submitted to *Geophys. J. Int.*

Imaging the subsurface using induced seismicity and ambient noise: 3D Tomographic Monte Carlo joint inversion of earthquake body wave travel times and surface wave dispersion

Xin Zhang[1], Corinna Roy[2], Andrew Curtis[1,3], Andy Nowacki[2], Brian Baptie[4]

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

[2] *School of Earth and Environment, University of Leeds, United Kingdom*

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

[4] *British Geological Survey, Edinburgh, United Kingdom*

7 May 2020

SUMMARY

Seismic body wave travel time tomography and surface wave dispersion tomography have been used widely to characterise earthquakes and to study the subsurface structure of the Earth. Since these types of problem are often significantly non-linear and have non-unique solutions, Markov chain Monte Carlo (McMC) methods have been used to find probabilistic solutions. Body and surface wave data are usually inverted separately to produce independent velocity models. However, body wave tomography is generally sensitive to structure around the sub-volume in which earthquakes occur and produces limited resolution in the shallower Earth, whereas surface wave tomography is often sensitive to shallower structure. To better estimate subsurface properties, we therefore jointly invert for the seismic velocity structure and earthquake locations using body and surface wave data simultaneously. We apply the new joint inversion method to a mining site in the U.K. at which induced seismicity occurred and was recorded on a small local network of stations, and where ambient noise recordings are available from the same stations. The ambient noise is processed to obtain inter-receiver surface wave dispersion measurements which are inverted jointly with body wave arrival times from local earthquakes. The results show that by using both types of data, the earthquake source parameters and the velocity structure can be better constrained than in independent inversions. To further understand and interpret the results, we conduct synthetic tests to compare the results from body wave inversion and joint inversion. The results show that trade-offs between source parameters and velocities appear to bias results if only body wave data are used, but this issue is largely resolved by using the joint inversion method. Thus the use of ambient seismic noise and our fully non-linear inversion provides a valuable, improved method to image the subsurface velocity and seismicity.

Key words: Markov chain Monte Carlo, joint inversion, induced seismicity

1 INTRODUCTION

Seismic tomography is a method to estimate the spatial distribution of properties of the subsurface, and is used in order to understand heterogeneity and processes in the Earth's interior. In seismic tomography one usually parameterizes subsurface properties in some way to form a subsurface *model*, then solves the parameter estimation problem given observed data and a relationship between the data and the parametrized physical properties.

Seismic tomography problems are traditionally solved using linearised methods to estimate the model parameter values which minimize the misfit between observed and synthetically predicted data. These methods first approximate the non-linear physical relation by a linear relation that is valid close to a reference model, and the model is updated to minimize the misfit predicted by that linearisation. This process is iterated until the model update is sufficiently small (Aki & Lee 1976; Dziewonski & Woodhouse 1987; Iyer & Hirahara 1993; Tarantola 2005). Since the problem is often under-determined and ill-posed, regularization is added to the process to enforce particular properties on the model (e.g., smoothness or minimal deviation from a reference model). However, the form of regularization is arbitrary and the strength of regularization is chosen by trial and error by invoking ad hoc criteria. Valuable information can therefore be concealed by regularization (Zhdanov 2002). Moreover, it is difficult if not impossible to estimate accurate uncertainties in solutions of non-linear problems when using linearised methods since the family of model parameter values that fit the data is defined by the true non-linear physics, and not by the linearised relations.

Markov chain Monte Carlo (McMC) methods have been introduced to geophysics to resolve some of these issues (Mosegaard & Tarantola 1995; Malinverno et al. 2000; Malinverno 2002; Malinverno & Briggs 2004; Bodin & Sambridge 2009; Galetti et al. 2015, 2017; Zhang et al. 2018). These methods solve the problem in a Bayesian sense by generating a set (or chain) of samples whose density approximates a *posterior* probability density function (pdf): this describes the probability of the model given both the observed data and any available *prior* information. The method has been extended to trans-dimensional inversions by using reversible jump McMC (rj-McMC – Green 1995; Bodin & Sambridge 2009) such that the dimensionality of the parameter space (the number of parameters and indeed their meaning) can vary in the inversion. This has the advantage that the parameterization can be adapted and simplified so as to best represent information in the data and prior information without over-parameterizing the model, which significantly improves performance in otherwise high dimensional problems (Malinverno & Briggs 2004; Bodin & Sambridge 2009; Bodin et al. 2012; Galetti et al. 2015; Zhang et al. 2018). The rj-McMC method has been used to estimate 2D phase or group velocity maps of the crust (Bodin & Sambridge 2009; Zulfakriza et al. 2014; Galetti et al. 2015; Zheng et al. 2017; Crowder et al. 2019b) and to estimate seismic velocity profiles with respect to depth in the Earth (Bodin et al. 2012; Shen et al. 2012, 2013; Young et al. 2013; Galetti et al. 2017; Zhang et al. 2019, 2020). The method was recently extended to estimate 3D velocity models using body wave travel time data (Piana Agostinetti et al. 2015; Hawkins & Sambridge 2015; Burdick & Lekić 2017) and surface wave dispersion data (Zhang et al. 2018, 2019, 2020).

In the above studies, body waves and surface waves are used separately to construct velocity models. Seismic body waves are generally sensitive to deeper structure where earthquake sources occur, and produce limited resolution closer to the surface. This is because we usually have a relatively sparse station array compared to the density of sources, which results in relatively sparse body wave ray coverage in the shallower Earth. In comparison, fundamental mode surface waves are generally more sensitive to shallower rather than to deeper structure. Body and surface wave data can therefore usefully be combined to better constrain the subsurface velocity structure.

Such joint inversions have already been used widely to study the crust and upper mantle structure (West et al. 2004; Reiter & Rodi 2008; Obrebski et al. 2011, 2012; Rawlinson & Fishwick 2012; Zhang et al. 2014; Syracuse et al. 2015; Fang et al. 2016; Liu & Zhao 2016; Roecker et al. 2017). However, these studies were performed using linearised inversion methods which renders associated uncertainty estimates questionable at best. In this study we apply the rj-McMC algorithm to fully non-linear joint inversion using both body wave arrival times and surface wave dispersion data. We show that results are significantly improved over independent body or surface wave inversions, both in terms of velocity structure and earthquake source location uncertainties.

In the next section we summarise the rj-McMC algorithm and describe how it is applied to the joint inversion problem. In section 3 we apply the new McMC joint inversion method to data from an ex-mining site located to the north of New Ollerton, U.K, and compare the results with those from individual inversions in section 4. Finally we discuss the implication of this work in section 5 before concluding.

2 METHODOLOGY

2.1 Parametrization

As in Zhang et al. (2018, 2020), in order to perform trans-dimensional inversion in three spatial dimensions we use Voronoi cells to parameterize our seismic velocity models (Figure 1). A Voronoi cell is defined by a generating point (called a site) and its volume which consists of all points that are closer to that site than to any other. Each cell has associated seismic properties, e.g., P-wave velocity and shear wave velocity. In this study, we use constant velocities within each cell. Our velocity model can therefore be parameterized as $(\mathbf{c}, \mathbf{v}_p, \mathbf{v}_s)$, where $\mathbf{c}$ is the vector of positions of Voronoi sites, and $\mathbf{v}_p$ and $\mathbf{v}_s$ are vectors of the associated P-wave velocity and shear wave velocity in each cell. Each earthquake source (number i) is parameterized as $s_i = (x_i, y_i, z_i, t_i)$, where x_i, y_i, z_i is the source location and t_i is the origin time. Our model $\mathbf{m}$ therefore can be represented as $(\mathbf{c}, \mathbf{v}_p, \mathbf{v}_s, \mathbf{s})$.

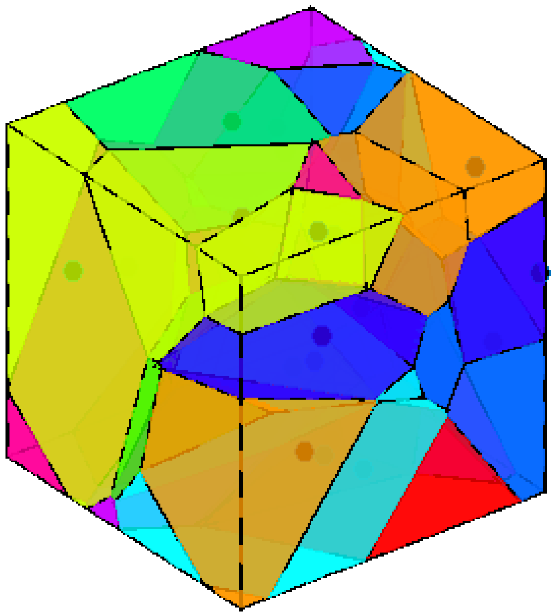

Figure 1. Example 3D Voronoi tessellation of a velocity model. Colours represent seismic velocities which are constant across each cell. Black dots (which appear grey in the 3D rendering) are the sites that generated each cell.

2.2 Reversible jump Markov chain Monte Carlo (rj-McMC)

We use rj-McMC to perform 3D tomographic inversion following the approach of Zhang et al. (2018). Rj-McMC is a generalized Metropolis-Hastings algorithm which generates a chain of samples distributed according to a target probability density. The algorithm allows the number of parameters to change along the chain (Green 1995), which makes the parameterization adaptable to the data and avoids the need to specify it exactly prior to the inversion (Bodin & Sambridge 2009). In seismic tomography we are interested in the posterior pdf of model $\mathbf{m}$ given the observed data $\mathbf{d}_{obs}$,

$$p(\mathbf{m}|\mathbf{d}_{obs}) = \frac{p(\mathbf{d}_{obs}|\mathbf{m})p(\mathbf{m})}{p(\mathbf{d}_{obs})} \tag{1}$$

where $p(\mathbf{d}_{obs}|\mathbf{m})$ is the *likelihood* which describes the probability of data given a specified model $\mathbf{m}$; $p(\mathbf{m})$ is the prior pdf which describes information that is independent of data and $p(\mathbf{d}_{obs})$ is a normalization factor called the *evidence*. We use a Gaussian distribution for the likelihood, for which the data variance is estimated in a hierarchical way in the inversion (Malinverno & Briggs 2004; Bodin et al. 2012; Zhang et al. 2018, 2019, 2020). The prior $p(\mathbf{m})$ is chosen to be a Uniform distribution.

Within each chain a new model $\mathbf{m}^{'}$ is drawn from a so-called 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)

$$\alpha(\mathbf{m}^{'}|\mathbf{m}) = min[1, \frac{p(\mathbf{m}^{'})}{p(\mathbf{m})} \times \frac{q(\mathbf{m}|\mathbf{m}^{'})}{q(\mathbf{m}^{'}|\mathbf{m})} \times \frac{p(\mathbf{d}_{obs}|\mathbf{m}^{'})}{p(\mathbf{d}_{obs}|\mathbf{m})} \times |\mathbf{J}|] \tag{2}$$

where $\mathbf{J}$ is the Jacobian matrix of the transformation from $\mathbf{m}$ to $\mathbf{m}^{'}$ and is used to account for the volume change of parameter space during jumps between different dimensionalities, but where in this case the Jacobian is an identity matrix (Bodin & Sambridge 2009). The new model $\mathbf{m}^{'}$ is accepted or rejected by generating a random number γ from a Uniform distribution on (0, 1] and comparing it with α. If $\gamma < \alpha$, the new model $\mathbf{m}^{'}$ is accepted; otherwise the new model is discarded and the current model is repeated as a new sample in the chain. The acceptance ratio α guarantees that the density of samples converges to the posterior pdf asymptotically as the number of samples tends to infinity (Green 1995).

Monitoring the convergence of Markov chains is an important component of McMC methods. In this study, we use the absolute misfits and the number of cells to monitor convergence as used in several previous studies (Bodin & Sambridge 2009; Bodin et al. 2012; Dosso et al. 2014; Galetti et al. 2015; Hawkins & Sambridge 2015; Zhang et al. 2018, 2019, 2020). For example, when the misfit value and the number of cells become approximately stationary, we assume the chain has reached some sort of dynamic equilibrium. Since consecutive samples are correlated (McMC is a random walk process and only converges to the posterior distribution as the number of samples tends to infinity), the estimated probability pdf from any finite set of samples is often biased (Chan & Geyer 1994). Therefore, we retain every 50^{th} sample along the chain once equilibrium has been reached, and only those retained samples are used to calculate parameter statistics (mean, standard deviation, etc.).

2.3 Joint inversion of body waves and surface waves

In seismic body wave tomography, the earthquake source locations are generally unknown within some volumetric region of uncertainty as are origin times. We therefore include these source parameters in our inversion. This produces a trade-off between source parameters and the seismic velocity model. To reduce this effect, Piana Agostinetti et al. (2015) updated the source origin times in an optimization for each velocity model. However that approach may cause errors in the results since sources may converge to incorrect locations and times, and it does not allow correct uncertainty analysis for source parameters. In this study we therefore also include origin times as parameters to be varied in the Markov chain. We start the chains with initial source parameter values obtained using a standard linearised optimization, whereafter they can vary freely within the prior pdf (defined below).

To forward model body wave travel time data we use a 3D fast marching method (Rawlinson & Sambridge 2004; Valero-Gomez et al. 2013). Due to source-receiver reciprocity, fast marching can be conducted either from sources to receivers or vice versa. Therefore, in practice one chooses the more efficient option based on the minimum number of sources and receivers, and we model from receivers to sources. The grid spacing affects the accuracy of travel times modelled by the fast marching method. In this study we use a spacing of 100 m which our tests showed is sufficient to produce accurate travel times (Rawlinson & Sambridge 2004).

For surface wave dispersion data, we use the two step forward modelling method described in Zhang et al. (2018) and applied in Zhang et al. (2019, 2020). First, for each geographical point the local phase velocity at each frequency is computed using the 1D velocity profile beneath that point using a modal approximation (Herrmann 2013) to create a 2D phase velocity map across the surface. Then, since our dispersion measurements are made between two receivers, for each receiver-to-receiver pair the phase travel time at a specific frequency can be calculated using a 2D ray tracing method (Rawlinson & Sambridge 2004). Group velocity travel times can be calculated by integrating over the ray path traced through phase velocity maps (Cerveny 2005; Reiter & Rodi 2008).

In joint inversion, the relative weights between different data types usually affect the results significantly (Bodin et al. 2012; Shen et al. 2012). In linearised methods, the weight is generally determined by subjective choices which could cause errors in results. In this study we set the data noise level of both data types to be free parameters so that the relative importance of different data types is determined by their own noise level (Bodin et al. 2012; Shen et al. 2012). As in Galetti et al. (2017) and Zhang et al. (2018), we hyper-parameterize the noise level using a linear relation with respect to travel times $\sigma = \sigma_0 \times traveltime + \sigma_1$, for each of body and surface wave travel times independently, where σ_0 and σ_1 are free hyperparameters.

In our rj-McMC algorithm there are six types of perturbation: adding a Voronoi cell, removing a cell, moving a cell, changing a cell's seismic velocity, changing the source parameters, and changing the data noise hyperparameters. This results in the following algorithm:

(i) Select an initial model $\mathbf{m}$ from the prior pdf (for seismic velocities) or from a linearised inversion (for source locations and times)

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

(iii) Calculate the acceptance ratio α in equation (2) and accept or reject model $\mathbf{m}^{'}$ with probability α. if $\mathbf{m}^{'}$ is accepted, let $\mathbf{m} = \mathbf{m}^{'}$.

(iv) Repeat from (ii).

For the fixed-dimensional step (moving a cell, changing velocities, changing source parameters and changing the hyperparameters), we use a Gaussian distribution which is centred at the current model as the proposal distribution. The width of the Gaussian is a parameter which needs to be tuned for each inversion (Hawkins & Sambridge 2015; Zhang et al. 2018). For trans-dimensional steps (adding or deleting a cell), the prior is used as the proposal distribution which usually gives a higher acceptance ratio than using a Gaussian proposal distribution as noted in Dosso et al. (2014).

3 APPLICATION TO THE NEW OLLERTON MINING SITE

We applied the method to a mining site located to the north of New Ollerton, Nottinghamshire, U.K (Figure 2) which operated from 1925 to 2015. A network of seven stations was deployed at the site and recorded 291 microseismic events in 2014. Figure 2 shows the location of the stations, event locations from the initial linearised inversion, and a histogram of the depth distribution of event locations. The events mainly occurred around 0.9 km depth with a few occurring significantly shallower or deeper. We used a total of 1725 P-wave arrival times and 923 S-wave arrival times obtained from the British Geological Survey (BGS) (Butcher et al. 2017).

We applied ambient noise interferometry (Campillo & Paul 2003; Curtis et al. 2006; Bensen et al. 2007) to obtain surface wave dispersion data for each inter-receiver pair. The data are first band-filtered between 0.8 s and 3.0 s to filter out earthquake signals which comprise higher frequencies. Cross-correlations between each receiver pair are then calculated using 24-hour long time segments, which are then stacked over the whole year to improve the signal-to-noise ratio (SNR). The group velocity dispersion of each receiver pair is picked using the frequency-time analysis (FTAN) method (Dziewonski et al. 1969; Levshin et al. 1972, 1992; Herrin & Goforth 1977; Russell et al. 1988; Ritzwoller & Levshin 1998; Levshin & Ritzwoller 2001; Nicolson et al. 2012; Yanovskaya et al. 2012). Figure 3 shows an example of the FTAN image used to pick group velocities. We discarded station-pairs for which the SNR is smaller than 5, and those whose inter-receiver distances are smaller than twice the wavelength at any frequency due to far-field surface wave approximation that is implicit within ambient-noise surface wave tomography (Yao et al. 2006; Lin et al. 2009). The SNR is calculated using the spectrum of the signals of interest and the spectrum of an interval of noise extracted from the end of the virtual source records. This results in a total of 12 inter-receiver dispersion curves across the New Ollerton area (Figure 2). Since Nicolson et al. (2012) and Galetti et al. (2017) both showed that uncertainties estimated directly from the FTAN images tend to be poor, uncertainties in dispersion curves were estimated hierarchically within the Markov chain. Note that since we use the same relationship between data uncertainties and travel times for all dispersion curves, unusually poor measurements probably cannot be assigned high uncertainties and can bias the results. In practice this issue can be reduced by removing potentially poor measurements.

We performed three different inversions: first using only body wave travel times, second using only surface wave dispersion data, and a third, joint inversion using both types of data. In body wave inversion and joint inversion we invert for both P and S wave velocity. In the surface wave inversion we only invert for S-wave velocity; in that case P-wave velocity is linked to S-wave velocity using a typical ratio 1.73, and density is computed from the P-wave velocity using a typical crustal relationship $\rho = 2.35 + 0.036 \times (v_p - 3.0)^2$ where v_p is in km·s^{-1} and ρ is given in g·cm^{-3} (Kurita 1973). The latter relationship is also used to calculate density in the joint inversion. Since surface waves are much more sensitive to shear velocity than P-wave velocity or density, the approximation should be sufficient in our case. For each inversion the prior pdf of shear velocity is set to be a Uniform distribution between 1.0 km/s and 4.0 km/s at all 3D locations. For body wave inversion and joint inversion the prior pdf of P-wave velocity is set to be a Uniform distribution between 1.6 km/s and 6.0 km/s. The prior pdf of the number of Voronoi cells is chosen to be a Uniform distribution between 20 and 300. For each event location we use a Uniform distribution across a 2 km box centred at the initial location estimated by BGS using linearised methods (Butcher et al. 2017) as the prior pdf, and for the origin time we used a Uniform distribution with 1 second width centred at the initial origin time. For body wave travel times the prior pdf of the hyperparameters σ_0 and σ_1 are chosen to be Uniform distributions between 0 and 0.1. Similarly for surface wave group travel times the prior of the two hyperparameters are set to be Uniform distributions between 0 and 0.2. Since seismic velocity generally varies more rapidly in the vertical direction than horizontally, we scaled the vertical direction by a factor of 5 larger to ensure vertical and horizontal directions are balanced as demonstrated by Zhang et al. (2018). For each chain an initial velocity model is generated from the prior, whereas initial source locations and times are set to values from the linearised inversion of Butcher et al. (2017).

For a fixed-dimensional step (moving a cell, changing velocities, changing source parameters and hyperparameters) we use a Gaussian perturbation centred at the current value as the proposal distribution. The width of the Gaussian distribution is chosen by trial and error to give an accep-

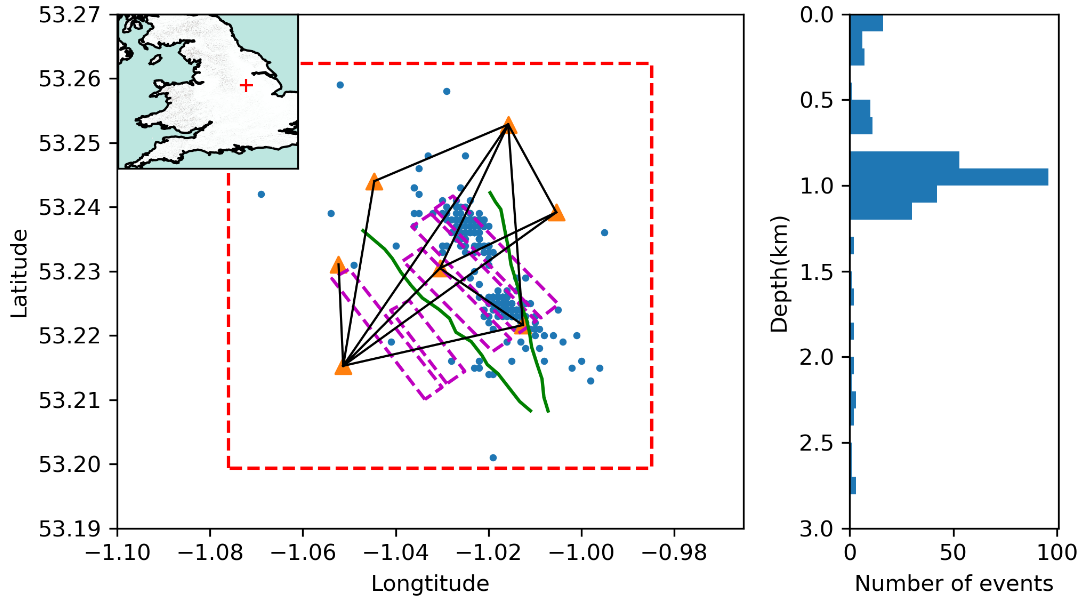

Figure 2. Microseismic events (blue dots) recorded with seven seismic stations (orange triangles) at New Ollerton mining site. The event locations are those found in the linearised inversion of Butcher et al. (2017). Black lines between stations show approximate paths along which surface wave dispersion data are available. Green lines show faults that appear in the study area (Bishop et al. 1993). Purple boxes show the location of coal seams located between 800 and 900 m depth. The red dashed-line box shows the extent of the other maps herein. The red plus in the inset map denotes the location of the mining site in England. The right panel shows a histogram of the event depths.

tance ratio between 20 and 50 percent. For a trans-dimensional step (adding or deleting a cell) the proposal distribution is chosen to be the prior pdf (Dosso et al. 2014; Zhang et al. 2018). For each inversion we used 16 chains; each of which contains 1,600,000 samples including a burn-in of 800,000 to reach apparent equilibrium. To reduce correlations between successive samples we only retain every 50^{th} sample in the chain post burn-in. Those sample are used to calculate parameter means and standard deviations. Final maps of statistics (mean and standard deviation) of solutions are presented without additional imposed smoothing.

4 RESULTS

4.1 Source parameters

Figure 4 shows the mean and standard deviation of each event location calculated using all collected samples from body wave inversion (Figure 4a and b) and from joint inversion (Figure 4c and d). Both results show that events occur deeper (majority $>$ 1 km) than the initial locations

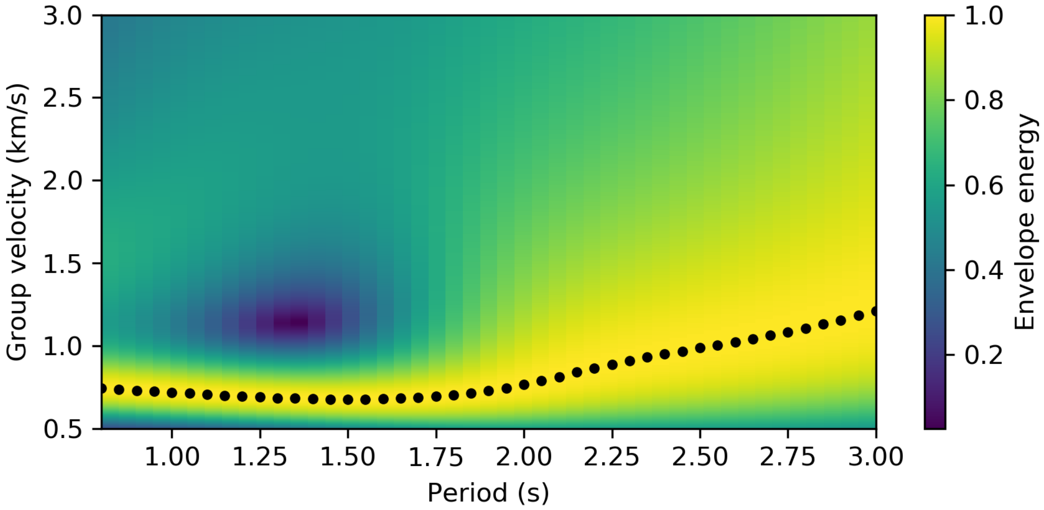

Figure 3. An example of a frequency-time analysis (FTAN) envelope image which is used to pick group velocities. The black dots show the group velocities picked in this case.

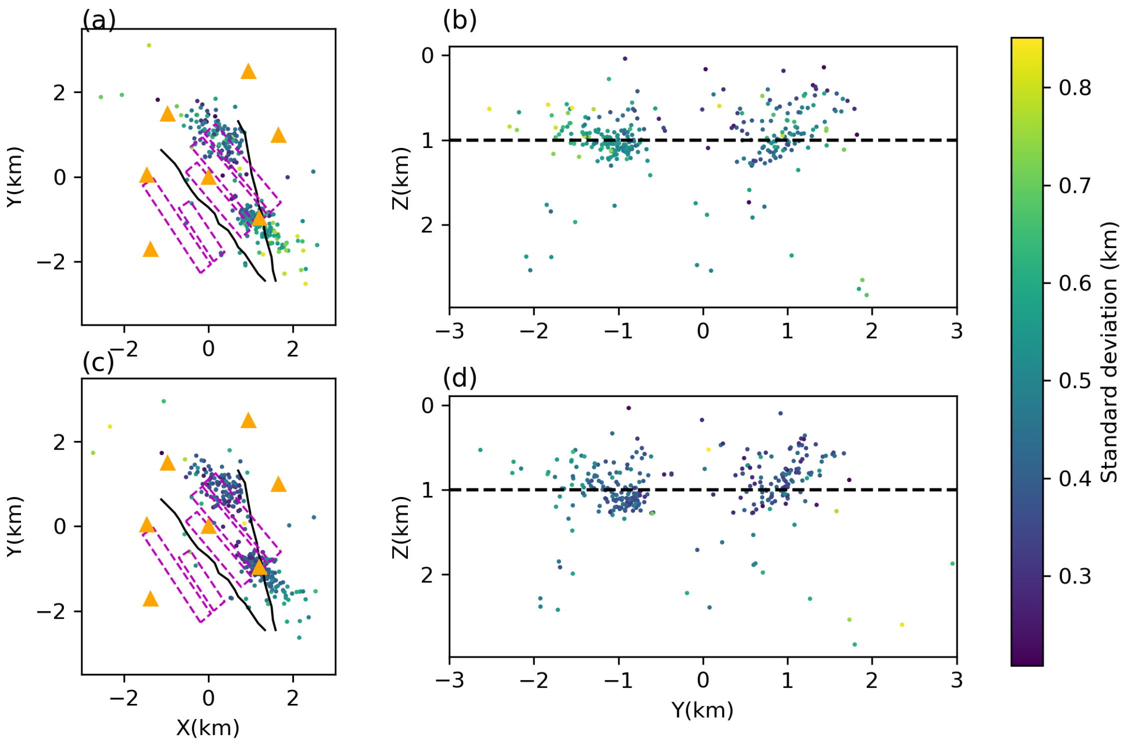

Figure 4. Source location results. **(a)** and **(b)** are map view and a latitudinal cross-section of source locations obtained using body wave travel time data only. **(c)** and **(d)** are map view and a cross-section of source locations obtained using both body wave travel time data and surface wave dispersion data. The orange triangles show the location of stations. The color of each dot reflects the standard deviations of each source location. Black lines show faults that appear in the study area obtained from Bishop et al. (1993). The purple boxes show the location of coal seams located between 800 and 900 m depth.

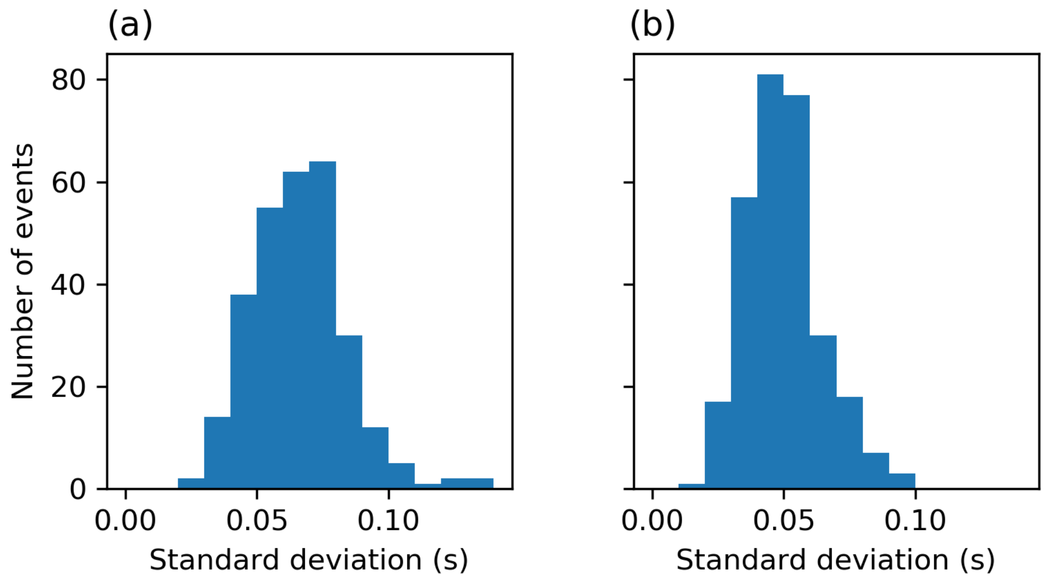

Figure 5. Histograms of the standard deviations of source origin time obtained using **(a)** body wave tomography and **(b)** joint body and surface wave inversion.

from BGS (majority $<$ 1 km). The results show two clusters: one in the southeast and the other in the north. The southeastern cluster has slightly higher uncertainties than those in the north, which is probably caused by the fact that the stations are distributed to one side of the southeastern cluster. Compared to the standard deviation from body wave inversion (around 0.5 km), the location results from joint inversion show lower uncertainties (around 0.4 km). This suggests that by including even only 12 surface wave dispersion curves the event locations can be better constrained since dispersion data provides additional information about the velocity model between stations. Figure 5a and b show histograms of the standard deviations of source origin time obtained using body wave tomography and joint inversion respectively. Most standard deviations from body wave tomography are higher than 0.05 s, while those from joint inversion are centred around 0.05 s. Therefore, by including surface wave dispersion data in the inversion, the source origin time can also be better constrained since this helps to resolve the trade off between origin time and velocity structure.

Verdon et al. (2017) showed that the seismicity is directly induced by the mining, as opposed to being caused by activation of pre-existing tectonic features due to stress transfer. However, in our results events of the southern cluster occur at the end of and beyond the coal seam, which

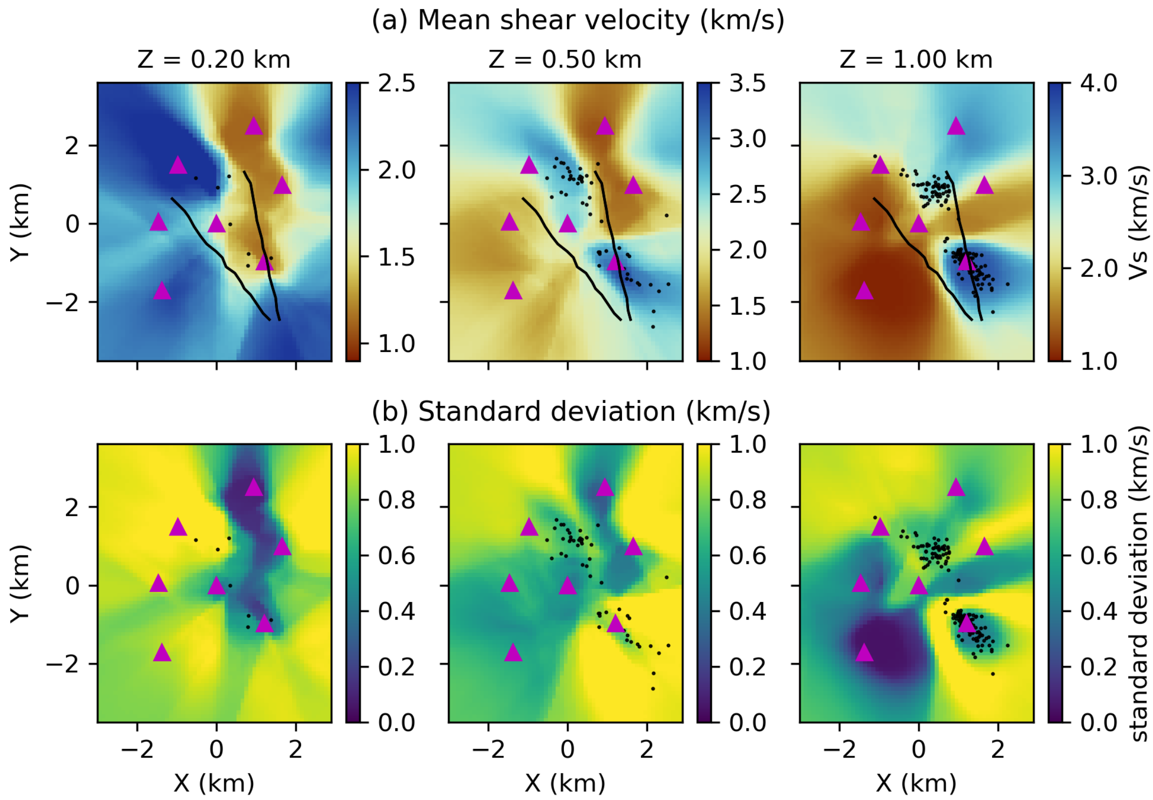

Figure 6. Horizontal slices through the 3D shear velocity model at depth of 0.2 km (left), 0.5 km (middle) and 1.0 km (right) obtained using body wave travel time data only. The upper panels **(a)** and the bottom panels **(b)** show the mean velocity maps and standard deviation maps, respectively. At each slice events within 0.2 km of the depth are plotted. Black lines show faults that appear in the study area.

suggests that those events might not be directly induced by the mining. Since in our results and the results of Verdon et al. (2017) the events of the souther cluster occur at greater depths than the coal seam and there is no correlation between the rate of excavation and the rate of seismicity (Verdon et al. 2017), it is possible that the events of the southern cluster can be caused by activation of pre-existing tectonic features, for example, the fault that crosses the cluster (Figure 2).

4.2 Velocity models

Figure 6 shows horizontal slices through the 3D mean and standard deviation maps of shear wave velocity at depths of 0.2 km, 0.5 km and 1.0 km obtained using body wave travel time data only. The standard deviation map at 0.2 km shows that only a small part of the model is well constrained, which is associated with lower velocities in the mean velocity map. Most of the other maps show the same values as the standard deviation of the prior pdf, suggesting that body waves

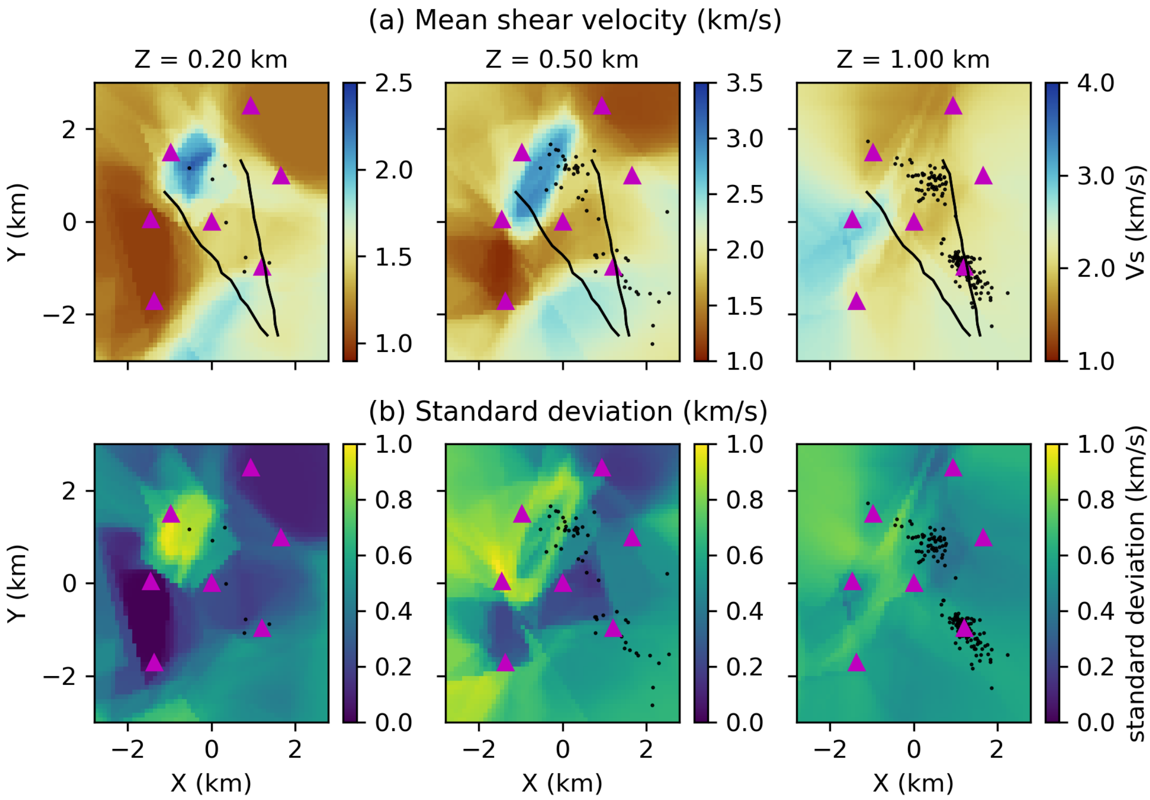

Figure 7. Horizontal slices through the 3D shear velocity model obtained using surface wave dispersion data only. Key as in Figure 6.

offer very limited information about the near surface as expected. The mean velocity map at 0.5 km depth shows that the shear velocities in the southwest and northeast are lower than elsewhere. The standard deviation map suggests that most of the structure within the boundary of the array is reasonably well constrained by the data, other than in the southeast which has higher velocities and higher uncertainties, probably caused by the limited data coverage in that area. At 1 km depth the mean velocity map shows that the velocity in the west is lower than the east, the northern earthquake cluster occurs at the boundary of velocity anomalies, and the southeastern earthquake cluster is correlated with a clearly-defined high velocity anomaly. Between the two clusters there are low velocity anomalies. The standard deviation map shows very low uncertainties (< 0.2 km/s) in the southwest associated with the low velocity anomaly, which suggests that the low velocity anomaly is well constrained, whereas slightly higher uncertainties (about 0.4 km/s) are observed elsewhere. There are loops of higher uncertainty around the southeastern high velocity anomaly

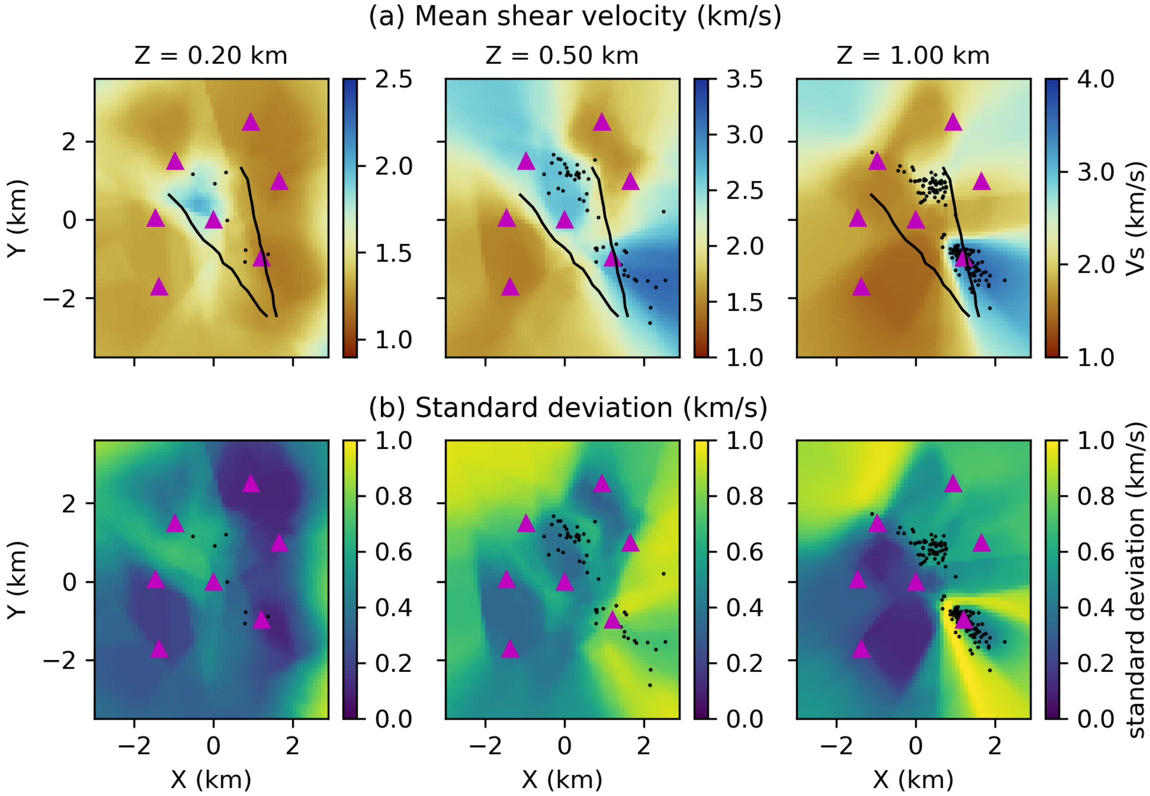

Figure 8. Horizontal slices through the 3D shear velocity model obtained from joint body and surface wave inversion. Key as in Figure 6.

and around the low velocity anomaly between the two clusters. These loops occur where there are strong velocity gradients or velocity contrasts whose locations are not well constrained. They rep-

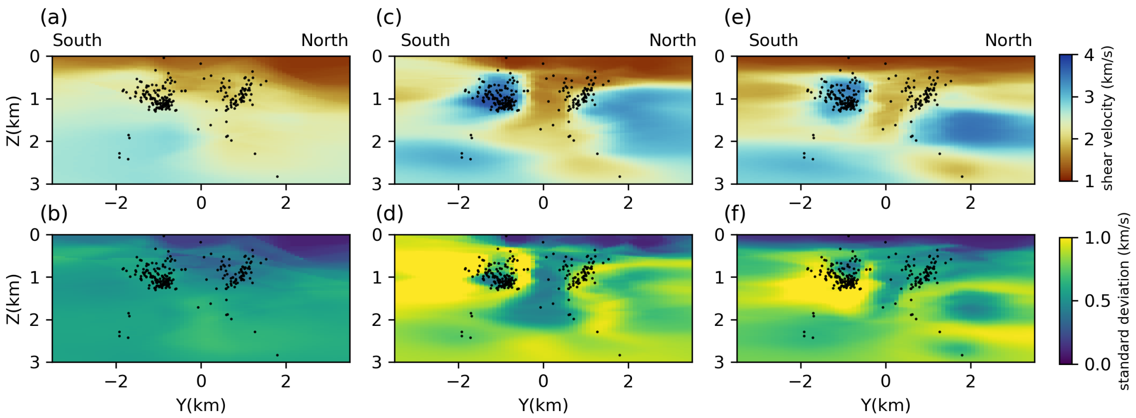

Figure 9. Vertical Cross sections of the mean (top) and standard deviation (bottom) of shear wave velocity at X=1 km obtained using surface wave tomography (**a** and **b**), body wave tomography (**c** and **d**) and joint body and surface wave inversion (**e** and **f**). Black dots are events lying within 0.8 km of the cross-section.

resent uncertainty due to the trade-off between the velocity and the location of velocity anomalies, and hence describe uncertainty in the anomalies' shapes (Galetti et al. 2015; Zhang et al. 2018).

Figure 7 shows horizontal slices through the mean and standard deviation obtained from surface wave tomography at the same depths as above. Compared to the results from body wave tomography, the mean shear velocity map at 0.2 km shows lower velocities (about 1.0 km/s) than the results from body wave tomography ($>$ 2.0 km/s), and the standard deviation is also much lower (about 0.2 km/s) than that from body wave tomography (about 1.0 km/s). This is due to the fact that surface waves are more sensitive to near surface structure than body waves. There is a higher velocity anomaly in the northwest which is probably caused by poor data coverage (Figure 2). At 0.5 km depth the mean velocity map shows similar patterns of structure to those obtained from body wave tomography: the velocity in the southwest and in the northeast is lower and the velocity in the southeast is higher. The mean velocity map at 1 km depth shows very different results compared to those from body wave tomography and its standard deviation is higher (about 0.6 km/s). This is probably caused by the fact that the frequency range of the surface waves used in the inversion has very low sensitivity at this depth.

Figure 8 shows horizontal slices through the mean and standard deviation obtained using joint inversion. Similarly to the results of surface wave tomography, the mean velocity map at 0.2 km shows lower velocity values than those from body wave tomography with lower standard deviations: near surface structure can be better constrained by including surface wave dispersion data in the inversion. There is still a higher velocity anomaly between Y=0 km and Y=2 km which is associated with high standard deviations: neither body waves nor surface waves have much resolution in this area so the velocity tends towards the mean of the prior pdf (2.5 km/s). The mean velocity maps at 0.5 km and 1.0 km are very similar to the results from body wave tomography: we have more body wave data than surface wave data that are sensitive to these depths so the body wave data dominate the results. Nevertheless, the velocity magnitudes are slightly different from the results of body wave tomography which is due to the contribution of surface waves, and the standard deviation map shows lower uncertainties within the station array which suggests that surface waves improve the resolution across that entire area. Similarly to the results of body wave tomography, the standard deviation map also shows a higher uncertainty loop around the southeastern high velocity anomaly.

Note that all of the standard deviation models show lower uncertainties than those of the prior pdf in the area outside of the station array where there is no obvious data. This is probably because the velocity in this area is correlated with the velocity within the station array through large Voronoi cells, but also for some models surface and body wave ray paths may assume trajectories that travel outside of the array. This phenomenon has also been observed in several previous studies (Galetti et al. 2015; Zhang et al. 2018; Zhang & Curtis 2020a).

In Figure 9 we show vertical cross sections through the mean and standard deviation maps from the three inversions along the X=1 km profile which lies between the two earthquake clusters. The mean velocity model from surface wave tomography (Figure 9a) shows that there is a low velocity anomaly between the two clusters. The standard deviation model (Figure 9b) shows that the near surface structure (< 0.8 km) is well constrained while the deeper part has very limited resolution. Figure 9c and d show the mean and standard deviation cross sections from body wave tomography. The velocity model also shows a low velocity anomaly between the two clusters, however the low velocity anomaly extends to deeper levels and the velocity at either side of the low velocity anomaly is much higher ($>$ 3km/s) than that from surface wave tomography ($\sim$ 2.0 km/s). The standard deviation model shows a low uncertainty area associated with the middle low velocity anomaly suggesting that the anomaly is well determined. There are also higher uncertainty loops around the high velocity anomalies at the two sides of the low velocity anomaly. Figure 9e and f show the results from joint inversion. The mean model is similar to that from surface wave tomography at shallow levels, and to that from body wave tomography at depth. However the velocity magnitude of the southern high velocity anomaly is lower than that from body wave tomography, and the velocity of the northern low velocity anomaly around 1 km depth is much lower than that from body wave tomography, both due to the contribution of surface waves. Similarly the standard deviation model shows lower uncertainties in the near surface, and higher uncertainty loops around high velocity anomalies.

Appendix A shows P-wave velocity models from both body-wave only inversion and joint inversion. The key finding of those results is that the addition of surface waves also helps to constrain P-wave velocities even though surface waves are not directly very sensitive to P-wave velocity. This is because P-wave velocities are correlated with shear velocities through the source locations and the latter are better estimated with the addition of surface wave data.

For one chain the body-wave only inversion takes 396 hours when parallelized using 9 cores, whereas the joint inversion takes 502 hours using the same number of cores. Therefore the joint inversion requires only 27% more computational time than the body-wave only inversion, while producing source locations and velocity models with notably lower uncertainties.

4.3 Synthetic tests

In the above results there is a high velocity anomaly at the location of the southern earthquake cluster (Figure 6, 8, 9): in the results from joint inversion the magnitude of the velocity anomaly is slightly lower, but is nevertheless clearly identifiable. Similar features have been observed previously and are generally interpreted as earthquake asperities that concentrate stress (Lees 1990; Eberhart-Phillips & Michael 1998; Chiarabba & Amato 2003; Tajima et al. 2009; Li et al. 2013; Zhang et al. 2013). However it is also possible that this correlation is caused by the trade off between source parameters and velocity values.

To better understand the correlation of the high velocity anomaly and the earthquake cluster we performed a simple synthetic test in which the ”true” model contains three horizontal layers and event locations are taken to be their mean values from the joint inversion above (Figure 10). We computed synthetic versions of the same body wave travel time data and surface wave dispersion data as used in the above inversion, and added 1 percent noise to the data. We then conducted three inversions: simultaneous inversion for source parameters and velocity model using only body wave data, inversion for velocity only using body wave data with sources fixed at their true locations, and joint inversion for sources and velocities using both types of data. The initial source parameters (event locations and origin times) are the same as were used in the real data inversions above. The prior pdfs are chosen to be the same as in the real data example except that the prior for the number of cells is chosen to be a Uniform distribution between 5 and 100 since the model is

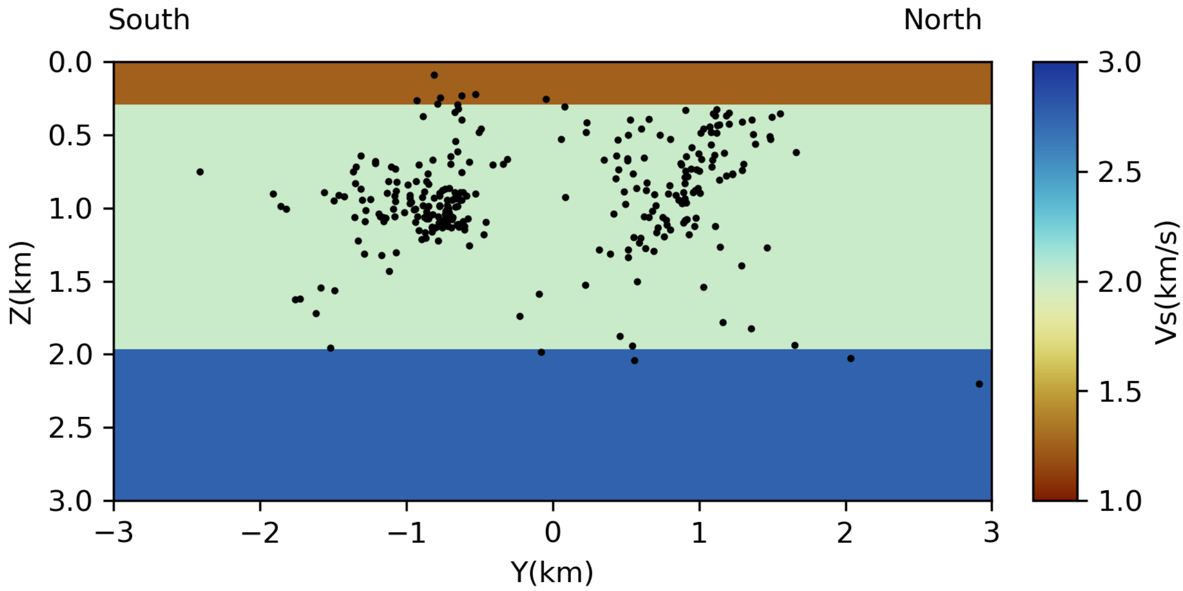

Figure 10. Cross section through the synthetic model at X=1 km. Black dots show the event locations which are taken from joint inversion of the real data in Figure 4d.

relatively simple. The proposal pdfs are also tuned in the same way as above. The burn-in and total samples for each chain and the number of chains are also set to be the same as in the real data inversions.

Figure 11 shows cross sections through the mean and standard deviation at X=1 km obtained using only body waves by simultaneously inverting for source parameters and the velocity model. Though the mean velocity model shows three layers which are to some extent similar to the true model, the velocity value around the southern cluster (around 1.75 km/s) is lower than the true value (2.0 km/s). This suggests that body wave tomography may produce biases in the results around the location of event clusters, caused by the trade off between event locations and velocity values (see Figure 12): shallower event depths are generally associated with a lower velocity value to fit the data. The standard deviation model shows low uncertainties from the surface down to around 1.5 km including in the low velocity areas around the southern cluster. This low velocity anomaly is due to the fact that the initial source locations are shallower than the true locations, so in order to fit the data the model decreases the velocity value at the location of event clusters (see Figure 12); this trade off creates complex multimodality in the posterior pdf (see Figure 16a, b and c below), and since random walk McMC algorithms are generally inefficient for multimodal

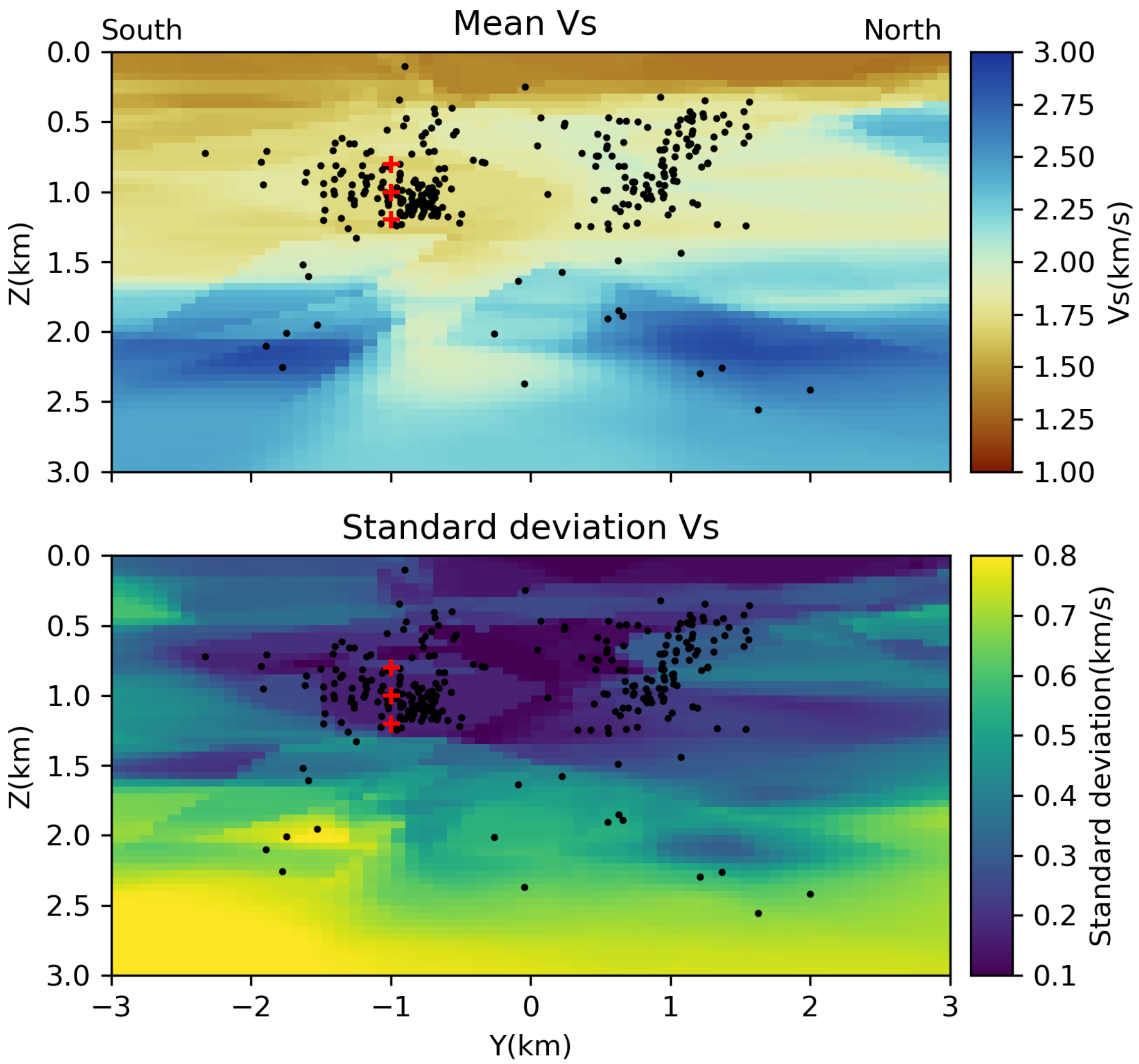

Figure 11. Cross sections of the mean and standard deviation at X=1 km obtained by inverting for source parameters and velocity model simultaneously using body wave data only. Black dots show the mean event locations. The red pluses show point locations which are referred to in the text.

distributions the chains likely get stuck in modes that have lower velocities. By contrast, in Figure 13 we show the results obtained from an inversion with source parameters fixed at their true values. The mean velocity model shows almost the same structure as the true model which again suggests that the non-uniquenesses in the posterior velocity pdfs in the previous results are caused by non-linear trade off between source parameters and velocity values. The standard deviation shows very low uncertainties (< 0.2 km/s) across the whole section except in a small area in the left

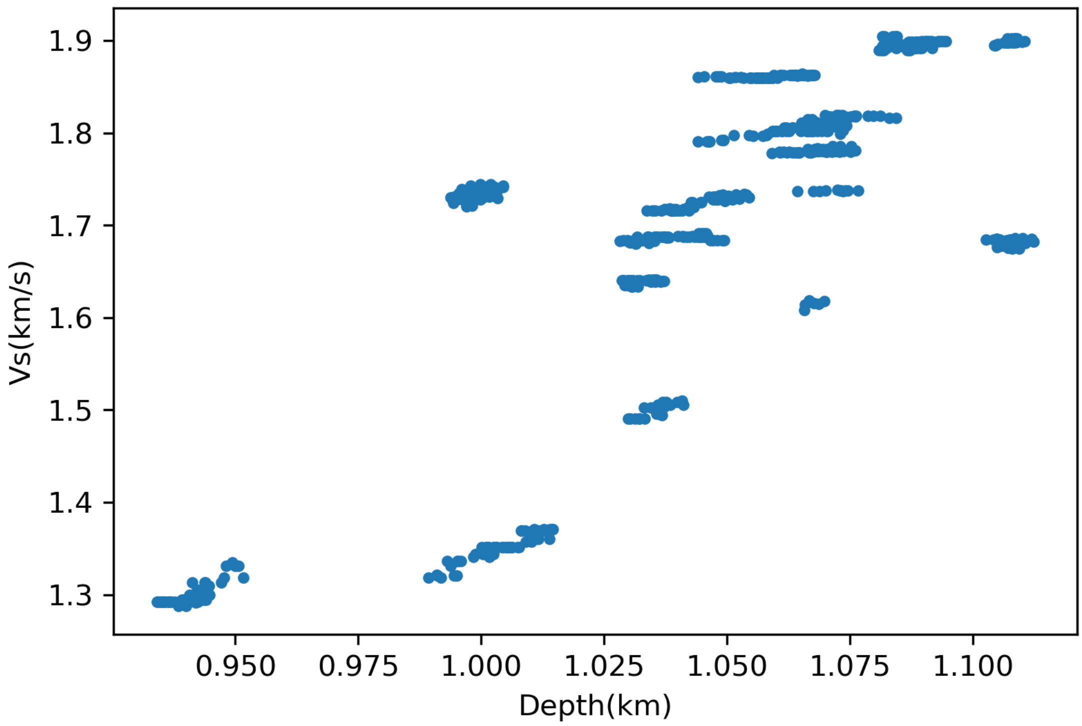

Figure 12. Average shear velocity at the location of the southern cluster versus average depth of events of the southern cluster.

corner where there are no events. It also shows slightly higher uncertainties at the boundaries between layers which reflect uncertainty in layer boundary locations similar to the uncertainty loops observed above and in Zhang et al. (2018). To give an idea of fit to the data, the simultaneous inversion of source parameters and velocity model produces an average residual of 0.81 s while the fixed-source inversion produces a residual of 0.80 s. Thus the two inversions produce almost the same average fit to the data even though they give different estimates of the velocity model; therefore one cannot discriminate between the two models based on data fit.

Figure 14 shows cross sections of the mean and standard deviation obtained using joint inversion of both data types for both velocities and source parameters. Though the mean velocity model shows slightly different velocity values in the second and third layer compared to the previous two models, it is significantly closer to the true model than that obtained from body wave tomography by simultaneously inverting for source parameters and the velocity model. The standard deviation model shows similar structures to those from the fixed-source inversion, including

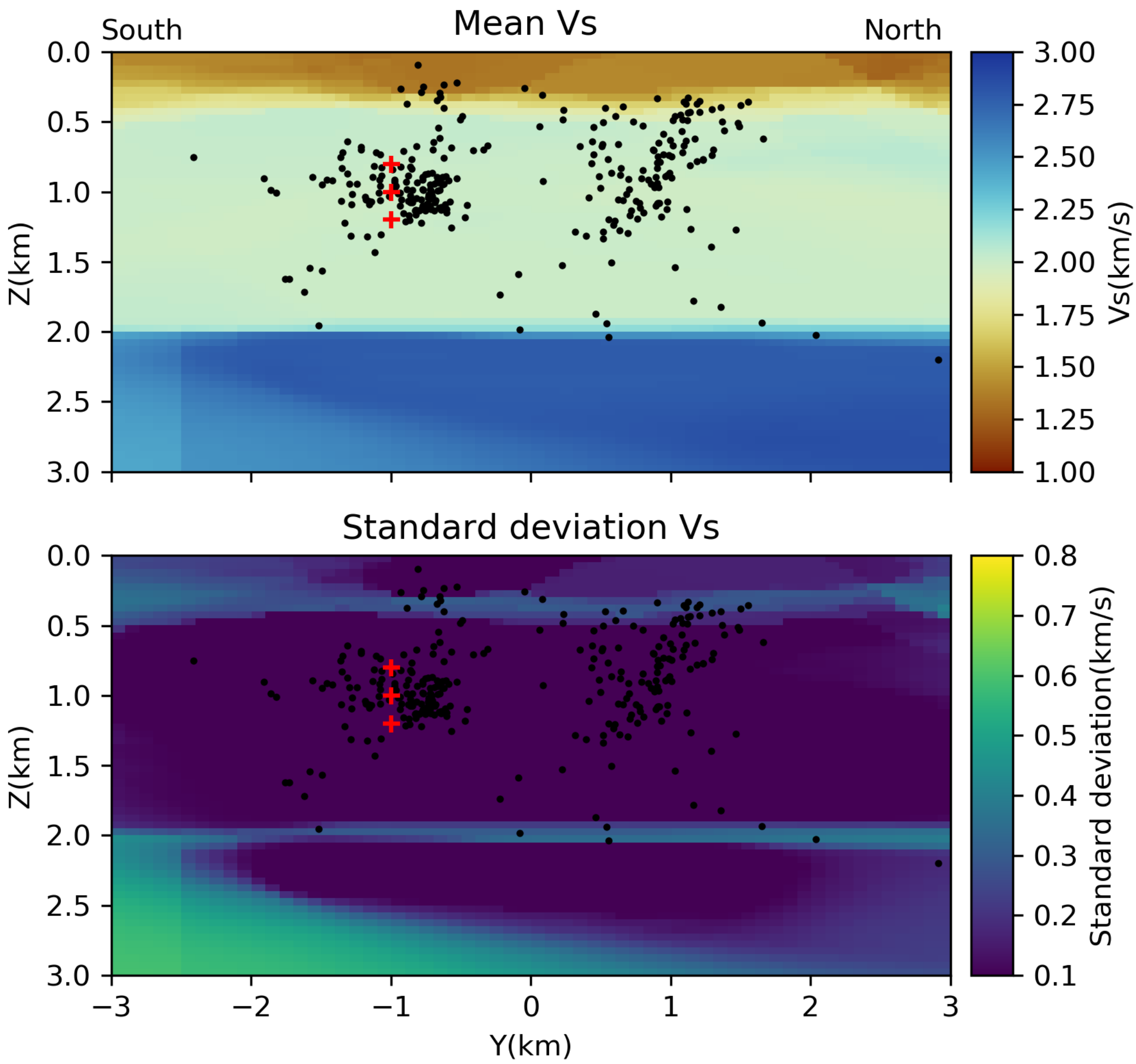

Figure 13. Cross sections of the mean and standard deviation at X=1 km obtained by inverting for velocities, and fixing the source parameters at their true values. Black dots show the event locations. Red pluses show point locations which are referred to in the text.

higher uncertainties at the boundary of layers. Figure 15 shows the true dispersion curve and the average dispersion curve calculated using the mean velocity model. At longer periods ($>$ 1.2 s) the average group velocities fit the true values, whereas at shorter periods the average group velocities are higher than true values which is probably caused by the bias produced by body wave data (see Figure 11), or because the mean model is not a good representation of the subsurface structure (Zhang & Curtis 2020a).

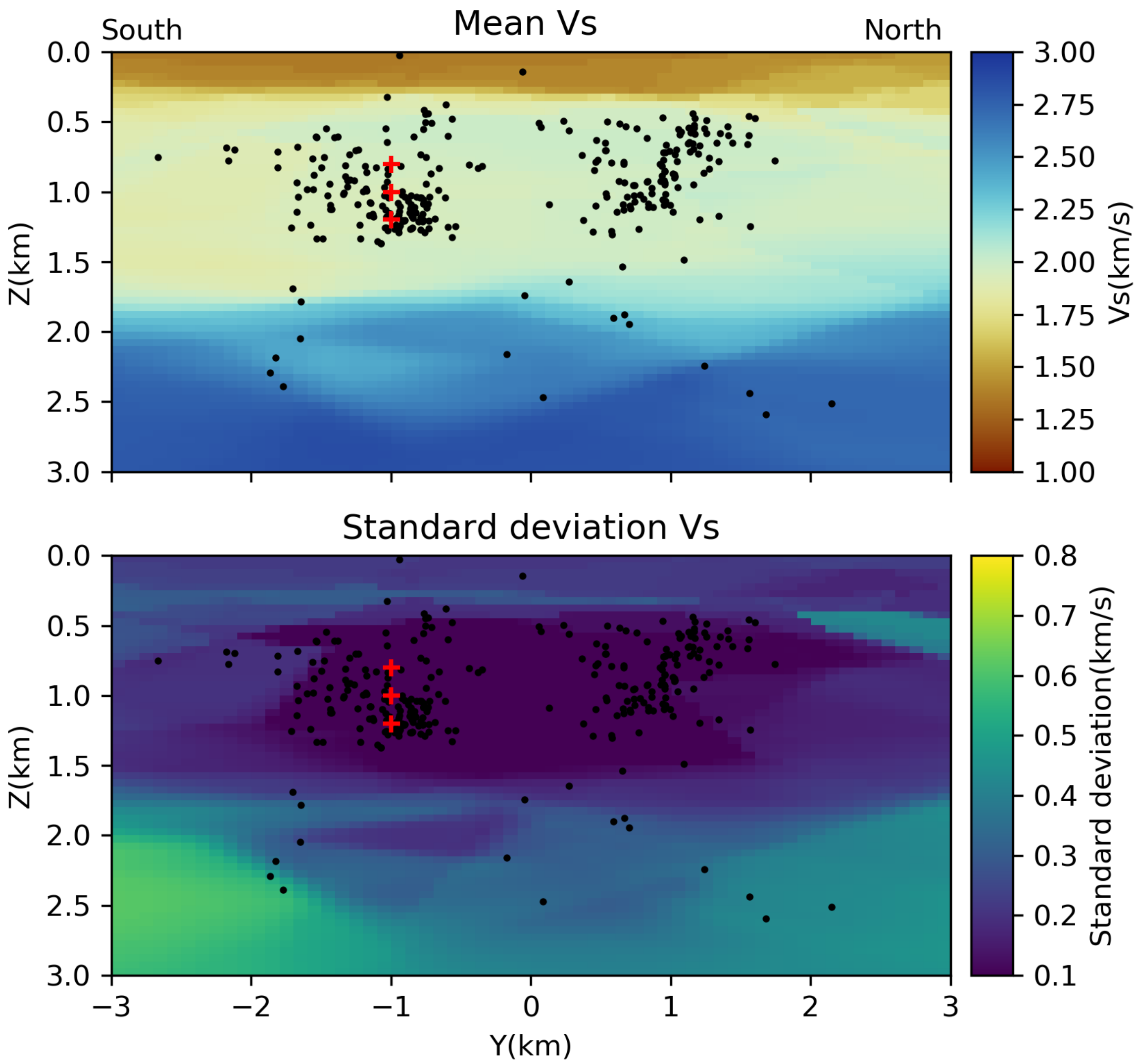

Figure 14. Cross sections of the mean and standard deviation at X=1 km obtained by inverting for both velocity and event locations using both body wave and surface wave data. Black dots show the mean event locations. Red pluses show point locations which are referred to in the text.

To better understand the results, in Figure 16 we show marginal posterior pdfs obtained using the three methods at three different points (1, -1, 0.84 km), (1, -1, 1.0 km) and (1, -1, 1.2 km) which cross the southern earthquake cluster in the above cross sections. The marginal distributions obtained from body wave tomography show complex multimodal distributions (Figure 16a, b and c) and are distributed away from the true value (2 km/s). By contrast, most of the marginal distributions obtained from joint inversion show a unimodal distribution concentrated around the

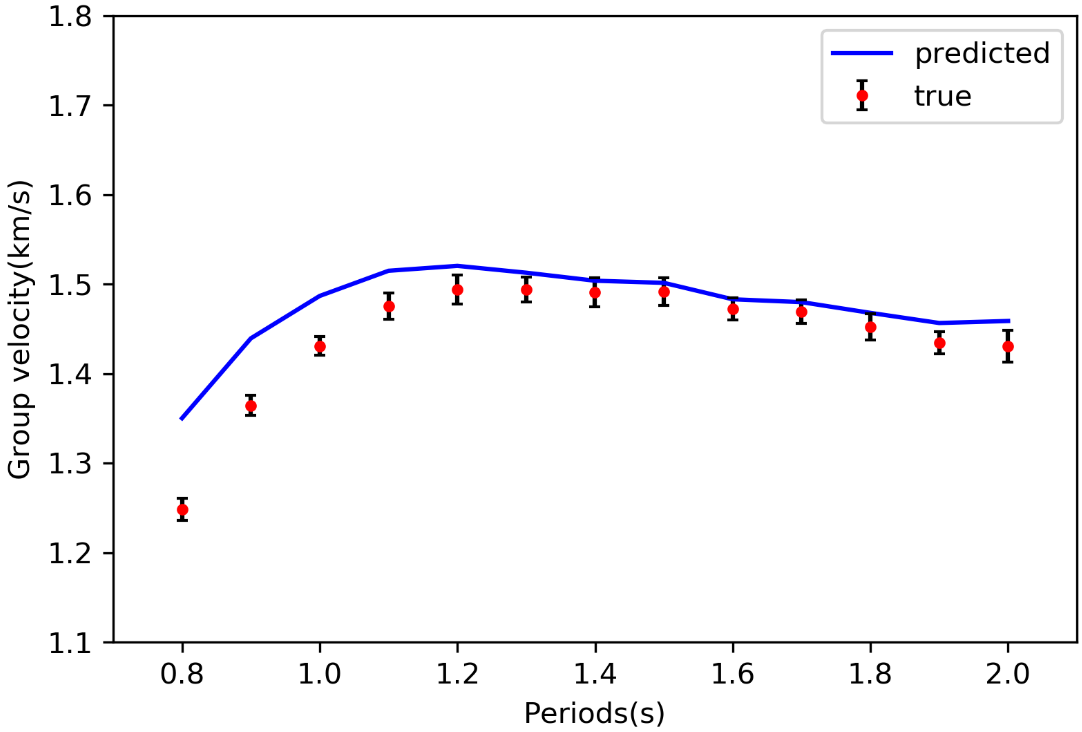

Figure 15. Group velocities used in the joint inversion (red dot) plotted with error bars and the average dispersion curve calculated from the mean velocity model (blue line). Since the true model is a 1D model, dispersion curves between different receiver pairs are almost the same except for random noise. Error bars show the standard deviation of group velocities of different dispersion curves.

true value (Figure 16 e and f) other than a remaining multimodality in Figure 16d. The marginal distributions obtained from fixed-source inversion focus to a unimodal distribution around the true value (Figure 16g, h and i). Thus, the simultaneous inversion for source parameters and velocity model using only body wave data can bias the results due to the trade off between source parameters and velocity values. By including surface wave dispersion data in the inversion, this problem can be resolved since surface wave data improve the velocity estimate.

Apart from joint inversion using both types of data, the results obtained using surface wave tomography are frequently used as prior information for body wave tomography to produce a more realistic velocity model. For example, velocity models from surface wave tomography were used as a starting model for body wave tomography by Rawlinson & Fishwick (2012) and Nunn et al. (2013). However in those studies since surface wave data are not used in the later inversion, trade-offs between source parameters and velocity models could still bias the results and the resulting

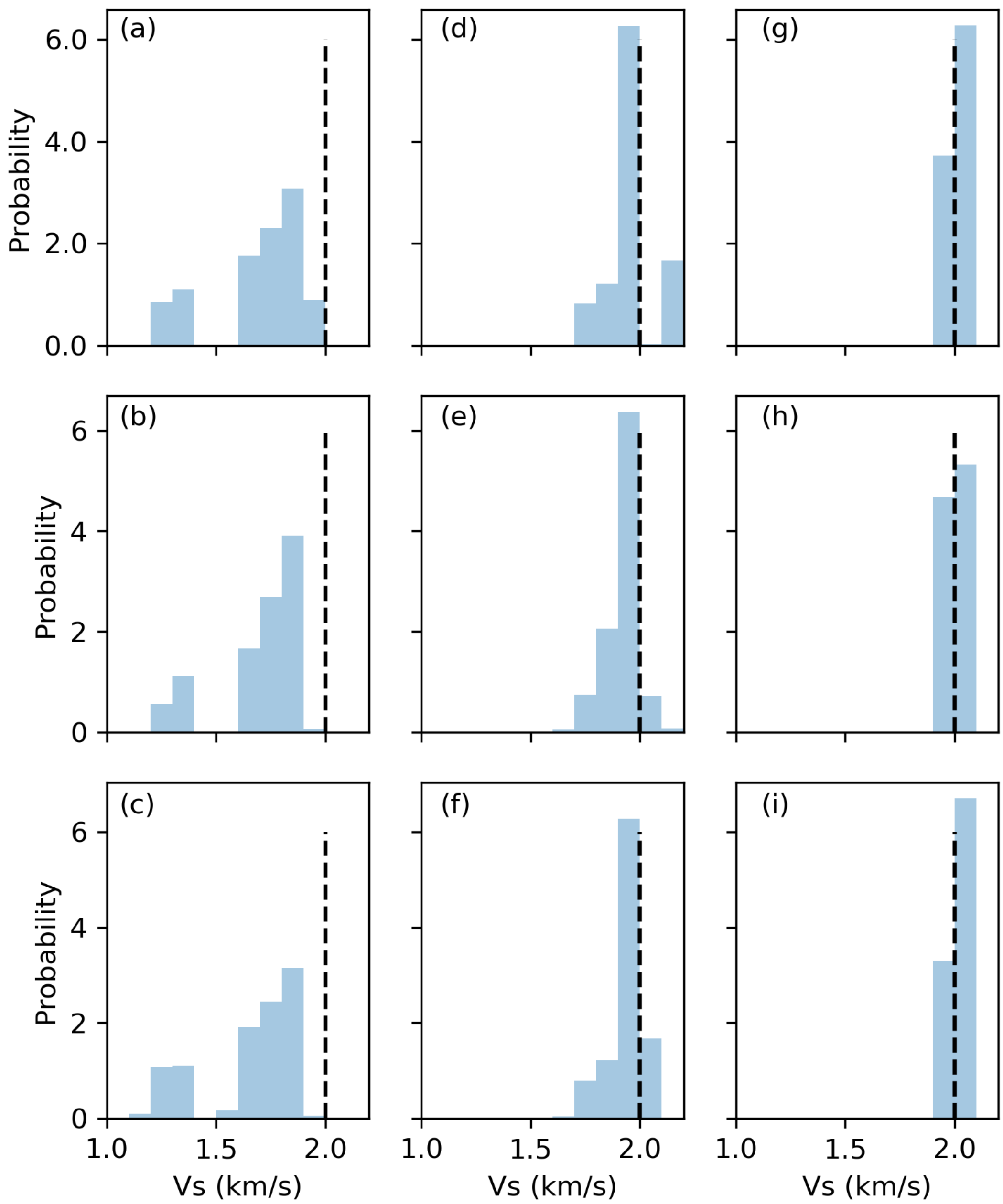

Figure 16. The marginal posterior pdfs of shear velocity at three points (pluses in Figure 11, 13 and 14). **(a)**, **(b)** and **(c)** show the marginal posterior pdfs at three points at depth 0.8 km, 1.0 km and 1.2 km obtained by inverting source parameters and velocity model simultaneously using body wave data. **(d)**, **(e)** and **(f)** show the marginal posterior pdfs at three points obtained by joint inversion using both body wave data and surface wave data. **(g)**, **(h)** and **(i)** show the marginal posterior pdfs at three points obtained by fixing the source parameters at their true values. The dashed black line denotes the true shear velocity value.

model can be inconsistent with surface wave data. Thus it is better to invert for a unified model of velocity and source locations jointly using both types of data.

In the real data results, the high velocity anomaly at the location of the southern cluster therefore may reflect the true structure of the subsurface, e.g., earthquake asperities following previous interpretations (Lees 1990; Eberhart-Phillips & Michael 1998; Chiarabba & Amato 2003; Tajima et al. 2009; Li et al. 2013; Zhang et al. 2013). However, since we still observe subtle multimodilities in the joint inversion results, and the real Earth may have a more complex structure, there is still the possibility that the details of the recovered model are obscured by the trade off between source parameters and velocity models. The synthetic test suggests that the trade off mainly affects the velocity structure at the location of the southern earthquake cluster, so our results at least remain valid for most of the subsurface.

5 DISCUSSION

In this study we used Voronoi cells to parametrize the subsurface. While the Voronoi parameterization is effective if we wish to image discontinuities, it can introduce difficulties when we attempt to recover a smooth model (Hawkins et al. 2019). For example, in our results there remain some signs of Voronoi cell shapes which we choose to leave un-smoothed (in contrast to other studies that use fewer McMC iterations and heavy smoothing - e.g. Young et al. 2013; Crowder et al. 2019a). It has also been found that the Voronoi parametrization can cause multimodalities in the posterior pdf, and produce uncertainty estimates that differ from those that one would normally associate with a pixelated image (Zhang & Curtis 2020a), both of which make interpretation of uncertainties difficult. To produce a smoother model, other parametrizations, such as wavelets (Hawkins & Sambridge 2015), Johnson-Mehl tessellation (Belhadj et al. 2018) and Delaunay and Clough-Tocher parametrization (Hawkins et al. 2019) may be used.

In this study we did not take into account any anisotropy that may exist in the area. This may cause errors in our results. However, as Verdon et al. (2017) showed that the anisotropy is not particularly strong in the area, our results should remain valid as a first-approximation. In future it is possible to include anisotropy in our method to produce more accurate results.

The rj-McMC algorithm is generally not efficient for exploring complex multimodal distributions (Green & Hastie 2009). In our body-wave only tomography synthetic test, by simultaneously inverting for source parameters and velocity models the chains may have got stuck at local modes and failed to find the true solution. To reduce this issue one could use a better initial model (as required by linearised inversion) if available to ensure that the chains explore around the globally optimal solution. In the real data inversions we used initial source parameters obtained using a 1D initial model, so to further improve the results one could adopt locations found using a good 3D model instead. If no better model exists (as is often the case) then that is not an option, so methods that are more effective for multimodal distributions might be used, for example grid search (Sen & Stoffa 2013), non-Markovian importance sampling (Lomax & Curtis 2001; Lomax et al. 2009), parallel tempering (Sambridge 2013) and variational inference methods (Zhang & Curtis 2020a,b).

6 CONCLUSION

We implemented a Monte Carlo method to perform joint inversion using both body wave arrival time data and surface wave dispersion data, and applied it at a mining site located to the north of New Ollerton, Nottinghamshire, U.K, at which induced seismicity occurs. The results show that by including surface wave dispersion data the shallow structure can be better constrained because surface waves are generally sensitive to the shallow structure, and this further improves estimate of source parameters. We also observed a high velocity anomaly at the location of one of the event clusters which may be related to earthquake asperities that concentrate stress. To further understand this correlation, we performed inversions using synthetic data generated using the same source and receiver distribution as in the real data experiment. The results show that due to the trade off between source parameters and velocity values, the inversion using only body wave data can produce biases; by including surface wave dispersion data in the inversion the problem can be resolved. We conclude that it is better to include surface wave data in seismic travel time tomographic inversions for velocity structure and earthquake source locations.

ACKNOWLEDGMENTS

The authors thank the Edinburgh Interferometry Project sponsors (Equinor, Schlumberger and Total) for supporting this research. CR and AN were supported by NERC standard grant NE/001154/1. This work used the Cirrus UK National Tier-2 HPC Service at EPCC (http://www.cirrus.ac.uk). The data used in this study can be accessed from British Geological Survey (`https://earthquakes.bgs.ac.uk/monitoring/data_archive.html`).

REFERENCES

Aki, K. & Lee, W., 1976. Determination of three-dimensional velocity anomalies under a seismic array using first P arrival times from local earthquakes: 1. a homogeneous initial model, *Journal of Geophysical research*, **81**(23), 4381–4399.

Belhadj, J., Romary, T., Gesret, A., Noble, M., & Figliuzzi, B., 2018. New parameterizations for bayesian seismic tomography, *Inverse Problems*, **34**(6), 065007.

Bensen, G., Ritzwoller, M., Barmin, M., Levshin, A., Lin, F., Moschetti, M., Shapiro, N., & Yang, Y., 2007. Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements, *Geophysical Journal International*, **169**(3), 1239–1260.

Bishop, I., Styles, P., & Allen, M., 1993. Mining-induced seismicity in the nottinghamshire coalfield, *Quarterly Journal of Engineering Geology and Hydrogeology*, **26**(4), 253–279.

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).

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

Butcher, A., Luckett, R., Verdon, J. P., Kendall, J.-M., Baptie, B., & Wookey, J., 2017. Local magnitude discrepancies for near-event receivers: Implications for the UK traffic-light scheme, *Bulletin of the Seismological Society of America*, **107**(2), 532–541.

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.

Chiarabba, C. & Amato, A., 2003. Vp and Vp/Vs images in the Mw 6.0 Colfiorito fault region (central Italy): A contribution to the understanding of seismotectonic and seismogenic processes, *Journal of Geophysical Research: Solid Earth*, **108**(B5).

Crowder, E., Rawlinson, N., Cornwell, D., Sammarco, C., Galetti, E., & Curtis, A., 2019a. Controls on the development and termination of failed continental rifts: Insights from the crustal structure and rifting style of the north sea via ambient noise tomography, *EarthArXiv*.

Crowder, E., Rawlinson, N., Pilia, S., Cornwell, D., & Reading, A., 2019b. Transdimensional ambient noise tomography of Bass Strait, southeast Australia, reveals the sedimentary basin and deep crustal structure beneath a failed continental rift, *Geophysical Journal International*, **217**(2), 970–987.

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

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.

Dziewonski, A., Bloch, S., & Landisman, M., 1969. A technique for the analysis of transient seismic signals, *Bulletin of the seismological Society of America*, **59**(1), 427–444.

Dziewonski, A. M. & Woodhouse, J. H., 1987. Global images of the Earth's interior, *Science*, **236**(4797), 37–48.

Eberhart-Phillips, D. & Michael, A. J., 1998. Seismotectonics of the Loma Prieta, California, region determined from three-dimensional Vp, Vp/Vs, and seismicity, *Journal of Geophysical Research: Solid Earth*, **103**(B9), 21099–21120.

Fang, H., Zhang, H., Yao, H., Allam, A., Zigone, D., Ben-Zion, Y., Thurber, C., & van der Hilst, R. D., 2016. A new algorithm for three-dimensional joint inversion of body wave and surface wave data and its application to the Southern California plate boundary region, *Journal of Geophysical Research: Solid Earth*, **121**(5), 3557–3569.

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.

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

Hawkins, R., Bodin, T., Sambridge, M., Choblet, G., & Husson, L., 2019. Trans-dimensional surface reconstruction with different classes of parameterization, *Geochemistry, Geophysics, Geosystems*, **20**(1), 505–529.

Herrin, E. & Goforth, T., 1977. Phase-matched filters: application to the study of Rayleigh waves, *Bulletin of the Seismological Society of America*, **67**(5), 1259–1275.

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.

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.

Lees, J. M., 1990. Tomographic P-wave velocity images of the Loma Prieta earthquake asperity, *Geophysical Research Letters*, **17**(9), 1433–1436.

Levshin, A. & Ritzwoller, M., 2001. Automated detection, extraction, and measurement of regional surface waves, in *Monitoring the Comprehensive Nuclear-Test-Ban Treaty: Surface Waves*, pp. 1531–1545, Springer.

Levshin, A., Ratnikova, L., & Berger, J., 1992. Peculiarities of surface-wave propagation across central Eurasia, *Bulletin of the Seismological Society of America*, **82**(6), 2464–2493.

Levshin, A. L., Pisarenko, V., & Pogrebinsky, G., 1972. On a frequency-time analysis of oscillations, in *Annales de Geophysique*, vol. 28, pp. 211–218, Centre National de la Recherche Scientifique.

Li, Z., Tian, B., Liu, S., & Yang, J., 2013. Asperity of the 2013 lushan earthquake in the eastern margin of Tibetan Plateau from seismic tomography and aftershock relocation, *Geophysical Journal International*, **195**(3), 2016–2022.

Lin, F.-C., Ritzwoller, M. H., & Snieder, R., 2009. Eikonal tomography: surface wave tomography by phase front tracking across a regional broad-band seismic array, *Geophysical Journal International*, **177**(3), 1091–1110.

Liu, X. & Zhao, D., 2016. P and S wave tomography of Japan subduction zone from joint inversions of local and teleseismic travel times and surface-wave data, *Physics of the Earth and Planetary Interiors*, **252**, 1–22.

Lomax, A. & Curtis, A., 2001. Fast, probabilistic earthquake location in 3d models using oct-tree importance sampling, in *Geophys. Res. Abstr*, vol. 3, p. 955.

Lomax, A., Michelini, A., & Curtis, A., 2009. Earthquake location, direct, global-search methods, *Encyclopedia of complexity and systems science*, pp. 2449–2473.

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.

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

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.

Nunn, C., Roecker, S. W., Tilmann, F. J., Priestley, K. F., Heyburn, R., Sandvol, E. A., Ni, J. F., Chen, Y. J., Zhao, W., & Team IV, t. I., 2013. Imaging the lithosphere beneath NE Tibet: teleseismic P and S body wave tomography incorporating surface wave starting models, *Geophysical Journal International*, **196**(3), 1724–1741.

Obrebski, M., Allen, R. M., Pollitz, F., & Hung, S.-H., 2011. Lithosphere–asthenosphere interaction beneath the western United States from the joint inversion of body-wave traveltimes and surface-wave phase velocities, *Geophysical Journal International*, **185**(2), 1003–1021.

Obrebski, M., Allen, R. M., Zhang, F., Pan, J., Wu, Q., & Hung, S.-H., 2012. Shear wave tomography of China using joint inversion of body and surface wave constraints, *Journal of Geophysical Research: Solid Earth*, **117**(B1).

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. & Fishwick, S., 2012. Seismic structure of the southeast Australian lithosphere from surface and body wave tomography, *Tectonophysics*, **572**, 111–122.

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.

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.

Roecker, S., Ebinger, C., Tiberi, C., Mulibo, G., Ferdinand-Wambura, R., Mtelela, K., Kianji, G., Muzuka, A., Gautier, S., Albaric, J., et al., 2017. Subsurface images of the Eastern Rift, Africa, from the joint inversion of body waves, surface waves and gravity: investigating the role of fluids in early-stage continental rifting, *Geophysical Journal International*, **210**(2), 931–950.

Russell, D. R., Herrmann, R. B., & Hwang, H.-J., 1988. Application of frequency variable filters to surface-wave amplitude analysis, *Bulletin of the Seismological Society of America*, **78**(1), 339–354.

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

Sen, M. K. & Stoffa, P. L., 2013. *Global optimization methods in geophysical inversion*, Cambridge University Press.

Shen, W., Ritzwoller, M. H., Schulte-Pelkum, V., & Lin, F.-C., 2012. Joint inversion of surface wave dispersion and receiver functions: a Bayesian Monte-Carlo approach, *Geophysical Journal International*, **192**(2), 807–836.

Shen, W., Ritzwoller, M. H., & Schulte-Pelkum, V., 2013. A 3-D model of the crust and uppermost mantle beneath the central and western US by joint inversion of receiver functions and surface wave dispersion, *Journal of Geophysical Research: Solid Earth*, **118**(1), 262–276.

Syracuse, E. M., Maceira, M., Zhang, H., & Thurber, C., 2015. Seismicity and structure of Akutan and Makushin Volcanoes, Alaska, using joint body and surface wave tomography, *Journal of Geophysical Research: Solid Earth*, **120**(2), 1036–1052.

Tajima, R., Tajima, F., & Kato, A., 2009. Seismic structure in and around the source area of the 2004 mid-Niigata, Japan, earthquake: 3-D waveform modelling based on local tomography images, *Geophysical Journal International*, **177**(1), 145–160.

Tarantola, A., 2005. *Inverse problem theory and methods for model parameter estimation*, vol. 89, siam.

Valero-Gomez, A., Gomez, J. V., Garrido, S., & Moreno, L., 2013. The path to efficiency: fast marching method for safer, more efficient mobile robot trajectories, *IEEE Robotics & Automation Magazine*, **20**(4), 111–120.

Verdon, J. P., Kendall, J.-M., Butcher, A., Luckett, R., & Baptie, B. J., 2017. Seismicity induced by longwall coal mining at the thoresby colliery, nottinghamshire, uk, *Geophysical Journal International*, **212**(2), 942–954.

West, M., Gao, W., & Grand, S., 2004. A simple approach to the joint inversion of seismic body and surface waves applied to the southwest US, *Geophysical research letters*, **31**(15).

Yanovskaya, T., Levshin, A., Its, E., Lander, A., Bukchin, B., Barmin, M., & Ratnikova, L., 2012. *Seismic surface waves in a laterally inhomogeneous Earth*, vol. 9, Springer Science & Business Media.

Yao, H., van Der Hilst, R. D., & De Hoop, M. V., 2006. Surface-wave array tomography in SE Tibet from ambient seismic noise and two-station analysis–i. phase velocity maps, *Geophysical Journal International*, **166**(2), 732–744.

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.

Zhang, H., Maceira, M., Roux, P., & Thurber, C., 2014. Joint inversion of body-wave arrival times and surface-wave dispersion for three-dimensional seismic structure around SAFOD, *Pure and Applied Geophysics*, **171**(11), 3013–3022.

Zhang, X. & Curtis, A., 2020a. Seismic tomography using variational inference methods, *Journal of Geophysical Research: Solid Earth*, **125**(4), e2019JB018589.

Zhang, X. & Curtis, A., 2020b. Variational full-waveform inversion, *Geophysical Journal International*, ggaa170.

Zhang, X., Zhang, H., Wang, H., & Pei, S., 2013. Seismic velocity imaging of the aftershock zone of the 2013 Mw 6.6 lushan Earthquake, China, in *AGU Fall Meeting Abstracts*.

Zhang, X., Curtis, A., Galetti, E., & de Ridder, S., 2018. 3-D Monte Carlo surface wave tomography, *Geophysical Journal International*, **215**(3), 1644–1658.

Zhang, X., Hansteen, F., & Curtis, A., 2019. Fully 3D Monte Carlo ambient noise tomography over Grane field, in *81st EAGE Conference and Exhibition 2019*.

Zhang, X., Hansteen, F., Curtis, A., & de Ridder, S., 2020. 1D, 2D and 3D Monte Carlo ambient noise tomography using a dense passive seismic array installed on the north sea seabed, *Journal of Geophysical Research: Solid Earth*, **125**(2), e2019JB018552.

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.

Zhou, Y., Dahlen, F., & Nolet, G., 2004. Three-dimensional sensitivity kernels for surface wave observables, *Geophysical Journal International*, **158**(1), 142–168.

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: P-WAVE VELOCITY MODELS

Figure A1 shows the horizontal slices of the mean and standard deviation of P-wave velocity using body waves only. Similar to the results for shear velocity, at the depth of 0.2 km higher velocities are associated with higher uncertainties since the near surface structure cannot be well constrained by body waves, and at 0.5 km depth the mean P-wave velocity model exhibits very similar patterns to those of shear velocity. Similarly to the results for shear velocity, the standard deviation map shows higher uncertainties at the location of the southeastern higher velocity anomaly due to limited data coverage. At the depth of 1.0 km the mean velocity map also shows similar structures to

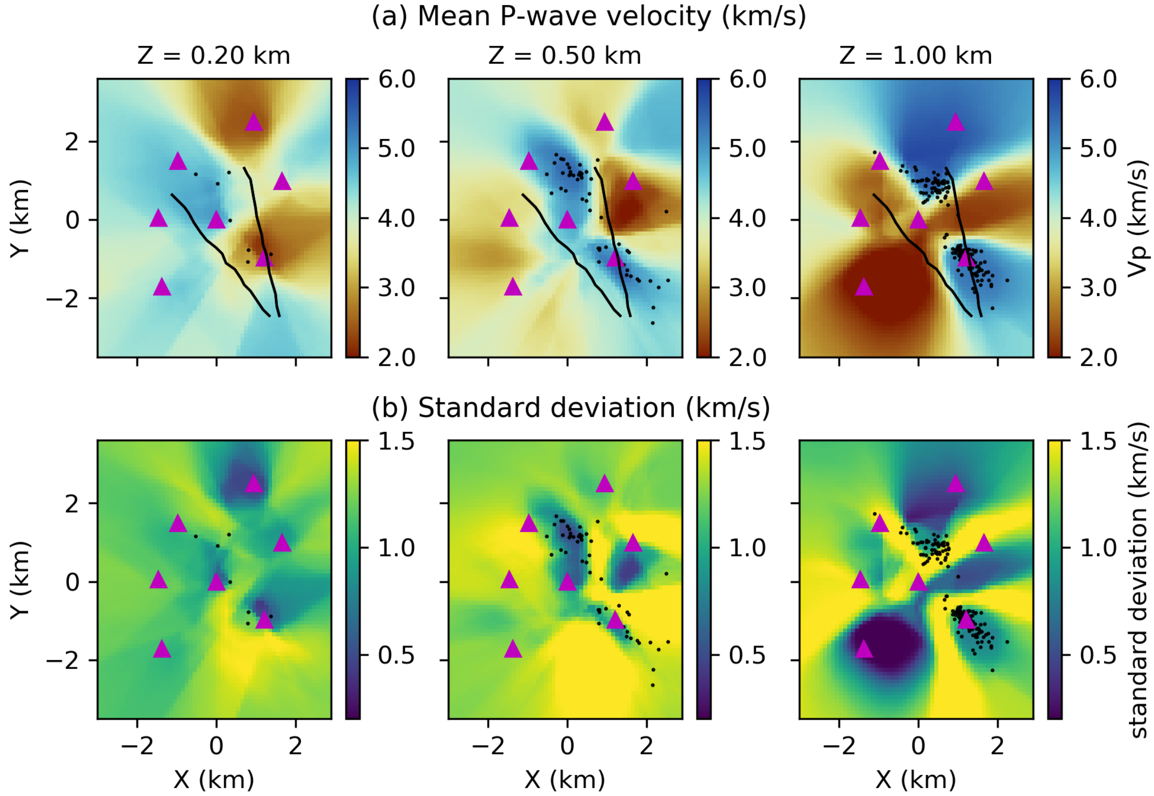

Figure A1. Horizontal slices through the 3D P-wave velocity model obtained using body wave travel time only. Key as in Figure 6.

those in the shear velocity results, and the standard deviation map shows higher uncertainty loops around velocity anomalies.

For comparison, we show the results of P-wave velocity from joint inversion in Figure A2. The P-wave velocity model at the depth of 0.2 km is better constrained by including surface wave dispersion data: most of the model has lower velocities ($<$ 3 km/s) compared to those from body wave tomography ($\sim$ 4 km/s) and lower uncertainties ($<$ 0.5 km/s) than those from body wave tomography ($>$ 1.0 km/s). This is due to the fact that shear velocity is better estimated by including surface waves, so the P-wave velocity can also be better constrained since P and S velocities are correlated through the common earthquake source parameters, and surface waves are also partly sensitive to P-wave velocity at near surface (Zhou et al. 2004; Fang et al. 2016). At greater depths (0.5 km and 1.0 km) the mean velocity model is similar to that from body wave tomography since surface waves have very little sensitivity to P-wave velocity structure there.

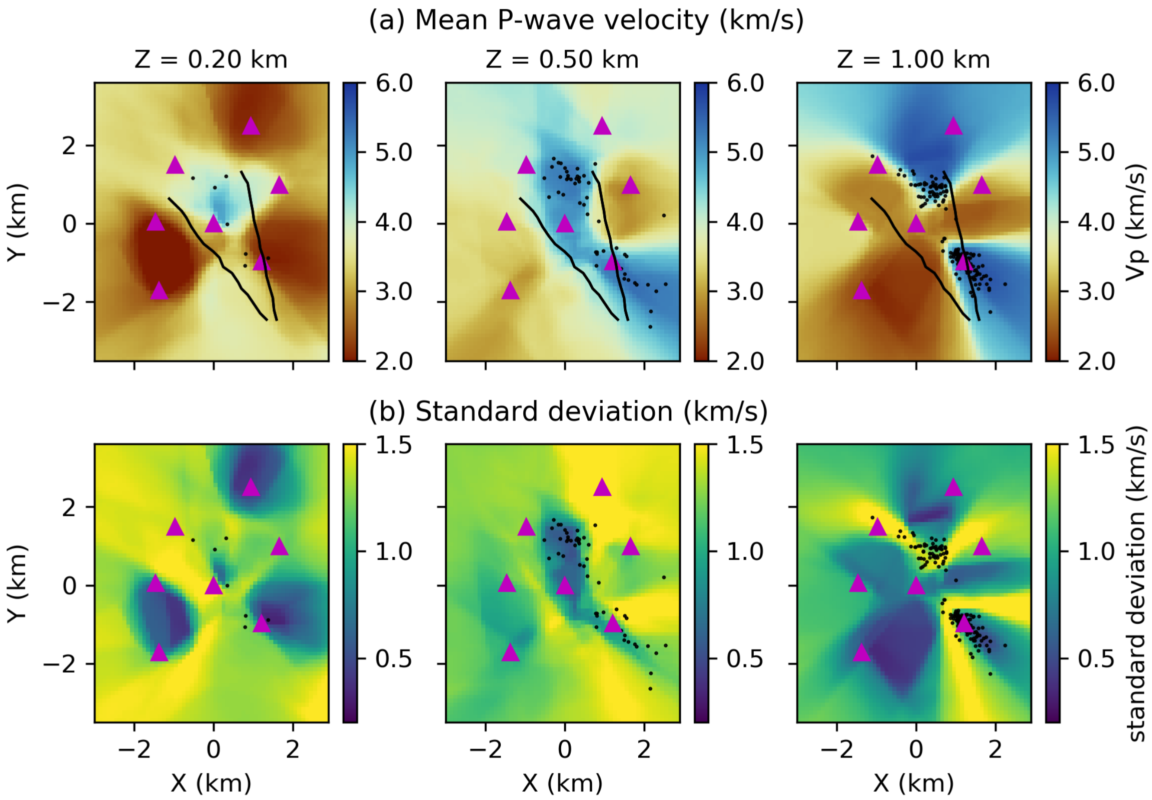

Figure A2. Horizontal slices through the 3D P-wave velocity model obtained from joint body and surface wave inversion. Key as in Figure 6.