Geophysical Journal International

Geophys. J. Int. (2023) **232,** 1600–1618
Advance Access publication 2022 October 18
GJI Seismology

https://doi.org/10.1093/gji/ggac410

Optimal stacking of noise cross-correlation functions

Xiaotao Yang [1,2] Jared Bryan [3] Kurama Okubo [1,*] Chengxin Jiang,[1,†] Timothy Clements[1,§] and Marine A. Denolle[1,¶]

[1]*Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA* 02138*, USA. E-mail: xtyang@purdue.edu*
[2]*Department of Earth, Atmospheric, and Planetary Sciences, Purdue University, West Lafayette, IN* 47907*, USA*
[3]*Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA* 02139*, USA*

Accepted 2022 October 14. Received 2022 October 5; in original form 2022 March 28

SUMMARY

Cross-correlations of ambient seismic noise are widely used for seismic velocity imaging, monitoring and ground motion analyses. A typical step in analysing noise cross-correlation functions (NCFs) is stacking short-term NCFs over longer time periods to increase the signal quality. Spurious NCFs could contaminate the stack, degrade its quality and limit its use. Many methods have been developed to improve the stacking of coherent waveforms, including earthquake waveforms, receiver functions and NCFs. This study systematically evaluates and compares the performance of eight stacking methods, including arithmetic mean or linear stacking, robust stacking, selective stacking, cluster stacking, phase-weighted stacking, time–frequency phase-weighted stacking, N^{th}-root stacking and averaging after applying an adaptive covariance filter. Our results demonstrate that, in most cases, all methods can retrieve clear ballistic or first arrivals. However, they yield significant differences in preserving the phase and amplitude information. This study provides a practical guide for choosing the optimal stacking method for specific research applications in ambient noise seismology. We evaluate the performance using multiple onshore and offshore seismic arrays in the Pacific Northwest region. We compare these stacking methods for NCFs calculated from raw ambient noise (referred to as Raw NCFs) and from ambient noise normalized using a one-bit clipping time normalization method (referred to as One-bit NCFs). We evaluate six metrics, including signal-to-noise ratios, phase dispersion images, convergence rate, temporal changes in the ballistic and coda waves, relative amplitude decays with distance and computational time. We show that robust stacking is the best choice for all applications (velocity tomography, monitoring and attenuation studies) using Raw NCFs. For applications using One-bit NCFs, all methods but phase-weighted and N^{th}-root stacking are good choices for seismic velocity tomography. Linear, robust and selective stacking methods are all equally appropriate choices when using One-bit NCFs for monitoring applications. For applications relying on accurate relative amplitudes, the linear, robust, selective and cluster stacking methods all perform well with One-bit NCFs. The evaluations in this study can be generalized to a broad range of time-series analysis that utilizes data coherence to perform ensemble stacking. Another contribution of this study is the accompanying open-source software package, *StackMaster*, which can be used for general purposes of time-series stacking.

Key words: Time-series analysis; Coda waves; Earthquake ground motions; Seismic interferometry; Seismic noise; Seismic tomography; Ambient noise cross-correlations.

1 INTRODUCTION

Cross-correlations of ambient seismic noise have been widely used to image the Earth's elastic (Shapiro *et al.* 2005; Yang & Gao 2018; Feng & Ritzwoller 2019; Yang & Gao 2020; Wu *et al.* 2021) and anelastic structure (e.g. Prieto *et al.* 2009; Liu *et al.* 2021a), model ground motions (Denolle *et al.* 2013, 2014; Viens *et al.* 2017; Kwak

*Now at: National Research Institute for Earth Science and Disaster Resilience, Tsukuba, Ibaraki 305-0006, Japan.
†Now at: Research School of Earth Sciences, Australian National University, Acton, ACT 0200, Australia.
§Now at: Earthquake Science Center, United States Geological Survey, Moffet Field, CA 94035, USA
¶Now at: Department of Earth and Space Sciences, University of Washington, Seattle, WA 98195, USA

et al. 2017; Denolle *et al.* 2018; Viens & Denolle 2019) and monitor transient velocity changes in the shallow subsurface (Brenguier *et al.* 2008a; Wang *et al.* 2017; Clements & Denolle 2018; Donaldson *et al.* 2019; Olivier *et al.* 2019; Feng *et al.* 2021; Yang *et al.* 2022c). The noise cross-correlation functions (NCFs) are typically computed in short-time windows (such as hours or days) after diverse pre-processing on the raw ambient noise waveforms (Bensen *et al.* 2007; Feng & Ritzwoller 2019). Stacking of NCFs over a longer period (such as weeks, months or years) is a common procedure for most applications utilizing NCFs (e.g. Seats *et al.* 2012). For seismic tomography, the final stack over the entire time period is used to make waveform or traveltime measurements. Seismic monitoring requires stacking over a subset of time windows, balancing the temporal resolution and the coherence of the NCFs (Hadziioannou *et al.* 2011). These applications rely on the phase information of ballistic (first arrival) and/or coda waves. Studies of ground motion prediction and attenuation tomography, on the other hand, require accurate relative amplitude measurements between station pairs. Because NCFs consist primarily of dispersive surface waves, preserving the relative amplitude also preserves the spectral content. Therefore, the performance of stacking and the preservation of either phase or amplitude information are important to ensure the robustness of the scientific results.

The most commonly used stacking method is the arithmetic mean of the data over temporal or spatial samples, referred to as Linear stacking. For NCFs, substantial temporal variation may exist across individual short time windows, which could result from the seasonal change of the location and strength of the microseismic sources (Bensen *et al.* 2007; Ermert *et al.* 2016, 2021; Tian & Ritzwoller 2017; Yang *et al.* 2019; Li *et al.* 2020; Liu *et al.* 2021b), disruption from tectonic signals (Baig *et al.* 2009) or unknown instrumental failures. It is important to use an appropriate stacking method to extract coherent NCFs, particularly when strong variations of the data quality are present. Many more advanced methods have been developed to improve the stacking of NCFs, and time-series data in general and overcome the limitations of Linear stacking. Rückemann (2012) provides a summary of the theoretical background of several stacking methods to improve the common-mid-point gathers in seismic reflection data. A systematical evaluation of the performance of different stacking methods in processing real NCF data is highly desirable and required for optimizing processing strategies.

In this paper, we compare eight algorithms used for stacking NCFs and evaluate their performance for canonical research applications. These methods include: robust stacking (Pavlis & Vernon 2010), selective stacking (modified from Liu *et al.* 2009; Olivier *et al.* 2015), cluster stacking (a new method), Phase-Weighted Stacking (PWS; Schimmel & Paulssen 1997), time–frequency Phase-Weighted Stacking (tf-PWS; Schimmel & Gallart 2007; Baig *et al.* 2009; Schimmel *et al.* 2011; Thurber *et al.* 2014; Zeng & Thurber 2016; Li *et al.* 2017), N^{th}-root stacking (Rückemann 2012; Millet *et al.* 2019) and linear stacking after applying an adaptive covariance filter (ACF; Nakata *et al.* 2015). We exclude the stacking methods that use the curvelet transform (Stehly *et al.* 2011) and singular value decomposition (Moreau *et al.* 2017), as these focus specifically on short-term convergence alone. We assess the performance of the eight algorithms using six different metrics relevant to specific research applications. Because the one-bit pre-processing is designed to reduce the influence of earthquake-like transient signals and to improve the quality of individual NCFs (Bensen *et al.* 2007), we benchmark our tests using both the raw and one-bit normalization pre-processing techniques. The development of new methods and computer codes and the performance tests from this study aim to guide the community in choosing the appropriate stacking method for ambient-noise seismology use cases, though they can also be generalized to process other time-series data.

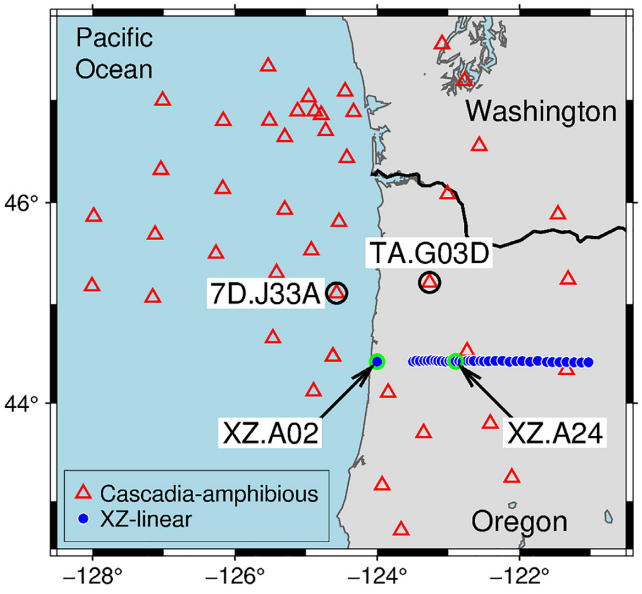

Figure 1. Seismic stations used in this study. The USArray Transportable Array and Cascadia Initiative Ocean Bottom Seismographs (red open triangles) form a composite data set that is referred to as the 'Amphibious' data set. The 1993–1994 Cascadia broad-band XZ linear array (blue dots) is referred to as the 'XZ' data set. Labelled stations are used as virtual sources and receivers in the examples in this paper.

2 AMBIENT NOISE CROSS-CORRELATIONS

This study uses the Pacific Northwest as a natural laboratory (Fig. 1). We gather data from 29 Cascadia Initiative Ocean Bottom Seismographs (OBS; network code: 7D; IRIS OBSIP 2011), 15 US Transportable Array stations (network code: TA; IRIS Transportable Array 2003) and 43 stations from the 1993 to 1994 Cascadia subduction zone experiment (network code: XZ; Nabelek *et al.* 1993). We group the data into two data sets: (1) the composite data set from the Cascadia Initiative OBSs and onshore TA stations that is referred to as the 'Amphibious' data set; (2) the data from the east–west XZ linear array in northern Oregon that is referred to as the 'XZ' data set. All data are publicly accessible from the seismic data archive of the Incorporate Research Institutions for Seismology Data Management Center (IRIS DMC).

We use the *SeisGo* Python toolbox to download the continuous seismic waveforms and compute the NCFs (Yang *et al.* 2022b). The cross-correlation function in *SeisGo* was modified from *NoisePy* (Jiang & Denolle 2020). We download the vertical component of the continuous waveforms in 12-hr segments. We select data from 10/1/2011 to 7/31/2012 for the Amphibious data set and from 6/1/1993 to 5/31/1994 for the XZ data set. We remove the instrument responses and convert the waveforms to displacements, followed by downsampling to 5 Hz. We compute the NCFs in 6-hr windows sliding every 3 hr. The data are demeaned and detrended for each 6-hr segment. We attempt to remove transient signals in the ambient noise that will contaminate the noise cross-correlations by removing the windows with anomalous amplitudes. For that, we calculate the maximum absolute raw noise amplitude of the segment and

the standard deviation of the raw noise amplitudes of all segments. We discard waveform segments with peak amplitudes greater than 10 times the standard deviation, as implemented in *NoisePy*.

We construct two sets of cross-correlations using the raw waveforms (Viens *et al.* 2017) and waveforms normalized in the time domain using the `sign` function (e.g. Bensen *et al.* 2007; Shen *et al.* 2012). We denote the two data sets as Raw NCFs and One-bit NCFs, respectively. To avoid spectral leakage, the short 6-hr time windows are tapered using a Tukey window with a cosine fraction of 0.05. We then apply a Fourier transform to the short-window segment and compute the cross-correlation in the frequency domain using the following equation:

$$X = \mathcal{F}^{-1}\left(\mathcal{F}^{*}(d_s)\mathcal{F}(d_r)\right), \tag{1}$$

where X is the NCF between time-series d_s (virtual source) and d_r (receiver), $\mathcal{F}$ is the Fourier transform operator, $\mathcal{F}^{*}$ stands for the complex conjugate of the Fourier transform and $\mathcal{F}^{-1}$ is the inverse Fourier transform operator. For stacking analyses, we average the computed NCFs within each 12-hr window to reduce the computational cost.

We show examples of the computed NCFs (bandpass filtered at 0.1–0.4 Hz) in Fig. 2 for both pre-processing strategies. The asymmetry of the NCF is a known effect of noise source directionality (Stehly *et al.* 2006). The first-order observation is that the individual NCFs contain incoherent noise that may overwhelm the coherent signals. The NCFs from the XZ land station pair all show clear ballistic phases over the entire period range of this analysis (Figs 2c and d). For the amphibious station pair 7D.J33A-TA.G03D, clear ballistic phases are present between 50 and 100 s for most of the NCFs between October 2011 and April 2012, for both Raw (Fig. 2a) and One-bit (Fig. 2b) NCFs. However, the ballistic phases are almost invisible for the NCFs after May 2012. These 'noisy' NCFs may contaminate the linear stack of the NCFs. The simple time-domain representation in Fig. 2 only provides a qualitative visual assessment. We later quantify the performance using different stacking methods with multiple metrics.

3 STACKING METHODS

This section describes the algorithms beyond the arithmetic mean (or Linear stacking) for optimally stacking NCFs. Most of these methods are adapted from the published literature. All methods can be generalized as weighted stacking with various ways of computing the stacking weight, which could be either trace weight or sample weight. We express the weighted stacking scheme as:

$$\mathbf{b} = \sum_{i=1}^{N} w_i \mathbf{d}_i, \tag{2}$$

where $\mathbf{b}$ is the final stack. For trace-weighted stacking, $\mathbf{d}_i$ is the ith NCF trace, w_i is the stacking weight for the ith NCF trace satisfying $\sum w_i = 1$ and N is the total number of NCF traces. For sample-weighted stacking, $\mathbf{d}_i$ is the ith sample of the arithmetic average or linear stack of all NCF traces, w_i is the stacking weight for the ith sample satisfying $\sum w_i = 1$ and N is the total number of samples in each NCF trace. Under the stacking formulation of eq. (2), Linear stacking is a special case of trace-weighted stacking where all traces have the same weight ($w_i = 1$). In the methods described in this section, the Robust (Section 3.1), Selective (Section 3.2) and Cluster (Section 3.3) stacking methods are all trace-weighted stacking methods. The phase-weighted (Section 3.4), time–frequency phase-weighted (Section 3.5), N^{th} -root (Section 3.6) and adaptive covariance filter (Section 3.7) stacking methods are all sample-weighted stacking methods. In the following subsections, we describe the computation of the stacking weights for each method in detail.

3.1 Robust stacking

Pavlis & Vernon (2010) first introduced Robust stacking to improve the accuracy of picking phase arrival times in seismograms from distant earthquakes. We adapt this generic, iterative, weighted stacking method to process NCFs. The weight in Robust stacking is based on waveform decoherence and penalizes the NCF that deviates too much from a reference stack. The weight w_i of the ith NCF, $\mathbf{d}_i$, for the jth iteration is calculated as (Pavlis & Vernon 2010):

$$w_i^j = \frac{|\mathbf{b}_j \cdot \mathbf{d}_i|}{||\mathbf{d}_i||||\mathbf{r}_i||}, \tag{3}$$

where $\mathbf{b}_j$ is the reference stack of the jth iteration, $|*|$ is the L1-norm operator, $||*||$ is the L2-norm operator and $\mathbf{r}_i = \mathbf{b}_j - (\mathbf{b}_j \cdot \mathbf{d}_i)\mathbf{d}_i$ is a modified residual quantifying the difference between the current reference stack $\mathbf{b}_j$ and the ith NCF scaled by the dot product of the two time-series. The weight w_i penalizes twice the data that does not resemble the reference stack. The first element is the dot product between the NCF and the stack: $\mathbf{b}_j \cdot \mathbf{d}_i/||\mathbf{d}_i||$. The second is the modified residual r_i: if it is high, the $1/||r_i||$ is small and w_i is low. The weights can be calculated over either the entire NCF or a specific time window that contains the seismic phase of interest. For the examples in this paper, we compute the weights over the entire NCF, though our software package allows the user to specify the time window. After computing the weights for all NCFs, we normalize them to their sum and use them to produce the updated stack.

The initial reference stack $\mathbf{b}_1$ is chosen as the median NCF, following the observation of Pavlis & Vernon (2010) that the median is more representative of the concentration of data in the presence of outliers. In our implementation, the users may specify another trace as the initial stack, such as the Linear stack. The iterative procedure stops when a convergence criterion is satisfied. We adopt the measure proposed by Pavlis & Vernon (2010):

$$\frac{|\mathbf{b}_j - \mathbf{b}_{j-1}|}{||\mathbf{d}_i||M} < \epsilon, \tag{4}$$

where M is the number of samples in each NCF time-series and ϵ is a small number, with a default value of 10^{-5}.

3.2 Selective stacking

In the presence of strong noise, it could be effective to stack only a subset of the NCF ensemble that exceeds a quality threshold and to ignore the low-quality NCFs. This concept, called selective stacking, has been implemented in previous studies (e.g. Olivier *et al.* 2015; Thangraj & Pulliam 2021). The signal-to-noise ratio (SNR) or the correlation to the reference can be used to evaluate the quality of each NCF. Olivier *et al.* (2015) used the SNR around the expected S-wave arrival time to reconstruct S waves in the stacked NCF. Liu *et al.* (2009) proposed the weighted stack of common-midpoint gather using the local correlation within the moving window to improve the SNR and suppress the random noise in the stacked trace. To enhance the coherence of the signal, we use a criterion based on the similarity, quantified as the Pearson correlation coefficient in our implementation, between each NCF and the reference NCF. In

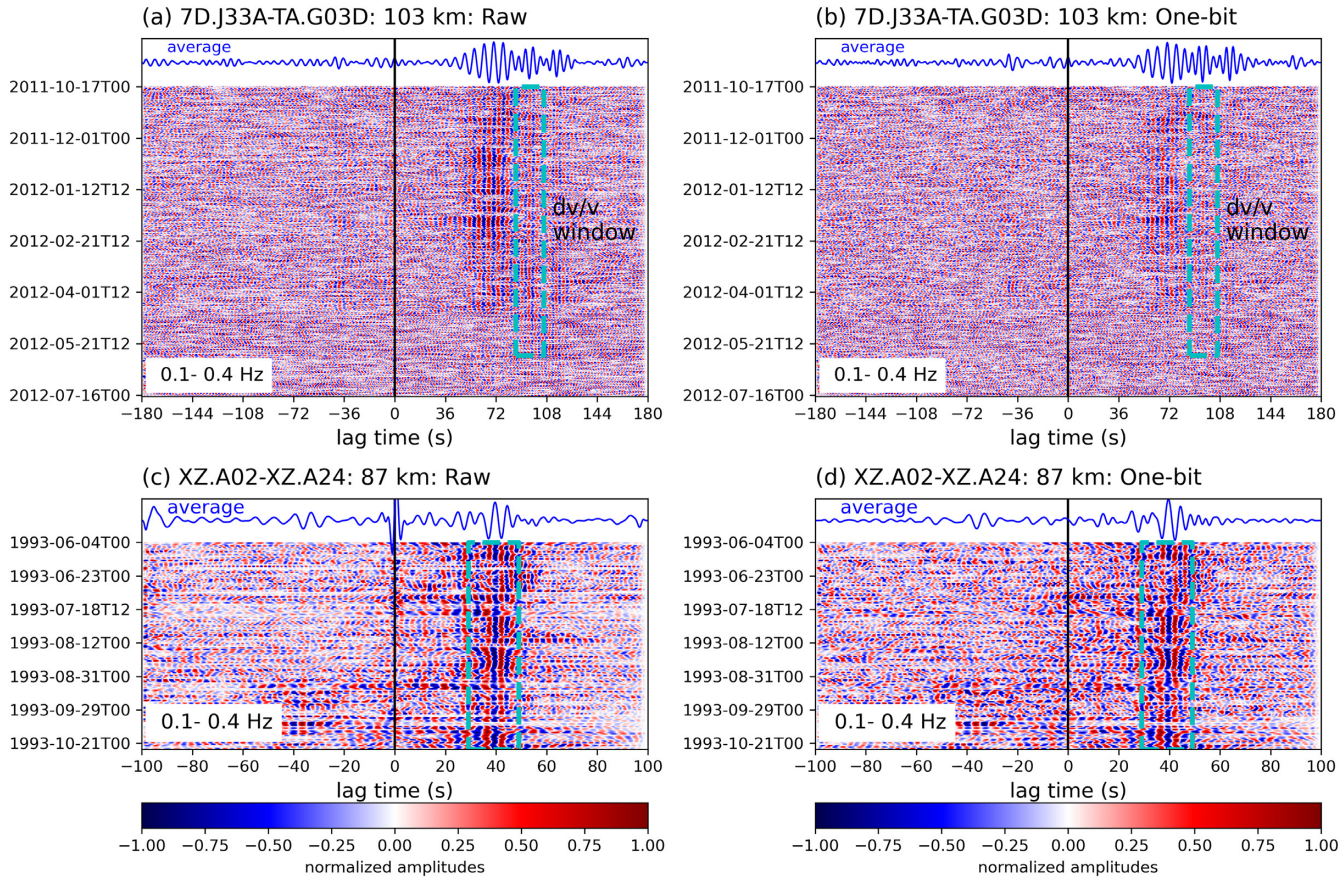

Figure 2. Examples of ambient noise cross-correlation functions (NCFs) between (a, b) OBS station 7D.J33A and onshore station TA.G03D and (c, d) XZ linear array stations XZ.A02 and XZ.A24. The NCFs are filtered at 0.1–0.4 Hz. The left-hand panels are NCFs computed with raw waveforms and the right-hand panels are NCFs computed after applying one-bit time-domain normalization. The dashed cyan boxes mark the windows used for the analysis of the preservation of transient phase changes (Section 4.4) using (a, b) coda waves and (c, d) ballistic phases. On the top of each panel, we show the linear stack as arithmetic averages. The NCFs are colour-coded by the normalized amplitudes (red is 1, blue is –1).

our implementation, by default, the reference NCF is initialized as the Linear stack of the entire ensemble (similar to Liu *et al.* 2009). As in Robust stacking, the user can specify any trace, such as the median NCF, as the reference stack and the window used to estimate the similarity. Selective stacking sets the weights of low-similarity NCFs to zero, computed as:

$$w_i = \begin{cases} 0, & \rho_{\mathbf{b},\mathbf{d}_i} \leq \rho_t \\ 1, & \rho_{\mathbf{b},\mathbf{d}_i} > \rho_t \end{cases}, \tag{5}$$

where $\rho_{\mathbf{b},\mathbf{d}_i}$ is the Pearson correlation coefficient between the individual trace and the reference NCF and ρ_t is the user-defined threshold. Note that the criterion used in this study is equivalent to the global correlation described in Liu *et al.* (2009) rather than the local correlation for simplicity in thresholding. A similarity threshold allows us to discriminate the set of NCFs to keep or discard (weight is set to zero). We then compute the Linear stack over the subset of NCFs with a Pearson correlation coefficient above the threshold and discard the NCFs below the threshold. Similar to robust stacking, we implement the selective stacking method in an iterative scheme, with the same convergence criterion as in eq. (4). In our examples, we use a correlation coefficient threshold of 0.

3.3 Cluster stacking

An alternative method to ranking waveforms according to a specific similarity metric is to group them into clusters and perform the stack for each cluster. Viens & Iwata (2020) proposed the idea of clustering the NCFs using a dimensionality reduction technique (principal component analysis), as exploited by Toghramadjian *et al.* (2021). In our implementation, we use the *Tslearn* Python toolkit (Tavenard *et al.* 2020) for k-means clustering of the NCF waveforms based on the Euclidean distance. We impose two clusters to separate NCFs with higher quality from those with lower quality.

After clustering the NCFs into two clusters, the final stack **b** is computed as a weighted stack of the two cluster centres

$$\mathbf{b} = \sum_{i=1,2} w_i C_i, \tag{6}$$

where C_i ($i = 1, 2$) are the centres (sample averages or linear stacks) of the two clusters and w_i ($i = 1, 2$) are the weights of the cluster centres. The centre stacking weights, w_i ($i = 1, 2$), are computed as

$$w_{i=1,2} = \begin{cases} \begin{cases} 0, & p_i < \max(p_1, p_2) \\ 1, & p_i >= \max(p_1, p_2) \end{cases}, & cc < h \\ p_i/(p_1 + p_2), & cc \geq h \end{cases}, \tag{7}$$

where cc is the correlation coefficient between the two cluster centres C_1 and C_2, h is the similarity threshold specified by the user,

and p_i ($i = 1, 2$) is the relative peak amplitude of the cluster centre, computed as the ratio of the maximum and root-mean-square of the absolute amplitudes. The relative peak amplitude we use here is an equivalent measure to the SNR of the cluster centre. In our examples, p_1 and p_2 are computed over the entire NCF. Our implementation also allows the user to specify the window of interest to compute the maximum of the absolute amplitudes. According to eq. (7), the two centres are weighted by the relative peak amplitude when their correlation coefficient exceeds the threshold ($h = 0.75$ for our examples). Otherwise, we use the centre with a higher peak amplitude as the final stack.

3.4 Phase-weighted stacking

Phase-weighted stacking (PWS) computes the weight for each trace by the coherency of the instantaneous phases (Schimmel & Paulssen 1997). Our implementation of the PWS method follows the description in Schimmel & Paulssen (1997). Readers are referred to Schimmel & Paulssen (1997) for detailed equations and procedures of the PWS method. The final stack is constructed as a non-linear weighted stack where each sample is weighted by the instantaneous phase. The key parameter for PWS is a harshness parameter that quantifies the fall-off of the weight of each NCF with decreasing similarity. We set the harshness to 2, as in Schimmel & Paulssen (1997).

3.5 Time–frequency phase-weighted stacking

Schimmel & Gallart (2007) and Schimmel *et al.* (2011) proposed an improved phase-weighted stacking method that computes the weight in both the time and frequency domains. This method, referred to as tf-PWS, projects each seismic trace into the time–frequency domain through the Stockwell transform (S-transform; Stockwell *et al.* 1996), which can be thought of as a short-time Fourier transform with a frequency-dependent time window. Baig *et al.* (2009) adapted the stacking method based on a more efficient discrete orthogonal S-transform (DOST) that is introduced by Stockwell (2007). Readers are referred to Schimmel & Gallart (2007) and Schimmel *et al.* (2011) for detailed formulation of the tf-PWS stacking method. We implement tf-PWS with both the original *S*-transform and the DOST. The implementation of DOST-based tf-PWS in this study follows the description by Schimmel *et al.* (2011) using a Python translation of the implementation of DOST by Battisti & Riba (2016). The two implementations produce comparable stacking results overall (Fig. S1), though the tf-PWS method with the original *S*-transform takes more than 10^3 times longer CPU time. Therefore, considering the significantly lower efficiency using the original *S*-transform, the comparison and evaluation in this study use the DOST-based tf-PWS. Similar to PWS, the tf-PWS is primarily controlled by a harshness parameter that determines the sharpness of the transition between phase similarity and dissimilarity. We set the harshness to 2 for our examples, as in Schimmel *et al.* (2011).

3.6 N^{th}-root stacking

*N*th-root stacking method is commonly used in array seismology (Muirhead 1968; Kanasewich *et al.* 1973; Schimmel & Paulssen 1997; Rost & Thomas 2002; Rückemann 2012; Millet *et al.* 2019). It is a non-linear stacking method that takes the N^{th}-root of the absolute amplitudes of each individual trace and sums them together. The summation is then raised to the power of N to assign the sample weight. The polarity of each sample is recovered with a sign function of the summation. Our implementation follows the description in Millet *et al.* (2019; their eq. 18) and can be summarized as:

$$\mathbf{b} = \text{sign}(\mathbf{r})|\mathbf{r}|^n, \tag{8}$$

where $\mathbf{b}$ is the final stack, n is the specified order of root, $\mathbf{r} = \frac{1}{N}\sum_{i=1}^{N} \text{sign}(\mathbf{d}_i)|\mathbf{d}_i|^{1/n}$, $\mathbf{d}_i$ is the ith NCF trace, and N is the total number of NCF traces. We use the square root (i.e. $n = 2$) in our examples.

3.7 Adaptive covariance filter stacking

Nakata *et al.* (2015) introduced an adaptive covariance filter (ACF) to suppress incoherent noise in seismic data based on the adaptive polarization filter (Samson & Olson 1981; Du *et al.* 2000). We implement the ACF stacking method following Nakata *et al.* (2015), which the readers are referred to for detailed formulations of the method. The final stack is the Linear stack after applying the ACF. We use 1 as the harshness of the filter.

4 EVALUATION AND COMPARISON OF STACKING ALGORITHMS

We implement the stacking algorithms described in Section 3 as a standalone Python package *StackMaster* (Yang *et al.* 2022a). The package is available as a repository on Python Package Index (PyPI). It can be installed with: `pip install stackmaster`. In this paper, we install *StackMaster* with the *SeisGo* toolbox (Yang *et al.* 2022b) under the same Anaconda environment. We apply the stacking techniques to the Amphibious and the XZ data sets. Fig. 3 shows the stacking results for the two station pairs in Fig. 2. Figs 4–6 and S2–S4 in the supplement are the stacking results of all station pairs from the virtual sources at 7D.J33A (Figs 4, 5, S2 and S3) and at XZ.A02 (Figs 6 and S4).

We then evaluate and compare the performance of these stacking methods. Recognizing that NCFs are used in multiple applications (e.g. seismic velocity tomography, monitoring and attenuation/ground motion analysis), we evaluate the performance of the algorithm over a range of metrics: (1) SNRs, (2) surface wave dispersion, (3) convergence of short-term stacks to the long-term stack, (4) transient phase changes, (5) peak amplitudes of ballistic phases and (6) computational expense. We compare the methods below according to each metric. It is worth noting that there are multiple parameters that can be tuned for each stacking method, though we only focus on the most commonly used or suggested parameters in this study. In Section 5, we discuss the choice of stacking methods for different categories of applications using NCFs.

4.1 SNRs

We use the SNR of the ballistic phase as a proxy for the quality of the NCF stack. In this study, we define the SNR, R, as

$$R = \left[\frac{\text{rms}(|A_{\text{signal}}|)}{\text{rms}(|A_{\text{noise}}|)}\right]^2, \tag{9}$$

where *rms*() is the root-mean-square operator, $|A_{\text{signal}}|$ are the absolute amplitudes within the signal window and $|A_{\text{noise}}|$ are the absolute amplitudes within the noise window. For the NCFs from amphibious station pairs (Figs 3a, b, 4 and S2), the signal window is defined by a fast wave travelling at 4.5 km s^{-1}, a slow wave travelling at 2 km s^{-1},

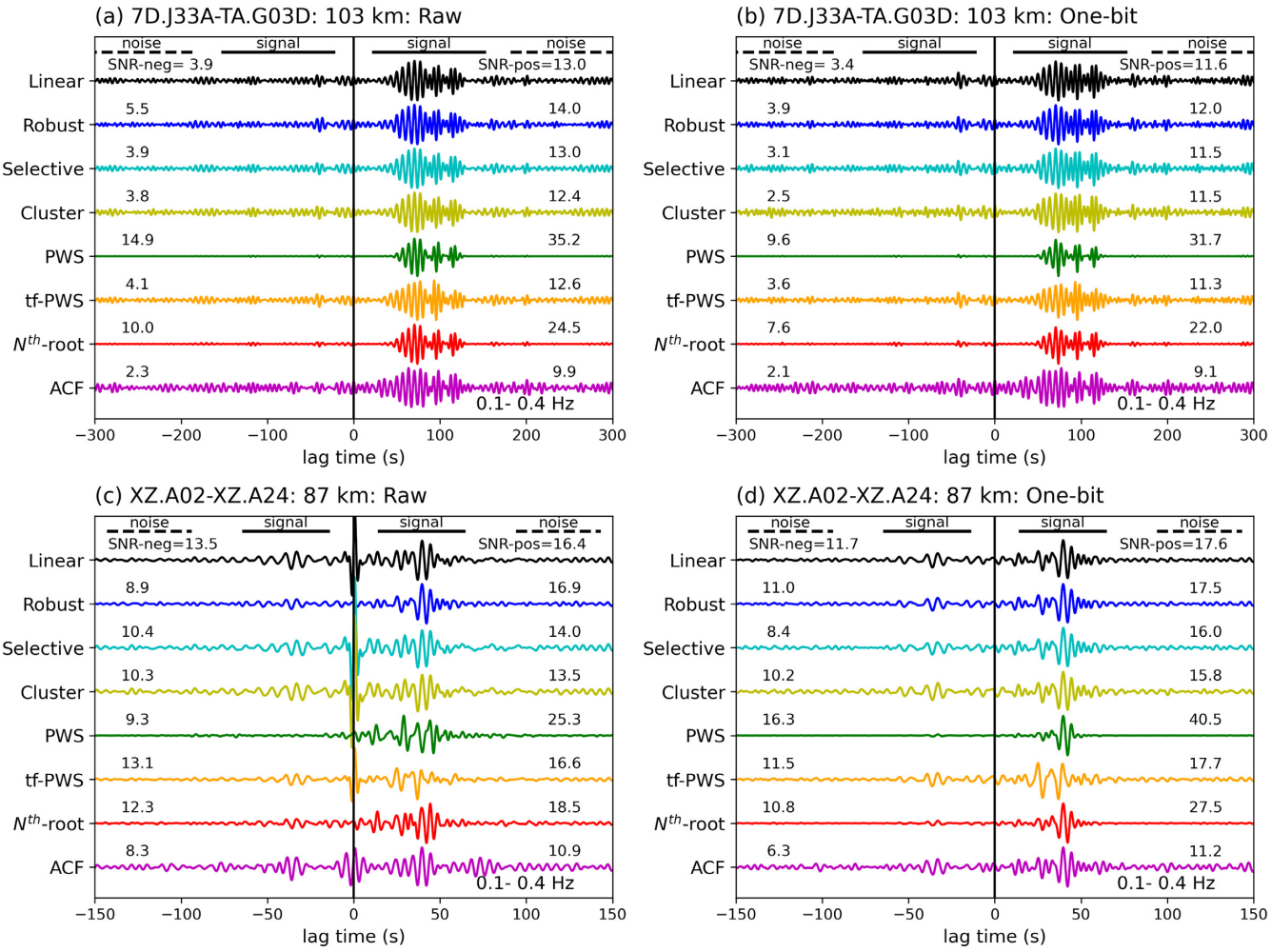

Figure 3. Examples of the stacked NCFs between station pairs (a, b) 7D.J33A-TA.G03D and (c, d) XZ.A02-XZ.A24, using different methods. The left-hand panels are the stacking results for Raw NCFs and the right panels are the stacking results for One-bit NCFs. All NCFs are filtered at 0.1–0.4 Hz after stacking. The stacking algorithm is labelled on the left of each panel. The SNR in decibel scale (eq. 10) is indicated above each trace on each side of the correlations. The thick solid and dashed lines mark the starts and ends of the signal and noise windows, respectively.

and we add 100 s to the end of the window. For OBS station pairs (Figs 5 and S3), we use a velocity range of 0.5–1.0 km s^{-1} to predict the signal window and add 60 s after the latest predicted arrival time. For the XZ data set (Figs 3c, d, 6 and S4), the signal window is defined by waves travelling between 2 and 3.7 km s^{-1}, extending for an additional 20 s. For both data sets, the noise window has the same length as the signal window with an offset of 30 s after the end of the signal window. Considering the relatively large dynamic range of the computed SNRs, we convert the ratios to decibel scales for all examples in this paper to assist the comparison and visualization, with

$$\mathrm{SNR} = 10 \log_{10} R. \tag{10}$$

It is worth noting that the SNR is unitless and the conversion in eq. (10) is mainly for scaling.

We observe notable variations in the SNRs of the ballistic phases retrieved using different stacking methods. From the single-pair examples (Fig. 3), the ballistic phases are most prominent on the positive lags from all stacking results, with relatively weaker phases on the negative lags, shown as lower SNRs overall. For the Cascadia amphibious station pair, the SNRs range from 9.9 to 35.2 for the Raw NCFs and 9.1–31.7 for the One-bit NCFs (Figs 3a and b). For the XZ station pair, the SNRs range from 10.9 to 25.3 for the Raw NCFs and 11.2-40.5 for the One-bit NCFs (Figs 3c and d). For all data sets, the highest SNRs are achieved with the PWS method, while the ACF method produces the lowest SNRs. The stacking results using the N^{th}-root method also show relatively high SNRs. The SNRs of other stacking results are at a comparable level.

The stacking results of all station pairs from the same virtual source provide a more holistic comparison of the performance of different methods (Figs 4–6 and S2–S4). We use the mean SNRs across all station pairs to quantify the comparison of stacks with different methods (Fig. 7). From the visual inspection of the moveout plots (time-lags versus interstation distances), the stacks of Raw (Figs 4 and 6) and One-bit (Figs S2 and S4) NCFs have comparable quality overall for both the Amphibious and the XZ data sets. This is also evident from the comparable mean SNR values of the two pre-processing methods (Fig. 7). For the Cascadia Amphibious data set, most of the NCFs from the OBS receivers are relatively noisy (Figs 5 and S3), with mean SNRs lower than those from the onshore receivers (Figs 7a and b). This may result from the contamination of tilt and compliance noise at most OBS stations, a well-known problem in offshore data (Tian & Ritzwoller 2017). Most of the

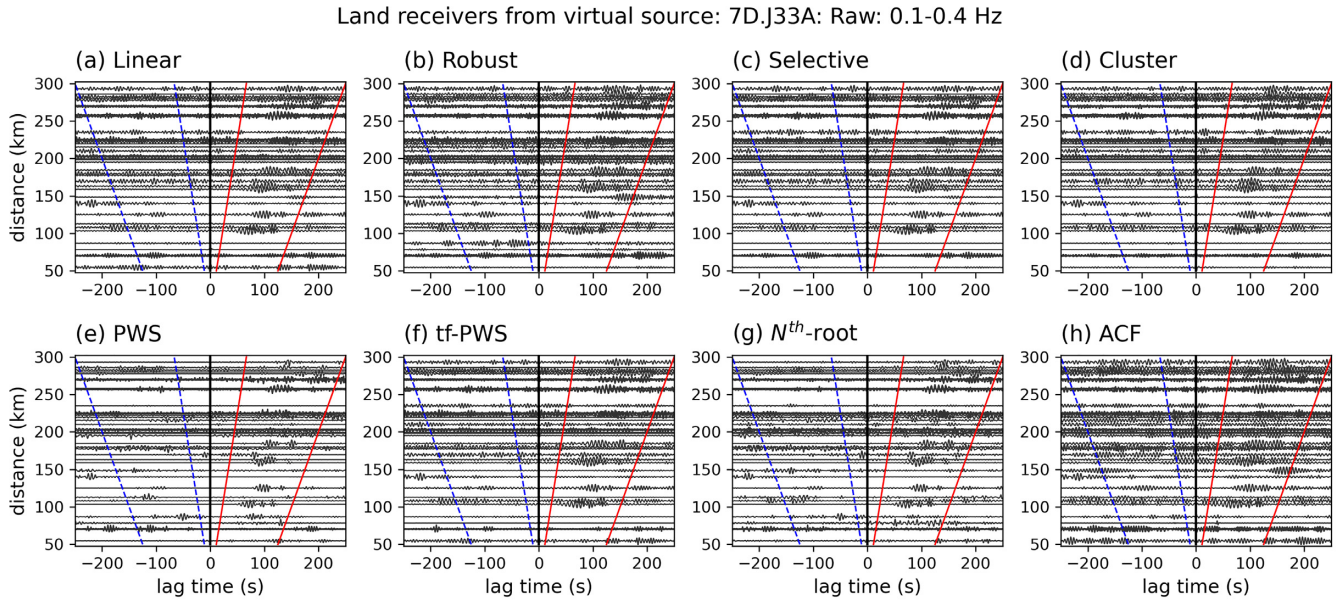

Figure 4. Stacked Raw NCFs of the Cascadia amphibious array from 7D.J33A to other land receivers using different stacking methods, filtered at 0.1–0.4 Hz. (a–h) The results using Linear, Robust, Selective, Cluster, PWS, tf-PWS, N^{th} -root and ACF stacking methods, respectively. The red solid lines and the blue dashed lines outline the positive-lag signal window and the negative-lag signal window, respectively, used to compute the SNR in Fig. 7. The signal and noise windows are determined with the same method as in Figs 3(a) and (b). See Fig. 1 for station locations.

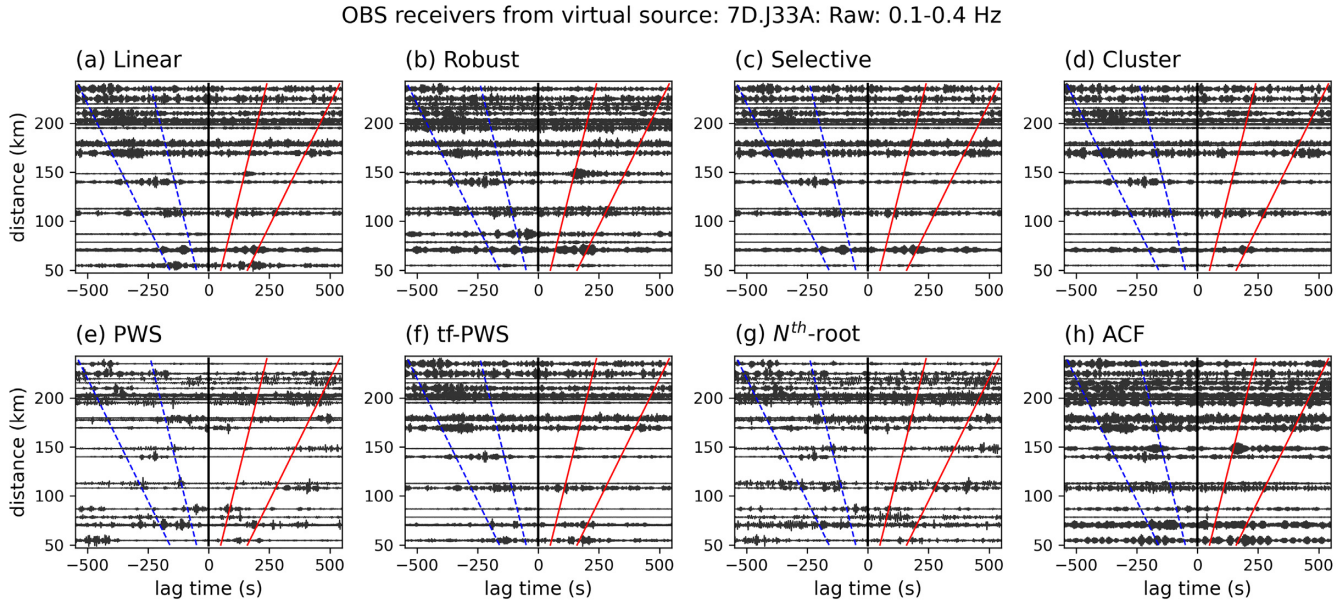

Figure 5. Same as Fig. 4 but for NCFs from 7D.J33A to other OBS receivers. To contain the visually identified ballistic phases from these OBS station pairs, we use a different velocity range (0.5–1.0 km s^{-1}) here to predict the signal window of the weakly coherent signals. We extend the window for an additional 60 s after the latest predicted arrival. See Fig. 1 for locations of the OBS receivers.

stacked XZ NCFs show clear ballistic phases (Figs 6 and S4), with much higher mean SNRs than those of the Amphibious data set (Fig. 7). In all examples with land receivers, the ballistic phases are dominantly visible at positive lag times, representing surface waves propagating away from the virtual sources, which we focus on for the following description. For all four data sets (land receivers only for the Amphibious NCFs), the PWS method produces the highest SNRs, while the standard deviation of the SNRs is also the largest. For the Cascadia Amphibious NCFs (Figs 7a and b), N^{th} -root stacking yields the second highest SNRs, while the SNRs of other stacking results are at a comparable level of 0–10 for both the Raw and One-bit NCFs. For the XZ NCFs (Figs 7c and d), the lowest SNRs are observed with ACF stacking, though the ballistic phases are still clearly retrieved (Figs 6h and S4h).

4.2 Surface wave dispersion

Surface waves are dispersive, which means that their wave speed depends on the frequency of the wave. Lower-frequency (longer period) surface waves generally travel faster than higher-frequency (shorter period) waves and are more sensitive to greater depths.

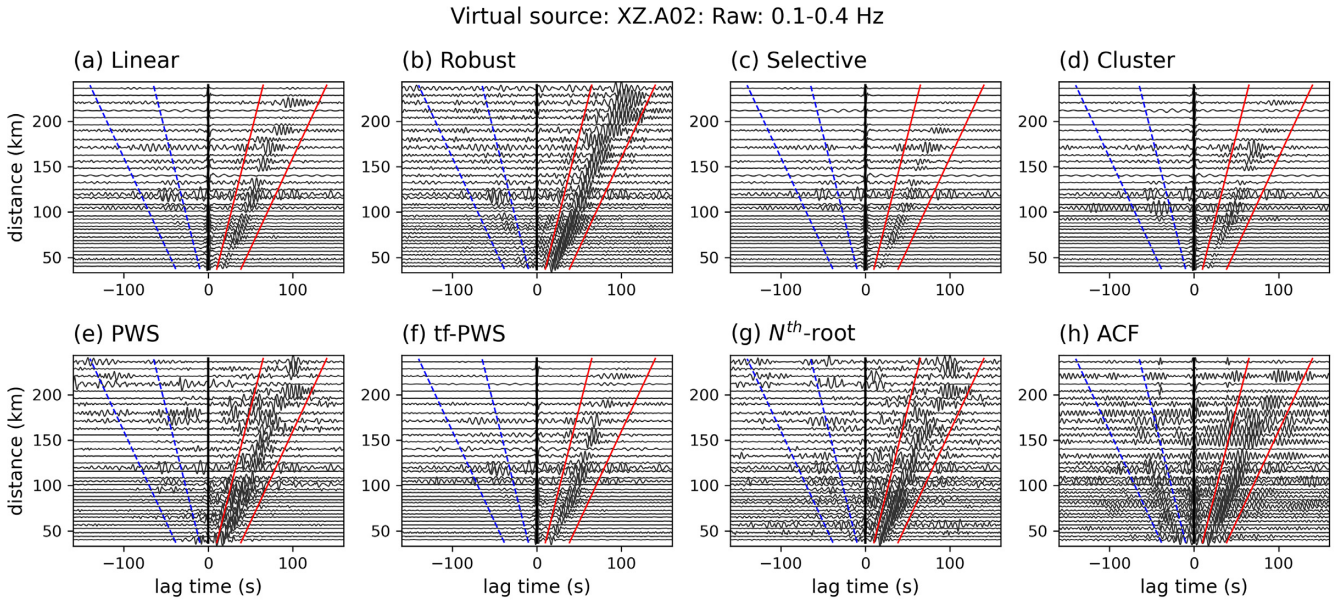

Figure 6. Stacked Raw NCFs between the XZ linear array stations from XZ.A02 to other receivers, filtered at 0.1–0.4 Hz. (a–h) The results using Linear, Robust, Selective, Cluster, PWS, tf-PWS, N^{th} -root and ACF stacking methods, respectively. The red solid lines and the blue dashed lines outline the positive signal window and the negative signal window, respectively, used to compute the SNR in Fig. 7. The signal and noise windows are determined with the same method as in Figs 3(c) and (d). See Fig. 1 for station locations.

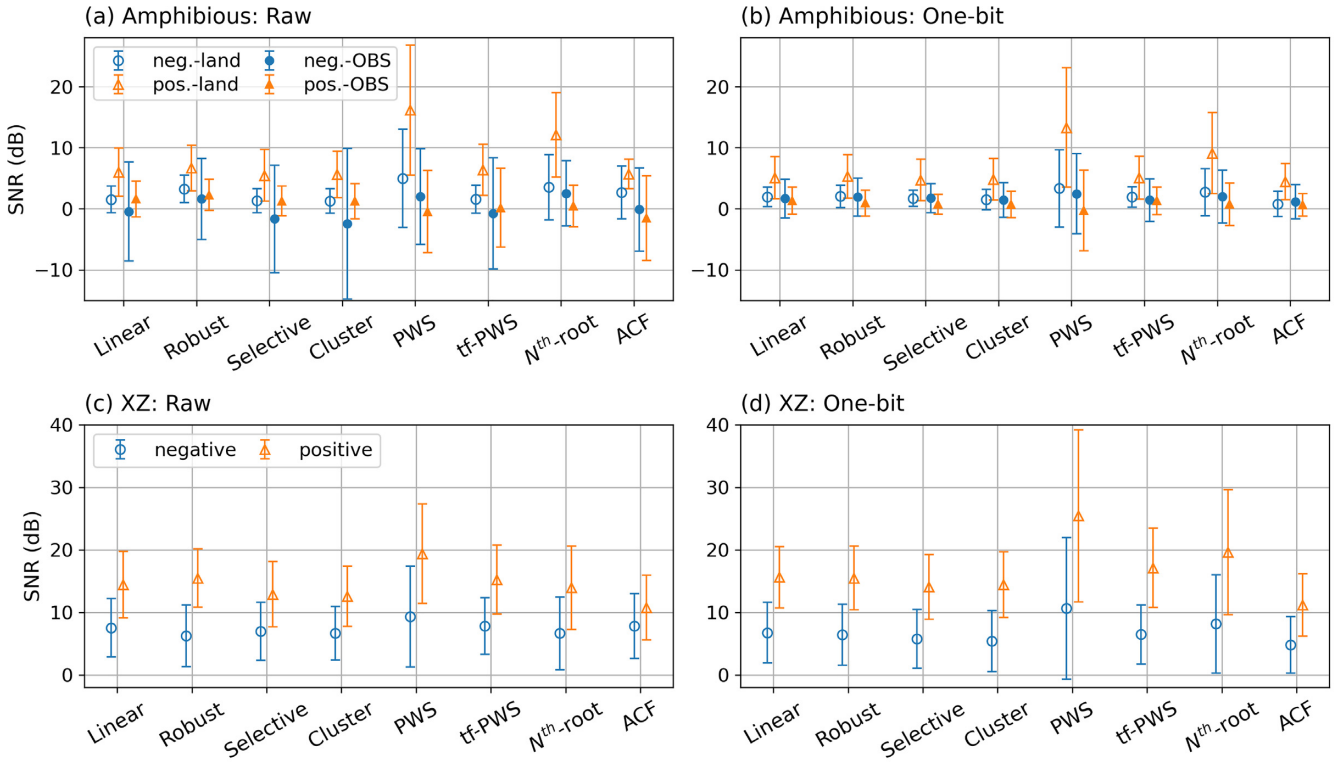

Figure 7. Comparison between the mean signal-to-noise ratios (SNRs) for the NCFs shown in Figs 4–6 and S2–S4. (a, b) Mean SNRs for the Cascadia amphibious array Raw and One-bit NCFs, respectively. The SNRs for the land (open circles and triangles) and OBS (dots and filled triangles) stations are plotted separately. (c, d) Results for the XZ linear array NCFs from Raw and One-bit NCFs, respectively. The error bars show the standard deviations of the SNRs.

This frequency–depth characteristic makes surface waves a popular choice for elastic and anelastic seismic tomography of the subsurface. Preserving the surface wave dispersion in the stacked NCFs is, therefore, one of the critical metrics to evaluate the performance of different stacking methods. We assess this performance by extracting and comparing the phase-velocity dispersion images. We use the Raw and One-bit NCFs from the XZ linear array between the virtual source at XZ.A02 and all other receivers (Fig. 6) to evaluate this metric.

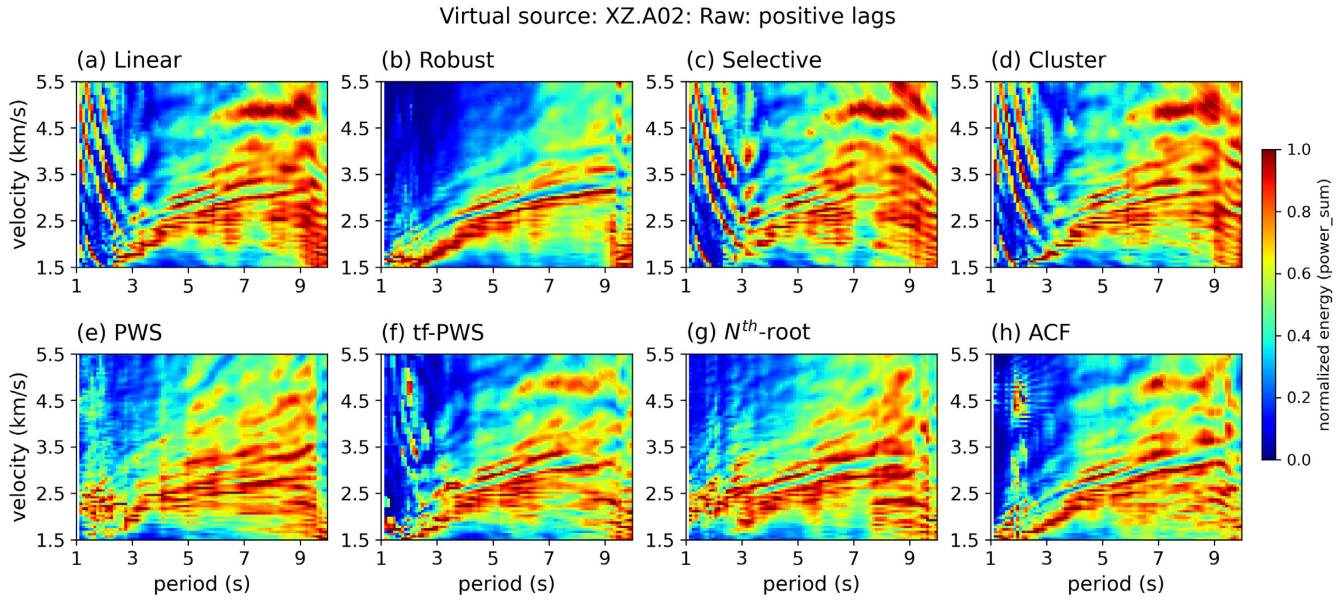

Figure 8. Dispersion images extracted from the Raw NCFs between the virtual source at XZ.A02 and other receivers using the time-domain phase-shift method through multichannel analysis of surface waves (e.g. Park *et al.* 1998, 1999; Ryden *et al.* 2004). (a–h) Results from the stacks using Linear, Robust, Selective, Cluster, PWS, tf-PWS, N^{th} -root and ACF stacking methods, respectively. The colour scale shows the power sum normalized at each velocity value. The NCFs are shown in Fig. 6.

We follow the phase-shift dispersion analysis procedure described in Park *et al.* (1998, 1999) and Ryden *et al.* (2004). We implement the procedure in the time domain through narrow-band filters. We narrow bandpass the seismograms using a Butterworth filter progressively between 1 and 10 s period, with a moving period band of 2 s and a step size of 0.1 s. We then shift the seismogram with a phase shift of r/v, where r is the interstation distance and v is the phase velocity that varies between 1.5 and 5.5 km s^{-1}, with a step of 0.05 km s^{-1}. We trim the data over an adaptive window length, which scales with the central period, computed as

$$L_i = a_i T_i, \tag{11}$$

where L_i is the window length for the ith period T_i. a_i is the scaling factor, which is determined by

$$a_i = a_{\min} + (i-1)\frac{a_{\max} - a_{\min}}{N_T}, \tag{12}$$

where $a_{\min}$ and $a_{\max}$ are the minimum and maximum of the scaling range and N_T is the number of period steps. For our examples, $a_{\min} = 1$ and $a_{\max} = 2$. Finally, to extract the dispersion image, we calculate the energy (sum of squared amplitude) of the windowed, filtered, shifted and stacked seismograms. The images are shown in Figs 8 and 9.

We apply this procedure to all station pairs with the common virtual source XZ.A02, with only results from the positive lags of the NCFs being shown here. We only analyse the results from station pairs with interstation distances of at least 1.5 times the wavelength for each velocity-period pair to remain in the far-field regime.

The dispersion images from the One-bit NCF stacks are largely similar to each other (Fig. 9). This contrasts with the great variability of the dispersion images from the Raw NCF stacks (Fig. 8). We use the average dispersion image of the One-bit results (Fig. 10a) as a reference. We compute the Structural Similarity Index (SSI; Wang *et al.* 2004) between the reference (Fig. 10a) and individual dispersion images from the NCF stacks using different methods (Figs 8 and 9). The SSI is widely used in the analysis of image degradation or alteration, with 1 meaning the two images are identical and 0 meaning they are completely different. The SSI results over the entire period range of 1–15 s for both the Raw and One-bit dispersion images are shown in Fig. 10(b). To examine the performance of different stacking methods at different period bands, we also compute the frequency-dependent SSI, as shown in Fig. 11.

The reference dispersion image (Fig. 10a) resembles major features of the One-bit dispersion images using different stacking methods (Fig. 9). Over the entire period range of 1–10 s, the dispersion images using the PWS and N^{th} -root methods for One-bit NCFs differ the most from the reference, with SSI below 0.8 (Fig. 10b). All other stacking methods show higher SSI values. In contrast with the One-bit results, most of the dispersion images for the Raw NCFs possess relatively lower SSI values (<0.6), except for the Robust stacking method, which shows an SSI of about 0.9. The frequency-dependent results (Fig. 11) show that all One-bit NCF stacks (Fig. 11b) and the Robust stack of Raw NCFs (Fig. 11a) retrieve the most consistent dispersion images at the period range of 4–9 s. For the One-bit NCFS, the SSI values outside this period for the PWS and N^{th} -root stacks decrease dramatically. All stacking results of Raw NCFs, except for the PWS result, show SSI values of >0.6 in a narrow band of 3–6 s (Fig. 11a). In summary, the Robust stacking method preserves the dispersion information the best for the Raw NCFs, with a dispersion image closest to the average of One-bit NCF stacks. For the One-bit NCFs, the Linear, Robust, Cluster and tf-PWS stacks perform comparably well in preserving the dispersion information over the entire examined period range from 1 to 10 s.

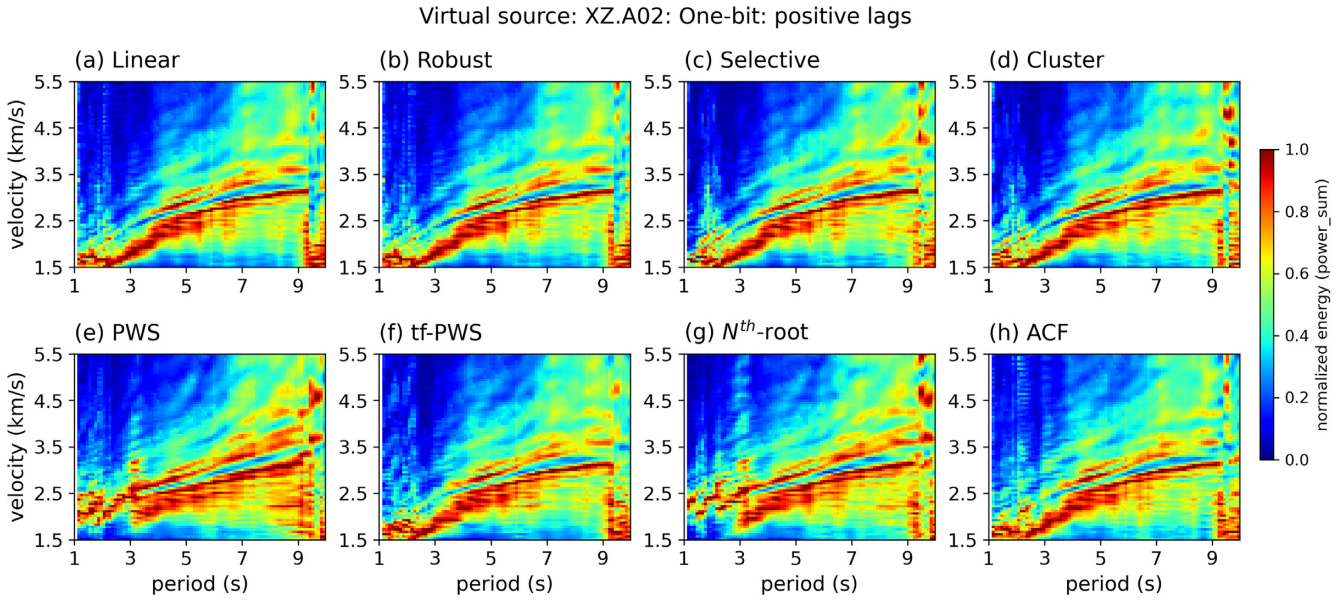

Figure 9. Same as Fig. 8 but for stacked One-bit NCFs as shown in Fig. S4 in the supplement.

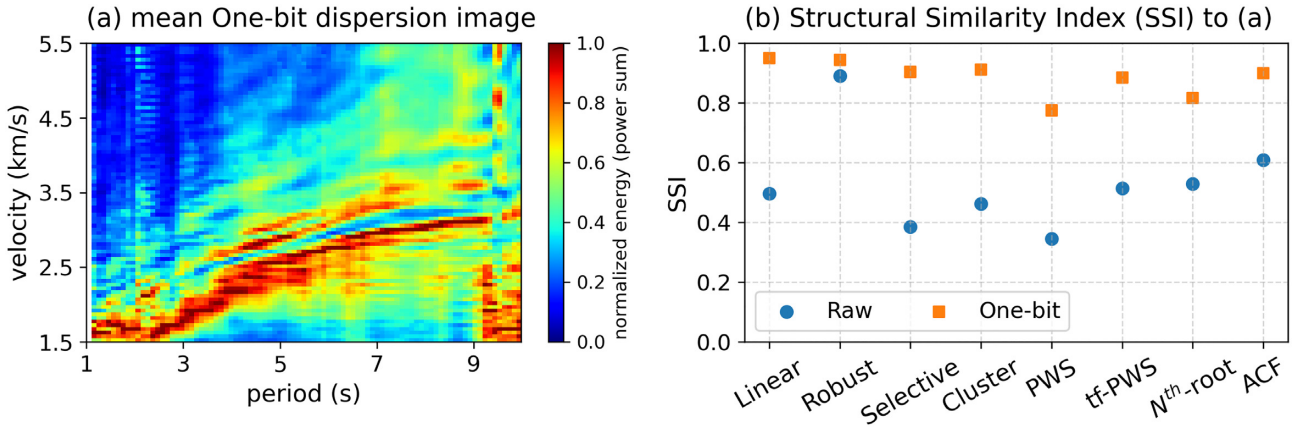

Figure 10. Comparison of the dispersion images in Figs 8 and 9. (a) The average dispersion image of the One-bit NCF stacks in Fig. 9, which is used as the reference for the comparison. (b) Structural Similarity Index (SSI) between individual dispersion images for Raw (dots) and One-bit (squares) NCF stacks and the reference dispersion image as in (a). The dispersion images are shown in Figs 8 and 9.

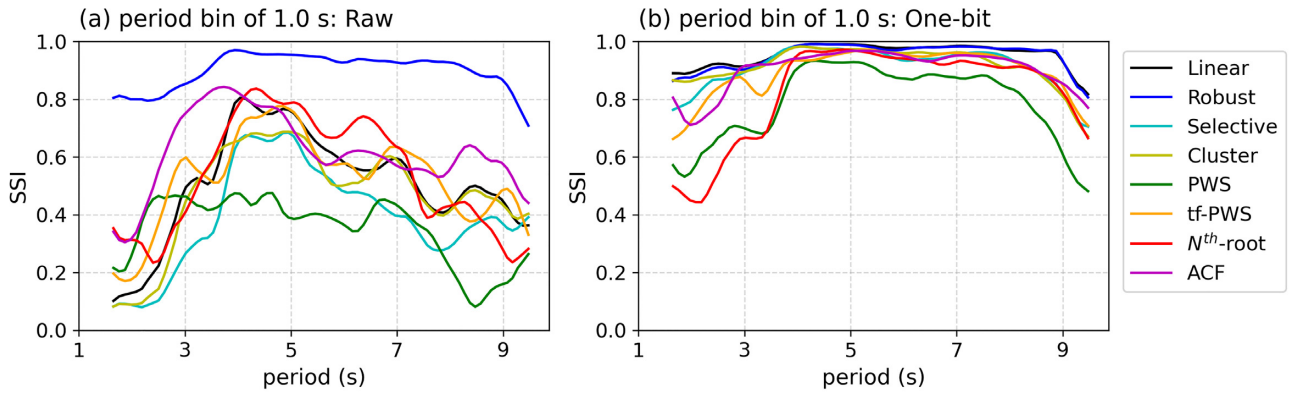

Figure 11. Frequency-dependent comparison of the dispersion images with the average dispersion image of the One-bit NCF stacks (Fig. 10a). (a) Structural Similarity Index (SSI) between individual dispersion images for Raw NCF stacks with a period bin of 2 s and a step of 0.1 s. The *x*-axis shows the central period of each bin. (b) Same as (a) but for One-bit NCF stacks. The dispersion images are shown in Figs 8 and 9.

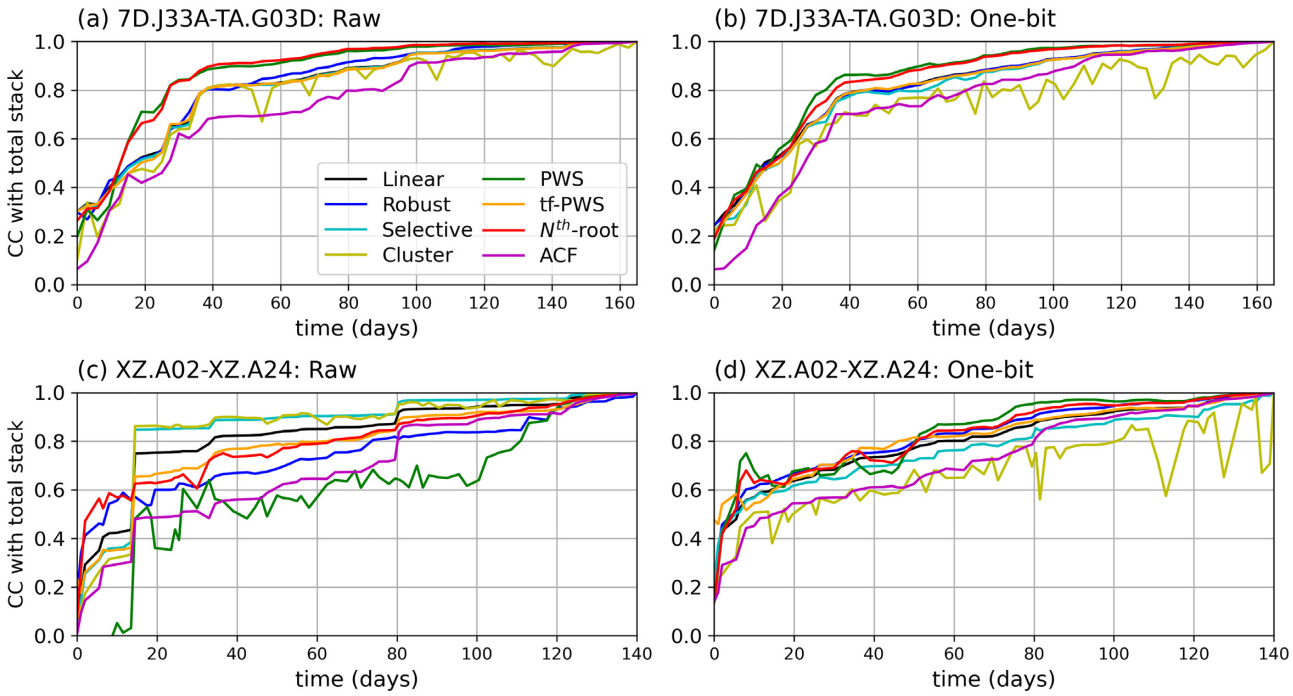

Figure 12. Convergence to long-term stacks with the increasing number of NCFs, shown as the equivalent time span of NCFs in days. (a, b) Correlation coefficients (CC) between the progressive stacking results and the stacks over the entire time period for Raw and One-bit NCFs between station pair 7D.J33A-TA.G03D. The pre-stack NCFs are shown in Figs 2(a) and (b). We only use NCFs before 4/1/2012 for the convergence analysis to exclude the much noisier data at later dates. All stacks are filtered at 0.1–0.4 Hz. (c, d) Same as (a) and (b) but for the XZ station pair XZ.A02-XZ.A24. The pre-stack NCFs are shown in Figs 2(c) and (d). We use all NCFs for this station pair.

4.3 Convergence of short-term stacks to the long-term stack

In measuring the temporal changes associated with the elastic properties of the crust, we often stack a subset of the short-term NCFs over an intermediate time period (e.g. Brenguier *et al.* 2008a; Hadziioannou *et al.* 2011; Seats *et al.* 2012; Clements & Denolle 2018; Donaldson *et al.* 2019; Olivier *et al.* 2019; Yuan *et al.* 2021). The time needed to reach convergence limits the temporal resolution of the monitoring of the change in elastic properties. Here, we quantify the *rate of convergence* as the time duration of NCFs needed to achieve a certain threshold of the correlation coefficient between the stack of the subset NCFs and the reference stack of NCFs over the entire period. We compute the correlation coefficient between the reference stack and the stack of a progressively growing subset of NCFs ordered by date and time with an increment of 5 NCFs (Fig. 12).

The rate of convergence varies among methods (Fig. 12). Overall, the PWS and N^{th}-root stacking results have the fastest convergence to the reference stack (Figs 12a, b and d), except for the Raw NCFs from the XZ station pair (Fig. 12c) where the PWS method converges the slowest. The ACF stacking result converges the slowest for the Raw NCFs of the Cascadia amphibious station pair (Fig. 12a). The Cluster stacking method converges the slowest for the One-bit NCFs from both station pairs (Figs 12b and d), though it shows a fast convergence rate for the Raw NCFs of the XZ station par (Fig. 12c). The strong fluctuation of the correlation coefficient using Cluster stacking may be due to the fact that the k-means clustering method involves some random processes in assigning the clusters. Most stacking results achieve relatively high ($\geq$0.7) correlation coefficients with more than about 20–40 d (Figs 12a and b). For the XZ NCFs to achieve a similar correlation coefficient, it takes about 15–70 d depending on the methods (Figs 12c and d). From the results shown here, the rate of convergence depends strongly on the stacking method and the quality of individual short-term NCFs.

4.4 Transient phase changes

Seismic monitoring using ambient noise interferometry relies on the phase difference of seismic waves measured at different lag times (e.g. Brenguier *et al.* 2008a; Hadziioannou *et al.* 2011; Seats *et al.* 2012; Clements & Denolle 2018; Donaldson *et al.* 2019; Olivier *et al.* 2019; Yuan *et al.* 2021). We analyse how these phase changes are preserved with different stacking methods by estimating the velocity changes (dv/v) over time. We utilize the trace stretching method (e.g. Lobkis & Weaver 2003; Brenguier *et al.* 2008b; Obermann *et al.* 2014; Yuan *et al.* 2021) to measure the dv/v between the two example station pairs: 7D.J33A-TA.G03D and XZ.A02-XZ.A24 (Fig. 13). The trace stretching method estimates dv/v by maximizing the correlation coefficient between the individual NCF coda with the reference coda through linearly stretching or squeezing the waveform within the specified time window. The Python function for trace stretching is built in the *SeisGo* interface (Yang *et al.* 2022b). For all results shown in Fig. 13, we measure the velocity changes in the frequency range of 0.1–0.4 Hz. For all data sets, we use a sub-stacking window length of 96 hr (4 d). The measuring time windows are 86–106 s for 7D.J33A-TA.G03D to capture the coda waves and 29–49 s for XZ.A02–XZ.A24 containing the ballistic phases. The dv/v measuring windows for both data sets are labelled in Fig. 2.

For each station pair, different stacking methods produce similar overall patterns of changes in dv/v over the entire data duration

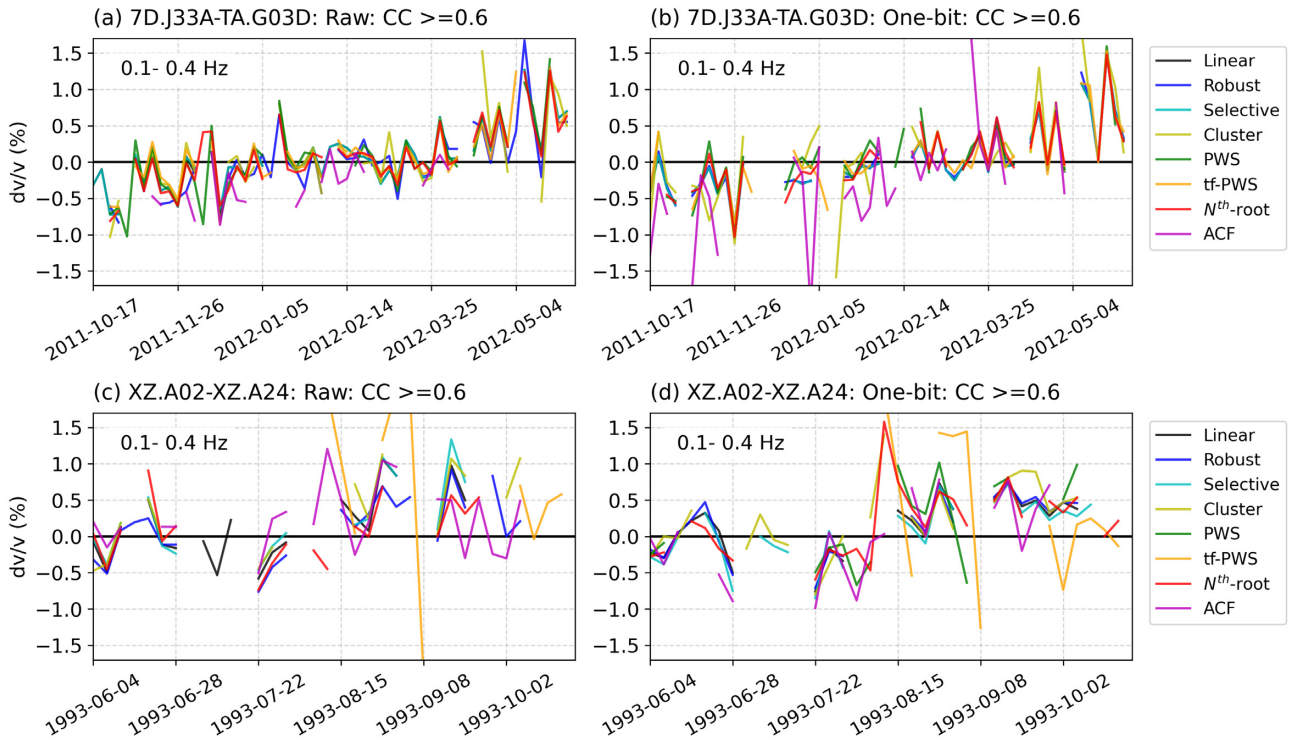

Figure 13. Transient velocity changes (dv/v) using the trace stretching method measured from NCFs from (a) 7D.J33A-TA.G03D and (b) XZ.A02-XZ.A24. The pre-stack short-term NCFs and the measuring windows are shown in Fig. 2. We use coda waves for the amphibious station pair (7D.J33A-TA.G03D) and ballistic phases for the XZ station pair (XZ.A02-XZ.A24). The coda waves for the XZ station pair (both Raw and One-bit NCFs) are not coherent and produce unstable dv/v measurements, regardless of the stacking method used. For 7D.J33A-TA.G03D, we only analyse the NCFs before 6/2/2012, during which we have clear coda phases on most of the NCFs. We substack the NCFs over 96 hr (4 d) for all station pairs. We only plot dv/v results with the post-stretching correlation coefficient (CC) of $\geq$0.6.

(Fig. 13). An exception is the dv/v measurements from the PWS and tf-PWS stacking results of the Raw XZ NCFs, with stronger variations and much lower after-stretching correlation coefficients compared to other stacking results. For 7D.J33A-TA.G03D, a general increase in seismic velocity from –1 per cent to 1.5 per cent is observed from both Raw and One-bit NCFs using all stacking methods (Figs 13a and b). For XZ.A02-XZ.A24, the dv/v fluctuates between –0.5 and 0.5 per cent from June to July of 1993 (Figs 13c and d). From August 1993 to the end of the observational period, the dv/v varies between 0 and 1 per cent (Figs 13c–d). While we do not intend to interpret these dv/v measurements, we note the great similarity of the measurements across methods. All stacking methods show comparable results in preserving the phase differences for coda waves from both Raw and One-bit NCFs and the ballistic phases from One-bit NCFs.

4.5 Peak amplitudes of ballistic phases

Important information about the Earth's anelastic structure can be retrieved from the relative amplitudes of NCFs (e.g. Prieto *et al.* 2009; Liu *et al.* 2021a). Here, we focus on the amplitude decay with distance and ignore the spatial variations in relative amplitudes that are subject to additional uncertainties. The NCFs chosen in this example are the same as in Fig. 6 and supplementary Fig. S3, filtered at 0.1–0.4 Hz. The NCFs at each station pair are stacked over the entire duration of data availability. We then measure the peak absolute amplitudes of the predicted ballistic waves for each station pair NCF stack within the time windows computed with a velocity range of 2.0–3.7 km s^{-1} (Figs 6 and S3). The XZ broad-band network is an east–west trending linear array in central Oregon, United States (Fig. 1). It is located at the active Cascadia convergent margin, spanning from the Cascadia Volcanic Arc in the east to the coastal mountains in the West. Considering that noise is dominantly generated by the oceanic microseisms in this frequency range (e.g. Webb 1998; Yang & Ritzwoller 2008), the XZ linear array provides an appropriate data set for analysing the attenuation of surface waves extracted from NCFs, minimizing the azimuthal dependence of noise sources. In this work, we only aim to compare the stability of the relative amplitude information with different stacking methods and do not intend to advocate for an estimate of attenuation, which might still depend on the distribution of noise sources (Stehly & Boué 2017). The microseismic noise is generated by the ocean and dominantly propagates eastward. Therefore, XZ.A02 is an appropriate choice for a virtual source (Fig. 1). We only measure the attenuation from the positive lags of the NCFs.

Assuming that the main signals of the NCF are fundamental-mode surface waves, the maximum absolute amplitudes of the stacked NCFs decay with distance following the relation:

$$A(D) = \frac{A_0}{\sqrt{D}} \exp(-\alpha D), \tag{13}$$

where D is the interstation distance, A_0 is a reference amplitude at the virtual source and α is a measure of 'attenuation' that could be attributed to intrinsic and scattering attenuation that further reduces the ground motion. To fit α, we correct for the geometrical spreading by scaling the amplitudes with a factor of $\sqrt{D}$. The peak absolute amplitudes are then fit to an exponential function $y =$

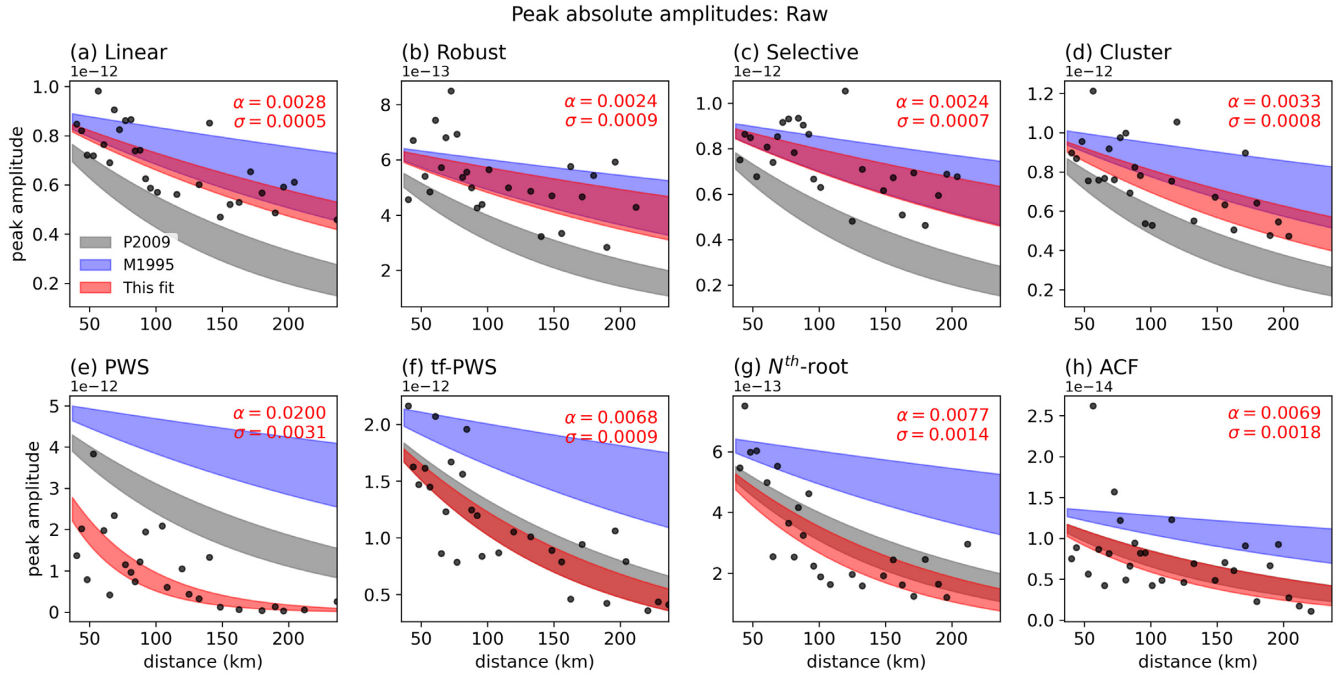

Figure 14. Peak absolute amplitudes of the positive lags of the XZ Raw NCFs within the predicted ballistic arrival windows computed using a velocity range of 2–3.7 km s^{-1} and corrected by $\sqrt{D}$, D is the interstation distance. In each panel, we only show the attenuation parameter, α, and its standard deviation, σ. The exponential fit from this study is shown as the red-shaded area. For reference, we also show the amplitude decay estimated by Prieto *et al.* (2009) (P2009; $\alpha = 0.0064 \pm 0.0013$; grey shaded area) and Mitchell (1995) (M1995; $\alpha = 0.002 \pm 0.001$; blue shaded area). See Fig. 6 for the moveout plots of the stacked NCFs.

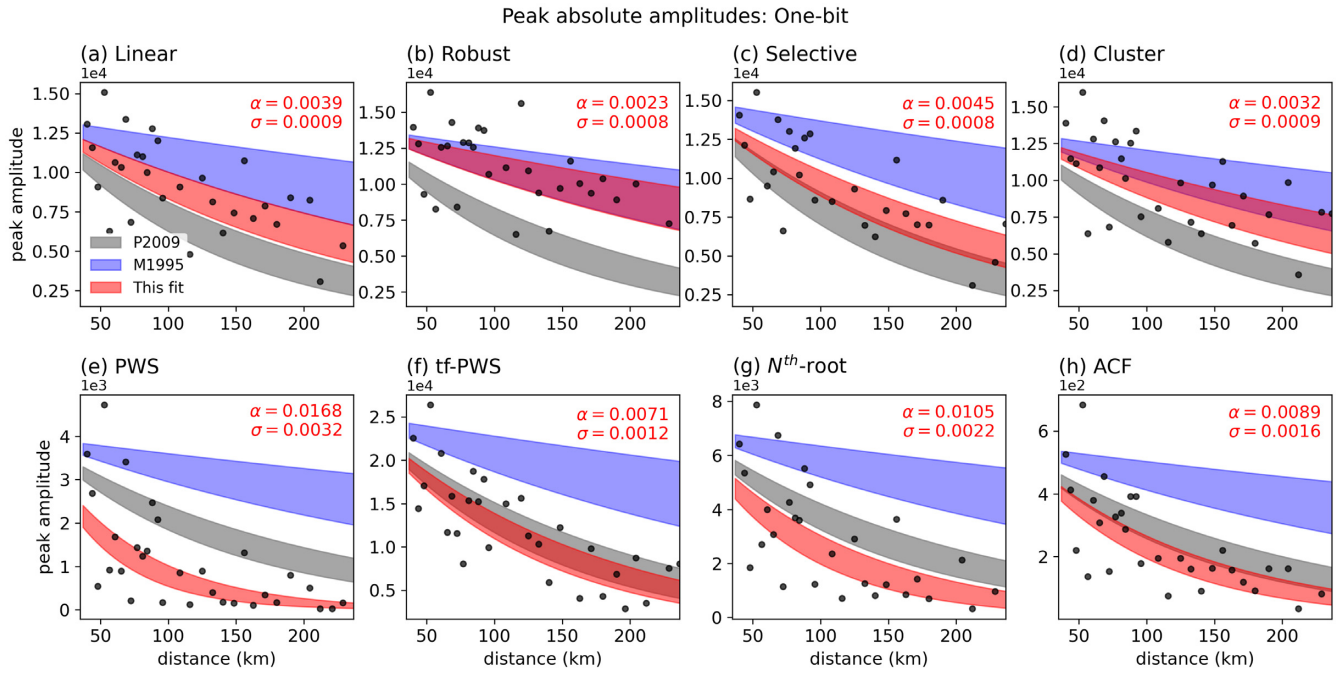

Figure 15. Same as Fig. 14 but for measurements from the XZ One-bit NCFs. See Fig. S3 in the supplement for the moveout plots of the stacked NCFs.

$e^{-\alpha x}$. We estimate the best fit through least squares linear regression of the peak amplitudes in a natural logarithmic space using the following steps: (1) calculate the natural log peak absolute amplitudes, (2) fit all data points using a least square linear regression (`scipy.stats.linregress`), (3) correct the data with the best-fitting model and measure misfit, (4) compute the mean and standard deviation of the misfit, (5) remove the outliers (data points that are more than one standard deviation away from the mean) and (6) use the remaining subset of the data to repeat step-2 to get the final linear fit parameters. The slope parameter in the linear regression from Step-2 is the attenuation factor α. The standard error in the slope, estimated by `scipy.stats.linregress`, is then used as the uncertainty of α, as shown in Figs 14 and 15.

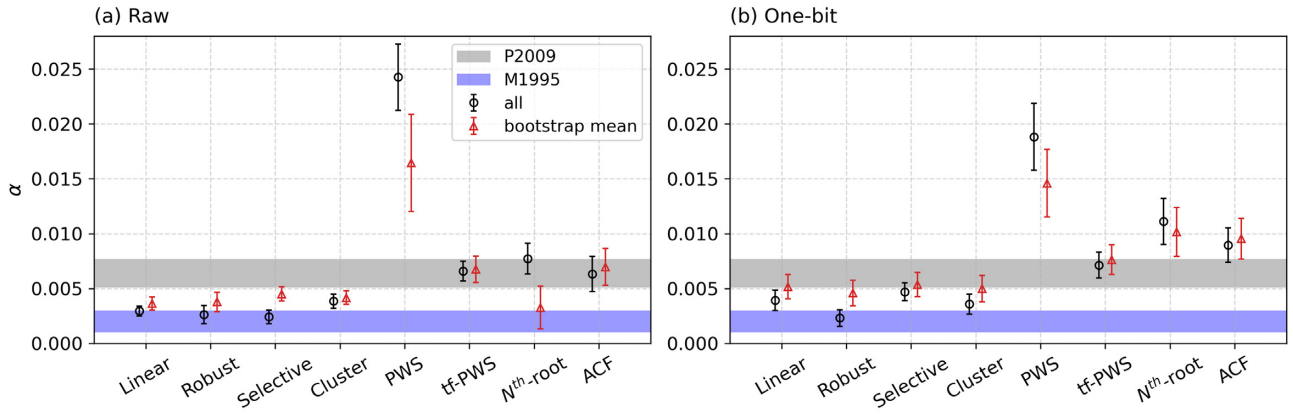

Figure 16. Comparison of the amplitude decay parameters (α) estimated using all-time NCF stacks (circles) and the mean bootstrap NCF stacks (triangles) for (a) Raw and (b) One-bit data sets. See the amplitude fit results in Figs 14, S7 and S8. For reference, we also show the ranges of amplitude decays estimated by Prieto *et al.* (2009) (P2009; $\alpha = 0.0064 \pm 0.0013$; grey shaded area) and Mitchell (1995) (M1995; $\alpha = 0.002 \pm 0.001$; blue shaded area).

We only use subsets of the NCFs in Figs 6 and S4 that have interstation distances between 37 km, which is the wavelength of a 0.1 Hz wave travelling at a velocity of 3.7 km s^{-1}, and 240 km. The exponential fit and the uncertainties are shown as the red shaded areas in Figs 14 and 15. For comparison, we also plot the independent attenuation measurements from Prieto *et al.* (2009) ($\alpha = 0.0064 \pm 0.0013$), who used spectral methods to measure attenuation in the Los Angeles basin in southern California using ambient noise NCF filtered at 0.2 Hz, and Mitchell (1995) ($\alpha = 0.002 \pm 0.001$) for active tectonic regions measured at 0.1–0.2 Hz using a global compilation of earthquake surface waves.

The temporal variation in location and strength of noise sources may affect the stability of the NCFs phases and amplitudes (e.g. Li *et al.* 2020). To examine these effects with respect to stacking methods, we conduct 1000 bootstrap stacking tests by randomly choosing 80% of the NCFs for each station pair without repetition. In Figs S5 and S6 in the supplement, we show the mean NCF stacks from the bootstrap test for all station pairs between XZ.A02 and other receivers. We also analyse the variability of NCF stacks compared to the bootstrapping mean stacks (see Text S1 and Fig. S7 in the supplement). We demonstrate that after taking into account the variation of temporal weighting when stacking, the all-time stacks are stable representations of the NCFs mostly within the standard deviations. Following the same procedures and parameters as using the all-time NCF stacks, we estimate the amplitude fit for the average stacks from the bootstrapping test (Figs S7 and S8). In Fig. 16, we compare the estimated amplitude decay coefficients, α, for both the all-time NCF stacks and the bootstrap mean NCF stacks.

From the all-time NCF stacks, the decay coefficient α can be measured with all stacking methods and for all pre-processing techniques. There is great variability in α among the methods (0.0024–0.02 for Raw NCFs and 0.0023–0.0168 for One-bit NCFs; Figs 14–16). For the Raw NCFs (Figs 14 and 16), the range of attenuation parameters measured from the stacking results using the Linear, Robust, Selective and Cluster methods all overlap with the values by Mitchell (1995). The coefficients using the N^{th}-root and ACF methods are comparable to that found by Prieto *et al.* (2009). The coefficient from the tf-PWS method falls in between the values provided by Mitchell (1995) and Prieto *et al.* (2009). For the estimates from One-bit NCFs (Figs 15 and 16), the result using the Robust method overlaps the most with the value proposed by Mitchell (1995). The Linear, Selective, Cluster and tf-PWS stacking results all fall in between the values by Mitchell (1995) and Prieto *et al.* (2009). The estimates of α from the N^{th}-root and ACF stacks are similar to that used by Prieto *et al.* (2009). Compared to the attenuation measurements from the Raw NCFs, the attenuation measurements from the One-bit NCFs seem to be more sensitive to the choice of stacking methods, shown as larger variations among different methods. For both data sets, the PWS method tends to overestimate the attenuation factors, α, with much higher values than other methods.

The deviation of the all-time NCF stacks from the bootstrap mean NCF stacks lead to variability in the amplitude decay measurements (Figs S7–S9 and 16). Fig. 16 shows that the amplitude decay parameters vary the most using the PWS method for both Raw and One-bit NCFs, followed by the results using the tf-PWS, N^{th} -root and ACF methods. The measurements using Linear, Robust, Selective and Cluster methods fluctuate within the ranges of about 0.001–0.005 for Raw NCFs (Fig. 16a) and about 0.0005–0.007 for One-bit NCFs. Considering the time-dependent variability of the NCF stacks, the Linear, Robust, Selective and Cluster methods produce comparable amplitude decay measurements for both Raw and One-bit NCFs.

4.6 Computational expenses

The computational efficiency varies notably among different stacking methods ranging from 10^{-4} s to a few seconds. The computing times are estimated from a single-core 3.6 GHz Intel Core i9 CPU. We compare the compute time spent stacking subsets of the NCFs (with an increment of 5 NCFs) using the different stacking methods (Fig. 17). Linear stacking is the fastest method. The ACF stacking method uses the most CPU time with about 4 s to stack 300 NCFs (Fig. 17). This is because the ACF method needs to compute the spectrum of each individual trace as well as the cross-spectrum in moving windows. As discussed in Section 3.5 and shown in Fig. S1, the tf-PWS method with the original Stockwell transform is much slower than the methods shown in Fig. 17. Most of the time in tf-PWS is spent on computing the Stockwell transform on individual NCFs. The performance of the tf-PWS method implemented with the more efficient DOST algorithm is comparable with that of the Cluster method (Fig. 17). The N^{th}-root method is the second fastest method following the Linear stack. These methods all scale nearly linearly with the number of NCFs to stack.

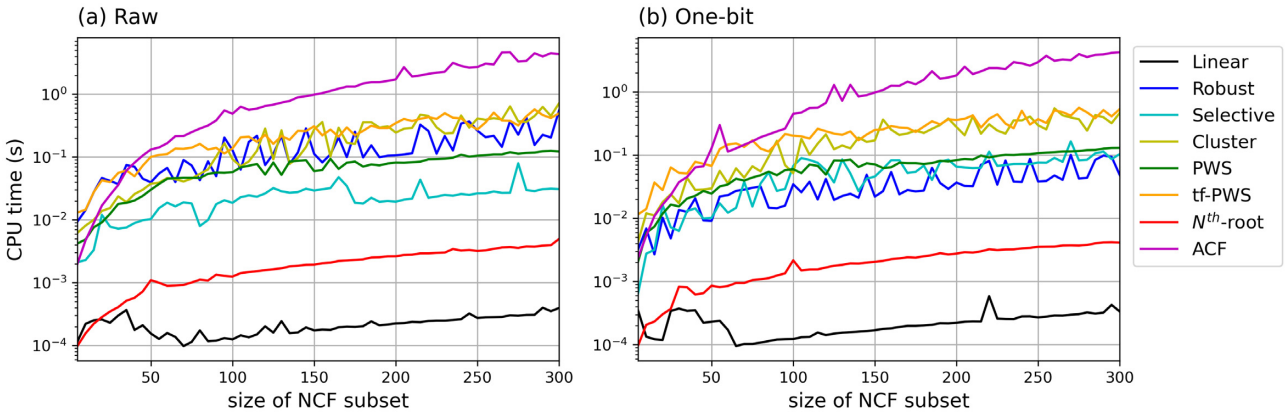

Figure 17. Computing times spent on stacking as a function of the number of NCFs to stack in a 5-step increment for (a) Raw and (b) One-bit NCFs from station pair 7D.J33A-TA.G03D.

Table 1. Our recommendations of stacking methods for major research applications using NCFs based on the evaluation metrics in Section 4. For methods denoted with (*), see the appropriate sections for additional discussion.

Applications	Evaluation metrics	Recommendations
Velocity tomography (Section 5.1)	Signal-to-noise ratio (Section 4.1), surface wave dispersion (Section 4.2) and computational expenses (Section 4.6)	Raw: Robust One-bit: Linear, Robust, Selective, ACF, Cluster
Monitoring (Section 5.2)	Signal-to-noise ratio (Section 4.1), surface-wave dispersion (Section 4.2), convergence of short-term stacks to the long-term stack (Section 4.3), transient phase changes (Section 4.4) and computational expenses (Section 4.6)	Raw: Robust One-bit: Linear, Robust, Selective, tf-PWS
Anelastic properties (Section 5.3)	Signal-to-noise ratio (Section 4.1), surface wave dispersion (Section 4.2), peak amplitudes of ballistic phases (Section 4.5) and computational expenses (Section 4.6)	Raw: Robust One-bit: Linear(*), Robust, Selective(*), Cluster(*)

5 CHOICE OF STACKING METHODS

In this section, we discuss the choice of stacking methods for different applications that utilize NCFs. We focus on the following major applications: (1) tomography of velocity structures, (2) monitoring of transient velocity changes and (3) characterization of anelastic properties. These three research applications extract different information from the NCFs: (1) seismic velocity tomography using surface waves requires the dispersion information of the ballistic surface wave phases, (2) seismic monitoring requires dispersion and slight variations in phase information of the ballistic phases and/or the coda waves and (3) the imaging of seismic attenuation and prediction of strong ground motion relies on relative amplitudes of the ballistic phases. Although the performance could vary with different data sets, we provide here the advantages and limitations of the stacking methods for each application in Table 1.

5.1 Stacking for tomography of velocity structures

Ambient noise tomography typically requires extracting the surface wave dispersion information from the stack of all NCFs for each station pair (e.g. Bensen *et al.* 2007). The fundamental mode surface waves are most commonly visible in the data, though there are some studies using higher modes for imaging (e.g. Wu *et al.* 2020; Jiang & Denolle 2022). Our recommendation for seismic velocity tomography is based on the following metrics: SNR (Section 4.1), surface wave dispersion (Section 4.2) and computational expense (Section 4.6).

Based on our examples in Section 4.1, except for the OBS-OBS station pairs, the PWS method produces stacks with the highest average SNRs, though with a large variance (see Figs 4–7 and S2–S4 in the supplement). The stacks using the N^{th}-root method also have relatively high SNRs. Other stacking methods perform at a comparable level in terms of the SNRs of the ballistic phases. The phase dispersion is better recovered using the Robust method than others for Raw NCFs over the entire examined period of 1–10 s (Fig. 10b), especially in the period range of 4–9 s (Fig. 11b). For One-bit NCFs, all but the PWS and N^{th}-root methods perform well overall in the dispersion analysis (Figs 9, 10b and 11b). In summary, our recommendations for tomographic imaging are the Robust stacking for Raw NCFs and the Linear, Robust, Selective, Cluster, DOST-based tf-PWS and ACF methods for One-bit NCFs.

5.2 Stacking for monitoring of transient velocity changes

Seismic monitoring uses ballistic or coda wave interferometry to infer small changes in the subsurface from short-time stacks of NCFs (e.g. Lobkis & Weaver 2003; Brenguier *et al.* 2008a; Clements & Denolle 2018; Donaldson *et al.* 2019; Obermann & Hillers 2019). A faster convergence of the NCF would lead to a higher temporal resolution in seismic monitoring. The convergence is often hindered and thus limits the temporal resolution (Hadziioannou *et al.* 2011). It is therefore important to find the optimal length of data that yields a reasonable convergence of the NCF stack. Meanwhile, time-lapse imaging requires that the dispersion of surface waves remains stable through time (e.g. Bergamo *et al.* 2016). Therefore, our recommendation for seismic monitoring using ambient noise interferometry is based on the following additional metrics on top of those for tomography (Sections 4.1, 4.2 and 4.6): convergence of short-term stack to long-term stack (Section 4.3) and transient phase changes (Section 4.4).

Regarding the convergence metric, all stacking methods show comparable performance (Fig. 12). The relatively large variability in the rate of convergence using the Cluster stacking method may lead to artifacts in monitoring. The ACF method is also relatively slow in converging to the long-term stack. It is worth noting again that the rate/time of convergence also depends on the quality of specific data sets. The dispersion analyses are most stable using the Robust method for Raw NCFs. For One-bit NCFs, all but the PWS and N^{th} -root methods are appropriate choices (Figs 8–11).

For transient phase changes, all but the PWS and tf-PWS methods perform comparably well for all of the four examined examples (Fig. 13). In summary, our recommendations for seismic monitoring are the Robust method for Raw NCFs and the Linear, Robust and Selective methods for One-bit NCFs.

5.3 Stacking for characterization of anelastic properties

The relative amplitudes of NCFs have been used to characterize the seismic attenuation properties of the Earth's lithosphere (Prieto *et al.* 2009) and the ground motion patterns (e.g. Denolle *et al.* 2013, 2014; Viens *et al.* 2017). Therefore, in addition to the metrics pertinent to tomography (Sections 4.1, 4.2 and 4.6), our recommendation for the characterization of anelastic properties using NCFs also evaluates the stability of peak absolute amplitudes of ballistic phases (Section 4.5).

Mitchell (1995) measured attenuation from a compilation of global earthquake records from a variety of tectonic settings, including active margins, which are similar to the setting for the XZ linear array at the Cascadia margin. Because it is measured from earthquake surface wave data, we argue that the attenuation parameter by Mitchell (1995) can serve as an independent benchmark in our evaluation. With this criterion, the Linear, Robust, Selective and Cluster stacking methods are ranked as the top four among all of the methods for Raw NCFs (Figs 14 and 16a). However, the requirement of a reliable measurement of frequency-dependent attenuation and dispersion (Figs 8, 10a and 11a) would narrow the choice down to Robust stacking for Raw NCFs. For One-bit NCFs, the amplitude decay measured from the all-time NCF stacks using the Robust stacking method overlaps the most with the value range by Mitchell (1995) (Figs 15 and 16b). However, when considering the bootstrapping mean NCF stacks and the uncertainties in amplitude decay estimates, the performance of Linear, Robust, Selective and Cluster methods are fairly comparable. Our recommendations for attenuation and ground motion studies, or any applications utilizing relative amplitudes, are the Robust method for both Raw and One-bit NCFs, though the Linear, Selective and Cluster methods are all reasonable choices for One-bit NCFs.

6 CONCLUSIONS

NCFs are widely used in seismic velocity and attenuation imaging, monitoring and ground motion analyses. The stacking of NCFs over longer time periods is needed for most applications utilizing NCFs to increase the strength of the coherent signals. There have been many temporal stacking methods developed to improve the stacking of NCFs. We compare eight temporal stacking methods, including Linear, Robust, Selective, Cluster, Phase-Weighted, time–frequency Phase-Weighted, N^{th}-root and Adaptive Covariance Filter stacking methods, to investigate their performance to enhance the quality of the stacked NCF. We examine the performance of these methods with six metrics, including SNRs, surface wave phase velocity dispersion, the convergence of short-term stacks to the long-term stack, wave speed changes, peak amplitudes and computational expenses. Our analyses demonstrate that although all methods are able to retrieve clear ballistic phases, their spectral contents and peak amplitudes vary strongly across methods. Based on multiple evaluation metrics, we recommend the Robust method for all three categories of applications using Raw NCFs, including velocity tomography, monitoring and attenuation studies. For tomography using One-bit NCFs, all methods except for Phase-Weighted, time–frequency phase-weighted and N^{th}-root stacking methods are good choices. For monitoring using One-bit NCFs, the Linear, Robust and Selective stacking methods are preferred choices, with the possibility of using time–frequency Phase-Weighted stacking for processing small data sets. For applications utilizing One-bit NCFs to extract relative amplitude information, both the Robust and Cluster stacking methods perform well. The findings in this study provide a practical guideline for choosing the appropriate stacking method for major applications utilizing NCFs. This work did not address the validity and effects of pre-processing techniques on our various use cases, but future work might address this (e.g. Fichtner *et al.* 2020). The open-source computer codes produced in this study can also be used for general time-series stacking analyses.

ACKNOWLEDGMENTS

This research has been funded by the startup funding of Purdue University for Xiaotao Yang, ARC grant DE220100907 to Chengxin Jiang and NSF grant EAR1749556 to Marine Denolle. Data from the TA network were made freely available as part of the EarthScope USArray facility, operated by Incorporated Research Institutions for Seismology (IRIS) and supported by the National Science Foundation, under Cooperative Agreements EAR-1261681.

Author contributions: X. Yang contributed to conceptualization, supervision, funding acquisition, data curation, methodology, formal analysis, software development and writing of the original draft. M. Denolle contributed to the conceptualization, supervision and funding acquisition of the project. J. Bryan and K. Okubo contributed to methodology development. All authors contributed to software development and reviewing and editing of the manuscript.

DATA AVAILABILITY

The seismic waveforms were downloaded from IRIS Data Management Center and processed using *SeisGo* (https://doi.org/10.5281/zenodo.5873724), built upon *ObsPy* (Beyreuther *et al.* 2010) and *NoisePy* (Jiang & Denolle 2020). The computer codes for the stacking methods are available as a standalone Python package, *StackMaster*, which is archived on Zenodo (https://doi.org/10.5281/zenodo.5951013) and is accessible on Python Package Index (https://pypi.org/project/stackmaster). All Python scripts and Jupyter notebooks used in the processing and visualizing of the data and stacking results are available at https://github.com/xtyangpsp/Paper_OptimalStacking and are archived on Zenodo (https://doi.org/10.5281/zenodo.6363585).

REFERENCES

Baig, A.M., Campillo, M. & Brenguier, F., 2009. Denoising seismic noise cross correlations, *J. geophys. Res.,* **114**(8), 1–12.

Battisti, U. & Riba, L., 2016. Window-dependent bases for efficient representations of the Stockwell transform, *Appl. Comput. Harmon. Anal.,* **40**(2), 292–320.

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**(3), 1239–1260.

Bergamo, P., Dashwood, B., Uhlemann, S., Swift, R., Chambers, J.E., Gunn, D.A. & Donohue, S., 2016. Time-lapse monitoring of climate effects on earthworks using surface wavestime-lapse seismic monitoring with SW, *Geophysics,* **81**(2), EN1–EN15.

Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y. & Wassermann, J., 2010. ObsPy: a Python toolbox for seismology, *Seismol. Res. Lett.,* **81**(3), 530–533.

Brenguier, F., Campillo, M., Hadziioannou, C., Shapiro, N.M., Nadeau, R.M. & Larose, E., 2008a. Postseismic relaxation along the San Andreas Fault at Parkfield from continuous seismological observations, *Science,* **321**(5895), 1478–1481.

Brenguier, F., Shapiro, N.M., Campillo, M., Ferrazzini, V., Duputel, Z., Coutant, O. & Nercessian, A., 2008b. Towards forecasting volcanic eruptions using seismic noise, *Nat. Geosci.,* **1**(2), 126–130.

Clements, T. & Denolle, M.A., 2018. Tracking groundwater levels using the ambient seismic field, *Geophys. Res. Lett.,* **45**(13), 6459–6465.

Denolle, M.A., Boué, P., Hirata, N. & Beroza, G.C., 2018. Strong shaking predicted in Tokyo from an expected M7+ Itoigawa-Shizuoka earthquake, *J. geophys. Res.,* **123**(5), 3968–3992.

Denolle, M.A., Dunham, E.M., Prieto, G.A. & Beroza, G.C., 2013. Ground motion prediction of realistic earthquake sources using the ambient seismic field, *J. geophys. Res.,* **118,** 2102–2118.

Denolle, M.A., Dunham, E.M., Prieto, G.A. & Beroza, G.C., 2014. Strong ground motion prediction using virtual earthquakes, *Science,* **343**(6169), 399–403.

Donaldson, C., Winder, T., Caudron, C. & White, R.S., 2019. Crustal seismic velocity responds to a magmatic intrusion and seasonal loading in Iceland's Northern Volcanic Zone, *Sci. Adv.,* **5**(11), doi:10.1126/sciadv.aax6642.

Du, Z., Foulger, G.R. & Mao, W., 2000. Noise reduction for broad-band, three-component seismograms using data-adaptive polarization filters, *Geophys. J. Int.,* **141**(3), 820–828.

Ermert, L., Villasenor, A. & Fichtner, A., 2016. Cross-correlation imaging of ambient noise sources, *Geophys. J. Int.,* **204**(1), 347–364.

Ermert, L.A., Sager, K., Nissen-Meyer, T. & Fichtner, A., 2021. Multifrequency inversion of global ambient seismic sources, *Geophys. J. Int.,* **225**(3), 1616–1623.

Feng, K.-F., Huang, H.-H., Hsu, Y.-J. & Wu, Y.-M., 2021. Controls on seasonal variations of crustal seismic velocity in Taiwan using single-station cross-component analysis of ambient noise interferometry, *J. geophys. Res.,* **126**(11), e2021JB022650, doi:10.1029/2021JB022650.

Feng, L. & Ritzwoller, M.H., 2019. A 3-D shear velocity model of the crust and uppermost mantle beneath alaska including apparent radial anisotropy, *J. geophys. Res.,* **124**(10), 10 468–10 497.

Fichtner, A., Bowden, D. & Ermert, L., 2020. Optimal processing for seismic noise correlations, *Geophys. J. Int.,* **223**(3), 1548–1564.

Hadziioannou, C., Larose, E., Baig, A., Roux, P. & Campillo, M., 2011. Improving temporal resolution in ambient noise monitoring of seismic wave speed, *J. geophys. Res.,* **116**(B7), doi:10.1029/2011JB008200.

IRIS OBSIP, 2011. Cascadia Initiative Community Experiment - OBS Component, International Federation of Digital Seismograph Networks, doi:10.7914/SN/7D_2011.

IRIS Transportable Array, 2003. USArray Transportable Array [Data set], International Federation of Digital Seismograph Networks, doi:10.7914/SN/TA.

Jiang, C. & Denolle, M.A., 2020. NoisePy: a new high-performance Python tool for ambient-noise seismology, *Seismol. Res. Lett.,* **91**(3), 1853–1866.

Jiang, C. & Denolle, M.A., 2022. Pronounced seismic anisotropy in Kanto sedimentary basin: a case study of using dense arrays, ambient noise seismology, and multi-modal surface-wave imaging, *J. geophys. Res.,* **127**(8), e2022JB024613, doi:10.1029/2022JB024613.

Kanasewich, E.R., Hemmings, C.D. & Alpaslan, T., 1973. Nth-root stack nonlinear multichannel filter, *Geophysics,* **38,** 327–338.

Kwak, S., Song, S.G., Kim, G., Cho, C.S. & Shin, J.S., 2017. Investigating the capability to extract impulse response functions from ambient seismic noise using a mine collapse event, *Geophys. Res. Lett.,* **44**(19), 9653–9662.

Li, G., Niu, F., Yang, Y. & Xie, J., 2017. An investigation of time-frequency domain phase-weighted stacking and its application to phase-velocity extraction from ambient noise's empirical Green's functions, *Geophys. J. Int.,* **212**(2), 1143–1156.

Li, J., Weaver, R.L., Yoritomo, J.Y. & Song, X., 2020. Application of temporal reweighting to ambient noise cross-correlation for improved seismic green's function, *Geophys. J. Int.,* **221,** 265–272.

Liu, G., Fomel, S., Jin, L. & Chen, X., 2009. Stacking seismic data using local correlation, *Geophysics,* **74**(3), V43–V48.

Liu, X., Beroza, G.C., Yang, L. & Ellsworth, W.L., 2021a. Ambient noise love wave attenuation tomography for the lassie array across the Los Angeles basin, *Sci. Adv.,* **7**(22), doi:10.1126/sciadv.abe1030.

Liu, Y., Xia, J., Xi, C., Dai, T. & Ning, L., 2021b. Improving the retrieval of high-frequency surface waves from ambient noise through multichannel-coherency-weighted stack, *Geophys. J. Int.,* **227**(2), 776–785.

Lobkis, O.I. & Weaver, R.L., 2003. Coda-wave interferometry in finite solids: recovery of *p*-to-*s* conversion rates in an elastodynamic billiard, *Phys. Rev. Lett.,* **90**(25), doi:10.1103/PhysRevLett.90.254302.

Millet, F., Bodin, T. & Rondenay, S., 2019. Multimode 3-D Kirchhoff migration of receiver functions at continental scale, *J. geophys. Res.,* **124**(8), 8953–8980.

Mitchell, B.J., 1995. Anelastic structure and evolution of the continental crust and upper mantle from seismic surface wave attenuation, *Rev. Geophys.,* **33**(4), 441–462.

Moreau, L., Stehly, L., Boué, P., Lu, Y., Larose, E. & Campillo, M., 2017. Improving ambient noise correlation functions with an SVD-based wiener filter, *Geophys. J. Int.,* **211**(1), 418–426.

Muirhead, K.J., 1968. Eliminating false alarms when detecting seismic events automatically, *Nature,* **217,** 533–534.

Nabelek, J., Li, X.Q., Azevedo, S., Braunmiller, J., Fabritius, A., Leitner, B., Trehu, M. A. & Zandt, G., 1993. A high-resolution image of the Cascadia subduction zone from teleseismic converted phases recorded by a broadband seismic array, *EOS, Trans. Am. geophys. Un.,* **74**(43), 431.

Nakata, N., Chang, J.P., Lawrence, J.F. & Boué, P., 2015. Body wave extraction and tomography at Long Beach, California, with ambient-noise interferometry, *J. geophys. Res.,* **120**(2), 1159–1173.

Obermann, A., Froment, B., Campillo, M., Larose, E., Planès, T., Valette, B., Chen, J.H. & Liu, Q.Y., 2014. Seismic noise correlations to image structural and mechanical changes associated with the Mw 7.9 2008 Wenchuan earthquake, *J. geophys. Res.,* **119**(4), 3155–3168.

Obermann, A. & Hillers, G., 2019. Chapter two - Seismic time-lapse interferometry across scales, in *Recent Advances in Seismology, Vol. 60 of Advances in Geophysics,* pp. 65–143, ed. Schmelzbach, C., Elsevier.

Olivier, G., Brenguier, F., Campillo, M., Lynch, R. & Roux, P., 2015. Body-wave reconstruction from ambient seismic noise correlations in an underground mine, *Geophysics,* **80**(3), KS11–KS25.

Olivier, G., Brenguier, F., Carey, R., Okubo, P. & Donaldson, C., 2019. Decrease in seismic velocity observed prior to the 2018 eruption of Kılauea volcano with ambient seismic noise interferometry, *Geophys. Res. Lett.,* **46**(7), 3734–3744.

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

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

Pavlis, G.L. & Vernon, F.L., 2010. Array processing of teleseismic body waves with the USArray, *Comput. Geosci.,* **36**(7), 910–920.

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**(7), 1–15.

Rost, S. & Thomas, C., 2002. Array seismology: methods and applications, *Rev. Geophys.,* **40**(3), 2–1.

Rückemann, C.-P.C., 2012. Comparison of stacking methods regarding processing and computing of geoscientific depth data, in *Proceedings of the*

Fourth International Conference on Advanced Geographic Information Systems, Applications, and Services (GEOProcessing 2012), pp. 35–40.
Ryden, N., Choon, Park, B., Ulriksen, .P. & Miller, R.D., 2004. Multimodal approach to seismic pavement testing, *J. Geotech. Geoenviron. Eng.,* **130**(6), doi:10.1061/(ASCE)1090-0241(2004)130:6(636).
Samson, J.C. & Olson, J.V., 1981. Data-adaptive polarization filters for multichannel geophysical data, *Geophysics,* **46**(10), 1423–1431.
Schimmel, M. & Gallart, J., 2007. Frequency-dependent phase coherence for noise suppression in seismic array data, *J. geophys. Res.,* **112**(4), 1–14.
Schimmel, M. & Paulssen, H., 1997. Noise reduction and detection of weak, coherent signals through phase-weighted stacks, *Geophys. J. Int.,* **130**(2), 497–505.
Schimmel, M., Stutzmann, E. & Gallart, J., 2011. Using instantaneous phase coherence for signal extraction from ambient noise data at a local to a global scale, *Geophys. J. Int.,* **184**(1), 494–506.
Seats, K.J., Lawrence, J.F. & Prieto, G.A., 2012. Improved ambient noise correlation functions using Welch' s method, *Geophys. J. Int.,* **188**(2), 513–523.
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, Y., Ren, Y., Gao, H. & Savage, B., 2012. An improved method to extract very-broadband empirical Green's functions from ambient seismic noise, *Bull. seism. Soc. Am.,* **102**(4), 1872–1877.
Stehly, L. & Boué, P., 2017. On the interpretation of the amplitude decay of noise correlations computed along a line of receivers, *Geophys. J. Int.,* **209**(1), 358–372.
Stehly, L., Campillo, M. & Shapiro, N., 2006. A study of the seismic noise from its long-range correlation properties, *J. geophys. Res.,* **111**(B10), doi:10.1029/2005JB004237.
Stehly, L., Cupillard, P. & Romanowicz, B., 2011. Towards improving ambient noise tomography using simultaneously curvelet denoising filters and sem simulations of seismic ambient noise, *Comp. Rend. Geosci.,* **343**(8–9), 591–599.
Stockwell, R.G., 2007. A basis for efficient representation of the S-transform, *Digit. Signal Process.: Rev. J.,* **17**(1), 371–393.
Stockwell, R.G., Mansinha, L. & Lowe, R.P., 1996. Localization of the complex spectrum: the S transform, *IEEE Trans. Signal Process.,* **44**(4), 998–1001.
Tavenard, R. *et al.*, 2020. Tslearn, a machine learning toolkit for time series data, *J. Mach. Learn. Res.,* **21**(118), 1–6.
Thangraj, J.S. & Pulliam, J., 2021. Towards real-time assessment of convergence criteria in seismic interferometry: selective stacking of cross-correlations at the San Emidio Geothermal Field, *J. appl. Geophys.,* **193,** doi:10.1016/j.jappgeo.2021.104426.
Thurber, C.H., Zeng, X., Thomas, A.M. & Audet, P., 2014. Phase-weighted stacking applied to low-frequency earthquakes, *Bull. seism. Soc. Am.,* **104**(5), 2567–2572.
Tian, Y. & Ritzwoller, M.H., 2017. Improving ambient noise cross-correlations in the noisy ocean bottom environment of the Juan de Fuca plate, *Geophys. J. Int.,* **210**(3), 1787–1805.
Toghramadjian, N., Ermert, L.A. & Denolle, M.A., 2021. A tale of urban seismology: ambient seismic noise, machine learning methods, and seismic hazard analysis at the seattle basin edge, in *Poster Presentation at 2021 SCEC Annual Meeting,* Poster #214, SCEC Contribution #11436, SCEC.
Viens, L. & Denolle, M.A., 2019. Long-period ground motions from past and virtual megathrust earthquakes along the Nankai Trough, Japan, *Bull. seism. Soc. Am.,* **109**(4), 1312–1330.
Viens, L., Denolle, M.A., Miyake, H., Sakai, S. & Nakagawa, S., 2017. Retrieving impulse response function amplitudes from the ambient seismic field, *Geophys. J. Int.,* **210**(1), 210–222.
Viens, L. & Iwata, T., 2020. Improving the retrieval of offshore-onshore correlation functions with machine learning, *J. geophys. Res.,* **125**(8), 1–19.
Wang, Q.-Y., Brenguier, F., Campillo, M., Lecointre, A., Takeda, T. & Aoki, Y., 2017. Seasonal crustal seismic velocity changes throughout Japan, *J. geophys. Res.,* **122**(10), 7987–8002.
Wang, Z., Bovik, A.C., Sheikh, H.R. & Simoncelli, E.P., 2004. Image quality assessment: from error visibility to structural similarity, *IEEE Trans. Image Process.,* **13**(4), doi:10.1109/TIP.2003.819861.
Webb, S.C., 1998. Broadband seismology and noise under the ocean, *Rev. Geophys.,* **36**(1), 105–142.
Wu, G., Pan, L., Wang, J. & Chen, X., 2020. Shear velocity inversion using multimodal dispersion curves from ambient seismic noise data of USArray Transportable Array, *J. geophys. Res.,* **125**(1), 1–14.
Wu, S.-M., Lin, F.-C., Farrell, J., Keller, W.E., White, E.B. & Hungerford, J.D., 2021. Imaging the subsurface plumbing complex of steamboat geyser and cistern spring with hydrothermal tremor migration using seismic interferometry, *J. geophys. Res.,* **126**(4), doi:10.1029/2020JB021128.
Yang, X., Bryan, J., Okubo, K., Jiang, C., Clements, T. & Denolle, M.A., 2022a. *StackMaster: A Collection of Methods for Data Stacking, This replaces StackPy/StackMaster,* v1.0.3.
Yang, X. & Gao, H., 2018. Full-wave seismic tomography in the northeastern United States: new insights into the uplift mechanism of the Adirondack Mountains, *Geophys. Res. Lett.,* **45**(12), 5992–6000.
Yang, X. & Gao, H., 2020. Segmentation of the Aleutian-Alaska subduction zone revealed by full-wave ambient noise tomography: implications for the along-strike variation of volcanism, *J. geophys. Res.,* **125**(11), 1–20.
Yang, X., Gao, H., Rathnayaka, S. & Li, C., 2019. A comprehensive quality analysis of empirical Green's functions at ocean-bottom seismometers in Cascadia, *Seismol. Res. Lett.,* **90**(2A), 744–753.
Yang, X., Zuffoletti, I.D., D'Souza, N.J. & Denolle, M.A., 2022b. *SeisGo: A Ready-To-Go Python Toolbox for Seismic Data Analysis,* v0.7.0, Zenodo. doi:10.5281/zenodo.5873725.
Yang, Y. & Ritzwoller, M.H., 2008. Characteristics of ambient seismic noise as a source for surface wave tomography, *Geochem. Geophys. Geosyst.,* **9**(2), doi:10.1029/2007GC001814.
Yang, Z., Yuan, C. & Denolle, M.A., 2022c. Detecting elevated pore pressure due to wastewater injection using ambient noise monitoring, *Seismic Record,* **2**(1), 38–49.
Yuan, C., Bryan, J. & Denolle, M.A., 2021. Numerical comparison of time-, frequency-, and wavelet-domain methods for coda wave interferometry, *Geophys. J. Int.,* **226**(2), 828–846.
Zeng, X. & Thurber, C.H., 2016. A graphics processing unit implementation for time-frequency phase-weighted stacking, *Seismol. Res. Lett.,* **87**(2A), 358–362.

SUPPORTING INFORMATION

Supplementary data are available at *GJI* online.

Figure S1. Comparison of stacked noise cross-correlation functions (NCFs) between station XZ.A02 and other stations in the XZ network, using the time–frequency Phase-Weighted Stacking (tf-PWS) method based on the original S-transform (ST; Stockwell *et al.* 1996) and the discrete orthogonal S-transform (DOST; Stockwell 2007). We show results using both Raw (top) and One-bit (bottom) NCFs. The number in the title of each panel is the CPU time used in stacking of all station pairs.

Figure S2. Stacked One-bit noise cross-correlation functions of the Cascadia amphibious array from 7D.J33A to other land receivers using different stacking methods, filtered at 0.1–0.4 Hz. (a–h) The results using the Linear, Robust, Selective, Cluster, PWS, tf-PWS, N^{th} -root and ACF stacking methods, respectively. The red solid lines and the blue dashed lines outline the positive-lag signal window and the negative-lag signal window, respectively, used to compute the signal-to-noise ratios in Fig. 7 in the main text. The signal and noise windows are determined with the same method as in Figs 3(a) and (b) in the main text.

Figure S3. Same as Fig. S2 but for NCFs from 7D.J33A to other Ocean Bottom Seismograph (OBS) receivers. To contain the visually identified ballistic phases from these OBS station pairs, we use a different velocity range (0.5–1.0 km s^{-1}) here to predict the signal window of the weakly coherent signals. We extend the window for an additional 60 s after the latest predicted arrival. See Fig. 1 in the main text for locations of the OBS receivers.

Figure S4. Stacked One-bit NCFs between the XZ.A02 and other receivers, filtered at 0.1–0.4 Hz. (a–h) The results using the Linear, Robust, Selective, Cluster, PWS, tf-PWS, N^{th} -root and ACF stacking methods, respectively. The red solid lines and the blue dashed lines outline the positive signal window and the negative signal window, respectively, used to compute the signal-to-noise ratios in Fig. 7 in the main text. The signal and noise windows are determined with the same method as in Figs 3(c) and (d) in the main text. See Fig. 1 in the main text for station locations.

Figure S5. The average of the Raw NCF stacks between XZ.A02 and other receivers filtered at 0.1–0.4 Hz, with bootstrapping of 1000 times and 80 per cent resampling ratio without replacements. (a–h) The results using the Linear, Robust, Selective, Cluster, PWS, tf-PWS, N^{th} -root and ACF stacking methods, respectively. The red solid lines and the blue dashed lines outline the positive signal window and the negative signal window, respectively. The signal and noise windows are determined with the same method as in Fig. S4. See Fig. 1 in the main text for station locations.

Figure S6. Bootstrap mean NCF stacks same as Fig. S5 but for One-bit NCFs.

Figure S7. Deviation index (DI) between the all-time NCF stacks and the bootstrap mean NCF stacks for (a) Raw and (b) One-bit NCFs. The deviation index of the NCF stack for each station pair is computed with eq. (S1). The thick line shows the baseline value when the difference is comparable to the standard deviation of the bootstrapping test. We only compute the DIs for station pairs with $\geq$5 NCFs.

Figure S8. Amplitude fit same as Fig. 14 in the main text but with the mean bootstrap stacks of Raw NCFs as shown in Fig. S5. See Fig. 14 and the main text for amplitude fitting procedures and parameters. The exponential fit from this study is shown as the red shaded area. For reference, we also show the amplitude decay estimated by Prieto *et al*. (2009) (P2009; $\alpha = 0.0064 \pm 0.0013$; grey shaded area) and Mitchell (1995) (M1995; $\alpha = 0.002 \pm 0.001$; blue shaded area).

Figure S9. Amplitude fit the same as Fig. S8 but with mean bootstrap stacks of One-bit NCFs (see Fig. S6 for the NCF stacks).

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the paper.