

Near-real-time near-surface 3D seismic velocity and uncertainty models by wavefield gradiometry and neural network inversion of ambient seismic noise

Ruikun Cao¹, Stephanie Earp¹, Sjoerd A. L. de Ridder², Andrew Curtis¹, and Erica Galetti¹

ABSTRACT

With the advent of large and dense seismic arrays, novel, cheap, and fast imaging and inversion methods are needed to exploit the information captured by stations in close proximity to each other and produce results in near real time. We have developed a sequence of fast seismic acquisition for dispersion curve extraction and inversion for 3D seismic models, based on wavefield gradiometry, wave equation inversion, and machine-learning technology. The seismic array method that we use is Helmholtz wave equation inversion using measured wavefield gradients, and the dispersion curve inversions are based on a mixture of density neural networks (NNs). For our approach, we assume that a single surface wave mode dominates the data. We derive a nonlinear relationship among the unknown true seismic wave velocities, the measured seismic wave velocities, the interstation spacing,

and the noise level in the signal. First with synthetic and then with the field data, we find that this relationship can be solved for unknown true seismic wave velocities using fixed point iterations. To estimate the noise level in the data, we need to assume that the effect of noise varies weakly with the frequency and we need to be able to calibrate the retrieved average dispersion curves with an alternate method (e.g., frequency wavenumber analysis). The method is otherwise self-contained and produces phase velocity estimates with tens of minutes of noise recordings. We use NNs, specifically a mixture density network, to approximate the non-linear mapping between dispersion curves and their underlying 1D velocity profiles. The networks turn the retrieved dispersion model into a 3D seismic velocity model in a matter of seconds. This opens the prospect of near-real-time near-surface seismic velocity estimation using dense (and potentially rolling) arrays and only ambient seismic energy.

INTRODUCTION

With the recent advent of so-called large N seismic arrays, where N refers to the number of seismometers (IRIS, 2018), a new class of seismic imaging and inversion methods has been developed to estimate subsurface properties from closely spaced observations of the full seismic wavefield. Array recordings provide high spatial and temporal resolution recordings of the wavefield, potentially allowing the recovery of high-resolution models of the subsurface. However, conventional imaging and inversion methods do not fully exploit the spatial proximity of individual recordings. For instance, closely spaced observations of the wavefield allow estimates of the spatial

gradients of the wavefield to be calculated. We show here that these can be used to constrain the near-surface seismic velocities in the vicinity of the stations in novel ways. The field data presented in this study are recorded by conventional 10 Hz corner frequency geophones planted in a field site southeast of Edinburgh.

It has long been known that relating observations of seismic wavefields made by proximate stations can reveal subsurface properties. One of the earliest approaches was to extract surface wave dispersion curves by spatial autocorrelation (Aki, 1957). Other early approaches included the inference of the seismic wave velocity between two stations from differential traveltime observations (Gerver and Markushevich, 1966, 1972) and the characterization of the local

Manuscript received by the Editor 31 July 2018; revised manuscript received 22 June 2019; published ahead of production 14 September 2019; published online 22 November 2019.

¹University of Edinburgh, School of GeoSciences, Grant Institute of Earth Science, Edinburgh EH9 3FE, UK. E-mail: ruikuncao@foxmail.com; stephanie.earp@ed.ac.uk; andrew.curtis@ed.ac.uk; erica.galetti@ed.ac.uk.

²Formerly University of Edinburgh, School of GeoSciences, Grant Institute of Earth Science, Edinburgh EH9 3FE, UK; presently University of Leeds, School of Earth and Environment, Leeds LS2 9JT, UK. E-mail: s.deridder@leeds.ac.uk (corresponding author).

© 2020 Society of Exploration Geophysicists. All rights reserved.

wave slownesses using a linear Radon transformation (Kennett, 1981). The reader is referred to Rost and Thomas (2002) for a good review on common array-based techniques.

A general strategy to estimate local elastodynamic properties from gradients of volumetric observations (observations spanning a small volume) of an elastodynamic wavefield is based on inverting the wave equation locally (i.e., wave equation inversion) (Curtis and Robertsson, 2002; Muijs et al., 2003). For body waves that have no particular known structure (any number may arrive from any angle with any amplitude), the general method requires stations to be located not only on the surface, but also to be buried in the subsurface. For surface waves, this requirement can be circumvented (de Ridder and Biondi, 2015) because surface waves propagate predominantly along the surface where observations are available. If wavefields can be assumed to be composed of nonoverlapping planar waves, then Langston (2007a, 2007b, 2007c) showed that the wavefield can be described by a plane-wave equation, which further simplifies the wave equation inversion problem. More generally, gradiometry refers to the measurement of gradients, irrespective of how these measurements are used in subsequent analysis. By plane-wave inversion, we can determine wave attributes such as horizontal slowness, changes in geometric spreading, changes in the radiation pattern, and the wave propagation azimuth. This plane-wave inversion approach has been refined with time-frequency analysis using wavelets (Poppeliers, 2010, 2011) and analysis in 3D to include the information from wave polarization (Poppeliers et al., 2013). This approach has been applied on a continental scale, along boreholes, and for terrestrial and lunar near-surface site studies (Liang and Langston, 2009; Liu and Holt, 2015; Edme and Yuan, 2016; Langston and Ayele, 2016; Maeda et al., 2016; Sollberger et al., 2016).

The propagation of surface waves is commonly approximated as being governed by a 2D Helmholtz equation with a frequency-dependent phase velocity map (Aki and Richards, 2002). Wielandt (1993) discusses the implications of lateral velocity variations on the interpretation of surface waves through *apparent* phase velocity maps. A strategy to extract wave velocities based on the Helmholtz equation description for surface wave propagation is to use the eikonal equation obtained using a ray theoretical approximation (Lin et al., 2009; Lin and Ritzwoller, 2011; de Ridder et al., 2015; Liu and Holt, 2015). This became a popular approach that found applications on scales ranging from continents to reservoir overburdens (Lin et al., 2009, 2013; de Ridder and Dellinger, 2011; Gouéard et al., 2012; Mordret et al., 2013a, 2013b).

De Ridder and Biondi (2015) and de Ridder and Curtis (2017) use measured wavefields of ambient seismic noise to directly invert the Helmholtz equation for apparent local phase velocity maps. This circumvented the need for cross-correlating recordings of a relatively long time duration to estimate surface-wave traveltimes. The ambient seismic wavefield can be composed of plane waves incident from different directions and overlapping (de Ridder and Biondi, 2015). A comparative analysis between eikonal tomography and Helmholtz wave equation inversion showed that using Helmholtz wave equation inversion can yield superior results when used together with compressive sensing techniques (Zhan et al., 2018).

The methods that we use to estimate near-surface earth properties from ambient seismic data are based on Helmholtz wave equation inversion using measured wavefield gradients. We make no attempt to separate the energy of different wave modes when analyzing the wavefield, which we assume to be dominated by fundamental mode

surface waves. By neglecting the wavefield energy of higher order wave modes and body waves, we assume that all of the recorded energy can be described by a Helmholtz equation with a single apparent phase velocity map that we aim to estimate. Furthermore, we develop specialized pre- and postprocessing steps that were required to handle the effect of noise and approximation errors when using finite-difference methods to compute the required wavefield derivatives.

Neural networks (NNs) approximate a nonlinear mapping between two parameter spaces. By presenting the network with a set of data-model pairs, it can be trained to create a mapping from the data to the model parameter space. This is particularly useful in geophysical inverse problems in which the forward mapping, from the model to the data parameter space, is well known or simple to calculate (to construct training data) but the inverse mapping is complex or difficult to determine directly. Once a network has been trained, it can be given previously unseen data and will output a new model estimate in seconds. Using a specific class of NN, called a mixture density network (MDN), we can also output uncertainties in the network (Bishop, 1995). Meier et al. (2007a, 2007b) use MDNs to invert regional surface-wave dispersion curves to give fully probabilistic estimates of global crustal thickness models. MDNs have also been used to perform petrophysical inversion of seismic data sets for subsurface porosity and clay content.

We start by setting out the theory for seismic wavefield gradiometry and Helmholtz wave equation inversion, and we establish a correction procedure to handle the errors resulting from noise and finite-difference approximations. We first test this theory and correction procedure with a synthetic data example, and then we present the results of a field experiment over a buried landfill to investigate the efficacy of this acquisition and processing method.

METHODS

Helmholtz wave equation inversion for surface wave phase velocities

We estimate phase velocity maps from wavefields dominated by surface waves by measuring the wavefield's temporal and spatial derivatives and inverting a wave equation. It is common practice to approximate the far-field behavior of surface waves as a superposition of wave modes traveling with frequency-dependent velocities. Each mode is composed of a superposition of plane waves that obey a scalar wave equation with a mode-dependent and frequency-dependent phase velocity, $\hat{c}(x, y, \omega, m)$, where (x, y) are the horizontal coordinates on the earth's surface, ω is the frequency, and m is the mode number. Consequently, when a recording is dominated by far-field surface waves of a single wave mode (potentially interfering simultaneous arrivals from different azimuths), the wavefield observed in the vertical component of particle velocity (or displacement), here denoted with U , obeys the 2D scalar Helmholtz equation with frequency-dependent velocity:

$$\nabla^2 \hat{U}(x, y, \omega) + \hat{s}^2(x, y, \omega) \omega^2 \hat{U}(x, y, \omega) = -\hat{f}(x, y, \omega), \quad (1)$$

where U is the scalar wavefield changing with time and space and \hat{U} is its Fourier transform dual in the frequency domain, f is a generalized source term, \hat{f} is its Fourier transformed dual, and $\hat{s} = 1/\hat{c}$ is the slowness where we have dropped the wave mode dependency

for notation simplicity. The term ∇^2 is the Laplace operator, so for a spatial 2D wavefield (in the subsequent field data examples), $\nabla^2 = \partial_x^2 + \partial_y^2$, whereas for a spatial 1D wavefield (in the subsequent synthetic data example), $\nabla^2 = \partial_x^2$. Although we could have followed the method of Allmark (2018) and estimated local sources of energy, here, we neglect local sources acting within the domain of recording (the array carpet) or by choosing a subdomain that omits strong localized sources, so we assume $\hat{f}(x, y, \omega) = 0$. We follow de Ridder and Biondi (2015) to find an appropriate time-domain scalar wave equation: We band-pass the recordings using a narrow frequency-domain Hann window and central frequency ω' to obtain $\hat{U}_{w'}(x, y)$. We ignore the frequency dependency of the phase velocity within the passed frequency bandwidth, so $\hat{c}(\omega) = c_{w'}$. With an inverse Fourier transformation, we then obtain a time-domain wave equation:

$$\nabla^2 U_{w'}(x, y, t) = \partial_t^2 U_{w'}(x, y, t) s_{w'}^2(x, y), \quad (2)$$

where ∂_t is the temporal derivative operator.

Equation 2 states that the spatial and temporal derivatives of the wavefield, after a narrow band-pass filter for a particular central frequency, are related through the local phase velocity. Hypothetically, if we had direct observations of the second-order spatial and temporal gradients of the continuous wavefield in space and time, we could solve this relationship for the slowness:

$$s(x, y) = \sqrt{\frac{\int_{t_0}^{t_n} dt \{ \partial_t^2 U_{w'}(x, y, t_i) \} \{ \nabla^2 U_{w'}(x, y, t_i) \}}{\int_{t_0}^{t_n} dt \{ \partial_t^2 U_{w'}(x, y, t_i) \} \{ \partial_t^2 U_{w'}(x, y, t_i) \}}}, \quad (3)$$

where the (x, y) coordinates correspond to the coordinates of a station in the array and the recording time started at $t = t_0$ and ended at $t = t_n$. Mapping this relationship as a function of space yields the phase velocity map for that particular central frequency. We can interpret the properties of the subsurface of the survey area using phase velocity maps at one or more central frequencies.

Seismic wavefield gradiometry

Equation 3 requires wavefield derivatives of U to be available. Although accelerometers could be used to observe the second-order temporal derivative of the wavefield, we are not usually able to observe the spatial gradients of the wavefield directly. The wavefield's particle velocity is usually observed discretely in space and time. We therefore estimate the spatial derivatives of the wavefield using finite-difference approximations to the derivative operators. This latter process is now often referred to as seismic wavefield gradiometry.

We assume that seismic stations are located on a regular grid with spacing of Δx and Δy , respectively, in the x - and y -directions, and measurements are made with a constant sampling rate with spacing Δt . Each recorded data point is labeled as $U(x_i, y_j, t_k)$. Finite-difference coefficients are found from Taylor series expansions of the function — see LeVeque (2007) for background on the finite-difference method. Higher order terms in the Taylor series are neglected to curtail the number of coefficients required to estimate the continuous derivative from discrete observations. In this work, we use finite-difference approximations with second-order accuracy. We estimate the second-order temporal derivatives of the wavefield using

$$\begin{aligned} \{ \partial_t^2 U \}_M(x_i, y_j, t_k) \\ = \frac{U(x_i, y_j, t_{k-1}) - 2U(x_i, y_j, t_k) + U(x_i, y_j, t_{k+1}))}{\Delta t^2}. \end{aligned} \quad (4)$$

That is, for each time sample at each station, the second-order temporal derivatives are estimated using two adjacent time-samples. We estimate the second-order spatial derivatives of the wavefield using

$$\begin{aligned} \{ \nabla^2 U \}_M(x_i, y_j, t_k) \\ = \frac{U(x_{i-1}, y_j, t_k) - 2U(x_i, y_j, t_k) + U(x_{i+1}, y_j, t_k)}{\Delta x^2} + \end{aligned} \quad (5)$$

$$\frac{U(x_i, y_{j-1}, t_k) - 2U(x_i, y_j, t_k) + U(x_i, y_{j+1}, t_k)}{\Delta y^2}. \quad (6)$$

That is, for each time sample, the second-order spatial derivatives at one station are measured using the corresponding time-sample at the four adjacent stations in a cross-shaped template. Using adjacent samples means that near the boundary of the array (where no adjacent stations are available), we cannot measure the spatial derivatives.

The measured wavefield derivatives can be used to invert the wave equation directly and obtain a measured estimate for the local wave slowness (similar to equation 3):

$$s_M(x_i, y_j) = \sqrt{\frac{\sum_{k=1}^{N_t} dt \{ \partial_t^2 U \}_M(x_i, y_j, t_k) \{ \nabla^2 U \}_M(x_i, y_j, t_k)}{\sum_{k=1}^{N_t} dt \{ \partial_t^2 U \}_M(x_i, y_j, t_k) \{ \partial_t^2 U \}_M(x_i, y_j, t_k)}}, \quad (7)$$

where N_t is the number of available measurements of second-order spatial and temporal derivatives for each station at (x_i, y_j) using all samples in time. In the least-squares regression formulation in equations 3 and 7, we choose to minimize the least-squares error in the spatial wavefield gradient because errors in the measured spatial wavefield gradients are usually larger than those in the measured temporal wavefield gradients (Allmark, 2018).

Correcting finite-difference approximation errors in wavefield gradiometry

Any finite-difference approximation is only accurate to a certain order in the interstation distance and depends on the wavelength of the underlying function. The approximation error becomes larger with increased distance and shorter wavelengths, but the effect of noise may dominate at small sampling distances (see the “Field data experiment” section). In Appendices A and B, we analyze the approximation error in finite differences and the effect of noise in our observations and we derive the following relationship between the slowness s_M , measured using wavefield gradiometry or wave equation inversion and the true slowness s_T (equation B-15):

$$s_T = \gamma(s_T) \sqrt{1 - \epsilon} s_M, \quad (8)$$

where $\epsilon < 1$ is a parameter that depends on the signal-to-noise level of the measured spatial gradients of the wavefield recordings and

$$\gamma(s_T) = \frac{\sqrt{|1 - \cos\{2\pi f \Delta t\}|} \Delta x}{\sqrt{|1 - \cos\{s_T 2\pi f \Delta x\}|} \Delta t} s_T. \quad (9)$$

This equation has two unknowns (ϵ and s_T). We use fixed point iterations to invert (solve) the nonlinear relationship between s_T and s_M given a particular ϵ . We need an observation or estimate of the average background dispersion to calibrate the noise level ϵ in this relationship, illustrated in the “Field data experiment” section. We will solve the nonlinear relationship in equation 8 with fixed point iterations (Deuhlhard, 2012):

$$s_j = \gamma(s_{j-1}) \sqrt{1 - \epsilon s_M} \quad j = 1, 2, \dots, n, \quad (10)$$

starting with $s_0 = s_M$ and finishing with $s_T \approx s_n$. Finally, an estimate for the true velocity is recovered as $c_T = s_T^{-1}$. We expect that $\gamma(s)$ has multiple fixed points and that there is no guarantee of convergence to the fixed point corresponding to the correct s_T . A complete analysis of the criteria on the starting slowness, for which the fixed-point iterations converge to the desired true slowness, is beyond the scope of this paper. However, given a slowness estimate reasonably close to the true slowness (perhaps a better estimate than $s_0 = s_M$), we can expect convergence to the correct s_T . We found that $n = 20$ iterations are sufficient to converge (see Appendix B). The unknown noise level is more challenging to estimate. One way is to ensure that the final corrected dispersion curves place energy in the same region of a dispersion image of all of the data (using, for example, a frequency-wavenumber spectrum or knowledge of a few points along an average dispersion curve). Despite requiring some knowledge about the average dispersion curve, we can still map lateral phase-velocity changes with high resolution (and assuming that ϵ does not change significantly across the array). This will be illustrated in the synthetic and field data examples that follow.

Mixture density networks

NNs are nonlinear mathematical models that can be used to express a relationship between data \mathbf{d} and a model \mathbf{m} . By presenting the network with a set of data-model (\mathbf{d}, \mathbf{m}) pairs, it can be trained to create a mapping from the data to the model parameter space by minimizing a cost function (usually a sum of squared errors) that measures the difference between the network model-space output for each data vector \mathbf{d} and the true model \mathbf{m} in each pair. An MDN is a class of NNs developed by Bishop (1995) to represent arbitrary probability distributions in the same way that an NN represents functions. They are trained in the same way as a standard

NN but instead of outputting a model estimate, they provide an estimate of parameters that describe a desired probability distribution — in our case, the Bayesian posterior distribution over \mathbf{m} given some recorded data \mathbf{d} as input. In our networks, we parameterize the output by a sum (mixture) of Gaussians, which has been shown to be adequate in other surface wave dispersion inversions (Meier et al., 2007a, 2007b). For a detailed description of MDNs and their application in geophysical contexts, we refer the reader to Meier et al. (2007b) or Shahraneini and Curtis (2011).

We forward modeled 100,000 randomly generated S-wave depth-velocity models for dispersion curves with frequencies 18, 20, 22, and 24 Hz using the DISPER80 subroutines by Saito (1988). The depth-velocity models each have 11 layers that increase in thickness with depth (summarized in Table 1) because we expect our phase velocity data to provide less resolution the deeper the layer (Shapiro and Ritzwoller, 2002). For each layer in the model, the S-wave velocities V_S were uniformly randomly selected from a velocity range; for the top layer, this was 50–300 m/s, and for subsequent layers, we used 300–2000 m/s. The pressure-wave velocity V_P and density ρ were set as $V_P = 1.16V_S + 1.36$, and $\rho = 1.74\sqrt[4]{V_P}$. Once the dispersion curves were generated for each model, the original models were averaged vertically to provide mean velocities in five depth layers with thicknesses of 3, 7, 10, 10, and 20 m (Table 1). We will invert the data for velocities in these upscaled models to increase the degree of constraint on each layer offered by the measured dispersion. The upscaled models \mathbf{m} and their associated dispersion curves \mathbf{d} were used as the data-model pairs to train the MDNs. Separate MDNs were trained to output the distribution of the S-wave velocity in each layer.

Training was implemented using Google’s TensorFlow open-source software library. Multiple MDNs were trained for each layer, and the network with the lowest cost value was selected as the final network for that layer. The outputs of each network are not true Bayesian posterior distributions because we do not know the absolute uncertainties on the phase velocity, but they do represent some measure of relative uncertainty on the S-wave velocity at each depth. They have the huge advantage over other nonlinear inversion methods (such as Monte Carlo; Shapiro and Ritzwoller, 2002; Galletti et al., 2015; Zhang et al., 2018) of computational speed: Once the networks are trained, they can be applied to new data sets in seconds (Meier et al., 2007b).

TWO-LAYER SYNTHETIC DATA EXPERIMENT

We first demonstrate the method to correct finite-difference approximation errors in wavefield gradiometry, followed by wave equation inversion using synthetic data. The model has two spatial dimensions (x, z), and the data were generated with a full elastic simulation. The model is a layer over a half-space with the interface at 50 m in depth. The layer has $V_P = 1500$ m/s and $V_S = 750$ m/s, whereas the half-space has $V_P = 2600$ m/s and $V_S = 1300$ m/s. Both layers have density of 1000 kg/m^3 . The velocities in the half-space are higher than those in the top layer, as is common in near-surface field scenarios. The simulated source was placed halfway along the spatial axis and just under the free surface inside the top layer. The source wavelet has frequency support over the range 1–15 Hz. The wavefield was simulated using a very dense spatial grid, so that numerical dispersion at frequencies below 15 Hz can be neglected in the following analysis. We record the vertical component of particle

Table 1. Summary of the parameterization of S-wave velocity-depth models used for training the MDN.

Velocity range in layer (m/s)	50-300	300-2000									
Layer thickness (m)	1	2	3	4	5	5	10	10	10	10	10
Up-scaled layer thickness (m)	3	7	10	10	20	-					

velocity at the surface, $u(x, t)$, which is a scalar field that has one spatial dimension and one temporal dimension (a sampling rate of $\Delta t = 0.1$ ms).

Figure 1 shows the simulated wavefield. Besides strong fundamental mode surface waves, we recognize at least two overtones. Refracted and reflected waves cannot be recognized, but they may be hidden beneath the surface waves. The wave motion of surface waves attenuates with depth, and lower frequencies penetrate deeper than higher frequencies. Because the velocity of the medium increases with depth, we often find that surface waves travel with faster velocities at lower frequencies than do higher frequencies. This phenomenon is known as dispersion. From here on, we neglect all body wave energy and higher mode surface-wave energy in the simulated wavefield and assume that the dispersive Helmholtz equation provides a good description of the dominant fundamental mode surface waves.

The frequency wavenumber spectrum of the shot record is shown in Figure 1b. The velocity with which the various wave-modes are observed by the array is given by the slope of the anomaly in the f - k space with respect to the origin. Their apparent velocity is much easier to observe in a dispersion image, which shows the energy as a function of phase velocity and frequency. To obtain this, we transform the f - k spectrum to an f - s spectrum given the relationship $s = k/f$. In the gray-shaded backgrounds of Figure 2, we thus identify the relationship between the frequency (f) and phase velocity (c) for the fundamental mode surface waves (the dominant wave mode) in the simulated wavefield.

We first study a case in which the observations are spaced 11 m apart. This recording is extracted from the fully sampled wavefield by subsampling. We perform a series of band-passes to the data with different central frequencies. These band-passes are 1 Hz wide, and the central frequencies are spaced 0.5 Hz apart. We use wavefield gradiometry to estimate the spatial and temporal derivatives, and we perform wave equation inversion by regression to extract the phase slownesses $s_\omega(x)$. We average this phase slowness over space but exclude the measurements made in the source region. The preliminary phase slownesses are shown as a light-blue curve in Figure 2a. This retrieved dispersion curve follows the dispersion image of the data at low frequencies, but it deviates at higher frequencies at which we underestimated the slowness of the surface waves. This is because at high frequencies, the wavelength of the waves is closer to the Nyquist frequency, so the approximation errors of the spatial and temporal finite differences used in the wavefield gradiometry increase. An underestimation of the spatial derivatives results in an overestimation of the velocity (an underestimation of the slowness). We repeat this procedure for the case in which the observations are spaced 22 m apart (shown in Figure 2b). As expected, the finite-difference approximation errors increase with increased interstation spac-

ings. The preliminary phase slowness curve deviates significantly from the true dispersion and corrupts the slowness-frequency dispersion relationship that one would use to reveal the subsurface layering.

To correct the dispersion curves, we use equation 8 and omit the effect of noise in this synthetic data example; thus, $\epsilon = 0$. Even though the effect of the temporal finite-difference approximation is small, we take the spatial and temporal finite-difference approximation errors into account and use the correction factor in equation 9 (see also equation B-13 in Appendix B). Solving the relationship in equation 8 (with $\epsilon = 0$), we obtain the corrected phase slowness curves shown as the red curves in Figure 2a. The corrected dispersion curves are in agreement with the dispersion of the fundamental surface-wave mode.

FIELD DATA EXPERIMENT

Acquisition

We acquired field data on land beside Newton Farm located in southeast Edinburgh on 17 November 2016, a windy and wet day. The field is now a picturesque meadow complete with a rather prolific flock of sheep, but it used to be a landfill. This farm is flanked

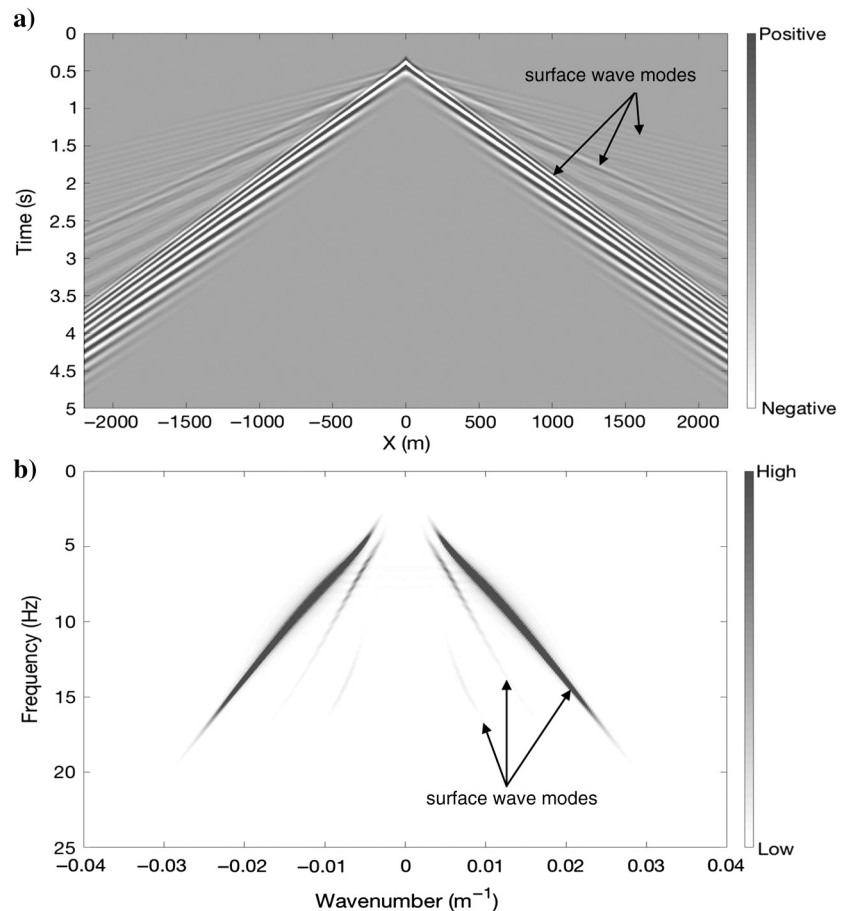


Figure 1. (a) Seismogram of the two-layer synthetic data set. Several surface wave modes can be recognized in the recordings. (b) Frequency-wavenumber spectrum of the two-layer synthetic data set. The strongest amplitudes correspond to the fundamental-mode surface waves.

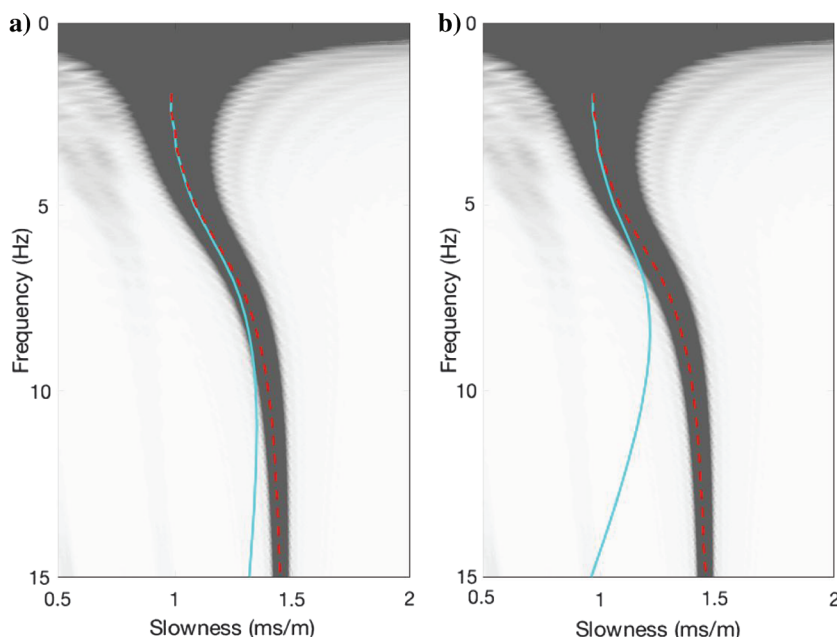


Figure 2. Frequency-slowness spectrum (background shading) and phase slowness measurements (curves) obtained by gradiometry for the two-layer synthetic data set. The strongest amplitudes correspond to the fundamental mode surface wave. The spatial sampling Δx used in the finite differences are (a) $\Delta x = 11$ m and (b) $\Delta x = 22$ m. The light-blue curves show the phase slowness measurements obtained by gradiometry before applying the finite-difference error correction procedure, whereas the red curves show the phase slowness measurements after applying the correction procedure.

by two roads, the Old Craighall Road on the northwestern side and the A720 Ring Road on the southeastern side. Figure 3 shows the main elements of the field site and the deployed seismic array.

We mapped out a square with measure tapes so we could place the geophones with care on a regular rectangular grid. The local coordinate axes mapped out in the field are shown in Figure 3 and are the reference axes in subsequent figures. The azimuth of the y-axis points 37.37° anticlockwise with respect to north. The array consisted of 88 geophones, deployed on an 8×11 grid with 5 m spacing in the x- and y-directions. Each geophone had a 10 Hz corner frequency and recorded virtually no energy of less than 1 Hz (see Figure 4).

We recorded the seismic energy continuously for a little more than an hour with a temporal sampling rate of 125 Hz. Traffic on both roads provided a source of ambient seismic energy. In addition, we recorded seismic energy due to our own footsteps and perhaps the wind and water waves from the nearby coast. A recording geometry deployed as in this field survey could easily be moved around in a rolling fashion until the entire area of interest is covered.

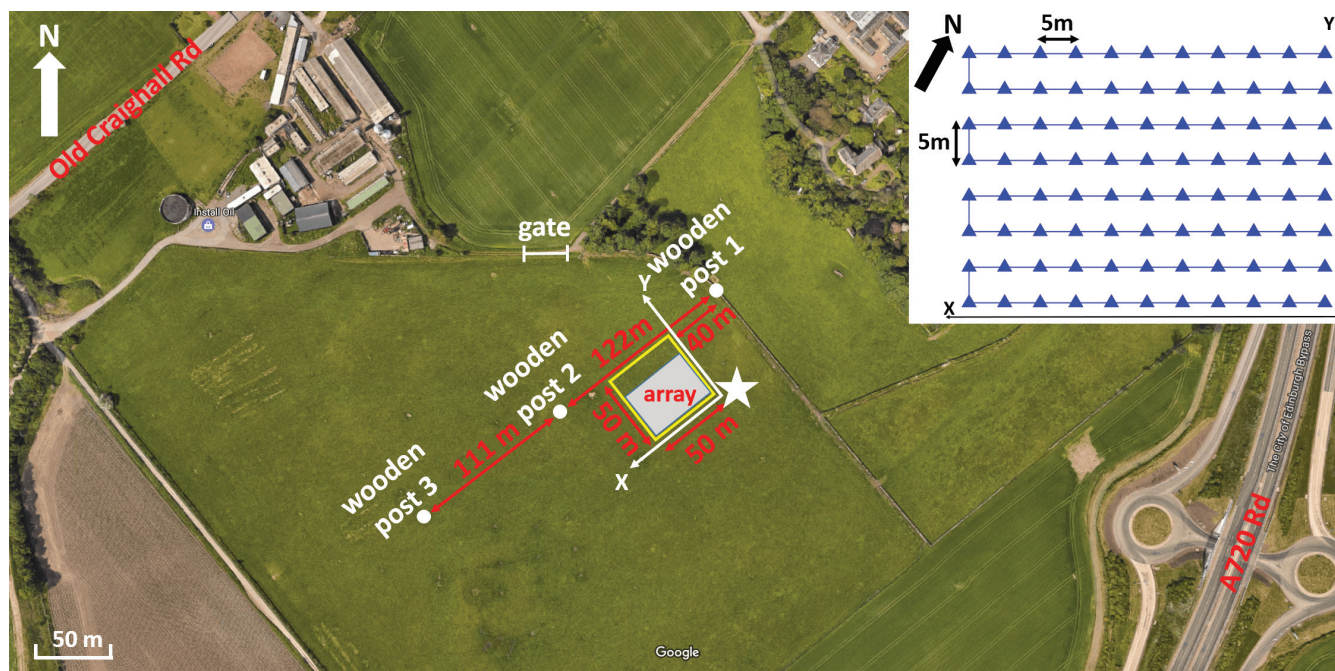


Figure 3. Satellite map of the acquisition field site southeast of Edinburgh. The GPS coordinates of the white star, the origin of our local reference grid, are (55.914766N, 3.070977W). The azimuth of the y-axis points to 37.37° anticlockwise with respect to north. There are two main roads that provide traffic noise: Old Craighall Road to the northwest and the A720 (Edinburgh ring road) to the southeast. The small white circles show wooden posts, and the yellow square corresponds to the local reference coordinate grid that we use to deploy the array. The array geometry within the yellow square is depicted in the top-right corner.

Data characterization and preprocessing

The recorded data had widely varying amplitudes for different stations due to differences in coupling for each geophone. Noise bursts and spikes further complicated processing and analysis. We first applied an automatic gain control (AGC) filter with a short window length (0.25 s) to attenuate amplitude bursts and spikes. To ensure a similar frequency content across stations, we reweighed the complex spectrum by a running average of the complex amplitudes in a 0.12 Hz wide running window. The final result of preprocessing and the effect of the AGC in the time and frequency domains is shown in Figure 4a and 4b. The average absolute amplitude of the data in Figure 4a and 4b is compared in Figure 4c.

Spatially coherent wavefield energy emerges only after this fairly aggressive preprocessing, suggesting that receiver coupling and very near receiver effects dominate over amplitudes of the wavefield that we need for gradiometry. The AGC in time corrects long period trends (low frequencies), but the gradiometry measurements are taken as the second derivatives in time (sensitive to the

high-frequency content). However, differences in coupling between nearby stations are now largely removed. We apply a Hann window band-pass filter with a central frequency of $f_0 = 19$ Hz and width of $\Delta f = 5$ Hz to the data, and we find a clear pattern of waves propagating from across the array (Figure 5a). From an animation of the recorded amplitudes as a function of time, we observe that these waves propagate predominantly in a northwestern direction.

To further analyze the character of the directional wave propagation across the array, we perform a beamform experiment (see, for example, Rost and Thomas, 2002). We filter the data between 18 and 20 Hz and perform a slant-stack transform (stacking the gathers over planes with the given intercept time at zero offset and apparent slownesses along the x - and y -axes). We average the absolute value of the resulting (τ, p_x, p_y) cube over arrival times τ , resulting in an image as a function of the x - and y -direction slowness, p_x and p_y , respectively (Figure 5b). The analysis confirms that the dominant source of the ambient seismic noise comes from the southeast, suggesting that the dominant energy is generated by traffic on the A720 Ring Road. Some seismic energy originates from the northwest,

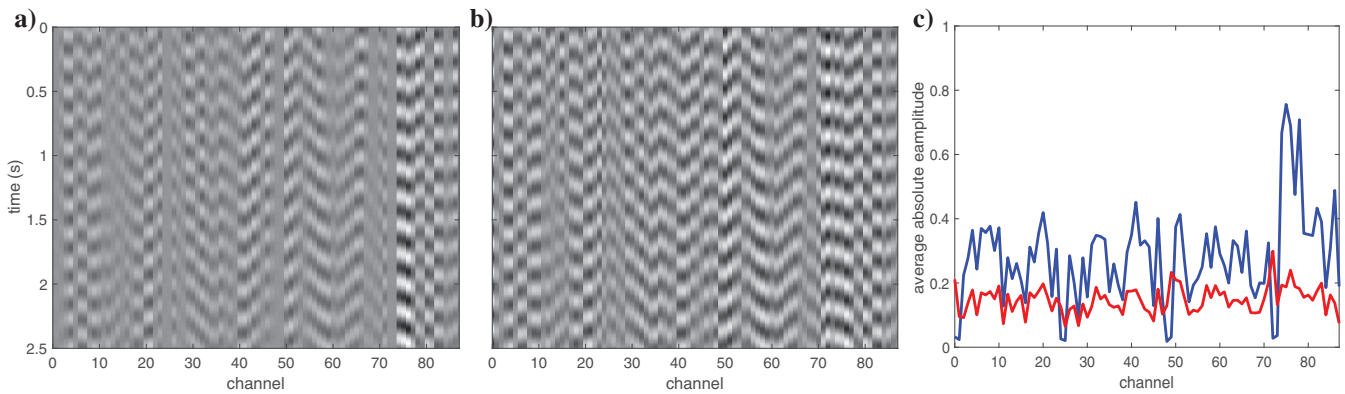


Figure 4. (a) Gather of 2.5 s of raw data. (b) The same gather after preprocessing with AGC in the time domain and reweighing in the frequency domain with a running average. (c) The same gather as in (b) after a Hann window band-pass filter with a central frequency of $f_0 = 18$ Hz and width of $\Delta f = 2$ Hz.

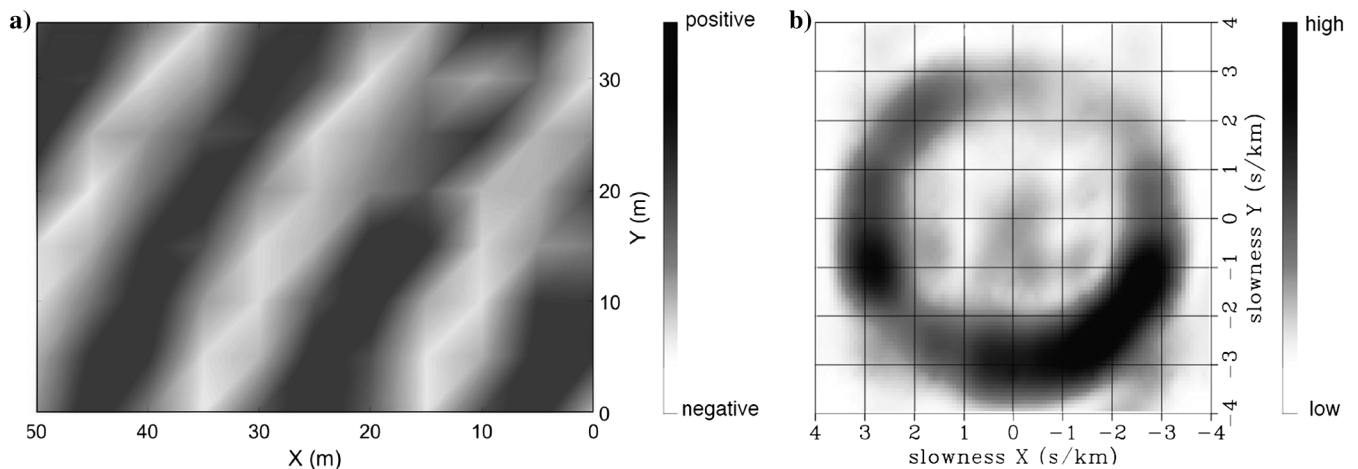


Figure 5. (a) A snapshot of processed recorded ambient seismic noise for data filtered between 18 and 20 Hz. (b) Phase slowness images obtained by beamforming for the same data, again filtered between 18 and 20 Hz.

which may be generated by traffic on Old Craighall Road. It is less clear what generates the seismic energy originating from due west. Possibly, this is seismic energy generated on the A720 Ring Road that is backscattered from a series of railroad tracks west of the field site.

Surface wave dispersion from wavefield gradiometry and wave equation inversion

We further analyzed the data to obtain dispersion curves using independent shorter portions of the full recording. After the above described preprocessing procedures, we successively band-passed the data with 5 Hz wide Hann windows spaced 1 Hz apart from 3 to 36 Hz. For each window and each band-passed recording, we measured spatial and temporal derivatives and inverted the wave equation. This yields dispersion curves for each station location in the interior of the array. We will verify these dispersion curves by comparison with the frequency wavenumber spectrum of the data (f, k_r), where $k_r^2 = k_x^2 + k_y^2$, which shows one dominant wave mode (the background shading in Figure 6).

We processed all of the data, used the neighboring stations to measure the spatial second-order derivative with $\Delta X = 5$ m, and made average dispersion curve measurements across the entire array from 11 recordings, each lasting 3 min. We also made dispersion curve measurements using the full hour of recording. The dispersion curves were averaged over the interior of the array, discarding the station locations at the edges at which no spatial derivative could be measured. This procedure yields average dispersion curves for the entire array, which we normalize by frequency for comparison with the f - k_r spectrum (shown as the light blue curves in Figure 6).

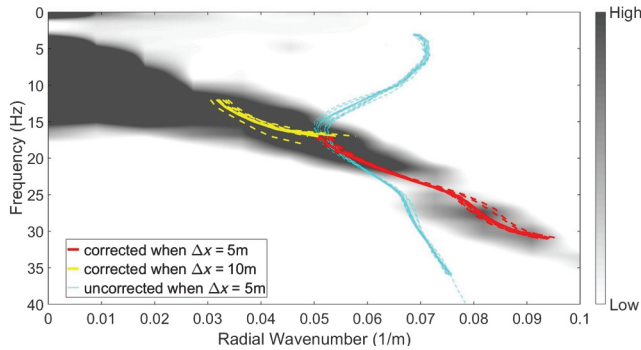


Figure 6. Frequency-wavenumber spectrum (background shading) and corresponding phase slowness measurements (curves) obtained by gradiometry for the ambient-noise field data set. The highest amplitudes correspond to the fundamental surface wave mode. Each set of curves shows the phase slowness measurements obtained by gradiometry and consists of a series of dotted curves (each obtained from single-data files that last approximately 130 s) and one solid curve (a combination of all such files lasting approximately half an hour). The light-blue curves show the phase slowness measurements obtained by gradiometry with a spatial sampling used in the finite differences of $\Delta x = 5$ m without applying the finite-difference error correction procedure. The red curves show the phase slowness measurements obtained with a spatial sampling of $\Delta x = 5$ m after applying the correction procedure. The yellow curves show the phase slowness measurements obtained with a spatial sampling of $\Delta x = 10$ m, after applying the correction procedure.

The dispersion curves obtained by wavefield gradiometry and wave equation inversion do not accord with the array-averaged dispersion image obtained by the frequency wavenumber transform. At higher frequencies, errors due to the finite-difference approximations dominate and the retrieved velocity is too high, whereas at low frequencies, noise dominates the wavefield gradiometry and the retrieved velocity is too low. To correct the dispersion curves, we use equation 8 and include the effect of noise, so $\epsilon \neq 0$. We take only spatial finite-difference approximation errors into account, and the correction factor in equation 9 simplifies to (see also equation B-14 in Appendix B):

$$\gamma_2(s_i) = \sqrt{\frac{1}{\alpha_s^2}} = \frac{s_T 2\pi |f| \Delta x}{\sqrt{2|1 - \cos\{s_T 2\pi f \Delta x\}|}}. \quad (11)$$

This correction factor is valid for the case that $\Delta x \approx \Delta y$ (which is valid for the geometry of the array that we deployed); in the event that $\Delta x \neq \Delta y$, a derivation of this correction factor would be significantly more complicated and a procedure as proposed by de Ridder and Curtis (2017) may be more suitable. We tested several values of ϵ and finally selected $\epsilon = 0.2$, which produced good agreement above 15 Hz between the measured dispersion (the red curves in Figure 6) with the frequency-wavenumber spectrum, both averaged across the entire array. At less than 15 Hz, this agreement breaks down (not shown in Figure 6). Although we cannot assume that the effect of the S/N in the measured gradients is constant over the entire frequency range, we note that even with a constant ϵ , we achieve good results within a certain finite frequency range. This is analyzed further in the “Discussion” section.

To move the support of the signal closer to the Nyquist frequency, we repeat the analysis (again using $\epsilon = 0.2$) by measuring the spatial second-order derivative using the same second-order finite-difference stencil but decimating to a station spacing of $\Delta X = 10$ m. This set of retrieved dispersion curves agrees with the frequency-wavenumber spectrum at less than 15 Hz (the yellow curves in Figure 6).

Finally, we are interested in how the phase velocity changes as a function of space. Using $\epsilon = 0.2$, a value that results in maximum agreement between the array-averaged frequency wavenumber spectrum and the dispersion curves from wavefield gradiometry and wave equation inversion, we correct the dispersion curves for each point in space (with no spatial averaging). This produces phase-velocity maps, shown for 18, 20, 22, and 24 Hz in Figure 7. These phase-velocity maps are obtained by processing the central 30 min of an hour-long recording. We note similar velocity structures for phase velocity maps at different frequencies. This is the first indication that our results are reliable.

To further test the reliability of these maps, we obtained similar maps using the first and second half-hour recordings independently: There is good agreement between maps of phase velocity despite using independent recordings (Figure 8). The velocity anomalies are of the same order of magnitude as the finite-difference error correction to the phase velocities. This might raise concerns that the correction procedure itself generates different structure. We therefore present maps of phase velocities obtained without applying the finite-difference error correction procedure and verify that the correction procedure did not alter the main geometric features

identified in the maps (Figure 9, where we used cubic spline interpolation to aid the visual comparison).

Mixture density network inversion

Using the phase velocity maps for frequencies 18, 20, 22, and 24 Hz shown in Figure 7, we extracted the dispersion curve at each grid location and inverted it for a 1D depth-velocity model beneath that location. Figures 10 and 11 are maps of the mean result and the corresponding standard deviation of the output probability distribution at each depth level. As an example, the 1D depth profile for location $(x, y) = (35 \text{ m}, 10 \text{ m})$ is shown in Figure 12a, and the individual probability density functions at each depth level are shown in Figure 12b–12f. Note that the probability density functions show that by summing multiple Gaussian kernels with different means and standard deviations, non-Gaussian solutions can be recovered, demonstrating that, in principle, the MDN method can give full probabilistic solutions. We used six kernels when training our final MDNs, but we found that no matter how many kernels we used, the networks assigned negligible amplitudes to all but two or three kernels in each case. In other words, the networks were able to

represent the probability density function for each layer using fewer kernels than we allow in the training.

DISCUSSION

We propose to use the scalar wave equation to extract frequency-dependent velocity maps for the fundamental (or dominant) surface-wave mode from dense observations of an elastodynamic wavefield (ambient noise, controlled sources, etc.). The synthetic example shows that when a wavefield is dominated by fundamental-mode surface waves, the body waves and higher order surface waves can be neglected. In the field data example, we learned from the $f-k_r$ spectrum (Figure 6) that our field data are dominated by a single surface wave mode. However, wave equation inversion using wavefield gradiometry to measure the spatial derivatives requires careful analysis of the effect of the spatial sampling and noise (aka that component of the recordings that does not correspond to a noise-free observation of the dominant surface wave mode) in the recordings. We derived an iterative procedure to correct our dispersion curves for the effect of spatial sampling and to include the effect of noise.

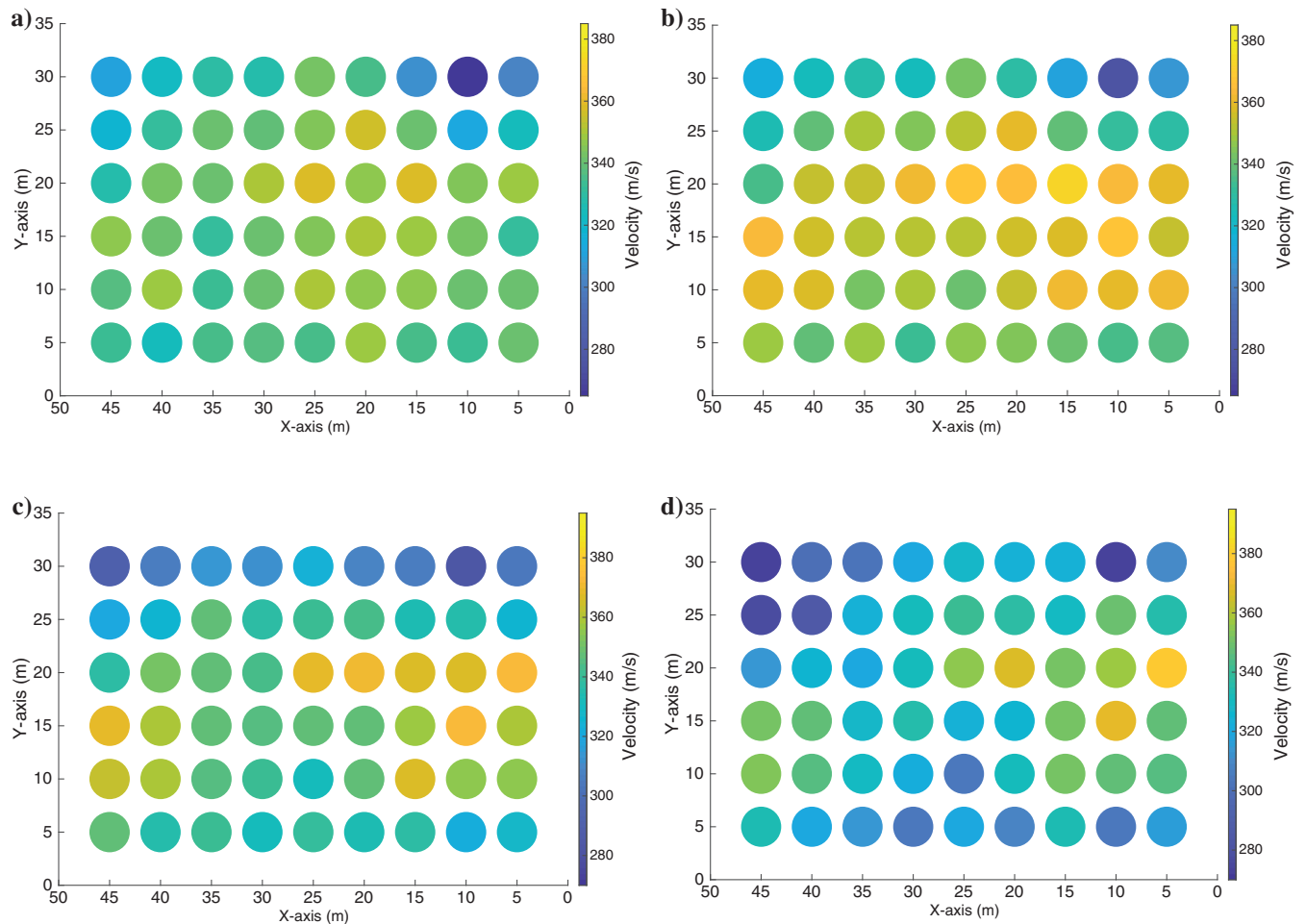


Figure 7. Phase velocity maps obtained from the ambient-noise field data set by gradiometry at four center frequencies: (a) 18 Hz, (b) 20 Hz, (c) 22 Hz, and (d) 24 Hz. These maps were obtained using half an hour of recordings extracted from the middle of the recording period (which lasted a total of 1 h). The axis limits correspond to the complete survey dimensions; the lack of velocity information near the edges is because the finite-difference method requires four neighboring stations (corresponding to locations outside of the acquisition area).

Because we use finite differences using neighboring receivers to measure the spatial derivatives of the wavefield (i.e., wavefield gradiometry), it is interesting to discuss what occurs when the wavefield is sampled beyond the Nyquist criterion. The error in the

measurement of spatial derivatives using finite differences increases when the spacings between observations increases. The expression for the spectrum of the finite difference operator (equation A-5 in Appendix A) holds for spacing larger than the Nyquist criterion;

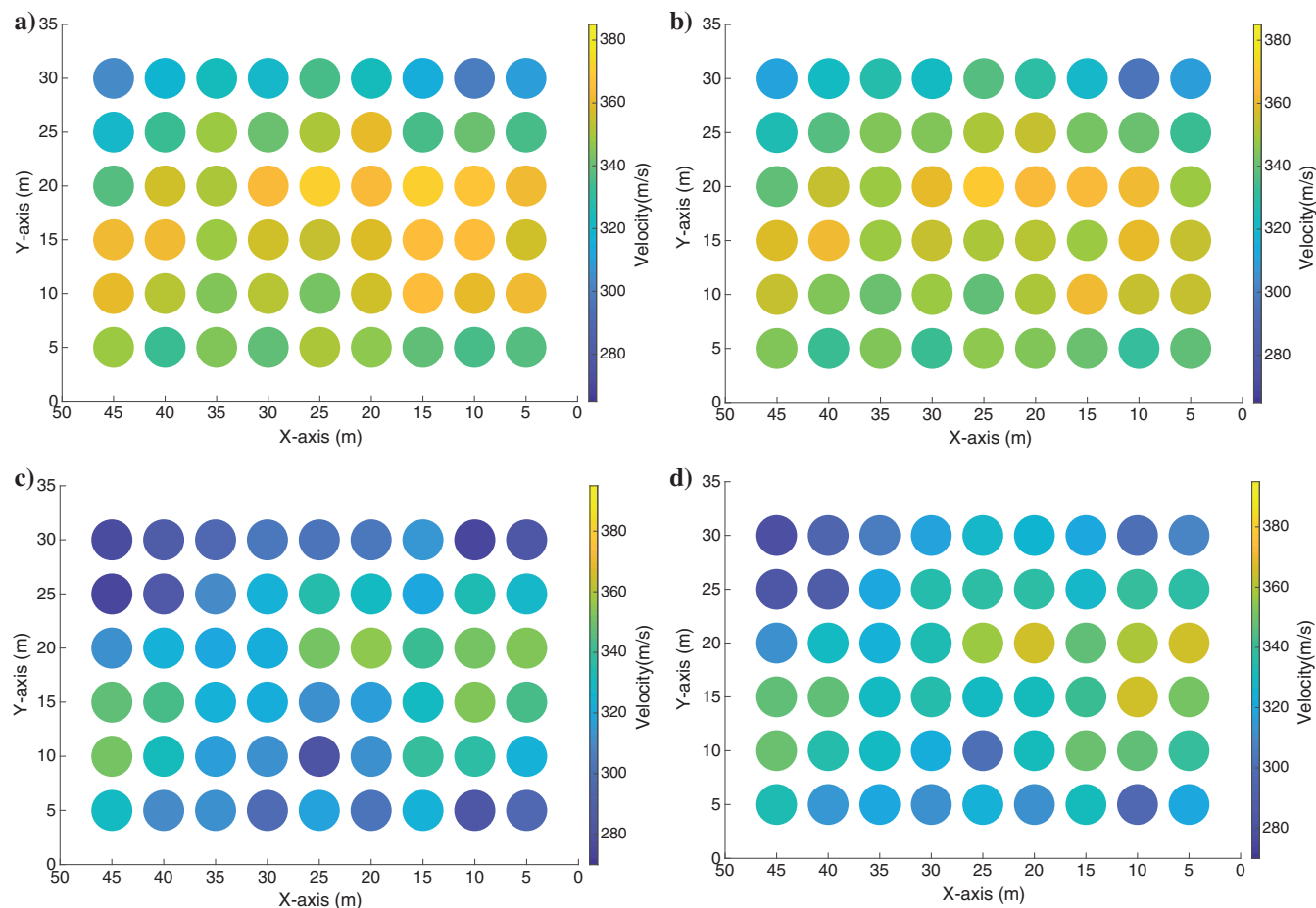


Figure 8. Phase-velocity maps obtained from the ambient-noise field data set by gradiometry at two center frequencies after applying the finite-difference error correction procedure: (a and b) 20 Hz and (c and d) 24 Hz. The maps for each center frequency are obtained from independent data, respectively, (a and c) from the first half hour of recordings and (b and d) from the second half hour of recordings.

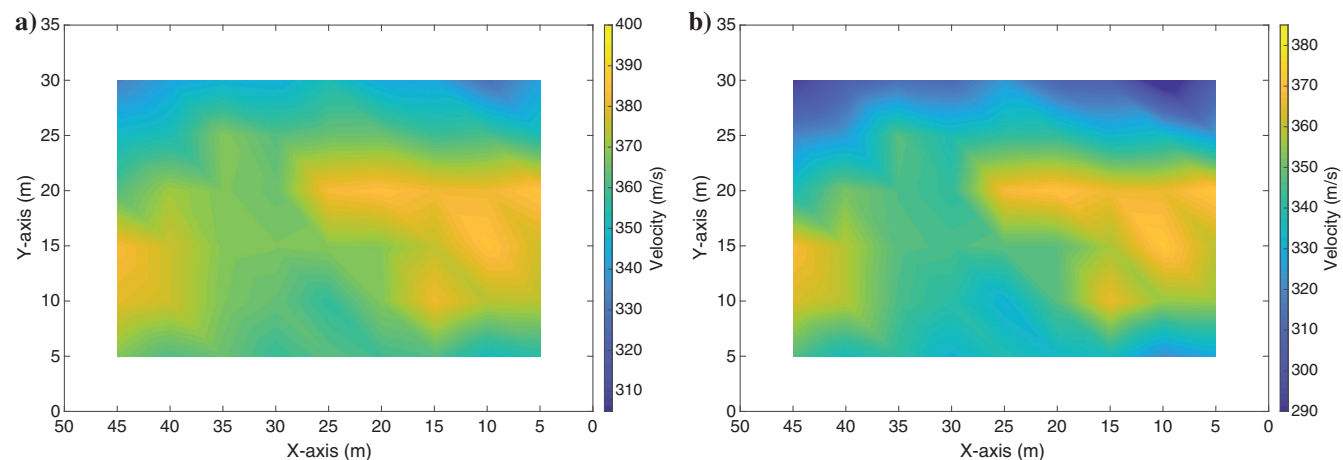


Figure 9. Phase-velocity maps obtained from the ambient-noise field data set by gradiometry at a 22 Hz center frequency, (a) before and (b) after applying the finite-difference error correction procedure.

however, at twice the Nyquist criterion sampling, this spectrum has zero amplitude, as opposed to the correct value of $(2k)^2$, and the correction factor becomes infinite. The (nonlinear) expressions for the error in measured velocities resulting from the approximation error in the spatial finite differences (equations B-13 and B-14 in Appendix B) also hold for station spacings beyond the Nyquist criterion. However, for wavefields that are sampled beyond the Nyquist criterion, we do not expect the fixed-point iterations to converge to the correct slowness, when the starting slowness for the fixed-point iterations is far from the true slowness. More importantly, the effect of noise becomes dominant as the station spacing exceeds the Nyquist criterion.

There are several causes of noise in the recordings. For example, the sensor coupling with the subsurface and electronic noise in the cables typically vary from station to station. We applied an AGC filter to the field data to attenuate coupling artifacts. There may be signals in the recordings resulting from scattering with wavenumbers beyond the Nyquist criterion. Noise typically has support over a wide (perhaps even the full) wavenumber spectrum; however, the signal in which we are interested has only limited support. Therefore, the application of the spatial derivatives (an operator with the spectrum wavenumber-squared) changes the signal-to-noise ratio (S/N). Because the noise near the Nyquist criterion is amplified the most, the S/N is usually improved when the signal has wavenumber support near the Nyquist criterion. However, this also means that ϵ should be dependent on the wavenumber (and therefore frequency). This poses an additional challenge because of the effect of noise in the iterative correction procedure with a factor that has to be estimated from the data (ϵ in Appendix B).

Unfortunately, because the true velocity is unknown, it is challenging to estimate the ϵ factor for a single frequency. Instead, we choose a factor for which we retrieve a dispersion that overlies the dispersion observed in a frequency-wavenumber plot averaged over all of the data. This means that we assume that the effect of noise in the measurements of the spatial wavefield derivatives varies only

weakly as a function of frequency. However, our analysis in the previous paragraph has shown that this is incorrect because the S/N of the signal changes as the wavenumber support of the signal varies with frequency. From the results in Figure 6, we nevertheless see that this assumption is approximately true over small frequency ranges. The assumption breaks down (especially at lower frequencies) as the support of the signal moves away from the Nyquist criterion. Therefore, in the example of Figure 6, we subsampled the wavefield to twice the grid spacing so that the support of the signal remained closer to the Nyquist criterion and the noise does not completely corrupt the analysis. In this study, we manage to extract dispersion curves between 11 and 30 Hz; this bandwidth is limited (especially on the lower end) by not averaging over the entire patch. The upside is that we are able to recover spatially varying dispersion curves.

The required matching process for the noise impact correction may impede the quantitative interpretation of the recovered dispersion curves as a function of frequency, but this does not prevent the quantitative interpretation of lateral variations in the dispersion velocity maps. For example, the phase velocity maps in Figure 7c can be divided into a high-velocity region and a low-velocity region, and these were shown to be robust to using different sections of recorded noise. The boundary may be a partitioning of the buried landfill, and the three peaks in the phase-velocity maps may indicate further heterogeneities.

A principal advantage of the methods presented here is that it can potentially be used on very short noise records, as suggested in Figure 6, with the dashed blue obtained from only 2 min. This would allow the possibility to “roll” a relatively spatially limited array around an area of interest as fast as the equipment can possibly be installed, moved, and reinstalled in new locations; because the sensors would in any case be installed in each set of locations for more than a few minutes, this would allow the surface-wave velocity structure to be estimated across the entire area. We thus create a practical method for mapping and characterizing subsurface properties over extended areas with relatively small and affordable

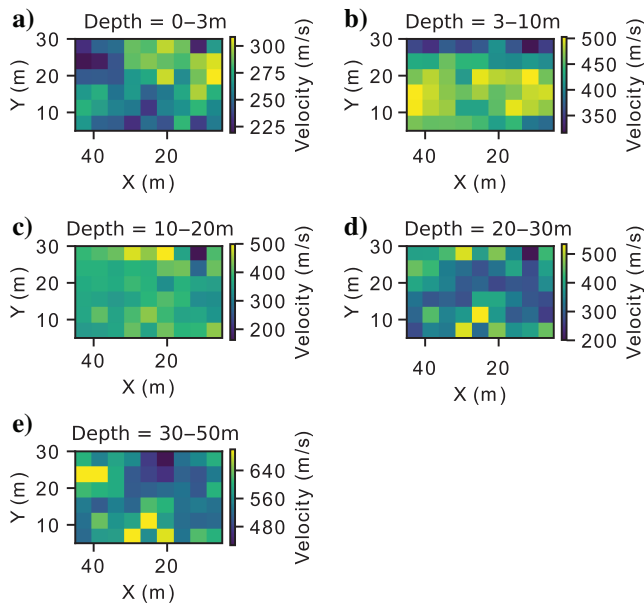


Figure 10. The mean S-wave velocity maps from MDN inversion for (a) 0–3 m, (b) 3–10 m, (c) 10–20 m, (d) 20–30 m, and (e) 30–50 m.

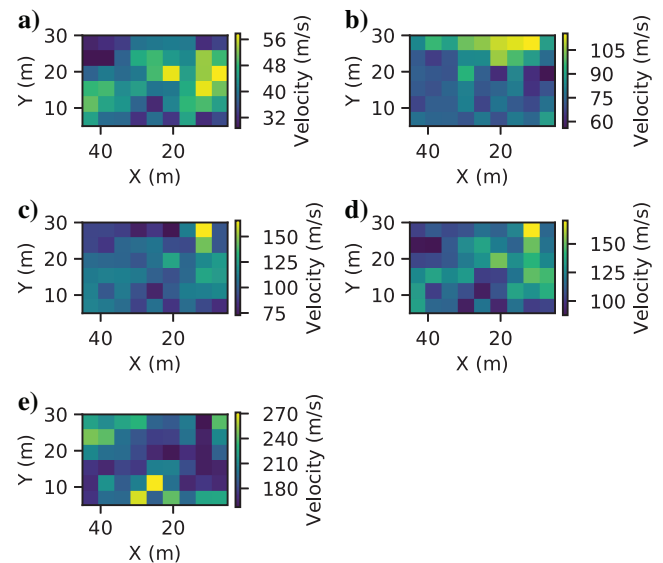


Figure 11. The standard deviation of the S-wave velocity maps from MDN inversion for (a) 0–3 m, (b) 3–10 m, (c) 10–20 m, (d) 20–30 m, and (e) 30–50 m.

sets of equipment. Another advantage of the methods presented here over approaches relying on for example beamforming using smaller subarrays (Löer et al., 2018) is that those approaches rely on stacking wavefields spatially over several wavelengths; its effect of averaging phase velocity estimates over space is certainly stronger than gradiometry and wave equation inversion. If point measurements of spatial wavefield gradients would be available, then the spatial resolution of the wave equation inversion could be as high as the point sensitivity.

We produced five depth maps for S-wave velocity and an indication of their associated uncertainties from inversion by MDN. The mean velocity maps (Figure 10) show a correlation with the phase velocity maps produced in Figure 7. The large uncertainties in the deepest (30–50 m) layer shown in Figures 11e and 12 imply that we are approaching the depth limit to which the frequencies in the data are sensitive.

Standard nonlinear 3D inversion methods, such as Monte Carlo methods (Bodin and Sambridge, 2009; Galetti et al., 2015; Zhang et al., 2018), that are commonly used to produce probabilistic results are computationally expensive. In contrast, an NN takes a couple of hours to train and can then be used repeatedly, without further training, to produce results in a matter of seconds for any data provided as input. By using a mixture of Gaussians, the MDN can represent any shape of distribution as long as a sufficient number of kernels is used. Therefore, using this method, it is possible to produce a non-Gaussian, nonlinear, fully probabilistic solution at a speed that is compatible with the rate at which phase dispersion data can be acquired using gradiometry (of the order of minutes).

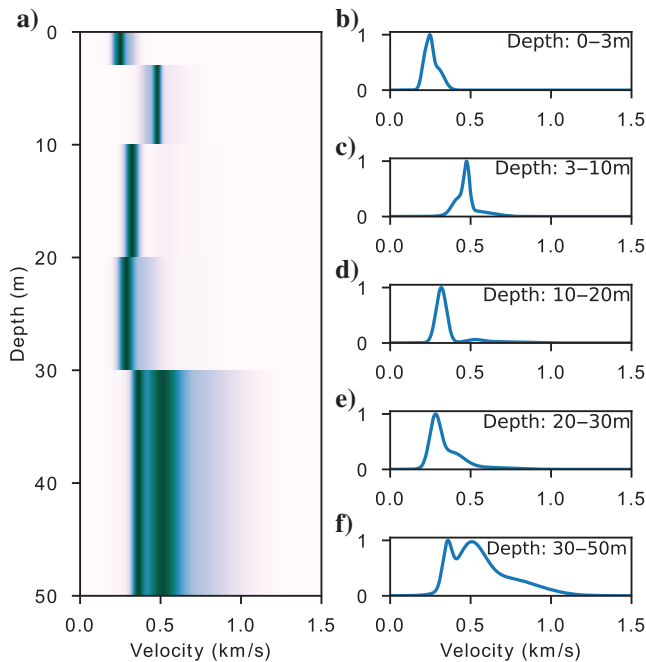


Figure 12. The 1D depth inversion result for location $(x, y) = (35 \text{ m}, 10 \text{ m})$. (a) The MDN posterior density function result: Dark colors represent areas of higher probability. The normalized posterior density function at each depth level is shown in panels (b) 0–3 m, (c) 3–10 m, (d) 10–20 m, (e) 20–30 m, and (f) 30–50 m.

CONCLUSION

We propose a sequence of fast seismic acquisition for dispersion curve extraction and inversion for 3D seismic models, based on wavefield gradiometry, wave equation inversion, and machine-learning technology. We can use short noise recordings, even if they contain directional seismic noise only, made in rapidly deployed rectangular acquisition grids and invert a dispersive scalar Helmholtz equation for phase-velocity maps using measured wavefield gradients. In turn, these phase velocity maps are turned into 3D seismic velocity models, with uncertainty estimates, using pretrained MDNs. We establish a nonlinear relationship between the unknown true seismic wave velocities, the measured seismic wave velocities, the interstation spacing, and the noise level in the signal (with a control coefficient related to the data noise). Our synthetic data example shows that this relationship can be inverted using fixed-point iterations. This procedure corrects for the frequency-dependent approximation error in the finite differences used to measure the wavefield gradients. To correct for the effect of noise in the S/N of the measured wavefield spatial gradients, we have to assume that the control coefficient noise level varies only weakly with frequency, and we need to be able to independently verify the retrieved dispersion curves. We propose to do so using the array-averaged data analysis demonstrated herein. The field data experiment revealed a relatively sharp boundary between low- and high-velocity regions and three more high-velocity anomalies. The lateral variations in the phase velocity maps are not altered by the correction procedures and are reproducible using independent portions of the ambient noise recordings. This indicates that the features identified in the velocity maps are reliable. These maps are obtained using 30 min of noise recordings and computationally cheap processing and inversion techniques. The results of this study open the prospect of near-real-time velocity estimation using dense (and potentially rolling) arrays.

ACKNOWLEDGMENTS

The authors thank A. Shaiban and X. Zhang for help with and advice for synthesizing the synthetic test data. The authors thank editors and reviewer V. Socco, J. Blanch, P. Edme, and two anonymous reviewers for their comments and suggestions that significantly improved the paper. A. Curtis and S. Earp thank the Edinburgh Interferometry Project sponsors (Schlumberger, Equinor, and Total) for supporting this research.

DATA AND MATERIALS AVAILABILITY

Data associated with this research are available and can be obtained by contacting the corresponding author.

APPENDIX A

FOURIER SPECTRA OF FINITE-DIFFERENCE STENCILS

The spectrum of the continuous second-order derivative operator is $|-2\pi k^2|$. The finite-difference approximation to that operator is a stencil applied as a convolution filter. The filter coefficients for the finite-difference approximation to a spatial second-order derivative operator are given as \mathbf{f} (with elements f_i , with $i = 1, \dots, NF$), and the stencil spacing is Δx . We use the filter \mathbf{f} to compute the derivative \mathbf{h} of the discrete function \mathbf{g} (with elements h_j and g_j):

$$h_j = \sum_{i=1}^{NF} f_i g_{j-k+l} \quad \text{where } l = (NF - 1)/2. \quad (\text{A-1})$$

The spectrum of that filter can be computed using the discrete Fourier transform:

$$\mathcal{F}\{\mathbf{f}\}(k) = \Delta x \sum_i^N f_i \exp\{-i2\pi k x_i\}, \quad (\text{A-2})$$

where k is the wavenumber with the unit $(1/m)$. We find the energy normalized spectrum as

$$\frac{1}{\Delta x} |\mathcal{F}\{\mathbf{f}\}(k)|. \quad (\text{A-3})$$

The spatial axis spans the stencil length; hence, $x_i = x_0 + (i - 1)\Delta x$, where $x_0 = -N - 1/2\Delta x$.

For the familiar set of finite-difference coefficients approximating the second-order derivative with second-order accuracy $1/(\Delta x)^2[1, -2, 1]$, we have

$$\begin{aligned} \mathcal{F}\{[1, -2, 1]\}(k) &= \frac{1}{(\Delta x)^2} (\Delta x \exp\{i2\pi k \Delta x\} \\ &\quad - 2\Delta x \exp\{i2\pi k 0\} + \Delta x \exp\{-i2\pi k \Delta x\}) \\ &= \frac{2}{\Delta x} (\cos\{2\pi k \Delta x\} - 1). \end{aligned} \quad (\text{A-4})$$

The energy-normalized amplitude spectrum of the approximation is therefore

$$\frac{1}{\Delta x} |\mathcal{F}\{\mathbf{f}\}(k)| = \frac{|2(\cos\{2\pi k \Delta x\} - 1)|}{(\Delta x)^2}. \quad (\text{A-5})$$

This relationship and a comparison to the amplitude spectrum of the true second-order operator are shown in Figure A-1.

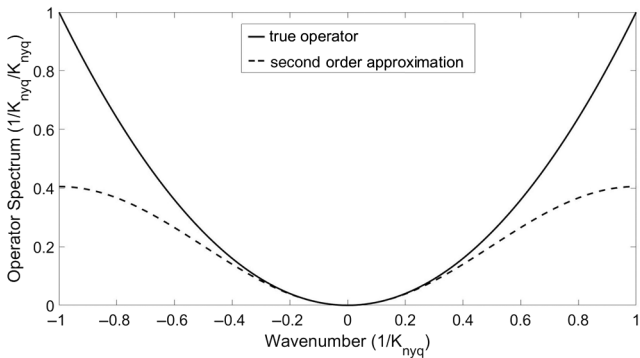


Figure A-1. Wavenumber spectrum of the continuous second-order derivative operator and the spectrum of the finite-difference approximation to that operator (with second-order accuracy). The wavenumber K_{nyq} is normalized to take the value of ± 1 at the Nyquist wavenumbers.

APPENDIX B

FINITE-DIFFERENCE APPROXIMATION ERRORS AND EFFECTS OF NOISE IN WAVE EQUATION INVERSION AND WAVEFIELD GRADIOMETRY

We aim to extract the local wave slowness as the equivalent of the ratio between spatial and temporal derivatives of the wavefield (see equation 2):

$$s^2 \simeq \frac{\nabla^2 U}{\partial_t^2 U}, \quad (\text{B-1})$$

where $\nabla^2 U \triangleq \partial_x^2 U + \partial_y^2 U = \{\partial_x^2 U\}_T + \{\partial_y^2 U\}_T$, $\partial_t^2 U = \{\partial_t^2 U\}_T$, and the subscript $\{\}_T$ denotes the *true* spatial and temporal derivatives. We consider that the total wavefield U is divided into two parts $U = U_S + U_N$, a signal wavefield U_S and a noise field U_N . The correct true velocity in the wave equation inversion would ideally be extracted from noise-free signal-only waveforms as

$$s_T^2 \simeq \frac{\nabla^2 U_S}{\partial_t^2 U_S}. \quad (\text{B-2})$$

In most cases, we cannot measure the signal wavefield independently from the noise, so we estimate the velocity from measured quantities:

$$s_M^2 \simeq \frac{\{\nabla^2 U_S\}_M}{\{\partial_t^2 U_S\}_M}, \quad (\text{B-3})$$

where $\{\}_M$ denotes that these are derivatives *measured* by deploying wavefield gradiometry.

We shall consider the effects of the finite-difference approximation error in the spatial and temporal wavefield derivative computations that are implicit in equation B-1. In this study, we consider that noise mainly affects the spatial derivatives. The measured temporal derivative of the wavefield is assumed to be related to the true temporal derivative of the wavefield by a factor β as

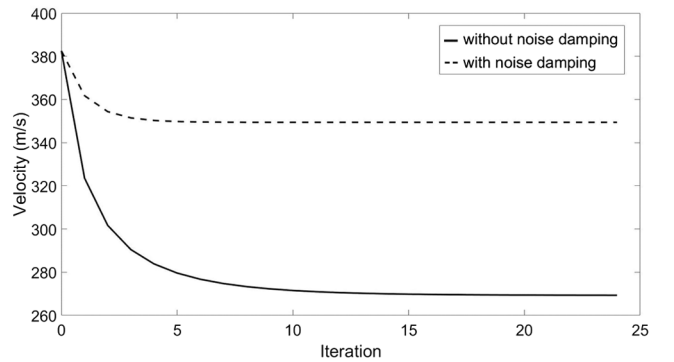


Figure B-1. An example of the iterative correction procedure. The initial measurement (380 m/s) was made at a frequency of 20 Hz for an average over the entire array in the field data example. We plot the phase velocity as a function of iteration number. The solid curve shows the phase velocities during the correction procedure without considering the effect of noise in wavefield gradiometry. The dashed curve shows the phase velocities during the correction procedure while taking into account the effect of noise in wavefield gradiometry (with $\epsilon = 0.2$).

$$\{\partial_t^2 U\}_M = \{\partial_t^2 U_S\}_M = \beta^2 \{\partial_t^2 U_S\}_T. \quad (\text{B-4})$$

The effect of noise on the spatial derivative measurements in equation B-1 cannot be neglected. The derivative is linear, so the spatial derivatives of the total wavefield can be written

$$\partial_x^2 U = \partial_x^2 U_S + \partial_x^2 U_N, \quad (\text{B-5})$$

$$\partial_y^2 U = \partial_y^2 U_S + \partial_y^2 U_N. \quad (\text{B-6})$$

However, measuring the spatial derivative of the wavefield accurately is further complicated by the error of the finite-difference approximation. Our measured results for the spatial wavefield derivatives are therefore related to the wavefield derivatives of the true signal and the derivatives of the noise field by two factors α_S and α_N as

$$\begin{aligned} \{\partial_x^2 U\}_M &= \{\partial_x^2 U_S\}_M + \{\partial_x^2 U_N\}_M \\ &= \alpha_S^2 \{\partial_x^2 U_S\}_T + \alpha_N^2 \{\partial_x^2 U_N\}_T, \end{aligned} \quad (\text{B-7})$$

$$\begin{aligned} \{\partial_y^2 U\}_M &= \{\partial_y^2 U_S\}_M + \{\partial_y^2 U_N\}_M \\ &= \alpha_S^2 \{\partial_y^2 U_S\}_T + \alpha_N^2 \{\partial_y^2 U_N\}_T. \end{aligned} \quad (\text{B-8})$$

Here, we implicitly assumed that when evaluating the spatial derivatives by finite differences $\Delta x = \Delta y$, the characteristics of the effect of noise are directionally invariant. This does not imply that the seismic wavefield itself is directionally invariant.

The desired true slowness in equation B-2 can then be found by

$$\begin{aligned} s_T^2 &\simeq \frac{\beta^2 [\{\nabla^2 U\}_M - \alpha_N^2 \{\nabla^2 U_N\}_T]}{\alpha_S^2 \{\partial_t^2 U\}_M} \\ &= \frac{\beta^2}{\alpha_S^2} \left[1 - \alpha_N^2 \frac{\{\nabla^2 U_N\}_T}{\{\nabla^2 U\}_M} \right] \frac{\{\nabla^2 U\}_M}{\{\partial_t^2 U\}_M}. \end{aligned} \quad (\text{B-9})$$

We define a new parameter $\epsilon = \alpha_N^2 (\{\nabla^2 U_N\}_T / (\{\nabla^2 U\}_M))$. Because ϵ depends on the (dominant) wavelength of the noise (and hence of the noise amplitude) and ϵ is a “noise-to-signal ratio.” This term should be small, so we consider $\epsilon < 1$. If we assume that the (dominant) wavelength of noise and its amplitude remain constant during a recording, ϵ is an unknown constant.

From equations B-3 and B-9, we find that the relationship between the true slowness s_T and the measured slowness s_M can be written as

$$s_T \simeq \frac{\beta}{\alpha_S} \sqrt{1 - \epsilon} s_M, \quad (\text{B-10})$$

where $\epsilon < 1$. When $\epsilon = 0$, or when the effect of noise is neglected, we correct a zero noise level.

We defined the relationship between the measured and the true spatial and temporal derivatives of the signal wavefield as, respectively, $\{\partial_x^2 U_S\}_t = \alpha_S^2 \{\partial_x^2 U_S\}_m$ and $\{\partial_t^2 U_S\}_t = \alpha_S^2 \{\partial_t^2 U_S\}_m$. Using the expressions for the Fourier spectra of the true continuous operator and the measured finite-difference approximation to the operator from Appendix A, we find

$$\alpha_S^2 = \frac{-(2\pi|k_M|)^2}{-(2\pi|k_S|)^2} = \frac{\frac{2|\cos\{2\pi k_S \Delta x\} - 1|}{(\Delta x)^2}}{-(2\pi|k_S|)^2} \quad (\text{B-11})$$

and

$$\beta^2 = \frac{-(2\pi|f_M|)^2}{-(2\pi|f|)^2} = \frac{\frac{2|\cos\{2\pi f \Delta t\} - 1|}{(\Delta t)^2}}{-(2\pi|f|)^2}. \quad (\text{B-12})$$

For the case where we consider spatial and temporal finite-difference approximation errors, we define a new parameter γ_1 (using $k_S = s_T f$) as

$$\gamma_1(s_T) = \sqrt{\frac{\beta^2}{\alpha_S^2}} = \frac{\sqrt{|1 - \cos\{2\pi f \Delta t\}|}}{\sqrt{|1 - \cos\{s_T 2\pi f \Delta x\}|}} \frac{\Delta x}{\Delta t} s_T. \quad (\text{B-13})$$

For the case in which we neglect temporal finite-difference errors and only consider spatial finite-difference approximation errors, we have $\beta = 1$ and define a new parameter γ_2 as

$$\gamma_2(s_T) = \sqrt{\frac{1}{\alpha_S^2}} = \frac{s_T 2\pi|f|\Delta x}{\sqrt{2|1 - \cos\{s_T 2\pi f \Delta x\}|}}. \quad (\text{B-14})$$

Finally, with either of these definitions, the relationship between the true slowness s_T and the measured slowness s_M can be written as

$$s_T = \gamma(s_T) \sqrt{1 - \epsilon} s_M. \quad (\text{B-15})$$

This is a nonlinear relationship, which we attempt to solve by fixed-point iterations:

$$s_j = \gamma(s_{j-1}) \sqrt{1 - \epsilon} s_M \quad j = 1, 2, \dots, n, \quad (\text{B-16})$$

starting with $s_0 = s_M$ and finishing with $s_T \simeq s_n$ after n iterations. Finally, the true velocity is given by $c_T = s_T^{-1}$. An example of this procedure, for the measurement at 20 Hz averaged across the entire array in the field data study, is shown in Figure B-1.

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.
- Aki, K., and P. G. Richards, 2002, *Quantitative seismology*, 2nd ed.: University Science Books.
- Allmark, C., 2018, *Analysing the Earth's near surface using ambient seismic noise*: Ph.D. thesis, University of Edinburgh.
- Bishop, C. M., 1995, *Neural networks for pattern recognition*: Oxford University Press.
- Bodin, T., and M. Sambridge, 2009, Seismic tomography with the reversible jump algorithm: *Geophysical Journal International*, **178**, 1411–1436, doi: [10.1111/j.1365-246X.2009.04226.x](https://doi.org/10.1111/j.1365-246X.2009.04226.x).
- Curtis, A., and J. O. A. Robertsson, 2002, Volumetric wavefield recording and wave equation inversion for near-surface material properties: *Geophysics*, **67**, 1602–1611, doi: [10.1190/1.1512751](https://doi.org/10.1190/1.1512751).
- de Ridder, S. A. L., and B. L. Biondi, 2015, Near-surface Scholte wave velocities at Ekofisk from short noise recordings by seismic noise gradiometry: *Geophysical Research Letters*, **42**, 7031–7038, doi: [10.1002/grl.v42.17](https://doi.org/10.1002/grl.v42.17).
- de Ridder, S. A. L., B. L. Biondi, and D. Nichols, 2015, Elliptical-anisotropic eikonal phase velocity tomography: *Geophysical Research Letters*, **42**, 758–764, doi: [10.1002/2014GL062805](https://doi.org/10.1002/2014GL062805).
- de Ridder, S. A. L., and A. Curtis, 2017, Seismic gradiometry using ambient seismic noise in an anisotropic earth: *Geophysical Journal International*, **209**, 1168–1179, doi: [10.1093/gji/ggx073](https://doi.org/10.1093/gji/ggx073).

- de Ridder, S. A. L., and J. Dellinger, 2011, Ambient seismic noise eikonal tomography for near-surface imaging at Valhall: *The Leading Edge*, **30**, 506–512, doi: [10.1190/1.3589108](https://doi.org/10.1190/1.3589108).
- Deuffhard, P., 2012, A short history of Newton's method: *Documenta Mathematica*, Extra Volume ISMP, 25–30.
- Edme, P., and S. Yuan, 2016, Local dispersion curve estimation from seismic ambient noise using spatial gradients: *Interpretation*, **4**, no. 3, SJ17–SJ27, doi: [10.1190/INT-2016-0003.1](https://doi.org/10.1190/INT-2016-0003.1).
- Galetti, E., A. Curtis, G. A. Meles, and B. Baptie, 2015, Uncertainty loops in travel-time tomography from nonlinear wave physics: *Physical Review Letters*, **114**, 148501, doi: [10.1103/PhysRevLett.114.148501](https://doi.org/10.1103/PhysRevLett.114.148501).
- Gerver, M. L., and V. M. Markushevich, 1966, Determination of a seismic wave velocity from the travel-time curve: *Geophysical Journal International*, **11**, 165–173, doi: [10.1111/j.1365-246X.1966.tb03498.x](https://doi.org/10.1111/j.1365-246X.1966.tb03498.x).
- Gerver, M. L., and V. M. Markushevich, 1972, Determining seismic-wave velocity from travel-time curves, in V. I. Keilis-Borok and E. A. Flinn, eds., *Computational seismology*: Springer US, 123–147.
- Gouédon, P., H. Yao, F. Ernst, and R. D. van der Hilst, 2012, Surface wave eikonal tomography in heterogeneous media using exploration data: *Geophysical Journal International*, **191**, 781–788, doi: [10.1111/j.1365-246X.2012.05652.x](https://doi.org/10.1111/j.1365-246X.2012.05652.x).
- IRIS, 2018, Recording the full seismic wavefield, <https://www.iris.edu/>, accessed 15 May 2018.
- Kennett, B. L. N., 1981, Slowness techniques in seismic interpretation: *Journal of Geophysical Research: Solid Earth*, **86**, 11575–11584, doi: [10.1029/JB086iB12p11575](https://doi.org/10.1029/JB086iB12p11575).
- Langston, C. A., 2007a, Spatial gradient analysis for linear seismic arrays: *Bulletin of the Seismological Society of America*, **97**, 265–280, doi: [10.1785/0120060100](https://doi.org/10.1785/0120060100).
- Langston, C. A., 2007b, Wave gradiometry in the time domain: *Bulletin of the Seismological Society of America*, **97**, 926–933, doi: [10.1785/0120060152](https://doi.org/10.1785/0120060152).
- Langston, C. A., 2007c, Wave gradiometry in two dimensions: *Bulletin of the Seismological Society of America*, **97**, 401–416, doi: [10.1785/0120060138](https://doi.org/10.1785/0120060138).
- Langston, C. A., and M. M. Ayele, 2016, Vertical seismic wave gradiometry: Application at the San Andreas Fault Observatory at depth: *Geophysics*, **81**, no. 3, D233–D243, doi: [10.1190/geo2015-0404.1](https://doi.org/10.1190/geo2015-0404.1).
- LeVeque, R., 2007, Finite difference methods for ordinary and partial differential equations: SIAM.
- Liang, C., and C. A. Langston, 2009, Wave gradiometry for USArray: Rayleigh waves: *Journal of Geophysical Research: Solid Earth*, **114**, B02308, doi: [10.1029/2008JE003095](https://doi.org/10.1029/2008JE003095).
- Lin, F.-C., D. Li, R. W. Clayton, and D. Hollis, 2013, High-resolution 3D shallow crustal structure in Long Beach, California: Application of ambient noise tomography on a dense seismic array: *Geophysics*, **78**, no. 4, Q45–Q56, doi: [10.1190/geo2012-0453.1](https://doi.org/10.1190/geo2012-0453.1).
- Lin, F.-C., and M. H. Ritzwoller, 2011, Helmholtz surface wave tomography for isotropic and azimuthally anisotropic structure: *Geophysical Journal International*, **186**, 1104–1120, doi: [10.1111/j.1365-246X.2011.05070.x](https://doi.org/10.1111/j.1365-246X.2011.05070.x).
- Lin, F.-C., M. H. Ritzwoller, and R. Snieder, 2009, Eikonal tomography: Surface wave tomography by phase front tracking across a regional broad-band seismic array: *Geophysical Journal International*, **177**, 1091–1110, doi: [10.1111/j.1365-246X.2009.04105.x](https://doi.org/10.1111/j.1365-246X.2009.04105.x).
- Liu, Y., and W. E. Holt, 2015, Wave gradiometry and its link with Helmholtz equation solutions applied to USArray in the eastern U.S.: *Journal of Geophysical Research: Solid Earth*, **120**, 5717–5746, doi: [10.1002/2015JB011982](https://doi.org/10.1002/2015JB011982).
- Löer, K., E. H. Saenger, and N. Riahi, 2018, Three-component ambient noise beamforming in the Parkfield area: *Geophysical Journal International*, **213**, 1478–1491, doi: [10.1093/gji/ggy058](https://doi.org/10.1093/gji/ggy058).
- Maeda, T., K. Nishida, R. Takagi, and K. Obara, 2016, Reconstruction of a 2D seismic wavefield by seismic gradiometry: *Progress in Earth and Planetary Science*, **3**, 31, doi: [10.1186/s40645-016-0107-4](https://doi.org/10.1186/s40645-016-0107-4).
- Meier, U., A. Curtis, and J. Trampert, 2007a, Fully nonlinear inversion of fundamental mode surface waves for a global crustal model: *Geophysical Research Letters*, **34**, L16304, doi: [10.1029/2007GL030989](https://doi.org/10.1029/2007GL030989).
- Meier, U., A. Curtis, and J. Trampert, 2007b, Global crustal thickness from neural network inversion of surface wave data: *Geophysical Journal International*, **169**, 706–722, doi: [10.1111/j.1365-246X.2007.03373.x](https://doi.org/10.1111/j.1365-246X.2007.03373.x).
- Mordret, A., N. Shapiro, S. Singh, P. Roux, and O. I. Barkved, 2013a, Helmholtz tomography of ambient noise surface wave data to estimate Scholte wave phase velocity at Valhall Life of the field: *Geophysics*, **78**, no. 2, WA99–WA109, doi: [10.1190/geo2012-0303.1](https://doi.org/10.1190/geo2012-0303.1).
- Mordret, A., N. Shapiro, S. Singh, P. Roux, J.-P. Montagner, and O. I. Barkved, 2013b, Azimuthal anisotropy at Valhall: The Helmholtz equation approach: *Geophysical Research Letters*, **40**, 2636–2641, doi: [10.1002/grl.50447](https://doi.org/10.1002/grl.50447).
- Muijs, R., J. O. A. Robertsson, A. Curtis, and K. Holliger, 2003, Near-surface seismic properties for elastic wavefield decomposition: Estimates based on multicomponent land and seabed recordings: *Geophysics*, **68**, 2073–2081, doi: [10.1190/1.1635061](https://doi.org/10.1190/1.1635061).
- Poppeliers, C., 2010, Seismic wave gradiometry using the wavelet transform: Application to the analysis of complex surface waves recorded at the Glendora Array, Sullivan, Indiana, USA: *Bulletin of the Seismological Society of America*, **100**, 1211–1224, doi: [10.1785/0120090304](https://doi.org/10.1785/0120090304).
- Poppeliers, C., 2011, Multiwavelet seismic-wave gradiometry: *Bulletin of the Seismological Society of America*, **101**, 2108–2121, doi: [10.1785/0120100226](https://doi.org/10.1785/0120100226).
- Poppeliers, C., P. Punoševac, and T. Bell, 2013, Three-dimensional seismic-wave gradiometry for scalar waves: *Bulletin of the Seismological Society of America*, **103**, 2151–2160, doi: [10.1785/0120120224](https://doi.org/10.1785/0120120224).
- Rost, S., and C. Thomas, 2002, Array seismology: Methods and applications: *Reviews of Geophysics*, **40**, 2-1–2-27, doi: [10.1029/2000RG000100](https://doi.org/10.1029/2000RG000100).
- Saito, M., 1988, Disper80: A subroutine package for the calculation of seismic normal mode solutions: Academic Press.
- Shahraeeni, M. S., and A. Curtis, 2011, Fast probabilistic nonlinear petrophysical inversion: *Geophysics*, **76**, no. 2, E45–E58, doi: [10.1190/1.3540628](https://doi.org/10.1190/1.3540628).
- Shapiro, N., and M. Ritzwoller, 2002, Monte-Carlo inversion for a global shear-velocity model of the crust and upper mantle: *Geophysical Journal International*, **151**, 88–105, doi: [10.1046/j.1365-246X.2002.01742.x](https://doi.org/10.1046/j.1365-246X.2002.01742.x).
- Sollberger, D., C. Schmelzbach, J. O. A. Robertsson, S. A. Greenhalgh, Y. Nakamura, and A. Khan, 2016, The shallow elastic structure of the lunar crust: New insights from seismic wavefield gradient analysis: *Geophysical Research Letters*, **43**, 10078–10087, doi: [10.1002/2016GL070883](https://doi.org/10.1002/2016GL070883).
- Wielandt, E., 1993, Propagation and structural interpretation of non-plane waves: *Geophysical Journal International*, **113**, 45–53, doi: [10.1111/j.1365-246X.1993.tb02527.x](https://doi.org/10.1111/j.1365-246X.1993.tb02527.x).
- Zhan, Z., Q. Li, and J. Huang, 2018, Application of wavefield compressive sensing in surface wave tomography: *Geophysical Journal International*, **213**, 1731–1743, doi: [10.1093/gji/ggy082](https://doi.org/10.1093/gji/ggy082).
- Zhang, X., A. Curtis, E. Galetti, and S. de Ridder, 2018, 3-D Monte Carlo surface wave tomography: *Geophysical Journal International*, **215**, 1644–1658, doi: [10.1093/gji/ggy362](https://doi.org/10.1093/gji/ggy362).