Geophysical Journal International

Geophys. J. Int. (2023) **232,** 523–536 https://doi.org/10.1093/gji/ggac331
Advance Access publication 2022 August 23
GJI Seismology

Surface wave dispersion inversion using an energy likelihood function

Xin Zhang,[1] York Zheng[2] and Andrew Curtis[1]

[1]*School of GeoSciences, University of Edinburgh, Edinburgh EH9 3FE, United Kingdom. E-mail: X.Zhang2@ed.ac.uk*
[2]*BP, Chertsey Road, Sunbury-on-Thames, Middlesex TW16 7LN, United Kingdom*

Accepted 2022 August 17. Received 2022 August 15; in original form 2021 November 24

SUMMARY

Seismic surface wave dispersion inversion is used widely to study the subsurface structure of the Earth. The dispersion property is usually measured by using frequency–phase velocity (*f*–*c*) analysis of data recorded on a local array of receivers. The apparent phase velocity at each frequency of the surface waves travelling across the array is that at which the *f*–*c* spectrum has maximum amplitude. However, because of potential contamination by other wave arrivals or due to complexities in the velocity structure the *f*–*c* spectrum often has multiple maxima at each frequency for each mode. These introduce errors and ambiguity in the picked phase velocities, and consequently the estimated shear velocity structure can be biased, or may not account for the full uncertainty in the data. To overcome this issue we introduce a new method which directly uses the spectrum as the data to be inverted. We achieve this by solving the inverse problem in a Bayesian framework and define a new likelihood function, the energy likelihood function, which uses the spectrum energy to define data fit. We apply the new method to a land data set recorded by a dense receiver array, and compare the results to those obtained using the traditional method. The results show that the new method produces more accurate results since they better match independent data from refraction tomography. This real-data application also shows that it can be applied with relatively little adjustment to current practice since it uses standard *f*–*c* panels to define the likelihood, and efficiently since it removes the need to pick phase velocities. We therefore conclude that the energy likelihood function can be a valuable tool for surface wave dispersion inversion in practice.

Key words: Inverse theory; Probability distributions; Surface waves and free oscillations; Statistical methods.

1 INTRODUCTION

Seismic surface waves travel along the surface of the Earth while oscillating over depth ranges that depend on their frequency of oscillation (Aki & Richards 1980). This in turn makes surface waves dispersive—different frequencies travel at different speeds, and these speeds are sensitive to different parts of the Earth. By measuring speeds at different frequencies this dispersion property can be used to constrain subsurface structures over different depth ranges on global scale (Trampert & Woodhouse 1995; Shapiro & Ritzwoller 2002; Meier *et al.* 2007a, b), regional scale (Zielhuis & Nolet 1994; Curtis *et al.* 1998; Simons *et al.* 2002; Yao *et al.* 2006) and industrial scale (Park *et al.* 1999; Xia *et al.* 2003; de Ridder & Dellinger 2011; Zhang *et al.* 2020a).

Surface wave dispersion property (phase or group velocities at different frequencies) can be measured in different ways depending on different acquisition systems. In the case of a single station or a sparse receiver array, as is often the case in seismology, the dispersion property can be measured by using the frequency–time analysis (FTAN) method (Dziewonski *et al.* 1969; Levshin *et al.* 1972; Herrin & Goforth 1977; Russell *et al.* 1988; Levshin *et al.* 1992; Ritzwoller & Levshin 1998; Levshin & Ritzwoller 2001; Bensen *et al.* 2007). In FTAN one constructs a frequency–time domain envelope image for each seismic trace by using a set of narrow bandpass Gaussian filters, and measures the group velocity using the arrival time of the maximum envelope at each frequency. The phase velocity can be derived using the phase of the signal at the time of the maximum envelope plus a phase ambiguity term (appropriate integer multiple of 2π) and a source phase term, or by using an image transformation technique (Yao *et al.* 2006).

In the case where a dense receiver array is deployed, the phase velocity can be determined using frequency–phase velocity (*f*–*c*) analysis, in which wavefields recorded by the array are slant stacked to obtain a *f*–*c* spectrum (Park *et al.* 1998; Xia *et al.* 2003). The phase velocity is then determined as the velocity associated with the highest energy at each frequency in the spectrum. Because different modes are separated in the *f*–*c* domain, the method can also be applied to determine phase velocities for higher modes. However, the obtained spectrum may still suffer from mulimodality which can be caused by multipathing effects (Evernden 1953, 1954), strong lateral heterogeneity (Hou *et al.* 2016), lack of energy in certain frequencies, or interference of different modes. Consequently, the velocity associated with the highest energy does not necessarily represent the most appropriate phase velocity measurement. This issue can be overcome by imposing additional prior information to the phase velocity, for example, by forcing the picked phase velocity dispersion curve to be continuous. Unfortunately, on one hand such additional constrains are usually achieved by picking the phase velocity deliberately and manually for each spectrum, which cannot be applied to large data sets. On the other hand an automatic procedure can easily introduce errors to the picked phase velocities because of the complexity of the spectrum.

To overcome the above issues in phase velocity estimation, in this study we introduce a new method which directly uses the spectrum as data rather than explicit picks of the phase velocities. The spectrum of data has been used in wave-equation dispersion inversion in the framework of full-waveform inversion (Li *et al.* 2017). However, that method is computationally expensive and the problem is solved using a deterministic method which cannot provide uncertainty estimates. To quantify uncertainty we solve the dispersion inversion problem using Bayesian inference. In Bayesian inference one constructs a so-called posterior probability density function (pdf) that describes the remaining uncertainty of models post inversion, by combining prior information with the new information contained in the data as represented by a pdf called the likelihood. The likelihood function describes the probability of observing data given a specific model, and is traditionally assumed to be a Gaussian distribution centred on the picked phase velocities. In this study we propose a new likelihood function, called the *energy* likelihood function which directly uses the spectrum based on the intuition that higher energy in the spectrum reflects higher probability of observing the associated phase velocity.

We apply the new method to a land data set recorded by a dense array and compare the results with those obtained using the traditional method. The data set consists of raw shot records taken from a subarea of a nodal land seismic survey that was conducted in a desert environment (Ourabah & Crosby 2020). This data set offers ultrahigh trace density with over 180 million traces per km^2 on a 12.5 m × 12.5 m receiver grid and a 100 m ×12.5 m source grid. The increased trace density greatly improves the spatial sampling of the wavefield, which in turn benefits the recording and analysis of surface waves. As a part of the depth model building process, refraction tomography was performed to yield a shallow *P*-wave velocity model (Buriola *et al.* 2021), which is then used here for qualitative comparison with the shallow *S*-wave velocity model obtained from our method.

To solve the Bayesian inference problem, we use the reversible-jump Markov chain Monte Carlo (rj-McMC) method. The rj-McMC method is a generalized McMC method which allows a trans-dimensional inversion to be carried out, meaning that the dimensionality of parameter space (the number of parameters) can vary in the inversion (Green 1995). Thus the parametrization itself can be dynamically adapted to the data and to the prior information. The method has been used to estimate phase or group velocity maps of the crustal structure (Bodin & Sambridge 2009; Zulfakriza *et al.* 2014; Galetti *et al.* 2015; Saygin *et al.* 2015; Zheng *et al.* 2017; Rosalia *et al.* 2020) and to estimate shear velocity structures of the crust and upper mantle using surface wave dispersion data (Bodin *et al.* 2012; Shen *et al.* 2012; Young *et al.* 2013; Galetti *et al.* 2017; Killingbeck *et al.* 2018; Yuan & Bodin 2018; Zhang *et al.* 2020a; Hallo *et al.* 2021).

In the following section, we first perform frequency–phase velocity analysis for the recorded data to obtain the *f–c* spectrum around each geographic location. In Section 3, we introduce the new energy likelihood function and give an overview of the rj-McMC algorithm. We then apply the new likelihood function to the obtained spectra to estimate the shear velocity structure, and compare the results with those obtained using the traditional method. The results demonstrate that the new method can generate more accurate results than the traditional method, and can be applied efficiently to large data sets. We therefore conclude that the energy likelihood function provides a valuable tool for surface wave dispersion inversion.

2 SURFACE WAVE DISPERSION ANALYSIS

Fig. 1(a) shows the locations of all 100 627 sensors which are deployed in a regular grid with a spacing of approximately 12.5 m in both directions, and record samples at 250 Hz. In total, 70 261 active sources are fired with 100 m spacing in *X* direction and 12.5 m spacing in *Y* direction to generate seismic surface waves (Fig. 1b). Fig. 1(c) shows an example of a shot gather which mainly contains surface waves.

To analyse the surface wave dispersion, we performed frequency–phase velocity (*f–c*) analysis of the recorded data. For a given geographic location *p*, the *f–c* spectrum $U_p(c, w)$ can be computed using the data recorded by a receiver array around the location:

$$U_p(c, w) = \int_{\mathcal{C}_p} e^{-j\frac{w}{c}x} u(x, w)/|u(x, w)| \mathrm{d}x \approx \sum_{i=1}^{N_p} e^{-j\frac{w}{c}x_i} u(x_i, w)/|u(x_i, w)| \Delta x, \quad (1)$$

where $\mathcal{C}_p$ denotes that the integration is performed around the location *p*, *x* is the source–receiver distance, *w* is frequency in radian, *c* is phase velocity, $j = \sqrt{-1}$, *i* is the index of records and N_p is the number of receivers around location *p*; $u(x, w)$ is the Fourier transform of the wavefield $u(x, t)$. For a given receiver array a larger N_p improves resolution of phase velocity, but reduces spatial resolution. In this study for a given location *p* we stack all the records whose receiver and source locations are, respectively, within 300 and 1500 m to the location *p*. These threshold distances are selected such that the phase velocity dispersion curve can be clearly identified in the spectrum without increasing the number of records unnecessarily. This process is repeated for every geographic location on a regular grid with a 12.5 m spacing in both directions across the survey area.

Fig. 2(a) shows an example spectrum obtained using the above method (displayed as phase velocity versus period) at one specific geographic location (red star in Fig. 1a). The spectrum shows three modes. The phase velocity of the fundamental mode varies from 500 to 800 m s^{-1} and contains two different branches at periods shorter than 0.25 s. This multimodality might be caused by effects described above, or represents two different surface wave modes. To further understand this, we performed an inversion using one of the branches (black dots in Fig. 2a) and modelled the first overtone using the obtained shear velocity profile (see the Appendix). The results show that the modelled first overtone is close to the mode with velocity higher than 600 m s^{-1} (Fig. A1). This therefore demonstrates that the two branches are probably not associated with different modes. The first overtone mainly appears in the period range from 0.1 to 0.25 s with phase velocities varying from 620 to 1000 m s^{-1}.

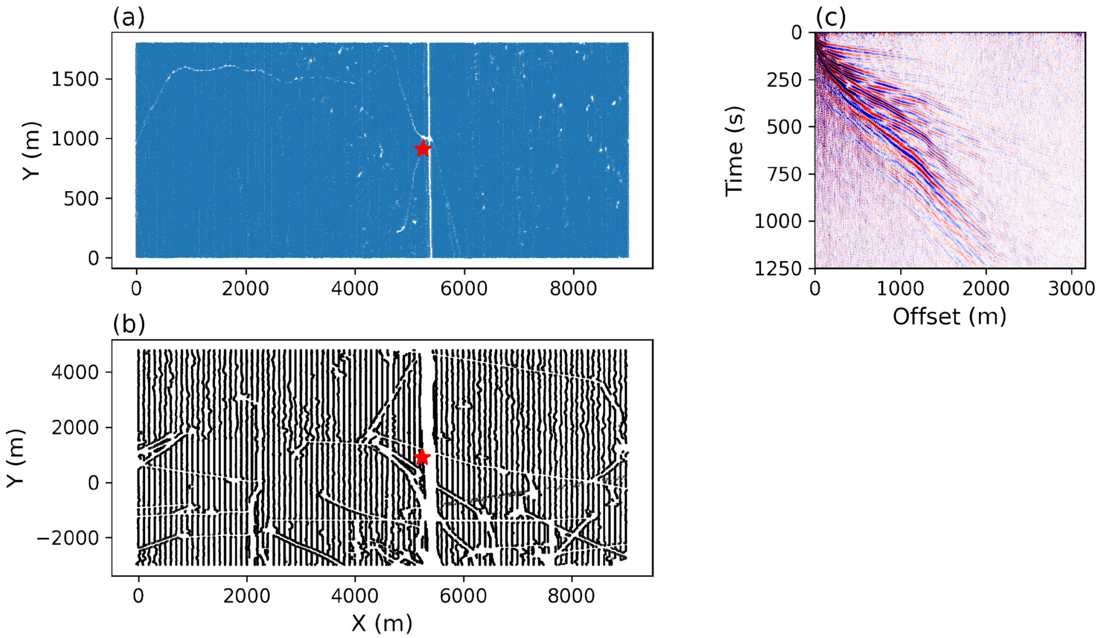

Figure 1. The acquisition system. Panels (a) and (b) show receiver (blue dots) and source (black dots) locations, respectively. The white lines and regions show locations where no receiver or sources is deployed. The red star is referred to in the text. Panel (c) shows an example of a shot gather, which displays the waveform data generated by a specific shot and recorded by receivers at different offsets.

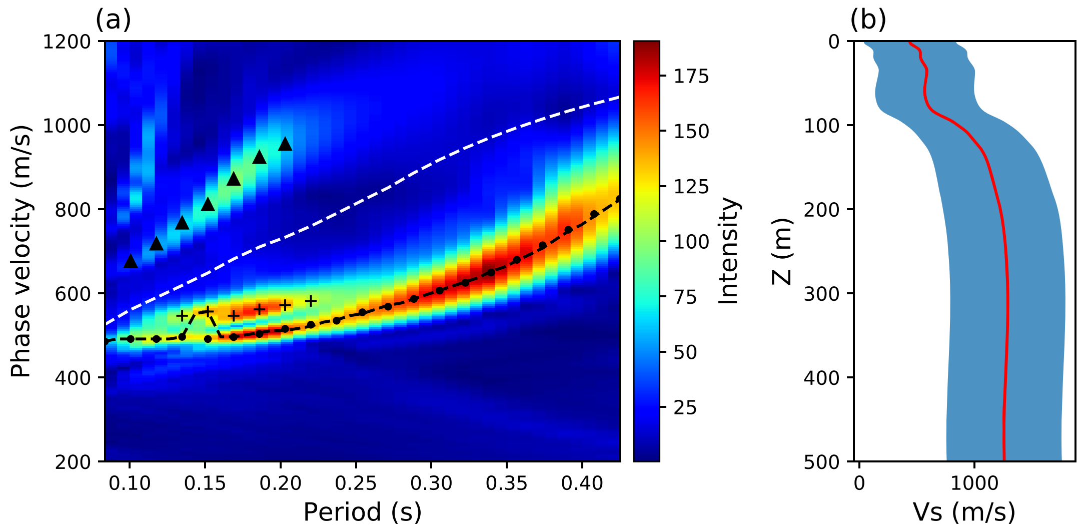

Figure 2. (a) An example spectrum obtained using the *f*–*c* analysis at a specific location (red star in Fig. 1a). Black dots and triangles denote the picked phase velocities for the fundamental mode and the first overtone, respectively, black pluses show the phase velocities associated with the second branch for the fundamental mode and the black dashed line shows the phase velocity dispersion curve associated with the maximum energy at each period. The white line is used to separate the two modes. Panel (b) shows the prior information, which is a uniform distribution with an interval up to 1000 m s^{-1} at each depth. The red line shows the mean of the prior pdf.

The second overtone has much lower energy compared to the other two modes, and we discarded this mode in the inversion.

Traditionally for each mode those phase velocities associated with the peak energy are used as data to constrain the subsurface shear velocity. However, for those modes that have complex structures in the spectrum (such as the fundamental mode in Fig. 2a), it becomes difficult to determine the correct phase velocity dispersion curve to use as these may have apparent jumps between neighbouring periods (e.g. the dashed black line in Fig. 2a). One way to reduce this issue is to impose continuity or smoothness constrains on the dispersion curves. For example, in Fig. 2 the black dots are determined by forcing the dispersion curve to be smooth. Unfortunately the picked dispersion curve is then forced to follow one of the branches at shorter periods ($<$0.25 s) which does not take account of the full data information, and consequently the inverted results do not reflect the full uncertainty in the data. In addition, for large data sets the dispersion curves need to be determined automatically which introduces difficulties to balance higher energy against the smoothness of dispersion curves, and therefore may cause errors in the estimated phase velocities. Alternatively one may identify and exclude the specific frequency ranges of complex data in the inversion. However, there is no easy way to achieve this especially when dispersion curves need to be determined automatically (e.g. Trampert & Woodhouse 1995; van Heijst & Woodhouse 1997; Curtis *et al.* 1998; O'Neill & Matsuoka 2005), and valuable information can be lost in the process. In this study, we therefore automatically pick one single fundamental dispersion curve for each spectrum by imposing smoothness constraints (black dots in Fig. 2) when performing traditional inversions, as is done in many other applications (Trampert & Woodhouse 1995; van Heijst & Woodhouse 1997; Zhang & Chan 2003; Xia *et al.* 2004; Grandjean & Bitri 2006; Cercato 2009; Xia 2014; Olafsdottir *et al.* 2018; Granados *et al.* 2019). In the next section, we propose a method which directly uses the *f–c* spectrum as data to constrain the subsurface velocity structure, thus avoiding these issues.

3 SHEAR WAVE VELOCITY INVERSION

The surface wave dispersion information obtained as above can be used to constrain the subsurface shear velocity structure, which involves solving a non-linear and non-unique inverse problem. In this study, we use Bayesian inference to characterize the fully non-linear uncertainty of the solution.

3.1 Bayesian inference

In Bayesian inference one constructs a so-called posterior pdf $p(\mathbf{m}|\mathbf{d}_{\text{obs}})$ of velocity model $\mathbf{m}$ given the observed data $\mathbf{d}_{\text{obs}}$, by combining prior information with new information contained in the data. According to Bayes' theorem,

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

where $p(\mathbf{m})$ describes the prior information of model $\mathbf{m}$ that is independent of the current data, $p(\mathbf{d}_{\text{obs}}|\mathbf{m})$ is called the *likelihood* which describes the probability of observing $\mathbf{d}_{\text{obs}}$ if model $\mathbf{m}$ was true and $p(\mathbf{d}_{\text{obs}})$ is a normalization factor called the *evidence*.

The prior information is critical for Bayesian inference. To construct a more informative prior distribution than the commonly-used Uniform distribution with little or no depth dependence, we first conduct a set of inversions at multiple geographic locations using a Uniform prior pdf from 300 to 1500 m s^{-1}, which spans the range of shear velocities in the upper 500 m according to a variety of similar studies (Lee & Collett 2008; Mordret *et al.* 2014; Chmiel *et al.* 2019; Zhang *et al.* 2020a). The average of the mean models from these inversions are then used as the mean of the prior pdf, and we construct a Uniform distribution with a width of 1000 m s^{-1} (larger than four standard deviations obtained from the previous inversions) at each depth (Fig. 2b). This prior information improves the depth resolution and constrains the subsurface structure better than an identical Uniform distribution across the depth ranges (Yuan & Bodin 2018). The prior information also avoids numerical instatility in surface wave dispersion modelling methods which can significantly bias the results using a variety of common forward modelling codes (Galetti *et al.* 2017).

3.2 Energy likelihood function

In traditional methods one often uses a Gaussian distribution for the likelihood function:

$$p(\mathbf{d}_{\text{obs}}|\mathbf{m}) = \frac{1}{\sqrt{(2\pi)^k|\Sigma|}}\exp\left(-\frac{1}{2}(\mathbf{d}_{\text{obs}}-\mathbf{d})^{\text{T}}\Sigma^{-1}(\mathbf{d}_{\text{obs}}-\mathbf{d})\right) \tag{3}$$

where $\mathbf{d}_{\text{obs}}$ is a real k-dimension data vector which in this case is the phase velocities picked from the *f–c* spectrum, $\mathbf{d}$ is the model predicted data vector and Σ is the data covariance matrix which usually is set to be a diagonal matrix. Note that this likelihood function is based on the assumption that data noise has a Guassian distribution which is not always true in reality.

As discussed above in this study we define a new likelihood function which directly uses the *f–c* spectrum as data. In the *f–c* spectrum the intensity (energy) at each frequency and each phase velocity is obtained by stacking the back-propagated signals according to each specific phase velocity and frequency (eq. 1). Assuming that the signal consists of surface waves, higher intensity in the spectrum means that the associated phase velocity is closer to the true phase velocity of a surface wave mode. Hence the intensity acts as a measure of the consistency between the associated phase velocity and the true phase velocity, which can be used as a measure of probability of observing the specific phase velocity. Based on this observation, we can directly use the spectrum as data and write a new likelihood function. Define $\mathbf{E}$ as the matrix representing the *f–c* spectrum and $E(T, c)$ as the energy at period $T = 1/f$ and phase velocity c, and assuming that the energy at each value of T and c has exponentially decaying probability away from the maximum energy value at that period, the likelihood can be expressed as:

$$p(\mathbf{E}|\mathbf{m}) = \frac{1}{Z}\exp\left[-\sum_i \frac{\max(\mathbf{E}(T_i,\cdot)) - E(T_i, c_i(\mathbf{m}))}{\sigma_i^2}\right], \tag{4}$$

where T_i is the ith period, $\max(\mathbf{E}(T_i,\cdot))$ is the maximum energy at period T_i which guarantees that the exponent is negative, $c_i(\mathbf{m})$ is the phase velocity at period T_i predicted using model $\mathbf{m}$, σ_i is a scaling factor and Z is a normalization factor. The normlization factor Z is calculated numerically by integrating the function in eq. (4) over the velocity range for each period. The scaling factor σ_i is generally unknown, so we treat it as an additional parameter and estimate it hierarchically (Malinverno & Briggs 2004). The above likelihood function takes a similar form as the Boltzmann distribution which describes the probability that a system will be in a certain state as a function of that state's energy (Landau & Lifshitz 2013). In this case the state is the set of phase velocities at different frequencies, and the scaling factor σ_i acts as the 'temperature' which is related to the noise in the original data. For example, more noise in the data means the energy in the *f–c* spectrum is less focused, and hence the scaling factor will be higher.

For multimodal inversion the energy of the fundamental mode may dominate the likelihood function in eq. (4) (for example, see Fig. 2a), and consequently the inverted results can be biased because models may have apparently larger likelihood values if their predicted higher modes also fit the fundamental mode energy. We therefore separate each mode by windowing out other modes. For example, define $\mathbf{E}_j$ to be the spectrum of the jth mode after other modes have been windowed out. Then the likelihood function becomes:

$$p(\mathbf{E}|\mathbf{m}) = \frac{1}{Z}\exp\left[-\sum_{ij} \frac{\max(\mathbf{E}_j(T_i,\cdot)) - E_j(T_i, c_i(\mathbf{m}))}{\sigma_{ij}^2}\right]. \tag{5}$$

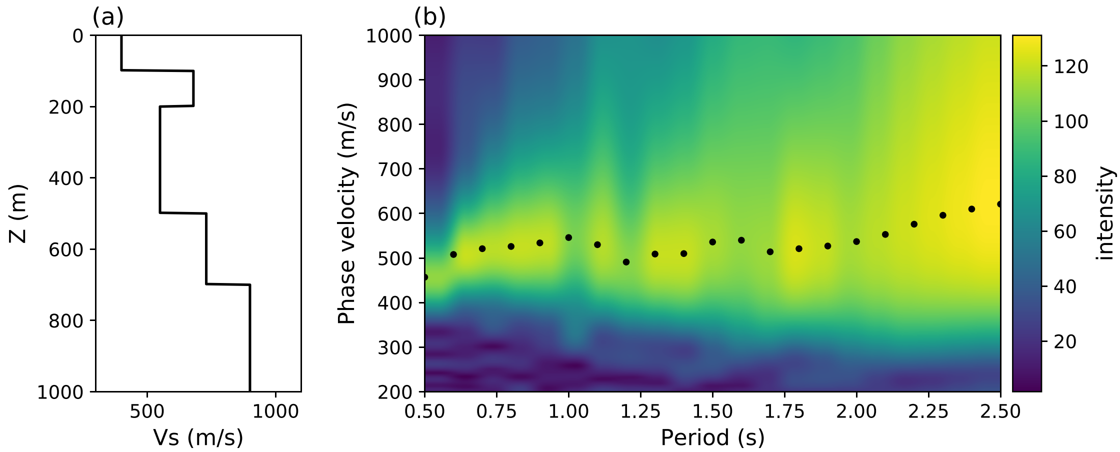

Figure 3. (a) The true model. (b) The spectrum obtained using eq. (1) by stacking the data simulated from the true model. Black dots show the phase velocities picked from the spectrum.

Although this requires that we define a window function for each mode, this process is usually straightforward. For example, the white dashed line in Fig. 2(a) shows the boundary used to separate the first two modes; this same line is used for all other spectra across the survey area since it appeared appropriate for a large number of spectra examined manually.

3.3 Reversible-jump Markov chain Monte Carlo

We use reversible-jump Markov chain Monte Carlo (rj-McMC) to generate samples from the posterior pdf. The rj-McMC method is a generalized version of the Metropolis–Hastings algorithm (Metropolis & Ulam 1949; Hastings 1970), which allows the number of model parameters to be variable in the inversion (Green 1995). Thus the parametrization of the seismic velocity model can itself be determined by the data and prior information. The method has been applied in a range of geophysical applications (Malinverno 2002; Bodin & Sambridge 2009; Dettmer *et al.* 2010; Minsley 2011; Ray & Key 2012; Young *et al.* 2013; Piana Agostinetti *et al.* 2015; Saygin *et al.* 2015; Galetti *et al.* 2017; Burdick & Lekić 2017; Biswas & Sen 2017; Zhu & Gibson 2018; Xiang *et al.* 2018; Zhang *et al.* 2020a, b; Estève *et al.* 2021; Mousavi *et al.* 2021; Hallo *et al.* 2021). In this study, we use the method to solve the surface wave dispersion inversion problem.

In rj-McMC one constructs a (Markov) chain of samples by perturbing the current model $\mathbf{m}$ using a proposal distribution $q(\mathbf{m}'|\mathbf{m})$ to generate a new model $\mathbf{m}'$, and by accepting or rejecting this new model with a probability $\alpha(\mathbf{m}'|\mathbf{m})$ called the acceptance ratio:

$$\alpha(\mathbf{m}'|\mathbf{m}) = \min\left[1, \frac{p(\mathbf{m}')}{p(\mathbf{m})} \times \frac{q(\mathbf{m}|\mathbf{m}')}{q(\mathbf{m}'|\mathbf{m})} \times \frac{p(\mathbf{d}_{\mathrm{obs}}|\mathbf{m}')}{p(\mathbf{d}_{\mathrm{obs}}|\mathbf{m})} \times |\mathbf{J}|\right], \quad (6)$$

where $\mathbf{J}$ is the Jacobian matrix of transforming $\mathbf{m}$ to $\mathbf{m}'$ and is used to account for any volume changes of parameter space during jumps between dimensionalities. In our case the Jacobian matrix is an identity matrix (Bodin & Sambridge 2009).

For surface wave dispersion inversion beneath each geographical location we use a set of layers to parametrize the subsurface, which can be changed in different ways within the rj-McMC algorithm (Green 1995; Bodin & Sambridge 2009): adding a new layer, removing a layer, changing layer positions and changing layer velocities. There is also another perturbation related to the hyperparameters of the likelihood function: changing the scaling factor σ_{ij} in eq. (5) or standard deviation in the Gaussian likelihood function. After each perturbation of the current model $\mathbf{m}$, the acceptance ratio α is computed using eq. (6) and is compared with a random number γ generated from the Uniform distribution on [0,1]. 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. This process guarantees that the generated samples are distributed according to the posterior pdf if the number of samples tends to infinity (Green 1995).

3.4 A synthetic example

To verify the energy likelihood function, we first conduct a synthetic test and compare the results with those obtained using the traditional method. The true shear velocity model contains five layers (Fig. 3a), and is used to calculate the phase velocity dispersion curve using the modal approximation method (Herrmann 2013). The modelled dispersion curve is then used to generate synthetic signals in time domain with a 50 Hz sampling rate using a plane wave approximation method for 200 uniformly spaced sensors with a 5 m spacing (Park & Miller 2008; Naskar & Kumar 2022). Here we only generated signals using the fundamental mode and kept amplitudes the same for different frequencies. We added Gaussian noise to the signals whose standard deviation is set to be 50 per cent of the medium of the maximum amplitudes of all sensors. Those signals are then stacked according to eq. (1) to generate a spectrum (Fig. 3b) which is used as data for energy likelihood function. The phase velocities associated with maximum intensity at each period (black dots in Fig. 3b) are picked as data for the Gaussian likelihood function. For the inversion we use 21 equally spaced periods from 0.5 to 2.5 s.

The prior pdf of shear velocity is set to be a Uniform distribution between 300 and 1100 m s^{-1} and the prior pdf of the number of layers is chosen to be a Uniform distribution between 3 and 20. The prior pdf of the scaling factor σ_i for the energy likelihood function is set to be a Uniform distribution between 0.05 and 10, and the prior pdf of the standard deviation for the Gaussian likelihood function is a Uniform distribution between 0.3 and 60. We use Gaussian distributions as the proposal distribution: the width of the Gaussian distribution for the fixed-dimensional steps (changing layer positions, changing layer velocities and changing likelihood

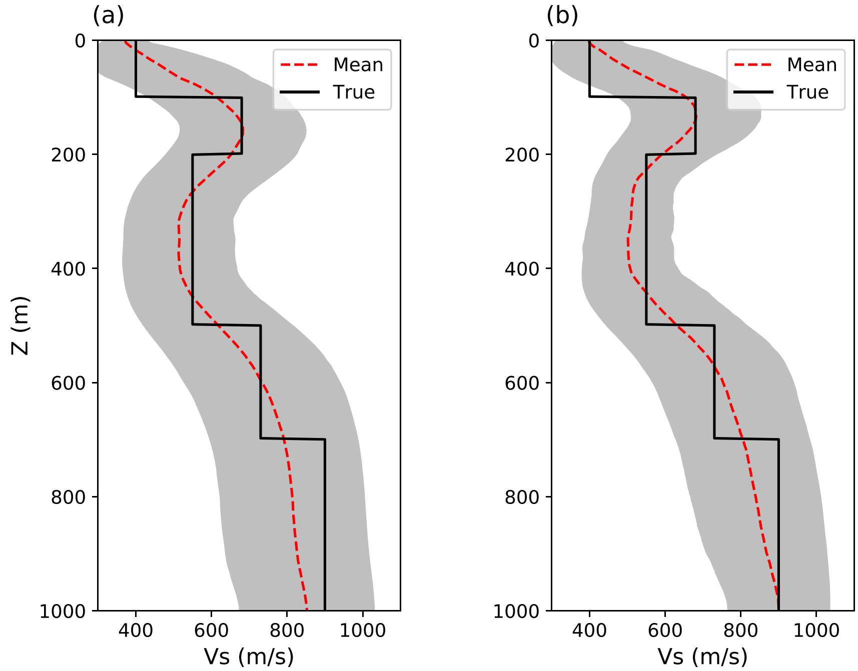

Figure 4. The results obtained using (a) the energy likelihood function and (b) the Gaussian likelihood function. Red and black lines show the mean and true models, respectively. The grey area shows the standard deviation around the mean model.

hyperparameters) is chosen by trial and error to ensure an acceptance ratio between 20 and 50 per cent after burn-in period; the width of the trans-dimensional steps (adding or removing a layer) is selected to produce the maximum acceptance ratio. For each inversion we run six chains, and each chain contains 800 000 iterations with a burn-in period of 300 000 during which all samples are discarded. After the iteration each chain is thinned (decimated) by a factor of 50, and the remaining samples are used for the subsequent inference.

Figs 4(a) and (b) show the results obtained using the energy likelihood function and the Gaussian likelihood function, respectively. Overall the two results show similar mean and standard deviation models even though they are obtained using very different likelihood functions. However, due to the difference between the two likelihood functions there are still slightly different features in the standard deviation models. We also note that because the actual noise distribution remains unknown, it is impossible to find the true posterior pdf. Nevertheless this example demonstrates that the energy likelihood function is a valid choice in the sense that it generates similar results to the traditional method.

3.5 A 1-D example

We now apply the above method to the dispersion data in Fig. 2(a) using both the energy likelihood function and the Gaussian likelihood function and compare their results. We use both the fundamental mode and the first overtone with 21 equally spaced periods from 0.0835 to 0.425 s. The prior pdf of shear velocity is shown in Fig. 2(b) and the prior pdf of the number of layers is set to be a discrete Uniform distribution between 2 and 25. For the energy likelihood function the prior pdf of the scaling factor σ_{ij} is chosen to be a Uniform distribution between 0.5 and 20, and for the Gaussian likelihood function the prior pdf of the data noise (standard deviation of the Gaussian distribution) is a Uniform distribution between 0.05 and 50. The inversions are conducted in the same way as described in the above section.

Figs 5(a) and (c) show the marginal distributions of shear velocity obtained using the energy likelihood function and the Gaussian likelihood function, respectively. The corresponding phase velocity distributions generated by those posterior samples are shown in Figs 5(b) and (d), respectively. The shear velocity marginal distribution obtained using the energy likelihood function shows clearly multimodal distributions in the near surface (<100 m), which are associated with the two branches in the *f*–*c* spectrum (Figs 2a and 5b). In comparison the marginal distribution obtained using the Gaussian likelihood function shows a unimodal distribution and the predicted data only fit the single branch on which the phase velocities were picked. Since we do not know which branch reflects the most appropriate phase velocity a priori, the shear velocity obtained using the Gaussian likelihood function is biased and the estimated uncertainty failed to take account of the full, multimodal uncertainty in the data. In contrast, by directly using the spectrum as data one can embed all data uncertainty in the likelihood function and therefore obtain a less biased result. In addition, the phase velocity distribution of the first overtone obtained using the energy likelihood function (Fig. 5b) fits the picked phase velocity better than that obtained using the Gaussian likelihood function. This suggests that there is inconsistency between the picked phase velocities of the fundamental mode and the first overtone, and therefore further demonstrates the necessity of including the second branch in the likelihood function. In the deeper part (>100 m) the velocity obtained using the energy likelihood function increases from 600 m s^{-1} at 100 m depth to 1500 m s^{-1} around 220 m and stays almost constant down to 500 m; whereas the velocity obtained using the Gaussian likelihood function is around 1200 m s^{-1} across the whole depth from 100 to 500 m. This is probably because the picked phase velocities (black dots and triangles in Fig. 2a) are not sensitive to the deeper part (>200 m), and consequently the shear velocity in

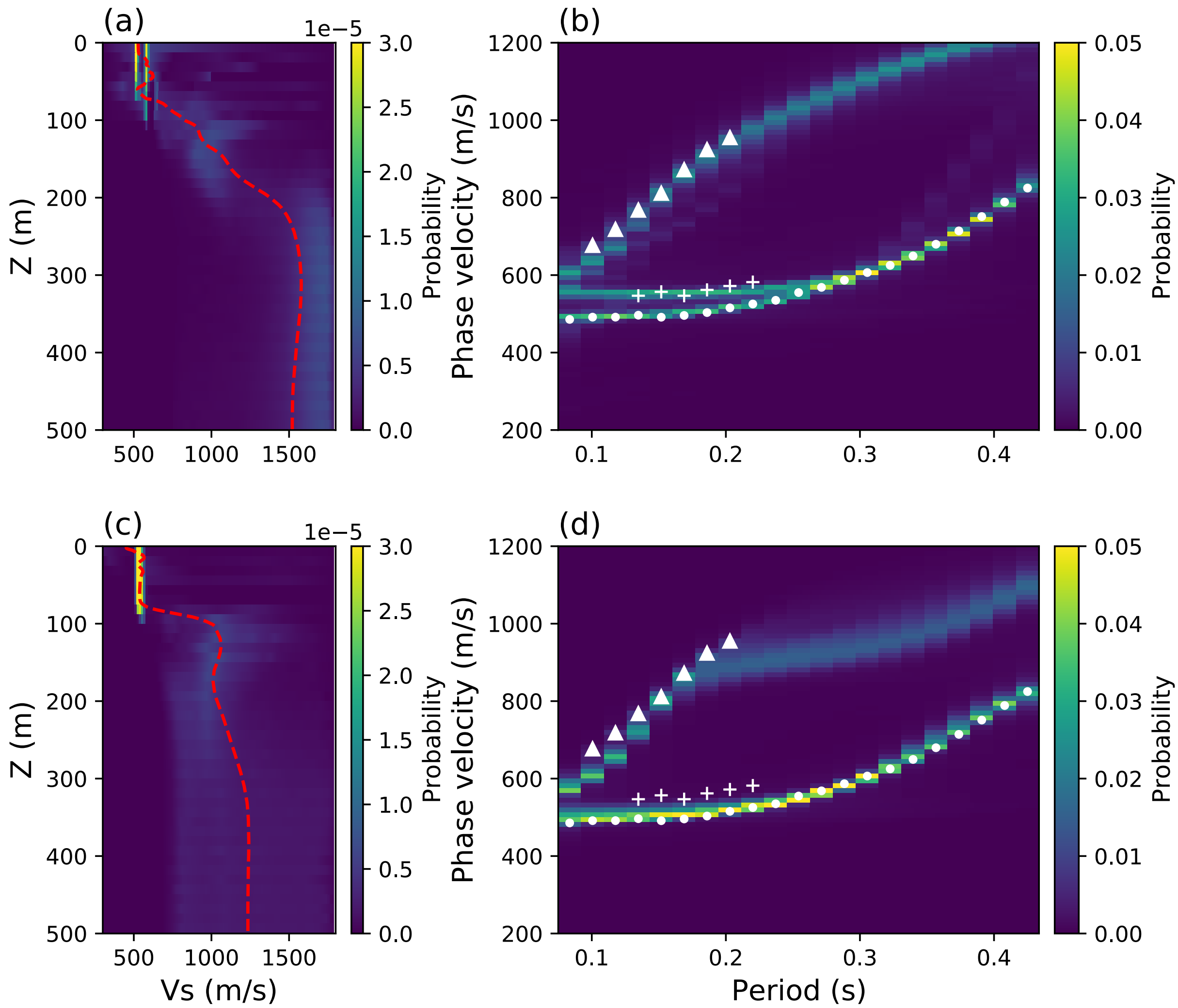

Figure 5. (a) The shear velocity and (b) the phase velocity marginal posterior distributions obtained using the posterior samples of the energy likelihood function for the spectrum in Fig. 2(a). The phase velocity distribution is obtained by modelling the phase velocity for all posterior samples. Panels (c) and (d) show the associated shear velocity and phase velocity marginal distributions obtained by the traditional method of using the picked phase velocities shown by black dots and triangles in Fig. 2(a). Red dashed lines show the mean velocity profile. White dots, pluses and triangles are associated with the black dots, pluses and triangles in Fig. 2, respectively.

the deeper part are dominated by the prior pdf. In comparison the energy likelihood function uses all the information contained in the spectrum and constrains deeper structure better. For example, at periods longer than 0.22 s phase velocities are not determined for the first overtone because of its lower energy, whereas the information is still used in the energy likelihood function to constrain deeper structures. This can also be observed in the phase velocity distributions predicted from posterior samples (Figs 5b and d), where the first overtone distribution obtained using the energy likelihood function at longer periods (>0.22 s) is more similar to the spectrum than that obtained using the Gaussian likelihood function. Thus, by directly using the spectrum as data the new energy likelihood function can use more information in the data and can obtain more accurate results than the traditional method.

4 3-D RESULTS

To obtain 3-D shear velocity models we perform 1-D inversions to all of the spectra across the survey area. For each inversion at each geographic location the inversion is conducted in the same way as described in the previous section with the same prior pdf. For the proposal pdf we tune the width of the Gaussian proposal distribution according to the criteria described above at a few randomly selected locations, and then use the same proposal pdf for all other inversions. To compare the results we use both the energy likelihood function and the Gaussian likelihood function around picked phase velocities. For each spectrum we automatically determine the phase velocity from longer periods to shorter periods: at each period the phase velocity is determined as the local maximum energy point whose phase velocity is closest to the already picked phase velocity at the neighbouring period, such that the picked dispersion curve is as continuous as possible. Alternatively, one may parametrize the dispersion curves in terms of splines to ensure smoothness (Ekström *et al.* 1997; Ekström 2011). To ensure the quality of picked phase velocities, we only determine the phase velocity at frequencies with sufficiently high energy, such that the picked energy is at least 1.8 times higher than the average energy at each frequency.

Fig. 6 shows the mean and standard deviation models at the depth of 20, 80, 140 and 200 m obtained using the energy likelihood

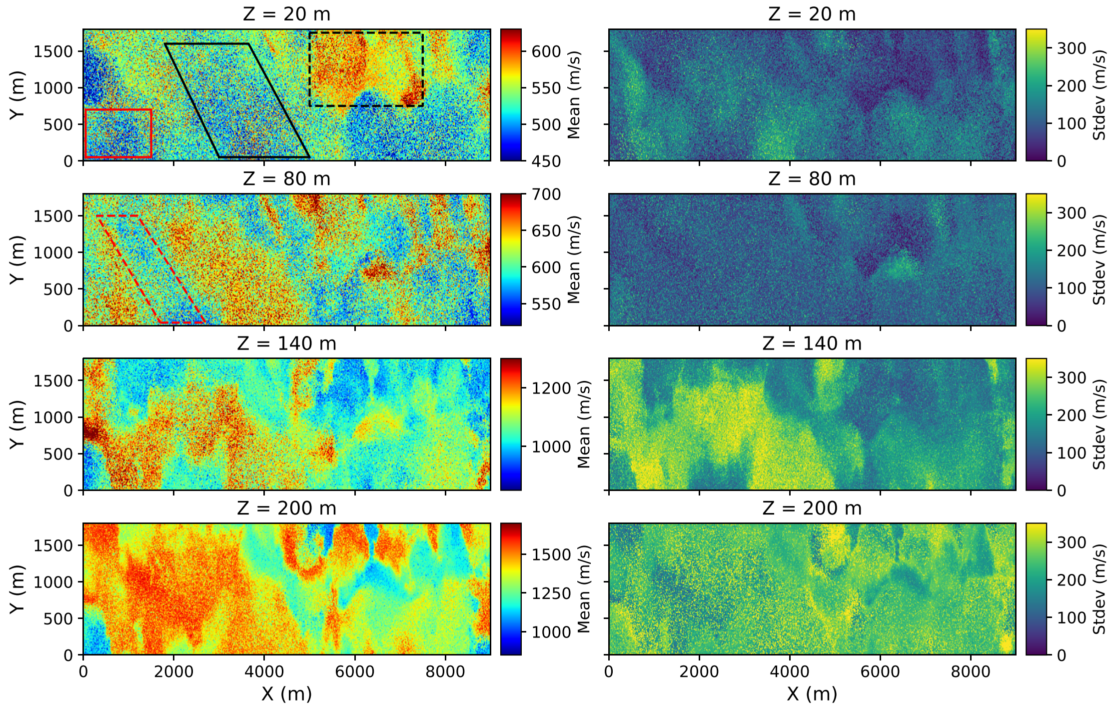

Figure 6. The mean (left-hand panel) and standard deviation (right-hand panel) of shear velocities obtained using the energy likelihood across horizontal slices at depths 20, 80, 140 and 200 m. Boxes highlight velocity anomalies which are referred to in the text.

function. Among various structures we have highlighted several velocity anomalies using boxes which we will refer to below. Overall the standard deviation model shows lower uncertainty at the shallower part (20 and 80 m) and higher uncertainty at the deeper part (140 and 200 m) due to the fact that seismic surface waves are more sensitive to near surface structure. At all the depths the standard deviation shows similar features to the mean model, but note that at 20 m depth higher velocity anomalies are associated with lower uncertainties, whereas at 140 m depth higher velocity anomalies correspond to higher uncertainties. This phenomenon has also been observed previously (Zhang & Curtis 2020b; Gebraad *et al.* 2020; Zhang & Curtis 2021a), and the different correlation between velocity anomalies and uncertainties at different depths suggests that there is a complex, non-linear relationship between seismic velocity and phase velocity data.

Fig. 7 shows the mean and standard deviation models obtained using the Gaussian likelihood function at the same depths. Overall the mean model obtained using the Gaussian likelihood function shows more small scale variations than that obtained using the energy likelihood function. This is probably caused by errors in the picked phase velocities, which is inevitable when the phase velocities are estimated automatically. At 20 m depth the mean model shows similar structures to those observed in the previous results in Fig. 6, for example the northern high velocity anomaly denoted by black dashed boxes and the low velocity anomalies in the southeast and northwest. However, the low velocity anomalies that are denoted by black and red solid line boxes in Fig. 6 are not present in Fig. 7. Similar to the previous results, at 80 m depth there are small scale variations in the east, but the low velocity anomaly denoted by red dashed line box and the high velocity anomaly next to this low anomaly are not visible. At greater depths (140 and 200 m) the mean velocity model is significantly different from the previous results and contains many small scale structures which probably do not reflect the true geologic structure as the scale of these structures is smaller than the scale (300 m) used for stacking which implicitly imposes smoothness to the velocity structure. This suggests that the traditional method can cause bias in the inverted velocity structure because of errors in the picked phase velocity and loss of useful information in the data. However, we note that there are small-scale structures in both *Vs* models, which may reflect the true structure of the subsurface, or may be caused by the failure of 1-D inversions to represent the true structure in the presence of substantial lateral heterogeneities. Similarly to the previous results, the standard deviation model shows similar features to the mean model, and the uncertainty is lower in shallower parts.

To further understand the results we compare the above shear velocity models with *P*-wave velocity model (Fig. 8) obtained using refraction tomography from the same data set (Buriola *et al.* 2021). At 20 m depth the *Vp* model shows similar features to the shear velocity models. For example, there is a high velocity anomaly in the east between $X = 6000$ and 8000 m, which may be related to the high velocity anomaly (black dashed line box) observed in the shear velocity models even though they are not at exactly the same location. Similarly to the shear velocity model obtained using the energy likelihood function, there is a low velocity anomaly between $X = 2000$ and 5000 m in the southeastern direction (black solid line box) and a low velocity anomaly in the southwestern corner (red solid line box). This strongly suggests that the new energy likelihood method is effective since those low velocity anomalies are not visible in the results obtained using the traditional method (Fig. 7). Even though at depths of 80 and 140 m the *Vp* model is different from both shear velocity models, there are still similarities between the *Vp* model and the shear velocity model obtained using the energy likelihood function. For example, at 80 m depth there is a low velocity channel in the southeastern direction (denoted by red

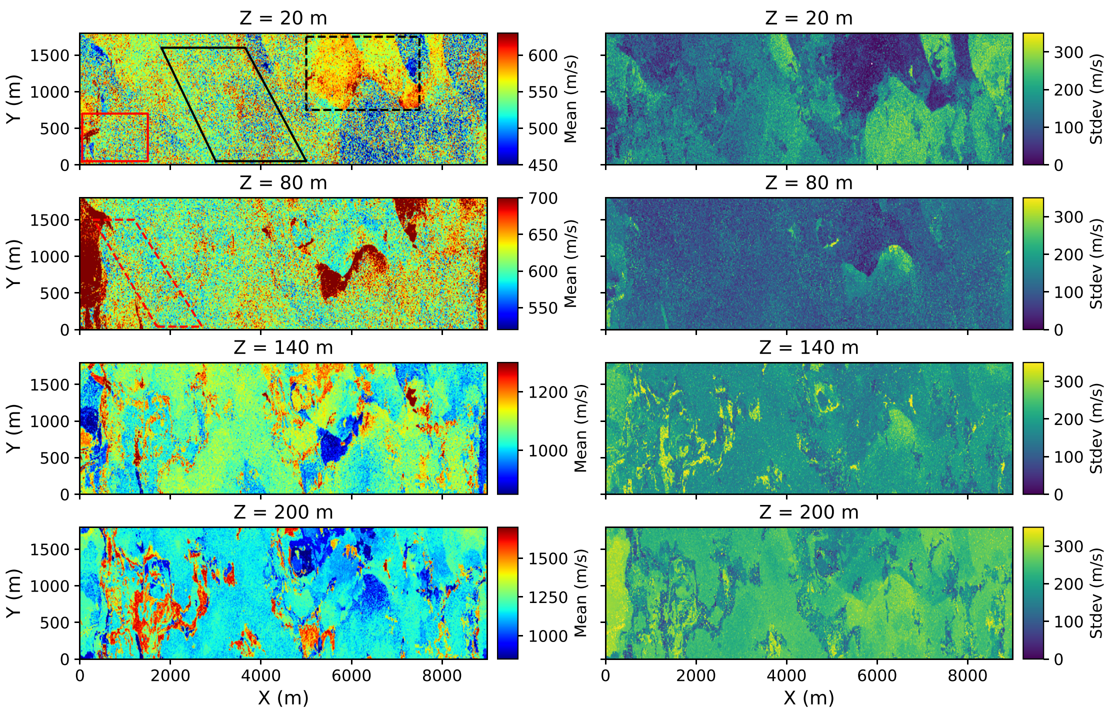

Figure 7. The mean (left-hand panel) and standard deviation (right-hand panel) of shear velocities obtained using the traditional method with picked phase velocities across horizontal slices at depths 20, 80, 140 and 200 m. Boxes are the same as in Fig. 6.

dashed line boxes) in the west of both models. At 200 m depth the *Vp* model shows more similarities to the shear velocity model obtained using the energy likelihood function: both models show a high velocity anomaly between $X = 2000$ and 5000 m in the southeastern direction across the area (the same location as denoted by the black solid line box at 20 m depth) and a low velocity channel to the east of this high velocity anomaly. In the east ($X > 4000$ m) there are similar small scale high velocity anomalies in both models. In comparison the shear velocity model obtained using the traditional method is very different from the *Vp* model. To further analyse the similarity between these models, we computed the structural similarity index measure (SSIM, Wang *et al.* 2004) between the two *Vs* models and the *Vp* model. The obtained SSIM values are 0.13 and 0.08 for the models obtained using the energy likelihood function and the Gaussian likelihood function, respectively. This suggests that the model obtained using the energy likelihood function is more similar to the *Vp* model than the model obtained using the Gaussian likelihood function. This indicates that the new energy likelihood function may generate more accurate shear velocity models than the traditional method.

Fig. 9 shows vertical sections through the different models at $Y =$ 875 m. To further compare these models, in Fig. 9(f) we show the average velocity profile across the section of each model with the *P*-wave velocity converted to shear velocity using a factor of 1.73 (Pickett 1963). In the near surface (<100 m) all three models show a lower velocity layer (Fig. 9f), although the two shear velocity models show more complex structures than the *P*-wave velocity. Between depths of 100 and 230 m the shear velocity obtained using the energy likelihood function increases from 600 to 1500 m s^{-1} which is consistent with the trend of *P*-wave velocity, and both models show a high velocity layer below 200 m. In comparison the shear velocity obtained using the traditional method is around 1300 m s^{-1} below 100 m. A similar phenomenon was observed in Section 3 in the 1-D example, which is caused by the fact that the picked phase velocities are not sensitive to the deeper structure. For example, the standard deviation model obtained using the traditional method shows higher uncertainties (>300 m s^{-1}) below 200 m, whereas that obtained using the energy likelihood function shows a much lower uncertainty (<100 m s^{-1}) between 200 and 400 m. In addition, the shear velocity model obtained using the traditional method shows higher lateral spatial variations in deeper parts of the section (>200 m) than the other two models. This further demonstrates the importance of using all information contained in the spectrum, rather than using only picked phase velocities. Note that both shear velocity standard deviation models show higher uncertainties at the layer boundary, which has been observed previously (Galetti *et al.* 2015; Zhang *et al.* 2018) and reflects the uncertainty of boundary positions.

5 DISCUSSION

The McMC method is generally computationally expensive. For the above inversion each chain took 0.173 CPU hours using one core of an Intel Xeon CPU, which results in a total of 1.04 CPU hr for each geographic location and 107,015.4 CPU hr for all 102 817 inversions across the survey area. However, the inversions are fully parallelizable since the inversion at each geographical location is independent of other inversions. For example, the above inversion across the whole survey area took 15 hr using 7200 CPU cores. We also note that other methods can be used to further improve the computational efficiency, for example, Hamiltonian Monte Carlo (Duane *et al.* 1987; Fichtner *et al.* 2018; Kotsi *et al.* 2020), Langevin Monte Carlo (Roberts *et al.* 1996; Siahkoohi *et al.* 2020), variational inference (Nawaz & Curtis 2018; Zhang & Curtis 2020a, b; Zhao *et al.* 2021; Smith *et al.* 2022) and neural network inversion (Meier *et al.* 2007a, b; Earp *et al.* 2020; Zhang & Curtis 2021b).

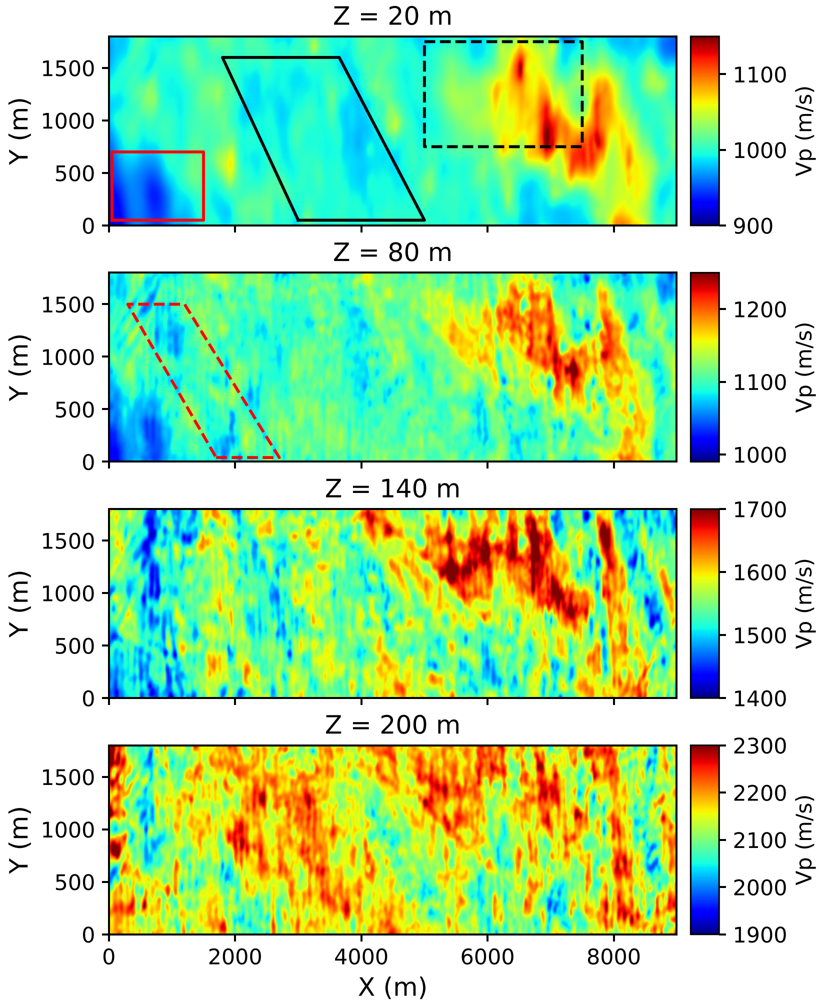

Figure 8. The *P*-wave velocity model obtained using refraction tomography across horizontal slices at depths 20, 80, 140 and 200 m. Boxes are the same as in Fig. 6.

In the above results the phase velocities used for the traditional method are determined automatically from the *f*–*c* spectrum, and these may contain errors due to the complexity of the *f*–*c* spectrum. We therefore note that the results obtained using the traditional method may be improved if the phase velocities can be determined more accurately. However, this usually requires deliberately and manually picking of a dispersion curve for each spectrum, which restricts its application to a relatively small number of spectra. By contrast the new energy likelihood function directly using the spectrum as data and can easily be applied to larger data sets.

The above method requires that we separate different modes in the *f*–*c* spectrum, which is not always easy since higher modes may overlap the fundamental mode. In such cases the phase-matched filter theory may be used to iteratively suppress the effect of interfering overtones (Ekström *et al.* 1997). Alternatively one can directly model apparent phase velocities rather than modelling phase velocities for each mode separately (Tokimatsu *et al.* 1992; Lai *et al.* 2014), in which case the energy likelihood function can still be used.

In this study, we treated the multimodality that appears in a surface wave mode in the *f*–*c* spectra as uncertainties in the data. This uncertainty is propagated into the posterior distribution using Bayesian inference. We note that it might also be possible to further analyse the cause of the multimodality, or use other methods like the two plane wave approach (Forsyth *et al.* 2005) to improve the measurements or interpretation of phase velocities in the presence of multipathing effects. We performed 1-D inversions using a plane wave approximation for a 3-D structure, which may cause some bias in the inverted velocity structure in the presence of strong lateral heterogeneities as noted in Wielandt (1993). In such cases a 3-D wave-equation based inversion method could be used in future to improve the results (Li *et al.* 2017; Liu *et al.* 2018).

We performed the surface wave dispersion inversion independently at each geographic location, which loses lateral spatial correlations between neighbouring velocities and can cause biases in the final results (Zhang *et al.* 2018). For example, there are horizontal discontinuities in the shear velocity mean and standard deviation models (Fig. 9), which probably do not reflect the real subsurface structure. To include the lateral spatial correlation and to further improve the results a 3-D parametrization may be used instead of the 1-D parametrization, for example, 3-D Voronoi tessellation can be used in surface wave tomography (Zhang *et al.* 2020a).

In this study, we used the spectrum obtained by stacking wavefields in the *f*–*c* domain, which requires a dense receiver array. In cases where only a sparse array is available, the energy likelihood can still be used to perform the inversion. For example, one could

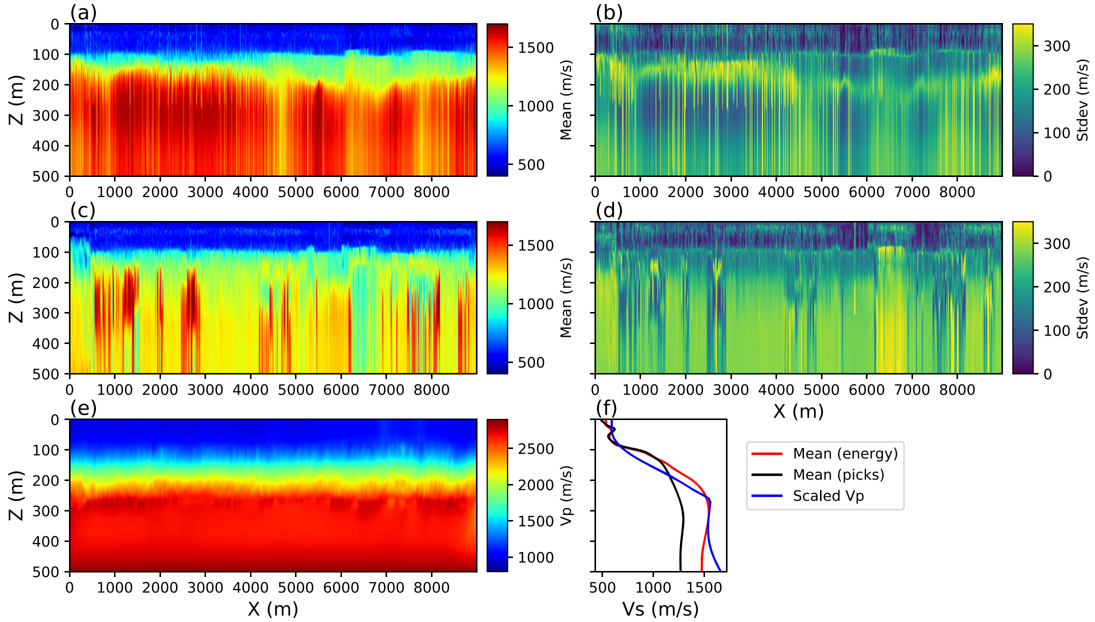

Figure 9. The mean and standard deviation model obtained using the energy likelihood function (a and b) and the traditional method (c and d) along a vertical section at $Y = 875$ m. (e) shows the vertical section of the *P*-wave velocity model obtained using refraction tomography. Panel (f) shows the average shear velocity profiles across the section obtained using the suite of methods. The *P*-wave velocity is converted to shear velocity using a Vp/Vs ratio of 1.73.

directly use the spectrum obtained between each source-receiver or receiver–receiver as the data to conduct phase or group velocity tomography using earthquake-generated surface waves or ambient noise data (Cauchie & Saccorotti 2013; Panning *et al.* 2015; Gaudot *et al.* 2021).

6 CONCLUSION

In this study, we introduced a new likelihood function for seismic surface wave dispersion inversion, called the energy likelihood function which directly uses the spectrum as data. We applied the new likelihood function to image the subsurface shear velocity structure using surface wave data recorded by a dense array, and compared the results with those obtained using the traditional method. The results showed that the new likelihood function can take account of all information contained in the spectrum and produce a less biased result than that obtained using the traditional method. In addition, the velocity model obtained using the new likelihood function is more similar to the *P*-wave velocity structure than that obtained using the traditional method. Because the new likelihood function directly uses the spectrum as data, it requires less effort to apply to large data sets than the traditional method, since the latter requires that we determine the phase velocity for each spectrum either manually or automatically and cannot be applied easily to large data sets. We thus conclude that the new energy likelihood function provides a powerful way to conduct surface wave dispersion inversion.

ACKNOWLEDGMENTS

The authors thank the Edinburgh Imaging Project sponsors (BP and Total) for supporting this research. We would like thank BP and ADNOC for providing the seismic data and CGG for the refraction tomography model. For the purpose of open access, the author has applied a ’Creative Commons Attribution (CC BY)’ licence to any Author Accepted Manuscript version arising.

DATA AVAILABILITY

The data underlying this paper were provided by BP and ADNOC by permission. Data can be accessed on request to BP and its partner ADNOC.

REFERENCES

Aki, K. & Richards, P.G., 1980. *Quantitative Seismology,* 2nd edn, University Science Books.

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

Biswas, R. & Sen, M., 2017. 2D full-waveform inversion and uncertainty estimation using the reversible jump Hamiltonian Monte Carlo, in *Proceedings of the SEG Technical Program Expanded Abstracts 2017,* pp. 1280–1285, Society of Exploration Geophysicists.

Bodin, T. & Sambridge, M., 2009. Seismic tomography with the reversible jump algorithm, *Geophys. J. Int.,* **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, *J. geophys. Res.,* **117**(B2), doi:10.1029/2011JB008560.

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

Buriola, F., Mills, K., Cooper, S., Hollingworth, S., Crosby, A. & Ourabah, A., 2021. Ultra-high density land nodal seismic—processing challenges and rewards, in *Proceedings of the 82nd EAGE Annual Conference and Exhibition,* 18–21 October 2021, Amsterdem, Netherlands.

Cauchie, L. & Saccorotti, G., 2013. Probabilistic inversion of Rayleigh-wave dispersion data: an application to Mt. Etna, Italy, *J. Seismol.,* **17**(2), 335–346.

Cercato, M., 2009. Addressing non-uniqueness in linearized multichannel surface wave inversion, *Geophys. Prospect.,* **57**(1), 27–47.

Chmiel, M. *et al.*, 2019. Ambient noise multimode Rayleigh and Love wave tomography to determine the shear velocity structure above the Groningen gas field, *Geophys. J. Int.,* **218**(3), 1781–1795.

Curtis, A., Trampert, J., Snieder, R. & Dost, B., 1998. Eurasian fundamental mode surface wave phase velocities and their relationship with tectonic structures, *J. geophys. Res.,* **103**(B11), 26 919–26 947.

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

Dettmer, J., Dosso, S.E. & Holland, C.W., 2010. Trans-dimensional geoacoustic inversion, *J. acoust. Soc. Am.,* **128**(6), 3393–3405.

Duane, S., Kennedy, A.D., Pendleton, B.J. & Roweth, D., 1987. Hybrid Monte Carlo, *Phys. Lett. B,* **195**(2), 216–222.

Dziewonski, A., Bloch, S. & Landisman, M., 1969. A technique for the analysis of transient seismic signals, *Bull. seism. Soc. Am.,* **59**(1), 427–444.

Earp, S., Curtis, A., Zhang, X. & Hansteen, F., 2020. Probabilistic neural network tomography across Grane field (North Sea) from surface wave dispersion data, *Geophys. J. Int.,* **223**(3), 1741–1757.

Ekström, G., 2011. A global model of Love and Rayleigh surface wave dispersion and anisotropy, 25-250 s, *Geophys. J. Int.,* **187**(3), 1668–1686.

Ekström, G., Tromp, J. & Larson, E.W., 1997. Measurements and global models of surface wave propagation, *J. geophys. Res.,* **102**(B4), 8137–8157.

Estève, C., Gosselin, J., Audet, P., Schaeffer, A., Schutt, D. & Aster, R., 2021. Surface-wave tomography of the northern Canadian cordillera using earthquake Rayleigh wave group velocities, *J. geophys. Res.,* **126**(8), e2021JB021960, doi:10.1029/2021JB021960.

Evernden, J.F., 1953. Direction of approach of Rayleigh waves and related problems (Part I), *Bull. seism. Soc. Am.,* **43**(4), 335–374.

Evernden, J.F., 1954. Direction of approach of Rayleigh waves and related problems (Part II), *Bull. seism. Soc. Am.,* **44**(2A), 159–184.

Fichtner, A., Zunino, A. & Gebraad, L., 2018. Hamiltonian Monte Carlo solution of tomographic inverse problems, *Geophys. J. Int.,* **216**(2), 1344–1363.

Forsyth, D.W., Li, A., Levander, A. & Nolet, G., 2005. Array analysis of two-dimensional variations in surface wave phase velocity and azimuthal anisotropy in the presence of multipathing interference, in *Seismic Earth: Array Analysis of Broadband Seismograms,Geophysical Monograph Series,* Vol. **157,** pp. 81–97, eds Levander, A. & Nolet, G., American Geophysical Union.

Galetti, E., Curtis, A., Meles, G.A. & Baptie, B., 2015. Uncertainty loops in travel-time tomography from nonlinear wave physics, *Phys. Rev. Lett.,* **114**(14), doi:10.1103/PhysRevLett.114.148501.

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, *Geophys. J. Int.,* **208**(1), 36–58.

Gaudot, I., Beucler, É., Mocquet, A., Drilleau, M., Haugmard, M., Bonnin, M., Aertgeerts, G. & Leparoux, D., 2021. 3-D crustal Vs model of western France and the surrounding regions using Monte Carlo inversion of seismic noise cross-correlation dispersion diagrams, *Geophys. J. Int.,* **224**(3), 2173–2188.

Gebraad, L., Boehm, C. & Fichtner, A., 2020. Bayesian elastic full-waveform inversion using Hamiltonian Monte Carlo, *J. geophys. Res.,* **125**(3), e2019JB018428, doi:10.1029/2019JB018428.

Granados, I., Calò, M. & Ramos, V., 2019. Noisy Dispersion Curve Picking (NDCP): a Matlab package for group velocity dispersion picking of seismic surface waves, *Comput. Geosci.,* **133,** doi:10.1016/j.cageo.2019.104315.

Grandjean, G. & Bitri, A., 2006. 2M-SASW: Multifold multichannel seismic inversion of local dispersion of Rayleigh waves in laterally heterogeneous subsurfaces: application to the Super-Sauze earthflow, france, *Near Surf. Geophys.,* **4**(6), 367–375.

Green, P.J., 1995. Reversible jump Markov chain Monte Carlo computation and Byesian model determination, *Biometrika,* **82**(4), 711–732.

Hallo, M., Imperatori, W., Panzera, F. & Fäh, D., 2021. Joint multizonal transdimensional Bayesian inversion of surface wave dispersion and ellipticity curves for local near-surface imaging, *Geophys. J. Int.,* **226**(1), 627–659.

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

Herrin, E. & Goforth, T., 1977. Phase-matched filters: application to the study of Rayleigh waves, *Bull. seism. Soc. Am.,* **67**(5), 1259–1275.

Herrmann, R.B., 2013. Computer programs in seismology: an evolving tool for instruction and research, *Seismol. Res. Lett.,* **84**(6), 1081–1088.

Hou, S., Zheng, D., Miao, X. & Haacke, R., 2016. Multi-modal surface wave inversion and application to North Sea OBN data, in *Proceedings of the 78th EAGE Conference and Exhibition 2016,* Vol. **2016,** pp. 1–5, European Association of Geoscientists & Engineers.

Killingbeck, S., Livermore, P., Booth, A. & West, L., 2018. Multimodal layered transdimensional inversion of seismic dispersion curves with depth constraints, *Geochem. Geophys. Geosyst.,* **19**(12), 4957–4971.

Kotsi, M., Malcolm, A. & Ely, G., 2020. Uncertainty quantification in time-lapse seismic imaging: a full-waveform approach, *Geophys. J. Int.,* **222**(2), 1245–1263.

Lai, C.G., Mangriotis, M.-D. & Rix, G.J., 2014. An explicit relation for the apparent phase velocity of Rayleigh waves in a vertically heterogeneous elastic half-space, *Geophys. J. Int.,* **199**(2), 673–687.

Landau, L.D. & Lifshitz, E.M., 2013. *Statistical Physics,* Vol. **5,** Elsevier.

Lee, M. & Collett, T., 2008. Integrated analysis of well logs and seismic data to estimate gas hydrate concentrations at Keathley Canyon, Gulf of Mexico, *Mar. Petrol. Geol.,* **25**(9), 924–931.

Levshin, A. & Ritzwoller, M., 2001. Automated detection, extraction, and measurement of regional surface waves, in *Monitoring the Comprehensive Nuclear-Test-Ban Treaty: Surface Waves,* 1531–1545, eds Levshin, A.L. & Ritzwoller, M.H., Pageoph Topical Volumes, Birkhäuser.

Levshin, A., Ratnikova, L. & Berger, J., 1992. Peculiarities of surface-wave propagation across central Eurasia, *Bull. seism. Soc. Am.,* **82**(6), 2464–2493.

Levshin, A.L., Pisarenko, V. & Pogrebinsky, G., 1972. On a frequency-time analysis of oscillations, *Ann. Geophys.,* **28,** 211–218.

Li, J., Feng, Z. & Schuster, G., 2017. Wave-equation dispersion inversion, *Geophys. J. Int.,* **208**(3), 1567–1578.

Liu, Z., Li, J., Hanafy, S.M. & Schuster, G., 2018. 3D wave-equation dispersion inversion of surface waves, in *Proceedings of the 2018 SEG International Exposition and Annual Meeting,* 14-19 October 2018, Anaheim Convention Center, Anaheim, CA, USA, OnePetro.

Malinverno, A., 2002. Parsimonious Byesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem, *Geophys. J. Int.,* **151**(3), 675–688.

Malinverno, A. & Briggs, V.A., 2004. Expanded uncertainty quantification in inverse problems: Hierarchical Byes and empirical Byes, *Geophysics,* **69**(4), 1005–1016.

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

Meier, U., Curtis, A. & Trampert, J., 2007b. A global crustal model constrained by nonlinearised inversion of fundamental mode surface waves, *Geophys. Res. Lett.,* **34**(16), doi:10.1029/2007GL030989.

Metropolis, N. & Ulam, S., 1949. The Monte Carlo method, *J. Am. Stat. Assoc.,* **44**(247), 335–341.

Minsley, B.J., 2011. A trans-dimensional Bayesian Markov chain Monte Carlo algorithm for model assessment using frequency-domain electromagnetic data, *Geophys. J. Int.,* **187**(1), 252–272.

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, *Geophys. J. Int.,* **198**(3), 1514–1525.

Mousavi, S., Tkalčić, H., Hawkins, R. & Sambridge, M., 2021. Lowermost mantle shear-velocity structure from hierarchical trans-dimensional Bayesian tomography, *J. geophys. Res.,* **126**(9), e2020JB021557, doi:10.1029/2020JB021557.

Naskar, T. & Kumar, J., 2022. MATLAB codes for generating dispersion images for ground exploration using different multichannel analysis of surface wave transforms, *Geophysics,* **87**(3), F15–F24.

Nawaz, M.A. & Curtis, A., 2018. Variational Bayesian inversion (VBI) of quasi-localized seismic attributes for the spatial distribution of geological facies, *Geophys. J. Int.,* **214**(2), 845–875.

Olafsdottir, E.A., Erlingsson, S. & Bessason, B., 2018. Tool for analysis of multichannel analysis of surface waves (masw) field data and evaluation of shear wave velocity profiles of soils, *Can. Geotech. J.,* **55**(2), 217–233.

Ourabah, A. & Crosby, A., 2020. A 184 million traces per km^2 seismic survey with nodes-acquisition and processing, in *Proceedings of the 90th SEG International Exposition and Annual Meeting,* pp. 41–45, Society of Exploration Geophysicists, doi:10.1190/segam2020-3426358.1.

O'Neill, A. & Matsuoka, T., 2005. Dominant higher surface-wave modes and possible inversion pitfalls, *J. Environ. Eng. Geophys.,* **10**(2), 185–201.

Panning, M.P., Beucler, É., Drilleau, M., Mocquet, A., Lognonné, P. & Banerdt, W.B., 2015. Verifying single-station seismic approaches using Earth-based data: Preparation for data return from the InSight mission to Mars, *Icarus,* **248,** 230–242.

Park, C.B. & Miller, R.D., 2008. Roadside passive multichannel analysis of surface waves (MASW), *J. Environ. Eng. Geophys.,* **13**(1), 1–11.

Park, C.B., Miller, R.D. & Xia, J., 1998. Imaging dispersion curves of surface waves on multi-channel record, in *Proceedings of the SEG Technical Program Expanded Abstracts 1998,* pp. 1377–1380, Society of Exploration Geophysicists.

Park, C.B., Miller, R.D. & Xia, J., 1999. Multichannel analysis of surface waves, *Geophysics,* **64**(3), 800–808.

Piana Agostinetti, N., Giacomuzzi, G. & Malinverno, A., 2015. Local three-dimensional earthquake tomography by trans-dimensional Monte Carlo sampling, *Geophys. J. Int.,* **201**(3), 1598–1617.

Pickett, G.R., 1963. Acoustic character logs and their applications in formation evaluation, *J. Petrol. Technol.,* **15**(06), 659–667.

Ray, A. & Key, K., 2012. Bayesian inversion of marine CSEM data with a trans-dimensional self parametrizing algorithm, *Geophys. J. Int.,* **191**(3), 1135–1151.

Ritzwoller, M.H. & Levshin, A.L., 1998. Eurasian surface wave tomography: group velocities, *J. geophys. Res.,* **103**(B3), 4839–4878.

Roberts, G.O., Tweedie, R.L. *et al.*, 1996. Exponential convergence of Langevin distributions and their discrete approximations, *Bernoulli,* **2**(4), 341–363.

Rosalia, S., Cummins, P., Widiyantoro, S., Yudistira, T., Nugraha, A.D. & Hawkins, R., 2020. Group velocity maps using subspace and transdimensional inversions: ambient noise tomography in the western part of Java, Indonesia, *Geophys. J. Int.,* **220**(2), 1260–1274.

Russell, D.R., Herrmann, R.B. & Hwang, H.-J., 1988. Application of frequency variable filters to surface-wave amplitude analysis, *Bull. seism. Soc. Am.,* **78**(1), 339–354.

Saygin, E. *et al.*, 2015. Imaging architecture of the Jakarta Basin, Indonesia with transdimensional inversion of seismic noise, *Geophys. J. Int.,* **204**(2), 918–931.

Shapiro, N. & Ritzwoller, M., 2002. Monte-Carlo inversion for a global shear-velocity model of the crust and upper mantle, *Geophys. J. Int.,* **151**(1), 88–105.

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

Siahkoohi, A., Rizzuti, G. & Herrmann, F.J., 2020. Uncertainty quantification in imaging and automatic horizon tracking—a Bayesian deep-prior based approach, in *Proceedings of the SEG Technical Program Expanded Abstracts 2020,* pp. 1636–1640, Society of Exploration Geophysicists.

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, *Geophys. J. Int.,* **151**(3), 738–754.

Smith, J.D., Ross, Z.E., Azizzadenesheli, K. & Muir, J.B., 2022. HypoSVI: hypocentre inversion with Stein variational inference and physics informed neural networks, *Geophys. J. Int.,* **228**(1), 698–710.

Tokimatsu, K., Tamura, S. & Kojima, H., 1992. Effects of multiple modes on rayleigh wave dispersion characteristics, *J. Geotech. Eng.,* **118**(10), 1529–1543.

Trampert, J. & Woodhouse, J.H., 1995. Global phase velocity maps of Love and Rayleigh waves between 40 and 150 seconds, *Geophys. J. Int.,* **122**(2), 675–690.

van Heijst, H.J. & Woodhouse, J., 1997. Measuring surface-wave overtone phase velocities using a mode-branch stripping technique, *Geophys. J. Int.,* **131**(2), 209–230.

Wang, Z., Bovik, A.C., Sheikh, H.R. & Simoncelli, E.P., 2004. Image quality assessment: from error visibility to structural similarity, *IEEE Trans. Image Process.,* **13**(4), 600–612.

Wielandt, E., 1993. Propagation and structural interpretation of non-plane waves, *Geophys. J. Int.,* **113**(1), 45–53.

Xia, J., 2014. Estimation of near-surface shear-wave velocities and quality factors using multichannel analysis of surface-wave methods, *J. appl. Geophys.,* **103,** 140–151.

Xia, J., Miller, R.D., Park, C.B. & Tian, G., 2003. Inversion of high frequency surface waves with fundamental and higher modes, *J. appl. Geophys.,* **52**(1), 45–57.

Xia, J., Chen, C., Li, P. & Lewis, M., 2004. Delineation of a collapse feature in a noisy environment using a multichannel surface wave technique, *Geotechnique,* **54**(1), 17–27.

Xiang, E., Guo, R., Dosso, S.E., Liu, J., Dong, H. & Ren, Z., 2018. Efficient hierarchical trans-dimensional Bayesian inversion of magnetotelluric data, *Geophys. J. Int.,* **213**(3), 1751–1767.

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, *Geophys. J. Int.,* **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.

Yuan, H. & Bodin, T., 2018. A probabilistic shear wave velocity model of the crust in the central West Australian craton constrained by transdimensional inversion of ambient noise dispersion, *Tectonics,* **37**(7), 1994–2012.

Zhang, S.X. & Chan, L.S., 2003. Possible effects of misidentified mode number on Rayleigh wave inversion, *J. appl. Geophys.,* **53**(1), 17–29.

Zhang, X. & Curtis, A., 2020a. Seismic tomography using variational inference methods, *J. geophys. Res.,* **125**(4), e2019JB018589, doi:10.1029/2019JB018589.

Zhang, X. & Curtis, A., 2020b. Variational full-waveform inversion, *Geophys. J. Int.,* **222**(1), 406–411.

Zhang, X. & Curtis, A., 2021a. Bayesian full-waveform inversion with realistic priors, *Geophysics,* **86**(5), 1–20.

Zhang, X. & Curtis, A., 2021b. Bayesian geophysical inversion using invertible neural networks, *J. geophys. Res.,* **126**(7), doi:10.1029/2021JB022320.

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

Zhang, X., Hansteen, F., Curtis, A. & de Ridder, S., 2020a. 1D, 2D and 3D Monte Carlo ambient noise tomography using a dense passive seismic array installed on the North Sea seabed, *J. geophys. Res.,* **125**(2), e2019JB018552, doi:10.1029/2019JB018552.

Zhang, X., Roy, C., Curtis, A., Nowacki, A. & Baptie, B., 2020b. Imaging the subsurface using induced seismicity and ambient noise: 3-D tomographic Monte Carlo joint inversion of earthquake body wave traveltimes and surface wave dispersion, *Geophys. J. Int.,* **222**(3), 1639–1655.

Zhao, X., Curtis, A. & Zhang, X., 2021. Bayesian seismic tomography using normalizing flows, *Geophys. J. Int.,* **228**(1), 213–239.

Zheng, D., Saygin, E., Cummins, P., Ge, Z., Min, Z., Cipta, A. & Yang, R., 2017. Transdimensional Byesian seismic ambient noise tomography across SE Tibet, *J. Asian Earth Sci.,* **134,** 86–93.

Zhu, D. & Gibson, R., 2018. Seismic inversion and uncertainty quantification using transdimensional Markov chain Monte Carlo method, *Geophysics,* **83**(4), R321–R334.

Zielhuis, A. & Nolet, G., 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, *Geophys. J. Int.,* **197**(1), 630–635.

APPENDIX: MODE VALIDATION

To understand the two branches that appear at short periods (<0.25 s) in the *f*–*c* spectrum in Fig. 2, and in particular to test whether the two branches represent two different modes, we first perform an inversion using the rj-McMC algorithm using one of the branches as data (black dots in Fig. A1). The prior pdf of the shear velocity is set to be a Uniform distribution between 300 and 1500 m s^{-1}. For the likelihood function we use the traditional Gaussian distribution. The inversion is then conducted in the same way as described in Section 3.5 with the same prior pdf for the number of layers and noise hyperparameters and the same proposal pdf. Fig. A1(b) shows the obtained mean and the marginal distribution of the shear velocity. We then use the mean shear velocity profile to simulate phase velocities of the fundamental mode and the first overtone. While the modelled fundamental-mode phase velocity (black dashed line in Fig. A1a) fits the data used, the modelled phase velocity of the first overtone (while line in Fig. A1a) is significantly closer to the mode with velocity higher than 600 m s^{-1} (black triangles) than to the other branch appearing in the fundamental mode. This clearly demonstrates that the two branches are unlikely to represent two different modes, and instead represent an effect such as the multipathing of the seismic energy of the fundamental model or the strong lateral heterogeneity.

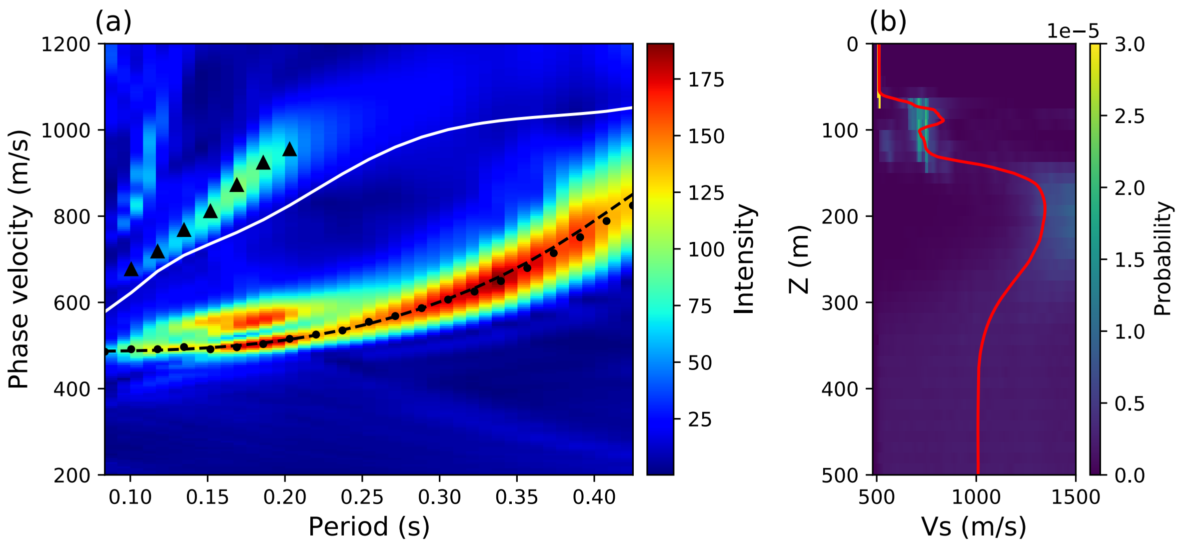

Figure A1. (a) The spectrum obtained using *f*–*c* analysis for the location in Fig. 1(b) (red star). Black dots show the picked phase velocities. (b) The marginal posterior distribution of shear velocities obtained using only the picked phase velocities of the fundamental mode. The red line shows the mean shear velocity profile. The black dashed line and the white line in (a) show the predicted phase velocities for the fundamental mode and the first overtone, respectively, modelled using the posterior mean model.