Automatic detection of low-frequency earthquakes (LFEs) based on a beamformed network response

W. B Frank, N M Shapiró

Geophysical Journal International

Geophys. J. Int. (2014) **197,** 1215–1223
Advance Access publication 2014 March 11

doi: 10.1093/gji/ggu058

Automatic detection of low-frequency earthquakes (LFEs) based on a beamformed network response

W. B. Frank and N. M. Shapiro

Équipe de Sismologie, Institut de Physique du Globe de Paris, Paris Sorbonne Cité, CNRS, F-75238 Paris, France. E-mail: frank@ipgp.fr

Accepted 2014 February 12. Received 2014 February 11; in original form 2013 November 25

SUMMARY

Low-frequency earthquakes (LFEs), which frequently originate from multiplet-generating sources that are closely linked with tectonic tremor in subduction zones around the world, are difficult to observe and characterize due to their low signal-to-noise ratios. This obstacle can be sidestepped by detecting and then stacking all of the multiplets of a master LFE event, or template, using a matched-filter search; the difficulty however lies in finding an LFE event to use as a template. We implement here an automated beamforming algorithm to detect LFEs within the Mexican subduction zone that can then be used as templates in a matched-filter search. Seismograms recorded on a network of seismic stations are aligned to match the moveout of a potential source at depth and their energies are then summed; any spikes in the summed energy indicate an event originating from that potential source. We apply this method to a 1-d test case and we are able to detect 381 unique, potential LFE templates. We then compare our method to a previously introduced LFE detection scheme based on multiplet correlations for three test cases and find that the two methods are complementary.

Key words: Time-series analysis; Seismicity and tectonics; Subduction zone processes.

1 INTRODUCTION

Tectonic tremors (TTs), also known as non-volcanic tremor, are long-duration, emergent seismic signals that have been observed in many subduction zones around the world, including southwestern Japan, Cascadia, Costa Rica and Mexico, as well as the San Andreas strike-slip fault (e.g. Obara 2002; Rogers & Dragert 2003; Brown *et al.* 2005; Nadeau & Dolenc 2005; Payero *et al.* 2008). Slow-slip events (SSEs), transient slip too slow to be observed seismically, have been observed concurrently in time and space with TTs, suggesting a link between the two phenomena (e.g. Rogers & Dragert 2003). Recent studies of the Mexican subduction zone have observed changes in seismic velocity near the subduction interface during periods of intense TT activity during SSEs, suggesting that TT activity can be used to track transient strain at depth (Rivet *et al.* 2011, 2014). These observations highlight the need to better understand TTs and their effect on the subduction cycle; however, due to their emergent nature, TTs are difficult to locate and characterize. A potential TT proxy, small, impulsive events called low-frequency earthquakes (LFEs) have been observed within tremor signal and have been used to better locate and characterize TT (e.g. Shelly *et al.* 2006; Bostock *et al.* 2012; Frank *et al.* 2013).

Low LFE signal-to-noise ratios (SNRs) make it very difficult to analyse a single LFE event. Swarms of LFEs have been observed within TT, with several sources producing many LFEs over time that we will call here multiplets (Shelly *et al.* 2007). Exploiting this multiplet source behaviour, an LFE master event, or LFE template, consisting of the LFE waveforms recorded on a seismic network along with the moveout with which it is observed on each station can be used in a matched-filter search to find other events originating from the same source as the LFE template (e.g. Shelly *et al.* 2006).

However, without an established catalogue of events, finding the LFE templates to use in the matched-filter search can be difficult. It is possible to identify them visually, but this restricts the detection of LFE templates to the 'loudest' events visible above the tremor signal as well as not being a time-efficient detection method. An LFE template detection method based on multiplet correlations was proposed by Brown *et al.* (2008) that relies on the repetitiveness of an LFE waveform on a given seismic station. We propose here an alternative LFE template detection method that relies on the response of the entire seismic network to a theoretical source.

We will first describe the proposed method in the following section, then apply it to the detection of LFEs in the Mexican subduction zone, and finally compare our method to the detection of LFEs based on multiplet correlations (inspired by Brown *et al.* 2008).

2 AUTOMATIC BEAMFORMING DETECTION METHOD

Our method can be broken down into three basic parts: (1) the preprocessing of the data, (2) the calculation of the traveltime tables for a 3-D grid of theoretical LFE sources and (3) the alignment and summation of the energies of the seismograms. The resulting

summed energy trace, or network response, can then be analysed to pick out the potential events originating from one of the theoretical sources of the 3-D grid. In this section, we will first detail each of the three principal steps, we will then describe how potential LFEs are detected within the network response, and to finish, we will briefly discuss the limitations inherent to this method.

2.1 Data pre-processing

We first pre-process daily seismograms from 10 Meso-America Subduction Experiment (MASE) broad-band stations (the data set used in this study will be further described in the following section), chosen for their high SNRs, their proximity to the LFE source region and the continuity of their data sets, by removing the daily mean and linear trend of each seismogram. The highest SNR frequency bandwidth of TT and LFEs in the Mexican subduction zone has been observed to be between 1 and 2 Hz (e.g. Payero *et al.* 2008; Frank *et al.* 2013); the seismograms are consequently bandpass filtered in this frequency range.

The strongest LFE ground motions have been observed on the horizontal components in many subduction zones (e.g. Shelly *et al.* 2006); we therefore choose only to use the NS component of our broad-band seismic stations.

2.2 3-D grid of theoretical LFE sources

The region where one would expect LFEs to originate from can be constrained horizontally and in depth by the previous locations of TT and LFEs (Payero *et al.* 2008; Kostoglodov *et al.* 2010; Husker *et al.* 2012; Frank *et al.* 2013). We divide this source region into a 3-D grid of potential point sources with dimensions of 100 by 100 by 75 sources in latitude, longitude and depth, respectively, the outer limits of which are shown in Fig. 1; the longitude grid interval is 0.025° or ∼3 km; the latitude grid interval is 0.02° or ∼2 km and the depth grid interval is 1 km. Supposing a regional velocity model (Iglesias *et al.* 2010), we can then calculate the traveltimes from each of the theoretical sources to the seismic network with a ray tracing technique. The *P* waves of LFEs are rarely observable in Mexico without stacking many similar events together due to their low amplitudes (Frank *et al.* 2013); going forward, we only calculate and consider the *S*-wave traveltimes as the aligned *S* waves will stack constructively much quicker than aligned *P* waves (see following section). We do not use absolute arrival times and instead use the relative moveout as it is necessary for the matched-filter search that is used to form a family of similar LFEs. We define the relative moveout at a given station i, δ_i, and originating from a source located at geographical coordinates (ϕ, λ, z), where ϕ is latitude, λ is longitude and z is depth, as the time delay between the arrival time of an event on a given station and the earliest arrival time for the same event observed on the entire network:

$$\delta_i(\phi,\lambda,z) = \alpha_i(\phi,\lambda,z) - \min_i\left[\alpha_i(\phi,\lambda,z)\right], \quad (1)$$

where α_i are the theoretical arrival times of a source at geographical coordinates (ϕ, λ, z), observed at station i. For simplicity's sake, we will henceforth use the general term moveout to refer to the relative moveout.

Due to the nearly linear geometry of the MASE seismic network, there are many sources within the 3-D grid that produce extremely similar moveouts. Given that the LFE's location will be determined by other means, the group of theoretical sources that produce similar moveouts can be reduced to a single theoretical source with a representative moveout. The computational time gained by removing the redundant theoretical sources depends on what quantity the user uses to determine if two moveouts are sufficiently similar to be considered redundant. In this study, we considered any theoretical source redundant if the cumulated time difference across the entire network between its moveout and any other moveout from the 3-D grid of theoretical sources is less than 0.5 s; we represent the total time difference as the following:

$$\Delta = \sum_i \left[\delta_i(\phi_k,\lambda_k,z_k) - \delta_i(\phi_l,\lambda_l,z_l)\right], \quad (2)$$

where Δ is the total time difference and k and l represent two different theoretical sources. A total time difference of 0.5 s spread across 10 stations would imply an average arrival time difference of 0.05 s on each of the stations, which we consider to be negligible given this study's frequency band. Using this method, we reduced the number of redundant sources by ∼90 per cent, from 750 000 to 70 476. In Supporting Information Fig. S1, we show how this optimization impacts the location precision of the proposed beamforming method.

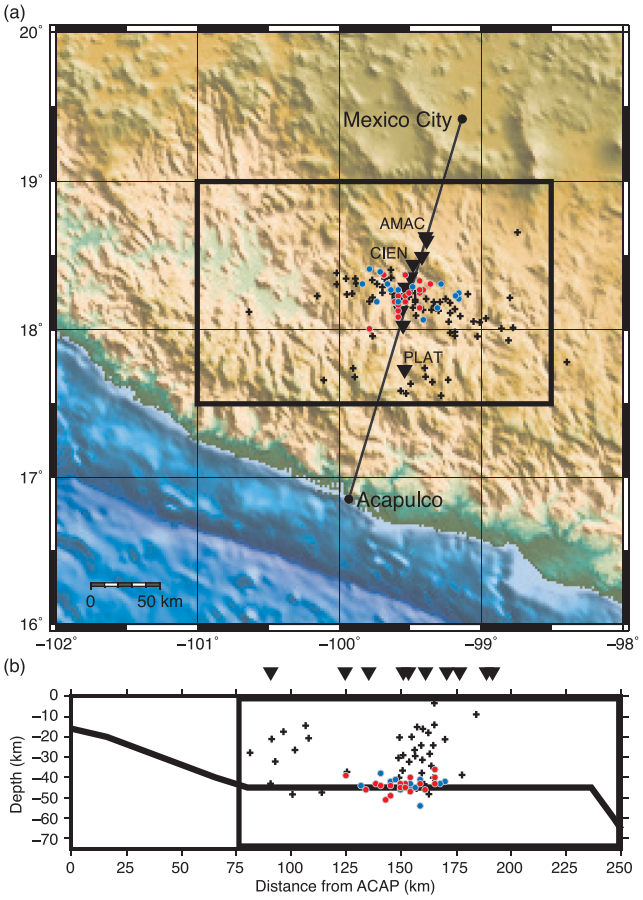

Figure 1. Horizontal and vertical slices of the Guerrero, Mexico study area. Locations of TT determined by Payero *et al.* (2008) are shown as small, black crosses. The 15 LFE families located by Frank *et al.* (2013) are shown as blue points. The 14 LFE families that were detected with our proposed method and characterized in this study are shown as red points. The 10 MASE stations that were used in this study are shown by the inverted triangles in (a) and (b). The outer limits of the 3-D grid of theoretical sources for this study is shown by the black box in (a) and (b); the longitude grid interval is 0.025°; the latitude grid interval is 0.02°; the depth grid interval is 1 km. (b) The thick black line represents the subducting Cocos plate geometry (Kim *et al.* 2010).

2.3 Alignment and network response calculation

The automatic detection method is based on two simple hypotheses: (1) if the hypocentre of a given LFE lies close to one of the theoretical sources on our 3-D grid, S_0 located at (ϕ, λ, z), it should be observed on the seismic network with a moveout very close to the one associated with S_0; (2) if the average interstation distance is relatively small compared to the distance travelled by an arriving wave from S_0, the travel paths of the arriving waves should not differ greatly, ensuring that each of the stations on the network record similar energy envelopes.

The alignment of the seismograms, $s_i(t)$ where t is time and i is the station, given the theoretical source S_0, can be performed by aligning the seismograms with S_0's associated moveout, $\delta_i(\phi, \lambda, z)$:

$$\tau = t - \delta_i(\phi, \lambda, z), \tag{3}$$

$$s_i(\tau, \phi, \lambda, z) = s_i(t), \tag{4}$$

where $s_i(\tau, \phi, \lambda, z)$ is the aligned seismogram at station i. Fig. 2 shows the unaligned and aligned seismograms of an LFE.

Once aligned, the seismograms are then normalized by station; we normalize to ensure that network response reflects the coherence of a signal originating from S_0 and is not dominated by the loudest seismogram. We then stack the instantaneous energy, defined as the square of the velocity, instead of the velocity because the geometry of the seismic network with respect to the 3-D grid of theoretical sources is such that there are potential focal mechanisms that could produce waveforms on two different stations that are of opposite polarity. When aligned, these waveforms would stack destructively and therefore not be detectable. Using the instantaneous energy not only overcomes this obstacle, it also increases the amplitude range of the network response, which we define here as the difference in amplitude between a non-detection and a detection, further facilitating the detection of LFEs within the network response.

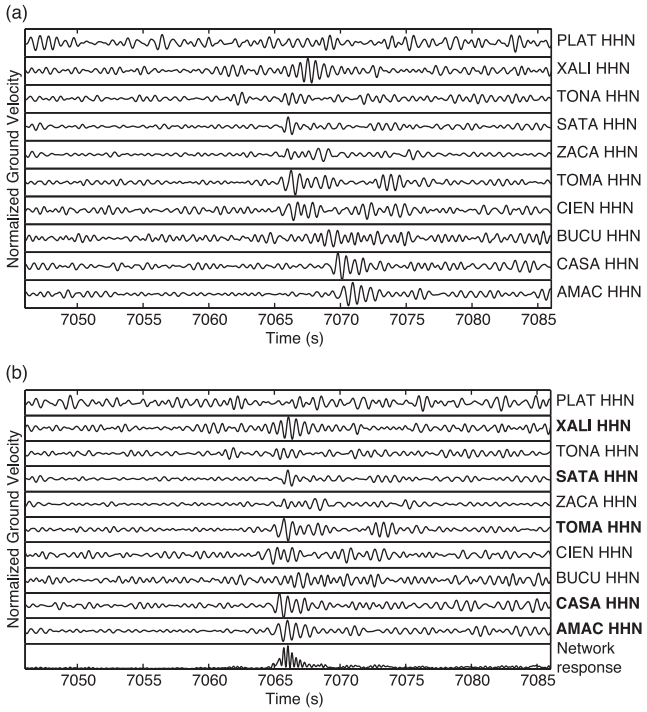

Figure 2. Unaligned and aligned LFE waveforms with the resulting network response from a theoretical source located at 18.19°N 99.41°W at 41 km depth. All seismograms have been filtered between 1 and 2 Hz and normalized. (b) The bolded station names indicate the stations that observed the highest observed amplitudes during the LFE arrivals; we then use these stations in the matched-filter search using this LFE as a template.

If the two hypotheses described at the beginning of the section are satisfied, the instantaneous energy originating from S_0 will stack constructively. The signal at each time t of the resulting stacked energy envelope, called the network response, will reflect the degree to which an event originating from S_0 is observed in-phase across the network. This calculation of the network response to the source S_0 at (ϕ, λ, z), defined as $NR(\tau, \phi, \lambda, z)$, is mathematically represented as the following:

$$NR(\tau, \phi, \lambda, z) = \sum_i \frac{s_i(\tau, \phi, \lambda, z)^2}{\mathrm{rms}\left[s_i(\tau, \phi, \lambda, z]^2\right)}. \tag{5}$$

Consequently, the largest value of the network response will occur during the time window containing the LFE waveform. This is demonstrated in Fig. 2(b) where the maximum in the network response is observed during the aligned LFE waveforms.

This however only reflects the network response to one theoretical source. To obtain the network response to all of the unique moveouts produced by the 3-D grid of theoretical sources, we repeat this alignment and summation for each of the unique moveouts, recording only the largest network responses at each time t and the associated moveout that produced that network response. We represent this composite network response, NR', as the following:

$$NR'(\tau) = \max_{\phi,\lambda,z}\left[NR(\tau, \phi, \lambda, z)\right]. \tag{6}$$

The composite network response provides the largest network response of the entire 3-D grid of theoretical sources at each time t; from here on out, we will simply write network response when referring to the composite network response. An example of a network response calculated over several hours is shown in Fig. 3. Each of the spikes in the network response indicates a spike of energy originating from one of the theoretical sources defined in the 3-D grid.

2.4 Finding LFE templates in the network response

Let us suppose that during a given day several LFEs originate from the defined LFE source region and are recorded on the MASE network. The network response for that day would have several peaks associated with each of the LFEs. It is then simple enough to set up a simple algorithm to search for the peaks within the network response; the moveout associated with each peak is then easily determined. With the waveforms and the moveout associated with the network response peak, we now have the two essential ingredients for an LFE template and we can use a matched-filter search on the entire data set to find the family of similar events.

Although the network response analysis is simple in theory, there are a few practical considerations that need to be taken into account. The vast majority of LFEs in Mexico have been observed within TT that ostensibly originates from a source nearby to that of the LFEs (Frank *et al.* 2013). This means that the long-duration TT signal, along with the LFE, will stack constructively during the network response calculation. This is however, not a serious issue due to the emergent nature of TT and the impulsive nature of LFEs; although the network response of TT will have increased amplitudes with respect to pure ambient noise, the impulsive nature of LFEs

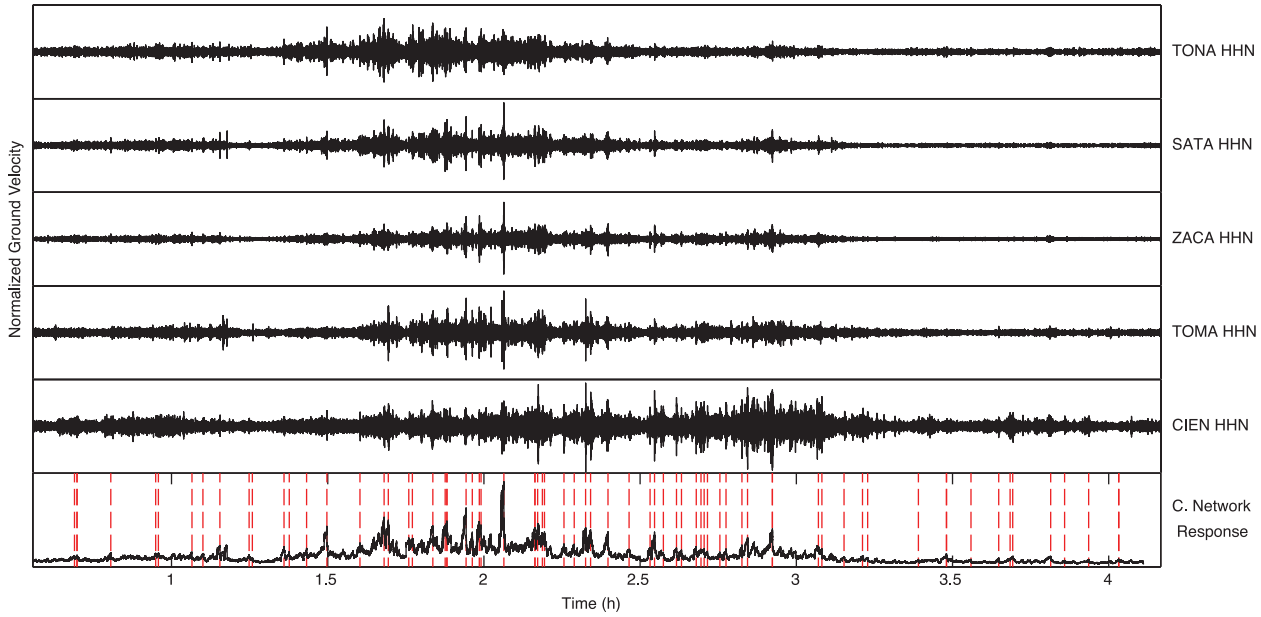

Figure 3. Composite network response calculated over several hours on 2005 March 20 with detected LFEs. All seismograms have been filtered between 1 and 2 Hz and normalized. The red, vertical dashed lines correspond to peaks of instantaneous energy originating from one of the theoretical sources defined in the 3-D grid that represent detected LFEs that can be used as templates. Using a coherence threshold of 0.2, there are 72 detected potential LFE templates.

will always ensure that their seismograms will stack constructively faster than TT.

The possibility of false detections must also be taken into account given the near-monochromatic nature of the pre-processed seismograms in the 1–2 Hz frequency band. The network response reflects the stacked instantaneous energy across the network at time t; it does not reflect the network coherence of a several second–long LFE waveform. In the case of random instantaneously coherent noise, the instantaneous coherent energy might be large enough to trigger a detection but its network coherence over several seconds would be much lower than that of an LFE. Taking this into account, false detections can be removed by calculating the network coherence of the aligned seismograms over a several second–long time window centred at the time of a peak network response. In this study, we defined the network coherence to be the average absolute value of the correlation coefficient between time windows of 4 s, centred on the peak network response time, at every possible station combination. Given N stations, there are $\frac{N(N-1)}{2}$ unique station combinations; in our case with 10 stations, the coherence is defined as the average absolute value of the correlations of 45 unique combinations between the network stations. Mathematically, we define the network coherence, C, as the following:

$$C = \frac{\sum_{n_{ij}=1}^{\frac{N(N-1)}{2}} |\langle s_i(t,\phi,\lambda,z), s_j(t,\phi,\lambda,z)\rangle_{n_{ij}}|}{N}, \tag{7}$$

where n_{ij} is the nth unique combination of station i and station j and $\langle s_i^a(t,\phi,\lambda,z), s_j^a(t,\phi,\lambda,z)\rangle_{n_{ij}}$ is the correlation coefficient between the pair of aligned seismograms at stations n_{ij}. The absolute value of the correlation coefficient is taken because an LFE whose focal mechanism's radiation pattern that produces positive and negative polarity waveforms on the network still reflects a coherent LFE detection. It is then simple enough to set a network coherence threshold and to consider all peak network responses that have a network coherence under the threshold as false detections. We use a network coherence threshold of 0.2 for this study.

Finally, if any false detections of LFEs do manage to make it through the network response analysis and are used as LFE templates, the resulting matched-filter search will only detect a very small number of similar events (<150 over the 2.5-yr–long MASE data set) that will be Poissonianly distributed in time, an extremely distinct signature of random noise templates.

2.5 Limitations of the beamforming method

This method has several limitations that must be taken into account. The first is that the investigated source region must be characterized with a velocity model that will allow the prediction of traveltimes. In addition, only events originating from the defined 3-D grid of theoretical sources can be detected. This is fairly evident but can be consequential when the source region is not well constrained and the 3-D grid must be enlarged, potentially increasing the computational cost of the method. A well-constrained source region on the other hand can be beneficial by intrinsically not detecting any events originating from outside the source region, avoiding the need to remove local earthquake events that could cause false positives in other detection schemes. Another limitation is that the medium underlying the seismic network cannot be too heterogenous; otherwise, regardless of the interstation distance, the recorded waveforms at each of the stations will not be similar enough to constructively stack when summed.

We have previously discussed how the 3-D grid of theoretical sources intrinsically filters out any unwanted events (such as local earthquakes) outside of the 3-D grid; there is the chance however, that non-LFE events that originate from within the 3-D grid are stacked constructively. Similar to false detections of random instantaneously coherent noise, these can be avoided as results of the matched-filter search using these local non-LFE events would be distinctly different from the results of a matched-filter search using an LFE. In our study, our source region does not include the seismogenic portion of the subduction interface, so the chance of repeater earthquakes (which could possibly resemble the multiplet behaviour

of LFEs) is reduced. The other principal undesirable non-LFE event that could be detected with this method, local crustal earthquakes, are not very likely to be reproduced hundreds to thousands of times at one source over a period of several years. The possibility of local, non-LFE events being detected with this method varies, however, with the area of study, and therefore the 3-D grid of theoretical sources must be adapted to the region of study. A final limitation is that the proposed beamforming method is sensitive to the instantaneous energy of seismograms: signals buried within incoherent noise will not sum constructively and will therefore not be detectable.

3 APPLICATION OF THE BEAMFORMING LFE DETECTION METHOD IN THE MEXICAN SUBDUCTION ZONE

We apply this method to the MASE data set obtained from a dense (∼5-km spacing) linear network of broad-band seismic stations installed between Acapulco and Tempoal, running through Mexico City, and operated between 2005 and mid-2007 (see Fig. 1). The MASE network produced a high-quality data set intended to characterize the subduction geometry of the Guerrero seismic gap, but has also been used to characterize TTs and LFEs and their relationship to SSEs (Payero *et al.* 2008; Kostoglodov *et al.* 2010; Rivet *et al.* 2011; Husker *et al.* 2012; Frank *et al.* 2013). The 10 stations used in this study (from south to north: PLAT, XALI, TONA, SATA, ZACA, TOMA, CIEN, BUCU, CASA and AMAC), chosen according to the criteria discussed in the section on data pre-processing, are shown as well in Fig. 1.

We will focus here on a single day: 2005 March 20. There are three observations of TT on 2005 March 20: the first is early in the morning (and is shown in Fig. 3), the second is during the afternoon and the last occurs in the early evening. Each burst was identified visually and automatically and lasts for about an hour and a half (Payero *et al.* 2008; Husker *et al.* 2010). We apply our detection method using the 3-D grid of theoretical sources described in the previous section and shown in Fig. 1. This resulted in 391 detections through the entire day. Due to the multiplet behaviour of LFEs, several of these detections could be a second or a third detection of a previously detected event. If we only consider the events whose moveouts are unique, there are then 381 detections for 2005 March 20. Due to the temporally correlated nature between LFEs and TT, one would expect more LFE detections during TT. Evidence of this relationship is demonstrated by the cumulative number of LFE detections during the entire day, as shown in Fig. 4(a): the rate of LFE detection increases during previously identified TTs (Payero *et al.* 2008). The LFE detection rate can then be used to modify the time window associated with previously identified TT (such as moving the start of the second TT in Fig. 4 several hours earlier) or even to detect previously unobserved TTs. The time window roughly between 5 and 7 a.m. on Fig. 4(a) contains a higher rate of LFE detections, indicating an increased probability of TT; taking a closer look at the data, tremor is easily visually identifiable during that time window, as shown in Fig. 4(b). We use 14 of the 381 detected events as LFE templates to perform the entire LFE detection workflow described in Frank *et al.* (2013) and find 75 465 robust detections. We then characterize the resulting LFE families and plot their locations on Fig. 1 alongside the LFE families from Frank *et al.* (2013). The combination of our proposed method along with the method of Frank *et al.* (2013) is nearly fully automated;

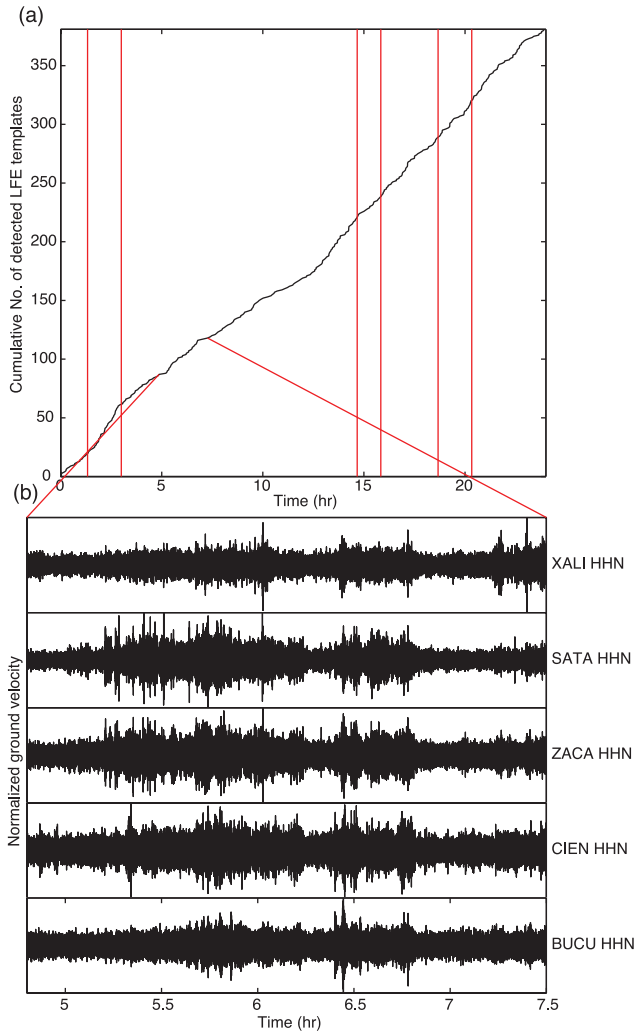

Figure 4. Cumulative distribution of detected LFEs on 2005 March 20 and seismograms of previously undetected TT. (a) Vertical red lines indicate start and stop times of detected TTs (Payero *et al.* 2008). (b) Seismograms recorded at five MASE stations between 4:50 and 7:30 a.m. on 2005 March 20. An increased rate of LFE detection using our proposed method hints at an undetected TT, clearly visible on the seismograms.

the user needs only to pick the arrivals on the stacked waveforms of the LFE family (Frank *et al.* 2013); efforts only need to be made on developing an automated picking algorithm to fully automate the entire workflow from raw seismogram to characterized LFE family. Given adequate computational resources, this workflow permits one to gather a consequent catalogue of LFE families that is necessary to investigate the activity and evolution of LFEs and their relationship to TT and SSEs.

4 COMPARISON WITH MULTIPLET CORRELATION LFE DETECTION

We will now compare the performance of our proposed LFE detection method to an established LFE detection algorithm.

4.1 Multiplet correlation detection

Brown *et al.* proposed, in 2008, a method to detect LFEs based on the repetition of waveforms. We modified the Brown *et al.* method

to make it completely automated and therefore directly comparable to our method; we call this modified version of the Brown *et al.* method the multiplet correlation method.

We first pre-process the seismograms in the same manner as described in Section 2.1. We then divide the period of time to be analysed into time windows; we define a given time window to start at time t and to finish at time $t + T$, where T is the duration of the time window. Keeping T constant for all time windows, we then refer to the discrete seismic signal recorded at station i during a time window that starts at time t_k as $s_i(t_k)$; we note for emphasis that t_k does not vary station to station. Unlike our proposed method, we use all station components for the multiplet correlation method; for simplicity's sake however, we use the index i for all different individual traces (different components at different stations) in the following description. We lag the start of each of the time windows by 0.5 s (the same amount as Brown *et al.* 2008).

When two different time windows, $s_i(t_k)$ and $s_i(t_l)$, are correlated station by station, there are I resulting correlation coefficients, where I is equal to the total number of stations; the resulting I correlation coefficients are represented as

$$CC_i^{kl} = \langle s_i(t_k), s_i(t_l) \rangle \,, \tag{8}$$

where CC_i^{kl} represents the correlation coefficient on station i. After defining K time windows, we then define the matrix a_{kl} with dimensions K by K that represents the summed correlation coefficients between all unique combinations of time windows $s_i(t_k)$ and $s_i(t_l)$. We define this as

$$a_{kl} = \sum_i CC_i^{kl}. \tag{9}$$

Each column of a_{kl} represents a time window correlated with all of the other time windows. To reduce a_{kl} to a simple detector trace covering the entire time period, we choose the largest correlation coefficient for each column; the resulting vector is considered the detector trace and is represented as

$$D(t_k) = \max_l [a_{kl}] \,. \tag{10}$$

The correlation coefficient at $D(t_k)$ represents the strongest correlation between $s_i(t_k)$ and a given $s_i(t_l)$. We establish a threshold of five times the median absolute deviation (MAD) of the detector trace; a summed correlation coefficient $D(t_k)$ greater than the threshold indicates a pair of time windows that have similar waveforms, one of which being $s_i(t_k)$; Brown *et al.* (2008) called the pair of similar time windows candidate events. We sum the pair of waveforms together for all stations and components and use the resulting waveforms as an LFE template in a matched-filter search, similar to LFE templates detected with our proposed method.

4.2 Method comparison

An important parameter in the repetitive waveform method is the size of the time windows that are then correlated. To evaluate the best time window size for our data set, we tested several different window sizes on a 15-min segment on 2005 March 20. Figs 5(a) and (b) show two different window sizes, 6 and 15 s, respectively, and highlight a problem with small window sizes: there is a risk of missing part of the waveform on some stations that have large moveouts. This is why we chose to use 15-s time windows: based on the moveouts of previously characterized Mexican LFE families, it is necessary to allocate enough time in the time windows to fully observe the waveforms on all stations. We then compared the resulting normalized detector traces during this 15-min time segment

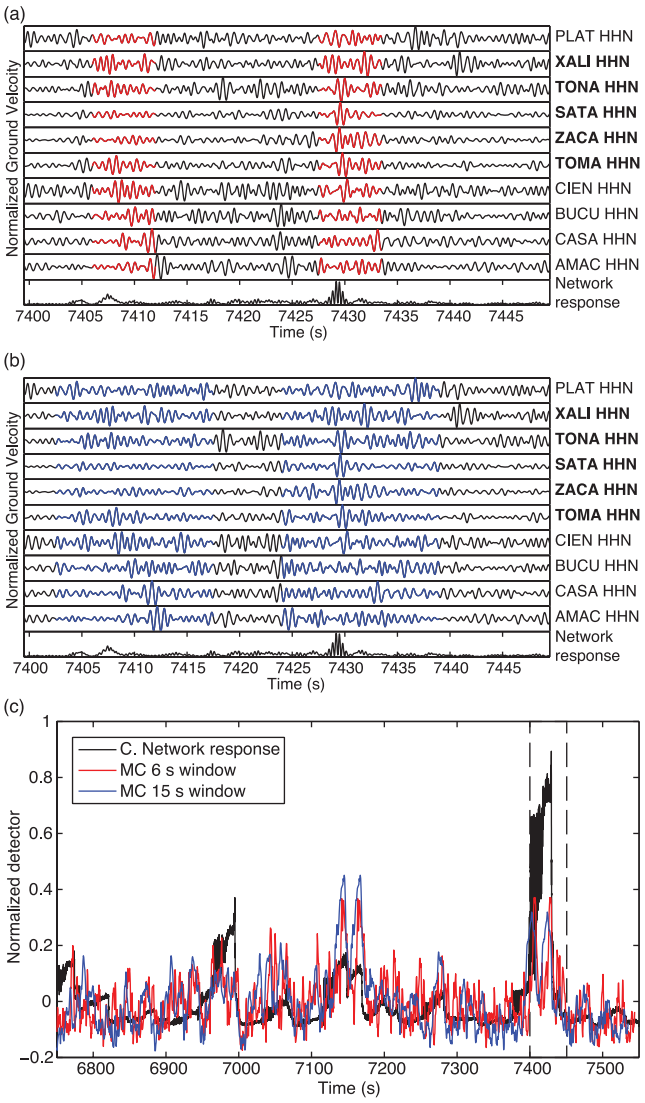

Figure 5. Different time window sizes for the multiplet correlation method and comparison to our proposed method during a 15 min-long segment on 2005 March 20. Red or blue signal in (a) and (b) indicates two time windows that correlate to a value above a threshold of five times the MAD and can be used as an LFE template. The resulting network response in (a) and (b) is calculated with the moveout from a theoretical source located at 18.27°N 99.56°W at 39 km depth. (a) Two time windows of 6 s, shown in red, are not able to capture the entire moveout of the LFE over the 10 MASE stations. (b) The entire LFE moveout can be observed within the two 15 s time windows in blue. (c) The normalized composite network response in black is compared to the resulting normalized multiplet correlation detector signal for 6 and 15 s time windows, in red and blue, respectively. The vertical dashed lines indicate the time period plotted in (a) and (b). The large amplitude range of the composite network response facilitates the detection of LFE templates. In addition, some events are more easily detected using our method, while others are more easily detected using the multiplet correlation method, suggesting that both methods are necessary to detect a maximum of LFE templates.

for the two methods: the series of correlation coefficients for each time window for the multiplet correlation method and the composite network response for our proposed method, as shown in Fig. 5(c). The first thing to note is that the amplitude range of our proposed method is larger than that of the repetitive waveform method, facilitating detection. In addition, while both methods detect LFEs

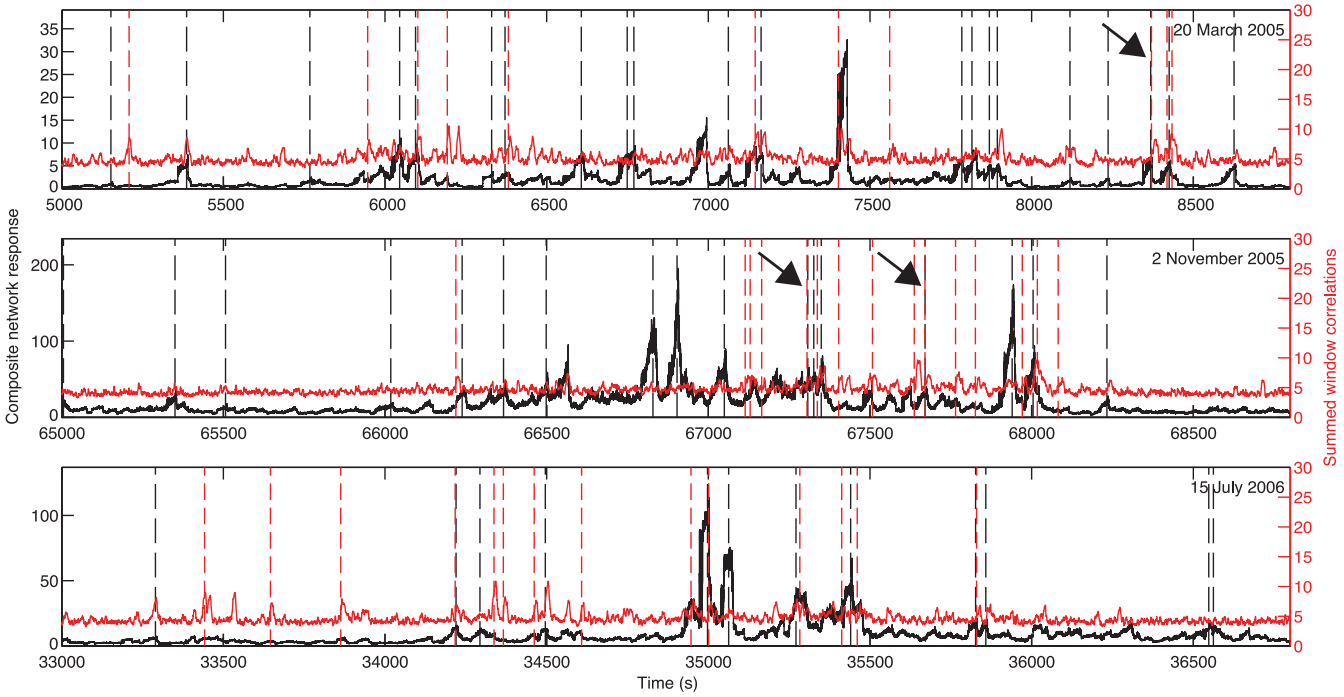

Figure 6. Network composite response compared to the repetitive waveform detector trace during three different hour-long time periods on 2005 March 20, 2005 November 2 and 2006 July 15, respectively. The vertical dashed lines represent the LFE templates that were detected with both methods: black for the 50 non-redundant LFE templates detected with our proposed beamforming method; red for the 40 non-redundant LFE templates detected with the repetitive waveform method. The black arrows indicate the three LFE templates that were detected with both methods. The red, horizontal dashed line indicates the five times the MAD detection threshold for the repetitive waveform method. The increased amplitude range of the composite network response facilitates LFE template detection.

during the same time periods, each method detects several events during time periods that the other method misses.

We then apply our method and the multiplet correlation method to 3 hr-long segments recorded, respectively, on 2005 March 20, 2005 November 2 and 2006 July 15; they were chosen because of increased TT activity during the three time periods, increasing the chance of detecting LFEs. We plot the resulting detector traces for each of the time segments in Fig. 6. As before, we can see that our proposed method has a greater amplitude range. After applying a five times the MAD threshold to the repetitive waveform detector trace and the detection picking process previously described to our proposed method, one can see that although both methods detect events at similar times, there are periods of time where only one method detects events. We note one example from each method to illustrate: between 65 000 and 66 100 s on 2005 November 2 for our proposed method and between 33 400 and 34 000 s on 2006 July 15 for the multiplet correlation method.

For each of the time periods, we then run a matched-filter search on the entire 2.5 yr-long data set using as templates each of the LFEs detected by our proposed method and the multiplet correlation method. The stacked waveforms of the LFEs detected by the matched-filter search can be considered to represent the LFE family (defined by its template). To identify any redundant LFE families, we calculated the average correlation coefficient between every unique pair of stacked waveforms for each method. Any average correlation coefficient between the stacked waveforms of a pair LFE families that is >0.6 indicates a pair of duplicate families. Of the 59 LFE templates detected with our proposed method, nine redundant templates were removed; six redundant templates were removed from the set of 46 LFE templates detected by the repetitive waveform method. After removing the 15 duplicate LFE templates and performing the same analysis on the entire set of LFE families detected by both methods, three pairs of LFE families, one LFE family in each pair from each method, qualified as redundant. Given this small overlap in template detection, this suggests that the two methods are in fact complementary and that when used in tandem, a maximum number of LFE templates can be detected.

The quality of the stacked waveforms of a given LFE family can also be used as a proxy for the quality of the detected LFE used as the family's template. To facilitate the classification of LFE template quality, we define three different qualities: an *A* quality template has stacked waveforms that exhibit clear *P*- and *S*-wave arrivals and that can be used to robustly locate the source of the LFE family; a *B* quality template has stacked waveforms that only exhibit clear *S*-wave arrivals with non-existent or difficult to pick *P*-wave arrivals and therefore are not adequate to robustly locate the LFE family; a *C* quality template has stacked waveforms that do not exhibit either clear *P*- or *S*-wave arrivals and are not at all adequate to locate the LFE family. We show an example of each quality LFE family in Fig. S2. Looking at the results in Table 1, one notes right away that both methods generate a similar number of *A* quality families and detections. Our method generates slightly more *A* quality families (both in number and proportionally to the total number of detected templates), but the difference is not significant.

5 DISCUSSION AND CONCLUSIONS

We have developed a new, fully automated method of detecting LFEs that is based on the network response to a theoretical point source. Our method is based on relatively easily satisfied hypotheses and its parameters can be adapted to any local region of study.

Table 1. Comparison of resulting LFE families use templates detected by the multiplet correlation method and our proposed beamforming method.

Method	No. unique families	Total detections	*A* families	*B* families	*C* families
Beamforming	50	34 167	38.0 per cent OR 19	42.0 per cent OR 21	20.0 per cent OR 10
Multiplet correlation	40	43 145	30.0 per cent OR 12	42.5 per cent OR 17	27.5 per cent OR 11

In regions of LFE activity that do not have established catalogues with LFEs that can be used as templates, this method can be used to efficiently generate a robust catalogue of LFE templates which can then be used in a matched-filter search.

Previous studies have demonstrated the strong temporal correlation between LFEs and TTs (e.g. Shelly *et al.* 2007). This can be further evidenced with the catalogue of LFEs detected with our proposed method: increased rates of detected LFEs coincide with detected TTs and can even potentially signal previously undetected TTs.

We then compared our method with a previously established LFE detection method (Brown *et al.* 2008). Our proposed method is based on the similarity of instantaneous energy across the entire network and how it stacks constructively when aligned with a move-out from a given theoretical source. The modified method of Brown *et al.* is based on the repetition of multiplets on one seismogram and if this repetition is reproduced on the other stations and components within a given time window size. The two main limitations of the multiplet correlation method are: (1) LFEs that do not repeat within the observational time period (i.e. an hour-long record) are not detectable; (2) if the length of the time window (15 s in our above comparison) is smaller than the sum of the largest move-out and the duration of the LFE waveform, a portion of the LFE waveforms will not be correlated as they will be outside the time window. These two different approaches are both capable of detecting a similar number of robust LFEs during the 3-hr–long test cases performed in this study. The multiplet correlation method, however, is theoretically able to detect smaller amplitude LFEs than our proposed beamforming method, whose detection efficiency based on constructively stacking instantaneous energy is proportional to the SNR of a given LFE. Only six of the 90 total LFE templates detected with both methods are considered redundant; we suggest here that the two methods are complementary as their detection criteria are different (ours is based on constructively stacked instantaneous energy and the multiplet correlation is based on repetitive waveforms) and both are necessary to detect a maximum number of LFEs.

We also note that our method is similar to two TT detection techniques: Wech & Creager (2008) and Kao & Shan (2004). Although these two methods along with our proposed method implement grid searches of theoretical sources, they do not optimize the same quantities: Wech & Creager (2008) optimize correlograms of 5-min–long time windows and Kao & Shan (2004) maximize the sum of migrated amplitudes within several second–long time windows. The method of Wech & Creager (2008) would not be applicable to our LFE data set because the 5-min–long time windows are an order of magnitude too long to detect several-second–long LFEs. Due to the fact that the source scanning algorithm of Kao & Shan (2004) sums amplitudes over several seconds, we would lose the phase information of the LFE waveforms which is on a shorter timescale given this study's frequency band of 1–2 Hz.

The characterization of LFEs and their relationship to TTs is complicated by the need for a large, robust catalogue of events that can be difficult to generate due a lack of previously established catalogues and computational obstacles. Our proposed method offers a straightforward approach based on simple hypotheses that can facilitate the creation of a catalogue of LFE templates and their associated families that is necessary for an in-depth study of the LFE and TT source region.

ACKNOWLEDGEMENTS

We thank the California Institute of Technology for the Meso-American Subduction Experiment data set used in this study. We also thank Aldo Zollo and Claudio Satriano for their programme that was used to calculate the theoretical traveltimes. We also acknowledge the use of resources provided by the European Grid Infrastructure. For more information, we refer the reader to the EGI-InSPIRE paper (http://go.egi.eu/pdnon). This work was supported by the Agence Nationale de la Recherche (France) under the contract RA0000CO69 (G-GAP) and by the European Research Council under the contract FP7 ERC Advanced grant 227507 (WHISPER). Finally, one of the figures was made with Generic Mapping Tools (Wessel & Smith 1998) .

REFERENCES

Bostock, M.G., Royer, A.A., Hearn, E.H. & Peacock, S.M., 2012. Low frequency earthquakes below southern Vancouver Island, *Geochem. Geophys. Geosyst.,* **13**(11), Q11007, doi:10.1029/2012GC004391.

Brown, J.R., Beroza, G.C. & Shelly, D.R., 2008. An autocorrelation method to detect low frequency earthquakes within tremor, *Geophys. Res. Lett.,* **35**(16), L16305, doi:10.1029/2008GL034560.

Brown, K.M., Tryon, M.D., DeShon, H.R., Dorman, L.M. & Schwartz, S.Y., 2005. Correlated transient fluid pulsing and seismic tremor in the Costa Rica subduction zone, *Earth planet. Sci. Lett.,* **238**(1), 189–203.

Frank, W.B., Shapiro, N.M., Kostoglodov, V., Husker, A.L., Campillo, M., Payero, J.S. & Prieto, G.A., 2013. Low-frequency earthquakes in the Mexican Sweet Spot, *Geophys. Res. Lett.,* **40,** 2661–2666.

Husker, A.L., Peyrat, S., Shapiro, N.M. & Kostoglodov, V., 2010. Automatic non-volcanic tremor detection in the Mexican subduction zone, *Geofis. Int.,* **49**(1), 17–25.

Husker, A.L., Kostoglodov, V., Cruz-Atienza, V.M., Legrand, D., Shapiro, N.M., Payero, J.S., Campillo, M. & Huesca-Pérez, E., 2012. Temporal variations of non-volcanic tremor (NVT) locations in the Mexican subduction zone: finding the NVT sweet spot, *Geochem. Geophys. Geosyst.,* **13**(3), Q03011, doi:10.1029/2011GC003916.

Iglesias, A., Clayton, R.W., Pérez-Campos, X., Singh, S.K., Pacheco, J.F., García, D. & Valdés González, C., 2010. S wave velocity structure below central Mexico using high-resolution surface wave tomography, *J. geophys. Res.,* **115**(B6), B06307, doi:10.1029/2009JB006332.

Kao, H. & Shan, S.-J., 2004. The source-scanning algorithm: mapping the distribution of seismic sources in time and space, *Geophys. J. Int.,* **157**(2), 589–594.

Kim, Y., Clayton, R.W. & Jackson, J.M., 2010. Geometry and seismic properties of the subducting Cocos plate in central Mexico, *J. geophys. Res.,* **115**(B6), B06310, doi:10.1029/2009JB006942.

Kostoglodov, V., Husker, A.L., Shapiro, N.M., Payero, J.S., Campillo, M., Cotte, N. & Clayton, R.W., 2010. The 2006 slow slip event and nonvolcanic tremor in the Mexican subduction zone, *Geophys. Res. Lett.,* **37**(24), L24301, doi:10.1029/2010GL045424.

Nadeau, R.M. & Dolenc, D., 2005. Nonvolcanic tremors deep beneath the San Andreas Fault, *Science,* **307**(5708), 389–389.

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

Payero, J.S., Kostoglodov, V., Shapiro, N.M., Mikumo, T., Iglesias, A., Pérez-Campos, X. & Clayton, R.W., 2008. Nonvolcanic tremor observed in the Mexican subduction zone, *Geophys. Res. Lett.,* **35**(7), L07305, doi:10.1029/2007GL032877.

Rivet, D., Campillo, M., Shapiro, N.M., Cruz-Atienza, V.M., Radiguet, M., Cotte, N. & Kostoglodov, V., 2011. Seismic evidence of nonlinear crustal deformation during a large slow slip event in Mexico, *Geophys. Res. Lett.,* **38**(8), L08308, doi:10.1029/2011GL047151.

Rivet, D. *et al.*, 2014. Seismic velocity changes, strain rate and non-volcanic tremors during the 2009–2010 slow slip event in Guerrero, Mexico, *Geophys. J. Int.,* **196,** 447–460.

Rogers, G. & Dragert, H., 2003. Episodic tremor and slip on the Cascadia subduction zone: the chatter of silent slip, *Science,* **300**(5627), 1942–1943.

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

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

Wech, A.G. & Creager, K.C., 2008. Automated detection and location of Cascadia tremor, *Geophys. Res. Lett.,* **35**(20), L20302, doi:10.1029/2008GL035458.

Wessel, P. & Smith, W.H.F., 1998. New, improved version of generic mapping tools released, *EOS, Trans. Am. geophys. Un.,* **79**(47), 579.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Figure S1. Location error associated with proposed detection method and relative moveout grid-search for an example LFE family detected on 2005 March 20.

Figure S2. Examples of different quality stacked waveforms.
(http://gji.oxfordjournals.org/lookup/suppl/doi:10.1093/gji/ggu058/-/DC1)

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.