Isolating and Suppressing the Spurious Non-Diffuse Contributions to Ambient Seismic Field Correlations

Xin Liu[1], Gregory C. Beroza[1] and Nori Nakata[2]

1. Department of Geophysics, Stanford University

2. Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology

Journal of Geophysical Research

In review, 2018

Short title:

Separating non-diffuse noise

Key points

1. We develop a method to isolate and suppress the spurious non-diffuse contributions in noise interferometry.

2. Beamforming of spurious non-diffuse and preferred Rayleigh components shows distinct patterns for primary and secondary ocean microseisms.

3. Our method reduces the bias in seismic velocity estimations due to spurious non-diffuse noise.

Abstract

Cross-correlation of fully diffuse wavefields averaged over time should converge to the Green's function; however, the ambient seismic field in the real Earth is not fully diffuse, which interferes with that convergence. We apply blind signal separation to reduce the effect of spurious non-diffuse components on the cross-correlation tensor of the ambient seismic field. We describe the diffuse component as having uncorrelated neighboring frequencies and equal intensity at all azimuths, and an independent (i.e., statistically uncorrelated) non-diffuse component arising from a spatially isolated point source for which neighboring frequencies are correlated. Under the assumption of linear independence of the spurious non-diffuse wave outside the stationary phase zone and the constructive interference of noise waves within that zone, we can suppress the spurious non-diffuse component from the noise interferometry. Our numerical simulations show good separation of one spurious non-diffuse noise source component for either non-diffuse Rayleigh or Love waves. We apply this separation to the Rayleigh-wave component of the Green's function for 136 cross-correlation pairs from 17 stations in Southern California. We perform beamforming over different frequency bands for the cross-correlations before and after the separation, and find that the reconstructed Rayleigh waves are more coherent. We also estimate the bias in Rayleigh wave phase velocity for each receiver pair due to the spurious non-diffuse contribution.

1. Introduction

Extracting the Green's function from seismic noise interferometry has led to many new developments and successful applications in travel time tomography *(Shapiro & Campillo,* 2004; *Shapiro,* 2005; *Curtis et al.,* 2009; *Nakata et al.*, 2015), monitoring velocity changes (*Brenguier et al.,* 2008), and site effect/attenuation measurements *(Prieto et al.,* 2009*; Tsai,* 2011*; Lin et al.,* 2012*; Liu et al.,* 2015). Based on normal-mode analyses, retrieval of the exact Green's function between two locations requires that the noise wavefield be fully diffuse, which implies that the wave modes of different eigen-frequencies are not correlated *(Weaver,* 1982*; Lobkis & Weaver,* 2001*; Weaver & Lobkis,* 2004) and propagate in all directions with equal energy. This theory only applies to an

object of finite size with standing wave modes (e.g. a disk or the entire Earth); however, for a locally defined wavefield in an open system, a diffuse field can be generated by localized sources and/or scatterers provided that the sources at different locations are not correlated and the noise intensity is the same in all directions (*Snieder*, 2004; *Weaver and Lobkis*, 2004).

A non-diffuse wavefield would result from isolated point forces (e.g. ocean storms) that are not continuously distributed in different directions or are not stationary in time. Moreover, these point sources will have finite source spectral widths, which will introduce correlations across neighboring frequencies in the ambient seismic field (*Liu and Ben-Zion*, 2016). Despite these contrasts in statistical properties compared with a diffuse field, separating the non-diffuse and diffuse fields is difficult because they usually mix in the continuous recordings and the non-diffuse source is non-stationary. In more realistic cases, the well-distributed ocean noise sources contribute to the reconstruction of the ambient noise empirical Green's function, while the responses at the receivers may be influenced by the signature of the sources in the form of correlated neighboring frequencies. We demonstrate that separating a strong localized source contribution from the contributions of more broadly distributed ocean noise sources is a practical and useful approach.

In this paper, we introduce a blind signal separation method to separate the spurious non-diffuse wave component from the Rayleigh wave contributions with redundant observations (polarizations) of ambient field cross-correlations. This blind signal separation method does not require the diffuse/non-diffuse property and it assumes that the separable components are independent of each other. Our synthetic test demonstrates that we can effectively separate the non-diffuse contribution for one spurious source outside the stationary phase zone. We find that the cross-correlation functions determined for the preferred Rayleigh-wave component contain fewer spurious acausal arrivals before the main Rayleigh wave packet, and that the preferred Rayleigh component has smoother azimuthal source power variation than for the cross-correlation functions determined without this separation. We refer to the spurious arrivals as spurious non-diffuse contributions to the ambient noise cross-correlation, as they are clearly related to

the spatially isolated and localized sources in the beamforming. We also demonstrate bias reduction in phase velocity measurements by reducing the non-diffuse contributions to the observed cross-correlations.

2. Methodology and numerical simulation

The theory of seismic noise interferometry requires that the noise sources be continuously distributed with azimuth and that the noise sources at different locations not be correlated. In reality, these two assumptions may not be valid and that can lead to artifacts in the noise cross-correlation.

The first assumption breaks when, for example, a distant ocean storm acts as an isolated source and appears as a spurious signal in the noise interferometry at a different lag time than the preferred signal due to the coherent interference of noise waves propagating within the stationary phase zone of the station pair. The second assumption partly breaks when the neighboring noise sources have spatial correlations. In either case, the resulting noise field is non-diffuse.

An ocean storm, or more generally a point force, corresponds to source spectra with finite bandwidth. If the point force repeats, the statistics of the recorded noise power spectra will exhibit neighboring frequency (up to the bandwidth) correlations (Liu & Ben-Zion, 2018), which is the definitive characteristic of non-diffuse noise. Diffuse noise, in contrast, has zero correlation among neighboring frequencies (*Weaver & Lobkis*, 2004) and equal noise intensity from all directions. The non-diffuse and diffuse components are by definition uncorrelated in the noise recordings and they do not have cross-terms in the noise cross-correlation.

Direct separation of the non-diffuse wavefield from the ambient noise wavefield is challenging. In this work, we demonstrate a special case where we separate a non-diffuse contribution that appears as a spurious signal from the noise interferometry assuming that the spurious signal is outside the stationary phase zone and therefore independent of the preferred interferometry signal produced by the constructive interference of noise wavefields within the stationary phase zone of the station pair.

In blind signal separation, one important condition for successful separation is that each time series observation contains a different linear combination of independent components and that there is no time shift for each independent component on different time series observations (Hyvärinen & Oja, 2000). A matrix containing the coefficients of those linear combinations is the mixing matrix (eq. S1 in Supplementary Text S1) and it has to be invertible for the independent components to be separable. Therefore, ideally the number of observations should be equal to the number of independent components. In the following numerical test, we apply the blind signal separation to separate two independent components from two observations.

In the numerical simulations, we assume that the expected spectral amplitude ratio of non-diffuse/diffuse noise is a Gaussian function, which fits the observed data in southern California (Liu & Ben-Zion, 2016). When the diffuse noise is white noise, the spectral amplitude of the spurious non-diffuse noise is a Gaussian function. We generate the diffuse component with uncorrelated complex Gaussian random variables (expected amplitude equals one; random phase) at different frequencies, and the non-diffuse component from a repeating point force with Gaussian-shaped spectral amplitude. We then compute the corresponding correlation matrix of neighboring frequencies for the mixed diffuse and non-diffuse fields according to the expectation of the cross-spectral amplitudes of noise data (Liu & Ben-Zion, 2016) (Fig. 1a). For a diffuse field, the correlation matrix will be diagonal. The presence of non-diffuse noise will manifest as off-diagonal correlations. In Fig. 1b, we simulate the diffuse noise using evenly distributed sources (blue dots) in a ring. The diffuse noise sources at different locations are not correlated. We follow *Liu et al.* (2015) for diffuse noise simulations in the frequency domain using time-shifted white noise models. In addition, we simulate the non-diffuse noise from one non-diffuse point source (red dot) outside the stationary phase (constructive interference) zone. We assume a simple 1D southern California velocity model (Shaw et al., 2015) to compute the wave propagation between sources and receivers. The non-diffuse source time function is based on the cross-frequency vector (the spectral amplitude) in Fig. 1a, which we convert to the time domain with the inverse Fourier transform, and insert into the source time function at random times with average

141 spacing of 100 s. The total duration of the simulated noise is 15 days for both diffuse and
142 non-diffuse components on the two receivers (Fig. 1b).

143 We consider polarization for Rayleigh and Love wave simulations. As an example,
144 the Rayleigh and Love wave modes propagating from the non-diffuse source to receiver *a*
145 can be found on R_s and T_s components, respectively (Fig. 1b). When the signal is
146 recorded on the receiver, the two horizontal axes R_s and T_s are projected to receiver
147 coordinates R (radial) and T (transverse). The Z (vertical) axis is not rotated. Therefore,
148 the non-diffuse Rayleigh wave will show up on Z, R and T components while the non-
149 diffuse Love wave will show up on R and T components.

150 In the first example, we simulate non-diffuse Rayleigh waves from one source outside
151 the stationary phase zone. We sum the synthetic diffuse and non-diffuse noise recordings
152 and cross-correlate the noise data for 5 days between two receivers *a* and *b* that are 60 km
153 apart. The raw cross-correlations for ZZ and RR components (Fig. 1c) show differences
154 due to the projection of the Rayleigh wave polarization, as the spurious non-diffuse
155 Rayleigh component is evidently smaller on RR than that on ZZ. For the secondary ocean
156 microseism in southern California, the bandwidth of non-diffuse noise is small (Liu &
157 Ben-Zion, 2018) and the spurious non-diffuse component has a wide pulse width in the
158 time domain.

159 We consider the simplest case in which there is one major spurious non-diffuse
160 source outside the stationary phase zone. The spurious non-diffuse and the preferred
161 Rayleigh components are assumed statistically independent and have different amplitudes
162 in the ZZ and RR cross-correlations. As a result, we can separate those two mixed
163 components using Independent Component Analysis (ICA), which is based on the
164 rotation of two orthogonal axes of the input ZZ and RR cross-correlations and minimizes
165 the statistical dependence between the output independent components. ICA works by
166 maximizing the non-Gaussianity and minimizing the mutual dependence between the
167 resulting components. The underlying assumption is that the independent components are
168 non-Gaussian and their linear mixtures (cross-correlation observations) are closer to
169 Gaussian according to Central Limit Theorem (Supplementary Text S1). Here we apply
170 the FastICA algorithm (*Hyvärinen and Oja*, 2000), which is a blind signal separation

technique. This method first pre-whitens the input data to uncorrelated data vectors by applying Singular Value Decomposition (SVD). Then it estimates a 2×2 de-mixing matrix (the inverse of the mixing matrix) by maximizing the non-Gaussianity of the separated (rotated) independent components. The non-Gaussianity is measured by negentropy, which is a differential measurement of entropy relative to a Gaussian. We use the FastICA algorithm in the scikit-learn Python library (http://scikit-learn.org). Details of this analysis are described in the Supplementary materials.

In order to separate the spurious non-diffuse component from the raw cross-correlations using ICA, an important condition is that the spurious non-diffuse source is outside the stationary phase zone and generates an independent response to that of the sources in the stationary phase zone in seismic interferometry. The sources within the stationary phase zone interfere constructively and a non-diffuse source within that zone is inseparable from the interferometric phases produced by constructive interference of the diffuse noise sources.

In the case of our numerical simulation, the only non-diffuse source is outside of the stationary phase zone and it generates a spurious signal on the noise cross-correlation. Therefore we can separate the spurious non-diffuse contribution to the cross-correlation function. The preferred Rayleigh component due to diffuse sources and the spurious non-diffuse Rayleigh component due to an isolated non-diffuse source are well separated after applying ICA (Fig. 1d). The correlation coefficients between the true and the recovered signals are greater than 0.98. In addition, we measure the phase velocity dispersion curves before and after separating the non-diffuse wave (Supplementary Text S2 and Fig. S2). Before the separation, the non-diffuse component causes ~2.3 % bias in velocity measurements at 0.18 Hz. After removing the non-diffuse component, the bias in phase velocity measurements reduces to ~0.2 %, which is comparable to the background fluctuation level from the stacking of 5 days data.

Another synthetic test is for non-diffuse Love waves instead of Rayleigh waves. We place the spurious non-diffuse Love source at the same location as in the previous example (Fig. 1b). The spurious non-diffuse Love wave contribution only appears on the horizontal component RR, not the vertical component ZZ (Fig. 1e). The ICA method still

works in this case even though it has zero amplitude on the ZZ component. The recovered preferred Love and spurious non-diffuse signals are also well-correlated (>0.98) with the true signals (Fig. 1f).

For a non-diffuse source located within the stationary phase zone, we cannot separate the non-diffuse component from the noise cross-correlation (Supplementary Text S2) because the waves propagating within the stationary phase zone have coherent phases and they are not independent. Therefore, the preferred Rayleigh/Love component from ICA output may contain non-diffuse contributions within the stationary phase zone.

We perform additional numerical simulations to test more realistic cases by including multiple spurious non-diffuse sources while still applying ICA to ZZ and RR observations (Supplementary Text S3, Figure S3). We show that for two spurious non-diffuse sources with equal amplitude, the ICA algorithm fails due to an insufficient number of observations. Additionally, for three spurious non-diffuse sources with different amplitudes, the major non-diffuse source with amplitude twice the others can be recovered by ICA algorithm from ZZ and RR observations, while the smaller non-diffuse sources cannot.

3. Application to regional network data in Southern California

We analyze 3-component data on 17 broadband stations in Southern California for days 100-161 of 2014 (Fig. 2b). We choose these stations to form a relatively small (less than 120×120 km) square-shaped array outside the Los Angeles Basin (low velocity) to ensure consistent array response in different directions and to ensure the validity of the plane wave assumption for beamforming. We compute both ZZ and RR component cross-correlations for all station pairs and aim to extract the Rayleigh wave. Our data are pre-processed based on a statistical outlier exclusion algorithm (*Liu et al.*, 2016; *Liu and Ben-Zion*, 2016), which removes time windows containing earthquakes and other transient signals that potentially contaminate the noise statistics without having to apply nonlinear operations such as pre-whitening or one-bit normalization (*Bensen et al.*, 2007) to the data. We use evenly spaced time windows of 200 s length, a 40 s gap between adjacent windows, and compute the cross-spectra for corresponding time windows on

two receivers. We set a criterion of 4 times the Median Absolute Deviation (MAD) to exclude anomalous cross-spectral values for each frequency. A time window is removed if it contains more than 10% outliers from 0.05-0.6 Hz.

We show an example for the station pair CHF-SBB2 with the ZZ and RR components in Fig. 2c. The ZZ component is evidently noisier with relatively higher amplitude near 0 s than the RR component. We apply ICA to the ZZ and RR components to obtain two independent components (Fig. 2d): the preferred Rayleigh wave component (top) and the non-diffuse noise component (bottom). We label an independent component as the preferred Rayleigh component because it has the same sign on the input ZZ and RR cross-correlations and has the closest group delay time to the reference group velocity dispersion based on the 1D velocity model in the center of the region from the Southern California Earthquake Center (SCEC) Community Velocity Model CVM-H (Shaw et al., 2015). Despite these labels, the preferred Rayleigh component may contain some non-diffuse contributions from smoothly distributed ocean noise sources, while one could extract information from the spurious non-diffuse component if the spurious source is well located. These two independent components are scaled such that their sum equals the ZZ component. The recovered preferred Rayleigh component is, in principle, less biased than ZZ and RR because it combines common information on both components with the independent, major spurious non-diffuse contribution removed. The spurious non-diffuse component has anomalous acausal arrivals near 0 s and overlaps with the Rayleigh wave arrivals. A significant non-diffuse component biases the Rayleigh wave phase and amplitude measurements. The RR component is equal to the sum of 0.86 times the preferred Rayleigh component and - 0.68 times the non-diffuse component. Therefore, the RR component contains less non-diffuse noise for CHF-SBB2. In addition, the spurious non-diffuse component may contain both Rayleigh and Love wave modes, which can complicate the separation.

We compare the Frequency-Time ANalysis (FTAN) results for the ZZ and preferred Rayleigh components in Figs. 2(e & f), respectively. The ZZ component shows a spurious early arrival (around 5s) between 0.45-0.57 Hz, while the preferred Rayleigh

component removes the spurious early arrival and results in less distorted wave packet energy peaks.

We apply ICA to all 136 station pairs and separate preferred Rayleigh and spurious non-diffuse components from the ZZ and RR components for each pair (Fig. 3). The cross-correlations are low-pass filtered at 0.25 Hz before ICA separation and the maximum absolute amplitude for each cross-correlation is normalized to one. Compared with the preferred Rayleigh component (Fig. 3c), the ZZ (Fig. 3a) and RR (Fig. 3a) components are noisier with spurious arrivals near 0 s, and the ZZ component contains larger spurious arrival amplitudes than the RR component. The spurious non-diffuse component figure (Fig. 3d) contains high amplitude spurious arrivals mainly between the expected Rayleigh arrival times for both causal and anti-causal sides of the result. Some spurious arrivals also arrive after the Rayleigh wave, suggesting temporal correlation structure in the non-diffuse noise sources (Liu et al., 2016), which can contaminate the coda of ZZ and RR cross-correlations. There are weak low-frequency waves arriving around the expected Rayleigh wave travel time in the spurious non-diffuse component figure. One possible explanation would be that the non-diffuse sources within the stationary phase zone constructively interfere with the delta-correlated diffuse sources and that they are difficult to separate based on observations of ZZ and RR cross-correlations. We measure the Root Mean-Square (RMS) of the spurious non-diffuse waveform before the Rayleigh wave arrivals (between two dashed lines in Fig. 3a) normalized by its maximum absolute amplitude for ZZ, RR and preferred Rayleigh panels (Fig. 3e). The normalized RMS values are binned in 10-km distance bins and the results suggest that the preferred Rayleigh components have the lowest spurious non-diffuse contributions for almost all distance bins.

We apply beamforming to the different components to examine the azimuthal distribution of the incoming wavefield before and after separating the spurious non-diffuse component (Fig. 4) to examine the effectiveness of the non-diffuse component suppression method. We first apply a band-pass filter to the cross-correlations in Fig. 3 between 0.05-0.10 Hz that includes the primary microseism. We then apply conventional Bartlett beamforming based on matched field processing (Porter & Tolstoy, 1994) to all

the cross-correlation components. The beamforming results from 0.05-0.1 Hz for the four components, ZZ, RR, preferred Rayleigh and spurious non-diffuse are shown in Fig. 4(a). The primary microseism energy comes mainly from the southwest, at azimuths of 160º-320º, with slowness around 0.28 s/km. The ZZ and RR beamforming results both show strong noise sources in different directions. The preferred Rayleigh component beamforming recovers common features between the ZZ and RR results and the source (beam) power is smoother and has better coverage for both ZZ and RR. The spurious non-diffuse component beamforming result contains a strongly localized source with a peak power comparable to the Rayleigh component at 210º and some background energy from 180º-320º. The presence of this strongly localized non-diffuse source will bias phase velocity and amplitude measurements of individual cross-correlation functions.

We filter the cross-correlation data between 0.12-0.18 Hz to focus on the secondary microseism energy (Fig. 4b). The beamforming results of the ZZ and RR components have a narrower azimuthal response of 160º-290º with significant variation in source power with azimuth. The beamforming of the preferred Rayleigh wave component has less variation in source power with azimuth. The beamforming of the spurious non-diffuse component contains many noise sources, with a few strong isolated sources from the southwest arriving at the same speed as the Rayleigh wave. The other sources, however, appear in random directions at different apparent velocities and may contain body waves. This distinction from the beamforming results of the primary microseism suggests a different source generating mechanism for the primary and the secondary microseisms (*Longuet-Higgins*, 1950; *Hasselmann*, 1963; *Cessaro*, 1994; *Stutzmann et al*., 2009).

The spurious non-diffuse contributions to the ZZ and RR cross-correlations will introduce bias into phase velocity measurements. The preferred Rayleigh component still contains travel time bias due to non-uniform azimuthal noise illuminations (*Weaver et al.*, 2009) that can be compensated using the beamforming results, but the bias is less than that for the ZZ and RR cross-correlations. Here we estimate the bias caused by the spurious non-diffuse contributions alone on the ZZ and RR cross-correlations. We first measure the phase velocity dispersion curves for ZZ, RR and preferred Rayleigh

components based on FTAN analysis (*Liu et al.*, 2015). Taking the phase velocity dispersion results from the preferred Rayleigh component as the reference, we compute relative phase velocity bias $(v - v_{ref})/v_{ref}$ on the ZZ and RR components, where v is the biased velocity and v_{ref} is the reference velocity. We divide all 136 station pairs into 3 distance bins (0-44 km, 44-87 km, and 87-131 km) and 18 azimuth bins (bin size 20º). For the ZZ component velocity bias around the primary microseism (Fig. 5a), we compute the median and the median absolute deviation (MAD) values of relative phase velocity bias for each bin. A -5% negative velocity bias occurs at 280º-300º, and a 5% positive velocity bias at 320º-340º for the distance bin 0-44 km. These two cases coincide with the inter-station path directions nearly perpendicular to the spurious non-diffuse noise source direction (Fig. 5a). For the RR component in the shortest distance bin, 0-44 km, the inter-station paths from 180º-220º show the most positive velocity bias of up to 8%, while the negative bias is much smaller (~1%) for station pairs in the same azimuth (Fig. 5b). This one-sided velocity bias is probably due to the narrower azimuthal coverage of strong noise sources in the RR component compared with the preferred Rayleigh component. Overall, the magnitude of bias decreases as the inter-station distance increases as can be observed from the results of different distance bins.

We analyze the relative phase velocity bias for the secondary microseism data (0.12-0.18 Hz) based on the same procedures as for the primary microseism data. For the shortest distance bin 0-44 km, the ZZ component shows 2% median phase velocity bias between 260º-280º with 1.3% MAD of phase velocity bias and -2% median phase velocity bias from 280º-300º (Fig. 5c), which are probably related to the strong source power in the 180º direction on the ZZ component (Fig. 4b). At the same distance range, the RR component results in median phase velocity bias are again one-sided with 2.8% positive phase velocity bias at 300º-320º (Fig. 5d), but the MAD of phase velocity bias is around 5%, which is much greater than the median value of the bias and indicates that there are strong fluctuations in the phase velocity measurements from the RR component due to spatio-temporal complexity of non-diffuse noise sources. For the secondary microseism, the magnitude of bias also decreases as the inter-station distance increases.

4. Discussion

We introduced a new method for separating the major spurious non-diffuse source contribution from cross-correlations of two stations that have 3-component sensors, and by doing so we reduce bias in Rayleigh wave phase velocity measurements from the ZZ and RR components. Our method assumes that the target wave component (e.g. Rayleigh wave) due to noise sources in the stationary phase zone is independent of the spurious non-diffuse component due to an isolated point source. The major non-diffuse source should be outside the stationary phase zone in order to be independent. The key to the successful separation is that these independent components be linearly mixed with different weighting coefficients on multiple recordings (receivers) without time shift. Here we choose the observations of the Rayleigh wave from different cross-correlation components due to a lack of redundant sensors at the same location. We assume that the above assumption regarding the linear combination of independent preferred Rayleigh and spurious non-diffuse components on ZZ and RR cross-correlations holds at least for a limited frequency bandwidth.

The numerical simulations show that the spurious non-diffuse Rayleigh or Love component has different proportions on ZZ and RR cross-correlations due to the projection of its polarization onto the receiver coordinates, and it demonstrates that the spurious non-diffuse cross-correlation component can be separated because the non-diffuse source is outside the stationary phase zone. For real data, there can be non-diffuse sources located within the stationary phase zone of some station pairs and those non-diffuse contributions cannot be separated due to constructive interference in that zone. As a result, we label the separated component containing interferometric Green's function as "preferred Rayleigh", instead of "the diffuse component".

The spurious non-diffuse sources outside the stationary phase zone bias the phase velocity and amplitude measurements because they generate spurious arrivals that sample different paths than the main arrivals produced by noise sources within the stationary phase zone. The non-diffuse sources within the stationary phase zone, however, sample similar ray paths to the diffuse waves traveling within the stationary phase zone and they all have coherent phases in the cross-correlation function. The synthetic test results

(Supplementary Text S2) suggest that the in-phase non-diffuse component cannot be separated based on the method introduced here, and the non-diffuse component causes ~1.5% velocity bias due to slightly different phase for the non-diffuse wave. Due to the similar ray paths for non-diffuse and diffuse waves within the stationary phase zone, this bias in phase velocity can be accounted for using spatial derivative in Eikonal equation (e.g. *Lin et al.*, 2009) or inverting for the source distribution (*Ermert et al.*, 2016). Similar bias in amplitude measurements can be addressed by taking amplitude ratios of linear triplet of stations (*Liu et al*., 2015).

After removing the spurious non-diffuse component for the noise cross-correlations in southern California, the beamforming results of the preferred Rayleigh component show less variation of source power with azimuth and therefore better source illumination. The beamforming results of the spurious non-diffuse noise, however, show more isolated point sources with significant azimuthal variations in source power, which will bias phase velocity measurements. As is shown in Fig. 5, the bias due to the spurious non-diffuse noise is complex, depending on the spatial distribution of those spurious non-diffuse sources in a particular time period, and the analytical correction term based on the 2nd derivative of the source power (*Weaver et al*., 2009) does not hold for those singular sources that are not continuous in all directions. We suggest our approach might be used first to separate the spurious non-diffuse component, then to measure phase velocity from the preferred Rayleigh component, and finally to estimate the bias due to a non-isotropic source distribution (*Weaver et al.*, 2009).

A complicating factor is that the degree of Rayleigh wave ellipticity – the amplitude on the ZZ vs. RR components, can vary with frequency. Choosing a narrow frequency band and assuming constant ellipticity within this band before applying the ICA algorithm could help to mitigate this effect. Another limitation of this algorithm based on ZZ and RR components is the limited number of independent observations. The ICA assumes that the number of independent components is less than or equal to the number of input observations (in Eq. S1). Therefore, we can only compute for one spurious non-diffuse component in addition to the preferred Rayleigh component. This limits our ability to separate less significant spurious sources that have different ratios on the ZZ

and RR components than the most significant spurious non-diffuse source. Combining ZR-RZ components by equalizing the phase with that of the ZZ/RR component (*van Wijk et al*., 2011), we can get one additional input observation. Redundant stations at the same location (e.g. the shallow borehole stations in Supplemental Text S4) could help assess and resolve both of these issues.

The temporal correlations (*Liu et al*., 2016) in non-diffuse noise can lead to weak spurious arrivals after the expected Rayleigh wave arrival time (Figs. 2d & 3d), causing problems for applications using the coda of noise cross-correlation as sources of diffuse waves (*Brenguier et al*., 2008; *Stehly et al*., 2008). This approach for non-diffuse component separation could potentially mitigate similar biases in the coda of cross-correlation (*Sheng et al*., 2018).

5. Acknowledgement

We appreciate the constructive comments from two reviewers Toshiro Tanimoto and Victor Tsai, and the associate editor that help improve the paper. This work is supported by Southern California Earthquake Center (SCEC) contribution 17227. The continuous waveform data are obtained from the Southern California Earthquake Data Center (SCEDC, http://scedc.caltech.edu).

Reference

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

Brenguier, F., Campillo, M., Hadziioannou, C., Shapiro, N. M., Nadeau, R. M., & Larose, E. (2008). Postseismic Relaxation Along the San Andreas Fault at Parkfield from Continuous Seismological Observations. *Science*, *321*(5895), 1478–1481. https://doi.org/10.1126/science.1160943

Cessaro, R. K. (1994). Sources of primary and secondary microseisms. *Bulletin of the Seismological Society of America*, *84*(1), 142–148.

Curtis, A., Nicolson, H., Halliday, D., Trampert, J., & Baptie, B. (2009). Virtual seismometers in the subsurface of the Earth from seismic interferometry. *Nature Geoscience*, *2*(10), 700–704. https://doi.org/10.1038/ngeo615

Ermert, L., Villaseñor, A., & Fichtner, A. (2016). Cross-correlation imaging of ambient noise sources. *Geophysical Journal International*, *204*(1), 347–364. https://doi.org/10.1093/gji/ggv460

Hasselmann, K. (1963). A statistical analysis of the generation of microseisms. *Reviews of Geophysics*, *1*(2), 177. https://doi.org/10.1029/RG001i002p00177

Hyvärinen, A., & Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural Networks*, *13*(4–5), 411–430. https://doi.org/10.1016/S0893-6080(00)00026-5

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*, *177*(3), 1091–1110.

Lin, F.-C., Tsai, V. C., & Ritzwoller, M. H. (2012). The local amplification of surface waves: A new observable to constrain elastic velocities, density, and anelastic attenuation. *Journal of Geophysical Research*, *117*(B6), B06302. https://doi.org/10.1029/2012JB009208

Liu, X., & Ben-Zion, Y. (2016). Estimating correlations of neighbouring frequencies in ambient seismic noise. *Geophysical Journal International*, *206*(2), 1065–1075. https://doi.org/10.1093/gji/ggw196

Liu, X., & Ben-Zion, Y. (2018). Analysis of non-diffuse characteristics of the seismic noise field in southern California based on correlations of neighbouring frequencies. *Geophysical Journal International*, *212*(2), 798–806. https://doi.org/10.1093/gji/ggx441

Liu, X., Ben-Zion, Y., & Zigone, D. (2015). Extracting seismic attenuation coefficients from cross-correlations of ambient noise at linear triplets of stations. *Geophysical Journal International*, *203*(2), 1149–1163. https://doi.org/10.1093/gji/ggv357

Liu, X., Ben-Zion, Y., & Zigone, D. (2016). Frequency domain analysis of errors in cross-correlations of ambient seismic noise. *Geophysical Journal International*, *207*(3), 1630–1652. https://doi.org/10.1093/gji/ggw361

Lobkis, O. I., & Weaver, R. L. (2001). On the emergence of the Green's function in the correlations of a diffuse field. *The Journal of the Acoustical Society of America*, *110*, 3011.

Longuet-Higgins, M. S. (1950). A Theory of the Origin of Microseisms. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *243*(857), 1–35. https://doi.org/10.1098/rsta.1950.0012

Nakata, N., Chang, J. P., Lawrence, J. F., & Boué, P. (2015). Body wave extraction and tomography at Long Beach, California, with ambient-noise interferometry. *Journal of Geophysical Research: Solid Earth*, *120*(2), 1159–1173. https://doi.org/10.1002/2015JB011870

Porter, M. B., & Tolstoy, A. (1994). THE MATCHED FIELD PROCESSING BENCHMARK PROBLEMS. *Journal of Computational Acoustics*, *02*(03), 161–185. https://doi.org/10.1142/S0218396X94000129

Prieto, G. A., Lawrence, J. F., & Beroza, G. C. (2009). Anelastic Earth structure from the coherency of the ambient seismic field. *Journal of Geophysical Research*, *114*(B7), B07303.

Shapiro, N. M. (2005). High-Resolution Surface-Wave Tomography from Ambient Seismic Noise. *Science*, *307*(5715), 1615–1618. https://doi.org/10.1126/science.1108339

Shapiro, N. M., & Campillo, M. (2004). Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise. *Geophysical Research Letters*, *31*(7), L07614. https://doi.org/10.1029/2004GL019491

Shaw, J. H., Plesch, A., Tape, C., Suess, M. P., Jordan, T. H., Ely, G., et al. (2015). Unified Structural Representation of the southern California crust and upper mantle. *Earth and Planetary Science Letters*, *415*, 1–15. https://doi.org/10.1016/j.epsl.2015.01.016

Sheng, Y., Nakata, N., & Beroza, G. C. (2018). On the Nature of Higher-Order Ambient Seismic Field Correlations. *Journal of Geophysical Research: Solid Earth*, *123*(9), 7969–7982. 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*, *69*(4), 046610. https://doi.org/10.1103/PhysRevE.69.046610

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*, *113*(B11). https://doi.org/10.1029/2008JB005693

Stutzmann, E., Schimmel, M., Patau, G., & Maggi, A. (2009). Global climate imprint on seismic noise: GLOBAL CLIMATE IMPRINT ON SEISMIC NOISE. *Geochemistry, Geophysics, Geosystems*, *10*(11), n/a-n/a. https://doi.org/10.1029/2009GC002619

Tsai, V. C. (2011). Understanding the amplitudes of noise correlation measurements. *Journal of Geophysical Research: Solid Earth*, *116*(B9), B09311. https://doi.org/10.1029/2011JB008483

Weaver, R., Froment, B., & Campillo, M. (2009). On the correlation of non-isotropically distributed ballistic scalar diffuse waves. *The Journal of the Acoustical Society of America*, *126*(4), 1817. https://doi.org/10.1121/1.3203359

Weaver, R. L. (1982). On diffuse waves in solid media. *The Journal of the Acoustical Society of America*, *71*(6), 1608. https://doi.org/10.1121/1.387816

Weaver, R. L., & Lobkis, O. I. (2004). Diffuse fields in open systems and the emergence of the Green's function (L). *The Journal of the Acoustical Society of America*, *116*(5), 2731. https://doi.org/10.1121/1.1810232

van Wijk, K., Mikesell, T. D., Schulte-Pelkum, V., & Stachnik, J. (2011). Estimating the Rayleigh-wave impulse response between seismic stations with the cross terms of the Green tensor: THE RAYLEIGH-WAVE GREEN TENSO. *Geophysical Research Letters*, *38*(16), n/a-n/a. https://doi.org/10.1029/2011GL047442

Figure captions

Figure 1. (a) Expected spectral amplitude ratios of non-diffuse over diffuse noise (upper). Correlation coefficient matrix of neighboring frequency amplitudes (lower), in which the neighboring frequency correlation is related to the amplitude spectra of the non-diffuse source. (b) Simulation geometry. Blue dots: delta-correlated sources for diffuse noise; red dot: non-diffuse noise source with correlated neighboring frequency statistics in panel a. T and R are transverse and radial receiver axes, respectively. (c) The ZZ and RR cross-correlation containing the mixture of diffuse and non-diffuse Rayleigh waves from simulated noise data. (d) The separated preferred Rayleigh and spurious non-diffuse Rayleigh wave components compared with the ground truth. (e) The ZZ and RR cross-correlation containing the mixture of diffuse Rayleigh and non-diffuse Love waves from simulated noise data. (f)) The separated preferred Rayleigh and spurious non-diffuse Love wave components compared with the ground truth.

Figure 2. (a) Model of two linearly independent components in noise interferometry: the Green's function produced by interference of multiply scattered waves (scatterers in the ring), and the spurious non-diffuse component produced by isolated physical sources that are non-stationary (stars). The dark gray shaded area is the stationary phase zone for direct waves. (b) Map of 17 broadband stations in southern California used in this study. (c) The ZZ and RR components of cross-correlations used as input for ICA source separation. (d) The preferred Rayleigh wave (top) and spurious non-diffuse components (bottom) as output from ICA algorithm. (e) FTAN analysis for the ZZ component cross-correlation of stations CHF-SBB2. (f) FTAN analysis for the preferred Rayleigh component cross-correlation of stations CHF-SBB2.

Figure 3. Different components of noise cross-correlations among the 17 stations. Data filtered between 0.05-0.25 Hz. The amplitude of each cross-correlation is normalized between -1 and 1. The different panels are: (a) ZZ, (b) RR, (c) preferred Rayleigh and (d) spurious non-diffuse components. The two dashed lines in panel a correspond to move-out velocity of 4.7 km/s. (e) RMS for the spurious early arrival (acausal) waveform between two dashed lines are estimated and normalized by the maximum absolute amplitude in each cross-correlation.

Figure 4. (a) Beamforming of ZZ, RR, preferred Rayleigh and spurious non-diffuse components for the primary ocean microseism (0.05-0.10 Hz). The slowness is 0.5 s/km at the circumference and 0 at the center. (b) Beamforming for the secondary ocean microseism filtered between 0.12-0.18 Hz.

Figure 5. (a) & (b) Relative phase velocity bias on ZZ and RR components, respectively, compared with the phase velocity on the preferred Rayleigh component for three distance bins: 0-44 km, 44-87 km and 87-131 km. The * represents median dv/v and the error bar equals plus/minus one MAD. The frequency band is 0.05-0.10 Hz for the primary ocean microseism. (c) & (d) Similar to panels a & b, but the frequency band is 0.12-0.18 Hz for the secondary ocean microseism.

Figure 1.

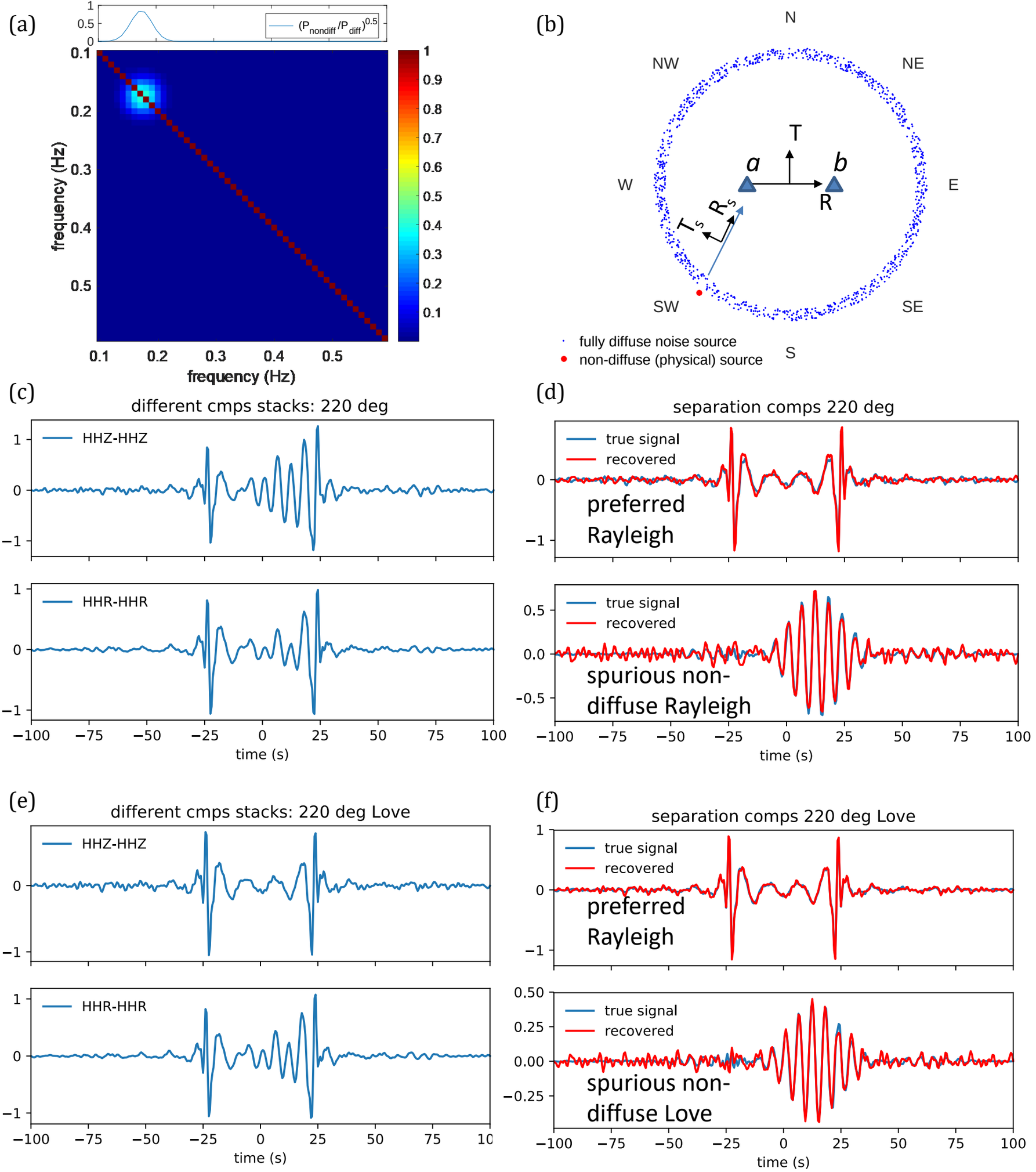

(a)
(P_nondiff /P_diff)^0.5
frequency (Hz)
frequency (Hz)
(b)
N
NW
NE
W
E
SW
SE
S
T
a
b
R
T_s
R_s
fully diffuse noise source
non-diffuse (physical) source
(c)
different cmps stacks: 220 deg
HHZ-HHZ
HHR-HHR
time (s)
(d)
separation comps 220 deg
true signal
recovered
preferred
Rayleigh
spurious non-
diffuse Rayleigh
time (s)
(e)
different cmps stacks: 220 deg Love
HHZ-HHZ
HHR-HHR
time (s)
(f)
separation comps 220 deg Love
true signal
recovered
preferred
Rayleigh
spurious non-
diffuse Love
time (s)

Figure 2.

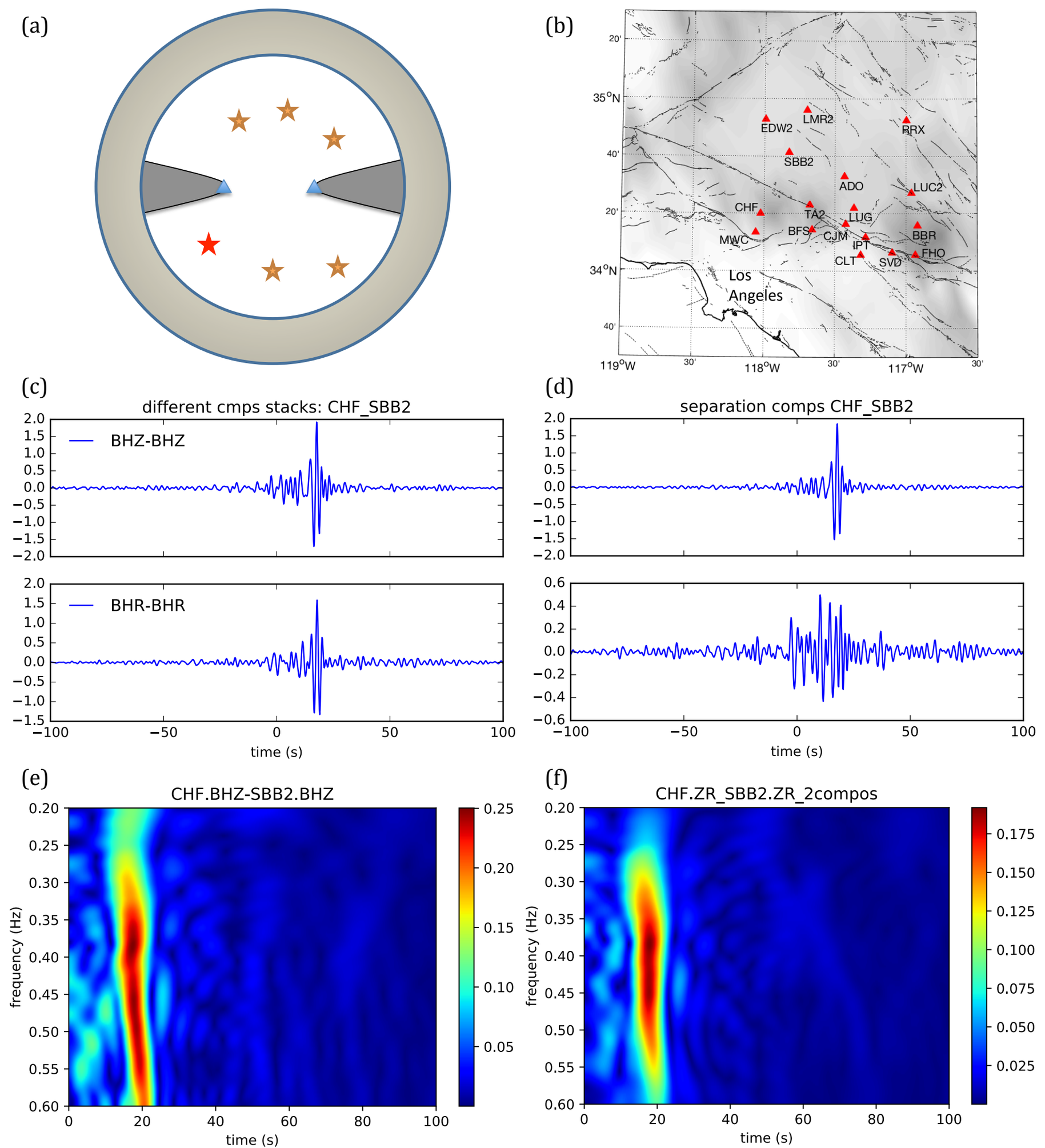

(a)
(b)
35°N
34°N
119°W
118°W
117°W
EDW2
LMR2
RRX
SBB2
ADO
LUC2
CHF
TA2
LUG
BFS
CJM
BBR
MWC
IPT
FHO
CLT
SVD
Los
Angeles
(c)
different cmps stacks: CHF_SBB2
BHZ-BHZ
BHR-BHR
time (s)
(d)
separation comps CHF_SBB2
time (s)
(e)
CHF.BHZ-SBB2.BHZ
frequency (Hz)
time (s)
(f)
CHF.ZR_SBB2.ZR_2compos
frequency (Hz)
time (s)

Figure 3.

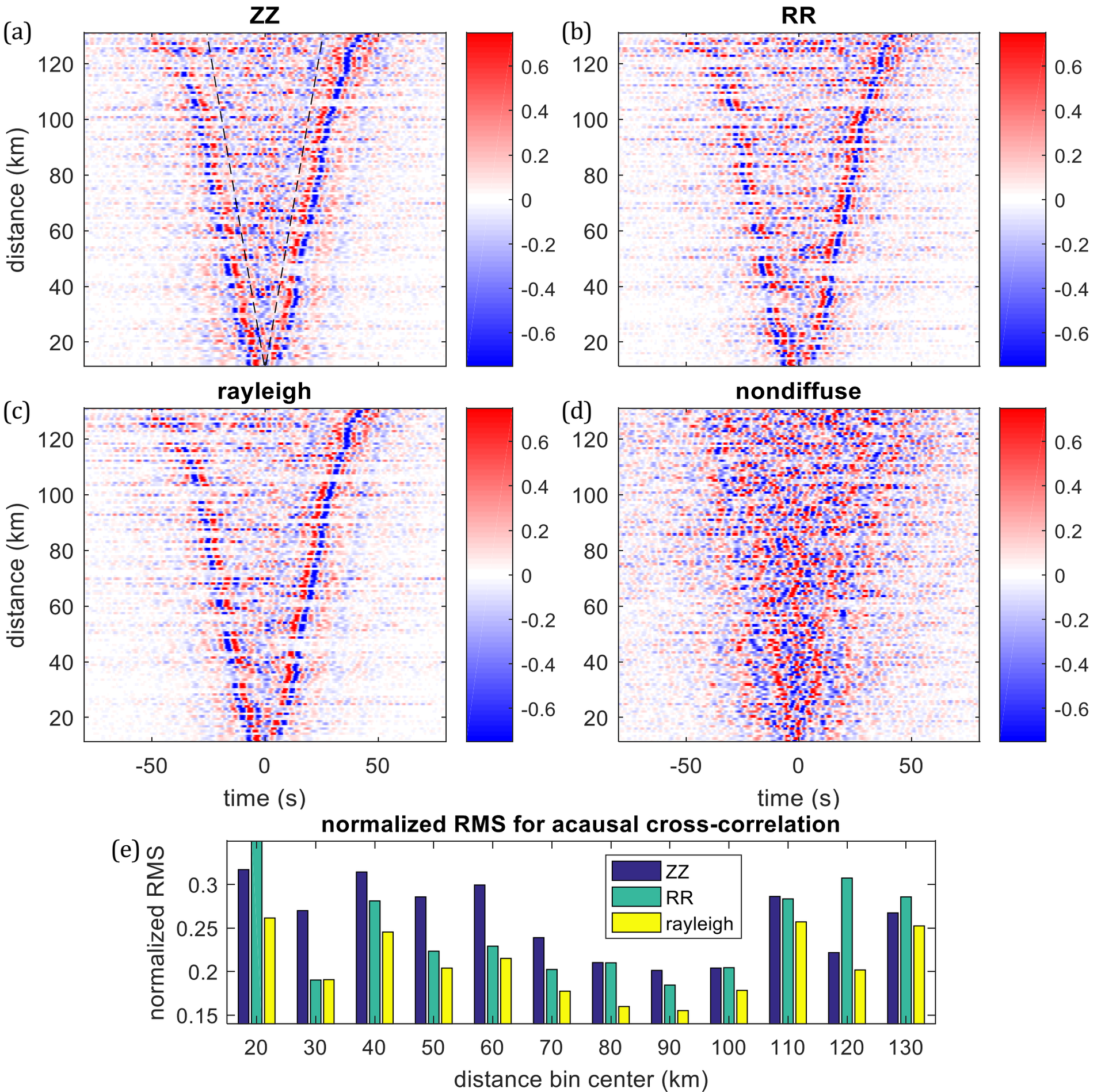

(a)
ZZ
(b)
RR
(c)
rayleigh
(d)
nondiffuse
distance (km)
time (s)
(e)
normalized RMS for acausal cross-correlation
normalized RMS
distance bin center (km)
ZZ
RR
rayleigh

Figure 4.

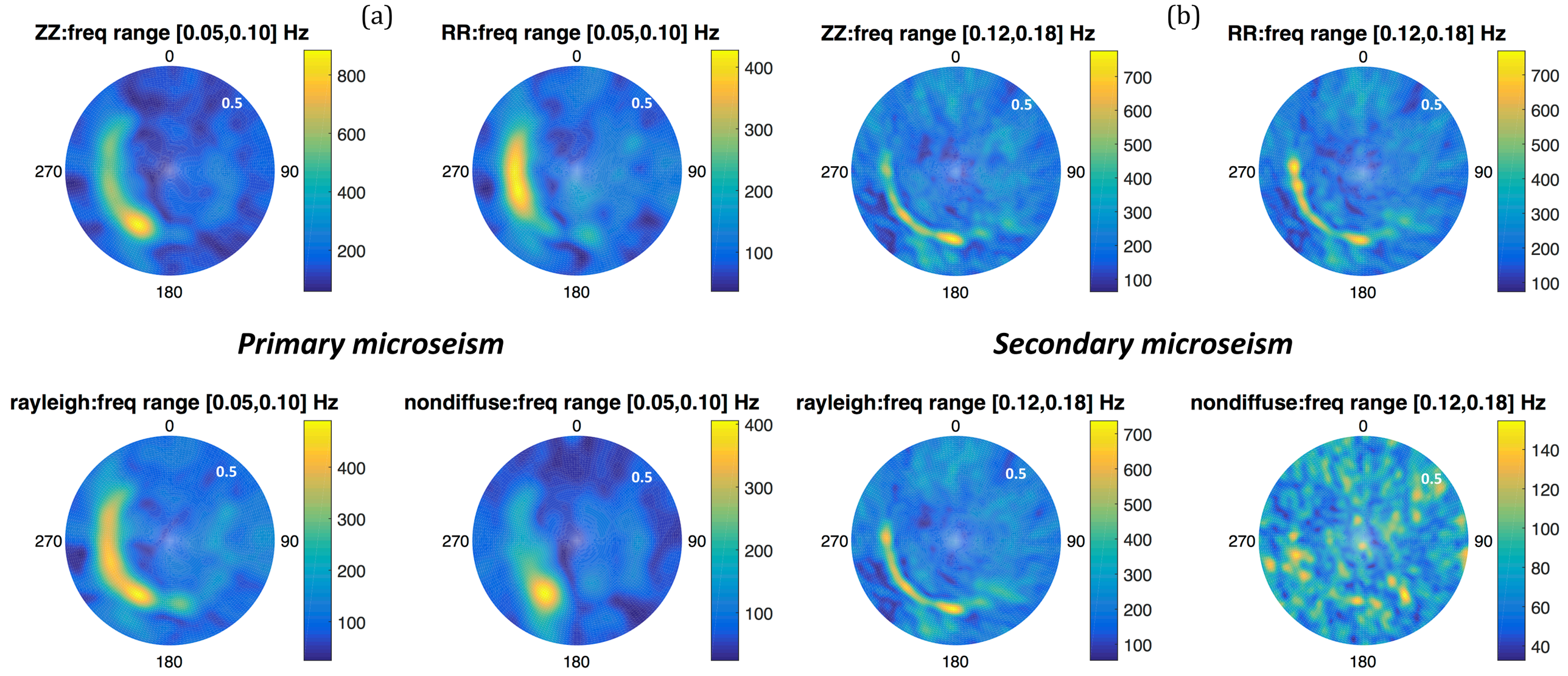
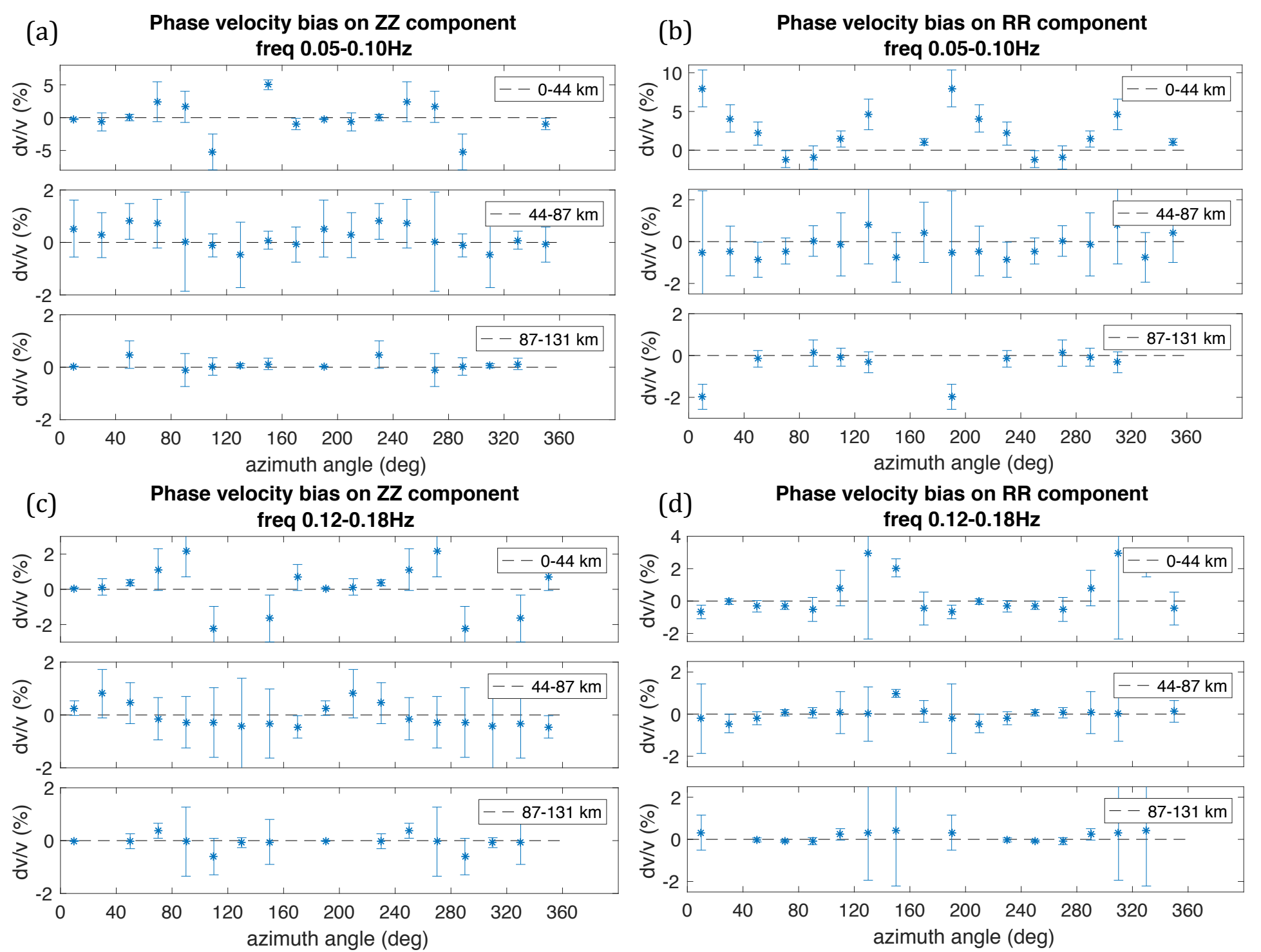

(a)
ZZ:freq range [0.05,0.10] Hz
RR:freq range [0.05,0.10] Hz
(b)
ZZ:freq range [0.12,0.18] Hz
RR:freq range [0.12,0.18] Hz
Primary microseism
Secondary microseism
rayleigh:freq range [0.05,0.10] Hz
nondiffuse:freq range [0.05,0.10] Hz
rayleigh:freq range [0.12,0.18] Hz
nondiffuse:freq range [0.12,0.18] Hz
0
90
180
270
0.5

Figure 5.


(a)
Phase velocity bias on ZZ component
freq 0.05-0.10Hz
(b)
Phase velocity bias on RR component
freq 0.05-0.10Hz
(c)
Phase velocity bias on ZZ component
freq 0.12-0.18Hz
(d)
Phase velocity bias on RR component
freq 0.12-0.18Hz
dv/v (%)
0-44 km
44-87 km
87-131 km
azimuth angle (deg)