Near-surface imaging with Rayleigh and Love waves extracted from DAS ambient noise data

Yumin Zhao[1], Yunyue Elita Li[2,1], and Bei Li[1]

[1]Department of Civil and Environmental Engineering, National University of Singapore, Singapore
[2]Department of Earth, Atmospheric, and Planetary Sciences, Purdue University, West Lafayette, Indiana, U.S.A.

Key Points:

- We simulate DAS ambient noise data with the source propagation direction and Rayleigh-to-Love waves ratio estimated from the observed data.
- We invert the shear-wave velocity model from the Rayleigh and Love waves extracted from the 2-minute DAS ambient noise data simultaneously.
- The resolved velocity model matches the local geological structure.

Corresponding author: Y. E. Li, `elitali@purdue.edu`

Abstract

The horizontally deployed distributed acoustic sensing (DAS) array records both Rayleigh and Love waves generated by the ambient noise sources. Common virtual-shot gathers and dispersion curves extracted from the DAS ambient noise data indicate the time lags and phase velocities of a combination of those of Rayleigh and Love waves. However, with the assumption that ambient noise sources are uniformly distributed or mainly travel along the axial direction of the linear DAS array, current studies suggest that only Rayleigh waves are extracted from the DAS ambient noise data. As a result, the near-surface shear-wave velocity model is obtained from Rayleigh wave inversion, which is unreliable. To solve this problem, we first estimate the source propagation direction and Rayleigh-to-Love waves (R/L) ratio from the DAS data. With this information, we then simulate the DAS ambient noise data given different velocity models in the forward modeling process and invert the shear-wave velocity model from Rayleigh and Love waves simultaneously, using the dispersion spectrum inversion method. We resolve the marginal distribution of each parameter (i.e., thickness and shear-wave velocity of each layer) of the velocity model instead of a single best-fitting one. Synthetic data examples show that our method recovers the velocity model well from the mixed surface waves. The velocity models obtained from the 2-minute DAS ambient noise data recorded by the Stanford DAS-1 array demonstrate that our approach provides reliable shear-wave velocity models, which are more consistent with the local geological structure than the inversion results assuming predominant Rayleigh waves.

Plain Language Summary

Surface waves, mainly containing Rayleigh and Love waves, are the main component of seismic waves generated by ambient noise sources. The horizontally deployed distributed acoustic sensing (DAS) array records both Rayleigh and Love waves. As ambient noise sources are not perfectly uniformly distributed in urban environments and do not precisely travel along the axial direction of the array, the extracted surface waves contain both Rayleigh and Love waves. Thus, inverting the near-surface shear-wave velocity model from the data containing both Rayleigh and Love waves with a standard Rayleigh wave inversion is not reliable. With the source propagation direction and Rayleigh-to-Love waves (R/L) ratio estimated from the DAS ambient noise data, we propose to invert the near-surface shear-wave velocity model from Rayleigh and Love waves simultaneously using the dispersion spectrum inversion method. We investigate the reliability of this approach with several synthetic data examples. Finally, we apply the method to the field DAS ambient noise data recorded by the Stanford DAS-1 array. The results show more consistency with the local geological structure than that obtained from a standard Rayleigh wave inversion.

1 Introduction

Seismic ambient noise, generated by both natural and anthropological activities at the Earth's surface, mainly contains surface waves (e.g., Friedrich et al., 1998; Ekström, 2001). In seismology, surface waves are divided into Rayleigh and Love waves based on the relationship between particle motion direction and propagation direction. Rayleigh waves move in an elliptical retrograde motion that consists of longitudinal and transverse motions. Love waves move in transverse motion in the horizontal plane. The longitudinal motion is parallel to the wave propagation direction, while the transverse motion is perpendicular to the wave propagation direction. Surface waves usually travel along the Earth's surface and penetrate to a depth that depends on their wavelength and carries information about the shear-wave velocities within the depth range.

High-frequency (above 1 Hz) seismic ambient noise recorded in urban environments is mainly generated by traffic (Díaz et al., 2017). Previous studies show that the traffic generates both Rayleigh and Love waves. Behm et al. (2014) demonstrated that the traffic noise (2-6 Hz) generated by the motor vehicles running on a nearby state road contains both Rayleigh and Love waves by analyzing the seismic ambient noise recorded by a three-component geophone. Nakata (2016) extracted Love waves (4-20 Hz) from the traffic noise recorded by a single horizontal-component geophone array and inverted the shear-wave velocity model with Love waves. Song et al. (2021) extracted a mix of Rayleigh and Love waves from the traffic noise recorded by a distributed acoustic sensing (DAS) array and inverted the shear-wave velocity model with Rayleigh and Love waves simultaneously.

Traditional seismic sensors, such as geophones, are only sensitive to the particle motions parallel to their axes. All components of a three-component geophone record Rayleigh waves. In comparison, only the two horizontal components of the geophone record Love waves. Thus, each horizontal component records a mixture of Rayleigh and Love waves. In ambient noise seismology, surface waves recorded by the vertical component of a geophone array are commonly used for near-surface characterization (e.g., Brooks et al., 2009; Zhang et al., 2019; Nilot et al., 2019) since they only contain Rayleigh waves which can be generated with any given subsurface structure (Behm & Snieder, 2013). Seismic data recorded by the two horizontal components of a three-component geophone are also getting increasing attention since they provide more information that prevents mode misidentification (Boaga et al., 2013) and can be used as constraints on model parameters to improve the inversion results (Ikeda et al., 2015). With the spectra of the seismic ambient noise data recorded by the horizontal components and the vertical component, the horizontal-to-vertical spectral ratio (Nakamura, 1989) can be calculated and used to invert the shear-wave velocity model (e.g., Fäh et al., 2003; Arai & Tokimatsu, 2004; Picozzi et al., 2005; Lunedei & Albarello, 2015; Sivaram et al., 2018). Rayleigh and Love waves can be used for joint inversion of the shear-wave velocity model (Joh et al., 2006; Hamimu et al., 2011; Chmiel et al., 2019; Yin et al., 2020), which reduces the uncertainties of the shear-wave velocity model parameters. Besides, Rayleigh and Love waves recorded by the two horizontal components of geophones can be separated by rotating them in the radial/transverse direction. Love waves may have better quality than Rayleigh waves and are dominant in the horizontal recordings (e.g., Lin et al., 2008; Jay et al., 2012; Behm & Snieder, 2013; Behm et al., 2014). Previous studies separated the Rayleigh and Love waves recorded by the horizontal components of the geophones and estimated Rayleigh and Love waves phase or group velocity maps (e.g., Lin et al., 2008; Chmiel et al., 2019).

DAS is a recently developed technology in which fiber-optic cables can be taken as densely distributed sensors that provide high-resolution seismic recordings. The fiber-optic cable is usually horizontally deployed when used for urban near-surface imaging and monitoring. Thus, both Rayleigh and Love waves are recorded by the DAS system. The extracted common virtual-shot gather contains both Rayleigh and Love waves since the sources in the urban environments usually come from fixed locations (Dou et al., 2017; Zhang et al., 2019; Nilot et al., 2019). Unlike geophones, Rayleigh and Love waves in the DAS ambient noise cannot be separated without enough information about the source. Previous studies assume the ambient noise sources either are uniformly distributed or mainly travel along the axial direction of the linear DAS array (e.g., Dou et al., 2017; Zeng et al., 2017; Martin, 2018). As a result, Rayleigh waves are assumed to be extracted and used to invert the shear-wave velocity model. However, the assumptions may lead to two kinds of errors. One is the overestimation of the velocities since the virtual source may not propagate along the axial direction of the DAS array. Besides, the extracted surface waves contain Love waves, of which the phase velocities usually are higher than those of Rayleigh waves (Behm & Snieder, 2013). The other one is the inaccurate estimation of the layer thickness. Yin et al. (2014) shows that the penetration depth of fundamental mode Love waves is 1.3-1.4 times deeper than that of fundamental mode Rayleigh

waves, given the same wavelength. Ignoring the Love waves may lead to the estimation error of the layer thickness. To the best of our knowledge, only one study on near-surface imaging with DAS ambient noise data considered Love waves in the extracted common virtual-shot gathers and inverted for the shear-wave velocity model from Rayleigh and Love waves simultaneously (Song et al., 2021). However, they did not specify the content of Rayleigh and Love waves in the DAS data simulation process.

Near-surface shear-wave velocity models are commonly inverted from the dispersion curves, which are extracted from the common virtual-shot gather with the phase-shift method (Park et al., 1998). However, picking errors may occur in the extracted dispersion curve, especially at lower frequencies where the resolution of the dispersion spectrum is low. These errors would lead to the unreliability of the inverted velocity model. In addition, mode misidentification may emerge when "mode-kissing" presents in the dispersion curves and leads to bias in the subsequent velocity model inversion. To overcome these challenges, dispersion spectrum inversion methods were proposed to invert the shear-wave velocity model (e.g., Ryden & Park, 2006; Dou & Ajo-Franklin, 2014; Song et al., 2021). Instead of finding the velocity model(s) which minimizes the difference between the observed and theoretical dispersion curves, the dispersion spectrum inversion method is designed to resolve the velocity model(s) by minimizing the difference between the observed and theoretical dispersion spectra.

In this paper, we invert the shear-wave velocity model from the Rayleigh and Love waves extracted from the DAS ambient noise data simultaneously with the dispersion spectrum inversion method. In the forward modeling procedure, we simulate the DAS ambient noise that contains both Rayleigh and Love waves with the ambient noise propagation direction and R/L ratio estimated from the observed data (Zhao et al., 2023), extract the common virtual-shot gather, and calculate the dispersion spectrum. Due to the nonlinearity of this problem, we invert the distribution of the shear-wave velocity and thickness of each layer with the Markov Chain Monte Carlo (McMC) algorithm. Synthetic data examples show that the proposed method yields a reliable shear-wave velocity model from a mix of Rayleigh and Love waves. Field data results demonstrate that the proposed method provides reliable shear-wave velocity models which are consistent with the local geological structure.

2 Methodology

2.1 Modelling the DAS ambient noise

DAS ambient noise data can be simulated with known R/L ratio, source propagation direction, and array geometry. We assume that N_s ambient noise sources outside the horizontally deployed close L-shape DAS array randomly generate plane surface waves during the recording period. Each source generates Rayleigh and Love waves simultaneously, and both are recorded by the DAS array. Let m denote the R/L ratio, and a_{Li} denotes the pulse amplitude of the Love waves generated by the i^{th} source at time t_i, $i = 1, 2, ..., N_s$. a_{Ri} and t_i follow a uniform distribution. Then the pulse amplitude of the Rayleigh waves generated by the i^{th} source is ma_{Li}. We also assume that all the ambient noise sources propagate in the same direction and the intersection angle between the propagation direction of the ambient noise sources and one leg of the array is $theta$. Then, the amplitude response of the point-wise Rayleigh and Love waves strain recorded by this leg of the array can be expressed as:

$$s_R(x,\omega) = A(\omega)\sum_{i=1}^{N_s} -\frac{i\omega\cos^2\theta_i}{v(\omega)} m a_{Li} e^{-i\omega(t_i + x\cos\theta_i / v_R(\omega))}, \tag{1}$$

and

$$s_L(x,\omega) = A(\omega)\sum_{i=1}^{N_s} -\frac{i\omega\cos\theta_i\sin\theta_i}{v(\omega)} a_{Li} e^{-i\omega(t_i + x\cos\theta_i / v_L(\omega))}, \tag{2}$$

where $A(\omega)$ is the amplitude spectrum of the source wavelet, and x denotes the distance between the current channel and the one at the intersection of the L-shape DAS array. $v_R(\omega)$ and $v_L(\omega)$ are Rayleigh and Love waves phase velocity at angular frequency ω, respectively.

Let L_g denotes gauge length, then the DAS Rayleigh waves measurement $\bar{s}_R(x,\omega)$ can be expressed as:

$$\bar{s}_R(x,\omega) = \frac{1}{L_g}\int_{x-L_g/2}^{x+L_g/2} s_R(l,\omega)dl, \tag{3}$$

and DAS Love waves measurement can be expressed as:

$$\bar{s}_L(x,\omega) = \frac{1}{L_g}\int_{x-L_g/2}^{x+L_g/2} s_L(l,\omega)dl, \tag{4}$$

Then, the DAS ambient noise data recorded by one leg of the L-shape DAS array can be expressed as:

$$\bar{s}(x,\omega) = \bar{s}_R(x,\omega) + \bar{s}_L(x,\omega), \tag{5}$$

Similarly, we can obtain the DAS ambient noise data recorded by the other leg of the array. Finally, we can get the DAS ambient noise data recorded by the whole array.

2.2 Seismic interferometry

Seismic interferometry is a technique that extracts coherent signals from the seismic ambient noise data by cross-correlating the data recorded by two receivers. Assuming a wavefield is excited at s and recorded by two receivers A and B deployed at locations x_A and x_B, then $U(x_A,s)$ and $U(x_B,s)$ are the amplitude spectra of the signals recorded by the two receivers, respectively. The cross-correlation between these two signals in the frequency domain is as follows:

$$C_{AB} = U^*(x_A,s)U(x_B,s), \tag{6}$$

where the asterisk denotes the complex conjugate. The cross-correlation result is equivalent to the source generated at receiver A (virtual source) and recorded by receiver B, which is often referred to as empirical Green's function (Claerbout, 1968; Wapenaar, 2004).

Cross-coherence (Aki, 1957) is another method for empirical Green's function extraction from seismic ambient noise data. Let equation 6 be divided by the product of the amplitude spectra of the signals recorded at receivers A and B, and then we can obtain the cross-coherence between the two signals:

$$H_{AB} = \frac{U^*(x_A,s)U(x_B,s)}{|U(x_A,s)||U(x_B,s)|}. \tag{7}$$

Equation 7 shows that this method preserves the phase information while discards the amplitude information so that the amplitude variations among traces are eliminated (Nakata et al., 2011). We use cross-coherence to extract the empirical Green's function for both synthetic and field data examples.

2.3 Phase-shift method

We use the Phase-Shift method (Park et al., 1998) to calculate the phase velocity spectrum from the correlogram. Let x_k and $\bar{s}_k$ represent the offset and seismic recording of the k^{th} channel, respectively. The method can be implemented using the following two steps:

(a) Calculate the amplitude spectrum via Fourier Transform and then apply amplitude normalization.

$$S_k(f, x_k) = \int_{-\infty}^{\infty} \bar{s}_k e^{-i2\pi ft} dt, \tag{8}$$

$$S_{k,norm}(f, x_k) = S_k(f, x_k)/|S_k(f, x_k)|. \tag{9}$$

(b) Apply offset-dependent phase shift $2\pi f x_k/v$ to the normalized amplitude spectrum $S_{k,norm}(f, x_k)$, and sum over all the offsets of the correlogram:

$$\tilde{S}(f, v) = \sum_{k=1}^{N} S_{k,norm}(f, x_k) e^{i2\pi f x_k/v}. \tag{10}$$

The spectrum peaks in the dispersion image correspond to the signals at a specific frequency that travels with the same velocity.

2.4 Dispersion spectrum inversion

We assume the observed normalized dispersion spectrum can be represented by the superposition of the normalized theoretical dispersion spectrum with random noise:

$$\mathbf{D}_{obs} = \mathbf{D}(\mathbf{m}) + \mathbf{N}, \tag{11}$$

where $\mathbf{m}$ is the parameters to be solved and $\mathbf{m} = [\mathbf{V_S}, \mathbf{d}]$. $\mathbf{V_S} = [v_{S1}, v_{S2}, ..., v_{Sn}]$ and $\mathbf{d} = [d_1, d_2, ..., d_n]$ (n is the layer number) denote the shear-wave velocity and thickness of each layer, respectively. $\mathbf{D}_{obs}$, $\mathbf{D}(\mathbf{m})$, and $\mathbf{N}$ denote the normalized observed dispersion spectrum, normalized theoretical dispersion spectrum, and the random noise, respectively.

We invert the near-surface shear-wave velocity model by minimizing the misfit between the observed and theoretical dispersion spectrum. The misfit can be formulated as

$$J(\mathbf{m}) = ||\mathbf{D}_{obs} - \mathbf{D}(\mathbf{m})||_2^2. \tag{12}$$

The optimization of equation 12 is a non-linear inversion problem and generally has non-unique solutions. Thus, one or multiple best data fitting velocity models can be arbitrary and not represent all the acceptable models (Bodin et al., 2012). Therefore, we use Bayesian inference to solve the marginal distribution of each parameter with the McMC method.

According to Bayesian inference (Bayes, 1763), the posterior distribution $p(\mathbf{m}|\mathbf{D}_{obs})$ is the probability that the inverted velocity model $\mathbf{m}$ is true, given the observed dispersion spectrum $\mathbf{D}_{obs}$. It can be estimated from the prior information and the likelihood. The posterior distribution can be expressed as

$$p(\mathbf{m}|\mathbf{D}_{obs}) = \frac{p(\mathbf{D}_{obs}|\mathbf{m})p(\mathbf{m})}{p(\mathbf{D}_{obs})}, \tag{13}$$

where $p(\mathbf{D}_{obs}|\mathbf{m})$ denotes the probability of observing the data $\mathbf{D}_{obs}$ with the model $\mathbf{m}$ and is often referred to as the likelihood function. $p(\mathbf{m})$ is the prior distribution which contains the prior information of the model, and $p(\mathbf{D}_{obs}|\mathbf{m})$ is a normalization factor called evidence.

Assuming the random noise $\mathbf{N}$ is uncorrelated Gaussian noise. Its mean is zero and its variance is σ. Then the likelihood function can be formulated as (Theune et al., 2010):

$$p(\mathbf{D}_{obs}|\mathbf{m}) = exp(-\frac{||\mathbf{D}_{obs} - \mathbf{D}(\mathbf{m})||^2}{2\sigma^2}). \tag{14}$$

We assume the prior distribution $p(\mathbf{m})$ follows uniform distribution and apply an adaptive Metropolis-Hastings algorithm (Haario et al., 2001; Salvatier et al., 2016) to sample from the posterior distribution.

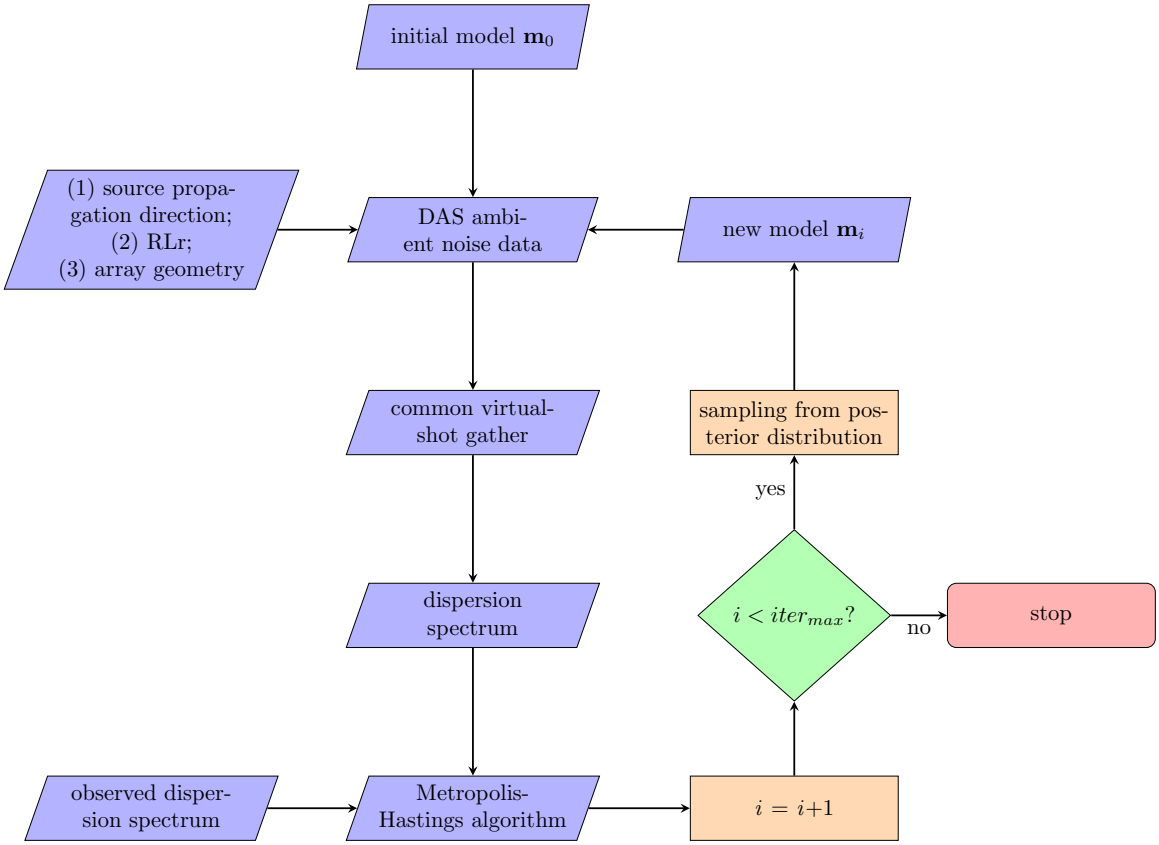

Figure 1. Inversion workflow

2.5 Inversion workflow

The inversion workflow is displayed in Figure 1. In the forward modeling process, we first simulate the DAS ambient noise data with the source propagation direction, R/L ratio, and array geometry. We then extract the virtual-shot gather and dispersion spectrum from it. The posterior distribution of the velocity model is resolved with the McMC method.

3 Synthetic data examples

We define the East as the 0° direction of propagation, and the angle increases from East to North. We assume there are 100 sources during the 2-minute recording period, and the source propagation direction is 110°. Each source generates Rayleigh and Love waves simultaneously at random times. The time sampling interval is 0.02 s, and the gauge length is 7 meters. These parameters are set based on the field data to be addressed in the next section. Table 1 shows the velocity model we use to simulate the DAS ambient noise data. We set the P wave velocity as $V_P = 2V_S$. The density has a constant value of 1800 kg/m^3.

In the forward modeling process, when the time sampling interval is fixed, the longer the period of the ambient noise data to be simulated, the more expensive the computation will cost. Thus, to explore if the extracted dispersion spectrum is influenced by the recording period, we simulate a 2-minute DAS ambient noise data and the other with a shorter period, and extract common virtual-shot gather and dispersion spectrum from

Table 1. Velocity model for DAS ambient noise simulation

Layer	**Thickness** (m)	V_S (m/s)	**Thickness range** (m)	V_S **range** (m/s)	V_P (m/s)	**Density** (kg/m^3)
1	20	350	[10, 30]	[200, 400]	700	1800
2	60	500	[30, 100]	[300, 600]	1000	1800
3	–	800	–	[600, 1000]	1600	1800

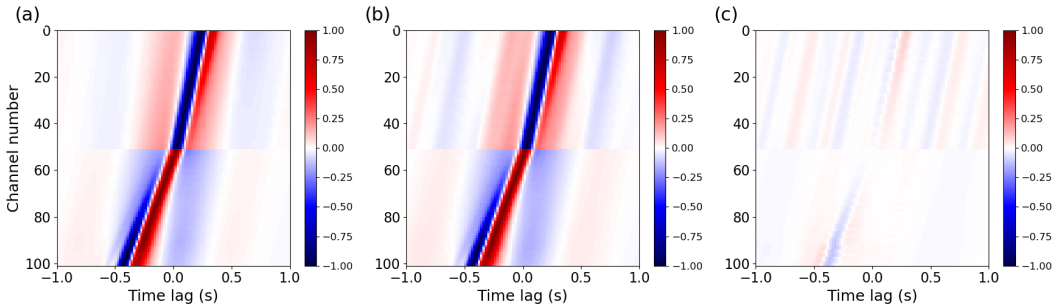

Figure 2. Common virtual-shot gathers obtained from synthetic DAS ambient noise data with different recording periods: (a) 2 minutes, and (b) 20 seconds. (c) is the misfit between (a) and (b)

.

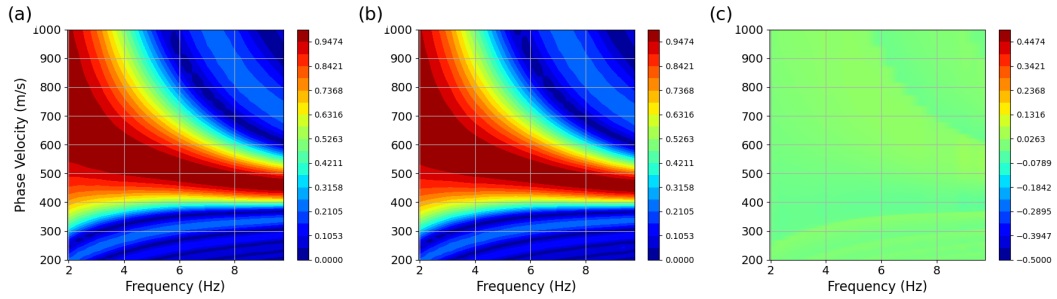

Figure 3. Dispersion spectra obtained from synthetic DAS ambient noise data with different recording periods: (a) 2 minutes, and (b) 20 seconds. (c) Misfit panel between (a) and (b)

.

both. Figures 2a and 2b show the common virtual-shot gathers extracted from the 2-minute and 20-second DAS ambient noise data with an R/L ratio of 0.5, respectively. Figure 2c shows the misfit panel between Figures 2a and 2b, which suggests that the errors between the strong events are small. Figures 3a and 3b show the dispersion spectra extracted from the common virtual-shot gathers displayed in Figures 2a and 2b, respectively. Figure 3c shows the misfit panel between Figures 3a and 3b. The misfit between the dispersion spectra is negligible. Thus, we only simulate 20-second DAS ambient noise data in the forward modeling process.

Although we reduce the period of the simulated data in the forward modeling process to reduce the computational cost, it is still computationally complex if the data is simulated by calculating the averaged strain within the gauge length. To compute the averaged strain, we first need to generate spatial samplings multiple times the number of original spatial samplings, average them within the gauge length, and put the result

at the original spatial sampling position. Computing the average is fast, however, generating such a large number of data points in space is time-consuming. On the other hand, the extracted dispersion spectrum is not influenced by this averaging procedure. Thus, in the forward modeling process, we simulate point-wise strain.

To test if simulating only 20-second point-wise strain in the forward modeling process influences the inversion results, we simulate a 2-minute DAS Rayleigh waves strain and then invert the near-surface shear-wave velocity model with the Rayleigh waves inversion from it. For the Metropolis-Hastings algorithm, we use four chains. Each includes 5000 samples, of which 1000 samples are used in the burn-in period. Figures 4a and 4b show the comparison between the true shear-wave velocity model (red line) and the posterior mean velocity model (blue line), and the marginal distributions for each parameter, respectively. The dispersion spectrum inversion method recovers the shear-wave velocity model well. Figures 4b and 4c are the dispersion spectra extracted from the original 2-minute DAS ambient noise data and the synthetic 20-second point-wise strain data with the posterior mean velocity model (blue line), respectively. Figure 4 shows the misfit between Figures 4b and 4c, which is small and indicates a good match between the original and inverted dispersion spectra. Thus, it is reliable to model 20-second point-wise strain data in the forward modeling process to reduce the computation complexity since it does not influence the extracted dispersion spectrum and inversion results.

3.1 Error induced by ignoring Love waves

To illustrate the error resulting from ignoring the existence of Love waves in the data, we simulate two 2-minute DAS ambient noise data, of which the R/L ratios are 0.5 and 0.1, respectively. And we invert the shear-wave velocity model with Rayleigh waves inversion. Figures 5a and 5b show the comparison between the true shear-wave velocity model (red line) and the posterior mean velocity model (blue line), and the marginal distributions for each parameter, respectively, when the R/L ratio is 0.5. Figures 5c and 5d are the same as Figures 5a and 5b except the results are inverted from the data with R/L ratio of 0.1. Figure 5a and 5c show that the deeper part of the shear-wave velocity models is not well recovered for both cases. As the R/L ratio decreases, it tends to be difficult to recover the shallow part of the velocity model (Figure 5c). Figure 5 demonstrates the non-uniqueness of the dispersion spectrum inversion problem, where different shear-wave velocity models fit the observed dispersion spectra equally well. Therefore, it is important to include the amplitude and polarity information in the inversion to constrain the phase-based dispersion inversion better.

3.2 Inversion with Rayleigh and Love waves simultaneously

To avoid errors induced by neglecting Love waves when inverting the shear-wave velocity model from the DAS ambient noise data, we use both Rayleigh and Love waves for the inversion. Figures 6a and 6b show the comparison between the true shear-wave velocity model (red line) and the posterior mean velocity model (blue line), and the marginal distributions for each parameter, respectively, when the R/L ratio is 0.5. Figures 6c and 6d are the same with Figures 6a and 6b except the results are inverted from the data with R/L ratio of 0.1. By considering both Rayleigh and Love waves in the inversion, the shear-wave velocity model is well recovered, although the uncertainties of the parameters get slightly larger when the R/L ratio is smaller.

4 Field data results

The field DAS ambient noise data in this section are the same as those introduced and processed in Zhao et al. (2023). We resolved the source propagation directions (111° and 112°) and R/L ratios (0.68 and 0.47) for these two datasets (Zhao et al., 2023). With

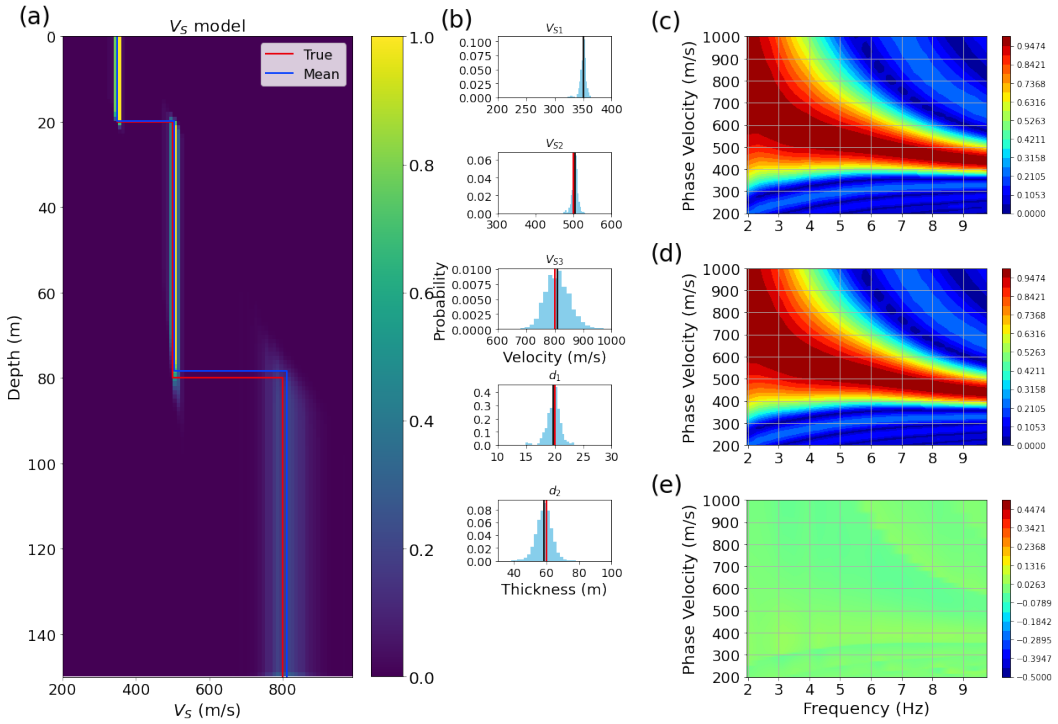

Figure 4. (a) Marginal posterior distributions of the near-surface shear-wave velocity model obtained with the McMC. The red and blue lines show the true shear-wave velocity model and the posterior mean velocity model, respectively. (b) Marginal distributions for each parameter v_1, v_2, v_3, d_1, and d_2. The red line and blue lines represent the true value and the posterior mean of each parameter, respectively. The range of the horizontal axis represents the searching range of each parameter. (c) and (d) are the dispersion spectra obtained from the 2-minute DAS ambient noise data and the posterior mean velocity (blue line displayed in (a)). (e) Misfit panel between (c) and (d). Note that the data used for the inversion is the 2-minute DAS strain which only contains Rayleigh waves. To reduce the computational cost, we only simulate 20-second point-wise strain for the forward modeling in the inversion process. The number of ambient noise sources is 100 for both the 2-minute and 20-second data.

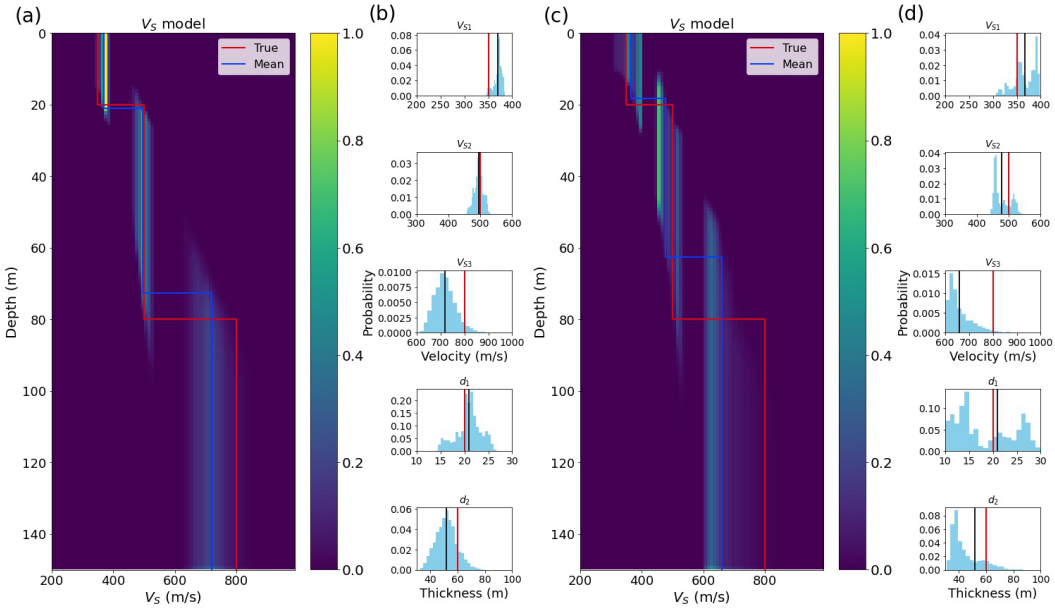

Figure 5. The same as Figures 4a and 4b except that the data used for the inversion is 2-minute DAS strain which contains both Rayleigh and Love waves. (a) and (b) are inverted from the data with an R/L ratio of 0.5, while (c) and (d) are inverted from the data with an R/L ratio of 0.1. However, only Rayleigh waves are simulated in the forward modeling process for the inversion.

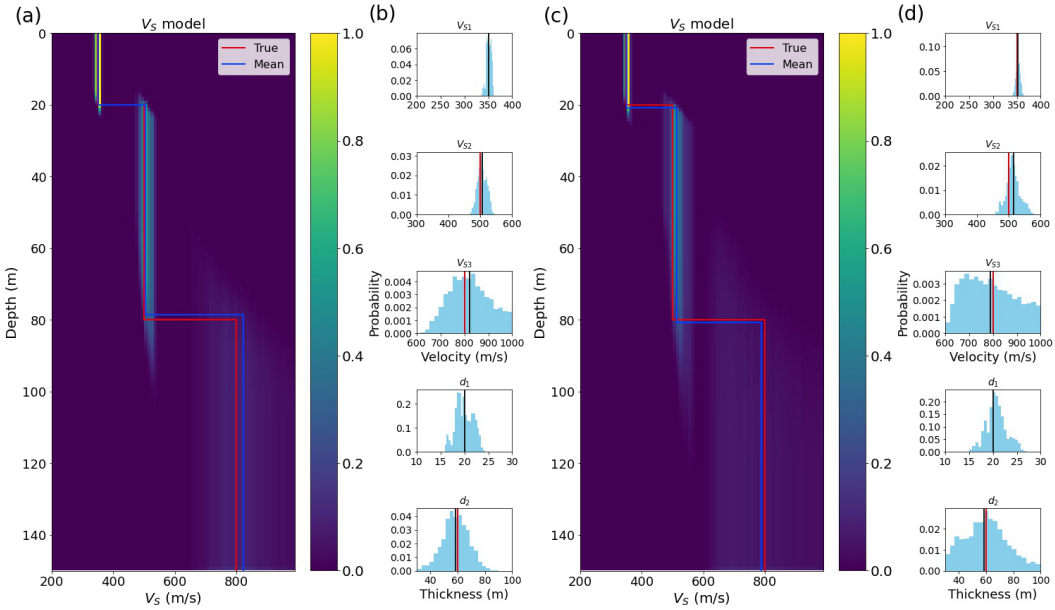

Figure 6. The same as Figures 4a and 4b except that the data used for the inversion is 2-minute DAS strain which contains both Rayleigh and Love waves. (a) and (b) are inverted from the data with the R/L ratio of 0.5, while (c) and (d) are inverted from the data with the R/L ratio of 0.1. We simulate a mix of Rayleigh and Love waves with the corresponding R/L ratios in the forward modeling process for the inversion.

this information, we invert the near-surface shear-wave velocity model with the dispersion spectrum inversion method. We use four chains, each containing 5000 samples, for the Metropolis-Hastings algorithm. There are 1000 samples used in the burn-in period for each chain. To validate our inversion results, we compare the inverted velocity model with those obtained by Thomas et al. (2014) and Spica et al. (2020). Figure 7a shows the posterior mean velocity model (blue line) inverted from the 2-minute DAS ambient noise data recorded on 3 Jan 2017 with the dispersion spectrum inversion method, best velocity model (red line) inverted from the month-long DAS ambient noise data recorded by the same DAS array (Spica et al., 2020), and the velocity model (magenta curve) inverted with an independent geotechnical study for the central campus of Stanford University ((Thomas et al., 2014)). Figure 7b shows the marginal distributions of each parameter. The uncertainty of the velocity of the half-space is big, which is reasonable due to the lack of low-frequency data (below 2.7 Hz) (Figure 7c). However, the thicknesses and velocity of the first two layers are smaller than those obtained from month-long DAS data by Spica et al. (2020). One possible reason is that the dispersion curve used for the velocity inversion in Spica et al. (2020) are apparent phase velocities, which lead to larger shear-wave velocities. Figure 8 shows the results inverted from the 2-minute DAS ambient noise data recorded on 12 April 2017. The inverted results are generally consistent with those inverted from the 2-minute DAS ambient noise data recorded on 3 Jan 2017, indicating the stability and reliability of our inversion method.

According to Thomas et al. (2014), the underground of the campus is stiff late Pleistocene alluvial deposits, of which the thickness varies from a few meters to 37 m, with the thickness increasing from the southwest end to the northeast end of the campus. Our resolved thicknesses for the first layer from the two datasets are both in the range of 10-20 m, consistent with the local geology. Besides, the velocities at the first layer of the inverted models agree with that inverted by Thomas et al. (2014). The Santa Clara formation underlies these alluvials, and its depth is about 130 m (Knudsen et al., 2000; Witter et al., 2006). For this formation, the resolved shear-wave velocity is about 500 m/s, which is also consistent with the result obtained by Thomas et al. (2014).

5 Discussions

It is commonly assumed that the ambient noise sources are uniformly distributed or mainly travel along the axial direction of the DAS array, however, these assumptions are not applicable to the short-period seismic ambient noise data recorded in the urban environment. Ambient noise sources in the urban environments are mainly generated at fixed locations (Dou et al., 2017; Zhang et al., 2019; Nilot et al., 2019; Li et al., 2020), which indicates that both Rayleigh and Love waves can be extracted from the DAS ambient noise recorded by the horizontally deployed DAS array. Based on the resolved source propagation direction and R/L ratio from the preprocessed DAS ambient noise data (Zhao et al., 2023), we invert the marginal distribution of the near-surface shear-wave velocity model with Rayleigh and Love waves simultaneously. The proposed approach could contribute to a more reliable velocity model than that obtained only with the apparent phase velocities of Rayleigh waves for the inversion.

We invert the near-surface shear-wave velocity model by minimizing the misfit between the normalized theoretical dispersion spectrum and the observed one. The forwarding modeling process consists of all the steps, from simulating the DAS ambient noise to extracting the dispersion spectrum. In the forward modeling process, we simulate a 20-second DAS ambient noise data with a temporal sampling of 0.02 s, where a dispersion curve with 1001 frequency samplings is required. Although we adopted the accelerated dispersion curve simulation open-source package called *disba* (Luu, 2021), it takes more than half a second for each iteration of the inversion. Besides, for the McMC algorithm, a large number of samples are essential to obtain a reliable marginal distribution of the inverted models. Thus, our dispersion spectrum inversion method is more time-

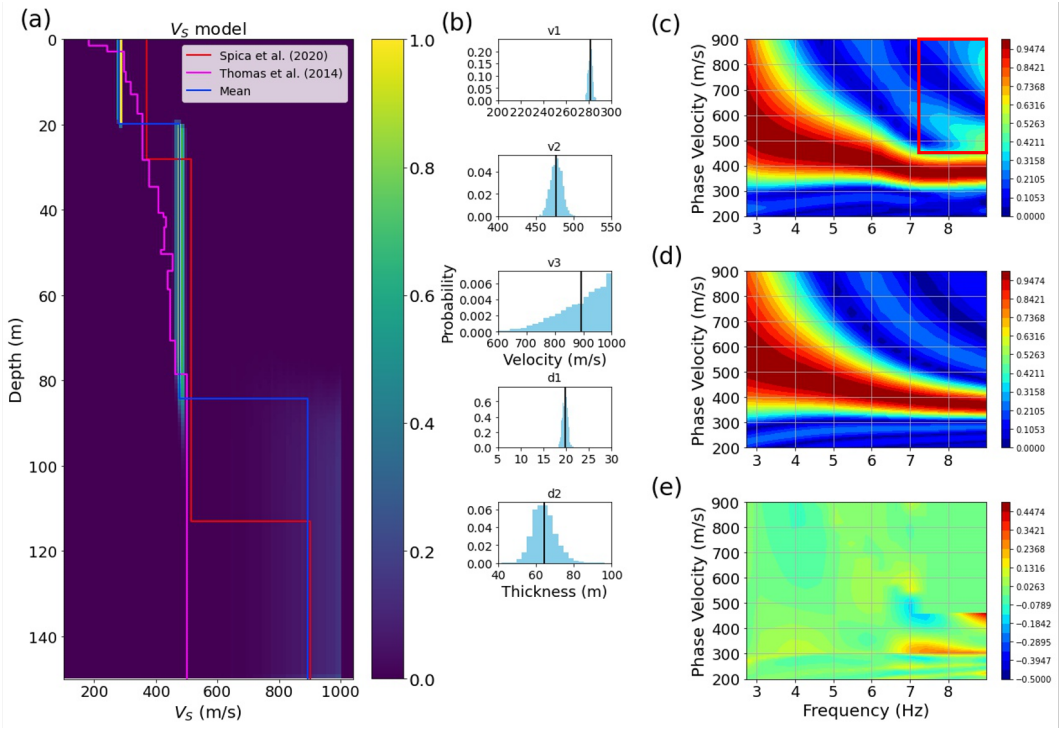

Figure 7. (a) Marginal posterior distributions of the near-surface shear-wave velocity model inverted from the 2-minute DAS ambient noise data recorded by the Stanford DAS-1 array on 3 Jan 2017. The red, magenta, and blue lines denote the shear-wave velocity model published in Spica et al. (2020), Thomas et al. (2014), and the posterior mean velocity calculated by the proposed method. (b) Marginal distributions for each parameter v_1, v_2, v_3, d_1, and d_2 are displayed on the right side of the shear-wave velocity panel. (c) Dispersion spectrum obtained from the 2-minute field DAS data. We mute the noisy part on the upper right of the panel marked by the red box. (d) Dispersion spectrum calculated from the posterior mean velocity (blue line displayed in (a)). (e) Misfit panel between (c) and (d).

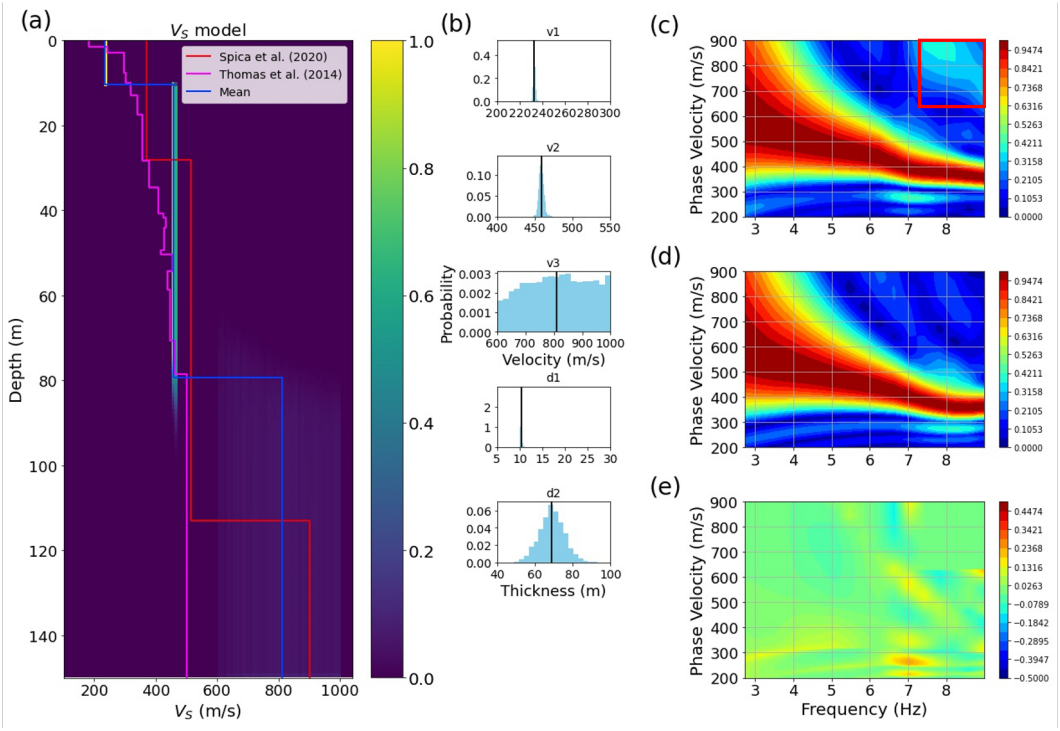

Figure 8. The same as Figure 7 except that the data was recorded on 12 April 2017.

consuming than the traditional dispersion curve inversion. In this case, with four chains and each containing 5000 samples, the whole inversion process takes about one hour.

Due to the very short period (2-minute) of our DAS data, the lowest frequency information that is extracted from our data is about 2.7 Hz, which is higher than that (2 Hz) extracted from the month-long DAS data (Martin, 2018; Spica et al., 2020). Since low-frequency signals mainly contain information about the geology structure in the deeper part, it is hard for our data to illuminate the geology structure at a depth as deep as that published in (Spica et al., 2020). It is commonly assumed the penetration depth of the surface waves is smaller than half of its smallest wavelength. Therefore, the deepest depth that our data can resolve is less than 90 m, and the inverted velocity model is unreliable at a depth larger than 90 m. Nonetheless, processing a short period of DAS data, with the rapidly developing data streaming technology and computational hardware, could significantly contribute to the real-time monitoring of near-surface hazards, such as sinkholes and landslides, where only shallow geology structures need to be investigated.

6 Conclusions

We proposed a novel method to invert the near-surface shear-wave velocity model from a short period of DAS ambient noise data recorded in an urban environment. Considering that the ambient noise sources in the urban environments are not perfectly uniformly distributed and both Rayleigh and Love waves are extracted from the DAS ambient noise data, we invert the near-surface shear-wave velocity model with both surface waves simultaneously with the R/L ratio and source propagation direction estimated from the data. The proposed method could contribute to a more reliable near-surface geological structure than that obtained from the inversion with a common assumption that Rayleigh waves are predominant. Moreover, as reliable surface waves information at higher fre-

quencies is extracted from short-period DAS ambient noise data, the proposed method has the potential to be used for near-surface real-time monitoring.

7 Open Research

Seismic data and the associate codes used in this study are available and can be downloaded from this link: `https://doi.org/10.5281/zenodo.7513878`.

Acknowledgments

We thank Professor Biondo Biondi for sharing the DAS data. We acknowledge Cambridge Sensing Pte Ltd for its financial support. We thank the open-source software Madagascar (`http://www.ahay.org/`), disba (`https://pypi.org/project/disba/0.1.2/`), and PyMC3 (`https://docs.pymc.io/en/v3/`), with which we preprocess the data, simulate the dispersion curve, and invert the near-surface shear-wave velocity model.

References

Aki, K. (1957). Space and time spectra of stationary stochastic waves, with special reference to microtremors. *Bulletin of the Earthquake Research Institute*, *35*, 415–456.

Arai, H., & Tokimatsu, K. (2004). S-wave velocity profiling by inversion of microtremor H/V spectrum. *Bulletin of the Seismological Society of America*, *94*(1), 53–63.

Bayes, T. (1763). LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFR S. *Philosophical transactions of the Royal Society of London*(53), 370–418.

Behm, M., Leahy, G. M., & Snieder, R. (2014). Retrieval of local surface wave velocities from traffic noise– an example from the La Barge basin (Wyoming). *Geophysical Prospecting*, *62*(2), 223–243.

Behm, M., & Snieder, R. (2013). Love waves from local traffic noise interferometry. *The Leading Edge*, *32*(6), 628–632.

Boaga, J., Cassiani, G., Strobbia, C. L., & Vignoli, G. (2013). Mode misidentification in Rayleigh waves: Ellipticity as a cause and a cure. *Geophysics*, *78*(4), EN17–EN28.

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

Brooks, L. A., Townend, J., Gerstoft, P., Bannister, S., & Carter, L. (2009). Fundamental and higher-mode Rayleigh wave characteristics of ambient seismic noise in New Zealand. *Geophysical Research Letters*, *36*(23).

Chmiel, M., Mordret, A., Boué, P., Brenguier, F., Lecocq, T., Courbis, R., ... Van der Veen, W. (2019). Ambient noise multimode Rayleigh and Love wave tomography to determine the shear velocity structure above the Groningen gas field. *Geophysical Journal International*, *218*(3), 1781–1795.

Claerbout, J. F. (1968). Synthesis of a layered medium from its acoustic transmission response. *Geophysics*, *33*(2), 264–269.

Díaz, J., Ruiz, M., Sánchez-Pastor, P. S., & Romero, P. (2017). Urban seismology: On the origin of earth vibrations within a city. *Scientific reports*, *7*(1), 1–11.

Dou, S., & Ajo-Franklin, J. B. (2014). Full-wavefield inversion of surface waves for mapping embedded low-velocity zones in permafrost. *Geophysics*, *79*(6), EN107–EN124.

Dou, S., Lindsey, N., Wagner, A. M., Daley, T. M., Freifeld, B., Robertson, M., ...

Ajo-Franklin, J. B. (2017). Distributed Acoustic Sensing for Seismic Monitoring of The Near Surface: A Traffic-Noise Interferometry Case Study. *Scientific Reports*, *7*(1), 1–12.

Ekström, G. (2001). Time domain analysis of Earth's long-period background seismic radiation. *Journal of Geophysical Research: Solid Earth*, *106*(B11), 26483–26493.

Fäh, D., Kind, F., & Giardini, D. (2003). Inversion of local S-wave velocity structures from average H/V ratios, and their use for the estimation of site-effects. *Journal of Seismology*, *7*(4), 449–467.

Friedrich, A., Krüger, F., & Klinge, K. (1998). Ocean-generated microseismic noise located with the Gräfenberg array. *Journal of Seismology*, *2*(1), 47–64.

Haario, H., Saksman, E., & Tamminen, J. (2001). An adaptive Metropolis algorithm. *Bernoulli*, 223–242.

Hamimu, L., Safani, J., & Nawawi, M. (2011). Improving the accurate assessment of a shear-wave velocity reversal profile using joint inversion of the effective Rayleigh wave and multimode Love wave dispersion curves. *Near Surface Geophysics*, *9*(1), 1–14.

Ikeda, T., Matsuoka, T., Tsuji, T., & Nakayama, T. (2015). Characteristics of the horizontal component of Rayleigh waves in multimode analysis of surface waves. *Geophysics*, *80*(1), EN1–EN11.

Jay, J. A., Pritchard, M. E., West, M. E., Christensen, D., Haney, M., Minaya, E., ... Zabala, M. (2012). Shallow seismicity, triggered seismicity, and ambient noise tomography at the long-dormant Uturuncu Volcano, Bolivia. *Bulletin of Volcanology*, *74*(4), 817–837.

Joh, S.-H., Stokoe, K. H., II, Lee, I.-W., Kang, T.-H., Rosenbld, B., & Bay, J. A. (2006). Joint Inversion for Apparent Phase Velocities of Rayleigh and Love waves. In *Geocongress 2006: Geotechnical engineering in the information technology age* (pp. 1–6).

Knudsen, K. L., Sowers, J. M., Witter, R. C., Wentworth, C. M., & Helley, E. J. (2000). Description of mapping of quaternary deposits and liquefaction susceptibility, nine-county San Francisco Bay Region, California. *Rep. No. United States Geologic Survey Open-File Report 00*, *444*.

Li, Y. E., Nilot, E., & Feng, X. (2020). Observation of guided and reflection P-waves in urban ambient noise cross-correlograms. In *Seg technical program expanded abstracts 2020* (pp. 2100–2104). Society of Exploration Geophysicists.

Lin, F.-C., Moschetti, M. P., & Ritzwoller, M. H. (2008). Surface wave tomography of the western United States from ambient seismic noise: Rayleigh and Love wave phase velocity maps. *Geophysical Journal International*, *173*(1), 281–298.

Lunedei, E., & Albarello, D. (2015). Horizontal-to-vertical spectral ratios from a full-wavefield model of ambient vibrations generated by a distribution of spatially correlated surface sources. *Geophysical Journal International*, *201*(2), 1142–1155.

Luu, K. (2021). *disba: Numba-accelerated computation of surface wave dispersion.* Retrieved from `https://github.com/keurfonluu/disba` doi: 10.5281/zenodo.3987395

Martin, E. R. (2018). *Passive imaging and characterization of the subsurface with distributed acoustic sensing*. Stanford University.

Nakamura, Y. (1989). A method for dynamic characteristics estimation of subsurface using microtremor on the ground surface. *Railway Technical Research Institute, Quarterly Reports*, *30*(1).

Nakata, N. (2016). Near-surface S-wave velocities estimated from traffic-induced Love waves using seismic interferometry with double beamforming. *Interpretation*, *4*(4), SQ23–SQ31.

Nakata, N., Snieder, R., Tsuji, T., Larner, K., & Matsuoka, T. (2011). Shear wave

imaging from traffic noise using seismic interferometry by cross-coherence. *Geophysics*, *76*(6), SA97–SA106.

Nilot, E., Zhang, Y., Li, Y. E., & Feng, X. (2019). Deep bedrock detection based on ambient noise recorded by a short geophone array: A Singapore case study. In *Seg technical program expanded abstracts 2019* (pp. 3121–3125). Society of Exploration Geophysicists.

Park, C. B., Miller, R. D., & Xia, J. (1998). Imaging dispersion curves of surface waves on multi-channel record. In *Seg technical program expanded abstracts 1998* (pp. 1377–1380). Society of Exploration Geophysicists.

Picozzi, M., Parolai, S., & Richwalski, S. (2005). Joint inversion of H/V ratios and dispersion curves from seismic noise: Estimating the S-wave velocity of bedrock. *Geophysical Research Letters*, *32*(11).

Ryden, N., & Park, C. B. (2006). Fast simulated annealing inversion of surface waves on pavement using phase-velocity spectra. *Geophysics*, *71*(4), R49–R58.

Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. (2016). Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*, *2*, e55.

Sivaram, K., Gupta, S., Kumar, S., & Prasad, B. (2018). Shear velocity structural characterization around the Lonar crater using joint inversion of ambient noise HVSR and Rayleigh wave dispersion. *Journal of Applied Geophysics*, *159*, 773–784.

Song, Z., Zeng, X., & Thurber, C. H. (2021). Surface-wave dispersion spectrum inversion method applied to Love and Rayleigh waves recorded by distributed acoustic sensing. *Geophysics*, *86*(1), EN1–EN12.

Spica, Z. J., Perton, M., Martin, E. R., Beroza, G. C., & Biondi, B. (2020). Urban Seismic Site Characterization by Fiber-Optic Seismology. *Journal of Geophysical Research: Solid Earth*, *125*(3), e2019JB018656.

Theune, U., Jensås, I. Ø., & Eidsvik, J. (2010). Analysis of prior models for a blocky inversion of seismic AVA data. *Geophysics*, *75*(3), C25–C35.

Thomas, P., Wong, I., Zachariasen, J., Darragh, R., & Silva, W. (2014). *2013 update to the sites-pecific seismic hazard analyses and development of seismic design ground motions* (Tech. Rep.). California: Stanford University.

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

Witter, R. C., Knudsen, K. L., Sowers, J. M., Wentworth, C. M., Koehler, R. D., Randolph, C. E., ... Gans, K. D. (2006). *Maps of Quaternary deposits and liquefaction susceptibility in the central San Francisco Bay region, California* (Tech. Rep.). US Geological Survey.

Yin, X., Xia, J., Shen, C., & Xu, H. (2014). Comparative analysis on penetrating depth of high-frequency Rayleigh and Love waves. *Journal of Applied Geophysics*, *111*, 86–94.

Yin, X., Xu, H., Mi, B., Hao, X., Wang, P., & Zhang, K. (2020). Joint inversion of Rayleigh and Love wave dispersion curves for improving the accuracy of near-surface S-wave velocities. *Journal of Applied Geophysics*, *176*, 103939.

Zeng, X., Lancelle, C., Thurber, C., Fratta, D., Wang, H., Lord, N., ... Clarke, A. (2017). Properties of Noise Cross-Correlation Functions Obtained from a Distributed Acoustic Sensing Array at Garner Valley, California. *Bulletin of the Seismological Society of America*, *107*(2), 603–610.

Zhang, Y., Li, Y. E., Zhang, H., & Ku, T. (2019). Near-surface site investigation by seismic interferometry using urban traffic noise in Singapore. *Geophysics*, *84*(2), B169–B180.

Zhao, Y., Li, Y. E., & Li, B. (2023). On beamforming of DAS ambient noise recorded in an urban environment and Rayleigh-to-Love waves ratio estimation. *submitted to Journal of Geophysical Research*, *x*, x-xx.