Geophysical Journal International

Geophys. J. Int. (2015) **203,** 1207–1226 doi: 10.1093/gji/ggv337
GJI Seismology

Using a modified time-reverse imaging technique to locate low-frequency earthquakes on the San Andreas Fault near Cholame, California

Tobias Horstmann,[1] Rebecca M. Harrington[2] and Elizabeth S. Cochran[3]

[1]*Federal Institute for Geosciences and Natural Resources, Stilleweg* 2, *D*-30665 *Hannover, Germany. E-mail: tobias.horstmann@bgr.de*
[2]*Department of Earth & Planetary Sciences, McGill University,* 3450 *University St. Montreal, Quebec* H3A 0E8 *Canada*
[3]*Earthquake Science Center, U.S. Geological Survey,* 525 *S. Wilson Ave, Pasadena, CA* 91106, *USA*

Accepted 2015 August 13. Received 2015 August 11; in original form 2014 August 19

SUMMARY

We present a new method to locate low-frequency earthquakes (LFEs) within tectonic tremor episodes based on time-reverse imaging techniques. The modified time-reverse imaging technique presented here is the first method that locates individual LFEs within tremor episodes within 5 km uncertainty without relying on high-amplitude *P*-wave arrivals and that produces similar hypocentral locations to methods that locate events by stacking hundreds of LFEs without having to assume event co-location. In contrast to classic time-reverse imaging algorithms, we implement a modification to the method that searches for phase coherence over a short time period rather than identifying the maximum amplitude of a superpositioned wavefield. The method is independent of amplitude and can help constrain event origin time. The method uses individual LFE origin times, but does not rely on *a priori* information on LFE templates and families. We apply the method to locate 34 individual LFEs within tremor episodes that occur between 2010 and 2011 on the San Andreas Fault, near Cholame, California. Individual LFE location accuracies range from 2.6 to 5 km horizontally and 4.8 km vertically. Other methods that have been able to locate individual LFEs with accuracy of less than 5 km have mainly used large-amplitude events where a *P*-phase arrival can be identified. The method described here has the potential to locate a larger number of individual low-amplitude events with only the *S*-phase arrival. Location accuracy is controlled by the velocity model resolution and the wavelength of the dominant energy of the signal. Location results are also dependent on the number of stations used and are negligibly correlated with other factors such as the maximum gap in azimuthal coverage, source–station distance and signal-to-noise ratio.

Key words: Earthquake source observations; Broad-band seismometers; Seismicity and tectonics; Transform faults.

1 INTRODUCTION

The low-amplitude signals and lack of distinctive phase arrivals make locating tectonic tremor challenging. Numerous methods have been developed to locate tectonic tremor (hereafter referred to simply as tremor) since it was discovered over a decade ago (Obara 2002; Rogers & Dragert 2003). Some methods locate minute(s)-long tremor bursts with coherence of waveforms near predicted arrival times (e.g. the Source-Scanning Algorithm of Kao & Shan 2004) and envelope cross-correlation techniques (e.g. Obara 2002; McCausland *et al.* 2005; Wech & Creager 2008). However, many early methods are often unable to determine precise arrival times, especially for data sets with low signal-to-noise ratios. Studies later established that longer (∼100 s) tremor episodes are composed of individual bursts of energy, known as low-frequency earthquakes (LFEs) that are 1–2 s in duration, suggesting that locating the source of tremor waveforms requires locating individual LFEs (e.g. Shelly *et al.* 2007). Subsequently, several methods were developed to identify and locate the LFEs. In particular, much work centred on exploiting the similarity between closely located LFE waveforms by cross-correlating waveforms and then stacking them. Phase arrivals are then identified on the stacked waveforms and standard ray-tracing algorithms are applied (Shelly & Hardebeck 2010). Stacking techniques have been used to locate large numbers of tremor event families with high accuracy, enabling new insights into deep slip processes (Obara *et al.* 2004; Wech & Creager 2007; Brown *et al.* 2009; Kim *et al.* 2011). However, they require *a priori* information of LFE master templates, hundreds or thousands of repeating LFEs to identify phase arrivals, and all stacked events must occur within a small source region. Often correlation coefficients between LFEs in

a given event family stack can be as low as 0.33 on average, raising the question of how far apart assumed co-located events may be (Shelly *et al.* 2007).

Advances in computational resources in recent years have enabled the application and development of alternative methods for locating seismic signals, such as time-reverse-imaging (TRI) methods and back-projection methods (Larmat *et al.* 2006, 2009; Lokmer *et al.* 2009; O'Brien *et al.* 2011; Haney 2014). TRI methods back propagate time-reversed seismograms at the recording position into a velocity model to recover source information. Ideally, the back-propagated wavefields interfere constructively at the source position and origin time, resulting in large amplitudes. The method is also capable of recovering additional source details such as duration, provided the seismic source dimension is smaller than 1/2 of the dominant wavelength used in the back-propagated waveform (Anderson *et al.* 2011). In theory, one should be able to use TRI methods to locate tremor by exploiting the full waveform without prior knowledge of phase information. However, TRI methods are limited to areas with dense station coverage and a high-resolution velocity model. Furthermore, standard TRI methods do not consider phase coherence in the rebroadcast signal, which may be important when determining the locations of events recorded in short time windows, such as the individual LFEs within a tremor episode.

Here, we apply a modified TRI method to seismic data recorded on the San Andreas Fault (SAF) near Cholame, California to locate individual LFEs within tremor episodes. We refine the standard TRI method by searching for phase coherence of the time-reversed wavefield over short time periods with a matched filter approach, rather than examining individual snapshots of the time-reversed wavefield for constructive interference (e.g. large amplitudes). The advantage of searching for phase coherence is that it makes the waveform shape more important, making it suitable for low-amplitude signals and a more sparse station configuration. The matched filter approach also enables constraining the hypocentre solution in time. The matched filter therefore removes amplitude dependence and helps constrain event origin time. The novel feature and fundamental advantage of our modified TRI method is that it is able to locate individual LFEs (as opposed to stacks of LFEs with assumed common locations), unlike current location algorithms that either rely on stacking multiple LFEs (with assumed common locations) or locate entire tremor episodes, rather than individual second-long LFE bursts. The computational costs are similar to conventional LFE location methods, or to standard TRI methods, however the advance of our method is that it is the first of its kind that has the capability to locate individual LFEs with errors ∼5 km. The method is ideally suited to source studies of higher-amplitude LFEs, so that source parameter values can be estimated for single events (e.g. magnitude, radiated energy, stress drop, etc.), instead of proxies of source parameters based on LFE stacks.

In the following, we describe the waveforms used in this study, as well as the details of the modified TRI method used to locate individual LFEs. We first perform a synthetic test of the method's capability to determine a double-couple source location followed by a calculation of the hypocentre and origin time of an M_L 1.4 earthquake recorded by the same array of stations used to locate tremor (we calculate the local magnitude M_L). We then present the results of our location calculations of 34 individual LFEs, as well as the locations for a set of LFEs within a single event family (Shelly & Hardebeck 2010). We describe the estimates of location errors and discuss how parameters such as station geometry and number of stations used affect the results and end with the conclusions of our study.

2 DATA

We use seismograms from permanent surface and borehole stations as well as data recorded during a temporary seismic deployment near the Cholame segment of the SAF. The temporary deployment included 13 surface stations from the Karlsruhe Institute of Technology's instrument pool (termed the PERMIT array) at seven sites around Cholame, California. PERMIT stations (KIT01–KIT13) were equipped with STS-2 sensors and recorded data continuously at a 200 Hz sampling frequency. Data from permanent stations were recorded on station PKD from the Berkeley Digital Seismic Network (BDSN), stations PHOB and PHF from the Northern California Seismic Network (NCSN), borehole stations GHIB, EADB, FROB, VCAB, VARB, MMNB and JCSB of the Berkeley High Resolution Station Network (HRSN) and stations B072, B076, B078, B079 and B901 of the Plate Boundary Observatory (PBO). The station distribution shown in Fig. 1 and Table B1 covers approximately 50 km of the Cholame segment of the SAF. The study covers the time period in which the temporary PERMIT array was active, from May 2010 to July 2011.

3 METHODS

The premise of the time-reverse imaging technique is to locate seismic signals using constructive interference of rebroadcasted seismic waveforms. The technique uses numerical modelling to propagate a time-reversed seismic signal recorded at discrete points at the surface back into the subsurface. The gridpoint and time window at when the constructive interference of the rebroadcasted wavefield is maximum represent the source location and origin time of the signal. Fig. 2 illustrates the technique schematically in three parts: (I) time reversal of the data, (II) back propagation into the subsurface and (III) evaluation of coherence in time and space.

First, we preprocess the three-component seismograms with instrument response removed to recover true ground motions and apply a four-pole, two-pass bandpass filter between 1 and 5 Hz. The filter frequency band chosen accommodates computational demands while including the dominant frequencies of the observed tremor. Tremor energy observed near Cholame is typically concentrated in the 2–8 Hz band (Obara 2002; Schwartz & Rokosky 2007; Beroza & Ide 2011), but we use a lower frequency limit of 1 Hz, as several studies suggest that tremor energy is visible in the 1–2 Hz range (Shelly *et al.* 2007; Brown *et al.* 2008; Wech & Creager 2008). In addition to drastically increasing computational time, we expect that frequencies above 5 Hz (corresponding to a wavelength ∼ 1 km) adversely affect the back propagation, because corresponding wavelengths are smaller than the true resolution of the 3-D velocity model (Thurber *et al.* 2006). We test the effect of including higher frequencies when testing the method by locating an M_L 1.4 earthquake.

Next, we resample the seismograms to match the time step used in the back-propagation calculation. We then rotate the horizontal component seismograms to fault parallel and fault perpendicular orientations. Finally, we time-reverse the seismograms in order to use them as a source function in the back-propagation calculation. We use the staggered grid, finite-difference (FD) code of Bohlen (2002) with the Thurber *et al.* (2006) model for the back propagation (Fig. 2II). The velocity model uses a coordinate system with origin at 35°57.60′ N and 120°30.28′ W and *Y*-axis rotated to an azimuth of 139.2°. There are 240 gridpoints in the fault perpendicular direction (*X*-direction), 520 gridpoints in the fault parallel direction (*Y*-direction), 280 gridpoints in depth, all with 100 m spacing for a

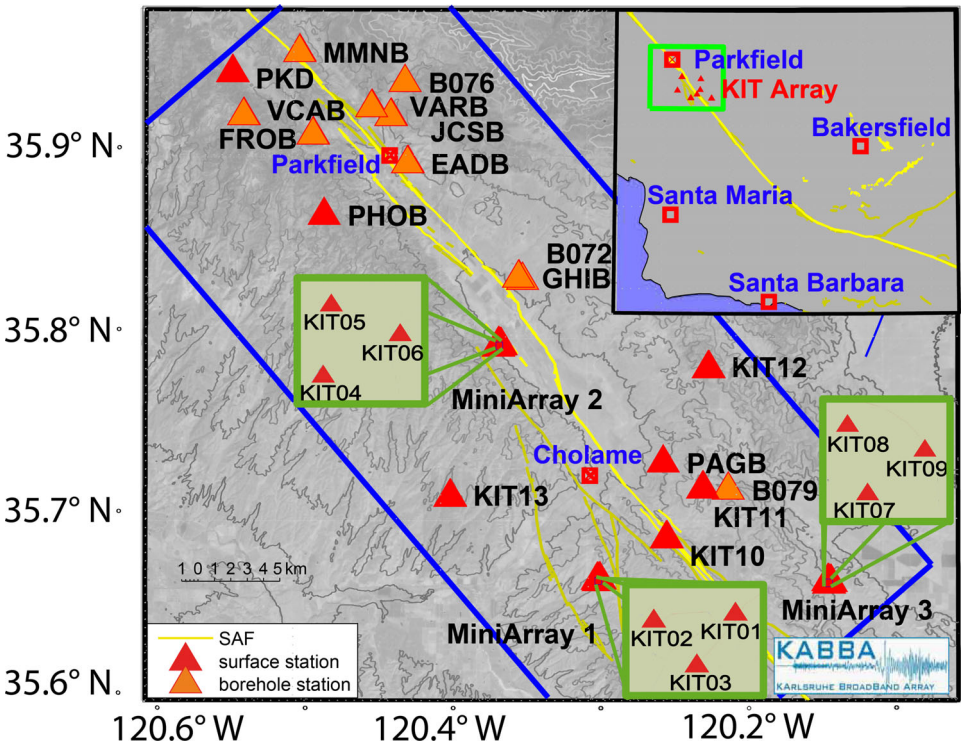

Figure 1. Seismic stations used in this study, including surface (red triangles) and borehole (orange triangles) stations. Yellow lines indicate the trace of the San Andreas Fault and blue lines indicate the extent of the velocity model volume. Grey shading and contour lines indicate elevation.

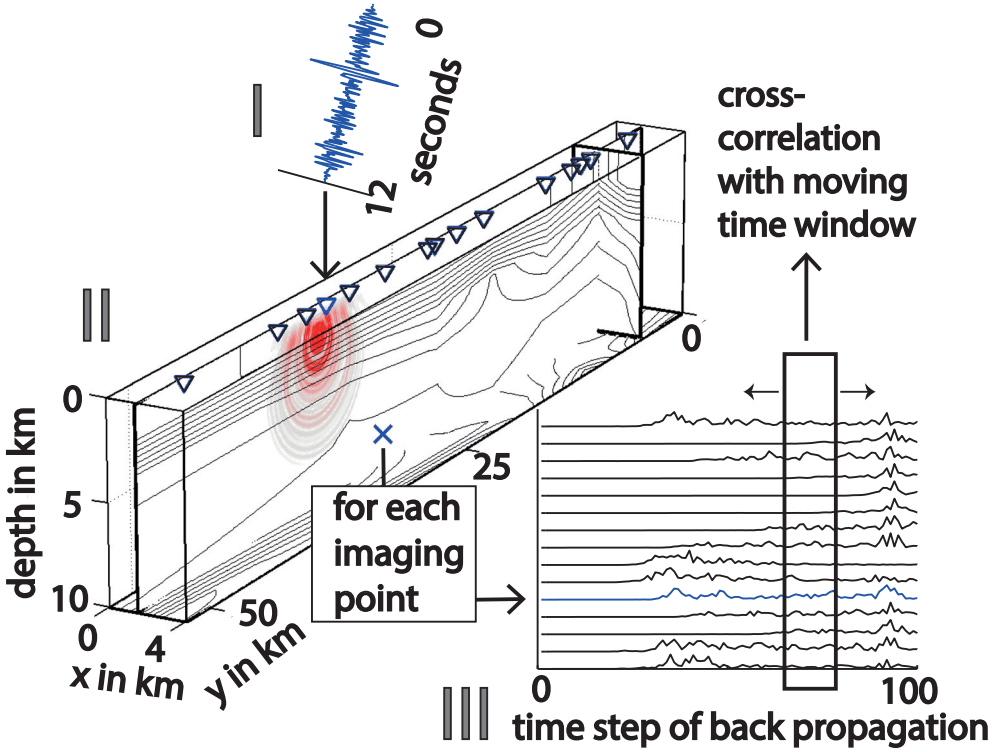

Figure 2. Schematic of the method: (I) time-reversed seismogram (note the reversed time axis) for a receiver (blue inverted triangle in II) , (II) curl field (red contours) from a single station in a slice of the model produced by back propagation and (III) the curl field from all 15 stations for an individual gridpoint (blue × in II). All curl fields from each receiver are summed to give the curl field function in time at a gridpoint. Hypocentre and origin time are determined by the maximum cross-correlation coefficient of the curl field function at each gridpoint calculated in moving time windows (III). The curl field for the gridpoint × at a receiver (blue inverted triangle in II) is shown by the blue trace. Contours along the fault plane in part (II) of the figure show the velocity model at the orientation described in Thurber *et al.* (2006) with positive *y* trending Southeast.

total of 34 944 000 gridpoints spanning 24 km × 52 km × 28 km. The upper left corner is located at −14 km in the *X*-direction and −2 km in the *Y*-direction (Thurber *et al.* 2006). The model geometry is shown in Fig. 1.

The expression for v_s is given by

$$v_s = \frac{v_p}{\sqrt{3}}. \quad (1)$$

Previous work suggests that the fault zone and surrounding damage zone near the fault have local variations in the ratio between v_p and v_s, therefore the assumption of a constant ratio is a simplifying assumption that will likely contribute to some of the location error (e.g. Zhang *et al.* 2009). Testing the method with separate v_p and v_s models will be the subject of a future study once separate models at similar resolution become available (Zeng *et al.* 2014).

Brocher (2005) determined an empirical relationship between v_p and density (ρ) using velocity measurements on rock samples, boreholes and seismic studies from California. The relationship is valid 1.5 km s^{-1} < v_p < 8.5 km s^{-1} and is given by

$$\rho = 1.6612 \cdot v_p - 0.4721 \cdot v_p^2 + 0.0671 \cdot v_p^3 - 0.0043 \cdot v_p^4 + 000106 \cdot v_p^5, \quad (2)$$

with v_p in km s^{-1} and ρ in g cm^{-3}. We use eqs (1) and (2) to calculate an *S*-wave and density model from the *P*-wave model of Thurber *et al.* (2006) that are used in the FD wave propagation calculation.

The calculation of elastic wave propagation employs an eighth order spatial Holberg FD operator and perfectly matched layers (PMLs) to absorb the energy at the boundaries as described by Komatitsch & Martin (2007). We set the spatial dimension of the PML to 15 gridpoints, the dominant frequency to 5 Hz and the velocity near the model boundary to 3.5 km s^{-1}. We interpolate the velocity model of Thurber *et al.* (2006) to an even grid spacing of 100 m and add the first layer 15 times to the top of our model to account for the PML absorbing frame. The reason for adding an absorbing boundary layer to the top of the model is to focus on direct arrival times and reduce spurious reflections from the free surface. A grid spacing of 100 m is chosen based on the minimum *S*-wave velocity in our model and the maximum frequency of the source signal to avoid grid dispersion due to inaccurate spatial sampling. We use a time step of 5 ms in the elastic wave propagation calculation based on the maximum *P*-wave velocity and the grid spacing. The 5 ms time step meets the Courant–Friedrichs–Lewy criterion (Courant *et al.* 1967), which ensures the stability of the FD code.

During the FD calculation, we write out the relative shear energy amplitudes for each rebroadcasted seismogram $s \in [1, N]$, $N \in \mathbb{N}$ every 50 ms at every gridpoint $j \in [1, G]$, $G \in \mathbb{N}$, where N is the number of stations and G denotes the number of gridpoints. The relative shear energy amplitudes function as an approximation for the curl field. We calculate them using displacement records, obtained from the integrated velocity records, and the FD code following the method of Dougherty & Stephen (1988). Thus, the result is a time series with $K \in \mathbb{N}$ elements of shear energy amplitudes for each station which approximate the curl field $X_{j,s} \in \mathbb{R}$ (Fig. 2III) at each point in the grid. In the remaining text we simply use the term 'curl field' to refer to the relative shear energy amplitude calculation.

In practice, one can use the wavefield itself, the divergence, or the curl of the wavefield (Larmat *et al.* 2009). The divergence and the curl field will be most sensitive to isotropic and shear displacement sources respectively. Here, we search for maximum coherence of the curl field because *S*-waves are more easily identified in tremor than *P*-waves (Wech & Creager 2007; La Rocca *et al.* 2009; Miyazawa & Brodsky 2008; Payero *et al.* 2008). Using the curl field (approximation) will prevent our method from effectively detecting the components of any seismic sources with isotropic character.

In the final step, we search for coherence in the curl field function in space and time by cross-correlating all combinations of the curl field time series in a sliding time window. We can write the absolute value of the cross-correlation between the curl fields of two rebroadcasted seismograms $s1 \in s$ and $s2 \in s$ for a given time window $w = [t; t + L]$ starting at time step $t \in [1; K - L]$ with a time window length of $L \in [1, K]$ at a given gridpoint j as

$$\mathrm{corr}_{j,t,s1,s2} = \left| \frac{\sum_{i=t}^{t+L} X_{j,s1}[i] \cdot X_{j,s2}[i]}{\sum_{i=t}^{t+L} (X_{j,s1}[i])^2 \cdot (X_{j,s2}[i])^2} \right|. \quad (3)$$

We then calculate a normalized median cross-correlation value $A_{j,t}$ between all combinations of N rebroadcasted seismograms for each gridpoint j and each time window w:

$$A_{j,t} = \mathrm{median}\{\mathrm{corr}_{j,t,s1,s2} \mid s1 < s2, s1 = 1...N, s2 = 1...N\} \quad (4)$$

We use a time window length L = 1.5 s with an increment of 0.1 s. While shorter windows may reduce the location uncertainty, we choose a time window length that is long enough to include the lowest dominant frequency energy at 1 Hz. We use the absolute value of the cross-correlation coefficient to account for 180° phase shifts due to radiation patterns of double-couple sources. We also impose a cross-correlation condition: if the local maximum value within the time window is lower than 50 per cent of the global maximum of the entire curl-field function, the cross-correlation value is set to zero. Mathematically it can be expressed as:

$$\mathrm{corr}_{j,t,s1,s2} = \{0 \mid \max(X_{j,s1}) > 2 \cdot \max(X_{j,s1}[f] \mid f \in [t, t+L]) \vee \max(X_{j,s2}) > 2 \cdot \max(X_{j,s2}[f] \mid f \in [t, t+L])\} \quad (5)$$

Setting correlation values for time windows with a maximum below 50 per cent of the global maximum to zero avoids numerical artefacts caused by division by small numbers in eq. (3) (zero terms are subsequently ignored), decreases computation time and avoids artefacts produced by locally high cross-correlation coefficients at random gridpoints. The source position is indicated by the highest median cross-correlation value over each time window centred on each gridpoint. We use the median value, as it is a more robust statistical indicator than the average, and is less sensitive to outliers. Similarly, the origin time is indicated by the time window containing the highest mean cross-correlation value. The 100 m grid spacing and the 0.1 s time window step constrain the resolution limit of the source location and origin time. We restrict the source position to the sub-volume of the model excluding the absorbing layers and the receiver positions, thereby ignoring the 15 gridpoints adjacent to the model boundaries and the top 30 gridpoints.

Determining an effective error estimation for the modified TRI method is less straightforward than for standard location methods that solve an inverse problem and have a quantitative measure of misfit. To understand the error estimation, consider the correlation coefficient values assigned to each gridpoint at the time step corresponding to the origin time: the source location has the highest correlation coefficient, and gridpoints surrounding the source location will also have higher correlation coefficients relative to gridpoints far away. Gridpoints having correlation values close to the maximum would correspond to gridpoints near the source location in classical location problems having similar traveltime residuals. We therefore define the correlation coefficients that are within 90 per cent of the

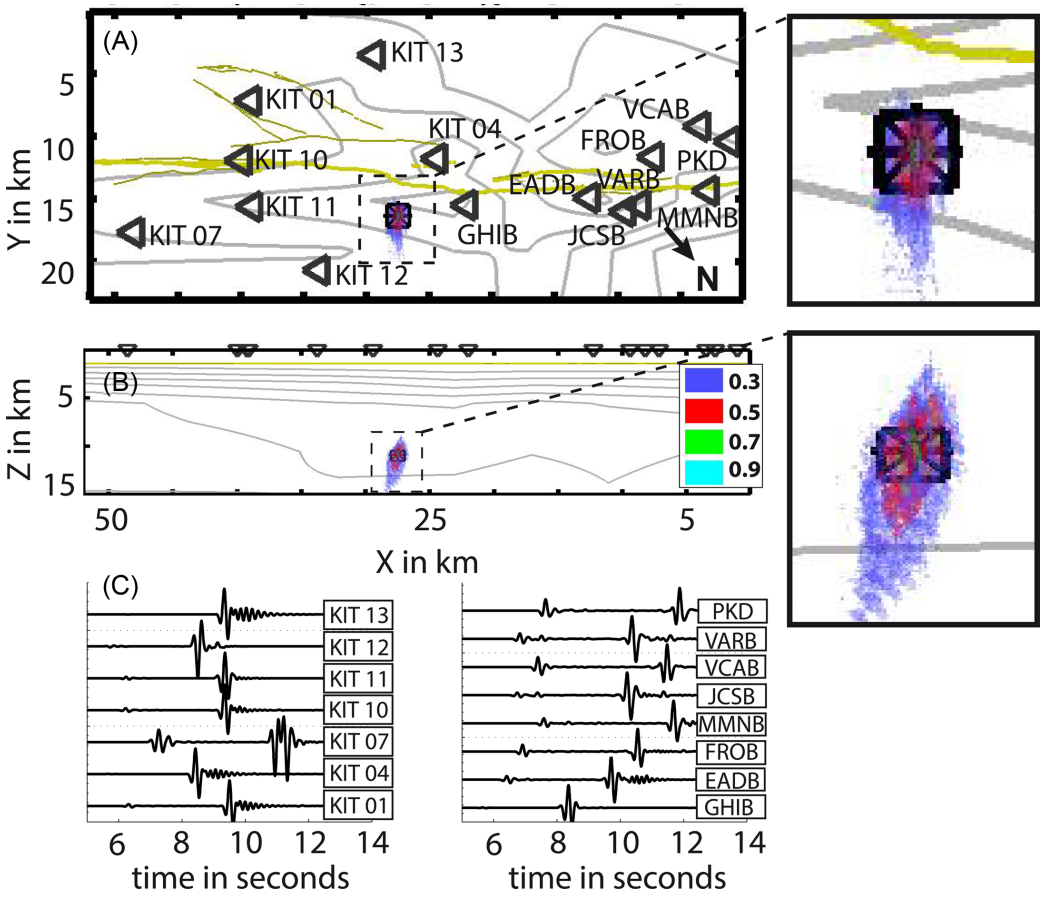

Figure 3. Hypocentral location of a synthetic double-couple source using the modified TRI method. (A) Map view with X denoting the fault parallel direction and Y denoting the fault perpendicular direction. (B) Fault-parallel cross-section with Z denoting depth. Coloured dots in panels (A) and (B) indicate the median cross-correlation value within the model volume, with the colour scale shown in panel (B). The centre of the black box marks the original source position of the synthetic source, the black star indicates the TRI calculated source location. Grey lines in panels (A) and (B) are contours of the 3-D shear wave velocity model, with contours showing slices through the model at 9 km depth (panel A) and 9 km in the Y-direction (panel B). Black triangles denote station locations and the yellow line is the surface trace of the San Andreas Fault. Panel (C) shows example vertical component synthetic waveforms for the stations shown in panel (A).

maximum value as the spatial cut-off for the error. One could also define the error in origin time in the same way, and we find when doing so, that our origin times are in agreement with the catalogued values. In general, the temporal error will not be greater than 2X the time window length (i.e. 3 s, in our case). We emphasize the hypocentral location comparison however, as the events chosen to test the method were selected from the catalogue (i.e. we selected windows containing LFEs from known origin times).

3.1 Method test: synthetic double-couple source

In order to verify that the modified TRI method is able to reliably recover point sources with more complex radiation patterns than isotropic sources, we generate synthetic seismograms of an earthquake with a double-couple mechanism. We use the 3-D velocity model described in Section 2 and evaluate the complete 3-D model volume for both the forward modelling and the waveform rebroadcasting in the synthetic test.

Fig. 3 shows the median cross-correlation value distribution within the model volume for the time step at $t = 7.9$ s, at which the highest median cross-correlation value occurs. The results depicted in Fig. 3 show high cross-correlation values concentrated around $X = 16$ km and $Y = 27$ km in model coordinates. The highest median cross-correlation value of 0.97 occurs at $X = 16.3$ km, $Y = 27.5$ km, $Z = 10.9$ km. The TRI determined location is 100 m shallower than the original source position but has the same epicentre. The relative error (based on the 90 per cent cut-off criterion discussed in Section 3) is ± 400 m horizontally and ± 800 m vertically. We repeat the test at depths similar to LFEs ($Z = 22.0$ km), finding that the method is able to recover the event with a decreased accuracy, due to the greater depth (1 and 1.5 km error in the X- and Y-directions, respectively, and 2.3 km error in depth). The highest median cross-correlation value for the test at 22 km is 0.93.

In addition to the test shown in Fig. 3, we tested the impact of signal-to-noise ratio on our ability to recover the source position using a 2-D model geometry (to save computational time). We were able to recover the source position with a signal to noise ratio of 1.66. We also tested the effect of adding random noise to the velocity model, finding that we were able to recover the source position with random perturbations of up to 20 per cent of the model values.

3.2 Method test: M_L 1.4 earthquake location

We test the performance of the modified TRI method with real data by locating a small (M_L 1.4) local earthquake (Fig. 4). Testing the method initially with earthquake data as opposed to LFE or tremor

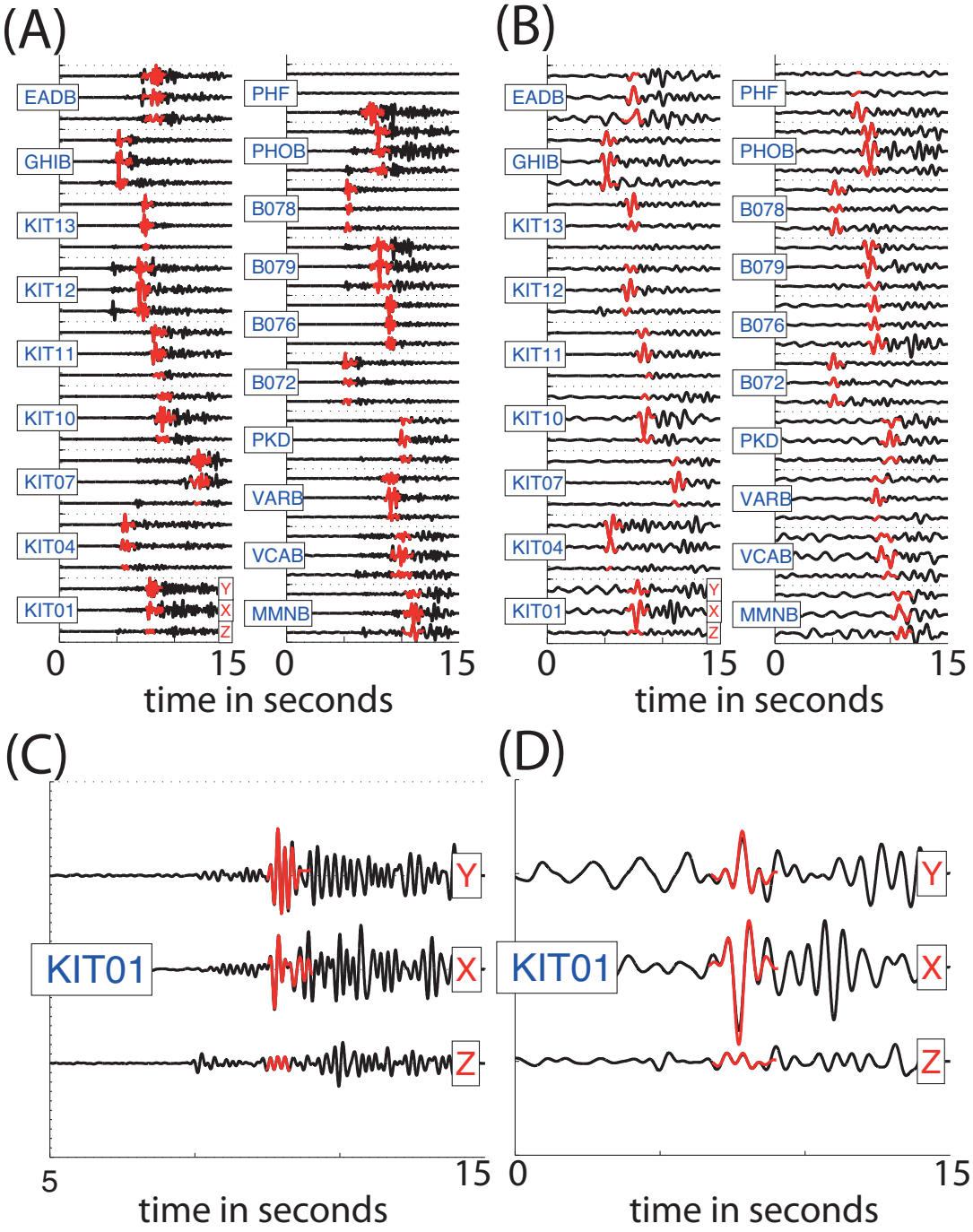

Figure 4. Waveforms of an M_L 1.4 earthquake that occurred at 04:27:06.11 on 2010 June 23. Waveforms filtered between (A) 1–5 Hz and (B) 0.5–2 Hz. Complete waveforms shown in black, tapered *S*-wave used in waveform rebroadcasting shown in red (magnified examples of tapered waveforms shown in panels C and D). *X*, *Y*, and *Z* labels represent fault parallel, fault normal, and vertical components of motion.

data has three advantages: (1) the signal-to-noise ratio is better, (2) a location obtained with ray-tracing methods is available for comparison and (3) the earthquake has energy in a wider useable frequency band than the tremor data, allowing us to evaluate the effect of the frequency band in the TRI locations. The test event is an M_L1.4 earthquake that occurred on 2010 June 23 at 04:27:06 am, with an epicentre at 35.793°N, 120.347°W and at 8.5 km depth (McClement *et al.* 2013). McClement *et al.* (2013) used *tomoDD* with additional seismic event data to produce an updated version of the velocity model used in this study and the location from the ray-tracing location are shown in Figs 5 and 6. The updated velocity model is currently a work in progress, and is not used here. Fig. 4 shows seismograms of the M_L 1.4 event. We perform two tests of the TRI method to evaluate the effect of different frequency bands on the location, the first using data filtered from 1 to 5 Hz and the second using data filtered from 0.5 to 2 Hz (Figs 4A and B, respectively). The magnified waveforms in Figs 4(C) and (D) show the tapered windows in red. We also perform tests using synthetic data to evaluate the effects of velocity model resolution and noise.

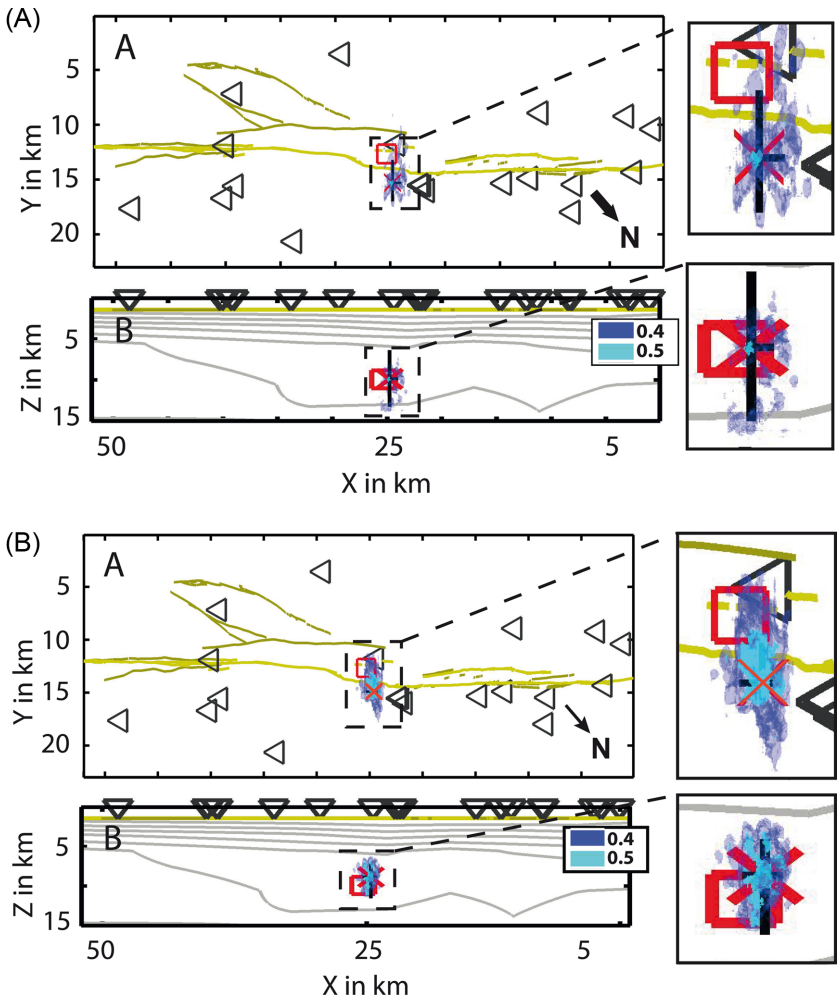

Figure 5. Hypocentral location of an M_L 1.4 earthquake recorded near Cholame, CA, shown in Fig. 4 determined using seismograms filtered between 0.5 and 2 Hz (A) using the complete seismograms and (B) using the tapered *S*-phase only. Colour code indicates the median cross-correlation value of the curl field within the model volume. The centre of the red box marks the solution at 35.793°N, 120.347°W and depth of 8.5 km (McClement *et al.* 2013). Red ×s indicates the modified TRI source location. Black crosses indicate horizontal and vertical location uncertainty. Black triangles denote station locations and yellow lines indicate the San Andreas Fault surface trace.

Figs 5 and 6 show the TRI location results using waveform data in the 0.5–2 Hz band and the 1–5 Hz band, respectively. Fig. 5 shows the distribution of the median cross-correlation values within the model space at $t = 14.0$ s, which includes the maximum median cross-correlation value of 0.56. The earthquake origin time of $t = 14.0$ s is relative to the start time of the seismograms at 04:27:00. The calculated source position is approximately 1 km east of the main SAF trace at $X = 15.3\,\mathrm{km}^{+1.7\,\mathrm{km}}_{-2.1\,\mathrm{km}}$, $Y = 24.8\,\mathrm{km}^{+0.2\,\mathrm{km}}_{-0.8\,\mathrm{km}}$, and $Z = 8.4\,\mathrm{km}^{+3.4\,\mathrm{km}}_{-3.5\,\mathrm{km}}$ depth. Note that we subtract the 1.5 km thickness of the absorbing layer on top of the model and report the true depth; however the figures show the full extent of the model. Figs 5 and 6 also show the location calculated by McClement *et al.* (2013) for comparison obtained by using a ray-tracing method with a velocity model based on an expansion the 3-D model of Thurber *et al.* (2006) using additional *S*-wave velocity values and deeper crustal earthquakes (indicated by red boxes in the figures; Thurber *et al.*, personal communication, 2015). The difference between the TRI source location and the McClement *et al.* (2013) solution is 2.7 km horizontally and 0.1 km in depth. (The ANSS catalogue location is similar to that obtained by McClement *et al.* (2013) at 35.7947°N, 120.3460°W and 9.29 km depth, however, we compare our location result to that of McClement *et al.* (2013), given that they also use a 3-D velocity model).

The maximum cross-correlation coefficient of 0.56 obtained for the test of the method using local earthquake data is lower than that obtained in the synthetic test (0.97) presented in Section 3.1. There are several reasons why the synthetic data should produce a higher cross-correlation coefficient. First, errors in the velocity model are not accounted for in the synthetic test, and slight shifts in phase arrivals due to velocity model inaccuracies are likely to smear the results. Second, the signal-to-noise ratio is lower for the earthquake data than for the synthetic test, so correlation values of the curl field are likely to be correspondingly lower in the short time windows used for correlation.

In an effort to reduce spurious correlation values resulting from errors in the velocity model, we taper the *S*-phases and repeat the location calculation. We determine the *S*-phase arrival by searching for the maximum vector amplitude on the horizontal component seismograms for each station and taper around the maximum amplitude using a three-second long cosine window. Fig. 4(B) depicts the tapered waveform in the frequency band of 0.5–2 Hz in red.

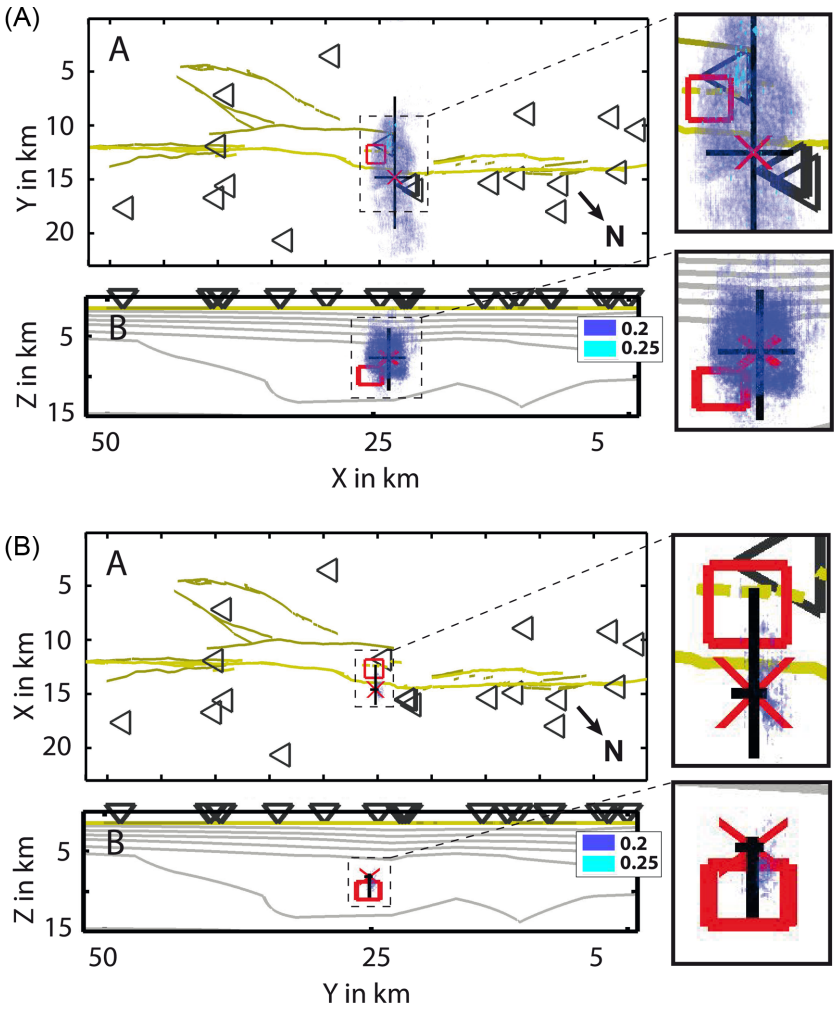

Figure 6. Hypocentral location of an M_L 1.4 earthquake recorded near Cholame, CA, shown in Fig. 4 determined using seismograms filtered between 1 and 5 Hz (A) using the complete seismograms and (B) using the tapered *S*-phase only. Colour code indicates the median cross-correlation value of the curl field within the model volume. The centre of the red box marks the solution at 35.793°N, 120.347°W and depth of 8.5 km (McClement *et al.* 2013). Red ×s indicate the modified TRI source location. Black crosses indicate horizontal and vertical location uncertainty. Black triangles denote station locations and yellow lines indicate the San Andreas Fault surface trace.

The location using the tapered waveforms shows a negligible change in hypocentral location compared to the un-tapered results, with the location differing by 0.4 km horizontally and 1 km vertically. We note an interesting feature in the curl field functions at the source position, that is, the gridpoint with the highest cross-correlation value (Fig. A1 in the auxiliary material). The curl field from the tapered *S*-phases pulses are partially shifted in time on select stations, up to one wavelength at the gridpoint with the highest value. The shift suggests the solution is spatially accurate only to within one wavelength. For example, a wave with 2 Hz dominant frequency travelling at 4 km s^{-1} implies that the location accuracy will be greater than 2 km for the source and receiver distribution used here. One reason for the shift could be that the window for the *S*-phase is slightly offset between stations where a large noise burst occurs. However, the signal-to-noise ratios are high, so the more likely reason for the offset of some stations might be localized inaccuracies in the velocity model.

We now locate the earthquake using seismograms filtered within a 1–5 Hz frequency band, the band with high tremor energy. Fig. 6(A) shows the median curl field cross-correlation value distribution at $t = 11.6$ s (i.e. origin time). The highest median cross-correlation value of 0.3 occurs at $X = 14.8\,\text{km}^{+4.8\,\text{km}}_{-7.5\,\text{km}}$, $Y = 23.6\,\text{km}^{+1.9\,\text{km}}_{-1.6\,\text{km}}$, and $Z = 6.2\,\text{km}^{+4.1\,\text{km}}_{-3.7\,\text{km}}$ depth. This solution differs from the solution in the 0.5–2 Hz band by 1.55 km horizontally, and 2.2 km vertically. The uncertainty in the location increases by a factor of approximately 3, and the error bars using the 0.5–2 Hz data and 1–5 Hz data locations overlap. Although the higher frequency data could theoretically provide smaller location errors than data filtered using a lower bandpass, the larger errors we observe here may be because the resolution of the velocity model is not sufficient. Consequently, the maximum median cross-correlation coefficient decreases to 0.3, as the defocusing can be multiple wavelengths in this frequency band.

When we use a tapered *S*-phase to reduce the effect of noise at the ends of the seismogram window, the location improves significantly. Fig. 6(B) shows the median curl field cross-correlation values at $t = 12.3$ s. The overall highest median cross-correlation value of 0.26 occurs at $t = 11.7$ s, similar to the un-tapered case. However, we have chosen to display the snapshot at $t = 12.3$ s. The reason for displaying the correlation values at 12.3 s is that the pattern of correlation values at any given time step may appear somewhat patchy, and the pattern at 12.3 s looks more similar to

the cumulative distribution of median cross-correlation values surrounding the global maximum. The calculated source position is at $X = 14.6\,\mathrm{km}^{+1.4\,\mathrm{km}}_{-2.3\,\mathrm{km}}$, $Y = 25.2\,\mathrm{km}^{+0.5\,\mathrm{km}}_{-0.3\,\mathrm{km}}$, and $Z = 6.7\,\mathrm{km}^{+2.6\,\mathrm{km}}_{-0.4\,\mathrm{km}}$ depth, similar to the location using the 0.5 Hz–2 Hz data. Thus, by tapering around the *S*-phase arrival, we reduce the influence of high amplitude spurious effects in the *S*-wave coda. Note that a higher-resolution velocity model would likely improve the median cross-correlation coefficients near the source position, as suggested by the high correlation coefficients in the synthetic case.

We include in the auxiliary material a figure that summarizes how the *X*, *Y*, *Z*, and cross-correlation values change through time for each of the four cases: tapered and untapered seismograms band-passed from 0.5–2 Hz and 1–5 Hz (Fig. A2). Overall, we observe higher median cross-correlation values for data within the 0.5–2 Hz band compared to data within the 1–5 Hz band. The location around the time of highest median cross-correlation value is stable for at least 5 s in all four cases. The maximum value remains stable for up to 5 s, partially because of the overlapping moving time window used in the cross-correlation of the curl field functions. The stability of the maximum correlation value (and thus the source position) over an extended time window highlights one advantage of the modified TRI method. The modified TRI method examines information over a finite time period rather than a particular time step as is used in the conventional TRI method, and the location results are based on waveform coherence rather than a maximum amplitude. Such attributes are helpful for locating low-amplitude LFE signals in a limited frequency band, but do not constrain the origin time precisely. In fact, when we initially tried a synthetic test of the conventional TRI method to locate an LFE source at 5 km depth we were unable to successfully locate the source without using the source time as *a priori* information. As discussed at the end of Section 3, the time errors can be as large as 2X the sliding-window length used in the cross-correlation. We use a time window length of 1.5 s; thus, the timing errors can be as large as 3 s. The reason is that different parts of the window contain high amplitude noise in a similar frequency band as the signal, which can obscure the cross-correlation peak in time. In addition, Fig. A1 shows that the curl field functions at the source gridpoint can be shifted up to one wavelength, which also contributes to temporal uncertainty. However, we again emphasize that the strength of the methods is the spatial resolution of individual LFEs, and that we use the catalogue LFE origin times as the more accurate measure of the origin time.

4 RESULTS

We apply the modified TRI method to locate individual LFEs within tremor episodes on the SAF near Cholame, California. This study focuses on LFE locations rather than detection, so we use the LFE event catalogue compiled by Shelly & Hardebeck (2010). We select 34 events with high signal-to-noise ratios that span the entire PERMIT recording period to account for changes in local noise conditions at stations. As a proxy for high signal-to-noise ratio, we use events with catalogue family cross-correlation coefficient $\geq$0.5 (as reported by Shelly & Hardebeck 2010) along with clearly visible *S*-phase energy. We limit the number of events to 34 due to the computational cost of the location calculation.

In the sections that follow, we first present the results of a single representative LFE location calculation, followed by the 34 events locations. We then discuss a subset of locations from a single LFE family (Shelly & Hardebeck 2010), as well as the location error estimation, and the implications of the location results.

4.1 Locating an individual low-frequency earthquake

We first apply the modified TRI method to a single representative LFE within a tremor episode. Fig. 7 shows a tremor episode recorded on 2010 September 2 with a 25 s long time window that includes a high-amplitude LFE. Before rebroadcasting to the subsurface we taper the waveforms around the maximum value, which is assumed

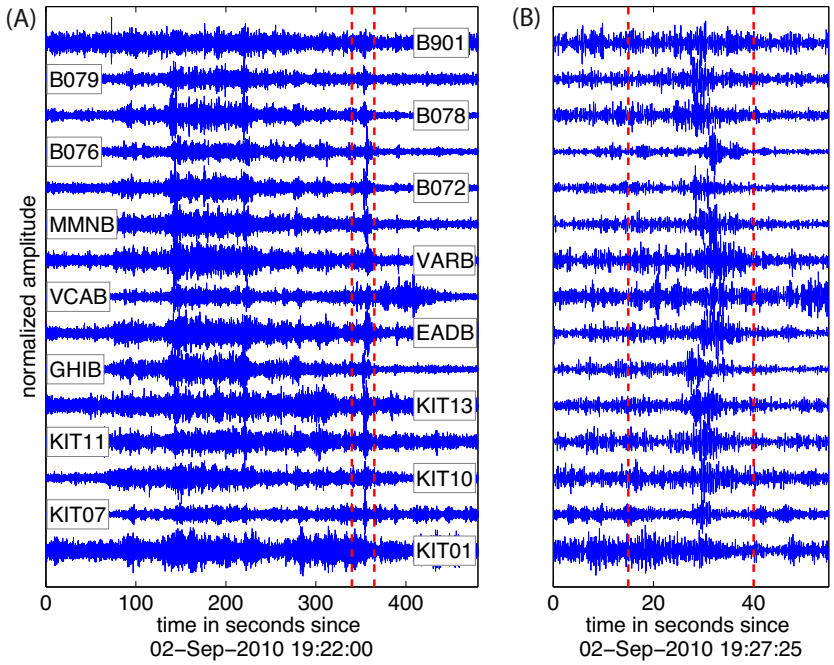

Figure 7. Tremor episode recorded on 2010 September 2, filtered between 1 and 5 Hz. (A) North component seismogram of the complete tremor episode. Red dashed lines mark the boundaries of the 25 s time window used for rebroadcasting. (B) Magnification of the time window in panel (A) that includes an LFE event.

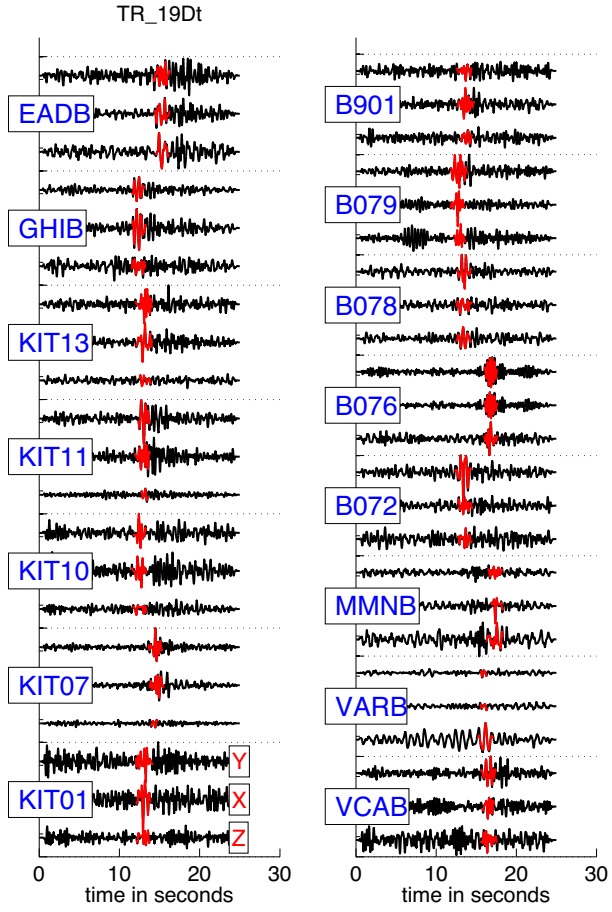

Figure 8. LFE waveforms filtered between 1 and 5 Hz recorded on 2010 September 2. Complete waveforms shown in black, tapered *S*-phase signals used for waveform rebroadcasting shown in red. Waveforms are identical to those shown in the time window indicated by red dashed lines in Fig. 7.

to represent the *S*-phase. We find that for our data set, tapering windows determined automatically require visual inspection to confirm the maximum amplitude is associated with the LFE due to lower signal-to-noise ratios at some stations. Fig. 8 shows the waveforms of all components used for rebroadcasting as well as the tapered waveforms.

Fig. 9(I) shows the hypocentre calculation using the tapered *S*-wave filtered between 1–5 Hz. The calculated source location is at $X = 11.7\,\mathrm{km}^{+4.3\,\mathrm{km}}_{-3.7\,\mathrm{km}}$, $Y = 33.3\,\mathrm{km}^{+1.5\,\mathrm{km}}_{-0.6\,\mathrm{km}}$, and $Z = 21.3\,\mathrm{km}^{+1.8\,\mathrm{km}}_{-5.7\,\mathrm{km}}$ depth. Again, note we have subtracted the 1.5 km thickness of the absorbing layer on top of the model from the depth. The highest median curl field cross-correlation value is 0.4, which is 0.1 higher than for the M_L 1.4 earthquake in the 1–5 Hz frequency band, in spite of the signal-to-noise ratio being lower. This difference is likely a reflection of the smaller number of stations used to locate the LFE compared to the earthquake location (15 versus 19 stations, respectively). The mean cross-correlation value decreases as the number of stations increases, likely due to the variation in waveforms across stations. The error bars for the LFE (Fig. 9I) are twice as large as the error bars for the earthquake (Fig. 6B), likely due to the greater depth of the LFE event (resulting in longer travel paths), and/or the lower signal-to-noise ratio.

We compare the calculated LFE source location with the locations determined by Shelly & Hardebeck (2010). We note that a direct comparison of the two methods is inappropriate, because the TRI method locates individual LFEs, while Shelly & Hardebeck (2010) stack up to 400 LFEs to obtain *P*- and *S*-phase picks, and locate the LFE stack using ray-tracing methods. Nevertheless, the stacked LFE family location serves as a reference point. The Shelly & Hardebeck (2010) location is at $X = 11.63$ km, $Y = 33.71$ km and $Z = 23.25$ km (Fig. 9I). The horizontal difference between the locations is 0.4 km horizontally and 2 km vertically.

A detailed look at the distribution of the median cross-correlation values in Fig. 9(I) reveals two concentrated spots with relatively high cross-correlation values extended in the *X*-direction perpendicular to the fault (causing the large uncertainty in the *X*-direction). We speculate that the spread of the high cross-correlation values could be reduced in the *X*-direction with increased station coverage in the fault-perpendicular direction. We note that two simultaneous events could also cause a spreading of high correlation values, however, examination of the time series suggests that only one event occurs during the time window.

Locating the LFE using un-tapered tremor waveforms increases the error by a factor of approximately 2 to 3 as was observed in the test case of the M_L 1.4 earthquake when using data in the 1–5 Hz frequency band (Fig. 9II). The source position for the un-tapered case is $X = 19.2\,\mathrm{km}^{+1.6\,\mathrm{km}}_{-14.5\,\mathrm{km}}$, $Y = 31.5\,\mathrm{km}^{+5.1\,\mathrm{km}}_{-4.1\,\mathrm{km}}$, and $Z = 26.7\,\mathrm{km}^{+0.6\,\mathrm{km}}_{-11.3\,\mathrm{km}}$, with the highest median cross-correlation coefficient of 0.33 at $t = 18.1$ s. The error bars for the tapered and untapered cases overlap.

4.1.1 Locating multiple LFEs

Next, we apply the modified TRI method with *S*-phase tapering to 34 LFEs from the Shelly & Hardebeck (2010) catalogue. Table 1 and Fig. 10(I) show the location results. Fig. 10(II) shows the locations based on the spatial mean error, which we will discuss below in Section 4.2. The number of stations used for the localization of each LFE event differs depending on data availability and the signal-to-noise ratio of the LFE seismogram. The number of stations used in each location, as well as other parameters such as azimuthal gap, signal-to-noise ratio, cross-correlation, etc., are listed in Table C1 of the appendix.

4.2 LFE locations within one family

We also examine how the TRI method locates 12 LFEs within a single event family identified by Shelly & Hardebeck (2010). Shelly & Hardebeck (2010) locate an LFE family by stacking waveforms from up to 400 LFE with similar waveforms in order to increase the signal-to-noise ratio, thereby making *P*- and *S*-phase arrivals more visible. The assumption is that all events within a family have similar waveforms because they repeatedly rupture the same patch of the fault. However, due to the low amplitude of tremor and corresponding low cross-correlation values, the individual source locations may not be identical, and it is unknown how large the differences in location are between individual LFEs in a family. The advantage of the modified TRI method used here is that it is capable of locating individual LFEs, and potentially investigating the variation in source locations.

Figure 11(I) shows the source locations as determined using the highest median cross-correlation value for the 12 LFEs within an event family. The source locations cluster around the SAF trace and the stacked family location, and apart from three of the LFEs, all calculated depths lie between 20 km and 26 km. The median difference between the modified TRI source locations and the LFE

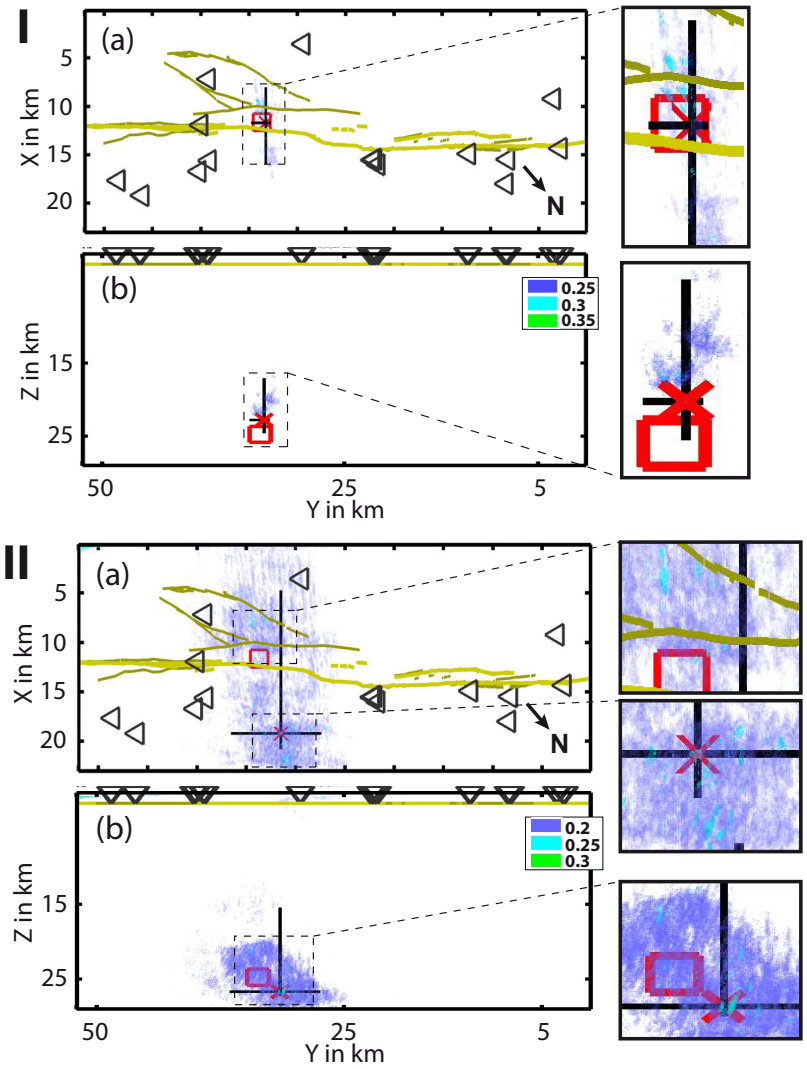

Figure 9. Hypocentral location of an individual LFE event in the frequency band of 1–5 Hz (I) with and (II) without tapering. Tremor episode is shown in Fig. 7 and waveforms used for rebroadcasting are shown in Fig. 8. Colour code indicates the median cross-correlation value of the curl field within the model volume that determines the source location (red 'x'). Centre of the red box marks the solution of Shelly & Hardebeck (2010) for reference. Black lines show estimated error range. Black triangles denote station positions and yellow lines indicate the San Andreas fault surface trace.

family location of (Shelly & Hardebeck 2010) is 1.25 km in the *X*-direction, 2.25 km in *Y*-direction and 5.1 km vertically. However, we are unable to distinguish between errors in the modified TRI locations and real perturbations within the source location without a higher resolution velocity model.

We also introduce a second approach to estimate the source position. Instead of only considering the global maximum value, we determine the source position by using median of the error range. As stated in the Methods section, the error range is determined by the positions of all median curl field cross-correlation value maxima exceeding 90 per cent of the global maximum median cross-correlation value. Figs 10(II) and 11(II) show the LFE locations using the median of the error range. The locations calculated using the median of the error range are more tightly clustered and centred around the SAF than the locations that use the highest median cross-correlation value. We expect the LFE locations for a single event family to be tightly clustered; thus, using the median of the error range to determine the source position may be preferable. While it may seem as if all correlation values above 90 per cent have an equal chance of being the true solution, it is not the case. Under the assumption that the cross-correlation values decrease with distance from the true source, we can determine the source location by using the shape of the volume that is defined by the gridpoints with cross correlation values that exceed 90 per cent. Using the median value allows us to include more gridpoints for a more statistically robust solution. LFE source location information determined with the median error extent is included in Table B2 in the appendix.

5 DISCUSSION

The calculated LFE locations in Fig. 10 lie mainly south of Cholame ($Y \geq 32$ km) with most events at depths between 15 and 25 km. The median uncertainty is 5.0 km in the *X*-direction, 2.6 km in the *Y*-direction and 4.8 km in depth. The horizontal uncertainties are

Table 1. LFE source locations with estimated error range in the local grid coordinate system *X*, *Y* and *Z*, and the corresponding latitude and longitude values. The last two columns give the horizontal and vertical differences to the stacked LFE family source locations of Shelly & Hardebeck (2010).

	Start of time window	*X* (km)	*Y* (km)	*Z* (km)	Latitude (°N)	Longitude (°W)	Hor. diff. (km)	Vert. diff. (km)
1	2010-07-01 17:42:50	$17.60^{+1.3}_{-9}$	$37.00^{+3.5}_{-1.3}$	19.00^{+8}_{0}	35.743	−120.221	6.11	−7.50
2	2010-07-01 17:42:50	$17.60^{+1.3}_{-9}$	$37.00^{+3.5}_{-1.3}$	19.00^{+8}_{0}	35.743	−120.221	6.11	−7.50
3	2010-07-02 04:42:45	$16.10^{+4.3}_{-10.6}$	$26.00^{+1.7}_{-1.4}$	$24.50^{+2.5}_{-3.3}$	35.809	−120.313	3.20	−2.00
4	2010-07-02 13:59:20	$6.40^{+0.4}_{-0.8}$	$36.10^{+1.1}_{-0.3}$	$23.80^{+2.5}_{-0.6}$	35.683	−120.322	5.57	0.05
5	2010-07-03 08:21:15	$13.60^{+5.2}_{-1.9}$	$39.00^{+1.3}_{-1.3}$	$20.60^{+1.3}_{-2.2}$	35.706	−120.240	5.55	−3.65
6	2010-07-04 11:08:35	$15.60^{+0.2}_{-8.8}$	$34.30^{+0.6}_{-2.3}$	$24.30^{+1.8}_{-2.1}$	35.750	−120.257	4.02	−0.45
7	2010-08-04 10:32:05	$8.80^{+3.6}_{-1.5}$	$43.80^{+0.8}_{-3.9}$	$25.60^{+1.1}_{-8.4}$	35.645	−120.246	3.97	0.10
8	2010-08-04 10:40:00	$10.10^{+0}_{-0.1}$	$44.90^{+0.3}_{-0.1}$	$25.60^{+0.4}_{0}$	35.645	−120.227	2.37	−0.90
9	2010-08-06 00:35:25	$14.40^{+0}_{-3.7}$	$42.80^{+2.6}_{0}$	$17.70^{+4.7}_{0}$	35.685	−120.206	6.20	−8.80
10	2010-08-06 00:35:45	$11.60^{+3.5}_{-8.4}$	$48.40^{+1.5}_{-4.2}$	$26.90^{+0.4}_{-8.1}$	35.630	−120.189	0.60	0.90
11	2010-09-02 10:31:40	$7.00^{+6.5}_{-2.2}$	$46.80^{+2.9}_{-4.5}$	$27.10^{+0.2}_{-4.3}$	35.614	−120.239	5.73	0.60
12	2010-09-02 17:32:40	$7.90^{+1.5}_{-0.1}$	$37.20^{+0.3}_{-0.3}$	$23.60^{+2.5}_{-1.5}$	35.685	−120.301	2.58	1.10
13	2010-09-02 19:27:40	$11.70^{+4.3}_{-3.7}$	$33.30^{+1.5}_{-0.6}$	$22.80^{+1.8}_{-5.7}$	35.733	−120.297	0.42	−1.95
14	2010-09-10 02:29:05	$12.90^{+0.1}_{0}$	$40.90^{+0}_{-1.3}$	$24.00^{+0}_{-1.2}$	35.689	−120.232	2.61	−2.50
15	2010-10-09 08:52:50	$18.70^{+0.4}_{-10}$	$41.30^{+3.5}_{-0.4}$	$23.80^{+0.2}_{-7}$	35.720	−120.181	7.80	−2.70
16	2010-10-12 12:51:45	$13.80^{+0.4}_{-0.3}$	$28.40^{+0.1}_{-0.1}$	$22.50^{+0.3}_{-0.1}$	35.779	−120.315	2.70	−1.75
17	2010-10-12 21:07:45	$3.50^{+15.6}_{-0.3}$	$32.80^{+9.7}_{-1.2}$	$11.40^{+13.8}_{-5}$	35.689	−120.370	12.46	−15.10
18	2010-10-12 21:07:50	$10.30^{+1.1}_{-0.3}$	$39.30^{+0.1}_{-0.1}$	$15.20^{+0.7}_{0}$	35.684	−120.266	3.50	−11.30
19	2010-10-22 14:31:30	$20.20^{+0.6}_{-2.7}$	$23.60^{+0.1}_{-0.7}$	$23.70^{+0.7}_{-2.6}$	35.849	−120.296	7.69	−2.80
20	2010-10-26 12:17:15	$4.10^{+10}_{-0.5}$	$37.30^{+0.6}_{-1.9}$	$18.90^{+0.5}_{-5.3}$	35.662	−120.332	6.38	−3.60
21	2010-11-07 02:45:45	$11.40^{+0}_{-5.1}$	$37.10^{+6.4}_{-1.3}$	$16.50^{+6.3}_{-4.9}$	35.706	−120.272	11.23	−10.00
22	2010-11-16 08:54:10	$14.80^{+0.2}_{-5.5}$	$43.20^{+3.7}_{-3.2}$	$23.50^{+3.8}_{-5}$	35.684	−120.200	3.80	−3.00
23	2010-11-17 10:51:25	17.10^{+2}_{-1}	$33.50^{+1.2}_{-1.4}$	$27.40^{+0.1}_{-4.2}$	35.764	−120.251	5.52	3.65
24	2010-11-17 10:57:55	$9.60^{+0}_{-1.3}$	$41.00^{+0.5}_{-2.6}$	$26.10^{+0.4}_{-3.7}$	35.669	−120.259	2.18	0.60
25	2010-11-24 11:26:15	14.10^{+3}_{0}	$16.30^{+0.6}_{-0.2}$	$10.40^{+1.9}_{-3}$	35.863	−120.400	3.66	−15.35
26	2010-12-01 12:36:10	$10.50^{+3.4}_{-5.1}$	$44.90^{+0.7}_{-4.9}$	$24.70^{+2.2}_{-10.4}$	35.647	−120.224	2.24	−1.80
27	2010-12-01 12:39:15	$13.00^{+2.7}_{-5.3}$	$39.30^{+2.4}_{-3.3}$	$25.10^{+2.2}_{-7.5}$	35.700	−120.243	3.95	−1.40
28	2010-12-28 14:46:35	$13.10^{+0.3}_{-1.9}$	$40.90^{+0.5}_{-0.5}$	$23.30^{+3.3}_{-2.5}$	35.690	−120.231	3.31	0.05
29	2010-12-28 22:27:55	$8.50^{+2.1}_{-0.4}$	$44.50^{+0.3}_{-4.2}$	$22.70^{+0.1}_{-10}$	35.638	−120.243	2.56	−0.55
30	2011-01-03 03:45:00	$2.30^{+6.1}_{0}$	$14.60^{+2.1}_{0}$	$13.60^{+6.6}_{-8.9}$	35.805	−120.512	8.80	−7.40
31	2011-01-23 19:41:20	$10.20^{+0.7}_{-2.6}$	$42.20^{+2.3}_{0}$	$21.10^{+1.5}_{-0.1}$	35.664	−120.246	0.98	−5.40
32	2011-02-03 01:42:45	$7.80^{+4.1}_{-1.3}$	$50.00^{+0}_{-2.9}$	$24.10^{+2.8}_{-1.4}$	35.597	−120.210	4.14	−2.40
33	2011-03-12 07:41:25	$9.10^{+5.2}_{-0.5}$	$49.90^{+0}_{-9.9}$	$22.10^{+5.2}_{-7.4}$	35.605	−120.199	7.43	−4.40
34	2011-04-04 21:05:55	$7.00^{+4.2}_{-4.3}$	$43.20^{+0.1}_{-3.2}$	$17.80^{+0.7}_{-6.5}$	35.638	−120.265	4.06	−8.70

larger in the *X*-direction likely because the station distribution is elongated in the *Y*-direction.

While the source determination from the median of the error range (Figs 10II and 11II) may provide a preferred horizontal source location, it may underestimate depth in cases when the error range extends to the base of the model. For example, consider a case in which the shallow extent of the high cross-correlation values is at a depth of 15 km. Assuming the actual source depth is 25 km, then the deeper high cross-correlation values would extend to 35 km. However, 35 km depth exceeds model dimensions. In such cases, the depth will be underestimated. Therefore the source depths may be underestimated in some cases due to the model geometry. In order to reduce computation time, we do not extend the model to greater depths because the velocity model of Thurber *et al.* (2006) does not extend beyond 30 km. Here we validate the method using LFEs with estimated depths ranging from 16–30 km (Shelly & Hardebeck 2010). Future work will entail estimating LFE locations with an improved *S*-wave velocity model that extends to greater depths based on LFE arrivals (Zeng *et al.* 2014).

The LFE locations exhibit larger location uncertainties compared with the test cases using earthquake and synthetic data, as expected. The larger uncertainty in the LFE location compared to the earthquake location is likely due to the lower signal-to-noise ratio of the LFE waveforms. Another factor that may increase uncertainty is that we use the 3-D *P*-velocity model of Thurber *et al.* (2006) and interpolate it to 0.1 km grid spacing. The true model resolution is sparser, ranging between 2 and 20 km across the model. Given the frequency band used to locate LFEs (1–5 Hz), complex velocity structure will not be fully accounted for resulting in decreased cross-correlation values and increased location errors. Other factors contributing to error may be the derivation of an S-wave velocity model from the *P*-wave model assuming a Poisson ratio of 1.73,

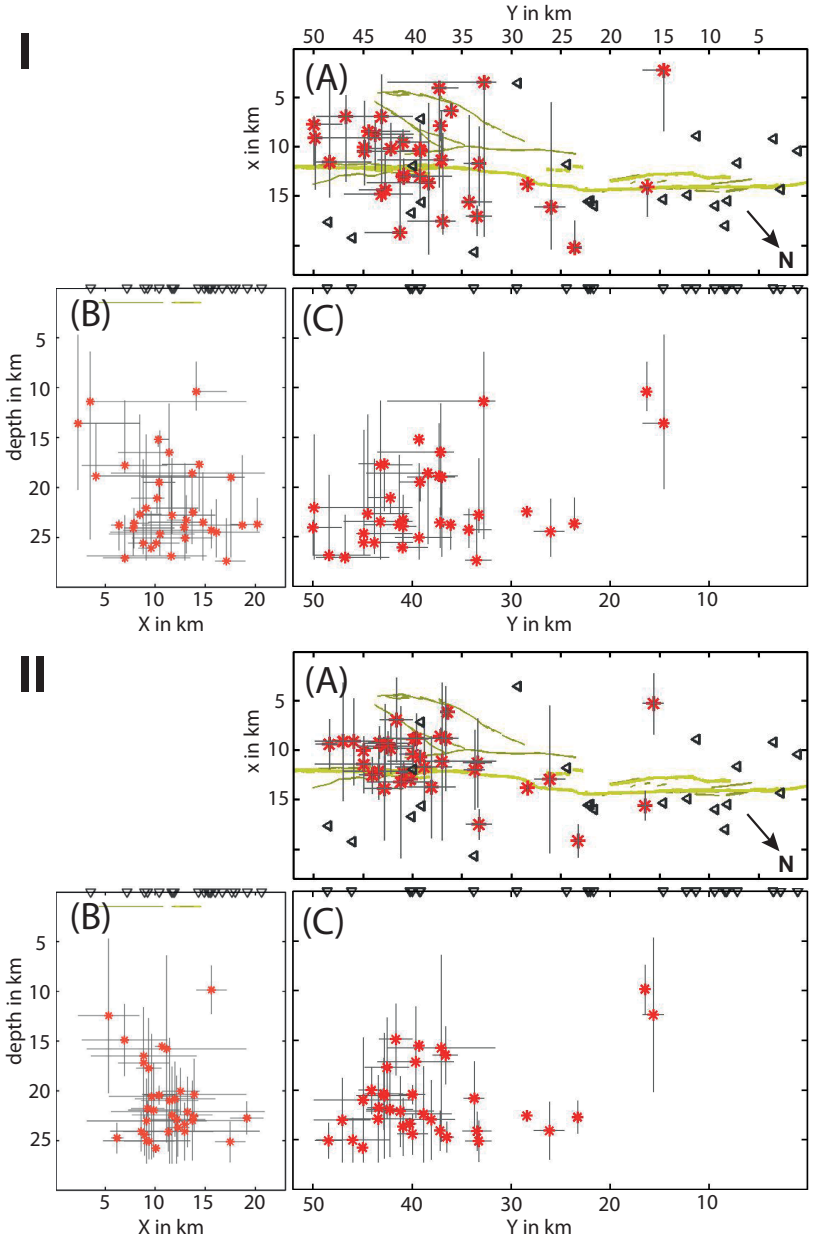

Figure 10. Hypocentral locations of 34 LFEs determined by (I) the maximum cross-correlation value and (II) the median error range (see the text) using the tapered *S*-phases. (A) map view, (B) vertical section in the *X*–*Z* plane and (C) vertical section in the *Y*–*Z* plane. Depth axis includes the thickness of the absorbing layer (1.5 km). Red stars show calculated source positions, black lines show the estimated error range. Black triangles denote all possible station locations and yellow lines indicate the San Andreas fault surface trace. Note that the stations used in the location calculation may differ for each LFE. North is indicated by the black arrow in panel (A).

and using an empirically determined relationship between v_p and density (ρ) (Brocher 2005). The highly clastic rock near the fault likely has a wide range of v_p/v_s ratios. Similarly, the contrasting geological units on both sides of the fault likely have different v_p/v_s ratios. Finally, the wave propagation calculations assume the ideal case of an isotropic and elastic medium.

We also examine how individual parameters such as the signal-to-noise ratio of the seismograms, the number of stations used, the source–station distance, and the azimuthal gap affect the location result (Fig. C1). The dependence of horizontal uncertainties and the maximum cross-correlation coefficient on each of the individual factors provides a measure of the influence each factor has on the LFE location quality. Both the signal-to-noise ratio and the number of stations show little correlation with the location error. We speculate that the main reason may be that tapering the *S*-phase effectively removes most of the noise, however, we were unable to constrain hypocentre solutions robustly without tapering. There is also a slight decrease in location uncertainty as the maximum azimuthal gap between stations decreases. The only unambiguous (negative) correlation we do observe is between the highest median cross-correlation value and number of stations, with the maximum cross-correlation decreasing as the number of stations increases. The negative correlation may result from the degrees of freedom decreasing as more station data are used, making it more difficult for the algorithm to find a gridpoint that optimizes the coherence between all rebroadcasted wavefields. The effect of individual

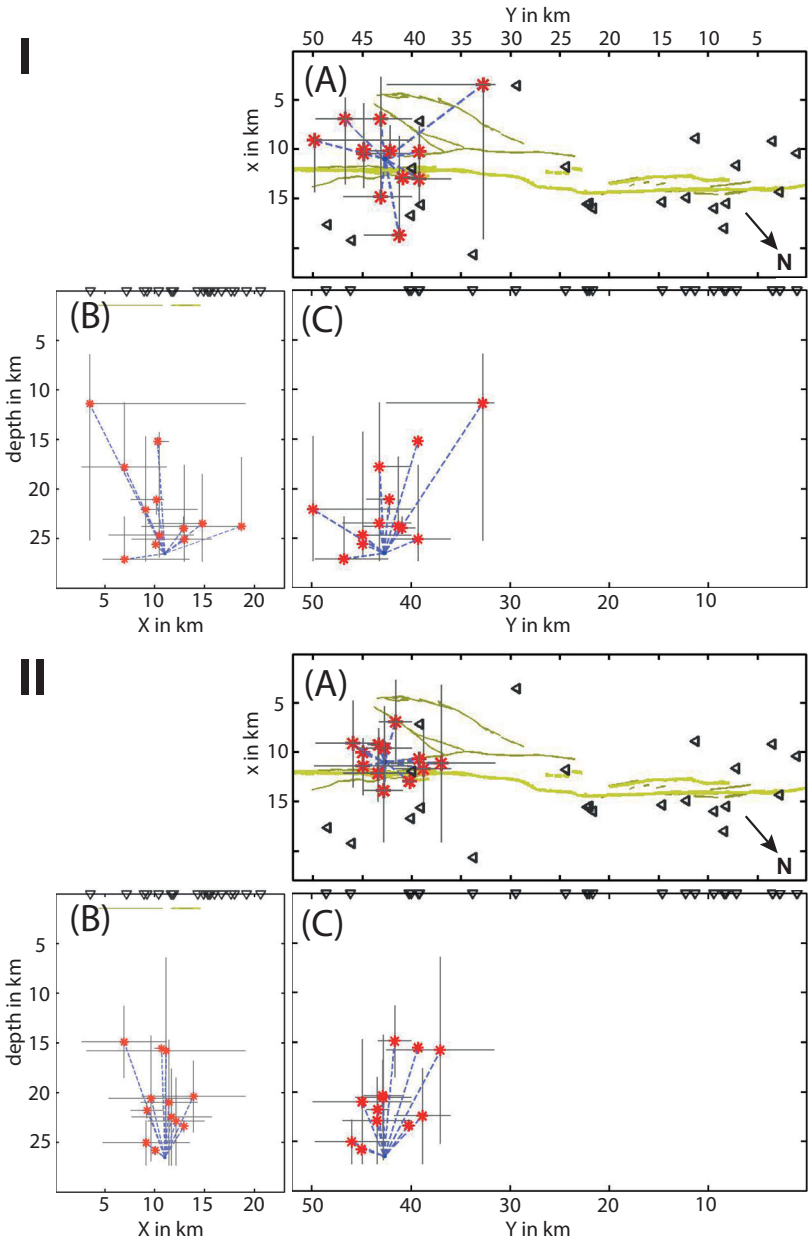

Figure 11. Hypocentral locations of 12 LFE family members determined by (I) the maximum cross-correlation value and (II) the median of the error range. (A) map view, (B) *X*–*Z* plane and (C) *Y*–*Z* plane. Red stars indicate source locations, black lines show the estimated error range and black triangles indicate the station locations. Number of stations used for the TRI location calculation may differ for each LFE depending on data quality. Grey dashed lines connect individual LFE locations to the stacked LFE locations of Shelly & Hardebeck (2010) for reference.

parameters on location errors is discussed in detail in the appendix (Appendix C).

Though not directly evident through any individual parameter, other studies suggest that the most influential factor affecting the location quality are the maximum amplitude of the seismograms and the velocity model resolution (including the assumption of a Poisson solid for determining *S*-wave speeds) (e.g. Larmat *et al.* 2010; Kremers *et al.* 2011). If so, an improved *S*-wave velocity model with high resolution may make location of individual LFEs possible so that the temporal progression within a tremor episode can be observed without having to apply a taper to individual *S*-wave arrivals. The subject of future work will be to see if with improved data, the modified TRI method may be able to locate (high amplitude) tremor using the rebroadcasted waveforms of the entire episode.

6 CONCLUSIONS

We present a modified TRI algorithm as a new method for locating LFEs within tremor episodes. The modified TRI method searches for phase coherence over a short time period between back-propagated wavefields. Thus, the advantage of the modified TRI method is the ability to include temporal information, making the method robust for locating low-amplitude signals using a sparse station configuration. Moreover, in theory, the method does

not require precise phase arrival times if the signal-to-noise ratio and velocity model resolution are high.

We successfully locate an M_L 1.4 earthquake using data filtered between 0.5–2 Hz and 1–5 Hz. However, the highest median cross-correlation coefficient value was slightly lower for the 1–5 Hz frequency band. The location results are significantly improved by tapering around the *S*-phase pulse with a 3 s long cosine taper. We also locate 34 LFEs with average location errors of 5 km in the *X*-direction, 2.6 km in the *Y*-direction and 4.8 km in depth. By comparison, the method of Wech & Creager (2008) was used by Horstmann *et al.* (2013) to estimate the location of a tremor episode using the same data set and found a horizontal uncertainty of 30 km. Other methods that locate individual LFEs of similar amplitudes report errors on the order of 10 km or more (Payero *et al.* 2008). The calculated source positions are in good agreement with stacked LFE family source positions determined by Shelly & Hardebeck (2010). We also locate 12 individual LFEs from an individual family, calculating source locations that cluster around the family stack location, however, with considerable scatter. With the present velocity model and station configuration, we are not yet able to resolve how much of the scatter can be attributed exclusively to variation in individual LFE locations.

We present two methods for calculating the source position: the maximum median cross-correlation coefficient and the median of the error range. The tight clustering of the median error range suggest that it is the preferred method for the particular station configuration and velocity model used here, however, source locations near the model boundaries may be subject to a bias. As for the location uncertainties, we find that the uncertainty in the location is only weakly influenced by the maximum gap in azimuthal coverage and closest station distance. The number of stations and the median signal-to-noise ratio do not show a significant influence on the location results. Also, the highest median cross-correlation coefficient does not correlate with how well a source location is determined. However, we find a clear relationship between the number of stations and the highest median cross-correlation value. The highest median cross-correlation value decreases with an increasing number of stations, likely due to the decrease in the degrees of freedom compounded with inaccuracies in the velocity model.

In summary, the modified TRI method successfully locates individual low-amplitude LFEs with higher precision compared to other methods that locate entire tremor episodes or individual LFEs. The location errors are higher than those reported for methods that use stacks of hundreds of LFEs to determine a source location, as expected. The rebroadcasting of individual recorded seismograms is computational demanding and requires a high-resolution velocity model. However, the modified TRI method is a powerful technique to provide event locations in regions where repeating LFEs have not been detected and thus stacking is not possible.

ACKNOWLEDGEMENTS

This work was supported by funding from the 'Concept of the Future' of the Karlsruhe Institute of Technology (KIT) within the framework of the national 'Initiative for Excellence'. Additional support came from the Southern California Earthquake Center. SCEC is funded by NSF Cooperative Agreement EAR-0529922 and U.S. Geological Survey Cooperative Agreement 07HQAG0008. The SCEC contribution number for this paper is 1639. The authors would like to thank David R. Shelly for the use of his LFE event catalogue, which was crucial to the analysis, as well as Thomas Bohlen for the use of his finite-difference code. The Karlsruhe Broadband Array (KABBA) provided seismic data from the temporary station deployment, and the permanent seismic data are distributed by the Northern California Earthquake Data Center (NCEDC). The HRSN data as well as the BDSN data are contributed by the Berkeley Seismological Laboratory, University of California, Berkeley. The ANSS Worldwide Earthquake Catalog is provided by Advanced National Seismic System (ANSS). The authors would like to thank the landowners for permission to install seismic stations on their property. In addition, we would like to thank Werner Scherer, Kayla Kroll, Tien-Hui Wang, and Peter Duffner for their help during station services and deployment. Any use of trade, firm or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. We would like to thank Honn Kao and one other anonymous reviewer for constructive comments that significantly improved the manuscript.

REFERENCES

Anderson, B.E., Griffa, M., Ulrich, T.J. & Johnson, P.A., 2011. Time reversal reconstruction of finite sized sources in elastic media, *J. acoust. Soc. Am.,* **130**(4), EL219–EL225.

Beroza, G. & Ide, S., 2011. Slow earthquakes and nonvolcanic tremor, *Annu. Rev. Earth Planet. Sci.,* **39,** 271–296.

Bohlen, T., 2002. Parallel 3-D viscoelastic finite difference seismic modelling, *Comput. Geosci.,* **28**(8), 887–899.

Brocher, T.M., 2005. Empirical relations between elastic wavespeeds and density in the Earth's crust, *Bull. seism. Soc. Am.,* **95**(6), 2081–2092.

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

Brown *et al.*, 2009. Deep low-frequency earthquakes in tremor localize to the plate interface in multiple subduction zones, *Geophys. Res. Lett.,* **36**(19), L19306, doi:10.1029/2009GL040027.

Courant, R., 1967. Friedrichs, K. & Lewy, H., On the partial difference equations of mathematical physics, *IBM J. Res. Dev.,* **11**(2), 215–234.

Dougherty, M.E. & Stephen, R.A., 1988. Seismic energy partitioning and scattering in laterally heterogeneous ocean crust, *Pure Appl. Geophys.,* **128**(1–2), 195–229.

Haney, M.M., 2014. Backprojection of volcanic tremor, *Geophys. Res. Lett.,* **41**(6), 1923–1928.

Horstmann, T., Harrington, R.M. & Cochran, E.S., 2013. Semiautomated tremor detection using a combined cross-correlation and neural network approach, *J. geophys. Res.,* **118**(9), 4827–4846.

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

Kim, M., Schwartz, S. & Bannister, S., 2011. Non-volcanic tremor associated with the March 2010 Gisborne slow slip event at the Hikurangi subduction margin, New Zealand, *Geophys. Res. Lett.,* **38**(14), L14301, doi:10.1029/2011GL048400.

Komatitsch, D. & Martin, R., 2007. An unsplit convolutional perfectly matched layer improved at grazing incidence for the seismic wave equation, *Geophysics,* **72**(5), SM155–SM167.

Kremers, S., Fichtner, A., Brietzke, G.B., Larmat, C.S., Huang, L. & Kaeser, M., 2011. Exploring the potentials and limitations of the time-reversal imaging of finite seismic sources, *Solid Earth,* **2,** 95–105.

Larmat, C., J.-P. Montagner Fink, M., Capdeville, Y., Tourin, A. & Clévédé, E., 2006. Time-reversal imaging of seismic sources and application to the great sumatra earthquake, *Geophys. Res. Lett.,* **33,** doi: 10.1029/2006GL026336.

Larmat, C.S., Guyer, R.A. & Johnson, P.A., 2009. Tremor source location using time reversal: selecting the appropriate imaging field, *Geophys. Res. Lett.,* **36**(22), doi:10.1029/2009GL040099.

Larmat, C.S., Guyer, R.A. & Johnson, P.A., 2010. Time-reversal methods in geophysics, *Phys. Today,* **63**(8), 31–35.

La Rocca, M., Creager, K.C., Galluzzo, D., Malone, S., Vidale, J.E., Sweet, J.R. & Wech, A.G., 2009. Cascadia tremor located near plate interface constrained by *S* minus *P* wave times, *Science,* **323**(5914), 620–623.

Lokmer, I., O'Brien, G.S., Stich, D. & Bean, C.J., 2009. Time reversal imaging of synthetic volcanic tremor sources, *Geophys. Res. Lett.,* **36**(12), L12308, doi:10.1029/2009GL038178.

McCausland, W., Malone, S. & Johnson, D., 2005. Temporal and spatial occurrence of deep non-volcanic tremor: from Washington to northern California, *Geophys. Res. Lett.,* **32**(24), L24311, doi:10.1029/2005GL024349.

McClement, K., Thurber, C., Shelly, D.R., Sumy, D., Bennington, N., Peterson, D., Cochran, E.S. & Harrington, R.M., 2013. Extending seismic tomography along the San Andreas fault to the lower crust with low frequency earthquakes, *Eos Trans. AGU, Fall Meet. Suppl., Abstract,* (T41A-2561).

Miyazawa, M. & Brodsky, E.E., 2008. Deep low-frequency tremor that correlates with passing surface waves, *J. geophys. Res.,* **113**(B1), B01307, doi:10.1029/2006JB004890.

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

Obara, K., Hirose, H., Yamamizu, F. & Kasahara, K., 2004. Episodic slow slip events accompanied by non-volcanic tremors in southwest Japan subduction zone, *Geophys. Res. Lett.,* **31**(23), L23602, doi:10.1029/2004GL020848.

O'Brien, G.S., Lokmer, I., L. De Barros Bean, C.J., Saccorotti, G., Metaxian, J. & Patane, D., 2011. Time reverse location of seismic long-period events recorded on Mt. Etna, *Geophys. J. Int.,* **184,** 452–462.

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

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

Schwartz, S. & Rokosky, J., 2007. Slow slip events and seismic tremor at circum-Pacific subduction zones, *Rev. Geophys.,* **45**(3), doi:10.1029/2006RG000208.

Shelly, D. & Hardebeck, J., 2010. Precise tremor source locations and amplitude variations along the lower-crustal central San Andreas fault, *Geophys. Res. Lett.,* **37**(14), L14301, doi:10.1029/2010GL043672.

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

Thurber, C., Zhang, H., Waldhauser, F., Hardebeck, J., Michael, A. & Eberhart-Phillips, D., 2006. Three-dimensional compressional wavespeed model, earthquake relocations, & focal mechanisms for the Parkfield, California, region, *Bull. seism. Soc. Am.,* **96**(4B), S38–S49.

Wech, A. & Creager, K., 2007. Cascadia tremor polarization evidence for plate interface slip, *Geophys. Res. Lett.,* **34**(22), L22306, doi:10.1029/2007GL031167.

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

Zeng, X., Thurber, C., Shelly, D.R., Bennington, N., Cochran, E.S. & Harrington, R.M., 2014. 3D P and S wave velocity structure and tremor locations in the Parkfield region, *Eos Trans. AGU, Fall Meet. Suppl., Abstract,* (S21B-0565).

Zhang, H., Thurber, C. & Bedrosian, P., 2009. Joint inversion for v_p, v_s, and v_p/v_s at SAFOD, Parkfield, California, *Geochem. Geophys. Geosyst.,* **10**(11), doi:10.1029/2009GC002709.

APPENDIX A: AUXILIARY FIGURES

Fig. A1 shows the curl field functions at the calculated source location of the M_L 1.4 earthquake, i.e. the grid point with the highest cross-correlation value. The curl field from the tapered *S*-phases pulses are partially shifted in time on select stations, potentially due to noise bursts and/or local inaccuracies in the velocity model.

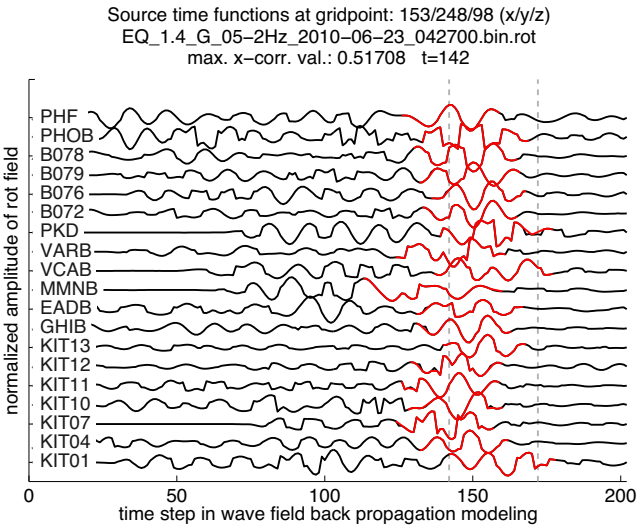

Figure A1. Relative shear energy amplitudes give an approximation of the curl field as a function of time at gridpoint $X = 15.3$ km, $Y = 24.8$ km and $Z = 9.8$ km for the seismograms shown in Fig. 4(B) (see the text). The grey dashed lines denote the time window containing the maximum median cross-correlation coefficient. Red indicates position of the rebroadcasted tapered *S*-phase relative to the time window. Note the shifted positions of the *S*-phase pulses, suggesting inaccuracies within the velocity model. Apparent discontinuities in the function are a result of writing out the field at 50 ms intervals (each 10th sample).

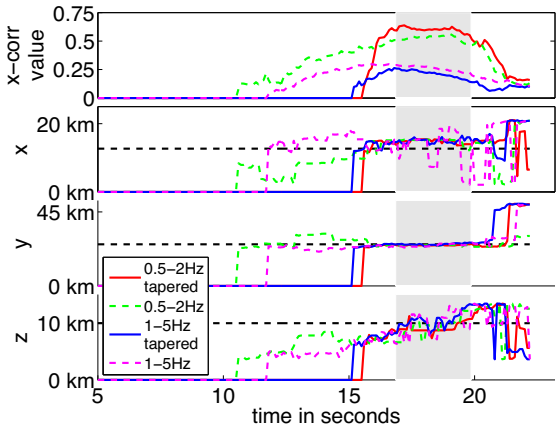

Figure A2. Temporal progression of the curl field cross-correlation values (top) and the corresponding implied source location (bottom three panels) of the M_L 1.4 earthquake. Coloured lines show values for the frequency bands indicated in the legend that were used for the locations, and solid and dashed lines indicate tapered versus untapered data. The upper panel denotes the maximum median cross-correlation value and the lower three panels show the position of the particular maximum within each time step. Black dashed lines and the grey box indicate the location of McClement *et al.* (2013) for comparison. The middle two panels indicate a stable horizontal position of the maximum correlation value in time.

Fig. A2 illustrates the stability of the location solutions by showing the temporal progression of the *X*, *Y*, *Z*, and curl field cross-correlation values around the time of highest median cross-correlation value.

Table B1. Seismic stations' locations used in the study. The PERMIT stations (KIT station names) recorded data continuously from May 2010 to July 2011. Station PKD is from the Berkeley Digital Seismic Network (BDSN), stations PHOB and PHF are from the Northern California Seismic Network (NCSN), stations GHIB, EADB, FROB, VCAB, VARB, MMNB and JCSB are from the Berkeley High Resolution Station Network (HRSN), and stations B072, B076, B078, B079 and B901 are from the Plate Boundary Observatory (PBO).

Station	Latitude (°N)	Longitude (°W)	Elevation (m)
KIT01	35.6663	−120.2925	511.1
KIT04	35.7943	−120.3611	374.9
KIT07	35.664	−120.1374	389.3
KIT10	35.6889	−120.2472	605
KIT11	35.7164	−120.2226	618.4
KIT12	35.7823	−120.2187	579.6
KIT13	35.7119	−120.3937	448.1
GHIB	35.8322	−120.3473	330
FROB	35.9109	−120.4869	231
EADB	35.8952	−120.4226	224
VCAB	35.9216	−120.5339	555
VARB	35.9261	−120.4471	177
JCSB	35.9212	−120.434	299
MMNB	35.9565	−120.496	480
PKD	35.9452	−120.5416	583
B072	35.831	−120.345	397.7
B076	35.9398	−120.4248	445
B078	35.8377	−120.3452	386.8
B079	35.7157	−120.2057	436.6
B901	35.6897	−120.142	275.3
PHF	35.8816	−120.4016	457
PHOB	35.8666	−120.4796	796

APPENDIX B: LFE LOCATIONS DETERMINED USING THE MEDIAN ERROR RANGE

The seismic stations locations used in the study are included in Table B1. The LFE source locations determined using the median error range are summarized in Table B2.

APPENDIX C: CORRELATION OF INDIVIDUAL PARAMETERS ON LOCATION RESULTS

Here we investigate if any specific parameter has a greater influence on the location quality. Fig. C1 examines how the location quality is affected by parameters such as the median signal-to-noise ratio of the LFE seismograms, the number of stations used, the maximum azimuthal gap in station coverage and the distance to the nearest station. The parameter values for each LFE location calculation are given in Table C1.

Figs C1(A)–(D) show different parameters over the range of horizontal uncertainties and their correlation with the uncertainties. The coloured stars denote the values for each individual LFE location result. The grey lines indicate the best fit and the red number in each panel gives the corresponding coefficient of determination (the R^2 value), which is a measure of the correlation. The R^2 value is based on the residuals to the best-fit regression line. An R^2 value of 1 reflects a purely linear relationship whereas a value of 0 indicates a horizontal line as best fit (hence, low correlation).

Panel A indicates a very weak relationship between the azimuthal gap and the location errors, as the points scatter around the trend of the line. Not surprisingly, the horizontal uncertainty increases with an increasing azimuthal gap. Panel B shows the influence of source–station distance. The R^2 value of 0.080 indicates a weak trend of horizontal uncertainty decreasing with source–receiver distance. Panel C suggests that the median signal-to-noise ratio of the LFE seismograms seems to have little effect on the location quality (see the R^2 value of 0.011). The reason may be that tapering of the *S*-phase removes all noise outside the tapering window. Nevertheless, one would still expect only a weak relationship between location error and signal-to-noise ratio, as a strength of the modified TRI method is the capability of identifying phase coherence within the superposed wavefield, even for small amplitude signals. In addition, panel D suggests that the number of stations used does not affect the horizontal error range. With the result indicated by panels A, B and D, one could argue that the distribution of the stations is more important than the number of stations.

In addition to testing the influence of individual parameters, we also want to determine if the maximum median cross-correlation coefficient is indicative of the location quality. Fig. C1(E) shows the maximum median cross-correlation coefficient versus the horizontal uncertainty. The majority of the source locations with different horizontal uncertainty have a maximum median cross-correlation value around 0.4, resulting in a low R^2 value of 0.03. Thus, the maximum median cross-correlation coefficient does not seem to be an adequate measure of location quality. However, the cross-correlation value also correlates with the number of stations used. Fig. C1(F) shows a clear relationship between the number of stations used and the maximum median cross-correlation coefficient, supported by the R^2 value of 0.487. The most likely reason for the correlation is that the number of degrees of freedom decreases with an increasing number of stations, making it more difficult for the algorithm to find a gridpoint that optimizes the coherence between all rebroadcasted wavefields.

The low R^2 values suggest that no one single parameter heavily influences the quality of the location results. We also check whether a combination of parameters (P_s) influences the LFE locations results. The value P_s is given as: $P_s = w_1 * P_1 + w_2 * P_2 + ...w_N * P_N$, where P denotes the individual parameters and w individual weighting factors between 0 and 1. The parameters are normalized in the formulation to values between 0 and 1, and parameters with a negative expected trend, such as the maximum gap in azimuthal coverage, are calculated as $1 - P$. We perform a grid search for a combination that optimizes the R^2 value. The best result provides weighting factors of 0.3 for the signal-to-noise ratio, 0.9 for the azimuthal gap, 0.5 for the number of stations used and 1 for the minimum source–station distance. However, the resulting R^2 value of 0.098 is on the same order of the R^2 values obtained for the maximum gap in azimuthal coverage and nearest station distance as single parameters. Hence, there is no evidence to suggest that a combination of the individual parameters controls the quality of the location result. We therefore conclude that the most likely factors affecting the location quality are the maximum amplitude of the *S*-phase within the taper window and the resolution of the velocity model.

Table B2. LFE locations calculated using the median error extent in local grid coordinates (X, Y) and latitude and longitude values. The last two columns indicate horizontal and vertical differences to the stacked LFE locations (Shelly & Hardebeck 2010).

	Start of time window	X (km)	Y (km)	Z (km)	Latitude (°N)	Longitude (°W)	Hor. diff. (km)	Vert. diff. (km)
1	2010-07-01 17:42:50	13.75 ± 5.15	38.10 ± 2.4	23.00 ± 4	35.713	−120.245	2.13	−3.50
2	2010-07-01 17:42:50	13.75 ± 5.15	38.10 ± 2.4	23.00 ± 4	35.713	−120.245	2.13	−3.50
3	2010-07-02 04:42:45	12.95 ± 7.45	26.15 ± 1.55	24.10 ± 2.9	35.789	−120.339	0.15	−2.40
4	2010-07-02 13:59:20	6.20 ± 0.6	36.50 ± 0.7	24.75 ± 1.55	35.679	−120.320	5.83	1.00
5	2010-07-03 08:21:15	15.25 ± 3.55	39.00 ± 1.3	20.15 ± 1.75	35.715	−120.226	6.22	−4.10
6	2010-07-04 11:08:35	11.30 ± 4.5	33.45 ± 1.45	24.15 ± 1.95	35.730	−120.300	0.42	−0.60
7	2010-08-04 10:32:05	9.85 ± 2.55	42.25 ± 2.35	21.95 ± 4.75	35.662	−120.248	2.21	−3.55
8	2010-08-04 10:40:00	10.05 ± 0.05	45.00 ± 0.2	25.80 ± 0.2	35.644	−120.227	2.48	−0.70
9	2010-08-06 00:35:25	12.55 ± 1.85	44.10 ± 1.3	20.05 ± 2.35	35.665	−120.212	4.34	−6.45
10	2010-08-06 00:35:45	9.15 ± 5.95	47.05 ± 2.85	23.05 ± 4.25	35.625	−120.220	2.92	−2.95
11	2010-09-02 10:31:40	9.15 ± 4.35	46.00 ± 3.7	25.05 ± 2.25	35.632	−120.227	3.78	−1.45
12	2010-09-02 17:32:40	8.60 ± 0.8	37.20 ± 0.3	24.10 ± 2	35.689	−120.295	1.88	1.60
13	2010-09-02 19:27:40	12.00 ± 4	33.75 ± 1.05	20.85 ± 3.75	35.732	−120.292	0.37	−3.90
14	2010-09-10 02:29:05	12.95 ± 0.05	40.25 ± 0.65	23.40 ± 0.6	35.693	−120.237	3.13	−3.10
15	2010-10-09 08:52:50	13.90 ± 5.2	42.85 ± 1.95	20.40 ± 3.6	35.681	−120.210	2.87	−6.10
16	2010-10-12 12:51:45	13.85 ± 0.35	28.40 ± 0.1	22.60 ± 0.2	35.779	−120.315	2.74	−1.65
17	2010-10-12 21:07:45	11.15 ± 7.95	37.05 ± 5.45	15.80 ± 9.4	35.705	−120.275	5.68	−10.70
18	2010-10-12 21:07:50	10.70 ± 0.7	39.30 ± 0.1	15.55 ± 0.35	35.687	−120.262	3.44	−10.95
19	2010-10-22 14:31:30	19.15 ± 1.65	23.30 ± 0.4	22.75 ± 1.65	35.845	−120.307	6.81	−3.75
20	2010-10-26 12:17:15	8.85 ± 5.25	36.65 ± 1.25	16.50 ± 2.9	35.694	−120.297	1.70	−6.00
21	2010-11-07 02:45:45	8.85 ± 2.55	39.65 ± 3.85	17.20 ± 5.6	35.673	−120.275	9.10	−9.30
22	2010-11-16 08:54:10	12.15 ± 2.85	43.45 ± 3.45	22.90 ± 4.4	35.667	−120.220	1.33	−3.60
23	2010-11-17 10:51:25	17.50 ± 1.5	33.30 ± 1.3	25.15 ± 2.05	35.767	−120.249	5.96	1.40
24	2010-11-17 10:57:55	8.95 ± 0.65	39.95 ± 1.55	24.45 ± 2.05	35.672	−120.272	3.07	−1.05
25	2010-11-24 11:26:15	15.60 ± 1.5	16.50 ± 0.4	9.85 ± 2.45	35.871	−120.386	3.00	−15.90
26	2010-12-01 12:36:10	9.65 ± 4.25	42.80 ± 2.8	20.60 ± 6.3	35.657	−120.246	1.38	−5.90
27	2010-12-01 12:39:15	11.70 ± 4	38.85 ± 2.85	22.45 ± 4.85	35.696	−120.257	3.93	−4.05
28	2010-12-28 14:46:35	12.30 ± 1.1	40.90 ± 0.5	23.70 ± 2.9	35.685	−120.237	2.78	0.45
29	2010-12-28 22:27:55	9.35 ± 1.25	42.55 ± 2.25	17.75 ± 5.05	35.657	−120.250	1.45	−5.50
30	2011-01-03 03:45:00	5.35 ± 3.05	15.65 ± 1.05	12.45 ± 7.75	35.816	−120.478	5.63	−8.55
31	2011-01-23 19:41:20	9.25 ± 1.65	43.35 ± 1.15	21.80 ± 0.8	35.651	−120.245	1.89	−4.70
32	2011-02-03 01:42:45	9.40 ± 2.5	48.45 ± 1.35	25.10 ± 1.8	35.617	−120.207	2.19	−1.40
33	2011-03-12 07:41:25	11.45 ± 2.85	44.95 ± 4.95	21.00 ± 6.3	35.653	−120.215	2.27	−5.50
34	2011-04-04 21:05:55	6.95 ± 4.25	41.65 ± 1.65	14.90 ± 3.6	35.649	−120.277	4.22	−11.60

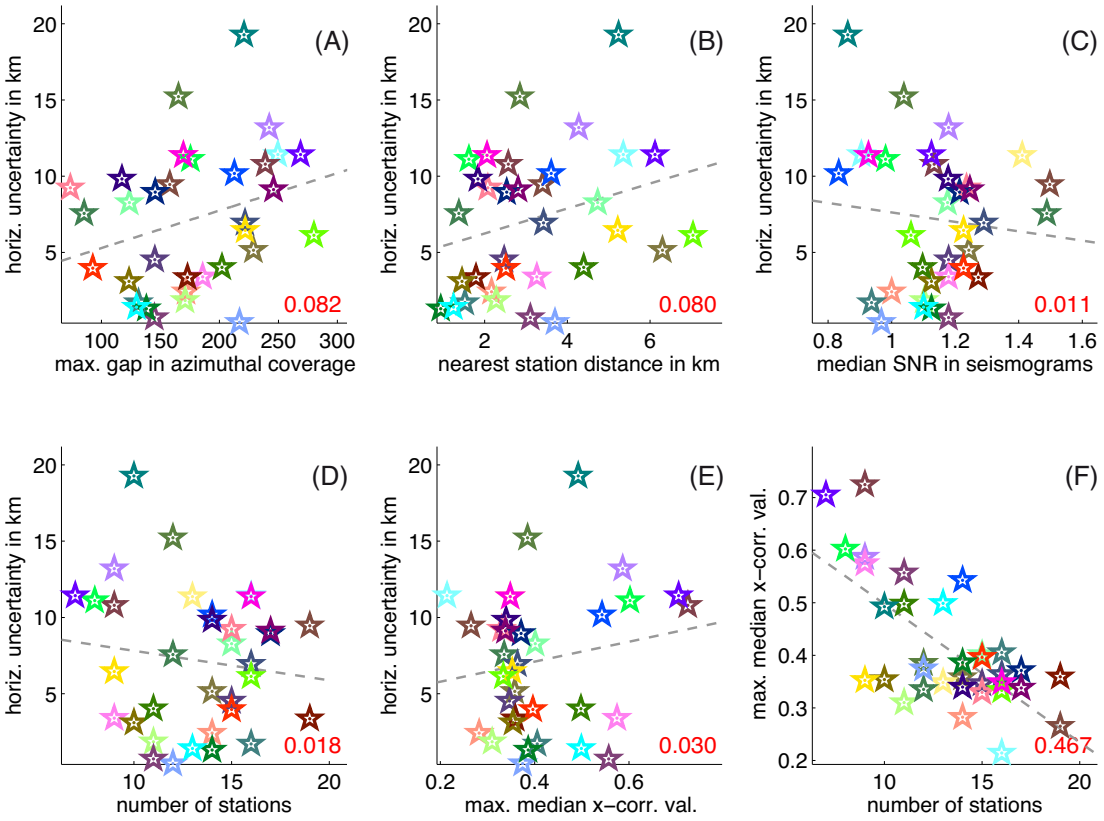

Figure C1. Correlation between location results and individual parameters such as the maximum gap in azimuthal coverage, nearest station distance, median signal-to-noise ratio in the seismograms, number of stations used and the maximum median cross-correlation coefficient in the model. (Individual parameter values shown in Table C1). Horizontal uncertainty provides a proxy for location quality. Stars indicate the values determined in each of the localization of the individual LFEs shown in Fig. 10(I). Grey dashed line displays the best-fit line, and the red number is the corresponding coefficient of determination (R^2 value).

Table C1. Individual parameter values such as number of stations used, median signal-to-noise ratio in the LFE seismograms, maximum azimuthal gap in station coverage, distance from LFE source to the nearest station and the maximum median cross-correlation coefficient for each LFE.

	Start of time window	No. of stat.	SNR	Azimuth. gap (°)	Nearest station (km)	Max. *X*-corr. coeff.
1	2010-07-01 17:42:50	13	1.41	110.73	1.5	0.35
2	2010-07-01 17:42:50	16	0.93	110.73	1.5	0.35
3	2010-07-02 04:42:45	12	0.98	133.67	1.5	0.39
4	2010-07-02 13:59:20	11	1.07	179.25	2.1	0.31
5	2010-07-03 08:21:15	12	1.50	103.50	0.3	0.34
6	2010-07-04 11:08:35	19	1.50	120.46	4.6	0.27
7	2010-08-04 10:32:05	16	1.29	189.56	2.2	0.36
8	2010-08-04 10:40:00	12	0.99	218.50	3.8	0.37
9	2010-08-06 00:35:25	15	1.18	180.30	2.9	0.35
10	2010-08-06 00:35:45	9	1.18	245.03	5.7	0.59
11	2010-09-02 10:31:40	16	0.88	234.64	4.7	0.21
12	2010-09-02 17:32:40	16	0.93	112.08	1.8	0.41
13	2010-09-02 19:27:40	15	1.18	125.39	4.4	0.40
14	2010-09-10 02:29:05	14	1.13	129.50	0.7	0.39
15	2010-10-09 08:52:50	8	0.98	151.60	2.7	0.60
16	2010-10-12 12:51:45	11	1.24	145.44	3.1	0.56
17	2010-10-12 21:07:45	10	0.86	113.85	2.2	0.49
18	2010-10-12 21:07:50	13	1.10	128.01	1.0	0.50
19	2010-10-22 14:31:30	9	1.24	176.38	2.5	0.57
20	2010-10-26 12:17:15	9	1.14	147.75	2.2	0.72
21	2010-11-07 02:45:45	15	1.28	148.72	1.2	0.33
22	2010-11-16 08:54:10	17	1.21	177.46	2.4	0.37
23	2010-11-17 10:51:25	15	1.23	97.62	2.2	0.40
24	2010-11-17 10:57:55	19	1.27	157.40	1.4	0.36
25	2010-11-24 11:26:15	10	1.13	156.73	1.3	0.35
26	2010-12-01 12:36:10	14	0.83	199.59	2.6	0.54
27	2010-12-01 12:39:15	14	1.18	116.67	2.8	0.34
28	2010-12-28 14:46:35	14	1.00	178.42	2.7	0.28
29	2010-12-28 22:27:55	14	1.24	211.11	5.0	0.36
30	2011-01-03 03:45:00	9	1.23	201.84	4.0	0.35
31	2011-01-23 19:41:20	11	1.10	215.98	5.4	0.50
32	2011-02-03 01:42:45	16	1.06	268.57	5.8	0.33
33	2011-03-12 07:41:25	7	1.13	203.09	5.0	0.71
34	2011-04-04 21:05:55	17	1.25	240.57	1.7	0.34