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 Mandic[5],G Pavlis[6],T Prestegard[5],V.C. Tsai[4]

[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

Ambient seismic noise is a rich source of information commonly used to study the Earth. The oceanic microseisms, generated by the interaction between ocean waves and both the seafloor and coastlines, is the strongest source of this ambient seismic noise. It is sometimes important for these studies to understand the wave composition of these microseisms. Using a dense seismic array at the Sanford Underground Research Facility in the former Homestake mine in South Dakota, we use correlations between seismometers to investigate the composition of the wave field. Specifically, we find evidence that body waves dominate oceanic microseisms at 0.2 Hz during times when the spectral density approaches the global low-noise model; this body wave background is then interrupted by typically week-long fundamental Rayleigh-wave transients. An idea closely related to seismic correlations is that of "Wiener filters," which are optimal linear filters designed to cancel noise defined in terms of correlations between sensors. Using Wiener filters, we show the significant benefits of underground seismic wave field measurements, making predictions of array sensor readings at the better than 1% level. 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 efforts have been made in the wider seismological community to understand and exploit background ambient seismic noise. The strongest 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*, 1950]. These mechanisms most strongly generate surface waves 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 *Gerstoft et al.* [2008]. Efforts to image these noise sources use array processing methods that consider the coherence of wavefronts incident upon the array, referred to as beamforming or k-f analysis [*Capon*, 1969], similar in many ways to the efforts here. Special attention has been paid to understanding the effect that the inhomogeneous distribution of noise sources would have on measured coherence or cross-correlation [*Tsai*, 2011], with the goal of determining whether such measurements can be reliably used for the study of seismic velocities [*Bensen et al.*, 2007] or attenuation [*Lawrence and Prieto*, 2011].

One of the advantages to interdisciplinary study is the relevance of techniques in other fields, in this case the gravitational-wave community, which can contribute to these efforts. 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], interest is being renewed in the development of new technologies for improving sensitivity, either as upgrades of the existing gravitational-wave detectors or for future-generation detectors. The network of advanced gravitational-wave detectors consists of the two Advanced LIGO interferometers in the United States [*J Aasi et al*, 2015], the Advanced Virgo interferometer in Italy [*F Acernese et al.*, 2015], GEO-HF in Germany [*Lück et al.*, 2010], the KAGRA interferometer in Japan [*Aso et al.*, 2013], and the LIGO-India detector in India [*Unnikrishnan*, 2013]. In the future, third-generation detectors, such as Cosmic Explorer [*B P Abbott et al.*, 2017] and the Einstein Telescope [*M Punturo et al.*, 2010], will be an order of magnitude more sensitive than these. Sophisticated seismic-isolation systems are used in order to limit the effect of seismic disturbances. The Advanced LIGO seismic isolation is a combination of active and passive isolation stages [*F. Matichard et al.*, 2015], while the Advanced Virgo seismic isolation is a mostly passive system [*S. Braccini et al.*, 2005]. Additional suppression of seismic noise can be achieved with feed-forward cancellation [*Kokeyama et al.*, 2014]. While these systems limit the effect of ground vibrations disturbing the detector sensitivity by coupling through ground motion, fluctuations in the gravitational field at the test mass from local seismic noise and temperature and pressure fluctuations in the atmosphere, which are known as Newtonian

noise (NN), will be a future limiting noise source below about 20 Hz [*Saulson*, 1984, *Hughes and Thorne*, 1998, *Creighton*, 2008, *Harms*, 2015]. For this reason, future gravitational-wave detectors are likely to be built at sites with low levels of ambient seismic noise [*Harms et al.*, 2010] or constructed underground [*M Punturo et al.*, 2010]. Feed-forward techniques similar to those used for seismic disturbances can also be used to subtract NN [*Cella*, 2000]. Feed-forward cancellation employs "Wiener filters," which are optimal linear filters designed to cancel noise defined in terms of correlations between sensors. We will discuss below how Wiener filtering is complementary to typical cross-correlation and beamforming techniques below.

For this analysis, we will use data from a new array installed at the Homestake mine. The Sanford Underground Research Facility in the former Homestake mine in South Dakota and surrounding land contained an array of 24 seismometers studying the properties of the underground seismic environment [*Mandic et al.*, 2017]. In the new seismic array, there are 24 seismometers, 15 underground and 9 above ground [*Mandic et al.*, 2017]. These broad-band instruments are sensitive to seismic noise between about 10 mHz and 50 Hz, which covers both the primary and secondary microseism appearing between 30 – 100 mHz, and 0.1 – 0.5 Hz respectively, as well as the anthropogenic band between 1 Hz and a few tens of Hz. In this array, the horizontal distances between some of the seismometers cover more than 6000 m, which is about 5 times the extent of the original array, while the vertical distances extend about 1500 m.

Previously, the Homestake mine was identified as a world-class, low-noise environment [*Harms et al.*, 2010, *M Coughlin*, 2014, *Mandic et al.*, 2017]. In previous work [*M Coughlin*, 2014], we implemented feed-forward noise cancellation using an array of 3 seismometers located at the Sanford Underground Research Facility in the Black Hills of South Dakota [*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].

In section 2, we present correlation measurements at the Homestake site, which lead to conclusions about the composition of the wave-field at the mine. In section 3, the Wiener filtering method is described, and we show how Wiener filter measurements provide predictive power in the wave field at the better than 1% level. Our conclusions are summarized in section 4.

2 Correlation Measurements

In this analysis, we improve upon the analysis of *M Coughlin* [2014] using a significantly expanded seismic array. We use a correlation method to extract information from the ambient seismic field. The first step was 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.

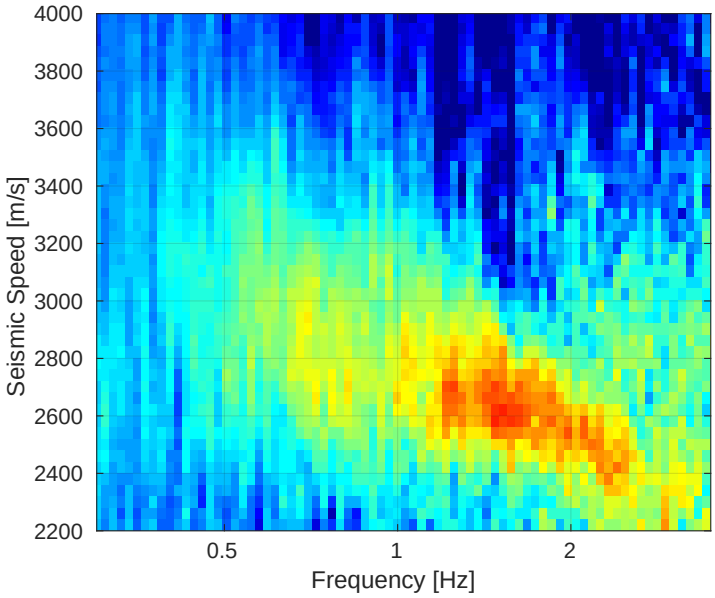


Figure 1. The plot shows a histogram of seismic speeds between 0.3 – 3.5 Hz. Red color means that the respective speed value was measured for a large number of k-f maps, while blue color means that the speed value was measured rarely.

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. 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.

We will use two different coherence-based analyses of the data lead to make seismic velocity measurements and therefore inference about the type of wave propagating. The first uses "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. The second uses a comparison of measured coherences to models for the seismic field, and inferences about seismic velocities are made using these model comparisons. For the speed measurements, 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

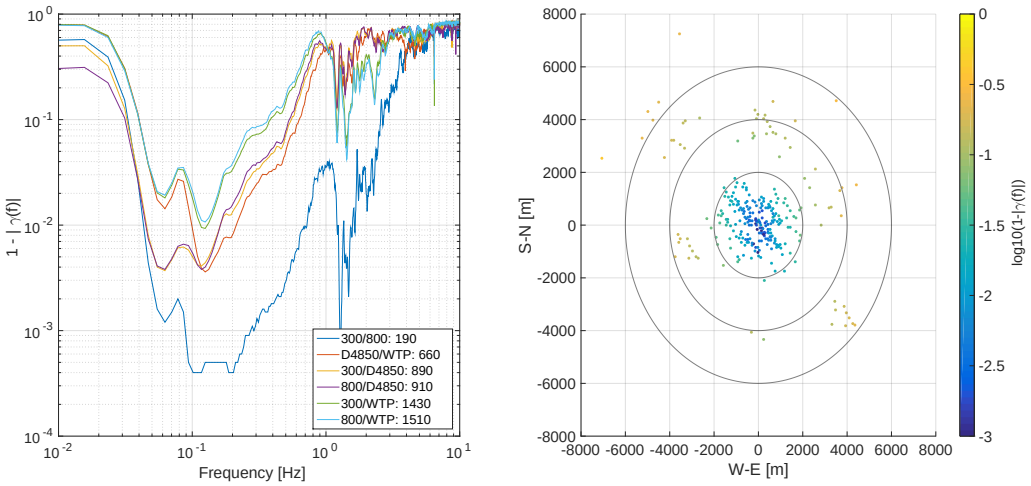


Figure 2. Left plot: $1 - |\gamma(f)|$ between a variety of seismometer pairs averaged over 6 months of coincident data divided into 128 s segments. The legend indicates the horizontal distance in meters between each pair shown, and the pairs are shown in ascending order of horizontal distance. Right 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.

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.

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, similar coherence measurements allow us to further investigate the wavefield at lower frequencies.

We now explore using correlation measurements to make conclusions about the composition of the Homestake wave field. Let us start by taking a look at the complex coherence $\gamma(f)$ defined in equation (1) between seismometers of the Homestake array. Here, coherence was calculated with 50% overlap, and in this form also used later for the Wiener filter. The left plot in Figure 2 shows the difference $1 - |\gamma(f)|$. 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

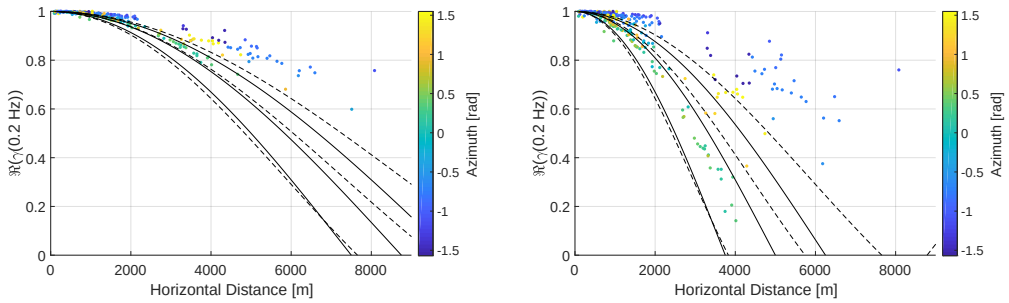


Figure 3. The RPCC as a function of distance between the vertical channels of all seismometers at 0.2 Hz (left: day 154 of 2015; right: day 191 of 2015). The colors correspond to the azimuth with respect to the east direction of the line connecting two seismometers. On the left, models are shown for 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).

the seismometer pairs are shown in the legend. At most frequencies, the shorter the horizontal distance, the higher the coherence. The 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.

We now show how we can use coherence to place constraints on the seismic velocities, and therefore on the composition of the wavefield. To do so, we will explore two orthogonal models for the wave field. The first is under the assumption that it is 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.

To give confidence in the method, we now show an alternate version of this analysis, which uses a time-domain cross-correlation based method to achieve the same result. A similar conclusion to that derived from RPCC calculations can be reached from inspection of bandpassed noise-correlation functions in the time domain. Figure 4 shows the envelope of these correlations between one of the Homestake seismometers (SHL) and a nearby instrument from the Global Seismograph Network (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 geophysics community *Bensen et al.* [2007] to reduce the influence of earthquakes or other spurious noise sources. 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

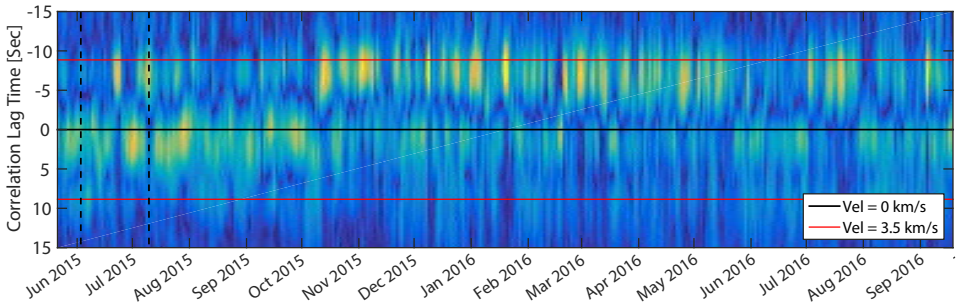


Figure 4. The plot shows the bandpassed noise-correlation function between one of the Homestake seismometers (SHL) and a nearby instrument from the Global Seismograph Network (RSSD) roughly 31 km away.

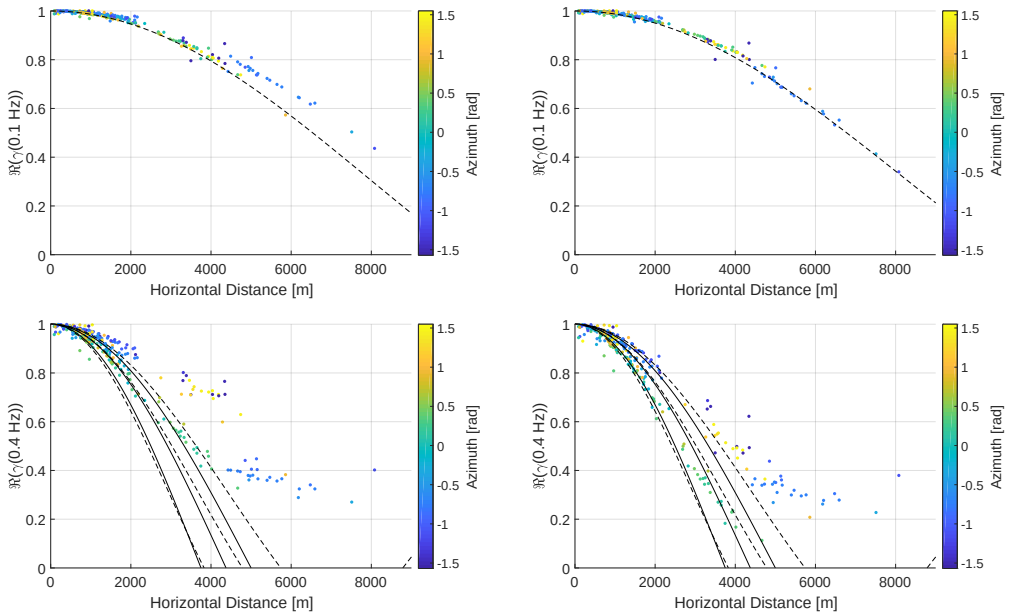
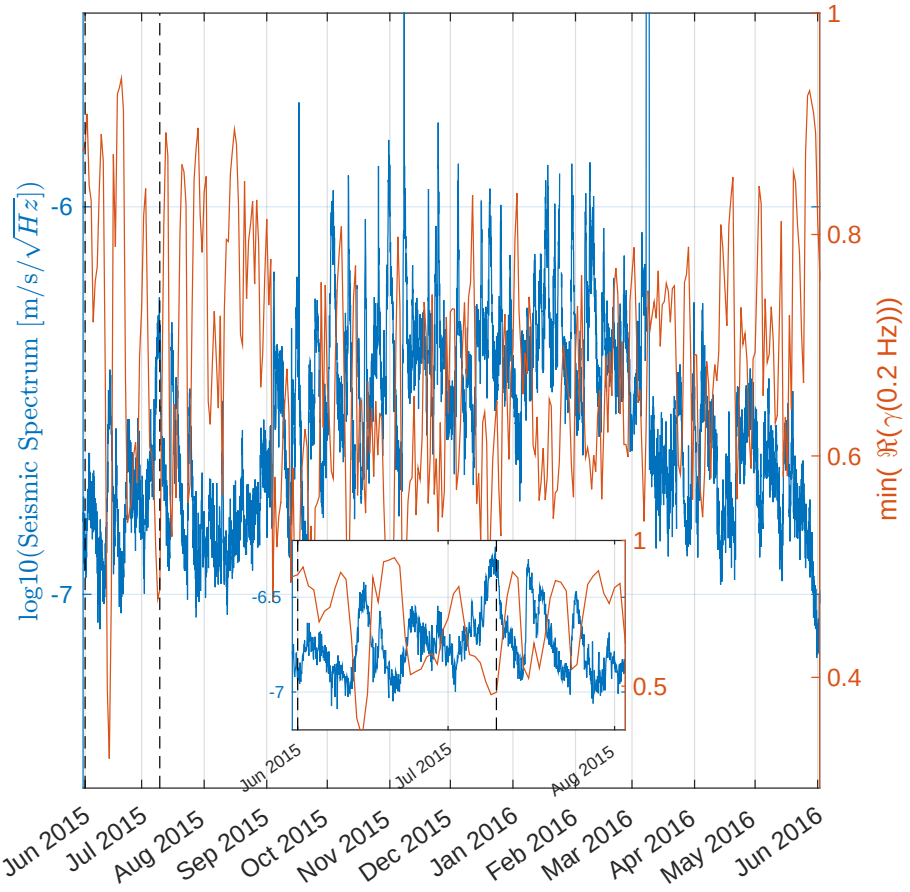


Figure 5. The RPCC as a function of distance between the vertical channels of all seismometers at 0.1 Hz (top row) and 0.4 Hz (bottom row) analogous to Figure 3. In the top row, an isotropic correlation model is shown with speed value 2.7 km/s (left plot), and 2.8 km/s (right plot), and in the bottom 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).

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.

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 5, 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 fun-



290 **Figure 6.** The plot shows the power spectral density (PSD) of the 800 ft station at 0.2 Hz and the minimum
291 coherence among all station pairs whose distance is less than 3 km where the dashed vertical lines mark the
292 two days used for the coherence plots corresponding to days 154 and 191 of year 2015.

287 damental Rayleigh waves as in the case of 0.2 Hz. On the other hand, 0.4 Hz, at the high fre-
288 quency end of the microseism, is significantly more complicated. It has contributions from
289 both fundamental Rayleigh waves and body waves, and the trend is similar to that of 0.2 Hz.

293 These observations imply that during day 154, the dominant contribution to the seis-
294 mic field comes from body waves, while Rayleigh waves dominate on day 191. Figure 6
295 shows the PSD at 0.2 Hz over one-year together with the minimal coherence observed be-
296 tween all seismometer pairs closer than 3 km to each other. The inset plot zooms onto the
297 first 60 days. The expected coherence from an isotropic fundamental Rayleigh-wave field
298 with a speed value of 3.5 km/s (among all plane-wave models, the isotropic model has the
299 highest minimal coherence value) between two seismometers at 3 km distance to each other
300 is 0.73 (assuming negligible instrumental noise). Coherence exceeds this value significantly
301 during many days, and interestingly, a significant decline of coherence is always accompa-
302 nied with a significant increase of the microseismic amplitude.

306 As we are confident that the observations relate to distant oceanic microseism sources,
307 the following discussion is potentially generalizeable to other locations and arrays. To check
308 that the anti-correlation between PSDs and minimal coherence at 0.2 Hz is not only present
309 at Homestake, we performed the same analysis for the Sweetwater array *Barklage et al.* [2014].

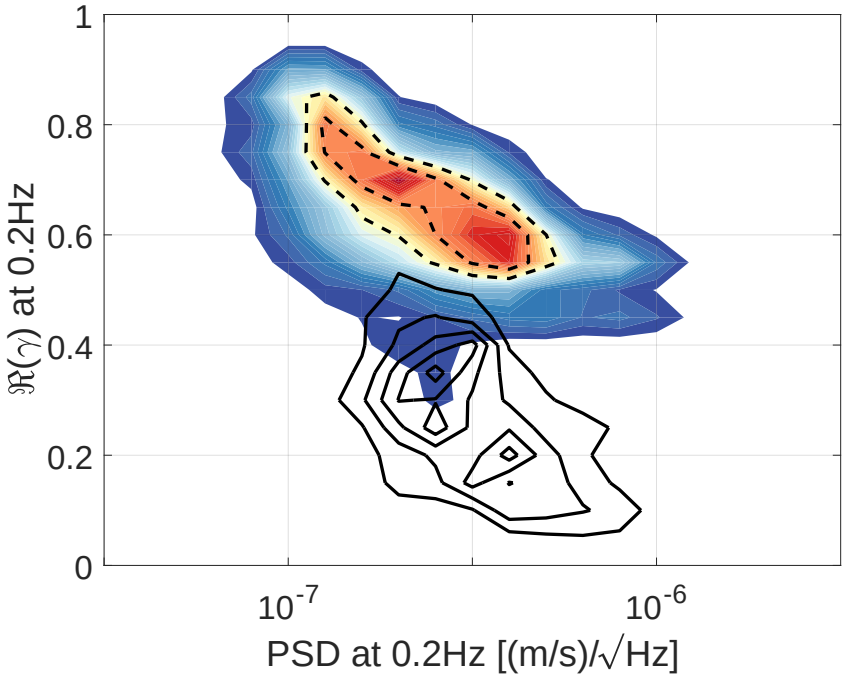


Figure 7. The plot shows the data in Figure 6 as a density plot for the Homestake array (colored contours with contour lines at 0.5 and 0.7) as well as the Sweetwater array (only contour lines at 0.1, 0.3, 0.5, 0.7, and 0.9.).

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].

3 Wiener Filtering

We now borrow a method common in gravitational-wave studies to further explore the coherence between seismometers in an array. In the gravitational-wave case, 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]:

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

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 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 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. In other words, a sub-1% prediction of the underground seismic wavefield is not possible with only surface sensors.

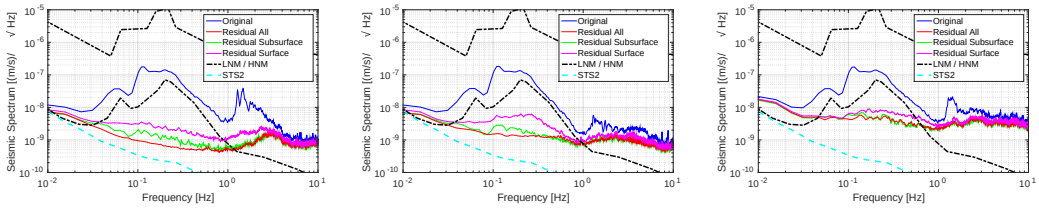


Figure 8. The plot on the left is the subtraction achieved using the seismometer on the 800 ft level as the target channel, up to 10 Hz. The middle and the right plots are the same for the seismometers on the 4850 ft level and the surface, respectively. In each plot, it is shown how the subtraction varies 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.

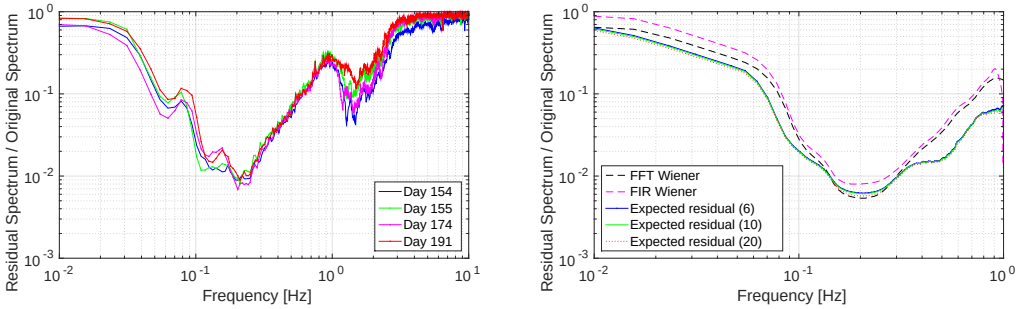


Figure 9. On the left, we show the performance of the Wiener filter over a few timescales using the vertical channel of the 800 ft station seismometer as the target. This result shows that Wiener filters are efficient in this band over long timescales. On the right, we show the expected residuals given 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.

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 in [*M Coughlin*, 2014], almost reaching the limit set by the sensor noise of the Kinemetrics 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 ig-

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

nored 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).

4 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, Wiener filters were calculated for the cancellation of seismic signals in several target sensors using data from the array as input. The residuals achieved here improve over previous results at Homestake with a smaller array. The spectra of residual noise approach the limit set by the instrumental noise over a broad frequency band between about 10 mHz and 0.5 Hz reducing seismic signals by more than 2 orders of magnitude. These results demonstrate the capability of seismic arrays to extract information about ambient seismic fields efficiently to provide coherent estimates of ground displacement at other points in the fields.

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.

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 Tollef-

449 son, and Bryce Pietzyk. This work was supported by National Science Foundation INSPIRE
450 grant PHY1344265. This paper has been assigned LIGO document number LIGO-P1700422.

451 References

452 Abbott, B. P. et al. (2016), Binary black hole mergers in the first advanced ligo observing
453 run, *Phys. Rev. X*, *6*, 041,015, doi:10.1103/PhysRevX.6.041015.
454 Abbott, B. P. et al. (2017), Gw170817: Observation of gravitational waves from a binary
455 neutron star inspiral, *Phys. Rev. Lett.*, *119*, 161,101, doi:10.1103/PhysRevLett.119.161101.
456 Aso, Y., Y. Michimura, K. Somiya, M. Ando, O. Miyakawa, T. Sekiguchi, D. Tatsumi, and
457 H. Yamamoto (2013), Interferometer design of the kagra gravitational wave detector, *Phys.*
458 *Rev. D*, *88*, 043,007, doi:10.1103/PhysRevD.88.043007.
459 B P Abbott et al. (2017), Exploring the sensitivity of next generation gravitational wave de-
460 tectors, *Classical and Quantum Gravity*, *34*(4), 044,001.
461 Barklage, M., D. Hollis, J. M. Gridley, R. Woodward, and N. Spriggs (2014), A Large-N
462 Mixed Sensor Active + Passive Seismic Array near Sweetwater, TX, *AGU Fall Meeting*
463 *Abstracts*.
464 Bensen, G. D., M. H. Ritzwoller, M. P. Barmin, A. L. Levshin, F. Lin, M. P. Moschetti, N. M.
465 Shapiro, and Y. Yang (2007), Processing seismic ambient noise data to obtain reliable
466 broad-band surface wave dispersion measurements, *Geophysical Journal International*,
467 *169*, 1239–1260, doi:10.1111/j.1365-246X.2007.03374.x.
468 Capon, J. (1969), High-resolution frequency-wavenumber spectrum analysis, *Proceedings of*
469 *the IEEE*, *57*(8), 1408–1418, doi:10.1109/PROC.1969.7278.
470 Cella, G. (2000), Off-line Subtraction of Seismic Newtonian Noise, in *Recent Developments*
471 *in General Relativity*, pp. 495–503, Springer.
472 Coughlin, M., and J. Harms (2012), Seismic topographic scattering in the context of GW
473 detector site selection, *Classical and Quantum Gravity*, *29*, 075,004.
474 Creighton, T. (2008), Tumbleweeds and airborne gravitational noise sources for LIGO,
475 *Class. Quantum Grav.*, *25*, 125,011.
476 F Acernese et al. (2015), Advanced Virgo: a second-generation interferometric gravitational
477 wave detector, *Classical and Quantum Gravity*, *32*(2), 024,001, doi:10.1088/0264-9381/
478 32/2/024001.
479 F. Matichard et al. (2015), Advanced LIGO two-stage twelve-axis vibration isolation and
480 positioning platform. Part 2: Experimental investigation and tests results, *Precision Engi-*
481 *neering*, *40*, 287 – 297, doi:http://dx.doi.org/10.1016/j.precisioneng.2014.11.010.
482 Feroz F., Hobson M.P., and Bridges M. (2009), Multinest: an efficient and robust bayesian
483 inference tool for cosmology and particle physics, *Mon. Not. Roy. Astron. Soc.*, *398*, 1601–
484 1614.
485 Foster, A., G. EkstrÃűm, and M. Nettles (2014), Surface wave phase velocities of the West-
486 ern United States from a two-station method, *Geophysical Journal International*, *196*(2),
487 1189–1206, doi:10.1093/gji/ggt454.
488 Gerstoft, P., P. M. Shearer, N. Harmon, and J. Zhang (2008), Global p, pp, and pkp wave mi-
489 croseisms observed from distant storms, *Geophysical Research Letters*, *35*(23), n/a–n/a,
490 doi:10.1029/2008GL036111, l23306.
491 Harms, J. (2015), Terrestrial Gravity Fluctuations, *Living Reviews in Relativity*, *18*(3), doi:
492 10.1007/lrr-2015-3.
493 Harms, J., and H. J. Paik (2015), Newtonian-noise cancellation in full-tensor gravitational-
494 wave detectors, *Phys. Rev. D*, *92*, 022,001, doi:10.1103/PhysRevD.92.022001.
495 Harms, J., F. Acernese, F. Barone, I. Bartos, M. Beker, J. van den Brand, N. Christensen,
496 M. Coughlin, R. DeSalvo, S. Dorsher, et al. (2010), Characterization of the seismic envi-
497 ronment at the Sanford Underground Laboratory, South Dakota, *Classical and Quantum*
498 *Gravity*, *27*(22), 225,011.
499 Hughes, S. A., and K. S. Thorne (1998), Seismic gravity-gradient noise in interferometric
500 gravitational-wave detectors, *Phys. Rev. D*, *58*, 122,002.

J Aasi et al (2015), Advanced LIGO, *Classical and Quantum Gravity*, *32*(7), 074,001, doi: 10.1088/0264-9381/32/7/074001.

Kokeyama, K., K. Izumi, W. Z. Korth, N. Smith-Lefebvre, K. Arai, and R. X. Adhikari (2014), Residual amplitude modulation in interferometric gravitational wave detectors, *J. Opt. Soc. Am. A*, *31*(1), 81–88, doi:10.1364/JOSAA.31.000081.

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.

Lück, H., C. Affeldt, J. Degallaix, A. Freise, H. Grote, M. Hewitson, S. Hild, J. Leong, M. Prijatelj, K. A. Strain, B. Willke, H. Wittel, and K. Danzmann (2010), The upgrade of geo 600, *Journal of Physics: Conference Series*, *228*(1), 012,012, doi:10.1088/1742-6596/228/1/012012.

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.

M Punturo et al. (2010), The einstein telescope: a third-generation gravitational wave observatory, *Classical and Quantum Gravity*, *27*(19), 194,002.

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*.

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

Unnikrishnan, C. S. (2013), Indigo and ligo-india: Scope and plans for gravitational wave research and precision metrology in india, *International Journal of Modern Physics D*, *22*(01), 1341,010, doi:10.1142/S0218271813410101.

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