Title

Denoising Surface Waves Extracted From Ambient Noise Recorded by 1-D Linear Array Using Three-Station Interferometry of Direct Waves

Journal

Journal of Geophysical Research: Solid Earth, 126(8)

ISSN

2169-9313

Authors

Qiu, Hongrui
Niu, Fenglin
Qin, Lei

Publication Date

2021-08-01

DOI

10.1029/2021jb021712

Peer reviewed

Denoising surface waves extracted from ambient noise recorded by 1-D linear array using three-station interferometry of direct waves

Hongrui Qiu[1], Fenglin Niu[1], and Lei Qin[2]

[1]Department of Earth, Environmental and Planetary Sciences, Rice University, Houston, TX, USA

[2]Hubei Subsurface Multi-Scale Imaging Key Laboratory, Institute of Geophysics and Geomatics, China University of Geosciences, Wuhan, China

Corresponding author: Hongrui Qiu (hq7@rice.edu; qiuhonrui@gmail.com)

Key points:

- Surface waves from ambient noise cross correlations are significantly enhanced at high frequencies using three-station interferometry
- Phase travel times are extracted reliably between 0.3-1.6 s for a 1.6-km-long linear array and are used to perform surface wave tomography
- Phase velocity models of Rayleigh and Love waves derived via eikonal tomography reveal high-resolution fault zone images

Abstract

We develop an automatic workflow for enhancing surface wave signals in ambient noise cross correlations (ANCs) calculated for a 1-D linear array. The proposed array-based method is applied to a 1.6-km-long dense linear nodal array crossing surface traces of the San Jacinto fault near Anza, California. Fundamental and higher modes of surface waves are observed in ANCs of the nodal array. After attenuating the surface wave overtones by applying a frequency-dependent tapering window to the ANCs, signals dominated by the fundamental mode surface wave are then enhanced through a denoising process based on three-station interferometry of direct waves. The signal to noise ratio is significantly increased at high frequencies (> 2 Hz) after denoising. Phase travel times are extracted reliably in the frequency domain for the period ranges of 0.3-1.2 s and 0.3-1.6 s for Rayleigh and Love waves, respectively. The corresponding period-dependent phase velocity profiles derived from the eikonal equation reveal high-resolution details of fault zone internal structures beneath the array. A broad (500-1000 m) low-velocity zone that narrows with increasing period is observed, illuminating a flower-shaped structure of the San Jacinto fault damage zone.

Plain Language Summary

Properties of fault damage zone (width of 100-1000’s meters), such as its geometry and velocity reduction compared to the surrounding host rock, can have a profound impact on our understandings of earthquake ruptures and the long-term behavior of the fault. Several dense nodal arrays with 10-100 m spacing and aperture of a few kilometers were deployed crossing surface traces of major faults, to provide high-resolution images of the fault zone internal structures. Surface waves travel at frequency-dependent speeds between every two sensors are observed in ambient noise cross correlations. We can infer structures at different depth using surface wave, as the velocity at higher frequency is more sensitive to shallower structures. However, surface waves extracted from ambient noise at high frequencies (> 1 Hz), that are essential to image fault zone in the top 100’s meters, are often very noisy. Here, we develop a denoising method that utilizes three-station interferometry to effectively suppress non-surface wave signals in a linear 1-D array. The quality of surface waves is significantly improved after the denoising, especially at high frequencies (> 2 Hz). Reliable measurements at high frequencies provide better constraints on fault zone internal structures at shallow depth.

1. Introduction

Noise-based surface wave tomography has been widely used to resolve crustal structures at various scales (e.g., Lin et al., 2009; Qiu et al., 2019; Wang et al., 2019; Zigone et al., 2019). In an effort to study high-resolution internal structures of major faults in southern California (SC), several dense arrays with station spacing less than 100 m were deployed crossing surface traces of the San Jacinto fault (e.g., Qin et al., 2021) and rupture zone of the 2019 Ridgecrest earthquake sequence (Catchings et al., 2020) for about one month. Analysis of high frequency (e.g., > 1 Hz) surface waves extracted from ambient noise cross correlations (ANCs) of these dense arrays can provide crucial information on the shallow (top 10s to 100s of meters) materials near faults with unprecedented spatial resolution (e.g., Wang et al., 2019) and thus improves our understanding of the local seismic hazard.

The quality of surface waves reconstructed from ANC depends on the duration of the continuous data and noise source distribution. Previous studies have shown that proper preprocessing steps (e.g., Bensen et al., 2007) can improve signal to noise ratio (SNR) of surface waves extracted from ANC. However, in contrast to high-quality signals at long periods (e.g., > 1 s), extraction of surface waves from ANCs calculated at high frequencies (e.g., > 1 Hz) for these linear arrays remains a challenging topic due to its low SNR (e.g., Wang et al., 2019), even if proper preprocessing procedures (e.g., Bensen et al., 2007) were implemented. This is likely due to the short recording time (e.g., one month) of these dense arrays and complicated pattern of noise sources at high frequencies and near faults (e.g., Hillers et al., 2013, 2014).

To achieve better reconstruction of surface waves from ambient seismic noise with inhomogeneous source distribution, Stehly et al. (2008) used correlations of the coda of the ANC (C3) calculated for a 2-D seismic network. This is achieved by stacking higher-order correlations for triplets of stations with two common virtual receivers. Froment et al. (2011) then investigated Green’s function reconstructed by correlations of the coda of C3 and concluded that the C3 method is helpful in suppressing source effects associated with non-isotropic source distribution. Sheng et al. (2018) later proposed an alternative prestack procedure that constructs C3 from the coda of each daily ANC first and then stack over the entire recording period. They showed that the prestack C3 reveals faster convergence and better recovery of arrivals at higher frequencies. The C3 method has also been applied to data recorded by asynchronous seismic networks to enhance the spatial resolution of noise-based surface wave tomography (e.g., Spica et al., 2016).

Different from C3, we use correlations of the entire waveform of the ANCs (later referred to as "three-station interferometry of direct waves") in this study to enhance signals of surface waves. Zhang et al. (2020) compared surface waves extracted from C3 and three-station interferometry of direct waves using data from the EarthScope USArray. They found that surface waves retrieved from three-station interferometry of direct waves show considerably higher SNR and broader bandwidth but yields small biases in dispersion measurements. Such bias arises from the geometry of the stations used to compute C3, and becomes zero or negligible when these stations align along a nearly straight line (i.e., 1-D linear array), even if noise sources are not evenly distributed (Lin et al. 2008). Therefore, we adopt the idea of using three-station interferometry of direct waves (hereinafter "three-station interferometry" for simplicity) to denoise the fundamental mode surface waves extracted from ANCs for 1-D linear arrays, which has not been done before.

We note that the configuration of 1-D linear array is commonly used in many dense deployments, such as linear arrays across major faults in SC (e.g., Catchings et al., 2020; Qin et al., 2021) and fiber optic cables (e.g., Cheng et al., 2021). The deployment periods of these 1-D linear arrays are in the order of a few tens of days to a few months, which are much shorter than those of 2-D broadband arrays that are used in ambient noise tomographic studies (e.g., Qiu et al., 2019). We apply this method to data recorded by a dense linear array deployed at the Ramona Reservation (RR) site across surface traces of the San Jacinto fault, near Anza (Fig. 1), California. Seismic waveforms from the RR array have been analyzed for fault zone internal structures in Qin et al. (2021). ANCs were computed for each station pair of the RR array and the corresponding Rayleigh wave phase velocities for periods from 0.3 s to 0.8 s were derived from double beamforming tomography in Wang et al. (2019).

Since signals of surface wave overtones were observed in ANCs of the RR array (Wang et al., 2019), in addition to the description of station configuration (Figs. 1a-b) and ANC data (Figs. 1c-d), we first analyze and attenuate signals of higher-mode surface waves for both components (Fig. 2) in section 2. Following the flow chart illustrated in Figure 3a, we present the theoretical formulation for denoising the fundamental mode surface waves using three-station interferometry with a 1-D linear array in section 3.1. The denoising process is demonstrated for an example station pair (Fig. 4) in section 3.2, and the comparison between surface wave signals before and after denoising is illustrated for the linear segment of the RR array (Fig. 5) in section 3.3. In section 4, following the flow chart in Figure 3b, surface wave phase travel times are first extracted from the denoised wavefield and then inverted for phase velocity dispersion models via the eikonal equation (Figs. 6-7). Discussion of the performance of the denoising method and comparison between the resulting phase velocity profiles and fault zone images from previous studies (Wang et al., 2019) are presented in section 5.

2. Data and Preprocessing

The RR array (red triangle in Fig. 1b) is located at north of Anza (blue square in Fig. 1b), California, and crosses surface traces of the Clark segment of the San Jacinto fault (Fig. 1a). The array consists of 94 three-component 5-Hz Fairfield geophones (balloons in Fig. 1a) that were set to record continuously for a month with a sampling rate of 500 Hz. ANC is obtained by first computing cross correlations of ambient noise data in 5-min windows, and then stacking them over the entire recording period for each station pair (Wang et al., 2019). The positive and negative time lags of the monthlong stacked ANC are fold and averaged to reduce the effects of the asymmetric noise source distribution. We use ANCs of a sub-array RR01-RR47 (yellow, blue, and red balloons in Fig. 1a) to demonstrate the surface wave denoising process (Fig. 3a) developed in this study. The sub-array has 47 stations with an average station spacing of ~30 m and an aperture of ~1.6 km.

We project stations in the sub-array to the straight line connecting RR01 and RR47 (cyan dashed line in Fig. 1a) and compare interstation distances calculated using station locations before and after the projection. The comparison yields negligible differences (<

1%) suggesting that the sub-array RR01-RR47 is in a 1-D linear configuration (later referred to as “the linear RR array”). In Wang et al. (2019), a period-dependent velocity threshold was applied to taper off the contamination of body waves or potential higher-mode surface waves. In this study, however, we first only apply a tapering window, with a maximum and minimum moveout velocities of 2 km/s and 0.1 km/s (black dashed lines in Figs. 1c-d), to the raw ANCs for both Transverse-Transverse (TT) and Vertical-Vertical (ZZ) components. The tapered ANCs are then filtered between 0.2 Hz and 10 Hz and depicted as colormaps in Figures 1c-d.

The aim of this study is to demonstrate that the proposed denoising procedure (Fig. 3a; Section 3.1) is effective in denoising surface waves, when only one mode (i.e., fundamental mode) of surface waves is present in the ANCs. Thus, we first analyze the surface wave overtones in the ANCs of the linear RR array by resolving the array-mean group and phase dispersion images (Fig. 2). Details of the analysis are described in supplementary materials (Text S1 and Fig. S1). Figures 2a-b show the array-mean group dispersion images for Love and Rayleigh waves, respectively. The blue dashed curves, connecting points with the largest values of the image obtained at different periods, denote the average group velocity dispersion of the fundamental mode signal. This suggests that the fundamental mode surface wave is the dominating signal in the ANCs filtered between 0.3 s and 1.6 s. This is in a good agreement with the observation of only one peak per period in the array-mean phase dispersion diagrams (MASW; Park et al., 1999) at long periods (> 0.6 s in Fig. 2c and > 0.7 s in Fig. 2d).

At short periods, however, higher-mode signals are clearly observed with much higher phase velocities in the array-mean phase dispersion diagrams (< 0.6 s in Fig. 2c and < 0.7 s in Fig. 2d). This is consistent with the weaker energy (above white dashed lines) observed at these short periods in the group dispersion images (Figs. 2a-b) that travels at group speeds much higher than that of the fundamental mode (blue dashed curves). The absence of higher mode signals at long periods is likely the result of low amplitude due to subsurface structures and excitation pattern of the ambient noise field. Another possibility is that the group velocity of higher modes at long periods is faster than 2 km/s. To suppress the surface wave overtones observed at short periods, we first highpass filter the ANCs (< 0.6 s for TT and < 0.7 s for ZZ; black dashed lines in Fig. 2) and apply a second tapering window to the filtered data that removes signals with group speeds above 0.55 km/s before performing surface wave denoising (Fig. 3a). The attenuation of higher-mode signals at short periods is effective, as the signatures of surface wave overtones that are clearly observed in Figures 2c-d are almost missing in the updated phase dispersion diagrams computed for the ANCs after the second tapering (Figs. 2e-f).

3. Surface Wave Denoising

Let $G_{i\_j}(t)$ be the positive lag of ANC for the station pair of *i*-th (virtual source) and *j*-th (virtual receiver) sensors in the linear RR array (yellow, green, and red triangles in Fig. 1a), we can expand it as

$$G_{i\_j}(t) = S_{i\_j}(t) + B_{i\_j}(t) + N_{i\_j}(t), \tag{1}$$

where $S_{i\_j}(t)$ and $B_{i\_j}(t)$ represent signals traveling on the surface (i.e., surface waves) and at depth (i.e., diving *P* or *S* body waves) between the source *i* and receiver *j*, respectively. $N_{i\_j}(t)$ is the residual (later referred to as "background noise"). This section aims to develop a denoising process that preserves $S_{i\_j}(t)$ while suppressing $B_{i\_j}(t)$ and $N_{i\_j}(t)$ in equation 1.

3.1 Three-station interferometry for a 1-D linear Array

Since surface waves are dispersive, let $\tilde{G}_{i\_j}(\omega)$ be the Fourier transform of $G_{i\_j}(t)$ at the angular frequency ω, we can rewrite equation 1 in the frequency domain

$$\tilde{G}_{i\_j}(\omega) = A_{G\_ij} \cdot e^{i \cdot \varphi_{G\_ij}} = \tilde{S}_{i\_j}(\omega) + \tilde{B}_{i\_j}(\omega) + \tilde{N}_{i\_j}(\omega)$$
$$= \sum_S A_{S\_ij} \cdot e^{-i\left(\omega \cdot T_{ij}^S + \varphi_s\right)} + \sum_B A_{B\_ij} \cdot e^{-i\left(\omega \cdot T_{ij}^B + \varphi_B\right)} + \tilde{N}_{i\_j}(\omega), \tag{2a}$$

where A_{S_ij} and T_{ij}^S are amplitude spectrum and phase travel time of surface wave signals in ANC at the angular frequency ω that propagate between the *i*-th and *j*-th stations, while A_{B_ij} and T_{ij}^B represent those of body waves that travel at depth. φ_s and φ_B are initial phases of surface- and body-wave signals in the ANC, respectively, and dependent on the distribution of ambient noise sources (e.g., Lin et al., 2008). Assuming the higher-mode surface waves are negligible or have already been removed from ANC (Section 2), we, therefore, can simplify equation 2a as:

$$\tilde{G}_{i\_j}(\omega) = \tilde{F}_{i\_j}(\omega) + \tilde{O}_{i\_j}(\omega) = A_{F\_ij} \cdot e^{-i\left(\omega \cdot T_{ij}^{F} + \varphi_F\right)} + \tilde{O}_{i\_j}(\omega), \qquad (2b)$$

where the symbol or subscript F stands for the fundamental mode surface wave. $\tilde{O}_{i\_j}(\omega) = \tilde{B}_{i\_j}(\omega) + \tilde{N}_{i\_j}(\omega)$, that consists of signals from body waves traveling at depth and background noise, is the term we want to suppress in the denoising process. It is interesting to note that $\varphi_F = \pi/4$ for an azimuthally homogenous ambient noise source distribution (Snieder, 2004), whereas $\varphi_F = 0$ when noise sources are only present in line with the station pair i and j (Lin et al., 2008).

For surface waves of a certain (e.g., fundamental) mode traveling between three stations $i < j < k$ in a 1-D linear array, the travel times satisfy the following relation

$$T_{ik}^{S} = T_{ij}^{S} + T_{jk}^{S}, \qquad (3a)$$

whereas

$$T_{ik}^{B} < T_{ij}^{B} + T_{jk}^{B}, \qquad (3b)$$

for body waves traveling at depth. Therefore, we introduce a third station k and perform three-station interferometry following Zhang et al. (2020):

$$\tilde{I}_{i\_j}(\omega; k) = \begin{cases} \tilde{G}_{i\_k}^{*}(\omega) \cdot \tilde{G}_{j\_k}(\omega), & k < i \\ \tilde{G}_{i\_k}(\omega) \cdot \tilde{G}_{j\_k}(\omega), & i < k < j. \\ \tilde{G}_{i\_k}(\omega) \cdot \tilde{G}_{j\_k}^{*}(\omega), & k > j \end{cases} \qquad (4a)$$

In equation 4a, we cross correlate $G_{i\_k}(t)$ and $G_{j\_k}(t)$ in the time domain, when $k < i$ or $k > j$ (later referred to as "outer-source zone"). The interferometry becomes equivalent to the convolution of $G_{i\_k}(t)$ and $G_{j\_k}(t)$ in the time domain for station k located within the two virtual sources (i.e., $i < k < j$; later referred to as "inter-source zone"). For the case $k = i$ or j, we define $\tilde{I}_{i\_j}(\omega; k) = A_{G\_ij}^{2} \cdot e^{i \cdot \varphi_{G\_ij}}$ that approximates the convolution of $G_{i\_j}(t)$ and $G_{i\_i}(t)$ or $G_{j\_j}(t)$, by assuming the amplitude spectrum of the auto-correlation $G_{i\_i}(t)$ or $G_{j\_j}(t)$ is similar to that of $G_{i\_j}(t)$, i.e., $A_{G\_ii} \approx A_{G\_jj} \approx A_{G\_ij}$.

Combining equations 2b, 3a, and 4a, if the fundamental mode surface wave is the dominant signal in ANC (i.e., $\tilde{O}_{i\_j}$ in Equation 2b is negligible), the phase term of the interferogram $\tilde{I}_{i\_j}(\omega; k) = A_{ij\_k} \cdot e^{i \cdot \varphi_{ij\_k}}$ is given by

$$\varphi_{ij\_k}(\omega) = \begin{cases} -\omega \cdot T_{ij}^{F} - 2\varphi_F, & i < k < j \\ -\omega \cdot T_{ij}^{F} - \varphi_F, & k = i \text{ or } j. \\ -\omega \cdot T_{ij}^{F}, & k < i \text{ or } k > j \end{cases} \qquad (4b)$$

T_{ij}^{F} and φ_F denote the phase travel time and initial phase of the fundamental mode surface wave signal (Equation 2b) extracted from the ANC of station pair *i* and *j*. Considering the amplitude spectrum, A_{G_ij}, of the original ANC usually peaks at certain frequencies (e.g., microseism frequency band), the three-station interferometry also acts as a bandpass filter that amplifies signals around those spectral peaks in A_{G_ij}, as the amplitude term of the interferogram $\tilde{I}_{i\_j}(\omega;k)$ is approximately given by $A_{G\_ij}^2$. Therefore, we take the square root of the amplitude term A_{ij_k} while preserving the phase term φ_{ij_k} of the original interferogram $\tilde{I}_{i\_j}(\omega;k)$ to suppress the effect of source spectra multiplication introduced in the three-station interferometry (Equation 4a), i.e., $\tilde{I}_{i\_j}^{c}(\omega;k) = \sqrt{A_{ij\_k}} \cdot e^{i \cdot \varphi_{ij\_k}}$. This is based on the assumption that amplitude spectra of the input ANCs are similar for all station pairs, i.e., $A_{ij\_k} = A_{G\_ik} \cdot A_{G\_jk} \approx \hat{A}_G^2$.

Equation 4b suggests that the interferograms within either the inter- or outer-source zones share the same phase, whereas interferograms from different zones are only aligned in phase when φ_F is zero. In cases when the term $\tilde{O}_{i\_j}$ is significant, we can divide the interferogram $\tilde{I}_{i\_j}^{c}(\omega;k)$ into two components: $\tilde{I}_{i\_j}^{cF}(\omega;k)$ and $\tilde{I}_{i\_j}^{cO}(\omega;k)$. $\tilde{I}_{i\_j}^{cF}(\omega;k)$ represents the interferogram that only involves the fundamental mode surface wave signal, i.e., when $\tilde{O}_{i\_j}$ is set to zero in equation 2b. The phase of $\tilde{I}_{i\_j}^{cF}(\omega;k)$, given by equation 4b, is independent of k when φ_F is zero. Therefore, we can simply stack the interferograms $\tilde{I}_{i\_j}^{c}(\omega;k)$ over all available station k to enhance the fundamental mode surface wave and the denoised waveform is given by:

$$\tilde{C}_{i\_j}^{3}(\omega) = \sum_{k=1}^{N} \tilde{I}_{i\_j}^{c}(\omega;k) \Big/ N \approx \sum_{k=1}^{N} e^{i \cdot \varphi_{ij\_k}} \sqrt{A_{ij\_k}} \Big/ N. \quad (5)$$

Here, *N* is the number of stations in the 1-D linear array, and we assume $\varphi_F = 0$. We note that the other component, $\tilde{I}_{i\_j}^{cO}(\omega;k)$, is suppressed through the stacking. This is because the phase term of $\tilde{I}_{i\_j}^{cO}(\omega;k)$ varies significantly with *k*, as it involves contributions from diving body waves and background noise that do not satisfy equation 3a and thus equation 4b. If $\varphi_F \neq 0$, we need to correct the interferogram $\tilde{I}_{i\_j}^{c}(\omega;k)$ for station k within the inter-source zone following equation 4b, i.e., $\tilde{I}_{i\_j}^{c}(\omega;k) \cdot e^{i \cdot 2\varphi_F}$, before stacking.

3.2 Surface wave denoising of the station pair RR10 and RR40

As described in section 2, we observe higher-mode Love and Rayleigh waves only at short periods that travel at group speeds higher than 0.55 km/s (Fig. 2) in ANCs of the linear RR array (Figs. 1c-d). Thus, the denoising process (Fig. 3a) is performed directly on the lowpass filtered ANCs for each component, as the surface wave overtones are negligible at low frequencies (> 0.6 s at TT and > 0.7 s at ZZ; Fig. 2). For ANCs at high frequencies, a second tapering window that effectively attenuates higher-mode signals is applied to the highpass filtered data before denoising. All results in later discussions are thus associated with the fundamental mode surface wave.

Figure 4 shows results of the three-station interferometry applied to ANCs of TT component for an example station pair RR10 (i = 10) and RR40 (j = 40). As surface waves are dispersive, we compare the input ANCs and the interferograms, $\tilde{I}^{c}_{i\_j}(\omega;k)$, after narrow bandpass filtering at two example periods, 0.8 s (Figs. 4a-c) and 0.3 s (Figs. 4d-f), to better demonstrate the performance of the denoising process in the low and high frequency bands, respectively. The bandpass filter is generated following section 3.1 of Qiu et al. (2019). Figures 4a and 4b show the ANCs filtered at 0.8 s with RR10 and RR40 as the virtual source, $G_{i\_k}(t)$ and $G_{j\_k}(t)$, respectively. Black and blue waveforms denote the filtered ANCs with the virtual receiver station k inside the outer-source (k < 10 or k > 40) and inter-source (10 < k < 40) zones (i.e., k is the y-axis of Fig. 4), respectively. The filtered ANC of the example station pair RR10 and RR40 is depicted in red.

Figure 4c demonstrates in colors the resulting interferograms $\tilde{I}^{c}_{i\_j}(\omega;k)$ transformed to the time domain through inverse Fourier transform. The interferograms show coherent surface waves filtered at 0.8 s that are well aligned in phase for all k values. This is consistent with our derivations in section 3.1 and the observation of high-quality surface wave signals in the filtered ANCs (Figs. 4a-b). The fact that the interferograms $\tilde{I}^{c}_{i\_j}(\omega;k)$ are well aligned between the inter- (blue vertical arrow) and outer-source (black vertical arrow) zones suggests the initial phase $\varphi_F \approx 0$ (Equation 4b) for ANCs filtered at 0.8 s. Therefore, we obtain the denoised waveform for the example station pair RR10 and RR40 filtered at 0.8 s (black waveform; Equation 5a) by stacking all the interferograms in Figure 4c. The surface wave signals (between the dashed lines) are almost identical between waveforms before (in gray) and after (in black) denoising, with the denoised waveform showing smaller coda waves.

Coherent surface waves are also observed in Figure 4f for interferograms narrow bandpass filtered at 0.3 s. Again, signals of surface wave in the interferograms filtered at high frequency are also well aligned between the inter- and outer-source zones, indicating the assumption of $\varphi_F \approx 0$ is valid. We also verified that the initial phase $\varphi_F \approx 0$ for all periods (from 0.3 s to 1.6 s) and components (ZZ and TT) analyzed in this study, by estimating the systematic phase difference between interferograms in the inter- and outer-source zones computed for the example station pair RR10 and RR40 (not shown here). The SNR of surface waves in interferograms at high frequency is much lower than those at low frequency, as incoherent arrivals with large amplitudes that vary significantly with the choice of station k are seen (Fig. 4f). This is because of the lower SNR for surface waves (i.e., $\tilde{O}_{i\_j}$ in Equation 2b is non-negligible) in the input ANCs filtered at high frequency (Figs. 4d-e). The SNR of surface wave is increased after denoising (in black), as the large coda waves present in the filtered ANC (in gray) are significantly suppressed.

Similar observations are obtained for Rayleigh waves extracted from ANCs at ZZ component (Fig. S2). Although the surface wave signals are generally aligned well between different interferograms filtered at both the low and high frequencies, small fluctuations in the travel times are still observed in Figures 4c and 4f. This is likely due to variations in SNR, i.e., amplitude ratio between surface and coda waves, of the interferograms. In addition, although a narrow bandpass filter is applied, the peak frequency of the filtered interferograms can still deviate from the center frequency of the filter. Thus, variations in peak frequency can also lead to visible changes in surface wave travel times, particularly for long interstation distances, as surface waves are dispersive (e.g., blue dashed curves in Fig. 2). We note that, based on our derivations in section 3.1, the denoising method is applicable to any arrivals if (1) the wave propagation satisfies equation 3a (e.g., teleseismic arrivals, body waves traveling along the surface or refracted from a subsurface impedance contrast) and (2) only one such arrival is present in the input wavefield $G_{i\_j}(t)$.

3.3 Surface wave denoising of the linear RR Array

In section 3.2, we demonstrate the workflow and effectiveness of the three-station-interferometry-based denoising process for an example station pair RR10 and RR40. As illustrated in the flow chart (Fig. 3a), we can further enhance surface waves extracted from ANCs of the entire linear array by performing the denoising process for multiples times: first self-normalize the output wavefield of the current iteration, and then use the normalized wavefield as the input for the next iteration. The number of iterations is determined so that the difference between input and output wavefields of the last iteration is negligible. As surface waves are dispersive, we use symbol $C_{i\_j}^{2+n}(t;T_c)$ to represent the waveform of station pair i and j, after first applying n (≥ 1) iterations of the denoising process and then narrow bandpass filtering at period T_c for better illustration of the denoised results.

Figure 5a shows the comparison between the TT component ANC (black) of the example station pair RR10 and RR40 filtered at 0.8 s and the corresponding denoised waveforms (in red) of the first $C_{i\_j}^{3}(t;0.8)$ and second $C_{i\_j}^{4}(t;0.8)$ iteration. Since the waveforms C^3 and C^4 are almost identical, this suggests that only two iterations are needed for the denoising results to converge at low frequencies (> 0.6 s for TT and > 0.7 s for ZZ; Fig. 2). On the other hand, Figure 5e suggests that four iterations (red waveforms) are needed to ensure a convergence of the denoising process at high frequencies, as the SNR of surface wave is much lower in the filtered ANC (black waveform) after the attenuation of higher-mode signals (Section 2). Although the surface wave SNR gradually increases with the number of denoising iterations (from bottom to top; Figs. 5a and 5e), the surface wave signals are always coherent and aligned in phase.

The ANC data $G_{i\_j}(t)$ of TT component filtered at 0.8 s and the corresponding denoised waveforms $C_{i\_j}^{4}(t;0.8)$ are illustrated in Figures 5b-c, respectively, for all station pairs. Although waveforms before and after denoising filtered at low frequency are similar, the difference is still noticeable (Fig. 5c). For instance, the background fluctuations with irregular arrival patterns in the coda waves are reduced; the arrival prior to surface wave with a phase velocity of ~2 km/s at long interstation distances (> 1.2 km), which are likely related to the tapering window (with an upper limit velocity of 2 km/s; Section 2), are suppressed after the denoising. Figure 5d compares the array-mean amplitude spectra averaged over all station pairs for data before (in black) and after (in red) the denoising. The similarity between the two average amplitude spectra is consistent with the high SNR of surface waves in the raw ANCs filtered at low frequency.

The difference between wavefields before and after denoising is much larger at high frequency (0.3 s; Figs. 5f-g). Although coherent surface wave signals are seen propagating at a group speed slightly slower than 0.5 km/s in the filtered ANC data (Fig. 5f), wavelets with large amplitudes are observed in coda waves (e.g., black waveform in Fig. 5f). These wavelets sometimes have arrival times similar to those of surface waves and thus can interfere with and bias the surface wave dispersion measurements. After four iterations of denoising, the background noise is greatly suppressed in the wavefield C^6 (Fig. 5g). Figure 5h shows the array-mean amplitude spectra, with the one averaged over data after denoising being smoother (in red). This is likely due to the interference between the surface and coda waves that contributes to the complicated array-mean amplitude spectrum calculated for the data before denoising (in black).

In addition to comparisons between the input ANCs and denoised wavefield at low (0.8 s) and high (0.3 s) frequencies in Figure 5, we also compute the array-mean group and phase dispersion images for the denoised wavefield at TT (Figs. S4a-b) and ZZ (Figs. S5a-b) components following the procedures described in Text S1. The array-mean group and phase velocity dispersion curves (white markers in Figs. S4-S5) are determined as the period-dependent velocity that yields the largest amplitude of the curve extracted from the image at each corresponding period. The dispersion relations obtained for data before (blue dashed curve) and after (white markers) denoising are compared in Figure 2. Differences between results before and after denoising are generally smaller than 5% for array-mean group velocities but much larger (~10%) for phase velocities. This is because array-mean phase velocities obtained from MASW (e.g., Figs. 2c-f) are inferred in the frequency domain (Text S1) and thus very sensitive to background noise.

4. Surface Wave Tomography

In this section, we use denoised waveforms of TT and ZZ components (e.g., Figs. 5c, 5g, S3c, and S3g) to infer phase velocity structures of Love and Rayleigh waves beneath the array, respectively. Following the flow chart shown in Fig. 3b, we first determine cycle-skipped phase travel times of surface waves propagating between all available station pairs at each period (e.g., Fig. 6a) in the frequency domain, which is much simpler than measuring in the time domain but requires high SNR (Section 4.1). Second, we infer phase velocity structures beneath the linear RR array, using travel time measurements after cycle-skipping correction from section 4.1, via the eikonal equation in section 4.2 (e.g., Fig. 6b). The aim of this section is to demonstrate that robust surface wave phase velocity models can be resolved from the denoised waveforms.

4.1 Determination of phase travel time

Frequency time analysis (FTAN) is widely used in previous studies to determine phase travel time of surface wave signal in ANC (e.g., Bensen et al., 2007; Lin et al., 2008; Qiu et al., 2019). First, Gaussian narrow bandpass filters centered on a series of consecutive frequencies are applied to the ANC, then the phase travel time dispersion is measured using the envelope and phase functions of the filtered ANC in the time domain. The advantage of FTAN is that reliable phase travel times can still be extracted when SNR is low at high frequencies. However, ad hoc criteria and thresholds are required to automate the FTAN. Additional details on the FTAN method can be found in section 3 of Qiu et al. (2019). Since our goal is to verify that the signals after denoising are representative of surface waves and high SNR is achieved for all frequencies, we thus measure phase travel times from the denoised waveforms in the frequency domain, which is much simpler than the FTAN method and described in detail below.

Although surface wave SNR is high in the denoised waveform, we still observe waves with very small amplitudes before and after the surface wave (e.g., black waveforms in Figs. 5c and 5g). This is because we can only attenuate rather than remove completely signals that are not surface waves. Here, we apply a frequency-dependent tapering window (e.g., black dashed lines in Figs. S4d-S5d) centered on the surface wave to further remove these background fluctuations. Width of the tapering window is set to six times the dominant period of the array-mean amplitude spectrum (e.g., red curve in Fig. 5d), whereas the center is determined by the array-mean phase and group velocities of the array at the target period (e.g., white markers in Fig. 2). We note that this tapering process is to ensure the accuracy of phase travel times measured in the frequency domain, which is unnecessary if FTAN is implemented.

Since $\varphi_F \approx 0$ (e.g., Fig. 4; Section 3.2) and the term O_{i_j} in equation 2b is negligible after denoising (e.g., Fig. 5) and tapering (e.g., Fig. S4d-S5d), we have

$$\tilde{C}_{i\_j}^{DT}(\omega; f_c) = A_{F\_ij}(\omega; f_c) e^{-i\cdot\omega T_{ij}^F(\omega)}, \tag{6}$$

where $\tilde{C}_{i\_j}^{DT}(\omega; f_c)$ is the spectrum of the denoised and tapered waveform filtered at the center frequency f_c for station pair i and j (e.g., blue waveform in Fig. 6a). Equation 6 suggests that we can extract cycle-skipped phase travel time from the phase spectrum of the tapered waveform. It is important to note that the peak frequency $f_{\max}$ of the tapered waveform, i.e., the peak of A_{F_ij}, may deviate from f_c, the center frequency of the filter. Therefore, we measure the wrapped phase (i.e., between -2π and 0) of the spectrum $\tilde{C}_{i\_j}^{DT}(\omega; f_c)$ at $\bar{f}_{\max}$, the peak frequency of the amplitude spectrum averaged over all station pairs, where the array-mean SNR of the surface wave is the highest. Then, the cycle-skipped phase travel time is computed as the wrapped phase divided by $-2\pi\bar{f}_{\max}$.

Figure 6a shows the cycle-skipped phase travel times (white circles) measured at $\bar{f}_{\max} = 3.14$ Hz for surface waves filtered and tapered at 0.3 s from station pairs associated with a common virtual source RR10 (y-axis of 0 km). To obtain the actual phase travel time, we perform a simple cycle-skipping correction as follows:

(1) As illustrated in Figure 6a, we first extract all the cycle-skipped phase travel times for surface waves of a virtual shot gather and arrange them as a function of the location to the virtual source.

(2) We perform cycle-skipping correction for surface waves traveling NE (toward RR47) and SW (toward RR01) separately.

(3) For surface waves traveling in the same direction, the principle of the cycle-skipping correction is to ensure that the travel time of any virtual receiver is larger than those of receivers that are closer to the virtual source after the correction.

(4) In practice, we examine measurements T_i and T_{i+1} of every two adjacent virtual receivers with the i-th station being closer to the virtual source. If $T_i \geq T_{i+1}$, we use T_i as the reference and add $N/\bar{f}_{\max}$ (N is an integer) to T_{i+1} so that $T_{i+1} + N/\bar{f}_{\max} > T_i \geq T_{i+1} + (N-1)/\bar{f}_{\max}$. The correction is performed for closer-to-source pairs first.

Travel times, for the virtual shot gather of RR10, after the correction are illustrated as red stars in Figure 6a. We note that a more sophisticated cycle-skipping correction (e.g., using phase velocity structure inferred at a longer period as the reference) is needed when station spacing is larger than one wavelength.

4.2 1-D eikonal tomography

We use the eikonal equation to derive phase velocity structures using travel time measurements of all station pairs in the linear RR array (Section 4.1). First, we project all stations to the straight line connecting RR01 and RR47 (cyan dashed line in Fig. 1a). Second, travel time measurements associated with each virtual source *i* at the target frequency $\bar{f}_{\max}$ are extracted and interpolated (e.g., black curve in Fig. 6a) with a regular grid size of Δ=50 m. Since variations in topography (Fig. 2b of Qin et al., 2021) have a negligible effect (< 0.5%) on the results, the eikonal tomography can be simplified as:

$$\tilde{v}_i(x;\bar{f}_{\max}) = 2\cdot\Delta/\left[T_i(x+\Delta;\bar{f}_{\max}) - T_i(x-\Delta;\bar{f}_{\max})\right], \qquad (7)$$

where $\tilde{v}_i(x;\bar{f}_{\max})$ and $T_i(x;\bar{f}_{\max})$ are the local phase velocity and interpolated phase travel time, respectively, of the grid cell at location *x*. Since the local phase velocity $\tilde{v}_i$ only varies with the grid cell location, it is independent of virtual source *i*. Thus, we can average the 1-D phase velocity profiles resolved from all available virtual sources at the same frequency $f_{\max}$ to achieve a more reliable phase velocity model:

$$\bar{v}(x;f_{\max}) = \sum_{i=1}^{N_x} \tilde{v}_i(x;\bar{f}_{\max}) \Big/ N_x, \qquad (8a)$$

and estimate the corresponding uncertainty as the standard deviation:

$$\delta(x;\bar{f}_{\max}) = \sqrt{\sum_{i=1}^{N_x}\left[\tilde{v}_i(x;\bar{f}_{\max}) - \bar{v}(x;\bar{f}_{\max})\right]^2 \Big/ N_x}, \qquad (8b)$$

where N_x is the number of virtual sources available for stacking at location *x*.

In surface wave studies, phase velocities derived at near-virtual-source grid cells are often excluded to satisfy the far-field approximation (e.g., Bensen et al., 2007). The size of the exclusion zone is usually multiples of the analyzed wavelength (e.g., one wavelength in Wang et al., 2019). Here, however, we set an exclusion zone with a fixed size of 100 m, i.e., discard phase velocities derived at the four grid cells closest to the virtual source, to avoid any potential bias in travel time gradient estimation near the virtual source. Figure 6b shows the 1-D phase velocity profile, in white dots, derived using measurements associated with the virtual source RR10 (black curve in Fig. 6a) for Love waves at $\bar{f}_{\max} = 3.14$ Hz, whereas phase velocity profiles resolved from all virtual sources are illustrated in gray curves and as the colormap. The average phase velocity and uncertainty profiles are calculated via equation 8 and demonstrated as red stars and error bars, respectively, in Figure 6b.

Figure 7 shows phase velocity models resolved at periods ranging from 0.3 s to 1.6 s for Love waves (Fig. 7a) and 0.3 s to 1.2 s for Rayleigh waves (Fig. 7b), together with the corresponding uncertainty estimations (Figs. 7c-d). The period range in the plot is determined so that the resolved maximum uncertainty is smaller than 0.1 km/s. In general, the uncertainties are smaller than 0.03 km/s for both Rayleigh and Love waves at all analyzed periods, indicating the resolved phase velocity structures are robust and reliable. The small uncertainty at low frequencies also justifies our choice of an exclusion zone with a 100-m-radius. This is because one wavelength at low frequency (e.g., ~800 m for Rayleigh wave at ~0.8 s; white markers in Fig. 2f) is much larger than 100 m. If the one wavelength exclusion zone is necessary, phase velocities calculated at a target grid cell should vary significantly for virtual sources within and outside the exclusion zone, and thus yield large uncertainty values.

Phase velocity models of both Love and Rayleigh waves show a ~500- to 1000-m-wide low-velocity zone at low frequencies (e.g., > 0.8 s) that gradually narrows with the period. Combined with the fact that phase velocity at lower frequency is more sensitive to structures at greater depth, this observation likely indicates a flower-shaped (i.e., width decreases with depth) fault damage zone beneath the linear RR array. We also see several ~100-m-wide narrow zones, that are close to the mapped fault surface traces (black dashed lines in Figs. 7a-b), with extremely low phase velocities (< 500 m/s) at high frequencies (e.g., 0.3-0.6 s). However, the shape and location of these low-velocity zones are different between Figures 7a and 7b. Structure patterns that are inconsistent between models of Love and Rayleigh waves may indicate the existence of radial anisotropy or complicated structures of Vp/Vs ratio.

5. Discussion

We present a denoising method based on three-station interferometry that effectively enhances surface waves extracted from ANCs of a 1-D linear array. This array-based denoising method complements the existing tools that utilize denoising filters (e.g., Baig et al., 2009; Moreau et al., 2017) or phase weighted stacking (e.g., Schimmel et al., 2011; Ventosa et al., 2017) to improve the quality of ANC calculated for a single station pair. There are three assumptions in the theoretical derivations of this method (Section 3.1):
(1) The wave propagation satisfies equation 3a, i.e., the array configuration is 1-D and linear.
(2) The phase of the target signal is given by $\omega \cdot T_{ij}^{F} + \varphi_F$ (Equation 2b) for source i and receiver j, i.e., the bias φ_F irrelevant to wave propagation is a constant.
(3) Only one such arrival is present in the input wavefield.
While the derivations are applicable to other signals (e.g., teleseismic and refracted body waves) and datasets (e.g., earthquake data), we focus on the demonstration of enhancing the fundamental mode surface waves extracted from ANCs of the linear RR array in this paper. Performance of the proposed denoising process is dependent on how well these three assumptions are satisfied using the field data.

To verify the first assumption, we estimate the error between the linear RR array and an ideal 1-D linear configuration. The station-configuration error is approximately given by the difference (in percentage) between interstation distances calculated using station locations before and after projecting the array to a straight line (green dashed line in Fig. 1a). For a rough estimation, the mean and maximum of the station-configuration error for the linear RR array are ~0.1% and 1%, respectively. The uncertainties estimated from eikonal tomography (Figs. 7c-d) suggest ~1% and ~3% for the mean and maximum perturbations in the resolved phase velocities, which is larger than the estimated error in array geometry. Therefore, we conclude that this denoising method is robust when the station-configuration error is less than the allowable uncertainty of the resulting phase velocity model (e.g., mean and maximum of 1% and 3% in this study).

Regarding the second assumption, as shown in Lin et al. (2008), the phase term φ_F is related to the effect of noise source distribution. In section 3.2, we analyze the term φ_F systematically for triplets of stations with two common virtual sources, RR10 and RR40, at various periods between 0.3 and 1.6 s (e.g., Figs. 4c and 4f). The observation that the interferograms calculated for different triplets of stations are well aligned suggests that not only the second assumption is valid but also $\varphi_F \approx 0$. It is interesting to note that $\varphi_F \approx 0$ indicates the noise sources recorded by the linear RR array are not isotopically distributed, as otherwise φ_F would be $\pi/4$ (Snieder, 2004). Since φ_F can be calculated for any given noise source distribution, measurements of φ_F using ANCs of sub-arrays aligned in a straight-line taken from a 2-D array along different angles may provide a new way of resolving the noise source distribution.

Although clear higher-modes of surface waves are observed in ANCs of the linear RR array at high frequencies (Figs. 2a-d), we find that these higher-mode signals can easily be separated from the fundamental mode surface waves and effectively attenuated through a second tapering process (Section 2) to satisfy the third assumption. However, if two modes (*F* and *M*) of surface waves are present in ANCs, following equation S3c derived in Text S2, we can still apply three-station-interferometry-based denoising to the ANCs by stacking the interferogram $\tilde{I}_{i\_j}(\omega;k)$ defined in Equation 4a:

$$\tilde{C}^3_{i\_j}(\omega) = \sum_{k=1}^{N} \tilde{I}_{i\_j}(\omega;k) \Big/ N \approx A_F^2 \cdot \left[e^{-i\omega\cdot T_{ij}^F} + R^2 \cdot e^{-i\omega\cdot T_{ij}^M}\right], \quad (9)$$

where T_{ij}^F and T_{ij}^M are the phase travel times of the mode *F* and *M*, respectively. A_F denotes the array-mean amplitude spectrum of the mode *F*, whereas the constant *R* indicates the amplitude ratio between the two modes. As demonstrated by derivations in Text S2 and the synthetic test in Figures S6-S7, waveforms after denoising via equation 9 still preserves accurate phase travel times of both modes, as cross terms between the two modes are suppressed through stacking. We use equation 9 in the denoising process when two modes are present in the wavefield, as the application of equation 5 to such data leads to pseudo arrivals (Nakata, 2020; Figs. S8-S9).

However, unlike equation 5, the denoising process defined by equation 9 also acts as a bandpass filter (A_F) around the peak frequency, $f_{\max}$, of the input wavefield and thus inherently attenuates signals at frequencies away from $f_{\max}$ (e.g., Fig. S7d). It is interesting to note that the amplitude ratio between the two modes changes to R^2 in the denoised waveform $\tilde{C}^3_{i\_j}$ in equation 9, i.e., the weaker mode in the input wavefield (e.g., the mode *M* if $R < 1$) is attenuated (by a factor of *R*) after the denoising. Therefore, even if higher-mode signals are present in the field data, the third assumption is still valid if the ratio R between the fundamental mode and any higher modes in the input wavefield satisfies $R^2 << 1$ (e.g., $R < 0.5$; Figs. S6-S7). Although a constant ratio R is assumed in the derivations of Text S2 and R may vary with many factors (e.g., distance), a good approximation of such constant ratio would be the array-mean of R values measured from all station pairs.

To further evaluate the surface wave signals in the denoised waveforms, we derive phase velocity dispersion models for Love (0.3-1.6 s; Fig. 7a) and Rayleigh (0.3-1.2 s; Fig. 7b) waves extracted from the denoised wavefield at TT and ZZ component in section 4, respectively. Uncertainties of the dispersion model is calculated as the standard deviation of phase velocities derived for the same grid cell from different virtual sources (Equation 8b). The small median uncertainties of ~20 m/s for both Love (Fig. 7c) and Rayleigh (Fig. 7d) waves are consistent with our derivations in section 3.1, as the denoising process aims at enhancing arrivals that have a propagation time between two receivers independent of the source location (Equation 3a). We note that, although the uncertainties are much larger (> 50 m/s) at longer periods (> 1.3 s in Fig. 7c and > 1.1 s in Fig. 7d), the errors in percentage are still small (~2-3%).

We compare Rayleigh wave phase velocity models derived from this study and Wang et al. (2019) in the overlapping period (0.3-0.8 s) and spatial (RR01-RR47) ranges (Figs. 8a-b). In their study, the double beamforming technique is applied to the ANCs. The local phase velocity is first obtained through grid search for each sub-array (three nearby stations) pair: first sum all nine ANCs of the two sub-arrays through slant-stacking using different slowness values, then determine the local phase velocity of each sub-array based on the maximum amplitude of the envelope function for the stacked waveform. The final phase velocity is given by the average value of phase velocities derived for the same receiver but different source sub-arrays. They did not derive phase velocities for Rayleigh wave at low frequencies (> 0.8 s) as the size of their exclusion zone (one wavelength) is comparable to the array aperture (~1.6 km; Fig. 1a).

Wang et al. (2019) used the standard deviation of the median as the uncertainty. This is because their phase velocities obtained from different sources are statistically independent, whereas such redundant information has already been implemented in our denoising process (Equation 5). We find extremely large uncertainty values (> 0.15 km/s) in regions with high phase velocities (left bottom corner) and near fault surface traces (i.e., F1-F3) from their results (Fig. 8d). Since surface waves are dispersive, these large values may be the result of the violation of the assumption required by the double beamforming method that peak frequencies of the applied bandpass filter and waveform after slant stacking are the same. For instance, such deviation between peak frequencies of the applied filter (3.33 Hz) and stacked amplitude spectrum (3.16 Hz) is clearly shown in Figure 5g. Except for regions with errors > 0.15 km/s, the uncertainties of Wang et al. (2019) (~30-40 m/s in Fig. 8d) are comparable with (slightly larger than) those of this study (~20 m/s in Fig. 8c).

Similarly, consistent velocity values and structural patterns are seen in both models, such as an ultra-low velocity (< 0.4 km/s) zone on the NE side of the middle fault surface trace (F2) and faster velocities (> 1.2 km/s) at the bottom left of Figures 8a-b, suggesting both phase velocity models are generally reliable. On the other hand, small-scale differences between the two models are clearly observed. For instance, we find several high-velocity anomalies with narrow width (50-100 m) near fault surface traces F2 and F3, outlined by blue dashed lines in Fig. 8b, in the model of Wang et al. (2019), which are missing from our model (Fig. 8a). Considering uncertainties associated with these anomalies resolved from Wang et al. (2019) are extremely large (> 0.15 km/s; blue dashed lines in Fig. 8d), we conclude that these features are likely artifacts, and the phase velocity model derived in this study (Fig. 8a) from the denoised wavefield is a better representation of fault zone structures beneath the linear RR array.

Different from Wang et al. (2019), we do not construct a shear wave velocity model for structures beneath the linear RR array using the resolved phase velocity dispersion profiles (Fig. 7). The reasons are twofold: (1) the phase velocity model derived from eikonal tomography assumes that wave amplitudes are varying smoothly in space (e.g., Lin et al., 2009; Lin & Ritzwoller 2011) and (2) the piecewise 1-D Vs inversion scheme adopted in Wang et al. (2019) is only accurate when structures are varying smoothly laterally. Considering the array is deployed crossing fault damage zones that are laterally heterogenous (Figs. 1a and 7) and can significantly amplify seismic motions (Qin et al., 2021), we propose performing full waveform tomography (e.g., Zhang et al., 2018) on the denoised surface waves to infer accurate fault zone structures beneath the array as the subject of a future study.

In the present study, we only show phase velocities resolved at frequencies up to ~3 Hz (Fig. 7) from the denoised ANCs by requiring a minimum wavelength of 150 m (i.e., the maximum station spacing) for eikonal tomography (Section 4). However, this denoising method can be applied to ANCs at even higher frequencies. Using high-quality surface waves denoised from ANCs of linear arrays crossing major fault zones for a wide range of frequencies (e.g., 0.6-3 Hz in this study; 2-40 Hz in Zigone et al., 2019), a shear wave velocity model that extends to both shallow (top few tens of meters) and deep (top 1-2 km) structures can be derived. Such fault zone velocity model with unprecedented high-resolution will complement the qualitative and semi-quantitative models inferred from traditional fault zone analyses (e.g., Qin et al., 2018, 2021; Qiu et al., 2017, 2020; Share et al., 2017, 2019). An integration of both quantitative and qualitative fault zone models can have significant implications for seismic hazard evaluations (e.g., Ben-Zion & Shi, 2005; Spudich & Olsen, 2001) and long-term behavior of the fault (e.g., Thakur et al., 2020).

6. Conclusions

We develop a simple workflow that enhances surface waves extracted from ANCs of a 1-D linear array by taking advantage of the redundant information of surface wave propagation along the same straight-line, i.e., the amount of time that surface waves travel between two stations is independent of the source location. We demonstrate the effectiveness and robustness of the three-station-interferometry-based surface wave denoising method in improving SNR of surface waves extracted from ANCs of the linear RR array, particularly at high frequencies (e.g., > 2 Hz). The proposed surface wave denoising method can be applied to a wide range of topics in the future:

1. Reduce the minimum duration of ambient noise recording and preprocessing steps needed to achieve high-quality surface waves from ANCs.
2. Provide high-quality surface wave signals both at high (> 2 Hz) and low frequencies (< 1 Hz) for better constraints of shallow (top 10s to 100s of meters) materials through full waveform tomography of surface waves.

3. Investigate the initial phase φ_F (Equation 2b) for 1-D linear arrays extracted from a 2-D deployment at different angles, and the possibility to infer the ambient noise source distribution from measurements of φ_F.
4. Since higher mode surface wave signals are observed in ANCs of the RR array and provide extra constraints on the subsurface structure, the surface wave overtones at short periods (e.g., < 0.6 s) can also be enhanced and analyzed by first attenuating the fundamental mode signal (i.e., suppress waves traveling at a speed slower than 0.55 km/s; Figs. 2a-b) and then applying the same denoising process.

Acknowledgements

We thank Fan-Chi Lin and Elizabeth Berg for providing the ambient noise cross correlation data and Rayleigh wave phase velocity model of Wang et al. (2019), and Benxin Chi for useful discussions. The manuscript benefits from useful comments from two anonymous reviewers, the Associate Editor Nori Nakata and the Editor Michael Bostock. Continuous seismic recordings of the RR array are available through the International Federation of Digital Seismography Networks (Allam, 2015; https://www.fdsn.org/networks/detail/9K_2015/). This work was supported by Rice University and the National Science Foundation (Award EAR-1251667). L. Q. is supported by the National Key R&D Program of China (2017YFC1500303).

References

Allam, A.A. (2015), San Jacinto damage zone imaging arrays. International Federation of Digital Seismograph Networks. Dataset/Seismic Network. Doi: https://doi.org/10.7914/SN/9K_2015

Baig, A. M., Campillo, M., & Brenguier, F. (2009). Denoising Seismic noise cross correlations. *Journal of Geophysical Research: Solid Earth*, *114*(8). https://doi.org/10.1029/2008JB006085

Ben-Zion, Y., & Shi, Z. (2005). Dynamic rupture on a material interface with spontaneous generation of plastic strain in the bulk. *Earth and Planetary Science Letters*. https://doi.org/10.1016/j.epsl.2005.03.025

Bensen, G. D., Ritzwoller, M. H., Barmin, M. P., Levshin, A. L., Lin, F.-C., Moschetti, M. P., 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

Catchings, R. D., Goldman, M. R., Steidl, J. H., Chan, J. H., Allam, A. A., Criley, C. J., et al. (2020). Nodal Seismograph Recordings of the 2019 Ridgecrest Earthquake Sequence. *Seismological Research Letters*. https://doi.org/10.1785/0220200203

Cheng, F., Chi, B., Lindsey, N. J., Dawe, T. C., & Ajo-Franklin, J. B. (2021). Utilizing distributed acoustic sensing and ocean bottom fiber optic cables for submarine structural characterization. *Scientific Reports*. https://doi.org/10.1038/s41598-021-84845-y

Froment, B., Campillo, M., & Roux, P. (2011). Reconstructing the Green's function through iteration of correlations. *Comptes Rendus - Geoscience*. https://doi.org/10.1016/j.crte.2011.03.001

Hillers, G., Campillo, M., Ben-Zion, Y., & Landès, M. (2013). Interaction of microseisms with crustal heterogeneity: A case study from the San Jacinto fault zone area. *Geochemistry, Geophysics, Geosystems*, *14*(7), 2182–2197. https://doi.org/10.1002/ggge.20140

Hillers, G., Campillo, M., Ben-Zion, Y., & Roux, P. (2014). Seismic fault zone trapped noise. *Journal of Geophysical Research: Solid Earth*. https://doi.org/10.1002/2014JB011217

Lin, F.-C., & Ritzwoller, M. H. (2011). Apparent anisotropy in inhomogeneous isotropic media. *Geophysical Journal International*. https://doi.org/10.1111/j.1365-246X.2011.05100.x

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*. https://doi.org/10.1111/j.1365-246X.2008.03720.x

Lin, F.-C., Ritzwoller, M. H., & Snieder, R. (2009). Eikonal tomography: Surface wave tomography by phase front tracking across a regional broad-band seismic array. *Geophysical Journal International*. https://doi.org/10.1111/j.1365-

246X.2009.04105.x

Moreau, L., Stehly, L., Boué, P., Lu, Y., Larose, E., & Campillo, M. (2017). Improving ambient noise correlation functions with an SVD-based Wiener filter. *Geophysical Journal International*, *211*(1), 418–426. https://doi.org/10.1093/GJI/GGX306

Nakata, N. (2020). Pseudo arrivals generated by frequency normalization for seismic interferometry with scattered waves – Stationary-phase analysis. In *SEG Technical Program Expanded Abstracts* (pp. 2085–2089). Society of Exploration Geophysicists. https://doi.org/10.1190/segam2020-3427903.1

Park, C. B., Miller, R. D., & Xia, J. (1999). Multichannel analysis of surface waves. *Geophysics*. https://doi.org/10.1190/1.1444590

Qin, L., Ben-Zion, Y., Qiu, H., Share, P. E., Ross, Z. E., & Vernon, F. L. (2018). Internal structure of the San Jacinto fault zone in the trifurcation area southeast of Anza, California, from data of dense seismic arrays. *Geophysical Journal International*, *213*(1), 98–114. https://doi.org/10.1093/gji/ggx540

Qin, L., Share, P.-E., Qiu, H., Allam, A. A., Vernon, F. L., & Ben-Zion, Y. (2021). Internal structure of the San Jacinto fault zone at the Ramona Reservation, north of Anza, California, from dense array seismic data. *Geophysical Journal International, 224*(2). https://doi.org/10.1093/gji/ggaa482

Qiu, H., Ben-Zion, Y., Ross, Z. E., Share, P. E., & Vernon, F. L. (2017). Internal structure of the San Jacinto fault zone at Jackass Flat from data recorded by a dense linear array. *Geophysical Journal International*, *209*(3), 1369–1388. https://doi.org/10.1093/gji/ggx096

Qiu, H., Lin, F.-C., & Ben-Zion, Y. (2019). Eikonal tomography of the Southern California plate boundary region. *Journal of Geophysical Research: Solid Earth*, 2019JB017806. https://doi.org/10.1029/2019JB017806

Qiu, H., Allam, A. A., Lin, F., & Ben-Zion, Y. (2020). Analysis of Fault Zone Resonance Modes Recorded by a Dense Seismic Array Across the San Jacinto Fault Zone at Blackburn Saddle. *Journal of Geophysical Research: Solid Earth*, *125*(10). https://doi.org/10.1029/2020jb019756

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. https://doi.org/10.1111/j.1365-246X.2010.04861.x

Share, P. E., Ben-Zion, Y., Ross, Z. E., Qiu, H., & Vernon, F. L. (2017). Internal structure of the San Jacinto fault zone at Blackburn Saddle from seismic data of a linear array. *Geophysical Journal International*, *210*(2), 819–832. https://doi.org/10.1093/gji/ggx191

Share, P. E., Allam, A. A., Ben-Zion, Y., Lin, F.-C., & Vernon, F. L. (2019). Structural Properties of the San Jacinto Fault Zone at Blackburn Saddle from Seismic Data of a Dense Linear Array. *Pure and Applied Geophysics*, *176*(3), 1169–1191. https://doi.org/10.1007/s00024-018-1988-5

Sheng, Y., Nakata, N., & Beroza, G. C. (2018). On the Nature of Higher-Order Ambient Seismic Field Correlations. *Journal of Geophysical Research: Solid Earth*. https://doi.org/10.1029/2018JB015937

Snieder, R. (2004). Extracting the Green's function from the correlation of coda waves: A derivation based on stationary phase. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*. https://doi.org/10.1103/PhysRevE.69.046610

Spica, Z., Perton, M., Calò, M., Legrand, D., Córdoba-Montiel, F., & Iglesias, A. (2016). 3-D shear wave velocity model ofMexico and South US: Bridging seismic networks with ambient noise cross-correlations (C1) and correlation of coda of correlations (C3). *Geophysical Journal International*. https://doi.org/10.1093/gji/ggw240

Spudich, P., & Olsen, K. B. (2001). Fault zone amplified waves as a possible seismic hazard along the Calaveras fault in central California. *Geophysical Research Letters*. https://doi.org/10.1029/2000GL011902

Stehly, L., Campillo, M., Froment, B., & Weaver, R. L. (2008). Reconstructing Green's function by correlation of the coda of the correlation (C3) of ambient seismic noise. *Journal of Geophysical Research: Solid Earth*. https://doi.org/10.1029/2008JB005693

Thakur, P., Huang, Y., & Kaneko, Y. (2020). Effects of Low-Velocity Fault Damage Zones on Long-Term Earthquake Behaviors on Mature Strike-Slip Faults. *Journal of Geophysical Research: Solid Earth*. https://doi.org/10.1029/2020JB019587

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

Wang, Y., Allam, A. A., & Lin, F.-C. (2019). Imaging the Fault Damage Zone of the San Jacinto Fault Near Anza With Ambient Noise Tomography Using a Dense Nodal Array. *Geophysical Research Letters*. https://doi.org/10.1029/2019GL084835

Zhang, C., Yao, H., Liu, Q., Zhang, P., Yuan, Y. O., Feng, J., & Fang, L. (2018). Linear Array Ambient Noise Adjoint Tomography Reveals Intense Crust-Mantle Interactions in North China Craton. *Journal of Geophysical Research: Solid Earth*. https://doi.org/10.1002/2017JB015019

Zhang, S., Feng, L., & Ritzwoller, M. H. (2020). Three-station interferometry and tomography: Coda versus direct waves. *Geophysical Journal International*, *221*(1), 521–541. https://doi.org/10.1093/gji/ggaa046

Zigone, D., Ben-Zion, Y., Lehujeur, M., Campillo, M., Hillers, G., & Vernon, F. L. (2019). Imaging subsurface structures in the San Jacinto fault zone with high-frequency noise recorded by dense linear arrays. *Geophysical Journal International, 217*(2), 879–893. https://doi.org/10.1093/gji/ggz069

Figure captions

Figure 1. (a) Google map for the RR array (colored balloons) deployment that crosses surface traces of the San Jacinto fault (colored lines). The stations colored in white are not analyzed in this study, whereas the green balloons denote three sensors closest to each corresponding fault surface trace. Surface wave denoising procedure is demonstrated for an example station pair RR10 and RR40 (red balloons). (b) Zoom out map of the San Jacinto fault zone. The background colors indicate topography. The red star and blue square denote locations of the RR array and the town of Anza. The black lines illustrate surface traces of major faults in this area. EF – Elsinore Fault; SAF – San Andreas Fault; SJF – San Jacinto Fault. (c) Ambient noise cross correlations at TT component of all station pairs for the sub-linear-array RR01-RR47. The cross correlations

are arranged according to interstation distance with red and blue colors representing positive and negative values. All the waveforms are first tapered using a velocity range of 2 km/s and 0.1 km/s (dashed lines), and then bandpass filtered between 0.2 and 10 Hz. (d) Same as (c) for the ZZ component. The black waveforms in (c) and (d) are the TT and ZZ component correlation functions of the station pair RR10 and RR40, respectively.

Figure 2. (a)-(b) Array-mean group velocity dispersion images (Text S1) for Love and Rayleigh waves extracted from ambient noise cross correlations (ANCs) at TT and ZZ component, respectively. The white dashed line denotes a group speed of 0.55 km/s that separates the energy of the fundamental mode surface waves from that of the higher-mode signals. (c)-(d) Multichannel analysis of surface waves (MASW; Park et al., 1999) for Love and Rayleigh waves, respectively. The black dashed line splits the target period range with the higher-mode surface waves only visible at short periods. (e)-(f) Same as (c)-(d) for ANCs after attenuating signals of higher-mode surface waves (Section 2). The background colors of each panel illustrate the likelihood of the array-mean surface wave velocity dispersion, with blue and red representing values between 0 and 1. The blue dashed curve and white markers illustrate the array-mean surface wave velocity dispersions inferred from data before and after denoising (Figs. S4-S5), respectively.

Figure 3. (a) Flow chart of the surface wave denoising and imaging procedure developed in this study. The dashed box outlines the part of the diagram that performs surface wave tomography. The workflow adopted in this study for surface wave tomography is shown in (b). ANC – Ambient Noise Cross-correlation. The filters applied to the ANCs before denoising are determined by the black dashed line shown in Figure 2. The removal of higher-mode surface waves at high frequencies are described in section 2.

Figure 4. (a) Ambient noise cross correlations (ANCs) of TT component narrow bandpass filtered at 0.8 s associated with the virtual source RR10 (red star). Waveforms are arranged by the station number of the virtual receiver. (b) Same as (a) for virtual source RR40. Waveforms in black and blue represent ANCs of virtual receivers in the outer- and inter-source zones, respectively, while the red waveform denotes the ANCs of the station pair RR10 and RR40 (red balloons in Fig. 1a). (c) Interferograms (colors) calculated via three-station interferometry, i.e., $\tilde{I}_{i\_j}^{c}(\omega;k)$ in equation 5. The gray

waveform denotes the filtered ANC of the station pair RR10 and RR40, whereas the linear stack of all the interferograms is shown in black. The black dashed lines (same as Fig. 5a) outline the surface wave signal. (d)-(f) same as (a)-(c) for F-ANC, i.e., data after the attenuation of higher-mode signals (Section 2; Fig. 3a), narrow bandpass filtered at 0.3 s.

Figure 5. (a) Comparison between the ANC (in black) and denoised waveforms (in red) of the station pair RR10-RR40 narrow bandpass filtered at 0.8 s for TT component. The black dashed lines outline the surface wave signal ($\pm$2 periods centered on the envelope peak of C^4). (b) ANCs narrow bandpass filtered at 0.8 s for TT component, with red and blue colors representing positive and negative values. The three white dashed lines illustrate moveout velocities of 2 km/s, 0.5 km/s, and 0.1 km/s. The black waveform is the same as the bottom black waveform in (a). (c) The denoised C^4 wavefield narrow bandpass filtered at 0.8 s. The white dashed line denotes a moveout velocity of 0.5 km/s. The black waveform is the same as the top red waveform in (a). (d) Array-mean amplitude spectra of waveforms shown in (b) and (c) are depicted in black and red, respectively. The peak frequency of the red amplitude spectrum (red dashed line) is labeled in the top right corner of the panel (c). (e)-(h) same as (a)-(c) for F-ANC, i.e., data after the attenuation of higher-mode signals (Section 2; Fig. 3a), narrow bandpass filtered at 0.3 s.

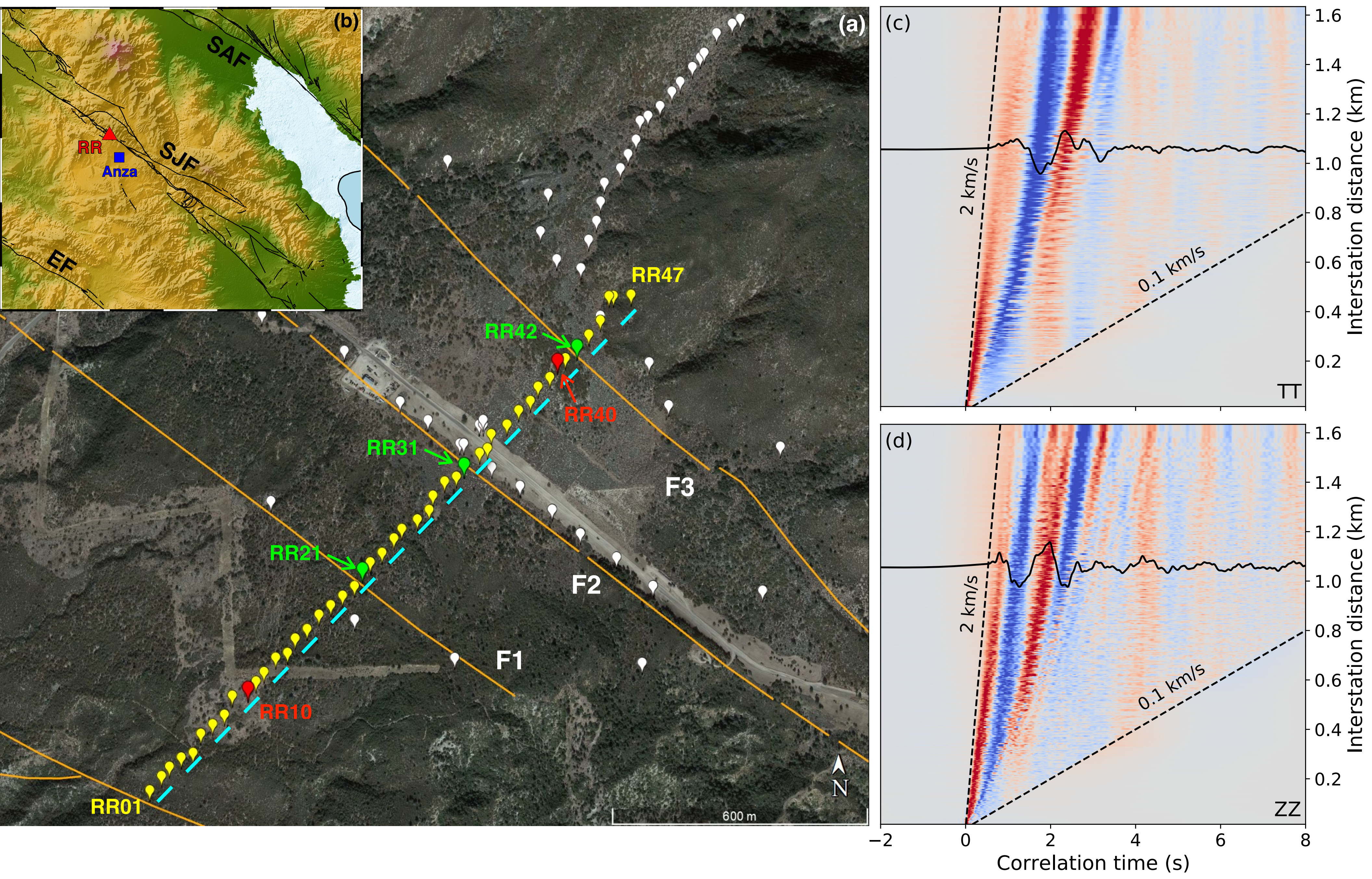
Figure 6. (a) Love waves associated with the virtual source RR10 extracted from the wavefield after denoising and tapering. The narrow bandpass filter centered at 0.3 s is applied. Red and blue colors represent positive and negative values, respectively. White circles denote cycle-skipped phase travel time measurements extracted at the peak frequency of the filtered wavefield (3.14 Hz), whereas red stars indicate travel times after cycle skipping correction (Section 4.1). The black curve illustrates the corrected phase travel time after interpolation using a grid size of 50 m. (b) Phase velocity profiles resolved for Love waves at 3.14 Hz. White circles depict the 1-D phase velocity profile derived via eikonal equation for virtual source RR10, i.e., using the black curve in (a). The colormap illustrates phase velocity profiles obtained using different stations as virtual sources (*x*-axis), with white space illustrating the near-source exclusion zone. The

phase velocity profile averaged over results of all virtual sources are depicted as red stars, with the error bar representing the corresponding standard deviation. The black vertical dashed line denotes the array-mean phase velocity estimated at 0.3 s (red star in Fig. S4b).

Figure 7. Phase velocity dispersion profiles for (a) Love and (b) Rayleigh waves beneath the RR array. The vertical dashed lines denote locations of the mapped fault surface traces (Fig. 1a). Uncertainties of the resolved phase velocity profiles are shown in (c) and (d) for Love and Rayleigh waves, respectively.

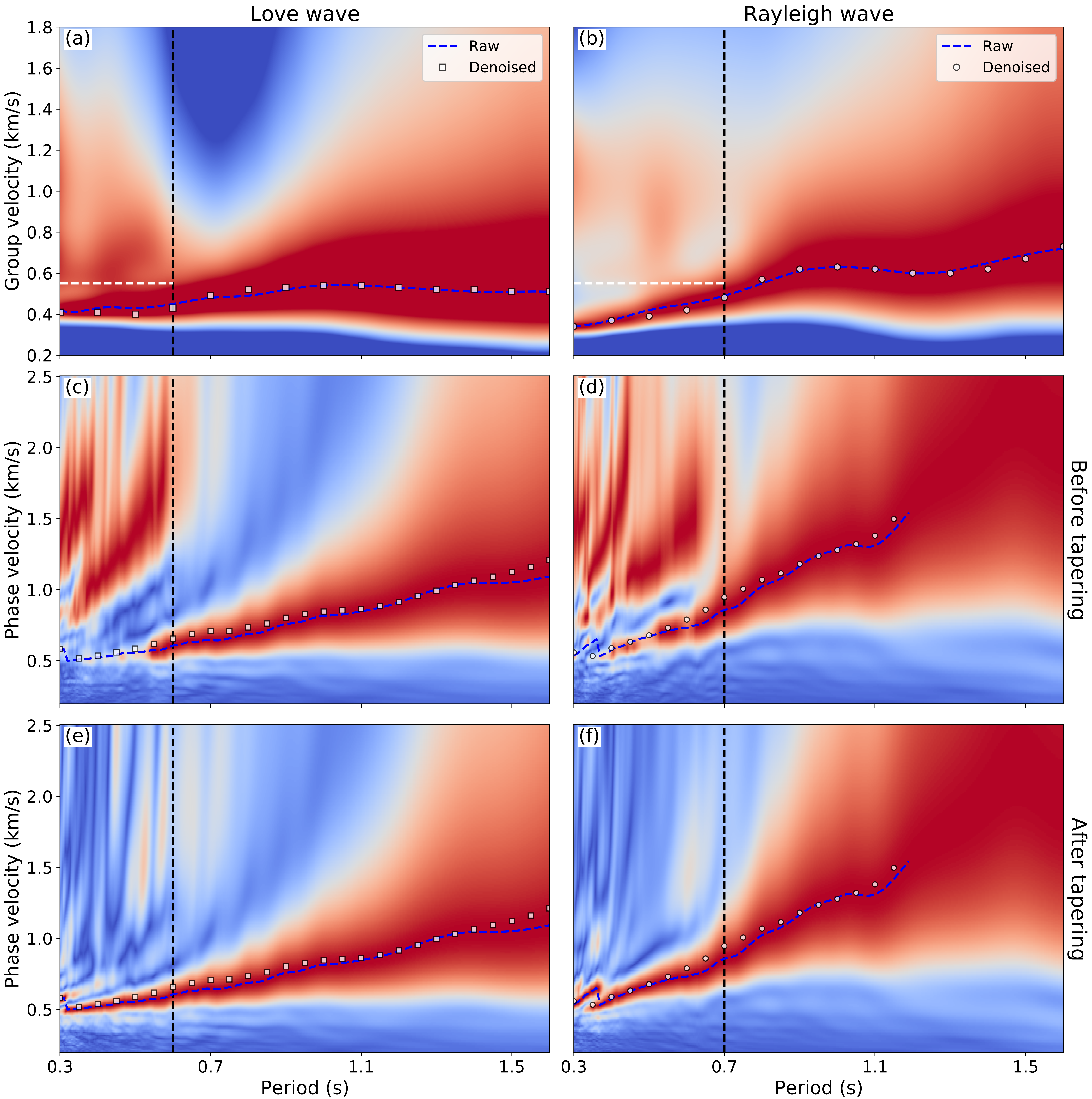
Figure 8. Comparison of (a) phase velocity and (b) uncertainty profiles of this study and those, (b) and (d), of Wang et al. (2019) in the overlapping period, 0.3-0.8 s. and spatial, RR01-RR47, ranges. The white space indicates the area not covered by the final model.

Figure 1.



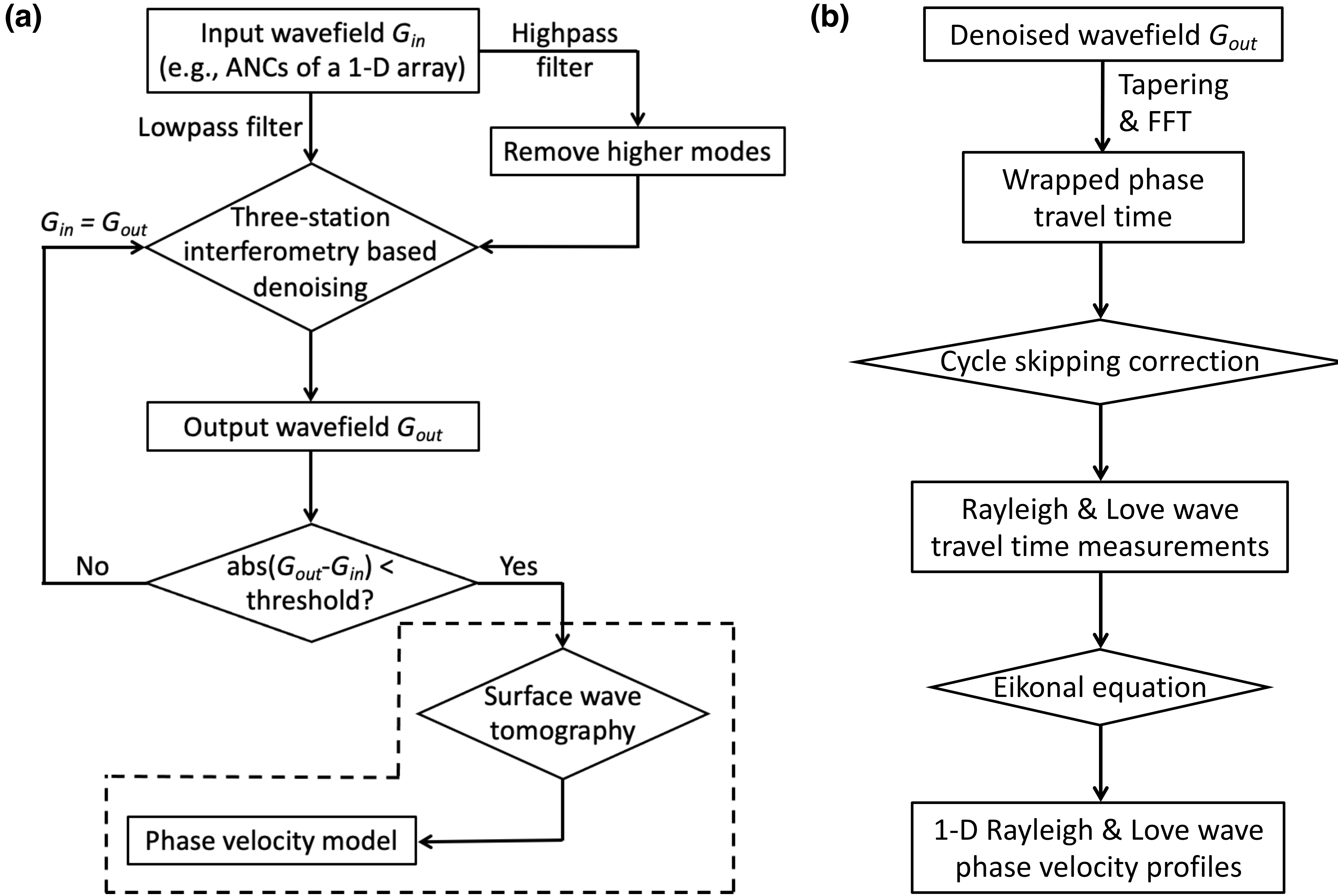
(a)
(b)
SAF
SJF
EF
RR
Anza
RR47
RR42
RR40
RR31
RR21
RR10
RR01
F1
F2
F3
N
600 m
(c)
(d)
2 km/s
0.1 km/s
TT
ZZ
Interstation distance (km)
Correlation time (s)

Figure 2.



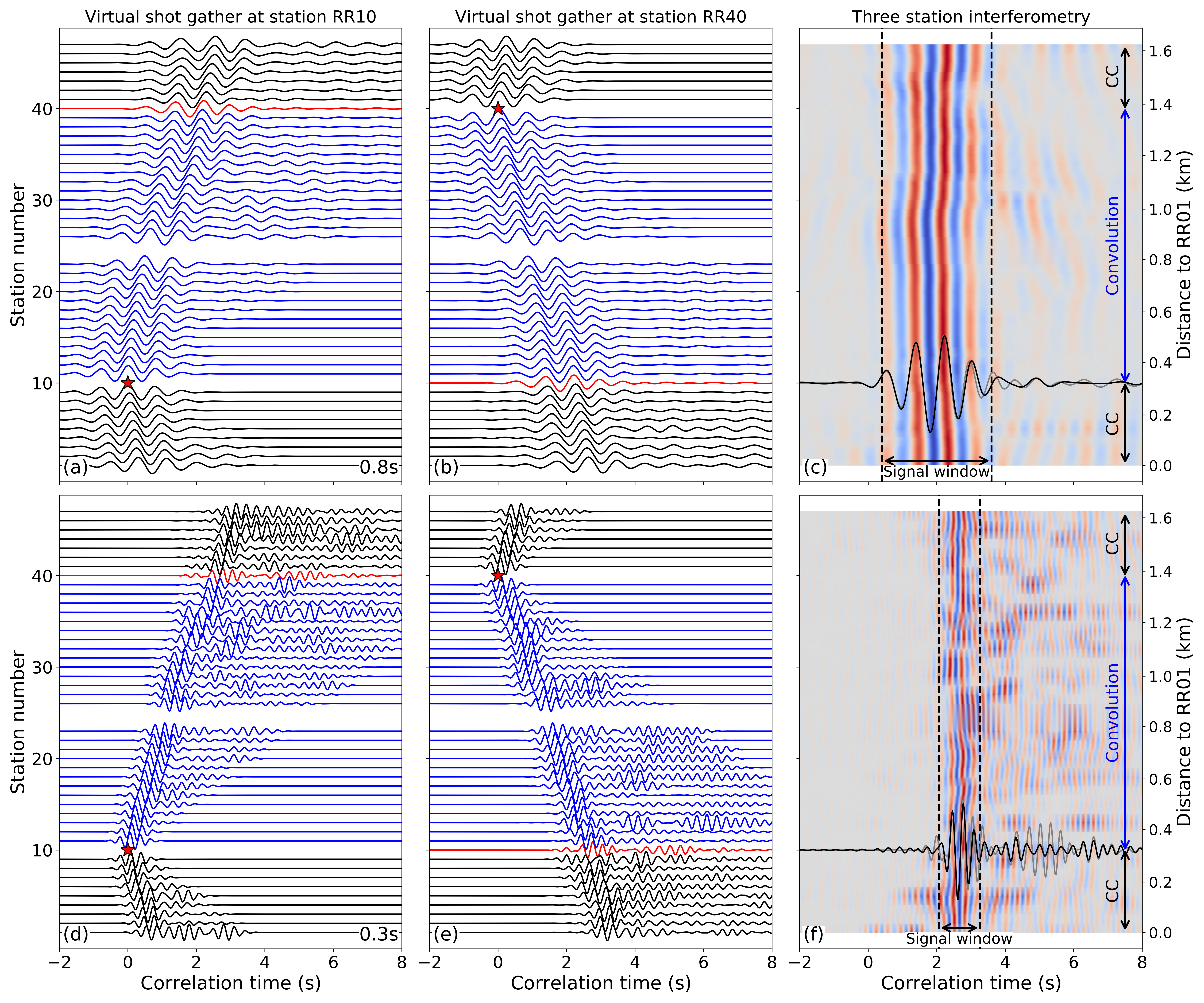
Love wave
Rayleigh wave
(a)
(b)
(c)
(d)
(e)
(f)
Raw
Denoised
Group velocity (km/s)
Phase velocity (km/s)
Before tapering
After tapering
Period (s)
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
2.5
2.0
1.5
1.0
0.5
0.3
0.7
1.1
1.5

Figure 3.



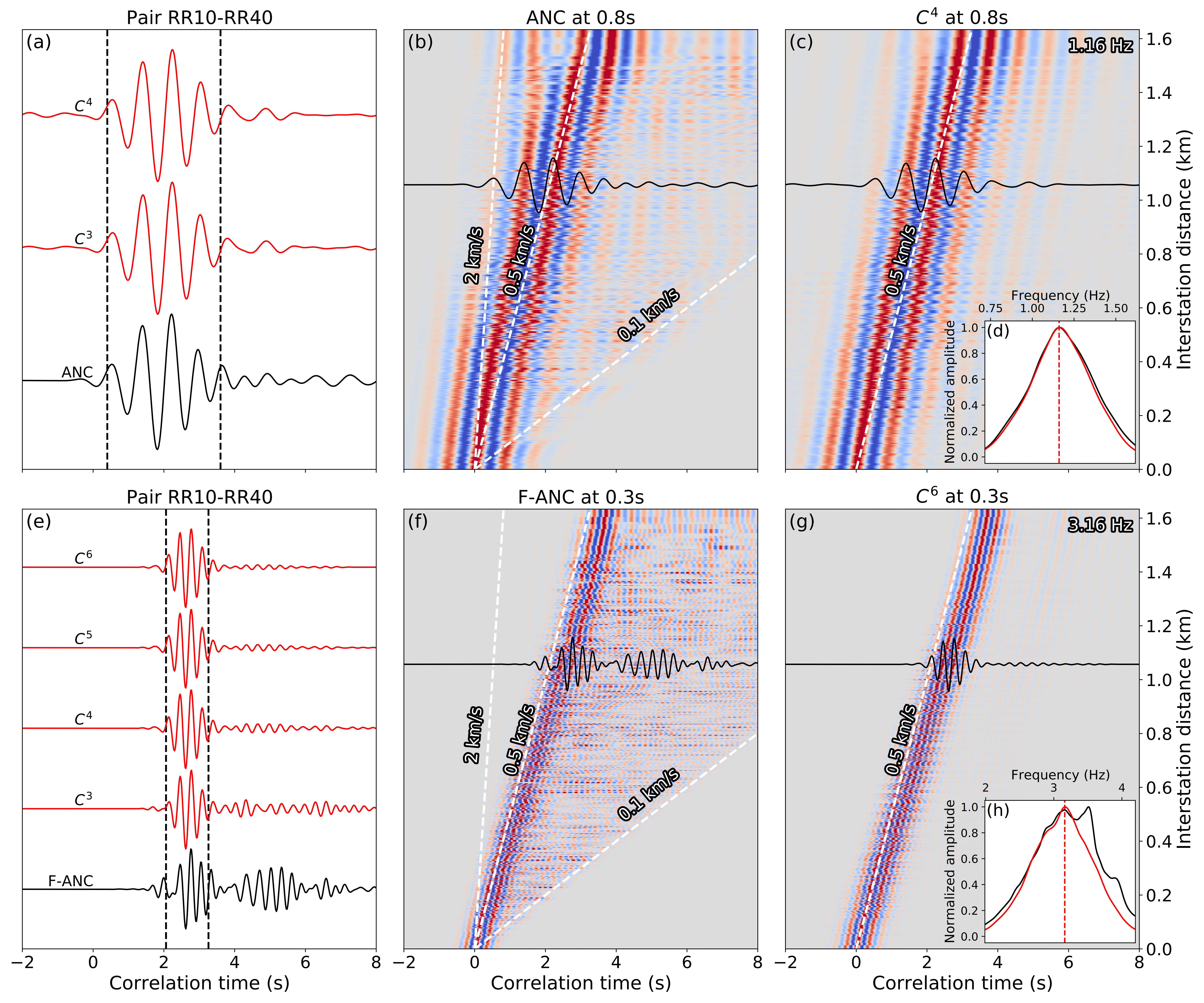
(a)
Input wavefield G_{in}
(e.g., ANCs of a 1-D array)
Highpass
filter
Lowpass filter
Remove higher modes
$G_{in} = G_{out}$
Three-station
interferometry based
denoising
Output wavefield G_{out}
No
Yes
abs(G_{out}-G_{in}) <
threshold?
Surface wave
tomography
Phase velocity model
(b)
Denoised wavefield G_{out}
Tapering
& FFT
Wrapped phase
travel time
Cycle skipping correction
Rayleigh & Love wave
travel time measurements
Eikonal equation
1-D Rayleigh & Love wave
phase velocity profiles

Figure 4.



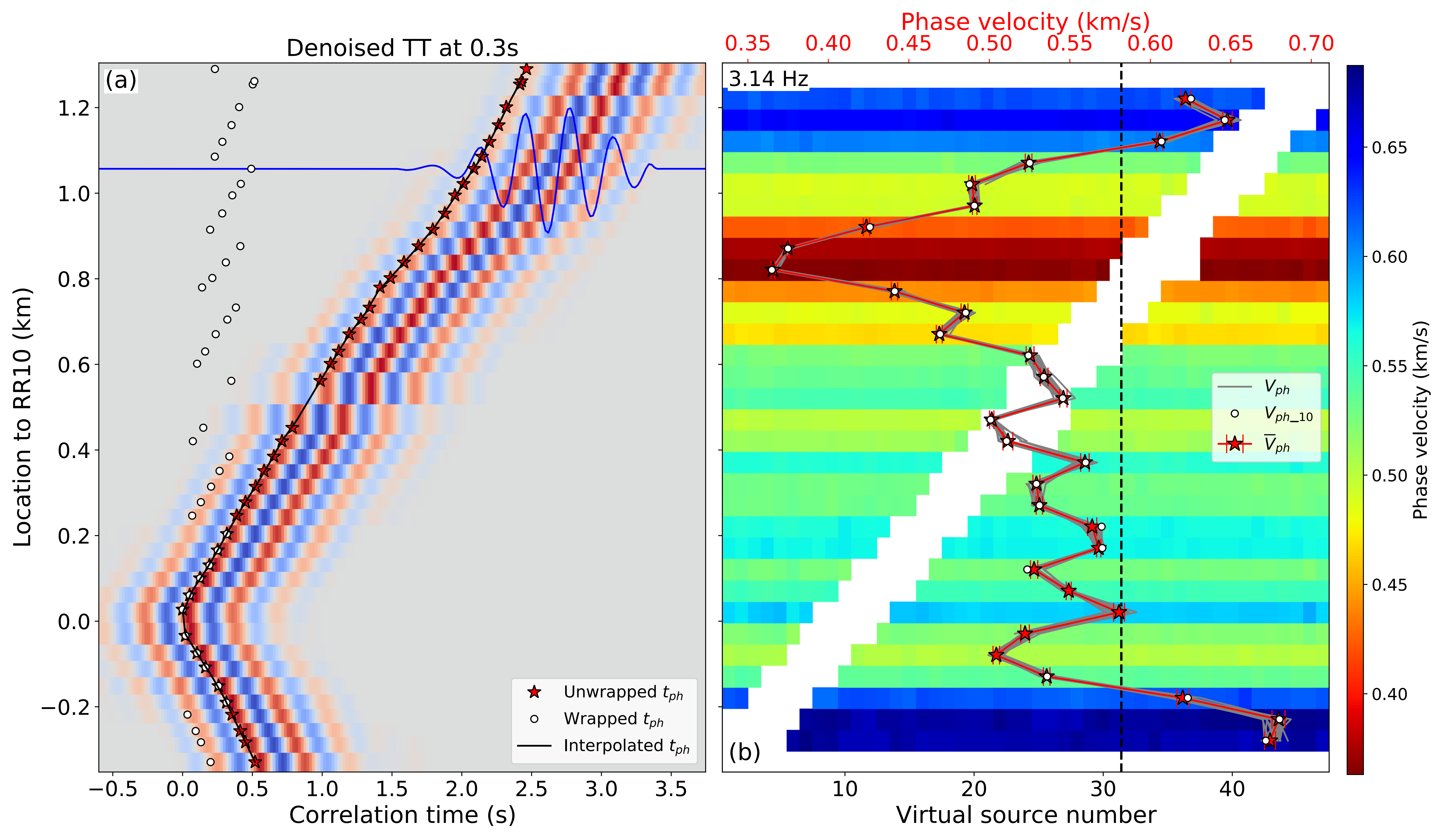
Virtual shot gather at station RR10
Virtual shot gather at station RR40
Three station interferometry
Station number
Correlation time (s)
Distance to RR01 (km)
(a)
0.8s
(b)
(c)
(d)
0.3s
(e)
(f)
CC
Convolution
Signal window

Figure 5.



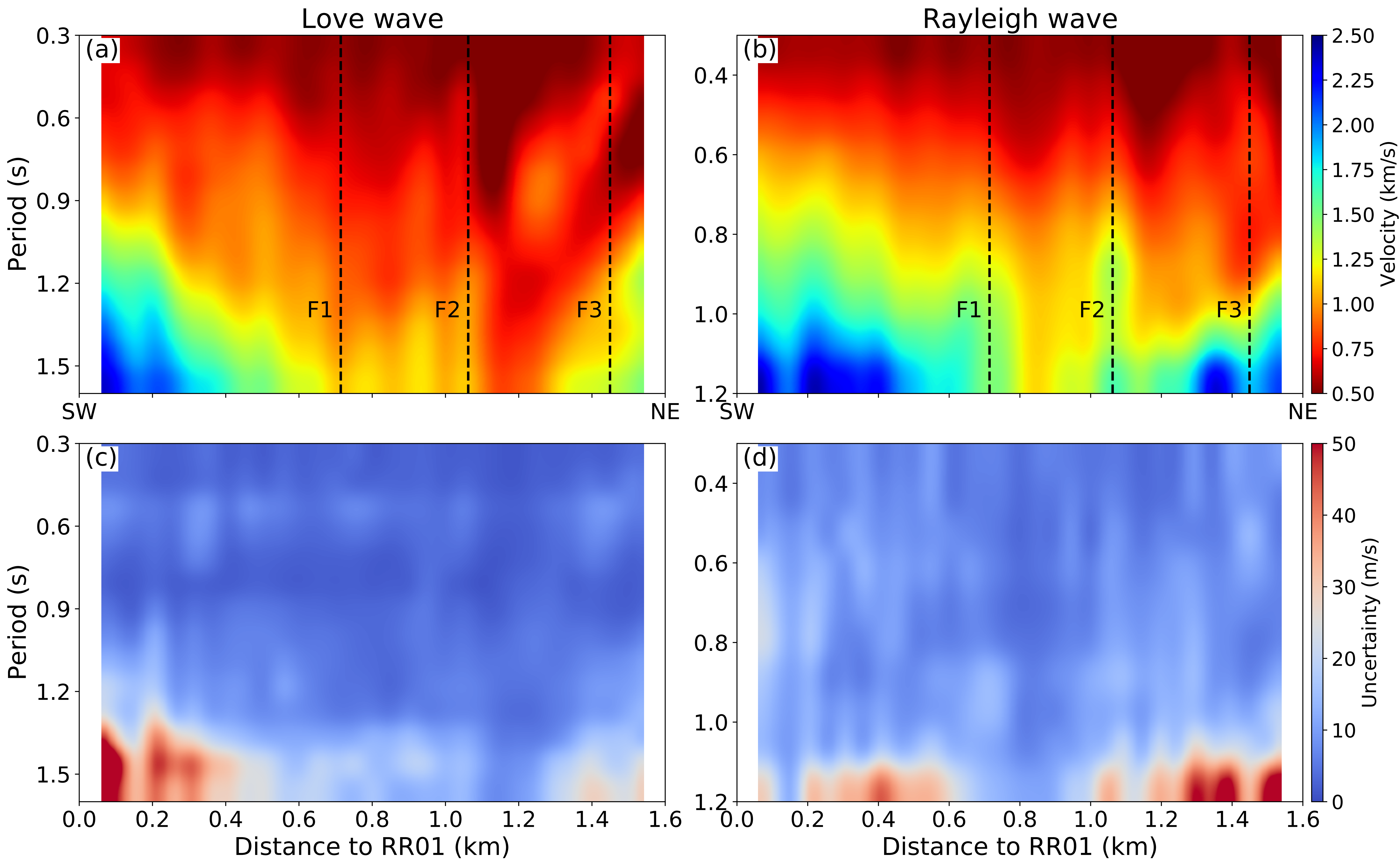
Pair RR10-RR40
ANC at 0.8s
C^4 at 0.8s
(a)
(b)
(c)
1.16 Hz
C^4
C^3
ANC
2 km/s
0.5 km/s
0.1 km/s
Frequency (Hz)
0.75
1.00
1.25
1.50
(d)
Normalized amplitude
Interstation distance (km)
Pair RR10-RR40
F-ANC at 0.3s
C^6 at 0.3s
(e)
(f)
(g)
3.16 Hz
C^6
C^5
C^4
C^3
F-ANC
(h)
Correlation time (s)

Figure 6.



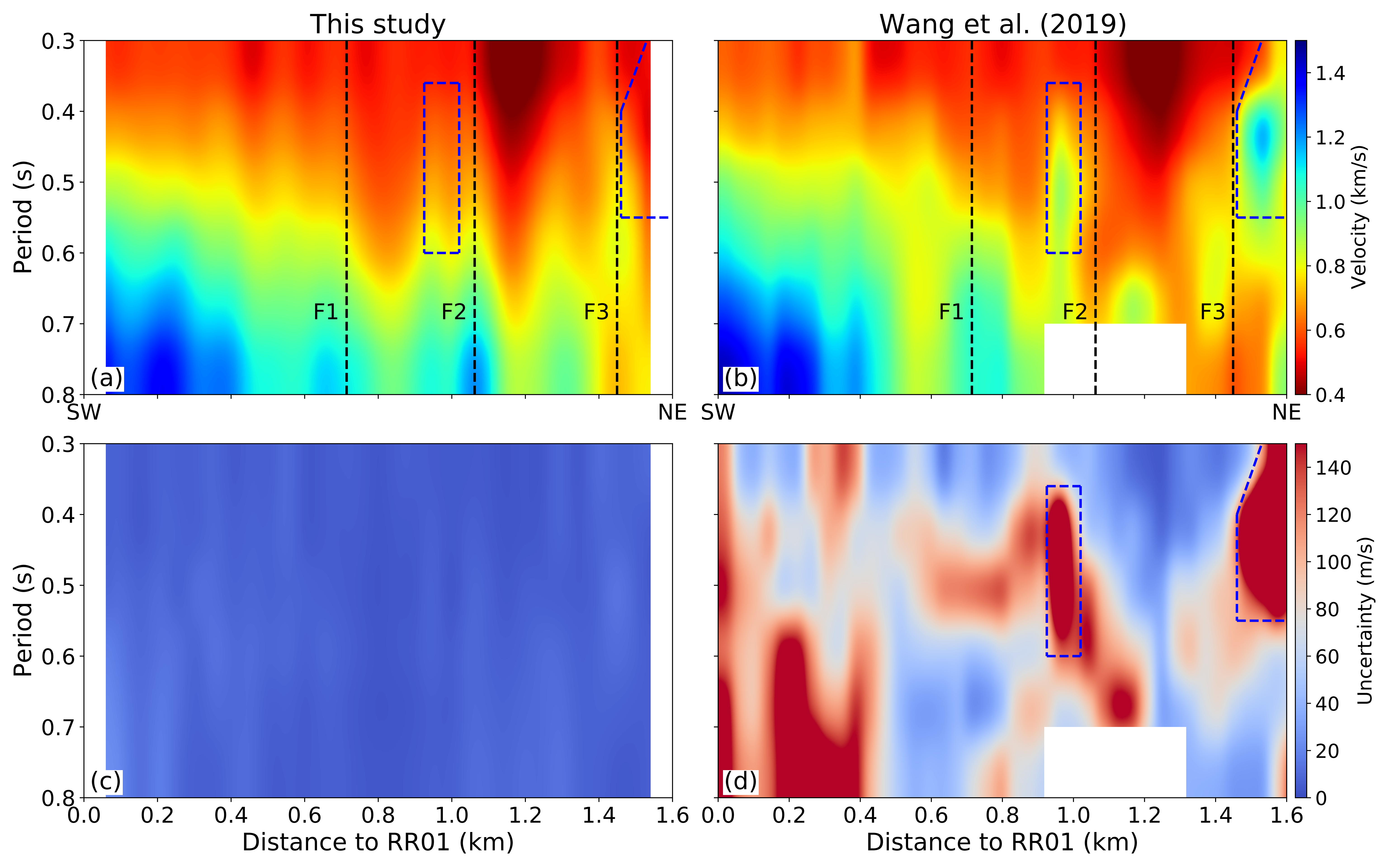
Denoised TT at 0.3s
(a)
Location to RR10 (km)
Correlation time (s)
Unwrapped t_{ph}
Wrapped t_{ph}
Interpolated t_{ph}
Phase velocity (km/s)
3.14 Hz
(b)
V_{ph}
V_{ph_10}
$\overline{V}_{ph}$
Virtual source number
Phase velocity (km/s)

Figure 7.



Love wave
Rayleigh wave
(a)
(b)
(c)
(d)
F1
F2
F3
SW
NE
Period (s)
Velocity (km/s)
Uncertainty (m/s)
Distance to RR01 (km)

Figure 8.



This study
Wang et al. (2019)
Period (s)
0.3
0.4
0.5
0.6
0.7
0.8
F1
F2
F3
(a)
(b)
(c)
(d)
SW
NE
Velocity (km/s)
0.4
0.6
0.8
1.0
1.2
1.4
Uncertainty (m/s)
0
20
40
60
80
100
120
140
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
Distance to RR01 (km)