase or group velocity and then inverting for the 3D spatial velocity structure preserves little information about lateral spatial correlations, and introduces additional uncertainties and errors into the 3D result. We introduce a 1-step 3D non-linear surface wave tomography method that removes these effects by inverting for 3D spatial structure directly from frequency-dependent travel-time measurements. We achieve this using the reversible jump Markov chain Monte Carlo (McMC) algorithm with a fully 3D model parameterization. Synthetic tests show that the method estimates the velocity model and associated uncertainties significantly better than the conventional 2-step McMC method, and that the computational cost seems to be comparable with 2-step McMC methods. The resulting uncertainties are more intuitively reasonable than those from the 2-step method, and provide directly interpretable uncertainty on volumetrics of structures of interest.

Key words: McMC, Tomography, Surface wave

1 INTRODUCTION

Seismic surface waves travel along the surface of the Earth while oscillating over depth ranges that depend on the period of oscillation. Measured speeds of travel are sensitive to Earth properties within those depth ranges. Consequently, surface waves have been used to study the subsurface structure of the Earth on global scales (Trampert & Woodhouse 1995; Shapiro & Ritzwoller 2002; Meier et al. 2007a, b), regional scales (Zielhuis & Nolet 1994; Curtis et al. 1998; Simons et al. 2002) and reservoir scales (de Ridder & Dellinger 2011; Mordret et al. 2013, 2014; Allmark et al. 2018). In those studies, subsurface information is deduced from the dispersion properties of surface wave phase or group velocities, with different frequency components constraining structures over different depth ranges in the subsurface.

Seismic surface wave tomography is often conducted using a two-step inversion scheme (Nakanishi & Anderson 1983; Trampert & Woodhouse 1995; Ritzwoller et al. 2002; Snoke & Sambridge 2002; Bodin & Sambridge 2009; Bodin et al. 2012; Galetti et al. 2017). First, a series of 2D phase or group velocity maps for different periods are estimated tomographically at each geographical point of interest using the arrival times of each period as data; the 1D dispersion curve at each geographical location is then inverted to estimate a 1D shear velocity structure beneath that location. Those 1D shear velocity structures placed side-by-side are interpolated to construct a 3D model.

The surface wave inversion problem is usually solved using a linearised procedure which involves approximating the true, nonlinear relation between data and parameters by a linearised relation; that approximate relationship is then used to seek an approximate solution by minimizing the data misfit while applying some regularization (Trampert & Woodhouse 1995; Ritzwoller et al. 2002). However, the regularization is often chosen by ad hoc means (often trial and error), and valuable information in the data can be concealed by the regularization (Zhdanov 2002). In addition, due to the irregular distribution of seismic sources and receivers, the subsurface is usually unevenly sampled, which limits the resolution of those region with poor data coverage (Curtis & Snieder 2002). The introduction of ambient noise interferometry (Campillo & Paul 2003; Wapenaar 2004; Van Manen et al. 2005, 2006; Wapenaar & Fokkema 2006; Curtis et al. 2006) and ambient noise tomography (Shapiro et al. 2005; Lin et al. 2007; Yang et al. 2007; Bensen et al.

2009; Behr et al. 2010) has partly resolved this issue because earthquakes are no longer necessary in the region of interest as stations play the role of both receivers and (virtual) sources. However, the seismic stations on the Earth's surface are themselves far from uniformly distributed in many areas, and as a result it is difficult to quantify the uncertainties in information derived from linearised ambient noise tomography (Shapiro & Ritzwoller 2002; Bensen et al. 2009; Yao & Van der Hilst 2009; Weaver et al. 2011; Nicolson et al. 2012, 2014). This limits the ability to determine the uncertainty of inferred subsurface shear-wave velocity structures.

To resolve these issues, Bodin & Sambridge (2009) proposed a method using the Markov chain Monte Carlo (McMC) algorithm to sample models from a posterior probability distribution (pdf), based on a Bayesian framework. McMC methods were introduced to Geophysics by Mosegaard & Tarantola (1995). Thereafter, Malinverno & Leaney (2000) introduced an extension of the standard method called *reversible jump McMC* (Green 1995; Green & Hastie 2009), which allows the number of model parameters (the dimensionality of parameter space) to change during the inversion, and applied it on an inversion of zero-offset vertical seismic profiles and an electrical resistivity sounding inversion (Malinverno 2002). Bodin & Sambridge (2009) further applied this method to a seismic tomography problem, in which Voronoi tessellations were used to allow spatially irregular model cells to parametrise seismic wave-speed variations, and trans-dimensional inversion allowed the dimensionality of parameter space (the number of cells in the model) to vary. Thus, the model parameterization is dynamically adapted to the irregular data coverage (Curtis & Snieder 2002), avoiding arbitrary regularization. The method was used successfully to estimate the Rayleigh-wave velocity models across Australia (Bodin & Sambridge 2009), the upper crustal structure of central Java in Indonesia (Zulfakriza et al. 2014) and SE Tibet in China (Zheng et al. 2017). Young et al. (2013) extended this method to include the second depth-inversion step of surface wave tomography to obtain a 3D shear-wave velocity structure of the Tasmanian crust and its uncertainty. Galetti et al. (2015, 2017) further generalized the method by making it fully nonlinear, and observed the emergence of loop-like topologies of the uncertainty structure around velocity anomalies which define the spatial resolution of those structures.

However, when the two steps of surface wave tomography are conducted separately and sequentially, the solution to the 1-D depth inversion cannot interact directly with the 2D phase and group velocity tomography step. In the second step, usually only the mean and standard deviations of the phase and group velocity maps are used as input (Young et al. 2013; Galetti et al. 2017). This may cause the valuable information contained in the pdf of phase and group velocity maps to be lost, and bias the estimate of the pdf of the 3D shear wave velocity structures. As a result, the lateral neighbours in the final 3D model typically preserve little of the 2D lateral spatial correlation information in the phase and group velocity maps.

To overcome both problems, we introduce a fully 3D model parameterization to implement 3D non-linearised surface wave tomography in one step, directly from period-dependent phase or group travel-time measurements, using the reversible jump Markov chain Monte Carlo (rj-McMC) method. This preserves the spatial correlation information in 3D, and consequently also in 2D dispersion maps. It also naturally avoids the loss of pdf information from the 2D phase or group velocity maps in the 3D models. Several previous efforts have been made to extend the trans-dimensional tomography method to 3D. For example, Hawkins and Sambridge (2015) used tree structures and wavelets with rj-McMC to implement 3D trans-dimensional tomography and applied it on a synthetic teleseismic body wave tomography problem. Piana Agostinetti et al. (2015) implemented a 3D trans-dimensional local body wave tomography method using 3D Voronoi tessellation. However, to-date, the fully 3D trans-dimensional scheme has not been used in any surface wave tomography problem, and its properties have not been studied.

In the following we first describe our method, as well as the standard linearised method and the 2-step nonlinear McMC method for surface wave tomography. We then compare our new method with the other methods using a synthetic test. We use the results to conclude that from a seismological point of view there seems little reason to use the 2-step nonlinear methods in future as they seem to be just as computationally demanding as the 3D method, and produce intuitively unreasonable solutions.

2 METHODOLOGY

Following Bodin & Sambridge (2009), we use the reversible jump Markov chain Monte Carlo (rj-McMC) algorithm (Green 1995) to generate samples from the posterior probability density function (pdf) in our seismic velocity model space. In this section, we first describe and compare three possible methods which can be used for seismic surface wave tomography and which are compared below. Then we provide an overview of the rj-McMC algorithm.

2.1 2-step Linearised method

Conventionally surface wave tomographic problems are solved using linearised inversion methods (Iyer & Hirahara 1993). For example, in a first step, phase or group velocity maps can be estimated by minimizing an objective function:

$$\phi = ||\mathbf{Gm} - \mathbf{d}|| + \lambda||\mathbf{m}|| + \mu||\mathbf{Lm}|| \tag{1}$$

where $\mathbf{G} = [\frac{\partial d_i}{\partial m_j}]$ is the first order derivative matrix which defines the assumed linearised physics between parameters $\mathbf{m}$ which describe the tomographic map of localized surface wave speeds, and $\mathbf{d}$ which is the inter-receiver or source-receiver measured dispersion data, $\mathbf{L}$ is a finite-difference derivative matrix and λ and μ are parameters that define the strength of damping and smoothing of the map, respectively. The resulting maps estimated at a number of frequencies can be used to estimate the shear velocity structure beneath each geographical point in a second step that uses a similar linearised expression. Since such tomographic problems are usually under-determined, some forms of regularization is necessary in order to construct a map. Though the regularization parameters λ and μ can be estimated by some accepted methods (e.g. the L-curve method - Hansen 1992), their values are still ad hoc and valuable information in the data can be destroyed by any applied regularization. In reality, the subsurface is often unevenly sampled, which may lead to poor resolution in those areas with poor data coverage, rendering structure in those areas more susceptible to regularisation than in better-constrained areas.

2.2 2-step McMC method

Sampling based methods like McMC are often used to allow non-linearised 2-step inversion in surface wave tomography (Bodin & Sambridge 2009; Bodin et al. 2012; Khan et al. 2013; Young et al. 2013; Rawlinson et al. 2014; Zulfakriza et al. 2014; Saygin et al. 2015; Galetti et al. 2015, 2017; Zheng et al. 2017). In the first step the 2D plane is usually partitioned into convex polygons using a Voronoi tessellation (Sambridge 1995). Each polygon contains one point (called a site) and is defined by the region of space consisting of all of the points nearer to that site than to any other (Figure 1b). Similarly, the model in the second step (depth inversion) can be parameterised using 1D Voronoi tessellation as in Fig 1a (Young et al. 2013; Galetti et al. 2017). In both steps, the McMC method simulates many samples of model space that are consistent with both the data and any available prior information, and varies the tessellation geometry such that it is consistent with the data in a probabilistic sense.

Generally the 1D depth inversions in the second step are run independently at each geographical location without interaction, as this allows perfect parallelisation of what is a computationally demanding task. As a result, the lateral neighbours in the final 3D velocity model may preserve little of the 2D lateral spatial correlation information, and uncertainty estimates based on the set of samples may therefore be incorrect. One option to correct the above errors would be to make each step-2 Markov chain interact with the chains of the lateral neighbours in such a way as to preserve lateral correlations observed in the step-1 inversion. However, this would remove or diminish the principal advantage of parallelisation as different chains would need to pass information to each other during the inversion. Alternatively, we can use a fully 3D Voronoi tessellation to parameterise our seismic velocity model (Figure 1c). 3D Voronoi tessellations have been used in seismic tomography problems to overcome the uneven distribution of data (Zhang et al. 2007). They have also been used within rj-McMC schemes to implement non-linearised 3D body wave tomography (Piana Agostinetti et al. 2015; Burdick and Lekic 2017). Here, we introduce the same for 3D single-step, non-linearised surface wave tomography.

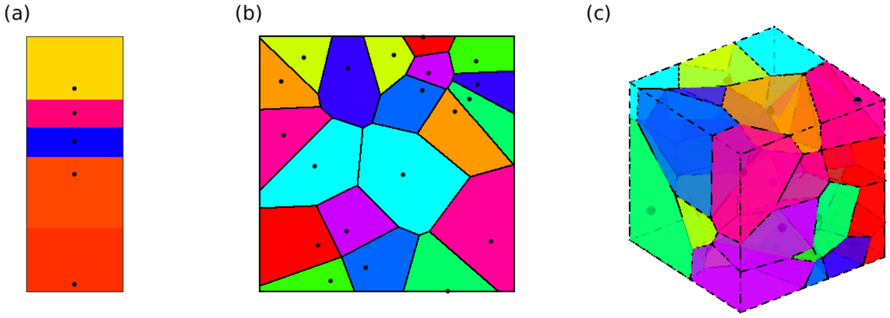
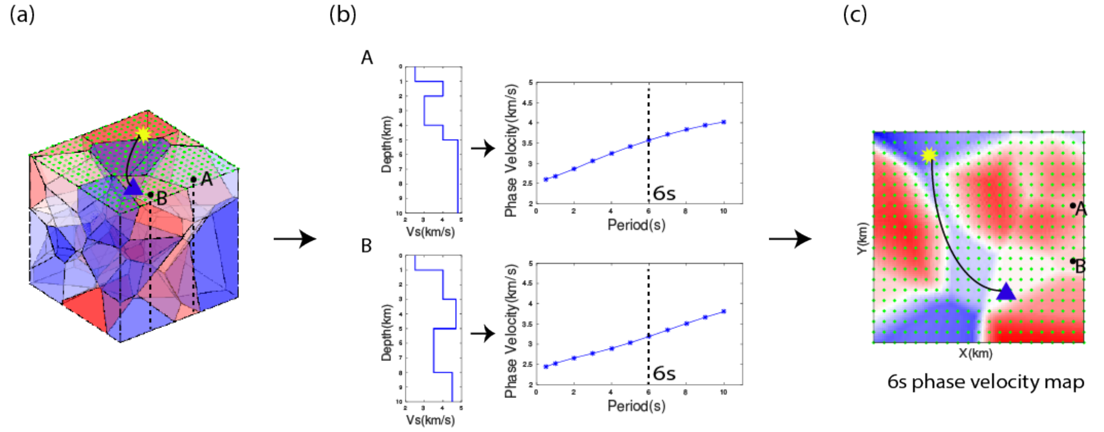


Figure 1. Examples of (a) a 1D Voronoi tessellation, (b) a 2D Voronoi tessellation and (c) a 3D Voronoi tessellation of velocity models. Colours represent seismic velocities in each cell. Black dots are the sites that generated each cell.

2.3 Fully 3D Voronoi tessellation

Our 3D seismic velocity field is discretized by a set of Voronoi polyhedral, each of which is determined by its 3D site location $\mathbf{c}_i$ and shear-wave velocity v_i (P-wave velocity and density are linked to the shear velocity - see below) since surface wave phase and group velocities are primarily sensitive to subsurface shear-velocity variations. Here, we use a constant velocity within each polyhedron but other interpolations between sites could be used if desired (Sambridge et al. 1995).

In order to perform 3-D inversion we need a forward modelling method to calculate the surface wave dispersion that would be measured along any source-receiver path in the case that any particular 3D model were true. Ideally a fully 3D wavefield simulation method could be used but these are generally computationally too expensive. We therefore use an approximate 2-step forward modelling method (Reiter & Rodi 2008). The first step is to compute phase or group velocity maps at each measurement period for our 3D earth model (Figure 2a, b); these can be determined by extracting the 1D shear velocity profile beneath each geographical point, and calculating the phase and group velocities for that 1D structure using a modal approximation (Saito 1988; Herrmann 2013). To calculate source-receiver phase travel times, we then use the fast marching method (Rawlinson & Sambridge 2004) to compute travel times through the phase velocity maps for each period (Ritzwoller & Levshin 1998; Stevens & Adams 2001). For group travel times we



(a) A 3D Voronoi model (b) 1D S-wave velocity profiles and simulated dispersion curves (c) 2D phase velocity map at 6s

Figure 2. A schematic figure of the forward modelling method. (a) A 3D Voronoi model discretized by Voronoi cells, with a dense, regular grid of points overlain on the surface. Colours represent different velocities. Points A and B show two example locations at the surface. (b) The 1D S-wave velocity profiles at points A and B extracted from the 3D Voronoi model (centre-left) and the simulated phase velocity dispersion curves using these 1D profiles (centre-right). (c) An example phase velocity map at 6s constructed by juxtaposing the phase velocity at 6s period calculated for all geographical points such as A and B. The source-receiver phase travel time can then be calculated, for example by tracing rays through the model or by using a fast marching method. The yellow star shows the location of one source and the blue triangle shows the location of one receiver.

integrate the group velocities along the ray path traced through the phase velocity map to determine the group travel times (Cerveny 2005; Reiter & Rodi 2008).

This forward simulation method is thus based on very similar approximations and assumptions to those made in 2-step inversion methods - that the dispersion properties of surface waves at each geographical point depend only on the velocity structure beneath that point. This has the advantage that it allows the 3D inversion method proposed herein to be compared fairly with the other methods (linearised and non-linearised 2-step inversion). Given this forward simulation scheme and measured dispersion data, the rj-McMC method can be used to perform 3D tomography.

2.4 Reversible jump Markov chain Monte Carlo

McMC is a class of algorithms to generate a set (or chain) of samples from a target probability density (Sivia 1996). The Metropolis-Hastings algorithm (Metropolis & Ulam 1949; Hastings 1957) is one such algorithms which is used in many fields. It was introduced to Geophysics over two decades ago (Mosegaard & Tarantola 1995; Malinverno & Leaney 2000; Malinverno 2002, 2004) and was first applied to a seismic tomography problem by Bodin & Sambridge (2009). In their method, a generalised version of the Metropolis-Hastings algorithm called reversible jump Markov chain Monte Carlo (rj-McMC) was used to allow a trans-dimensional inversion (Green 1995; Green & Hastie 2009) which means that the number of model parameters can change along the chain. This allows the parameterization of the seismic velocity model itself to be determined by data and any prior information, avoiding fixing the parameterisation before inversion (Bodin & Sambridge 2009). Following their lead, we apply the rj-McMC algorithm to 3D seismic surface wave tomography.

In our problem, the target probability density is the posterior probability density function (pdf) of our velocity model $\mathbf{m}$ given the observed data $\mathbf{d}_{obs}$, written $p(\mathbf{m}|\mathbf{d}_{obs})$. According to Bayes theorem,

$$p(\mathbf{m}|\mathbf{d}_{obs}) = \frac{p(\mathbf{d}_{obs}|\mathbf{m})p(\mathbf{m})}{p(\mathbf{d}_{obs})} \tag{2}$$

where $p(\mathbf{d}_{obs}|\mathbf{m})$ is called the *likelihood* which is the probability of observing the measured data conditional on a certain model $\mathbf{m}$ being true; $p(\mathbf{m})$ describes the prior information about model $\mathbf{m}$, and $p(\mathbf{d})$ is a normalization factor called the *evidence*. We choose a Gaussian noise distribution for our likelihood with the data variance as an additional parameter that is also estimated during the inversion in a hierarchical way (for more information see Malinverno & Briggs 2004; Bodin et al 2012; Galetti et al 2017). For the prior pdf, we use an uninformative prior - a Uniform distribution with wide bounds on each parameters.

In the rj-McMC algorithm, a new model $\mathbf{m}^{'}$ in the chain is drawn from a proposal distribution $q(\mathbf{m}^{'}|\mathbf{m})$ that depends on the current model $\mathbf{m}$, and is accepted or rejected with a probability $\alpha(\mathbf{m}^{'}|\mathbf{m})$ given by (Green, 1995)

$$\alpha(\mathbf{m}^{'}|\mathbf{m}) = min[1, \frac{p(\mathbf{m}^{'})}{p(\mathbf{m})} \times \frac{q(\mathbf{m}|\mathbf{m}^{'})}{q(\mathbf{m}^{'}|\mathbf{m})} \times \frac{p(\mathbf{d}_{obs}|\mathbf{m}^{'})}{p(\mathbf{d}_{obs}|\mathbf{m})} \times |\mathbf{J}|] \tag{3}$$

where $\mathbf{J}$ is the Jacobian matrix of the transformation from $\mathbf{m}$ to $\mathbf{m}^{'}$ and is used to account for the volume changes of parameter space during jumps between dimensionalities. In our case, it can be shown that the Jacobian is an identity matrix (Bodin & Sambridge 2009). Once a new model is generated via the proposal distribution, it is accepted or rejected by generating a random number γ

from the uniform distribution on [0, 1] and comparing it with the value of the acceptance ratio α. If $\gamma < \alpha$, the new model is accepted; otherwise, the new model is rejected and the current model is repeated as a new sample in the chain. The acceptance ratio α ensures that the density of samples in the Markov chain converges to the posterior probability distribution as the number of samples tends to infinity (Green 1995).

In seismic tomography problems we have five types of perturbations: adding a cell, removing a cell, moving a cell, changing the velocities and changing the data noise hyperparameters. Thus, our algorithm can be described as:

(i) Draw an initial model randomly from the prior pdf.

(ii) Generate a new model $\mathbf{m}'$ by randomly choosing one of the five possible perturbation types listed above, and then perturbing the current model according to the proposal probability.

(iii) Calculate the acceptance ratio α and accept or reject the model according to α.

(iv) Repeat from (ii).

For the proposal probability we choose a Gaussian distribution for the fixed-dimensional perturbation (moving a cell, changing velocities and changing data noise hyperparameters) as also chosen by Bodin & Sambridge (2009). For trans-dimensional perturbations (adding or deleting a cell) we choose to use the prior pdf as the proposal probability since that leads to a higher acceptance ratio compared to using a Gaussian distribution (Dosso et al. 2013). It is a property of McMC methods that in principle the choice of proposal distribution does not affect the fact that the final distribution of samples tends to the posterior pdf as the number of samples tends to infinity.

Note that successive models in a Markov chain are not independent, which for any finite set of samples might cause bias in the estimated posterior probability distribution (Chan & Geyer 1994). Thus, some thinning of the chain, retaining only every 100th sample of the chain, is applied to obtain a final ensemble of samples. Thereafter, statistical properties of the inverse problem solution (e.g. mean, standard deviation) can be calculated from the remaining ensemble of samples.

Monitoring of McMC convergence is important to ensure that the estimated posterior probability density becomes stationary. However, this appears to be a difficult problem, especially for trans-dimensional chains (Green & Hastie 2009). In this study, we chose several scalar statistics,

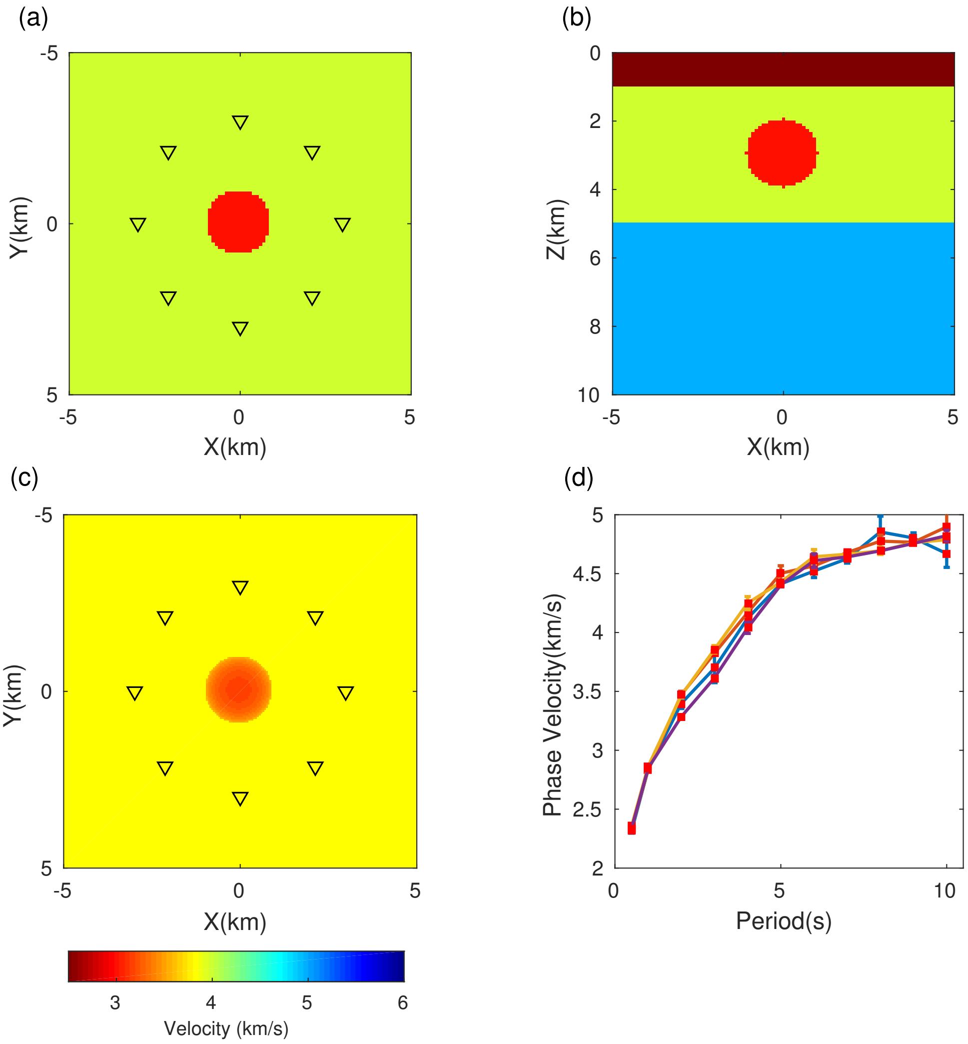


Figure 3. True model and data used for the synthetic test. Open black triangles show the locations of sources and receivers, which are colocated to simulate a typical ambient noise experiment. (a) Horizontal section of the true model at depth of 3km; (b) vertical section at Y=0km; (c) An example phase velocity map at 3s period; (d) Examples of modelled inter-receiver phase velocity dispersion curves.

such as the misfit and the number of cells, to diagnose apparent convergence. When the misfit and the number of cells become stationary, we assume that convergence is attained and begin retaining every hundredth sample from that point on in the chain, which seems to be sufficient in our synthetic test. However, we note that in other more complicated applications, using such scalar statistics may be insufficient such that a more sophisticated approach may be necessary (Green & Hastie 2009).

3 RESULTS

To validate our method, we conducted a synthetic test using Rayleigh wave phase velocity dispersion data. Our true model is composed of 3 layers with S-wave velocities of 2.5km/s, 4km/s and 5.0km/s respectively and a spherical low velocity anomaly of radius 1km within the second layer (Figure 3a, b), of which the velocity is 3.0km/s. A sphere was chosen deliberately because it definitely lies outside of the range of finite-dimensional, straight-sided Voronoi cell parameterisations that can be explored by the Markov chain. Sources and receivers occur at idealised locations to simulate a typical ambient noise experiment where receivers are also used as virtual sources. To calculate the corresponding Rayleigh wave phase velocity dispersion data we use the two-step simulation method described above. First, the phase velocity map is calculated for each period using a 200×200 regular grid on the surface (Figure 3c). The P-wave velocity is given by a fixed v_p/v_s ratio with a typical crustal value of 1.73, and density is assumed to be dependent on v_p through

$$\rho = 2.35 + 0.036 \times (v_p - 3.0)^2 \tag{4}$$

where v_p is given in km/s and ρ is given in g/cm^3 (Kurita 1973). Then, the phase velocity travel times of surface waves are calculated at each period between each station. We then added 0.01 seconds ($\sim 1\%$) Gaussian noise to the data. This gives 28 travel times in total for each period, and we use eleven periods between 0.5s to 10s (red dots in Figure 3d).

To demonstrate our 1-step 3D method, we compared it with the linearised 2-step method and the 2-step McMC method using the synthetic dataset. For the fully 3D McMC method, the velocity prior pdf is set to be a Uniform distribution between 2km/s and 6km/s, which encompasses the true model. The prior pdf on the number of Voronoi cells is selected to be a discrete Uniform distribution between 4 and 200. The prior pdf of the two noise parameters are set to be a Uniform distribution between 0.00001 and 0.01, and a Uniform distribution between 0 and 0.03, respectively. As described above, we use the *prior* distribution for the trans-dimensional perturbation (cell birth and death) and a Gaussian perturbation for the fixed-dimensional steps (changing velocity, position and noise hyperparameters). The width of each Gaussian perturbation is tuned to produce an acceptance rate between 20% to 50%. In this test, we used 16 independent Markov chains, each generating 4 million samples. After a burn-in period of 1 million samples, we thinned each chain by retaining only every 100th sample.

For the 2-step McMC method, the first step is to determine the phase velocity for each period. Here we use the same method as Galetti et al. (2015) which is a fully nonlinear McMC 2D tomographic method using 2D Voronoi tessellation. The velocity prior for each period is set to be a Uniform distribution with a 2km/s width centred at an averaged velocity of all of the ray paths. The prior for the number of Voronoi cells is selected to be a discrete Uniform distribution between 3 and 100 cells (considering that the true phase velocity maps are relatively simple - e.g., Figure 3c). Similarly to above, the width of each Gaussian perturbation for fixed-dimensional steps (changing velocity, position and noise hyperparameters) is tuned to produce an acceptance rate between 20% to 50%. The width of the Gaussian perturbation for trans-dimensional steps (cell birth or death) is selected to give the maximum possible acceptance ratio. For each period, we used 16 independent chains to generate samples from the posterior probability density each with 3 million iterations. After a burn-in of 0.5 million samples, we thinned these chains by retaining only every 100th model to create the final ensemble. Then the phase velocity mean and its standard deviation of the ensemble are calculated at each period on a 100×100 regular grid, and these are taken as the data for the second inversion step. In the second step, we use the McMC based method from Galetti et al. (2017) to determine a 1D shear-wave velocity model beneath each geographic point. For each inversion, we use the same shear-wave velocity prior as in the 3D McMC method – a Uniform distribution between 2km/s and 6km/s. The prior for the number of layers in each 1D depth profile is set to be a discrete uniform distribution between 2 to 20. The proposal distribution for velocity is chosen to be the same Gaussian perturbation as used in the 3D McMC inversion. Other proposal Gaussian distributions are tuned to provide an acceptance rate between 20% to 50%. Similarly, the Gaussian perturbation for the trans-dimensional step (layer birth and death) is selected to give a maximum acceptance ratio. For each 1D inversion beneath each geographical point, we use 6 independent chains, each generating 3 million samples. After a burn-in period of 1 million samples, each chain is thinned by retaining only every 100th model.

Table 1. Parametrization, priors and proposal distribution for all 3 inversions

Method	Parametrization	Priors	Proposal distributions
3D McMC	3D Voronoi cells	cells number: Uniform 4-200 velocities: Uniform 2 - 6km/s	Fixed-D: Gaussian Trans-D: prior
2-step McMC	2D Voronoi cells	cells number: Uniform 3-100 velocities: 2km/s width Uniform	Fixed-D: Gaussian Trans-D: Gaussian
	1D Voronoi layers	layers number: Uniform 2-20 velocities: Uniform 2-6km/s	Fixed-D: Gaussian Trans-D: Gaussian
Linearised inversion	20×20 grid	homogeneous initial model	NA
	1km layered model	estimated from phase velocities	NA

For the linearised inversion, we regard initial models as priors.

For the conventional 2-step linearised inversion, first we need to determine the phase velocity map for each period at which we have data. Here we use fast-marching surface wave tomography (Rawlinson & Sambridge 2005) to determine the phase velocity. For each period, the initial model is chosen to be a homogeneous model with an average velocity of all of the ray paths. The regularization damping and smoothing factors are selected using the standard L-curve method (Hansen 1992). The model is parameterized using a 20×20 regular square grid, which was chosen from a series of regular grids (e.g. 5×5, 10×10, 20×20, 30×30) to give the least model parameters that also produced a small misfit tolerance. For the second step, we use the Occam inversion method (Lai & Rix 1998) to determine the shear-wave velocity structure beneath each grid point. For each 1D inversion beneath each geographical point, the model is parameterized by an evenly spaced layered model with 1km thickness for each layer. The initial velocity of each layer is set to be 1.1 times the phase velocity at the most sensitive period for that particular depth (since for a homogeneous medium the Rayleigh phase velocity is 0.92 times the shear velocity of the medium). The misfit tolerance for the Occam inversion is tuned to give a balance between model complexity and data fitting. To conclude, we summarised the parameterization, priors and proposal distributions used in all three inversions in Table 1.

3.1 Model comparison

Figure 4 shows the mean, standard deviation and relative error ($|\mathbf{m}_i^{mean} - \mathbf{m}_i^{true}|/\sigma_i$) determined from the ensemble generated by the fully 3D 1-step McMC method. The mean velocity model clearly shows a low velocity anomaly at the center of the second layer. It is not a perfect sphere due to the way we parameterized our model (Voronoi tessellation) as cells have straight edges. We also observe three layers in the mean model, though they are not exactly the same as in the true model. Since surface waves are more sensitive to the shallow structures, the top layers are better determined than bottom layers and have smaller uncertainties. From the relative error map, we can see that the error is within 1 standard deviation across most of the model, except at the bottom boundary. This means we have successfully recovered the true model to within approximately correct uncertainty estimates. The uncertainty map at 3km depth in the upper-middle panel of Figure 4 shows low uncertainties in the middle area covered by the data, surrounded by high uncertainties due to lack of data coverage. There is a loop of higher uncertainty around the boundary of the low velocity sphere and at the boundaries between different layers. These loops (in this case actually 3D spherical shells as shown in the cross-section in the lower-middle panel of Figure 4) exist due to the multi-modality of the posterior probability density (essentially that the data can not define whether any point in the loop is inside or outside of that sphere, and hence whether it is a point of low or high velocity) and are mainly caused by ray bending - so second or higher order aspects of wave physics (Galetti et al., 2015, 2017). They define uncertainties in the boundary and shape of velocity anomalies, which could be used to help interpret the velocity map. These are the first 3D uncertainty loops that have been observed, though their existence in 3D was conjectured by Galetti et al. (2015).

The noise level generally affects the complexity of the models obtained (Bodin et al. 2012). Here we show some examples of the histograms of two noise parameters in Figure 5. The noise level is derived from these two parameters through a linear relationship with the ray length (Bodin et al. 2012, Galetti et al. 2017). However, the results do not converge to the true distribution ($\sigma_0 = 0, \sigma_1 = 0.01$). This is probably due to the fact that we used a step size of 0.001 for the proposal distribution, which makes it difficult for σ_0 to approach zero and consequently leads σ_1 to be smaller. In addition, although we added noise with a standard deviation of 0.01 second to the data, it is also quite reasonable that the post-inversion noise level is smaller or larger than the initial noise level because the posterior distribution also accounts for consistency with the prior

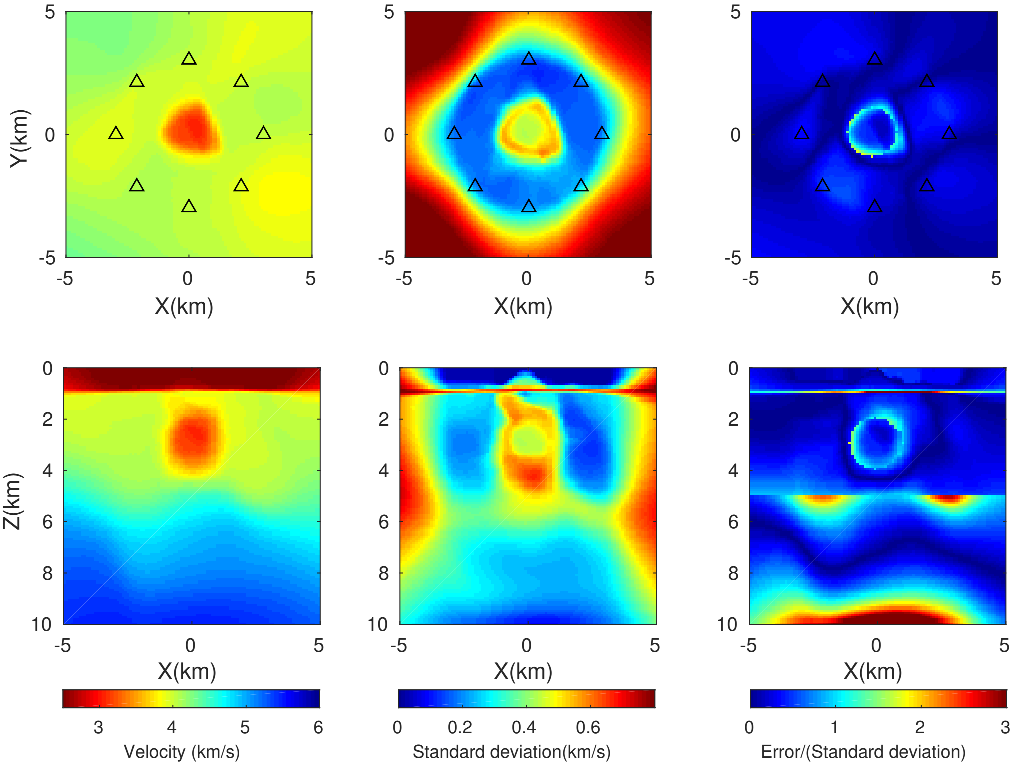


Figure 4. The mean velocity model (left), standard deviation (middle) and the number of standard deviations of the error (right) of the horizontal slice of Z=3km (top) and at the vertical slice of Y=0km (bottom) using 3D McMC method.

range of models. Note that here we added independent Gaussian noise for each datum. In reality the data noise might not be independent, especially at neighbouring frequencies.

To compare our new method with the more standard 2-step McMC method, we show the latter results in Figure 6. The mean velocity model suggests that the low velocity anomaly is clearly estimated, but similar to the results determined using the fully 3D McMC method, its shape is

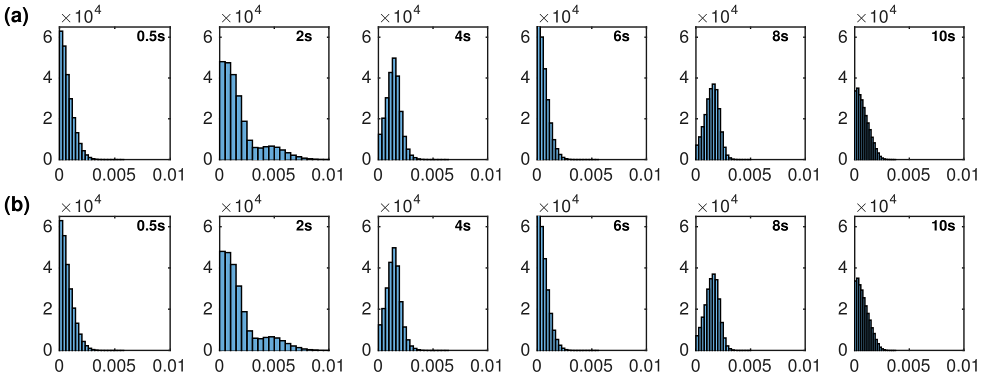


Figure 5. Examples of the histograms of two noise parameters at periods of 0.5s, 2s, 4s, 8s and 10s. The top row shows σ_0, the bottom row shows σ_1. The noise level is derived from them by $\sigma = \sigma_0 * raylength + \sigma_1$.

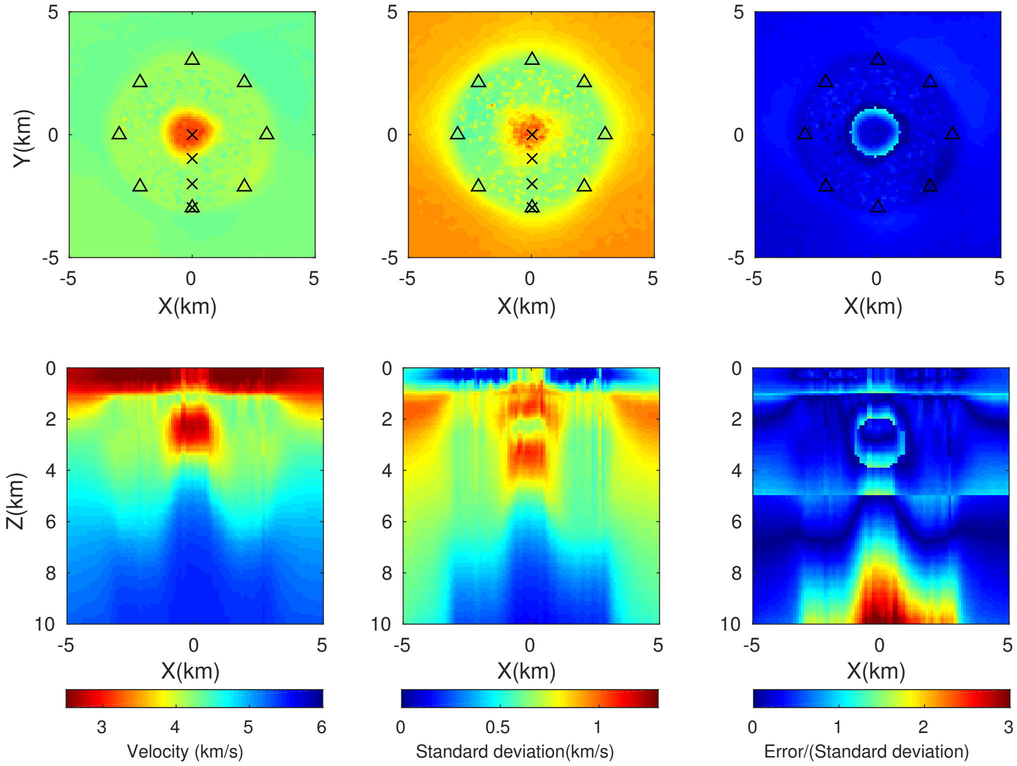


Figure 6. The mean velocity model (left), standard deviation (middle) and the number of standard deviations of the error (right) of the horizontal slice of Z=3km (top) and at the vertical slice of Y=0km (bottom) using 2-step McMC method. The crosses show the point location which are referred to in the text.

not perfect recovered. The top layer is clearly recovered due to high sensitivities at shallow depth, except that the layer boundary is deeper at either side which is likely to be a consequence of the prior information due to the lack of data in those areas. Across the model the magnitude of uncertainties are far higher from the 2-step McMC than those determined from the 1-step 3D McMC method. Even so, the relative error map suggests that much of the bottom layer still has errors of 2 to 3 standard deviations. Parts of the uncertainty loops are observed in the centre of the model at the boundary of the low velocity anomaly. However these do not join up to create an uncertainty shell in 3D as we could expect intuitively – indeed the uncertainty loop in the horizontal plane is not recovered at all by this method, even though these are clearly observed in the first step of the inversion (Galetti et al. 2015; see Figure 7 herein). This shows that the standard 2-step McMC method loses lateral spatial correlations in the second step of the inversion and hence loses the uncertainty loops around the lateral extremities of anomalies, while the fully 3D McMC method provides more intuitively correct results. In other words, the high uncertainty loops at the

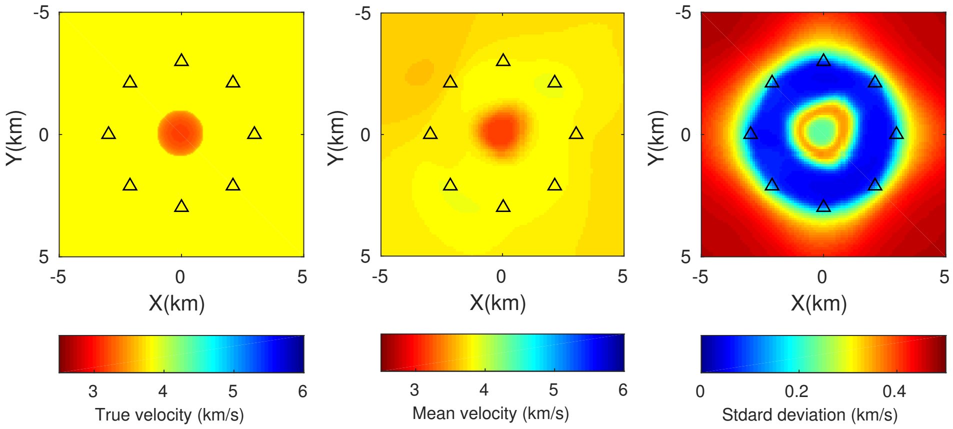


Figure 7. An example of the true phase velocity map (left), the mean phase velocity map estimate (middle) and the standard deviation map (right) at 3s period estimated using the first step of the 2-step McMC method.

boundaries of anomalies observed in the results of 3D McMC method spuriously disappear in the results of the 2-step McMC method.

Notice that there is a consistency problem in this comparison: by changing the parametrization of the model, we have also implicitly changed the prior information that is included in the Bayesian solution. This is almost inevitable in such nonlinear problem. It does not diminish the usefulness of comparing solutions, but does mean that we cannot interpret the results as a direct comparison of solutions to exactly the same problem.

To analyse the possible errors introduced by the 2-step McMC method, in Figure 8 we show the phase velocity posterior pdf as well as their mean and uncertainties determined in the first step along with the true phase velocities at four points shown in Figure 6 (top-left-panel). Those mean phase velocities at points which lie outside of the low velocity anomaly are consistent with the true phase velocities. However, phase velocities within the low velocity anomaly deviate to several standard deviations away from the true velocities at the same location, especially at periods which are sensitive to the low velocity anomaly. It is likely that this is because in this first step of the 2-step method we assume that the velocity at each period is completely decoupled from that of every other period, thus correlations across periods that are imposed by the physics of the problem are lost in this method. This leads to bias in the phase velocities, and subsequently to bias in the shear velocities in the second step when we perform a 1D inversion independently at each geographical

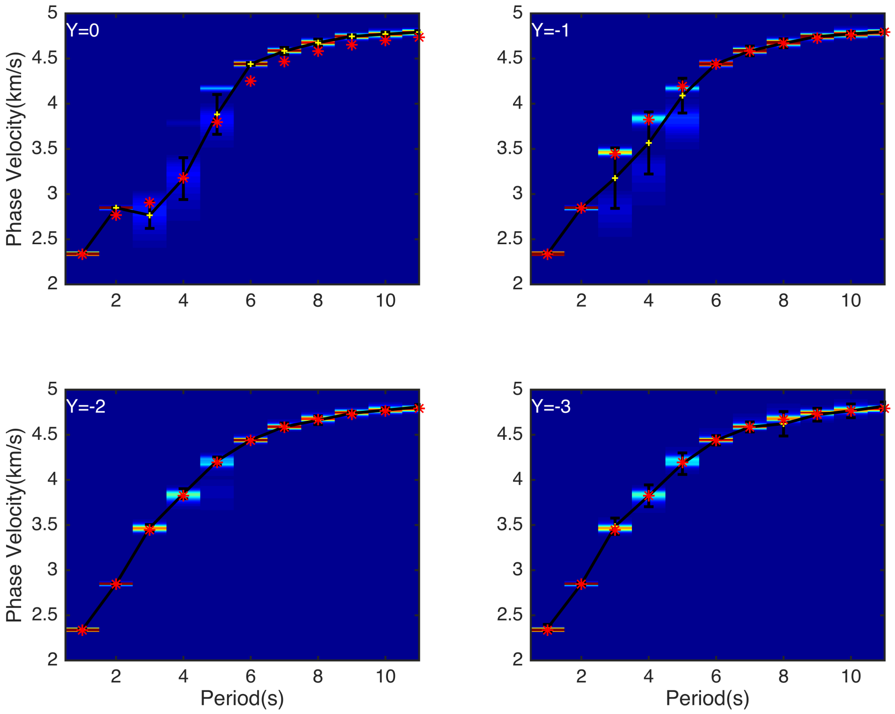


Figure 8. The phase velocity posterior probability density distribution (shading - lighter colours have higher probability) and their mean (yellow plus) and standard deviation (error bar) estimated using the first step of the 2-step McMC method at points Y=0, Y=-1, Y=-2 and Y=-3km shown in Figure 6 (top-left panel). Red stars represent the true dispersion curves.

point. By using a direct 3D method we naturally avoid this drawback because velocities at different periods are naturally coupled and correlated due to the structure of the model with depth: the shear velocity at each depth in the model contributes to the phase velocity over a continuous range of periods, thus imposing phase velocity correlations across periods that are consistent with the assumed physics of the problem. Note that the mean phase velocities at the boundary are far away from true velocities due to the multimodality in the posterior (top-right-panel in Figure 8). As a result, when using only the mean and the standard deviation as data in the second step (as is standard practice), the shear velocity structure is biased.

Figure 9 shows results from the standard 2-step linearised inversion. The velocity model clearly shows the low velocity anomaly and the three layers, though their velocity values are slightly biased which is probably caused by regularization. There are some artefacts around the low velocity anomaly and in the deeper parts of the model, which suggests that the data are overfitted. By comparison, in the McMC based inversions because the parameterization is determined by the data,

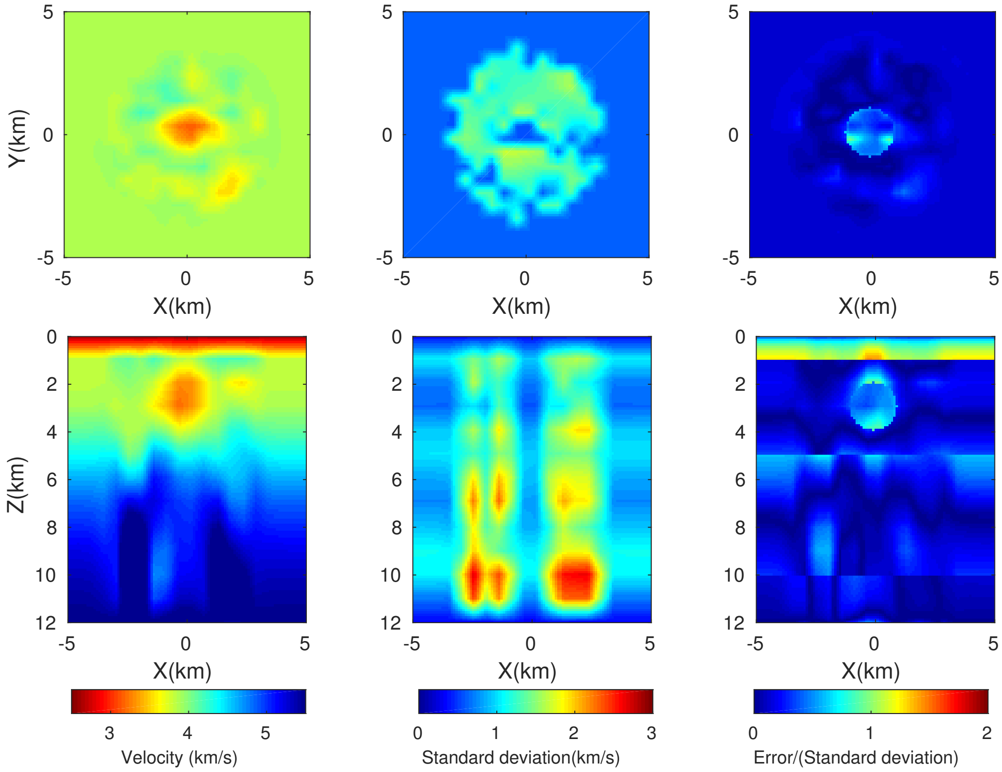
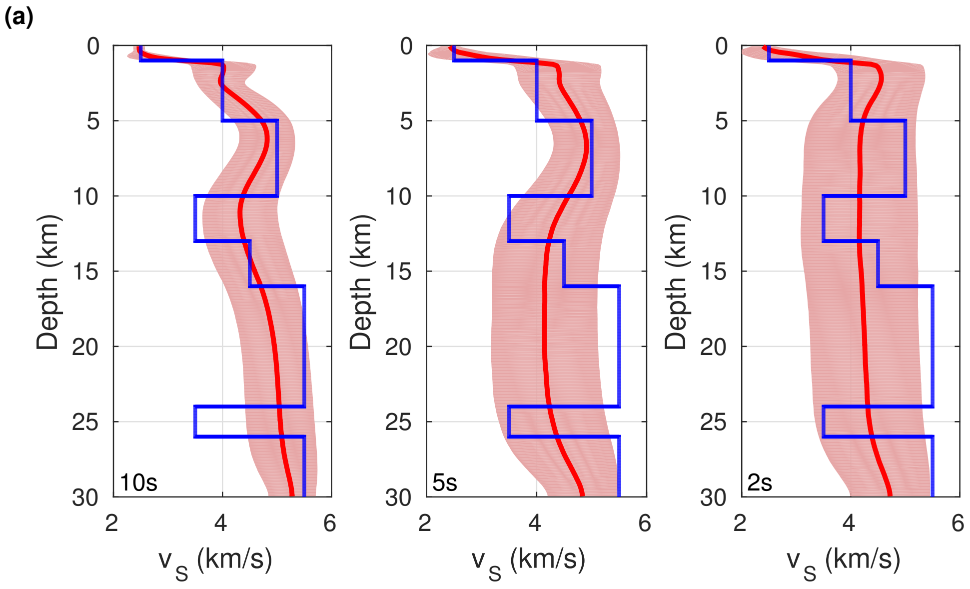
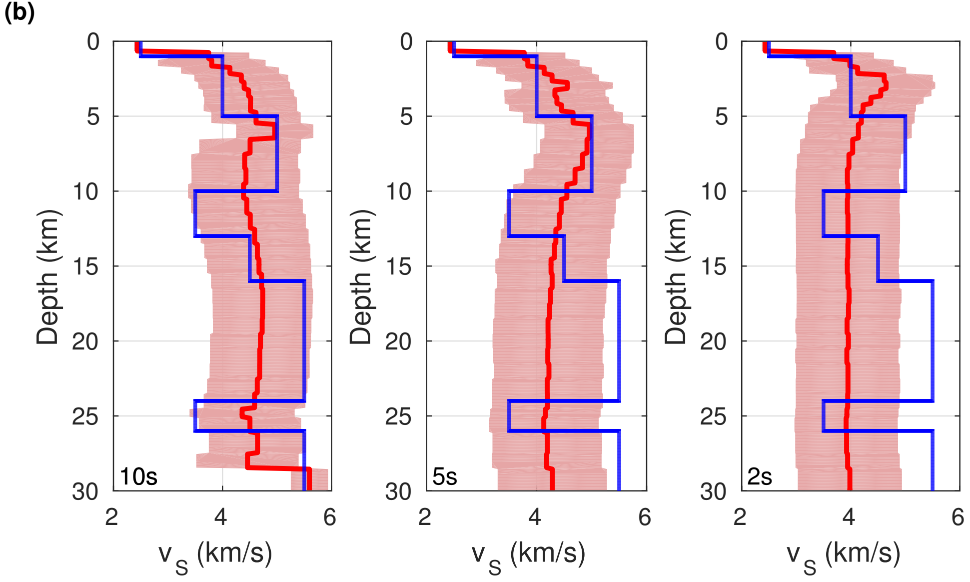


Figure 9. The mean velocity model (left), standard deviation (middle) and the number of standard deviations of the error (right) of the horizontal slice of Z=3km (top) and at the vertical slice of Y=0km (bottom) using standard linearised inversion.

overfitting is reduced which produces a better estimate of subsurface structure. The magnitude of uncertainties estimated by linearised inversion is significantly larger than those from either of the McMC based methods, which leads to small relative errors in the right panels of Figure 9. However, the linearised estimates of uncertainty are generally unreliable because they rely entirely on an estimate of the gradients of data with respect to parameters at only a single point in parameter space. Without prior knowledge of the structure of the problem and its gradients throughout this space, it is really just a chance event whether predicted uncertainties are larger or smaller than they should be in the nonlinear problem. Besides, regularization is often used in linearised problem, making it difficult to quantify the correct uncertainties. Thus, the linearised estimates of uncertainty provide little reliable information to aid interpretation of the velocity model.



(a) Trans-dimensional inversion



(b) Fixed-dimensional inversion

Figure 10. 1D dispersion curve inversion using (a) trans-dimensional inversion and (b) fixed-dimensional inversion. The blue line shows the true shear velocity profile, while the red line shows the posterior mean velocity with its uncertainties (pink shading). At the bottom-left of each figure we show the longest period used in the inversion; in each case the lowest period used was 0.5s and then equally spaced periods (spacing 1s) from 1s up to the maximum were included.

3.2 Uncertainty analysis

In results from the 1-step 3D and 2-step McMC methods, there are low velocity uncertainties in the very bottom layer, which is counter intuitive since surface waves are less sensitive to deep structures. In retrospect we notice that this was also exhibited in the result from the surface wave dispersion curve inversion of Bodin et al. (2012). To further understand this phenomenon we performed several 1D tests using both trans-dimensional inversion and fixed-dimensional inversion with different period ranges of data (Figure 10). The true shear velocity model is set to have 8 layers down to 30km (the blue structure in Figure 10). The period ranges used in these inversions are 0.5-10s, 0.5-5s and 0.5-2s for the left, central and right panels respectively. In the fixed dimensional inversion, the nodes are set to be a regular grid of size 0.5km above 5km depth and size 1km below 5km depth.

Both of the results from 10s inversion show low uncertainties in the deepest layer. The mean velocity shows that we have some resolution at deeper depths (below 10km) because it deviates from the prior mean model of 4km/s towards the true velocity. The relatively more correct mean velocities and smaller uncertainties show that trans-dimensional inversion has slightly higher resolution at deeper depths than the fixed-dimensional inversion. The low uncertainty at the deepest layer is probably due to the accumulated resolutions to deeper depths down to several tens of kilometres at long periods because within the modal approximation used in the forward modellers in our method and that of Bodin et al. (2012), the model is actually assumed to be a half-space below the deepest layers boundary at 30km. When we reduce the longest periods to 5s, there is no low uncertainty at the deepest layer using fixed-dimensional inversion. However, the low uncertainties still exists in trans-dimensional inversion, even if we reduce the longest periods to 2s. The mean velocity model and its uncertainty shows that in reality we do not have any resolution at depths below 15km in this case since the results simply reflect the prior information. Thus, the low uncertainty in the deepest layer is not geophysically interpretable and it always exists in trans-dimensional inversion using Voronoi cells. This is probably due to the fact that natural parsimony of trans-dimensional inversion prefers only one or a few nodes to represent the model where we have little or no resolution, so the lower most Voronoi cell always combines regions where we have no resolution with those in which some resolution exists.

3.3 Computational cost

It is clear that the linearised inversion method is substantially less computationally demanding than sampling based methods. However, it usually gives incorrect estimates of uncertainties since it neglects the non-linearity of the system. Therefore, here we compare the computational cost of the two McMC based methods. Generally, McMC methods need large computational power, especially in high dimensional parameter spaces. To overcome this issue, Bodin & Sambridge (2009) fixed the ray geometry during each McMC chain (thus partly linearising the problem), and updated the rays only between successive McMC chains. However, Galetti et al. (2015) showed that this may introduce artefacts and bias in the solution. Thus, in our study we update the ray geometry in every iteration both for the 2D map inversion in the 2-step method, and in our 1-step 3D McMC method. In our method, every new model is a small perturbation of the previous model, which almost always involves only a small number of cells (Jamin et al. 2018). Thus, in the first step of our two-step forward modelling scheme, we only need to update the phase or group velocity dispersion curves affected by these perturbations, which offers a very significant saving in computation. For example, in one million samples in the above examples, the fully 3D McMC method involves $\sim$3,000 million forward modal simulations of phase or group velocity from 1D shear velocity profiles, while the standard 2-step McMC method involves 10,000 million forward modal simulations. Given that the 'true' model that we use for tests herein is simple, this saving is expected to be substantially greater in a complicated earth structure which needs more cells to represent the model.

However, due to the higher parameter space dimensionality of a 3D model, our new method might require more samples to generate a stationary estimate of posterior probability distributions. For example, in the above synthetic test, the 3D McMC method needs 4 million samples including a 1 million sample burn-in period, while the 2D McMC method needs only 3 million samples including a 0.5 million sample burn-in period. Nevertheless, compared to the time saved in forward modelling described above, this increase in burn-in is not significant in this case. For example, in the above synthetic test the 2-step McMC method involved 30,000 million modal simulations and takes $\sim$4,000 cpu hours for one chain, while the 3D McMC method involves 12,000 million modal simulations and takes $\sim$1,600 cpu hours for one chain. To provide an idea of the overall computational time, the above 3D synthetic test costs 160 hours with each chain parallelized with 9 CPU cores.

Note that the computational cost in each case depends strongly on the method used to assess convergence, which is difficult and depends on some subjective choices. In turn, this introduces some subjectivity to the comparison of computational cost between these methods. Despite this it is at least true that the computational cost of the 1-step and 2-step MC method is comparable in our example. We note however that some other more efficient Bayesian inference methods could be used more easily in the 2-step method than in the 1-step method. For example, Meier et al. (2007) used a Gaussian mixture model to invert for 1D shear-velocity structure from phase velocities, which is more computationally efficient.

4 DISCUSSION

We have shown that using 3D Voronoi tessellation in a McMC surface wave tomography method preserves spatial correlations and better estimates the uncertainties of velocity structures. Because of computational restrictions, we used an approximate forward modelling method in our inversion which still uses the 1D modal approximation to estimate phase velocities at each geographical location across our model. It is certainly possible that this approximation might cause errors in our final model. However, this is the same approximation that is used in the 2-step inversion method, which renders our comparison between the 1-step and 2-step methods fair. Also, there is no impediment to using more precise forward modelling methods if sufficient computational power is available.

Voronoi cells are generally defined using a L2-norm distance metric which treats different directions equally, and in particular it has equal lateral and vertical scales. However, seismic velocities often vary more in the vertical direction than laterally; that is, a large aspect ratio model with horizontal major axes may be more likely than an equal aspect ratio model. In this case, the Voronoi cells might cause the inverted structure to be significantly distorted. One possible solution is to explicitly weight the vertical scale relative to the horizontal. To test this, we created a simple synthetic example by using a layered large aspect ratio model (see Appendix A). According to our test, in a large aspect ratio case, the unweighted Voronoi cells caused the structure to be distorted and led to a long convergence time. In comparison, the explicitly weighted Voronoi cells successfully recovered the true model and used less computational time. However, by doing this we inevitably add a new parameter (the weight) to the inversion, which might affect the results if using different weight values. In our example the results are not particularly sensitive to the choice of weight as our model is not particularly high in its aspect ratio. However, this issue may need to be considered for more complicated real problems.

In our method, Voronoi cells only need to be updated locally around any change in sites in each iteration. For geometry changes (i.e. move, birth and death) we implement a local change method (Jamin et al. 2018). We keep track of the underlying grid velocities at every iteration so that any velocity change can be updated efficiently. This leads to a fast algorithm. However, if global updating is needed, for example if one were to use a globally updating method like Hamiltonian Monte Carlo method, then updating the Voronoi structure itself would probably become inefficient, especially for large numbers of cells.

Though in our simple example the computational cost is comparable with the standard 2-step McMC method, it still requires large computational resources. In McMC, a key factor affecting efficiency is the proposal mechanism. However, it appears to be difficult to construct efficient trans-dimensional proposals, since the natural ideas of closeness or proximity in fixed-dimensional proposals is no longer intuitive or necessarily true (Hastie & Green 2009). In our method, we use the prior pdf for the trans-dimensional proposals (Dosso et al. 2013). However, some other efficient design might be used to further reduce the computational cost (Brooks et al. 2003; Ehlers & Brooks 2008; Green & Hastie 2009; Karagiannis & Andrieu 2016). Another possibility is to use some parallel interacting Markov chains technique such as parallel tempering to increase the model mixing (Earl & Deem 2005; Dettmer & Dosso 2012; Dosso et al. 2012; Sambridge 2014).

5 CONCLUSION

For the first time we implemented 3D fully non-linearised surface wave tomography directly from period-dependent travel-time measurements. We used the rj-McMC method and a parameterization based on 3D Voronoi tessellation. This method preserves the 3D horizontal and vertical spatial correlations in Earth properties and in uncertainties which are not preserved using other existing non-linearised methods. A synthetic test shows that the method provides better estimates of the velocity structure and of uncertainties than previous methods, and reproduces uncertainty loops around velocity anomalies in 3D as would be expected intuitively. It also does so at comparable cost to the standard 2-step Monte Carlo tomography method. This shows that our method is a valuable tool to investigate the shear-wave velocity structure of Earth. At least from the points of view of computation and accuracy of final uncertainty estimates, there seems to be little reason to persist in using 2-step method, although of course there can be other reasons to adopt it (ease of implementation and parallelization of the computation, familiarity etc.).

ACKNOWLEDGMENTS

The authors thank the Edinburgh Interferometry Project sponsors (Schlumberger, Statoil and Total) for supporting this research. This work used the Cirrus UK National Tier-2 HPC Service at EPCC (http://www.cirrus.ac.uk).

REFERENCES

Allmark, C., Curtis, A., de Ridder, S., & Galetti, E., 2018. Seismic attenuation from ambient noise across the north sea ekofisk field, *submitted*.

Behr, Y., Townend, J., Bannister, S., & Savage, M., 2010. Shear velocity structure of the northland peninsula, new zealand, inferred from ambient noise correlations, *Journal of Geophysical Research: Solid Earth*, **115**(B5).

Bensen, G., Ritzwoller, M., & Yang, Y., 2009. A 3-d shear velocity model of the crust and uppermost mantle beneath the united states from ambient seismic noise, *Geophysical Journal International*, **177**(3), 1177–1196.

Bodin, T. & Sambridge, M., 2009. Seismic tomography with the reversible jump algorithm, *Geophysical Journal International*, **178**(3), 1411–1436.

Bodin, T., Sambridge, M., Tkalčić, H., Arroucau, P., Gallagher, K., & Rawlinson, N., 2012. Transdimensional inversion of receiver functions and surface wave dispersion, *Journal of Geophysical Research: Solid Earth*, **117**(B2).

Brooks, S. P., Giudici, P., & Roberts, G. O., 2003. Efficient construction of reversible jump markov chain monte carlo proposal distributions, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **65**(1), 3–39.

Burdick, S. & Lekić, V., 2017. Velocity variations and uncertainty from transdimensional p-wave tomography of north america, *Geophysical Journal International*, **209**(2), 1337–1351.

Campillo, M. & Paul, A., 2003. Long-range correlations in the diffuse seismic coda, *Science*, **299**(5606), 547–549.

Cerveny, V., 2005. *Seismic ray theory*, Cambridge university press.

Chan, K. S. & Geyer, C. J., 1994. Discussion: Markov chains for exploring posterior distributions, *The Annals of Statistics*, **22**(4), 1747–1758.

Curtis, A. & Snieder, R., 2002. Probing the earth's interior with seismic tomography, *International Geophysics Series*, **81**(A), 861–874.

Curtis, A., Trampert, J., Snieder, R., & Dost, B., 1998. Eurasian fundamental mode surface wave phase velocities and their relationship with tectonic structures, *Journal of Geophysical Research: Solid Earth*, **103**(B11), 26919–26947.

Curtis, A., Gerstoft, P., Sato, H., Snieder, R., & Wapenaar, K., 2006. Seismic interferometryturning noise into signal, *The Leading Edge*, **25**(9), 1082–1092.

de Ridder, S. & Dellinger, J., 2011. Ambient seismic noise eikonal tomography for near-surface imaging at valhall, *The Leading Edge*, **30**(5), 506–512.

Dettmer, J. & Dosso, S. E., 2012. Trans-dimensional matched-field geoacoustic inversion with hierarchical error models and interacting markov chains, *The Journal of the Acoustical Society of America*, **132**(4), 2239–2250.

Dosso, S. E., Holland, C. W., & Sambridge, M., 2012. Parallel tempering for strongly nonlinear geoacoustic inversion, *The Journal of the Acoustical Society of America*, **132**(5), 3030–3040.

Dosso, S. E., Dettmer, J., Steininger, G., & Holland, C. W., 2014. Efficient trans-dimensional bayesian inversion for geoacoustic profile estimation, *Inverse Problems*, **30**(11), 114018.

Earl, D. J. & Deem, M. W., 2005. Parallel tempering: Theory, applications, and new perspectives, *Physical Chemistry Chemical Physics*, **7**(23), 3910–3916.

Ehlers, R. S. & Brooks, S. P., 2008. Adaptive proposal construction for reversible jump mcmc, *Scandinavian Journal of Statistics*, **35**(4), 677–690.

Galetti, E., Curtis, A., Meles, G. A., & Baptie, B., 2015. Uncertainty loops in travel-time tomography from nonlinear wave physics, *Physical review letters*, **114**(14), 148501.

Galetti, E., Curtis, A., Baptie, B., Jenkins, D., & Nicolson, H., 2017. Transdimensional love-wave tomography of the british isles and shear-velocity structure of the east irish sea basin from ambient-noise interferometry, *Geophysical Journal International*, **208**(1), 36–58.

Green, P. J., 1995. Reversible jump markov chain monte carlo computation and bayesian model determination, *Biometrika*, pp. 711–732.

Green, P. J. & Hastie, D. I., 2009. Reversible jump mcmc, *Genetics*, **155**(3), 1391–1403.

Hansen, P. C., 1992. Analysis of discrete ill-posed problems by means of the l-curve, *SIAM review*, **34**(4), 561–580.

Hastings, W. K., 1970. Monte carlo sampling methods using markov chains and their applications, *Biometrika*, **57**(1), 97–109.

Hawkins, R. & Sambridge, M., 2015. Geophysical imaging using trans-dimensional trees, *Geophysical Journal International*, **203**(2), 972–1000.

Herrmann, R. B., 2013. Computer programs in seismology: An evolving tool for instruction and research, *Seismological Research Letters*, **84**(6), 1081–1088.

Iyer, H. & Hirahara, K., 1993. *Seismic tomography: Theory and practice*, Springer Science & Business Media.

Jamin, C., Pion, S., & Teillaud, M., 2018. 3D triangulations, in *CGAL User and Reference Manual*, CGAL Editorial Board, 4.12 edn.

Karagiannis, G. & Andrieu, C., 2013. Annealed importance sampling reversible jump mcmc algorithms, *Journal of Computational and Graphical Statistics*, **22**(3), 623–648.

Khan, A., Zunino, A., & Deschamps, F., 2013. Upper mantle compositional variations and discontinuity topography imaged beneath australia from bayesian inversion of surface-wave phase velocities and thermochemical modeling, *Journal of Geophysical Research: Solid Earth*, **118**(10), 5285–5306.

Kurita, T., 1973. Regional variations in the structure of the crust in the central united states from p-wave spectra, *Bulletin of the Seismological Society of America*, **63**(5), 1663–1687.

Lin, F.-C., Ritzwoller, M. H., Townend, J., Bannister, S., & Savage, M. K., 2007. Ambient noise rayleigh wave tomography of new zealand, *Geophysical Journal International*, **170**(2), 649–666.

Malinverno, A., 2002. Parsimonious bayesian markov chain monte carlo inversion in a nonlinear geophysical problem, *Geophysical Journal International*, **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.

Malinverno, A., Leaney, S., et al., 2000. A monte carlo method to quantify uncertainty in the inversion of zero-offset vsp data, in *2000 SEG Annual Meeting*, Society of Exploration Geophysicists.

Meier, U., Curtis, A., & Trampert, J., 2007a. Global crustal thickness from neural network inversion of surface wave data, *Geophysical Journal International*, **169**(2), 706–722.

Meier, U., Curtis, A., & Trampert, J., 2007b. A global crustal model constrained by nonlinearised inversion of fundamental mode surface waves, *Geophysical Research Letters*, **34**, L16304.

Metropolis, N. & Ulam, S., 1949. The monte carlo method, *Journal of the American statistical association*, **44**(247), 335–341.

Mordret, A., Landès, M., Shapiro, N., Singh, S., Roux, P., & Barkved, O., 2013. Near-surface study at the valhall oil field from ambient noise surface wave tomography, *Geophysical Journal International*, **193**(3), 1627–1643.

Mordret, A., Landès, M., Shapiro, N., Singh, S., & Roux, P., 2014. Ambient noise surface wave tomography to determine the shallow shear velocity structure at valhall: depth inversion with a neighbourhood algorithm, *Geophysical Journal International*, **198**(3), 1514–1525.

Mosegaard, K. & Tarantola, A., 1995. Monte carlo sampling of solutions to inverse problems, *Journal of Geophysical Research: Solid Earth*, **100**(B7), 12431–12447.

Nakanishi, I. & Anderson, D. L., 1983. Measurement of mantle wave velocities and inversion for lateral heterogeneity and anisotropy: 1. analysis of great circle phase velocities, *Journal of Geophysical Research: Solid Earth*, **88**(B12), 10267–10283.

Nicolson, H., Curtis, A., Baptie, B., & Galetti, E., 2012. Seismic interferometry and ambient noise tomography in the british isles, *Proceedings of the Geologists' Association*, **123**(1), 74–86.

Nicolson, H., Curtis, A., & Baptie, B., 2014. Rayleigh wave tomography of the british isles from ambient seismic noise, *Geophysical Journal International*, **198**(2), 637–655.

Piana Agostinetti, N., Giacomuzzi, G., & Malinverno, A., 2015. Local three-dimensional earthquake tomography by trans-dimensional monte carlo sampling, *Geophysical Journal International*, **201**(3), 1598–1617.

Rawlinson, N. & Sambridge, M., 2004. Multiple reflection and transmission phases in complex layered media using a multistage fast marching method, *Geophysics*, **69**(5), 1338–1350.

Rawlinson, N., Fichtner, A., Sambridge, M., & Young, M. K., 2014. Seismic tomography and the assessment of uncertainty, in *Advances in Geophysics*, vol. 55, pp. 1–76, Elsevier.

Ray, A., Alumbaugh, D. L., Hoversten, G. M., & Key, K., 2013. Robust and accelerated bayesian inversion of marine controlled-source electromagnetic data using parallel tempering, *Geophysics*, **78**(6), E271–E280.

Reiter, D. T. & Rodi, W. L., 2008. A new regional 3-d velocity model for asia from the joint inversion of p-wave travel times and surface-wave dispersion data, Tech. rep., WESTON GEOPHYSICAL LEXINGTON MA.

Ritzwoller, M. H. & Levshin, A. L., 1998. Eurasian surface wave tomography: Group velocities, *Journal of Geophysical Research: Solid Earth*, **103**(B3), 4839–4878.

Ritzwoller, M. H., Shapiro, N. M., Barmin, M. P., & Levshin, A. L., 2002. Global surface wave diffraction tomography, *Journal of Geophysical Research: Solid Earth*, **107**(B12).

Saito, M., 1988. Disper80: A subroutine package for the calculation of seismic normal-mode solutions, *Seismological algorithms*, pp. 293–319.

Sambridge, M., 2013. A parallel tempering algorithm for probabilistic sampling and multimodal optimization, *Geophysical Journal International*, p. ggt342.

Sambridge, M., Braun, J., & McQueen, H., 1995. Geophysical parametrization and interpolation of irregular data using natural neighbours, *Geophysical Journal International*, **122**(3), 837–857.

Saygin, E., Cummins, P., Cipta, A., Hawkins, R., Pandhu, R., Murjaya, J., Irsyam, M., Widiyantoro, S., & Kennett, B., 2015. Imaging architecture of the jakarta basin, indonesia with transdimensional inversion of seismic noise, *Geophysical Journal International*, **204**(2), 918–931.

Shapiro, N. & Ritzwoller, M., 2002. Monte-carlo inversion for a global shear-velocity model of the crust and upper mantle, *Geophysical Journal International*, **151**(1), 88–105.

Shapiro, N. M., Campillo, M., Stehly, L., & Ritzwoller, M. H., 2005. High-resolution surface-wave tomography from ambient seismic noise, *Science*, **307**(5715), 1615–1618.

Simons, F. J., Van Der Hilst, R. D., Montagner, J.-P., & Zielhuis, A., 2002. Multimode rayleigh wave inversion for heterogeneity and azimuthal anisotropy of the australian upper mantle, *Geophysical Journal International*, **151**(3), 738–754.

Sivia, D., 1996. Data analysis: A bayesian tutorial (oxford science publications).

Snoke, J. A. & Sambridge, M., 2002. Constraints on the s wave velocity structure in a continental shield from surface wave data: Comparing linearized least squares inversion and the direct search neighbourhood algorithm, *Journal of Geophysical Research: Solid Earth*, **107**(B5).

Stevens, J. L., Adams, D. A., & Baker, G. E., 2001. Improved surface wave detection and measurement using phase-matched filtering with a global one-degree dispersion model, Tech. rep., SCIENCE APPLICATIONS INTERNATIONAL CORP SAN DIEGO CA.

Trampert, J. & Woodhouse, J. H., 1995. Global phase velocity maps of love and rayleigh waves between 40 and 150 seconds, *Geophysical Journal International*, **122**(2), 675–690.

van Manen, D.-J., Robertsson, J. O., & Curtis, A., 2005. Modeling of wave propagation in inhomogeneous media, *Physical Review Letters*, **94**(16), 164301.

van Manen, D.-J., Curtis, A., & Robertsson, J. O., 2006. Interferometric modeling of wave propagation in inhomogeneous elastic media using time reversal and reciprocity, *Geophysics*, **71**(4), SI47–SI60.

Wapenaar, K., 2004. Retrieving the elastodynamic green's function of an arbitrary inhomogeneous medium by cross correlation, *Physical review letters*, **93**(25), 254301.

Wapenaar, K. & Fokkema, J., 2006. Greens function representations for seismic interferometry, *Geophysics*, **71**(4), SI33–SI46.

Weaver, R. L., Hadziioannou, C., Larose, E., & Campillo, M., 2011. On the precision of noise correlation interferometry, *Geophysical Journal International*, **185**(3), 1384–1392.

Yang, Y., Ritzwoller, M. H., Levshin, A. L., & Shapiro, N. M., 2007. Ambient noise rayleigh wave tomography across europe, *Geophysical Journal International*, **168**(1), 259–274.

Yao, H. & Van Der Hilst, R. D., 2009. Analysis of ambient noise energy distribution and phase velocity bias in ambient noise tomography, with application to se tibet, *Geophysical Journal International*, **179**(2), 1113–1132.

Young, M. K., Rawlinson, N., & Bodin, T., 2013. Transdimensional inversion of ambient seismic noise for 3d shear velocity structure of the tasmanian crust, *Geophysics*, **78**(3), WB49–WB62.

Zhdanov, M. S., 2002. *Geophysical inverse theory and regularization problems*, vol. 36, Elsevier.

Zheng, D., Saygin, E., Cummins, P., Ge, Z., Min, Z., Cipta, A., & Yang, R., 2017. Transdimensional bayesian seismic ambient noise tomography across se tibet, *Journal of Asian Earth Sciences*, **134**, 86–93.

Zielhuis, A., Nolet, G., et al., 1994. Deep seismic expression of an ancient plate boundary in europe, *Science*, **265**(5168), 79–81.

Zulfakriza, Z., Saygin, E., Cummins, P., Widiyantoro, S., Nugraha, A. D., Lühr, B.-G., & Bodin, T., 2014. Upper crustal structure of central java, indonesia, from transdimensional seismic ambient noise tomography, *Geophysical Journal International*, **197**(1), 630–635.

APPENDIX A: SCALING OF VORONOI CELLS

Voronoi cells are usually defined by a L2-norm distance metric which treats different directions equally, and in particular which has equal lateral and vertical scales. However, seismic velocities often vary more in the vertical direction than laterally; that is, a large aspect ratio model with horizontal major axes may be more likely than an equal aspect ratio model. This will potentially cause models to be distorted when using 3D Voronoi cells. In order to address this potential issue, we tested explicitly weighting the vertical scale compared to lateral scales. Here we conducted some synthetic tests to show the potential limitation of 3D Voronoi cells and the effects of different scaling factors added on the vertical scale.

To focus on large aspect ratio models, we created a $10 \times 2 \times 1km$ layered model with a maximum 10:1 lateral and vertical aspect ratio (Figure A1). The model is composed of 4 layers with shear wave velocity of 1km/s, 1.5km/s, 2km/s and 2.5km/s, respectively. The P wave velocity is derived from shear wave velocity using a fixed ratio of 1.73. We used 10 receivers, each of which

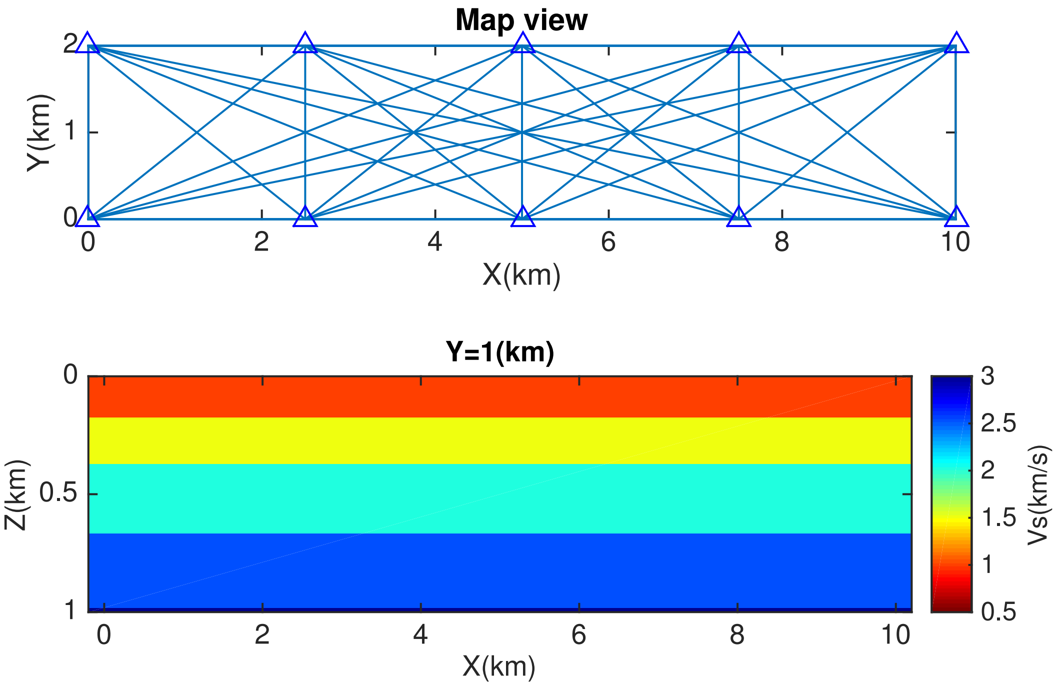


Figure A1. The map view of the true model plotted with receivers and data coverage (top panel) and a vertical slice of Y=1km (bottom panel).

is also used as a virtual source. We used phase velocity data at 9 periods from 0.5s to 2.1s with an equal spacing of 0.2s. To focus on the Voronoi parametrization, we fixed the noise level using a standard deviation of 2 percent of the data.

In order to test the potential effects of different weights added on the vertical scale, we explicitly weighted the vertical scale by factors of 1, 4 and 10. Figure A2 shows some results from those three inversions. For each inversion, we used 8 independent Markov chains. In the top panel we show the number of cells versus computational time. After about 300,000 seconds the inversion with scale factor 1 (left panel) has still not converged, while the other two inversions approximately converge. In terms of the number of cells, the inversion with scaling factor 10 converges fastest (converged after 120,000s). For all three inversions we started to collect posterior samples after a burn-in period of 150,000 seconds and thinned each sample chain by a factor of 100. We show the mean velocity models across a vertical slice (Y=1km) in the middle panel of Figure A2 and their uncertainties at the bottom panel. The mean velocity of the inversion with scaling factor

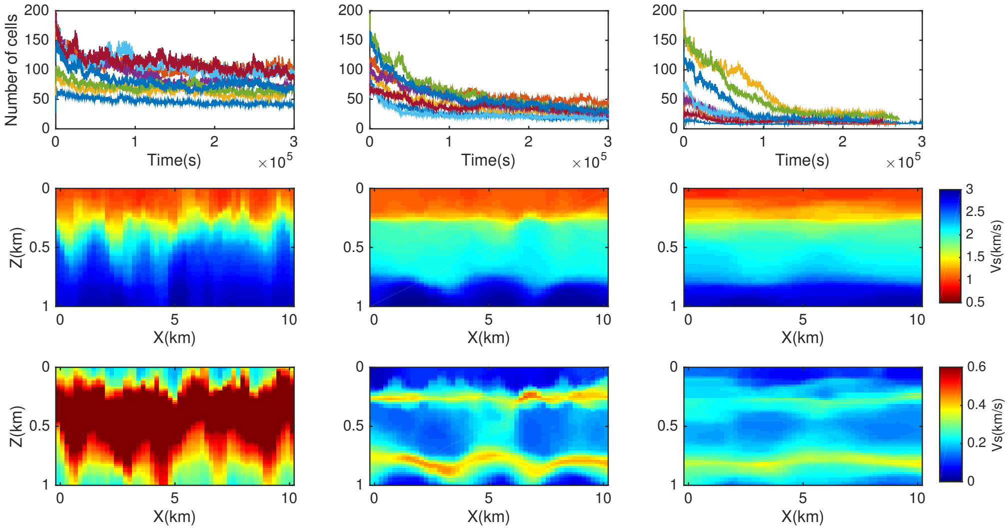


Figure A2. The number of cells versus computational time (top panel), the mean velocity model (middle panel) and standard deviation (bottom panel) across the vertical slice of Y=1km using 3 different vertical scaling factors 1 (left panel), 4 (middle panel) and 10 (right panel).

1 only contains 2 layers associated with relatively large uncertainties. The boundary between the layers also varies significantly across the model. The other two inversions recovered the model significantly better, though the second layer seems to be smeared out which is probably due to the limited resolution of the data used. In both case, the uncertainty maps show high uncertainties at the boundary of layers. For the boundary location of the bottom layer, the inversion with scaling factor 10 recovers the model slightly better with less variation across the model.

Since qualitatively the same results would be expected to occur in 2D or 3D, we conclude that using either Voronoi cells without scaling could cause distortion and bias in inverted results, but that this could be resolved by explicitly adding a scaling factor on the vertical scale. However, this introduces another parameter to the inversion (the scaling factor). In our simple example, this choice does not affect the results significantly since our model is not particularly high in its aspect ratio, but this issue may need to be considered for more complicated real problem.