Coherence-based approaches for estimating the composition of the seismic wavefield

M. Coughlin[1], J. Harms[2,3], D.C. Bowden[4], P. Meyers[5], V.C. Tsai[4], V. Mandic[5], G. Pavlis[6],T. Prestegard[5]

[1]Division of Physics, Math, and Astronomy, California Institute of Technology, Pasadena, CA 91125, USA
[2]Gran Sasso Science Institute (GSSI), I-67100 LâĂŹAquila, Italy
[3]INFN, Laboratori Nazionali del Gran Sasso, I-67100 Assergi, Italy
[4]Seismological Laboratory, California Institute of Technology, Pasadena, CA 91125, USA
[5]School of Physics and Astronomy, University of Minnesota, Minneapolis, Minnesota 55455, USA
[6]Department of Geological Sciences, Indiana University, Bloomington, IN 47405, USA

Key Points:

- Contrary to standard models, microseism observations in the former Homestake mine in South Dakota indicate dominance of body waves generating the global seismic low-noise model at 0.2 Hz interrupted by fundamental Rayleigh wave transients
- Inclusion of underground seismometers allow for the prediction of seismic array measurements at better than the 1% level

Corresponding author: Michael Coughlin, mcoughli@caltech.edu

Abstract

As new techniques exploiting the Earth's seismic ambient noise field are developed and applied, such as for the observation of temporal changes in velocity structure, it is crucial to quantify the precision with which measurements can be made. This work considers two aspects that control this precision: the type of seismic wave contributing to the ambient noise field at microseism frequencies and the effect of array geometry, specifically with the inclusion of underground stations. Both of these are quantified using measurements of wavefield coherence between stations, though this is explored in several ways: coherencies are examined in frequency-wavenumber domain, in time domain, as a function of station-station distance and by the construction of Wiener filters. Regarding the type of wave contributing to the ambient noise field at microseism frequencies (i.e., 0.2 Hz), we find a strong seasonal change between body-wave and surface-wave content that could be misinterpreted as a change in velocity structure for small-aperture arrays. Regarding the inclusion of underground stations, we quantify the lower limit to which the ambient noise field can be resolved; the application of Wiener filters is more or less successful in describing background ambient noise depending on the array geometry and aperture. This implies that there is some amount of random variability in coherence-based measurements that cannot be reliably used given the array in question. We discuss the implications of these results for the geophysics community performing ambient seismic noise studies, as well as for the cancellation of seismic Newtonian noise in ground-based, sub-Hz, gravitational-wave detectors.

1 Introduction

Significant effort has been made in the wider seismological community to understand and exploit background ambient seismic noise. One important mechanism for the generation of seismic noise relates to continuous harmonic forcing of ocean waves as they interact with both the seafloor and coastlines, and this varies strongly in time, frequency and azimuth [*Longuet-Higgins and Ursell*, 1948]. These mechanisms most strongly generate energy in the range of 0.06-0.13 Hz (8 to 16 second periods), but a much wider range of periods is also observed worldwide and there can also be strong body wave components as well [e.g., *Gerstoft et al.*, 2008]. Efforts to image these noise sources usually use array processing methods that consider the coherence of wavefronts incident upon the array, referred to as beamforming or k-f analysis [*Capon*, 1969, *Rost and Thomas*, 2002, *Gerstoft et al.*, 2008], similar in many ways to the effort described here.

Particular attention has been paid to understanding the effect that the inhomogeneous distribution of noise sources would have on the coherence or cross-correlation measured between stations, with the goal of determining whether measurements can be reliably used for the study of seismic velocities or attenuation [e.g., *Cupillard and Capdeville*, 2010, *Weaver*, 2011, *Tsai*, 2009, 2011, *Lawrence and Prieto*, 2011], with additional studies exploring the extent to which signal preprocessing can reduce the effect of imhomogeneous noise sources [e.g., *Viens et al.*, 2017, *Bensen et al.*, 2007]. Some of these velocity or attenuation measurements require a great amount of precision and stability over time, such as for the observation of material velocity changes [i.e., *Brenguier et al.*, 2008]; velocity variations on a daily or monthly timescale may be as small as a couple percent, but have been shown to yield valuable information regarding temperature or pore pressure changes. This paper explores two aspects of such cross-correlation or coherence based observations that affect the final precision with which measurements may be reliably made.

The first is an analysis of the types of waves that constitute the background ambient noise field. Should the relative contributions of body-wave energy compared to surface-wave energy change over time, this may bias the velocities measured from coherence or correlation techniques, especially when the inter-station distance is small enough that different seismic phases are not well separated. Coherence measurements are considered in wavenumber-frequency domain, as a function of station-station distance and in time-domain, with the conclusion that for the secondary microseism at 0.2 Hz, differing velocities are observed over the course of a year that can only be explained by differences in the type of wave dominating the measurements. This conclusion that body waves often dominate the wavefield at this frequency has strong implications for the reliability of coherence-based velocity observations and indicates that care should be taken if measurements are to be made in particular seasons.

The second analysis considers the geometry of the array being used, and the lower limit to which the wavefield can be adequately resolved. Specifically, we explore the utility of adding underground seismometers as compared to most seismic arrays which are constrained to observations at the Earth's surface. This characterization is done through the construction of "Wiener filters," which simultaneously use coherencies between all stations in an array rather than on a station-station basis as is usually done. These are optimal linear filters designed to cancel noise; the extent to which ambient noise can be predicted and subtracted from a given target station directly relates to the array's efficacy at describing the wavefield under changing conditions. Using underground stations is shown to improve Wiener filter predictions by a factor of 4, suggesting that the resolution of time-dependent seismology can be significantly improved by going underground.

For most of this analysis we focus on a new seismic array at the former Homestake mine in Lead, South Dakota. Since mining activity has ceased, the Sanford Underground Research Facility there has been demonstrated to be a world-class, low-noise environment [*Harms et al.*, 2010, *M Coughlin*, 2014, *Mandic et al.*, 2017]. In 2015 and 2016, a PASSCAL array of 24 broadband instruments (15 underground and 9 above ground) were deployed in and around the mine, covering horizontal distances of more than 6000 m, and vertical depths of about 1500 m. The quiet environment and 3D geometry make the array an ideal location to test the approaches and questions described above, though an data from an array in Sweetwater, Texas, is also briefly used as one example to show that conclusions regarding the wavefield composition are not solely constrained to the array South Dakota.

2 Velocity measurements and wavefield composition

This section considers velocity observations made through different approaches, with the conclusion that body waves and surface waves contribute energy at different amounts over the course of a year. At 0.2Hz observed velocities shift substantially depending on the season, indicating a dominance of either surface waves or body waves. This may be misinterpreted as a time-varying velocity change for small aperture arrays where seismic phases are not well separated, and implies care should be taken if observations are to be stacked over an entire year or short deployments are to be used in particular seasons.

Observations in this section are made by considering station-station coherence. This is similar in many ways to the cross-correlations used by other studies, and we define our observations formally here. The first step is to calculate the complex spectral coherence of all of the vertical channels of seismometer pairs using one hour of data. The one hour coherences between seismometers i, j were collected over several months in their complex form

$$\gamma_{ij}(f) = \frac{\langle x_i(f)\, x_j^*(f)\rangle}{\sqrt{\langle |x_i(f)|^2\rangle \langle |x_j(f)|^2\rangle}} \tag{1}$$

where $x_i(f)$ is the value of the Fourier Transform at a particular frequency f for the ith seismometer, $x_i^*(f)$ its complex conjugate, and $\langle\rangle$ indicate an average. This metric keeps information about relative phases between seismometers.

Assuming that all seismic sources are sufficiently distant, we can divide the seismic field into three components: plane shear waves, compressional waves, and surface Rayleigh and Love waves. Our goal is to obtain speed estimates by observing the ambient seismic field. In this case, an additional challenge (relative to methods using specific earthquake events) is that there can be multiple waves contributing simultaneously at all frequencies.

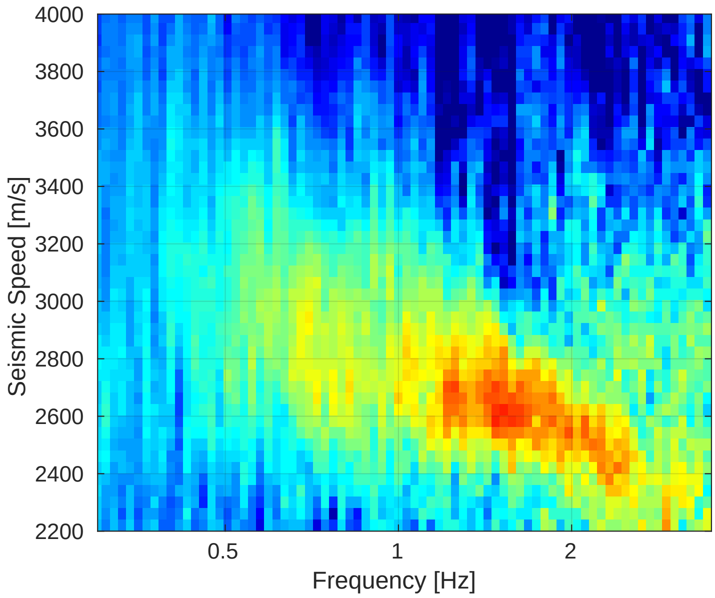

Figure 1. A histogram of seismic speeds between 0.3 – 3.5 Hz. Red color means that the respective speed value was measured for a large number of k-f maps, while blue color means that the speed value was measured rarely.

The array dimension, i.e. the array size and density of instruments, then sets a lower limit to the range of frequencies where multiple waves can be disentangled to obtain well-defined differential phases between sensors.

2.1 Observations in Frequency-wavenumber Domain

Our first estimate of wavespeed, is done in frequency domain using "k-f maps," which are a 3-dimensional data product with the two components of the horizontal wave-vector on two axes, $\vec{k}$, and frequency on the third. Each coherence value was calculated with 128 s FFT length, no overlap, and averaged over the course of a given day. The next step was to calculate the corresponding k-f map. This can be understood as a parameterized stacking method, where a plane-wave model is used to search over all possible phase shifts as a function of propagation directions and seismic speed using a high-dimensional sampling algorithm [*Feroz F., Hobson M.P., and Bridges M.*, 2009]:

$$m(\vec{k}, f) = \sum_{i,j} \gamma_{ij}(f)\, \mathrm{e}^{\mathrm{i}\,\vec{k}(f)\cdot\vec{r}_{ij}}, \tag{2}$$

where $\vec{r}_{ij}$ are the relative position vectors between seismometers, and the wave vector $\vec{k}(f)$ is determined by seismic speed and propagation direction. We make histograms of seismic speeds at each frequency bin proportional to $m(\vec{k}, f)$ and add them in order to construct probability distributions for the seismic speeds. With the weights from $\gamma_{ij}(f)$, this method also takes into account the degree of coherence, as contributions from low-coherence pairs are suppressed. If the dimensions of the array and seismometer spacing are favorable, then one can potentially find multiple distinct local maxima, which correspond to different, simultaneously present waves.

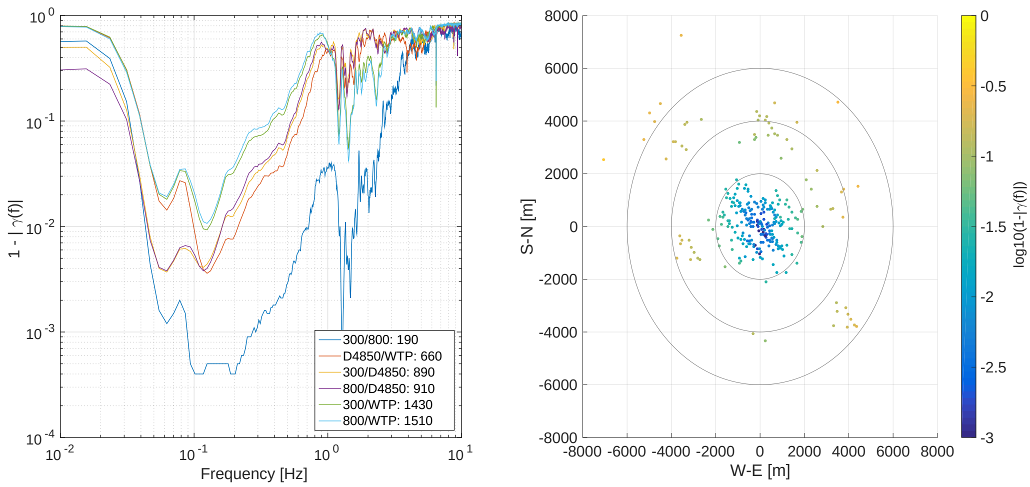

Figure 2. Left plot: $1 - |\gamma(f)|$ between a variety of seismometer pairs averaged over 6 months of coincident data divided into 128 s segments. The legend indicates the horizontal distance in meters between each pair shown, and the pairs are shown in ascending order of horizontal distance. Right plot: logarithm of $1 - |\gamma(f)|$ at 0.2 Hz between all seismometers, where the x,y-coordinates correspond to the relative horizontal position vector between two seismometers.

The k-f maps are calculated for each 1-day coherence matrix, and collected to produce histograms covering a period of about one year. One such histogram is shown in Figure 1. The plot shows seismic speeds in the range between 0.3 Hz to 3.5 Hz. The distribution of maxima tends to lower speed values at higher frequencies, which is the normal dispersion of Rayleigh waves. Between 1 Hz and 2 Hz, Rayleigh-wave speed is found to be about 2.6 km/s falling to lower values above 2 Hz. Speed estimates below 1 Hz were less accurate limited by the dimension of the array, and above 2.5 Hz because of loss of coherence between seismometers. We do not get meaningful speed estimates above 2.5 Hz since coherence between stations becomes very low above 2.5 Hz. At the same time, the array dimension prevented us from obtaining good estimates of seismic speeds below 1 Hz, where the width of the histogram is too large to clearly identify a specific mode. This arises from the difficulty in measuring well-defined differential phases between sensors at these frequencies. It was possible to decrease the width of the distribution by increasing correlation time, but 1-day averaging was about the maximum that could be done keeping a sufficiently high number of samples for the histogram. The histogram traces out a dispersion curve consistent in shape and absolute value with Rayleigh-wave models. While we are therefore confident that the wavefield above 0.3 Hz is dominated by surface waves, we must turn to alternative methods to further investigate the wavefield at lower frequencies.

2.2 Coherence Decay with Station-Station Distance

We can also explore the strength of coherence ($\gamma(f)$ in equation 1) as a function of frequency and station-station distance. Here, coherence was calculated with 50% overlap, and in this form also used later for the Wiener filter section. Coherence is considered for all station-station pairs, and the left panel of Figure 2 shows the difference $1 - |\gamma(f)|$ for a few pairs. Accordingly, coherence is generally high within the band of the primary and secondary oceanic microseismic peaks between a few tens of mHz and 1 Hz, and is insignificant above a few Hertz. Horizontal distances between the seismometer pairs are shown in the legend. At most frequencies, the shorter the horizontal distance, the higher the coherence. The

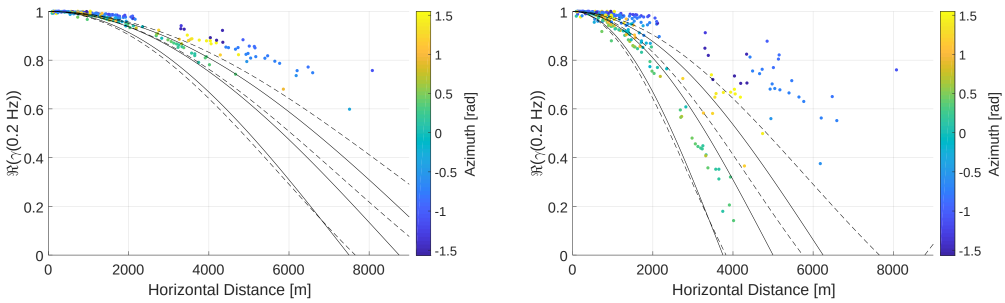

Figure 3. The RPCC as a function of distance between the vertical channels of all seismometers at 0.2 Hz (left: day 154 of 2015; right: day 191 of 2015). The colors correspond to the azimuth with respect to the east direction of the line connecting two seismometers. On the left, models are shown for the single plane wave (solid lines; speed values 6 km/s, 7 km/s, 8 km/s), and isotropic field (dashed lines; speed values 4 km/s, 5 km/s, 6 km/s). The same models were used in the right with speed values 3 km/s, 4 km/s, 5 km/s (solid lines; single plane wave), and 2 km/s, 3 km/s, 4 km/s (dashed lines; isotropic). The velocities are chosen to be consistent with body waves (left) and fundamental Rayleigh waves (right).

right plot shows the logarithm of $1 - |\gamma(f)|$ at 0.2 Hz for day 191 of year 2015 in a scatter plot where the two coordinates are the components of the relative horizontal position vector between two seismometers. We highlight 0.2 Hz because it is the most coherent frequency in the array, as can be seen on the left plot in Figure 2, and also is the strongest contributor of seismic noise [*Longuet-Higgins*, 1950]. We do not include a third coordinate for depth since Rayleigh waves, which are the dominant contribution at these frequencies on day 191 (see below), are known to produce displacement whose phase does not depend on depth (although the relative body wave contribution may change with depth). Coherence is well characterized by the horizontal distance between seismometers. There are no major inhomogeneities or outliers from the overall pattern, but close inspection of the plot reveals significant directional dependence approximately aligned with the north-northwest-south-southeast and west-southwest-east-northeast directions.

The rate at which coherence decays as a function of distance can also be used to place constraints on the seismic velocities [e.g., *Aki*, 1957], and therefore on the composition of the wavefield. However, the decay rate depends on assumptions about the background ambient noise field, so we explore two end-member, orthogonal models for the wave field. The first is under the assumption that it is composed of plane waves that are uniformly distributed in azimuth for a given phase velocity c, which would imply that the real part $\Re(\gamma)$ of the complex coherence (RPCC) is given by $J_0(2\pi r/\lambda)$, where J_0 is a Bessel function of order zero and λ is the wavelength of the waves [*Harms*, 2015]. The second is the possibility that the wave field is composed of a single plane wave where an angle θ is the azimuth of the source relative to the station pair. This results in a RPCC of $\cos(2\pi\cos(\theta)r/\lambda)$. We can take the point at which $\Re(\gamma) = 0.5$ as a diagnostic point for this function. For an isotropic Rayleigh-wave field, this value is observed at a distance $r = \lambda/4$. On the other hand, for the plane wave case, the distance between the seismometers needs to be $r = \lambda/6$ to observe $\Re(\gamma) = 0.5$. Wavelengths of $> \lambda/6$ are possible in the case of seismometer pairs separated along different directions.

We plot the RPCC in Figure 3 at 0.2 Hz for the two days 154 and 191 of year 2015. These days are during the summer time, but we have checked that the following results also hold in the winter time. The plots show a bimodal distribution, which is a consequence of the directional dependence of the seismic field together with the directional non-uniformity of the seismic array. The directional dependence of the seismic field is expected from the known distribution of sources of oceanic microseisms observed at Homestake [*Harms et al.*, 2010]. Extending the lower envelope of the scattered points in the left of Figure 3 to a coherence value $\Re(\gamma) = 0.5$, we find for day 154 that the minimal distance with $\Re(\gamma) = 0.5$ is about 7 km, and about 3 km for day 191. Assuming isotropy, we can infer for day 154 a seismic speed of about $4 \cdot 0.2\,\text{Hz} \cdot 7\,\text{km} = 5.6\,\text{km/s}$, or 8.4 km/s assuming maximal directional dependence. The corresponding values for day 191 are 2.4 km/s and 3.6 km/s. While the speed values of day 191 are consistent with expected fundamental Rayleigh-wave speeds, the inferred speeds of day 154 are too high. In this way, we have used the RPCC to place constraints on the composition of the wavefield on these days.

The directional dependence of the seismic field only explains a variation of coherence values at fixed distance as a function of azimuth; it does not, however, explain the bimodal distribution. The latter can be explained by the directional non-uniformity of the seismic array. Almost all of the pairs in Figure 3 with horizontal distance > 2 km include a surface station since surface stations are generally located at a greater distance from the main underground array. Surface stations TPK, WTP, and LHS lie on a line pointing approximately along the E-W direction, while the line DEAD-SHL is almost perpendicular to it. Identifying seismometer pairs of the > 2 km coherence values, we find that SHL and DEAD appear in the high-coherence part while TPK, LHS, and WTP appear in the low-coherence part. This is consistent with a directional dependence of a seismic field consisting mainly of waves propagating along the E-W direction, and the bimodal structure is enforced by the approximate cross-shape of the surface array.

We can also exclude any significant impact from transient local sources at 0.2 Hz irrespective of whether they produce coherent or incoherent disturbances between stations. Observations covering the entire area of the US showed that speeds of fundamental Rayleigh waves with a 25 s period are about 3.6 km/s [*Foster et al.*, 2014]. Together with our results in Figure 1, we can infer that Rayleigh-wave speed at 0.2 Hz should have a value around 3 km/s, which means that the lengths of all types of waves are larger than the array dimension. We also checked that coherence does not decrease systematically when increasing correlation time from one day to one month or longer, which means that there are no significant incoherent disturbances that would average out over long periods of time. Next, we know from our observation of seismic spectra that local disturbances must be weaker than oceanic microseisms by a factor 10 or more since there is no disturbance visible even when oceanic microseisms are close to their minimum. Finally, if local sources had such a big effect on correlations, then they would have an equally significant effect on our Wiener filters (see following sections). However, this can be excluded since the Wiener filters prove to be highly efficient with the cancellation of oceanic microseisms (reduction by more than two orders of magnitude in most cases), which is only possible if the filter is almost fully determined by correlations consistent with oceanic microseisms. This is because the phases measured by the filters are most likely to be different than the microseism; in a case where a local source produced plane waves consistent in phase with microseisms, these would be subtracted as well.

These observations imply that during day 154, the dominant contribution to the seismic field comes from body waves, while Rayleigh waves dominate on day 191. Figure 4 shows the PSD at 0.2 Hz over one-year together with the minimal coherence observed between all seismometer pairs closer than 3 km to each other. The inset plot zooms onto the first 60 days. The expected coherence from an isotropic fundamental Rayleigh-wave field with a speed value of 3.5 km/s (among all plane-wave models, the isotropic model has the highest minimal coherence value) between two seismometers at 3 km distance to each other is 0.73 (assuming negligible instrumental noise). Coherence exceeds this value significantly during many days, and interestingly, a significant decline of coherence is always accompanied with a significant increase of the microseismic amplitude.

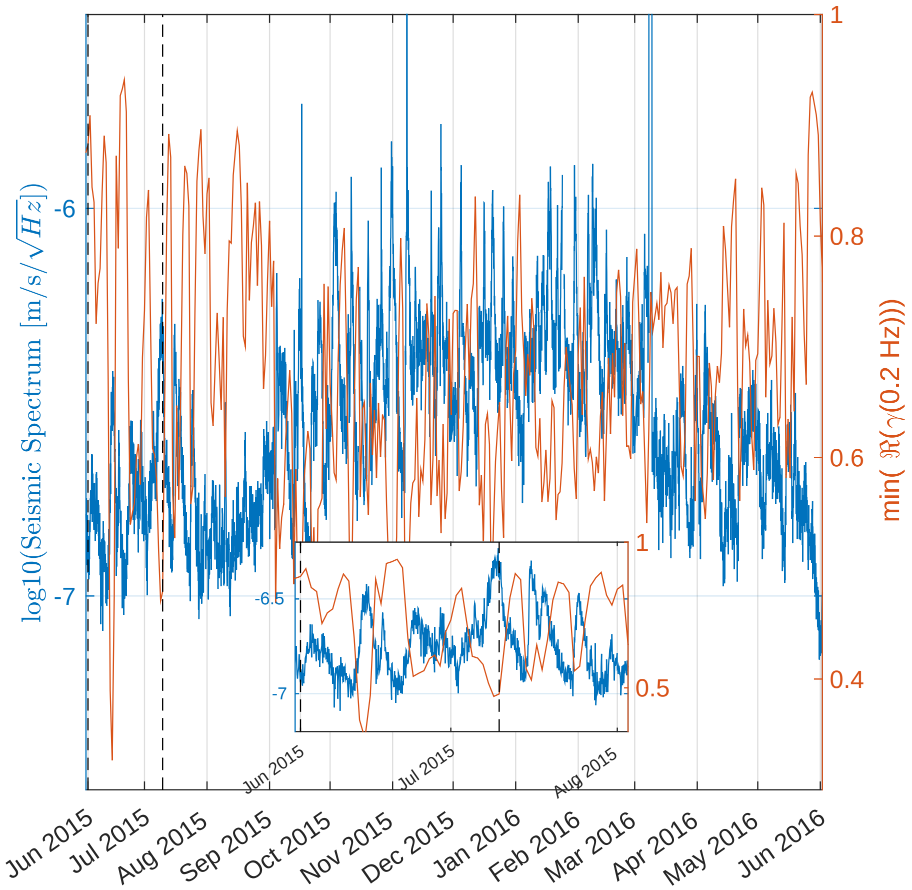

Figure 4. The plot shows the power spectral density (PSD) of the 800 ft station at 0.2 Hz and the minimum coherence among all station pairs whose distance is less than 3 km where the dashed vertical lines mark the two days used for the coherence plots corresponding to days 154 and 191 of year 2015.

2.3 Time-domain observations

To further confirm this observation, we show an alternate version of this analysis, which uses a time-domain cross-correlation based method to achieve the same result. Figure 5 shows the envelope of these correlations between one of the Homestake seismometers (station SHL) and a nearby instrument from the Global Seismograph Network (station RSSD) roughly 31 km away. The correlation for a given day is constructed by averaging hourly coherence measurements between the two vertical channels, including a time-domain running-mean normalization and a frequency-domain spectral whitening, both of which are common in the community to reduce the influence of earthquakes or other spurious noise sources [i.e., *Bensen et al.*, 2007]. The resulting correlation functions are bandpassed from 0.1 to 0.3 Hz. Both positive and negative lag times are plotted, corresponding to coherent signals traveling from RRSD to TPK or from TPK to RSSD, respectively. Horizontal red lines indicate the expected group arrival of surface waves (at either positive or negative correlation lag times) traveling at 3.5 km/s. While surface waves dominate in the winter months when 0.2 Hz microseism noise is strong, many times of the year are dominated by a very fast arrival that suggests rather body waves incident from below the two stations.

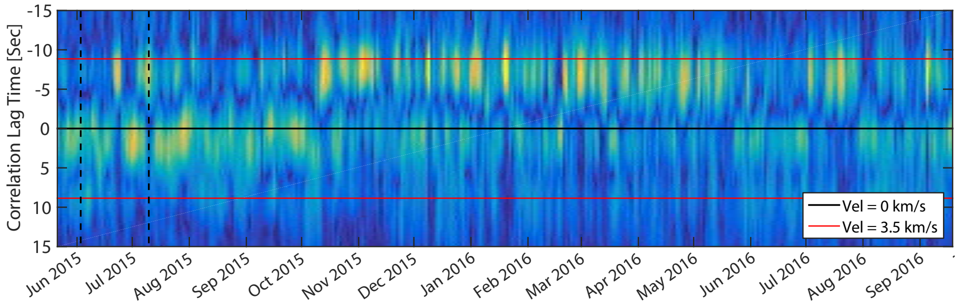

Figure 5. The plot shows the bandpassed noise-correlation function between one of the Homestake seismometers (SHL) and a nearby instrument from the Global Seismograph Network (RSSD) roughly 31 km away.

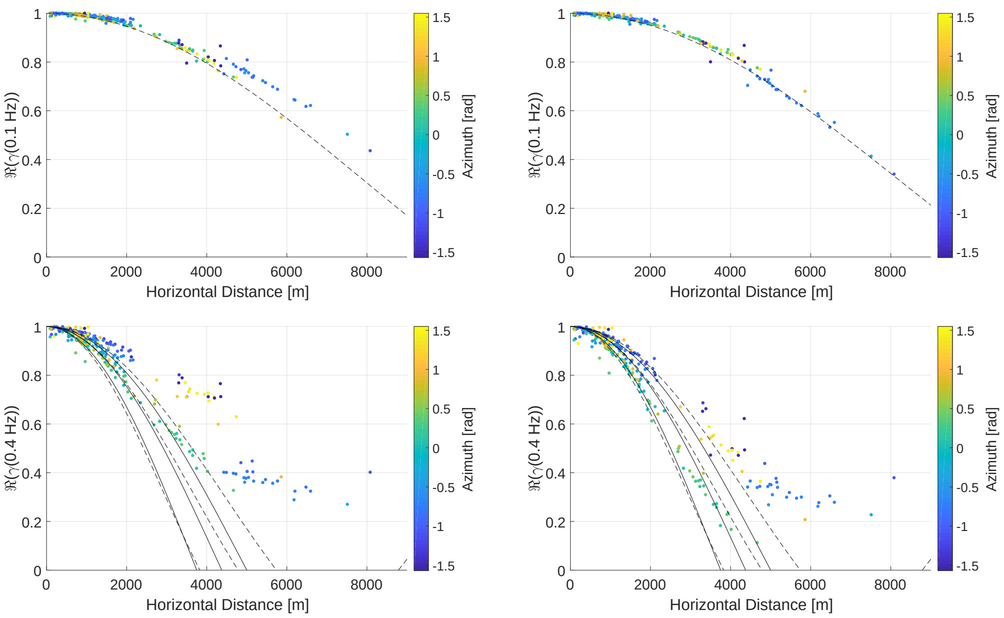

Figure 6. The RPCC as a function of distance between the vertical channels of all seismometers at 0.1 Hz (top row) and 0.4 Hz (bottom row) analogous to Figure 3. In the top row, an isotropic correlation model is shown with speed value 2.7 km/s (left plot), and 2.8 km/s (right plot), and in the bottom row, models are shown for the single plane wave (solid lines; speed values 6 km/s, 7 km/s, 8 km/s), and isotropic field (dashed lines; speed values 4 km/s, 5 km/s, 6 km/s).

2.4 Comparisons and discussion of wave content

To establish consistency with other analyses, we also plot the RPCC at 0.1 Hz and 0.4 Hz for the vertical channels. In the top row of Figure 6, we show the RPCC as a function of distance for 0.1 Hz, and in the bottom row, for 0.4 Hz. We can use the RPCC measurements to constrain seismic velocities at these frequencies as well. The seismic speeds, measured to be ≈ 3 km/s, are entirely consistent with fundamental Rayleigh waves at 0.1 Hz. There is no visible evolution between the days that were dominated by body waves and fundamental Rayleigh waves as in the case of 0.2 Hz. On the other hand, 0.4 Hz, at the high frequency end of the microseism, is significantly more complicated. It has contributions from both fundamental Rayleigh waves and body waves, and the trend is similar to that of 0.2 Hz.

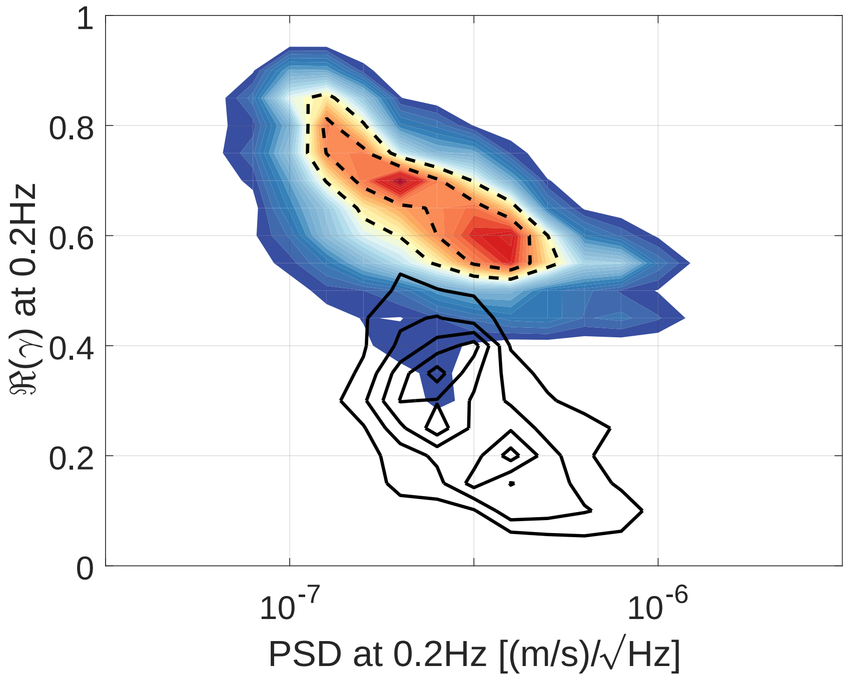

Figure 7. The plot shows the data in Figure 4 as a density plot for the Homestake array (colored contours with contour lines at 0.5 and 0.7) as well as the Sweetwater array (only contour lines at 0.1, 0.3, 0.5, 0.7, and 0.9.).

As we are confident that the observations relate to distant oceanic microseism sources, the following discussion is potentially generalizeable to other locations and arrays. To check that the anti-correlation between PSDs and minimal coherence at 0.2 Hz is not only present at Homestake, we performed the same analysis for the Sweetwater broadband array [*Barklage et al.*, 2014]. The seismometers in this analysis are from an array in Sweetwater, Texas, which is located at 32°28'5" N and 100°24'26" W. The array consists of two approximate circles, one with about a 10 km diameter, another with a 25 km diameter, with 23 stations with good data quality during March and April 2014. This array has significantly larger horizontal spacing than the Homestake array, with horizontal distances between the center of the array and other seismometers ranging between 2-14 km. It also has significant variation in elevation over the array, with a max elevation change between seismometers of about 250 m. We perform the same analysis with this array as in the Homestake case, computing the PSDs and coherences between the station pairs. Figure 7 shows $\Re(\gamma)$ vs. the PSDs for the Homestake and Sweetwater arrays. The difference in $\Re(\gamma)$ arises from the different sizes of the arrays.

This strongly points towards the following model of oceanic microseisms at Homestake at 0.2 Hz. When the oceanic microseisms are weak, i.e., approaching the global low-noise model, then the field is dominated by body waves. Typically week-long, strong transients of Rayleigh waves add to this background of body waves, decreasing RPCC values because of the slower velocities of fundamental Rayleigh waves. The existence of body waves in oceanic microseisms is well known and modeled [*Landès et al.*, 2010, *Obrebski et al.*, 2013]. However, the hypothesis that body waves define the microseismic spectrum at quiet times has not been formulated before to our knowledge. This link seems to exist at the Homestake site at least, and it would be very interesting to obtain direct confirmation using methods from [*Landès et al.*, 2010]. This method shows that it is possible to differentiate between fundamental Rayleigh waves and body wave contributions; this is a potentially important technique applicable in the field of time-dependent velocity measurement.

3 Testing recoverability of the wavefield with Wiener Filters

Given that the ambient noise field is constantly changing in direction and wavetype, there should be a limit in a given array's ability to describe the ambient noise wavefield. That is, should the ambient noise correlation functions from one time period be compared to that of another, there may exist some portion of the wavefield that must be attributed to random, variable processes that cannot be resolved given the geometry of instruments used. This section explores this for different array geometries by the construction of Wiener filters. The Wiener filter approach is in many ways similar to the work presented above, but rather than consider only two stations at a time, the Wiener filter simultaneously considers all available station-station coherencies from the array. A "target" station is defined, for which information from all others in the array are used to predict and subtract known signals. The extent to which signal remains after this subtraction at subsequent observation times indicates the lower limit to which coherence-based approaches can be reliably interpreted.

In previous work [*M Coughlin*, 2014], we implemented feed-forward noise cancellation using an array of 3 seismometers in the same general location as our current Homestake array [*Harms et al.*, 2010]. We used Wiener filters, which are the optimal linear filters to cancel noise of (wide-sense) stationary random processes defined in terms of correlations between witness and target sensors [*Vaseghi*, 2001]. We explored how to maximize subtraction, including exploring the rate at which the filters are updated and the number of filter coefficients. There were limits to this original study. Due to the fact that we only had three functional seismometers, we could not explore the effect of body waves on the coherence between the seismometers and thus the subtraction that we could achieve. In addition to the self-noise of the seismometers, topographic scattering and body waves in the seismic field could limit performance due to the filters only being able to subtract sources of noise that are always present [*Coughlin and Harms*, 2012].

The method is common in gravitational-wave studies, for which the interest is to use arrays of seismometers as witness sensors to the gravitational-wave interferometer to subtract the noise in the seismic field present in the detector. The idea of Wiener filtering is to make predictions of time-series of a single sensor (target sensor) based on observations of other sensors (witness sensors). Wiener filtering uses the correlation of all sensors of the array, including accounting for both correlations amongst the witness sensors and the target sensor, when making the predictions. This is different from beamforming techniques which predominantly depend on the coherence between only two stations and subsequently stack observations or estimate parameters based on a model. In general, studies of this type are of interest for extracting information about the intervening media. For example, amplitude relations between two stations could provide information about the intervening geologic structure, attenuation, and anisotropy. The high level of coherence between seismometers leads to the question of how well these predictions can be made. The Wiener filtered time-series, with the microseism removed, of are of interest for those studying the seismic environment not produced by this source. For example, one might straight-forwardly use them to find small earthquakes otherwise buried in the noise.

The method for computing the Wiener filters is as follows. For samples $y(t_i)$ from a single target channel, M input time series $\vec{x}(t_i) = (x_m(t_i))$ with $m = 1, \ldots, M$, and a Wiener filter $\vec{h}(i) = (h_m(i)), i = 0, \ldots, N$ that minimizes the residual error, the residual seismic time-series can be written symbolically as a convolution (symbol $*$) *Vaseghi* [2001]:

$$r(t_i) = y(t_i) - \sum_{m=1}^{M} (h_m * x_m)(t_i), \tag{3}$$

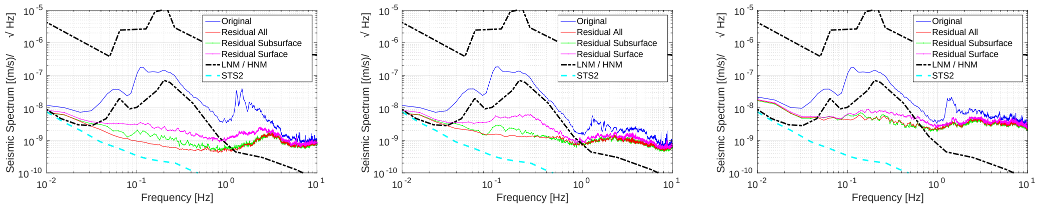

Figure 8. The plot on the left is the subtraction achieved using the seismometer on the 800 ft level as the target channel, up to 10 Hz. The middle and the right plots are the same for the seismometers on the 4850 ft level and the surface, respectively. In each plot, it is shown how the subtraction varies depending on what set of seismometers are used as witness sensors (subsurface, surface, and all). The dashed black lines correspond to Peterson's high and low noise models *Peterson* [1993]. The residual noise can be compared with the STS-2 sensor noise.

where the convolution is defined as

$$h_m * x_m(t_i) = \sum_{k=0}^{N} h_m(k) x_m(t_{i-k}), \tag{4}$$

where N is the order of the finite impulse-response filter h.

It is useful to compare the measured residuals to expected estimates. These can be computed as follows. If we denote C_{SS} as the matrix containing the cross spectral densities of witness seismometers, $\vec{C}_{\mathrm{ST}}$ as the vector containing the cross spectral densities between the witness and target sensors, and $\vec{C}_{\mathrm{TT}}$ as the PSD of the target seismometer, then the average relative noise residual R achieved is given by

$$R(f) = 1 - \frac{\vec{C}_{\mathrm{ST}}^{\mathrm{T}}(f) \cdot C_{\mathrm{SS}}^{-1}(f) \cdot \vec{C}_{\mathrm{ST}}(f)}{C_{\mathrm{TT}}(f)}. \tag{5}$$

When using just a single witness seismometer, this simply reduces to

$$R(f) = 1 - |\gamma(f)|^2 \tag{6}$$

where $\gamma(f)$ is the witness-target coherence as defined in equation (1).

In the following analysis, we will use a seismometer at the center of the array as our target and the remaining seismometers as witnesses. In Figure 8, we demonstrate the performance of the filter on the seismic array data using as targets the vertical channels of three seismometers on the 800 ft level, the 4850 ft level, and the surface. We achieve more than a factor of 100 reduction in noise at the microseism peak when using all available channels. To say this another way, we can predict the seismic time-series of the target sensor to better than 1%. We can also explore the loss in information from using only surface stations when measuring the seismic wave-field below ground. Using only surface stations as witness channels is worse than the configuration where all channels are used by a factor of ≈ 4; sub-1% prediction of the underground seismic wavefield is not possible with only surface sensors.

To exclude the possibility that the improvement is simply an increase in the number of channels, we show on the right of Figure 9 that the expected performance of the Wiener filter rapidly converges as a function of the number of witness sensors. We use equation (5) to determine the expected residuals for a few optimal subsets of seismometers taken from the total array. Optimal subsets are the ones that, given a number of seismometers, produce lowest subtraction residuals. Generally, there is no clearly visible residual microseismic peak except for the case of using surface seismometers as input channels to cancel noise in a 4850 ft seismometer (right plot in Figure 8). So we were able to improve over previous results reported

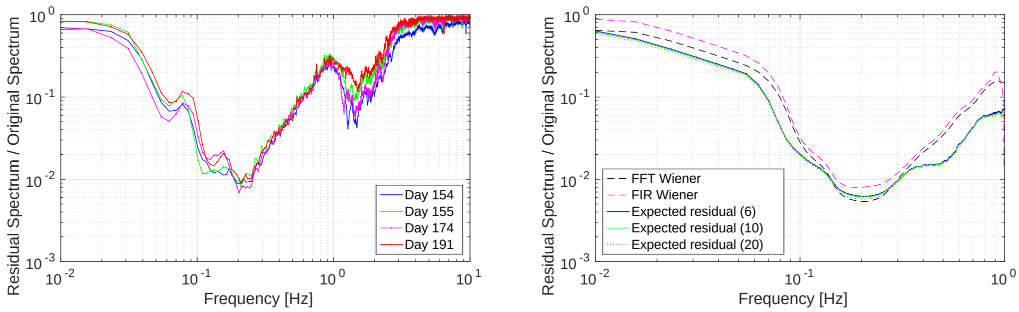

Figure 9. On the left, we show the performance of the Wiener filter over a few timescales using the vertical channel of the 800 ft station seismometer as the target. This result shows that Wiener filters are efficient in this band over long timescales. On the right, we show the expected residuals given the expression in equation (5) for a number of seismometer arrays and comparisons to both FFT Wiener and FIR Wiener filters for the vertical channel of the 800 ft station as the target channel.

in [*M Coughlin*, 2014], almost reaching the limit set by the sensor noise of the Kinematics STS-2 broadband seismometers up to 1 Hz used at the 800 ft station [1].

The right of figure 9 also shows that the achieved subtraction is in line with the expected residuals, indicating the efficacy of our implementation. Noise residuals are computed for two different implementations of Wiener filters. One is the frequency-domain filter. The other is the finite-impulse response (FIR) filter applied as shown in equations (3) and (4). The frequency-domain filter typically achieves slightly better cancellation performance since noise in neighboring frequency bins is only weakly correlated, and this correlation can be ignored simplifying the filter. The FIR filter, which is applied in time domain, has to cope with strong correlations potentially between all samples of the time series. This makes it numerically more challenging to calculate the Wiener filter mostly due to large, degenerate correlation matrices, which need to be inverted. In our case, differences between the performances of these two implementations are minor.

The left of figure 9 explores the efficiency of a Wiener filter applied to data on various time-scales. In general, a loss of up to a factor of 2 in the predictive power of the filter can be seen on month-long timescales. A loss in performance is unsurprising given the changing composition of the seismic field, but the relatively minimal loss in performance indicates that in general, the body-wave vs. fundamental Rayleigh wave content does not have a significant impact on the phase of the correlations measured between the seismometers (which is what determine the composition of the filters).

In summary, we can use this method to determine that the underground seismometers significantly increases the accuracy of the measurement of the underground wavefield. Measurements of this type show the utility of including underground seismometers in future arrays dedicated to time-dependent velocity measurements, allowing predictions at the 1% level of the wave-field (i.e., 8), whereas constraining observations to surface stations we are left with at least a 4% level residual.

The Wiener filter can be considered comparable to other coherence or cross-correlation type observations for the time at which it was trained. The filter applied at any other time period should then perform equally well if all aspects of the environment remained constant.

[1] STS-2 were used everywhere in the array except for stations DEAD, ROSS, YATES, 300, where Güralp 3T were deployed

When applied to another time period, the fact that some residual remains between prediction and the actual observation implies that either the intervening medium has changed, or that changes in the ambient noise field cannot be resolved by the array. In our case, we assume that any material velocity changes would be relatively constant over the extent of the different sub-arrays, but still find that different geometries used produces different residuals. The fact that surface-only 2D observations, for example, cannot describe more than 96% of the waveform implies that there is an upper limit to what we can expect to resolve or explain; that last 4% may be considered a random level of variability given the geometry used.

4 Implications for gravitational-wave observations

This work, and even the deployment of the Homestake array [*Mandic et al.*, 2017], is additionally motivated by open questions in the gravitational-wave community because vibrations in the Earth's crust are a significant source of noise in gravitational-wave observatories. In addition to the coupling of ground motion to the interferometers through the suspension systems, fluctuations in the gravitational field at the mirrors are another source of noise, referred to as Newtonian Noise (NN). This is important as with the recent detections of gravitational waves from binary black holes [*Abbott, B. P. et al.*, 2016] and binary neutron stars [*Abbott, B. P. et al.*, 2017], there is significant interest in the development of new technologies for improving sensitivity; these improvements may take the form either as upgrades of the existing gravitational-wave detectors or for future-generation detectors. Sophisticated seismic-isolation systems are used in order to limit the effect of seismic disturbances in gravitational-wave detectors [*F. Matichard et al.*, 2015, *S. Braccini et al.*, 2005]. Vibrations in the Earth hinder the precise measurements needed by gravitational-wave detectors, both in the form of ground motions that can be suppressed and in the form of gravity fluctuations as densities are redistributed. While sophisticated seismic-isolation systems are used in order to limit the effect of seismic disturbances [*F. Matichard et al.*, 2015, *S. Braccini et al.*, 2005], fluctuations in the gravitational field at the test mass from local seismic noise and temperature and pressure fluctuations in the atmosphere will be a future limiting noise source below about 20 Hz [*Saulson*, 1984, *Hughes and Thorne*, 1998, *Creighton*, 2008, *Harms*, 2015]. The Wiener filters, combined with knowledge of the body wave type, can be used to determine the NN contribution and mitigate its effects.

Understanding a seismic field in terms of its two-point spatial correlations, i.e., estimated from correlations between two seismometers, is fundamental to the understanding of NN and its cancellation [*Harms*, 2015]. Understanding the wave content of oceanic microseisms is of high priority for sub-Hz GW detectors where seismic fields produce NN about 1000 times stronger than the instrumental noise required to detect GWs [*McManus et al.*, 2017]. These measurements have significant implications for NN cancellation for potential future low-frequency gravitational-wave detectors. The assumption so far has been that the seismic field is dominated by Rayleigh waves, which greatly helps with the cancellation of the associated NN using off-line Wiener filter subtraction [*Harms and Paik*, 2015]. Given that NN cancellation in the presence of multiple wave polarizations is a complicated task even for modest cancellation goals [*Harms*, 2015], continuous body-wave content as observed at Homestake would be a substantial additional challenge for plans to suppress seismic NN at sub-Hz frequencies by large factors. Subtraction at the level of 1 % and below do give confidence though that in the case of body-wave and fundamental Rayleigh wave separation, significant mitigation of NN is possible. Such capabilities are essential to realize cancellation of terrestrial gravity noise in future gravitational-wave detectors.

5 Conclusion

In this paper, we have used one year of data from an underground and surface array deployed in 2015 at the Sanford Underground Research Facility (former Homestake mine) for correlation analyses of the ambient seismic field. The results include the year-long evolution of spectral density and seismometer correlations at 0.2 Hz and the broadband cancellation of seismic signals in the array using Wiener filters. The long-term study of PSDs and correlations at 0.2 Hz showed evidence of an incessant background of body waves frequently perturbed by week-long Rayleigh-wave transients. These findings are consistent with previous observations, but our findings go beyond previous results as the body-wave content seems to enforce the low-noise model at the Homestake site. This link has not been established before to our knowledge. Finally, while it has been previously known that array geometry plays an important role in a method's ability to resolve and recover the ambient noise field, our application of Wiener filters allows us to quantify the lower limit of this recovery. These Wiener filters are used to estimate and cancel seismic signals in a target sensor using data from other stations in the array, reducing seismic signals by more than 2 orders of magnitude. By comparing the estimate and residual of different subarrays we find that this can be improved by a factor of 4 by including underground stations to better capture the entire ambient noise wavefield.

Acknowledgments

MC was supported by the David and Ellen Lee Postdoctoral Fellowship at the California Institute of Technology. We thank the staff at the Sanford Underground Research Facility and PASSCAL for assistance, particularly the help of Tom Regan, Jaret Heise, Jamey Tollefson, and Bryce Pietzyk. This work was supported by National Science Foundation INSPIRE grant PHY1344265. This paper has been assigned LIGO document number LIGO-P1700422.

References

Abbott, B. P. et al. (2016), Binary black hole mergers in the first advanced ligo observing run, *Phys. Rev. X*, *6*, 041,015, doi:10.1103/PhysRevX.6.041015.

Abbott, B. P. et al. (2017), Gw170817: Observation of gravitational waves from a binary neutron star inspiral, *Phys. Rev. Lett.*, *119*, 161,101, doi:10.1103/PhysRevLett.119.161101.

Aki, K. (1957), Space and Time Spectra of Stationary Stocastic Waves, with Special Reference to Microtremors., *Bulletin of Earthquake Research Institute*, *35*, 415–457.

Barklage, M., D. Hollis, J. M. Gridley, R. Woodward, and N. Spriggs (2014), A Large-N Mixed Sensor Active + Passive Seismic Array near Sweetwater, TX, *AGU Fall Meeting Abstracts*.

Bensen, G. D., M. H. Ritzwoller, M. P. Barmin, A. L. Levshin, F. Lin, M. P. Moschetti, N. M. Shapiro, and Y. Yang (2007), Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements, *Geophysical Journal International*, *169*, 1239–1260, doi:10.1111/j.1365-246X.2007.03374.x.

Brenguier, F., N. M. Shapiro, M. Campillo, V. Ferrazzini, Z. Duputel, O. Coutant, and A. Nercessian (2008), Towards forecasting volcanic eruptions using seismic noise, *Nature Geoscience*, *1*(2), 126–130, doi:10.1038/ngeo104.

Capon, J. (1969), High-resolution frequency-wavenumber spectrum analysis, *Proceedings of the IEEE*, *57*(8), 1408–1418, doi:10.1109/PROC.1969.7278.

Coughlin, M., and J. Harms (2012), Seismic topographic scattering in the context of GW detector site selection, *Classical and Quantum Gravity*, *29*, 075,004.

Creighton, T. (2008), Tumbleweeds and airborne gravitational noise sources for LIGO, *Class. Quantum Grav.*, *25*, 125,011.

Cupillard, P., and Y. Capdeville (2010), On the amplitude of surface waves obtained by noise correlation and the capability to recover the attenuation: A numerical approach, *Geophysical Journal International*, *181*(3), 1687–1700, doi:10.1111/j.1365-246X.2010.04586.x.

F. Matichard et al. (2015), Advanced LIGO two-stage twelve-axis vibration isolation and positioning platform. Part 2: Experimental investigation and tests results, *Precision Engineering*, *40*, 287 – 297, doi:http://dx.doi.org/10.1016/j.precisioneng.2014.11.010.

Feroz F., Hobson M.P., and Bridges M. (2009), Multinest: an efficient and robust bayesian inference tool for cosmology and particle physics, *Mon. Not. Roy. Astron. Soc.*, *398*, 1601–1614.

Foster, A., G. EkstrÃűm, and M. Nettles (2014), Surface wave phase velocities of the Western United States from a two-station method, *Geophysical Journal International*, *196*(2), 1189–1206, doi:10.1093/gji/ggt454.

Gerstoft, P., P. M. Shearer, N. Harmon, and J. Zhang (2008), Global p, pp, and pkp wave microseisms observed from distant storms, *Geophysical Research Letters*, *35*(23), n/a–n/a, doi:10.1029/2008GL036111, l23306.

Harms, J. (2015), Terrestrial Gravity Fluctuations, *Living Reviews in Relativity*, *18*(3), doi:10.1007/lrr-2015-3.

Harms, J., and H. J. Paik (2015), Newtonian-noise cancellation in full-tensor gravitational-wave detectors, *Phys. Rev. D*, *92*, 022,001, doi:10.1103/PhysRevD.92.022001.

Harms, J., F. Acernese, F. Barone, I. Bartos, M. Beker, J. van den Brand, N. Christensen, M. Coughlin, R. DeSalvo, S. Dorsher, et al. (2010), Characterization of the seismic environment at the Sanford Underground Laboratory, South Dakota, *Classical and Quantum Gravity*, *27*(22), 225,011.

Hughes, S. A., and K. S. Thorne (1998), Seismic gravity-gradient noise in interferometric gravitational-wave detectors, *Phys. Rev. D*, *58*, 122,002.

Landès, M., F. Hubans, N. M. Shapiro, A. Paul, and M. Campillo (2010), Origin of deep ocean microseisms by using teleseismic body waves, *Journal of Geophysical Research: Solid Earth*, *115*(B5), n/a–n/a, doi:10.1029/2009JB006918, b05302.

Lawrence, J. F., and G. A. Prieto (2011), Attenuation tomography of the western united states from ambient seismic noise, *Journal of Geophysical Research: Solid Earth*, *116*(B6), n/a–n/a, doi:10.1029/2010JB007836, b06302.

Longuet-Higgins, M. S. (1950), A theory of the origin of microseisms, *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *243*(857), 1–35, doi:10.1098/rsta.1950.0012.

Longuet-Higgins, M. S., and F. Ursell (1948), Sea waves and microseisms, *Nature*, *162*(4122), 700–700, doi:10.1038/162700a0.

M Coughlin, e. a., J Harms (2014), Wiener filtering with a seismic underground array at the Sanford Underground Research Facility, *Classical and Quantum Gravity*, *31*, 215,003.

Mandic, V., V. C. Tsai, G. L. Pavlis, T. Prestegard, D. C. Bowden, P. Meyers, and R. Caton (2017), Three-dimensional, Broadband Seismometer Array Experiment at the Homestake Mine, *Submitted to Seismological Research Letters*.

McManus, D. J., P. W. F. Forsyth, M. J. Yap, R. L. Ward, D. A. Shaddock, D. E. McClelland, and B. J. J. Slagmolen (2017), Mechanical characterisation of the TorPeDO: a low frequency gravitational force sensor, *Classical and Quantum Gravity*, *34*(13), 135,002.

Obrebski, M., F. Ardhuin, E. Stutzmann, and M. Schimmel (2013), Detection of microseismic compressional (P) body waves aided by numerical modeling of oceanic noise sources, *Journal of Geophysical Research: Solid Earth*, *118*(8), 4312–4324, doi:10.1002/jgrb.50233.

Peterson, J. (1993), Observation and modeling of seismic background noise, *Open-file report*, *93-322*.

Rost, S., and C. Thomas (2002), Array seismology: Methods and applications, *Reviews of Geophysics*, *40*(3), doi:10.1029/2000RG000100.

S. Braccini et al. (2005), Measurement of the seismic attenuation performance of the VIRGO Superattenuator, *Astroparticle Physics*, *23*(6), 557 – 565, doi:http://dx.doi.org/10.1016/j.astropartphys.2005.04.002.

Saulson, P. R. (1984), Terrestrial gravitational noise on a gravitational wave antenna, *Phys. Rev. D*, *30*, 732.

Tsai, V. C. (2009), On establishing the accuracy of noise tomography travel-time measurements in a realistic medium, *Geophysical Journal International*, *178*(3), 1555–1564, doi:10.1111/j.1365-246X.2009.04239.x.

Tsai, V. C. (2011), Understanding the amplitudes of noise correlation measurements, *Journal of Geophysical Research*, *116*(B9), B09,311, doi:10.1029/2011JB008483.

Vaseghi, S. V. (2001), *Wiener Filters*, pp. 178–204, John Wiley & Sons, Ltd.

Viens, L., M. Denolle, H. Miyake, S. Sakai, and S. Nakagawa (2017), Retrieving impulse response function amplitudes from the ambient seismic field, *Geophysical Journal International*, *210*, 210–222, doi:10.1093/gji/ggx155.

Weaver, R. L. (2011), On the amplitudes of correlations and the inference of attenuations, specific intensities and site factors from ambient noise, *Comptes Rendus - Geoscience*, *343*(8-9), 615–622, doi:10.1016/j.crte.2011.07.001.