Bulletin of the Seismological Society of America, Vol. 106, No. 5, pp. –, October 2016, doi: 10.1785/0120150230

An Improved Algorithm for Real-Time *S*-Wave Picking with Application to the (Augmented) ANZA Network in Southern California

by Z. E. Ross, M. C. White, F. L. Vernon, and Y. Ben-Zion

Abstract We develop an automatic shear-wave picking algorithm suitable for real-time applications as well as with existing databases. The method can scan through packets of continuous waveforms and make picks without prior knowledge of whether earthquakes have occurred. This makes the algorithm suitable for detecting earthquakes at the same time. Expanding upon and improving the method of Ross and Ben-Zion (2014a), the algorithm first uses polarization filters to remove *P*-wave energy from the seismogram. Then, short-term average/long-term average and kurtosis detectors are applied to the data in tandem to lock in on the phase arrival. The method is tested by applying it to a full month of continuous waveform data recorded by a regional network at 123 stations and comparing the resulting automatic picks with 11,353 handmade picks. The automatic picks are found to be within 0.16 s of the analyst picks 75% of the time, and *S* picks are successful 92% of the time that a *P*-wave pick is made. The algorithm is then applied to an entire year of continuous data and detects 11,197 earthquakes. The hypocenters of these earthquakes are, on average, improved by more than 1 km when compared with the regional network's automated catalog.

Introduction

Phase picking algorithms are becoming increasingly valuable in seismology for automated processing of large seismic data sets. Historically, phase arrivals at seismic networks were picked manually by trained analysts, but the present rate of data accumulation worldwide now makes this unfeasible in most circumstances. Automatic picks are critical for many routine calculations made by regional seismic networks, including earthquake detection, hypocenter location, magnitude calculation, and derivation of source mechanisms. High-quality phase picks are necessary for numerous other purposes outside the normal scope of seismic network operations, such as in seismic tomography and analysis of shear-wave anisotropy.

A variety of reliable methods for automatic *P*-wave picking have been available for decades (e.g., Allen, 1982; Sleeman and van Eck, 1999; Saragiotis *et al.*, 2002), but *S*-wave picking algorithms have generally lagged behind in reliability. The majority of these methods operate in the time domain, using characteristic functions such as short-term average/long-term average (STA/LTA) (Allen, 1982) that give the ratio of a short-term moving average to a long-term moving average. A number of recent studies provided additional developments for *S*-wave picking algorithms. Baillard *et al.* (2014) used moving window kurtosis functions to identify transient signal onsets combined with polarization analysis to identify and distinguish *P* waves from *S* waves. Gentili and Michelini (2006) and Wang and Teng (1997) used machine learning algorithms with time-series detectors to identify *S*-wave arrivals. Hildyard *et al.* (2008) and Nippress *et al.* (2010) employed predominant period and/or kurtosis characteristic functions to pick *S* waves. Rawles and Thurber (2015) used a nearest-neighbor and similarity approach for automatic picking of *S* arrivals. Kurzon *et al.* (2014) employed an iterative singular value decomposition (SVD) algorithm to construct polarization filters for separating *P* and *S* energy in seismograms and an SVD-derived function to make *S* picks from the filtered seismograms in real time. Ross and Ben-Zion (2014a, hereafter referred to as RBZ14) utilized a sliding covariance matrix to construct polarization filters for separating *P* and *S* energy and STA/LTA combined with kurtosis detectors in tandem to lock in on a well-defined *S* arrival. The method, denoted as the DBSHEAR algorithm, was shown to perform well on various data sets recorded by regional and local seismic networks (RBZ14; Qiu *et al.*, 2015; Share *et al.*, 2015; Wu *et al.*, 2016) and is developed further in this work.

Many picking algorithms have been designed to work on event-style traces, which are short time segments cut around earthquake signals from the continuous waveform data. This requires that the earthquake was detected previously, such as by a regional network. They often use the earthquake's location to aid in identifying the arrival. These methods cannot

be used to detect new earthquakes and are further unsuitable for use in real time. A real-time picking and detection algorithm must be capable of scanning through continuous waveforms and identifying whether a given wavelet likely contains an earthquake without prior knowledge. Because the majority of recorded seismograms consist of background noise or transient signals not related to earthquakes, a real-time picking algorithm must determine when a seismic phase has likely arrived at the instrument. If too many picks are made on nonearthquake signals in a real-time context, this can ultimately lead to spurious event detections.

In this article, we describe an improved DBSHEAR algorithm for automatic *S*-wave arrival picking in data recorded by local seismic networks. The method is capable of operating in streaming real time on packets of continuous waveform data without prior knowledge about whether an earthquake has occurred. The algorithm is tested on data from the San Jacinto fault zone (SJFZ) in southern California and is shown to significantly improve the number and quality of *S*-wave picks for the real-time catalog. The improved *S*-wave picks lead to more detected events and improved locations.

Methodology

Our technique for automatic *S*-wave picking is based on the DBSHEAR algorithm of RBZ14. The method has three stages that can be summarized as follows: (1) use polarization analysis on a three-component time series to remove *P*-wave energy from the seismogram, (2) use STA/LTA detectors on the resulting polarized horizontal traces to make a trial *S* pick in the vicinity of the *S* wave, and (3) use kurtosis detectors in conjunction with the trial pick to lock on the *S* arrival. One problem with the method is that if the polarization analysis fails to remove enough *P*-wave energy, the trial *S* pick (in step 2) can mistakenly lock on the *P*-wave arrival. In this study, we introduce an alternative method for making the trial *S* pick more reliable, leading to significantly more *S* picks that are also more accurate. The study focuses on the SJFZ in southern California (Fig. 1). We use data from 123 stations around the SJFZ that are archived by the ANZA network (AZ) in southern California, which are referred to as the ANZA+ network for simplicity (see Data and Resources for more details).

We describe the three stages of the algorithm in detail so that the entire method, including the implemented modifications, is explicitly clear. The first stage of the algorithm is designed to identify and remove as much of the *P*-wave energy from the horizontal seismograms as possible. This part of the procedure is unchanged from RBZ14 and uses a polarization filter constructed from particle motion analysis. Figure 2 contains a step-by-step visual demonstration of the algorithm, which is referred to throughout the remainder of this section. In Figure 2a–c, there is an example three-component trace from a local earthquake (star, Fig. 1) recorded at station JFS1 (triangle, Fig. 1). To obtain the polarization filter, a covariance matrix is first calculated from a finite sample of three-component data as

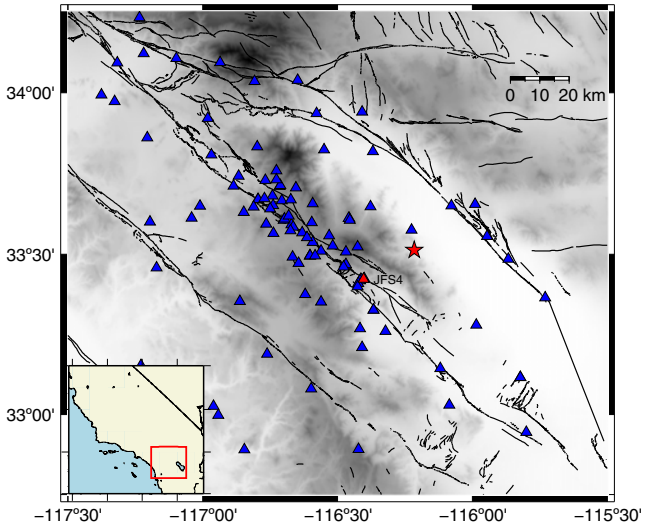

Figure 1. Map of the San Jacinto fault zone (SJFZ) in southern California. A total of 123 stations (triangles) in the region were used by this study, which consist of AZ, YN, PBO, SB, and CI networks (see Data and Resources). The star indicates the hypocentral location of the example event used in Figure 2. Black lines denote the surface expressions of faults in the region. The color version of this figure is available only in the electronic edition.

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

in which Z, N, and E are the vertical, north–south, and east–west components, respectively. The covariance between any two components X and Y is given by

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

in which M is the number of samples. RBZ14 used a 3-s sliding window for instruments recording with a sampling rate of 40–200 Hz. This parameter may need to be adjusted somewhat for significantly different data sets. From the covariance matrix for a particular window, the eigenvalues ($\lambda_1 \geq \lambda_2 \geq \lambda_3$) and eigenvector matrix $\mathbf{u} = (\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3)$ can be used to calculate various aspects of the particle motion (e.g., Vidale, 1986; Jurkevics, 1988). Specifically we use the rectilinearity

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

and the apparent vertical incidence angle

$$\phi = \cos^{-1}(u_{13}), \quad (4)$$

in which u_{13} is the direction cosine between the principal eigenvector and the vertical component. These two quantities are then used to construct an *S*-wave polarization filter:

$$s = r(1 - \cos(\phi)). \quad (5)$$

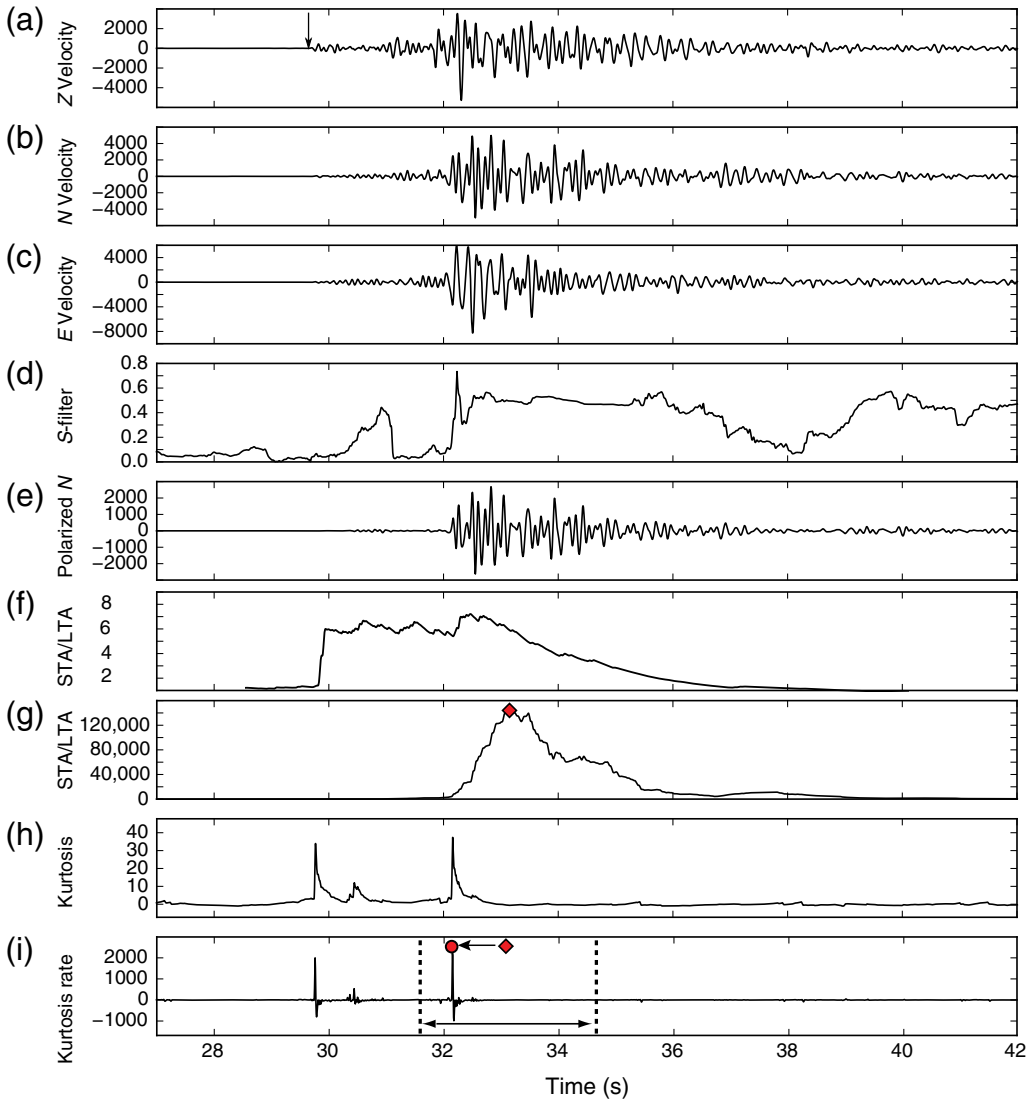

Figure 2. A step-by-step demonstration of the picking method. (a) Vertical-component velocity seismogram, (b) north–south (N–S) component velocity seismogram, (c) east–west (E–W) component velocity seismogram, (d) *S*-polarization filter, (e) *S*-polarized N component velocity seismogram, (f) standard STA/LTA calculated on (e), (g) short-term average/long-term average (STA/LTA) with LTA locked, (h) kurtosis function calculated on (e), and (i) kurtosis rate function. The diamond indicates the trial pick at the peak of the STA/LTA function, the dashed lines indicate the search range for the peak of the kurtosis rate function, and the circle indicates the final *S* pick. The color version of this figure is available only in the electronic edition.

The polarization filter has a range of [0,1] and is large when the particle motion is primarily in a horizontal plane and is small otherwise. Equation (5) is more appropriately thought of as a "not *P*" filter rather than an *S*-wave filter. This is because it tends to produce low values during the *P* wave and high values otherwise. In this method, equation (5) is used to dampen the *P*-wave arrival rather than enhance the *S*-wave arrival. Applying this process to the three-component data in Figure 2 results in the polarization filter shown in Figure 2d. Because this filter is calculated every sample as the window slides, it is a time series with the same length as the data traces. The polarization filter is used to modulate the horizontal traces by simple multiplication at each time step, attempting to remove as much of the *P*-wave motion as possible (Fig. 2e). Indeed, when comparing Figure 2e with Figure 2a–c, the *P*-wave energy is considerably weaker. For these reasons, RBZ14 called these *S*-polarized traces. For more discussion of this stage, see RBZ14.

With the polarization filter applied to the data, it generally removes enough of the *P* wave so that its amplitude is small relative to the *S* wave. RBZ14 calculated the STA/LTA (Allen, 1978, 1982) on the *S*-polarized horizontal traces (e.g., Fig. 2e). The STA/LTA effectively measures signal-to-noise ratio (SNR), or in some sense, the total energy in the window as a function of time. RBZ14 used an STA window length of 1 s and an LTA length of 10 s. Mulder *et al.* (unpublished manuscript, 2016, see Data and Resources) found that adding at least 1 count white noise to the polarized traces helps to

stabilize the STA/LTA function, which minimized spurious detections. An STA/LTA function with these parameters is shown for the *S*-polarized trace in Figure 2f. After calculating the STA/LTA function, RBZ14 applied a trigger mechanism that turned on when an STA/LTA threshold of 5 was met and turned off when the STA/LTA function fell to a threshold of 1. The trigger duration was required to be more than 1 s. This ensured that the trace contained a transient signal with sufficiently large amplitude. Then, they made a trial *S* pick when the STA/LTA function was maximized, which generally occurred in the vicinity of the *S* wave if the polarization analysis was successful. This is also the case in Figure 2f. If the polarization analysis is unsuccessful at removing the *P* wave, or the *P*-wave arrival is simply too strong relative to the *S*-wave arrival, the STA/LTA function's maximum potentially occurs during the *P* wave instead, leading to a mispick.

We modify this stage of the DBSHEAR algorithm to account for this shortcoming. This is done by locking the LTA so that it does not increase beyond the noise level. The LTA can be set to lock as soon as the STA/LTA reaches a value of 3 and unlock when the STA/LTA reaches a value of 1 (the noise level). Figure 2g shows an STA/LTA function with a locked LTA. By locking the LTA, the STA/LTA function becomes equal to a sliding SNR, which has a general tendency of being peaked around the *S* wave, because it has the most energy. This can be used to make a trial pick (diamond, Fig 2) from the time index in which the STA/LTA with locked LTA is maximized. If the LTA is not locked, as in RBZ14, it will continue to increase with the STA as the sliding window enters the earthquake seismogram. This can cause the STA/LTA function to have the largest peak at the *P* wave rather than at the *S* wave, even when the polarized traces are used.

With the trial pick obtained from stage 2, the last stage of the algorithm is now possible. As the trial pick tends to be located somewhere around the *S* wave, the pick must be refined to the location of the actual phase arrival. For this stage, we follow the exact procedure of RBZ14, who used a moving kurtosis function (Saragiotis *et al.*, 2002) to lock in on the arrival. Kurtosis is the fourth-order statistical moment and measures peakedness; it is quite effective at identifying abrupt changes in the amplitude of a time series, because it is extremely sensitive to outliers. Mathematically, it is defined as

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

in which M is the number of samples, $x = \{x_1, x_2, \ldots, x_M\}$ is the finite-length data sequence, $\bar{x}$ is the sample mean, and s_x is the standard deviation. RBZ14 used equation (6) with a sliding window length of 5 s on the *S*-polarized horizontal traces, which is shown for the example traces in Figure 2h. They found that the 5 s value was appropriate for sampling rates 50–200 Hz. Here, we use a 1 s window, because more extensive testing has shown that the resulting picks are slightly more accurate. From here, the kurtosis rate was calculated (Fig. 2i) and used to identify the phase arrival. This is accomplished using the trial *S* pick to define the center of a window on the kurtosis rate. RBZ14 also attempted to make a *P*-wave detection on the vertical component using an STA/LTA detector with a trigger threshold of 5. If a *P*-wave detection is successful, the window length used is equal to $t_S - t_P$, in which t_S is the trial *S* pick and t_P is the *P*-wave detection time (Fig. 2a, arrow). Otherwise, the window duration is 1.5 s. The pick is then refined so that the kurtosis rate is maximized within the 1 s window and adjusted slightly to the closest zero preceding the maximum. At this point, further *P*-wave discrimination could potentially be achieved by examining the apparent incidence angle φ in the vicinity of the final pick (e.g., Baillard *et al.*, 2014; Kurzon *et al.*, 2014). A value of $\varphi < 45^{\circ}$ is indicative of non-*S*-wave energy, and corresponding picks could be rejected; however, such a test has not been implemented in this work.

Real-Time Implementation and Picking Accuracy

RBZ14 demonstrated their *S*-picking algorithm only on waveforms of earthquakes that were detected and located previously. This was done with the intent of repicking phase arrivals for an event at different stations of interest, relying on data traces to be cut around the seismogram beforehand. To use the method for detecting and locating events, however, it must be capable of working on continuous data volume with no information *a priori*. Here, we demonstrate that the RBZ14 algorithm is suitable for real-time processing of continuous data and is furthermore capable of building a high-quality seismicity catalog from scratch when combined with the Antelope Environmental Monitoring Software package (BRTT Inc.). For these purposes, we use all the continuous waveform data recorded by the ANZA+ network (Fig. 1) from 1 May to 6 June 2014.

To aid in the task of detection and location, we use the Antelope platform, which is presently employed by a number of permanent seismic networks around the world, including the ANZA+ network. We first apply Antelope's DBDETECT software to all the raw continuous data and make *P*-wave arrival detections. The parameters used for this are the default settings provided by the manufacturer and are the same as those used internally by the AZ to detect earthquakes. For each *P*-wave detection made on the continuous data, the three-component data are cut starting 10 s before and 14 s after the *P*-wave detection. Then, the resulting trace is passed to the DBSHEAR algorithm, including our modifications to stage 2. The DBSHEAR algorithm attempts to make *S*-wave detections on each of the horizontal components, and if both components are successful, the one with the larger SNR is chosen. Alternatively, the one with a larger peak kurtosis value could be used, which would correlate with arrival sharpness. We note that a phase detection, as described here, is not the same as a phase arrival pick. This is because we are detecting new events from scratch on the continuous data, and only a subset of the phase detections will be later converted into actual phase arrival picks. The distinction between *P*- and *S*-wave

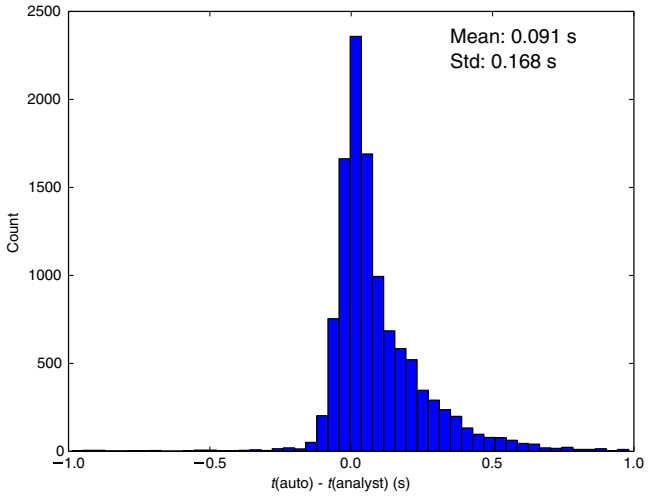

Figure 3. Histogram of picking errors compared to analyst picks for 2014 data set. A total of 11,353 manual *S* picks were used in the comparison, and the average automatic pick had an accuracy of 0.09 ± 0.168 s. The color version of this figure is available only in the electronic edition.

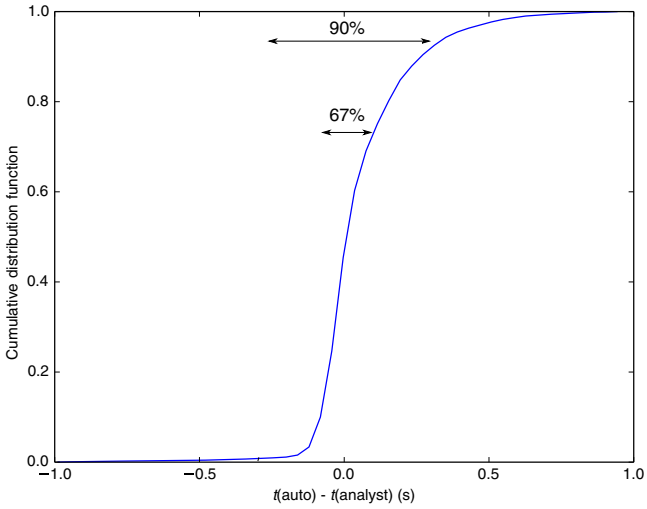

Figure 4. Empirical cumulative distribution function describing the same results in Figure 3. The 11,353 picks being compared here are within 0.06 s 50% of the time, 0.11 s 67% of the time, and 0.31 s 90% of the time. The color version of this figure is available only in the electronic edition.

detections is therefore a naming convention for whether the phase detection is made by DBDETECT or DBSHEAR, respectively. Using this process on all the continuous data for the analyzed period results in 302,346 *P* detections and 199,786 *S* detections.

It is unknown at this time whether any of these phase detections are actually associated with a seismic event. For example, a phase detection could turn out to be an instrument glitch or an arbitrary transient signal not related to an earthquake. The steps described are capable of working on a completed seismic database or on packets of data streaming in real time. In both situations, a piece of continuous data is passed to the phase detection algorithms to see if it contains something that looks like an earthquake; however most of the time it will just contain noise. We successfully implemented the DBSHEAR algorithm in real time on the AZ network, and configured DBSHEAR to run only when a phase detection is made by DBDETECT on a given station somewhere. This reduces the overall amount of computation considerably, because the polarization filters and kurtosis functions only need to be calculated for a small percentage of the total data. However, it could also be applied directly to the continuous data independently from DBDETECT. The DBSHEAR method works exactly the same regardless of whether it is working on archived data or packets of data streaming in real time.

After a sufficient number of phase detections are made, they are processed by Antelope's grid associator, DBGRASSOC, to check whether any are associated with a coherent earthquake-type source. The grid association is performed by back projecting all the phase detections to different nodes in a grid around the SJFZ by subtracting their predicted travel times from the phase detection time. An event is detected when a minimum of five back-projected detections cluster within a 1.5 s window. The travel times are calculated using a 1D velocity model. The clustered phase detections are then promoted to phase arrival picks at this point in the process. Whenever an event is detected, the phase arrivals associated with the coherent source location are inverted using the GENLOC algorithm of Pavlis *et al.* (2004) to get a formal hypocenter and origin time.

The described process resulted in 1399 earthquakes detected and located, along with 26,248 *P* picks and 23,717 *S* picks. During the examined time period, more than 20,000 *S*-wave arrivals were handpicked by examining every single earthquake detected by the ANZA+ network. We use these picks for direct comparison with our automatic ones to assess the statistical accuracy. For every automatic pick in common with this manual set of picks, which totaled 11,353 *S* picks, we calculated the error relative to the analyst pick. Figure 3 shows a histogram of the picking errors for this data set. The mean of this distribution is close to zero, and the picks are primarily delayed, which is the result of a combination of several factors. These include, for example, occasional *P*-wave coda overlapping *S*-wave onsets, a delayed rise time in the kurtosis function, or sharp secondary phase arrivals. Figure 4 shows the empirical cumulative distribution function (ECDF) for the values in Figure 3. From this ECDF, 50% of the 11,353 picks are within 0.061 s of the analyst pick. Furthermore, 75% of all picks are within 0.16 s of the analyst pick, and 95% are within 0.43 s. Analyst picks at the ANZA+ network typically have an error of around 0.1 s, and roughly two-thirds of the picks made by our algorithm are as accurate. We also note that there are missed picks in both the automatic set and the manual set, because only about half of the picks are common between the two. One major reason for this is that, during the process of

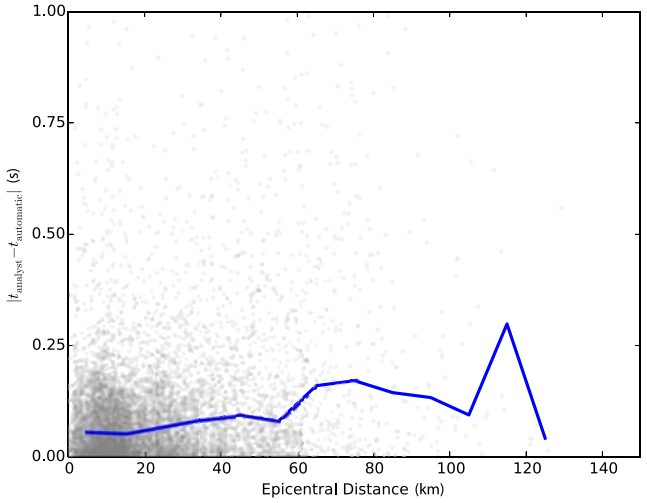

Figure 5. Picking method accuracy as a function of epicentral distance. The median absolute pick error relative to the analyst pick is given by the line. From 0 to 60 km epicentral distance, the median pick error is about 0.05 s and then slightly increases to about 0.15 s beyond this range. Overall the errors are quite stable over the full distance range tested. The color version of this figure is available only in the electronic edition.

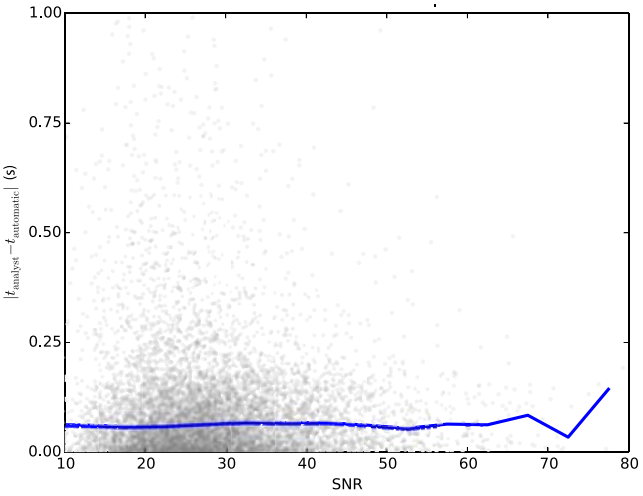

Figure 6. Picking method accuracy as a function of signal-to-noise ratio (SNR). The median pick accuracy is nearly constant at 0.07 s over the entire range of SNR values observed in the data set. The color version of this figure is available only in the electronic edition.

handpicking the data, ∼30% additional earthquakes were detected visually, and none of these extra picks have a match in the automatic set. The automatic set contains a significant number of new picks that were missed during the manual picking. We tried to be as thorough as possible during the handpicking process; however, with more than 100 three-component stations available, it is difficult to ensure every station possible is picked for each event.

The picking errors can further be studied as a function of SNR and epicentral distance. Figure 5 shows that the median pick error, indicated by the line, is about 0.1 s or less up to about 60 km. Beyond this distance, the median pick error is slightly higher but still remains <0.2 s up to about 100 km, at which distance the data become scarce. Figure 6 presents the same picks as a function of SNR, and the median pick error is seen to be essentially flat with regard to SNR. Taken together with the results in Figure 5, this indicates that the method is more affected by the changes produced in the waveforms due to longer propagation rather than by magnitude itself. However, these changes are not very pronounced, and the method works well over the full range of these different examined scales. The SJFZ has a wide diversity of focal mechanism types (e.g., Bailey *et al.*, 2010), leading to greater variability in the character of observed waveforms. We therefore have no reason to expect the quality of the picks to worsen on other data sets, and, indeed, applications to different regions support this expectation (e.g., Wu *et al.*, 2016).

Systematic Application to the ANZA+ Network

With the accuracy of the method established in the previous section, we now proceed to discuss the application of the algorithm to a data set over a longer time period. The data set used for this part consists of the entire year 2013 of continuous waveform data recorded by the ANZA+ network. We applied the same detection and location procedure from the previous section to this larger data set to detect and locate every single earthquake from scratch. In this manner, side-by-side analysis was possible in studying the effect that introducing DBSHEAR to the workflow had on the seismicity catalog. Two separate seismicity catalogs were produced for the entire year. The first catalog was produced using just the DBDETECT program alone. The resulting catalog is identical to the catalog produced by the AZ network's real-time system for 2013, void of any analyst interaction. It is therefore a fully automatic seismicity catalog. The second seismicity catalog used both DBDETECT and DBSHEAR, and as a result, the only differences are due to the DBSHEAR algorithm. In this manner, we can directly study how our method alters the locations and picks in a seismicity catalog.

The results of this process for the two catalogs are shown in Figure 7. In both catalogs, the number of events detected is 11,197. For the first catalog using DBDETECT alone, 290,339 *P* picks and 38,538 *S* picks are made, resulting in a *P* to *S* pick ratio of 7.53. In the second catalog the number of *P* picks decreases slightly to 290,155, due to various phases now being associated as *S* waves, as the result of having more information. The number of *S* picks, however, increases to 266,699 (∼600%). The resulting *P* to *S* ratio drops from 7.53 to 1.08. Another way of looking at this is that for each *P* pick that is successful on the continuous data, an *S* pick is successful 92% of the time.

With such a considerable increase in the number of *S* picks between the two catalogs, the most important question is how the locations and origin times changed as a result. Figure 8 contains information on how various location

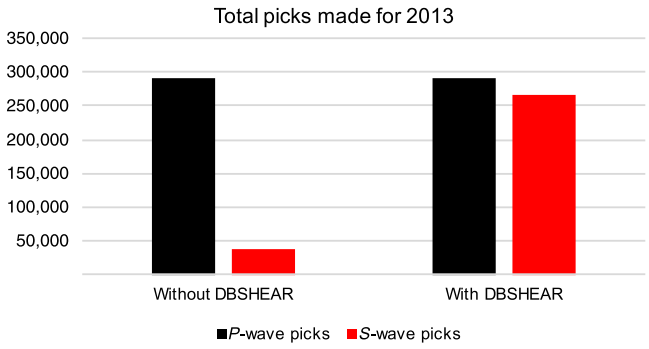

Figure 7. Comparison of number of phase picks made when using the stock Antelope DBDETECT program alone, as compared with DBDETECT and the DBSHEAR algorithm together. Nearly the same amount of *P*-wave arrivals were obtained with both methods. Roughly 38,000 *S* picks were made by DBDETECT alone; however, when adding in DBSHEAR, this number increased to 266,699 (~600%). The resulting *P* to *S* ratio has decreased from 7.53 to 1.08. The color version of this figure is available only in the electronic edition.

quantities changed between the two catalogs. The four quantities shown in Figure 8a are all from the GENLOC algorithm (Pavlis *et al.*, 2004) and are the major and minor axes of the covariance matrix, depth, and origin time resolution. Here, these quantities are the differences in the uncertainty estimates between the two locations. The major and minor axes of the covariance matrix provide estimates of the horizontal error. For each parameter, negative values indicate that the estimate of the uncertainty decreased and that the degree of fit increased. For all four of the parameters in Figure 8a, about 99% of the events have the same or better location accuracy. Conversely, addition of *S*-wave picks from DBSHEAR resulted in degradation of location accuracy for ~1% or less of the events. Because the location uncertainty decreased for the vast majority of cases, another question of interest is how these resolution changes map into 4D hypocentral shifts. Figure 8b shows the differences in latitude, longitude, depth, and origin time between the two catalogs. These histograms indicate that the average epicentral shift between the two catalogs is on the order of 0.5 km, and the average depth change is 1.1 km. This is measured over the 11,197 earthquakes tested. Furthermore, the average origin time shift is 0.2 s. The distributions of these quantities, however, have very long tails. A number of events, for example, have depth changes as large as 3–4 km and origin time shifts of 0.75 s. Because the only difference between the two data sets is the inclusion of the DBSHEAR algorithm, it is entirely responsible for the improvement to the locations. The changes in depth are often found to trade off with changes in the origin time. Figure 9 presents a scatter plot of these quantities, in which it is clear that they are related for this data set as well. Thus, the changes should not be considered completely independent.

The uncertainty estimates are derived from a covariance matrix (Pavlis *et al.*, 2004) by projecting the results onto horizontal and vertical planes. We focus on these quantities because they are what is provided by the GENLOC software and are the same quantities used in the AZ's real-time system. Many other techniques exist for quantifying the uncertainty in a hypocenter, such as bootstrap methods. However, the goal here is simply to demonstrate that the origin shifts are reinforced by similar changes in the provided uncertainty metrics, rather than to identify the most suitable metric for quantifying location uncertainty.

Finally, we compare the hypocentral locations between the two catalogs visually around the SJFZ. Figure 10 shows a map of the SJFZ with events from the catalog produced with DBDETECT alone, alongside events from the catalog produced with DBDETECT and DBSHEAR. A projection of all events in the region along the strike of the SJFZ is included as well. Compared to locations produced by a relocation algorithm such as hypoDD (Waldhauser and Ellsworth, 2000), the locations tend to shift toward the center of the network in a radial sense. Some of these changes are large; however, the majority are relatively small (Fig. 8). Some events are found to be located artificially near about 30 km depth using DBDETECT alone but are pulled up more than 10 km after including DBSHEAR. Some artifacts in the locations are visible on the edges of the network and at depth in the seismicity. This is due to the two-stage location process, in which the events are initially located using a grid. For the locations that are poorly constrained, the subsequent hypocenter inversion (GENLOC) will often converge near or at the gridded location. We found that these artifacts disappear after relocation via hypoDD. It is further noted that events that have been relocated with an algorithm like hypoDD are generally more reliable.

Discussion

The method described herein is capable of making high-quality *S*-wave picks on continuous data with no knowledge of whether earthquakes have occurred. It is thus a real-time picking algorithm and is appropriate for use by an operational seismic network as well as for processing existing archives of continuous waveform data to detect earthquakes. The method comes with several caveats that should be mentioned for completeness. It was designed for picking phases on local earthquakes and requires three-component instruments. The method has been successful at picking *S* waves up to about 120 km. Beyond this distance, the development of regional phases complicates the signal and makes automatic detection more challenging. It is possible that reparameterization of the method may produce useful results at regional scales; however, no testing at regional scales has been performed.

The original DBSHEAR algorithm typically led to a *P* to *S* pick ratio between 2 and 3. Wu *et al.* (2016) used DBSHEAR together with the method of Ross and Ben-Zion (2014b) on the TAIGER arrays deployed around Taiwan in 2009. They obtained a *P* to *S* pick ratio of 2.56 after detecting more than 8000 earthquakes. In this study, we showed that the proposed modifications decreased this ratio to 1.08 for

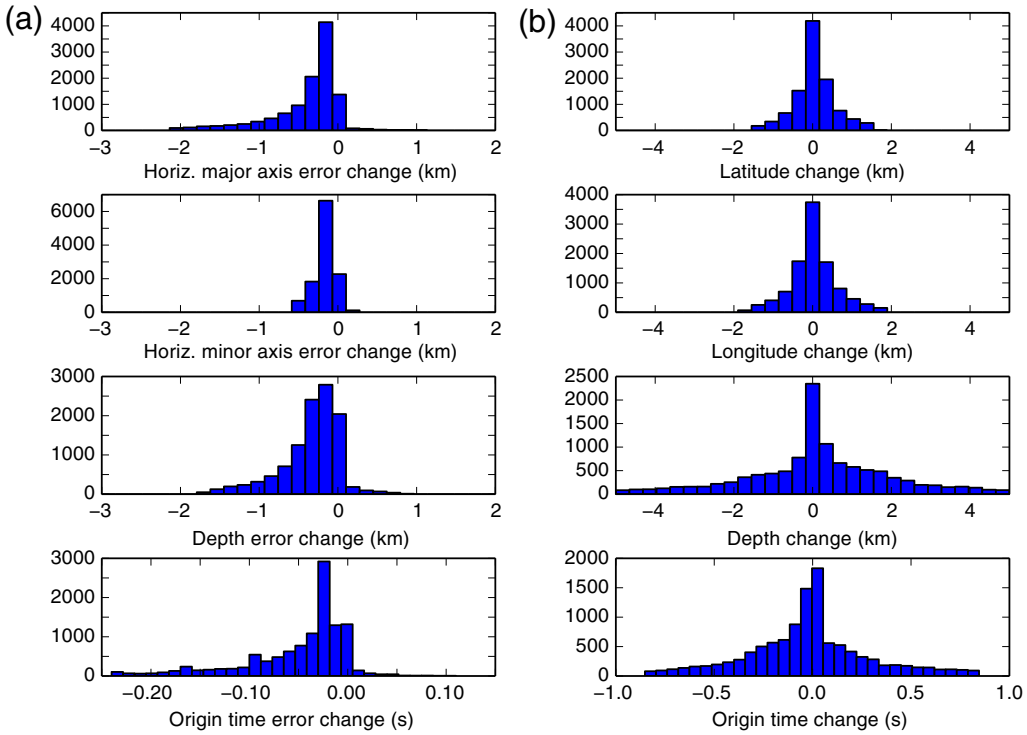

Figure 8. Comparison of location uncertainty statistics between the two 2013 data sets, which contained 11,197 events. (a) Histograms of the change in hypocenter resolution as estimated by the GENLOC algorithm (Pavlis *et al.*, 2004). For these histograms, a negative number indicates that the uncertainty in the fitted location decreased by including the DBSHEAR algorithm, compared to DBDETECT alone. In a very small number of events, the uncertainty increased, reflecting *S* picks that were of poor quality. (b) Histograms indicating how the actual hypocenter quantities changed before and after including DBSHEAR. The average depth changed by 1.1 km, the average epicenter changed by roughly 500 m, and the average origin time changed by 0.2 s. The color version of this figure is available only in the electronic edition.

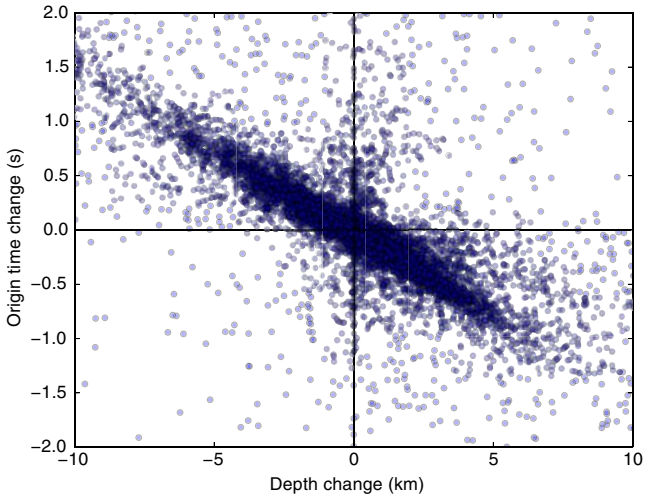

Figure 9. Comparison of depth changes against origin time changes. A clear trade-off exists between the two quantities, and thus the changes shown in Figure 8 should not be considered independent between the different quantities. The color version of this figure is available only in the electronic edition.

the 11,197 earthquakes detected for the year 2013. Furthermore, the average pick's accuracy is dramatically improved, and 75% of the picks tested are within 0.16 s of the corresponding analyst pick.

In this study, we make the formal distinction between a real-time algorithm and a re-picking algorithm. Many of the picking algorithms discussed in the literature (e.g., Wang and Teng, 1997; Gentili and Michelini, 2006; Baillard *et al.*, 2014) are re-picking methods and require additional *a priori* information, such as an earthquake location, approximate arrival time, or concrete knowledge that an event exists in a particular data window. The data are then cut around the earthquake, and the method is applied to pick the phases. In applying a picking method to continuous data, during the vast majority of the time there are no earthquakes, and there are often noise bursts and other transient signals from vehicles, animals, or cultural sources. A real-time picking algorithm must be capable of distinguishing between all of these types of signals and deciding when to make a pick. This includes the null case of no pick, which is the most common scenario, and if too many false picks are made, it can lead to false event detections.

Real-time picking algorithms are used by permanent networks to locate earthquakes as part of the detection process and are responsible for making the first picks available to a network. Further development of real-time picking algorithms can therefore have a significant impact on the locations provided in seismicity catalogs. In this study it was shown that by including our *S*-picking algorithm in the AZ's workflow,

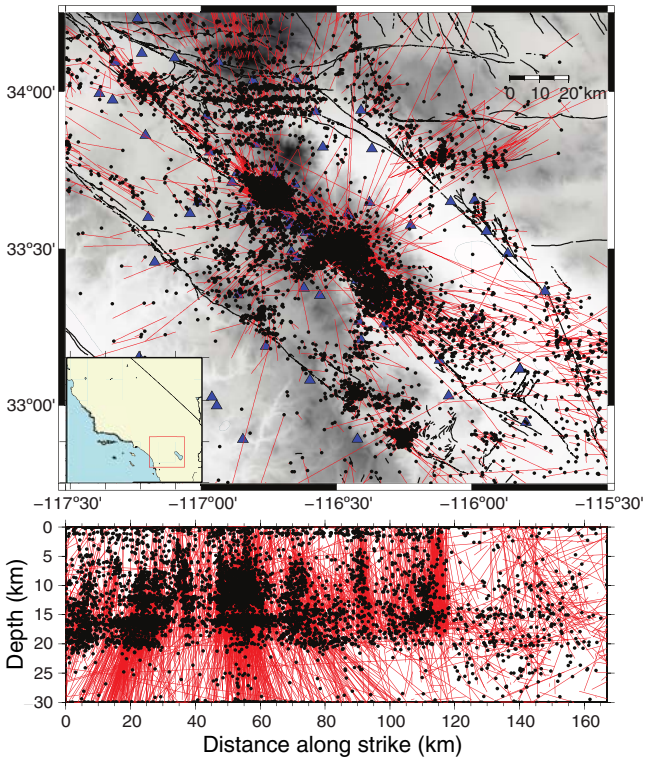

Figure 10. Location comparisons for the two 2013 data sets. For both data sets, the number of events is 11,197. The black dots are the locations of events using both DBDETECT and DBSHEAR to pick phases. The lines indicate the change in location from those using DBDETECT alone. For most events, the change is small, but occasionally an event has a large shift in location. This is most prominent on the edges of the network, and all location changes tend to be in a radial direction toward the center of the AZ+ network. Events which originally located near 30 km depth were almost all moved closer to the surface. The color version of this figure is available only in the electronic edition.

the average hypocenter was improved by more than 1 km. These improvements will propagate through to downstream applications, such as tomography studies.

An important aspect of picking phase arrivals before any events are detected or located is that it allows for a fully self-consistent database. If a re-picking algorithm is applied to events that were detected previously by a permanent network, the picks will be different from what the network produced internally. All of the information derived from the network's original picks—which includes locations, magnitudes, and source properties—will be inconsistent with the new picks produced. This introduces a form of uncertainty that is avoided if picks are made before locations are calculated. Otherwise, we suggest that with re-picking arrival times, new locations should be calculated to keep this inconsistency from unnecessarily being passed on to future analyses.

Conclusions

We developed a modification to the DBSHEAR algorithm introduced originally by Ross and Ben-Zion (2014a) for picking *S*-wave arrivals on seismic waveform data. The method is a three-stage procedure designed to account for a wide variety of expected waveform characteristics. It first uses polarization analysis to remove as much *P*-wave energy from a seismogram as possible. Next, STA/LTA detectors are run on *S*-polarized traces to identify the neighborhood of the *S* wave. Finally, a moving kurtosis function is used to refine the final pick to a well-defined region of the seismogram. In this study, we demonstrated for the first time that the method works on packets of data streaming in real time, with no prior knowledge of whether an event has occurred. We then applied the method to more than a year of continuous waveform data recorded at over 100 stations around the SJFZ and built a seismicity catalog from scratch. The resulting phase picks associated with events in this catalog were compared with more than 11,000 handmade phase picks to rigorously examine the accuracy of the method. In doing so, 75% of all *S* picks were found to be within 0.16 s of the corresponding analyst pick. We demonstrated that by using the DBSHEAR method together with the Antelope platform, the average event hypocenter improved by more than 1 km, and the number of *S* picks made increased by more than 600%.

Data and Resources

Seismic data used in this study are gathered and managed by the following networks: the ANZA network (AZ) and the San Jacinto fault zone (SJFZ) (YN) network are operated by Institute of Geophysics and Planetary Physics, University of California, San Diego (http://eqinfo.ucsd.edu/, last accessed August 2015); the SCSN network (CI) is the Southern California Seismic Network operated by Caltech and U.S. Geological Survey (http://www.scsn.org/, last accessed August 2015); the PBO network (PB) is the Plate Boundary Observatory operated by UNAVCO (http://www.earthscope.org/science/observatories/pbo, last accessed August 2015); the SB network is operated by University of California, Santa Barbara (http://nees.ucsb.edu/publications, last accessed August 2015); gaps in data were filled by the Incorporated Research Institutions for Seismology (IRIS) Data Management Center (DMC) (http://ds.iris.edu/ds/nodes/dmc, last accessed August 2015); and instrumentation for seismic stations was provided by IRIS Program for the Array Seismic Studies of the Continental Lithosphere (PASSCAL) (http://www.passcal.nmt.edu/, last accessed August 2015). The instruments (Fig. 1) used are a mixture of broadband and short-period seismometers and accelerometers. The network includes the core AZ network stations, a temporary PASSCAL deployment (YN) around the SJFZ, several PBO (PB) borehole stations, adjacent SCSN (CI) stations, and the BVDA and GVDA (SB) borehole arrays. This collective set of stations is denoted below as the ANZA+ network for simplicity. The employed data were recorded during two separate time periods. The first data set consists of all the continuous waveform data recorded from 1 May to 6 June 2014, chosen because it coincided with a 1108 instrument nodal array that was temporarily deployed

in the SJFZ (Ben-Zion *et al.*, 2015). More than 20,000 *S*-wave picks were handmade by analysts during this period, yielding an extensive high-quality pick and event database, and we used this set to systematically test the quality of our method's picks. The second data set consists of the entire year 2013 of continuous waveform data recorded by the ANZA+ network. This data set was used to systematically study how the method improves upon the existing/previous real-time detection capabilities and results of the ANZA+ network. The DBSHEAR software is available with both Antelope and ObsPy versions upon request. We used information from the unpublished manuscript "Analysis of the 28 October 2012 Haida Gwaii aftershock sequence using automated earthquake locations from singular value decomposition detections" by T. L. Mulder, D. Kilb, M. White, and F. L. Vernon (2016), submitted to Bulletin of the Seismological Society of America.

Acknowledgments

This work was funded by the National Science Foundation Grant Numbers EAR-0908903 and EAR-0908042. We thank Associate Editor Eric Chael and an anonymous reviewer for their constructive comments.

References

Allen, R. (1978). Automatic earthquake recognition and timing from single traces, *Bull. Seismol. Soc. Am.* **68,** 1521–1532.

Allen, R. (1982). Automatic phase pickers: Their present use and future prospects, *Bull. Seismol. Soc. Am.* **72,** no. 6, S225–S242.

Bailey, I. W., Y. Ben-Zion, T. W. Becker, and M. Holschneider (2010). Quantifying focal mechanism heterogeneity for fault zones in southern and central California, *Geophys. J. Int.* **183,** 433–450, doi: 10.1111/j.1365-246X.2010.04745.x.

Baillard, C., W. C. Crawford, V. Ballu, C. Hibert, and A. Mangeney (2014). An automatic kurtosis-based *P*- and *S*-phase picker designed for local seismic networks, *Bull. Seismol. Soc. Am.* **104,** 394–409, doi: 10.1785/0120120347.

Ben-Zion, Y., F. L. Vernon, Y. Ozakin, D. Zigone, Z. E. Ross, H. Meng, M. White, J. Reyes, D. Hollis, and M. Barklage (2015). Basic data features and results from a spatially-dense seismic array on the San Jacinto fault zone, *Geophys. J. Int.* **202,** 370–380, doi: 10.1093/gji/ggv142.

Gentili, S., and A. Michelini (2006). Automatic picking of *P* and *S* phases using a neural tree, *J. Seismol.* **10,** 39–63.

Hildyard, M. W., S. E. J. Nippress, and A. Rietbrock (2008). Event detection and phase picking using a time-domain estimate of predominant period Tpd, *Bull. Seismol. Soc. Am.* **98,** 3025–3032.

Jurkevics, A. (1988). Polarization analysis of three-component array data, *Bull. Seismol. Soc. Am.* **78,** 1725–1743.

Kurzon, I., F. L. Vernon, A. Rosenberger, and Y. Ben-Zion (2014). Real-time automatic detectors of *P* and *S* waves using singular value decomposition, *Bull. Seismol. Soc. Am.* **104,** no. 4, 1696–1708, doi: 10.1785/0120130295.

Nippress, S. E. J., A. Rietbrock, and A. E. Heath (2010). Optimized automatic pickers: Application to the ANCORP data set, *Geophys. J. Int.* **181,** 911–925.

Pavlis, G. L., F. L. Vernon, D. Harvey, and D. Quinlan (2004). The generalized earthquake-location (GENLOC) package: An earthquake-location library, *Comp. Geosci.* **30,** 1079–1091, doi: 10.1016/j.cageo.2004.06.010.

Qiu, H., Y. Ben-Zion, Z. E. Ross, P.-E. Share, and F. L. Vernon (2015). Internal structure of the San Jacinto fault zone at Jackass Flat from data recorded by a dense linear array, *Abstract of the annual SSA meeting*, 21–23 April, Pasadena, California.

Rawles, C., and C. Thurber (2015). A nonparametric method for automatic determination of *P*-wave and *S*-wave arrival times: Application to local microearthquakes, *Geophys. J. Int.* **202,** no. 2, 1164–1179, doi: 10.1093/gji/ggv218.

Ross, Z. E., and Y. Ben-Zion (2014a). Automatic picking of direct *P*, *S* seismic phases and fault zone head waves, *Geophys. J. Int.* **199,** 368–381, doi: 10.1093/gji/ggu267.

Ross, Z. E., and Y. Ben-Zion (2014b). An earthquake detection algorithm with pseudo probabilities of multiple indicators, *Geophys. J. Int.* **197,** no. 1, 458–463, doi: 10.1093/gji/ggt516.

Saragiotis, C., L. Hadjileontiadis, and S. Panas (2002). PAI-S/K: A robust automatic seismic *P* phase arrival identification scheme, *IEEE Trans. Geosci. Remote Sens.* **40,** 1395–1404.

Share, P.-E., Y. Ben-Zion, and C. Thurber (2015). Elucidating fault zone structures in the South-Central Transverse Ranges area using double difference tomography, Presented at the *SSA Annual Meeting*, April 2015, Pasadena, California.

Sleeman, R., and T. van Eck (1999). Robust automatic *P*-phase picking: An on-line implementation in the analysis of broadband seismogram recordings, *Phys. Earth Planet. In.* **113,** 265–275.

Vidale, J. E. (1986). Complex polarization analysis of particle motion, *Bull. Seismol. Soc. Am.* **76,** no. 5, 1393–1405.

Waldhauser, F., and W. L. Ellsworth (2000). A double-difference earthquake location algorithm: Method and application to the northern Hayward fault, California, *Bull. Seismol. Soc. Am.* **90,** no. 6, 1353–1368.

Wang, J., and T. Teng (1997). Identification and picking of *S* phase using an artificial neural network, *Bull. Seismol. Soc. Am.* **87,** no. 5, 1140–1149.

Wu, F. T., Z. E. Ross, D. Okaya, Y. Ben-Zion, C.-Y. Wang, H. Kuo-Chen, and W.-T. Liang (2016). Dense network, intense seismicity and tectonics of Taiwan, *Tectonophysics*, doi: 10.1016/j.tecto.2016.04.025.

Department of Earth Sciences
University of Southern California
Los Angeles, California 90089-0740
zross@usc.edu
(Z.E.R., Y.B.-Z.)

Institute of Geophysical and Planetary Sciences
Scripps Institution of Oceanography
University of California, San Diego
La Jolla, California 92093-0225
(M.C.W., F.L.V.)

Manuscript received 27 August 2015;
Published Online 19 July 2016