Auxiliary Material of "A window into the complexity of the dynamic rupture of the 2011 Mw 9 Tohoku-Oki earthquake"

Lingsen Meng[1*], Asaf Inbal[1,2], Jean-Paul Ampuero[1]

1. Seismological Laboratory, California Institute of Technology

2. Department of Geological and Environmental Sciences, Ben-Gurion University of the Negev, Beer-Sheva, Israel

1. Movies of back-projection results

The attached files, *jp_movie_EU_music.gif* and *jp_movie_US_music.gif*, are movies showing results of our MUSIC back-projection source imaging in the 0.5-1 Hz band. The quantities plotted and conventions are similar to those in Figure 3 of the main text.

2. Parameters of the array processing

theory

We applied a high resolution array processing technique called Multiple SIgnal

Classification (MUSIC) [*Goldstein and Archuleta*, 1991; *Schmidt*, 1986], which estimates the direction of arrival (DOA) of multiple waves propagating through an array by a subspace projection method. We first estimate the covariance matrix of the array data for each sliding time window and each frequency using a multitaper technique**Error! Reference source not found.**. The eigenvectors corresponding to its largest eigenvalues span the signal subspace. Its complement defines the noise subspace. For a given DOA, a steering vector a collects the Fourier-domain phase shift terms for each station resulting from the differential travel times relative to the hypocentral travel times. The MUSIC pseudo-spectrum, as a function of DOA, is defined as the inverse of the projection of the steering vector onto the noise subspace (through a projection matrix U_n):

$$Pm = \frac{a^H a}{a^H U_n U_n^H a} \qquad (1)$$

where the superscript H denotes the complex conjugate transpose. Ideally, the steering vector of a signal is orthogonal to the noise subspace. In practice, the DOAs of the multiple sources are identified by the largest peaks of the MUSIC pseudo-spectrum. We improve the estimate robustness by averaging over a frequency band the pseudo-spectra normalized by their respective maximum at each frequency. The absolute amplitude of the MUSIC pseudo-spectrum is not directly related to the strength of the source. Once the DOA of the sources are determined, we estimate their power by the projection of the principal eigenvalues (diagonal matrix Λ_s) onto the signal subspace (projection matrix U_s) excluding the noise noise power (square of the averaged noise eigenvalues σ) [*Mestre et al.*, 2007]

$$Pw = \frac{a^H a}{a^H U_s (\Lambda_s - \sigma^2 I)^{-1} U_s^H a} \qquad (2)$$

Data pre-processing and selection

We band-pass filtered the vertical component of the waveforms between 0.5 and 1 Hz. We then selected the stations with signal to noise ratio greater than 3. We aligned these on the first 8 seconds of P arrival by multi-channel cross-correlation[*Vandecar and Crosson*, 1990]. The seismograms are then normalized by the standard deviation of the first 200 s of P wave recordings.

Length of sliding windows

To achieve a balance between temporal resolution and imaging stability we fixed the length of the analysis windows to 5 times the longest period at which the data is analysed. For our primary frequency band (0.5-1 Hz) we slide a 10 seconds long analysis window by steps of 1 second.

Number of principal eigenvalues in MUSIC

In theory the number (M) of eigenvalues of the covariance matrix contained in the signal subspace is equal to the number of signal sources. In practice the signal is not perfectly coherent and its power is partially diverted into the subspace of smaller eigenvalues. We set M equal to the number of eigenvalues greater than the number of stations times the floating point precision scaled by the largest eigenvalue.

Number of tapers in cross-spectrum estimation

MUSIC achieves better resolution than beamforming in direction of arrival estimation problems with sufficiently long and stationary signals [*Krim and Viberg*, 1996]. However, seismic signals are highly transient. This hampers the robust estimation of the data covariance matrix. To circumvent this issue, we estimate the array covariance with the multitaper cross-spectrum estimator [*Thomson*, 1982], which averages the Fourier estimates tapered by a sequence of functions with optimal spectral concentration. The Slepian tapers maximize the ratio of energy within and outside a frequency band W for a window of duration T. We require the frequency smearing bandwidth to be at most 20% of the highest frequency (1 Hz), that is, $W = 0.2$ Hz. Following standard practice, we use the first $2TW - 1$ tapers. Hence, for the frequency band 0.5-1 Hz and $T = 10$ s we use 3 tapers.

3. Foreshocks and aftershocks

The M7.2 March 9th foreshock and the March 11th M7.1 aftershock have adequate signal to noise ratio to be aligned and back-projected by the USarray (Fig.S3). The two events appear to radiate coherent energies no longer than 30s with rupture length around 30km. The two events also have different rupture directions, indicating no systematic bias due to the array orientation. The deviation of the back-projection locations of the initial phases with respect to their epicenter is due to imperfect

alignment of the waveforms. We use the M7.1 aftershock as the empirical Green's function for the synthetic scenario tests in the next section.

4. Synthetic tests of back-projection source imaging

To understand the performance and potential bias of the back-projection we conduct tests over a series of hypothetical earthquake scenarios. To include key characteristics of the real wave field, such as the decay of the waveform coherence as a function of time, we employ empirical Green's functions (eGf). Synthetics based on theoretical Green's functions computed by ray theory and a spherically symmetric Earth model tend to over-emphasize the effect of interference between multiple sources. We chose the March 11th M7.1 aftershock as eGf event. The initial location error due to the waveforms misalignment is corrected by shifting the initial HF radiator to the event epicenter.

We first test a scenario of unilateral rupture to the South. We consider 10 uniform sub-sources located every 15 km along the strike direction (Fig. S4). Their rupture time corresponds to an assumed rupture speed of 1.5 km/s. The MUSIC back-projection recovers the location and timing of the scenario sources very well. The apparent secondary rupture to the West is an artifact due to the rupture pattern of the eGf event itself.

We also tested more complicated scenarios inspired by our final results of the back-projection of the Tohoku earthquake. Our second scenario is composed of five

different phases with various rupture speeds and directions (Fig. S5). The source process is well recovered, although perturbed by modulation by the rupture pattern of the eGf event. To minimize that effect, in our third scenario the inter-subevent spacing is exaggerated (Fig. S6). Overall, the back projection results image our synthetic scenarios adequately. In particular, the stages of different rupture speed are reproduced. One of potential bias of the back projection imaging is that when rupture temporally ceases, the coda of previous rupture would create apparent secondary sources (Fig. S7). However, these sources show substantially smaller power than the previously sources, which is unlikely to be the case of this earthquake, since we observed sources of almost uniform power during the rupture.

5. Array analysis of lower frequency recordings

To explore the potential frequency-dependence of the spatial distribution of the source radiation, we repeat the back-projection analysis of the Tohoku-Oki earthquake in two lower frequency bands, 0.125-0.25 Hz and 0.25-0.5 Hz. Fig. S8 shows a map of the peak energy radiators. Overall the lower frequency results follow the similar down-dip distribution and temporal evolution as the high frequencies (Figure 1 of the main text). The results from both arrays remain mutually consistent. However, the lower frequency results are more scattered due to the expected lower resolution of the array processing, which is inversely proportional to the wavelength.

6. Analysis of strong motion data

We examine velocity envelopes of strong motion data recorded along the north-eastern Honshu coastline in order to identify prominent phases observed at teleseismic distances. We bandpass filter the data, integrate, square and sum traces of horizontal components. The resulting velocity envelopes are averaged using a sliding median filter whose length is equal to five times the lower bound of the bandpass filter. Fig. S9a and Fig. S9b show scaled smoothed velocity envelopes for 0.5-1 Hz and 2.5-5 Hz frequency bands, respectively. We find that the prominent high-frequency bursts from locations that ruptured 33, 85, and 140 seconds after the mainshock are well resolved for frequencies as low as 0.5Hz. This observation is consistent with our results for the 5-10 Hz frequency band presented in the main text.

7. Velocity model

In the analysis of the ground motion data recorded by the K-net and KiK-net seismic networks we use the following 1D model derived from [*Takahashi et al.*, 2004]:

Depth (km)	S-wave velocity (km/s)
0	3.46
16	3.87

32	4.5

7. Data sources

We downloaded the data from the following sources:

USArray: http://www.usarray.org/

European seismic network: http://www.orfeus-eu.org/

K-net: http://www.k-net.bosai.go.jp/

Kik-net: http://www.kik.bosai.go.jp/

Italian network: http://iside.rm.ingv.it/

References

Goldstein, P., and R. J. Archuleta (1991), Deterministic Frequency-Wave-Number Methods and Direct Measurements of Rupture Propagation during Earthquakes Using a Dense Array - Theory and Methods, *J Geophys Res-Solid*, *96*(B4), 6173-6185.

Krim, H., and M. Viberg (1996), Two decades of array signal processing research - The parametric approach, *Ieee Signal Proc Mag*, *13*(4), 67-94.

Mestre, X., B. A. Johnson, and Y. I. Abramovich (2007), Source power estimation for array processing applications under low sample size constraints, *Int Conf Acoust Spee*, 897-900.1156.

172 Schmidt, R. O. (1986), Multiple Emitter Location and Signal Parameter-Estimation,
173 *Ieee T Antenn Propag*, *34*(3), 276-280.
174 Takahashi, N., S. Kodaira, T. Tsuru, J. O. Park, Y. Kaneda, K. Suyehiro, H. Kinoshita,
175 S. Abe, M. Nishino, and R. Hino (2004), Seismic structure and seismogenesis off
176 Sanriku region, northeastern Japan, *Geophys J Int*, *159*(1), 129-145.
177 Thomson, D. J. (1982), Spectrum Estimation and Harmonic-Analysis, *P Ieee*, *70*(9),
178 1055-1096.
179 Vandecar, J. C., and R. S. Crosson (1990), Determination of Teleseismic Relative
180 Phase Arrival Times Using Multi-Channel Cross-Correlation and Least-Squares, *B*
181 *Seismol Soc Am*, *80*(1), 150-169.
182
183

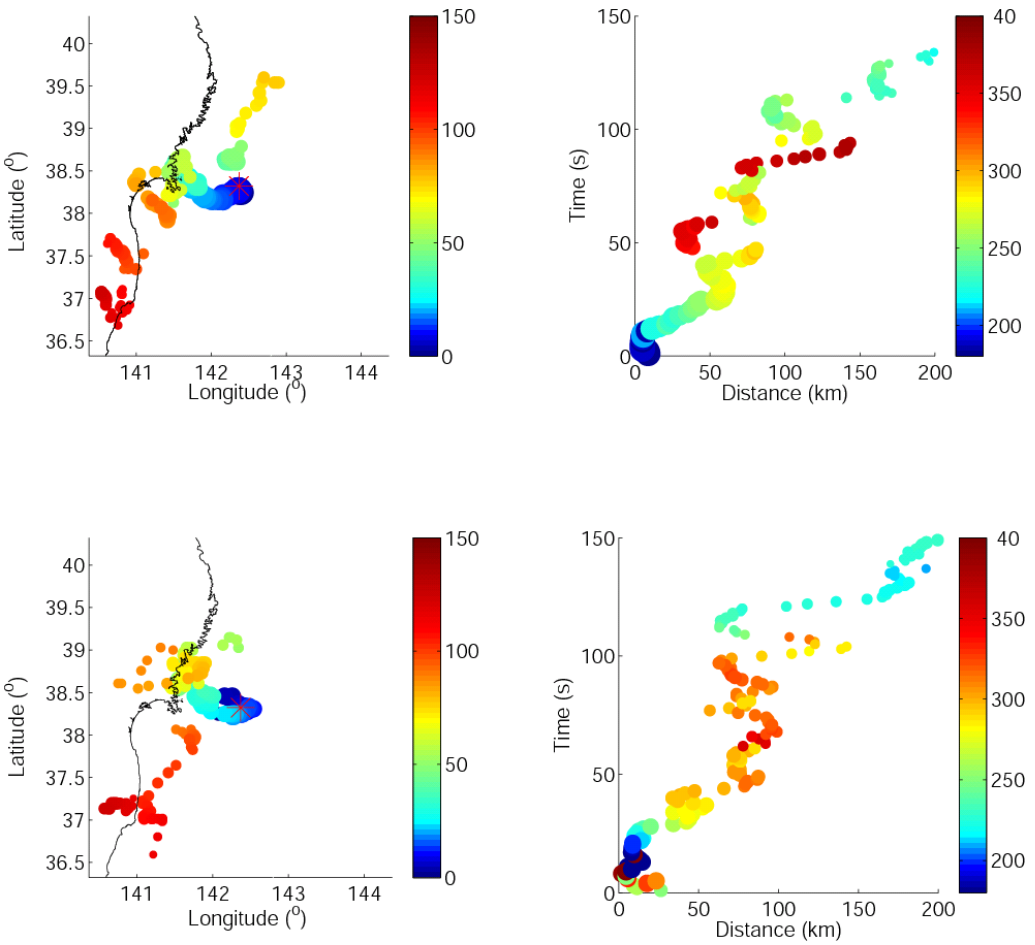


Figure S1. Back projections using correlation stacking at 0.5-1 Hz**.** *Left*: location of the strongest HF radiators color coded by rupture time. *Right*: rupture time versus epicentral distance, color coded by azimuth relative to the epicenter. *Top*: results based on European data. *Bottom*: results based on USarray data.

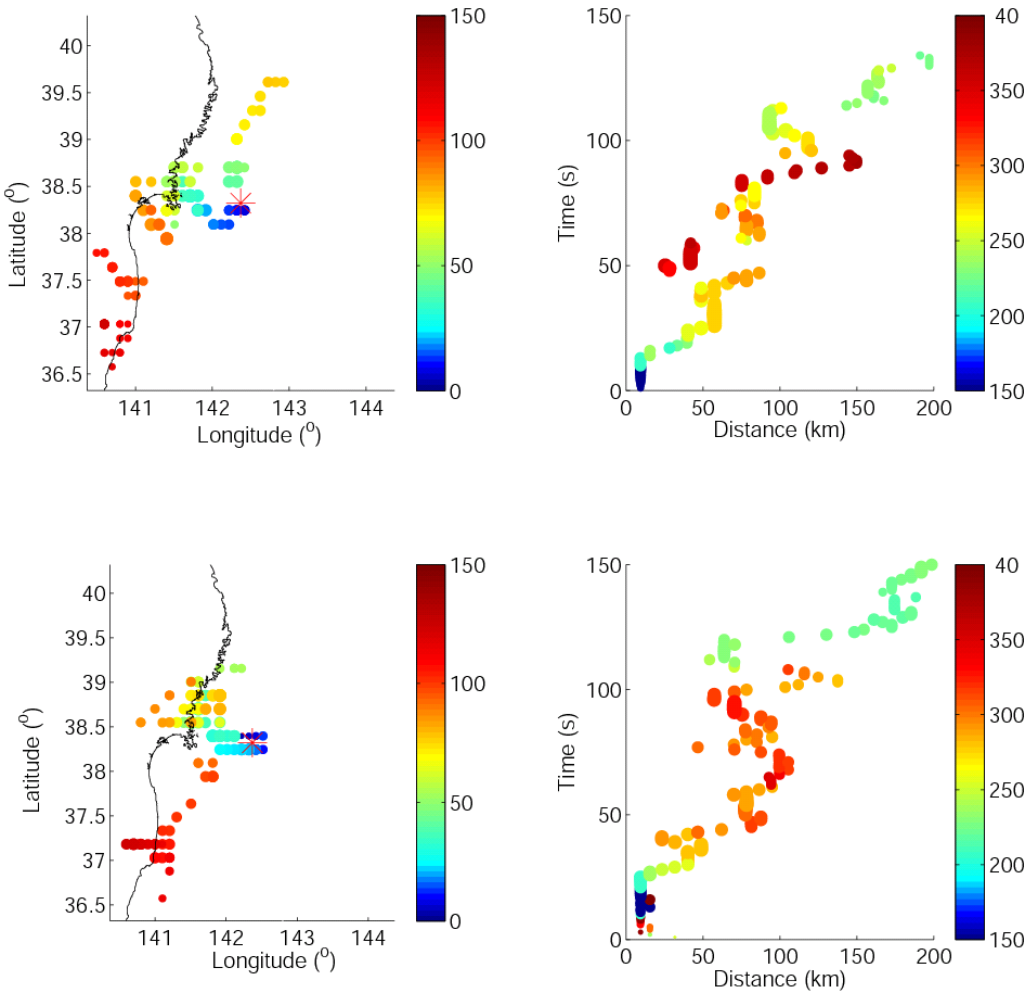


Figure S2. Back projections using cubic root stacking. *Left*: location of the strongest HF radiators color coded by rupture time. *Right*: rupture time versus epicentral distance, color coded by azimuth relative to the epicenter. *Top*: results based on European data. *Bottom*: results based on USarray data.

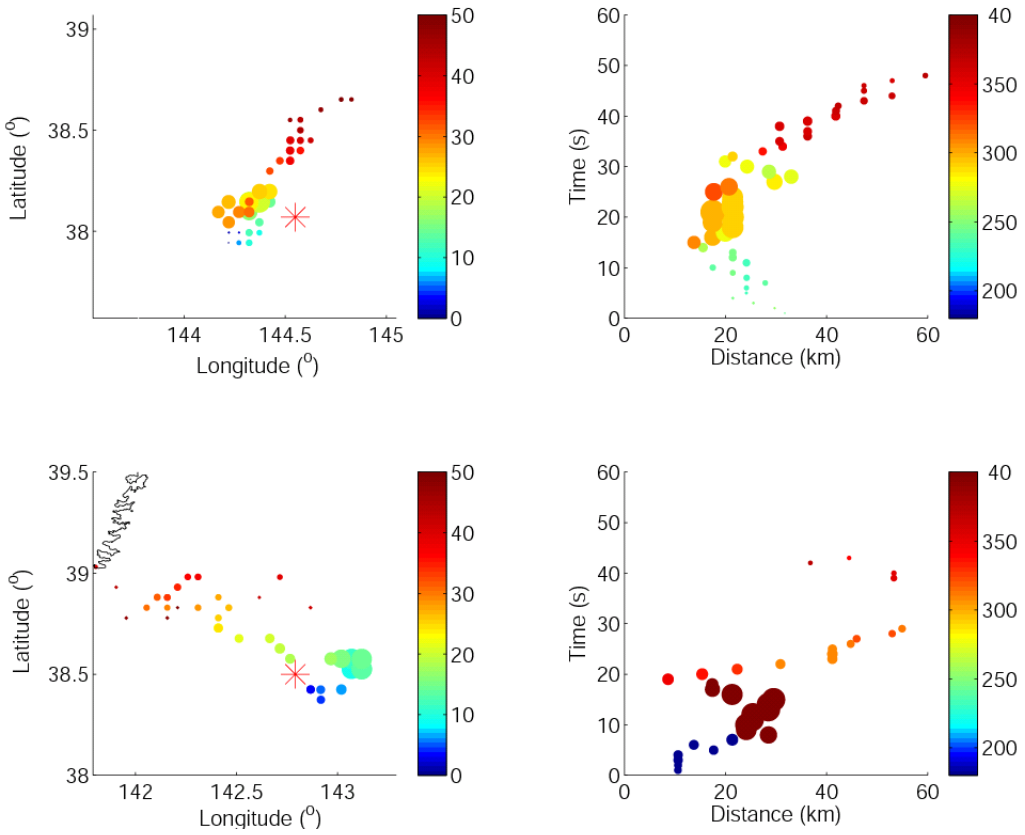


Figure S3. Foreshocks and aftershocks processed by the MUSIC technique at 0.5-1Hz, based on USArray data**.** Same convention as as in figures S1 and S2. *Top*: the March 9th M7.2 foreshock. *Bottom*: the March 11th M7.1 aftershock.

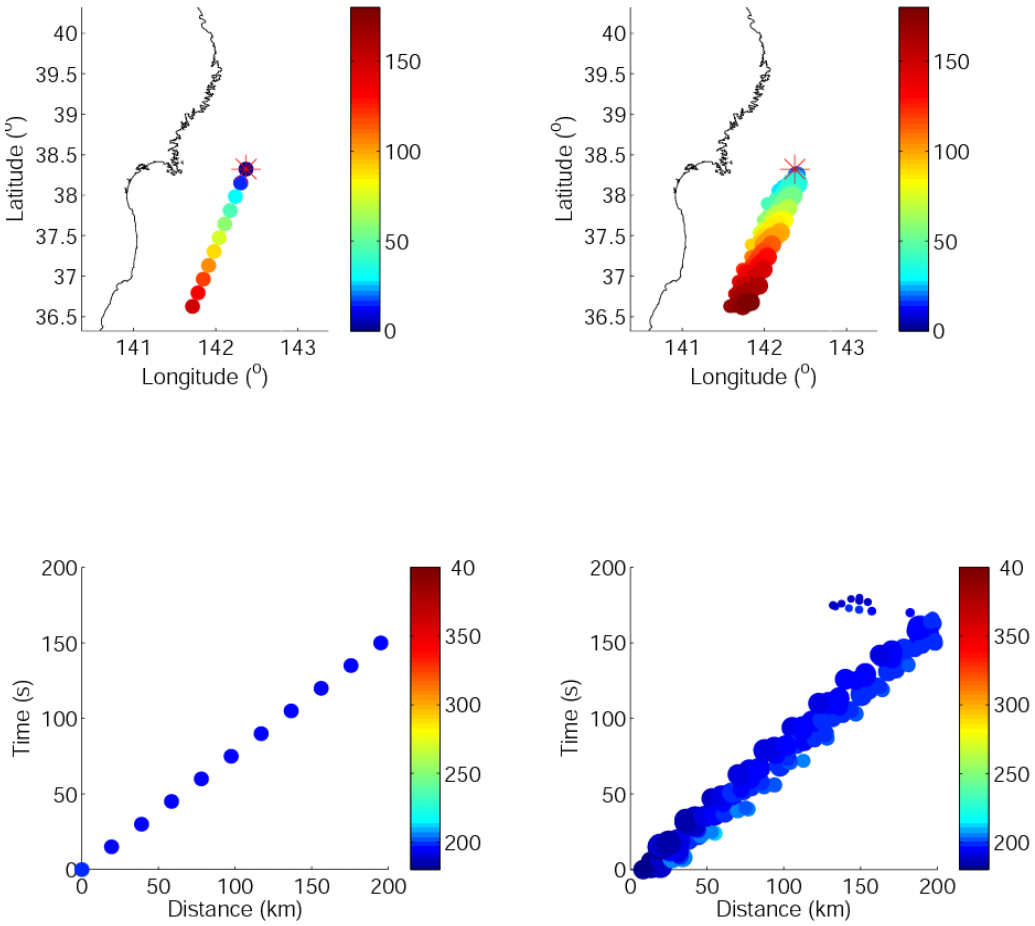


Figure S4. Synthetic test 1. Same conventions as in previous figures. *Left*: synthetic source scenario. *Right*: source model recovered by MUSIC back-projection.

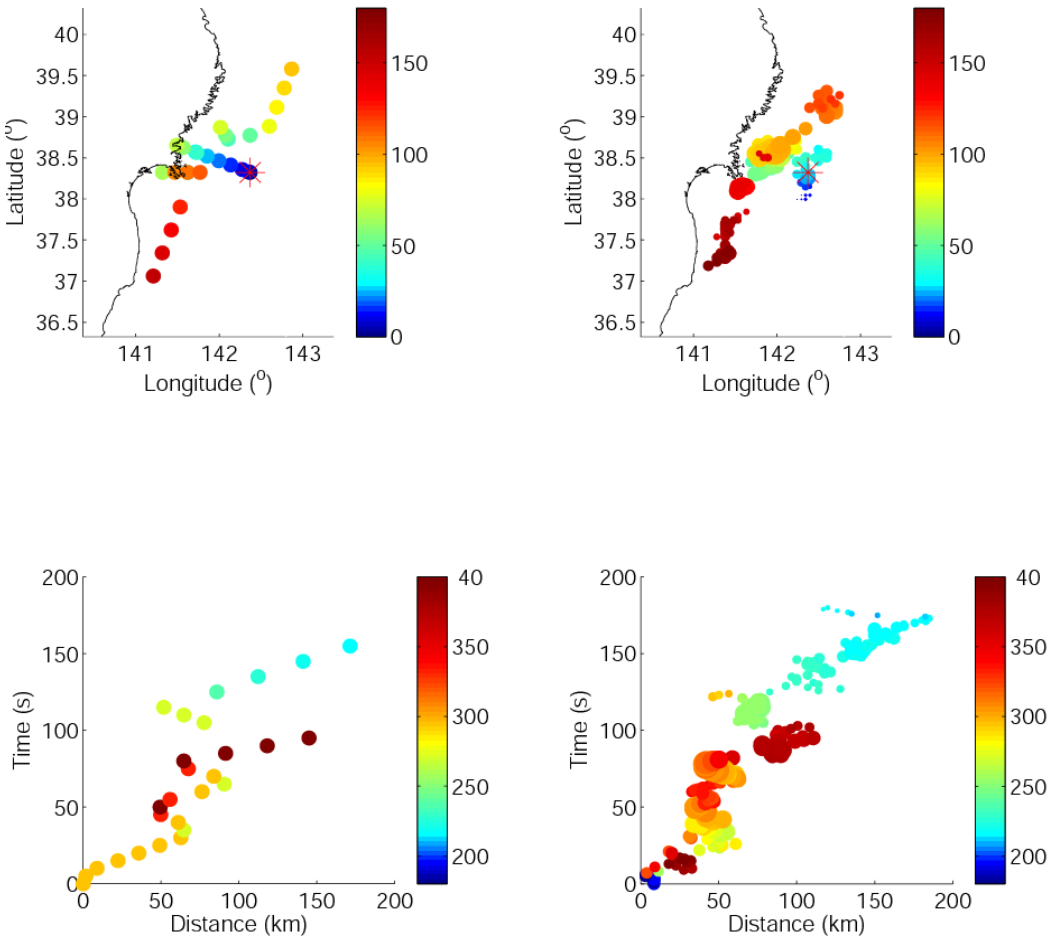


Figure S5. Synthetic test 2. Same conventions as in previous figures. *Left*: synthetic source scenario. *Right*: source model recovered by MUSIC back-projection.

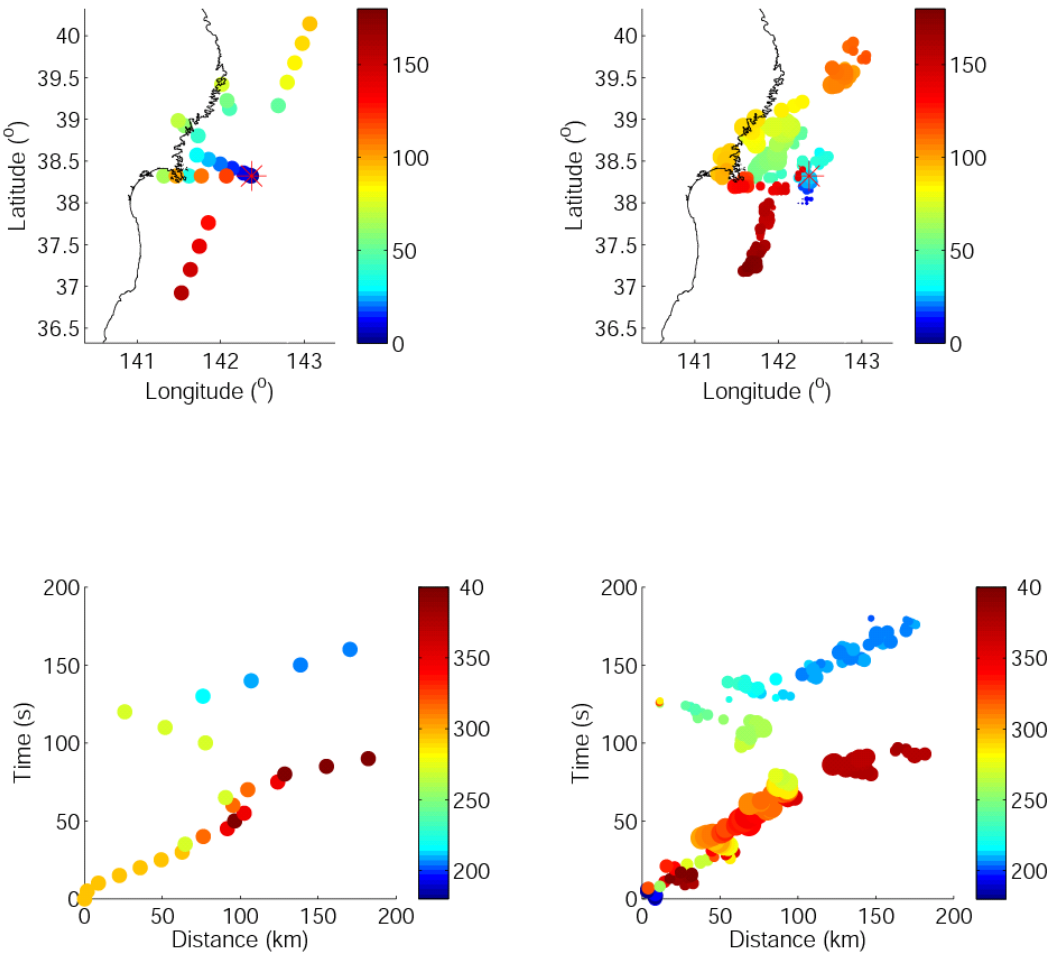
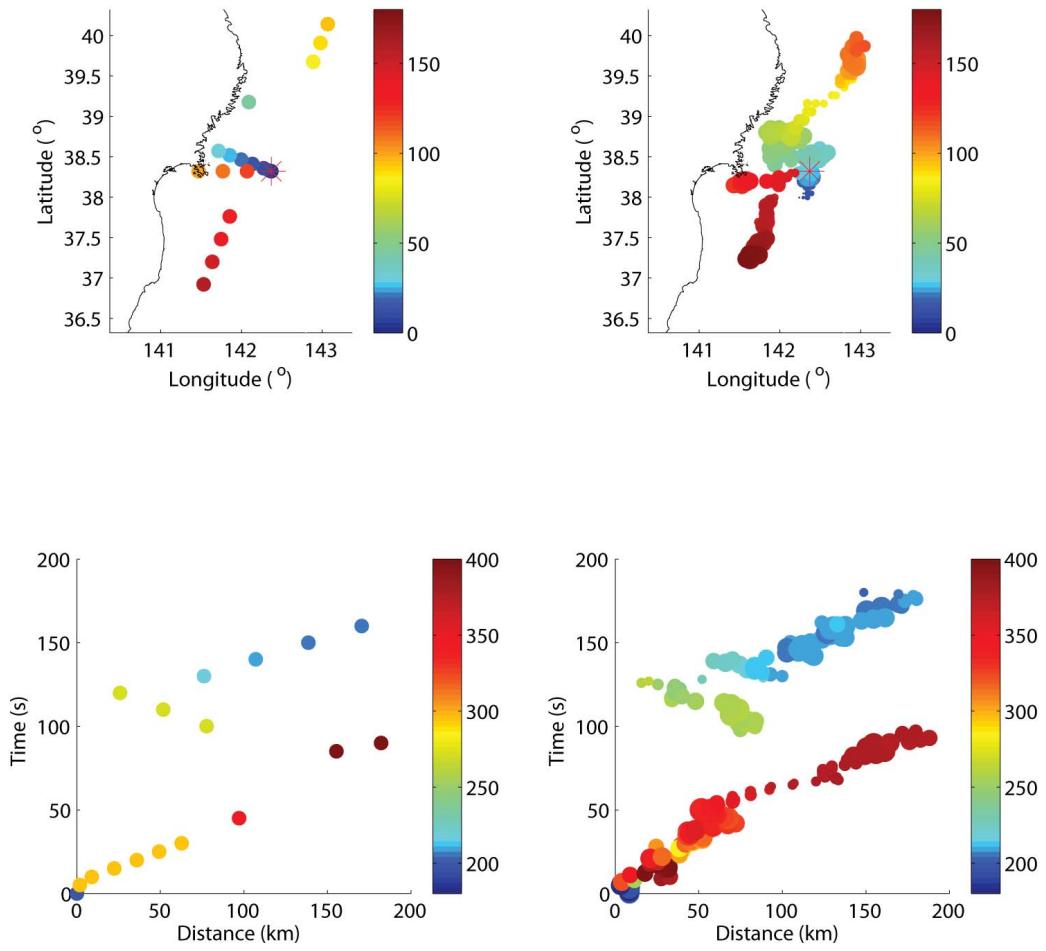


Figure S6. Synthetic test 3. Same conventions as in previous figures. *Left*: synthetic source scenario. *Right*: source model recovered by MUSIC back-projection.



208

209 **Figure S7.** Synthetic test 4. Same conventions as in previous figures. *Left*: synthetic

210 source scenario. *Right*: source model recovered by MUSIC back-projection.

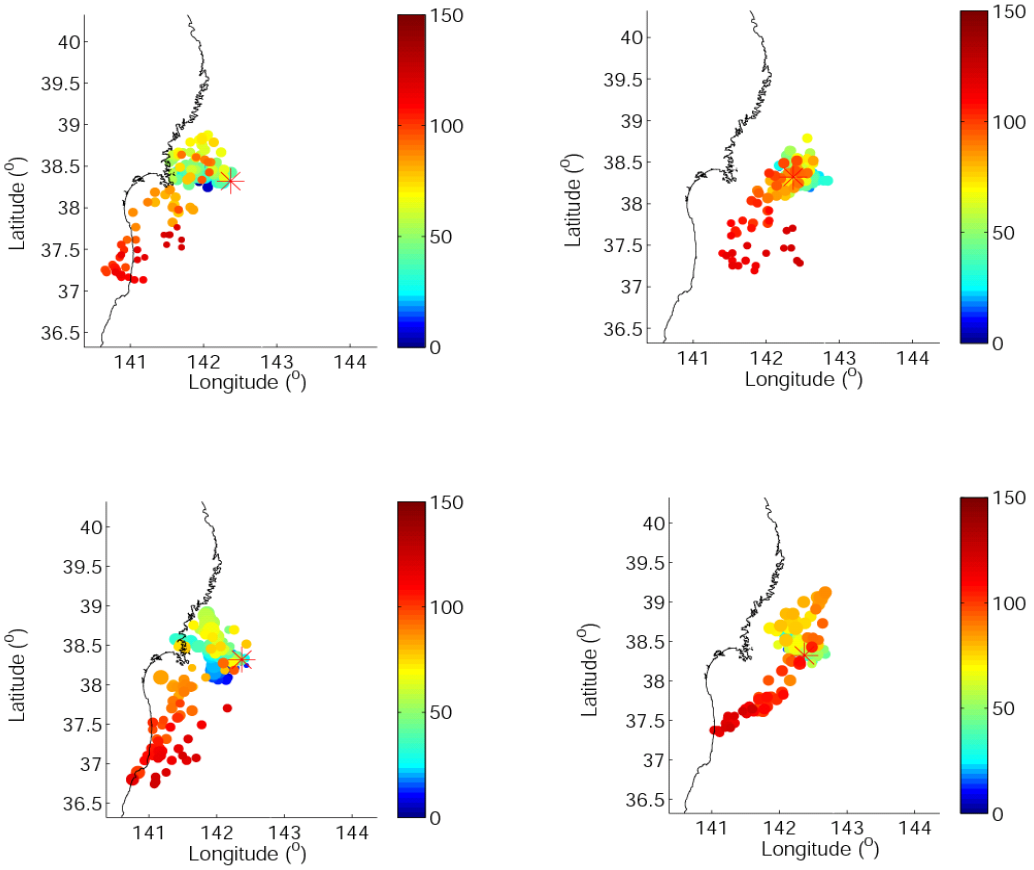


Figure S8. Array analysis of lower frequency recordings. Location of the strongest radiators color coded by rupture time. *Top*: 0.125 to 0.25 Hz. *Bottom*: 0.25 to 0.5 Hz. *Left*: USarray. *Right*: European network.

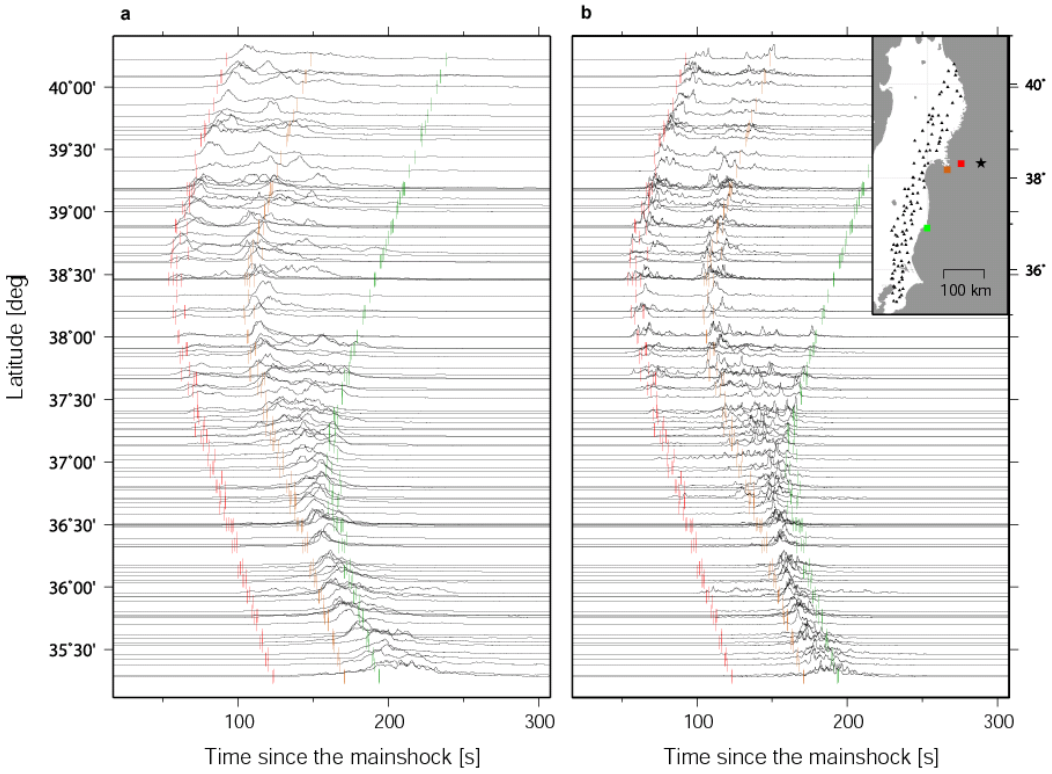


Figure S9. Strong motion recordings at lower frequency band. Smoothed 0.5-1 Hz (left) and 2.5-5 Hz (right) velocity envelopes of strong motion data recorded along the north-eastern Honshu coastline. Traces are scaled by their maximal amplitude. Star and triangles in the inset map denote the epicentral location of the Tohoku Mw9.0 mainshock (USGS catalog reference) and the recording stations, respectively. Red, yellow and green vertical bars along the time series indicate the the S-wave travel times from asperities that ruptured 33s, 85s and 140s after that mainshock, respectively. The asperities locations are indicated by squares of corresponding colors in the inset map.