JGR Earth Surface

RESEARCH ARTICLE
10.1029/2023JF007091

Processing Ambient Noise Data Using Phase Cross-Correlation and Application Toward Understanding Spatiotemporal Environmental Effects

H. B. Woo[1], S. L. Bilek[1], J. A. Gochenour[1], R. Grapenthin[2], A. J. Luhmann[3], and J. B. Martin[4]

[1]Department of Earth and Environmental Science, New Mexico Institute of Mining and Technology, Socorro, NM, USA, [2]Geophysical Institute and Department of Geosciences, University of Alaska Fairbanks, Fairbanks, AK, USA, [3]Department of Earth and Environmental Science, Wheaton College, Wheaton, IL, USA, [4]Department of Geological Sciences, University of Florida, Gainesville, FL, USA

Key Points:

- Short-period (<1 s) surface waves are used to extract the empirical Green's functions (EGFs) using two different interferometry methods
- Number of wavelengths has the greatest impact when constraining the range of reliable wave periods from surface wave dispersion measurements
- EGFs quality controlled using seismic phase show greater effectiveness in measuring complex velocity structures than velocity measurements

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

H. B. Woo,
hanbyulwoo@gmail.com

Citation:

Woo, H. B., Bilek, S. L., Gochenour, J. A., Grapenthin, R., Luhmann, A. J., & Martin, J. B. (2023). Processing ambient noise data using phase cross-correlation and application toward understanding spatiotemporal environmental effects. *Journal of Geophysical Research: Earth Surface*, *128*, e2023JF007091. https://doi.org/10.1029/2023JF007091

Received 30 JAN 2023
Accepted 5 JUL 2023

Abstract Typical use of ambient noise interferometry focuses on longer period (>1 s) waves for exploration of subsurface structure and other applications, while very shallow structure and some environmental seismology applications may benefit from use of shorter period (<1 s) waves. We explore the potential for short-period ambient noise interferometry to determine shallow seismic velocity structures by comparing two methodologies, the conventional amplitude-based cross-correlation and linear stacking (TCC-Lin) and a more recently developed phase cross-correlation and time-frequency phase-weighted-stacking (PCC-PWS) method with both synthetic and real data collected in a heterogeneous karst aquifer system. Our results suggest that the PCC-PWS method is more effective in extracting short-period wave velocities than the TCC-Lin method, especially when using data collected in regions containing complex shallow structures such as the karst aquifer system investigated here. In addition to the different methodologies for computing the cross correlation functions, we also examine the relative importance of signal-to-noise ratio and number of wavelengths propagating between station pairs to determine data/solution quality. We find that the lower number of wavelengths of 3 has the greatest impact on the network-averaged group velocity curve. Lastly, we test the sensitivity of the number of stacks used to create the final empirical Green's function, and find that the PCC-PWS method required about half the number of cross-correlation functions to develop reliable velocity curves compared to the TCC-Lin method. This is an important advantage of the PCC-PWS method when available data collection time is limited.

Plain Language Summary Seismic monitoring can provide important information about the Earth's interior structure. However, mapping shallow sedimentary geology can be challenging due to the complex structure and environmental effects, such as variable water tables that modify seismic waves. With seismic data measured above a karst aquifer system at O'Leno and River Rise Preserve State Parks, Florida, USA, we compare the quality and reliability of the results between a conventional seismic data processing method that uses the amplitude of the seismic signal and a relatively new method using the phase of the seismic signal in extracting information of the shallow subsurface. We also investigate the effectiveness of the two methods, determining the amount of seismic data required to derive reliable velocity structures. Although detailed signal characterization is necessary in selecting the best method, our study suggests that the seismic data processing method using seismic phase information will play an important role in mapping shallow structures as well as in monitoring applications.

1. Introduction

Theoretical and experimental studies have shown that the empirical Green's functions (EGFs) can be extracted from cross-correlation of diffuse seismic wave-fields between station-pairs (Bensen et al., 2007; Derode et al., 2003; Larose et al., 2005; Snieder, 2004; Wapenaar, 2004). The use of EGFs retrieved from ambient noise cross-correlations provides supportive and compatible information to body wave signals for improving seismic tomographic resolution, due to the wave ray paths constrained between two stations (Haned et al., 2016; Roux et al., 2005; Sabra et al., 2005; Shapiro & Campillo, 2004). Ambient noise data processing has been extensively and successfully applied in regional and global tomographic studies where EGFs have been determined from long

period ($T > 3$ s) waves associated with deep crustal and mantle structures (Bensen et al., 2007; Boué et al., 2014; Dias et al., 2015; Ekstrom et al., 2009; Fichtner et al., 2008; Haned et al., 2016; Sabra et al., 2005; Shapiro & Campillo, 2004).

There have been fewer attempts to measure velocity profiles from EGFs using short ($T < 1$ s) and intermediate ($1 \text{ s} < T < 3$ s) wave periods to identify shallow crustal and sedimentary structures (Chávez-García & Luzón, 2005; James et al., 2017; Pilz et al., 2012; Savage et al., 2013). In the use of short-period waves, it is important to establish a reliable criterion in selecting the randomized noise because reconstruction of velocity profiles that relies on the scattered wave energy at depths may incorporate biases from uneven distribution of wave sources (Poli et al., 2012). Nonetheless, cross-correlation functions (CCFs) aligned in different directions have been observed not to diverge considerably without the knowledge of wave source directions (Aki & Chouet, 1975), and succeeding studies have stated that the effects of inhomogeneous distribution of wave sources on ambient noise velocity measurements will be within the errors of isotropic velocity measurements (Bensen et al., 2007; Harmon et al., 2010; Pilz et al., 2012). Also, greater complexity is introduced to the velocity measurements by multi-scattering, high wave attenuation, and multimodal signals due to shallow subsurface heterogeneities that perturb the waveform in comparison to the less complex deep interior of the Earth (Aki & Chouet, 1975; Levshin & Panza, 2006; Pilz et al., 2012; Shapiro & Campillo, 2004; Vernon et al., 1998).

Development of ambient noise interferometry techniques has improved the estimation of EGFs by reducing incoherent signals between seismic traces (Campillo & Paul, 2003; Schimmel & Gallart, 2007; Schimmel et al., 2011; Ventosa et al., 2017, 2019). The methods, including cross-correlation with the seismic phase angle (PCC) and stacking the correlation functions weighted by the time-frequency dependent phase (tf-PWS), have been successfully applied to the global seismic network (GEOSCOPE) and regional datasets (Dantas et al., 2018; Haned et al., 2016; Schimmel et al., 2011; Toledo et al., 2022; Ventosa et al., 2017). However, the PCC and tf-PWS methods, advantageous in mitigating strong amplitude events compared to the traditional ambient noise interferometry method, have not yet been widely used with short-period waveforms so their performance has not been rigorously compared to traditional methods (De Plaen et al., 2019; Toledo et al., 2022).

With advancements in seismic data processing methods, measurements of seismic velocity changes, associated with perturbations in elastic properties, have also been widely used for environmental studies to understand the physical dynamics of various geological settings (Andajani et al., 2020; Clements & Denolle, 2018; De Plaen et al., 2019; Lecocq et al., 2017; Mao et al., 2020; Mordret et al., 2010; Ratdomopurbo & Poupinet, 1995; Rivet et al., 2015). The key for monitoring temporal variations in seismic velocity is to compare phase shifts between a reference EGF that represents an average background state for a longer time period and current EGFs that represent the situation at a certain time period (Clarke et al., 2011; Clements & Denolle, 2018; Ratdomopurbo & Poupinet, 1995). However, there has not been much examination on the quality of such EGFs, which could be critical in obtaining robust measurements of the sensitive response of seismic waves propagating through a variable elastic medium, especially with short-period waves.

The purpose of this study is to compare the traditional cross-correlation method implemented in the time-domain (TCC) with the phase cross-correlation (PCC) method by estimating EGFs from synthetic data over a network of less than 20 km interstation distance and short-period ($T < 1$ s) waves. We further examine the two methods with a real seismic data set measured in the Santa Fe River watershed in north-central Florida where surface water interacts with the semi-confined Floridan karst aquifer (Scott, 1992). Surface wave dispersion curves measured from the EGFs are assessed for quality, and these results provide insights on selecting specific ambient noise techniques and parameters for various studies based on the network scale, interstation distances, and geology within the seismic array network. The quality-controlled dispersion measurements from both synthetic and real data are inverted to compare with the initial velocity model (synthetic) and stratigraphic structure of the karst aquifer (real). Using data from the local field survey located within a complex karst aquifer system, we examine whether the high frequency ambient seismic noise can be used to measure reliable group velocity curves in a highly heterogeneous layered structure, and determine what set of parameters, such as the number of CCF stacks, provides reliable velocity structures. Because of the interest in applying these ambient noise techniques to a variety of temporally varying environmental processes, our study also provides preliminary guidelines for setting the criteria for surface wave dispersion measurement quality in order to acquire reliable velocity measurements and maximize the use of a given data set for short-period seismic waves.

2. Methods

EGFs acquired from cross-correlating ambient seismic noise of a station-pair contain the subsurface structural information between the paired stations. Symmetric Green's functions (G), from station A to B or B to A can be expressed as,

$$G_{AB}(t) = G_{BA}(t) = -\frac{d}{dt}\left[\frac{C_{AB}(t) + C_{BA}(t)}{2}\right] (0 \leq t), \tag{1}$$

where C are the cross correlation coefficient functions from station A to B or B to A and t is time (Bensen et al., 2007). The main ambient noise data processing steps include (a) pre-processing, (b) cross-correlation, and (c) stacking CCFs. One commonly used method down-weights the influence of large-amplitude signals through binarization of the data in the time-domain and spectral whitening in the frequency-domain to exclude earthquake signals and other heterogeneous noise in pre-processing steps, where meaningful signals might be lost (Bensen et al., 2007; Haned et al., 2016). An alternative method was developed to cross-correlate the instantaneous phase of station-pair data based on the analytical signal theory to extract phase coherent signals that are amplitude unbiased. Additionally, the time-frequency phase-weighted stacking (tf-PWS) method was developed to calculate the EGFs rather than linearly stacking the CCFs (Haned et al., 2016; Schimmel, 1999; Schimmel et al., 2011).

2.1. Phase Cross-Correlation

For each pair of seismic stations, phase cross-correlograms are computed as follows. From a single seismic waveform $u(t)$, the analytical signal is $s(t) = u(t) + iH(u(t)) = A(t)e^{i\phi(t)}$, where $H(u(t))$ is the Hilbert transform of $u(t)$ and $A(t)$ and $\phi(t)$ are the envelope and phase of the analytic signal, respectively. Considering two waveforms $u(t)$ and $v(t)$ with a starting time of τ_0, and their instantaneous phases $\Phi(t)$ and $\Psi(t)$, a phase cross-correlation (PCC) between two time series is:

$$C_{PCC}(\tau) = \frac{1}{2T}\sum_{t=\tau+\tau_0}^{T-\tau+\tau_0}\left(|e^{i\Phi(t+\tau)} + e^{i\Psi(t)}|^{\nu} - |e^{i\Phi(t+\tau)} - e^{i\Psi(t)}|^{\nu}\right), \tag{2}$$

where t is time, τ is the amount of time-lag, ν is the power factor value of PCC that controls the amplification of the CCFs, and T is the time-length of each correlation window (Haned et al., 2016; Schimmel, 1999). PCC produces the similarity property between the two time series as a function of τ allowing detection of weak or noisy seismic signals, unbiased by the amplitude, and is more efficient to detect phase-coherent ambient noise (Schimmel, 1999; Ventosa et al., 2017). C_{PCC} is used to determine whether the two signals are completely correlated ($C_{PCC} = 1$; auto-correlation), completely anticorrelated ($C_{PCC} = -1$), or not correlated ($C_{PCC} = 0$), thus $|C_{PCC}| \leq 1$. Performance of the PCC compared to the classical cross-correlation method has been proven to provide robust group-velocity measurements (Haned et al., 2016; Schimmel, 1999). However, PCC is much more computationally demanding to apply to large seismic data set with high sampling frequencies, requiring 2T sums and (2T-1) sums of complex numbers and modulus, and real numbers, respectively from the two analytic sequences. Therefore, a simplified derivation of the PCC equation is used to dramatically decrease the computational expense by simplifying the PCC equation with the power factor value ($\nu = 2$) to a complex cross-correlation employing the Fast Fourier Transform as,

$$C_{PCC2}[m] = \mathrm{Re}\frac{1}{N}\sum_{n=0}^{N-1} e^{i\Phi_1[n+m]}e^{i\Phi_2[n]}, \tag{3}$$

where Φ_1 and Φ_2 are the instantaneous phase of the two discrete time series, n is the sample number, m is the time-lag number, and Re is the operator retrieving the real part of a signal. Equation 3 has the advantage over Equation 2 in estimating CCFs because the application of the cross-correlation theorem reduces computational expense. The simplification of Equation 2 to Equation 3 has thus far only been shown to be possible for $\nu = 2$ (Ventosa et al., 2019).

2.2. Phase Weighted Stacking

While the cross-correlation functions were stacked linearly by each time window in the conventional method, the Stockwell-transform (S-transform), a time-frequency decomposition process, enables a different approach to detect weak phase coherent signals, which will herein be referred to as phase weighted stacking, or PWS (Bensen

et al., 2007; Haned et al., 2016; Schimmel & Paulssen, 1997; Schimmel et al., 2011; Stockwell et al., 1996). The PWS is a non-linear stack where CCFs computed at each time segment are weighted by a coherence measure, which is amplitude unbiased. The main step of the PWS is to compute the phase coherence from the phase cross-correlogram, $C_{PCC}(\tau)$ through the S-transform in both time and frequency domain. The S-transform obtains an analytic signal from a waveform ($u(t)$) that involves filtering through a Gaussian window function (w) centered at τ and window width proportional to $\left|\frac{1}{f}\right|$, where the | | symbol represents the absolute value function:

$$S(\tau,\ f)\ =\ \int_{-\infty}^{\infty} u(t)\, w(\tau - t,\ f)\, e^{-i2\pi f t}\, dt \tag{4}$$

The time-frequency instantaneous phase is defined as $\frac{S(\tau,f)e^{i2\pi f\tau}}{|S(\tau,f)|}$ and the PWS is then computed as the product of the mean of the time-frequency phase stack and the mean of the phase-correlation (C_{PCC}) stack decomposed in the time-frequency domain as:

$$S_{PWS}(\tau, f) = \left| \frac{1}{N} \sum_{j=1}^{N} \frac{S_j(\tau, f) e^{i2\pi f \tau}}{|S_j(\tau, f)|} \right|^{\nu} \cdot S\left(\frac{1}{N} \sum_{i=1}^{N} C_{PCC_i}(\tau) \right), \tag{5}$$

where N is the total number of stacks. The sharpness of phase coherency of the mean phase stack can be controlled with the exponent ν, which is equivalent to the power factor in Equation 2 (Haned et al., 2016; Schimmel & Paulssen, 1997). The inverse of the S-transform is computed from $S_{PWS}(\tau, f)$ to return to the time domain to obtain PWS as a function of time-lag. In the following sections, the conventional method will be referred to as the time cross-correlation and linear stacking (TCC-Lin) and the new method as phase cross-correlation and time-frequency phase weighted stacking (PCC-PWS).

2.3. Dispersion Measurement

In order to obtain dispersion measurements from the EGFs to extract the group velocity curves, time (time-lag) and frequency representation of the EGFs are required to locate the spectral peaks. Here, we use a different method from the traditional frequency-time analysis (FTAN) (Herrmann, 1973, 2013) and employ the continuous wavelet transform (CWT) analysis (Bensen et al., 2007; Fichtner et al., 2008; Jiang & Denolle, 2020; Torrence & Compo, 1998). Continuous wavelet transform is a convolution process of a time-series with a scaled and translated version of a specific wavelet $\psi_0(\eta)$ (η is non-dimensional time) (Torrence & Compo, 1998). The $\psi_0(\eta)$ has a zero mean, is localized in both time and frequency, and has a complex conjugate $\psi_0(\eta)$*. Various forms of $\psi_0(\eta)$ are available for convolution with the EGFs, but the most general "Morlet" wavelet function $\psi_0(\eta) = e^{-\eta^2/2}\cos(5\eta)$ is used, which is a plane wave modulated by a Gaussian function (Torrence & Compo, 1998). The CWT is processed in the frequency domain for faster calculation, and the amplitude is normalized for each frequency to detect the position of time-lag of the maximum peak that is picked to represent the group velocity curve of the EGFs (Jiang & Denolle, 2020).

3. Data

3.1. Synthetic Surface Wave Data

We use synthetic surface wave data modeled in a spatially heterogeneous medium to examine and illustrate the performance and limitations between the TCC-Lin and PCC-PWS methods. Using synthetic data allows us to constrain the true signals and their range in frequency and noise characteristics propagating through a predetermined half-space velocity structure. Thus, we can apply different ambient noise data processing methods and evaluate their performance against a known answer. We utilize a reference velocity structure model (Table S1 in Supporting Information S1) obtained from a previous karst tomographic study in Florida (James et al., 2017) and because we will later apply the techniques to real data, we limit the frequency distribution based on the corner frequency (0.2 Hz) of the seismic sensor (Sercel L22) deployed at our survey site above a karst conduit (O'Leno and River Rise Preserve State Parks, FL). We do not focus on the complicated noise generation from different sources nor the complex propagation and scattering of waves, but emphasize the extraction of group velocity curves based on the interstation distance and signal quality, which will be affected by the signal-to-noise ratio of the EGFs and the number of wavelengths that travel between the spacings between station-pairs for each center period.

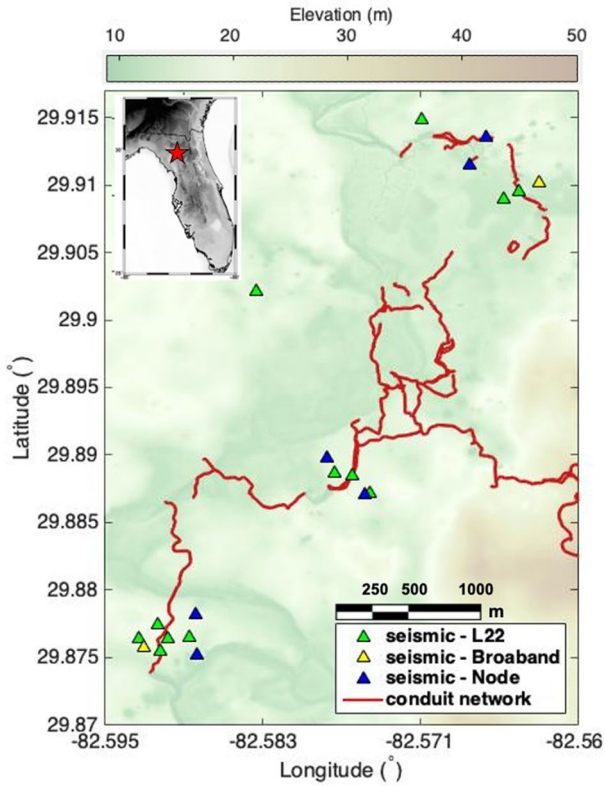

Figure 1. Map of O'Leno and River Rise Preserve State Parks (red star in inset map) and the seismic network with 20 seismic sensors (triangles) plotted above a digital elevation model (USGS). The red line refers to the network of water filled caves that is located approximately 30 m below the land surface (Conduit network provided by the Santa Fe River Rise Project of the National Speleological Society Cave Diving Section, mpoucher@cavesurvey.com).

From the Computer Programs in Seismology software (Herrmann, 2013), we adopt a surface wave synthetic modeling algorithm that generates a point source Green's Function (GF) propagating through a 1-dimensional elastic medium (Herrmann, 1973; Lin et al., 2008; Wapenaar, 2004). Within a source region of a 100 km × 100 km grid, we ran 5 different models with 2 receiver stations spaced 1, 2, 5, 10, and 20 km (Figure S1 in Supporting Information S1), respectively, a sampling frequency (f_{samp}) of 0.25 Hz, and frequency limits of $f_1 = 0.005$ Hz, $f_2 = 0.01$ Hz, $f_3 = 49$ Hz, and $f_4 = 50$ Hz. The four corner frequencies are used to perform a cubic cosine tapering ($|f_1, f_2|$ and $|f_3, f_4|$) to ensure that the deconvolved passband spectrum is between f_2 and f_3 (Herrmann, 1973). A single wave source is positioned 50 km west from the westernmost receiver aligned with both stations to generate force functions under a predetermined velocity and density model (Table S1 in Supporting Information S1). The integrated energy of Love (A_L) and Rayleigh (A_R) surface waves are defined in terms of integrals of the eigenfunctions, $A_L = \int_0^\infty \rho U_\phi^2 dz$ and $A_R = \int_0^\infty \rho \left[U_z^2 + U_r^2\right] dz$, where U is the eigenfunction and ϕ, z, and r represent azimuthal, vertical, and radial direction from the source, respectively (Snieder, 2004). The point force GF initiated at the source position is recorded at the two stations, where the fundamental mode is calculated between the source and 2 stations (Snieder, 2004). White Gaussian noise with the same frequency range as the raw seismic signal with a scale factor of 5 to completely mask the signal is added to the vertical component of the raw seismic waves measured at the 2 receivers in order to generate an authentic noise-like seismic signal for our tests of the efficiency and robustness between the TCC-Lin and PCC-PWS methods.

3.2. Real Observational Data

A subset of seismic data is used collected by a temporary seismic network (FDSN network code XS, https://doi.org/10.7914/SN/XS_2018) deployed at O'Leno and River Rise Preserve State Parks, FL (Figure 1). The seismic network includes 12 Sercel L-22 short-period sensors with a corner frequency of 2 Hz, 2 Guralp CMG-3T broadband sensors with corner frequencies of 0.0083–50 Hz, and 150 Fairfield ZLand 3C nodes with a corner frequency of 5 Hz. We select data from the short-period, broadband, and six nodes recorded over Julian days 236–263 in 2019. Subsequent traces of 20 min recordings were selected based on the realistic surface wave velocity range (0.4 − 2.5 km/s) and interstation distances to compute cross-correlation functions (CCFs) from the vertical component of the raw seismic data, which were band-pass filtered between 1 and 25 Hz in order to focus on the period range of interest ($T < 1$ s). Power spectral density has been computed and plotted for stations (Figure S4 in Supporting Information S1), for which spectral power peaks near 20 Hz, except OPN which peaks at ∼10 Hz. This suggests that additional sources of seismic noise exist near station OPN compared to other stations. Maximum lags for the CCFs were constrained to ±15 s to have flexibility in picking the correct signal window during the dispersion measurement process based on the maximum interstation distance ($\Delta = 4.77$ km).

4. Results

4.1. Synthetic Data

4.1.1. Dispersion Measurement

Preprocessing steps of the white-noise-added seismic data with a time window of approximately 80 s include removing the mean and trend, and applying a 5% cosine taper. TCC-Lin and PCC-PWS methods are applied to the preprocessed data to cross-correlate and stack the CCFs (900 stacks) and develop EGFs for each method. Record-section plots show the moveout of the surface wave as a function of interstation spacing, with similar time-lags of peak correlation among the two methods (Figure 2). Similarity in the moveout of peaks of the two

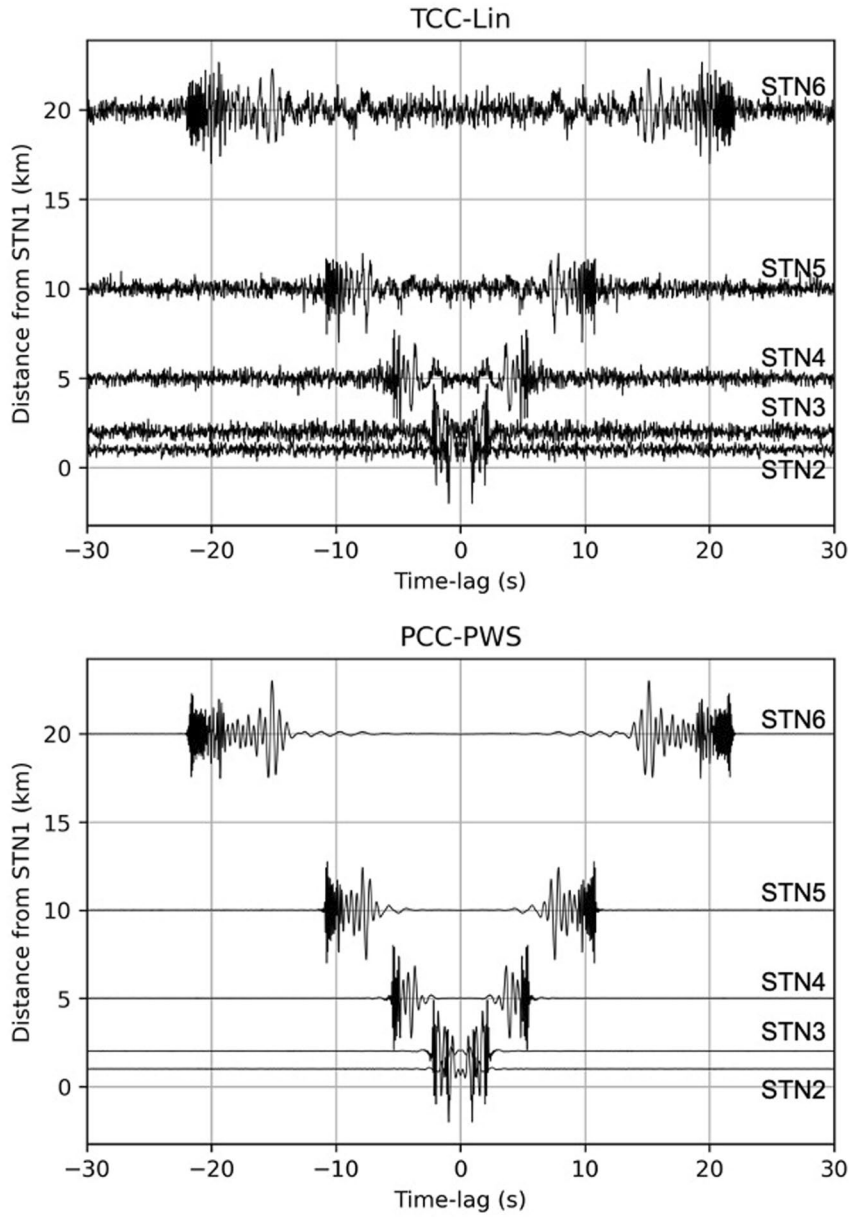

Figure 2. Record-section plot of cross correlation functions determined with two ambient noise data processing methods (TCC-Lin top, PCC-PWS bottom) applied to synthetic seismic data. Cross correlation functions in both panels are results of a total of 900 stacks.

methods indicates that an equivalent depth-averaged group velocity would be measured, but the noticeable difference in the level of noise implies a potential discrepancy in measured velocities. CCFs estimated from the same station pair, but cross-correlated and stacked in the opposite direction are summed and averaged to have uniform EGFs in both positive and negative time-lag directions. This process results in a 'symmetric' signal assuming the two waveforms traveling in opposite directions between a station-pair are homogeneously distributed along the azimuthal path (causal and acausal sequences will ideally be symmetric) (Bensen et al., 2007). The positive side of the symmetric EGFs are then transformed with the scaled and translated Morlet wavelet function, in order to measure the dispersion. From the wavelet transformed EGFs, a reference velocity model is applied to constrain the velocity range of each wave period to extract the velocities with peak spectral energy. The lower limit of period is adjusted based on the Nyquist frequency (50 Hz) of the modeled synthetic surface wave signal. Despite the appearance of noise within the signal window, group velocities are accurately measured from both methods

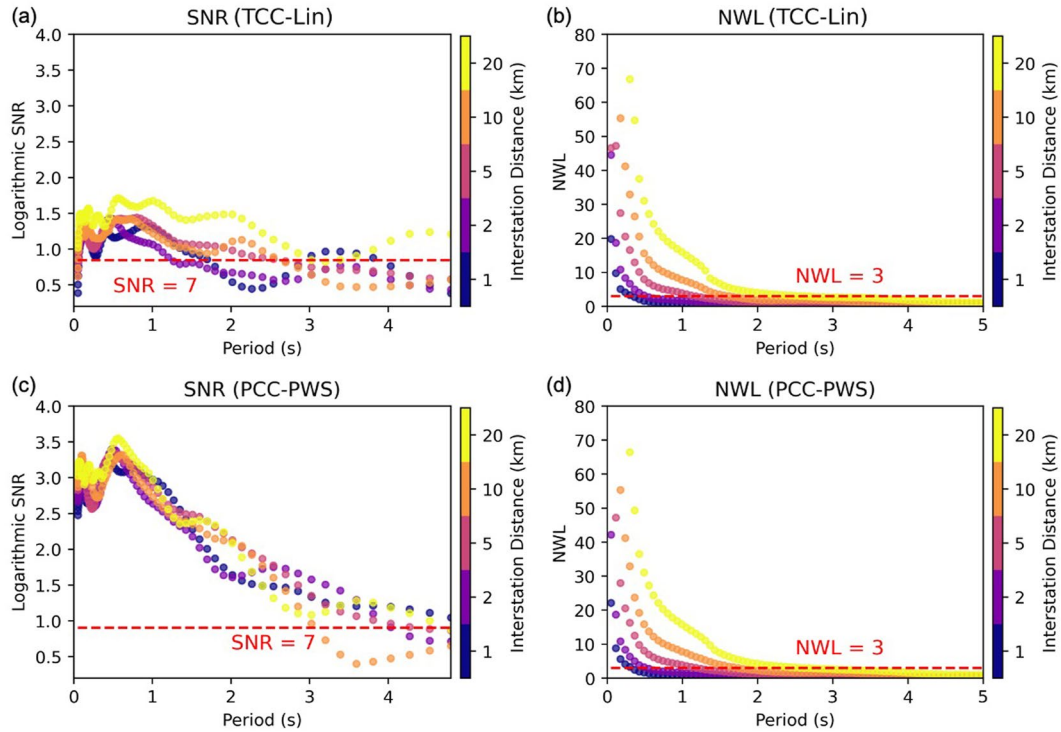

Figure 3. Quality control criteria for the dispersion measurement for each center period and interstation distance. (a and c) Signal-to-noise ratio over the center periods for the two ambient noise methods. (b and d) Number of wavelengths estimated for each center period with coinciding group velocity values.

over the given period distribution for both TCC-Lin and PCC-PWS methods and display velocity curves comparable with the input model (Figure S2 in Supporting Information S1).

4.1.2. Quality of Group Velocity Curves

Raw group velocity curves extracted from the EGFs through the wavelet method incorporate unreliable information that requires further examination of their quality. In the course of measuring surface wave dispersion for individual EGFs without any filtering, we calculate the signal-to-noise ratio (SNR) of the power spectrum at each period. A signal window is determined based on the time-lag (τ) and maximum period ($T_{\max}$) from the lower end of the corner frequencies of the seismic sensor. By selecting the signal window centered at the time-lag (τ_p) in the time domain, where the correlation coefficient is maximum, we set the window of the signal to be $\tau_p - T_{\max}$, $\tau_p + 2T_{\max}$ (Bensen et al., 2007). The noise windows for each EGF are positioned outside the signal window in the positive τ_p direction. Their window width is the same as the signal window without any overlap in time. SNR is calculated by dividing the peak value in the signal window by the standard deviation of a noise window throughout the period range (Figures 3a and 3c). SNR values greater than 7 have been considered high (or acceptable) values from previous studies that used the TCC-Lin method (Bremner et al., 2019; James et al., 2017). Both TCC-Lin and PCC-PWS exhibit the greatest SNR values near $T = 0.6$ s, and SNR values estimated from the EGFs using the PCC-PWS method have larger SNR values for most of the $T < 3$ s indicating stronger detection in coherent signals suggesting a need for an SNR cut-off value greater than 7. The SNR appears to have a linear relationship with interstation distance for 0.2 s $< T <$ 3 s for the TCC-Lin method (Figure 3a) that does not seem observable for the PCC-PWS method (Figure 3c). The number of wavelengths (NWL) of monochromatic waves that propagate between station-pairs are calculated to further test the quality of the dispersion measurement between the TCC-Lin and the PCC-PWS methods. We use the seismic velocity measured at each T to calculate the NWL for each station pair with discrete distances (NWL $= \Delta/(1\ \text{km/s} \times T)$). NWL exhibits a clear pattern with decreasing values with greater T and shorter distance (Δ) (Figures 3b and 3d). A generalized lower NWL (LNWL) cutoff value of 3 has been widely used, which is considered practical in obtaining a reliable group velocity curve, but a standard upper cutoff for NWL (UNWL) is not well established (Bensen et al., 2007; Luo et al., 2015; Toledo et al., 2022).

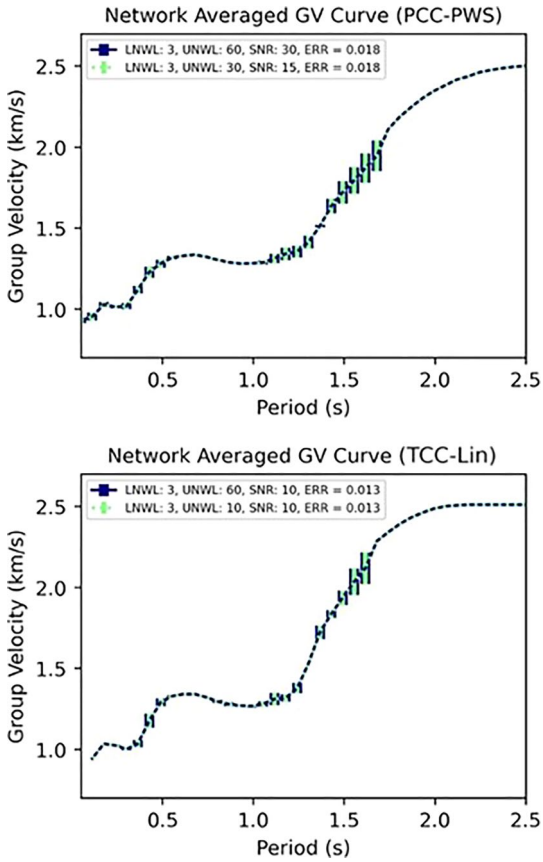

Figure 4. Network-averaged group velocity curves with the 2 smallest error values from a total of 80 combinations in quality control parameters. The velocities extracted using the TCC-Lin and PCC-PWS methods show similar profiles with greater error values at 1.5 s < T < 2 s. LNWL value remains constant for the 5 velocity curves indicating the dominant effect among the 3 parameters.

In order to find the optimal combination of SNR, LNWL, and UNWL values, we explore different values of SNR (5–10 for TCC-Lin and 5, 10, 15, 20, 25, 30 for PCC-PWS), LNWL (0.5, 1, 1.5, 2, 2.5, 3), and UNWL (10, 20, 30, 40, 50, 60) and estimate the error from the averaged group velocity curve over all the station-pairs as the mean of the standard deviation of the velocities over the period range. Selected range of quality control parameters are extended from single values used or discussed in previous studies (Bensen et al., 2007; James et al., 2017; Luo et al., 2015). The average group velocity curves with the 2 lowest errors are shown in Figure 4. Although the combinations (a total of 216) of the quality control parameters are different between the two methods, the error values are similar (TCC-Lin: 0.071–0.124, PCC-PWS: 0.072–0.12). However, the group velocity curves with the lowest error for both methods share the same NWL parameters, although with different SNR values, and they also have the smallest root-mean-square error (RMSE) away from the reference input model (PCC-PWS: 0.14, TCC-Lin: 0.16). The lack of station pairs (5 synthetic models) limits the number of velocity measurements for T > 2.7 s to provide a reliable averaged group velocity curve. Nevertheless, both TCC-Lin and PCC-PWS methods correspond well with the group velocity curves estimated from FTAN. The input depth-velocity model directly converted into a group velocity curve shows a very similar profile for T < 2 s (Figure S3 in Supporting Information S1). The comparable group velocity curves among the three measurements and models demonstrate the capacity of the ambient noise methods described above to adequately estimate group velocity structures with high frequency (>1 Hz) surface waves and regional (<20 km) interstation spacing.

4.2. Real Observational Data

4.2.1. Dispersion Measurement

From a total of 171 station-pairs, excluding the auto-correlations, 20-min CCFs were computed and stacked with TCC-Lin and PCC-PWS methods. The final stacked CCFs (EGFs) have an average of 1930 stacks with a standard deviation of 23 stacks, where erroneous or empty traces were ignored during the stacking process. Record section plots of the EGFs computed from the two methods (Figure 5) are more complicated than the synthetic data, likely a result of the more heterogeneous subsurface structure, but there is a clear noise contrast between the TCC-Lin and PCC-PWS methods, similar to the results with the synthetic data. The time-lag position of the peak coefficient from the raw EGFs are similar among the two data processing methods, indicating that the velocity measurements should be within an acceptable error range. Based on the alignment of the time-lags at the peaks of EGFs as a function of Δ (offset distance in km from station OPN in Figure 5), the moveouts that define the frequency and network-averaged group velocity of the seismic Rayleigh waves within the spatial coverage of the 19 stations is approximately 1.1 km/s.

In preparation for dispersion measurements, we follow equivalent steps to the synthetic data averaging the causal and acausal sequences of the EGFs. The CWT is applied to the causal sequence of the EGFs and is filtered based on a reference velocity structure to constrain the range of velocities for each frequency, avoiding unrealistic measurements outside the 0.4–2.5 km/s velocity range (Bremner et al., 2019; Fores et al., 2018; James et al., 2017). Although the corner frequencies of the L22 sensors and nodes are 2 and 5 Hz, respectively, we incorporate frequencies down to 1 Hz where more energy is required to produce the same response as the frequencies above the corner frequencies. Group velocity curves based on dispersion measurements of the EGFs estimated from the two methods share a similar profile above T ~ 0.2 for different station-pairs (Figure S5 in Supporting Information S1). Moreover, comparable group velocity profiles among station-pairs indicate a similar geologic structure shared within the survey area. However, velocities of periods less than 0.2 s often exhibit discontinuity and high complexity that depends on the station-pair distance and the analysis method. Discontinuities in the group velocity curves may arise from detection of multi-modal signals that incorporate phase shifts or may

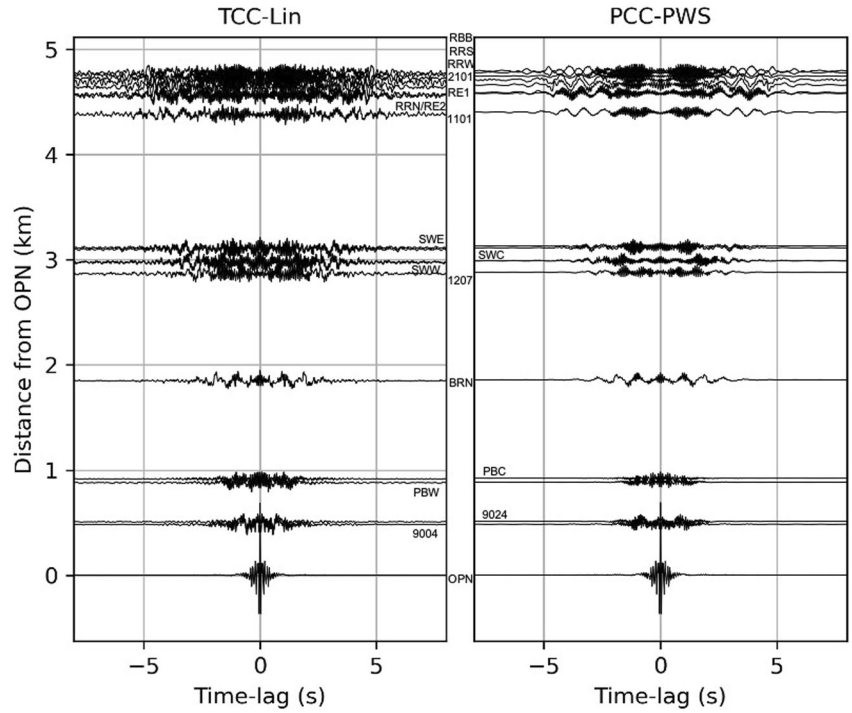

Figure 5. Record-section plot showing cross-correlation functions (CCFs) for a subset of the station-pairs using two different data processing methods (TCC-Lin left, PCC-PWS right). The CCFs are determined by averaging from approximately 1930 stacks over the 1 month survey period.

be caused by multiple scattering of surface waves from local heterogeneities (Aki & Chouet, 1975; Levshin & Panza, 2006; Shapiro & Campillo, 2004; Vernon et al., 1998), something that we would expect given our study area of complex karst structure.

4.2.2. Quality of Group Velocity Curves

Group velocity curves extracted from the dispersion measurements are examined for each period with 3 quality control parameters with an extended range of cutoff values implemented from the synthetic data of SNR (5–20 for TCC-Lin and 20–50 for PCC-PWS), LNWL (0.5–3.5), and UNWL (5–70) (Figure 6). Similar to the results calculated with the synthetic data set, SNR values of the PCC-PWS method can be several orders of magnitude greater than the TCC-Lin method. SNR values appear to increase for $T < 0.7$ s with decreasing Δ, while displaying an opposite trend for $T > 0.7$ s. $T \sim 0.1$ and 0.23 s for both TCC-Lin and PCC-PWS methods have peaks in SNR with a decreasing trend in both higher and lower periods. The lowest SNR values occur at $T \sim 0.04$ and 0.14 s for both methods, where the network-averaged group velocity will be altered the most by varying the SNR cutoff value. NWL values share an analogous trend between the TCC-Lin and PCC-PWS methods, where an increase in Δ and decrease in T results in an increase in NWL of a monochromatic seismic wave (Figure 6). Predicting the effect of adjusting the LNWL is clear, as most of the group velocity measurements at $T > 0.5$ s with short Δ (<1 km) will be discarded during the quality control process, although the effect of the UNWL requires examination of the quality controlled velocity curves. Over 92% and 86% of NWL for $\Delta < 1$ km reside in NWL less than 3 for TCC-Lin and PCC-PWS, respectively. However, the optimal value of LNWL may change based on the complexity of the subsurface structure, as surface waves with lower T have a greater chance of experiencing multiple scattering that may appear as multiple travel time arrivals. Station-pairs with 2 km $< \Delta <$ 3 km show a peak in SNR near NWL = 10, and two peaks at NWL = 7 and 19 for $\Delta > 3$ km. UNWL controls the lower threshold of T, and station-pairs with greater Δ are affected. UNWL should be distinguished between the two methods as PCC-PWS for $\Delta < 1$ km to incorporate more NWL points occupied in NWL > 3.

In order to evaluate changes in the network-averaged group velocity curves based on the selection of the quality control parameters, we constrain two parameters and explore the effect governed by a single parameter (Figure 7). Increasing SNR values results in a decrease in mean error (Table S2 in Supporting Information S1),

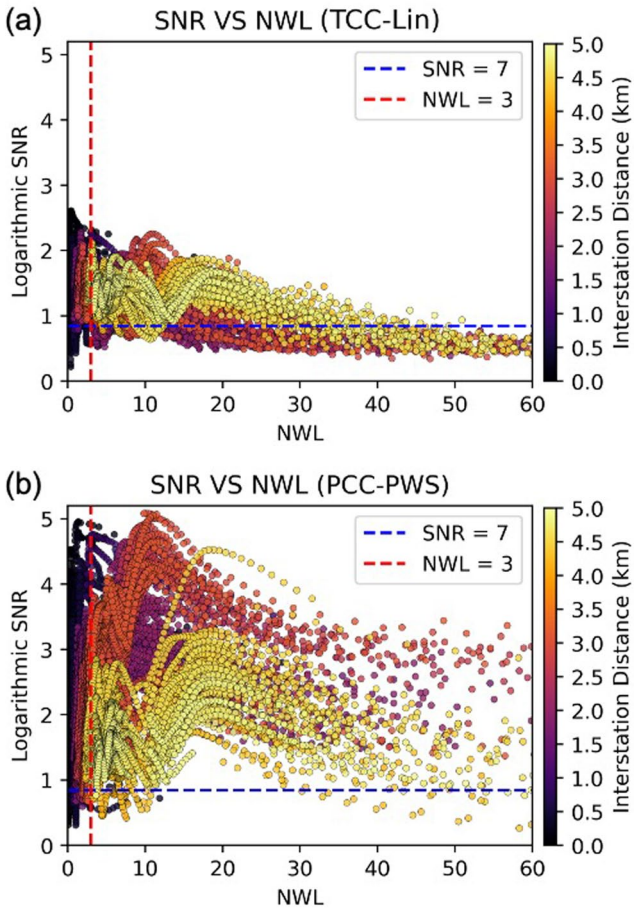

Figure 6. Quality control parameters (SNR and NWL) calculated for each station-pair over the period distribution. SNR = 7 and NWL = 3 are typical values used in previous ambient noise studies to limit empirical Green's functions by quality (Bensen et al., 2007; James et al., 2017).

and velocities at $T < 0.2$ s of the PCC-PWS method appear higher than the TCC-Lin method by changing the SNR value (Figures 7a and 7b). Averaged group velocity curves of the PCC-PWS method controlled by varying SNR and constraining the LNWL and UNWL at their lowest and highest extreme, respectively, exhibit nearly indistinguishable group velocity curves (Figure 7a). Hence, we select the lowest SNR cutoff of 20 to be a constant to incorporate dispersion measurements to the maximum extent while varying NWL. Group velocity curves of the TCC-Lin method display a marginal difference at $T < 0.18$ s and $T > 0.79$ s and slight divergences from the smooth curve at $T = 0.54$ and 0.57 s (Figure 7b). Therefore, the SNR value of 7 was selected as an invariable parameter, which displays the lowest error, and deviation is not observed. LNWL affects longer periods ($T > 0.6$ s) for both methods, and causes an increase in estimation of velocities with increasing LNWL values and decreasing error values (Figures 7c and 7d and Table S3 in Supporting Information S1). We continue varying the UNWL with a generalized constant LNWL value of 3 that produces one of the lowest mean errors (Table S4 in Supporting Information S1) (Bensen et al., 2007; Luo et al., 2015). Variance in velocities measured with different UNWL values appear at $T < 0.7$ s, and the velocity curve estimated with a UNWL value of 5 clearly departs from greater values of UNWL excluding velocities at $0.35 \text{ s} < T < 0.45$ s and $T > 0.7$ s (Figures 7e and 7f). Removing velocity measurements with NWL greater than 5 leads to removing approximately 31% of the capacity and drops to 15% when controlling the quality with a UNWL value of 10. Averaged velocity measurements appear to stabilize with UNWL values greater than 10.

Group velocity curves quality-controlled with constrained threshold values between the PCC-PWS (SNR: 20, LNWL: 3, UNWL: 10) and TCC-Lin (SNR: 7, LNWL: 3, UNWL: 10) methods depict comparable outputs, demonstrating the reliability of both PCC-PWS and TCC-Lin data processing methods in extracting EGFs with short-period waves ($T < 1$ s) (Figure 8). However, greater velocities up to 0.11 km/s are measured from the PCC-PWS method for $T > 0.38$ s and $T \sim 0.15$ s with the specific combination of quality control parameters. The differences in velocity measurements among the data processing methods indicates that careful decision making is necessary when selecting the quality control parameters, especially with seismic data collected above a complex geologic medium.

5. Discussion

5.1. Ambient Noise Interferometry of Short-Period Waves

We compare the robustness of the EGFs extracted from applying two ambient noise data processing methods of short-period ($T < 1$ s) waves. The philosophy of the PCC-PWS method is based on coherence of the time-frequency dependent phase of the analytic signals, and the traditional TCC-Lin method detects the largest sum of amplitudes, which is associated with energy, as coherent seismic signals (Bensen et al., 2007; Haned et al., 2016; Schimmel, 1999; Schimmel et al., 2011; Ventosa et al., 2017). The PCC method has greater waveform sensitivity than the TCC method, but has disadvantages when the phase signal is corrupted by high amplitude noise. Therefore, the most appropriate surface wave data processing method depends on the waveform signal characteristics (Schimmel & Paulssen, 1997; Schimmel et al., 2011). Application of the TCC-Lin method to synthetic data shows higher noise levels within the signal window compared to the PCC-PWS method (Figure 2). This demonstrates the effectiveness of the PCC-PWS method in detecting phase coherent signals and attenuating incoherent signals associated with the random noise input, which is also shown in the wavelet spectrogram (Figure S2 in Supporting Information S1). In addition to the amplitude-unbiased advantage of the PCC method and minimal alteration of the waveforms, the time-frequency PWS is also an important step which improves the signal extraction performance and azimuthal coverage (Baig et al., 2009; Schimmel et al., 2011).

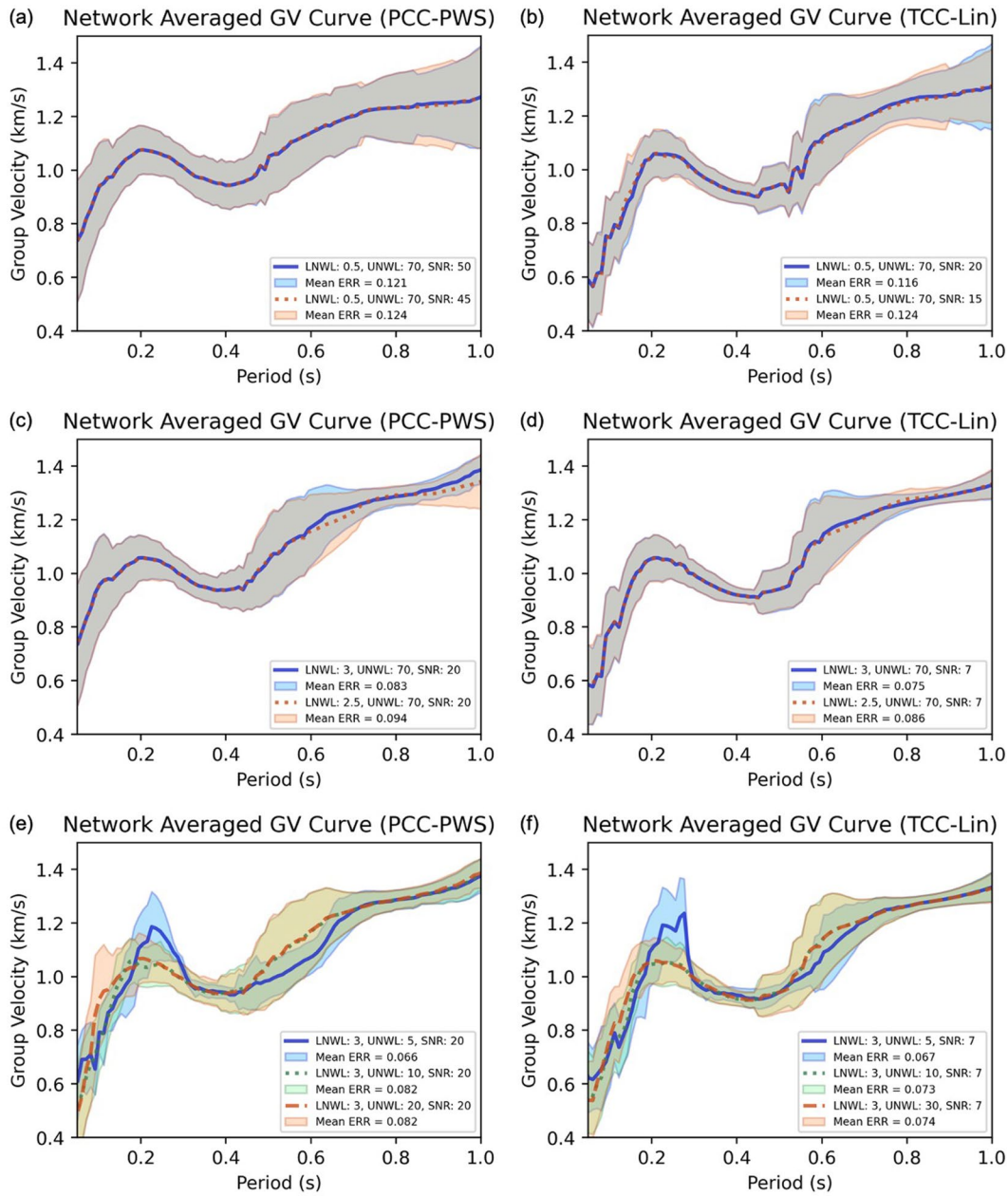

Figure 7. Quality-controlled network-averaged group velocity curves by varying (a and b) signal-to-noise ratio (SNR), (c and d) lower number of wavelengths (LNWL), and (e and f) upper number of wavelengths (UNWL).

Real data application using the TCC-Lin and PCC-PWS methods produced similar noise levels compared to the synthetic results, where the PCC-PWS data processing steps resulted in an improved detection in coherent signals with reduced noise than the TCC-Lin method (Figure 5). Measurement of dispersion within the short-period range (0.04 s < T < 1 s) was complicated, especially below 0.2 s, in contrast to the results of the synthetic data implementation where deviation of the dispersion curves among the two methods were greater in longer periods (2.8 s < T < 4 s) (Figures S2 and S5 in Supporting Information S1). However, the emergence of complexity and discontinuity of spectral peaks for T < 0.2 s in real data is expected due to the highly heterogeneous geologic structure of the karst aquifer. Such structure likely causes multi-scattering of surface waves, as well as multi-modal signals that incorporate shifts in phase of the seismic waveforms (Aki & Chouet, 1975; Levshin & Panza, 2006; Pedersen & Krüger, 2007; Pilz et al., 2012; Shapiro & Campillo, 2004; Vernon et al., 1998).

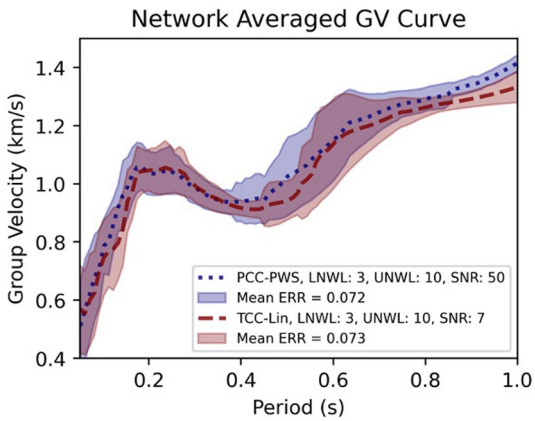

Figure 8. Results of network-averaged (171 station-pairs) group velocity curves examined with specific quality control parameter values for the PCC-PWS method (blue) and TCC-Lin method (red). Periods (T) greater than 0.38 s depict higher velocity measurements for the PCC-PWS method.

The erroneous dispersion measurements suggest the need for careful quality control (SNR and NWL), which is related to the surface wave velocity and interstation distances, prior to measuring the network-averaged group velocity curves. In addition to the complexity of dispersion measurements, the EGFs also incorporate directional and distance related biases based on the network position and the frequency-dependent azimuthal direction of the noise sources (Figure S6 in Supporting Information S1) (Luo et al., 2015). However, we assume the waveforms are symmetric along the path between station-pairs and focus on the comparison between the two methods and the effect of the quality control parameters on the network-averaged group velocity curves.

5.2. Control on Quality of Dispersion Measurements

From the spectral dispersion analysis, SNR and NWL values are estimated for each station-pair and period and evaluated based on the predefined threshold values (Figure 3). Both averaged group velocity curves between periods of 1.5 and 1.8 s exhibit the greatest errors (Figure 4), and the magnitude of velocity residual relative to average velocity for $T > 1.5$ s (Figure S3 in Supporting Information S1) could potentially result from (a) the lack of station-pairs that have interstation spacing between 10 and 20 km, and (b) random noise that altered the synthetic surface wave signal. Moreover, measurements of velocities of $T > 2.7$ s are limited due to the absence of station-pairs with distances above 20 km. The constant value of LNWL = 3 among the 5 velocity curves with the lowest errors suggests that the LNWL has the most impact on the velocity measurements. The constraint on the maximum wave period for each station-pair distance based on the LNWL values is shown in Figure 9. A general decreasing trend is observed between maximum T and LNWL for both PCC-PWS and TCC-Lin methods in synthetic and real data applications. For the synthetic data application, maximum T increases logarithmically with increasing interstation distances (Figures 9a and 9c), while a linear trend is observed for real data (Figures 9b and 9d), although a better distance resolution is required to confirm the data-specific relationship linked to the design of the seismic network.

For the field data analysis, measurements of SNR have shown that the more distant station-pairs produce velocity measurements with relatively greater SNR at higher periods ($T > 0.7$ s), although the SNR values decrease with distance for $T < 0.7$ s (Figures 7a and 7b). In general, the LNWL from our quality control results suggests that the cutoff of 3 provides the network-averaged group velocity curve with the lowest mean error along the period distribution, which agrees with the conventional three-wavelength requirement (Figures 7c and 7d) (Bensen et al., 2007). However, verified by previous studies, individual dispersion measurements processed with the PCC method, especially with short-separation paths, are capable of producing reliable measurements with lower LNWL threshold than 3 (Luo et al., 2015). Nevertheless, LNWL most likely limits station-pairs with shorter interstation distances controlling periods greater than $T > 0.6$ s. Control on UNWL value over the predefined range influences velocity measurements mostly for $T < 0.7$ s (Figures 7e and 7f). The observed differences can be associated with the highly heterogeneous shallow karst structure leading to a spatial variation in wave attenuation rates for the short-period ($T < 1$ s) surface waves (Hackert & Parra, 2003; He & Cai, 2012). Moreover, similar to the synthetic results, the lack of spatial coverage can result in deviations in velocities with larger errors. Therefore, constraint on the UNWL under this condition should not only be resolved with the mean error of the network-averaged group velocity curve, but should also be carefully resolved with the site-specific wave attenuation rates.

Network-averaged group velocity curves with the smallest and largest errors, extracted from varying all three quality control parameters, illustrate corresponding velocity profiles with a maximum difference of 0.12 km/s ($T \sim 0.2$ s), which is within the error range of both results (Figure S9 in Supporting Information S1). Therefore, smaller error values do not necessarily indicate a more reliable velocity curve, unless the subsurface medium is spatially homogeneous, and should be carefully selected based on the information of the geologic structure. The range of reliable wave periods differ based on the interstation spacing, and the deviations in group velocity measurements of different station-pairs reflect the path-dependent surface wave dispersion from extreme

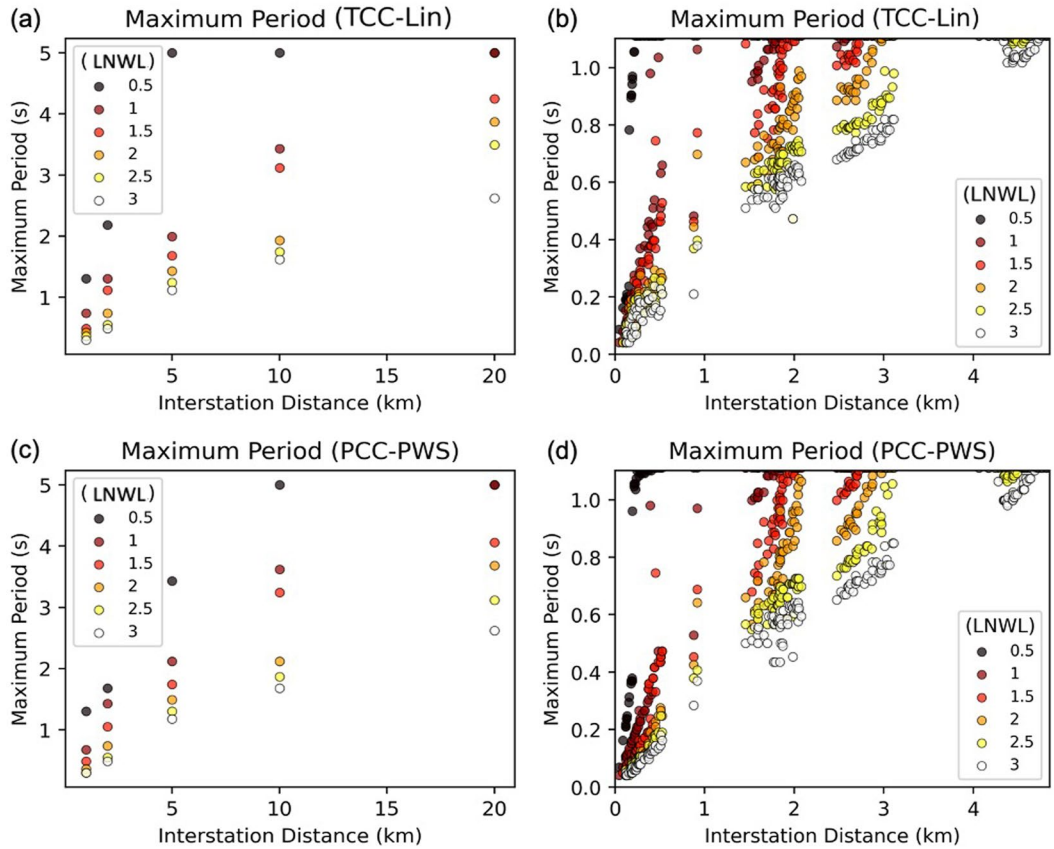

Figure 9. Maximum period associated with the lower threshold of the number of wavelengths (LNWL) required to propagate between the interstation spacings. The left column (a, c) are the results obtained from the synthetic data, and the right column (b, d) are results from the real data application.

heterogeneity of the carbonate crust in Florida. Moreover, the undulating velocity curves may be resulting from spurious precursory arrivals interfering with the primary signals (Bensen et al., 2007).

5.3. Inversion of Group Velocity Curves

Inverted seismic velocity models can provide basic information of the geology without a known velocity structure. We invert for 1-dimensional velocity models using the network-averaged group velocity curves derived from the synthetic data and compare with the input velocity model (Table S1 in Supporting Information S1) used to generate the synthetic Green's functions for the initial testing. The velocity models computed from the synthetic data are similar, with generally increasing velocity with depth (Figure S7 in Supporting Information S1). However, the difference between the original input and velocity models computed by inverting our group velocities indicate the effect of signal detection from the white noise introduced to the synthetic waveform. Moreover, the small number of seismic stations limits accurate recovery of the input velocity model.

We do the same inversion for a 1-dimensional velocity model determined using the network-averaged group velocity curves found for the real data set. This resulting velocity structure can be interpreted in light of the known geologic structures of the site (Williams & Kuniansky, 2015). Shallow layers (<15 m) required a steep decrease in velocity to produce a misfit smaller than 0.1 that would account for the sediment-bedrock boundary below O'Leno and River Rise Preserve State Parks, Florida, USA, producing greater misfit than deeper structures. Layers beneath the sediments down to ∼300 m, are comprised of highly permeable carbonate rocks of the Floridan aquifer system that results in a decrease in seismic velocity compared to deeper structure, which is also observed in our inversion models (Figure S8 in Supporting Information S1) (Fores et al., 2018; James et al., 2017; Williams & Kuniansky, 2015). However, the inversion process does not incorporate the variances

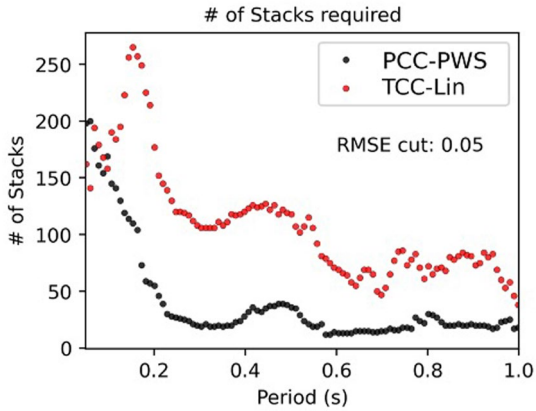

Figure 10. The average number of stacks required for the empirical Green's functions to stabilize for all station-pairs at each center period. $T < 0.2$ s from the TCC-Lin include high uncertainty, also shown in the dispersion measurements.

of the velocity measurements at different frequencies when estimating for the network-averaged velocity curve. Therefore, the same amount of errors in velocity should be added to the inversion model to better account for the spatial structure below the seismic network (XS).

5.4. Reproducibility of the Empirical Green's Functions

Another useful procedure to estimate reliability of group velocity measurements is to assess the temporal reproducibility as the physical properties of the elastic medium and the major source of ambient noise seismic signals may change in both space and time (Bensen et al., 2007; Yang et al., 2007). Variation in ambient noise conditions is non-trivial, and environmental seismology studies have utilized this occurrence to monitor spatiotemporal changes in seismic velocity (dv/v) (Clarke et al., 2011; Clements & Denolle, 2018; Ratdomopurbo & Poupinet, 1995). The fundamental procedure in measuring dv/v is to compare the phase shift between a reference EGF that represents the average background velocity of the study area and the current EGF that holds information of the current state of the velocity structure that varies through time (Clarke et al., 2011; Hillers et al., 2014; Mordret et al., 2010). In order to measure accurate dv/v, the reference and current EGFs must also have sufficient similarity (Clarke et al., 2011). In the course of cross-correlation and stacking, we output progressive EGFs after each additional stack to compare the emergence of EGFs among the TCC-Lin and PCC-PWS methods. RMSE between the progressive group velocity curves and the group velocity curve extracted from the fully stacked EGFs were estimated to find the average number of stacks needed to reach sufficient similarity (RMSE = 0.05) for each wave period (Figure 10). Assuming the 1-month stacked EGF represents a stable group velocity structure, the required number of stacks shows a decreasing trend with increase in wave period. The TCC-Lin method required an average of 125 stacks to reproduce a stable velocity curve sufficiently similar to the 1-month stacked result, while an average of 69 stacks was needed for the PCC-PWS method. Differences in the necessary number of stacks between the two methods are in line with the progressive SNR calculations (Figure S10 in Supporting Information S1), where the PCC-PWS method requires less number of stacks for stable signal emergence. Irregularities are found in several wave periods ($T = 0.47$ and 0.8 s for the PCC-PWS method and $T = 0.18$ s, 0.42 s, 0.76 s, and 0.9 s for the TCC-Lin method (Figure 10) that could be a result from lack of reliable velocity measurements at certain periods due to insufficient station-pair distances Figure S6 in Supporting Information S1). Nonetheless, our findings suggest that the PCC-PWS method would be a preferable method in the application of monitoring studies to track dv/v with a better time resolution.

6. Conclusion

To test the validity and effectiveness in detecting short-period ($T < 1$ s) ambient noise signals, two different cross-correlation methods based on the amplitude (TCC) and phase coherence (PCC) of ambient noise signals accompanied by distinct stacking methods (Lin and PWS) were examined with synthetic and real seismic data measured at O'Leno and River Rise Preserve State Parks, Florida, USA. Both methods were capable of generating the empirical Green's functions (EGFs) with a maximum peak positioned at a comparable time-lag, but the apparent noise amplitude within the signal window indicates that the PCC-PWS method is better in detecting coherent signals for both synthetic and real data. However, high heterogeneity of the shallow subsurface structure at the field site introduced complications to the dispersion measurement process, especially at the lowest periods ($T < 0.2$ s). The complexity of the carbonate aquifer could lead to erroneous velocity extraction from inclusion of multi-scattered and multimodal surface wave signals for both TCC-Lin and PCC-PWS methods, suggesting the necessity of the quality control process subsequent to the dispersion measurement.

From the spectral dispersion analysis, various combinations of signal-to-noise ratio (SNR) and the upper and lower limits of the number of wavelengths traveled between station-pairs (UNWL and LNWL) were explored to identify the effects of each quality control parameters on the extraction of group velocities over the period distribution. A constant value of LNWL = 3 applied to develop the averaged group velocity curves of the synthetic data resulted in the smallest mean errors and was most similar to the input reference model, suggesting that the

LNWL is the dominating effect on extracting reliable group velocities of our modeled surface wave data. Also, our results show larger errors for wave periods 1.5 s < T < 2 s for both PCC-PWS and TCC-Lin methods, which likely have resulted from (a) random noise inserted to the surface wave signal, and (b) insufficient interstation distance coverage. Although SNR is effective in examining the robustness of the empirical Green's functions (EGFs), NWL had greater impact on differences in the network-averaged group velocity curves from the ambient noise data measured at O'Leno State Park. Varying the UNWL in particular generated high divergence at 0.2 s < T < 0.3 and 0.45 s < T < 0.7 s potentially resulting from the high heterogeneity of the carbonate structure and the lack of spatial coverage, analogous to the synthetic data application. Regardless of the method in extracting coherent signals, both PCC-PWS and TCC-Lin methods output comparable group velocity curves of the lowest errors with the same value of LNWL = 3, indicating the effectiveness of the quality control process. This also results in a similar velocity structure of the two methods from inverting the dispersion curves. However, network-averaged group velocity curves with both low and high mean errors are within the standard deviation of velocities extracted from all station pairs. Therefore, smaller error does not always indicate a more reliable group velocity curve, especially for complex geologic structures with high heterogeneity. Error values of the averaged group velocity curves should also be applied when inverting for the velocity structure. Quality control parameters should be carefully selected based on the signal characteristic and information of the geology.

Reproducibility of EGFs was examined for the two data processing methods to account for the spatiotemporal changes in the complex elastic medium as well as the ambient noise sources. The PCC-PWS method required fewer stacks (or time) over the wave periods (69 stacks) to measure reliable group velocities than the TCC-Lin method (125 stacks), further demonstrating the efficiency and robustness in locating coherent ambient noise signals. Our results for the number of stacks needed to extract stable group velocities may be used as a baseline in producing timely EGFs to compare with a background EGF for environmental seismology studies. However, the site-specific reproducibility will vary based on the complexity of the geologic structure in monitoring changes in group velocities, and waveform signal biases based on the directions and distances of the seismic network need to be further investigated to improve uncertainties in measuring the time and frequency dependent surface wave velocities.

Data Availability Statement

The seismic data collected at O'Leno and River Rise Preserve State Parks, Florida, USA, are available through the International Federation of Digital Seismograph Networks website (FDSN network code XS, https://doi.org/10.7914/SN/XS_2018) (Bilek, 2018). The codes for controlling quality control parameters and computing reproducibility of dispersion measurements in this paper can be found online (at https://doi.org/10.5281/zenodo.7582572) (Woo, 2023).

Acknowledgments
We are grateful for Madison Flint, Andrew Oberhelman, Joel (Lucas) Cuevas, Tatiana Summerall, Krista Van Der Velde, Jared Ciarico, Lauren Madsen, Calvin Block, Riccardo Ansaldi, Alexander Janelle, Emily Morton, Pnina Miller, and Paul Bremner for their assistance in installation of equipment, conducting field surveys, and collecting data. We also thank the anonymous reviewers for their thorough feedback. We also acknowledge the Florida Park Service and personnel at O'Leno and River Rise Preserve State Parks who made this project possible. This material is based upon work supported by the National Science Foundation under Grant No.1850667. Seismic instruments were provided by the Incorporated Research Institutions for Seismology (IRIS) through the PASSCAL Instrument Center at New Mexico Tech. The facilities of the IRIS Consortium are supported by the National Science Foundation's Seismological Facilities for the Advancement of Geoscience (SAGE) Award under Cooperative Support Agreement EAR-1851048.

References

Aki, K., & Chouet, B. (1975). Origin of coda waves: Source, attenuation, and scattering effects. *Journal of Geophysical Research*, *80*(23), 3322–3342. https://doi.org/10.1029/jb080i023p03322

Andajani, R. D., Tsuji, T., Snieder, R., & Ikeda, T. (2020). Spatial and temporal influence of rainfall on crustal pore pressure based on seismic velocity monitoring. *Earth Planets and Space*, *72*(1), 1–17. https://doi.org/10.1186/s40623-020-01311-1

Baig, A. M., Campillo, M., & Brenguier, F. (2009). Denoising seismic noise cross correlations. *Journal of Geophysical Research*, *114*(B8), B08310. https://doi.org/10.1029/2008jb006085

Bensen, G., Ritzwoller, M., Barmin, M., Levshin, A. L., Lin, F., Moschetti, M., et al. (2007). Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements. *Geophysical Journal International*, *169*(3), 1239–1260. https://doi.org/10.1111/j.1365-246x.2007.03374.x

Bilek, S. L. (2018). Geophysical characterization of karst aquifers using dynamic recharge events [Dataset]. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/XS_2018

Boué, P., Roux, P., Campillo, M., & Briand, X. (2014). Phase velocity tomography of surface waves using ambient noise cross correlation and array processing. *Journal of Geophysical Research: Solid Earth*, *119*(1), 519–529. https://doi.org/10.1002/2013jb010446

Bremner, P. M., Panning, M. P., Russo, R., Mocanu, V., Stanciu, A. C., Torpey, M., et al. (2019). Crustal shear wave velocity structure of central Idaho and eastern Oregon from ambient seismic noise: Results from the IDOR project. *Journal of Geophysical Research: Solid Earth*, *124*(2), 1601–1625. https://doi.org/10.1029/2018jb016350

Campillo, M., & Paul, A. (2003). Long-range correlations in the diffuse seismic coda. *Science*, *299*(5606), 547–549. https://doi.org/10.1126/science.1078551

Chávez-García, F. J., & Luzón, F. (2005). On the correlation of seismic microtremors. *Journal of Geophysical Research*, *110*(B11), B11313. https://doi.org/10.1029/2005jb003671

Clarke, D., Zaccarelli, L., Shapiro, N., & Brenguier, F. (2011). Assessment of resolution and accuracy of the moving window cross spectral technique for monitoring crustal temporal variations using ambient seismic noise. *Geophysical Journal International*, *186*(2), 867–882. https://doi.org/10.1111/j.1365-246x.2011.05074.x

Clements, T., & Denolle, M. A. (2018). Tracking groundwater levels using the ambient seismic field. *Geophysical Research Letters*, *45*(13), 6459–6465. https://doi.org/10.1029/2018gl077706

Dantas, O. A. B., do Nascimento, A. F., & Schimmel, M. (2018). Retrieval of body-wave reflections using ambient noise interferometry using a small-scale experiment. *Pure and Applied Geophysics*, *175*(6), 2009–2022. https://doi.org/10.1007/s00024-018-1794-0

De Plaen, R. S., Cannata, A., Cannavo, F., Caudron, C., Lecocq, T., & Francis, O. (2019). Temporal changes of seismic velocity caused by volcanic activity at Mt. Etna revealed by the autocorrelation of ambient seismic noise. *Frontiers in Earth Science*, *6*, 251. https://doi.org/10.3389/feart.2018.00251

Derode, A., Larose, E., Tanter, M., De Rosny, J., Tourin, A., Campillo, M., & Fink, M. (2003). Recovering the Green's function from field-field correlations in an open scattering medium (l). *Journal of the Acoustical Society of America*, *113*(6), 2973–2976. https://doi.org/10.1121/1.1570436

Dias, R. C., Julià, J., & Schimmel, M. (2015). Rayleigh-wave, group-velocity tomography of the Borborema Province, NE Brazil, from ambient seismic noise. *Pure and Applied Geophysics*, *172*(6), 1429–1449. https://doi.org/10.1007/s00024-014-0982-9

Ekstrom, G., Abers, G. A., & Webb, S. C. (2009). Determination of surface-wave phase velocities across USArray from noise and Aki's spectral formulation. *Geophysical Research Letters*, *36*(18), L18301. https://doi.org/10.1029/2009gl039131

Fichtner, A., Kennett, B. L., Igel, H., & Bunge, H.-P. (2008). Theoretical background for continental-and global-scale full-waveform inversion in the time–frequency domain. *Geophysical Journal International*, *175*(2), 665–685. https://doi.org/10.1111/j.1365-246x.2008.03923.x

Fores, B., Champollion, C., Mainsant, G., Albaric, J., & Fort, A. (2018). Monitoring saturation changes with ambient seismic noise and gravimetry in a karst environment. *Vadose Zone Journal*, *17*(1), 1–12. https://doi.org/10.2136/vzj2017.09.0163

Hackert, C. L., & Parra, J. O. (2003). Estimating scattering attenuation from vugs or karstsscattering attenuation from vugs. *Geophysics*, *68*(4), 1182–1188. https://doi.org/10.1190/1.1598111

Haned, A., Stutzmann, E., Schimmel, M., Kiselev, S., Davaille, A., & Yelles-Chaouche, A. (2016). Global tomography using seismic hum. *Geophysical Journal International*, *204*(2), 1222–1236. https://doi.org/10.1093/gji/ggv516

Harmon, N., Rychert, C., & Gerstoft, P. (2010). Distribution of noise sources for seismic interferometry. *Geophysical Journal International*, *183*(3), 1470–1484. https://doi.org/10.1111/j.1365-246x.2010.04802.x

He, Y., & Cai, J. (2012). Q tomography towards true amplitude image and improve sub-karst image. In *2012 seg annual meeting*.

Herrmann, R. B. (1973). Some aspects of band-pass filtering of surface waves. *Bulletin of the Seismological Society of America*, *63*(2), 663–671. https://doi.org/10.1785/bssa0630020663

Herrmann, R. B. (2013). Computer programs in seismology: An evolving tool for instruction and research. *Seismological Research Letters*, *84*(6), 1081–1088. https://doi.org/10.1785/0220110096

Hillers, G., Campillo, M., & Ma, K.-F. (2014). Seismic velocity variations at TCDP are controlled by MJO driven precipitation pattern and high fluid discharge properties. *Earth and Planetary Science Letters*, *391*, 121–127. https://doi.org/10.1016/j.epsl.2014.01.040

James, S. R., Screaton, E. J., Russo, R. M., Panning, M. P., Bremner, P. M., Stanciu, A. C., et al. (2017). Hydrostratigraphy characterization of the Floridan aquifer system using ambient seismic noise. *Geophysical Journal International*, *209*(2), 876–889. https://doi.org/10.1093/gji/ggx064

Jiang, C., & Denolle, M. A. (2020). NoisePy: A new high-performance python tool for ambient-noise seismology. *Seismological Research Letters*, *91*(3), 1853–1866. https://doi.org/10.1785/0220190364

Larose, E., Derode, A., Clorennec, D., Margerin, L., & Campillo, M. (2005). Passive retrieval of Rayleigh waves in disordered elastic media. *Physical Review*, *72*(4), 046607. https://doi.org/10.1103/physreve.72.046607

Lecocq, T., Longuevergne, L., Pedersen, H. A., Brenguier, F., & Stammler, K. (2017). Monitoring ground water storage at mesoscale using seismic noise: 30 years of continuous observation and thermo-elastic and hydrological modeling. *Scientific Reports*, *7*(1), 1–16. https://doi.org/10.1038/s41598-017-14468-9

Levshin, A., & Panza, G. (2006). Caveats in multi-modal inversion of seismic surface wavefields. *Pure and Applied Geophysics*, *163*(7), 1215–1233. https://doi.org/10.1007/s00024-006-0069-3

Lin, F.-C., Moschetti, M. P., & Ritzwoller, M. H. (2008). Surface wave tomography of the Western United States from ambient seismic noise: Rayleigh and love wave phase velocity maps. *Geophysical Journal International*, *173*(1), 281–298. https://doi.org/10.1111/j.1365-246x.2008.03720.x

Luo, Y., Yang, Y., Xu, Y., Xu, H., Zhao, K., & Wang, K. (2015). On the limitations of interstation distances in ambient noise tomography. *Geophysical Journal International*, *201*(2), 652–661. https://doi.org/10.1093/gji/ggv043

Mao, S., Mordret, A., Campillo, M., Fang, H., & van der Hilst, R. D. (2020). On the measurement of seismic traveltime changes in the time–frequency domain with wavelet cross-spectrum analysis. *Geophysical Journal International*, *221*(1), 550–568. https://doi.org/10.1093/gji/ggz495

Mordret, A., Jolly, A., Duputel, Z., & Fournier, N. (2010). Monitoring of phreatic eruptions using interferometry on retrieved cross-correlation function from ambient seismic noise: Results from Mt. Ruapehu, New Zealand. *Journal of Volcanology and Geothermal Research*, *191*(1–2), 46–59. https://doi.org/10.1016/j.jvolgeores.2010.01.010

Pedersen, H. A., & Krüger, F. (2007). Influence of the seismic noise characteristics on noise correlations in the Baltic shield. *Geophysical Journal International*, *168*(1), 197–210. https://doi.org/10.1111/j.1365-246x.2006.03177.x

Pilz, M., Parolai, S., Picozzi, M., & Bindi, D. (2012). Three-dimensional shear wave velocity imaging by ambient seismic noise tomography. *Geophysical Journal International*, *189*(1), 501–512. https://doi.org/10.1111/j.1365-246x.2011.05340.x

Poli, P., Pedersen, H., & Campillo, M. (2012). Emergence of body waves from cross-correlation of short period seismic noise. *Geophysical Journal International*, *188*(2), 549–558. https://doi.org/10.1111/j.1365-246x.2011.05271.x

Ratdomopurbo, A., & Poupinet, G. (1995). Monitoring a temporal change of seismic velocity in a volcano: Application to the 1992 eruption of Mt. Merapi (Indonesia). *Geophysical Research Letters*, *22*(7), 775–778. https://doi.org/10.1029/95gl00302

Rivet, D., Brenguier, F., & Cappa, F. (2015). Improved detection of preeruptive seismic velocity drops at the Piton de La Fournaise volcano. *Geophysical Research Letters*, *42*(15), 6332–6339. https://doi.org/10.1002/2015gl064835

Roux, P., Sabra, K. G., Gerstoft, P., Kuperman, W., & Fehler, M. C. (2005). P-waves from cross-correlation of seismic noise. *Geophysical Research Letters*, *32*(19). https://doi.org/10.1029/2005gl023803

Sabra, K. G., Gerstoft, P., Roux, P., Kuperman, W., & Fehler, M. C. (2005). Extracting time-domain Green's function estimates from ambient seismic noise. *Geophysical Research Letters*, *32*(3), L03310. https://doi.org/10.1029/2004gl021862

Savage, M. K., Lin, F.-C., & Townend, J. (2013). Ambient noise cross-correlation observations of fundamental and higher-mode Rayleigh wave propagation governed by basement resonance. *Geophysical Research Letters*, *40*(14), 3556–3561. https://doi.org/10.1002/grl.50678

Schimmel, M. (1999). Phase cross-correlations: Design, comparisons, and applications. *Bulletin of the Seismological Society of America*, *89*(5), 1366–1378. https://doi.org/10.1785/bssa0890051366

Schimmel, M., & Gallart, J. (2007). Frequency-dependent phase coherence for noise suppression in seismic array data. *Journal of Geophysical Research*, *112*(B4), 494–506. https://doi.org/10.1029/2006jb004680

Schimmel, M., & Paulssen, H. (1997). Noise reduction and detection of weak, coherent signals through phase-weighted stacks. *Geophysical Journal International*, *130*(2), 497–505. https://doi.org/10.1111/j.1365-246x.1997.tb05664.x

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. *Geophysical Journal International*, *184*(1), 494–506. https://doi.org/10.1111/j.1365-246x.2010.04861.x

Scott, T. M. (1992). *A geological overview of Florida*. Florida Geological Survey.

Shapiro, N. M., & Campillo, M. (2004). Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise. *Geophysical Research Letters*, *31*(7), 1–4. https://doi.org/10.1029/2004gl019491

Snieder, R. (2004). Extracting the green's function from the correlation of coda waves: A derivation based on stationary phase. *Physical Review*, *69*(4), 046610. https://doi.org/10.1103/physreve.69.046610

Stockwell, R. G., Mansinha, L., & Lowe, R. (1996). Localization of the complex spectrum: The s transform. *IEEE Transactions on Signal Processing*, *44*(4), 998–1001. https://doi.org/10.1109/78.492555

Toledo, T., Obermann, A., Verdel, A., Martins, J. E., Jousset, P., Mortensen, A. K., et al. (2022). Ambient seismic noise monitoring and imaging at the theistareykir geothermal field (Iceland). *Journal of Volcanology and Geothermal Research*, *429*, 107590. https://doi.org/10.1016/j.jvolgeores.2022.107590

Torrence, C., & Compo, G. P. (1998). A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, *79*(1), 61–78. https://doi.org/10.1175/1520-0477(1998)079<0061:apgtwa>2.0.co;2

Ventosa, S., Schimmel, M., & Stutzmann, E. (2017). Extracting surface waves, hum and normal modes: Time-scale phase-weighted stack and beyond. *Geophysical Journal International*, *211*(1), 30–44. https://doi.org/10.1093/gji/ggx284

Ventosa, S., Schimmel, M., & Stutzmann, E. (2019). Towards the processing of large data volumes with phase cross-correlation. *Seismological Research Letters*, *90*(4), 1663–1669. https://doi.org/10.1785/0220190022

Vernon, F. L., Pavlis, G. L., Owens, T. J., McNamara, D. E., & Anderson, P. N. (1998). Near-surface scattering effects observed with a high-frequency phased array at pinyon flats, California. *Bulletin of the Seismological Society of America*, *88*(6), 1548–1560. https://doi.org/10.1785/bssa0880061548

Wapenaar, K. (2004). Retrieving the elastodynamic Green's function of an arbitrary inhomogeneous medium by cross correlation. *Physical Review Letters*, *93*(25), 254301. https://doi.org/10.1103/physrevlett.93.254301

Williams, L. J., & Kuniansky, E. L. (2015). *Revised hydrogeologic framework of the Floridan aquifer system in Florida and parts of Georgia, Alabama, and South Carolina*. US Department of the Interior, US Geological Survey.

Woo, H. B. (2023). Short-Period ambient noise quality control and reproducibility [Software]. Zenodo. https://doi.org/10.5281/zenodo.7582572

Yang, Y., Ritzwoller, M. H., Levshin, A. L., & Shapiro, N. M. (2007). Ambient noise Rayleigh wave tomography across Europe. *Geophysical Journal International*, *168*(1), 259–274. https://doi.org/10.1111/j.1365-246x.2006.03203.x