

## **Supplementary Methods**

This supplementary material provides detailed information on how we estimate and account for the array response, as well as additional figures, as referenced in the main text, describing the array stacking and detection procedure (Supplementary Figs 1, 2, 3, and 4), array response signature (Supplementary Fig. 5) and the fit to the seasonally averaged data of different distributions of sources (Supplementary Figs 6, 7, 8, and 9).

### **Array response**

In order to determine the distortions in the amplitude patterns as a function of back-azimuth due to the array shape, we compute synthetic stacks corresponding to a uniform distribution of sources of Rayleigh waves around the globe.

It is not possible to generate completely accurate synthetic waveforms for the hum, because the source mechanism is not known at present. However, several previous studies help us to define the properties of the waveforms: (1) waveforms mainly consist of fundamental mode Rayleigh waves; (2) mean energy over long time spans should be constant as a function of direction of arrival, if sources are completely random in time and space. Our synthetics are designed to satisfy these two conditions. We calculate 16 reference Rayleigh wave synthetics which correspond to sources at distances of 20 to 170 ° from the centre of the array (i.e. every 10°), an arbitrary moment tensor solution and shallow focal depth (15km). Construction of random source synthetics then involves three steps: 1) we randomly choose 1 out of the 16 synthetics and randomly perturb the source amplitude and source phase; 2) we randomly distribute the origin time, over an interval of 24 hours, of 4000 such synthetics. Each 24 hour waveform thus obtained is then assigned a specific direction in azimuth and then we repeat steps 1) and 2) for every 10 ° step in azimuth; 3) For each station in the array, we sum the 36 traces thus obtained, after correcting each for Rayleigh wave dispersion with respect to the centre of the array. We then perform the array stacking procedure in the same way as for real data and

obtain a distribution of amplitude as a function of azimuth. Ten realisations of this experiment are then averaged to obtain the array response shape shown in Supplementary Fig. 5c, d, which can be compared to the seasonal distribution of amplitudes (Supplementary Fig. 5a, b). The distributions shown in Supplementary Figs 6 and 7 are obtained by modifying the uniformly distributed source amplitudes in particular azimuth ranges.

### **Generation of regional source distributions.**

For the synthetic calculations shown in Supplementary Figs 8 and 9, we adopted a more computationally efficient approach. We first considered 642 points uniformly distributed over the sphere, using a sphere triangulation algorithm. For a one day period, and for each point we considered 30 sources with randomly distributed origin times. Source amplitudes and phases were obtained by adding a random perturbation to a reference source model, as was done in the array response analysis described above. Source amplitudes are constant in all azimuths (thus ignoring any possible non-uniform radiation pattern). Rayleigh waves are propagated to each station of each array, following which they are stacked using the same procedure as for real data, and the stacks are averaged over one day. To obtain a preferential distribution of sources in a particular region of the globe, the source amplitudes for the points located within that region are increased by a factor ranging between 100% and 400%, depending on the region, and so as to obtain an optimal fit to the data. To save computational time, we only considered a relatively sparse distribution of sources and a realisation of the experiment over only one day, whereas 10 days are required to obtain a completely stable pattern. However, we verified that the fluctuations between the realisations do not impact the key features of each regional distribution.

### **Supplementary Figure Captions**

**Supplementary Figure 1** Illustration of stacking procedure. **a**, Schematic diagram showing how waveforms are mapped to the centre of an array assuming plane wave propagation. Symbols indicate quiet BDSN stations used in this study (blue squares), other BDSN stations (red triangles) and centre of the array (green solid circle). In practice, we also use several additional stations from the TERRAscope network in southern California. **b**, Vertical velocity waveforms generated by the Jan. 8, 2000 Mw 7.2 event at 7 quiet BDSN stations, filtered using a Gaussian filter centred at 240 sec; **c**, Same as **b** after back-projecting the waveforms to the centre of the array, correcting for dispersion and attenuation using the reference PREM model<sup>27</sup>, and the known back-azimuth of the event.

**Supplementary Figure 2 a**, Gaussian filtered waveforms (with centre period 240 sec) recorded at F-net stations after back-projecting to centre of the array. Weak surface wave energy corresponding to the January 2, 2000 Mw 5.8 earthquake (12:58:45.2UTC; 51.54N 175.50W, distance: 36.31°) arrives around 3000 sec. **b**, Comparison between three different stacking methods: straight summation and mean (blue); 3<sup>rd</sup> root stacking method (green); phase weighted stack (red). By taking the envelope of every stack for all possible azimuths, we obtain the amplitude function as a function of time and back azimuth (Supplementary Fig. 3).

**Supplementary Figure 3** Amplitude of array stack as a function of back azimuth and time for a day with an earthquake. We pick the back-azimuth which corresponds to the maximum amplitude of the stack for the time interval considered, and define two functions of time: the “back-azimuth” function, and the corresponding “maximum amplitude” function (MAF). If that maximum exceeds a preset threshold, a detection is declared as illustrated in this example.

**a**, Plot of stack amplitude as a function of time and back azimuth at F-net. Green circles indicate significant amplitude in the time period considered. They correspond to R1, R2 and R3 for the January 2, 2000 Mw 5.7 earthquake (15:16:34.8UTC; 21.24S 173.30W, distance 74.17°). **b**, the corresponding maximum amplitude function as a function of time, highlighting the time of arrival of R1, R2 and R3.

**Supplementary Figure 4** Analysis of a detection during a quiet day (January 31, 2000) on the F-net array. **a**, Plot of amplitude (in m/s) of stack as a function of time and azimuth for a 6000 sec interval without earthquakes. Waveforms have been filtered with a Gaussian filter centred at 240 sec. Single station waveforms before and after correction for dispersion across the array at the back-azimuth of the maximum in **a** are shown in **b** and **c** respectively. Red lines show the time of best alignment. Corresponding stacks are shown in **d** before alignment and **e** after alignment. Note that the amplitude of the maximum is of the same order of magnitude as for the M5.8 event shown in Supplementary Fig. 2. **f**, Search for optimal phase velocity for dispersion correction before stacking, as a function of azimuth, for the time period corresponding to the maximum in the stack (54,500-55,500 sec). White lines indicate expected range of phase velocities for Rayleigh waves. **g**, Parameter search for phase velocity as a function of period for the same interval of time as in **f**. Here, the data have been bandpass filtered between 150 and 500 sec. The white line indicates the theoretical Rayleigh wave dispersion curve for the PREM model. For periods longer than 300 sec, phase velocity is not well resolved, as expected. The results of the parameter searches in **f** and **g** confirm that the observed energy in the stacks corresponds to Rayleigh waves.

**Supplementary Figure 5** Analysis of array response. **a**, Distribution of observed stack amplitude as a function of azimuth, averaged over winter (red) and summer (blue) for F-net. **b**, Same as **a** for BDSN. **c**, Normalised Array response for F-net computed as described in the Methods Section. **d**, Same as **c** for BDSN. **e**, Fourier spectrum in azimuth for winter (red), summer (blue) and array response (black) for F-net. Amplitudes are normalised by total power of spectrum. **f**, Same as **e** for BDSN.

**Supplementary Figure 6** Forward modelling of source distribution in azimuth for F-net **a** observed (red) and fitted (green) stack amplitude as a function of azimuth, for winter, compared to array response shape (black). All have been normalised to their respective maximum amplitudes. **b**, Observed (blue) and fitted (green) stack amplitude as a function of azimuth, for summer. **c**, Proportion (in percent) of excess sources as a function of azimuth needed to fit the observed amplitude variations in winter, compared to the corresponding degree 1 in the observed spectrum. **d**, Same as **c** for summer.

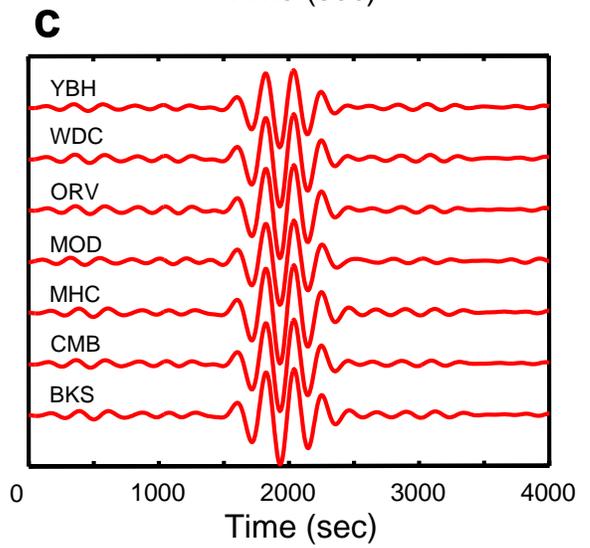
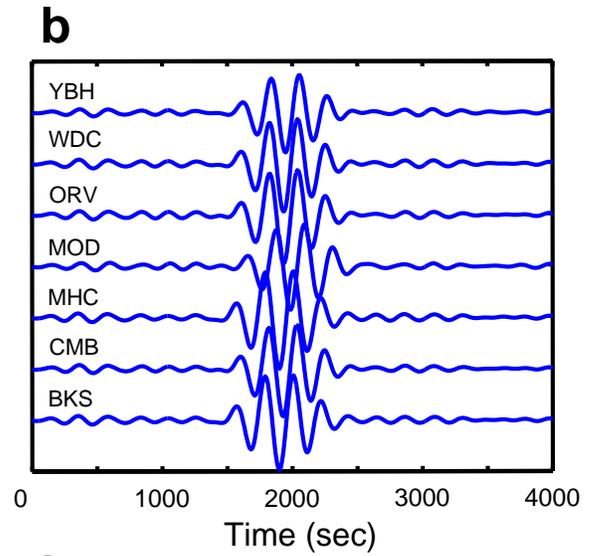
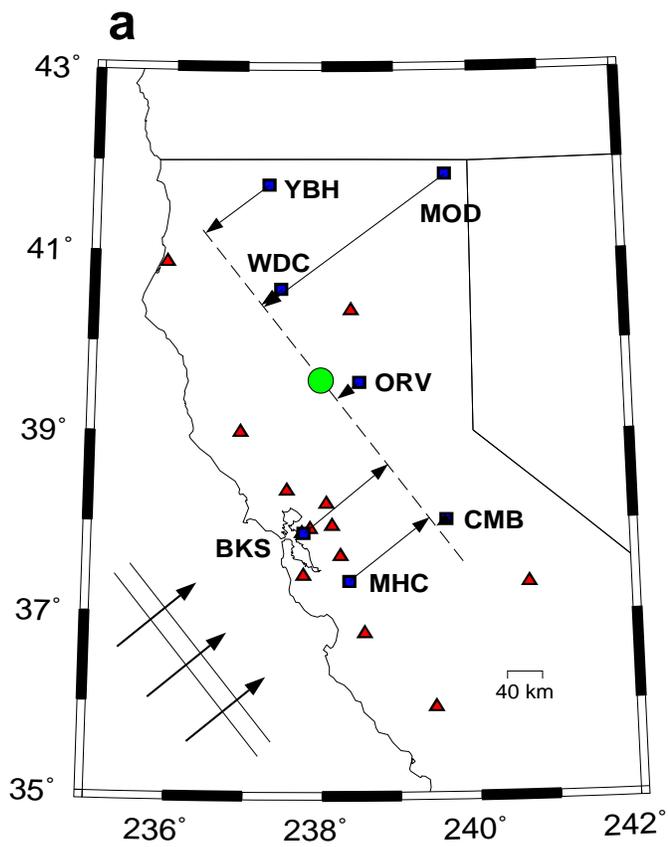
**Supplementary Figure 7** Same as Supplementary Fig. 6 for BDSN.

**Supplementary Figure 8** Results of forward modelling of stack amplitudes as a function of azimuth, for F-net (left) and BDSN (right) for a distribution of sources concentrated over different continents. Starting from a uniform distribution over the entire globe (black), the amplitude of sources of fundamental mode Rayleigh waves is increased by 100% over each region considered, successively. Model predictions (green) are compared with the predictions for a globally uniform distribution (black) and the observed distributions for winter (red) and summer (blue). Arrows point to the maxima in each distribution. **a** and **b**: sources in Eurasia (ER). This region is defined as spanning longitudes: 0°E-135°E for the

latitude range 30°N-70°N, and longitudes 45°E-120°E for the latitude range 15°N-30°N. **c** and **d**: Africa (AF). This region is defined as spanning longitudes 15°W to 45°E for the latitude range 0°N-30°N, and longitudes 15°E-45°E for the latitude range 30°S-0°S. **e** and **f**: North America (NAM). This region is defined as spanning longitudes 125°W-70°W and latitudes 30°N-70°N. **g** and **h**: South America (SAM). This region is defined as spanning longitudes 80°W-50°W for the latitude range 0°N-12°N, longitudes 80°W-35°W for the latitude range 15°S-0°S, and longitudes 75°W-50°W for the latitude range 45°S-15°S. We note that, for each of the continental regions, when the predicted maximum amplitude is compatible with one of the observed distributions (winter or summer) for one of the arrays, it is not compatible for the other array, ruling this out as a possible solution to explain the observed patterns. We also verified that the combination of several, or all continental areas does not predict distributions compatible with both arrays simultaneously.

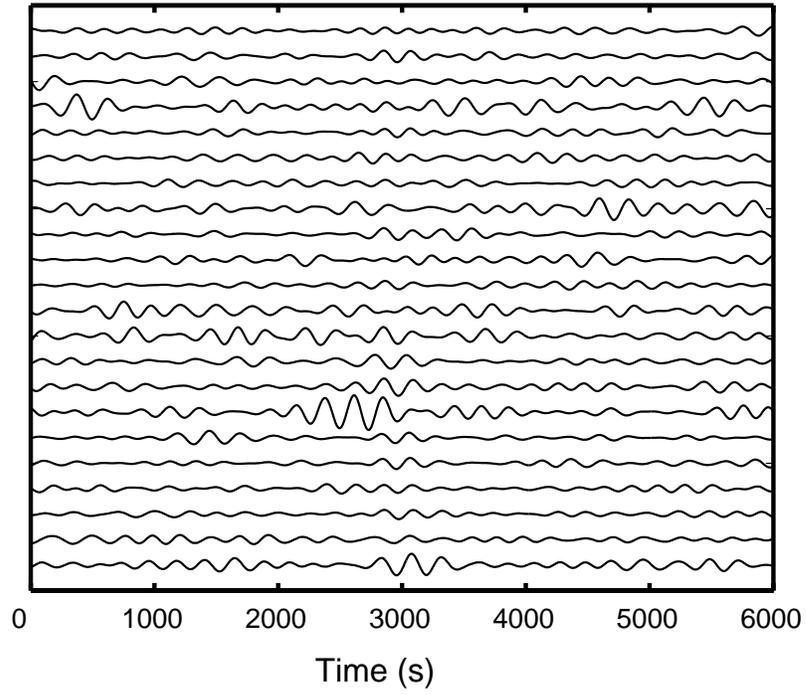
**Supplementary Figure 9** Results of forward modelling of stack amplitudes as a function of azimuth, for F-net (left) and BDSN (right) for a distribution of sources concentrated over selected oceanic areas. Predicted stack amplitudes as a function of azimuth (green curves) are compared to observed ones for winter (top, red curves) and summer (bottom, blue curves), as well as those predicted for a globally uniform distribution of sources (black curves). Starting from a globally uniform distribution of sources, source amplitudes are increased in specified oceanic areas. **a** and **b**: Winter. The region considered is the northern Pacific ocean defined as spanning latitudes 0N-60N and longitudes 150°E-135°W, with source amplitudes increased by 150% in this region. **c** and **d**: Summer. Here the region comprises part of the south Atlantic ocean (45°W-0°E and 75°S-10°S), with source amplitudes increased by 400% with respect to the

starting globally uniform distribution, and part of the south Pacific Ocean (160°E-90°W and 60°S-30°S) with source amplitudes increased by 200%. In both winter and summer, predicted amplitude variations as a function of back-azimuth are compatible with the observed amplitudes at both arrays. In order to obtain more accurate fits, we would need to perform much more refined modelling, which is not warranted, given that the observed stack amplitudes are averaged over 6 months in each season, and during this time span the distribution of sources is likely not completely stationary. There are also other sources of noise (non-Rayleigh wave related) which may influence the details of the observed curves, and which we do not take into account in our experiments.

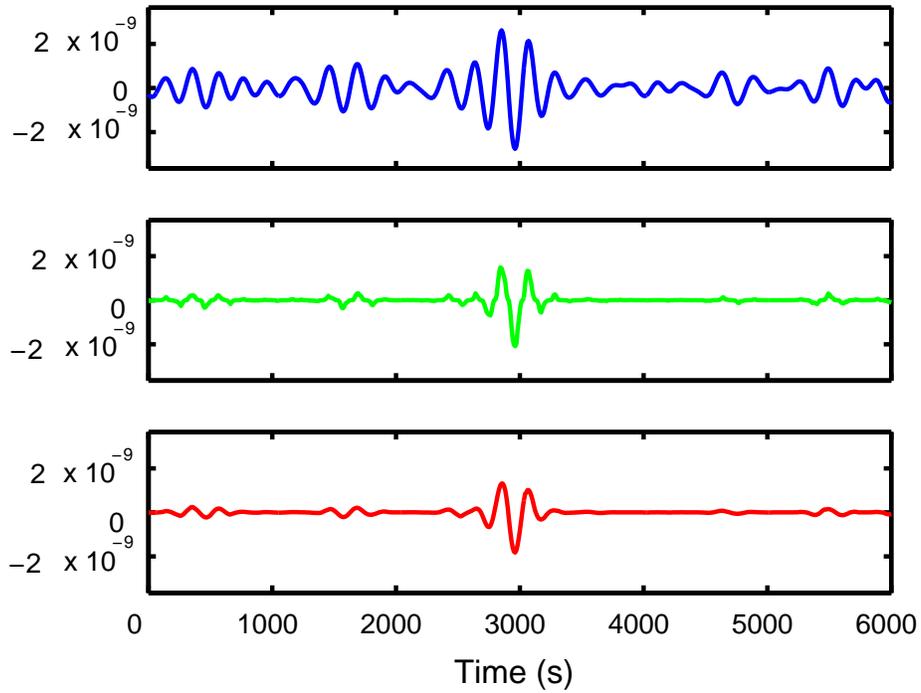


Supplementary Figure 1

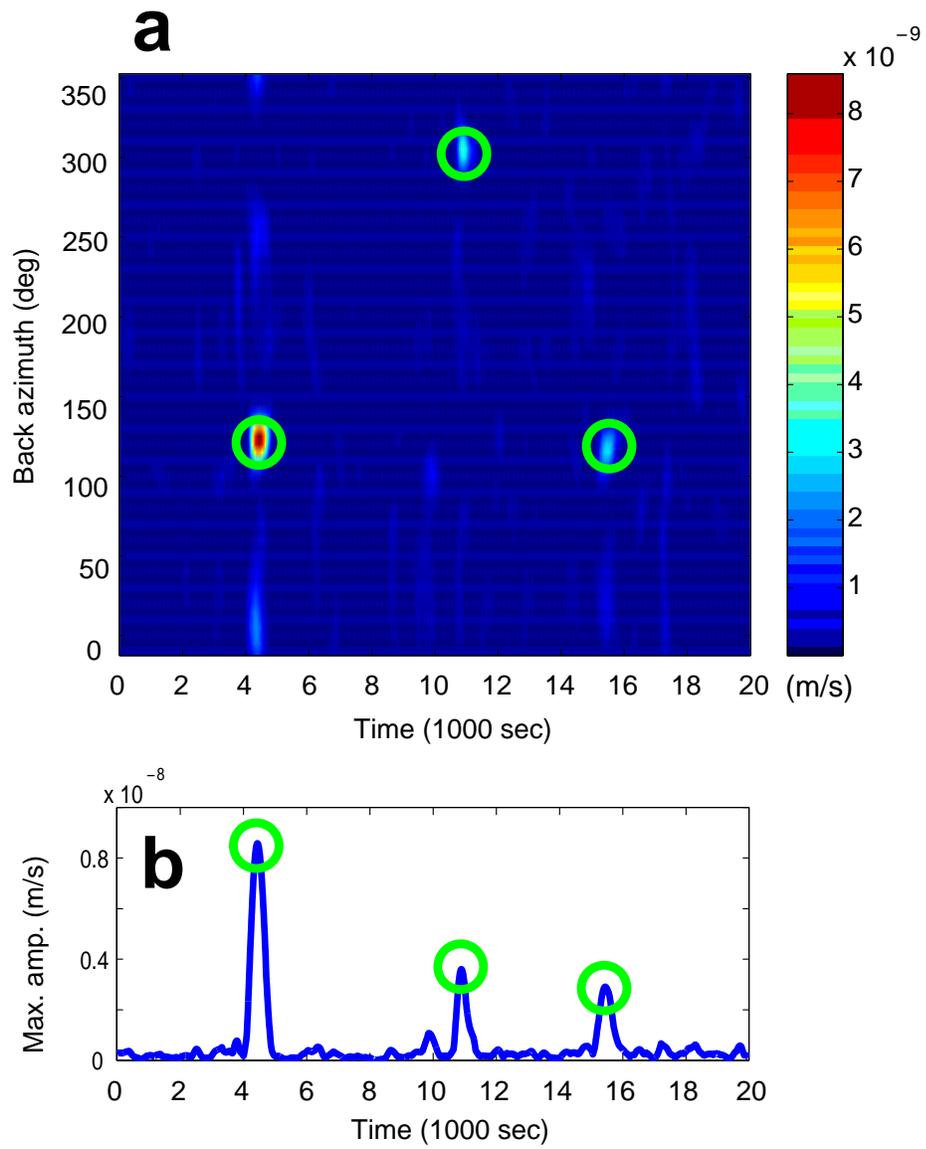
**a**



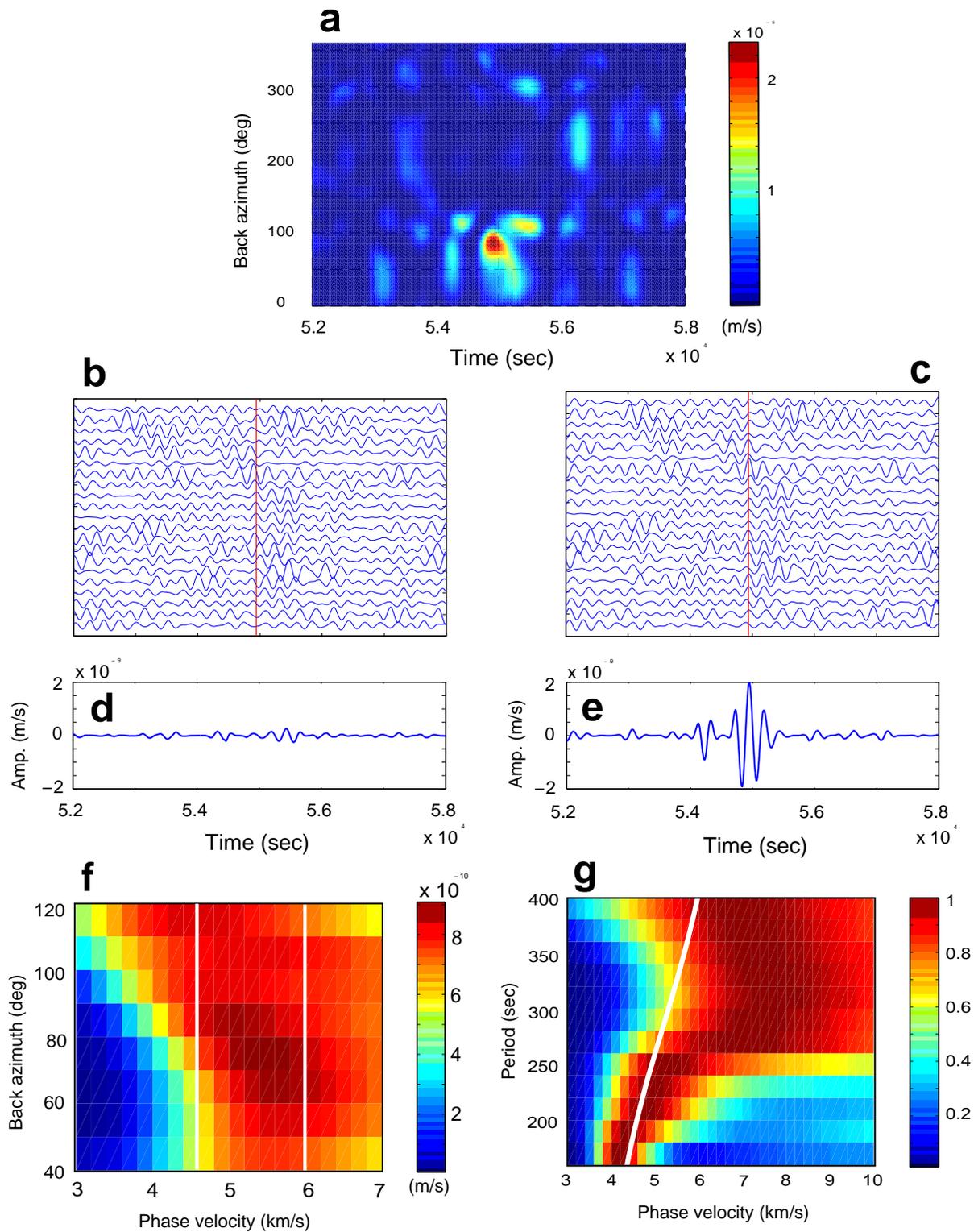
**b**



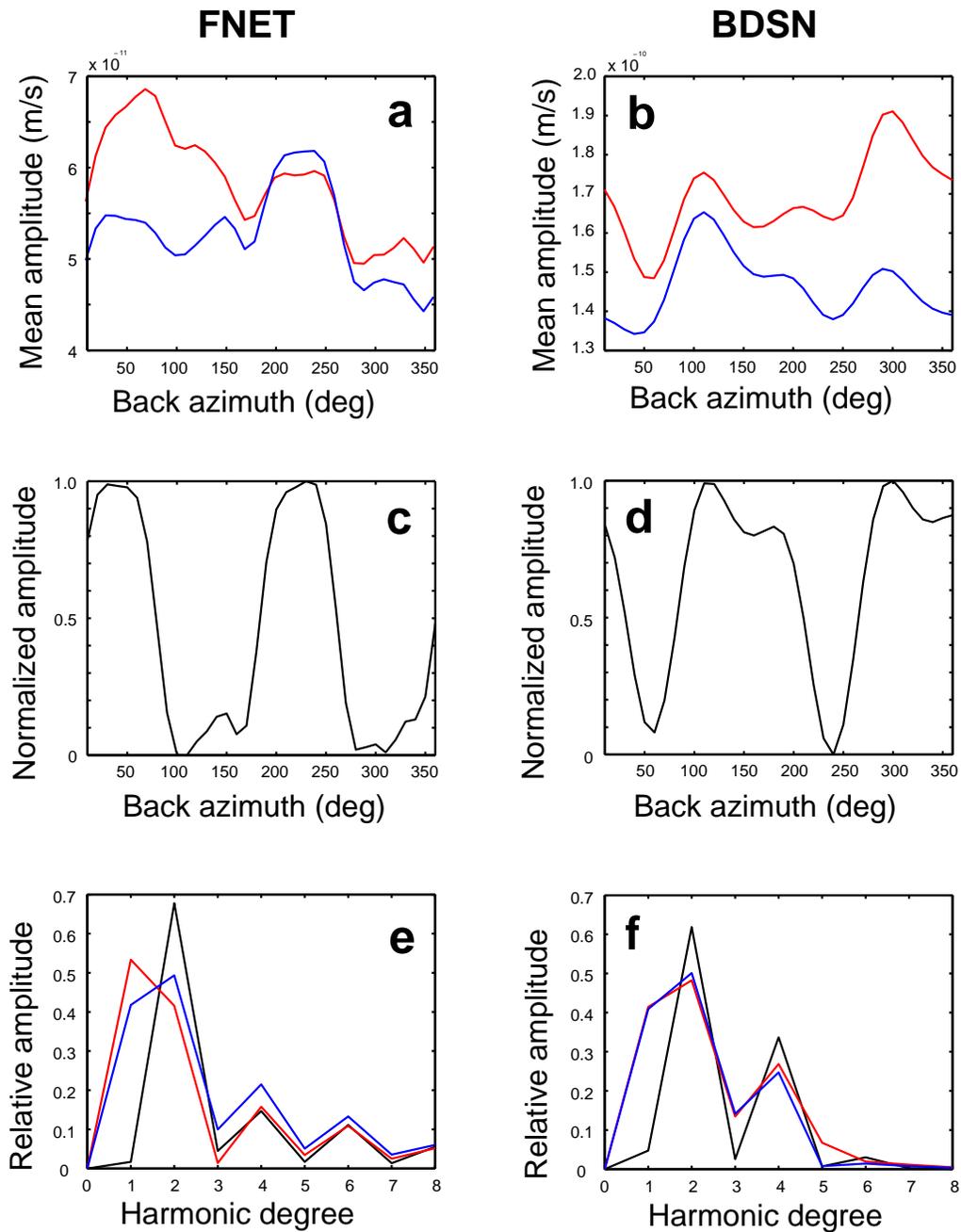
Supplementary Figure 2



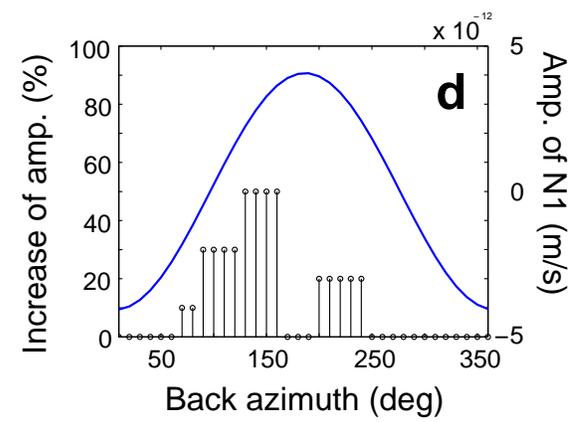
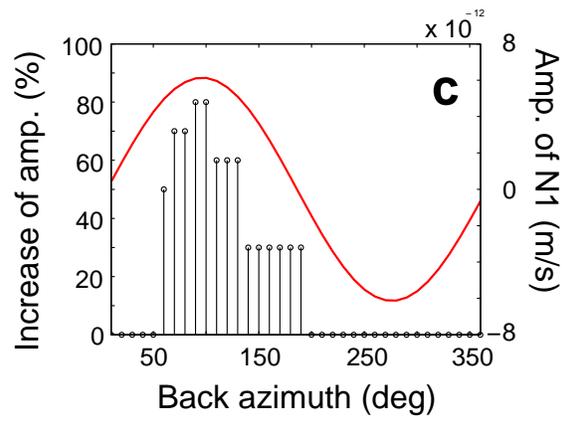
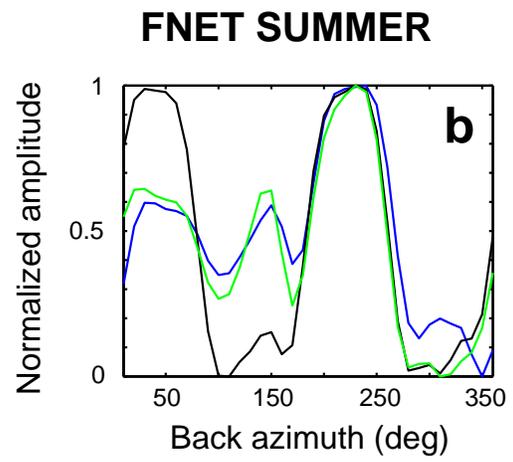
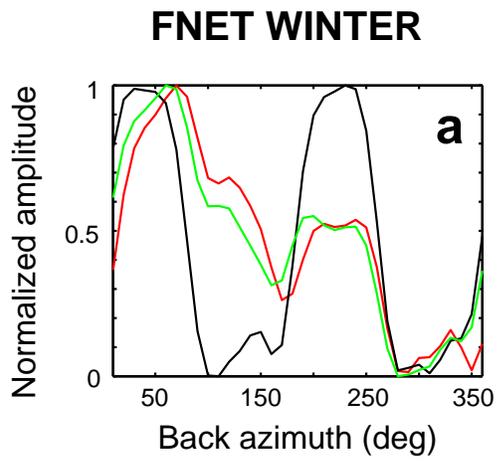
Supplementary Figure 3



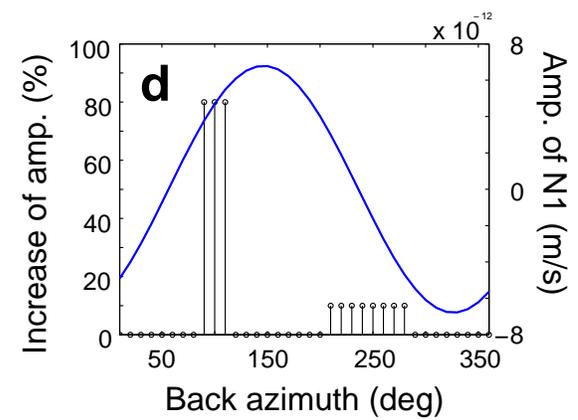
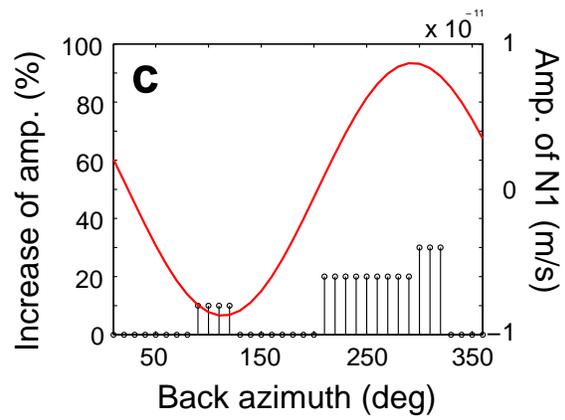
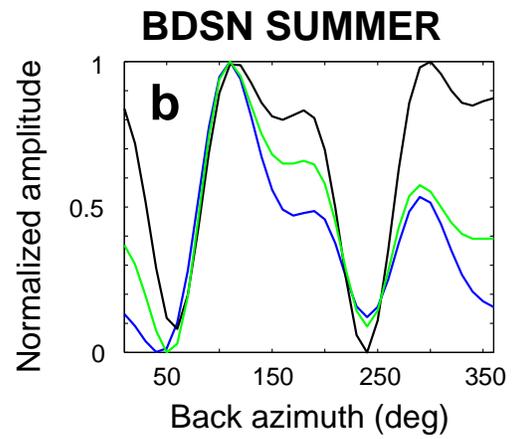
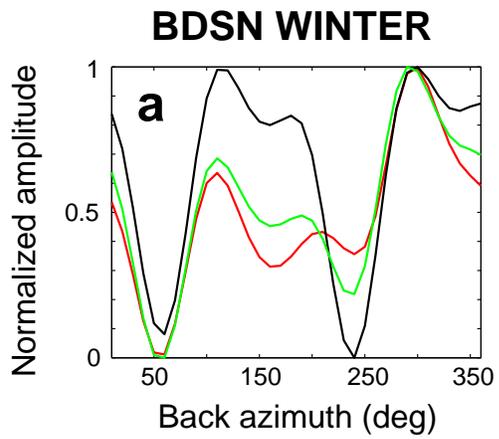
Supplementary Figure 4



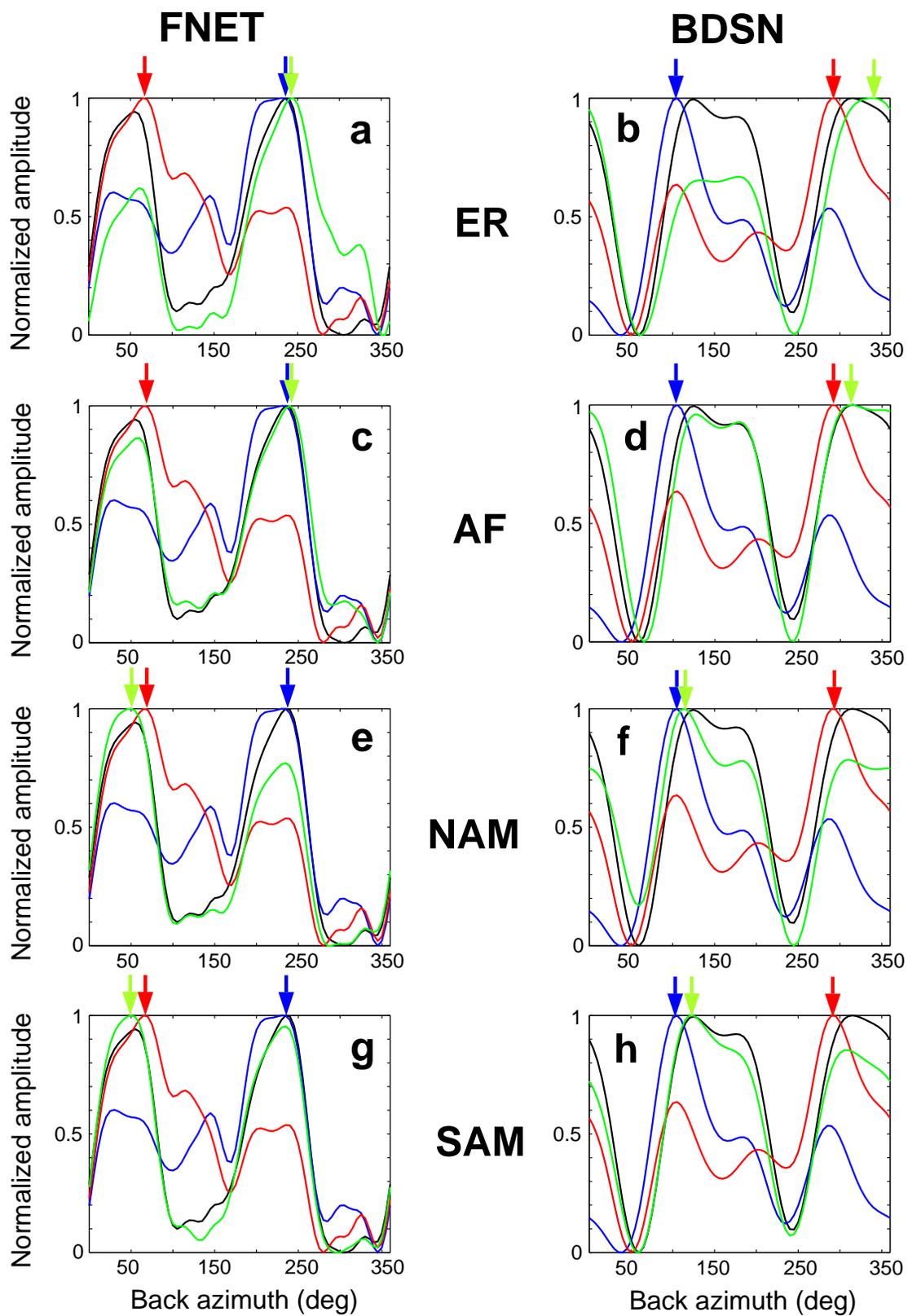
Supplementary Figure 5



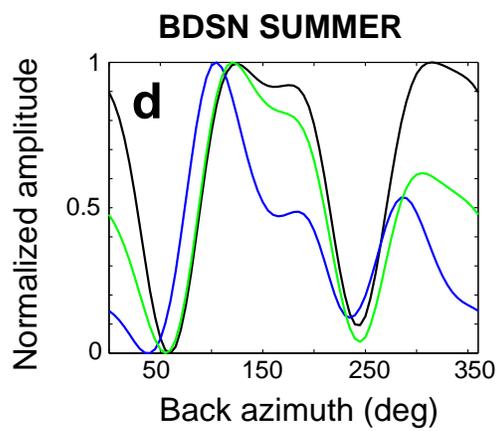
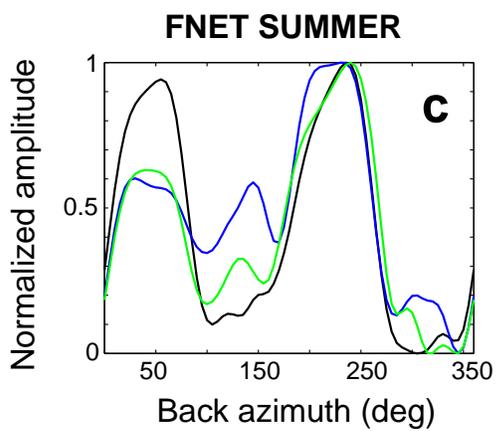
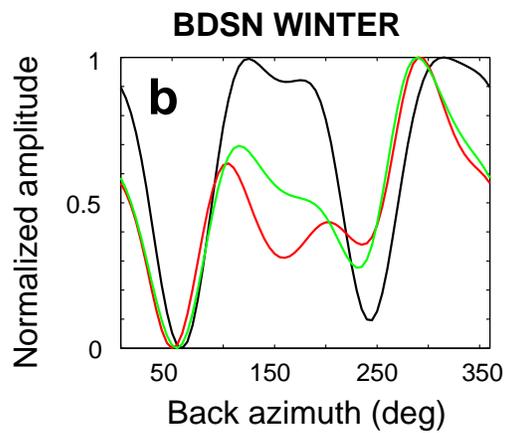
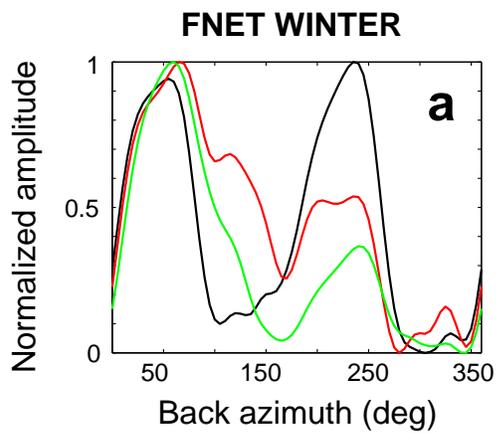
Supplementary Figure 6



Supplementary Figure 7



Supplementary Figure 8



Supplementary Figure 9