

An algorithm for automated identification of fault zone trapped waves

Z.E. Ross and Y. Ben-Zion

Department of Earth Sciences, University of Southern California, Los Angeles, CA 90089-0740, USA. E-mail: zross@usc.edu

Accepted 2015 May 11. Received 2015 May 8; in original form 2014 December 11

SUMMARY

We develop an algorithm for automatic identification of fault zone trapped waves in data recorded by seismic fault zone arrays. Automatic *S* picks are used to identify time windows in the seismograms for subsequent search for trapped waves. The algorithm calculates five features in each seismogram recorded by each station: predominant period, 1 s duration energy (representative of trapped waves), relative peak strength, arrival delay and 6 s duration energy (representative of the entire seismogram). These features are used collectively to identify stations in the array with seismograms that are statistical outliers. Applying the algorithm to large data sets allows for distinguishing genuine trapped waves from occasional localized site amplification in seismograms of other stations. The method is verified on a test data set recorded across the rupture zone of the 1992 Landers earthquake, for which trapped waves were previously identified manually, and is then applied to a larger data set with several thousand events recorded across the San Jacinto fault zone. The developed technique provides an important tool for systematic objective processing of large seismic waveform data sets recorded near fault zones.

Key words: Time-series analysis; Guided waves; Site effects; Wave propagation; Fractures and faults.

1 INTRODUCTION

Fault zone structures can produce, in addition to the usual *P* and *S* body waves, head waves that propagate along bimaterial interfaces and trapped waves that are associated with resonance modes in a low velocity fault zone waveguide (e.g. Ben-Zion & Aki 1990). Such waves have been observed along subduction zones (e.g. Fukao *et al.* 1983; Shapiro *et al.* 2000; Hong & Kennett 2003), large strike-slip faults (e.g. Li *et al.* 1990; Ben-Zion *et al.* 2003; McGuire & Ben-Zion 2005) and normal faults (e.g. Rovelli *et al.* 2002; Calderoni *et al.* 2012; Avallone *et al.* 2014). The observation and modelling of head and trapped waves provide high-resolution information on the internal structure of fault zones, motivating deployments of dense seismic arrays across faults (e.g. Li *et al.* 1994; Mamada *et al.* 2004; Ozakin *et al.* 2012).

A growing number of dense deployments near faults are recording vast amounts of data (e.g. Bulut *et al.* 2009; Kurzon *et al.* 2014; Ben-Zion *et al.* 2015). Systematic analyses of such data require automatic algorithms for detecting various phases. These types of algorithms can sift through large data sets and add objectivity to the results. Ross & Ben-Zion (2014) developed an automatic algorithm for detecting and phase picking of *P* and *S* body waves and fault zone head waves. In the present paper we introduce an algorithm for automatic identification of fault zone trapped waves (FZTW). The method searches for a set of features characteristic of FZTW in

time windows identified by the *S*-wave picking algorithm of Ross & Ben-Zion (2014). These features are examined in seismograms generated by many events and recorded at near-fault stations. Frequent occurrence of these features at given stations are identified as reflecting candidate FZTW. The technique is independent of array geometry and station position, and it assumes no prior knowledge about which stations (if any) are located inside low velocity fault zone layers. The algorithm is verified and demonstrated using data recorded by two dense linear arrays that cross large fault zones in southern California.

2 DATA AND PREPROCESSING

The data used in this study were recorded by two linear arrays deployed across the surface rupture of the 1992 Landers earthquake in the Eastern California Shear Zone and across the San Jacinto Fault Zone (SJFZ). The examined data from the Landers array were recorded during 1992 October 14–17, on 22 three-component short-period seismometers (Fig. 1). The instruments were spaced 25 m apart within 200 m of the surface rupture and 50–100 m further away. A total of 207 events recorded by the Landers array in that period and located by Peng *et al.* (2003) are considered in the automatic detection analysis performed below. Their analysis indicated that ~6–7 of the stations were located inside the fault damage zone

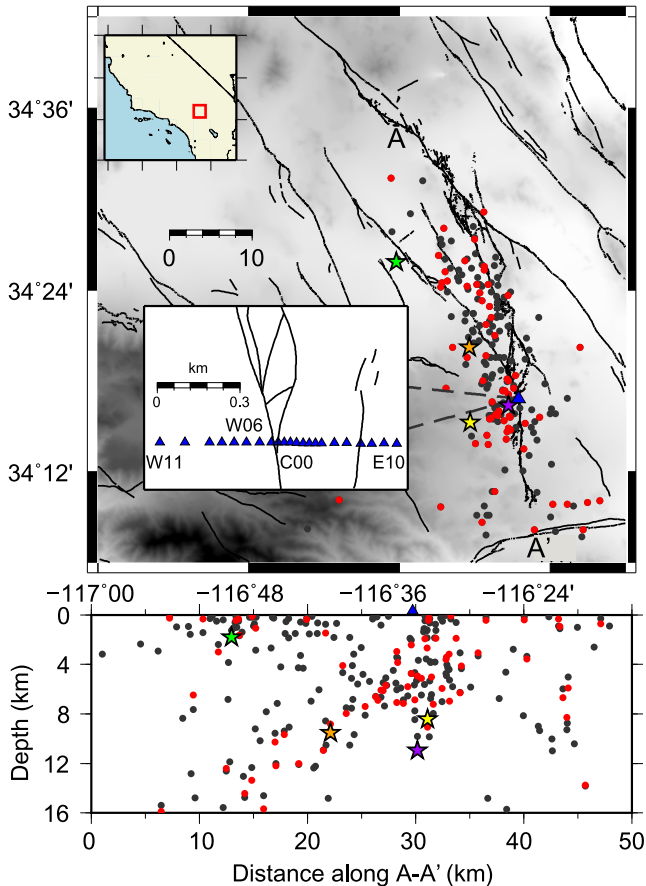


Figure 1. Location map for data (circles and stars) recorded during 1992 October 14–17, by a 22-station linear array (blue triangle and inset) across the rupture zone of the 1992 Lander's M 7.3 earthquake. The bottom plot shows projection on a vertical cross-section along the line A–A'. The locations of the 207 earthquakes shown are from Peng *et al.* (2003). Data generated by these events were analysed for trapped waves. Events with at least one detection at the array are coloured red. A total of 70 events were flagged by the algorithm. The purple, green, yellow and orange stars indicate location of events used in Figs 3, 4, 6 and 9, respectively. The shading indicates the regional topography, and black lines denote faults.

(stations C00–E06). The examined data from the SJFZ were recorded at Jackass Flats (JF) during 2013 January 1 to 2013 December 31 by nine three-component broad-band seismometers across the surface expression of the Clark fault (Fig. 2). This array has instrument spacing of 25–50 m and the events were detected and located by the ANZA network (eqinfo.ucsd.edu). Results from this study, which are described in detail subsequently, indicate that stations JF00–JFS3 are likely to be inside or on the edge of the fault damage zone. A total of 5203 earthquakes located within 100 km of the JF array are considered in the analysis below.

Basic pre-processing was performed on the data for both arrays. The mean and trend were removed and the horizontal components of motion were rotated to a fault-parallel, fault-normal coordinate system. This is because typical Love-type trapped waves energy exists preferentially in a component of motion oriented parallel to the fault plane (e.g. Ben-Zion & Aki 1990; Peng *et al.* 2003; Lewis & Ben-Zion 2010). For the Lander's array, the data are used in a raw unfiltered form to demonstrate the general applicability of the method. For the JF array data set, which is over $20\times$ larger, the instrument response was removed and a Butterworth bandpass filter was applied between 2 and 20 Hz to eliminate instrument noise.

3 METHODOLOGY

Fig. 3(a) shows an example set of waveforms recorded by the Lander's array containing FZTW at stations C00 through E04. The seismograms containing FZTW (black ellipse) have longer period, higher amplitude phases arriving shortly after the S -wave that are about half a second in duration. These features are common among FZTW, but the waveform shape, arrival time, and other characteristics can vary depending on the properties of the trapping structure (e.g. Ben-Zion 1998; Jahnke *et al.* 2002). Our methodology for automatic detection of FZTW can be generally stated as a procedure for detecting outliers of various waveform features across a fault zone array that indicate common aspects of FZTW.

As shown by Ben-Zion & Aki (1990) and Lewis & Ben-Zion (2010), FZTW start from the end of the S body wave, or slightly behind in structures with overall velocity contrast across the fault, so a reliable estimate of the S arrival is crucial. We therefore first run the S -wave picking algorithm of Ross & Ben-Zion (2014) on seismograms at each station of a given array for each event. The picking method first employs polarization analysis to remove P -wave energy from the seismograms. Then STA/LTA and kurtosis detectors are run in tandem to lock on a well-defined S -arrival. Picks are made on both horizontal components (N and E), if possible, and the median value is calculated over all picks and components to get a single, robust estimate of the S -arrival at the array. This provides a good approximation for relatively short arrays (length <1 km) of the type analysed here. For longer arrays, it may be desirable to use individual picks at each station instead due to a potential moveout across the array. For the Lander's and JF arrays, the method has been tested and works both ways, but using individual picks is in general less reliable since the picks vary in accuracy. After the picks are made, the traces are restored to the original raw form for detecting trapped waves, since the polarization process used during picking is unnecessary for FZTW detection. The algorithm of Ross & Ben-Zion (2014) requires a minimum signal-to-noise ratio (SNR) of 5 to make a valid S pick. We require further that more than half of the stations have S picks in order to calculate the median; otherwise, the event is skipped. The use of the median is to prevent outliers from dominating the result. Subsequent references to S picks in the paper denote the median S pick across the array.

The S pick for a given event is used to define the start of a 1.0 s long window (Fig. 3a). The duration of the trapped waves windows is expected to increase with the width of the trapping structure, propagation distance within the fault zone layer, and velocity contrast with the bounding rocks (e.g. Ben-Zion 1998). Previous studies indicate that observed FZTW in large strike-slip faults associated with low velocity zones with widths on the order of 100 m and velocity reduction of 30–50 per cent generally have a duration of about 1 s or less (e.g. Li *et al.* 1994; Ben-Zion *et al.* 2003; Lewis *et al.* 2005; Mizuno *et al.* 2008). We tested window lengths of 1.0, 1.5 and 2.0 s and found that, for the Lander's rupture zone and JF site at the SJFZ, a 1.0 s window identified the most trapped waves at the most reliable rate. This window is referred to below as the trapped wave window. Additional information about the window length is given in the discussion section.

The trapped wave window is used to look for four features common to FZTW at each station. These characteristics are: (1) longer period energy than S body waves expected for a resonance mode in a waveguide, (2) typically larger amplitudes following the S -wave arrival (e.g. Fig. 3a), (3) arrival time delay relative to S wave, and (4) peak amplitude of trapped waves window relative to the average amplitude (termed 'relative peak'). Other characteristics such

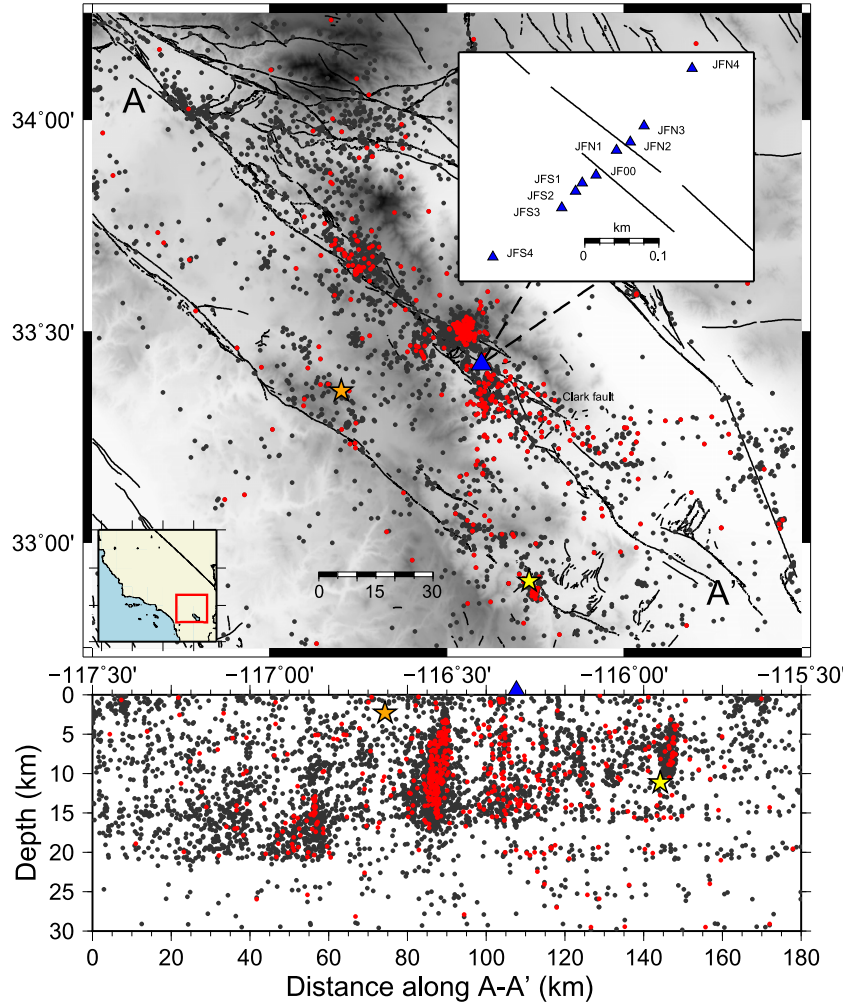


Figure 2. Location map for data (circles and stars) recorded during 2013 by a nine-station linear array (blue triangle and inset) across the San Jacinto fault zone at Jackass Flats. The bottom plot shows projection on a vertical cross-section along the line A–A'. The 5203 earthquakes shown were detected and located by the ANZA network. Data generated by these events are analysed for trapped waves. Events with at least one detection at the array are coloured red. A total of 540 events were flagged by the algorithm. The yellow and orange stars indicate location of events used in Figs 7 and 10, respectively. The shading indicates the regional topography, and black lines denote faults.

as dispersion are weaker second-order features for shallow trapping structures (e.g. Ben-Zion *et al.* 2003), and as such are not considered here. The features we use are analysed in the time domain, which is equivalent to but more computationally efficient than Fourier domain techniques (e.g. Kanamori 2005).

To measure the approximate period of the trapped waves window, we use the recursive predominant period algorithm of Nakamura (1988),

$$T_i^P = 2\pi \sqrt{\frac{X_i}{D_i}}, \quad (1a)$$

$$X_i = \alpha X_{i-1} + x_i^2, \quad (1b)$$

$$D_i = \alpha D_{i-1} + \left(\frac{\partial x}{\partial t}\right)_i^2, \quad (1c)$$

where T_i^P is the predominant period, x_i is the ground velocity at time index i and α is a damping constant equal to 0.999. These equations are recursively calculated for each sample over the trapped waves window to obtain the final measurement. Eqs (1a)–(1c) are used to

perform a single measurement of the predominant period for the trapped waves window in each examined seismogram.

The second feature is an estimate of the seismic energy in the trapped waves window. This is calculated simply as the sum of the squares of each data point in the time window. The third feature (relative peak) measures how strong the peak of the trapped waves window is relative to the entire window. It is calculated by taking the ratio between the absolute peak amplitude of the trapped waves window to the average amplitude of the window. This feature is small numerically when the entire trapped waves window has similar amplitude, and large when the peak value is considerably greater than the average value. If a FZTW is indeed present in the window, its amplitude should be generally larger than the surrounding S wave, leading to a large relative peak. The fourth feature measured is the time delay between the peak of the trapped waves window and the S -wave arrival. This property of FZTW can change due to a velocity contrast across the fault (Lewis & Ben-Zion 2010) and thus may not be present in all cases. As a result, it is used in our algorithm but given relatively low weight. More details are given about this feature in a subsequent section.

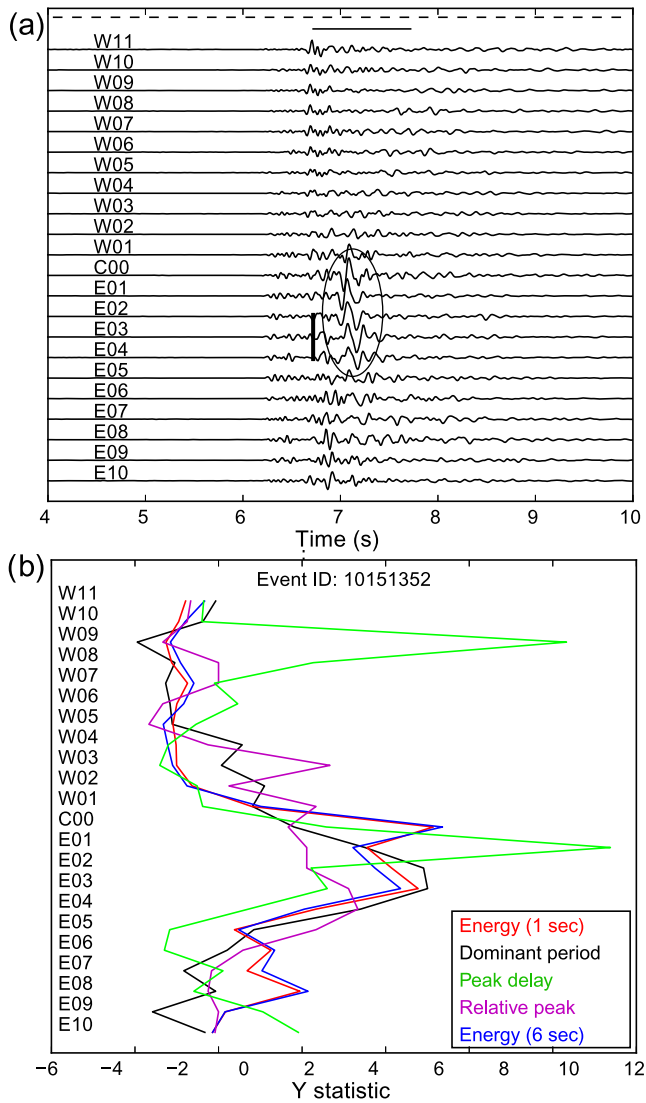


Figure 3. Example waveforms with FZTW (enclosed by ellipse) in the Landers array generated by the event denoted by the purple star in Fig. 1. (a) Fault-parallel component velocity seismograms across the array. FZTW are observed at stations C00–E04 shortly after the *S*-wave arrival (vertical black line; automatic pick). A 1.0 s window (solid horizontal line) starting at the *S* pick is used to calculate predominant period, wave energy, arrival delay, and relative peak strength at each station in the array. A 6.0 s window (dashed horizontal line) centred on the median *S* pick is used to calculate energy and identify records with possible site amplification. (b) Plot of the *Y* statistic values for each of the features. Stations E01–E05 have *Y* statistics for all of the 1 s window measurements that are 1 deviation above the median or more.

We found that using jointly these four characteristics is useful for identifying FZTW candidate waveforms, but similar features can be produced in some cases by other propagation and site effects. For example, highly localized site amplification can occur within an array, regularly yielding large energy measurements for a given site relative to the rest of the array. To identify situations where this occurs, a fifth measurement is made by calculating the energy in a longer time interval around the *S* arrival. Here we use a window duration of 6 s, centred on the *S* arrival (see Fig. 3a). The long duration of this window, combined with the fact that it is not restricted to just the *S* wave, makes this a suitable quantity

for identifying site amplification. This duration is also considerably longer than the expected duration of FZTW, so that if one is present it will not dominate the energy measurements in this window. The precise length of the window is not too important as long as the total duration is a significant portion of the entire seismogram and much longer than the duration of a trapped waves group. While the first four features described tend to indicate the likelihood of FZTW, the fifth one helps the algorithm to suppress erroneous cases with likely site amplification.

With the five features measured at each station for a given event, the problem becomes one of outlier detection to identify events with candidate FZTW. A common statistical method of outlier identification is to test whether a measurement is sufficiently far from the mean of a set of measurements, after normalizing by the sample's dispersion. This method is ineffective when the sample size is small and some expected values are statistical outliers themselves. Since the sample size here (the number of stations in a given array) is generally small, we must include all the stations available for each array in the calculations. Having possible unknown outliers decreases the detection sensitivity by biasing the mean and standard deviation. To address this limitation, we use a quantile based variation of the outlier identification scheme that uses instead the median and median-absolute-deviation (MAD). The MAD for the *j*th feature is defined as,

$$\text{MAD}^j = \text{median}(|X_i^j - \text{median}(X^j)|) \quad \text{for } i = 1, 2, \dots, N; \quad j = 1, 2, \dots, 5, \quad (2)$$

where X is the feature type and N is the total number of stations in the array. The MAD is intuitively a measure of dispersion that incorporates how far each measurement is from the median (rather than the mean) in an absolute value sense (e.g. Hampel 1974). It is thus insensitive to outliers that may be present, just like the median or interquartile range.

For a sample with a large number of measurements, the median and MAD would only need to be calculated once per event for each feature. However, since the number of stations may generally be small, and can still be affected if multiple stations are outliers, the median and MAD are calculated independently at each station. This is done by removing the station in question from these calculations, resulting for each feature in a slightly different MAD and median per station. Using these parameters the normalized statistic Y is calculated as,

$$Y_i^j = \frac{X_i^j - \text{median}_i(X^j)}{\text{MAD}_i^j}, \quad (3)$$

where X_i^j is the *j*th feature at the *i*th station. Y_i^j represents the number of deviations from the median for the *j*th feature at the *i*th station. This provides a simple convenient method for normalizing each measurement so that the same scale can be used for all features and events. The statistic Y is calculated for each of the five features at each station in the array, yielding a total of $5N$ Y -values. Fig. 3(b) illustrates the Y statistic values for each feature, at each station, for the waveform set shown in Fig. 3(a). For stations C00–E04, a clear bump is visible for all features indicating that the group of stations are collectively producing features that are above the median.

With the measurements transformed into a normalized form, detection thresholds can be easily set. We systematically tested many different combinations of thresholds for each of the five features, and found the set given in Table 1 to produce the most FZTW detections at the most reliable rate. Specific details about how these parameters

Table 1. *Y*-statistic thresholds used for trapped wave detection.

Feature type	Threshold
Energy (1 s)	0.75
Dominant period	1.25
Relative peak	0.0
Arrival delay	0.0
Energy (6 s)	2.75

were chosen are given in the next section. To make a detection, the four features measured on the 1 s window are all required to be above the listed threshold, whereas the energy measurement over the 6 s window is required to be below the listed threshold. The reason for the two different windows is now explicitly clear: the 1 s window is designed to target FZTW characteristics in a localized window after the *S*-arrival, while the 6 s window enables the method to avoid detection of site amplification that may be present in a particular trace. All records that satisfy the above criteria are flagged for further investigation.

The thresholds given in Table 1 may need to be changed to fit different data sets or applications. For example, since arrival delay for the trapped waves group is not always present, this parameter can be varied based on whether a velocity contrast is expected across the fault. For the example in Fig. 3, a detection is made on station E04. As several more stations would have been flagged if the 6 s energy requirement was relaxed or not imposed, it is clear that there is a trade-off between detection rate and ensuring that false detections by site amplification are kept to a minimum. We note that trapping structures and other low velocity fault zone layers generally produce themselves some site amplifications (e.g. Ben-Zion *et al.* 2003; Kurzon *et al.* 2014). This can be observed in the higher amplitude *P* waves for stations C00–E05 in Fig. 3(a), which explains why the energy measurements in the 6 s window are higher for these stations in Fig. 3(b). Nevertheless, the 6 s energy requirement prevents false detections due to other effects (e.g. topography and sedimentary basin) that may locally amplify the ground motion more strongly. One example of such non FZTW amplification is shown in Fig. 4.

4 APPLICATION TO DATA SETS AT THE LANDERS AND JF ARRAYS

We now demonstrate the method on the raw unfiltered data set recorded by the array across the Landers rupture zone. In total, 190 events met all of the quality criteria (minimum SNR and minimum number of *S*-picks) described in the previous section and were passed to the detector for investigation. First, we used the observations of Peng *et al.* (2003), who found that FZTW only were seen on stations W01–E06, to test all possible combinations of *Y*-statistic thresholds from 0 to 3.0 deviations above the median, in 0.25 deviation increments. The optimal parameter set, given in Table 1, was chosen by requiring the success rate to be greater than 90 per cent and then maximizing the number of detections on stations W01–E06. To give a sense for how sensitive the method is to the actual thresholds used, nearly 200 000 different combinations were tested during this process and yet the mean success rate over all of them was 88.3 per cent. As a result, we conclude that the method is quite robust with regards to the exact thresholds used.

Using the optimized set of thresholds provided in Table 1, we examined the corresponding detection results in greater detail. Fig. 5 displays a histogram with the number of detections at each station deemed to be true, as well as the number of false detections.

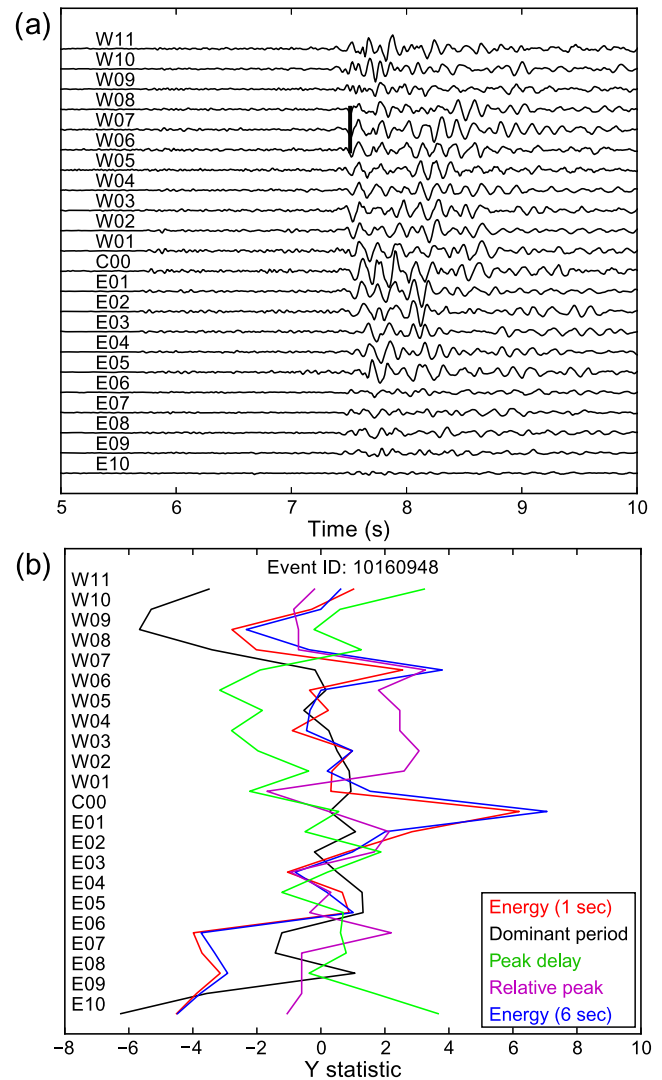


Figure 4. Example waveforms (fault-parallel component velocity seismograms) with large site amplification at station W07 generated by the event denoted by the green star in Fig. 1. The vertical black line indicates the median *S* pick across the array. Station W07 was found to have a *Y* statistic of 3.53 for the energy measurement in the 6 s window.

We visually inspected each detection in both the frequency and time domains to thoroughly examine which ones were false. In total, 70 events out of 190, which corresponds to 36.8 per cent, were flagged as FZTW candidates by the detector. The detected events are shown in map view and vertical cross-section as red circles in Fig. 1 on the background of other examined events (black circles). From the set of 108 seismograms flagged, 9 were deemed to be false detections (~8 per cent). Of the 99 detections we consider to be true, ~60 per cent were also identified by Peng *et al.* (2003) as having a large energy ratio between stations inside and outside the damage zone. It can be seen that the events with FZTW are broadly distributed in space rather than being highly localized near the rupture zone (Fig. 1). A deep low velocity fault zone layer that reaches the bottom of the seismogenic zone produces trapped waves only from events within the fault zone, since waves impinging on the fault zone from the outside are largely reflected away (e.g. Ben-Zion & Aki 1990; Igel *et al.* 1997; Jahnke *et al.* 2002). In contrast, a shallow low velocity layer can generate trapped waves from deeper regional events that inject wave energy into the fault zone from below (Ben-Zion *et al.*

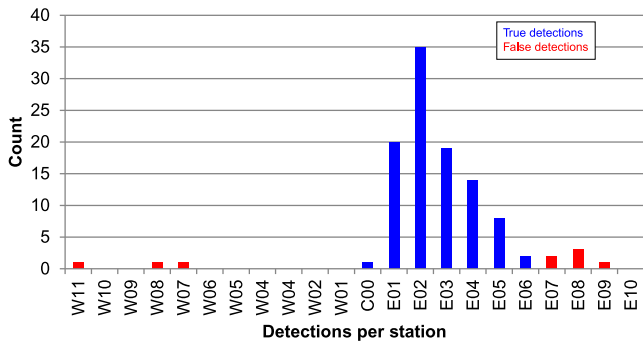


Figure 5. Results of the automatic detection algorithm applied to the Landers array data set. A total of 108 records were flagged for by the detector as likely containing trapped waves. Of these detections, 8.3 per cent were found to be false based on visual inspection. Nearly all detections occur between stations C00–E06.

2003; Peng *et al.* 2003; Fohrmann *et al.* 2004). The detection results in Fig. 1 therefore indicate that the Landers trapping structure extends primarily over the top few km of the crust.

One example set of FZTW generated by a given event and detected by the algorithm was given in Fig. 3. Another example is shown in Fig. 6; the stations with FZTW detections are denoted by a black star and the event location is marked by the yellow star in Fig. 1. These detections, as well as nearly all of the ones summarized in Fig. 5, are in close agreement with the general observations of FZTW detected by Peng *et al.* (2003). They found that FZTW in the Landers array were primarily observable at stations C00–E06. Our detection algorithm, like most other detection methods in seismology (e.g. Nippres *et al.* 2010; Langet *et al.* 2014; Ross & Ben-Zion 2014) is based on thresholds being met, and thus will miss FZTW with weaker attributes, especially for receivers near the edges of the trapping zone. This is clearly demonstrated in Fig. 6 by the fact that FZTW are visible at stations C00–E05, but were only identified by the algorithm at stations E04 and E05. The method is designed to detect for each examined event the clearest FZTW at a few stations, rather than detect weak candidate FZTW on many array stations, since the latter can lead to many false detections.

Next, the method is applied to the JF array data set across the SJFZ recorded during 2013. This data set consists of several thousand events, and with nine stations in the array (Fig. 2) is a very large volume of waveforms to sort through manually in a search for FZTW. Due to the large number of source–receiver combinations it is expected that there will also be a significant number of records with poor overall quality. To limit the number of seismograms with large noise and signal outside the frequency band of interest, we first apply to this data set a 2–20 Hz bandpass filter. Fig. 7 provides an example set of waveforms with FZTW detected at stations of the JF array. The location of the event generating this data set is marked with a yellow star in Fig. 2. It can be seen that long period, high amplitude FZTW about 0.5 s in duration start about half a second after the S arrival. The phases are only present clearly on JF00, JFS1, JFS2 and JFS3. These FZTW look visually overall similar to the ones in the Landers rupture zone (Figs 3 and 6), and are also overall similar to FZTW observed across a different location on the SJFZ (Lewis *et al.* 2005) as well as trapped waves at the Karadere-Duzce section of the North Anatolian fault (Ben-Zion *et al.* 2003) and the Parkfield section of the San Andrea fault (Li *et al.* 1990; Lewis & Ben-Zion 2010).

The ability of the algorithm to identify trapped waves for fault zone structures with different properties (here the Landers rupture

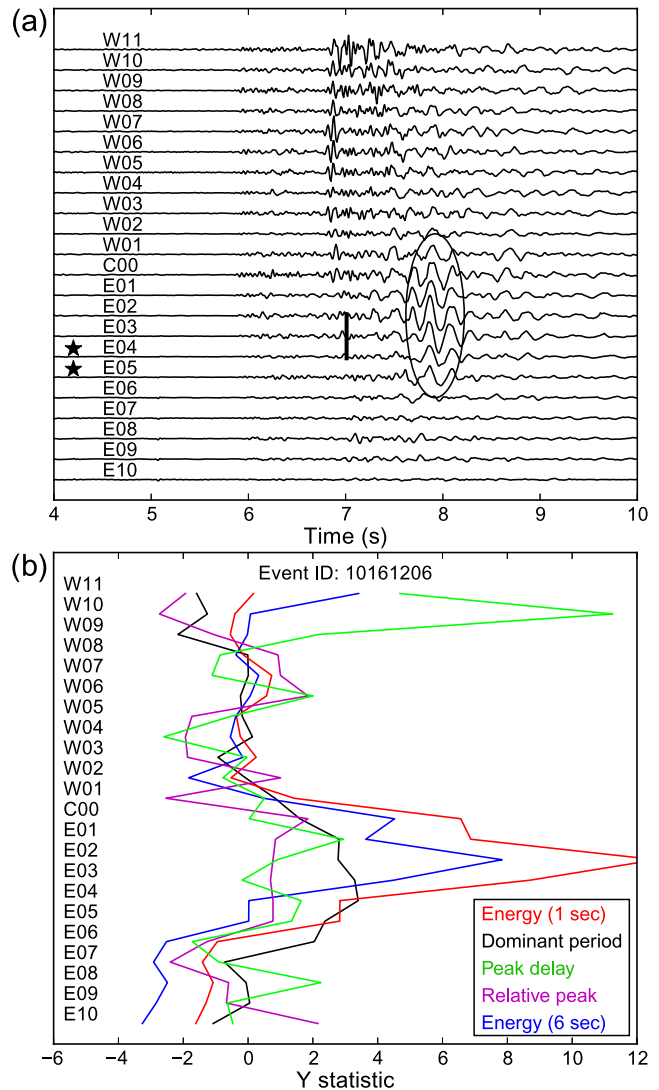


Figure 6. Fault-parallel component velocity seismograms with trapped waves detected at the Landers array generated by the event denoted by the yellow star in Fig. 1. The stations with detections are indicated by black stars and the vertical black line indicates the median *S* pick across the array. Visual inspection indicates candidate FZTW at stations C00–E05 (enclosed by ellipse).

zone and the JF site of the SJFZ) is important, as it allows the algorithm to be applied to different data sets with limited changing of parameters. Fig. 8 summarizes the number of detections in each station of the JF array for the examined data set. A total of 2255 events met all of the quality criteria (minimum SNR and minimum number of picks) needed for processing. In contrast to the Landers data set, we have no previous information on FZTW in this region to compare with. In total, 526 of the 582 detections (90.4 per cent) are concentrated at stations JF00–JFS3 as in the example shown in Fig. 7. These stations are also believed to span the extent of the damage zone. Station JFS1 is found to have the most detections by far; this is related to the fact that this station frequently records the most pronounced trapped waves across the array. The locations of all events generating detected FZTW at the JF array are shown in Fig. 2 (red dots). As with the Landers data set, the locations of events producing candidate FZTW at the JF array are distributed broadly, pointing again to a trapping structure that exists primarily over the top few km of the crust. While it appears from Fig. 2 that there are

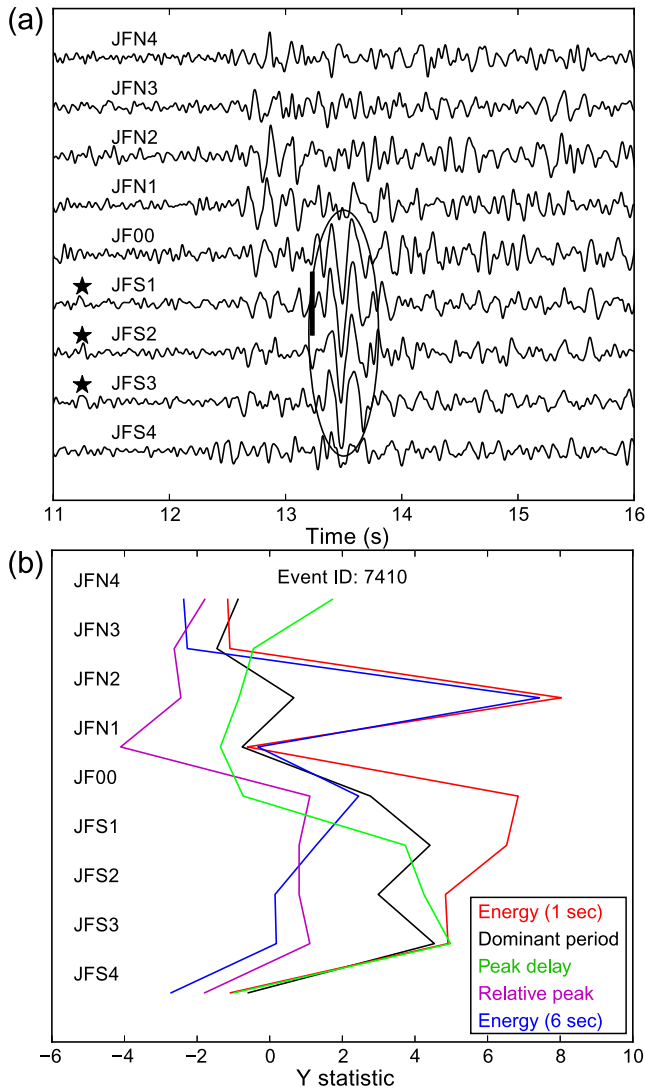


Figure 7. Fault-parallel component velocity seismograms with trapped waves detected at the JF array generated by the event denoted by the yellow star in Fig. 2. The stations with detections are denoted by black stars and the vertical black line indicates the median S pick across the array. Visual inspection indicates candidate FZTW (enclosed by ellipse) also at station JF00. Note the large 6 s energy measurement at station JFN2, which is not present at other stations, suggesting localized site amplification.

a few large clusters of detections, this is visually misleading as a substantial portion of all events are concentrated in these clusters. We have inspected visually about 300 randomly selected detections at the JF array and estimate our false detection rate on this data set to be around 10 per cent, which is in agreement with the results from the Landers array. None of the detections on the north side were found to contain clear trapped waves and are thus considered false detections. Visual inspection of the entire data set is the subject of future work; in particular the events with weakly generated FZTW require detailed examination to properly confirm. We further tested the method on raw JF data and found the general properties of the results to be unchanged. However, with raw data there are a larger number of additional detections for stations on the north side of the fault.

An important aspect of detection algorithms is to provide information on false detections. For the Landers data set, we examined visually all the automatic detections and also compared our detec-

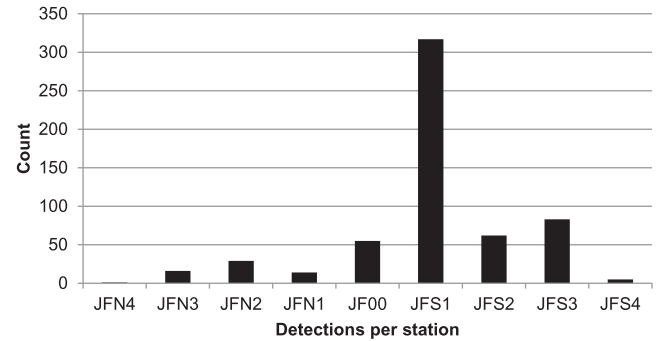


Figure 8. Results of applying the automatic detection algorithm to the JF array data set. A total of 582 records were flagged by the method with roughly 90 per cent of them concentrated between JF00–JFS3. FZTW are generally observed most strongly in the waveforms at station JFS1, explaining the large number of detections made by the algorithm. The detections at other stations are likely associated with various occasional site effects.

tions to the events identified by Peng *et al.* (2003) as having a large energy ratio between on- and off-fault stations. Fig. 9 shows an example of a detection at station W07 of the Landers array determined by visual inspection to be false. For this particular event, the five features just barely met the thresholds necessary for flagging the record at station W07 and no detections were made at other stations. An example of a false detection on the JF array is shown in Fig. 10. Here, the false detection occurred for station JFN2, and in this case the set of five features also just barely met all of the required thresholds. However, as evidenced by the histogram in Fig. 8, detections on the north side of the JF array are uncommon. For these types of isolated cases, highly localized site effects are the most probable cause of longer period S waves.

5 DISCUSSION

FZTW provide high-resolution information on internal components of fault zone structures, but identification of records containing FZTW has traditionally been a tedious task. While several automatic techniques have been used previously to aid in their identification, they were limited by the time needed to manually pick S phases and assumptions made. Peng *et al.* (2003) calculated energy ratios between different stations and Lewis *et al.* (2005) used a variant technique involving energy ratios over a particular bandwidth between different stations. The methods in these studies did not formally classify events or recordings as to whether or not they contained trapped waves, which is a key ingredient of a detection algorithm. With classification also comes success rates that provide an estimate of how reliable the method is. In this work we generalize and combine elements from these past techniques with the automatic S picking algorithm of Ross & Ben-Zion (2014), and use the method to classify each individual recording. The developed algorithm examines each seismogram in comparison to records at other stations of the array, and determines whether expected features of FZTW in the seismogram are statistical outliers of the array. Our detection algorithm adds to the energy estimates of previous techniques explicit measurements of the predominant period, relative peak strength, arrival delay of the considered phase and possible site amplification. These extra features provide important additional metrics for comparing the wavefields at different stations. By using the five features together, lower thresholds can be used, leading to more detections at a similar degree of reliability.

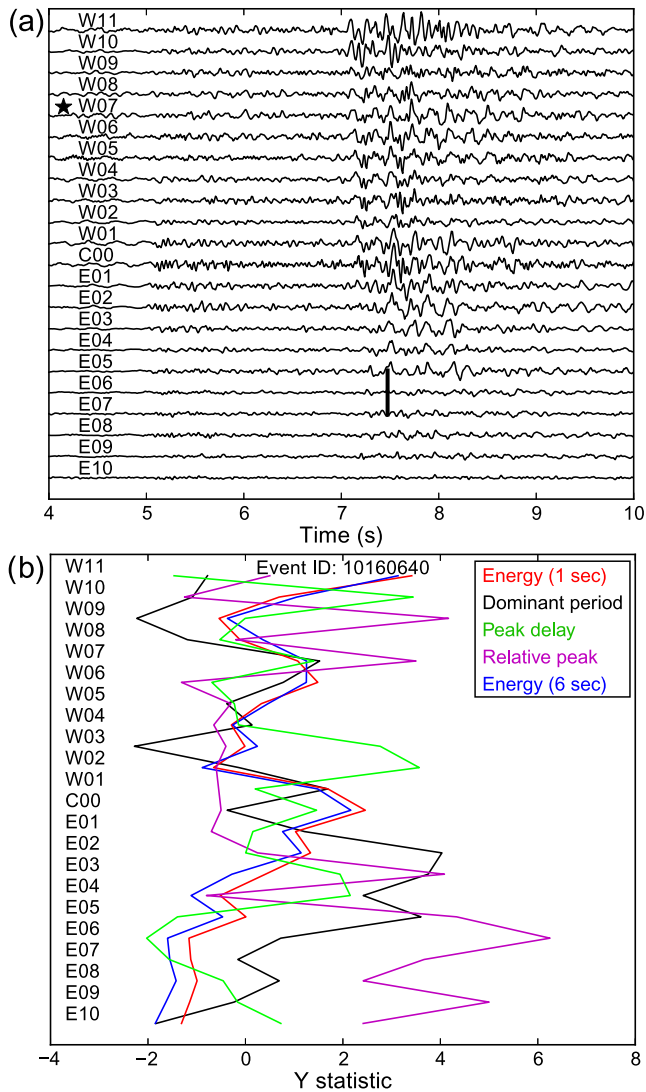


Figure 9. Fault-parallel component velocity seismograms with falsely detected trapped waves at the Landers array generated by the event denoted by the orange star in Fig. 1. The station with the false detection is denoted by a star and the vertical black line indicates the median S pick across the array. The four features measured on the 1 s window are just barely above the required thresholds and are not observed on nearby stations like in Figs 3 or 6.

Another challenge for identification of FZTW is that other wave propagation phenomena can occasionally produce (e.g. for some source mechanisms and source–receiver geometries) highly localized site effects that resemble FZTW. As seen in the histograms of Figs 5 and 8, we identify multiple occurrences of such signals (e.g. station JFN2 in Fig. 7). However, the systematic analysis performed by our algorithm recognizes that they are not generated by many events, whereas resonance modes associated with waveguides below the stations are expected to be frequently produced. One of the most serious limitations of past studies was the need for a trained analyst to pick S -wave arrivals to define the window for the FZTW search. The data volume collected now around the world is vast and rapidly increasing, making it unfeasible to sort through the data manually. The incorporation of automatic S picks in our algorithm allows for rapid processing of years of data, as demonstrated for the JF array.

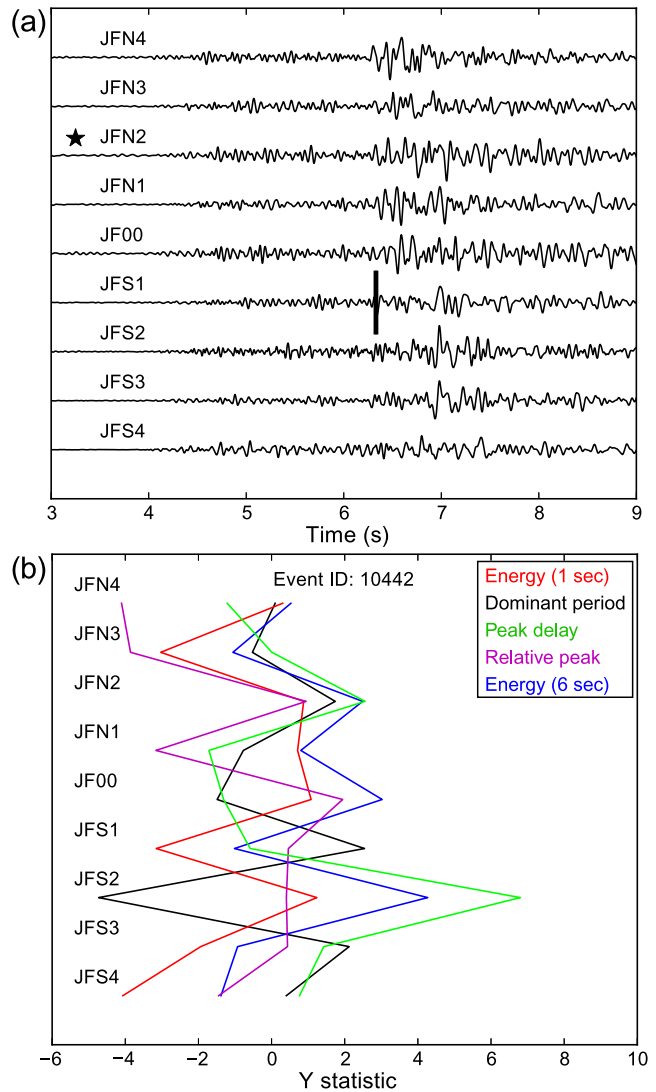


Figure 10. Fault-parallel component velocity seismograms with falsely detected trapped waves at the JF array generated by the event denoted by the orange star in Fig. 2. The station with the false detection is denoted by a star and the vertical black line indicates the median S pick across the array. The four features measured on the 1 s window are barely above the required thresholds at station JFN2 but visual inspection of the waveform at this and neighbouring stations leads to rejecting the detection.

The application of the automated algorithm should be followed by careful additional analyses to confirm that the identified phases are FZTW. The primary goal of the algorithm is to systematically and objectively analyse large data sets and reduce them to a dramatically smaller set of records that are likely to contain FZTW at certain stations. The method helps to identify which stations are worth examining first, which events are likely to be strong FZTW candidates, and which stations may have occasional signals with features similar to those of trapped waves (Figs 5 and 8). Additional follow up analyses of candidate FZTW include examining the moveout between the S wave and trapped waves group, analysing the dispersion and spectral properties of candidate FZTW, and performing waveform inversions for fault zone properties (e.g. Ben-Zion *et al.* 2003; Peng *et al.* 2003; Lewis *et al.* 2005). Identification of the highest-quality FZTW candidates within a set of detections is often an important first step for modelling waveforms to invert for structural properties of fault zones. We found the number of detections made per event

to be generally a reliable indicator of FZTW quality. By sorting the events based on the number of stations with detections, one can start with the events having most detections and work backwards. For both the Landers and JF arrays, events with two to three detections typically had FZTW clearly visible at many stations, but this number will vary depending on the number of stations within the damage zone.

Nearly all detection methods, including the one presented here, involve the use of some parameters. In the current method there are five detection thresholds and two window durations that may be adjusted for different applications. The values used here for these parameters (Table 1) were selected to provide many detections while keeping the false detection rate relatively low. We used the same parameter values for both the Landers and JF arrays and found these to work well in both cases. If the thresholds for the 1 s window features are lowered, additional weak candidate trapped waves will be recognized at the cost of more false detections.

The method has several limitations that should be noted. First, the number of detections that can be made per event depends on the number of stations both inside and outside the fault zone trapping structure. For future deployments of linear arrays with fixed number of stations, our results imply that fewer arrays with more stations may be better than more arrays with fewer stations for the purpose of detecting FZTW. Our testing suggests that less than ~30 per cent of the stations should be located inside the damage zone to facilitate identifying stations with trapped waves as outliers. Another limitation of the method is the accuracy of the automatic *S* picks. The algorithm of Ross & Ben-Zion (2014) is able to make *S* picks to within 0.25 s of the analyst pick roughly 90 per cent of the time. If the used trapped wave window length is 1 s, as in this work, this allows some room for error in the *S* pick, and should not be an issue unless the FZTW are arriving quite late (~1 s or more). Further, the *S*-wave picks tend to be late, rather than early, which helps lessen this problem.

Using the median *S* pick and requiring a majority of stations to have valid picks improves the stability of the detections. Most of the false detections in the examined data at the Landers and JF arrays were in fact due to mispicks of *S* arrivals. However, the overall false detection rate is reasonably low enough to produce overall robust results. The mean *S* pick across the array could alternatively be used instead of the median if outliers are discarded beforehand. In cases with strong overall velocity contrast across the fault, the variability of *S* picks at different stations as well as the delay between the *S* arrival and FZTW increase (e.g. Lewis & Ben-Zion 2010). In such cases it may be better to use *S* picks at individual stations and to lengthen the trapped wave window. The detection algorithm is not very sensitive to the precise number of array stations. We have tested the method on arrays with the number of stations varying from 7 to 21 and find similar success rates.

The decision to rotate seismograms to fault-parallel component of motion for performing the analysis may be viewed as a pre-processing step. Love-type trapped waves involve particle motion parallel to the fault zone layer (Ben-Zion & Aki 1990; Ben-Zion 1998), so FZTW are expected to be strongest on the fault-parallel (and for vertical fault zones to some extent also vertical) component seismograms. For fault zones with known geometry that is approximately linear, it is straight forward to rotate seismograms to the fault-parallel direction. For fault zones with curvature, such as the Landers rupture zone (Fig. 1), the choice of a trend is more difficult. We found, however, that using the north component of motion for all of the Landers data yielded results that were similar to an average fault parallel component by drawing a line through the rupture

end points. Alternatively, the vertical component seismograms can be used (e.g. Li *et al.* 1994). In any case, the choice of rotation is not critical and should be seen more as a technique for enhancing the contrast between trapped waves and *S* waves rather than a strict requirement.

The developed automated method for identification of fault zone trapped waves was tested in this work in the context of *S*-type FZTW, but it could in principle be also applied to *P*-type FZTW (Ellsworth & Malin 2011) with appropriate changes of parameters. Our algorithm was tested partially on several other linear arrays deployed across the SJFZ, and seems to produce satisfactory results on detection of *S*-type FZTW without changing the parameters (Qiu *et al.* 2015; Share *et al.* 2015). Work on comprehensive detection of FZTW in data of multiple dense arrays across the SJFZ and modelling the data for high-resolution information on the internal fault zone structure is in progress.

6 CONCLUSIONS

We developed a method for automated identification of fault zone trapped waves produced by constructive interference in low-velocity fault zone layers. The method identifies the *S*-wave group using an automatic *S*-wave arrival picking algorithm (Ross & Ben-Zion 2014), and computes a set of features that are representative of trapped waves on each record for a given event. The features are compared among all stations across a fault zone array, and are used to identify statistical outliers. Records that satisfy a set of criteria are flagged for further visual examination. A set of detection thresholds was optimized over thousands of different combinations to obtain the best results in the examined data sets. Applying the method to data sets in the Eastern California Shear Zone and San Jacinto Fault Zone yielded clear detections at subsets of array stations consistent with manual analyses. The method may be applied to data sets at other locations and other wave types with some changes of parameters.

ACKNOWLEDGEMENTS

We thank Hongrui Qiu and Pieter-Ewald Share for useful discussions. The study was supported by the National Science Foundation (grant EAR-0908903). The manuscript benefitted from constructive comments of two anonymous referees.

REFERENCES

- Avallone, A., Rovelli, A., Di Giulio, G., Improta, L., Ben-Zion, Y., Milana, G. & Cara, F., 2014. Waveguide effects in very high rate GPS record of the 6 April 2009, Mw 6.1 L'Aquila, central Italy earthquake, *J. geophys. Res.: Solid Earth*, **119**, 490–501.
- Ben-Zion, Y., 1998. Properties of seismic fault zone waves and their utility for imaging low-velocity structures, *J. geophys. Res.: Solid Earth*, **103**, 12 567–12 585.
- Ben-Zion, Y. & Aki, K., 1990. Seismic radiation from an SH line source in a laterally heterogeneous planar fault zone, *Bull. seism. Soc. Am.*, **80**, 971–994.
- Ben-Zion, Y. *et al.*, 2003. A shallow fault zone structure illuminated by trapped waves in the Karadere-Duzce branch of the North Anatolian Fault, western Turkey, *Geophys. J. Int.*, **152**, 699–717.
- Ben-Zion, Y. *et al.*, 2015. Basic data features and results from a spatially-dense seismic array on the San Jacinto fault zone, *Geophys. J. Int.*, **202**, 370–380.

- Bulut, F., Bohnhoff, M., Ellsworth, W.L., Aktar, M. & Dresen, G., 2009. Microseismicity at the North Anatolian Fault in the Sea of Marmara offshore Istanbul, NW Turkey, *J. geophys. Res.*, **114**, B09302, doi:10.1029/2008JB006244.
- Calderoni, G., Di Giovambattista, R., Vannoli, P., Pucillo, S. & Rovelli, A., 2012. Fault-trapped waves depict continuity of the fault system responsible for the 6 April 2009 MW 6.3 L'Aquila earthquake, central Italy, *Earth planet. Sci. Lett.*, **323–324**, 1–8.
- Ellsworth, W.L. & Malin, P., 2011. Deep rock damage in the San Andreas Fault revealed by P- and S-type fault-zone-guided waves, *Geol. Soc. Spec. Publ.*, **359**(1), 39–53.
- Fohrmann, M., Igel, H., Jahnke, G. & Ben-Zion, Y., 2004. Guided waves from sources outside faults: an indication for shallow fault zone structure?, *Pure appl. Geophys.*, **161**, 2125–2137.
- Fukao, Y., Hori, S. & Ukawa, M., 1983. A seismological constraint on the depth of basalt–eclogite transition in a subducting oceanic crust, *Science*, **303**, 413–415.
- Hampel, F.R., 1974. The influence curve and its role in robust estimation, *J. Am. Stat. Assoc.*, **69**, 383–393.
- Hong, T.-K. & Kennett, B.L.N., 2003. Modelling of seismic waves in heterogeneous media using a wavelet-based method: application to fault and subduction zones, *Geophys. J. Int.*, **154**, 483–498.
- Igel, H., Ben-Zion, Y. & Leary, P., 1997. Simulation of SH and P-SV wave propagation in fault zones, *Geophys. J. Int.*, **128**, 533–546.
- Jahnke, G., Igel, H. & Ben-Zion, Y., 2002. Three-dimensional calculations of fault-zone-guided waves in various irregular structures, *Geophys. J. Int.*, **151**(2), 416–426.
- Kanamori, H., 2005. Real-time seismology and earthquake damage mitigation, *Annu. Rev. Earth planet. Sci.*, **33**, 195–214.
- Kurzton, I., Vernon, F.L., Rosenberger, A. & Ben-Zion, Y., 2014. Real-time automatic detectors of P and S waves using singular value decomposition, *Bull. seism. Soc. Am.*, **104**, 1696–1708.
- Langet, N., Maggi, A., Michelini, A. & Brenguier, F., 2014. Continuous kurtosis-based migration for seismic event detection and location, with application to Piton de la Fournaise Volcano, La Reunion, *Bull. seism. Soc. Am.*, **104**, 229–246.
- Lewis, M.A. & Ben-Zion, Y., 2010. Diversity of fault zone damage and trapping structures in the Parkfield section of the San Andreas Fault from comprehensive analysis of near fault seismograms, *Geophys. J. Int.*, **183**, doi:10.1111/j.1365-246X.2010.04816.x.
- Lewis, M.A., Peng, Z., Ben-Zion, Y. & Vernon, F.L., 2005. Shallow seismic trapping structure in the San Jacinto fault zone near Anza, California, *Geophys. J. Int.*, **162**, 867–881.
- Li, Y.G., Leary, P., Aki, K. & Malin, P., 1990. Seismic trapped modes in the Oroville and San Andreas fault zones, *Science*, **249**, 763–766.
- Li, Y.G., Aki, K., Adams, D., Hasemi, A. & Lee, W.H.K., 1994. Seismic guided waves trapped in the fault zone of the Landers, California, earthquake of 1992, *J. geophys. Res.*, **99**, 11 705–11 722.
- Mamada, Y., Kuwahara, Y., Ito, H. & Takenaka, H., 2004. Discontinuity of the Mozumi–Sukenobu fault low-velocity zone, central Japan, inferred from 3-D finite-difference simulation of fault zone waves excited by explosive sources, *Tectonophysics*, **378**, 209–222.
- McGuire, J. & Ben-Zion, Y., 2005. High-resolution imaging of the Bear Valley section of the San Andreas fault at seismogenic depths with fault-zone head waves and relocated seismicity, *Geophys. J. Int.*, **163**, 152–164.
- Mizuno, T., Kuwahara, Y., Ito, H. & Nishigami, K., 2008. Spatial variations in fault-zone structure along the Nojima Fault, Central Japan, as inferred from borehole observations of fault-zone trapped waves, *Bull. seism. Soc. Am.*, **98**, 558–570.
- Nakamura, Y., 1988. On the urgent earthquake detection and alarm system (UrEDAS), in *Proceedings of the 9th World Conference on Earthquake Engineering, Japan*, Vol. VII, pp. 673–678.
- Nippres, S.E.J., Rietbrock, A. & Heath, A.E., 2010. Optimized automatic pickers: application to the ANCORP data set, *Geophys. J. Int.*, **181**, 911–925.
- Ozakin, Y., Ben-Zion, Y., Aktar, M., Karabulut, H. & Peng, Z., 2012. Velocity contrast across the 1944 rupture zone of the North Anatolian fault east of Isetmpasa from analysis of teleseismic arrivals, *Geophys. Res. Lett.*, **39**, L08307, doi:10.1029/2012GL051426.
- Peng, Z., Ben-Zion, Y., Michael, A.J. & Zhu, L., 2003. Quantitative analysis of seismic trapped waves in the rupture zone of the 1992 Landers, California earthquake: evidence for a shallow trapping structure, *Geophys. J. Int.*, **155**, 1021–1041.
- Qiu, H., Ben-Zion, Y., Ross, Z.E., Share, P.-E. & Vernon, F., 2015. Internal structure of the San Jacinto fault zone at Jackass Flat from data recorded by a dense linear array, in *2015 Annual Meeting Seismological Society of America*, 21–23 April, Pasadena, California.
- Ross, Z.E. & Ben-Zion, Y., 2014. Automatic picking of direct P, S seismic phases and fault zone head waves, *Geophys. J. Int.*, **199**, 368–381.
- Rovelli, A., Caserta, A., Marra, F. & Ruggiero, V., 2002. Can seismic waves be trapped inside an inactive fault zone? the case study of Nocera Umbra, Central Italy, *Bull. seism. Soc. Am.*, **92**, 2217–2232.
- Shapiro, N.M., Olsen, K.B. & Singh, S.K., 2000. Wave-guide effects in subduction zones: evidence from three-dimensional modeling, *Geophys. Res. Lett.*, **27**(3), 433–436.
- Share, P.-E., Ben-Zion, Y., Ross, Z.E., Qiu, H. & Vernon, F., 2015. Characterization of the San Jacinto Fault Zone northwest of the trifurcation area from earthquake data recorded by a dense linear array, in *2015 Annual Meeting Seismological Society of America*, 21–23 April, Pasadena, California.