"Monitoring remote ocean waves using P-wave microseisms."

Jennifer Neale[1], Nicholas Harmon[1] and Meric Srokosz[2]

Corresponding author: J. Neale, Ocean and Earth Science, National Oceanography Centre Southampton, University of Southampton Waterfront Campus, European Way, Southampton, SO14 3ZH, United Kingdom. (Jennifer.Neale@noc.soton.ac.uk)

[1]Ocean and Earth Science, National Oceanography Centre Southampton, University of Southampton, Southampton, U.K.

[2]National Oceanography Centre, University of Southampton Waterfront Campus, Southampton, U.K.

Abstract. Oceanic microseisms are generated by the interaction of opposing ocean waves and subsequent coupling with the seabed, so microseisms should contain information on the ocean conditions that generated them. This leads to the possibility of using seismic records as a proxy for the ocean gravity wavefield. Here we investigate the P-wave component of microseisms, which has previously been linked to areas of high wave interaction intensity in mid-ocean regions. We compare modeled P-wave microseismic sources with those observed at an array in California, and also investigate the relationship between observed sources and significant wave height. We found that the time-varying location of microseism sources in the North Pacific, mapped from beamforming and backprojection of seismic data, was accurate to $\leq 10^{\circ}$ in 90% of cases. The modeled sources were found to dominate at $\sim$0.2 Hz which was also reflected in the seismic observations. An empirical relationship between observed beampower and modeled source power allowed sources during an independent data period to be estimated with a correlation coefficient of 0.63. Likewise, significant wave height was also estimated with a correlation coefficient of 0.63. Our findings suggest that with improvements in resolution and amplitude retrieval from beamforming, correlations up to 0.78 should be possible between observed P-wave microseisms and significant wave height in remote ocean regions.

1. Introduction

Oceanic microseisms are tiny, continuous oscillations of the ground caused by the interaction of ocean waves with the solid earth beneath them. The most energetic microseisms are generated when ocean wave trains of similar frequency but opposite direction interact, producing a pressure fluctuation that is unattenuated with depth, has twice the frequency of the forcing waves, and a near-zero wave number [*Longuet-Higgins*, 1950; *Ardhuin and Herbers*, 2013; *Traer and Gerstoft*, 2014; *Ardhuin et al.*, 2015]. This pressure fluctuation couples to the seabed to produce microseisms with typical peak frequencies of about 0.14-0.20 Hz (5-7 seconds), which propagate as seismic surface waves and body waves (e.g. *Toksöz and Lacoss* [1968]; *Haubrich and McCamy* [1969]; *Roux et al.* [2005]; *Gerstoft et al.* [2006]; *Koper and de Foy* [2008]; *Koper et al.* [2010]; *Reading et al.* [2014]).

The requirement of opposing waves to generate double frequency microseisms means that microseism energy cannot be explicitly related to ocean wave height [*Kedar et al.*, 2008; *Ardhuin et al.*, 2011], as it depends on the directional characteristics of the wave trains as well as the wave energy. However, in cases where the opposing wave trains are related, as a result of directional spreading of one wave system (classified as Class I in *Ardhuin et al.* [2011, 2012]) or reflection from a coastline (Class II), a strong dependence on the ocean wave height is expected. As such, empirical relationships between microseism energy and significant wave height recorded at nearby coastal wave buoys have been identified [*Ardhuin et al.*, 2012; *Bromirski et al.*, 1999; *Ferretti et al.*, 2013]. These relationships were based on the full seismic spectrum which is dominated by seismic surface

waves, and work well for seismic stations that are most sensitive to local sea states or coastal reflection sources.

This study investigates the P-wave component of seismic noise, for which origins in the deep ocean have been inferred [*Toksöz and Lacoss*, 1968; *Haubrich and McCamy*, 1969; *Landès et al.*, 2010; *Gerstoft et al.*, 2008; *Zhang et al.*, 2010a; *Euler et al.*, 2014; *Koper et al.*, 2010]. Here Class I situations do occur, but the opposing wavefield could also result from the interaction of two independent wave systems: either the crossing of two swells, or the interaction of swell with locally generated wind waves (Class III in *Ardhuin et al.* [2011, 2012]). Given that two low-energy wave systems directly opposing could produce the same microseismic energy as two high-energy wave systems with more oblique interaction, the correspondence between microseism energy and wave heights is expected to be weaker. Over long averaging times, sources of P-wave microseisms do coincide with regions of high wave heights [*Euler et al.*, 2014], but it is unclear what microseisms can tell us about ocean wave energy on shorter time periods when the degree of opposing wavefield is unknown and likely to be highly variable. This has implications for using P-wave microseisms as a proxy for wave heights.

In this study we investigate the relationship between P-wave microseisms and significant wave height at short (3-hour) time periods. Significant wave height is a common and useful parameter to describe sea state, but measurements by wave-buoys far from the coast are very sparse, and coverage by satellites insufficient. Locating P-wave microseisms to these deep ocean regions, and being able to infer wave heights, would allow for the development of real-time monitoring of sea state from seismic records, adding a valuable new record to our current measurements of significant wave height in the deep ocean.

We first compare observed P-wave microseisms with modeled microseism sources in terms of location, frequency and amplitude, and derive an empirical relationship between the observed and modeled source. Then, we reconstruct the ocean wave energy spectrum from the estimated microseism source spectrum by applying assumptions about the degree of wavefield opposition. Finally, the significant wave height calculated from the estimated ocean wave energy spectrum is then compared with modeled significant wave height.

2. Data and Processing

Two years of vertical component seismic data (channel LHZ) from an array in California (Figure 1), were downloaded from the Southern California Earthquake Data Center [*SCEDC*, 2013] covering the period September 2012-September 2014. The Southern California Seismic Network (SCSN) that makes up the bulk of the array has been used previously to successfully locate P-wave microseisms [*Gerstoft et al.*, 2006, 2008; *Zhang et al.*, 2010a, b; *Obrebski et al.*, 2013]. Instrument response was removed, and each daily record was band-pass filtered between 0.002 and 0.400 Hz. Earthquakes were removed using an automated detection and removal algorithm. Earthquake events were identified using the ISC bulletin [*International Seismological Centre*, 2013] and removed by setting 1 hour of the waveform to zero if the RMS of that window was over 3 times the daily RMS. Daily spectra for each station were then individually examined and any bad quality data were discarded.

A similar beamforming procedure to *Gerstoft and Tanimoto* [2007] and *Gerstoft et al.* [2008] was used to examine microseisms as a function of frequency, azimuth and slowness. Each daily timeseries was split into 512-s chunks which were Fourier transformed to give a complex valued vector $v(f_2, t_i\ )$ containing the response from all stations in the array,

where f_2 is frequency in Hz and t_i refers to the start time of the Fourier transform. Only phase was retained by dividing v by its magnitude to reduce the influence of local site amplification effects [*Gerstoft and Tanimoto*, 2007]. The cross-spectral density matrix, $\mathbf{C} = \langle \mathbf{v}\mathbf{v}^{\mathbf{T}} \rangle$, where the brackets indicate temporal averaging over a 3-hour period, was calculated. The plane-wave response of the array is given by $\mathbf{p}(f_2, s, \theta) = exp[i2\pi f_2 s(\mathbf{r}e)]$ where $e = (sin\theta cos\theta)^T$is the directional cosines for a plane wave with given azimuth θ, $\mathbf{r}$ is the coordinates of the seismometers with respect to their mean and s is slowness. The beamformer output is then $b(f_2, s, \theta, t) = \mathbf{p}(f_2, s, \theta)^T \mathbf{C}(f_2, t)\mathbf{p}(f_2, s, \theta)$ which we adjusted for the number of stations N in the array by dividing by N^2 [*Euler et al.*, 2014]. Array response functions at frequencies of 0.1 and 0.3 Hz are shown in Figure 2a and b.

For each 3-hour time period we therefore have beamformer output as a function of frequency, slowness and azimuth. An examination of the full 2 years of data revealed a large contrast between the strongest arrivals during winter and summer (Figure S1). During winter months, the strongest arrivals came from the North Pacific and North Atlantic with slownesses corresponding to P or PP phases, whereas during summer months the strongest arrivals came from southerly azimuths with slownesses of PKP phases. *Gerstoft et al.* [2008] previously found P-waves from the Pacific to dominate at the SCSN network during winter, so here we limit our analysis to P-waves from the North Pacific from mid-October to mid-March. Additionally we reject any days when the number of stations falls below 170 to minimize errors and biases associated with too few stations. At each frequency of interest, we back-projected azimuth and slowness onto a 2° latitude by 2° longitude geographical grid assuming slownesses of direct P-waves with a source at the Earth's surface using the ak135 travel time tables of *Kennett et al.* [1995] for a

spherically symmetric Earth model. We projected for distances between 30° and 90° from the array centre which is the typical range for teleseismic P-waves [*Obrebski et al.*, 2013]. A synthetic test for a point source located at 35°N 169°E is shown in Figure 2c and d. An example output for 23rd December 2012 (00:00) is shown in Figure 3a.

Double frequency P-wave microseism sources were modeled over the same 2-year period as the seismic data. The method of *Ardhuin et al.* [2011] and *Ardhuin and Herbers* [2013] based on the numerical ocean wave model WAVEWATCH III [*Tolman*, 2014] was followed to calculate the vertical ground displacement associated with P-waves at each source location. From *Ardhuin et al.* [2011] and *Farra et al.* [2016], the second-order pressure spectrum at near-zero wavenumber and twice the ocean wave frequency f, due to the interaction of similar frequency waves traveling in opposite directions [*Hasselmann*, 1963], is given by:

$$F_P(f_2 = 2f) = [2\pi]^2[\rho_w g]^2 f_2 E^2(f) I(f) \tag{1}$$

which has units of $\mathrm{Pa^2m^2s}$ and where ρ_w is the density of water, g is gravitational acceleration, $E(f)$ is the ocean wave frequency spectrum and $I(f)$ is a non-dimensional function that depends on the wave energy distribution M over the directions θ :

$$I(f) = \int_0^{\pi} M(f,\theta) M(f,\theta+\pi) d\theta \tag{2}$$

WAVEWATCH III was forced with 6-hourly winds and sea-ice cover from ECMWF's ERA-interim reanalysis [*Dee et al.*, 2011] and the second order pressure spectrum was output hourly on a 0.5° longitude by 0.5° latitude global grid from 78° North to 78° South. The output was smoothed using a 7-by-7 Gaussian low-pass filter and resampled

to a 2° by 2° grid and averaged over 3 hour time periods for direct comparison with the beamforming. Coastal wave reflection, which is function of wave amplitude, wave frequency and beach slope, can significantly affect the value of F_P near the coast. Because we are mainly interested in deep-water events where reflection has less influence, we chose the reflection coefficient simply as R^2=0.1 for continents and large islands and 0.2 for small islands [*Ardhuin et al.*, 2011]. A second model run over a two-month period with lower reflection coefficients of R^2=0.02 for continents and large islands and 0.04 for small islands, which more closely correspond to the estimates of *Stutzmann et al.* [2012], did not produce significantly different values of F_P. The map of P-wave sources (Pa^2m^2s) is calculated by multiplying the second order pressure spectrum at each gridpoint by the squared source site effect $[2|C_P|\frac{\rho_c}{\rho_w}]$[*Farra et al.*, 2016]:

$$\begin{aligned} P(f_2) &= F_p(f_2) \times [2|C_P|\frac{\rho_c}{\rho_w}]^2 \\ &= [2\pi]^2[\rho_w g]^2 f_2 E^2(f) I(f) \times [2|C_P|\frac{\rho_c}{\rho_w}]^2 \end{aligned} \tag{3}$$

C_P is a non-dimensional amplification coefficient dependent on frequency, water-depth and P-wave take-off angle (distance from source to receiver) [*Ardhuin and Herbers*, 2013; *Gualtieri et al.*, 2014]. C_P was calculated using the water depth and take-off angle appropriate for each gridpoint using the formulation of *Gualtieri et al.* [2014], taking water-depth from the ETOP01 dataset [*Amante and Eakins*, 2009] and using the ObsyPy package [*Beyreuther et al.*, 2010] to calculate take-off angle based on the distance from the array center. Values for spatially-varying crustal density ρ_c and water density ρ_w were taken from the upper crust layer and water layer of the global crustal model CRUST1.0 (`http://igppweb.ucsd.edu/~gabi/rem.html`). An example of the modeled source on

23rd December 2012 (00:00) is shown in Figure 3b. A study by [*Obrebski et al.*, 2013] has previously validated the location of modeled sources with seismic P-wave observations.

We also output the modeled ocean wave energy spectrum $E(f)$ from which we calculated significant wave height, H_s, as observations of H_s (e.g. from wave buoys or satellites) are not available at the temporal resolution or spatial extent required for comparison with the seismic data. Again we smoothed and resampled the WAVEWATCH III output of $E(f)$ to a 2° by 2° grid and averaged over 3-hour time periods.

3. Results

3.1. Comparison between observed and modeled source

For each 3-hour time period over the first winter (mid-October 2012 to mid-March 2013) the beamformer output and modeled source were integrated over the main double-frequency microseism band between 0.1 and 0.3 Hz. In order to identify multiple peaks in the beamformer maps, caused by more than one source acting in the basin at a given time, we used a standard watershed algorithm that identifies peaks in the image and the regions associated with each of these peaks (see Figure 3c). For each beamformer peak identified we found the largest modeled source within it's watershed region to examine how the beampower varies with source power. By associating the beampower peak to a source located anywhere within its watershed region, rather than using a direct pixel comparison, we allow for errors in the beamformer location which result in it being offset from the true source location. We considered beamform peaks in the North Pacific (above 0°N).

The plot of modeled source power vs. beampower in Figure 4 shows that up to a certain limit, an increase in modeled source power had little to no effect on the observed beam-

power, but beyond this limit beampower increased rapidly with source power, although with a lot of scatter. The correlation coefficient calculated using Spearman's Rank was 0.71 (strong correlation, >99% significance). After binning the data (shown in red) we defined the noise level as the mean of the first bin plus 2 standard deviations, which corresponds to a beampower of -28.68 dB. The binned data was found to be best described by an exponential function (shown in green), which indicated that the beampower only exceeded the noise level when source power ($F_P[2|C_P|\frac{\rho_c}{\rho_w}]^2$) reached approximately 102 dB (1.6×10^{10} Pa^2m^2).

Next we calculated the distance between the beamformer peaks and their associated modeled sources, excluding all peaks below the noise threshold of -28.68 dB. The locations of the beamformer peaks and the sources are plotted in Figure 5a and b with color indicating the number of occurrences at each grid point. It can be seen from Figure 5a and b that both the observed and modeled sources occurred over the whole region, although the most energetic 100 events were concentrated around 40-50°N, 160-180°E (Figure 5c and d), which corresponds well with the locations observed by *Obrebski et al.* [2013] in the year 2010. The beamformer peak and associated modeled source were found to be located $\leq$10° (5 grid points) apart in 90% of occurrences, were $\leq$6° (3 grid points) apart in 67% of occurrences but only matched the same grid point in 4% of occurrences. *Obrebski et al.* [2013] found similar offsets (between 1.1° and 9.5°) for the 54 strongest modeled events in their study. Discrepancies may be due to 3-d velocity structure that we are not taking into account in the backprojection (up to 4° [*Euler et al.*, 2014]), contamination from other phases or earthquakes, contamination from array response [*Gal et al.*, 2016], or inaccuracies in the modeled source location (either due to inaccuracies in the wave model

or site effect). In some cases the discrepancy was found to arise from the resolution limits of the array, where two neighboring sources merged into one observed source centered on the average location (a limitation that has been observed previously [*Hillers et al.*, 2012; *Euler et al.*, 2014]).

To compare the frequency content of the modeled and observed sources, we examined the non-integrated output of each of the beamformer peaks and associated source peaks. Again we only considered beamformer peaks that were above the noise threshold, and in addition only the cases when the observed and modeled sources were $\leq 10^{\circ}$ apart to exclude any spurious observations.

The mean spectra of the observed and modeled sources were calculated (Figure 6a and b), and the mean frequency from these spectra was calculated as $\sqrt{\frac{m_2}{m_0}}$ where $m_n = \int_{f_2=0.1Hz}^{f_2=0.3Hz} (S(f_2) \times f_2^n) df$ and $S(f_2)$ is either the beamform spectrum b or the modeled source spectrum P. The mean frequency of the beamformer was 0.21 Hz, whilst the mean frequency of the modeled source was 0.19 Hz, indicating that the seismic observations represented well the frequency content of the sources acting in the ocean basin. To examine the variability amongst individual spectra, the peak frequency of each spectrum was picked and plotted as a histogram with color corresponding to number of occurrences in Figure 6c. Again it can be seen that both the beamform peak and modeled source peak occurred most often at about 0.2 Hz, with rare instances where the peak occurred down to 0.1 Hz or up to 0.3 Hz.

The results indicate that the seismic beamforming observations reflect the location, frequency content and amplitude of microseismic sources, with location accurate to 10°

in 90% of occurrences and with a strong correlation coefficient between beampower and source power of 0.71.

Next we used the relationship between beampower and modeled source power found during the first winter to estimate source power over the second winter (mid-October 2013 to mid-March 2014) using the observed beampowers over that period.

To estimate source power from observed beampower, we fitted another curve between beampower and source power this time with beampower as the predictor and source power as the response variable (Figure 7a), and only considering beampowers over the noise threshold. Again we also excluded from the regression any cases when the observed and modeled sources were >10° apart. The best-fitting logarithmic function was applied to the seismic observations of the second winter to estimate sources during the second winter (a linear fit between beampower and source power for source powers >102 dB was attempted but resulted in overestimation of sources at large beampowers). The timeseries of the modeled source amplitude (10^(source power/10)) is plotted along with estimated source amplitude in Figure 7b, where only the largest estimated source at each timestep is plotted if there were more than one source acting simultaneously. Pearson's linear correlation coefficient between the estimated source power and modeled source power was 0.63 (>99% significance).

3.2. Relating microseism source to significant wave height

The results presented have shown that we are able to estimate the integrated microseism sources in the North Pacific from seismic observations with some confidence. If we now consider observations at each frequency, we can reconstruct the full source spectrum, from

which the ocean wave energy spectrum can be estimated, and thus significant wave height. This is a rearrangement of equation 3:

$$E(f) = \sqrt{\frac{P(f_2)}{[2|C_P|\frac{\rho_c}{\rho_w}]^2[2\pi]^2[\rho_w g]^2 f_2 I(f)}} \tag{4}$$

where the source power spectral density spectrum $P(f_2)$ is estimated from the observed beampower spectrum, values of ρ_c , ρ_w and C_p are taken at the grid points where peaks in beampower are located, and $I(f)$ is unknown.

Significant wave height can then be calculated from the estimated ocean wave energy spectrum:

$$H_s(swell) = 4\sqrt{\int_{0.05Hz}^{0.15Hz} E(f)df} \tag{5}$$

Here we distinguish our estimate of H_s as $H_s(swell)$ because we are only considering the spectrum between seismic frequencies of 0.1-0.3 Hz (0.05-0.15 Hz ocean wave frequency which equals 6.7-20 s periods) whereas traditionally significant wave height is calculated from a wider band (e.g. 2-30 ocean wave periods).

To estimate the source spectrum $P(f_2)$ from the observed beampower, beampower and source power were correlated in a similar way as previously, but instead of plotting the integrated powers, the power at each frequency was plotted separately (Figure 8a). From Figure 8a it was found that values at the different frequencies all lay along the same curve, so we applied one relationship to all frequencies. Again we binned the data points according to source power and fitted an exponential function. The noise level of the beampower (as defined previously) was found to be -20.16 dB and the source exceeded the noise level at a value of 117 dB (5.0×10^{11} Pa^2m^2s). Taking only beampower values above this noise

threshold, we replotted using beampower as the predictor variable (Figure 8b) and found that source power at each frequency is best estimated from the observed beampower using a logarithmic function given by: $P_{(estimated)}(dB) = alog_e(b \times (beampower - x_0)) + y_0$ with a=10.3564, b=0.89105, x_0=-20.8629 and y_0=108.5138.

The estimated source spectrum was constructed from the beampower spectrum using the identified logarithmic relationship in the following way. If the peak of the beampower spectrum was above the noise level of -20.16 dB, the source power of this peak, $P_{(estimated)}$(dB), was estimated, and the source amplitude as $P_{(estimated)} = 10\text{^}(P_{(estimated)}(dB)/10)$. It was found that by setting the estimated source amplitude at all other frequencies (below the noise threshold) to zero, the source spectrum and resulting ocean wave spectrum became too narrow. Instead, the mean shape of the modeled source spectrum as a function of f/f_p (where f_p is the peak frequency of the spectrum) was calculated (Figure 8c), and $P_{(estimated)}$) was multiplied by this. An example of the construction of one spectrum in this way is shown in Figure S2. An assumption here is that the source spectrum only has one peak, and indeed it was found that the modeled source spectrum had a single peak (counting peaks as those above the mean + one standard deviation) in 99% of cases whilst the beamformer spectrum had a single peak in 80% of cases. We limit our estimation of $P_{(estimated)}$ to cases when both the beamformer spectrum and source spectrum had a single peak.

$I(f)$ depends on the degree of wave interaction at each frequency. When the opposing wavefield is a result of coastal reflection this can be related to the reflection coefficient, or when the opposing wavefield is a result of directional spread this be related to the wave energy [*Ardhuin et al.*, 2012]. However for two opposing wavefields in mid-ocean regions

$I(f)$ cannot necessarily be related to the ocean wave energy but depends on the (unknown) characteristics of both wavefields. For example, a given $I(f)$ could result from two low energy swells that directly oppose eachother, or from one large energy swell that meets a low energy wind-sea. The assumption made about $I(f)$ is therefore expected to be one of the largest sources of error in the estimation of ocean wave energy and significant wave height from seismic observations. We calculated the mean $I(f)$ from the modeled cases as a function of f/f_p (Figure 8d) and assumed this $I(f)$ spectrum in the calculations.

Because the previous results indicated that location was only accurate up to 10° in the majority of cases, the map of $[2|C_p|\frac{\rho_c}{\rho_w}]^2$ was first smoothed over 10° using a 11-by-11 low-pass Gaussian filter before taking the value at the beampower location.

The wave energy spectrum was estimated from the beamforming observations of the first winter using equation 4, and significant wave height estimated using equation 5. Figure 9 shows the modeled $H_s(swell)$ vs. the estimated $H_s(swell)$ which had a Pearson's linear correlation coefficient of 0.48. The scatter points are shaded by point density and binned by modeled $H_s(swell)$. It can be seen from the figure that at the most common wave heights (∼2-3m) the estimation was most accurate, whereas higher modeled wave heights were underestimated. We used the mismatch between the estimated $H_s(swell)$ bins and the line $y = x$ to define a calibration factor, which was then applied to estimates of $H_s(swell)$ during the second winter (a calibration factor based on a curve which smoothed out the uncertainties at the larger wave heights did not make any significant difference to the results). Figure 10 shows modeled vs. estimated $H_s(swell)$ for the second winter which had a Pearson's linear correlation coefficient of 0.63. With the calibration factor,

the means of each data bin lie much closer to the line $y = x$, with underestimation only occurring at $H_s(swell)$ values over about 8m.

4. Discussion

Errors in the estimate of $E(f)$ using equation 4 and consequently H_s will arise for three main reasons: 1) Inaccurate estimate of source amplitude $P(f_2)$ from beamformer amplitude 2) Inaccurate location of beamformer peak, therefore inaccurate site effect 3) inaccurate estimate of $I(f)$. Another expected source of error is that is it not possible to extract information about the ocean wave spectrum at frequencies for which the wavefield is unidirectional ($I(f) = 0$), because the observed beampower is only sensitive to opposing waves. Consequently, these parts of the ocean wave energy spectrum would be underestimated as would the derived significant wave height.

In order to understand how each of these errors influence the correlation between estimated and modeled $H_s(swell)$, we looked at each of these sources of error in turn. Figure 11a shows the ideal case when the source spectrum, site effect and $I(f)$ are known exactly for the estimation of E(f) using equation 4 (i.e. the modeled values were used). The correlation coefficient in this case between modeled $H_s(swell)$ and estimated $H_s(swell)$ was equal to 1.00. Underestimation would be expected if there was a large portion of unidirectional wave energy in addition to the opposing wavefields, but this wasn't the case as the modeled $I(f)$ was rarely (<1% of the time) exactly 0. Plot b examines the error due to inaccurate estimate of $I(f)$. In this estimate of $H_s(swell)$ the mean value of $I(f)$ shown in Figure 8d was used in the calculation. There was still moderately strong correlation of 0.78 but this was much lower than the ideal case (a). There also appears to be a tendency for underestimation which increases with modeled $H_s(swell)$. This would

be caused by the assumed $I(f)$ being too large for higher wave heights. An explanation for this may be related to the case for waves generated by local winds, in which $I(f)$ generally decreases with increasing $E(f)$ [*Ardhuin et al.*, 2012]. Plot c examines the error due to inaccurate location (site effect). In this estimate of $H_s(swell)$ the location of the beamformer peak was used to obtain the value of the smoothed site effect. The correlation of 0.90 indicates that the error introduced from inaccurate location is less than the error introduced by inaccurate $I(f)$. Plot d examines the combined error from $I(f)$ and location and as expected the correlation drops further, to 0.71.

Plots e-h are the same as plots a-d but with the addition of error caused by inaccurate estimate of the source amplitude. Plot e) shows that inaccurate amplitude estimation causes a larger reduction in correlation than errors in $I(f)$ and location combined. With inaccuracies in $I(f)$ and location correlation goes down to 0.48 (Figure 11h).

The large amount of scatter between source amplitude and beampower is unsurprising given the amplitude removal during the beamforming process and, as mentioned by *Obrebski et al.* [2013], because the beampower also depends on the size of the area the source is acting over as well as energy losses along the propagation path (including geometric spreading, attenuation and transmission through Earth structure such as the Moho, 410km and 660km discontinuities [*Nishida and Takagi*, 2016]). Nevertheless, our results show that a relationship between source amplitude and observed beampower does exist. Furthermore, Figure 8b suggests that with improvements in beampower and location estimation correlations of up to 0.78 are possible even with the uncertainty surrounding $I(f)$. A direction for these improvements may be found in the recent work of *Farra et al.* [2016] and *Nishida and Takagi* [2016]. *Farra et al.* [2016] used a ray-theoretical approach

to estimate P-wave ground displacement for a given source including site, receiver and propagation effects, and *Nishida and Takagi* [2016] used a similar formulation to estimate the pressure source by minimising the squared difference between observed and modeled ground displacement.

Finally, throughout the study, the modeled source has been considered the 'true' value. Scatter between the beampower and modeled source may be caused by inaccuracies in the wave model itself, for example due to the wind input or parameterization within the model. Although there is currently no other way of estimating wave-interaction intensity over such spatial and temporal scales with which to validate the model output, an idea of error within the WAVEWATCH III model could be found by analyzing the spread of results obtained from multiple runs with different wind inputs and parameterization. Scatter may also be caused by inaccuracies in the calculation of the site effect, which may not well represent each 2 by 2° pixel in regions of large bathymetric variability [*Hillers et al.*, 2012], and for which we have not taken into account the effect of sediments or earth structure below the upper crust [*Gualtieri et al.*, 2014]. A thick sediment layer at the source results in reduced amplitudes of land-recorded microseisms [*Gualtieri et al.*, 2015], and may be important for sources close to the coast where sediments are thicker.

It is important to remember however that estimates about significant wave height can only be made where there is wave interaction occurring. Sometimes this does correspond to the largest wave heights in the ocean basin (e.g. Figure 12a,b), but this is not necessarily the case (Figure 12c,d).

5. Conclusions

Observed seismic P-waves in California were found to relate to modeled microseismic sources in the North Pacific Ocean in terms of location, frequency content and amplitude. The observed P-waves were located through beamforming and backprojection, and were found to match the location of strong modeled sources by $\leq 10^\circ$ (5 grid points) in 90% of cases. Both the modeled sources and observed P-waves were dominated by microseisms with a frequency of approximately 0.20 Hz. Beampower was moderately to strongly correlated with the power of the modeled sources, and only exceeded the noise threshold when sources were $>1.6\times10^{10}$ Pa^2m^2 (integrated between 0.1 and 0.3 Hz) or $>5.0\times10^{11}$ Pa^2m^2s. The empirical relationship between beampower and source power allowed sources during the second winter to be estimated from observed beampower. The resulting estimated sources were found to correlate with the modeled sources with a correlation coefficient of 0.63.

After reconstructing the source spectrum from the beamformer spectrum, and making an assumption about the directional characteristics of the wavefield, the ocean wave energy spectrum was estimated, and from that, significant wave height. During the first year, the modeled and estimated significant wave height correlated by 0.48. An underestimation of wave height at higher modeled values appeared to be introduced in the assumption of the directional wave characteristics. A calibration factor between modeled and estimated significant wave height calculated from the first winter's results was applied to the second year and the underestimation was largely removed, giving a correlation of 0.63 between modeled and estimated significant wave height. Inaccuracy in beampower was found to be the largest source of error, followed by inaccuracy in directional assumption. With

improvements in the location and amplitude estimation of sources from beamforming, it should be possible to obtain estimates of significant wave height that correlate with modeled wave heights by up to 0.78.

Acknowledgments. All datasets used are freely available to the public and are listed in the references. Source code for WAVEWATCH III can be requested from `http://polar.ncep.noaa.gov/waves/wavewatch/`. J. Neale's PhD studentship was funded by the Natural Environment Research Council (NERC). M. Srokosz was supported by NERC National Capability funding.

References

Amante, C., and B. Eakins (2009), ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis, doi:10.7289/V5C8276M.

Ardhuin, F., and T. Herbers (2013), Noise generation in the solid Earth, oceans and atmosphere, from nonlinear interacting surface gravity waves in finite depth, *Journal of Fluid Mechanics*, *716*, 316–348, doi:10.1017/jfm.2012.548.

Ardhuin, F., E. Stutzmann, M. Schimmel, and A. Mangeney (2011), Ocean wave sources of seismic noise, *Journal of Geophysical Research: Oceans*, *116*, C09,004, doi:10.1029/2011jc006952.

Ardhuin, F., A. Balanche, E. Stutzmann, and M. Obrebski (2012), From seismic noise to ocean wave parameters: General methods and validation, *Journal of Geophysical Research: Oceans*, *117*, C05,002, doi:10.1029/2011JC007449.

Ardhuin, F., L. Gualtieri, and E. Stutzmann (2015), How ocean waves rock the Earth: Two mechanisms explain microseisms with periods 3 to 300 s, *Geophysical Research*

Letters, *42*, 765–772, doi:10.1002/2014GL062782.

Beyreuther, M., R. Barsch, L. Krischer, T. Megies, Y. Behr, and J. Wassermann (2010), ObsPy: A Python Toolbox for Seismology, *Seismological Research Letters*, *81*, 530–533, doi:10.1785/gssrl.81.3.530.

Bromirski, P. D., R. E. Flick, and N. Graham (1999), Ocean wave height determined from inland seismometer data: Implications for investigating wave climate changes in the NE Pacific, *Journal of Geophysical Research: Oceans*, *104*(C9), 20,753–20,766, doi:10.1029/1999JC900156.

Dee, D. P., S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger, S. B. Healy, H. Hersbach, E. V. Hólm, L. Isaksen, P. Kållberg, M. Köhler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J.-J. Morcrette, B.-K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.-N. Thépaut, and F. Vitart (2011), The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Quarterly Journal of the Royal Meteorological Society*, *137*, 553–597, doi:10.1002/qj.828.

Euler, G. G., D. A. Wiens, and A. A. Nyblade (2014), Evidence for bathymetric control on the distribution of body wave microseism sources from temporary seismic arrays in Africa, *Geophysical Journal International*, *197*, 1869–1883, doi:10.1093/gji/ggu105.

Farra, V., E. Stutzmann, L. Gualtieri, M. Schimmel, and F. Ardhuin (2016), Ray-theoretical modeling of secondary microseism P waves, *Geophysical Journal International*, *206*, 1730–1739, doi:10.1093/gji/ggw242.

Ferretti, G., A. Zunino, D. Scafidi, S. Barani, and D. Spallarossa (2013), On microseisms recorded near the Ligurian coast (Italy) and their relationship with sea wave height, *Geophysical Journal International*, *194*, 524–533, doi:10.1093/gji/ggt114.

Gal, M., A. Reading, S. Ellingsen, K. Koper, R. Burlacu, and S. Gibbons (2016), Deconvolution enhanced direction of arrival estimation using one- and three-component seismic arrays applied to ocean induced microseisms, *Geophysical Journal International*, *206*, 345–359, doi:10.1093/gji/ggw150.

Gerstoft, P., and T. Tanimoto (2007), A year of microseisms in southern California, *Geophysical Research Letters*, *34*, L20,304, doi:10.1029/2007GL031091.

Gerstoft, P., M. C. Fehler, and K. G. Sabra (2006), When Katrina hit California, *Geophysical Research Letters*, *33*, L17,308, doi:10.1029/2006GL027270.

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*, L23,306, doi:10.1029/2008GL036111.

Gualtieri, L., E. Stutzmann, V. Farra, Y. Capdeville, M. Schimmel, F. Ardhuin, and A. Morelli (2014), Modelling the ocean site effect on seismic noise body waves, *Geophysical Journal International*, *197*, 1096–1106, doi:10.1093/gji/ggu042.

Gualtieri, L., E. Stutzmann, Y. Capdeville, V. Farra, A. Mangeney, and A. Morelli (2015), On the shaping factors of the secondary microseismic wavefield, *Journal of Geophysical Research: Solid Earth*, *120*, 6241–6262, doi:10.1002/2015JB012157.

Hasselmann, K. (1963), A statistical analysis of the generation of microseisms, *Rev. Geophys.*, *1*, 177–210, doi:10.1029/RG001i002p00177.

Haubrich, R. A., and K. McCamy (1969), Microseisms: Coastal and pelagic sources, *Reviews of Geophysics*, *7*, 539–571, doi:10.1029/RG007i003p00539.

Hillers, G., N. Graham, M. Campillo, S. Kedar, M. Landès, and N. Shapiro (2012), Global oceanic microseism sources as seen by seismic arrays and predicted by wave action models, *Geochemistry, Geophysics, Geosystems*, *13*, Q01,021, doi:10.1029/2011GC003875.

International Seismological Centre (2013), On-line Bulletin, *Tech. rep.*, Int. Seis. Cent., Thatcham, United Kingdom.

Kedar, S., M. Longuet-Higgins, F. Webb, N. Graham, R. Clayton, and C. Jones (2008), The origin of deep ocean microseisms in the North Atlantic Ocean, *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *464*, 777–793.

Kennett, B. L. N., E. R. Engdahl, and R. Buland (1995), Constraints on seismic velocities in the Earth from traveltimes, *Geophysical Journal International*, *122*, 108–124, doi:10.1111/j.1365-246X.1995.tb03540.x.

Koper, K. D., and B. de Foy (2008), Seasonal Anisotropy in Short-Period Seismic Noise Recorded in South Asia, *Bulletin of the Seismological Society of America*, *98*, 3033–3045, doi:10.1785/0120080082.

Koper, K. D., K. Seats, and H. Benz (2010), On the Composition of Earth's Short-Period Seismic Noise Field, *Bulletin of the Seismological Society of America*, *100*, 606–617, doi:10.1785/0120090120.

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*, B05,302, doi:10.1029/2009JB006918.

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

Nishida, K., and R. Takagi (2016), Teleseismic S wave microseisms, *Science*, *353*, 919–921, doi:10.1126/science.aaf7573.

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*, 4312–4324, doi:10.1002/jgrb.50233.

Reading, A. M., K. D. Koper, M. Gal, L. S. Graham, H. Tkalčić, and M. A. Hemer (2014), Dominant seismic noise sources in the Southern Ocean and West Pacific, 2000-2012, recorded at the Warramunga Seismic Array, Australia, *Geophysical Research Letters*, *41*, 3455–3463, doi:10.1002/2014GL060073.

Roux, P., K. G. Sabra, P. Gerstoft, W. A. Kuperman, and M. C. Fehler (2005), P-waves from cross-correlation of seismic noise, *Geophysical Research Letters*, *32*(19), n/a–n/a, doi:10.1029/2005GL023803.

SCEDC (2013), Southern California Earthquake Center, doi:10.7909/C3WD3xH1.

Stutzmann, E., F. Ardhuin, M. Schimmel, A. Mangeney, and G. Patau (2012), Modelling long-term seismic noise in various environments, *Geophysical Journal International*, *191*(2), 707–722, doi:10.1111/j.1365-246X.2012.05638.x.

Toksöz, M. N., and R. T. Lacoss (1968), Microseisms: mode structure and sources., *Science*, *159*, 872–3, doi:10.1126/science.159.3817.872.

Tolman, H. L. (2014), User manual and system documentation of WAVEWATCH III version 4.18, *NOAA / NWS / NCEP / MMAB Technical Note*, *316*.

500 Traer, J., and P. Gerstoft (2014), A unified theory of microseisms and hum, *Journal of*
501 *Geophysical Research: Solid Earth*, *119*, 3317–3339, doi:10.1002/2013JB010504.
502 Zhang, J., P. Gerstoft, and P. D. Bromirski (2010a), Pelagic and coastal sources of P-
503 wave microseisms: Generation under tropical cyclones, *Geophysical Research Letters*,
504 *37*, doi:10.1029/2010GL044288.
505 Zhang, J., P. Gerstoft, and P. M. Shearer (2010b), Resolving P-wave travel-time anomalies
506 using seismic array observations of oceanic storms, *Earth and Planetary Science Letters*,
507 *292*, 419–427, doi:10.1016/j.epsl.2010.02.014.

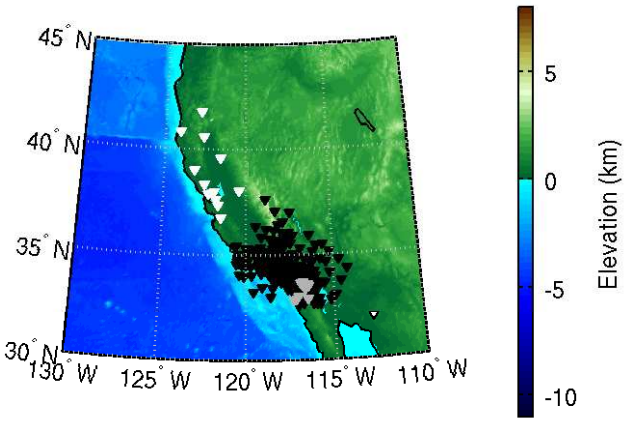


Figure 1. The seismic networks used in the study. White triangles=Berkeley Digital Seismograph Network, black triangles=Southern California Seismic Network, light gray triangles=ANZA regional network , white triangle with black border=USArray Transportable Array.

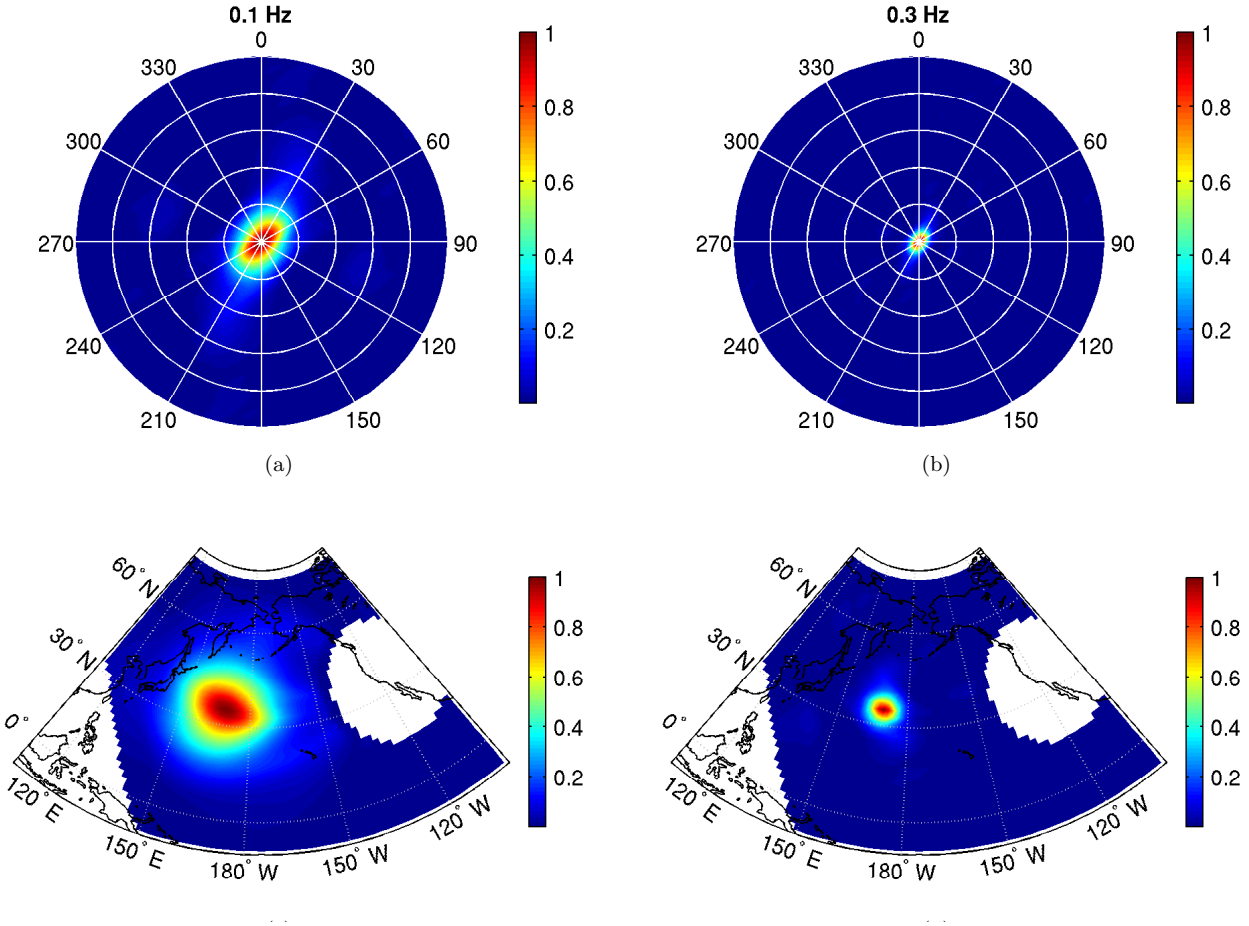
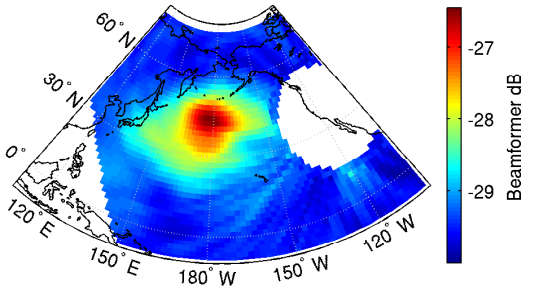
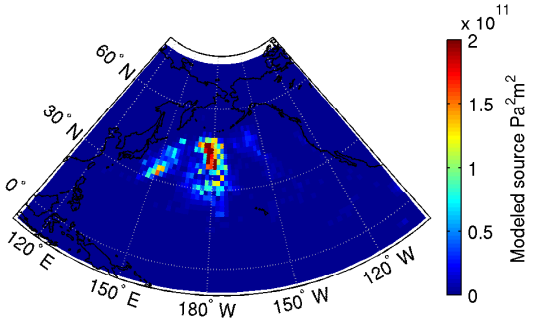


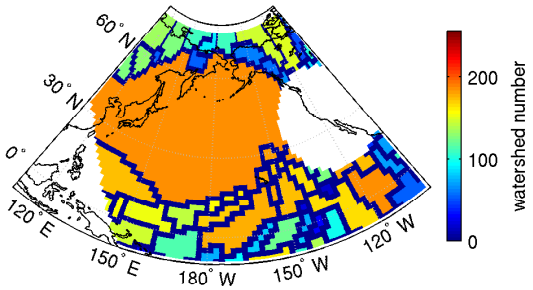
Figure 2. Top row: Slowness vs. azimuth array response functions at frequencies of a) 0.1 and b) 0.3 Hz. White circles are plotted at slowness values of 0.02, 0.04, 0.06 and 0.08 s/km. Bottom row: Array response for a point source located at 35° N 169°E for frequencies of c) 0.1 and d) 0.3 Hz.



(a)



(b)



(c)

Figure 3. Example output for 23rd December 2012 00:00:00. a) Beamformer output integrated between 0.1 and 0.3 Hz. b) Modeled source integrated between 0.1 and 0.3 Hz. c) Watershed regions identified from the beamformer image in (a).

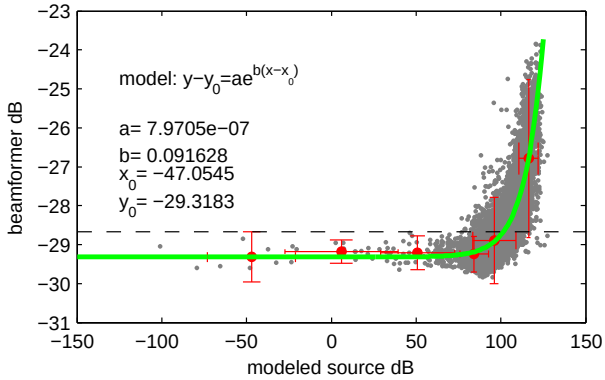


Figure 4. a) Modeled source power vs. observed beampower (both integrated between 0.1 and 0.3 Hz). The red circles are the mean of each data bin with error bars of ± two standard deviations. The dashed line is placed at the top of the error bar of the first bin and is equal to -28.68 dB, which we define as the noise threshold. The green curve is the best-fitting exponential of the form $y = ae^{b(x-x_0)} + y_0$ where values of a,b,x_0 and y_0 are given in the figure.

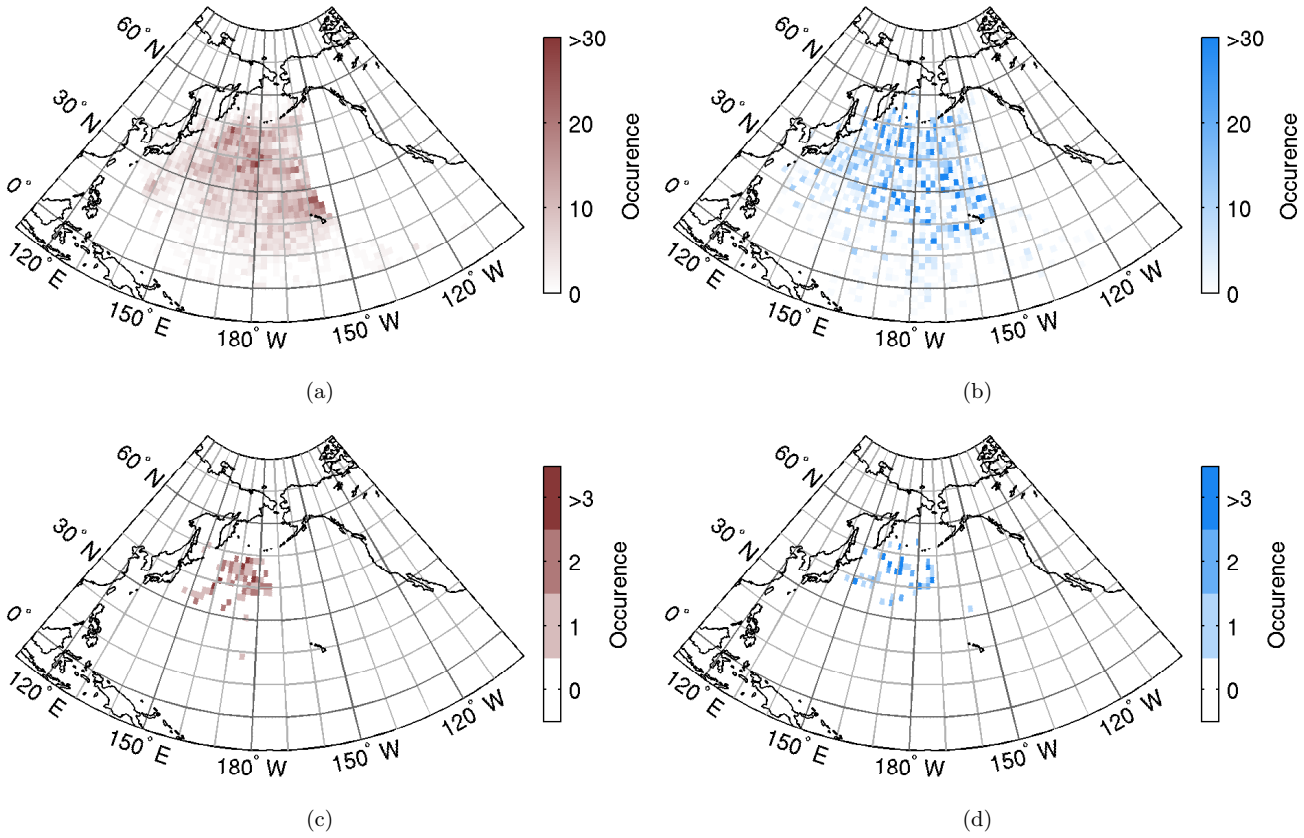
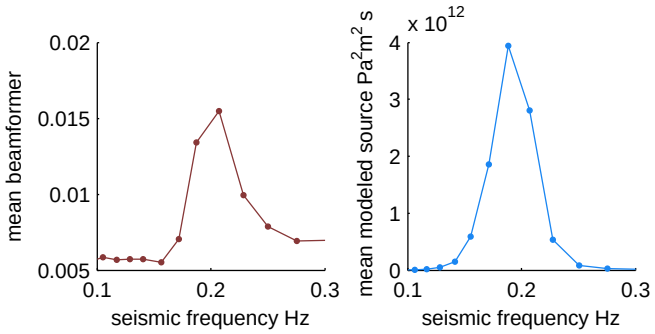
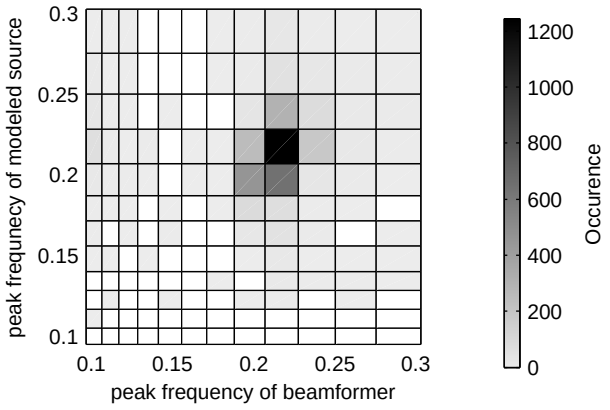


Figure 5. a) Located peaks in beamformer output. Beamformer output was integrated between 0.1 and 0.3 Hz before applying a watershed algorithm at each time step to identify the peaks. Only peaks that were above -28.68 dB (the noise threshold) are plotted. b) Locations of corresponding modeled source peaks. The modeled source was also integrated between 0.1 and 0.3 Hz and the maximum source within each beamformer watershed region of the peaks shown in (a) was identified as the source peak. Color indicates how many times the peak was located at each grid point. c) Same as (a) but for the 100 most energetic peaks. d) Same as (b) but only for the 100 events shown in (c).

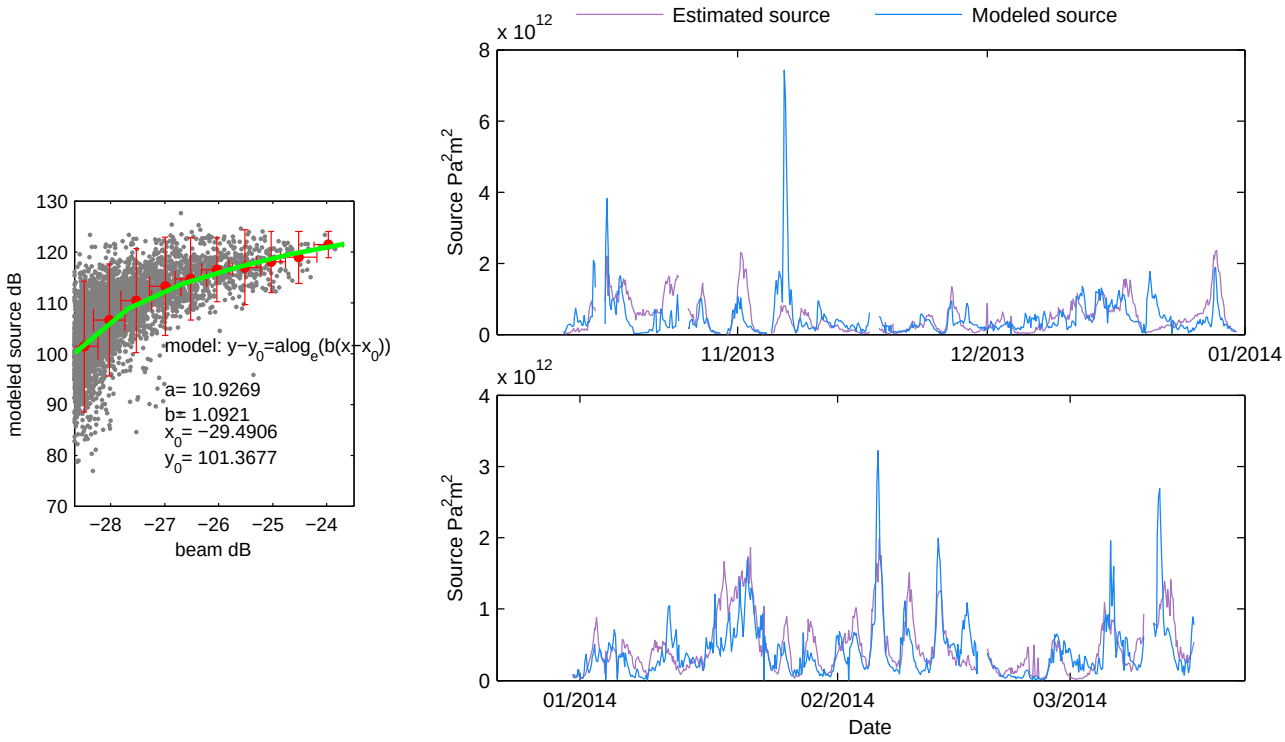


(a)



(b)

Figure 6. a) Mean beamformer (left) and modeled source (right) spectrum. Instances when the beamform peak and modeled source peak were $> 10°$ apart were excluded from the calculation of the mean. b) Peak frequency of beamform spectrum vs. peak frequency of modeled source spectrum. The grayscale shading indicates the number of timesteps the peak occurred at each given frequency.



(a) (b)

Figure 7. a) Modeled source power vs. observed beampower for the first winter. Both are integrated between 0.1 and 0.3 Hz. Values of beampower below the noise threshold of -28.68 dB and cases when the observed and modeled source locations were >10° apart were excluded. The red circles are the mean of each data bin with error bars of ± two standard deviations. The green curve is the best-fitting logarithmic function of the form $y = alog_e(b[x - x_0]) + y_0$ where values of a ,b ,x_0 and y_0 are given in the figure. b) Modeled source vs. estimated source for the second winter (mid-October 2013 to mid-March 2014). The estimated source was calculated from the observed beampower using the regression in (a). The timeseries has been split over two plots for clarity. Note the difference in y-axis limit between the plots.

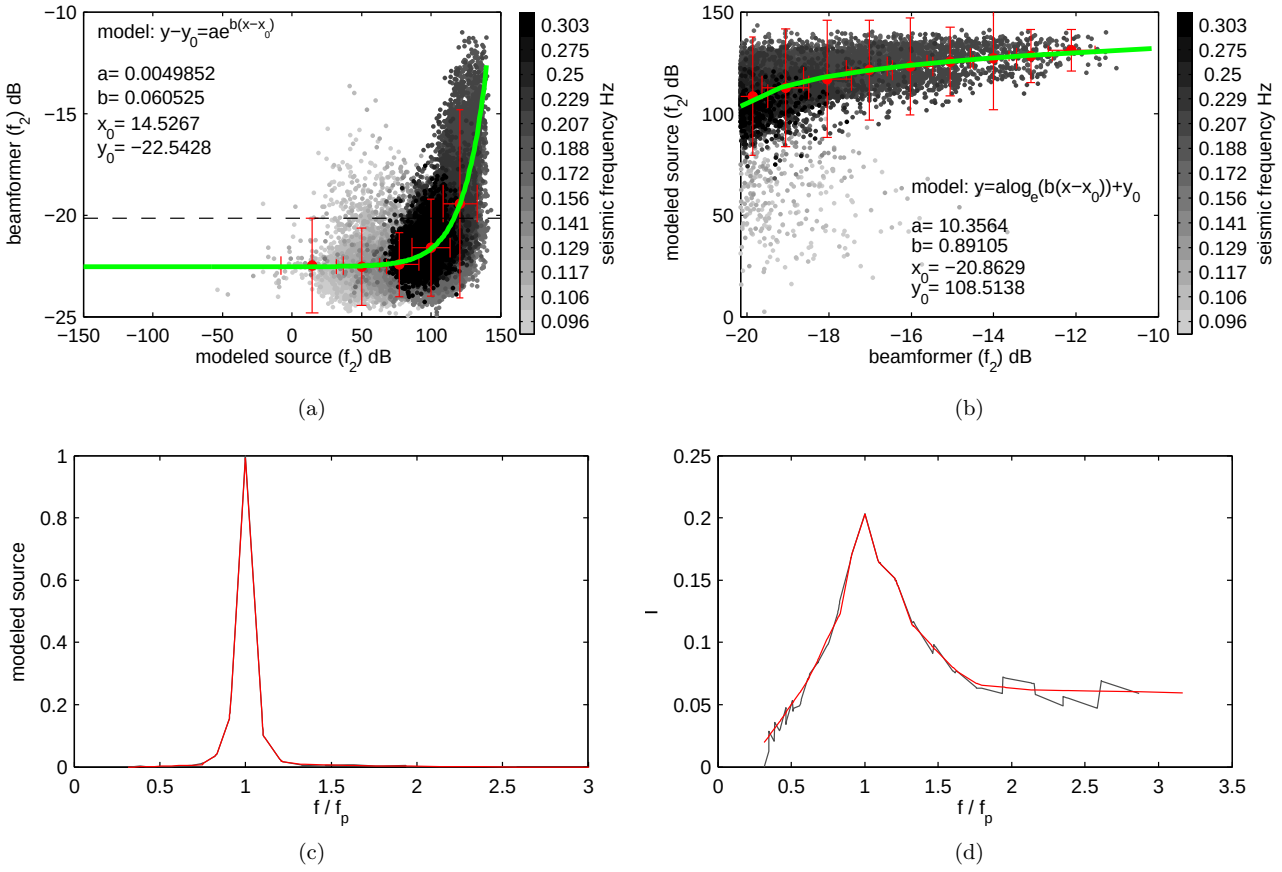


Figure 8. a) Similar to Figure 4 but with modeled source power vs. observed beampower plotted separately at each frequency. Only cases where the integrated value of the beampower was above the noise threshold of -28.68 dB are plotted, and only cases where the observed and modeled source were apart by $\leq 10°$. The red circles are the mean of each data bin with error bars of $\pm$ two standard deviations. The dashed line is placed at the top of the error bar of the first bin and is equal to -20.16 dB, which we define as the noise threshold. The green curve is the best-fitting exponential of the form $y = ae^{b(x-x_0)} + y_0$ where values of a, b, x_0 and y_0 are given in the figure. b) Same as (a) but with the axes switched and only including beampowers(f_2) $\geq$ -20.16 dB. The green curve is the best-fitting logarithmic function of the form $y = alog_e(b \times (x - x_0)) + y_0$. c) Mean shape of modeled source spectrum as a function of f/f_p which is the seismic frequency divided by the peak seismic frequency of the source spectrum. The data is plotted in black and the smoothed data in red (almost identical). d) Mean spectrum of the wave-interaction intensity.

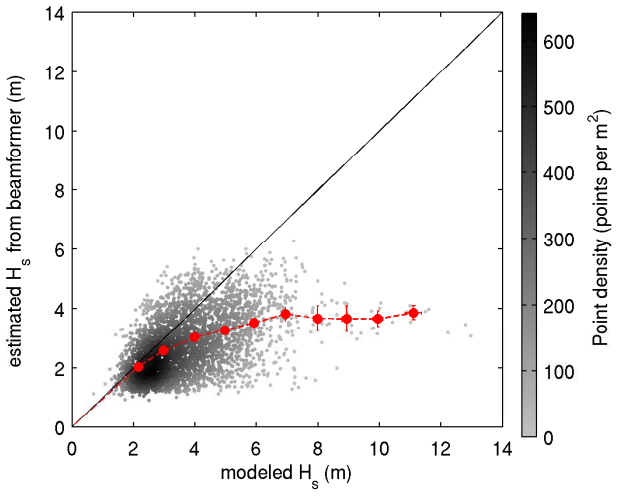
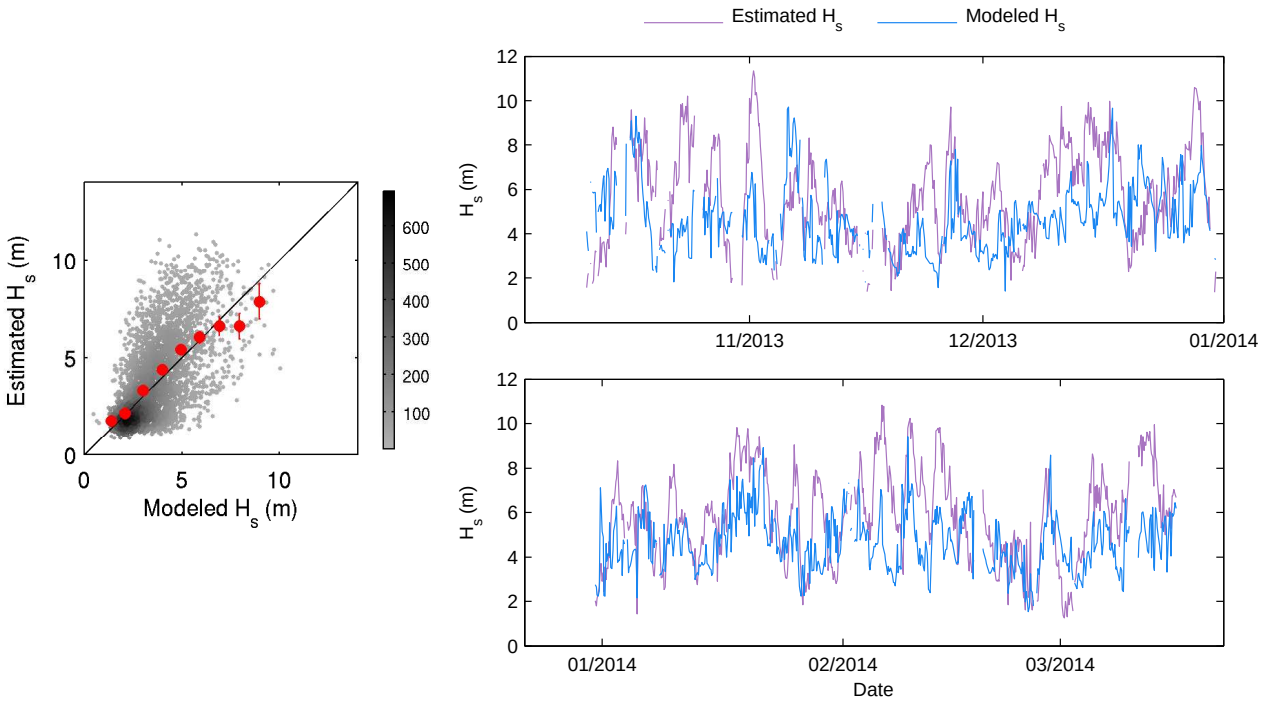


Figure 9. a) Modeled $H_s(swell)$ during the first winter (mid-October 2012 to mid-March 2013) vs. $H_s(swell)$ estimated from beamformer output. The red circles are the mean of each data bin with error bars of $\pm$ two standard errors. The grayscale shading indicates the point density (number of points per m^2). The offset between the mean of each data bin and the line $y = x$ was used as a calibration factor for the second winter.



(a) (b)

Figure 10. a) Modeled $H_s(swell)$ during the second winter (mid-October 2013 to mid-March 2014) vs. $H_s(swell)$ estimated from beamformer output. The red circles are the mean of each data bin with error bars of ± two standard errors. The grayscale shading indicates the point density (number of points per m^2). b) Timeseries of modeled and estimated $H_s(swell)$ during the second winter. Only the maximum estimated $H_s(swell)$ at each timestep is plotted.

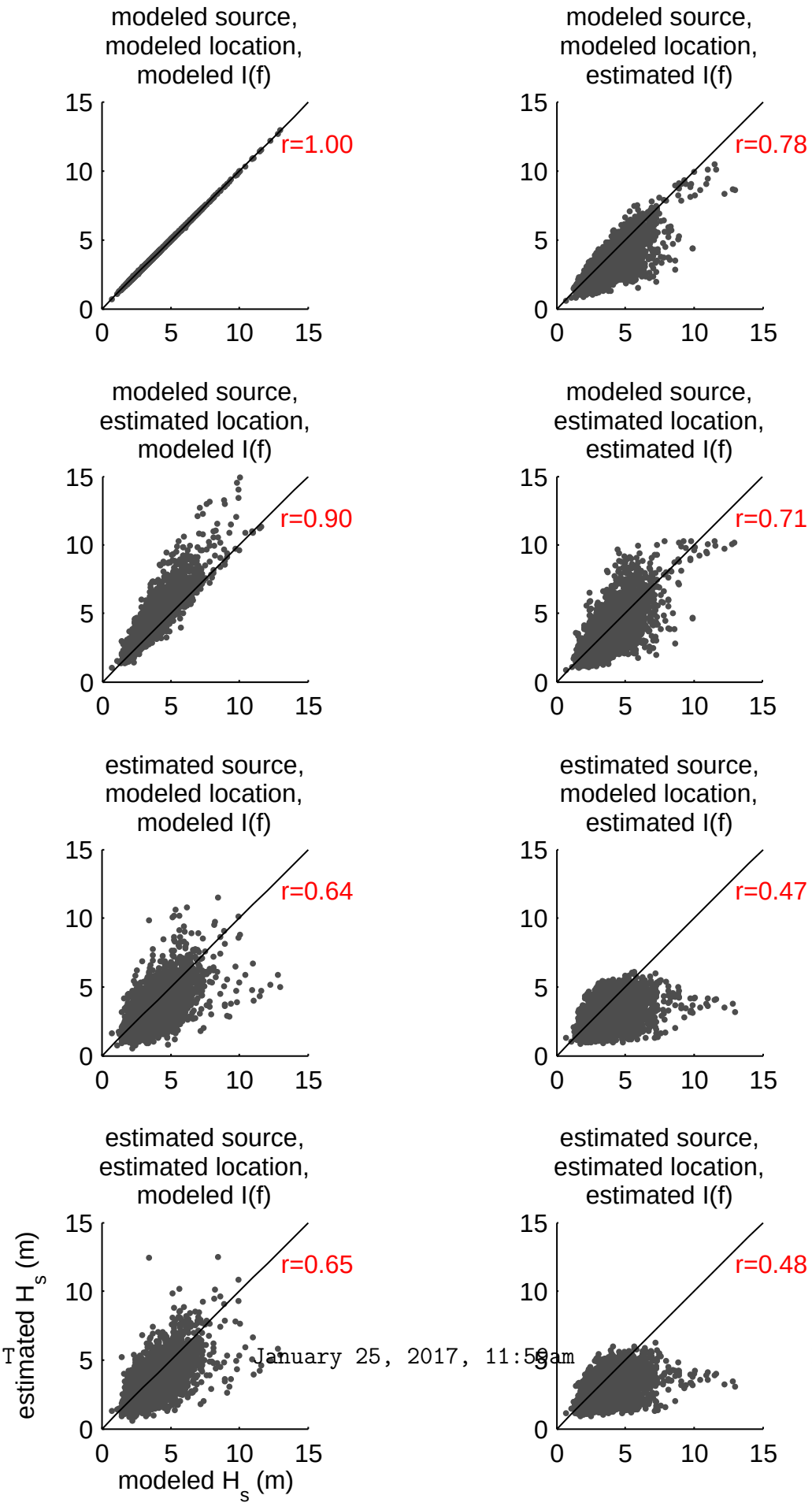

modeled source,
modeled location,
modeled I(f)
r=1.00
modeled source,
modeled location,
estimated I(f)
r=0.78
modeled source,
estimated location,
modeled I(f)
r=0.90
modeled source,
estimated location,
estimated I(f)
r=0.71
estimated source,
modeled location,
modeled I(f)
r=0.64
estimated source,
modeled location,
estimated I(f)
r=0.47
estimated source,
estimated location,
modeled I(f)
r=0.65
estimated source,
estimated location,
estimated I(f)
r=0.48
0
5
10
15
estimated H_s (m)
modeled H_s (m)

Figure 11. Causes of error in the estimation of $H_s(swell)$. In all plots modeled $H_s(swell)$ is on the x-axis and estimated $H_s(swell)$ is on the y-axis. r is the Pearson's linear correlation coefficient. 'Modeled source' means that modeled $P(f_2)$ was used in the calculation. 'Modeled location' means that the modeled source location was used for the value of the site effect. 'Modeled $I(f)$' means that the exact (modeled) value of $I(f)$ for each case was used. 'Estimated source' means that $P(f_2)$ was estimated from the beampower. 'Estimated location' means that the beamformer location was used to obtain the value of the smoothed site effect. 'Estimated $I(f)$' means that the $I(f)$ of Figure 8d was used. The plots therefore represent: a) The ideal case. b) error in $I(f)$. c) error in location. d) error in $I(f)$ and location. e) error in source amplitude. f) error in source amplitude and $I(f)$. g) error in source amplitude and location. h) error in source amplitude, $I(f)$ and location.

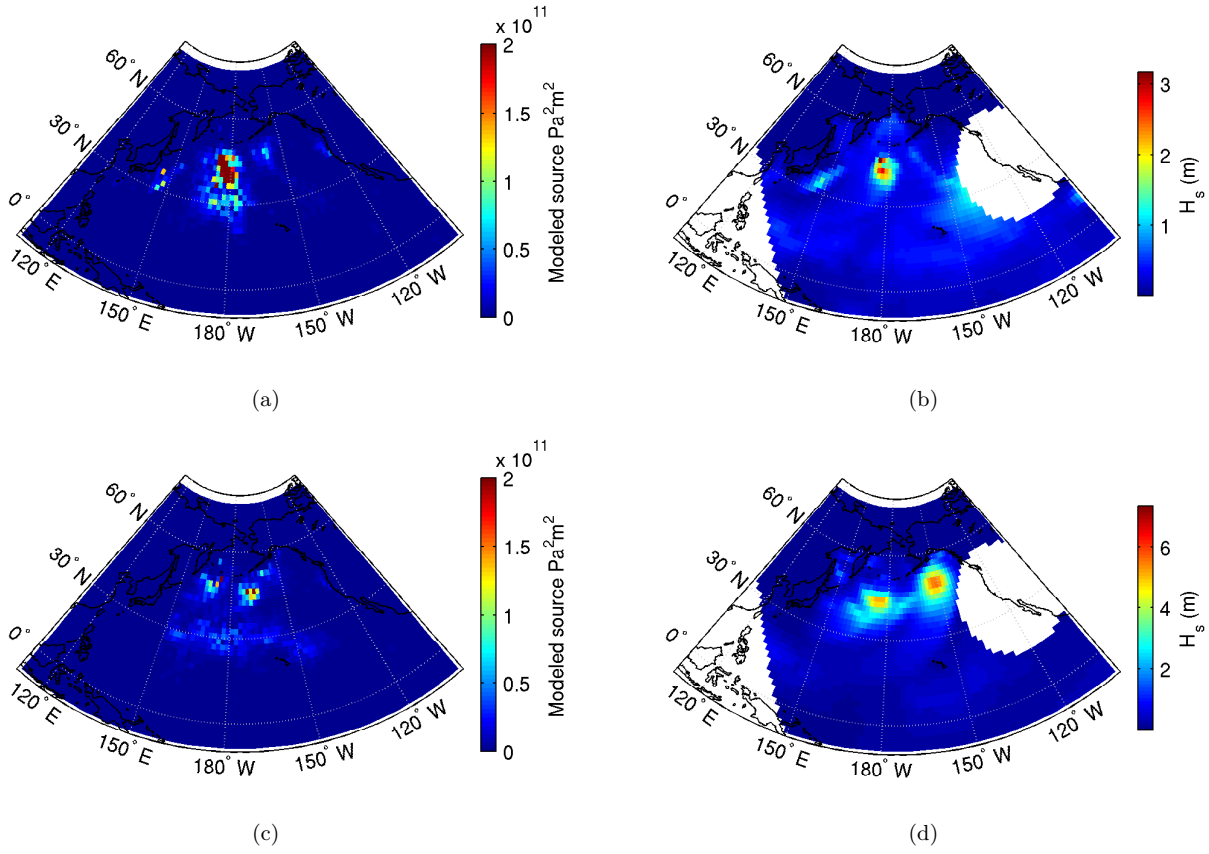


Figure 12. An example when maximum source corresponds to location of maximum H_s (top row) and when maximum source does not correspond to location of maximum H_s (bottom row). a) Modeled source on 22-Dec-2012 15:00:00. b) Modeled H_s on 22-Dec-2012 15:00:00. c) Modeled source on 27-Dec-2012 15:00:00. d) Modeled H_s on 27-Dec-2012 15:00:00. The modeled source has been integrated over f_2 =0.1-0.3 Hz whilst H_s has been calculated from the modeled ocean wave spectra over the full ocean wave frequency range f =0.0300-0.5758 Hz.