FINAL TECHNICAL REPORT

Location of Triggered Tremor Near the Calaveras Fault

Department of Geophysics
397 Panama Mall
Stanford University
Stanford, CA, 94305-2215

Telephone: (650) 723-4746; fax: (650) 725-7344

Principal Investigator: Gregory C. Beroza (beroza@stanford.edu)

Keywords: Tremor, Calaveras Fault

Program Element III: U. S. Geological Survey National Earthquake Hazards Reduction Program Award Number G10AP00016

Project Period: 1/2010-12/2011

Research supported by the U.S. Geological Survey (USGS), Department of the Interior, under USGS award number G10AP00016. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Government.

Technical Abstract

We developed a new approach to detecting similar seismic signals occurring at different times for application to extracting discrete LFEs from continuous tremor. We use the PageRank algorithm on seismogram windows for which we measure similarity using waveform cross correlation. PageRank allows us to identify multiply linked (i.e., more than pairwise) waveforms that repeat most frequently. By doing so we take advantage of the fact that LFEs repeat not once, but many times during tremor episodes. This allows us to extract weaker signals of lower *snr* than would otherwise be possible. We demonstrate the effectiveness of this approach by applying it to triggered tremor that has been independently, and previously studied. We conclude that it should be possible to detect and locate low frequency earthquakes on the Calaveras Fault, where tectonic tremor was triggered by waves from the magnitude 7.9 Denali, Alaska earthquake.

Publications resulting from this grant:

2014 Aguiar, A. C. and G. C. Beroza, PageRank for Earthquakes, *Seismol. Res. Lett.,* **85**, 344-350, doi: 10.1785/0220130162.

Award # G10AP00016 *Location of Triggered Tremor Near the Calaveras Fault*

Gregory C. Beroza
Department of Geophysics
397 Panama Mall
Stanford, CA, 94305-2215

Phone: (650) 723-4746
Fax: (650) 725-7344
beroza@stanford.edu

Non-Technical Abstract

We developed a new approach to detecting similar seismic signals occurring at different times. We use the PageRank algorithm, previously used to rank web pages, to identify patterns in seismic waveforms that repeat most frequently. By doing so we take advantage of the fact that certain earthquake signals that are difficult to identify (called low frequency earthquakes) repeat many times. We demonstrate the effectiveness of this approach by applying it to previously studied tremor and conclude that it should be possible to detect and locate low frequency earthquakes on the Calaveras Fault, where seismic tremor was triggered by waves from the 2002 *M* 7.9 Denali, Alaska earthquake.

Triggered Tremor on the Calaveras Fault.

Our research focuses on triggered tremor, which is deep, non-volcanic tremor triggered dynamically during the passage of surface waves from distant, large earthquakes. Triggered tremor was reported in several areas of Northern California, including an area centered on the Calaveras Fault (*Gomberg et al.,* 2008). The locations determined to date for triggered tremor are approximate – perhaps not even accurate enough to associate it unambiguously with the Calaveras Fault in this case. The goal of the proposed research is to apply tremor location techniques that our research group has developed while studying spontaneous tremor in Japan, to triggered tremor in California, and by doing so to develop a better understanding of the deep roots of fault zones, and what relationship tremor might have to ordinary earthquakes.

In the course of our research, we discovered that of the signals thought to have been triggered tremor on/near the Calaveras Fault during the wavetrain of large teleseisms, only the signals following the Denali earthquake are clearly legitimate tremor detections. We applied our autocorrelation technique to this signal, but were unable to extract reliable LFE signals. We attribute this to the fact that, unlike the case where we are matching known LFEs of high *snr* against tremor signals of low *snr*, in this case both signals we are comparing have low *snr.* To improve our chances for extracting LFE signals from tremor, we developed a new approach that uses the fact that LFEs repeat within tremor signals many times, rather than just once, to identify those signals most characteristic of LFEs. We show that this works component-by-component on known tremor data from another location. The PageRank statistics we use to accomplish this also provide a possible alternative tremor detection approach.

PageRank for Earthquakes.

Low-frequency earthquakes (LFEs) have proven to be critical to understanding tectonic tremor as repeated shear slip on the deep extension of known faults. *Shelly et al.* (2006) showed that LFEs correlate strongly with times of tremor and that they occur on the plate interface. *Shelly et al.* (2007a) showed that tremor consists of swarms of LFEs, and by inference occurred on the plate interface as well. Follow-up work (*Ide et al.*, 2007) demonstrated that LFEs per tremor occurred as shear slip and hence were part of a family of slow earthquakes (*Ide et al*., 2007).

Initially, LFEs were found through visual inspection by network analysts in Japan (*Beroza and Ide*, 2011). LFEs have been identified as periods of high-amplitude arrivals by the same visual inspection technique (*Shelly*, 2009). Other methods have been developed to identify them. First, LFEs from the Japan Meteorological Agency (JMA) catalog were used as templates to find other LFEs by cross correlating the LFE signal with tremor (*Shelly et al*., 2007a,b). For cases where no previous templates exist, *Brown et al.* (2008) developed an autocorrelation method to find the LFEs. This method considered each small window as a possible detection by cross correlating each window with all the others (*Brown et al*., 2008) and searching for repeats. These LFE detection methods do not exploit the fact that LFEs repeat, with similar waveforms, not once, but multiple times.

We take advantage of the similarity of LFEs within tremor by developing a test of significance that explicitly accounts for short-time correlations in the data, and the likelihood that the sources repeat more than once. To do this, we use the structure of how similar windowed waveforms are linked with one another. A complicating factor is that these links do not form closed loops that satisfy closure, but have links that may be incomplete. Despite this, we would like to take advantage of the structure of similar seismograms. Our problem maps very closely to one that has been solved previously.

One of the initial web page searching algorithms implemented by Google used links between pages to calculate the quality of a document. This data-mining algorithm is known as PageRank and was developed by *Page et al.* (1999). If we substitute candidate windows in place of web pages, and we know that these windows are linked to one another, we can use the PageRank algorithm to rank them. In applying PageRank to tremor data, we find the ranking of a specific window and show how windows rank relative to each other. This tells us which windows are most likely potential LFEs, and can be used to construct templates through waveform stacking. These stacks provide high *snr* templates that can be used to find other LFEs.

We demonstrate the technique on the April 2006 Shikoku, Japan, tremor episode due to the wealth of available information concerning tremor in this area (*Obara*, 2002; *Shelly et al*., 2006, 2007a; Ide, 2010). This specific tremor episode has been analyzed in great detail (*Shelly et al*., 2007b). By applying the PageRank approach we create robust LFE templates that match known LFEs from the JMA catalog. Using these templates we find detections in data for the April episode with similar results to *Shelly et al.* (2007b) for the same time periods. We also find detections for weaker segments of tremor that were not previously reported. Our approach suggests a new approach to differentiate tremor signals from noise for sparse data sets using the fact that PageRank behavior is qualitatively different for tremor versus noise.

Multiplicity of Repeats.

To find LFEs within tremor data, we begin with the autocorrelation method of *Brown et al.* (2008), which detects potential LFEs in a pair-wise manner. This method finds the signals by autocorrelating each window with all other windows in time during a tremor segment of interest. We divide the tremor data in 10 s windows lagged by 0.08 s or 2 samples for data with 25 samples per second. This creates a population of 44,900 windows for 1 hour of data. The population of values for the correlation coefficients (CC) between these windows follows a normal distribution (Figure 1). We applied a Fisher transform to the CC values, where:

$$z = \frac{1}{2}\left[\ln(1+\mathrm{CC}) - \ln(1-\mathrm{CC})\right]$$

With this, we can assume normal statistics for the distribution of the CC values. The transformation, while technically necessary, did not make too much difference to the distribution. We base our detection threshold on a Gaussian distribution with zero mean and focus on the large positive values to declare a positive detection. For a normal distribution, σ = 1.253 x mean absolute deviation (MAD). MAD is the mean of the absolute deviations of a set of data about the data's mean and this measure is not sensitive to outliers. Our objective is to find the positive outliers in the data, which represent the repeating signals within the tremor data. Given that the MAD is not affected by outliers it should not bias our estimate of σ, so we can use this as a measure to establish a threshold of detection. By assuming zero mean (Figure 1), we can calculate the MAD directly from the population of CC values and estimate σ from this. Here we have chosen to use the mean instead of the median which has been preferred on previous analysis. We found that for a very large set of window pairs the computational costs of calculating the median are much greater than to calculate the mean which can be easily computed as crosscorrelations are computed without the need of storing all values until the end of the processing. With this, we perform the autocorrelation to find detection pairs for each of the stations analyzed with a low positive threshold of 3σ, which corresponds to a two-sided significance level of 99.7%. This level, which we use to define a positive correlation, is a trade-off between a higher threshold, which will provide more confident matches at the cost of few positives for our low *snr* data, and a lower threshold, which will provide more positives, but less confidence in the reliability of individual matches. The noise level, and therefore the statistical behavior of the autocorrelations, varies between stations so we perform the analysis one station at a time.

Finding detection pairs within the tremor data does not signify closure because LFEs repeat multiple times during a single tremor episode (*Shelly et al.* 2007a) and the *snr* is low. Figure 2 shows this lack of closure schematically. Window A correlates significantly with Window B and Window D. Later in time, Window B matches with Window C, but Window A does not match with Window C, and Window B does not match with Window D. Such a lack of closure is inevitable with low *snr* signals, and complicates the detection statistics, yet we have to make the best of the information we have. That is, we want to use all the links, and the complex hierarchical relationships they express, to identify the windows most likely to represent LFE signals.

We use a tool developed for data mining to address this problem. Specifically we apply the PageRank algorithm to the detection pairs from the autocorrelation process to determine which windows have the most number of links. We then calculate a ranking for each window and if the probability is high, this suggests a robust LFE detection for that window within the data.

PageRank.

The PageRank method calculates the probability of importance of each web page in a set of web pages, assembled in a vector. These probabilities are based on the number of links each web page has, giving a high probability to web pages with high number of links and low probability to pages with low number of links. Specifically, each of the elements of this vector is the probability that a "random surfer" will visit a particular page on the web (*Page et al.*, 1999; *Moler*, 2011). Like a Markov Chain, PageRank is a random, iterative method where the probability at one stage in the iterations is computed from the previous stage, with an initial condition of equal probabilities for all pages.

To test our method, we use data from 8 stations in Shikoku Japan (Figure 3) during the April 2006 tremor event under Shikoku. This group of stations was selected due to the locations of a large number of well-studied LFEs that have occurred within this area. We analyzed 1 hour of data at 25 samples per second from April 16, 2006, where a large number of LFEs have been previously detected using templates. We chose a time period where the tremor signal is small at the beginning and increases in amplitude towards the end (Figure 4) to understand differences in behavior of PageRank. Using these data, we first calculated the autocorrelations using the population of 44,900 windows created from the 10 sec windows lagged 2 samples. We then applied the PageRank approach to the detection pairs to find the probabilities for each window. Figure 4 Shows the PageRank values for each window in one hour of Japan data for several stations.

Once we have computed the PageRank for each window, we know which windows have the highest probabilities of being repeating signals, and are the most likely to be LFE waveforms. A large PageRank value here signifies a window with a large number of links to other windows, both direct and indirect. Figure 4 shows that the PageRank values get higher as the tremor amplitude increases for all three stations shown. It is also noteworthy that there are high values towards the beginning where tremor amplitude is smaller.

Using the windows with the highest PageRank values we create template signals by stacking the matched windows. To exploit the multiplicity of the LFEs, we use both the direct and indirect matches of the high ranked windows (Figure 4) to create the stack for the template LFE. Figure 5 shows a stack with all the windows that were found to be a match to a window with a high number of matches for station YNDH.N during the one hour of tremor data analyzed. On the left, the stacked signal (Figure 5a) is only formed by windows with direct links. On the right (Figure 5b), the stacked signal includes widows from both direct and indirect links to the window with high PageRank.

The template will be formed from the windows that directly match the main window, and also all the windows that matched each one of those initial matches. These are the windows with indirect links to the main window. The advantage of this process can be

seen in Figure 6, which shows a comparison between the first, second and third level stacks for several stations used in the analysis. The signal of the stacked waveform improves significantly once a second level of detections is added to the stack (windows with indirect links). The difference is best seen from Level 1 to Level 2 (Figure 6), where each signal has less noise around the main peak of the wave.

Here, each level of stack was created with a different number of windows. In the case of Level 1, the lowest number of windows was used because these stacks only include the windows with direct links to the highest ranked window. Each window from the direct links has its own set of links. These would be the first level of windows with indirect links to the main window. Level 2 has a larger number of stacks than Level 1 because it also includes these windows, the first level of windows with indirect links. Adding these windows to each stack improves the signal for each template (Figure 6).

We can add another level of links by stacking the windows with links to the first level of windows with indirect links. These stacks are shown in Figure 6 as Level 3. We can see that adding one more level of windows does not change the stack significantly and the number of signals used is very similar to Level 2. The reason for this is that the windows added at this step are mostly windows that were already present as links to other windows. For that reason they do not contribute independent information to the stack. We note here that in some cases the number of windows used for Level 3 are slightly lower than in Level 2. This is caused by the way we have defined the near repeat window elimination in the processing. As we scan the list of windows, we eliminate the near repeats by finding the window that was paired with the best CC value. We do this in ascending numerical order for a 3 second time span. If the window with the highest CC was not the first on that group then this pushes the selected window number forward. These might cause the next group of windows to be considered now as a near repeat if the window number of the first of the group is now less than 3 seconds than the previously selected one. Adding one more level of windows only finds more similar windows to the signals used in Level 2 and some of these might fill some small gaps in, causing a few more near repeats to be eliminated. Given the similarity of the stacks from Level 2 and Level 3 (Figure 6), this suggests that it is only necessary to create the stacks using one level of windows with indirect links to get an optimum result for creating a template waveform.

We validate the templates created with these stacks by comparing the stacked signals to known LFEs from the JMA catalog. We selected various events from the catalog that locate to the same general area of the LFEs reported for the April 2006 sequence (*Shelly et al*., 2007b). Figure 7 shows the waveform comparison between the LFEs from the JMA catalog and the templates created in this study for the stations available in the catalog. The stacks created for the five stations, shown in Figure 7, match the three different events from the catalog and preserve the moveout across the network.

A significant advantage to the PageRank approach is that it can also be used to distinguish between tremor and noise. We compared an hour of tremor data from the April 16, 2006 episode to an hour of noise data, where very little seismicity is observed, for the same stations in Shikoku. Figure 8 shows the differences between the normalized PageRank histograms for three different stations used in this analysis. These histograms are significantly different between tremor and noise because tremor

has many more windows with high PageRank. If the data has a lower *snr*, then this characteristic of PageRank should help to distinguish tremor from noise. Because the differences show up in single-components at different stations independently, it should also help to distinguish tremor from noise for sparse data.

Application to Continuous data.

Having found robust signals to use as LFE templates, we applied these to search through continuous data during the Shikoku event in April of 2006. We picked data where it is clear that the strong tremor episode is getting started, so we can observe small tremor bursts but also larger amplitude, more significant bursts (Figure 9) to test the ability of our detector of finding LFEs within lower *snr* tremor data as well. We cross-correlate the templates with 10 second windows and move the window every 2 samples through the data to find matches. We perform this analysis one station at a time using a low threshold (3σ) and later compare the results between stations to distinguish between true and false detections.

Figure 9 shows the second half of April 16, 2006 for Shikoku Hi-Net data. This data set shows small tremor bursts between 12,000 and 24,000 seconds and larger amplitude tremor bursts towards the end of the day, around 32,000 seconds. Here we used only stations were it was possible to compare our LFE template to LFE picks present in the JMA catalog (Figure 7). To associate detections found for each station (Figure 9a), we compare all stations within a 2-s window and declare a positive detection if three or more of the stations show a detection within those 2 seconds. Using this simple association algorithm we find very similar results to Shelly et al. (2007b), particularly for the stronger tremor burst towards the end of April 16, 2006. We also find a large number of detections that were missed previously, during smaller tremor bursts (Figure 9b) between 12,000 and 24,000 seconds and between 32,000 and 36,000 seconds.

We have also found that PageRank values differ between tremor and noise. By looking at histograms of PageRank values we can differentiate between tremor and noise because unlike noise, tremor data has large numbers of high PageRank values due to the significant links between windows within a short period of time. This could prove to be a useful tool for automated tremor detection and for detecting tremor where it has not been detected previously, or where it has been found but the signal is not as prominent as it is on Hi-Net data. Finally, although we have applied it to detect LFEs within tremor, the PageRank approach may be useful for other situations such as swarms or aftershock sequences, for which many similar waveforms may be present.

References.

Aguiar, A. C., T. I. Melbourne, and C. W. Scrivner (2009). Moment release rate of Cascadia tremor constrained by GPS, *Journal of Geophysical Research: Solid Earth* **114**, B7, B00A05.

Beroza, G. C., and S. Ide (2011). Slow Earthquakes and Nonvolcanic Tremor, *Annual Review of Earth and Planetary Sciences* **39**, 1, 271-296.

Brown, J. R., G. C. Beroza, and D. R. Shelly (2008). An autocorrelation method to detect low frequency earthquakes within tremor, *Geophysical Research Letters* **35**, 16, L16305.

Fisher, R. A. (1973), *Statistical methods for research workers*, 14th ed., rev. and enl. ed., Hafner Pub. Co., New York.

Gomberg, J. J. Rubinstein, Z. Peng, K. C. Creager, J. E Vidale, and P. Bodin (2008) Widespread triggering of nonvolcanic tremor in California, *Science,* **319**, 10.1126/science.1149164.

Ide, S. (2010). Striations, duration, migration and tidal response in deep tremor, *Nature* **466**, 7304, 356-359.

Ide, S., D. R. Shelly, and G. C. Beroza (2007a). Mechanism of deep low frequency earthquakes: Further evidence that deep non-volcanic tremor is generated by shear slip on the plate interface, *Geophysical Research Letters* **34**, 3, L03308.

Ide, S., G. C. Beroza, D. R. Shelly, and T. Uchide (2007b). A scaling law for slow earthquakes, *Nature* **447**, 7140, 76-79.

Katsumata, A., and N. Kamaya (2003). Low-frequency continuous tremor around the Moho discontinuity away from volcanoes in the southwest Japan, *Geophysical Research Letters* **30**, 1, 1020.

McCausland, W., S. Malone, and D. Johnson (2005). Temporal and spatial occurrence of deep non-volcanic tremor: From Washington to northern California, *Geophysical Research Letters* **32**, 24, L24311, doi:24310.21029/22005GL024349.

Moler, C. (2011), Experiments with MATLAB, edited, MathWorks, Inc., www.mathworks.com/moler/exm/chapters.html.

Obara, K. (2002). Nonvolcanic deep tremor associated with subduction in Southwest Japan, *Science* **296**, 5573, 1679-1681.

Page, L., S. Brin, R. Motwani, and T. Winograd (1999), The PageRank Citation Ranking: Bringing Order to the Web, *Technical Report Rep.*, Stanford InfoLab.

Rogers, G., and H. Dragert (2003). Episodic tremor and slip: The chatter of slow earthquakes, *Science* **300**, 1942-1944.

Shelly, D. R. (2009). Possible deep fault slip preceding the 2004 Parkfield earthquake, inferred from detailed observations of tectonic tremor, *Geophysical Research Letters* **36**, 17, L17318.

Shelly, D. R., G. C. Beroza, and S. Ide (2007a). Non-volcanic tremor and low-frequency earthquake swarms, *Nature* **446**, 7133, 305-307.

Shelly, D. R., G. C. Beroza, and S. Ide (2007b). Complex evolution of transient slip derived from precise tremor locations in western Shikoku, Japan, *Geochemistry Geophysics Geosystems* **8**, Q10014, doi:10010.11029/12007GC001640.

Shelly, D. R., G. C. Beroza, S. Ide, and S. Nakamula (2006). Low-frequency earthquakes in Shikoku, Japan, and their relationship to episodic tremor and slip, *Nature* **442**, 7099, 188-191.

Wech, A. G., and K. C. Creager (2008). Automated detection and location of Cascadia tremor, *Geophysical Research Letters* **35**, L20302, doi:20310.21029/22008GL035458.

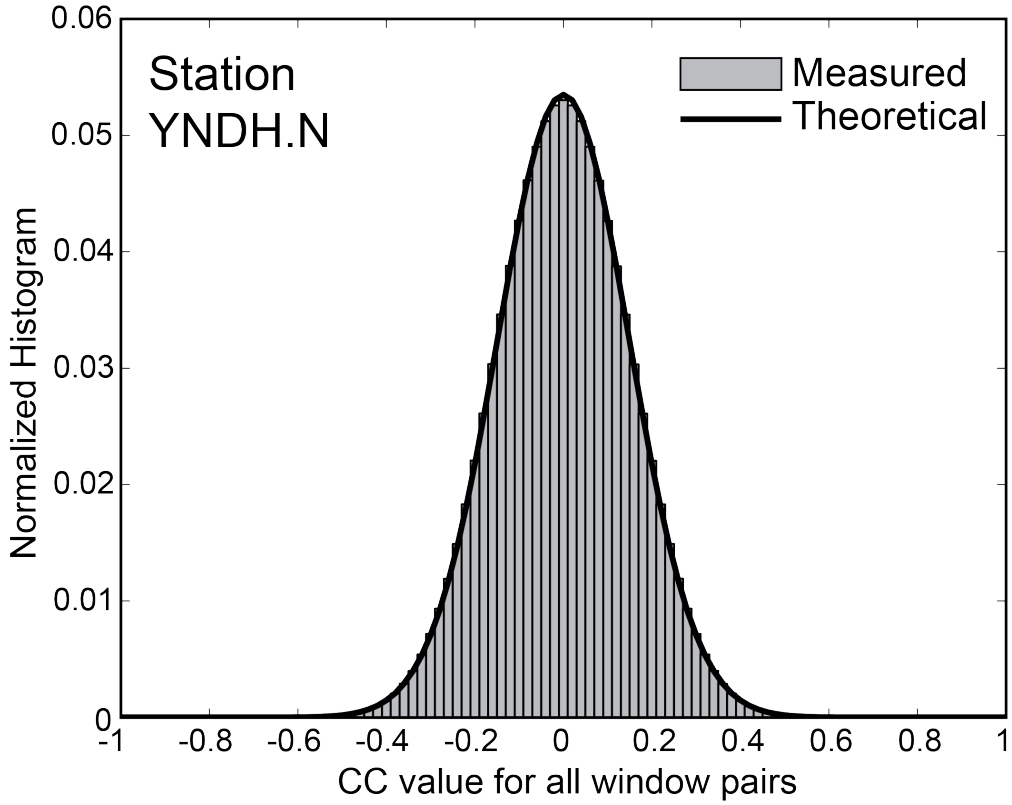

Figure 1. Distribution of CC values for all window pairs of 1 hour of tremor data on station YNDH.N. Black line represents the theoretical normal distribution calculated using σ = 1.253 x MAD. Here MAD is computed from all window pairs.

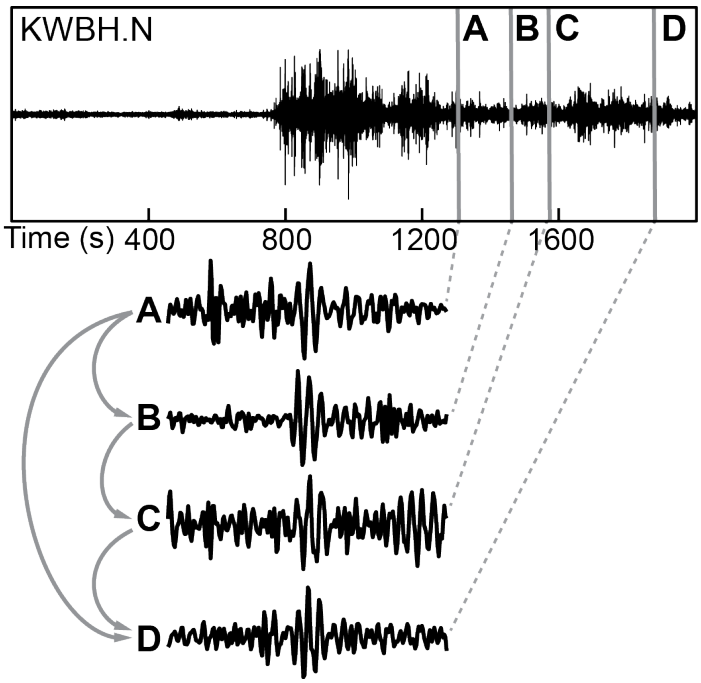

Figure 2. Window pair links: A, B, C, and D are windows that were found to be a match with another window given the threshold selected. This figure shows how all these pairs are linked together. Some windows are linked directly: A → B and B → C. Other windows are linked indirectly: A is linked indirectly to C (since B → C) and B is indirectly linked to D (since C → D).

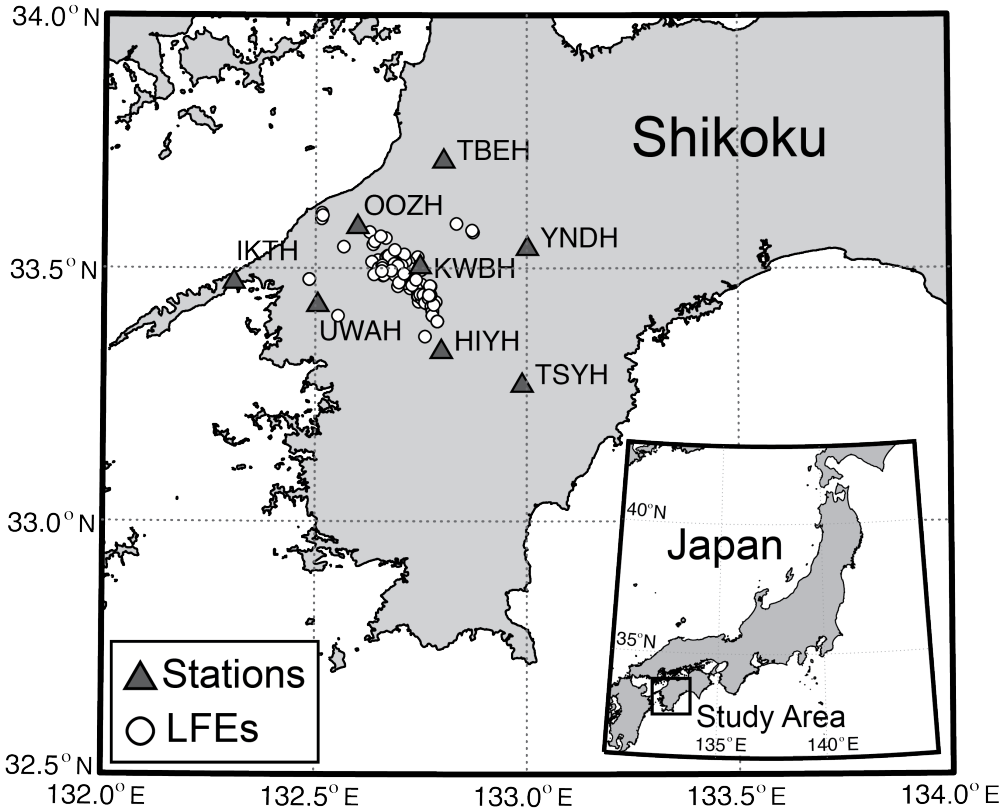

Figure 3. Map of Shikoku, Japan with the location of stations (dark gray triangles) used in our analysis. Circles represent LFE locations from Shelly et al. (2007b) during the tremor episode of April 16, 2006. Insert map shows the location of the study area in Japan.

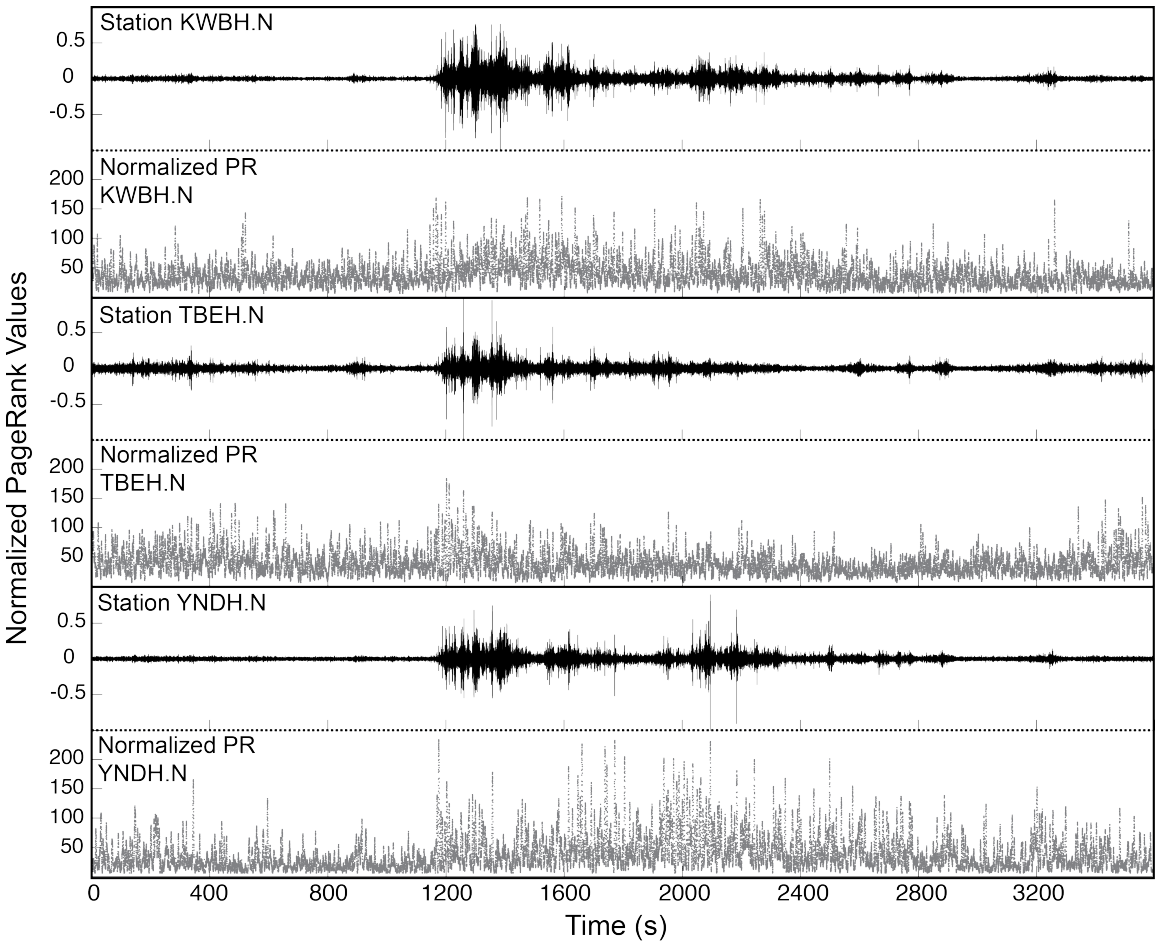

Figure 4. One hour of tremor data (black) during the tremor event in Shikoku, Japan, on April 16, 2006 for stations KWBH, TBEH, and YNDH associated to its PageRank values (gray). The PageRank values are normalized by the total number of windows analyzed.

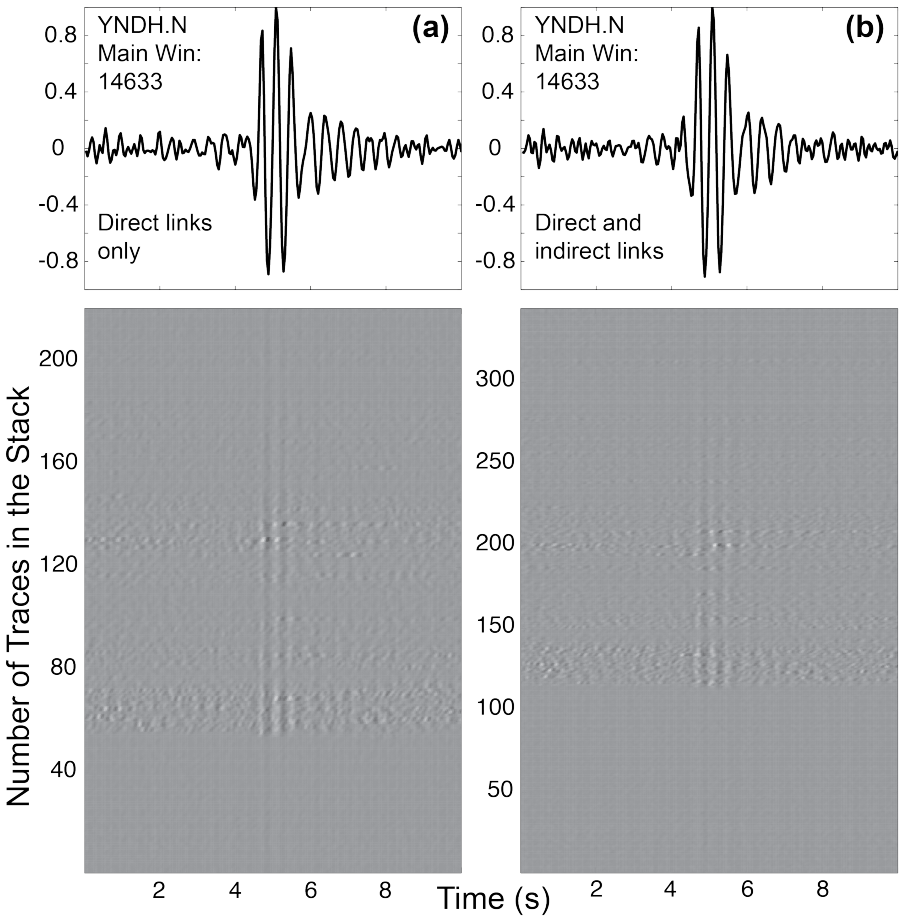

Figure 5. Stacked signal (top) formed by summing all the windows (a) using only windows with direct links (221 traces) and (b) using windows with direct and indirect links (345 traces) that were found as a match to the window with highest PageRank. Grayscale plot (bottom) of each one of the windows forming the stack.

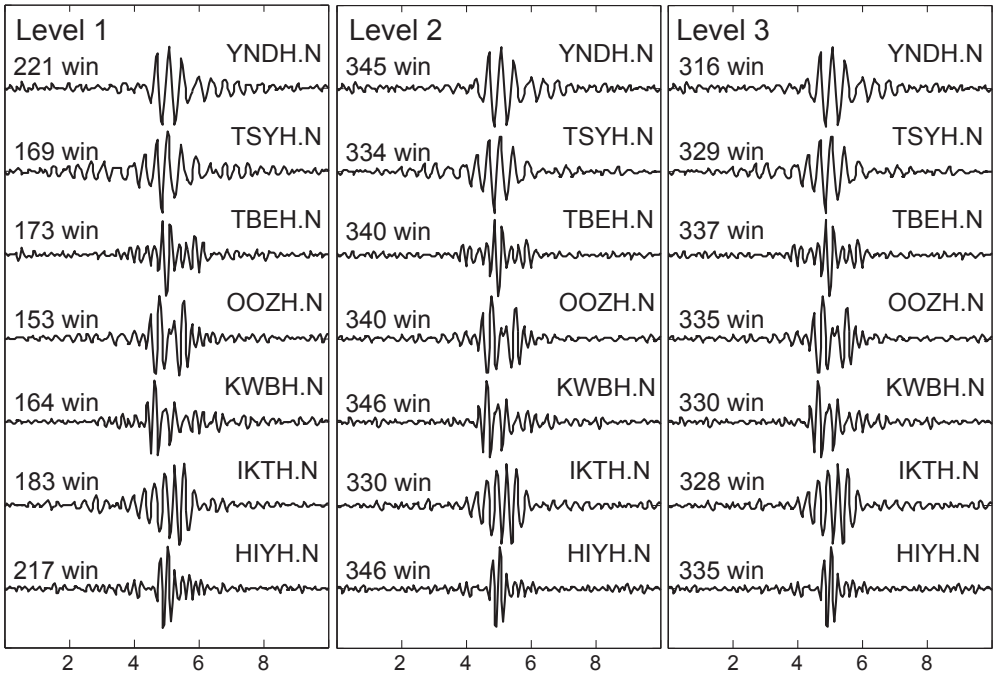

Figure 6. Plot of stacked signals formed by windows of three different levels of links. Level 1 shows stacks created using only the windows with direct links to the window with highest PageRank. Level 2 shows stacks created with windows with direct links and one level of indirect windows (windows matched to the matches of the main window). Level 3 plot shows the stacks created with direct windows and two levels of indirect windows. The number on the left of each plot shows the amount of traces that were used to create each stack.

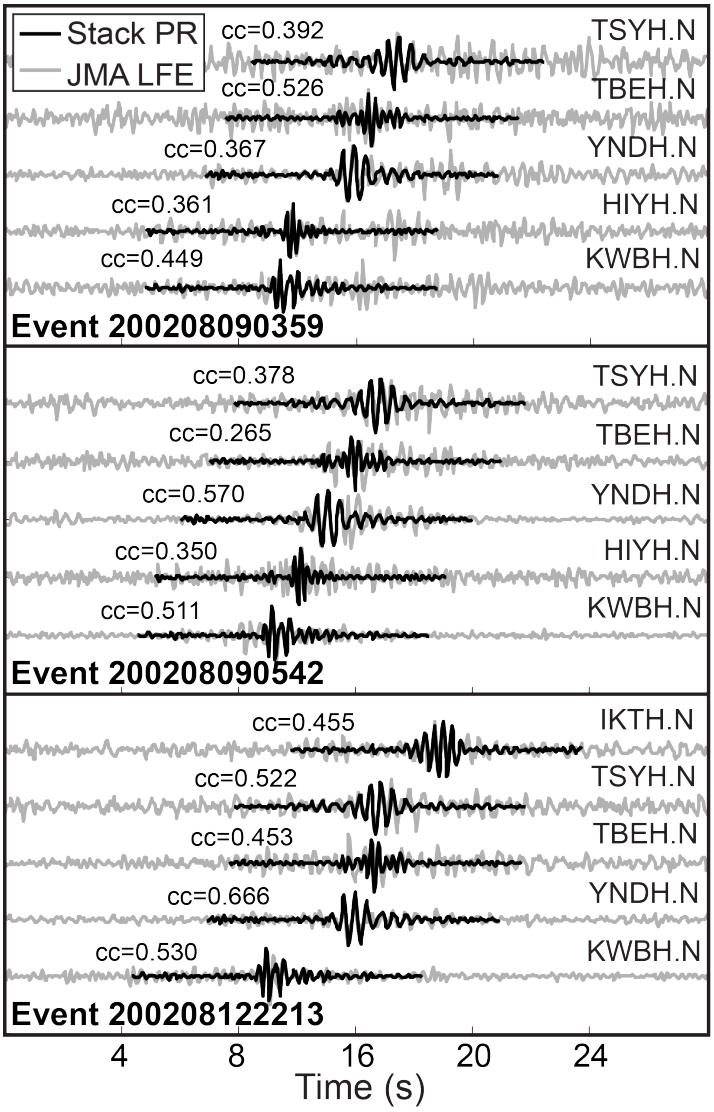

Figure 7. Three different LFEs from the JMA catalog (gray) compared with one template stack (black) created using the PageRank approach from an hour of April 16, 2006 of Shikoku tremor data.

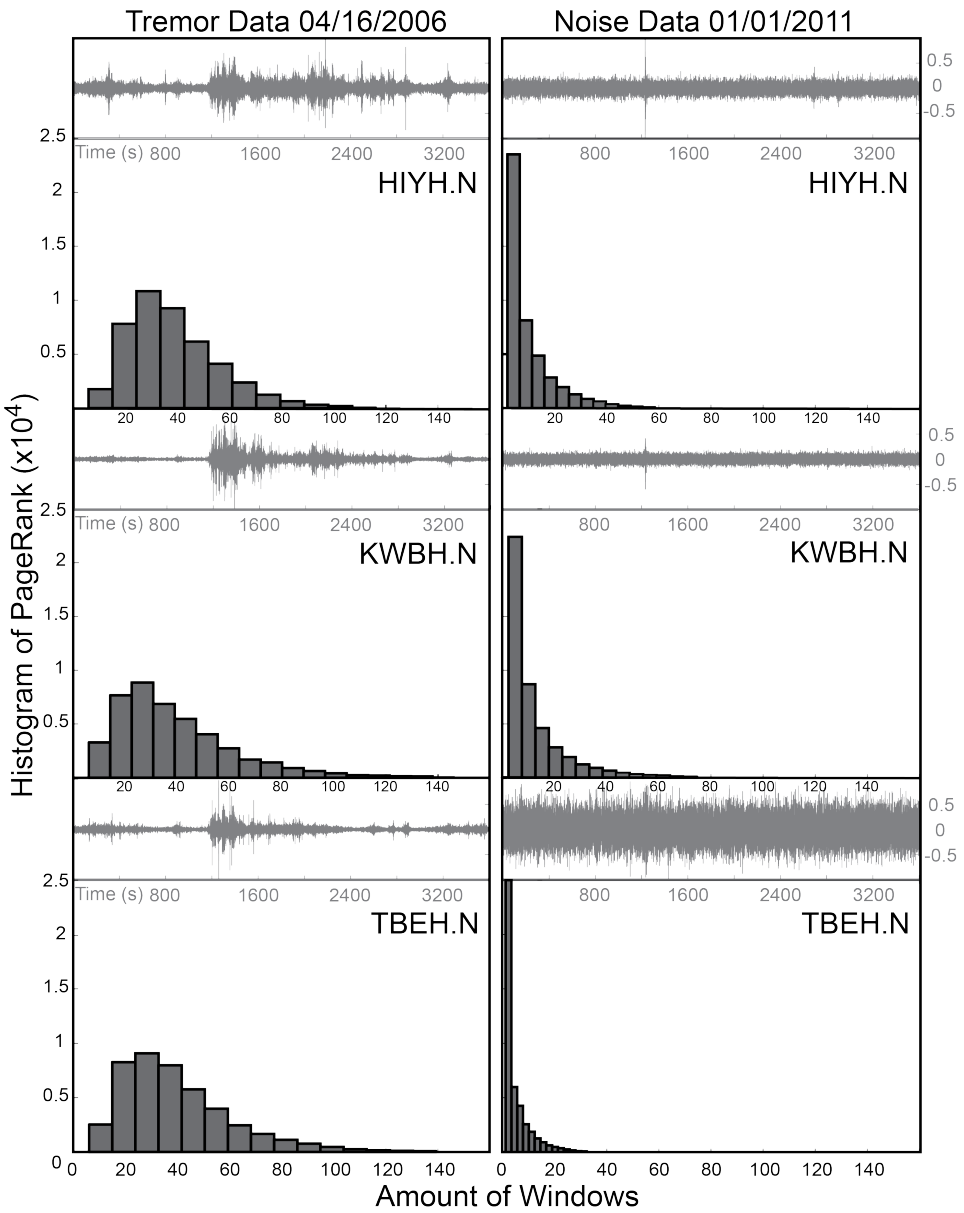

Figure 8. Histograms of normalized PageRank values for tremor data (left) and noise data (right) for three different stations used in the analysis. Noise data shows large numbers of low PageRank values whereas the tremor data have higher PageRank values.

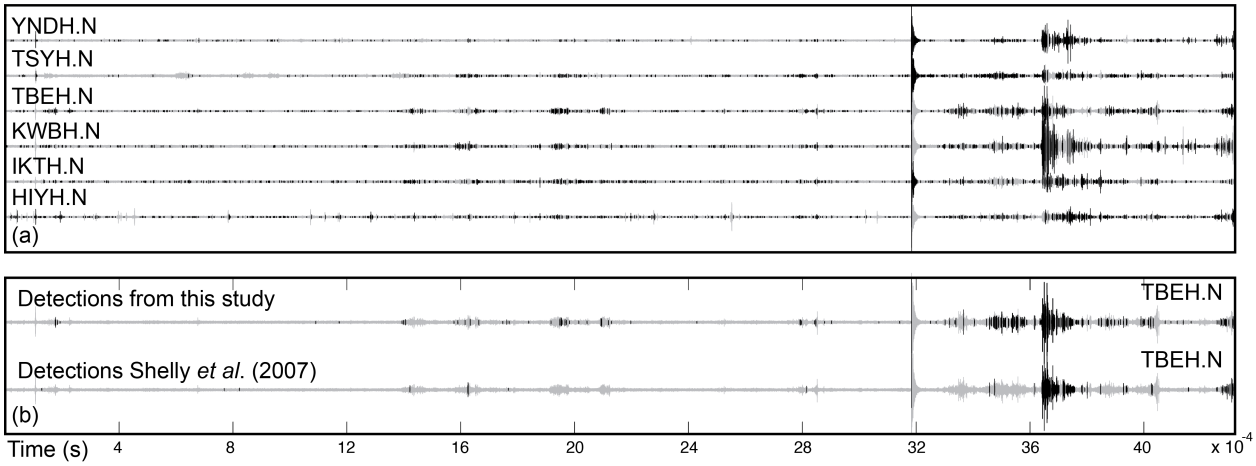

Figure 9. Half day of data during April 16, 2006 of several stations in Shikoku showing the beginning of the tremor episode. (a) Detections (black) found individually for each station using the stack created with the PageRank approach for 6 of the stations in the analysis. (b) Detections found on at least three of the stations within a 2 s window (top), compared to Shelly et al. (2007b) detections (bottom) using template matching.