

Title	Optimized workflows for high-frequency seismic interferometry using dense arrays
Creators	Xu, Yihe and Lebedev, Sergei and Meier, Thomas and Bonadio, Raffaele and Bean, Christopher J.
Date	2021
Citation	Xu, Yihe and Lebedev, Sergei and Meier, Thomas and Bonadio, Raffaele and Bean, Christopher J. (2021) Optimized workflows for high-frequency seismic interferometry using dense arrays. Geophysical Journal International, 227 (2). pp. 875-897. ISSN 0956-540X Online ISSN 1365-246X
URL	<a href="https://dair.dias.ie/id/eprint/1154/">https://dair.dias.ie/id/eprint/1154/</a>
DOI	<a href="https://doi.org/10.1093/gji/ggab260">https://doi.org/10.1093/gji/ggab260</a>

# Optimized workflows for high-frequency seismic interferometry using dense arrays

Y. Xu<sup>1,2</sup>, S. Lebedev<sup>1</sup>, T. Meier<sup>3</sup>, R. Bonadio<sup>1</sup> and C.J. Bean<sup>1</sup>

<sup>1</sup>*Geophysics Section, School of Cosmic Studies, Dublin Institute for Advanced Studies, 5 Merrion Square North, Dublin 2 D02Y006, Ireland. E-mail: xuyh@cp.dias.ie*

<sup>2</sup>*Institute of Geophysics, China Earthquake Administration, 5 Minzudaxue South Road, Beijing 100081, China*

<sup>3</sup>*Institute of Geosciences, Christian-Albrechts Universität Kiel, Otto-Hahn-Platz 1, D-24118 Kiel, Germany*

Accepted 2021 July 1. Received 2021 June 26; in original form 2021 March 5

## SUMMARY

High-frequency seismic surface waves sample the top few tens of meters to the top few kilometres of the subsurface. They can be used to determine 3-D distributions of shear-wave velocities and to map the depths of discontinuities (interfaces) within the crust. Passive seismic imaging, using ambient noise as the source of signal, can thus be an effective tool of exploration for mineral, geothermal and other resources, provided that sufficient high-frequency signal is available in the ambient noise wavefield and that accurate, high-frequency measurements can be performed on this signal. Ambient noise imaging using the ocean-generated noise at 5–30 s periods is now a standard method, but less signal is available at frequencies high enough for deposit-scale imaging (0.2–30 Hz), and few studies have reported successful measurements in broad frequency bands. Here, we develop a workflow for the measurement of high-frequency, surface wave phase velocities in very broad frequency ranges. Our workflow comprises (1) a new noise cross-correlation procedure that accounts for the non-stationary properties of the high-frequency noise sources, removes bandpass filtering, replaces temporal normalization with short time window stacking, and drops the explicit spectral normalization by adopting cross-coherence; (2) a new phase-velocity measurement method that extends the bandwidth of reliable measurements by exploiting the (resolved)  $2\pi$  ambiguity of phase-velocity measurements and (3) interstation-distance-dependent quality control that uses the similarity of subgroups of dispersion curves to reject outliers and identify the frequency ranges with accurate measurements. The workflow is highly automated and applicable to large arrays. Applying our method to data from a large-N array that operated for one month near Marathon, Ontario, Canada, we use rectangular subarrays with 150-m station spacing and, typically, 1 hr of data and obtain Rayleigh-wave phase-velocity measurements in a 0.5–30 Hz frequency range, spanning over 5.9 octaves, twice the typical frequency range of 1.5–3 octaves in previous studies. Phase-velocity maps and the subregion-average 1-D velocity models they constrain show a high-velocity anomaly consistent with the known, west-dipping gabbro intrusions beneath the area. The new structural information can improve our understanding of the geometry of the gabbro intrusions, hosting the Cu-PGE Marathon deposit.

**Key words:** Time-series analysis; Seismic interferometry; Seismic noise; Surface waves and free oscillations.

## 1 INTRODUCTION

Seismic interferometry uses cross-correlation of seismic records from receiver pairs as a means of investigating the properties of the medium (Lobkis & Weaver 2001; Campillo & Paul 2003; Snieder 2004; Wapenaar 2004; Roux *et al.* 2005b; Shapiro *et al.* 2005; Curtis *et al.* 2006). Seismic interferometry can use both surface and body waves, from both active and passive source data (e.g. Roux *et al.*

2005a; Draganov *et al.* 2007; Halliday *et al.* 2008). Passive source interferometry at higher frequencies typically uses signal from the ambient noise wave field.

The ambient noise surface wave tomography (ANSWT) is now an established method, using surface waves in the noise wave field to constrain shear wave velocity models of the crust and upper mantle (Shapiro *et al.* 2005; Yao *et al.* 2006; Lin *et al.* 2007; Yang *et al.* 2007; Saygin & Kennett 2010; Calkins *et al.* 2011; Zheng *et al.*

2011; Ekström 2014; Kästle *et al.* 2016; Shen *et al.* 2016). Theoretical studies suggest that the successful recovery of surface waves from noise depends on the distribution of noise sources (Wapenaar 2004; Halliday & Curtis 2008; Tsai 2009; Kimman & Trampert 2010; Kästle *et al.* 2016). The sources used in the crustal-scale imaging are ocean-generated microseisms, recorded globally in a period range from a few seconds to a few tens of seconds (typically, 0.03–0.2 Hz), allowing imaging in any location with an array of stations.

ANSWT is also applicable, in principle, at frequencies higher than the ocean microseism band, that is  $>0.2$  Hz. Surface waves at these frequencies sample depth ranges from a few tens of meters to a few kilometres below the surface and can constrain the distributions of seismic wave speeds and the locations of interfaces at depth (Brenquiere *et al.* 2007; Gouédard *et al.* 2008; Halliday *et al.* 2008; Picozzi *et al.* 2009; Renalier *et al.* 2010; Behm & Snieder 2013). The emergence of large, dense arrays with numerous sensors (also known as Large-N array) over the last decade makes passive seismic imaging an increasingly powerful tool for the mineral, geothermal, hydrocarbon and other natural resource exploration (de Ridder & Dellinger 2011; Lin *et al.* 2013; Mordret *et al.* 2013; Spica *et al.* 2018; Chmiel *et al.* 2019), monitoring (de Ridder & Biondi 2013) and investigation of the structure of basins, volcanoes and fault zones (Hannemann *et al.* 2014; Mordret *et al.* 2015; Li *et al.* 2016; Wang *et al.* 2017; Wu *et al.* 2017; Lehujeur *et al.* 2018; Chmiel *et al.* 2019; Inzunza *et al.* 2019; Mordret *et al.* 2019; Zigone *et al.* 2019). Its advantages include its relatively low cost, low environmental impact, and applicability over rugged terrain.

The approach, however, can only work under three conditions. First, sufficient high-frequency signal must be available in the ambient noise wavefield. This is not given at frequencies of a few Hz and above, where the ocean-generated noise dies out. Secondly, a sufficiently dense array of sensors must be deployed over the area. Finally, effective methods are required for accurate, broad-band surface wave measurements at high frequencies. For surface wave imaging, specifically, the problem is to develop robust methods and processing workflows for obtaining accurate dispersion measurements with maximum bandwidths.

A widely used workflow presented by Bensen *et al.* (2007) used bandpass filtering, temporal normalization, spectral whitening and cross-correlation to produce noise correlation functions (NCFs), which, in this study, refers to the result computed by either cross-correlation, or cross-coherence or deconvolution, without differentiation; see Wapenaar *et al.* 2010; Nakata *et al.* 2011), and group- or phase-velocity measurement via the multiple filter and matched filter techniques to retrieve dispersion curves from the NCFs. The workflow has been successfully applied to data from numerous broad-band seismic networks (e.g. Lin *et al.* 2007; Yang *et al.* 2007; Bensen *et al.* 2008; Zheng *et al.* 2008, 2011; Pawlak *et al.* 2012) and, also, to dense arrays, with minor modifications (e.g. Lin *et al.* 2013; Li *et al.* 2016; Wang *et al.* 2017). However, this workflow was developed to enhance the low-frequency ocean microseisms and may not be optimal in high-frequency cases.

Temporal normalization, for example, is used to suppress ‘intermittent large-amplitude perturbations’ and preserve the phase information of (optimally, stationary Gaussian-distributed) random noise (Groos *et al.* 2012; Seats *et al.* 2012; Hanasoge & Branicki 2013). This works well in crustal-scale noise studies when large earthquakes are the perturbations and ocean microseisms are the random noise. However, the high-frequency noise field can greatly deviate from the stationary Gaussian noise (e.g. Groos & Ritter 2009; Lavoué *et al.* 2021). Furthermore, it is inconclusive yet whether

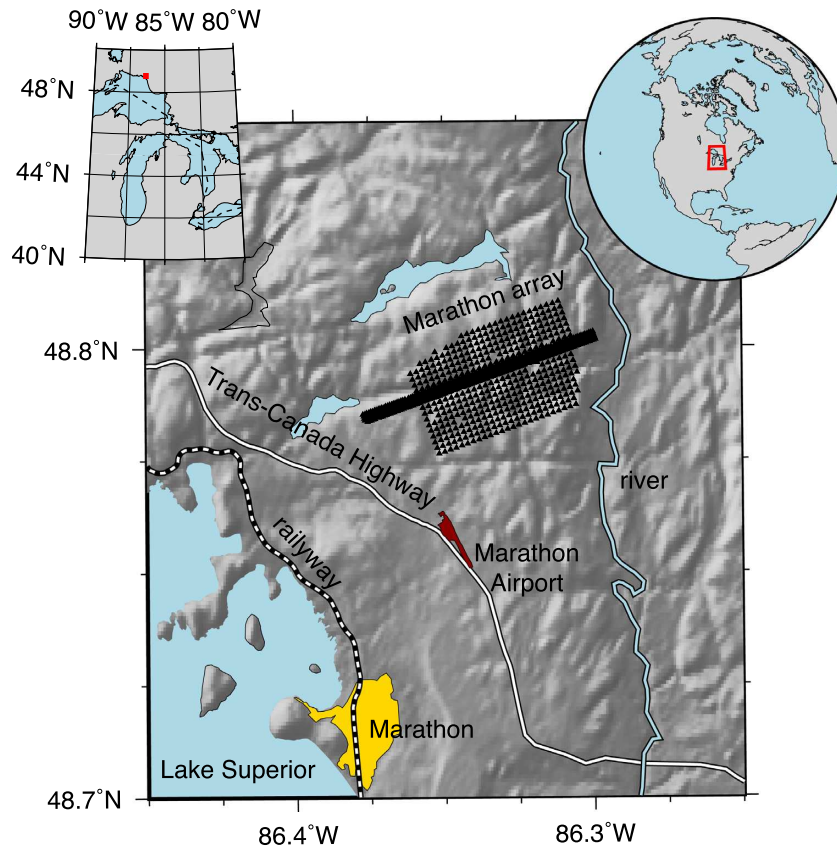
the high-frequency noise also features the pattern of short-duration large-amplitude perturbations and long-duration low-amplitude random noise, and whether the random noise contributes most to the NCFs. Stacking of correlation of short segments was proposed to be an alternative to temporal normalization (Prieto *et al.* 2011; Seats *et al.* 2012), which has been proved useful in studies at various scales (Saygin & Kennett 2010; Calkins *et al.* 2011; Lin *et al.* 2013; Ekström 2014; Chmiel *et al.* 2019). The short segment stacking approach not only can remove the effect of large earthquakes, but seems applicable to a wider variety of noise sources.

Another advantage of replacing temporal normalization with short segment stacking is that we can drop the bandpass filtering, so as to maximize the potential frequency range of recovered surface waves. Filtering has two main roles in the processing, namely, enhancing frequency bands of interest and reducing the distortion caused by temporal normalization (Pedersen *et al.* 2007; Groos *et al.* 2012). If we remove the nonlinear temporal normalization and keep other processing steps linear, bandpass filtering before cross-correlation is no longer necessary.

Spectral whitening can also be dropped if cross-coherence is used. Cross-coherence is one of the four interferometry methods (the other three are cross-correlation, deconvolution and convolution; Roux & Fink 2003; Slob *et al.* 2007; Wapenaar *et al.* 2010; Nakata *et al.* 2011; Chmiel *et al.* 2018). It is equivalent to cross-correlation after spectral whitening (Wapenaar *et al.* 2010), except that spectral whitening often uses smoothed amplitude spectra in seismic studies (Prieto *et al.* 2009, 2011; Seats *et al.* 2012; Chang *et al.* 2016), whereas cross-coherence uses the original amplitude spectra directly. Because cross-coherence has inherent spectral whitening, no explicit whitening is needed.

The convergence rate of the NCFs is another issue of practical importance because it determines the amount of data required to obtain reliable NCFs. It is relevant for determining the period of deployment and the temporal resolution of monitoring. Previous studies showed that fewer data are, generally, needed for smaller arrays and higher frequencies. For instance, three months to 1 yr of data is usually sufficient for region-scale networks (e.g. Yao *et al.* 2006; Moschetti *et al.* 2007; Yang *et al.* 2007; Bensen *et al.* 2008) and 1 month is commonly used for large-N arrays (e.g. Lin *et al.* 2013). When strong noise sources occur, the sufficient recording length can be as small as 6 hr (Mordret *et al.* 2013) or 2 hr (Ridder & Dellinger 2011). Even faster convergence can be found in engineering seismology studies, which showed that 1 hr of data is enough for dense arrays with a spacing of meters (e.g. Cheng *et al.* 2016). A recent quantitative study showed that NCFs can converge faster ( $<1$  month) than normally expected (1–2 yr) if using short segment stacking (Seats *et al.* 2012).

Phase- or group-velocity measurement is a classical observational approach in seismology (Sato 1955; Dziewonski *et al.* 1969; Levshin *et al.* 1992; Ritzwoller & Levshin 1998), but its methodology continues to develop (Meier *et al.* 2004; Kästle *et al.* 2016; Soomro *et al.* 2016; Bonadio *et al.* 2018, 2021). Application to high-frequency data from large, dense arrays requires particular workflows and a high level of automation. A widely used automatic measurement method is AFTAN (Levshin *et al.* 1992; Levshin & Ritzwoller 2001; Bensen *et al.* 2007), which was designed to measure group velocity. The phase-velocity measurement is generated by correcting the group velocity using the phase measured at the group traveltimes (Bensen *et al.* 2007). Alternatively, the automatic phase unwrapping method (Meier *et al.* 2004; Soomro *et al.* 2016; Bonadio 2019; Bonadio *et al.* 2021; El-Sharkawy *et al.* 2020) directly yields phase-velocity measurements. Furthermore, it combines multiple



**Figure 1.** The Marathon array and potential local noise sources. The red rectangle on the upper-right map shows the location of the upper-left map area. The small rectangle on the north shore of Lake Superior (upper-left map) shows the location of the main map. The Marathon array sensors are denoted by black dots. The nearest residential area is the town of Marathon (yellow). The dashed line crossing the town indicates the railway, the white line the trans-Canada highway and the dark red area the Marathon airport, all potential high-frequency noise sources. Lake Superior (the lower-left corner), along with several smaller lakes, generates noise at relatively low frequencies. The Pic River to the east of the array is a potential noise source as well.

ridges of frequency–time representation of the surface wave signal, which, in principle, allows broader band measurements.

The purpose of this study is to develop an optimal workflow for obtaining robust, wide-band, high-frequency dispersion measurements from the ambient noise recorded by dense arrays. We discuss in detail the key components of the processing and illustrate the effects of alternative implementations—and the advantages of the ones we identify as optimal—with real-data tests, aiming to provide a useful reference for future studies. We then illustrate the consecutive steps of the workflow using an application of the methods to the data from a new, large-N array deployed over a Cu-PGE mineral deposit in Ontario, Canada.

## 2 MARATHON ARRAY AND GEOLOGICAL BACKGROUND

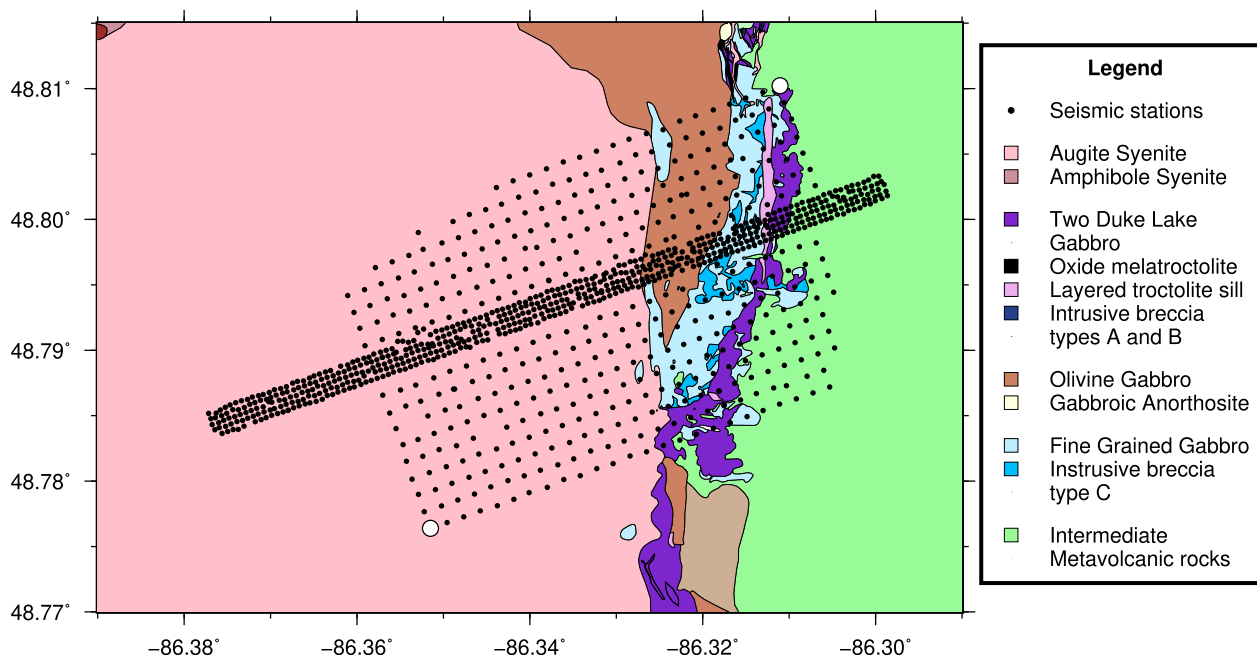
### 2.1 Marathon array

The Marathon array (Fig. 1) was deployed by the PACIFIC (Passive seismic techniques for environmentally friendly and cost efficient mineral exploration) international project (Dales *et al.* 2020, <https://www.pacific-h2020.eu/>, last accessed 2 January 2021) over a Cu-PGE (copper–platinum group elements) deposit northeast of the town of Marathon, Ontario, Canada. The primary purpose of the array deployment was to test the feasibility of imaging the deposit

and deposit-related structures with ambient noise methods and to develop the methods further.

The Marathon array comprised 1024 nodes, each equipped with a Zland GEN-2 1-C ARU sensor. The single-component sensor records vertical ground motions at a rate of 250 samples per second. The high-pass filter was switched off, which extended the usable frequency range down to  $\sim 0.125$  Hz. The array included a rectangular array of  $4 \text{ km} \times 2.5 \text{ km}$  total dimensions, with 414 sensors and 150-m spacing, and an overlapping ‘fat line’ array, with 610 sensors deployed in five parallel, closely spaced (50 m) lines across the middle of the rectangle. The spacing in each line was also 50 m. The two arrays were elongated northeast–southwest, towards a powerful potential noise source, Lake Superior. The array recorded continuously for 1 month, from 21 September 2018 to 22 October 2018, sufficient for retrieving surface wave signal according to previous studies (e.g. Lin *et al.* 2013; Chmiel *et al.* 2019).

A number of local noise sources around the Marathon array could be expected to provide signal for ambient noise imaging. Potential high-frequency noise sources include trains running along the shoreline of Lake Superior, the trans-Canada highway and the Marathon airport, located southwest of the array. These sources can provide noise at frequencies as high as 30 Hz (Dales *et al.* 2020; Lavoué *et al.* 2021), which can be used to image the shallow structure. The potential low-frequency noise sources are Lake Superior, Marathon town, the Pic River and several small lakes, producing noise mainly



**Figure 2.** Geological background. The region is covered by, from west to east, syenite (pink), gabbro intrusions (brown, blue and purple) and the Archean bedrocks (green). The black dots denote the Marathon array. The two white dots at the SW and NE extremities of the array indicate the station pair 1011.1030–1026.1111 ( $\Delta = 4.79$  km) used in the following experiments.

around 1 Hz (Xu *et al.* 2017), useful for imaging structure at greater depths.

## 2.2 Geological background

Situated to the northeast of Lake Superior, the Marathon mineral deposit can be thought of as a result of the North America Mid-Continent Rift System (MCR), which formed 1.1 Ga but eventually failed (Stein *et al.* 2018; Hinze & Chandler 2020). The failed rift left two long strips of mafic/ultramafic intrusion zones, extending southwest and southeast from Lake Superior. Whereas the southern part of MCR is buried under sedimentary layers, massive outcrops are exposed around Lake Superior. They include the Coldwell Alkaline Complex, the largest alkaline complex in North America and the site of the Marathon deposit.

The Coldwell Complex was formed during a cauldron subsidence, when magma intruded the Archean rocks subhorizontally, leaving a circular outcrop (Walker *et al.* 1993). On the northern and eastern margin of the complex lies the Eastern Gabbro Suite, which comprises multistage intrusions that were categorized into three series by Good *et al.* (2015), from oldest to youngest, the Fine-Grained Series, Layered Series and Marathon Series. The Marathon Series includes a gabbro intrusion named The Two Duck Lake Gabbro, which hosts the Marathon Cu-PGE deposit.

The area occupied by the Marathon temporary seismic array can be divided into three parts, according to the basement rock type: syenite, the Easter Gabbro Suite and Archean metavolcanic rocks (Fig. 2). Lab experiments on drill core samples show that gabbro has a higher velocity than the other two rock types (Gunawardana 2017). Thus, even though the velocity discontinuities within the gabbro suite, that is those between different stages of gabbro intrusions and between the deposit and its host, Two Duck Lake Gabbro, may be too subtle to detect, we can still image the boundary between the gabbro suite and the syenite or the Archean metavolcanic rocks (Hollis *et al.* 2019). The geometry of the latter boundary affects

the accumulation of the deposit-related sulphides, which is useful information for further mineral exploration (Good *et al.* 2015).

## 3 NOISE CROSS-CORRELATION

Noise cross-correlation processing shows a great diversity in details of implementation, despite following a common sequence: temporal normalization, spectral whitening, correlation and stacking. For example, at least 4 methods are commonly used for temporal normalization, that is one-bit normalization (Campillo & Paul 2003; Shapiro & Campillo 2004), running absolute mean normalization (Bensen *et al.* 2007), clipping (Sabra 2005) and short-time window stacking (Prieto *et al.* 2011). The diversity allows customizing the workflow for specific noise sources and array configurations.

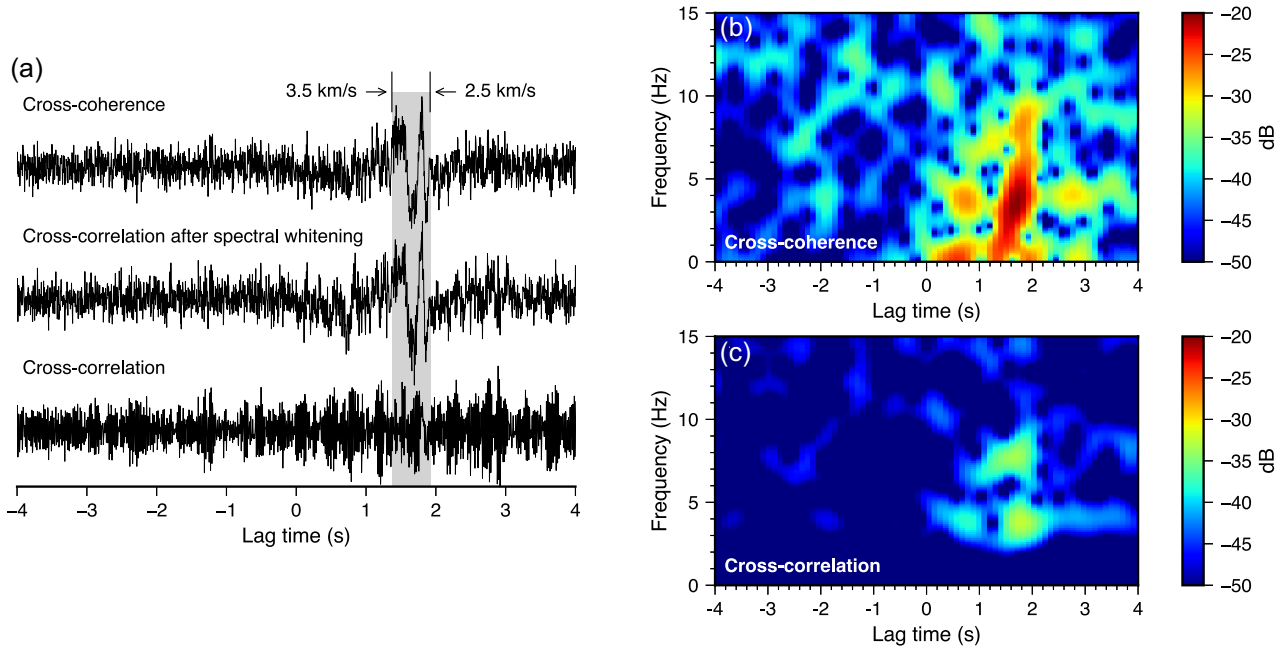
We optimized the workflow for high-frequency noise sources and dense arrays, which means the workflow yields the broadest bandwidth of accurate phase-velocity measurements, given the amount of data available. It was validated by experiments on the Marathon data set. We strived to reduce empirical parameters and provide guidelines on tuning the parameters when they are impossible to remove.

### 3.1 Basic pre-processing

Basic pre-processing comprises all the processing steps before the temporal and spectral normalization, which includes decimating the time-series, removing instrumental responses, splitting continuous seismograms into segments, demeaning, detrending, tapering and bandpass filtering. We categorized them into three groups: necessary, optional and unnecessary, according to their significance for obtaining reliable, wide-band dispersion measurements.

The ‘necessary’ group comprises splitting into segments, demeaning, detrending and tapering. Demeaning and detrending can





**Figure 3.** Comparison of correlation methods. (a) NCFs obtained by different correlation methods (cross-coherence, cross-correlation after spectral whitening and cross-correlation). Same pre-processing were used and no temporal normalization was applied for all the three methods. Seismograms from the station pair 1011.1030–1026.1111 in the first 10 min after 2018–10–01T04:00:00 were used in this example. Correlations were computed using 50 per cent-overlapped 1-min time windows. The surface wave window (2.5–3.5 km s<sup>-1</sup>) is shaded grey. (b) Spectrogram of the NCF computed by cross-coherence. The spectrogram is computed using a 1-s window with 90 per cent overlap. The amplitude of the spectrogram is converted to decibel using  $10 \log_{10}$ . Cross-correlation after spectral whitening yields a very similar spectrogram and thus not shown here. (c) Spectrogram of the NCF computed by cross-correlation.

remove long-term variations due to the instrumental drift, and tapering reduces the numerical oscillations known as the Gibbs effect when filtering and performing Fourier transforms. Splitting into segments has several advantages, one of which is efficiency, because the fast Fourier transform (FFT) is faster for multiple short time-series than for a long time-series. Other advantages—which, eventually, make splitting into segments one of the most critical pre-processing steps in the workflow—will be discussed below (see Section 3.3).

The ‘optional’ group contains instrumental response removal, which is needed only when different models of seismometers are involved or when aiming at frequencies lower than the corner frequency of the seismometer. The ‘unnecessary’ group comprises decimation and bandpass filtering, both can result in a loss of useful frequency content. Decimation should be used only when the frequency range of interest is pre-determined. Bandpass filtering, as discussed in Section 3.3, is unnecessary if temporal normalization is removed.

### 3.2 Cross-coherence

Cross-coherence follows the basic pre-processing in our workflow, with no explicit temporal normalization and spectral whitening. Assuming that  $A(\omega)$  and  $B(\omega)$  are the spectra of two seismogram segments at stations A and B, respectively, the cross-coherence can be written as (Wapenaar *et al.* 2010; Nakata *et al.* 2011).

$$H_{AB}(\omega) = \left\langle \frac{A(\omega) B^*(\omega)}{|A(\omega)| |B(\omega)|} \right\rangle, \quad (1)$$

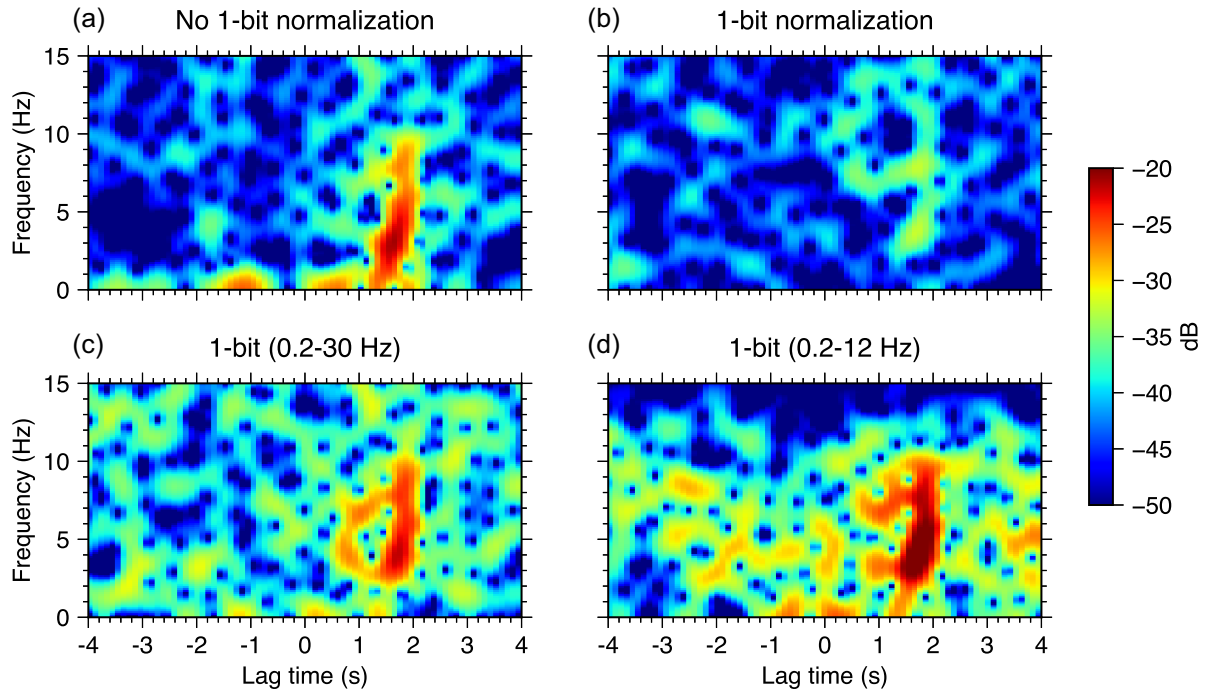
where  $|A(\omega)|$  is the amplitude spectrum,  $\langle \cdot \rangle$  represents averaging over segments and  $*$  represents a complex conjugate.

The intrinsic spectral normalization means that cross-coherence has a similar effect to cross-correlation after spectral whitening. We can see the similarity from the mathematical representation of the latter (Prieto *et al.* 2011; Seats *et al.* 2012; Chang *et al.* 2016),

$$C_{AB}(\omega) = \left\langle \frac{A(\omega) B^*(\omega)}{|\widehat{A(\omega)}| |\widehat{B(\omega)}|} \right\rangle, \quad (2)$$

where  $|\widehat{A(\omega)}|$  is the smoothed amplitude spectrum. The difference between (1) and (2) is whether the spectrum in the denominator is smoothed or not. To confirm the similarity, we compared the NCFs computed by cross-coherence, cross-correlation after spectral whitening and cross-correlation (Fig. 3a). A frequency window of 1 Hz is used in the spectral whitening. The former two methods yield nearly identical waveforms, especially in the surface wave window.

Another finding is that spectral normalization is critical for retrieving wide-band surface-wave signal in high-frequency cases. The surface waves obtained with spectral normalization (either spectral whitening or cross-coherence) is clearer than one obtained without normalization (Fig. 3a). Spectrogram analysis confirms the observation, which shows cross-coherence yields a continuous surface wave energy from ~0 to 10 Hz (‘~0’ here means that the surface waves have energy down to the low-frequency resolution limit of the analysis), whereas cross-correlation only yields two separate energy maxima at 4 and 8 Hz (Figs 3b and c). Normalizing the spectrum brings the amplitude of different frequencies to the same level, which enhances the frequencies with weak energy and thus removes potential spectral holes (eqs 1 and 2).



**Figure 4.** Spectrogram of NCFs with and without one-bit normalization. (a) The NCF is computed without one-bit normalization, (b) with one-bit normalization but no bandpass filtering. Panel c and d are filtered in 0.2–30 Hz and 0.2–12 Hz frequency bands, respectively, before one-bit normalization. All NCFs are computed using the same 10 min of seismogram from the station pair 1011.1030–1026.1111 as the one in Fig. 3. All NCFs are computed using 1-min-long segments.

### 3.3 Replacing temporal normalization with short segment stacking

Temporal normalization, which balances the energy from different noise sources, can be achieved in two approaches. First, classical temporal normalization methods, such as one-bit normalization (Campillo & Paul 2003) and its generalization, running absolute mean normalization (Bensen *et al.* 2007), adjust the amplitude within each segment. Secondly, stacking with a peak normalization adjusts the relative amplitude between segments, which also achieves the effect of temporal normalization. Studies showed that using short segments makes the second approach an alternative method to the classical temporal normalizations in suppressing large earthquakes (Prieto *et al.* 2011).

There have been concerns about the non-linearity of the classical temporal normalization methods (Pedersen *et al.* 2007; Seats *et al.* 2012), despite their successful applications. Theoretical studies suggested that one-bit normalization preserves phase only when seismic noise is dominated by Gaussian noise (Hanasoge & Brannicki 2013). In practice, when seismic noise strongly deviates from a Gaussian distribution, applying one-bit normalization may even give worse results (Comparing Figs 4b with a). Bandpass filters were recommended before one-bit normalization (Pedersen *et al.* 2007), because seismic noise in a narrower frequency band is believed to be more likely to follow a Gaussian distribution. However, the pre-filters can limit the frequency band that cross-correlation can recover.

Striving for wide-band surface wave dispersion measurements, we favour short segment stacking over the classic temporal normalization. Short segment stacking requires no pre-filters, thus with no risk of losing frequencies. Applying the two approaches to the 10-min-long data set in Fig. 3, we found that short segment stacking yield wider-band surface waves than one-bit normalization,

regardless of whether bandpass filtering was applied or not (Fig. 4). The best result for one-bit normalization (Fig. 4d) was obtained using a filter of 0.2–12 Hz. Despite being comparable to the short-segment-stacking result (Fig. 4a), finding the optimal filter for the one-bit normalization requires extra effort. Using a non-optimal filter can result in stronger noise and the loss of certain frequencies (e.g. <2 Hz in Fig. 4c).

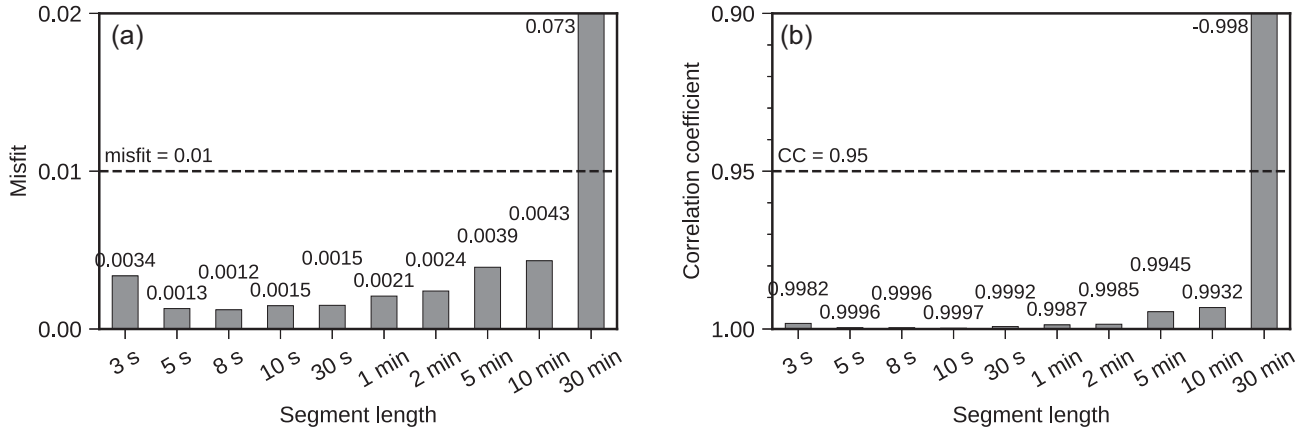
### 3.4 Optimal length of segments

Although the usage of short segments was recommended by various studies (e.g. Prieto *et al.* 2011; Seats *et al.* 2012), the optimal length was rarely discussed. The segment length used in the previous studies varies more than one order of magnitude in the crustal-scale studies (1–24 hr, e.g. Bensen *et al.* 2007; Lin *et al.* 2007; Villaseñor *et al.* 2007; Prieto & Beroza 2008; Zheng *et al.* 2008; Ekström *et al.* 2009; Li *et al.* 2009) and more than two orders of magnitude in high-frequency studies (1–120 min, e.g. Lin *et al.* 2013; Mordret *et al.* 2013; Chang *et al.* 2016; Wu *et al.* 2017). Although it is generally agreed that smaller arrays should use shorter segments, no clear rule exists to guide the selection of the optimal length.

The lower limit of the optimal length should allow a surface wave train to travel between station pairs (Groos *et al.* 2012),

$$T_{\text{optimal}} \geq 2(\Delta/c_{\min} + T_{\text{duration}}), \quad (3)$$

where  $\Delta$  is the interstation distance between the two stations,  $c_{\min}$  the minimum phase velocity and  $T_{\text{duration}}$  is the extra time that allows building up robust spectral estimation. For short-duration transient noise sources,  $T_{\text{duration}}$  can be the duration of the dominant noise source. The factor of two is added to include surface waves at both the causal and acausal branches. Using shorter segment will wrap the surface waves to the opposite branch, causing spurious arrivals.



**Figure 5.** Comparison of stacked NCFs using different segment lengths. (a) The quality of NCFs is evaluated by the relative misfit between dispersion curves of the NCFs and a reference dispersion curve in 3–15 Hz. The reference dispersion curve is from the month-long stacked NCF computed using 1-min segments. Segment lengths are 3 s, 5 s, 8 s, 10 s, 30 s, 1 min, 2 min, 5 min, 10 min and 30 min. The data of the first half-day of 2018/10/01 from the station pair 1011.1030–1026.1111 is used to compute the stacked NCF. (b) Same as (a), but uses the correlation coefficient between dispersion curves and the reference dispersion curve to evaluate the quality. The CC of 30-min segment is negative because it shows reverse dispersion.

For the upper limit, we determined it through experiments on the real data. We compared NCFs computed with different segment lengths in terms of quality, defined by (1) the correlation coefficients between the observed dispersion curves and a reference dispersion curve and (2) the root mean square of their relative difference (Fig. 5). The best NCF is obtained with the segment length of 8 s, according to the misfit, or 10 s, according to the correlation coefficient. The optimal segment length is close to the lower limit of optimal segment length (7.548 s) predicted by eq. (3), with  $\Delta = 4.79$  km,  $c_{\min} = 2.7$  km s<sup>-1</sup> and a lower frequency limit of 0.5 Hz. Therefore, we conclude that the optimal segment length is close to the lower limit given by eq. (3),

$$T_{\text{optimal}} \approx 2(\Delta/c_{\min} + T_{\text{duration}}). \quad (4)$$

For stationary sources, for example ocean microseisms,  $T_{\text{duration}}$  can be much shorter (e.g. 30 min) than the complete duration of a typical microseism (days or weeks). A possible explanation is that 30 min of seismic recording is sufficient to estimate the spectrum of microseisms. Hence, for stationary sources, we define  $T_{\text{duration}}$  in terms of multiples of their spectral period. The experiment in Seats *et al.* (2012) suggests, for ocean microseisms,  $T_{\text{duration}}$  should be at least 70 times of the spectral period, if we consider the longest period as 25 s (e.g. Möllhoff & Bean 2016; Le Pape *et al.* 2021). Nevertheless, our experiment also suggests that suboptimal lengths (e.g. 10 min, 60 times as large as the optimal length of 10 s) can yield reasonably good results as well.

### 3.5 Convergence of NCFs

Although stacking all the available data generates the best NCF in most cases, investigating the minimum amount of data required for convergence to a high-quality NCF can help minimize the unnecessary computational cost and also paves the way to time-lapse tomography. Studies suggested that 1 month was sufficient for crustal-scale studies, meaning a 90 per cent reduction in cost (Seats *et al.* 2012), compared with 1–2 yr that are usually used.

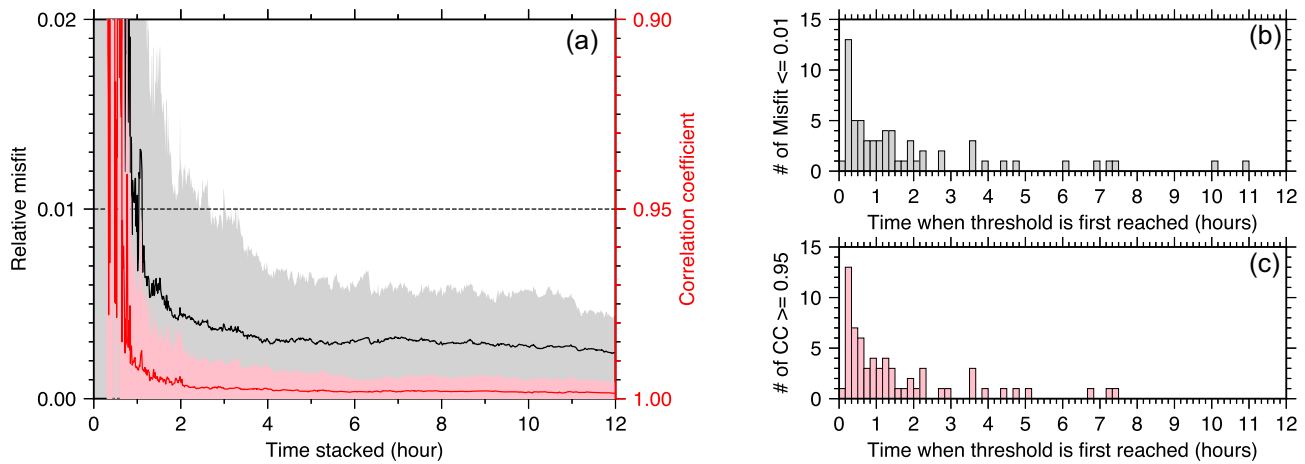
We analysed the convergence of NCFs by investigating the temporal variations of the quality of the cumulative-stack NCFs. Convergence is defined as the relative misfit of dispersion curves less than 0.01 or CC greater than 0.95. Using the same station pair as

in the optimal segment length experiment, we computed cumulatively stacked NCFs for 12 hr of continuous data, whose relative misfits or the correlation coefficients form a convergence curve. We repeated the same experiment for 66 half-days available in the Marathon data set. The median curve of the 66 convergence curves was used to estimate the average convergence rate and three times of the median absolute deviation (MAD)—the confidence interval. The result shows that the NCFs, on average, converge in the first 2 hr and become stable after 4 hr. Assuming a normal distribution, about 95 per cent of the 66 half-days converge within 3 hr. Our experiment shows that the data required to get a reliable NCF for the frequency range of 1–20 Hz and interstation distances of around 5 km is normally less than 3 hr and up to 11 hr in a few exceptional cases.

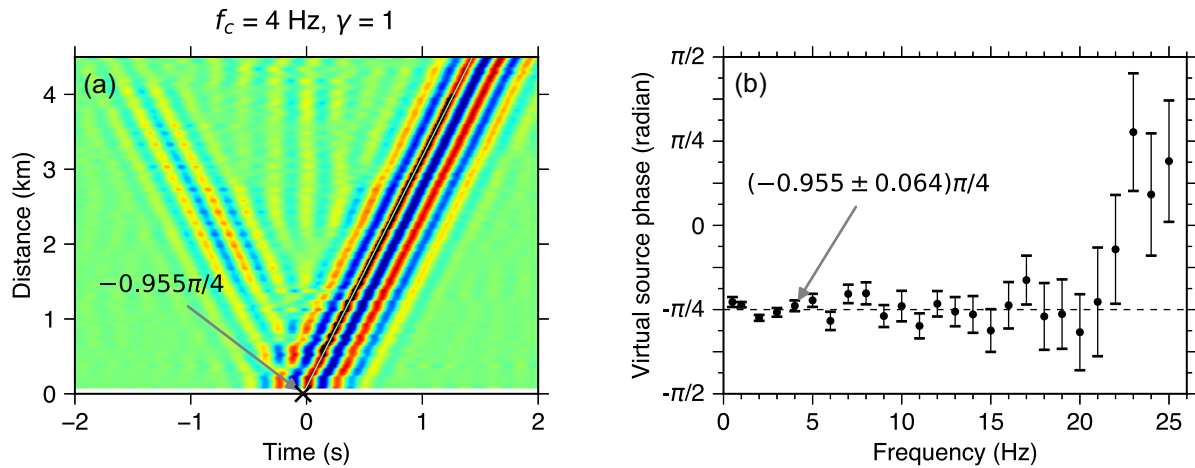
The substantial differences among the convergence curves are also an interesting observation: they indicate strong temporal variations of the noise sources. The wide confidence interval of the average convergence curve (shaded area in Fig. 6a) suggests that some NCFs converge after 3 hr and some NCFs converge almost immediately. Defining the convergence time as the first time when reaching the thresholds, we found that the fastest convergence occurs within 10 min and the slowest convergence takes as long as 11 hr (Figs 6b and c). The large difference in convergence can be explained by localized strong noise sources, which was also observed in previous noise studies in Valhall, where a storm enabled obtaining reliable NCF from only 6 hr of data (Mordret *et al.* 2013).

To further validate the reliability of the NCFs, we estimated the source phase that is expected to be equal to  $-\pi/4$  ( $\pi/4$  if using empirical Green's functions, EGFs) using a method suggested by Martins *et al.* (2019). The method first fits the phase traveltime picks as a linear function of interstation distance, then uses the time-intercept, after multiplying by  $\omega$ , as the source phase estimation. The fitting is done for each frequency individually. The closer to  $-\pi/4$  the source phase is, the less biased the phase velocity measurements are considered to be. In this experiment, we used the binning stack of 1-hr stacked NCFs, with the bin width of 50 m (below 10 Hz) and 2 m (above 10 Hz). The binning stacked NCFs are narrowband filtered, and then one ridge is selected and fit to get the source phase estimation (Fig. 7a). The source phases in 0.5–25 Hz are shown in Fig. 7(b) with their standard deviation. The phases are close to the expected  $-\pi/4$  in 0.5–21 Hz, supporting the reliability of NCFs at





**Figure 6.** Convergence of NCFs. (a) Quality of NCFs in 66 half-days from the station pair 1011.1030–1026.1111 as a function of stacked time. The quality of NCFs is evaluated by relative misfit and correlation coefficient of dispersion curves with respect to a month-long stacked NCF, thus forming 66 convergence curves. The median of convergence curves are denoted by black (relative misfit) and red lines (correlation coefficient). The shaded region is the error region defined by three times of the MAD. The dashed line marks the thresholds of high-quality dispersion curves (0.01 for misfit and 0.95 for CC). (b) Histogram of the first time (smallest stacking time) when the relative misfit reached 0.01. (c) Histogram of the first time (smallest stacking time) when the convergence curves of CC reached 0.95.



**Figure 7.** Source phase estimation. (a) An example of the estimation at 4 Hz. The binning stacked NCFs are filtered using a Gaussian filter centred at 4 Hz with  $\alpha = 2\pi f_c \gamma^2$ . The black dots denote the traveltime picks, and the green line across them is a linearfit of the time picks. The intercept at the time axis corresponds to  $-0.955\pi/4$  in phase, close to the expected value of  $-\pi/4$ . (b) Source phase estimated for 0.5–25 Hz. Black dots denote the source phases, and the error bars indicate their standard deviation.

that frequency range. As for the deviation from  $-\pi/4$  above 21 Hz, we believe it is mainly because of the irregular coverage of short distance, which greatly reduces the robustness of the fitting at high frequencies (Fig. B1).

In practice, a small pilot array can be useful for the appraisal of the available noise sources and determination of the optimal deployment length. Although our experiments clearly show that a few hours are sufficient to obtain NCFs from the Marathon data set, the applicability of this result to other arrays is not warranted automatically, because the convergence of NCFs is strongly affected by the temporal characteristics of local noise sources, which vary for different regions. Furthermore, the target frequency ranges can also affect the length of the deployment, as low-frequency NCFs may require a longer time stacking to achieve a certain signal-to-noise ratio (SNR).

#### 4 PHASE-VELOCITY MEASUREMENT

The phase velocity of surface waves can be measured either in the time domain, by selecting phase traveltime or correcting from the group phase traveltime (e.g. Yao *et al.* 2006; Bensen *et al.* 2007), or in the frequency domain, by unwrapping the phase of the cross-correlation function (Meier *et al.* 2004; Soomro *et al.* 2016; Bonadio 2019) or by using Aki's spectral formulation (Aki 1957; Ekström *et al.* 2009). Time-domain methods are popular, in large part, because of their intuitive simplicity.

We present a new time-domain method which is empowered by the idea of exploiting the resolved  $2\pi$  ambiguity of phase velocity measurement (Meier *et al.* 2004). Unlike the usual methods that select one solution from the multiple solutions given by the ambiguity as the correct phase velocity, the idea suggests that (1) every solution can be converted to a valid phase velocity measurement if the

phase offset is corrected, and (2) combining the multiple solutions can produce a wider-band measurement than selecting one. Applications to both teleseismic surface wave (Kästle *et al.* 2016; Soomro *et al.* 2016; Bonadio *et al.* 2018; El-Sharkawy *et al.* 2020) and ambient noise cross-correlation studies (Bonadio 2019) yield wide-band (>5 octaves) measurements. Our method can be basically seen as a time-domain alternative to the original frequency-domain implementation (Soomro *et al.* 2016). But we also added a modification named amplitude-guided ridge tracking, which allows the recovery of more high-frequency measurements. The method is highly automated and thus allows efficient processing of massive amounts of NCFs.

## 4.1 Automatic measurement

### 4.1.1 Multiple phase velocity solutions

Our method starts with converting the surface wave signal in a NCF to its frequency-time representation using a comb of narrow bandpass filters (Dziewonski *et al.* 1969),

$$F(\omega, \omega_c) = \exp\left(-\alpha\left(\frac{\omega}{\omega_c} - 1\right)^2\right), \quad (5)$$

where  $\omega$  is the angular frequency,  $\omega_c$  the centre frequency of the filter and  $\alpha$  an empirical parameter to control the bandwidth of the filter, balancing the spectral and temporal resolution. We used a frequency-dependent  $\alpha$  ( $= \gamma^2 \omega_c$ ), following Meier *et al.* (2004) and Soomro *et al.* (2016), where  $\gamma$  is an empirical parameter that replaces  $\alpha$  to control the filter's bandwidth.

In the time domain, the narrow-band filtered surface wave train resembles a cosine function (with certain phase shift), with one of the peaks correspond to the phase traveltime  $t_{sw}$  (Yao *et al.* 2006). Using the asymptotic far-field approximation of either the fundamental-mode surface wave or the NCF (Dahlen & Tromp 1998; Tsai 2009; Boschi *et al.* 2013; Kästle *et al.* 2016), the phase traveltime can be converted to the phase velocity by

$$c = \frac{\Delta}{t_{sw} + 1/(8f)}. \quad (6)$$

where  $c$  is the phase velocity,  $\Delta$  the interstation distance and  $f$  the centre frequency of the filter. If the time is measured in EGFs that are obtained from derivatives of the NCFs (Roux *et al.* 2005b), the sign before  $1/(8f)$  should be changed to minus.

However, it is not clear which peak corresponds to the right phase traveltime. Hence, the multiple peaks generate a group of possible phase velocities (separated by  $2\pi$  in phase), resulting in the so-called  $2\pi$  ambiguity of phase velocity measurement. A common approach is to select one phase velocity by comparing to a reference velocity. However, the arrival time of the peaks relates to the phase traveltime via a simple relation,

$$t_{\text{ridge}} = t_{sw} + n/f, n = 0, \pm 1, \pm 2, \dots, \quad (7)$$

where  $1/f$  corresponds to a  $2\pi$  cycle in phase. If the correct  $n$  is known for a peak, it can also generate the correct phase velocity. To differentiate the peaks, we define  $n$  as the order of the peaks. A set of peaks with the same order are referred to as a ridge. For example, the 0th ridge corresponds to the correct phase traveltime and the 1st ridge is  $1/f$ s later. Fig. 8(e) shows that the 0th and 1st ridges (white and cyan crosses in Fig. 8b, respectively) yield nearly identical dispersion curves at 3–10 Hz.

### 4.1.2 Amplitude-guided ridge tracking: a new approach to select ridges

Different ridges in the frequency-time plane yield somewhat different phase velocities (<3 Hz in Fig. 8e). We believe the reason is that the accuracy of the measurements depends on the amplitude of the ridges, because higher amplitude results in higher SNR. The ridges may be selected by large amplitudes in the time-frequency plane (Meier *et al.* 2004; Soomro *et al.* 2016; Bonadio *et al.* 2018, 2021; El-Sharkawy *et al.* 2020). However, we found that this approach results in larger errors in phase velocities than the ridge tracking method in the frequency-time plane when the surface wave signal becomes weaker (>10 Hz in Fig. 8f).

We developed a new method, named amplitude-guided ridge tracking, to select the ridges based on both amplitude and local continuity. The method is a modification of the ridge tracking method by allowing switching to adjacent ridges during the tracking process. The method is applied to the frequency-time representation of a NCF as follows.

- (1) Set a starting frequency  $f_i$ . For the starting frequency, select the ridge with the highest amplitude and record its arrival time  $t_i$ . Set the order of the selected ridge to 0.
- (2) For the next frequency, find the closest ridge  $t_c$  and its two adjacent ridges,  $t_l$  and  $t_r$ . From the three ridges, select the one with the highest amplitude and record its arrival time as  $t_{i+1}$ . Keep a record of the changes of the order.
- (3) Repeat step II until all the frequencies are processed.

The arrival time of the ridges is measured by fitting a parabola to 3 points around the peak. Using a starting frequency of 1 Hz, the 0th ridge is selected below 3.5 Hz and the 1st ridge is selected at 3.5–10 Hz, due to their high amplitude. An improvement can be seen above 10 Hz when the tracing stays within the surface-wave window instead of jumping to high-amplitude noise (e.g. at 2.8 s), extending the bandwidth of the accurate phase-velocity measurements beyond 10 Hz.

### 4.1.3 Combining ridges to form a dispersion curve

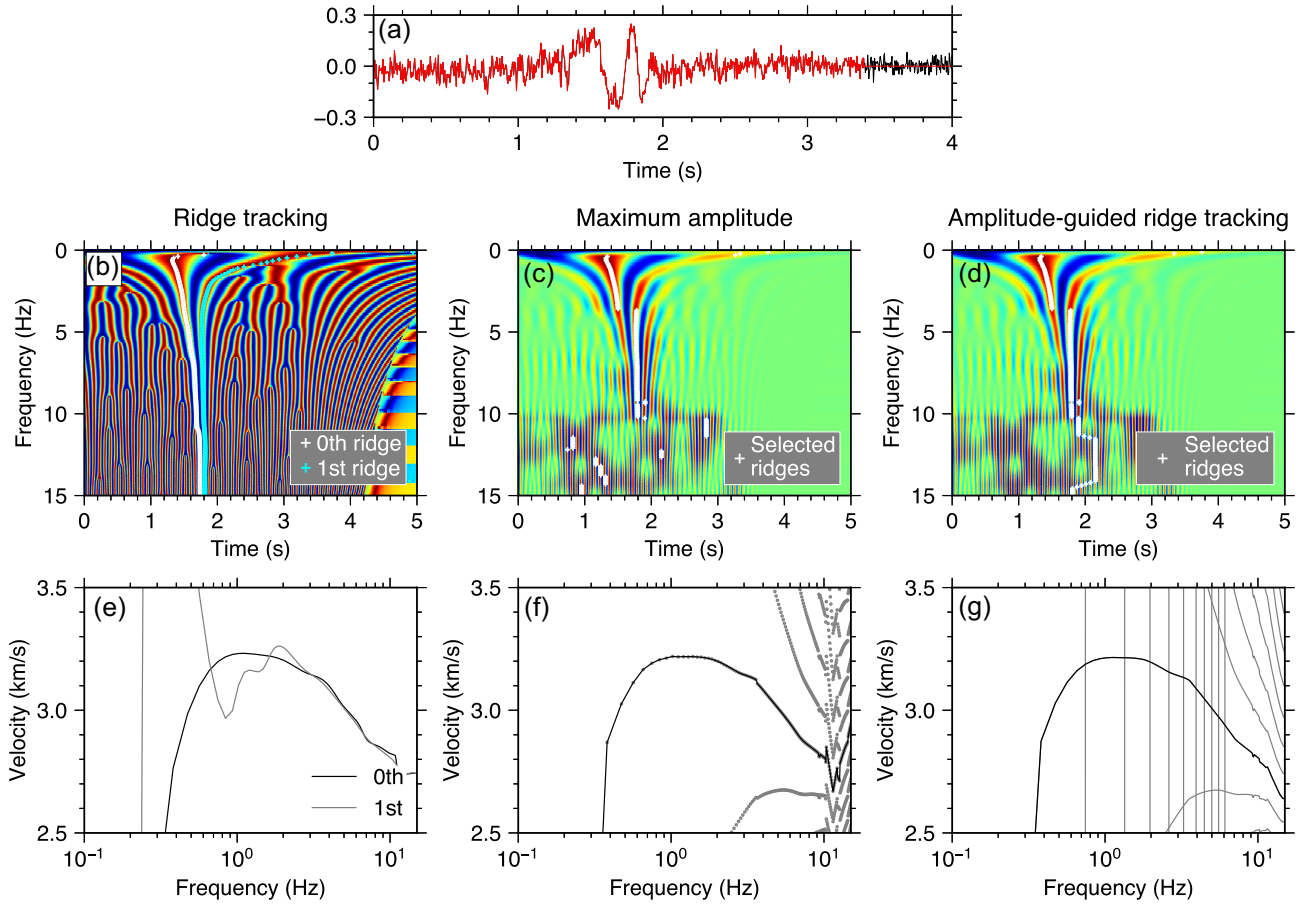
The arrival time of the selected ridges,  $t(f)$ , can be converted to a dispersion curve using the following formula,

$$c = \begin{cases} \frac{\Delta}{t(f) + 1/(8f) - n/f}, & \text{for NCF,} \\ \frac{\Delta}{t(f) - 1/(8f) - n/f}, & \text{for EGF.} \end{cases} \quad (8)$$

where  $n$  is the order of the ridges obtained during the tracking. The remaining problem is that the order of the starting frequency, despite being assumed to be 0, is actually unknown. Therefore, the  $2\pi$  ambiguity is not completely resolved yet (Fig. 8g).

## 4.2 Resolving the $2\pi$ ambiguity

The  $2\pi$  ambiguity of a phase-velocity dispersion curve is generally solved by comparing it to a reference dispersion curve at low frequencies and then tracking to high frequencies based on smoothness of dispersion curves (e.g. Soomro *et al.* 2016; Bonadio *et al.* 2018). However, for the local-scale imaging, finding an available reference dispersion curve can be challenging due to the lack of local-scale velocity models. A possible option is to use the ambiguity-free dispersion curves generated by array-based phase-velocity measurement methods, such as the  $f$ - $k$  transform, as a reference (e.g.



**Figure 8.** Comparison of phase-velocity measurement methods. (a) Symmetric NCF before (black) and after (red) windowing. A Tukey window of  $[\Delta/5-1, \Delta/2+1]$  s is used in this example with  $\Delta = 4.79$  km. (b) Normalized frequency-time representation of the windowed NCF. Each filtered NCF (each row of the image) is normalized by its envelope. The 0th (white crosses) and 1st (cyan crosses) ridges are traced by the original ridge tracking method. The corresponding dispersion curves are shown in (e), where the 1st ridge is corrected by  $-1/f$  in time before the conversion to phase velocity. (c) Frequency-time representation of the windowed NCF. White crosses denote the ridges selected for each frequency. The corresponding dispersion curve in (f), denoted by black dots, is computed by the phase unwrapping method. The grey dots denote other possible dispersion curves obtained by adding multiples of  $2\pi$  to the unwrapped phase. (d) Same as (c), but with ridges selected by the amplitude-guided ridge tracking method. The corresponding dispersion curves are shown in (g), where the direct output is in black and the ones generated by adding multiples of  $2\pi$  are in grey. The grey vertical lines are caused by velocity jumps from  $+\infty$  to  $-\infty$  km s $^{-1}$  when their corresponding phase traveltimes cross the zero time to negative.

Martins *et al.* 2019). Here, we present two alternative methods that are easier to implement. The first method is based on the density plot of all dispersion curves, which can generate a reference dispersion curve (Bonadio *et al.* 2018; Carvalho *et al.* 2019). The second approach is selecting the ridges with maximum amplitude at low frequencies and tracking the selected ridge to higher frequencies in the time-frequency plane. It is based on the fact that the amplitude of the ‘wrong’ ridges will drop quickly at low frequencies as they separate from the phase traveltimes by  $n/f$  (Fig. 8d), provided that the phase traveltimes are reasonably close to the group traveltimes that determines the arrival of the strongest surface wave energy.

#### 4.2.1 Forming a reference dispersion curve without a velocity model: the density plot approach

The density plot approach generates the reference dispersion curve by stacking all the dispersion curves together and identifying the high-density ridge as the reference dispersion curve (Bonadio *et al.* 2018). A similar idea was also used to determine a reasonable phase velocity range by Rawlinson *et al.* (2014). But the reason why the

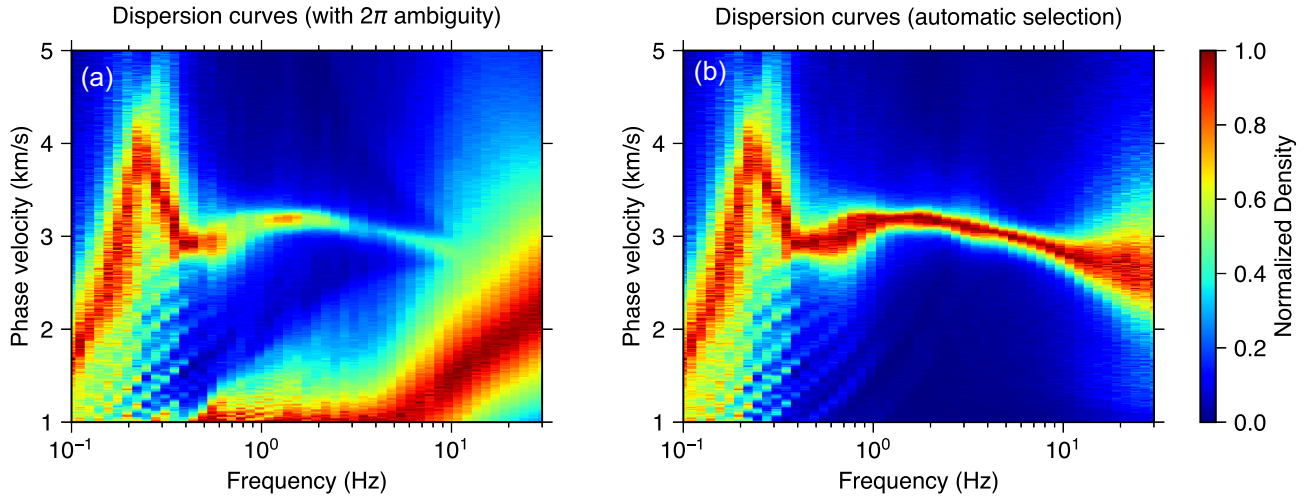
simple approach works has not been analysed formally, which we do here.

Assuming we have selected a ridge with the order of  $n_0$  and the traveltimes of  $t_0$ , the phase velocity  $c$  can be computed by

$$c = \begin{cases} c_0, & \text{if } n = n_0, \\ \frac{1}{\frac{1}{c_0} + \frac{n_0 - n}{f\Delta}}, & \text{otherwise,} \end{cases} \quad (9)$$

where  $c_0$  is the correct phase velocity and  $n$  is the order used to correct the  $2\pi$  cycles. Eq. (9) means that, when using a wrong order  $n$ , the resulting phase velocity depends on the interstation distance, whereas the one computed with the correct  $n$  ( $= n_0$ ) does not. When overlaying dispersion curves from station pairs with different interstation distance, only the correct dispersion curves will stack constructively. Therefore, the high-density area in the density plot should correspond to the correct dispersion curve.

Fig. 9(a) displays the density plot of the dispersion curves obtained in the Marathon site. Each column of the image is a normalized histogram of the phase-velocity measurements at a frequency. The high-density ridge around 3 km s $^{-1}$  can be used as the reference



**Figure 9.** (a) Density plot of all dispersion curves (the  $2\pi$  ambiguity not resolved). (b) Density plot of automatically selected curves (the  $2\pi$  ambiguity resolved). Each column in the density plot is a normalized histogram of the phase-velocity measurements at that frequency over the phase velocity. The bin width of the histogram is  $0.01 \text{ km s}^{-1}$ .

dispersion curve. The example demonstrates that the density plot approach, introduced previously at lower frequencies (e.g. Bonadio *et al.* 2018), is also applicable in high-frequency noise cross-correlation studies. When no *a priori* information of local geology is available, one can also use  $f$ - $k$  analysis in order to select the correct high-density ridge or obtain a reference dispersion curve (e.g. Martins *et al.* 2019).

#### 4.2.2 Automatic selection in the time–frequency plane

We found the amplitude-guided ridge tracking method can select the correct ridge without any reference dispersion curves, when sufficient low-frequency measurements are available. Since the ridges separate by  $1/f$  s in the frequency–time representation, the wrong ridges will be far away from the correct one when the frequency is sufficiently low. Considering the surface wave energy is localized around the correct ridge, the wrong ridges will have a low amplitude. Hence, if selecting a low starting frequency, the correct ridge will be automatically selected.

The example in Figs 8(d) and (g) shows that our method can reliably select the correct ridge using a starting frequency in the 0.5–3 Hz range. Fig. 9(b) shows the density plot of all the dispersion curves automatically selected (the  $2\pi$  ambiguity resolved), where the high-density ridge is consistent with but better resolved than the one in Fig. 9(a) (all potential measurements, with the  $2\pi$  ambiguity not resolved), validating the automatic selection of the curves, anchored at low frequencies.

#### 4.3 Correction to the measurement of the phase traveltime

The local maximum of a bandpass-filtered surface waveform does not necessarily correspond to the phase traveltime, due to the interference from neighbouring frequencies. The systematic shifts between them depend on centre frequencies and bandwidths of the filters, and interstation distances. They are generally small for narrow filters. Based on a 1-D velocity model obtained from the average dispersion curve (Fig. 9b), we computed the systematic shifts numerically and corrected our phase velocity measurements. Assuming a flat amplitude spectrum, we computed surface wave synthetics by summing harmonic waves over a wide frequency range

(0.1–30 Hz), for given interstation distance. The broad-band synthetics was filtered using the same comb of Gaussian filters in the actual measurements, and then we measured the difference between the measured phase traveltime and the theoretical phase traveltime,  $t = \Delta/c(\omega_c) - \pi/4\omega_c$ . We applied this time shift to the phase traveltimes measured in Section 4.1.2 and recomputed the phase velocity dispersion curves (Fig. 10). We can see that the dispersion curves after correction have no significant changes at high frequencies but are elevated by more than  $0.1 \text{ km s}^{-1}$  at low frequencies ( $<1 \text{ Hz}$ ). Detailed discussions of the correction are beyond the scope of this paper and will be the focus of a following companion paper.

## 5 QUALITY CONTROL

### 5.1 Control criteria

We assessed the quality of the dispersion curves and rejected the measurements deemed unacceptable via a series of quality checks, using consistent criteria with pre-defined thresholds. To limit potential biases, we only used three criteria, namely, the slope, probability and median absolute deviation (MAD), with relatively broad thresholds. The criteria are mostly based on the mutual consistency of a cluster of dispersion curves, which allows the effective thresholds to vary for different clusters, accommodating the strong heterogeneity of the shallow structure. The three criteria are defined as follows.

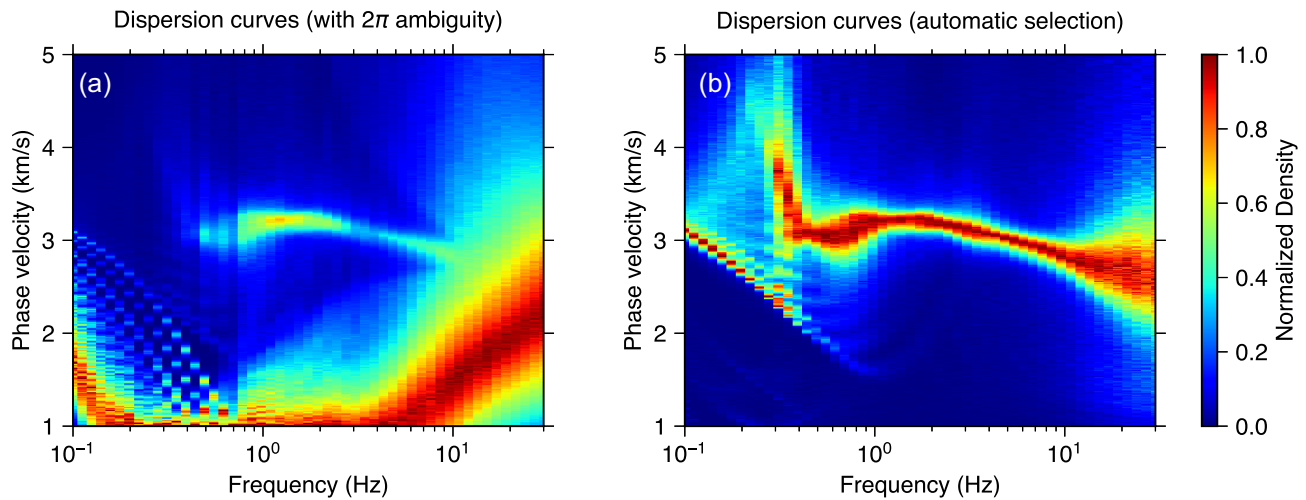
(1) Slope. The slope of each dispersion curve  $c$  is computed by the centre difference, similar to Soomro *et al.* (2016)

$$\text{slope}_i = (c_{i+1} - c_{i-1}) / (f_{i+1} - f_{i-1}). \quad (10)$$

The slope at the two boundary points  $c_0$  and  $c_N$  are computed by the forward and backward difference, respectively. Phase-velocity measurements with large slopes will be rejected.

(2) Probability. We use ‘probability’ to represent the extent to which the dispersion curve is consistent with a group of dispersion curves. Regarding the density plot as a 2-D probability density function, the probability of a dispersion curve is defined as the integral of the density function along the curve. The integral is normalized





**Figure 10.** Same as Fig. 9 but after applying the time shifts that account for the interference of neighbouring frequencies.

by the number of frequency samples to keep it within  $[0, 1]$ . Dispersion curves with higher probability are better representatives of the group.

(3) Median absolute deviation (MAD). The MAD criteria, defined as the MAD of a group of phase-velocity measurements for each individual frequency, is used to estimate the variance. We found the MAD stays almost constant for high-SNR frequencies and increases when the SNR drops. Therefore, we can remove less accurate measurements by rejecting high-MAD frequencies.

To perform the quality control, a group of dispersion curves should be selected. Within each group, the curves should have mutual similarity, e.g. those from station pairs that share similar paths or sample the same region. After that, the quality control is performed as follows,

- (1) Compute the probability for each dispersion curve. In this study, the probability ranged from 0 to 0.8.
- (2) Reject the segments of dispersion curves with a large slope. If a dispersion curve splits into segments after the rejection, the longest segment is kept. The threshold value was determined by the visual inspection of the high probability dispersion curves ( $> 0.6$ ). Using our data set, we found that retaining the slopes (eq. 10)  $[-3, 0.5]$  yielded a good balance between removing outliers and keeping quality measurements.
- (3) Recompute the probability using the remaining dispersion curves and reject the low probability dispersion curves. A relatively low threshold of  $< 0.1$  was used.
- (4) Truncate the dispersion curves based on the MAD criterion. We used twice the median of all MADs as the threshold. Frequencies with the MAD greater than the threshold were rejected.

## 5.2 Distance dependence of retrievable frequency range of NCFs

The retrievable frequency range of a NCF, defined as the frequency range of the accurate dispersion curve in this paper, is dependent on the interstation distance. The lower frequency limit is constrained by the far-field approximation used in the derivation of the mathematical representation of NCFs (Snieder 2004; Tsai 2009; Boschi *et al.* 2013; Kästle *et al.* 2016), which have been postulated, based on inspection of real data, to be 1–3 wavelengths (Shapiro *et al.* 2005; Yao *et al.* 2006; Bensen *et al.* 2007; Boschi *et al.* 2013;

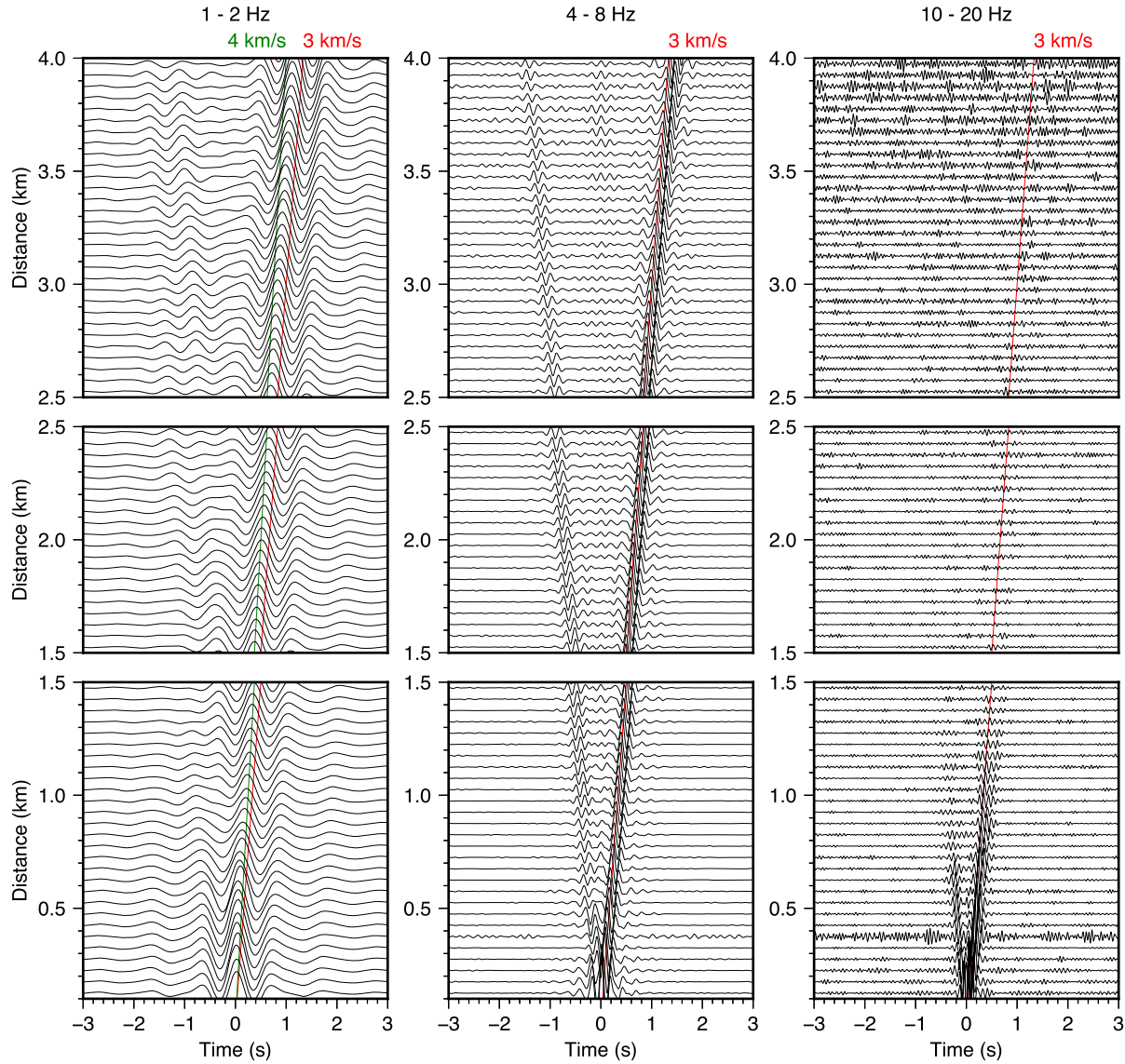
Luo *et al.* 2015). The effective upper frequency limit is due to the increasing attenuation with frequency that is caused by scattering and the intrinsic attenuation, which weakens the higher-frequency energy. By visual inspection of bandpass filtered NCF profiles, we found that the upper limit corresponds to 10–20 wavelengths for the Marathon data set. The limits of retrievable frequency ranges mean that long-distance station pairs are better at retrieving the low-frequency information and short-distance station pairs are better at retrieving the high frequencies.

We illustrate the dependence of the retrievable frequency range on the interstation distance in our data set in Fig. 11. For the long-distance group (2.5–4.0 km), surface waves can be clearly observed at 1–2 Hz or 4–8 Hz, and become less coherent in the high-frequency range (10–20 Hz). In contrast, for the short-distance group (0.5–1.5 km), clear surface waves appear in the higher frequency bands (4–8 or 10–20 Hz), and are less reliable in the low-frequency range (1–2 Hz), due to the interference between causal and acausal branches and the possible breakdown of the far-field approximation.

## 6 APPLICATION TO THE MARATHON ARRAY

### 6.1 Noise cross-correlation and phase-velocity measurement

We applied our cross-correlation method and phase-velocity measurement method (Sections 3 and 4) to the Marathon data set. In this study, we only adopted the pre-processing operations deemed necessary (Section 3.1). The continuous record from each station was split into 1-min segments with 50 per cent overlap. Then, we removed the linear trend of each segment and tapered at both ends within 10 per cent of the segment length. Cross-coherence waveforms from all segments were then linearly stacked, after peak normalization, to produce the stacked NCFs. No explicit bandpass filter, temporal normalization or spectral whitening was applied during the process. We selected only the first hour of the day 01/10/2018, out of the 1-month-long data set, to compute the NCFs. Although 1 hr seems short compared with 10–30 d typically used in previous studies, our analysis of the convergence rate has confirmed that the NCFs converge within half an hour during this time period. In total, 73 724 NCFs were obtained (Fig. 12). We can observe clear surface waves



**Figure 11.** Filtered NCFs with different interstation distances. For illustrative purposes, the waveforms are stacked in 50-m-wide distance bins. The stacked NCFs are divided into three distance ranges: 0.5–1.5 km, 1.5–2.5 km and 2.5–4.0 km and then filtered to three frequency ranges: 1–2 Hz, 4–8 Hz and 10–20 Hz. An apparent velocity of  $3 \text{ km s}^{-1}$  (red line) is used as a reference in all the frequency ranges and an additional reference line of  $4 \text{ km s}^{-1}$  (green) is used in the low-frequency range.

with an apparent velocity of around  $3 \text{ km s}^{-1}$ , which is consistent with the outcrop of high-speed mafic/ultramafic igneous rocks in the Marathon area.

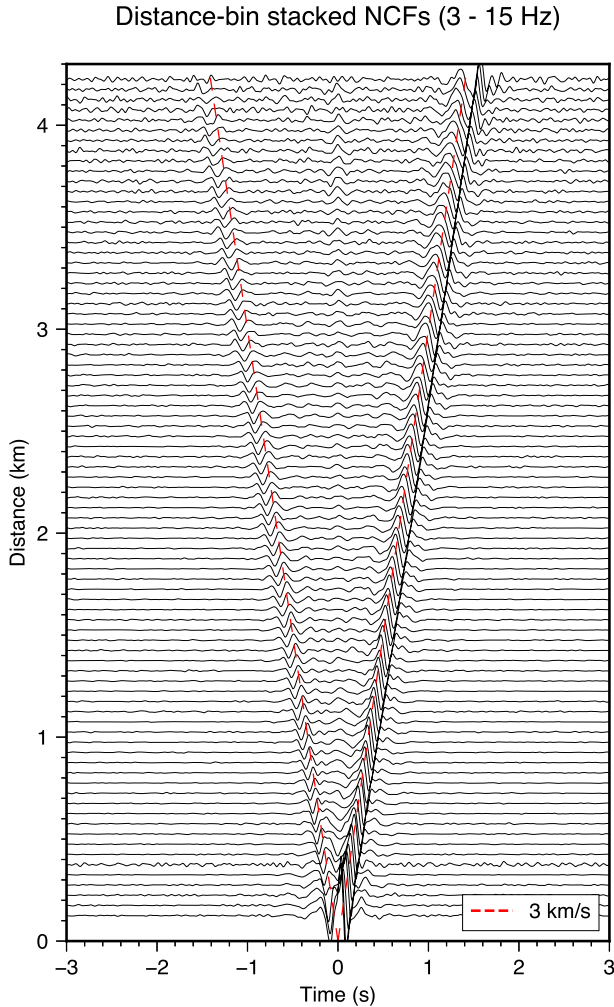
Our phase-velocity measurement method was applied to the symmetric NCFs computed by stacking the causal and acausal branches of the stacked NCFs, so as to balance the asymmetric distribution of the noise sources. A Tukey window, defined between  $[\Delta/c_{\max}-1, \Delta/c_{\min}+1]$  s, was applied to the symmetric NCFs to reduce the noise, where  $c_{\min}$  and  $c_{\max}$  are the upper and lower limit of surface wave velocity, respectively. The additional 2 s extends the window to include a complete cycle of the surface wave signal down to 0.5 Hz, the lowest frequency at which we aimed to make measurements. The frequency-time representations of the NCFs were computed at 50 centre frequencies, distributed logarithmically—so as to balance the structural sensitivity along the broad-band curves (e.g. Agius & Lebedev 2017)—between 0.1 and 30 Hz. The parameter  $\alpha$  that controls the width of narrow-band filters in eq. (5) was set to 1.0. The

amplitude-guided ridge tracking was initialized at the frequency of 1 Hz.

The density plot of automatically measured dispersion curves shows mutual consistency (Figs 13d, g and j), especially at 1–10 Hz. We know this apparent drop at low frequency is an artefact, probably due to the failure of the far-field approximation (Tsai 2009; Kästle *et al.* 2016), because it occurs at different frequencies for different station pairs (0.8, 1.3 and 2 Hz). The divergence of dispersion curves above 10 Hz is likely to be due to either the stronger heterogeneity at shallow depths or the reduced SNR at high frequencies.

## 6.2 Distance-dependent quality-control

For quality control, we divided all the dispersion curves into groups based on their spatial distribution of the stations and the interstation distance. As shown in Fig. 13(a), we covered the Marathon array with an  $11 \times 7$  array of virtual nodes. For each node, we selected



**Figure 12.** Stacked NCFs in 50-m-wide distance bins. Low-count stacked NCFs ( $<30$ ) are omitted in this figure. The NCFs are filtered to 3–15 Hz for illustration purposes. The apparent velocity of  $3 \text{ km s}^{-1}$  (red dashed lines) is plotted as reference.

the dispersion curves with their middle-path points that are in the vicinity of the node. The threshold of the vicinity was 300 m for the medium- ( $1.5 \leq \Delta < 2.5 \text{ km}$ ) and long-distance ( $\Delta \geq 2.5 \text{ km}$ ) groups, and 400 m for the short-distance group ( $\Delta < 1.5 \text{ km}$ ), to ensure a sufficient number of dispersion curves. The three groups of dispersion curves were processed separately, but with the same quality control thresholds. The thresholds are  $[-3, 0.5] \text{ km s}^{-1} \text{ Hz}^{-1}$  for the slope,  $<0.1$  for the low probability dispersion curves rejection, and two times of the median of the MADs for determining the cut-off frequencies.

Figs 13(c)–(k) show an example of the distance-dependent quality control at one node. Each row corresponds to one distance group. We can see a change in frequency ranges of the dispersion curves both before (Figs 13d, g and i) and after the quality control (Figs 13e, h and k), which confirms that shorter-distance station pairs provide higher frequencies. A summary of all the dispersion curves after quality control can be found in Fig. 14, which shows a frequency range of 0.5–30 Hz. As the number of measurements drops when approaching the boundaries of the frequency range (Fig. 14b), we focus our following discussion of phase maps primarily on the 1–15 Hz range.

### 6.3 Phase velocity maps

We used the average dispersion curves from the 77 nodes as the point-wise dispersion curves (Figs 13a and b), yielding phase-velocity maps directly, without any tomographic inversions. Visual inspection of individual dispersion curves in our shows high levels of noise in the individual, single-station-pair dispersion curves (Fig. C1). We address this by means of averaging all the dispersion curves sampling the vicinity of the same node. The situation when individual dispersion curves are much noisier than neighbourhood averages is not unique to high-frequency measurements. It is also encountered in long-period studies, especially when array deployment times are relatively short. Adam & Lebedev (2012) computed very-broad-band, smooth dispersion curves in southern Africa—constraining detailed  $V_s$  structure in broad depth ranges from the upper crust to deep upper mantle (Ravenna *et al.* 2018)—by means of averaging thousands of dispersion measurements within sub-regions. They also showed that strongly smoothed phase-velocity tomography with the original, very noisy station-station measurements yielded results consistent with those of the subregion averaging. In the presence of noise in the individual measurement, the optimal strategy (tomography versus local averaging) depends on the characteristics of the noise and the objectives of the study. Here, we chose the spatial averaging approach as it yielded robust, broad-band phase-velocity curves, well suited for the inversions for the depth ranges of the gabbro intrusions.

For an illustration of the phase-velocity maps, we selected six representative frequencies at around 15, 10, 5, 3, 2 and 1 Hz (Fig. 15). The average velocity at each frequency shows a steady increase as the frequency decreases, from  $2.67 \text{ km s}^{-1}$  at 15.1 Hz to  $3.24 \text{ km s}^{-1}$  at 1.23 Hz. Regarding the anomalies, the most prominent feature is the high-velocity anomaly that matches, roughly, the outcrops of the gabbro intrusions at higher frequencies (Figs 15a and b) and gradually moves westwards at lower frequencies (Figs 15c–f). The anomaly reaches the central and western parts of the array at 3.06 and 1.23 Hz, respectively.

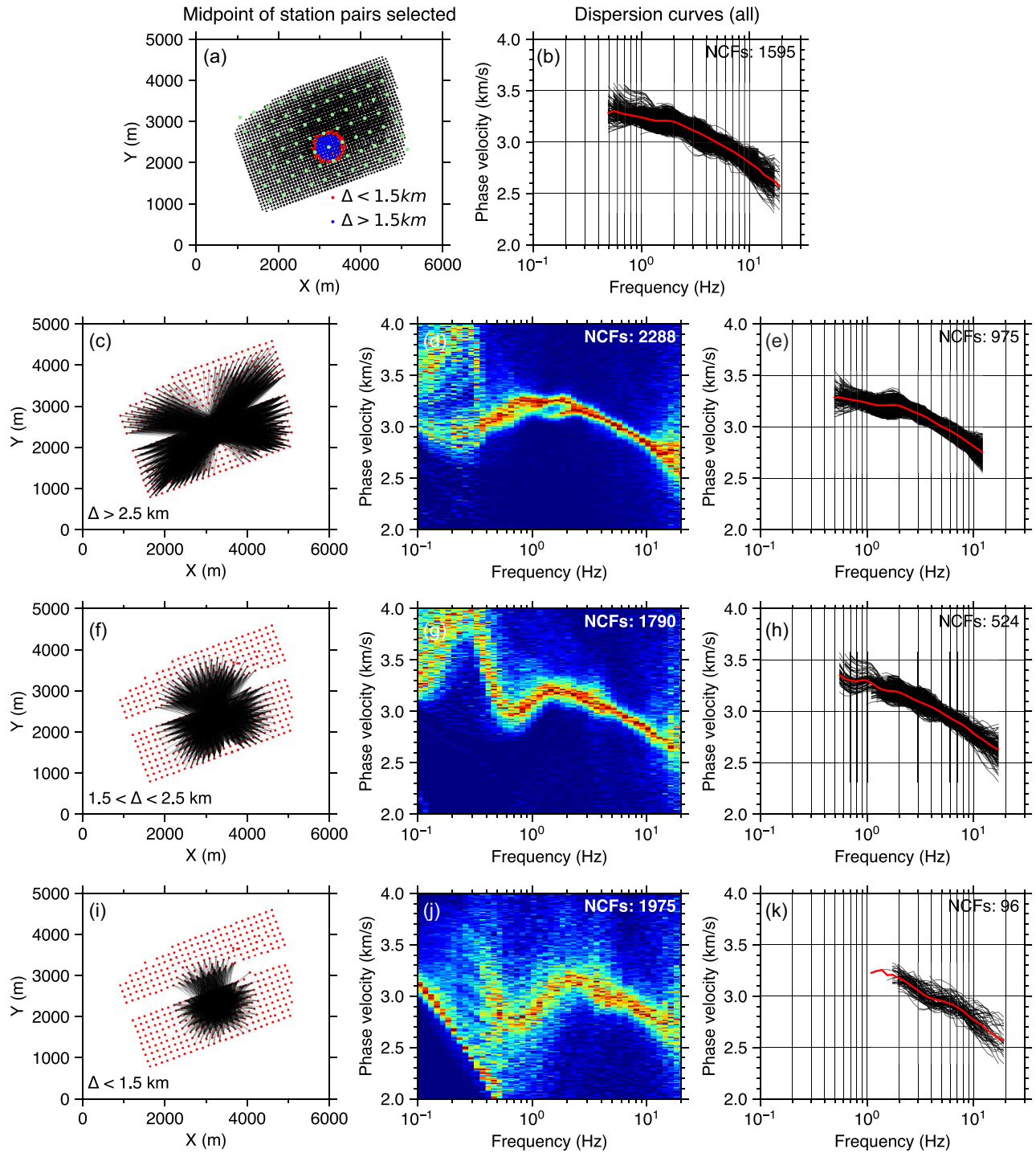
### 6.4 Subregion average dispersion curves and 1-D $S$ -wave velocity models

To see the lateral variations of the average dispersion curves even clearer, we divided them into the west, centre and east groups and computed three subregion average dispersion curves (Fig. 16). Among them, the east region has the highest phase velocity at 5–20 Hz, whereas the centre region has the highest phase velocity at 2–5 Hz, and the west region—below 2 Hz. Because the depth range sampled by the phase-velocity measurements gets shallower with the increasing frequency (Fig. 17), the result implies that a high-velocity anomaly starts at the shallower depth in the east region and deepens progressively westward.

To obtain the depth of the high-velocity anomaly, we inverted the three average dispersion curves for 1-D shear wave velocity models. Each 1-D model comprises 50 layers above 3 km depth and 1 extra layer that is a half-space below 3 km. Taking into account the broadening of the surface wave sensitivity kernels with depth (Fig. 17), the thickness of the 50 layers are logarithmically distributed, with thinner layers at shallow depths and thicker layers at greater depths. We inverted for the shear velocity model using the least square method. The objective function is defined as

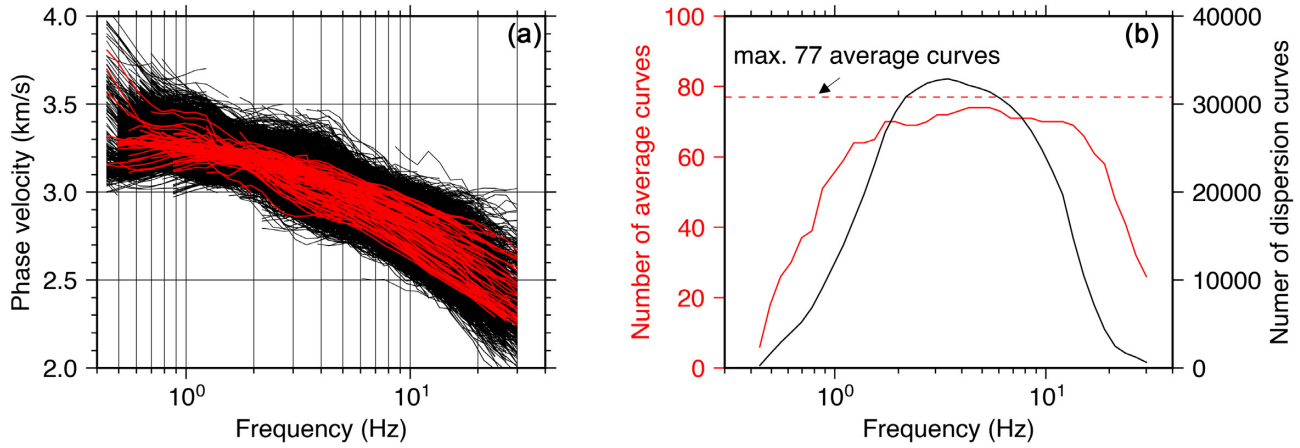
$$\chi^2 = \|\mathbf{G}(\mathbf{m}) - \mathbf{d}\|^2 + \lambda \|\mathbf{Lm}\|^2, \quad (11)$$



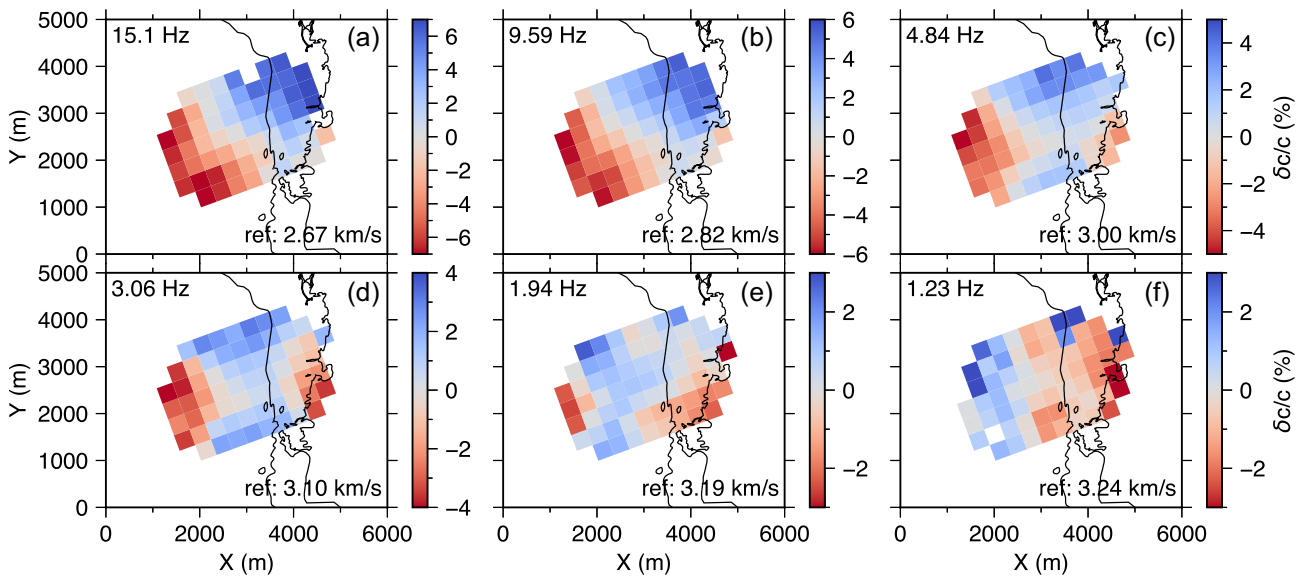


**Figure 13.** Construction of the broad-band phase-velocity curve. (a) Station pairs selected. The green dots denote the locations of the  $11 \times 7$  nodes. The black dots denote the middle-points of all station pairs, and the red and blue dots denote the ones selected. A larger radius is used for the short-distance group ( $\Delta < 1.5$  km) to include a sufficient number of station pairs. (b) Dispersion curves associated with selected station pairs after the quality control (black) and the median dispersion curve (red). (c, d, e) Station pairs selected for  $\Delta > 2.5$  km and the dispersion curves before and after quality control. Red dots in (c) denote the locations of all the stations and black lines denote the selected station pairs. Panel (d) shows the density plot of the corresponding, automatically measured dispersion curves. The number of NCFs is labelled at the upper right corner. Panel (e) shows the dispersion curves after quality control (black) and the average dispersion curve (red). (f, g, h) Station pairs selected for  $1.5 < \Delta < 2.5$  km and the corresponding dispersion curves before and after quality control. (i, j, k) Station pairs selected for  $\Delta < 1.5$  km and the corresponding dispersion curves before and after quality control.





**Figure 14.** Dispersion curves obtained after quality control. (a) Dispersion curves for each station pair (black lines) and average dispersion curves (red lines) for each node (green dots in Fig. 13a). (b) Number of dispersion curves (black) and average dispersion curves (red). The horizontal red dashed line marks the maximum number of average dispersion curves (77), as 77 nodes were used to obtain the average curves (Fig. 13a).



**Figure 15.** Phase-velocity maps constructed via the combination of dispersion curves at the 77 nodes. (a) Phase velocity map at the frequency of 15.1 Hz. The blue-red mosaic shows the phase velocity residuals relative to the reference velocity labelled at the lower-right-hand corner. The black lines outline the gabbro intrusions. (b) Same as (a) but at 9.59 Hz, (c) 4.84 Hz, (d) 3.06 Hz, (e) 1.94 Hz and (f) 1.23 Hz.

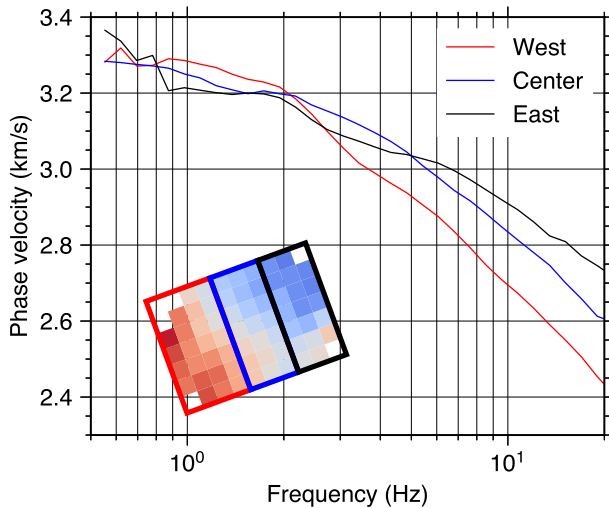
where  $\mathbf{m}$  is the shear wave velocity,  $\mathbf{d}$  the phase velocity,  $\mathbf{G}$  the forward modelling operator from shear-wave velocity to phase velocity,  $\mathbf{L} (= \mathbf{m}_{i+1} - \mathbf{m}_i)$  the differential operator and  $\lambda$  the smoothing factor that balances the data fitting and the model roughness. Using the initial model that was interpolated from a best-fitting two-layered model obtained by grid searching, we solve the least-square problem by the gradient descent method with a smoothing factor of 0.4.

The inverted models match the observed dispersion curves with the relative misfits of less than 0.5 per cent (Figs 18b and c) and reveal the depth of the high-velocity anomaly in the east, centre and west subregions (Fig. 18a), at about 0.2, 0.5 and 0.9 km, respectively.

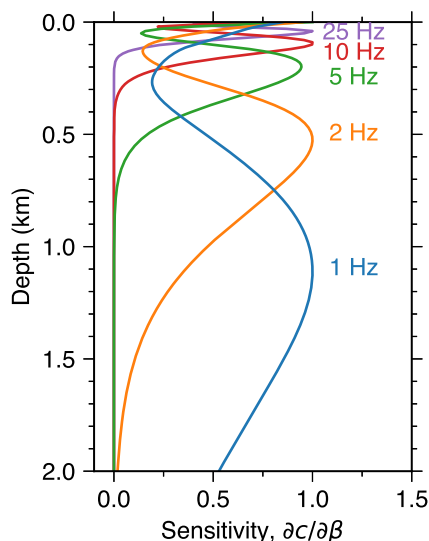
## 7 DISCUSSION

Our experiments show that the optimal noise cross-correlation workflow needs to be modified for high-frequency ( $>1$  Hz) noise studies, due to the changes in noise sources, specifically the stronger temporal variations of high-frequency noise sources compared with

ocean microseisms. Striving for the widest band phase velocity measurements, one of the most relevant changes in the workflow should be replacing the classical temporal normalization (e.g. one-bit or running absolute average normalization) with short segment stacking, which allows the dropping of the bandpass filtering step without reducing the quality of NCFs. Other changes include using cross-coherence, optimizing the segment length and choosing a fast convergence time period. Improvements in phase velocity measurement method and distance-dependent quality control, despite irrelevant to the changes in noise sources, also contribute to the extension of the frequency band of the measurements. Application of our workflow to the Marathon data set demonstrates its effectiveness and yields wide-band (0.5–30 Hz, and 5.9 in octaves) phase-velocity measurements, twice as wide as the typical bandwidth (1.5–3 octaves) of previous ambient noise cross-correlation studies at higher frequencies (e.g. Gouédard *et al.* 2008; Picozzi *et al.* 2009; Renalier *et al.* 2010; de Ridder & Dellinger 2011; Mordret *et al.* 2013; Lin *et al.* 2013; Szanyi *et al.* 2016; Zeng *et al.* 2017; Spica *et al.* 2018;



**Figure 16.** Subregion-average dispersion curves. The boundaries of the regions are shown in the inset map with the same colours as the dispersion curves.



**Figure 17.** Depth sensitivity kernels of the phase velocity of the fundamental-mode Rayleigh waves.

Mordret *et al.* 2019; Chmiel *et al.* 2019; Hollis *et al.* 2019; Inzunza *et al.* 2019; Zigone *et al.* 2019). The wide-band measurements will enable us to image the structure at the depths from 0.05 to 2.0 km.

According to previous geological studies (Good *et al.* 2015) and the lithological study of the drill core samples (Gunawardana 2017), the Marathon region can be divided into three subregions, with high-velocity gabbro intrusions in the centre, and the relatively low-velocity syenite and intermediate metavolcanic rocks in the west and east subregions, respectively. Our phase velocity maps reveal a west-dipping high-velocity anomaly. A spatial correspondence is found between the gabbro intrusions and the high-velocity anomaly in the phase maps ( $>10$  Hz), confirming that the multiple layers of gabbro intrusions have a higher shear-wave velocity than syenite and intermediate metavolcanic rocks. Similar conclusions were drawn by Hollis *et al.* (2019) as well, who used a sparser pilot array in the same region. The high-velocity anomaly in the 15-Hz phase map extends beyond the outcrop area of the gabbro intrusions (Fig. 15a), probably due to the lateral averaging of the phase-velocity maps.

Alternatively, the shallow dip angle of the gabbro intrusions and the relatively broad ranges of the depth sensitivity of surface waves at each frequency could explain the apparent inconsistency.

With the high-velocity anomaly interpreted as the body comprising gabbro intrusions, our results show that this body dips to the west at a shallow angle. The upper contact of the gabbro intrusions locates at the depths of 0.0, 0.2 and 0.4 km beneath the east, centre and west part of the array, respectively, while the lower contact is at the depths of 0.4, 0.9 and  $>1.5$  km. The lower contact beneath the west part of the array is not well constrained due to the lack of the low-frequency dispersion measurements ( $<0.5$  Hz), which is limited by the size of the array. The interpretation is consistent with previous geological studies which showed the gabbro intrusions have a westwards dipping structure beneath the eastern part of the array (Good *et al.* 2015), and provides new constraints on the structure in the centre and west part. Detailed mapping and interpretation of the 3-D shear velocity structure will be the focus of a future publication.

## 8 CONCLUSIONS

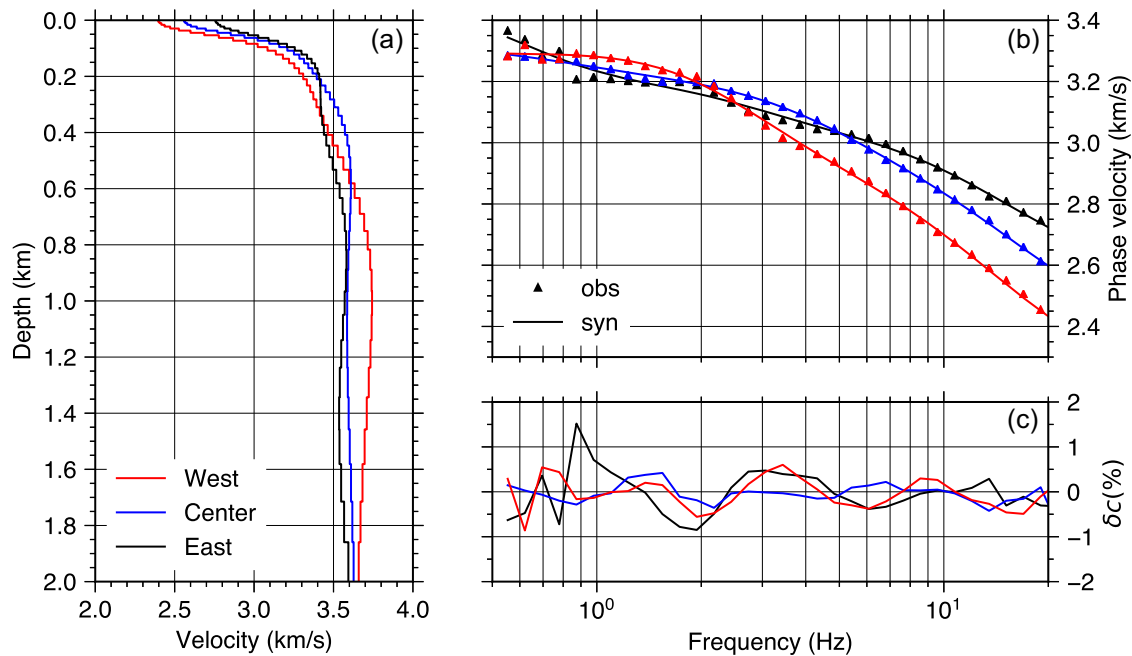
The optimal workflow for the high-frequency noise cross-correlation ( $>1$  Hz), defined as the one retrieving the broadest bandwidth of accurate dispersion measurements in this study, is found to be different from the classic workflow used in crustal-scale ambient noise studies. The differences are primarily due to the strong temporal variations of the high-frequency noise sources. Our analysis of the cumulative stacked NCFs shows that short bursts (e.g. 10 min) of strong noise sources can offer high-quality NCFs and, thus, more structural information than a long time (e.g. 10 hr) of weak sources.

The strong temporal variation of the noise sources suggests different pre-processing techniques for obtaining a high-quality NCF. We recommend the following workflow for the processing of the ambient noise data from dense arrays. Broadly, it comprises the noise cross-correlation, phase-velocity measurement and quality control. The noise cross-correlation is the most important step, which determines the quality of recovered surface waves. It comprises

- (1) Splitting continuous seismograms to *short* overlapping segments.
- (2) Basic pre-processing, including detrending and tapering. Removal of the instrumental response if applicable.
- (3) Computing the cross-coherence.
- (4) Stacking.

Compared with the typical workflow in previous noise cross-correlation studies, our workflow is simplified. Bandpass filtering, temporal normalization and explicit spectral whitening are removed. However, the quality of the NCFs is not compromised, because all the positive effects of the temporal and spectral normalization are achieved by the short segment stacking and the cross-coherence.

The phase-velocity measurement exploits the multiple solutions due to the  $2\pi$  ambiguity to extend its bandwidth. The amplitude-guided ridge tracking method we developed can automatically select high-quality measurements from different solutions and combine them to generate a dispersion curve, which facilitates efficient processing of the massive amount of dispersion curves generated from a large-N array. Furthermore, when low-frequency surface waves are available, the method can automatically resolve the  $2\pi$  ambiguity without the help of a reference dispersion curve.



**Figure 18.** Subregion-average 1-D  $V_s$  models. (a) 1-D  $V_s$  models obtained by inverting the average dispersion curves in Fig. 16 and (b). The top 2.0 km is shown here to highlight velocity changes at the shallow depths. (b) Comparison of the observed (triangle) and synthetic (solid line) dispersion curves. (c) Relative misfit between the observed and synthetic dispersion curves.

For quality control, dispersion curves can be divided into groups according to the interstation distance. Because the frequency of the retrievable surface waves generally decreases with an increasing interstation distance, grouping by distance produces bundles of dispersion curves in similar frequency ranges and makes the mutual similarity within the bundle of measurements an effective quality criterion.

Application of the newly developed workflow to the Marathon data set yielded Rayleigh-wave, phase-velocity measurements from 0.5 to 30 Hz. The bandwidth is 5.9 octaves, about twice as wide as the typical bandwidth (1.5–3 octaves) in high-frequency noise cross-correlation studies. The phase maps reveal a west-dipping high-velocity anomaly, which probably indicates the gabbro intrusions hosting the Marathon deposit.

## ACKNOWLEDGEMENTS

We thank Elmer Ruigrok, an anonymous reviewer and the Editor, Andrea Morelli, for constructive comments and suggestions that helped us to improve the manuscript. The work was supported by the European Union's Horizon 2020 research and innovation program under grant agreement No. 776622, PACIFIC. This publication reflects only the authors' view and the European Commission is not responsible for any use that may be made of the information it contains. It was also supported by the Science Foundation Ireland (SFI) grant 13/CDA/2192, SFI grant 16/IA/4598, cofunded by the Geological Survey of Ireland and the Marine Institute, and SFI grant 13/RC/2092, co-funded under the (0:funding-source 3:href="http://dx.doi.org/10.13039/501100008530")European Regional Development Fund/(0:funding-source) and by iCrag industry partners. Matplotlib (Hunter 2007) and Generic Mapping Tools (Wessel *et al.* 2013) were used to produce the figures.

## DATA AVAILABILITY STATEMENT

The Marathon data set (Brengruer *et al.* 2018) underlying this paper is subject to an embargo until 31 December 2021. Once the embargo expires, the data can be shared on reasonable request to the members of the PACIFIC consortium (<https://www.pacific-h2020.eu>, last accessed 2 January 2021).

## REFERENCES

- Adam, J.M.-C. & Lebedev, S., 2012. Azimuthal anisotropy beneath southern Africa, from very-broadband, surface-wave dispersion measurements, *Geophys. J. Int.*, **191**, 155–174.
- Agius, M.R. & Lebedev, S., 2017. Complex, multilayered azimuthal anisotropy beneath Tibet: evidence for co-existing channel flow and pure-shear crustal thickening, *Geophys. J. Int.*, **210**, 1823–1844.
- Aki, K., 1957. Space and time spectra of stationary waves with special reference to microtremors, *Bull. Earthq. Res. Inst. Univ. Tokyo*, **35**, 415–456.
- Behm, M. & Snieder, R., 2013. Love waves from local traffic noise interferometry, *Leading Edge*, **32**, 628–632.
- Bensen, G.D., Ritzwoller, M.H., Barmin, M.P., Levshin, A.L., Lin, F., Moschetti, M.P., Shapiro, N.M. & Yang, Y., 2007. Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements, *Geophys. J. Int.*, **169**, 1239–1260.
- Bensen, G.D., Ritzwoller, M.H. & Shapiro, N.M., 2008. Broadband ambient noise surface wave tomography across the United States, *J. geophys. Res.*, **113**, B05306, doi:10.1029/2007JB005248.
- Brengruer, F. & Hollis, D., Dublin Institute for Advanced Studies, Generation Mining, Sisprobe SAS, & Grenoble Alpes University, 2018. PACIFIC ambient noise dense seismic survey at Marathon deposits [Data set]. *International Federation of Digital Seismograph Networks*, available at: <https://doi.org/10.7914/SN/6L.2018>.
- Bonadio, R., 2019. Broadband surface wave tomography of Ireland, Britain, and Other Regions, *PhD thesis*, Trinity College Dublin, Dublin, Ireland.
- Bonadio, R. *et al.*, 2018. Hot upper mantle beneath the Tristan da Cunha Hotspot from probabilistic Rayleigh-wave inversion and petrological modeling, *Geochem. Geophys. Geosyst.*, **19**, 1412–1428.

- Bonadio, R. *et al.*, 2021. Optimal resolution tomography with error tracking: imaging the upper mantle beneath Ireland and Britain, *Geophys. J. Int.*, **226**, 2158–2188.
- Boschi, L., Weemstra, C., Verbeke, J., Ekström, G., Zunino, A. & Giardini, D., 2013. On measuring surface wave phase velocity from station–station cross-correlation of ambient signal, *Geophys. J. Int.*, **192**, 346–358.
- Brenguier, F., Shapiro, N.M., Campillo, M., Nercissian, A. & Ferrazzini, V., 2007. 3-D surface wave tomography of the Piton de la Fournaise volcano using seismic noise correlations, *Geophys. Res. Lett.*, **34**, L02305, doi:10.1029/2006GL028586.
- Calkins, J.A., Abers, G.A., Ekström, G., Creager, K.C. & Rondenay, S., 2011. Shallow structure of the Cascadia subduction zone beneath western Washington from spectral ambient noise correlation, *J. geophys. Res.*, **116**, B07302, doi:10.1029/2010JB007657.
- Campillo, M. & Paul, A., 2003. Long-range correlations in the diffuse seismic coda, *Science*, **299**, 547–549.
- Carvalho, J., Bonadio, R., Silveira, G., Lebedev, S., Mata, J., Arroucau, P., Meier, T. & Celli, N.L., 2019. Evidence for high temperature in the upper mantle beneath Cape Verde archipelago from Rayleigh-wave phase-velocity measurements, *Tectonophysics*, **770**, 228225.
- Chang, J.P., Ridder, S.A.L. & Biondi, B.L., 2016. High-frequency Rayleigh-wave tomography using traffic noise from Long Beach, California, *Geophysics*, **81**, B43–B53.
- Cheng, F. *et al.*, 2016. Multichannel analysis of passive surface waves based on crosscorrelations, *Geophysics*, **81**, EN57–EN66.
- Chmiel, M. *et al.*, 2019. Ambient noise multimode Rayleigh and Love wave tomography to determine the shear velocity structure above the Groningen gas field, *Geophys. J. Int.*, **218**, 1781–1795.
- Chmiel, M., Roux, P., Herrmann, P., Rondeleux, B. & Wathelet, M., 2018. Data-based diffraction kernels for surface waves from convolution and correlation processes through active seismic interferometry, *Geophys. J. Int.*, **214**, 1468–1480.
- Curtis, A., Gerstoft, P., Sato, H., Snieder, R. & Wapenaar, K., 2006. Seismic interferometry—turning noise into signal, *Leading Edge*, **25**, 1082–1092.
- Dahlen, F.A. & Tromp, J., 1998. *Theoretical Global Seismology*, Princeton Univ. Press.
- Dales, P. *et al.*, 2020. Virtual sources of body waves from noise correlations in a mineral exploration context, *Seismol. Res. Lett.*, **91**, 2278–2286.
- de Ridder, S. & Dellinger, J., 2011. Ambient seismic noise eikonal tomography for near-surface imaging at Valhall, *Lead. Edge*, **30**, 506–512.
- de Ridder, S. & Biondi, B.L., 2013. Daily reservoir-scale subsurface monitoring using ambient seismic noise, *Geophys. Res. Lett.*, **40**, 2969–2974.
- Draganov, D., Wapenaar, K., Mulder, W., Singer, J. & Verdel, A., 2007. Retrieval of reflections from seismic background-noise measurements, *Geophys. Res. Lett.*, **34**, L04305, doi:10.1029/2006GL028735.
- Dziewonski, A., Bloch, S. & Landisman, M., 1969. A technique for the analysis of transient seismic signals, *Bull. seism. Soc. Am.*, **59**, 427–444.
- Ekström, G., 2014. Love and Rayleigh phase-velocity maps, 5–40 s, of the western and central USA from USArray data, *Earth planet. Sci. Lett.*, **402**, 42–49.
- Ekström, G., Abers, G.A. & Webb, S.C., 2009. Determination of surface-wave phase velocities across USArray from noise and Aki's spectral formulation, *Geophys. Res. Lett.*, **36**, L18301, doi:10.1029/2009GL039131.
- El-Sharkawy, A., Meier, T., Lebedev, S., Behrmann, J.H., Hamada, M., Cristiano, L., Weidle, C. & Köhn, D., 2020. The slab puzzle of the Alpine-Mediterranean Region: insights from a new, high-resolution, shear wave velocity model of the upper mantle, *Geochem. Geophys. Geosystems*, **21**, e2020GC008993, doi:10.1029/2020GC008993.
- Good, D.J., Epstein, R., McLean, K., Linnen, R.L. & Samson, I.M., 2015. Evolution of the Main Zone at the Marathon Cu-PGE sulfide deposit, Mid-continent Rift, Canada: spatial relationships in a magma conduit setting, *Econ. Geol.*, **110**, 983–1008.
- Gouédaud, P., Roux, P. & Campillo, M., 2008. Small-scale seismic inversion using surface waves extracted from noise cross correlation, *J. acoust. Soc. Am.*, **123**, EL26–EL31.
- Groos, J.C., Bussat, S. & Ritter, J.R.R., 2012. Performance of different processing schemes in seismic noise cross-correlations: processing schemes in noise interferometry, *Geophys. J. Int.*, **188**, 498–512.
- Groos, J.C. & Ritter, J.R.R., 2009. Time domain classification and quantification of seismic noise in an urban environment, *Geophys. J. Int.*, **179**, 1213–1231.
- Gunawardana, H., 2017. Petrophysical Investigations of the Marathon Cu-PGE Deposit, Marathon, ON, *PhD thesis*, The University of Western Ontario, London, Canada.
- Halliday, D. & Curtis, A., 2008. Seismic interferometry, surface waves and source distribution, *Geophys. J. Int.*, **175**, 1067–1087.
- Halliday, D., Curtis, A. & Kragh, E., 2008. Seismic surface waves in a sub-urban environment: active and passive interferometric methods, *Leading Edge*, **27**, 210–218.
- Hanasoge, S.M. & Branicki, M., 2013. Interpreting cross-correlations of one-bit filtered seismic noise, *Geophys. J. Int.*, **195**, 1811–1830.
- Hannemann, K., Papazachos, C., Ohrnberger, M., Savvaidis, A., Anthymidis, M. & Lontsi, A.M., 2014. Three-dimensional shallow structure from high-frequency ambient noise tomography: new results for the Mygdonia basin-Euroseistest area, northern Greece, *J. geophys. Res.*, **119**, 4979–4999.
- Hinze, W.J. & Chandler, V.W., 2020. Reviewing the configuration and extent of the Midcontinent rift system, *Precambrian Res.*, **342**, 105688.
- Hollis, D., McBride, J., Beaupretre, S., Mordret, A., Brenguier, F., Arndt, N. & Good, D., 2019. Ambient noise surface wave tomography at the Marathon PGM-Cu deposit, Ontario, Canada, *Recorder*, **44**, 1–20.
- Hunter, J.D., 2007. Matplotlib: a 2D graphics environment, *Comput. Sci. Eng.*, **9**, 90–95.
- Inzunza, D.A., Montalva, G.A., Leyton, F., Prieto, G.A. & Ruiz, S., 2019. Shallow ambient-noise 3D tomography in the Concepción Basin, Chile: implications for low-frequency ground motions, *Bull. seism. Soc. Am.*, **109**, 75–86.
- Kästle, E.D., Soomro, R., Weemstra, C., Boschi, L. & Meier, T., 2016. Two-receiver measurements of phase velocity: cross-validation of ambient-noise and earthquake-based observations, *Geophys. J. Int.*, **207**, 1493–1512.
- Kimman, W.P. & Trampert, J., 2010. Approximations in seismic interferometry and their effects on surface waves, *Geophys. J. Int.*, **182**, 461–476.
- Lavoué, F. *et al.*, 2021. Understanding seismic waves generated by train traffic via modelling: implications for seismic imaging and monitoring, *Seismol. Res. Lett.*, **92**, 1–14.
- Le Pape, F., Craig, D. & Bean, C. J., 2021. How deep ocean–land coupling controls the generation of secondary microseism Love waves, *Nat. Commun.*, **12**, 2332.
- Lehujeur, M., Vergne, J., Schmittbuhl, J., Zigone, D. & Le Chenadec, A., EstOF Team, 2018. Reservoir imaging using ambient noise correlation from a dense seismic network, *J. geophys. Res.*, **123**, 6671–6686.
- Levshin, A., Ratnikova, L. & Berger, J., 1992. Peculiarities of surface-wave propagation across central Eurasia, *Bull. seism. Soc. Am.*, **82**, 2464–2493.
- Levshin, A.L. & Ritzwoller, M.H., 2001. Automated detection, extraction, and measurement of regional surface waves, *Pure appl. Geophys.*, **158**, 1531–1545.
- Li, H., Su, W., Wang, C.-Y. & Huang, Z., 2009. Ambient noise Rayleigh wave tomography in western Sichuan and eastern Tibet, *Earth planet. Sci. Lett.*, **282**, 201–211.
- Li, Z., Ni, S., Zhang, B., Bao, F., Zhang, S., Deng, Y. & Yuen, D.A., 2016. Shallow magma chamber under the Wudalianchi volcanic field unveiled by seismic imaging with dense array, *Geophys. Res. Lett.*, **43**, 4954–4961.
- Lin, F.-C., Li, D., Clayton, R.W. & Hollis, D., 2013. High-resolution 3D shallow crustal structure in Long Beach, California: application of ambient noise tomography on a dense seismic array, *Geophysics*, **78**, Q45–Q56.
- Lin, F.-C., Ritzwoller, M.H., Townend, J., Bannister, S. & Savage, M.K., 2007. Ambient noise Rayleigh wave tomography of New Zealand, *Geophys. J. Int.*, **170**, 649–666.
- Lobkis, O.I. & Weaver, R.L., 2001. On the emergence of the Green's function in the correlations of a diffuse field, *J. acoust. Soc. Am.*, **110**, 3011–3017.
- Luo, Y., Yang, Y., Xu, Y., Xu, H., Zhao, K. & Wang, K., 2015. On the limitations of interstation distances in ambient noise tomography, *Geophys. J. Int.*, **201**, 652–661.
- Martins, J.E., Ruigrok, E., Draganov, D., Hooper, A., Hanssen, R.F., White, R.S. & Soosalu, H., 2019. Imaging Torfajökull's magmatic plumbing



- system with seismic interferometry and phase velocity surface wave tomography, *J. geophys. Res.*, **124**, 2920–2940.
- Meier, T., Dietrich, K., Stöckhert, B. & Harjes, H.-P., 2004. One-dimensional models of shear wave velocity for the eastern Mediterranean obtained from the inversion of Rayleigh wave phase velocities and tectonic implications, *Geophys. J. Int.*, **156**, 45–58.
- Möllhoff, M. & Bean, C. J., 2016. Seismic noise characterization in proximity to strong microseismic sources in the Northeast Atlantic, *Bull. seism. Soc. Am.*, **106**, 464–477.
- Mordret, A., Landès, M., Shapiro, N.M., Singh, S.C., Roux, P. & Barkved, O.I., 2013. Near-surface study at the Valhall oil field from ambient noise surface wave tomography, *Geophys. J. Int.*, **193**, 1627–1643.
- Mordret, A., Rivet, D., Landès, M. & Shapiro, N.M., 2015. Three-dimensional shear velocity anisotropic model of Piton de la Fournaise Volcano (La Réunion Island) from ambient seismic noise, *J. geophys. Res.*, **120**, 406–427.
- Mordret, A., Roux, P., Boué, P. & Ben-Zion, Y., 2019. Shallow three-dimensional structure of the San Jacinto fault zone revealed from ambient noise imaging with a dense seismic array, *Geophys. J. Int.*, **216**, 896–905.
- Moschetti, M.P., Ritzwoller, M.H. & Shapiro, N.M., 2007. Surface wave tomography of the western United States from ambient seismic noise: rayleigh wave group velocity maps, *Geochem. Geophys. Geosyst.*, **8**, Q08010, doi:10.1029/2007GC001655.
- Nakata, N., Snieder, R., Tsuji, T., Lerner, K. & Matsuoka, T., 2011. Shear wave imaging from traffic noise using seismic interferometry by cross-coherence, *Geophysics*, **76**, SA97–SA106.
- Pawlak, A., Eaton, D.W., Darbyshire, F., Lebedev, S. & Bastow, I.D., 2012. Crustal anisotropy beneath Hudson Bay from ambient noise tomography: evidence for post-orogenic lower-crustal flow? *J. geophys. Res.*, **117**, 1–16.
- Pedersen, H.A. & Krüger, F., the SVEKALAPKO Seismic Tomography Working Group, 2007. Influence of the seismic noise characteristics on noise correlations in the Baltic shield, *Geophys. J. Int.*, **168**, 197–210.
- Picozzi, M., Parolai, S., Bindi, D. & Strollo, A., 2009. Characterization of shallow geology by high-frequency seismic noise tomography, *Geophys. J. Int.*, **176**, 164–174.
- Prieto, G.A. & Beroza, G.C., 2008. Earthquake ground motion prediction using the ambient seismic field, *Geophys. Res. Lett.*, **35**, L14304, doi:10.1029/2008GL034428.
- Prieto, G.A., Denolle, M., Lawrence, J.F. & Beroza, G.C., 2011. On amplitude information carried by the ambient seismic field, *Comptes Rendus Geosci.*, **343**, 600–614.
- Prieto, G.A., Lawrence, J.F. & Beroza, G.C., 2009. Anelastic Earth structure from the coherency of the ambient seismic field, *J. geophys. Res.*, **114**, B07303, doi:10.1029/2008JB006067.
- Ravenna, M., Lebedev, S., Fulla, J. & Adam, J.M.-C., 2018. Shear-wave velocity structure of southern Africa's lithosphere: variations in the thickness and composition of cratons and their effect on topography, *Geochem. Geophys. Geosyst.*, **19**, 1499–1518.
- Rawlinson, N., Arroucau, P., Musgrave, R., Cayley, R., Young, M. & Salmon, M., 2014. Complex continental growth along the proto-Pacific margin of East Gondwana, *Geology*, **42**, 783–786.
- Renalier, F., Jongmans, D., Campillo, M. & Bard, P.-Y., 2010. Shear wave velocity imaging of the Avignonet landslide (France) using ambient noise cross correlation, *J. geophys. Res.*, **115**, doi:10.1029/2009JF001538.
- Ritzwoller, M.H. & Levshin, A.L., 1998. Eurasian surface wave tomography: group velocities, *J. geophys. Res.*, **103**, 4839–4878.
- Roux, P. & Fink, M., 2003. Green's function estimation using secondary sources in a shallow water environment, *J. acoust. Soc. Am.*, **113**, 1406–1416.
- Roux, P., Sabra, K.G., Gerstoft, P., Kuperman, W.A. & Fehler, M.C., 2005a. P-waves from cross-correlation of seismic noise, *Geophys. Res. Lett.*, **32**, 1–4.
- Roux, P., Sabra, K.G., Kuperman, W.A. & Roux, A., 2005b. Ambient noise cross correlation in free space: theoretical approach, *J. acoust. Soc. Am.*, **117**, 79–84.
- Sabra, K.G., 2005. Extracting time-domain Green's function estimates from ambient seismic noise, *Geophys. Res. Lett.*, **32**, L03310, doi:10.1029/2004GL021862.
- Sato, Y., 1955. Analysis of dispersed surface waves by means of Fourier transform: part 1, *Bulletin of the Earthquake Research Institute, University of Tokyo*, **33**, 33–47.
- Saygin, E. & Kennett, B.L.N., 2010. Ambient seismic noise tomography of Australian continent, *Tectonophysics*, **481**, 116–125.
- Seats, K.J., Lawrence, J.F. & Prieto, G.A., 2012. Improved ambient noise correlation functions using Welch's method, *Geophys. J. Int.*, **188**, 513–523.
- Shapiro, N.M. & Campillo, M., 2004. Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise, *Geophys. Res. Lett.*, **31**, L07614, doi:10.1029/2004GL019491.
- Shapiro, N.M., Campillo, M., Stehly, L. & Ritzwoller, M.H., 2005. High-resolution surface-wave tomography from ambient seismic noise, *Science*, **307**, 1615–1618.
- Shen, W. et al., 2016. A seismic reference model for the crust and uppermost mantle beneath China from surface wave dispersion, *Geophys. J. Int.*, **206**, 954–979.
- Slob, E., Draganov, D. & Wapenaar, K., 2007. Interferometric electromagnetic Green's functions representations using propagation invariants, *Geophys. J. Int.*, **169**, 60–80.
- Snieder, R., 2004. Extracting the Green's function from the correlation of coda waves: a derivation based on stationary phase, *Phys. Rev. E*, **69**, 046610.
- Soomro, R.A., Weidle, C., Cristiano, L., Lebedev, S. & Meier, T., PASSEQ Working Group, 2016. Phase velocities of Rayleigh and Love waves in central and northern Europe from automated, broad-band, interstation measurements, *Geophys. J. Int.*, **204**, 517–534.
- Spica, Z., Pertou, M., Nakata, N., Liu, X. & Beroza, G.C., 2018. Shallow Vs imaging of the Groningen Area from joint inversion of multimode surface waves and H/V spectral ratios, *Seismol. Res. Lett.*, **89**, 1720–1729.
- Stein, S. et al., 2018. Insights from North America's failed Midcontinent Rift into the evolution of continental rifts and passive continental margins, *Tectonophysics*, **744**, 403–421.
- Szanyi, G., Grácz, Z., Györi, E., Kaláb, Z. & Lednická, M., 2016. Ambient seismic noise tomography of a loess high bank at Dunaszekcső (Hungary), *Pure appl. Geophys.*, **173**, 2913–2928.
- Tsai, V.C., 2009. On establishing the accuracy of noise tomography travel-time measurements in a realistic medium, *Geophys. J. Int.*, **178**, 1555–1564.
- Villaseñor, A., Yang, Y., Ritzwoller, M.H. & Gallart, J., 2007. Ambient noise surface wave tomography of the Iberian Peninsula: implications for shallow seismic structure, *Geophys. Res. Lett.*, **34**, L11304, doi:10.1029/2007GL030164.
- Walker, E.C., Sutcliffe, R.H., Shaw, C.S.J., Shore, G.T. & Penczak, R.S., 1993. Precambrian geology of the Coldwell Alkaline Complex, Ontario Geological Survey Open File Report 5868, Toronto, Canada.
- Wang, Y., Lin, F.-C., Schmandt, B. & Farrell, J., 2017. Ambient noise tomography across Mount St. Helens using a dense seismic array, *J. Geophys. Res. Solid Earth*, **122**, 4492–4508.
- Wapenaar, K., 2004. Retrieving the elastodynamic Green's function of an arbitrary inhomogeneous medium by cross correlation, *Phys. Rev. Lett.*, **93**, 254301.
- Wapenaar, K., Slob, E., Snieder, R. & Curtis, A., 2010. Tutorial on seismic interferometry: part 2 — Underlying theory and new advances, *Geophysics*, **75**, 75A211–75A227.
- Wessel, P., Smith, W.H.F., Scharroo, R., Luis, J. & Wobbe, F., 2013. Generic Mapping Tools: improved version released, *Eos Trans. Am. Geophys. Union*, **94**, 409–410.
- Wu, S.-M., Ward, K.M., Farrell, J., Lin, F.-C., Karplus, M. & Smith, R.B., 2017. Anatomy of Old Faithful from subsurface seismic imaging of the Yellowstone Upper Geyser Basin, *Geophys. Res. Lett.*, **44**, 10240–10247.
- Xu, Y., Koper, K.D. & Burlacu, R., 2017. Lakes as a source of short-period (0.5–2 s) microseisms: microseisms from lakes, *J. Geophys. Res. Solid Earth*, **122**, 8241–8256.

- Yang, Y., Ritzwoller, M.H., Levshin, A.L. & Shapiro, N.M., 2007. Ambient noise Rayleigh wave tomography across Europe, *Geophys. J. Int.*, **168**, 259–274.
- Yao, H., van der Hilst, R.D. & de Hoop, M.V., 2006. Surface-wave array tomography in SE Tibet from ambient seismic noise and two-station analysis - I. Phase velocity maps, *Geophys. J. Int.*, **166**, 732–744.
- Zeng, X., Lancelle, C., Thurber, C., Fratta, D., Wang, H., Lord, N., Chalari, A. & Clarke, A., 2017. Properties of noise cross-correlation functions obtained from a distributed acoustic sensing array at Garner Valley, California, *Bull. seism. Soc. Am.*, **107**, 603–610.
- Zheng, S., Sun, X., Song, X., Yang, Y. & Ritzwoller, M.H., 2008. Surface wave tomography of China from ambient seismic noise correlation, *Geochim. Geophys. Geosystems*, **9**, Q05020, doi:10.1029/2008GC001981.
- Zheng, Y., Shen, W., Zhou, L., Yang, Y., Xie, Z. & Ritzwoller, M.H., 2011. Crust and uppermost mantle beneath the North China Craton, northeastern China, and the Sea of Japan from ambient noise tomography, *J. geophys. Res.*, **116**, B12312, doi:10.1029/2011JB008637.
- Zigone, D., Ben-Zion, Y., Lehujeur, M., Campillo, M., Hillers, G. & Vernon, F.L., 2019. Imaging subsurface structures in the San Jacinto fault zone with high-frequency noise recorded by dense linear arrays, *Geophys. J. Int.*, **217**, 879–893.

## APPENDIX A: RAW SEISMOGRAMS AND FURTHER COMPARISON OF TEMPORAL NORMALIZATION METHODS

As shown in Fig. A1, raw seismograms of the Marathon data set commonly have short bursts of events. These large-amplitude short-duration (0.05 s) events may have negative effects on the noise cross-correlations, as they could dominate the resulting NCFs. Temporal normalization is usually used to reduce the detrimental effect of those events (Bensen *et al.* 2007). Alternatively, using short segments can also mitigate the negative effect (Prieto *et al.* 2011). For example, an experiment of suppressing large-amplitude earthquakes using short segment stacking is presented in Prieto *et al.* (2011).

We tested two commonly used temporal normalization methods, that is one-bit normalization and running absolute mean normalization (RAM), against the short segment stacking method. Two segment lengths, 60 and 20 s, are used by the short segment stacking, and one-bit and RAM normalization used the length of 60 s. The spectral smoothing length used in RAM was set to 0.05 Hz.

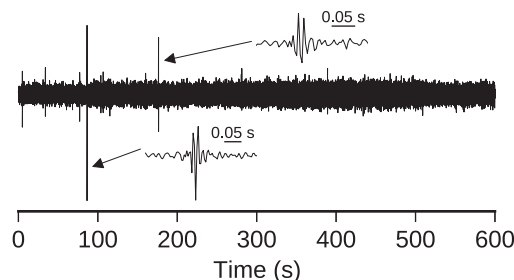
First, we evaluate the effect of these large-amplitude events on the NCFs by examining the pre-stack NCFs without temporal normalization (Fig. A2). For short segment stacking (Figs A2b and c), no general correlation is found between the events and the low-quality NCFs. Although the quality of the NCFs seems to drop at 180 s when two events also appear, the strong events at 90 and 120 s do not cause a strong decrease in quality. Additionally, the quality drop at 30 s is not associated with any strong glitches.

Then, we investigate the effect of temporal normalization. The pre-stack NCFs show that the RAM method is as good as the short segment stacking (60 s) in retrieving surface waves and performs better at 180 s (Figs A2b and e), whereas one-bit normalization, surprisingly, produces generally worse NCFs (Fig. A2d). The SNR of the stacked NCFs further confirms the observation (Fig. A3). RAM (SNR = 6.37) and short segment stacking with 60-s segments (6.64) are equally good, two times better than the one-bit normalization (3.94). The short segment stacking (20 s) outperforms all the other methods with the highest SNR of 7.70.

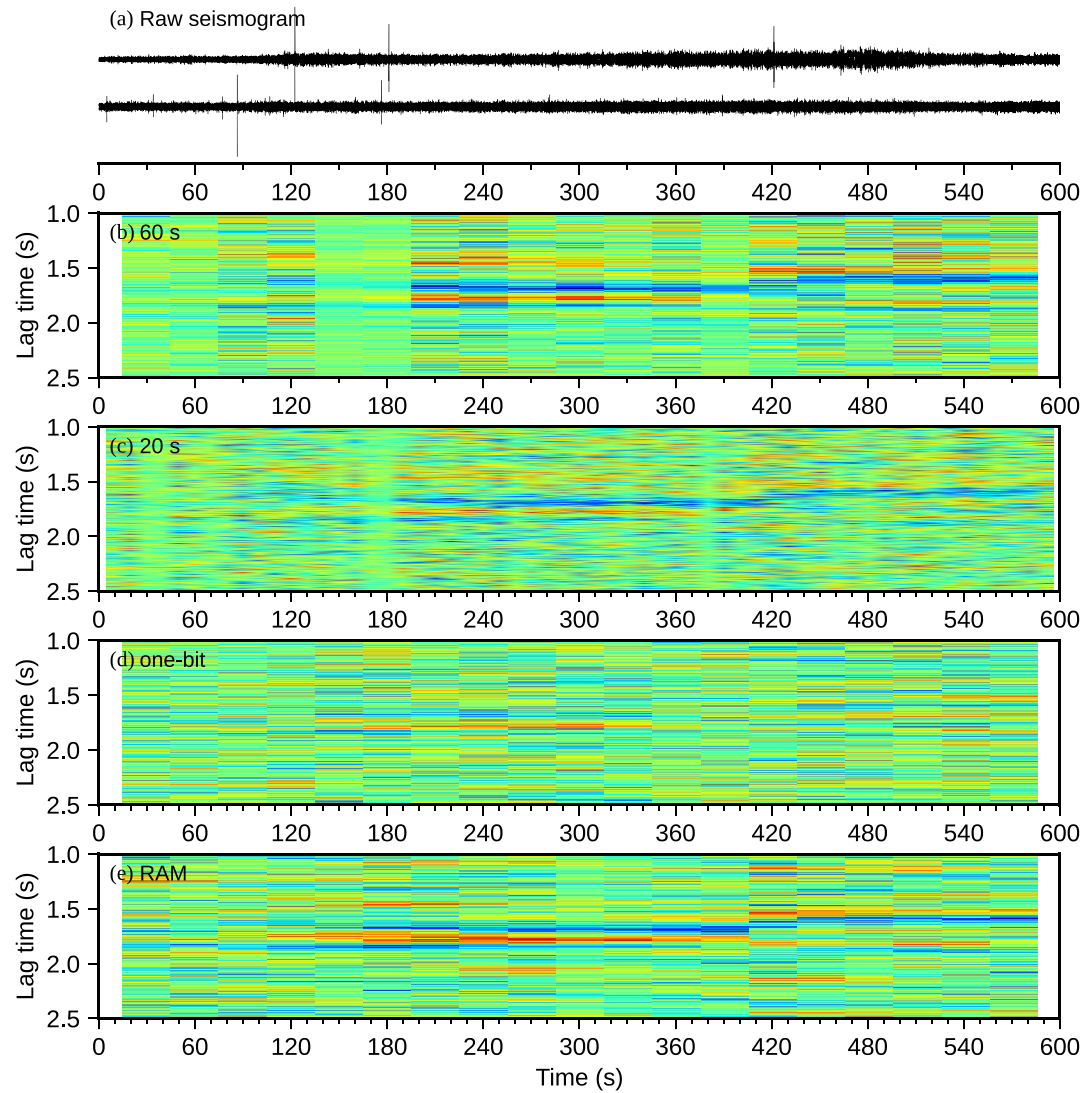
As we discussed in Section 3.3, the one-bit normalization works best with noise with a Gaussian distribution. Otherwise, it may even decrease the quality of the NCFs. The RAM method is more flexible, which can either degenerate to one-bit normalization with a smoothing length of 1 or mimics a peak amplitude normalization with a large smoothing length. RAM might be able to give a better NCF than short segment stacking; however, it may involve a considerable amount of work to tune the smoothing length. In contrast, short segment stacking is rather simple and robust, and the quality of NCFs can be easily improved by using shorter segments.

## APPENDIX B: INCOMPLETE COVERAGE OVER SHORT DISTANCE

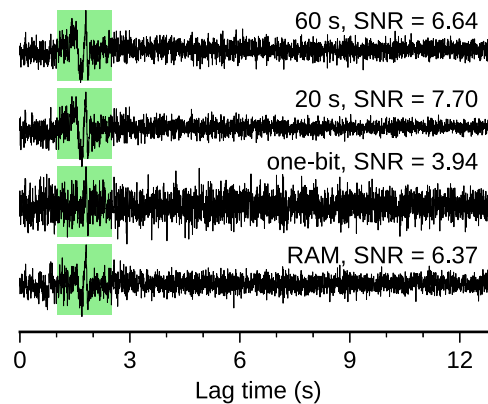
A good distance coverage is crucial for obtaining a robust estimation of the source phase (Martins *et al.* 2019). Because the station interval is 150 m for the data set we used, the distance coverage is incomplete for distance ranges close to 150 m, for example 0–500 m. Fig. B1 shows the distribution of interstation distance in 0–500 m (for high frequency (>20 Hz), NCFs with farther distances generally cannot be used because of their low quality due to scattering). We can see in Fig. B1(b) that (1) the distance distribution with 2 m bin width is sparse; (2) the number of station pairs in each bin is low (mostly at 3). These two problems reduce the robustness of the fitting, and thus the estimation of the source phase.



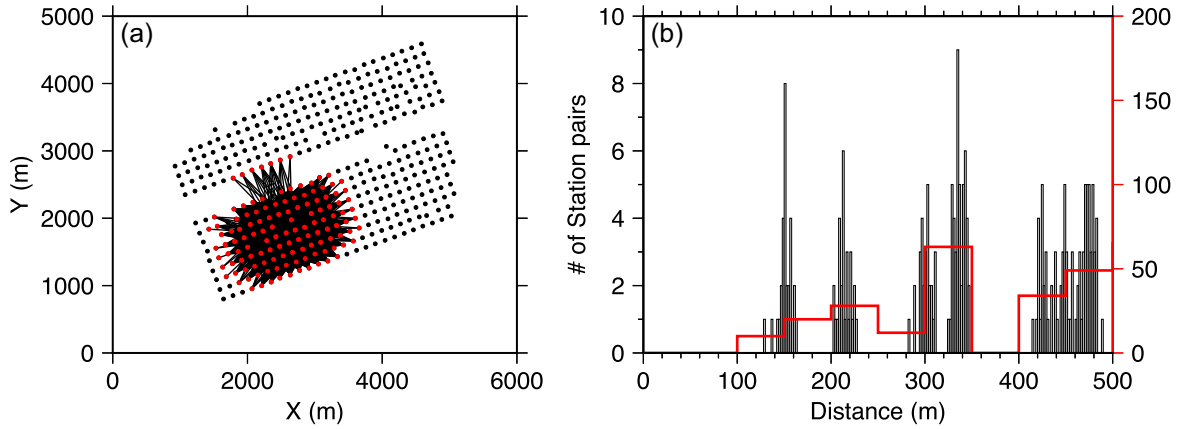
**Figure A1.** Raw seismogram at the station 1011.1030. The 10-min segment is cut after 2018–10–01T04:00:00, which also is used in the experiments in Figs 3 and 4.



**Figure A2.** (a) Raw seismograms at station 1011.1030 (top panel) and 1026.1111 (bottom panel). (b) NCFs computed using short segment stacking with a length of 60 s; (c) short segment stacking with a length of 20 s; (d) one-bit normalization; (e) running absolute mean normalization.



**Figure A3.** Stacked NCFs using a segment length of 60 s and 20 s without normalization, a segment length of 60 s with one-bit normalization and RAM normalization. The signal window used for computing SNR is marked green.

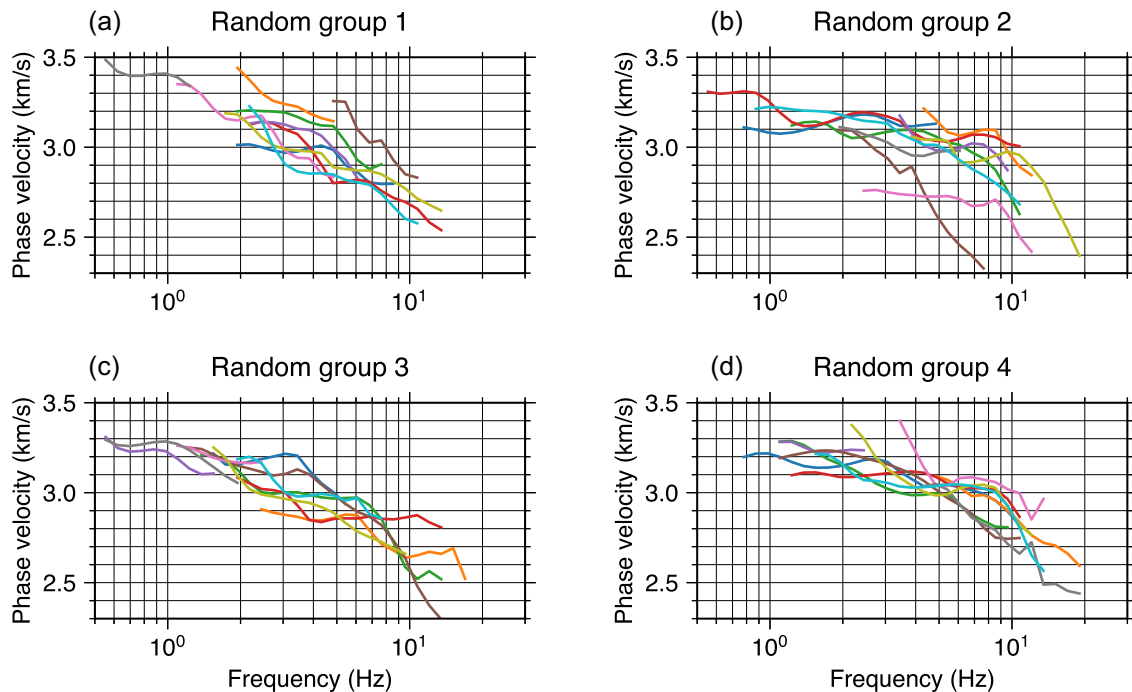


**Figure B1.** Station pairs selected for source phase estimation and the distance distribution. (a) Black dots denote the stations, and the red dots denote the station selected. Black lines indicate the station pairs available. (b) Histogram of interstation distance with the bin width of 2 m (black) and 50 m (red).

Using a wider bin (for instance, 50 m) could increase the number of stacked traces in each bin, but the spatial resolution reduces greatly, which also weakens the accuracy of the fitting (Fig. B1b). Alternatively, selecting station pairs over a wide area could also increase the number of traces without decreasing the spatial resolution. However, it may violate the underlying assumption of the source phase estimation method, that is that the phase velocity is constant among the station pairs due to spatial heterogeneity.

### APPENDIX C: EXAMPLES OF DISPERSION CURVES RANDOMLY SELECTED

We examined the quality of individual dispersion curves after quality control by checking dispersion curves selected randomly (Fig. C1). Although most of the dispersion curves follow a similar trend, some of them are bumpy. The degree of bumpiness is commonly observed in phase velocity dispersion curves (e.g. Soomro *et al.* 2016, Bonadio *et al.* 2018), which can be reduced by averaging a few dispersion curves with similar paths or using a relatively large smoothing factor during the tomographic inversion.



**Figure C1.** Dispersion curves randomly selected. Each group contains 10 dispersion curves randomly selected from the dispersion curves after quality control.