

Automatic picking of direct P , S seismic phases and fault zone head 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 2014 July 8. Received 2014 July 5; in original form 2014 March 24

SUMMARY

We develop a set of algorithms for automatic detection and picking of direct P and S waves, as well as fault zone head waves (FZHW), generated by earthquakes on faults that separate different lithologies and recorded by local seismic networks. The S -wave picks are performed using polarization analysis and related filters to remove P -wave energy from the seismograms, and utilize STA/LTA and kurtosis detectors in tandem to lock on the phase arrival. The early portions of P waveforms are processed with STA/LTA, kurtosis and skewness detectors for possible first-arriving FZHW. Identification and picking of direct P and FZHW is performed by a multistage algorithm that accounts for basic characteristics (motion polarities, time difference, sharpness and amplitudes) of the two phases. The algorithm is shown to perform well on synthetic seismograms produced by a model with a velocity contrast across the fault, and observed data generated by earthquakes along the Parkfield section of the San Andreas fault and the Hayward fault. The developed techniques can be used for systematic processing of large seismic waveform data sets recorded near major faults.

Key words: Time-series analysis; Earthquake source observations; Body waves; Interface waves; Wave propagation.

1 INTRODUCTION

Systematic analyses of seismic data sets require algorithms for automatic event detection and phase picking. This is especially important with the vast amount of data that are now continuously recorded by various seismic networks. Analysts' inspections and corrections can refine the automatic results, but are unfeasible with the large modern data sets and also less objective if done alone. In the present study we attempt to develop reliable algorithms for automatic detection of seismic phases in the early portions of P and S waveforms recorded in close proximity to large faults and earthquake rupture zones. The algorithms process earthquake waveform data sets that have already been selected by event detection algorithms (e.g. Withers *et al.* 1998; Nippress *et al.* 2010; Ross & Ben-Zion 2014).

Much effort has been devoted to algorithms for automatic picking of direct P waves which are often the first arrivals on seismograms. Proposed techniques and signals include the short-term average/long-term average (STA/LTA; Allen 1978), envelope functions (Baer & Kradolfer 1987), neural networks (e.g. Gentili & Michelini 2006) and autoregressive methods (e.g. Sleeman & van Eck 1999; Rastin *et al.* 2013). Higher order statistics including kurtosis and skewness have also been used for automatic phase picking (Saragiotis *et al.* 2002). The kurtosis was found to be more accurate than the traditional methods by a number of studies (e.g. Kuperkoch *et al.* 2010; Nippress *et al.* 2010; Langet *et al.* 2014) since it is primarily sensitive to abrupt changes in the character of

a time-series. A similar statement holds for the skewness. Nippress *et al.* (2010) utilized two picking techniques in tandem to obtain more reliable P -wave picks. They used an STA/LTA function or the damped predominant period T^{pd} function (Hildyard *et al.* 2008) to find an initial picking region, and then refined the pick using a kurtosis characteristic function (CF).

Automatic picking of S -wave arrivals is more challenging due to contamination of the direct S -wave arrival by the P coda and converted phases. A number of studies have incorporated polarization analysis into their picking algorithms to distinguish between P and S waves. Baillard *et al.* (2014) used for this purpose the kurtosis function with a 'dip-rectilinearity' function derived from the covariance matrix. Cichowicz (1993) calculated a set of polarization quantities from the covariance matrix, and constructed from them a CF used to pick phases. Kurzon *et al.* (2014) used a singular value decomposition based on the algorithm of Rosenberger (2010) to filter seismograms into polarized P and S channels, and picked direct P and S phases on CFs derived from them.

In various situations in seismology, the direct P wave is not the first arrival on the seismogram. For example, events recorded at regional and teleseismic distances often have first arriving head waves (P_n) that spend most of the travel path propagating (refracting) along the Moho with the faster velocity of the mantle (e.g. Aki & Richards 2002). Head waves are associated with emergent arrivals on the slower side of the lithology interface, in contrast to the more impulsive body waves. Similar waves that refract along

horizontal interfaces are encountered in exploration seismology and exploited for imaging purposes. Head waves may also refract along dipping or vertical velocity contrast (bimaterial) interfaces of major fault structures. Such waves generated by earthquakes located close to the bimaterial interface are termed fault zone head waves (FZHW), and they have opposite first motion polarity from the following direct P arrivals on the slower side of the fault (e.g. Ben-Zion 1989; Ben-Zion & Aki 1990).

In the last few decades numerous dense seismic networks have been deployed around large faults and earthquake rupture zones, in an effort to obtain high resolution results on structure and source properties (e.g. Bakun & Lindh 1985; Fletcher *et al.* 1987; Seeber *et al.* 2000). Seismic data recorded near faults allow the detection and use of FZHW in diverse tectonic settings including subduction zones (e.g. Fukao *et al.* 1983; Hori *et al.* 1985; Hong & Kennett 2003) and large strike-slip faults (e.g. Ben-Zion & Malin 1991; Hough *et al.* 1994; McGuire & Ben-Zion 2005; Bulut *et al.* 2012). Although FZHW exist only in relatively narrow regions on the slower velocity media (see eq. 6 in Section 4), they are recorded routinely by near-fault seismic networks (e.g. Ben-Zion & Malin 1991; Allam *et al.* 2014). Misidentification of FZHW as direct P arrivals can lead to systematic errors in derived velocity structures, event locations and focal mechanisms (e.g. McNally & McEvelly 1977; Oppenheimer *et al.* 1988), while proper use of FZHW can increase the resolution of the derived results (e.g. Ben-Zion *et al.* 1992; Bennington *et al.* 2013).

Ruptures along bimaterial interfaces have fundamentally different properties than classical ruptures in a homogenous solid (e.g. Weertman 1980; Andrews & Ben-Zion 1997; Ben-Zion 2001; Ranjith & Rice 2001). Significantly, bimaterial ruptures tend to evolve for various conditions to pulses with a preferred propagation direction and dynamic reduction of normal stress leading to small frictional heat and possible tensile component of faulting (e.g. Ben-Zion & Huang 2002; Shi & Ben-Zion 2006; Ampuero & Ben-Zion 2008; Brietzke *et al.* 2009). These and related observational studies (e.g. Rubin & Gillard 2000; Lewis *et al.* 2005; Dor *et al.* 2008; Lengline & Got 2011; Zaliapin & Ben-Zion 2011) highlight the importance of detecting and imaging properties of fault bimaterial interfaces. FZHW provide the clearest evidence for the existence, and highest imaging resolution, of fault bimaterial interfaces.

In the following sections we describe and illustrate a set of algorithms for automatic picking of P and S body waves and FZHW generated by local earthquakes and recorded near faults. The algorithms can be used as a single unit to pick all these phases or as a set of separate phase picking techniques. The picking of direct P and S phases utilizes polarization analysis to distinguish between these waves, while the picking of direct P and FZHW utilizes expected properties of these waves to make picks with a high degree of accuracy. We apply the techniques to waveforms generated by several thousand events at the Parkfield section of the San Andreas fault, as well as synthetic waveforms and data recorded near the Hayward fault. The algorithms separate P and S waveforms for a large portion of the data, and identify early FZHW preceding the direct P body waves for a significant fraction of the waveforms recorded on the slow sides of the San Andreas and Hayward faults.

2 DATA AND PRE-PROCESSING

The algorithm for picking the various phases of interest (direct P , S and FZHW) is primarily illustrated and tested using seismograms

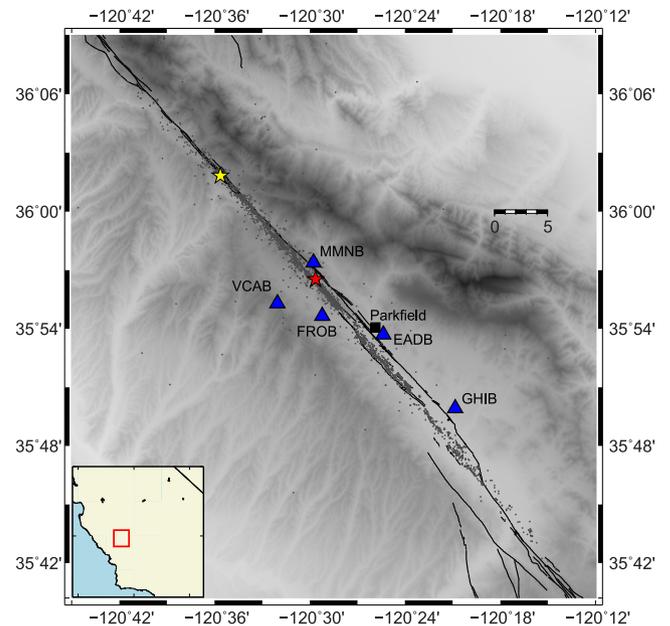


Figure 1. The Parkfield section of the San Andreas Fault with five HRSN stations (blue triangles) and epicentres of ~ 2500 earthquakes (grey circles and coloured stars) used in this study from the NCEDC catalogue. The events marked by stars are utilized to illustrate various aspects of the algorithm. The town of Parkfield, CA, and surface traces of faults are indicated by black square and lines.

recorded at five stations of the High Resolution Seismic Network (HRSN) at the Parkfield section of the San Andreas fault (Fig. 1). The data are recorded by short-period borehole seismometers operating at depths ranging from 100 to 300 m. In addition, we use synthetic seismograms in a model consisting of a vertical low velocity zone between two quarter spaces (Ben-Zion & Aki 1990; Ben-Zion 1998) to verify the performance of the direct P and head wave pickers, and we test the algorithm further on data of the Northern California Earthquake Data Center recorded around the Hayward fault.

The Parkfield section of the San Andreas fault separates generally faster granite rocks to the SW from slower Franciscan rock to the NE (e.g. Eberhart-Phillips & Michael 1993; Thurber *et al.* 2006). The five HRSN stations utilized in this work were chosen because they were used in several previous studies of FZHW near Parkfield (Ben-Zion & Malin 1991; Ben-Zion *et al.* 1992; Zhao *et al.* 2010). These studies demonstrated that the three stations on the NE side of the fault (MMNB, Ghib and EADB) frequently record FZHW, and the two stations located on the SW side (VCAB and FROB) do not. In Sections 3–5, we analyse seismograms generated by 2489 events that occurred in 2004 (grey dots in Fig. 1).

Prior to the phase picking we perform basic pre-processing involving removing the mean from each seismic record and band-pass filtering the data to remove long period noise and high frequency glitches that are sometimes present in the data. For the HRSN Parkfield records we use a (single-pass) causal band-pass filter between 0.5 and 30 Hz; these values may require modifications in other applications based on data quality. A causal filter is important because 2-pass acausal filters can introduce ringing to the front of the seismogram which can reduce the accuracy of picking the arrival of FZHW (Allam *et al.* 2014). We also note that since FZHW are emergent with lower frequencies than direct P waves, they can be significantly affected by filters above 0.5 Hz. The instrument response is not removed from the seismograms.

3 PICKING P AND S PHASES

The algorithm for picking P and S body waves relies on polarization analysis to facilitate their separation within recorded seismograms. Kurzon *et al.* (2014) recently used a singular value decomposition technique to separate P from S waves. Here we use a covariance matrix for the polarization analysis, because we find that the phase separation is more effective for some recordings with this method. However, in most cases both techniques of polarization analysis produce similar separation of P and S phases.

The covariance matrix for a finite sample of three-component data can be computed as,

$$\sigma = \begin{bmatrix} \text{Cov}(N, N) & \text{Cov}(N, E) & \text{Cov}(N, Z) \\ \text{Cov}(E, N) & \text{Cov}(E, E) & \text{Cov}(E, Z) \\ \text{Cov}(Z, N) & \text{Cov}(Z, E) & \text{Cov}(Z, Z) \end{bmatrix}, \quad (1)$$

where Z is the vertical component, N is north–south component, E is the east–west component, and the covariance between any two components X and Y is given by

$$\text{Cov}(X, Y) = \frac{1}{M} \sum_{i=1}^M x_i y_i, \quad (2)$$

with M being the number of samples. For separating the P and S phases, we compute the covariance matrix of the pre-processed seismograms for a 3 s sliding window. This duration has been found to be stable for use with covariance matrices by Baillard *et al.* (2014). The corresponding eigenvalues ($\lambda_1 \geq \lambda_2 \geq \lambda_3$) and eigenvector matrix $\mathbf{u} = (\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$ of σ can be used to calculate various aspects of particle motion (e.g. Jurkevics 1988).

The rectilinearity, r , or degree of linear polarization, is calculated as,

$$r = 1 - \left(\frac{\lambda_2 + \lambda_3}{2\lambda_1} \right), \quad (3)$$

and has a range of $[0, 1]$. For body waves, r should be approximately equal to 1. The apparent vertical incidence angle, φ , for the particle motion is defined as

$$\cos(\varphi) = u_{11}. \quad (4)$$

For local earthquakes, $\cos(\varphi)$ should be close to 1 for P waves and close to 0 for S waves. From these quantities, we construct two polarization filters,

$$p = r \cos(\varphi), \quad (5a)$$

$$s = r[1 - \cos(\varphi)]. \quad (5b)$$

These filters are used to modulate the three-component data (e.g. Rosenberger 2010) by simple multiplication at each time step. Specifically, p is used to modulate the vertical component and s is used to modulate the two horizontal components. We refer to these modulated signals as polarized traces since they have suppressed energy with a particular polarization. This part of the procedure is similar to that used by Cichowicz (1993), who also employed a third (transverse energy) quantity involving the ratio of the horizontal vector magnitude to the total vector magnitude.

An example of our P and S separation process is illustrated in Fig. 2 for an event recorded at station MMNB (red star, Fig. 1). The polarization filters p and s calculated in this case are shown in Figs 2(d) and (e), respectively. It is clear that p is large during the P wave and small during the S wave and vice versa for s . The polarized versions of the traces are plotted in Fig. 2(f) with the polarized Z

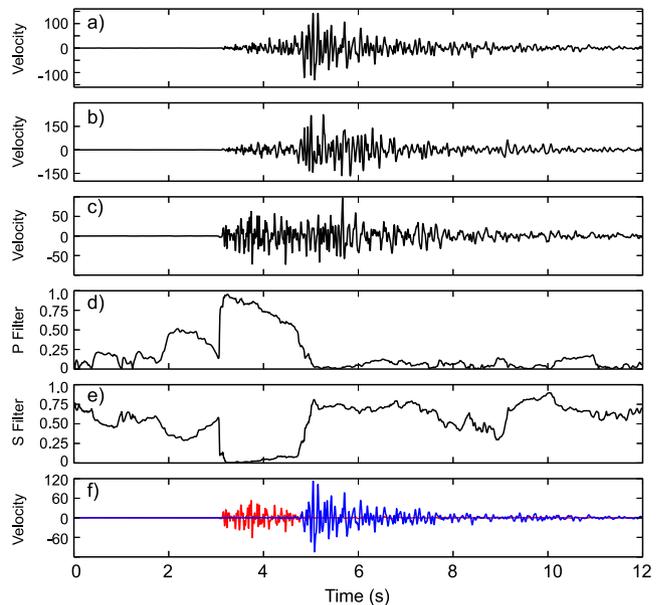


Figure 2. Application of the polarization analysis to separate a seismogram into P and S phases for the event denoted by red star in Fig. 1 and recorded at station MMNB. (a) North component velocity seismogram. (b) East component velocity seismogram. (c) Vertical component velocity seismogram. (d) A p polarization filter (eq. 5a) determined from the covariance matrix with high values during the P wave and low values during the S wave. (e) A corresponding s polarization filter determined from the covariance matrix (eq. 5b). (f) A P -polarized vertical trace (red) plotted together with the S -polarized east–west trace (blue).

component in red and the polarized N component in blue. The P -wave signal originally present in the N component has been almost entirely removed, and the S -wave signal is mostly removed from the Z component. This leads to a clear visual separation of the phases.

Next we run detectors on the polarized traces to pick various phases automatically. To obtain an initial pick of the P -wave arrival, we calculate a STA/LTA CF on the polarized vertical trace with STA and LTA windows of 1 and 30 s, respectively. These parameters were chosen so that the STA/LTA function would have general applicability to source–receiver distances and event magnitudes common to local earthquakes, without being overly sensitive to short transient signals not arising from earthquakes. The initial P arrival pick is made when the STA/LTA reaches an *ad hoc* level of 5. This value was found in exploratory analyses of several data sets to provide a good balance between the numbers of correct and false detections. The initial pick is used to approximately locate the P -wave arrival, but must be refined further as there is some amount of delay time associated with the STA/LTA recognizing the P arrival. This refinement stage is described in detail in the subsequent section where we also account for the possibility of FZHW. An example of the P -wave picking process discussed so far is illustrated in Fig. 3 for the same event used in Fig. 2, along with results for the S -wave pick described next. Figs 3(a)–(c) show the unpolarized three-component traces with P and S wave picks denoted by red and blue vertical lines, respectively. Fig. 3(d) displays the STA/LTA function calculated on the P -polarized vertical trace, and Figs 3(e)–(i) are associated with the S -wave pick.

To pick the S -wave arrival, we first calculate an STA/LTA CF on each of the S -polarized horizontal traces (Figs 3e–f). The STA and LTA windows are 1 and 30 s in duration, respectively. These values were chosen for general applicability to local earthquakes, and changing the durations somewhat around these values does

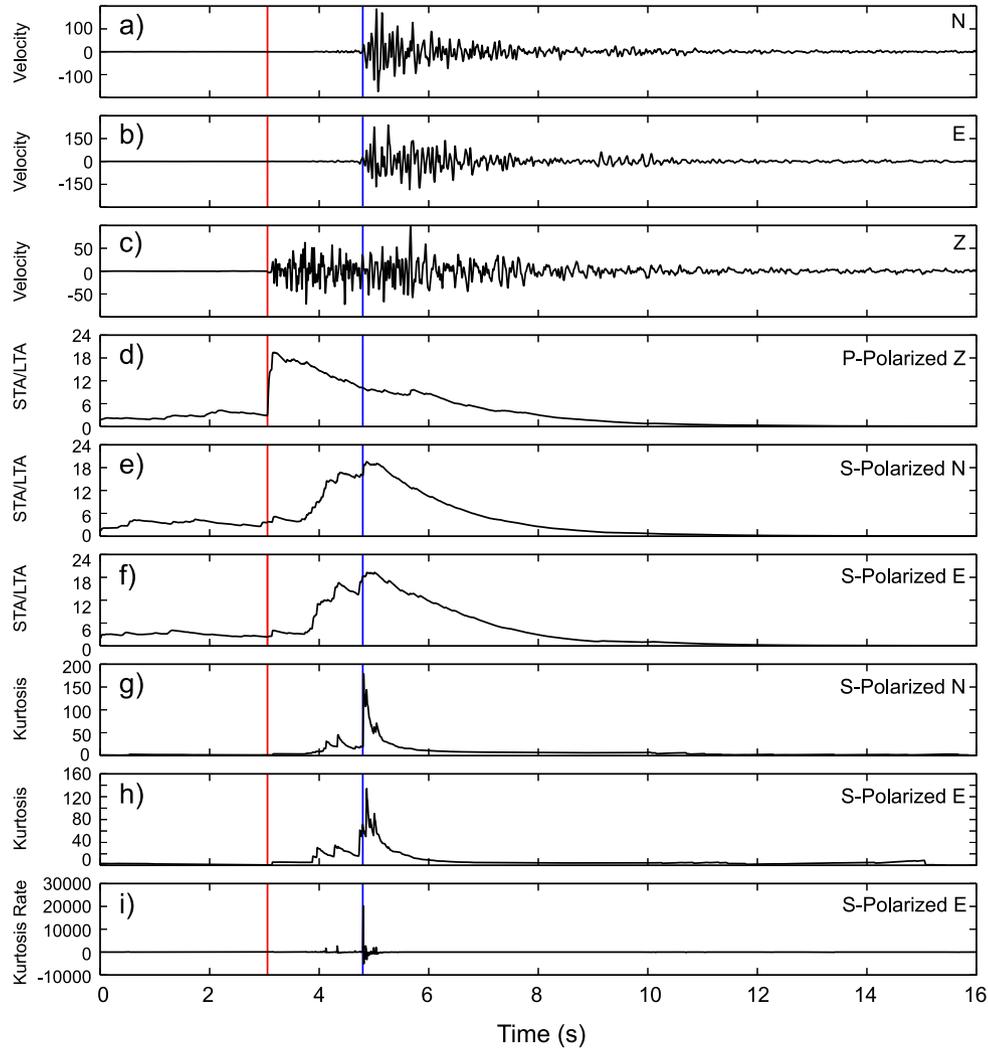


Figure 3. Demonstration of the *P* and *S* picking algorithm for the event denoted by red star in Fig. 1 and recorded a MMNB. The red and blue bars denote the automatic *P*- and *S*-wave picks, respectively. (a) North component velocity seismogram. (b) East component velocity seismogram. (c) Vertical component velocity seismogram. (d) STA/LTA function calculated on *P*-polarized vertical component. (e) STA/LTA function calculated on *S*-polarized *N* component. (f) STA/LTA function calculated on *S*-polarized *E* component. (g) Kurtosis function calculated on *S*-polarized *N* component. (h) Kurtosis function calculated on *S*-polarized *E* component. (i) Time derivative of kurtosis function calculated on *S*-polarized *E* component.

not have an appreciable effect on the results. If the polarization filters are working well, the *S*-polarized horizontal traces will have minor *P*-wave energy remaining (as in Fig. 2). In such cases, the STA/LTA detectors will primarily recognize the *S*-wave and can be used to make an initial pick. This is done on the *S*-polarized *N* and *E* components independently by making a pick at the location of the maximum of the CF. As seen in Figs 3(e)–(f), the *S* arrival is not clearly defined on the STA/LTA detectors, and we found that using the peak values works best at locating the initial picking region.

The STA/LTA detector is very effective at recognizing changes in amplitude, but less so at accurately identifying phase arrivals. For this purpose, we refine the initial *S* picks with a moving kurtosis function

$$K = \frac{\sum_{i=1}^M (x_i - \bar{x})^4}{(M-1)s_x^4} - 3, \quad (6)$$

where M is the number of samples, $x = \{x_1, x_2, \dots, x_M\}$ is the finite length *P*-wave sequence, \bar{x} is the sample mean, and s_x is the standard deviation. The kurtosis function measures peakedness and is useful for identifying abrupt signal changes such as those produced in

seismic records by earthquake phase arrivals. We calculate the kurtosis for a sliding window of 5 s duration on each of the *S*-polarized horizontal traces (Figs 3g and h), and then calculate the derivative of each (Fig. 3i). We tested sliding window durations in the range of 1–10 s and found that 5 s typically was the most accurate, but changing the window length somewhat around this value does not have a significant effect on the results. For each component, we search for the peak of the derivative in a 0.5 s window around the initial STA/LTA *S* pick and tentatively move the pick to this location. The derivative is used over the kurtosis because it helps to localize the arrival better. Since the kurtosis function is slightly delayed in time relative to the motion onset, we look for a local minimum in the kurtosis function preceding the tentative pick location by no more than 0.25 s. If multiple minima are present in this time window, the one with the smallest value is taken. The 0.25 s value was chosen to allow a small amount of flexibility in refining the pick, without moving it too far. If one is found, the pick is updated to the minima's location (blue vertical line), and if not the pick remains unchanged.

In some cases, the *P* and *S* phases do not separate well and considerable energy for each phase remains in the polarized traces.

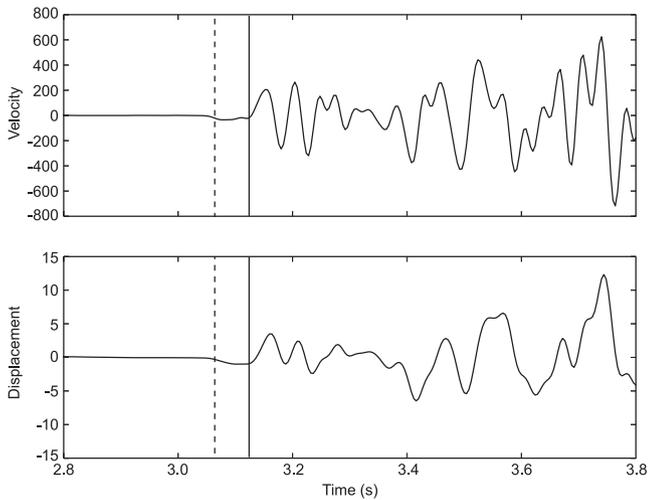


Figure 4. Close-up of vertical component velocity and displacement seismograms for the data used in Fig. 3. Dashed line indicates automatic P -wave pick, while solid line indicates analyst pick of the direct P wave. The automatic picker locks on a FZHW instead of the direct P wave that is delayed by ~ 0.1 s.

This can lead to the P -wave arrival being recognized strongly by the STA/LTA detectors on the N and E components, which can cause picking problems. To suppress such cases we check that the P and S picks are more than 0.3 s apart and discard any S picks which do not satisfy this requirement. The separation of P and S phases and quality of the S picks are generally better at stations on the faster side of the fault, without the additional complexity of fault zone waves that are present at near-fault stations on the slow side.

To demonstrate the accuracy of the technique in picking phase arrivals, the automatic picks should be compared quantitatively with manual ones. Towards this end, S -wave arrivals were picked manually for 196 events and the S -picking algorithm was run on the same set. These events were chosen because they have P , S and FZHW present in each (the P pick accuracy is discussed separately in the next section where the topic is treated in detail). The S arrivals were picked on each of the horizontal components independently. The median absolute time difference between automatic and manual picks of the S waves was 0.17 s, with a standard deviation of 0.72 s. Approximately 80 per cent of the automatic picks were within 0.5 s of the manual pick.

When FZHW are present they are likely to be misidentified by the STA/LTA and kurtosis pickers as direct P arrivals. Indeed, a closer examination of the first arrival picked in Fig. 3 indicates that it is a FZHW followed by a direct P arrival that is delayed by about 0.1 s (Fig. 4). This problem was noted by the early observational papers on FZHW (e.g. Ben-Zion & Malin 1991) and was demonstrated clearly by Allam *et al.* (2014) in the context of the Hayward fault; the P -wave picks of the Northern California Earthquake Data Center at near-fault stations on the slow side of the Hayward fault are commonly FZHW rather than direct P waves (figs 2 and 6 of Allam *et al.* 2014). Developing an automatic picking algorithm capable of distinguishing between these two phases can be very useful for processing seismic data recorded near large faults. This is described in the next section.

4 PICKING FZHW

FZHW are emergent phases that propagate along fault bimaterial interfaces with first motion polarity of the faster velocity medium,

and are radiated from the interface to the slow side of the fault (Ben-Zion 1989, 1990). They are the first arrivals at stations on the slower velocity medium located within a critical distance x_c from the fault given by (Ben-Zion & Malin 1991),

$$x_c = r \cdot \tan[\cos^{-1}(\alpha_s/\alpha_f)], \quad (7)$$

where r is the distance the FZHW travels along the fault and α_f , α_s are the average P -wave velocities of the faster and slower media, respectively. Several characteristics distinguish direct P body waves from FZHW: they are more impulsive, have opposite first motion polarity, and are usually larger in amplitude. However, since the amplitude of the FZHW may be significantly larger than the ambient noise (e.g. McGuire & Ben-Zion 2005; results below), they are routinely picked erroneously by automatic algorithms for detection of P body waves. Our algorithm for detection and picking of FZHW and direct P arrivals exploits the different characteristics of these phases, and the procedure is applied with several stages of quality control to minimize false detections. The algorithm for separating between and picking FZHW and direct P phases only needs one component seismograms and below we apply it to the vertical component of the unpolarized seismograms. This is because the polarization filters of the previous section can potentially change the character of the waveforms, leading to less accurate picks of FZHW and direct P arrivals. The initial stage of P -wave picking is used in our algorithm only to find the relevant portion of the seismogram for detailed analysis of direct and possible FZHW arrivals.

The first stage of the FZHW picking algorithm involves making a pick on the vertical component for the earliest onset of the seismic motion over the noise. This pick should lock on the head wave arrival if one is present or the P -wave arrival otherwise. It is performed using an STA/LTA detector with an STA length of 0.1 s, LTA length of 10 s, and trigger level of 4. These values were found to produce good results with the Parkfield data (see next section). The STA/LTA detector is used at this stage because it is more sensitive to the emergent onset of the FZHW than the kurtosis function. An example of this process applied to the data of Fig. 3 is shown in Fig. 5. To demonstrate the process we use raw data from station MMNB on the slow side with no bandpass filtering. Figs 5(a) and (b) show the raw velocity and displacement seismograms for the used event, respectively. The STA/LTA function calculated on the velocity seismogram is shown in Fig. 5(c) and the first-motion pick is indicated by a red vertical line.

We now attempt to specifically pick a direct P -wave arrival. This is achieved using two additional pickers in unison to find the sharpest arrival in the early portion of the seismogram. The first of these pickers uses the kurtosis function (6), and the second picker uses a moving skewness function (Saragiotis *et al.* 2002),

$$S = \frac{\sum_{i=1}^M (x_i - \bar{x})^3}{(M-1)s_x^3}, \quad (8)$$

where the various symbols are the same as in (6). The skewness measures statistical asymmetry and in particular indicates whether the sample is left-skewed or right-skewed. Both the kurtosis and skewness are quite sensitive to abrupt changes in the character of a time-series, which should occur more strongly at the P -wave arrival than for an emergent FZHW. We calculate both of these statistical functions with a 5 s sliding window (Figs 5d and e), and find the peak of each in the vicinity of the examined portion of the seismogram. The value of 5 s is used because it is half the duration of the LTA in the first-motion pick (using windows of 3–7 s lead to similar results). We then calculate the derivative of each function within one second to the left of the peak's location, and take the position of the

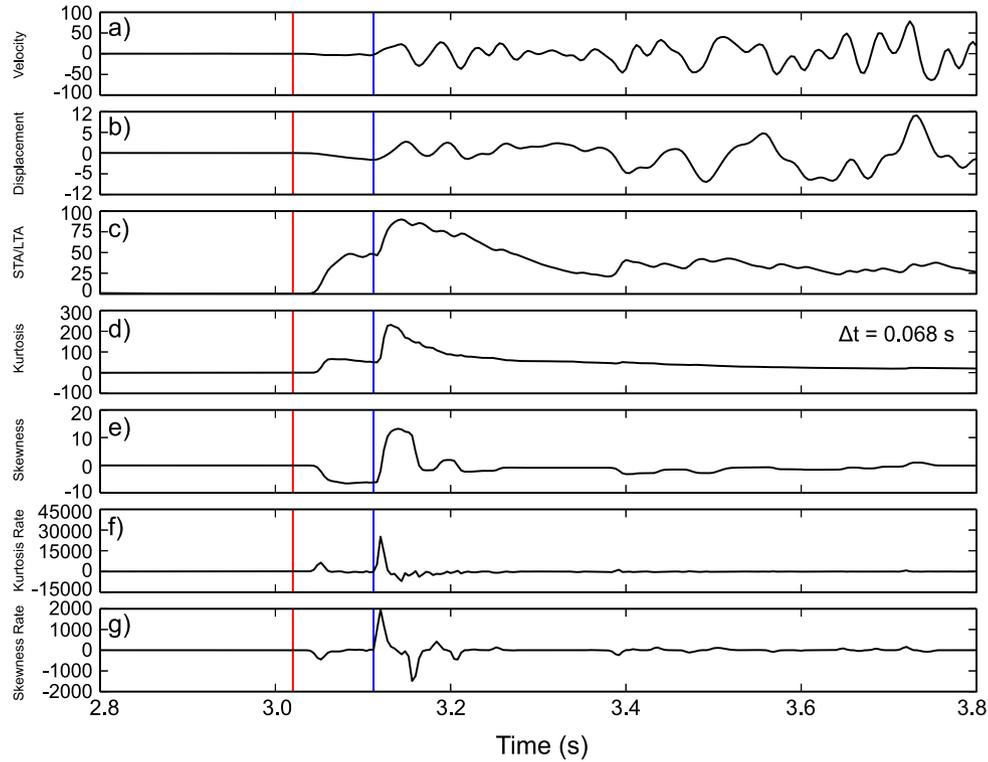


Figure 5. Demonstration of the FZHW and direct *P* picking algorithm for the event denoted by red star in Fig. 1 and recorded at MMNB. The red and blue lines indicate the automatic FZHW and direct *P* picks, respectively. No filtering is used in the algorithm for this example. The time difference between the FZHW pick and *P*-wave pick is 0.068 s. (a) Vertical component velocity seismogram. (b) Vertical component displacement seismogram with 0.5 Hz high-pass filter for plotting purposes only. (c) STA/LTA function calculated from (a). (d) Kurtosis function calculated from (a). (e) Skewness function calculated from (a). Note the reversed polarities between the head and *P* waves, indicated most clearly by the skewness. (f) Time derivative of kurtosis function calculated from (d). (g) Time derivative of skewness function calculated from (e).

maximum derivative as the pick of the direct *P* wave (Figs 5f and g). The derivative is used rather than the kurtosis/skewness functions themselves to indicate where these functions change most rapidly. The one second window restricts the pick to the vicinity of the seismogram onset; the precise value of the window does not have an appreciable effect on the results.

There may be various circumstances not associated with FZHW leading to several distinct phases in the early portion of seismograms (e.g. free-surface reflections at borehole instruments, arrivals from different earthquakes, etc.). To distinguish such early phases from FZHW we use the property that a FZHW has opposite first motion polarity from the direct *P* wave (Ben-Zion 1989, 1990). We include this constraint in the algorithm by examining the polarity of the calculated skewness function at both the first motion pick and the later skewness pick. The skewness function provides a reliable indicator of motion polarity that is insensitive to small fluctuations. In many cases it is more stable than measuring the polarities from the velocity or displacement seismograms, because it only recognizes polarity changes when the moving window fully changes its statistical asymmetry. The algorithm requires further that the skewness does not change polarity other than near the *P*-wave arrival. This is because a genuine FZHW should maintain the same polarity until the direct *P*-wave arrives.

As an example, the skewness in Fig. 5(e) has only a negative polarity during the portion associated with FZHW, while it has only a positive polarity during the *P* wave. In some cases, the direct *P* pick may be a few time samples earlier than the polarity reversal in the skewness function, and we account for this by looking for a nearby zero crossing within half the rise time around the pick (here about

0.02 s). The half-rise time is approximated as the time between the pick (where the derivative is greatest) and the peak of the CF. If a polarity reversal is present, we measure the polarity immediately after the zero crossing. Otherwise, the polarity is measured at the direct *P* pick. We then require that the polarities of the tentative FZHW and direct *P* arrival are opposite.

Next we calculate the time difference between each tentative pick of the direct *P* wave and the first motion pick, and compare these values to basic theoretical expectations for FZHW. If no head wave is present and there are no unusual early arrivals (e.g. prominent free-surface reflections), the differences between the picks should be near zero. For the example in Fig. 5, the obtained time difference is around 0.07 s for both the kurtosis and skewness signals. In a simple model consisting of two different homogeneous quarter spaces (Ben-Zion 1989), the FZHW traveltime is

$$t_H = \frac{r}{\alpha_f} + \sqrt{\alpha_s^{-2} - \alpha_f^{-2}}x, \quad (9)$$

where x is the perpendicular distance from the fault to the station and $r = \sqrt{R^2 - x^2}$ is the component of the source–receiver separation along the fault with R being the hypocentral distance. The traveltime for the direct *P* wave is

$$t_D = \frac{R}{\alpha_s}. \quad (10)$$

From (9) and (10), we can calculate the time difference between FZHW and direct *P* arrivals as

$$\Delta t = t_H - t_D. \quad (11)$$

We use (11) in conjunction with assumed P -wave velocities of the media across the fault to determine maximum allowed time differences between FZHW and direct P arrivals at different stations. For the Parkfield area we assume $\alpha_f = 5.5 \text{ km s}^{-1}$ and $\alpha_s = 4.95 \text{ km s}^{-1}$, associated with a velocity contrast of 10 per cent, which is at the upper end of imaging results for the region (e.g. Ben-Zion *et al.* 1992; Zhao *et al.* 2010; Bennington *et al.* 2013). A minimum allowed time difference is defined from the approximate length for two swings of a direct P wave, estimated here to be 0.065 s. This value is an effective resolution limit of the algorithm where a FZHW is no longer distinguishable from two swings of a direct P wave. Reducing or eliminating the minimum time would increase the number of detections, while increasing also the number of false FZHW detections. If source locations are unknown for data of interest, it is possible to simply replace the hypocentral distance with a maximum value expected for sources in the region. This will act to increase the values given by eq. (11), leading generally to the same number of correct picks and also more erroneous picks. Similarly, when analysing data of fault zone arrays or near-fault stations without precise information on the distance from the fault, it is best to set a fixed (small) distance from the fault for all stations (e.g. 0.25 km), which will increase the values estimated by eq. (11). Applications to other regions, or efforts to maximize the possible detections, may require changing the assumed velocity values and minimum allowed time separation.

To continue with the next stage of the algorithm, we require the time difference for the direct P picks based on the kurtosis and skewness to be within the range associated with the predicted difference for the FZHW. To ensure that the direct P picks based on the kurtosis and skewness are in good agreement, we further require that both are locked on the same portion of the seismogram by allowing them to be separated in time by no more than 0.03 s. The value was chosen to correspond roughly to one swing of a P wave. This restriction decreases the likelihood that the direct P picks are in the allowable range by chance, and additionally gives confidence to the picks as recognizing a genuine phase arrival. If the picks are separated in time by more than 0.03 s, it essentially indicates that the arrival is not well defined. The time value of 0.03 s was chosen based on inspection of the rise time in the kurtosis/skewness functions for numerous local earthquakes, and testing indicates that a majority of events will fall within this range. This is primarily used to suppress outliers at this stage. If all requirements have been satisfied, the first-motion pick is identified as a FZHW and the direct P pick is taken as the average of the kurtosis and skewness picks (blue vertical line, Fig. 5). Otherwise, the first motion pick is identified as a direct P pick, and the later P picks are discarded.

The algorithm up to this point focused on robust ‘identification’ of phases using finite-size time windows, leading to picks that are slightly later in time than the arrival onsets. In the final stage of the algorithm we attempt to refine the picks to lock on more sharply defined time points. For the first-motion pick, we start at its current value and move backward in time until an STA/LTA value of 2 is reached, at which point the pick is moved to that location. This is an attempt to get the pick as close to its departure from the noise level as possible, as STA/LTA is effectively a measure of signal-to-noise ratio. If the first-motion phase was flagged as a FZHW candidate, we also refine the later direct P pick by returning to the original picks made on the kurtosis and skewness. These picks were made at the peak of the derivative of the respective CF, which is roughly half-way between when the CF begins to increase and the maximum of the CF. The time between where the derivative is maximum and the CF is maximum is approximated as half the rise time. We search

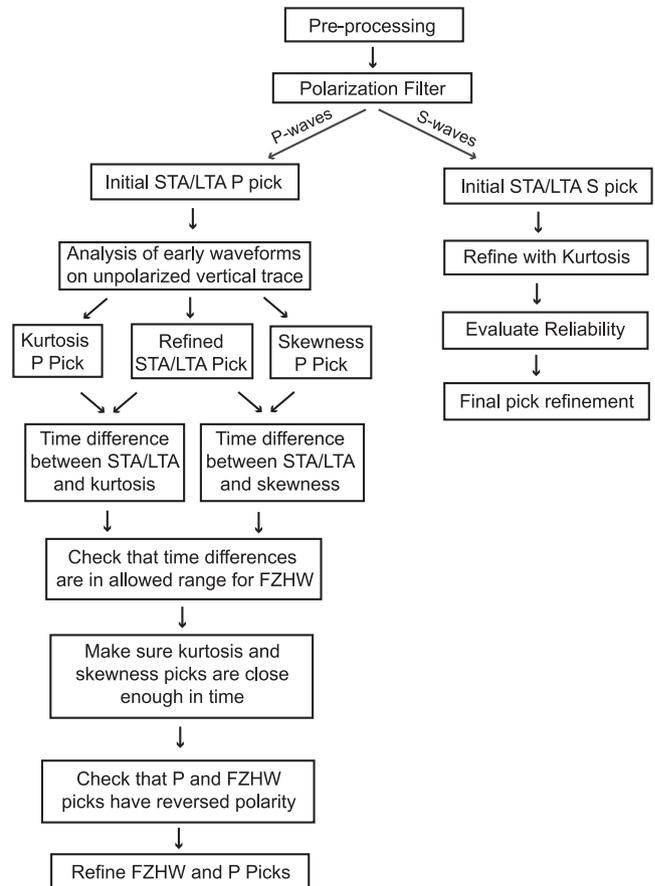


Figure 6. A flow chart describing the algorithm for identification and picking of P , S and FZHW phases.

for a local minimum on the kurtosis function that precedes the skewness polarity reversal by no more than the rise time. If one is found, the kurtosis pick is revised to this value and otherwise it remains unchanged. The skewness may have a local minimum or maximum preceding the polarity reversal as the polarity of the arrival is variable. Therefore a local extremum is searched for on the skewness function that precedes the polarity reversal by no more than the rise time. If one is found, the pick is updated to the location of the extremum; otherwise the pick remains unchanged. For the kurtosis function, if multiple minima are present then the one with the smallest value is taken. The skewness function has two possible cases because the polarity can be both positive and negative. If the polarity is positive, multiple minima are checked for and if present the one with the smallest value is taken. Similarly, if the polarity is negative, multiple maxima are checked for and if present the largest is taken. The final direct P pick is set to the average of the new kurtosis and skewness picks, and at this point the algorithm comes to an end. The various stages of the algorithm are summarized by a flow chart in Fig. 6.

To verify that the algorithm correctly identifies and correctly picks both FZHW and direct P arrivals, we test it on synthetic seismograms. The seismograms are calculated using the analytical solution of Ben-Zion & Aki (1990) for the scalar wave equation with a line dislocation source in a structure consisting of a 100-m-wide vertical low velocity zone between two different quarter spaces (Fig. 7). The results correspond to acoustic P or SH waves depending on the assumed material properties. Here we calculate P waveforms using for the three media, denoted by subscripts

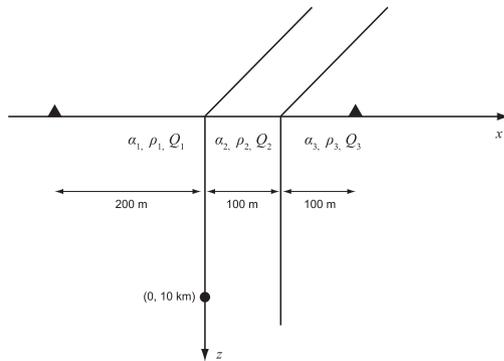


Figure 7. A vertical low velocity layer between two different quarter spaces with a line dislocation source on the fault (circle). Synthetic seismograms are calculated at the two receivers (triangles) on the opposite sides of the fault with the solution of Ben-Zion and Aki (1990). See text for media parameters.

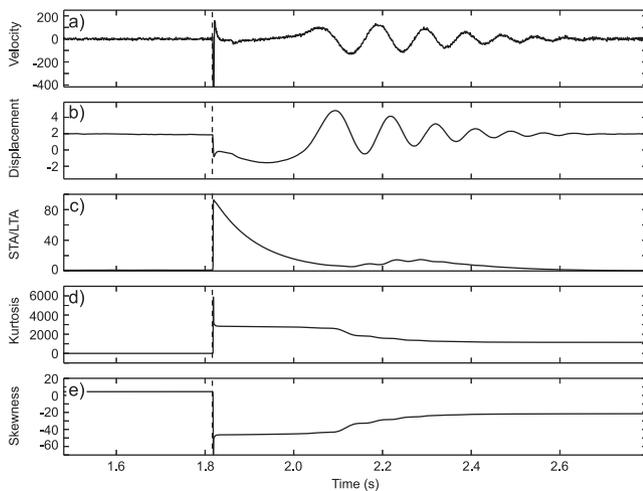


Figure 8. Testing the FZHW and direct P pickers on synthetic seismograms on the faster side of the fault, where only direct P and trapped waves are present. All picks made by the algorithm lock on the direct P wave (dashed line) and no FZHW is detected. (a) Synthetic velocity seismogram. (b) Displacement seismogram. (c) STA/LTA function calculated on (a). (d) Kurtosis function calculated on (a). (e) Skewness function calculated on (a).

1, 2, 3 from left to right, a common mass density of 2500 kg m^{-3} and the following wave velocities and attenuation factors: $\alpha_1 = 5500 \text{ km s}^{-1}$, $\alpha_2 = 3700 \text{ km s}^{-1}$, $\alpha_3 = 5000 \text{ km s}^{-1}$, $Q_1 = Q_3 = 1000$ and $Q_2 = 30$. The source is located at a depth of 10 km on the interface between the left quarter space and fault zone layer (Fig. 7). In the examples below, velocity and displacement seismograms are calculated at two receivers located on the opposite sides of the fault zone, at epicentral distances of 200 m from the overall (left-hand panel) velocity contrast interface. A small amount of Gaussian noise is added to the synthetic seismograms to produce more realistic records with small fluctuations mimicking ambient noise between the main arrivals.

Figs 8(a) and (b) show synthetic velocity and displacement seismograms calculated at a receiver on the faster quarter-space 200 m from the left interface. In this case the basic phases present in the seismograms are a direct P wave and trapped-like waves resulting from the low velocity layer. The algorithm correctly makes only a direct P pick (dashed line) within 0.01 s of the true arrival time. The three CF of the algorithm based on STA/LTA, kurtosis and skewness are given in Figs 8(c)–(e), respectively. Fig. 9 shows corresponding

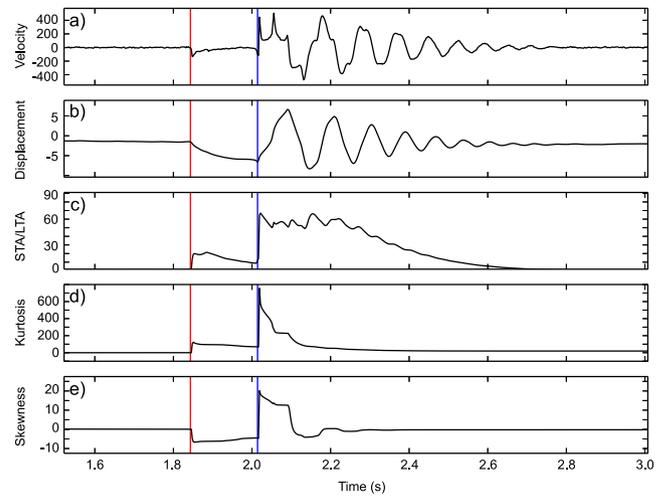


Figure 9. Testing the FZHW and direct P pickers on synthetic seismograms on the slower side of the fault, where a FZHW is present in addition to direct P and trapped waves. The red and blue lines indicate the picked FZHW and direct P wave, respectively. (a) Synthetic velocity seismogram. (b) Displacement seismogram. (c) STA/LTA function calculated on (a). (d) Kurtosis function calculated on (a). (e) Skewness function calculated on (a).

results associated with seismograms at a receiver on slower quarter-space 200 m from the left interface. In this case the first ballistic phase is a FZHW, followed by a direct P wave, fault zone reflected phases and trapped-like waves. The algorithm correctly identifies both the FZHW and direct P wave and makes picks (red and blue vertical bars) that are within 0.01 s of the true arrivals. We tested the algorithm on numerous additional synthetic seismograms, using cases that have an additional transition layer between the two quarter spaces (Ben-Zion 1998) and obtained similar reliable results as those shown in Figs 8 and 9.

5 SYSTEMATIC TESTS ON PARKFIELD DATA

To perform stronger tests on the performance of the algorithm we apply it to seismograms of ~ 2500 earthquakes recorded by five HRSN stations at the Parkfield section of the San Andreas fault (Fig. 1). The San Andreas fault near Parkfield provides a good testing ground, because it has both a prominent velocity contrast and local complexities that may include a reversal in the sense of the contrast (e.g. Eberhart-Phillips & Michael 1993; Thurber *et al.* 2006). It is known from Ben-Zion & Malin (1991) and Zhao *et al.* (2010) that stations MMNB and EADB on the generally slower NE block record commonly FZHW over a wide range of distances, while at stations FROB and VCAB on the generally faster SW block FZHW have not been observed. Station Ghib on the NE block is located above the hypocentre location of the 2004 M_6 event associated with a local reversal of the velocity contrast (Bennington *et al.* 2013), and does not record head waves from events in that portion of the seismogenic zone (Zhao *et al.* 2010).

We find that considerably more FZHW are identified and picked by the algorithm on the generally slow NE side of the fault than the opposite SW side. In particular, 10.8 per cent of the examined events recorded at MMNB are flagged as having candidates for FZHW, while the detection level at VCAB is only 0.2 per cent (54 times fewer). Fig. 5 illustrated the automatic detection and picking of FZHW at MMNB using data generated by an event close to and below the station (red star, Fig. 1). Fig. 10 provides another

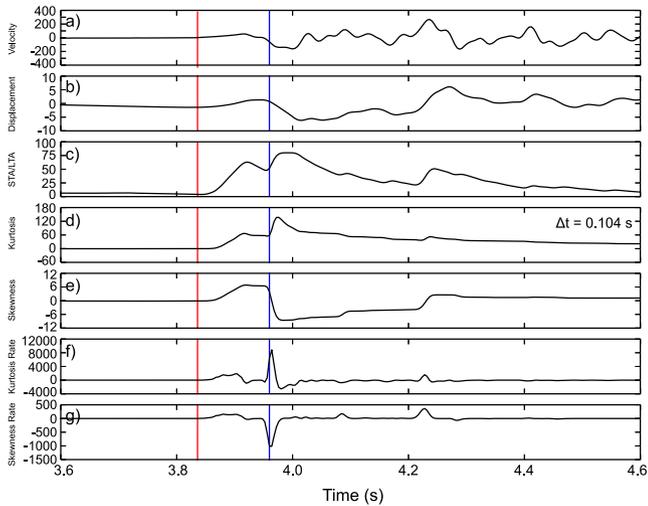


Figure 10. Illustration of the FZHW and direct P picking algorithm for the event denoted by yellow star in Fig. 1 and recorded at MMNB. The red and blue lines indicate the automatic FZHW and direct P picks, respectively. (a) Vertical component velocity seismogram. (b) Vertical component displacement seismogram (c) STA/LTA function calculated from (a). (d) Kurtosis function calculated from (a). (e) Skewness function calculated from (a). (f) Time derivative of kurtosis function calculated from (d). (g) Time derivative of skewness function calculated from (e).

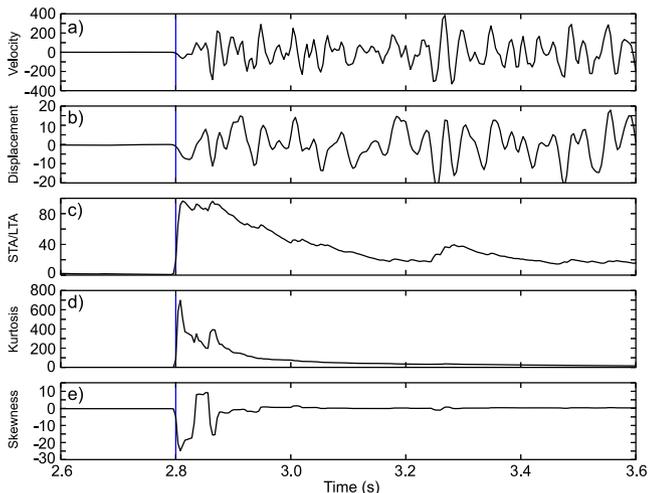


Figure 11. Corresponding results to Fig. 5 generated by the event denoted by red star in Fig. 1 and recorded at station VCAB. Only a direct P wave is picked by the algorithm (blue line). (a) Vertical component velocity seismogram. (b) Vertical component displacement seismogram (c) STA/LTA function calculated from (a). (d) Kurtosis function calculated from (a). Note the single immediate peak after the motion onset. (e) Skewness function calculated from (a).

example of automatic detection and picking of FZHW at MMNB using data generated by an event to the NW of the station (yellow star, Fig. 1). The seismograms recorded at VCAB and FROB for the events used in Figs 5 and 10, are shown in Figs 11 and 12, respectively. As seen in these examples, the recordings at FROB and VCAB typically have a single immediate peak in the kurtosis and skewness functions, while the recordings at MMNB often have two peaks that are about 0.1 s apart or more. We also note that these events have right lateral strike-slip focal mechanisms (Thurber *et al.* 2006), but they produce the same first motion polarity on both sides of the fault as Fig. expected for FZHW.

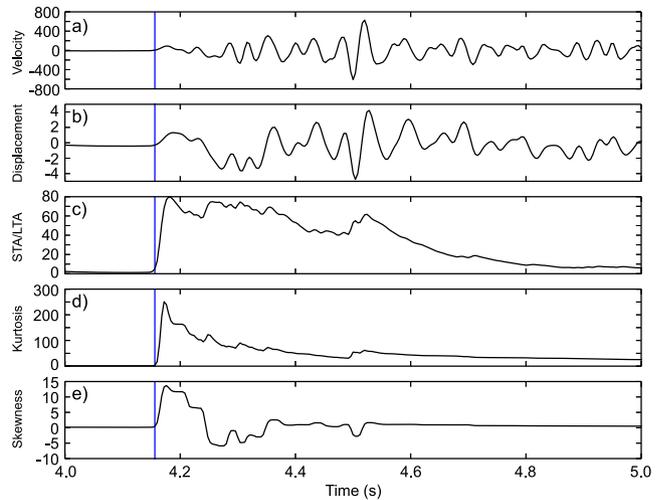


Figure 12. Corresponding results to Fig. 9 generated by the event denoted by yellow star in Fig. 1 and recorded at station FROB. Only a direct P wave is picked by the algorithm (blue line). (b) Vertical component displacement seismogram (c) STA/LTA function calculated from (a). (d) Kurtosis function calculated from (a). Note the single immediate peak after the motion onset. (e) Skewness function calculated from (a).

Fig. 13 summarizes the detection of FZHW at the five employed HRSN stations, using corresponding colours for stations and FZHW detections, in a map view (Fig. 13a) and vertical cross section (Fig. 13b). For station FROB on the generally fast SW side of the fault, the algorithm flags candidate FZHW in 1.6 per cent of the events. For stations GHIB and EADB on the nominally slow NE side, the FZHW detection rates are 9.1 and 4.6 per cent, respectively. We recall that the algorithm parameters (e.g. minimum and maximum time separation between FZHW and direct P arrivals) are chosen to reduce the number of false detections rather than maximize the detection rate. As seen in Fig. 13, the stations NE of the fault detect candidate FZHW (red, orange and yellow circles) along most of the fault, with notable exceptions in the region below and around station GHIB having a local reversal of the velocity contrast (Thurber *et al.* 2006; Bennington *et al.* 2013) and a region of structural complexity to the NW of station EADB. The locations of automatic FZHW detections within the seismicity patches are very consistent with the manual results of Zhao *et al.* (2010). The blue and green symbols, signifying FZHW detections by stations VCAB and FROB on the SW of the fault, are few in number and are not found to have FZHW upon closer inspection. Fig. 14 provides examples of false detections at these stations as well as at the three stations NE of the fault. As mentioned, the false detections are associated with spurious phases such as reflections from the free surface or other early arrivals that are reminiscent of FZHW. They may be identified by checking that the first motion polarities at different stations are consistent with expectations for direct body waves (i.e. opposite polarities across the fault) and lack of other characteristic features of FZHW.

As with the S picks, we test the accuracy of the technique in picking P and FZHW arrivals by comparing the automatic picks with manual ones. The same data set of 196 events was used, and each event was manually selected to ensure it had a FZHW as the first arrival. P and FZHW arrivals were picked manually on the vertical component. The median absolute time difference between automatic and manual FZHW picks was found to be 0.016 s, with a standard deviation of 0.035 s. For the P picks, the median absolute

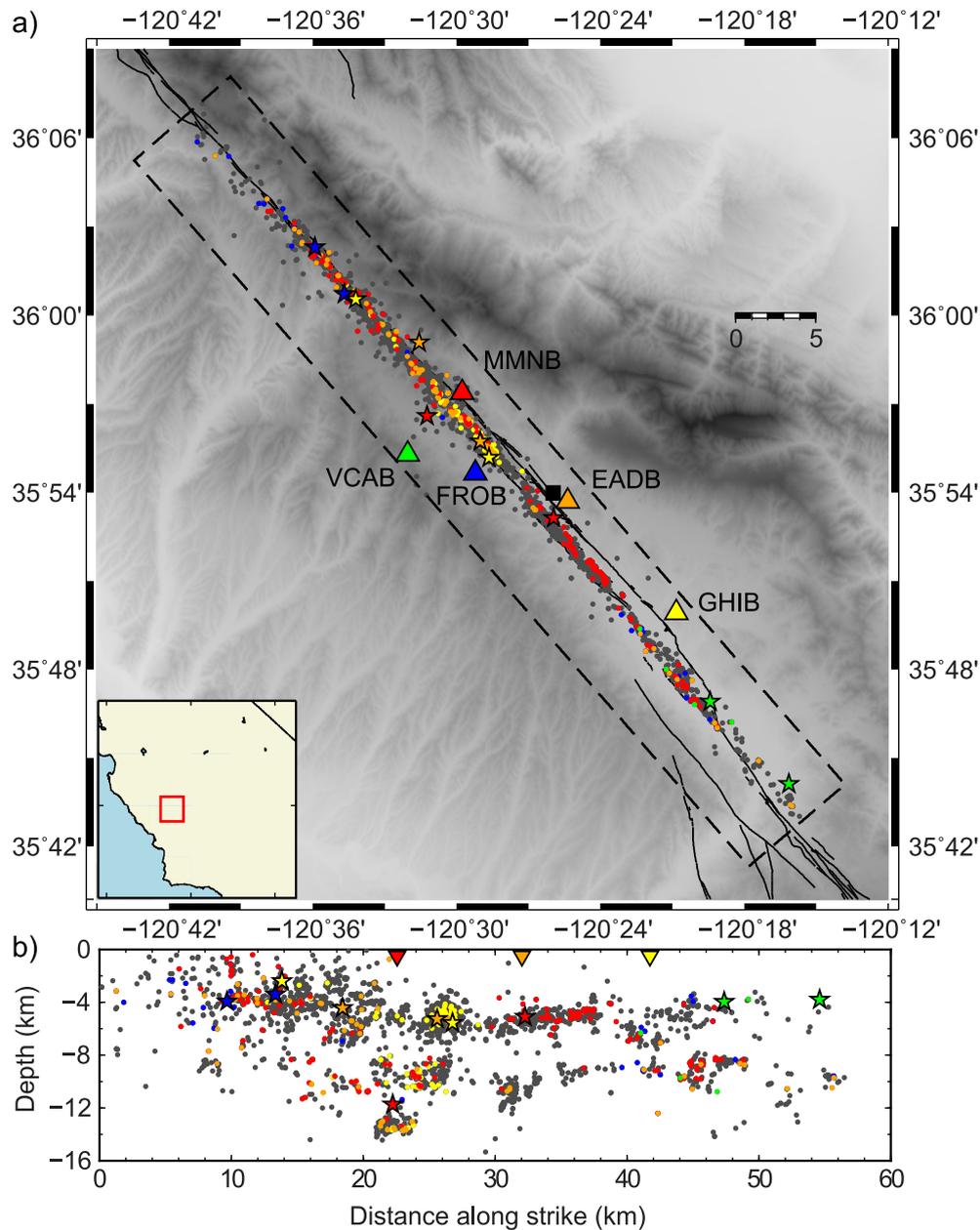


Figure 13. Summary of automatic detections for the 5 HRSN stations (coloured triangles). Events producing detections of FZHW at the various stations are marked by corresponding colours, while the other events are denoted by grey. (a) Results in a map view. (b) Results projected on a vertical plane. The stars denote events producing false detections shown in Fig. 14.

time difference between automatic and manual picks was found to be 0.004 s with a standard deviation of 0.023 s.

6 DISCUSSION

Reliable techniques for automatic identification and picking of seismic phases are becoming increasingly important for analysing large seismic data sets. We have developed a set of algorithms in this study which have the ability to accurately identify and pick arrivals of direct *P* and *S* body waves, as well as early arriving FZHW (Fig. 6). The algorithms utilize CFs based on STA/LTA, kurtosis, and skewness of waveforms in moving time windows. The *S*-wave picking algorithm uses polarization analysis of three-component waveforms to considerably reduce *P*-wave energy from the seismo-

grams, so that the employed detectors lock on the *S* arrival (Figs 2 and 3). The pick of the *S* phase itself is made in two stages so that the arrival is targeted more accurately.

The direct *P* wave and FZHW picking algorithm analyses the early portions of vertical component seismic waveforms. It is easy to analyse all three components, but this was not found to improve the results in the test cases examined here. The algorithm makes use of basic characteristics of the two phases (relative motion polarities, expected time difference, relative sharpness and amplitudes) to distinguish between them in a rigorous series of tests. In doing so, the first arrival is assumed to be a *P* wave unless evidence for a FZHW is provided. This gives more reliability to the FZHW picks, which are not assumed to be present in all recordings as direct *P* waves are. Testing the algorithm on synthetic waveforms indicates that it correctly identifies whether FZHW are present, and accurately

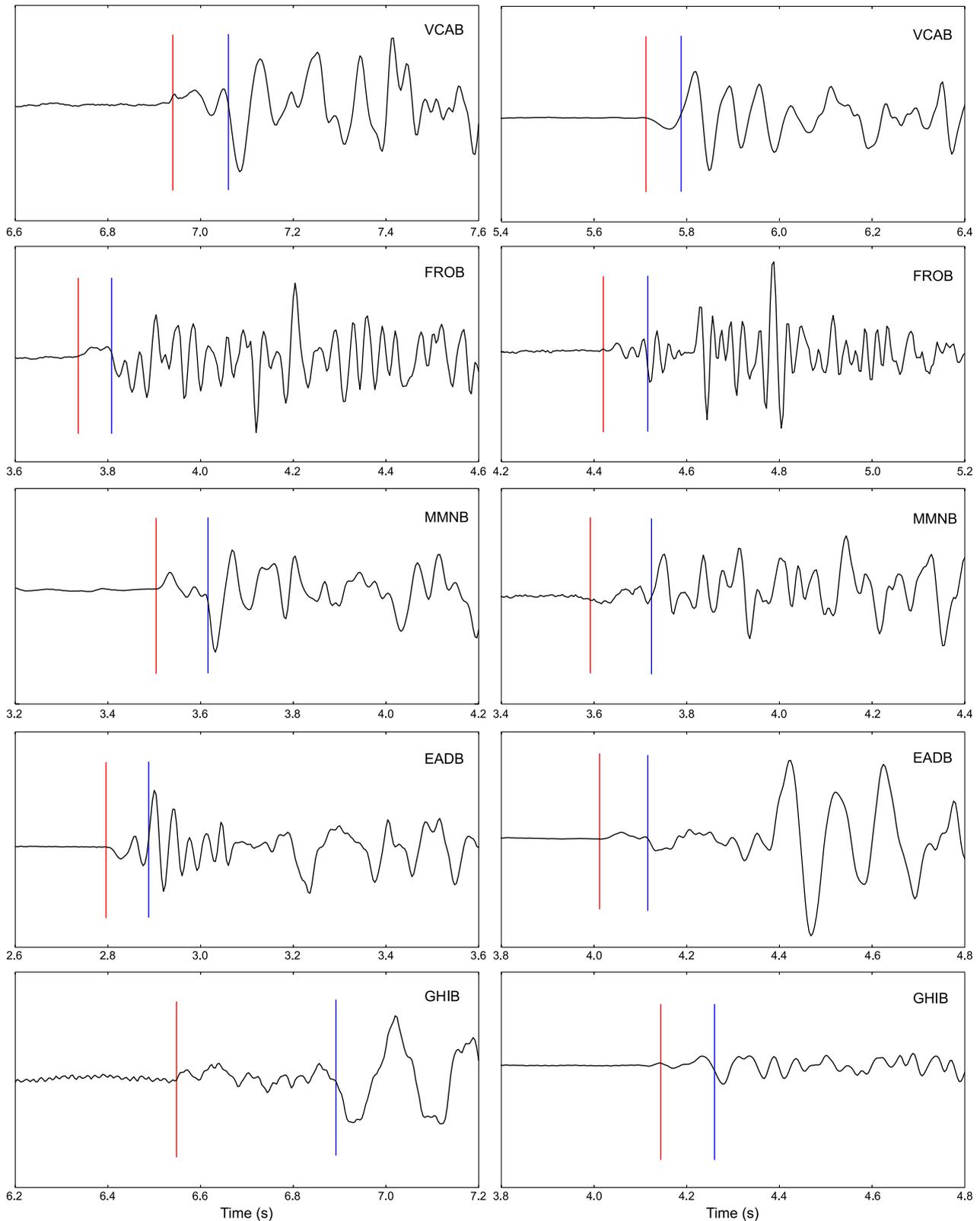


Figure 14. Examples of false detections at each of the stations generated by the events marked with stars in Fig. 13. The recordings are vertical component velocity seismograms. The red and blue lines indicate candidate FZHW and P wave picks, respectively.

picks the arrival times of the FZHW and direct P waves (Figs 8 and 9). Systematic analysis of waveforms recorded by the five employed HRSN Parkfield stations yields results that are consistent with previous studies on the velocity structure and FZHW in the

Parkfield region (e.g. Ben-Zion & Malin 1991; Thurber *et al.* 2006; Zhao *et al.* 2010; Bennington *et al.* 2013). The automatic algorithm identifies FZHW on the overall slow NE side of the fault generated by earthquakes along most of the study area (Fig. 13). Important

exceptions are the section associated with the hypocentre and largest slip of the 2004 *M*₆ Parkfield earthquake and a small additional area of complexity. Over 50 times as many events are identified as producing FZHW at station MMNB close to the fault on the slow side compared to station VCAB on the fast side. The percentage of all events that are identified as having FZHW on the SW side is less than 2 per cent for FROB and VCAB, which is reasonable for automatic processing. The percentage of near-fault events producing FZHW at the stations on the NE side is likely to be higher than the 10.8 per cent observed at MMNB, because the assumed velocity values were designed to provide a reliable set of parameters for automatic picking with relatively small number of false detections. The sensitivity of the algorithm can be increased by increasing the assumed velocity contrast and reducing the minimum allowed time separation between FZHW and direct *P* waves, leading to a greater number of false and true picks.

To test further the performance of the FZHW picking algorithm, we examine data recorded near the Hayward fault. Specifically, we run the algorithm on waveforms generated by a set of 116 events containing FZHW picks based on manual picks made by Allam *et al.* (2014). For stations BKS and CSP on the nominally slow side of the fault, we find that FZHW are detected in 19.5 and 11.9 per cent of the recordings, respectively. For stations CPM and RFSB that are directly across the fault on the fast side, the algorithm (falsely) detects FZHW in 2.3 and 3.4 per cent of the same events, respectively. In this analysis we use during the pre-processing stage a 1 Hz high-pass filter instead of 0.5 Hz, since the data quality at the used stations is not as good as in the HRSN Parkfield stations. From this we conclude that the algorithm can be applied successfully to other data sets with minimal tuning necessary.

The automatic identification and picking of FZHW should be followed by analysts confirmations and possible adjustments. In addition to the basic differences between FZHW and direct *P* waves assumed by the algorithm, there should be indicative patterns associated with analysis of many events. These include systematic relations between minimum propagation distance along the fault and the fault-station distance for head wave detections (e.g. Ben-Zion & Malin 1991; McGuire & Ben-Zion 2005) and systematic moveout between the FZHW and direct *P* arrivals (e.g. Lewis *et al.* 2007; Zhao *et al.* 2010). In addition, since the FZHW are radiated from the fault they should have horizontal particle motion with significant fault-normal component, while the particle motion of the direct *P* waves should point to the epicentre direction (Bulut *et al.* 2012; Allam *et al.* 2014). Similarly, an analyst should confirm and improve if needed the automatic results associated with the picking of the direct *S* waves.

Earthquake locations are generally calculated by comparing observed traveltimes with predicted ones using a given velocity model. The more *S* picks that are available, the better the location accuracy, and this is especially true for events that are near or outside the network. A high percent of *S* picks is also essential for tomographic inversions for V_p/V_s ratios. Misidentification of FZHW as direct *P* waves may produce a bias in arrival times and errors in first motion polarities. If several stations have first arriving FZHW that are erroneously picked as direct *P* waves, there will be a location bias towards the fast side of the fault (e.g. Ben-Zion & Malin 1991). Similarly, using stations with FZHW first arrivals to obtain first-motion polarities will produce errors in focal mechanisms (e.g. McNally & McEvilly 1977).

All the proposed picking algorithms were designed and calibrated specifically for local earthquakes recorded in fault zone environ-

ments, but they may have utility in other circumstances. We found that the *S* picking algorithm can work for regional earthquakes in some cases, but more work is needed to ensure that it performs consistently on these and other types of data. It is likely that some of the parameters used in this work will need to be changed to account for the different waveform character in larger scale source–receiver settings. The head wave picker was designed with the intention of identifying FZHW, and has not been tested on more classical head waves generated by horizontal interfaces such as *P*_n phase. A significant difference, in addition to first motion polarity, is that *P*_n and other such phases are expected in all events recorded beyond a certain distance. On the other hand, the basic operating principle of the FZHW picker is that no head wave is present unless a series of rigorous tests suggest otherwise, so it is not directly translatable to *P*_n type phases. Adaptations of the algorithms for identification of regional phases will be the subject of future work.

The accuracy of the *S*-wave picking algorithm is strongly tied to how well the polarization filter can remove *P*-wave energy. The filter used here was constructed using only the polarization quantities rectilinearity and vertical incidence angle. Additional polarization quantities, such as ratio of transverse energy (e.g. Cichowicz 1993) may help to more effectively separate the phases. In particular, this would be effective for earthquakes with vertical incidence angles that differ from the expected values for *P* and *S*. Including information on the expected radiation patterns of *P* and *S* waves and noise level (e.g. Rastin *et al.* 2013) can also improve the performance of the automatic pickers. Typical event detection scenarios (e.g. Joswig 1995; Gentili & Michelini 2006; Ross & Ben-Zion 2014) use ‘association’ to combine detection information from different stations and/or components before making a decision. Similar types of analysis could be used for the output of the FZHW picker as well as the direct *P* and *S* pickers. For example, this could involve requiring a minimum number of near-by stations to flag the same event before making a given pick. Ultimately, the number and types of additional analyses, as well as the set of algorithm parameters used, will depend on a balance between detection rate, reliability of having correct phase identification, and required accuracy of the picked arrival times.

ACKNOWLEDGEMENTS

The study was supported by the National Science Foundation (grants EAR-0908903 and EAR-1315340). The manuscript benefitted from constructive comments of two anonymous referees.

REFERENCES

- Aki, K. & Richards, P.G., 2002. *Quantitative Seismology*, 2nd edn, University Science Books.
- Allam, A.A., Ben-Zion, Y. & Peng, Z., 2014. Seismic imaging of a bimaterial interface along the Hayward fault, CA, with fault zone head waves and direct *P* arrivals, *Pure. appl. Geophys.*, **171**, doi:10.1007/s00024-014-0784-0.
- Allen, R., 1978. Automatic earthquake recognition and timing from single traces, *Bull. seism. Soc. Am.*, **68**, 1521–1532.
- Ampuero, J.-P. & Ben-Zion, Y., 2008. Cracks, pulses and macroscopic asymmetry of dynamic rupture on a bimaterial interface with velocity-weakening friction, *Geophys. J. Int.*, **173**, 674–692.
- Andrews, D.J. & Ben-Zion, Y., 1997. Wrinkle-like slip pulse on a fault between different materials, *J. geophys. Res.*, **102**, 553–571.
- Baer, M. & Kradolfer, U., 1987. An automatic phase picker for local and teleseismic events, *Bull. seism. Soc. Am.*, **77**, 1437–1445.

- Baillard, C., Crawford, W.C., Ballu, V., Hibert, C. & Mangeney, A., 2014. An automatic kurtosis-based P- and S-phase picker designed for local seismic networks, *Bull. seism. Soc. Am.*, **104**, doi:10.1785/0120120347.
- Bakun, W.H. & Lindh, A.G., 1985. The Parkfield, California earthquake prediction experiment, *Science*, **229**, 619–624.
- Bennington, N.L., Thurber, C., Peng, Z., Zhang, H. & Zhao, P., 2013. Incorporating fault zone head wave and direct wave secondary arrival times into seismic tomography: application at Parkfield, California, *J. geophys. Res.*, **118**, 1008–1014.
- Ben-Zion, Y., 1989. The response of two joined quarter spaces to SH line sources located at the material discontinuity interface, *Geophys. J. Int.*, **98**, 213–222.
- Ben-Zion, Y., 1990. The response of two half spaces to point dislocations at the material interface, *Geophys. J. Int.*, **101**, 507–528.
- 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., 2001. Dynamic rupture in recent models of earthquake faults, *J. Mech. Phys. Solids*, **49**, 2209–2244.
- 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. & Huang, Y., 2002. Dynamic rupture on an interface between a compliant fault zone layer and a stiffer surrounding solid, *J. geophys. Res.*, **107**(B2), doi:10.1029/2001JB000254.
- Ben-Zion, Y. & Malin, P., 1991. San-Andreas fault zone head waves near Parkfield, California, *Science*, **251**, 1592–1594.
- Ben-Zion, Y., Katz, S. & Leary, P., 1992. Joint inversion of fault zone head waves and direct-P arrivals for crustal structure near major faults, *J. geophys. Res.: Solid Earth*, **97**, 1943–1951.
- Brietzke, G.B., Cochard, A. & Igel, H., 2009. Importance of biomaterial interfaces for earthquake dynamics and strong ground motion, *Geophys. J. Int.*, **178**(2), 921–938.
- Bulut, F., Ben-Zion, Y. & Bohnhoff, M., 2012. Evidence for a bimaterial interface along the Mudurnu segment of the North Anatolian Fault Zone from polarization analysis of P waves, *Earth planet. Sci. Lett.*, **327**, 17–22.
- Cichowicz, A., 1993. An Automatic S-phase picker, *Bull. seism. Soc. Am.*, **83**, 180–189.
- Dor, O., Yildirim, C., Rockwell, T.K., Ben-Zion, Y., Emre, O., Sisk, M. & Duman, T.Y., 2008. Geologic and geomorphologic asymmetry across the rupture zones of the 1943 and 1944 earthquakes on the North Anatolian Fault: possible signals for preferred earthquake propagation direction, *Geophys. J. Int.*, **173**, 483–504.
- Eberhart-Phillips, D. & Michael, A., 1993. Three-dimensional velocity structure, seismicity, and fault structure in the Parkfield region, Central California, *J. geophys. Res.: Solid Earth*, **98**, 15 737–15 758.
- Fletcher, J., Haar, L., Vernon, F., Baker, L., Hanks, T. & Brune, J., 1987. The digital array at Anza, California: processing and initial interpretation of source parameters, *J. geophys. Res.*, **92**, 369–382.
- 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.
- Gentili, S. & Michelini, A., 2006. Automatic picking of P and S phases using a neural tree, *J. Seismol.*, **10**, 39–63.
- Hildyard, M.W., Nippress, S.E. & Rietbrock, A., 2008. Event detection and phase picking using a time-domain estimate of predominate period T_{pd}, *Bull. seism. Soc. Am.*, **98**, 3025–3032.
- 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.
- Hori, S., Inoue, H., Fukao, Y. & Ukawa, M., 1985. Seismic detection of the untransformed ‘basaltic’ oceanic crust subducting into the mantle, *Geophys. J. R. astr. Soc.*, **83**, 169–197.
- Hough, S.E., Ben-Zion, Y. & Leary, P., 1994. Fault zone waves observed at the southern Joshua Tree earthquake rupture zone, *Bull. seism. Soc. Am.*, **84**, 761–767.
- Joswig, M., 1995. Automated classification of local earthquake data in the BUG small array, *Geophys. J. Int.*, **120**, 262–286.
- Jurkevics, A., 1988. Polarization analysis of three-component array data, *Bull. seism. Soc. Am.*, **78**, 1725–1743.
- Kuperkoch, L., Meier, T., Lee, J., Friederich, W. & EGELADOS Working Group, 2010. Automated determination of P-phase arrival times at regional and local distances using higher order statistics, *Geophys. J. Int.*, **181**, 1159–1170.
- Kurzon, 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**, doi:10.1785/0120130295.
- 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.
- Lengline, O. & Got, J.L., 2011. Rupture directivity of microearthquake sequences near Parkfield, California, *Geophys. Res. Lett.*, **38**, L08310, doi:10.1029/2011GL047303.
- 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.
- Lewis, M.A., Ben-Zion, Y. & McGuire, J.J., 2007. Imaging the deep structure of the San Andreas Fault south of Hollister with joint analysis of fault zone head and direct P arrivals, *Geophys. J. Int.*, **169**, 1028–1042.
- 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.
- McNally, K.C. & McEvilly, T.V., 1977. Velocity contrast across the San Andreas fault in central California: small-scale variations from P-wave nodal plane distortion, *Bull. seism. Soc. Am.*, **67**, 1565–1576.
- Nippress, S.E.J., Rietbrock, A. & Heath, A.E., 2010. Optimized automatic pickers: application to the ANCORP data set, *Geophys. J. Int.*, **181**, 911–925.
- Oppenheimer, D.H., Reasenber, P.A. & Simpson, R.W., 1988. Fault plane solutions for the 1984 Morgan Hill, California, earthquake sequence: evidence for the state of stress on the Calaveras fault, *J. geophys. Res.*, **93**, 9007–9026.
- Ranjith, K. & Rice, J., 2001. Slip dynamics at an interface between dissimilar materials, *J. Mech. Phys. Solids*, **49**, 341–361.
- Rastin, S.J., Unsworth, C.P., Benites, R. & Gledhill, K.R., 2013. Using real and synthetic waveforms of the Matata Swarm to assess the performance of New Zealand GeoNet phase pickers, *Bull. seism. Soc. Am.*, **103**, 2173–2187.
- Rosenberger, A., 2010. Real-Time ground-motion analysis: distinguishing P and S arrivals in a noisy environment, *Bull. seism. Soc. Am.*, **100**, 1252–1262.
- Ross, Z.E. & Ben-Zion, Y., 2014. An earthquake detection algorithm with pseudo-probabilities of multiple indicators, *Geophys. J. Int.*, **197**, 458–463.
- Rubin, A. & Gillard, D., 2000. Aftershock asymmetry/rupture directivity along central San Andreas fault microearthquakes, *J. geophys. Res.*, **105**, 19 095–19 109.
- Saragiotis, C., Hadjileontiadis, L. & Panas, S., 2002. PAI-S/K: a robust automatic seismic P phase arrival identification scheme, *IEEE Trans. Geosci. Remote Sens.*, **40**, 1395–1404.
- Seeber, L., Armbruster, J.G., Ozer, N., Aktar, M., Baris, S., Okaya, D., Ben-Zion, Y. & Field, E., 2000. The 1999 earthquake sequence along the North Anatolia transform at the juncture between the two main ruptures, in *The 1999 Izmit and Duzce Earthquakes: Preliminary Results*, pp. 209–223, eds Barka, A., Kazaci, O., Akyuz, S. & Altunel, E., Istanbul Technical University.
- Shi, Z. & Ben-Zion, Y., 2006. Dynamic rupture on a bimaterial interface governed by slip-weakening friction, *Geophys. J. Int.*, **165**, 469–484.
- Sleeman, R. & van Eck, T., 1999. Robust automatic P-phase picking: an on-line implementation in the analysis of broadband seismogram recordings, *Phys. Earth planet. Inter.*, **113**, 265–275.
- Thurber, C., Zhang, H., Waldhauser, F., Hardebeck, J., Michael, A. & Eberhart-Phillips, D., 2006. Three-dimensional compressional wavespeed model, earthquake relocations, and focal mechanisms for the Parkfield, California, region, *Bull. seism. Soc. Am.*, **96**, S38–S49.

- Weertman, J., 1980. Unstable slippage across a fault that separates elastic media of different elastic-constants, *J. geophys. Res.*, **85**, 1455–1461.
- Withers, M., Aster, R., Young, C., Beiriger, J., Harris, M., Moore, S. & Trujillo, J., 1998. A comparison of select trigger algorithms for automated global seismic phase and event detection, *Bull. seism. Soc. Am.*, **88**, 95–106.
- Zaliapin, I. & Ben-Zion, Y., 2011. Asymmetric distribution of aftershocks on large faults in California, *Geophys. J. Int.*, **185**, 1288–1304.
- Zhao, P., Peng, Z., Shi, Z., Lewis, M.A. & Ben-Zion, Y., 2010. Variations of the velocity contrast and rupture properties of M6 earthquakes along the Parkfield section of the San Andreas fault, *Geophys. J. Int.*, **180**, 765–780.