Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements

G.D. Bensen[1†], M.P. Barmin[1], A.L. Levshin[1], F. Lin[1], M.P. Moschetti[1], N.M. Shapiro[2], Y. Yang[1], and M.H. Ritzwoller[1]

1 - Center for Imaging the Earth's Interior, Department of Physics, University of Colorado at Boulder, Campus Box 390, Boulder, CO 80309, USA

2 - Laboratoire de Sismologie, CNRS, IPGP, 4 place Jussieu, 75005 Paris, France.

† To whom correspondence should be directed: gbensen@ciei.colorado.edu

In preparation for submission to *Bull. Seism. Soc. Amer.*: July, 2006.

Abstract

1. Introduction

Theoretical studies have shown that the time-derivative of the cross-correlation of diffuse wavefields (e.g., ambient noise, scattered coda waves) can provide an estimate of the Green function between the stations (e.g., Weaver and Lobkis, 2001a, 2001b, 2004; Derode et al., 2003; Snieder, 2004; Wapenaar, 2004; Larose et al., 2005). Seismic observations based on cross-correlations between pairs of stations have confirmed the theory for surface waves using both coda waves (Campillo and Paul, 2003; Paul et al., 2005) and long ambient noise sequences (Shapiro and Campillo, 2004; Sabra et al., 2005a) and for crustal body waves using ambient noise (Roux et al., 2005). Oceanic applications are also feasible (Lin et al., 2006a). An example of a year-long cross-correlation between a station-pair in the Pacific filtered into several sub-bands is shown in Figure 1.

The first attempts to use ambient noise for surface wave tomography, called ambient noise surface wave tomography, were applied to stations in Southern California (Shapiro et al., 2005; Sabra et al., 2005b). These studies resulted in group speed maps at short periods (7.5 - 15 sec) that displayed striking correlations with the principal geological units in California with low-speed anomalies corresponding to the major sedimentary basins and high-speed anomalies corresponding to the igneous cores of the main mountain ranges.

Ambient noise tomography is now expanding rapidly. Recent applications have arisen across all of California and the Pacific Northwest (Moschetti et al., 2005), in South Korea (Cho et al., 2006), in Tibet (Yao et al., 2006), in Europe (Yang et al., 2006), across New Zealand (Lin et al., 2006b), as well as elsewhere in the world. Most of the studies, to date, like the earlier work of Shapiro et al. (2005), have been performed in the microseism band below 20 sec period. Broad-band applications extending to considerably longer periods are now emerging (e.g., Bensen et al., 2005; Yao et al., 2006; Yang et al., 2006) and the method is also being applied to increasingly large areas such as Europe (Yang et al., 2006). In spite of these developments, the data processing procedures that underlie ambient noise tomography remain poorly documented. The purpose of this paper is to summarize the state of data processing as it has developed since the first papers on the use of ambient noise to obtain surface wave dispersion measurements (Shapiro and Campillo, 2004).

In its current state, ambient noise data processing procedure divides into four principal phases that are applied roughly in order: (1) single station data preparation, (2) cross-correlation and temporal stacking, (3) measurement of dispersion curves, and (4) error analysis and selection of the acceptable measurements. These steps are presented schematically in Figure 2. After data processing is complete, tomography for group or phase speed maps (e.g., Yang et al., 2006) and inversion for a V_s model (e.g., Cho et al., 2006; Lin et al., 2006b) may follow, but discussion of these steps is beyond the scope of the present paper.

In judging between candidate components of the data processing procedure, we have assigned significant weight to flexibility and the applicability to a wide variety of observational situations. The procedures described here, therefore, are designed to be applied over a broad range of periods, inter-station distances, and geographical scales. Examples are shown in this paper from regional to continental scales, from very short to long periods, and are drawn from the Pacific, Europe, North America, and New Zealand. Applications are, however, taken exclusively from continental or ocean island stations. As discussed by Lin et al. (2006a), broad-band cross-correlations of ambient noise obtained at ocean bottom or sub-bottom seismometers (OBS) are contaminated at long periods (above ∼25 sec) by tilting under fluid flow and seafloor deformation under gravity waves. Crawford et al. (2006) argues that these effects can be reduced on the vertical component using horizontal component data and a co-located differential seafloor pressure gauge. The success of this process will be needed for broad-band ambient noise measurements to be obtained from OBS data. We are unaware of research that has tested this idea in the context of ambient noise measurements, however.

Our principal purpose, therefore, is to summarize the status of the ambient noise data processing procedure that we have developed over the past several years. The paper is intended to explain, justify, and present salient examples of this development. It is also intended to act as a primer to help provide guidance and act as a basis for future efforts in surface wave studies based on ambient seismic noise. Each of the four following sections presents a discussion of one phase of the data processing procedure, which ranges from processing data from a single station (section 2), cross-correlating and stacking data from station-pairs (section 3), measuring surface wave dispersion (section 4), and data quality control, particularly estimating uncertainties and selecting reliable measurements (section 5).

2. Single station data preparation

The first phase of data processing consists of preparing waveform data from each station individually. The purpose of this phase is to accentuate broad-band ambient noise by attempting to remove earthquake signals and instrumental irregularities that tend to obscure ambient noise. Obscuration by earthquakes is most severe above about 20 sec period, so this step of the data processing is most important at periods longer than the microseism band (~5 to ~17 sec period). In addition, because the spectral amplitude of ambient noise peaks in the microseismic band, methods have to be devised to extract the longer period ambient noise from seismic records. Figure 2 shows the steps that compose Phase 1 of data processing: removal of the instrument response, de-meaning, de-trending, and band-pass filtering the seismogram, time-domain normalization, and spectral whitening. This procedure is typically applied to a single day of data. Day data with less than 80% "on-time" are currently rejected, but this may be modified at the user's discretion. Some of the steps, such as the temporal normalization and spectral whitening, impose non-linear modifications to the waveforms, so the order of operations is significant. Because this phase of data processing is applied to single stations, rather than to station-pairs, it is much less time consuming and computationally intensive than subsequent cross-correlation, stacking, and measurement phases that are discussed in later sections of the paper. Our current applications involve from several dozen (e.g., 41 stations across New Zealand) to several hundred (e.g., 110 stations across Europe, ~250 stations across North America) stations.

2.1 Temporal normalization

The most important step in single-station data preparation is what we call "time-domain" or "temporal normalization". Time-domain normalization is a procedure for reducing the effect on the cross-correlations of earthquakes and instrumental irregularities. Earthquakes are among the most significant impediments to automated data processing. They occur irregularly and, although the approximate times and locations of large earthquakes can be found in earthquake catalogs, small earthquakes over much of the globe are missing from global catalogs. In addition, the time of arrival of surface wave phases at short periods is not well known. Thus, removal of earthquake signals must be data-adaptive, rather than prescribed from a catalog.

We have considered five different methods to identify and remove earthquakes and other contaminants automatically from seismic waveform data. An illustrative example is shown in Figure 3. The first and most aggressive method is called "one-bit" normalization (Figure 3b), which retains only the sign of the raw signal by replacing all positive amplitudes with a 1 and all negative amplitudes with a -1. This method has been shown to increase signal-to-noise ratio (SNR) when employed in acoustic experiments in the laboratory (Larose, et al. 2004) and has been used in a number of early seismic studies of coda waves and ambient noise (Campillo and Paul, 2003, Shapiro and Campillo 2004, Shapiro et al., 2005; Yao et al. 2006). The second method, employed for example by Sabra et al. (2005a), involves the application of a clipping threshold equal to the root-mean-square (rms) amplitude of the signal for the given day. An example is shown in Figure 3c. The third method, illustrated by Figure 3d, involves automated event detection and removal in which 30 minutes of the waveform are set to zero if the amplitude of the waveform is above a critical threshold. This threshold is arbitrary and its choice is made difficult varying amplitudes at different stations. Fourth, there is running-absolute-mean normalization, which is the method of time normalization that we promote here. This method computes the running average of the absolute value of the waveform in a normalization window of fixed length and weights the waveform at the center of the window by the inverse of this average. That is, given a discrete time-series d_j, we compute the normalization weight for time point n as:

$$w_n = \frac{1}{2N+1} \sum_{j=n-N}^{n+N} |dj| \tag{1}$$

so that the normalized datum becomes $\tilde{d}_n = d_n / w_n$. The width of the normalization window $(2N+1)$ determines how much amplitude information is retained. A one-sample window

($N = 0$) is equivalent to one-bit normalization, while a very long window will approach a re-scaled original signal as $N \to \infty$. After testing various time window widths, we find that about half the maximum period of the pass-band filter works well and that this length can be varied considerably and still produce similar results. An example result of the application of this method is shown in Figure 3e. This method is not without its faults, however. For example, it does not surgically remove narrow data glitches, as it will inevitably down-weight a broad time-interval around the glitch. One-bit normalization does not suffers from this shortcoming. Finally, there is a method that we call iterative "water-level" normalization in which any amplitude above a specified multiple of the daily rms-amplitude is down-weighted. The method is run repeatedly until the entire waveform is below the water-level, which is six times the daily rms level in the example shown here. An example of the application of this method is shown in Figure 3f. This method of time-domain normalization is the most time-intensive of the candidates considered here.

Figure 4 presents examples of year-long cross-correlations, band-pass filtered between 20 sec and 100 sec period, using each of these methods of time-domain normalization. The raw data (Fig. 4a), the clipped waveform method (Fig. 4c), and the automated even detection method (Fig. 4d) produce noisy cross-correlations in this period band. The one-bit normalization (Fig. 4b), the running-absolute-mean normalization (Fig. 4e), and the water-level normalization (Fig. 4f) methods produce very similar results with relatively high signal-to-noise ratio (SNR) waveforms displaying signals that arrive at nearly the same time. In this example, the one-bit and the running-absolute-mean normalizations are nearly identical. Results of a more systematic test performed using 15 GSN stations in North America are summarized in Table 1. We have used temporal SNR, defined in Figure 1, to compare the methods in five frequency bands. The resulting SNR values are similar for one-bit normalization, the running-absolute-mean normalization, and the water-level normalization methods. The running mean normalization provides a small enhancement to SNR values roughly 5% above one-bit normalization at all periods.

The principal reason we prefer running-absolute-mean normalization over the water-level or one-bit normalization methods is its greater flexibility and adaptability to the data.

For example, in areas with high regional seismicity it is desirable to tune the time-domain normalization to the frequency content of the seismicity. Figure 5 shows that if the temporal weights of the running-absolute-mean normalization are computed on the raw waveform data, small earthquakes can get through the procedure because they exist in the raw waveform near the background noise level. Earthquakes are revealed by a low-pass filter both in the raw waveform (Fig. 5b) and the temporally normalized waveform (Fig. 5d). Alternately, the temporal weights of the running-absolute-mean normalization can be computed on the waveform filtered in the earthquake band (Fig. 5b). In this case, if d_j is the raw seismogram and $\hat{d}_j$ is the seismogram band-pass filtered in the earthquake band, we define new temporal weights calibrated to the regional seismicity

$$\hat{w}_n = \frac{1}{2N+1} \sum_{j=n-N}^{n+N} |\hat{d}j|. \tag{2}$$

These weights are then applied to the raw data as before ($\tilde{d}_n = d_n/\hat{w}_n$). This procedure severaly down-weights time-series during earthquakes (Fig. 5e), which more effectively removes them from low-pass filtered seismograms (Fig. 5f). Contamination of earthquakes of the cross-correlations, therefore, should be ameliorated.

Earthquake signals that pass through the temporal normalization tend to appear on cross-correlations as spurious precursory arrivals, such as the high amplitude arrivals between 0 - 100 sec in the 12-month cross-correlation shown in Figure 6a. Defining the temporal normalization weights in the earthquake band, however, reduces the amplitude of the precursors, as Figure 6b illustrates. This process will be most important in regions with significant regional seismicity. The example shown in Figure 6 is from New Zealand where, because of high levels of seismicity in the Fiji and Tonga-Kermadec regions, the process is recommended strongly (Lin et al., 2006b).

2.2 Spectral normalization or whitening

The ambient noise is not flat in the frequency domain (i.e., is not spectrally white), but is peaked near the primary (around 15 sec) and secondary (around 7.5 sec) microseisms and rises at very long periods above 50 sec to form a signal now referred to as Earth "hum" (e.g., Rhie and Romanowicz, 2004). Figure 7a presents an example of an amplitude spectrum for a day long time series obtained after temporal normalization. Primary and secondary microseisms as well as Earth hum signatures can be seen clearly on this record that was band-pass filtered between 7 sec and 150 sec period. In addition to these signals, there is a smaller peak near 26 sec that is caused by a persistent narrow-band noise source in the Gulf of Guinea (Shapiro et al., 2006). Without the temporal normalization, which reduces the effect of earthquakes, the 26 sec resonance typically is not seen. Inversely weighting the complex spectrum by a smoothed version of the amplitude spectrum produces the normalized or whitened spectrum shown in Figure 6b. Spectral normalization acts to broaden the band of the ambient noise signal in cross-correlations and also combats degradation caused by persistent monochromatic sources such as the Gulf of Guinea source.

First, regarding the problem of an isolated, persistent monochromatic noise source, the grey box in Figure 7a highlights the noise peak at 26 sec period as observed at the station HRV on a northern summer day. As documented by Holcomb (1998), this signal is seasonal, being much stronger in the northern summer than in the winter. Figure 8a shows a 12-month cross-correlation between stations ANMO and CCM in which spectral normalization has not been applied. The 26 sec resonance appears with a broad envelope in the time domain and corrupts the cross-correlation at positive correlation lag. Shapiro et al. (2006) used the apparent arrival time of the 26 sec signal to locate the source in the Gulf of Guinea. The amplitude spectrum of this cross-correlation displays the prominent peak at $\sim$26 sec period ($\sim$0.038 Hz) as seen in Figure 8b. In contrast, Figures 8c and 8d show the cross-correlation and its amplitude spectrum where spectral normalization has been applied. The affect of the 26 sec resonance is greatly reduced. Shapiro et at. (2006) recommend eradicating this problem by applying a narrow band reject filter centered around 26 sec period. Figures 8e and 8f show the effect of this filter. The cross-correlation is largely unchanged. In many cases, therefore, the more gentle approach of spectral whitening is sufficient to eliminate the 26 sec problem from the cross- correlations. The band-reject filter also crates problems for automated dispersion measurement in a later stage of processing, so spectral whitening is preferable if it suffices to ameliorate the effect of the 26 sec microseism.

Second, spectral normalization seeks to reduce broad imbalances in single-station spectra to aid in the production of a broad-band dispersion measurement. Figures 9a and 9b show a one-month broad-band cross-correlation between stations CCM (Cathedral Cave, MO, USA) and SSPA (Standing Stone, PA, USA) for spectrally whitened and un-whitened data taken during the northern spring (when the 26 sec resonance is weak). Figures 9c and 9d display the amplitude spectra of the un-whitened and whitened cross-correlations, respectively. Without the whitening, Figure 9c shows that the resulting cross-correlation is dominated by signals in the microseism band, predominantly from 15 to 17 sec and from 6 to 9 sec period. Not surprisingly, spectral whitening produces a broader-band signal. In many cases, the cross-correlation amplitude spectrum is shaped with the longer periods having higher amplitudes than the shorter periods, as in Figure 9d. This is apparently because the longer period ambient noise, although naturally lower in amplitude than microseismic noise, propagates more coherently over long distances. Additional whitening of the cross-correlation prior to dispersion measurement is an option.

3. Cross-correlation, stacking, and signal emergence

After preparation of the daily time-series described in section 2, the next step in the data processing scheme (Phase 2) is cross-correlation and stacking. Although some inter-station distances may be either too short or too long to obtain reliable measurements, we perform cross-correlations between all possible station pairs and perform data selection later. This yields a total of $n(n-1)/2$ possible station pairs, where n is the number of stations. Obtaining tens of thousands of cross-correlations is common when ambient noise data processing is performed over large spatial scales.

Cross-correlation is performed daily in the frequency domain. After the daily cross-correlations are returned to the time-domain they are added to one another, or “stacked”, to correspond to longer time series. Alternately, stacking can be done in the frequency domain which would save the inverse transform. We prefer the organization that emerges from having daily raw time-series and daily stacks that are then stacked further into weekly, monthly, yearly, etc. time-series. In any event, the linearity of the cross-correlation procedure guarantees that this method will produce the same result as cross-correlation applied to the longer time series. The resulting cross-correlations are two-sided time functions with both positive and negative time coordinates; i.e., both positive and negative correlation lags. We typically store the correlations from -5000 to 5000 sec, but the length of the time series needed will depend on the period band of interest, the group speeds of the waves, and the longest inter-station distance.

The positive lag part of the cross-correlation is sometimes called the "causal" signal and the negative lag part the "acausal" signal. These waveforms represent waves traveling in opposite directions between the stations. Several examples of cross-correlations have been shown earlier in the paper. Figures 4, 8, and 9 display some two-sided cross-correlations for different time-series lengths. If sources of ambient noise are distributed homogeneously in azimuth, the causal and acausal signals would be identical. However, considerable asymmetry in amplitude and spectral content is typically observed, which indicates differences in both the source process and distance to the source in the directions radially away from the stations. We often compress the two-sided signal into a one-sided signal by averaging the causal and acausal parts. We call this the "symmetric" signal or component and examples have been shown in Figures 1 and 6. Another example of a broad-band symmetric component cross-correlation for a year-long time-series can be seen in Figure 10, which clearly shows the broad-band content of ambient noise.

Stacking over increasingly long time-series, on average, improves signal-to-noise ratio. An example is shown in Figure 11, which displays cross-correlations of different length time-series observed at the stations ANMO (Albuquerque, NM, USA) and DWPF (Disney Wilderness Preserved, FL, USA). The causal and acausal signals are seen to emerge as the time-series length increases in both of the period bands that are displayed in Figures 11a and 11b.

Measurements of the frequency dependence of the SNR are useful to quantify observations of the emergence of the signals with increasing time-series length. We also use it as part of data selection in Phase 4 of the data processing procedure. (The spectral SNR is an extension of the temporal SNR mentioned earlier in the paper (Fig. 1).) Figure 12 illustrates one way in which the frequency dependence of SNR may be measured. From the 3-D model of Shapiro and Ritzwoller (2002), we predict the maximum and minimum group arrival times (t_{min}, t_{max}) expected for the path between the station-pair over the period band of interest (τ_{min}, τ_{max}). We perform a series of narrow band-pass filters centered on a discrete grid of frequencies and measure the peak in the time domain in a signal window ($t_{min} - \tau_{max}, t_{max} + 2\tau_{max}$) shown with solid vertical lines in Figure 12. We also measure the root-mean-square (rms) noise level in a 500 sec noise window (vertical dashed lines) that trails the end of the signal window by 500 sec. This rms level is shown with dotted lines in Figure 12 in the noise window. The resulting ratio of peak signal in the signal window to rms noise in the trailing noise window on the grid of center frequencies is the "spectral" SNR measurement. Center frequencies and SNR are identified in each panel of Figure 12. Note that although we call this a spectral SNR measurement, it is, in fact, a measurement of SNR in the time-domain. It is "spectral" only in the sense that the measurements is a function of frequency.

A spectral SNR curve for the 12-month cross-correlation between stations ANMO and DWPF, shown in Figure 11a and 11b, is presented in Figure 11c. It is contrasted with the average SNR over all GSN station pairs within the US. For this example, spectral SNR, on average, peaks in the primary microseism band around 15 sec period, minimizes near 40 sec period, and then is fairly flat to much longer periods, although it rises slightly. The details of the curve, however, will vary geographically, with path length, and season. Figure 11d shows how spectral SNR increases with time-series length. The shapes of the SNR spectra also change subtly with time-series length.

Emergence of the signal..... Figure 13.

4. Dispersion measurement

After the daily cross-correlations have been computed and stacked, the group and phase speeds as a function of period can be measured. This is Phase 3 of the data processing procedure. As with Phases 1 and 2, because the number of inter-station pairs can be very large, the dispersion measurement process needs to be automated. A number of approaches are possible for the group velocity measurements. A new method devised by Guy Masters (REF?) involves only minimal interaction by an analyst. A principal innovation of the method is making measurements on clusters of waveforms that have similar properties, such as having emanated from the same event, propagated along the same path or at similar epicentral distances. This method may work well for the application to ambient noise cross-correlations, and deserves future analysis. Traditional frequency-time analysis (FTAN) (e.g., Levshin et al., 1972; 1992; Ritzwoller and Levshin, 1998) obtains measurements on single waveforms and involves significant analyst interaction. The computational structure of FTAN allows automation, however, and this is what we describe here. Although FTAN has been applied dominantly to measure group speeds, phase speed curves are also measured naturally in the process.

The discussion is facilitated by introducing a little notation. We roughly follow the notation and terminology of Bracewell (1978) (pages 268-272). If $d(t)$ is the waveform of interest, measurements are obtained by considering the “analytic signal” defined as

$$A(t) = d(t) - iH(t) = |A(t)| \exp(-i\phi(t)), \tag{3}$$

where $H(t)$ is the Hilbert transform of $d(t)$. The analytic signal is constructed for a set of narrow band-pass filtered waveforms with center frequencies ω_0. We suppress the ω_0 notation here and assume consideration of the narrow-band filtered waveform hereafter. The modulus of the analytic signal, $|A(t)|$, is a smooth envelope function and $\phi(t)$ is a smooth phase function. The smoothness of the analytic signal is a principal reason for its use. Group speed is measured using $|A(t)|$ and phase speed using $\phi(t)$. In considering the envelope and frequency at a particular instant in time t, we follow Bracewell and use $|A(t)|$ for the envelope function but introduce a refined frequency called the “instantaneous frequency” equal to the time rate of change of the phase of the analytic signal at time t. We, therefore, replace the center frequency of the narrow-band filter, ω_0, with the instantaneous frequency, ω: $\omega_0 \rightarrow \omega = d\phi(t)/dt$. This correction is most significant when the spectrum of the input waveform is not flat. Due to leakage in the frequency domain the central frequencies of the narrow-band filters do not accurately represent the frequency content of the output of the filters.

The FTAN procedure divides into eight steps. We discuss each step and then indicate how the analyst-driven and the automated FTAN processes differ. This will be done in the context of group velocity measurements in section 4.1 and then we will follow with a discussion of how FTAN measures phase speed curves in section 4.2. Figure 14 graphically illustrates the process. In this figure, all results are for the automated FTAN process.

4.1 Group speed measurements

Figure 14a shows a broadband signal obtained from a one-year cross-correlation between stations ANMO and SSPA in the US. In Step 1 of FTAN, a frequency (period) - time (group speed) or FTAN image is produced by displaying the logarithm of the square of the envelope of the analytic signal, $\log|A(t)|^2$, for a set of different filter center frequencies. Figure 14b shows the FTAN image of the waveform in Figure 14a. The envelope functions $|A(t)|^2$ are arrayed vertically on a grid of different values of ω_0 to produce a matrix that can be displayed as a 2-D image. There is a similar phase matrix not displayed here. Typically, group speed replaces time and period replaces filter center frequency. In Step 2, the dispersion ridge is tracked as a function of period to obtain a raw group speed curve. Figure 14b shows this curve as a solid line. This raw group speed measurement may be sufficient for many applications.

Steps 3-8 of FTAN involves phase-matched filtering to clean the waveform of potential contamination and generates an alternative group speed curve. This measurement may be preferable in some applications. In Step 3, an anti-dispersion or phase-matched filter is defined on a chosen period-band. Levshin and Ritzwoller (2001) discuss the phase-matched filtering method in detail. In Step 4, this anti-dispersion filter is applied to the waveform in the period band chosen to produce the undispersed signal. Figure 14c shows the undispersed or "collapsed" signal. In Step 5, contaminating noise is identified and removed from the undispersed signal. Typically, for earthquakes this noise is signal-generated, being composed of multi-pathed signals, seismic coda, body waves, and so forth. An example cut is also shown in Figure 14c. In Step 6, the cleaned collapsed waveform is redispersed. It is shown as the dashed line in Figure 14a. In Step 7, the FTAN image of the cleaned waveform is computed using the same process applied to the raw waveform in Step 1. Figure 14d shows the FTAN image of the cleaned waveform. Finally in Step 8, the dispersion ridge is tracked as a function of period on the cleaned FTAN image to obtain the cleaned group speed curve. Figure 14d shows this curve as a solid line.

The traditional analyst-driven FTAN procedure has been applied to earthquake data by analysts for more than 200,000 individual paths globally (e.g., Shapiro and Ritzwoller, 2002). The analyst, however, only enters the process in Steps 3 and 5. In Step 3, the analyst defines the phase-matched filter and the frequency band of interest, which usually depends on the band-width of the signal that is observed. The analyst either can use the group speed curve that is automatically produced on the raw FTAN image in Step 2 or can define a curve interactively. The latter approach is usually chosen as FTAN images of earthquake data commonly display spectral holes which vitiate the automated group speed measurement. The automated group speed measurements are also often tricked by scattered or multipathed arrivals and, therefore, do not track the dispersion branch of interest accurately. Multipathing and scattering is a problem mostly for large epicentral distances. In Step 5, the analyst interacts with the collapsed signal to remove noise. It is, therefore, only Steps 3 and 5 that require automation beyond the existing method.

To automate Step 3, the group speed measurements that result in Step 2 must be used to define the phase-matched filter. Therefore, these measurements must be robust to spectral holes and scattered or multipathed arrivals. Fortunately, FTAN images that result from cross-correlations of ambient noise tend to be much simpler than those from earthquakes, and spectral holes are rare. Inter-station spacing for ambient noise measurements is also typically less than epicentral distances, so multipathing is not as severe of a problem. The automated procedure, therefore, only differs from the raw group velocity procedure applied during interactive FTAN in that in Step 2 added measures are taken to ensure the continuity of the dispersion curve by rejecting spurious glitches or jumps in group times. Formal criteria are set to reject curves with distinctly irregular behavior or to interpolate through small glitches by selecting realistic local instead of absolute maxima. When glitches are too large in amplitude or persistent in period, the dispersion curve is rejected. Spectral whitening (section 2.2) helps to minimize jumps in the measured curve and incompleteness of measurements at the long period end of the spectrum. The raw group speed curve that emerges from Step 2 is one of two alternative curves that emerge from the automated process

To automate Step 5, the undispersed signal is selected from the surrounding noise automatically. Figure 14c illustrates this procedure graphically using the waveform from Figure 14a. In an ideal case, the anti-dispersed signal will collapse into a single narrow spike. The spike, in Figure 14c for example, is then cut from the surrounding time-series and re-dispersed to give the clean waveform shown in Figure 14a. The principal advantage of this method arises when there exists strong neighboring noise that can be removed from the undispersed signal. In the case of ambient noise cross-correlations, spurious precursory arrivals exist in many cases, particularly at long periods. A good example can be seen in Figure 6a. Such arrivals tend to interfere with the primary signals and the resulting group velocity curves are undulatory. Phase-matched filtering helps reduce the effect of precursory arrivals and smooths the measured group speed curve.

A problem occurs with phase-matched filtering, however, when the waveform of interest is narrow-band. In this case, the undispersed signal will possess prominent side-lobes that will need to be included in the cleaned collapsed signal cut from surrounding noise. If these side-lobes extend broadly enough in time, the cutting procedure may not effectly eliminate contaminating noise. Alternately, if the side-lobes are not included in the selected waveform, the redispersed signal will be biased and the dispersion curve will often be undulatory at the long period end of the measurement. For these reasons, phase-matched filtering (i.e., FTAN Steps 3-8) is only recommended for application to broad-band signals.

4.2 Phase speed measurements

By analysing the envelope function $|A(t)|$, the group speed curve, $U(\omega)$, is measured. Phase speed cannot be derived directly from group speed, but the group speed can be computed from phase speed. To see this, let $U = \partial\omega/\partial k$ and $c = \omega/k$ be group and phase speed, respectively, $s_u = U^{-1}$ and $s_c = c^{-1}$ be group and phase slowness, respectively, and k be wavenumber. Then $s_u = \partial k/\partial\omega = \partial(\omega s_c)/\partial\omega$. which gives the following first-order differential equation relating the group and phase slownesses at frequency ω:

$$\frac{\partial s_c}{\partial \omega} + \omega^{-1} s_c = \omega^{-1} s_u. \tag{4}$$

If the phase speed curve $c(\omega)$ is known, the group speed curve $U(\omega)$ can be found directly from this equation. If the group speed curve is known, this differential equation must be solved to find $c(\omega)$, which involves an integration constant that is generally unknown. The solution is

$$s_c(\omega) = \omega^{-1} \left(\int_{\omega_n}^{\omega} s_u(\omega) d\omega + \omega_n s_c^n \right), \tag{5}$$

where the constant of integration has been written in terms of a boundary condition that the phase speed curve is known at some frequency ω_n: $s_c(\omega_n) = s_c^n$. This is a condition that will generally not apply. Nevertheless, knowledge of the group speed can help to find the phase speed, as we now show.

Measurement of the phase speed curve requires information in addition to the envelope function on which the group speed has been measured. It derives from the phase $\phi(t)$ of the analytic signal which is composed of a propagation term, an initial source phase, and a phase ambiguity term this will be discussed further below. At instantaneous frequency ω, this can be written:

$$\phi(t) = k\Delta - \omega t - \phi_s - \phi_a, \tag{6}$$

where Δ is distance (inter-station or epicentral), k is wavenumber, ϕ_s is source phase, and ϕ_a is the phase ambiguity term. To proceed, we evaluate the observed phase at the group arrival time $t_u = \Delta/U$ and let $k = \omega s_c$ to find the expression for phase slowness:

$$s_c = s_u + (\omega\Delta)^{-1} \left(\phi(t_u) + \phi_s + \phi_a\right). \tag{7}$$

The group speed curves, therefore, enter this process by defining the point in time at which the observed phase is evaluated.

Equation (7) prescribes the phase slowness (and hence the phase speed) curve. Its use, however, depends on knowledge of the initial source phase and the extra phase ambiguity term. In earthquake seismology, ϕ_s is typically computed from Centroid Moment Tensor (CMT) solutions. One of the traditional advantages of studies of group speed over phase speed is that source phase plays a secondary role in group speed (Levshin et al, 1999) and, therefore, group speeds can be measured at short periods unambiguously using small earthquakes without prior knowledge of the CMT solution. For cross-correlations of ambient noise, however, the situation is considerably easier, as the source phase should be zero: $\phi_s = 0$.

For both earthquake and ambient noise studies, the phase ambiguity term contains a part derived from the 2π ambiguity inherent to any phase spectrum: $\phi_a = 2\pi N$, where $N = 0, \pm 1, \pm 2, \ldots$. Typically, this ambiguity can be resolved by using a global 3-D model (e.g., Shapiro and Ritzwoller, 2002) or phase velocity maps (e.g., Trampert and Woodhouse, 1995; Ekstrom et al., 1997) to predict phase speed at long periods. The value of N then is chosen to give the closest relation between theory and observation. If observations extend to long periods (e.g., greater than 40 sec at least, preferably longer), a global model or observed phase velocity maps may predict phase speed well enough to get N right in most cases. As discussed in section 5, we recommend making dispersion measurements only up to a period (in sec) equal to $\Delta/12$, where Δ is in km. To obtain a 40 sec measurement, therefore, requires an inter-station spacing of about 500 km. If resolution of the phase ambiguity requires 100 sec observations, then inter-station spacing of at least 1200 km is recommended. For ambient noise cross-correlations, if observations are limited to short periods or short inter-station distances, the phase ambiguity may not resolve in a straightforward way.

For ambient noise cross-correlations, the phase ambiguity appears to be exacerbated by another factor. Equation (21) of Snieder (2004) shows that the phase of the cross-correlation possesses a term proportional to $\pi/4$ that arises from the stationary phase integration (effectively over sources) in the direction transverse to the two stations. The sign of the term depends on the component of the seismometer, negative for the vertical component and positive for the radial component for a Rayleigh wave. The assumption, however, is that sources are homogeneously distributed with azimuth. If all sources occur only along the two stations, for example, this term would be zero. More theoretical work is needed on this problem, but it is reasonable to assume that for a realistic distribution of sources of ambient noise this term could vary considerably, running between 0 and $-\pi/4$ for a vertical component. Thus, for vertical component ambient noise cross-correlations the phase ambiguity term is $\phi_a = 2\pi N - \lambda\pi/4$, where $\lambda \in [0, 1]$ is a real unknown quantity that depends on the azimuthal distribution of ambient noise sources. Because this distribution may vary strongly with frequency, λ is probably also frequency dependent.

In summary, phase-slowness derived from a vertical component ambient noise cross-correlation can be written

$$s_c = s_u + (\omega\Delta)^{-1} \left(\phi(t_u) + 2\pi N - \lambda\pi/4\right). \quad (8)$$

where $N = 0, \pm 1, \pm 2, \ldots$ and $\lambda \in [0, 1]$. More theoretical work is need to quantify the uncertainty in λ. We recommend that phase speed measurements be made for long inter-station distances, greater than 500 - 1000 km. Shorter path dispersion measurements probably need to be confined to group speeds, but more work is needed to establish this definitively. Figure 15 shows the observed group and phase speed curves measured on the waveform shown in Figure 14a compared with the curves predicted by the 3-D model of Shapiro and Ritzwoller (2002). We set $\lambda = 0$ in this example.

5. Quality control

Because the number of inter-station paths grows as the square of the number of stations, the data processing procedure that is applied to ambient noise cross-correlations must be designed to need minimal human interaction. Erroneous dispersion measurements are more likely to arise than if analysts were providing guidance at strategic intervals along the process. Data quality control measures, therefore, must be devised to identify and reject bad measurements and compute quality assurance statistics for the accepted measurements.

First, we have found that a reliable dispersion measurement at period τ requires an inter-station spacing (Δ) of at least 3 wavelengths (λ): $\Delta > 3\lambda = 3c\tau$ or $\tau < \Delta/3c$. Because $c \sim 4$ km/sec, for measurements obtained at an inter-station spacing of Δ, there is a maximum cut-off period of about $\tau_{\max} = \Delta/12$. We have clearly observed the degradation of dispersion measurements at periods greater than about $\tau_{\max}$, at least for group speeds. This imposes a severe constraint on measurements obtained from small regional arrays such as PASSCAL experiments. A broad-band network 500 km in extent, for example, can only produce measurements up to about 40 sec period, and that only for the stations across the entire array which is only a small subset of the inter-station paths. Intermediate period measurements, such as at 40 sec, will be most likely to be obtained from the array to surrounding stations, which indicates the importance of permanent (back-bone) stations in the context of regional deployments. At present, we have less experience with phase speed measurements obtained on cross-correlations of ambient noise, so it is possible that the period cut-off may be able to be relaxed for phase speeds.

Second, we need means to determine the reliability of dispersion measurements that satisfy the period cut-off criterion. One way to estimate reliability is comparison with ground truth. The best case would be when an earthquake has occurred beneath one of the stations. Figure 16 presents an example comparison. Examples such as this give confidence to the data processing generally, but are too rare to be of specific use for data selection or uncertainty estimation.

The principal metric on which to base a judgment of the quality of the measurements is stability, the robustness of the measurement to perturbations in the conditions under which it is obtained. The stability of spatially clustered and temporally repeated measurements are particularly useful to identify erroneous measurements and to quantify uncertainties.

Clustering measurements obtained at a particular station from a set of earthquakes located near to one another is commonly used to assess uncertainties in earthquake dispersion measurements (e.g., Ritzwoller and Levshin, 1998). A similar cluster analysis can be applied to ambient noise data. For example, Figure 17 presents a spatial cluster analysis that exploits the high station density in southern California. Numerous measurements between southern California and distant stations can be obtained with similar paths. Cross-correlations between the southern California stations and the stations DUG (Dugway, UT) and HRV (Harvard, MA) provide one estimate of uncertainty. Note in this example that for inter-station paths from DUG (Dugway, UT) to southern California there is substantial difference in velocity compared to the CU-Boulder global model (Shapiro and Ritzwoller, 2002). Measurements from longer paths between southern California and HRV, as expected, are closer to the model prediction. Spatial cluster analyses such as this one are only possible under certain restrictive conditions. A tight cluster of stations is needed that subtends a small angle to a relatively distant station (located many inter-station spacings away from the cluster). These conditions typically will not hold for most measurements, although the growth in regional arrays like the Transportable Array component of USArray/EarthScope will help to make this method increasingly applicable. At present, however, cluster analysis provides only an assessment of average uncertainty for long path measurements or a data rejection criterion for a subset of the measurements.

A more useful method to estimate reliability is to assess temporal repeatability. The physical basis for this method is that sources of ambient noise change seasonally and provide different conditions for the measurements. The repeatability of the measurement given the changing conditions is a significant component of reliability. This standard is elevated to a high position in our assessment, as we equate seasonal repeatability with measurement uncertainty. It is one of the salutary features of ambient noise dispersion measurements regionally that uncertainties can be measured, unlike earthquake derived measurements.

Figure 18 presents an example..... DESCRIBE.....

In our previous applications (Yang et al., 2006; Lin et al., 2006b), dispersion measurements are obtained on 12-months of data. To estimate uncertainties in these measurements, dispersion is also measured on all sequential 3-month stacks if signal-to-noise (SNR) exceeds some threshold. Standard-deviation is computed if a sufficient number of the 3-month stacks exceeds the SNR criterion. In the high ambient noise environment of New Zealand, Lin et al. (2006b) required seven of the 3-month stacks to have SNR > 10. Yang et al. (2006), working with the lower ambient noise conditions that prevail across most of Europe, were forced to loosen the criteria (four 3-month stacks with SNR > 7). Both studies rejected any measurement for which an uncertainty measurement could not be determined. Yang et al. (2006), in particular, rejected many measurements because uncertainty could not be determined even with the loosened criteria. They argued, therefore, that at least across much of Europe, two years of data would be preferable to one in order to estimate uncertainties and reject far fewer measurements. Presumably this would be true on most other continents around the world.

If a seismic array is emplaced for a short duration or operates for a short period of time, temporal subsetting to estimate uncertainties may not be feasible. In this case, SNR can be used as a proxy for uncertainties. An example is shown in Figure 19.... DESCRIBE....

Third, we seek measurements that cohere as a whole; that is, that agree with other accepted measurements. This condition can be tested tomographically. Measurements that can be fit with a smooth tomographic map are said to agree with one another. Yang et al. (2006) presents a detailed discussion of the application of this criterion across Europe. He finds that, on average, dispersion measurements that derive from ambient noise tomography can be fit better than those that derive from earthquake data. Moreover, the distribution of misfit is tight. While erroneous measurements do pass the previous selection criteria, they are small in number. An example comparison between the misfit histograms of ambient noise and earthquake derived group speed measurements across Europe is shown in Figure 20.

6. Summary and conclusions

Acknowledgments

The authors thank Roel Snieder for a discussion of the $\pi/4$ phase term that appears in equation (8). All of the data used in this research were downloaded either from the IRIS Data Management Center, the European Orfeus Data Center, or the New Zealand GeoNet Data Center. This research was supported by a contract from the US Department of Energy, DE-FC52-2005NA26607, and grants from US National Science Foundation, EAR-0409217 and EAR-0450082.

References

Barmin, M.P., M.H. Ritzwoller, A.L. Levshin, A fast and reliable method for surface wave tomography, *Pure Appl. Geophys., 158*, 1351-1375, 2001.

Bassin, C., Laske, G. and Masters, G., 2000. The current limits of resolution for surface wave tomography in North America, EOS Trans AGU, 81, F897.

Bensen, G.D., M.H. Ritzwoller, N.M. Shapiro, A.L. Levshin, 2005. Extending ambient noise surface wave tomography to continental scales: Application across the United States. EOS Trans. AGU, 86(52), Fall Meeting Suppl., Abstract S31A-0274.

Bracewell, R.N., 1978. *The Fourier Transform and its Applications*, McGraw-Hill, 2nd edition, New York.

Campillo M. and A. Paul, 2003. Long-range correlations in the diffuse seismic coda, Science, 299,547-549.

Crawford, W.C., R.A. Stephen, and S.T. Bolmer, 2006. A second look at low-frequency marine vertical seismometer data quality at the OSN-1 stie off of Hawaii for seafloor, buried and borehole emplacements, Bull. Seism. Soc. Amer., in press.

Derode, A., E. Larose, M. Tanter, J. de Rosny, A. Tourim, M. Campillo, and M .Fink, 2003. Recovering the Green's function from filed-field correlations in an open scattering medium (L), J. Acoust. Soc. Am., 113,2973- 2976.

Dyman, T. S., Rice, D. D., Westcott, P. A. Geologic controls of deep natural gas resources in the United States.

Dziewonski, A. M., Bloch, S., and M. Landisman, A technique for the analysis of transient seismic signals, Bull. seism. Soc. Am., 59, 427 - 444, 1969.

Ekstrom, G., J. Tromp, and E.W.F.Larson, 1997. Measurements and global models of surface wave propagation, J. Geophys. Res., 8137-8157.

Herrin, E. E. and T. T. Goforth, Phase-matched filters: Application to the study of Rayleigh Waves, Bull. Seism. Soc. Am. 67: 1259-127, 1977.

Holcomb, L.G., 1998. Spectral structure in the Earth's microseismic background between 20 and 40 seconds, Bull. Seism. Soc. Am., 88, 744-757.

Larose, E., A. Derode, M. Campillo, and M. Fink, 2004. Imaging from one-bit correlations of wideband diffuse wavefields, J. Appl. Phys., 95, 8393-8399.

Larose, E., A. Derode, D. Corennec, L. Margerin, and M. Campillo, 2005. Passive retrieval of Rayleigh waves in disoredered elastic media, Phys. Rev. E., 72, 046607, doi:10.113/PhysRevE.72.046607.

Laske, G. and G. Masters, 1997. A Global Digital Map of Sediment Thickness, EOS Trans. AGU, 78, F483.

Levshin, A. L., Pisarenko, V. F., and G. A. Pogrebinsky, On a frequency-time analysis of oscillations, Ann. Geophys., 28, 211 - 218, 1972.

Levshin, A.L., L. Ratnikova, and J. Berger, 1992. Peculiarities of surface wave propagation across central Eurasia, Bull. Seism. Soc. Am., 82, 2464-2493, 1992.

Levshin, A.L. and M.H. Ritzwoller, 2001. Automated detection, extraction, and measurement of regional surface waves, Pure Appl. Geophys., 158(8), 1531 - 1545, 2001.

Levshin, A.L., M.P. Barmin, M.H. Ritzwoller, and J. Trampert, Minor-arc and major-arc global surface wave diffraction tomography, Phys. Earth Planet. Ints., 149, 205- 223, 2005.

Lin, F.-C., M.H. Ritzwoller, and N.M. Shapiro, 2006a. Is ambient noise tomography across ocean basins possible?, Geophys. Res. Lett., in press.

Lin, F.-C., M.H. Ritzwoller, J. Townend, M. Savage, and S. Bannister, 2006b. Ambient noise Rayleigh wave tomography of New Zealand, Geophys. J. Int., in preparation.

Mooney, W.D., G. Laske and G. Masters, CRUST5.1: A global crustal model at $5^\circ \times 5^\circ$. J. Geophys. Res., 103, 727-747, 1998.

Russell, D. W., R. B. Herrman, and H. Hwang, Application of frequency-variable filters to surface wave amplitude analysis, Bull. seism. Soc. Am., 78, 339 - 354, 1988.

Ritzwoller, M.H., and A.L. Levshin, 1998, Surface wave tomography of Eurasia: group velocities, J. Geophys. Research, 103, 4839-4878.

Ritzwoller, M.H., N.M. Shapiro, M.P. Barmin, and A.L. Levshin, Global surface wave diffraction tomography, J. Geophys. Res., 107(B12), 2335, 2002.

Ritzwoller, M.H., Y. Yang, N.M. Shapiro, G.D. Bensen, 2005, Broad-Band Ambient Noise Surface Wave Tomography Across Eurasia: Early Results. EOS Trans. sAGU, 86(52), Fall Meeting Suppl., Abstract S31A-0279.

Roux, P., K.G. Sabra, P. Gerstoft, W.A. Kuperman, and M.C. Fehler, 2005. P-waves from cross-correlation of seismic noise, Geophys. Res. Lett., 32, L19393, doi:10.1029/2005GL023803.

Sabra, K. G., P. Gerstoft, P. Roux, W. A. Kuperman, and M. C. Fehler 2005a. Extracting time-domain Green's function estimates from ambient seismic noise, Geophys. Res. Lett., 32, L03310, doi:10.1029/2004GL021862.

Sabra, K. G., P. Gerstoft, P. Roux, W. A. Kuperman, and M. C. Fehler 2005b. Surface wave tomography from microseism in southern California, Geophys. Res. Lett., 32, L14311, doi:10.1029/2005GL023155.

Shapiro, N.M. and M.H. Ritzwoller, Monte-Carlo inversion for a global shear velocity model of the crust and upper mantle, Geophys. J. Int., 151, 88-105, 2002.

Snieder, R., 2004. Extracting the Green's function from the correlation of coda waves: A derivation based on stationary phase, Phys. rev. E, 69, 046610.

Trampert, J. and J.H. Woodhouse, 1995. Global phase velocity maps of Love and Rayleigh waves between 40 and 150 sec, Geophys. J. Int., 122, 675-690.

Van der Hilst, R.D., and De Hoop, M.V., Banana-doughnut kernels and mantle tomography, Geophysical Journal International, v. 163, p. 956-961, 2005.

Wapenaar, K., 2004. Retrieving the elastodynamic Green's function of an arbitrary inhomogeneous medium by cross correlation, Phys. Rev. Lett., 93, 254301, doi:10.1103/PhysRevLett.93.254301.

Weaver, R.L. and O.I. Lobkis, 2001b. On the emergence of the Green's function in the correlations of a diffuse field, J. Acoust. Soc. Am., 110, 3011-3017.

Yao, H., R.D. van der Hilst, and M. V. de Hoop, 2006. Surface-wave tomography in SE Tibet from ambient seismic noise and two-station analysis: I. - Phase velocity maps, Geophys. J. Int., submitted.

Zhou Y., Dahlen, F.A. and Nolet, G., 2004. 3-D sensitivity kernels for surface-wave observables, Geophys. J. Int., 158, 142-168.

Received ____________________

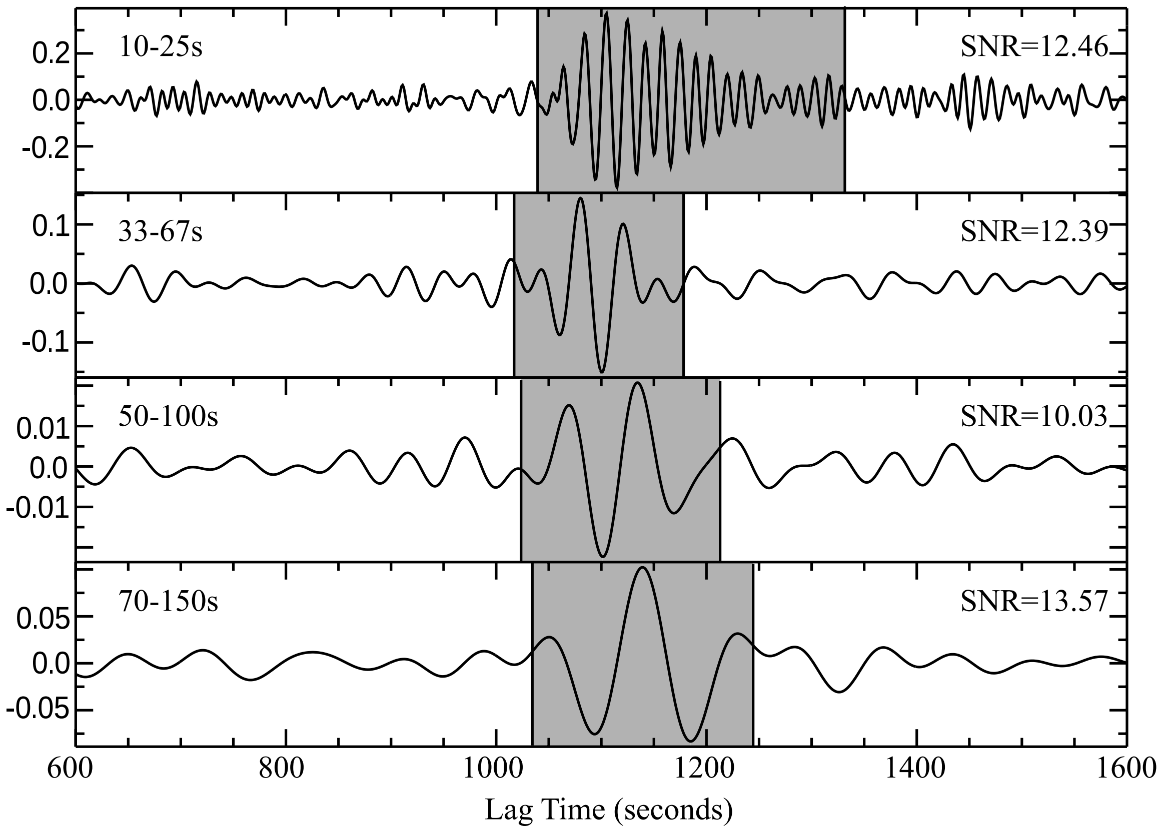

Figure 1. Example cross-correlation between 12-months of data recorded at two Pacific island stations (RPN, Rapanui, Easter Island; PPT, Papeete, Tahiti) computed using the data processing method described in this paper. The "symmetric component" cross-correlation is shown, which is the average of the cross-correlation at positive and negative correlation lag. Grey shaded regions mark the group arrival window predicted by the 3-D Vs model of Shapiro and Ritzwoller (2002), expanded by 75 sec in both directions. The temporal signal-to-noise ratio (SNR) is defined as the peak amplitude in the window divided by the root-mean-square of the trailing noise in a time window of similar length. Temporal SNR is identified on each of the panels, which present different sub-bands of the cross-correlation.

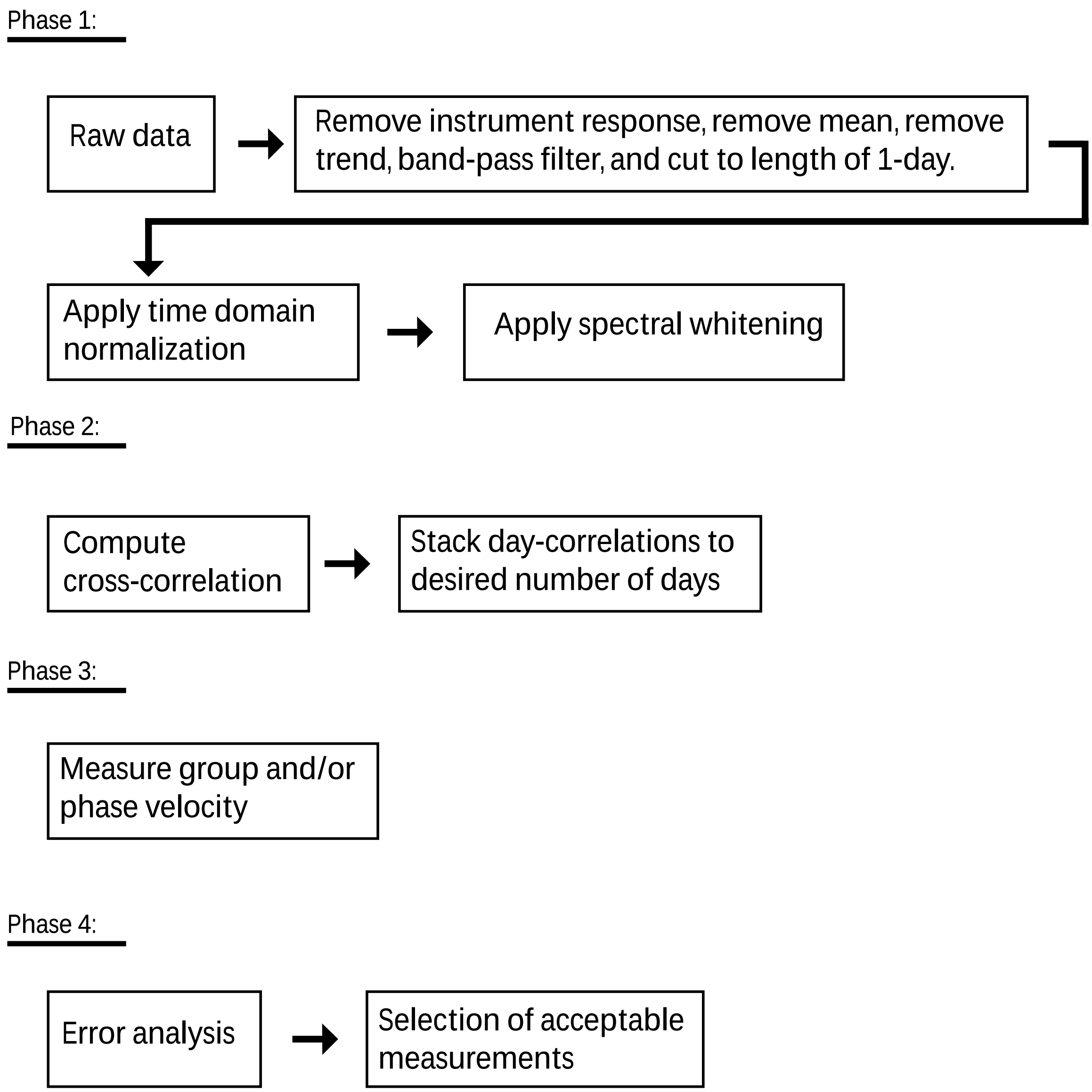

Figure 2. Schematic representation of the data processing scheme. Phase 1 (described in section 2 of the paper) shows the steps involved in preparing single-station data prior to cross-correlation. Phase 2 (section 3) outlines the cross- correlation procedure and stacking, Phase 3 (section 4) includes dispersion measurement and Phase 4 (section 5) is the error analysis and data selection process.

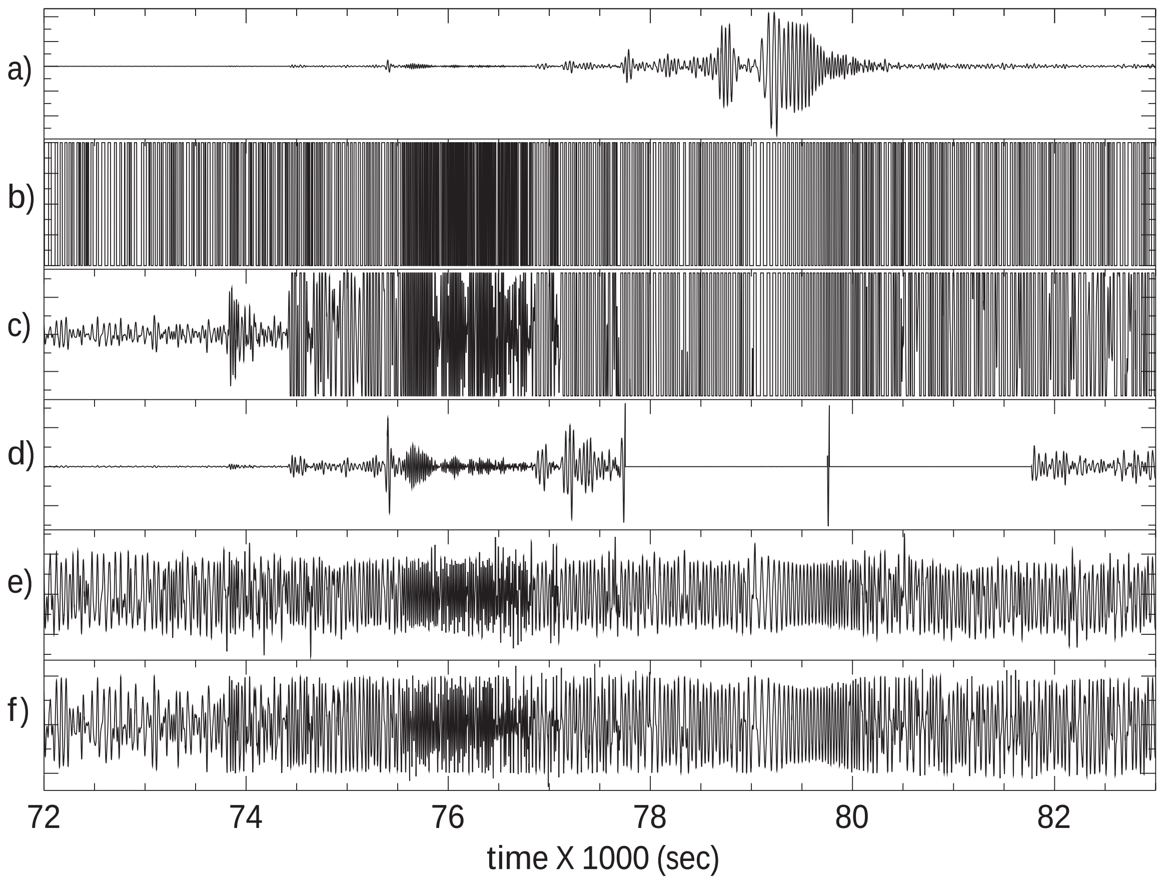

Figure 3. Waveforms displaying examples of the five types of time domain normalization tested. The examples are band-pass filtered between 20 and 100 sec period to clarify the contamination by the earthquake signal. (a) Raw data showing ∼3 hours of data windowed around a large earthquake (Ms = 7.2, Afghanistan-Tajikistan border region) recorded at station ANMO (Albuquerque, NM, USA). (b) One-bit normalized waveform, whereby the signal is set to ±1 depending on the sign of the original waveform. (c) Clipped waveform, where the clipping threshold is equal to the root-mean-square (rms) amplitude of the signal for the given day. (d) Automated event detection and removal. If the amplitude of the waveform is above a certain threshold, the next 30 minutes of it are set to zero. (e) Running absolute mean normalization whereby the waveform is normalized by a running average of its absolute value. (f) "Water level normalization" whereby any amplitude above a certain multiple of the daily rms-amplitude is down-weighted. It is run iteratively until the entire waveform is nearly homogeneous in amplitude.

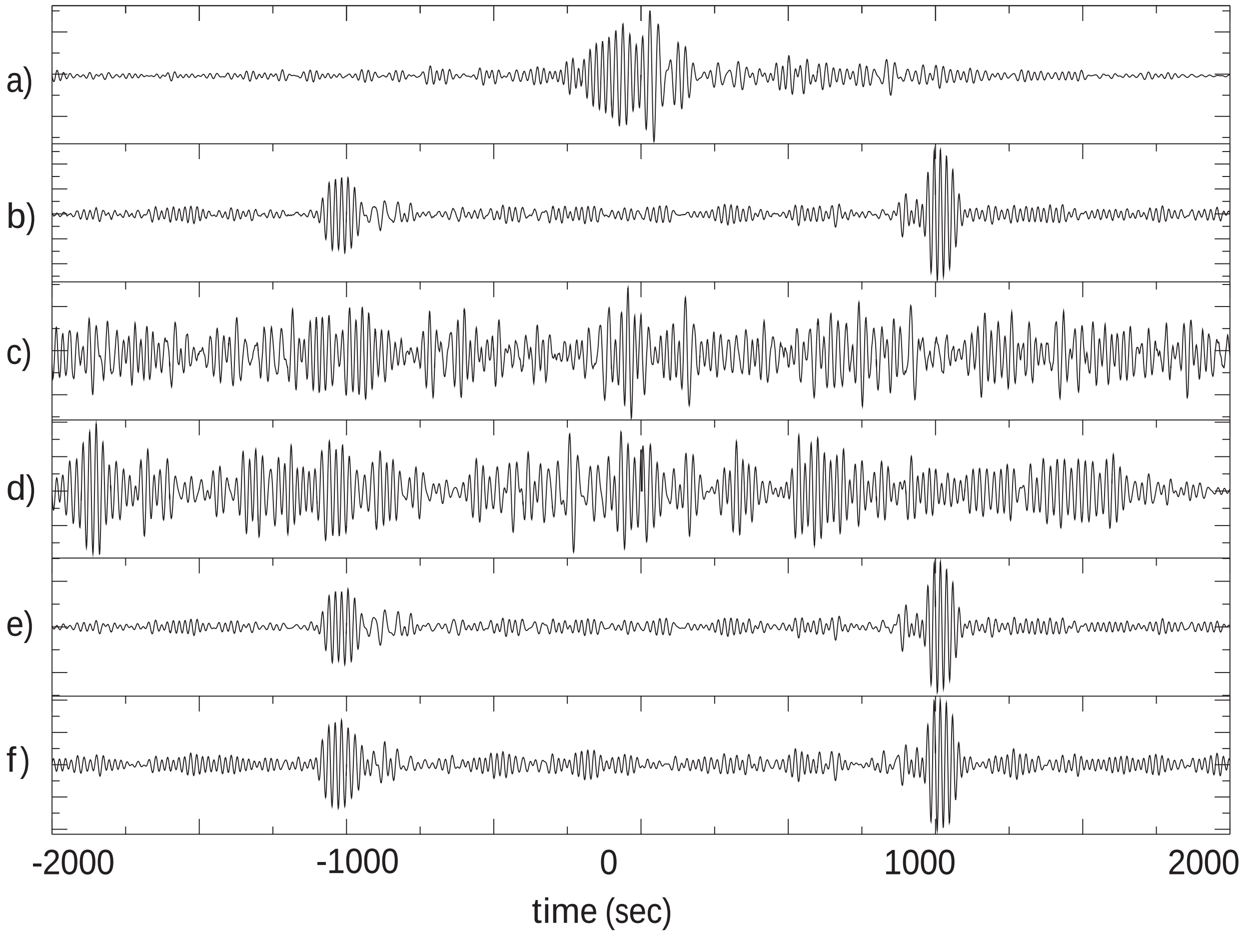

Figure 4. Twelve-month cross-correlations between the station-pair ANMO (Albuquerque, NM, USA) and HRV (Harvard, MA, USA) for the time-domain normalization methods shown in Figure 3. The pass-band is 20 - 100 seconds period. The panels of the figure (a-f) correspond to those in Figure 3.

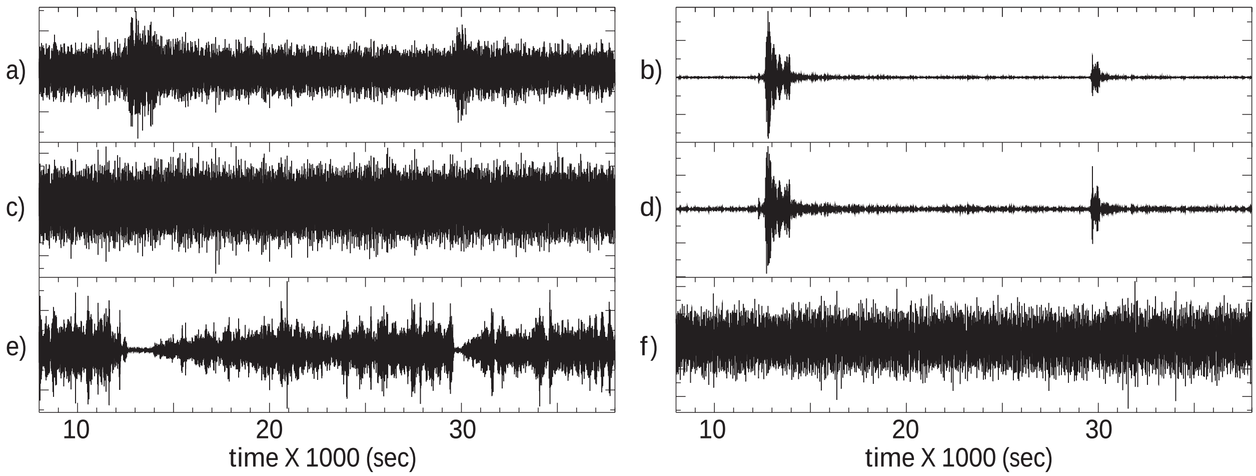

Figure 5. Example of the effect of tuning time-domain normalization to earthquake signals for data from GeoNet station CRLZ in New Zealand. (a) Raw broad-band data from Oct. 14, 2005 showing two earthquakes barely emerging above background noise. (b) Data from (a) band-pass filtered between 15 - 50 sec period, more clearly showing the two earthquake signals (first: S. Fiji, mb = 5.4; second: S. of Kermadec, mb = 5.1). (c) Data after temporal normalization using the running absolute mean method in which the weights are defined on the raw (unfiltered) data in (a). (d) Data from (c) band-pass filtered between 15 - 50 sec period, showing that the earthquake signals have not been removed by temporal normalization defined on the raw data. (e) Data after temporal normalization using the running absolute mean method in which the weights are defined on the band-pass filtered data in (b). (f) Data from (e) band-pass filtered between 15 - 50 sec period, showing that the earthquake signals have been removed by temporal normalization defined on the band-pass filtered data.

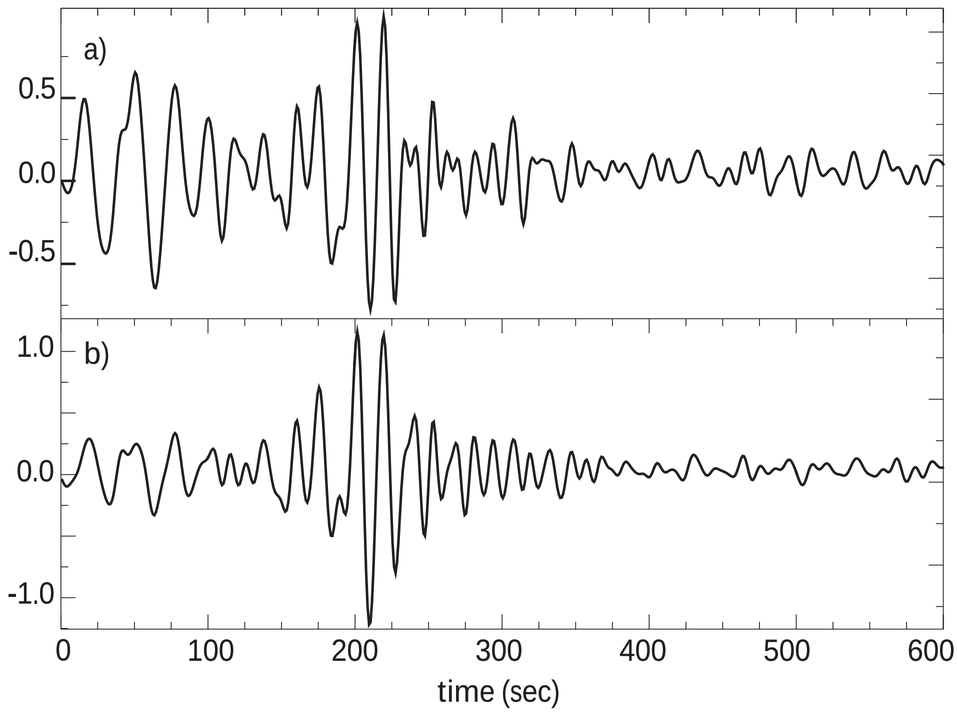

Figure 6. Example of the effect of tuning time-domain normalization to earthquake signals on cross-correlations computed between GeoNet stations CRLZ and HIZ in New Zealand. (a) Year-long cross-correlation in which the temporal normalization is defined on the raw data. (b) Year-long cross-correlation in which the temporal normalization is defined on data band-pass filtered between 15 and 50 sec period. Spurious precursory arrivals are substantially reduced in (b) relative to (a). Waveforms are band-pass filtered between 5 and 50 sec period.

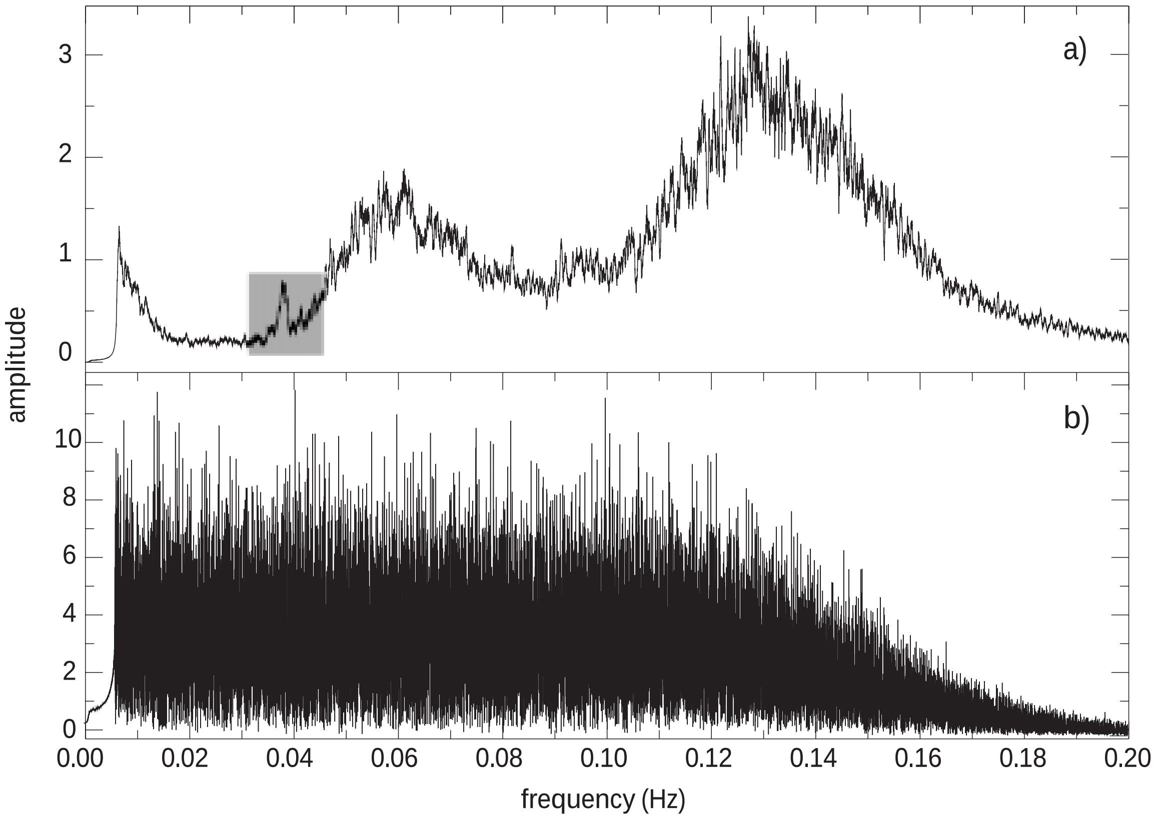

Figure 7. (a) Raw and (b) spectrally whitened amplitude spectra for 1 sample per second vertical component data at station HRV (Harvard, MA) for July 5, 2004. The shaded box indicates the location of the 26 second period signal originating from the Gulf of Guinea. The taper seen at both ends of the spectra is largely attributable to a 7 - 150 second band-pass filter.

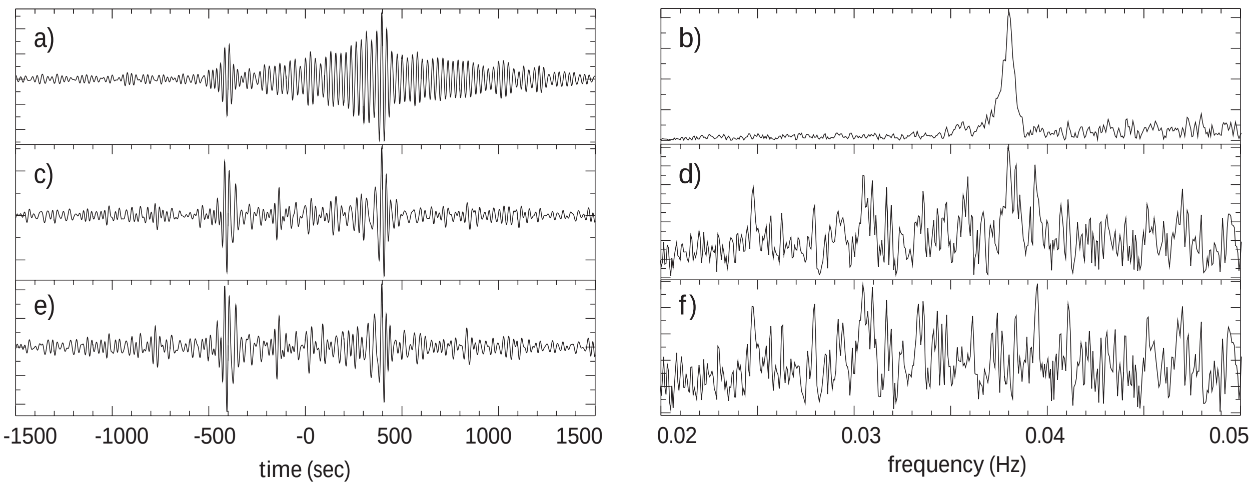

Figure 8. Affect of the 26 sec microseism on cross-correlations, and attempts to remove it. (a) Twelve-month cross-correlation between data from stations ANMO (Albuquerque, NM) and CCM (Cathedral Cave, MO). The broad, nearly monochromatic 26 signal at positive lag dominates the waveform. (b) Amplitude spectrum of the cross-correlation in (a) showing the spectral peak at about 26 sec period. (c) Cross-correlation between data from the same two stations that have been spectrally whitened prior to cross-correlation. (d) Amplitude spectrum of the cross-correlation in (c) showing that the 26 sec spectral peak is largely missing. (e) Cross-correlation between the data that have been spectrally whitened prior to cross-correlation with a notch filter applied around 26 sec period. (f) Amplitude spectrum of the cross-correlation in (e). Application of the notch filter changes the cross-correlation only minimally.

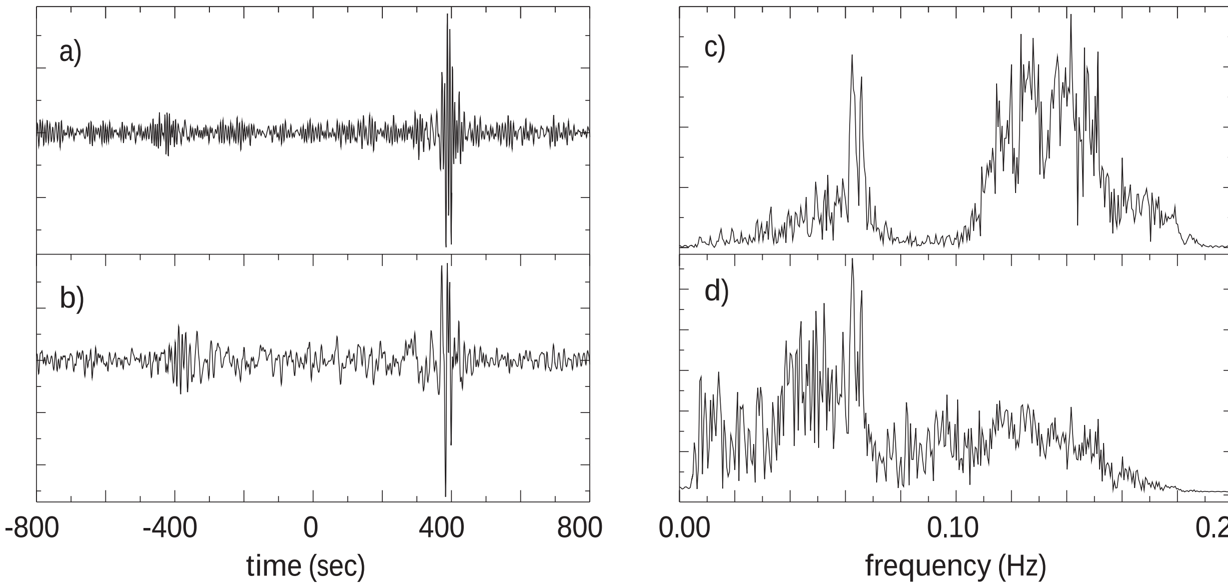

Figure 9. Comparison of cross-correlations with and without spectral whitening. Cross-correlation are for the month April, 2004 for data from stations CCM (Cathedral Cave, MO) and SSPA (Standing Stone, PA) band-pass filtered from 7 to 150 seconds period. (a) Cross-correlation without spectral whitening. (b) Cross-correlation with spectral whitening. (c) Amplitude spectrum of the unwhitened waveform in (a). The primary and secondary microseisms dominate the spectrum. (d) Amplitude spectrum of the prewhitened waveform in (b).

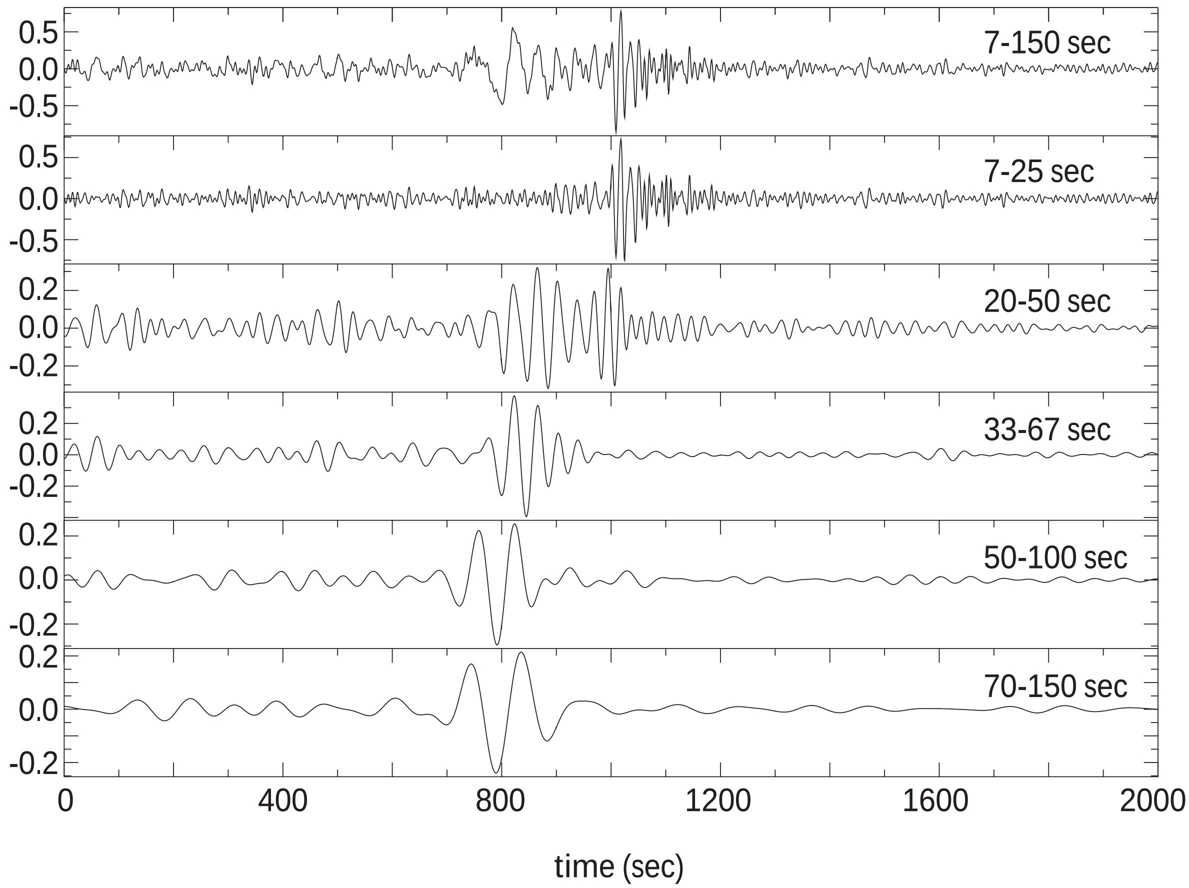

Figure 10. Example of a broad-band symmetric-component cross-correlation using 12-months of data from stations ANMO (Albuquerque, NM, USA) and HRV (Harvard, MA, USA). The broad-band signal (7 - 150 sec pass-band) is shown at top and successively longer period pass-bands are presented lower in the figure.

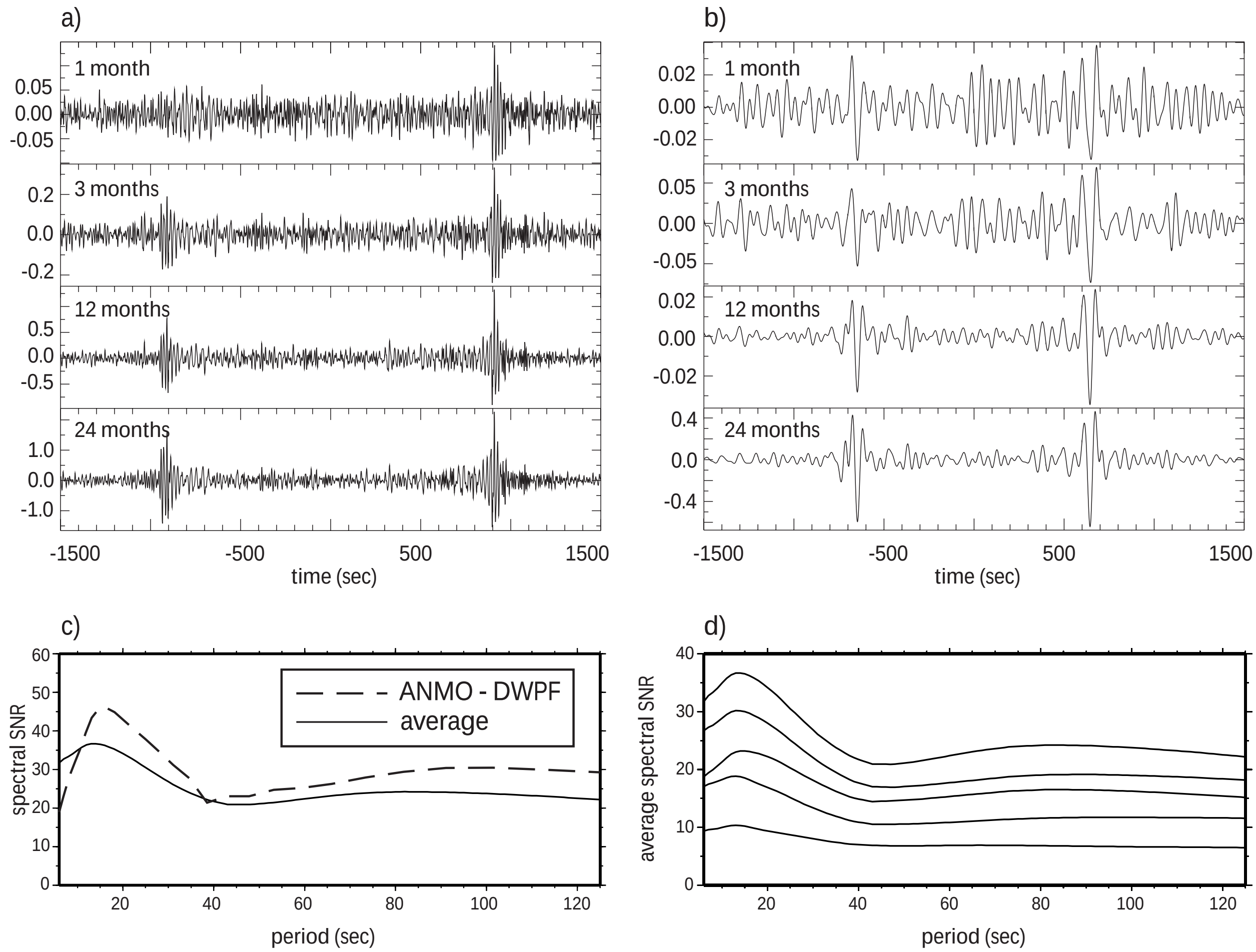

Figure 11. Example of the emergence of the Rayleigh waves for increasingly long time-series. (a) Cross-correlations at the specified time-series lengths for the station pair ANMO (Albuquerque, NM, USA) and DWPF (Disney Wilderness Preserve, FL, USA) band-passed between 5 to 40 sec period. (b) Same as (a), but for a pass-band between 40 sec to 100 sec period. (c) Spectral SNR for the 12-month ANMO-DWPF cross-correlation shown with a dashed line, and the spectral SNR averaged over all cross-correlations between GSN stations in the US shown with a solid line. (d) Spectral SNR averaged over all cross-correlations between GSN stations in the US for different time-series lengths of 1, 3, 6, 12, and 24 months.

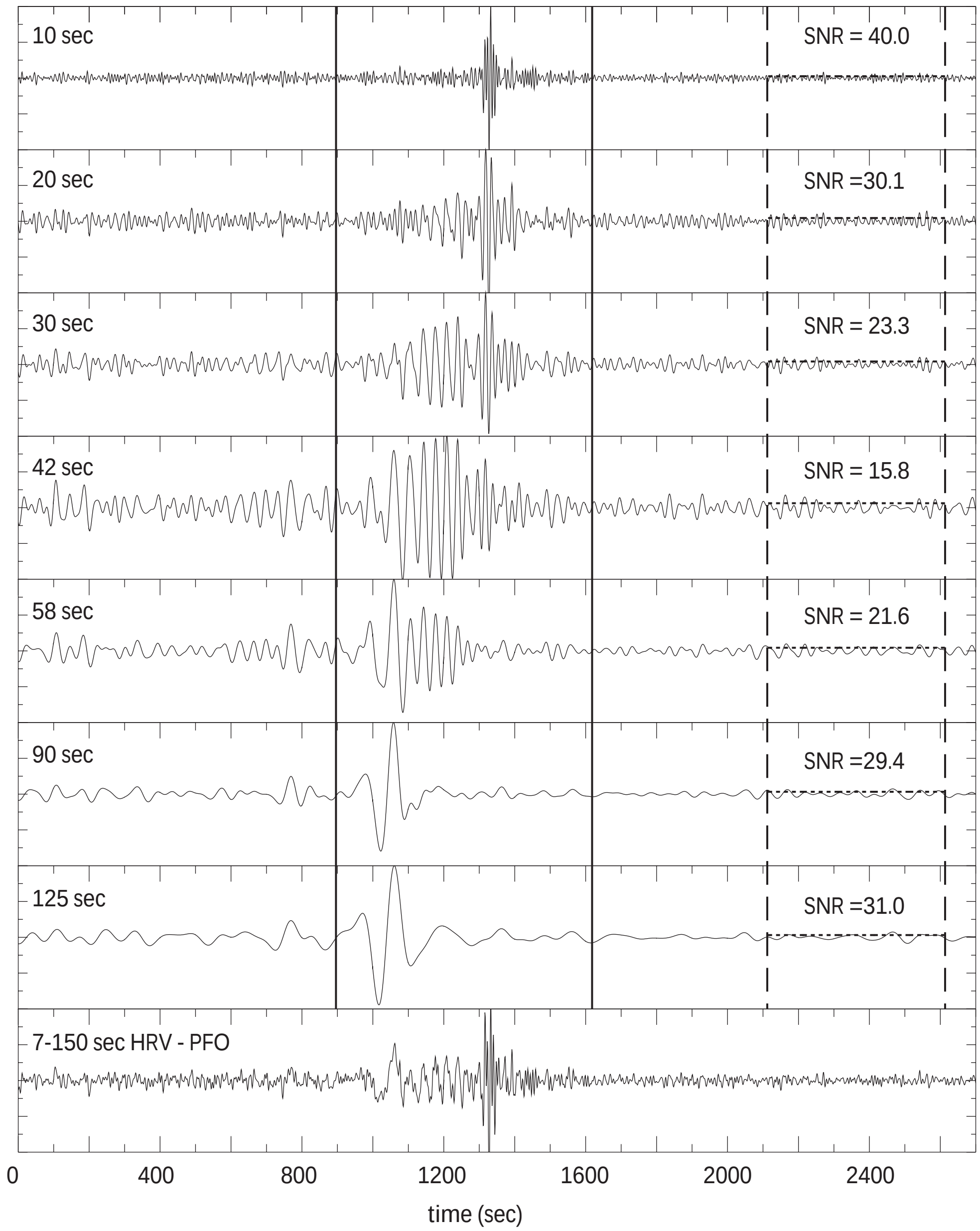

Figure 12. Example of how spectral SNR measurements are obtained on a 12-month cross-correlation between data from stations HRV and PFO (Pinyon Flat, CA, USA). Vertical solid lines indicate the signal windows and vertical dashed lines the noise windows. Waveforms are centered on the period indicated at left in each panel, and SNR is defined as the ratio of the peak within the signal window and rms-noise in the noise window. The noise level is presented as the horizontal dotted lines in the noise windows. SNR in each band is indicated at right in each panel.

Figure 13. Emergence of the signal. SNR vs time-series length

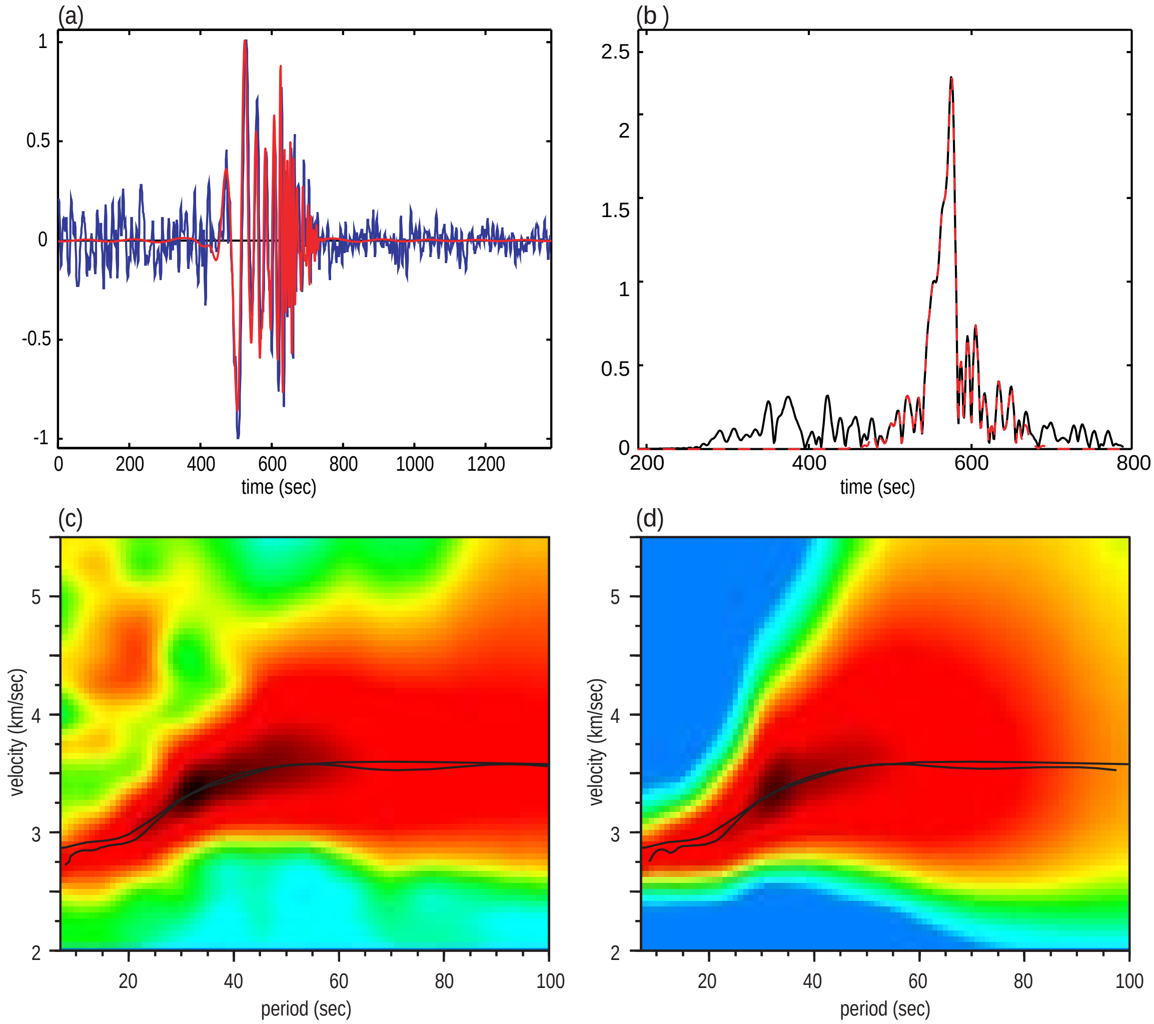

Figure 14. Graphical representation of FTAN. (a) Raw (black) and cleaned (red) waveforms for the 12-month stacked cross-correlation between stations ANMO and COR (Corvallis, OR, USA). (b) Raw FTAN diagram, measured group speed curve, and prediction from the 3-D model of Shapiro and Ritzwoller (2002). (c) Undispersed signal (black) and cleaned signal (red). (d) Cleaned FTAN diagram and measured group speed curve.

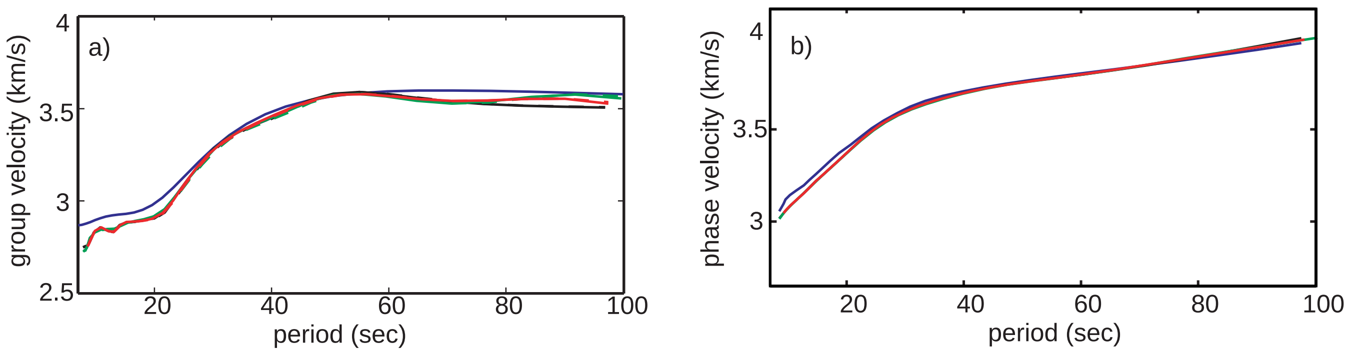

Figure 15. Observed (a) group speed and (b) phase speed curves obtained in Fig. 14 compared with predictions from the 3-D model of Shapiro and Ritzwoller (2002).

Figure 16. Comparison of cross-correlation with earthquake record.

Figure 17. Spatial stability of measurements. Cross correlation between single stations and station cluster.

Figure 18. Temporal stability of measurements. Seasonal variability.

Figure 19. SNR as a proxy for uncertainty.

Figure 20. Measurement coherence. Ambient noise and earthquake misfit histograms at 20 sec period across Europe.