

A Robust Algorithm for Automatic *P*-wave Arrival-Time Picking Based on the Local Extrema Scalogram

by Ting-Chung Huang* and Yih-Min Wu

Abstract A robust algorithm has been developed for the automatic picking of *P*-wave arrival times. Owing to the properties of the local extrema scalogram (LES), this algorithm finds all significant quasi-periodic peaks and valleys without selecting a specific frequency. Consequently, the *P*-wave arrival times can be accurately derived from the peaks and valleys of the seismic signal. A comparison of the proposed algorithm with the common short-term average/long-term average (STA/LTA) method and the Akaike information criterion (AIC) method is conducted using real data. The results show that our method consistently outperforms both methods, especially when substantial noise is present.

Introduction

The short-term average/long-term average (STA/LTA) method (Allen, 1978, 1982) has achieved remarkable success in automatically identifying *P*-wave arrivals in real-time scenarios situated within a quiet environment. In this approach, a characteristic function (CF) is first defined, after which the STA/LTA ratio is calculated. If this ratio exceeds a predefined threshold, the time will be designated as a picked *P*-wave arrival. STA/LTA is computational cost-effective due to its memoryless design, which will not look back to previous window, and best suited in the real-time detection. A few seconds of *P*-wave signal is all it needs to trigger a pick. However, some difficulties are encountered during the application of this method. First, it is limited by its one-way design for differentiating true signals from noise. As a consequence of this flaw, the STA/LTA ratio will be triggered by false fluctuations in some cases, thereby providing erroneous arrival times. Second, an excessive number of free parameters are available for selection. In his original paper, Allen (1978) presented a total of five parameters, and many more parameters appeared in recent variations of the STA/LTA method. This reflects the fact that higher order details of the CF require additional parameters. Moreover, in addition to the number of parameters, the meaning of each parameter provides another level of complexity; although the parameters are defined explicitly, it is not easy to discern their physical meanings.

The Akaike information criterion (AIC; Akaike, 1973; Sleeman and van Eck, 1999) constitutes another commonly

used picking approach. Although the AIC approach is capable of demonstrating good results, it often provides highly erroneous picks. Consequently, in practice, one often determines effective windows either by hand or by an automated method to ensure that the AIC technique outputs the correct result.

Many approaches have been proposed based on the above-mentioned principles to resolve problems associated with the picking of *P*-wave arrival times. However, although some of these techniques are mentioned later in this article, this list is not exhaustive because of the scope of this article. In the family of STA/LTA methods, Allen (1978, 1982) used a CF of $x[i]^2 + C(x[i] - x[i - 1])^2$, and Baer and Kradolfer (1987) later introduced a specific differential prefactor into the CF with the form of $CF = x[i]^2 + D\dot{x}[i]^2$, in which $D = \text{SUM}(x[i]^2)/\text{SUM}(\dot{x}[i]^2)$. Subsequently, Küperkoch *et al.* (2010) proposed a more sophisticated CF using higher order statistics, such as kurtosis and skewness, to identify the transition from Gaussian to non-Gaussian. In the family of AIC methods, Sleeman and van Eck (1999) combined the original AIC technique with autoregressive analysis to improve the picking results, and Scafidi *et al.* (2018) defined the search window for the AIC and updated the window recursively with the calculated location. Other methods beyond these two popular approaches have also been developed. For example, Zhang *et al.* (2003) used the discrete wavelet transform in conjunction with the AIC, Ait Laasri *et al.* (2013) used the cross-correlation operation, Bogiatzis and Ishii (2015) adopted continuous wavelet analysis, Ross and Ben-Zion (2014) used polarization analysis, and Zhu and Beroza (2018) used deep learning techniques.

In this article, we solve the above-mentioned problems through a different approach. Inspired by the automatic multiscale-based peak detection (AMPD) algorithm (Scholkmann *et al.*, 2012), we design a new technique that

*Also at NTU Research Center for Future Earth, National Taiwan University, Taipei 10617, Taiwan; Institute of Earth Sciences, Academia Sinica, No. 128, Section 2, Academia Road, Nangang, Taipei 11529, Taiwan; and National Center for Research on Earthquake Engineering, National Applied Research Laboratories, No. 200, Section 3, Xinhai Road, Taipei 106, Taiwan.

uses the local extrema scalogram (LES) to find quasi-periodic peaks within a noisy environment. Our method solves the picking problem by initially determining all of the peaks and valleys within the record. Then it determines the peaks and valleys corresponding to the seismic phase arrival and finally traces the signal back to the P -wave arrival time from the first apparent peak or valley, whose amplitude is of the same order of magnitude to the maximum amplitude. Our contributions are as follows:

- We describe how we derive the proposed technique from the AMPD algorithm to determine P -wave arrival times (see the [Algorithm](#) section).
- We compare our method with two widely implemented existing methods, namely, the generic STA/LTA method with optimum parameters (see the [STA/LTA Method](#) section) and the AIC method (see the [AIC Method](#) section). To this end, we use real waveform data from the National Research Institute for Earth Science and Disaster Resilience (NIED) and plot the corresponding results for analysis (see the [Comparisons among the Methods](#) section).
- We show that our method operates effectively even under noisy environments (see the [Different Groups of Signal-to-Noise Ratios](#) section).
- We also demonstrate that our method can avoid common picking errors, including noisy spikes resulting from offsets (see the [Common Picking Difficulties](#) section).

Algorithm

The proposed algorithm consists of three main parts. The first part finds all of the periodic or quasi-periodic peaks. The second part filters out all possible noise peaks and keeps only the apparent peaks. The third part traces the signal back to the P -wave arrival time. Figure 1 summarizes the algorithm used in this study, and Figure 2 illustrates each step using both real and synthetic data.

The first part of our method is based on the AMPD algorithm (Scholkmann *et al.*, 2012), which uses the local maxima scalogram, a construction inspired by the wavelet scalogram, to search for quasi-periodic peaks. Although the AMPD algorithm was originally designed to locate electrocardiography signals, it has been applied across a wide variety of disciplines; for example, Jena and Panigrahi (2015) used it with acoustic signals to detect rotary gear faults. In addition, because it was recommended that the signal be detrended before applying the original AMPD algorithm to locate the peaks and valleys, we detrend the signal in our algorithm.

Finding Peaks and Valleys

Let $x[i], i = 1, \dots, N$ be a uniformly sampled signal. First, we determine the maximum window length w_k in terms of L

$$w_k = 2k, \quad k = 1, 2, \dots, L, \quad (1)$$

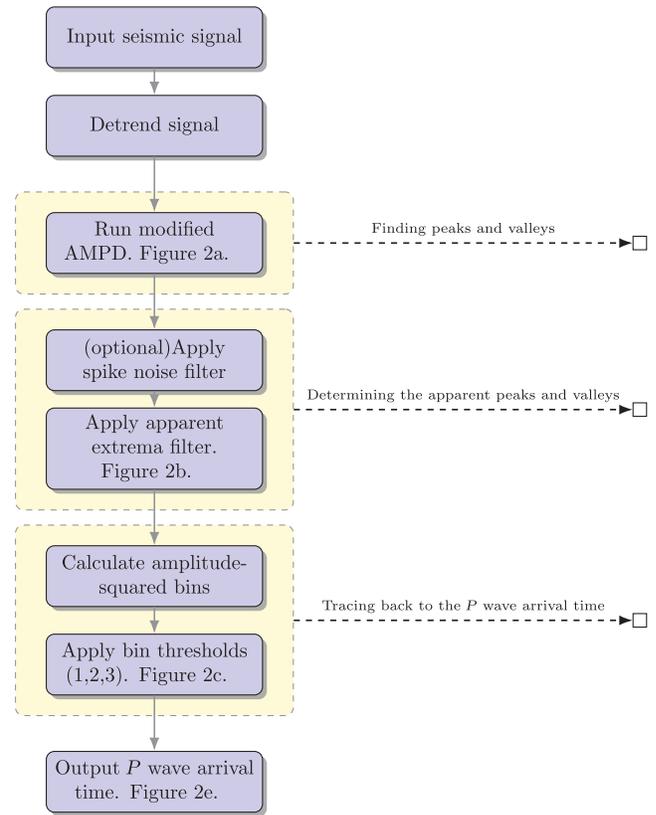


Figure 1. The flowchart of the proposed picking algorithm.

in which k is the scale in the wavelet formalism and L is the maximum analyzed scale. In other words, L corresponds to the minimum frequency of the periodic oscillation that is sought. In general, the value of L can be as large as $N/2$. For seismological research, we assign the value of L to correspond to the typical long-period limit of seismic P waves, that is, 13.33 s (0.075 Hz). Our choice of this higher period bound is more inclusive than a scale of 5 s, which represents the τ_c value of a magnitude 8 earthquake (Wu and Kanamori, 2005).

Second, an $L \times N$ LES matrix $m_{k,i}$ is constructed. It is different from the original AMPD. For every scale k , i runs from $2k + 1$ to $N - 2k$

$$m_{k,i} = \begin{cases} 0, & x[i] > x[i - 2k] \wedge x[i] > x[i + 2k] \\ 0, & x[i] < x[i - 2k] \wedge x[i] < x[i + 2k], \\ \mathcal{U}(0, 1) + 1, & \text{otherwise} \end{cases} \quad (2)$$

in which $\mathcal{U}(0, 1)$ is a continuous uniform random number distribution in the range of $[0, 1]$.

Third, a reduced LES matrix is found, for which a row-wise summation of $m_{k,i}$ is performed

$$\gamma_k = \sum_{i=1}^N m_{k,i}, \quad \text{for } k \in \{1, 2, \dots, L\}. \quad (3)$$

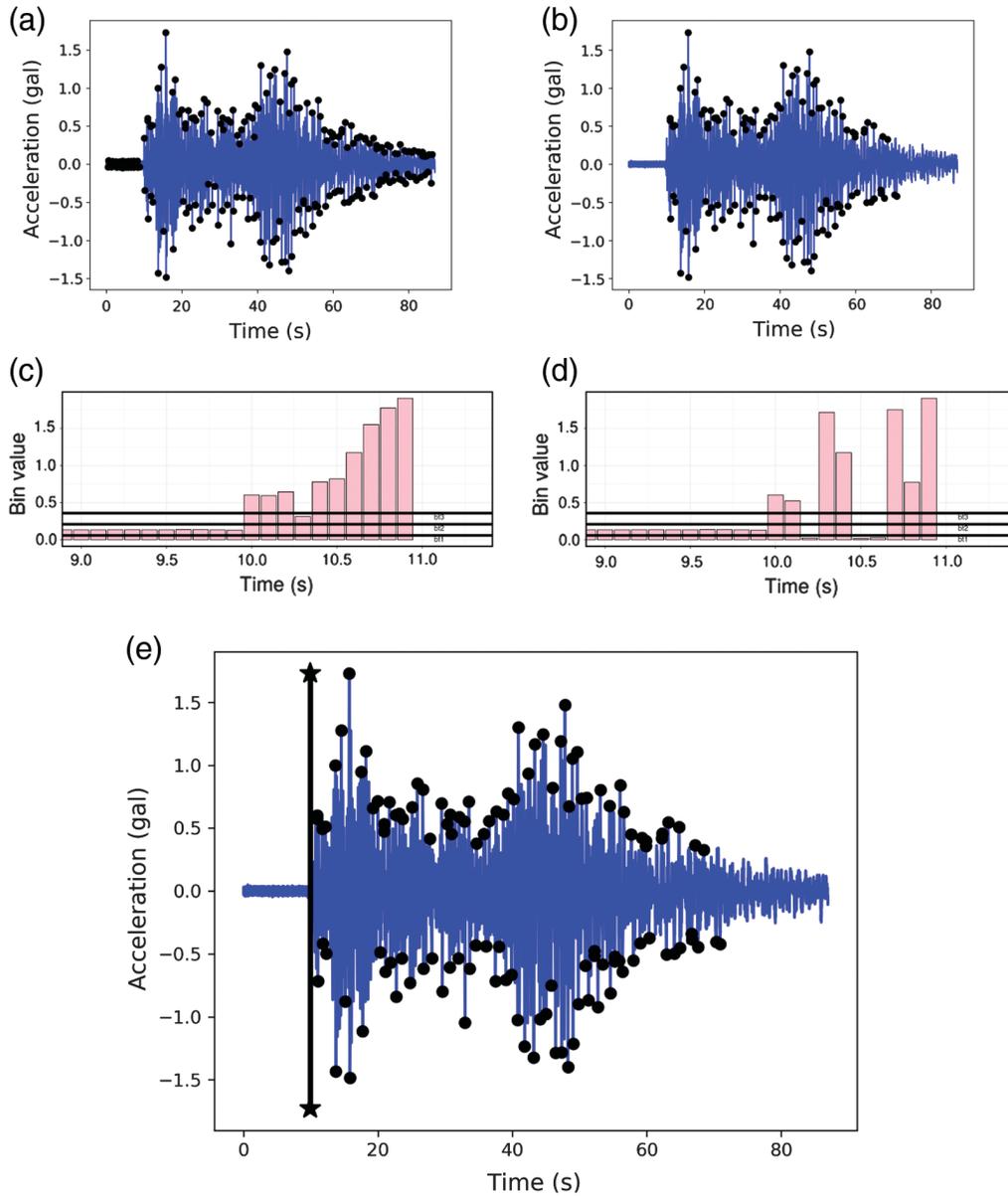


Figure 2. Figures illustrating the steps of our method. (a) Raw peaks and valleys detected using the automatic multiscale-based peak detection (AMPD) algorithm in black dots. (b) List of apparent peaks and valleys in black dots. (c) Bin distribution with time and the three designed bin thresholds. (d) A problematic bin distribution that will cause trouble in our method. (e) Final result. The solid line marked with stars is the final pick.

A global minimum of γ can be found from γ_k , and the corresponding k is called λ

$$\lambda = \arg(\min(\gamma_k)). \tag{4}$$

We then use λ to reshape the LES matrix into a $\lambda \times N$ matrix \bar{m}

$$\bar{m}_{k,i} = m_{k,i}, \text{ for } k \in \{1, 2, \dots, \lambda\} \text{ and } i \in \{1, 2, \dots, N\}. \tag{5}$$

Finally, the peaks and valleys are detected by taking the column-wise standard deviation of the matrix \bar{m}

$$\sigma_i = \sqrt{\frac{1}{\lambda - 1} \sum_{k=1}^{\lambda} \left(\bar{m}_{k,i} - \frac{1}{\lambda} \sum_{k=1}^{\lambda} \bar{m}_{k,i} \right)^2}, \tag{6}$$

for $i \in \{1, 2, \dots, N\}$,

in which all peaks and valleys are found at $\{i | \sigma_i = 0\}$. Figure 2a illustrates the located peaks and valleys as black dots.

Determining the Apparent Peaks and Valleys

From the above procedure, we obtained a full list of peaks and valleys. However, although the AMPD algorithm

has already filtered out any extrema with an incomplete wave shape, we still need to apply supplementary constraints to ensure that all of the extrema are evidently associated with the seismic wave and that the amplitude is on the same order of magnitude as the maximum amplitude. In this work, we determine the apparent peaks and valleys by keeping peaks and valleys with amplitudes that are larger than 20% of the maximum amplitude, that is,

$$\text{extrema}_{\text{apparent}} \geq 0.2 \times (\text{maximum amplitude}), \quad (7)$$

in which $\text{extrema}_{\text{apparent}}$ are the apparent peaks and valleys that are comparable to the maximum amplitude. Another possible option is to use the detector noise level to derive a threshold and eliminate peaks and valleys that are smaller than this threshold. Figure 2b illustrates the results of all apparent peaks and valleys.

When spike noise is expected, the algorithm should include an additional prefilter to handle it. Some of the options are as follows:

- Check the mean of the neighborhood of the extrema. For a seismic wave, we expect some cancellation within the random noise. For noisy spikes, the mean can be quite large. Consequently, the mean should be smaller than a certain threshold. For example, if $x[n_1]$ is located at an apparent peak and $x[n_2]$ is located at a noisy spike,

$$\left| \sum_{n_1-20}^{n_1+20} x[i] \right| < \text{threshold}_{\text{spike,mean}} < \left| \sum_{n_2-20}^{n_2+20} x[i] \right|, \quad (8)$$

in which +20 indicates that the next 20 elements are included; in signal traces with a frequency of 100 Hz, 20 elements correspond to 0.2 s.

- Check the zero-crossing property. For a seismic wave, we expect some regular zero crossing pattern, but this pattern for a noisy spike is quite different. For example, $x[n_1]$ is located at an apparent peak, $x[n_{1z,\text{left}}]$ and $x[n_{1z,\text{right}}]$ are its nearest zero-crossing points to the left and to the right, respectively; accordingly, $x[n_2]$ is located at a noisy spike, and $x[n_{2z,\text{left}}]$ and $x[n_{2z,\text{right}}]$ are its nearest zero-crossing points. We can write down a corresponding threshold

$$(n_{1z,\text{right}} - n_{1z,\text{left}}) < \text{threshold}_{\text{spike,zero}} < (n_{2z,\text{right}} - n_{2z,\text{left}}). \quad (9)$$

Tracing the Apparent Signal Back to the *P*-Wave Arrival Time

After the above process, we obtained a list of apparent peaks and valleys. The first peak or valley is located in the neighborhood of the first *P*-wave arrival. Thus, we need to trace the signal back to the *P*-wave arrival time from the first peak or valley. We collect the square of the amplitude (a^2) before the first peak or valley in bins of 0.1 s (or every 10

samples) in reverse, that is, for data with a sampling interval of 100 samples per second,

$$\begin{aligned} \text{bin}_{1s} &= \text{Amp}_{1.00s}^2 + \text{Amp}_{0.99s}^2 + \text{Amp}_{0.98s}^2 + \dots + \text{Amp}_{0.91s}^2, \\ \text{bin}_{0.90s} &= \text{Amp}_{0.90s}^2 + \text{Amp}_{0.89s}^2 + \text{Amp}_{0.88s}^2 + \dots + \text{Amp}_{0.81s}^2. \end{aligned} \quad (10)$$

After we acquire all of the bins, we establish a threshold to trace the waveform back to the *P*-wave arrival time. In this work, we assign three bin thresholds in terms of the means and standard deviations of the noise bins. These three bin thresholds (denoted bt1, bt2, and bt3) used in this work are

$$\begin{aligned} \text{bt1} &= \overline{\text{noise bin}} + 3\sigma(\text{noise bin}), \\ \text{bt2} &= \overline{\text{noise bin}} + 4\sigma(\text{noise bin}), \\ \text{bt3} &= \overline{\text{noise bin}} + 5\sigma(\text{noise bin}), \end{aligned} \quad (11)$$

in which $\overline{\text{noise bin}}$ is the mean of a noise bin and $\sigma(\text{noise bin})$ is the standard deviation of a noise bin. We then use all three acquired thresholds to find the arrival time. In practice, one can set any bin threshold based on the noise level in the record as long as it is possible to distinguish the noise from the signal. To identify the arrival time, we find the first bin satisfying the criterion in which the next three consecutive bins are all smaller than the bin threshold value (bt). If N is the found bin, then it must satisfy the following requirements:

$$\begin{aligned} \text{bin}_N &\geq \text{bt}, \\ \text{bin}_{N-0.1s} &< \text{bt}, \\ \text{bin}_{N-0.2s} &< \text{bt}. \end{aligned} \quad (12)$$

The details of the algorithm used to locate *P*-wave arrival times are shown in Algorithm 1.

Figure 2c illustrates this process. All of the bars represent the bin values at different times. The three thresholds are established as horizontal lines, in which bt1 is the lowest and bt3 is the highest. According to the above algorithm, we first try to use the lowest threshold (bt1) as the finding criterion. The signal is traced back to the first apparent bin centered on 10.9 s. Unfortunately, no bin can satisfy equation (12), and thus, the program traces the signal back to the very beginning of the record. Then the program selects the second threshold (bt2) as the finding criterion; this time, the program finds the bin centered on 10 s and returns it as the picked arrival time.

The quality of the data and the compatibility of the given bin width both affect the quality of the returned pick. Figure 2d illustrates a case of failure. In this case, the bins suffer from a random noise distribution and consequently fail to increase with time monotonically; as a result, the program will return a wrong pick at 10.7 s. One possible way to fix this problem is to increase the bin width, for example, from 0.1 to 0.2 s. Doing so will smooth the bins, and they will increase monotonically with time.

Data: Amplitude-squared bin values, bt1, bt2, bt3

Result: Arrival time

Set thresholdlist=[bt1, bt2, bt3];

Set n=index of first apparent peak/valley bin;

for *threshold* **In** *thresholdlist* **do**

while *n* is not at the beginning **do**

if found bin fits the requirements **then**

 Return time of the found bin as pick;

end

n = *n* - 1;

end

end

/* Fallback, if nothing is found

*/

Return time of the first apparent peak/valley bin;

Algorithm 1. Tracing back to arrival time.

manually. The total number of traces used in this study is 2000, and the details of each event can be found in Table 1.

Determining the Reference Arrival Time Using the AIC within a Manual Window

Here, we determine the reference arrival time to lower the impacts of manual biases on determining specific arrival times generated by assuming that human eyes are sufficiently trustworthy to determine a transition from noise to signal. After a window is assigned manually, the AIC can determine the best-fitting arrival time as the reference pick.

STA/LTA Method

To verify that our LES method can consistently outperform the generic STA/LTA technique, we compare our method with the optimum STA/LTA method. To locate the optimum STA/LTA ratio, we need to find the best-fitting parameters. Five parameters are used in the STA/LTA method, namely, C_1 , C_2 , C_3 , C_4 , and C_5 . For a trace $x[i]$, C_1 and C_2 control the CF through X_2 and dX_2 (Baer and Kradolfer, 1987)

$$\begin{aligned} X_2 &= C_1 X_1 + (x[i] - x[i - 1]), \\ dX_2 &= \frac{C_2(x[i] - x[i - 1])}{dt}, \\ CF &= X_2^2 + dX_2^2, \end{aligned} \tag{13}$$

Data and Results

Dataset

We collect strong-motion traces of 26 recent medium- to high-magnitude earthquakes ($M_L > 4.6$) from the NIED. We choose traces for which we are able to pick the arrival times

Table 1
Twenty-Six Events Used in This Study

Number	Origin Time (yyyy/mm/dd)	UTC (hh:mm)	Latitude (°N)	Longitude (°E)	Depth (km)	M_L	Number of Traces Used
1	2016/10/16	16:37	38.323	141.527	20	5.2	61
2	2016/10/20	11:50	35.862	140.523	37	5.3	63
3	2016/10/21	14:07	35.380	133.855	11	6.6	175
4	2016/11/12	06:43	38.463	141.607	58	5.9	92
5	2016/11/19	11:48	33.842	135.463	51	5.4	87
6	2016/11/21	06:59	35.877	140.960	17	5.0	59
7	2016/11/22	05:59	37.353	141.603	25	7.3	218
8	2016/11/22	06:39	37.243	141.377	22	5.7	25
9	2016/11/22	23:04	37.178	141.443	28	5.7	48
10	2016/11/24	06:23	37.173	141.345	24	6.2	83
11	2016/12/20	01:41	37.308	141.638	33	5.5	41
12	2016/12/28	21:38	36.720	140.573	11	6.3	147
13	2016/12/31	05:08	37.355	141.408	27	5.4	58
14	2017/01/05	00:44	36.862	140.978	56	5.3	78
15	2017/01/05	02:53	37.122	141.358	26	5.6	50
16	2017/01/28	22:40	39.760	140.767	151	5.0	46
17	2017/02/11	06:05	37.177	141.718	41	5.4	46
18	2017/02/19	18:19	35.730	140.663	52	5.4	82
19	2017/02/28	16:49	37.513	141.367	52	5.7	100
20	2017/03/02	23:53	32.645	132.132	37	5.3	60
21	2017/03/12	04:57	37.512	141.542	46	5.4	48
22	2017/04/12	03:10	36.160	140.100	54	4.6	86
23	2017/04/30	23:42	42.322	143.070	53	5.4	39
24	2017/05/27	22:50	36.065	135.543	17	5.1	48
25	2017/06/20	23:27	32.900	132.100	40	5.0	49
26	2017/06/25	07:02	35.900	137.600	0	5.7	111

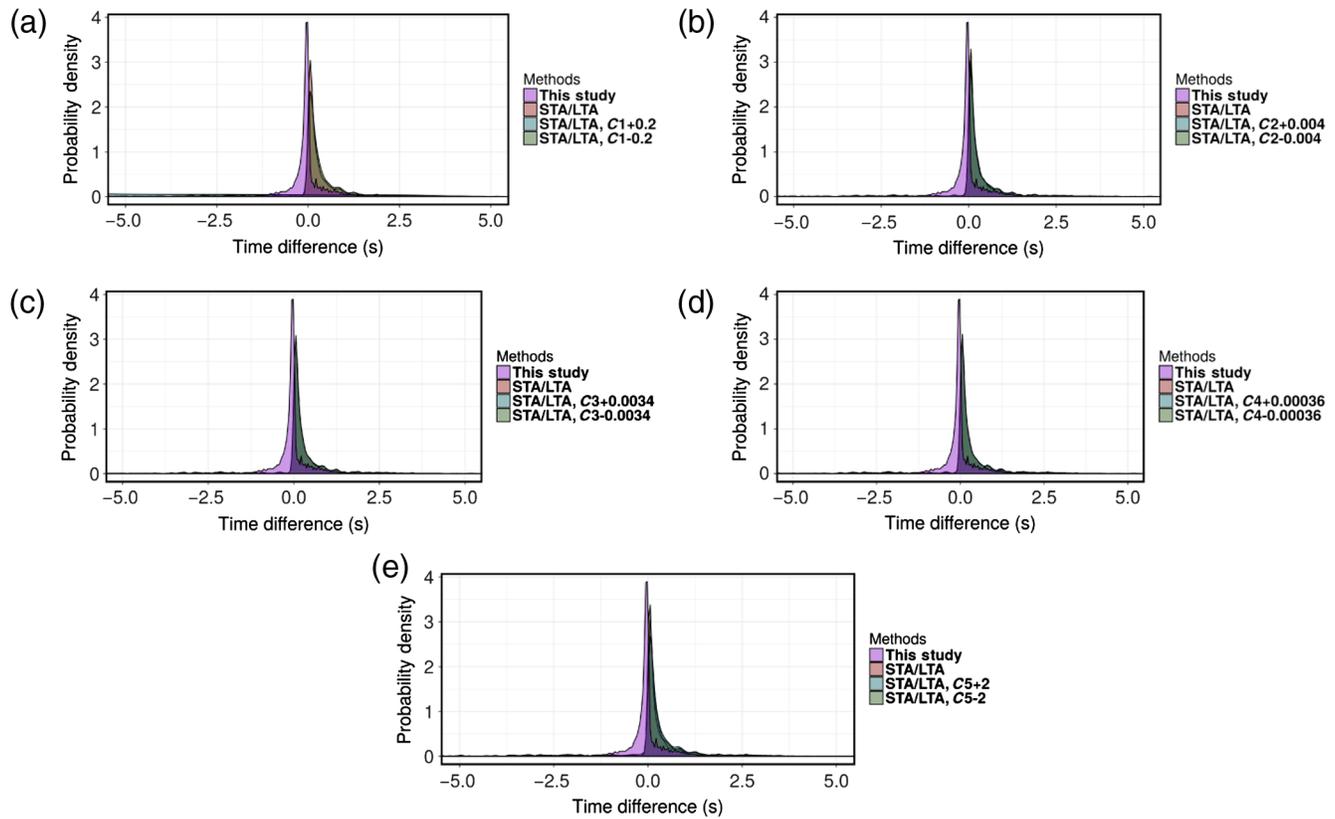


Figure 3. The probability distribution functions (PDFs) of the time differences for 20% perturbations of the optimal C parameter values. (a) C_1 , optimum value: 0.995; (b) C_2 , optimum value: 0.02; (c) C_3 , optimum value: 0.165; (d) C_4 , optimum value: 0.00181; and (e) C_5 , optimum value: 10.0.

in which X_1 and X_2 are iterative terms starting from $x[0]$, and dt corresponds to the sampling interval.

After the CF is defined, the STA/LTA ratio can be written as:

$$R = \frac{\text{STA}[i]}{\text{LTA}[i]} = \frac{\text{STA}[i-1] + C_3(\text{CF}[i] - \text{STA}[i-1])}{\text{LTA}[i-1] + C_4(\text{CF}[i] - \text{LTA}[i-1])}, \quad (14)$$

in which C_3 is designed to be larger than C_4 to provide a greater weight to short-term changes. If $R > C_5$ at any point, the value of i at that point will be retained and further considered to be the P -wave arrival time.

To find the optimum values of C_1 – C_5 that can yield best picking result based on the given dataset, we scan through all possible parameter combinations in the coefficient parameter space. The results are judged based on the shape of the probability distribution function (PDF) each parameter combination produced. We further plot a series of comparisons in Figure 3 between the proposed method and the optimum coefficients with a 20% perturbation. It shows that the optimum coefficients give the best STA/LTA result.

AIC Method

The AIC method considers the difference in the standard deviations between two groups (i.e., the noise and the signal)

and then defines a function to calculate the information criterion. In principle, the P -wave arrival time coincides with the minimum information criterion, at which the noise and seismic-wave signal separate into two groups. In this article, we will use the STA/LTA method to determine the effective window. Let a trace inside the effective window have N samples and assume that the autoregressive filter length is small compared with N . At the k th position, assuming the order is negligible with respect to the length of the noise or signal, the AIC can be written as (Sleeman and van Eck, 1999):

$$\text{AIC}(k) = k \log(\sigma_{\text{Noise}, 1 \text{ to } k}^2) + (N - k) \log(\sigma_{\text{Signal}, k+1 \text{ to } N}^2), \quad (15)$$

in which σ is the standard deviation of noise and signal, respectively. The P -wave arrival time appears at the k_p th position

$$\text{AIC}(k_p) \leq \text{AIC}(k), \quad \forall 1 \leq k \leq N. \quad (16)$$

PDF of the Arrival-Time Difference

Throughout this study, we present our results as a PDF of the time difference between the pick of the given method and the reference pick (Figs. 3–5). This approach effectively reveals the disparity between the reference arrival time and

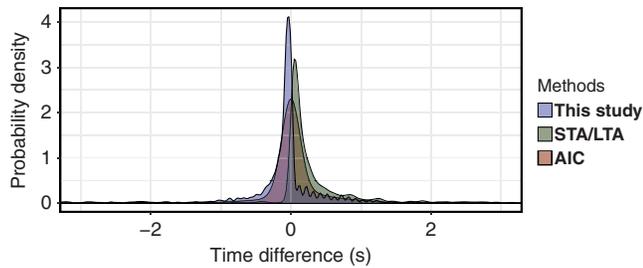


Figure 4. A comparison between the PDF of the deviation in the picking time from the reference arrival time for the proposed method and those for the short-term average/long-term average (STA/LTA) and Akaike information criterion (AIC) methods.

the result of each method. Whereas a positive time difference (pointing to the right) represents a late pick with respect to the reference pick, a negative time difference (pointing to the left) represents an early pick with respect to the reference pick.

Comparisons among the Methods

Figure 4 provides a direct comparison between the PDF of our method and the PDFs of the optimum STA/LTA and AIC methods. Table 2 describes the results of a statistical analysis among the three methods. Furthermore, Figures 6 and 7 show a series of traces with different magnitudes and different environments. Figure 7 also shows some common fail picks of all three methods. Figure 7c represents the STA/LTA at fallback when it cannot be triggered by any signal in the trace. Figure 7e represents the STA/LTA pick on the noise before the *P*-wave arrival. Figure 7g represents the inappropriate early pick of the AIC due to the early effective window. Figure 7i represents the inappropriate early pick of the proposed method. This inappropriate early pick happens when the emergent signal is long and all of bins are still above the threshold.

Different Groups of Signal-to-Noise Ratios

We define the signal-to-noise ratio (SNR) on a decibel scale as

$$\text{SNR}_{\text{dB}} = 20 \log_{10} \left(\frac{A_{\text{Max signal}}}{A_{\text{Max noise}}} \right), \quad (17)$$

in which $A_{\text{Max signal}}$ is the maximum amplitude after the reference arrival time and $A_{\text{Max noise}}$ is the maximum amplitude before the reference arrival time.

We further categorize the dataset into three groups based on the SNR, namely, high SNR ($\text{SNR}_{\text{dB}} > 60$), medium SNR ($30 \leq \text{SNR}_{\text{dB}} \leq 60$), and low SNR ($\text{SNR}_{\text{dB}} < 30$). Table 2 shows the statistical analysis of all three groups, and Figure 5 shows the results for these three SNR groups.

Common Picking Difficulties

Figure 8 shows a series of traces in which picking errors occurred because of some difficulties. In Figure 8a, the

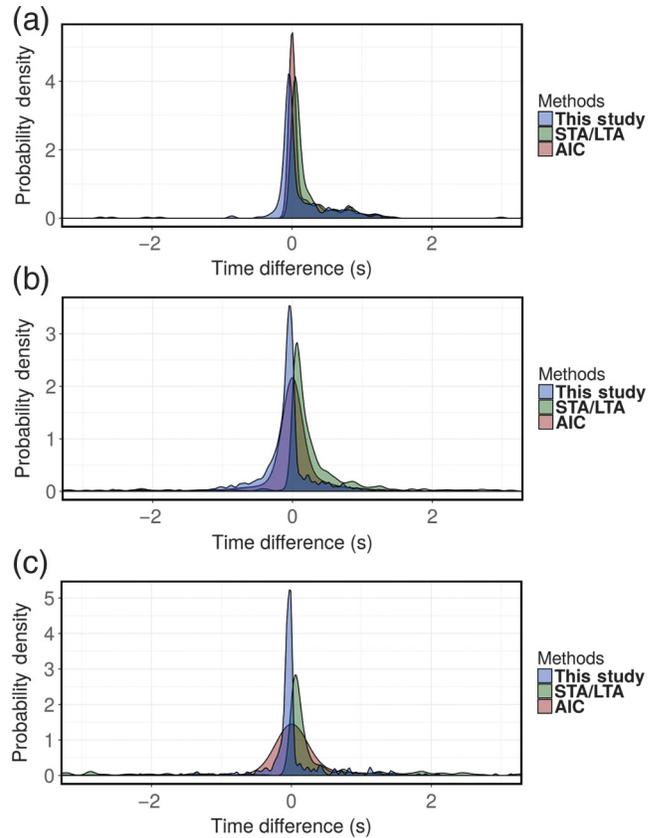


Figure 5. The PDFs of the time differences in three signal-to-noise ratio (SNR) groups. (a) High SNR, (b) medium SNR, and (c) low SNR.

STA/LTA method and AIC method pick the arrival time within the fluctuating noise before the arrival. In Figure 8c, a triangle-shaped emergent arrival exists within the trace, hindering the determination of the arrival time; because all three picks are similar, it is difficult to tell which method performs the best. In Figure 8e, a series of noisy spikes affects the trace. Our method uses a given threshold in phase one of the above-mentioned procedures to avoid errors in the picking associated with spikes, but the STA/LTA and AIC methods will both pick the wrong time.

Discussion

In this article, we developed an LES-based picking program that exhibits smaller standard deviations of time difference than the STA/LTA method and the AIC method (Fig. 4 and Table 2). In all three SNR groups, our method demonstrates a narrower standard deviation and a smaller mean (Table 2 and Fig. 5). High SNR of Figure 5 shows a higher peak at zero for AIC method. This high peak is due to many zeros in time difference for the AIC method; an artifact reflects the similarity between AIC picks and AIC-based reference picks. Also, this similarity can explain the bell shape in PDF of AIC and all the median of AIC are zeros.

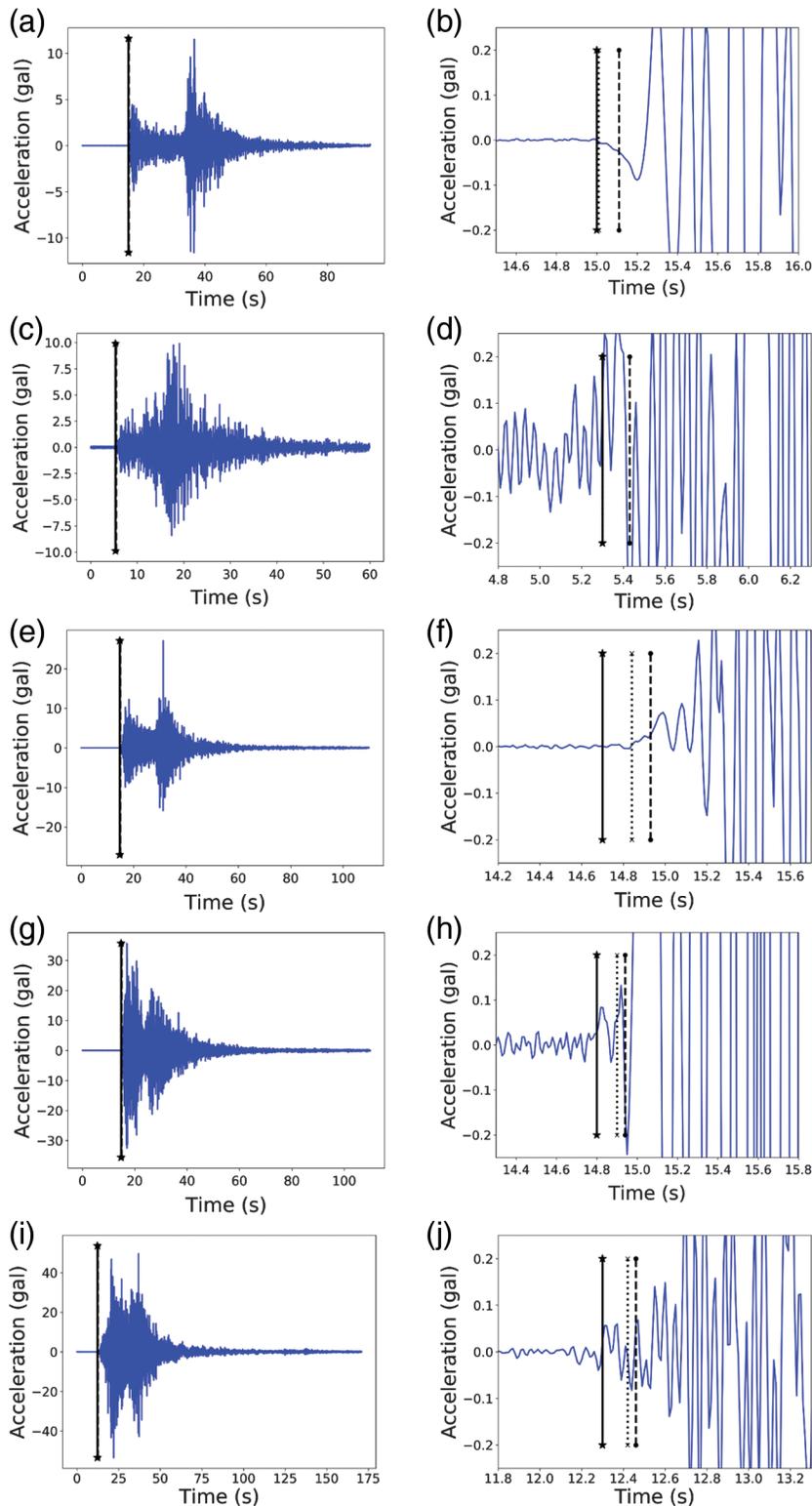


Figure 6. A gallery of traces with different magnitudes and different environments. The solid line marked with stars shows the picked arrival time using our method, the long dashed line marked with circles shows that using the STA/LTA method, and the dotted line marked x symbols shows that using the AIC method. All the subfigures (b,d,f,h,j) on the right are the zoom-in of the left. (a) M_L 5.2, quiet; (c) M_L 5.2, noisy; (e) M_L 6.6, quiet; (g) M_L 6.6, noisy; and (i) M_L 7.3, quiet.

Noise Reduction and Error Level Estimation

One advantage of our method is that it avoids the picking of P -wave arrival times within noise. During the first stage of our method, a full set of peaks and valleys is selected; this set is further used to ensure that we focus on true seismic signals rather than random noise. Starting from the first apparent peak or valley, the seismic waveform is traced back to the true arrival time. Although the arrival time in this study is in some cases farther away from the reference arrival time than the pick obtained using the STA/LTA method, the first peak or valley is almost always in the neighborhood of the P -wave arrival, and thus, the proposed technique rarely picks a noisy arrival.

Another interesting feature is the binning width, which automatically gives us a benchmark of the precision because it depends on the data quality (see Fig. 2c,d). In contrast, although the STA/LTA method offers two time scales, estimates of the precision of either the STA scale or the LTA scale remain unclear; moreover, the use of the STA scale alone is not appropriate because the LTA scale also plays an important role. Here, the difference between the STA/LTA method and the proposed technique originates from the number of stages used. In the STA/LTA method, there is only one stage, during which the signal is distinguished from the noise and the arrival time is simultaneously determined. The two time scales, which are used to serve multiple purposes, ultimately complicate the definition of the precision. In the proposed algorithm, we separate the procedure into three parts. The first part produces a list of all quasi-periodic peaks and valleys, the second part distinguishes the signal from the noise and depicts the trace as peaks and valleys, and the third part selects the arrival time. Therefore, it is easier to define the level of precision using the proposed method.

First Motions, Pick Uncertainty, and Pick Quality

The first motions calculated from this method can avoid possible false-positive arrival times. However, the quality of the first motions depends on many things, for

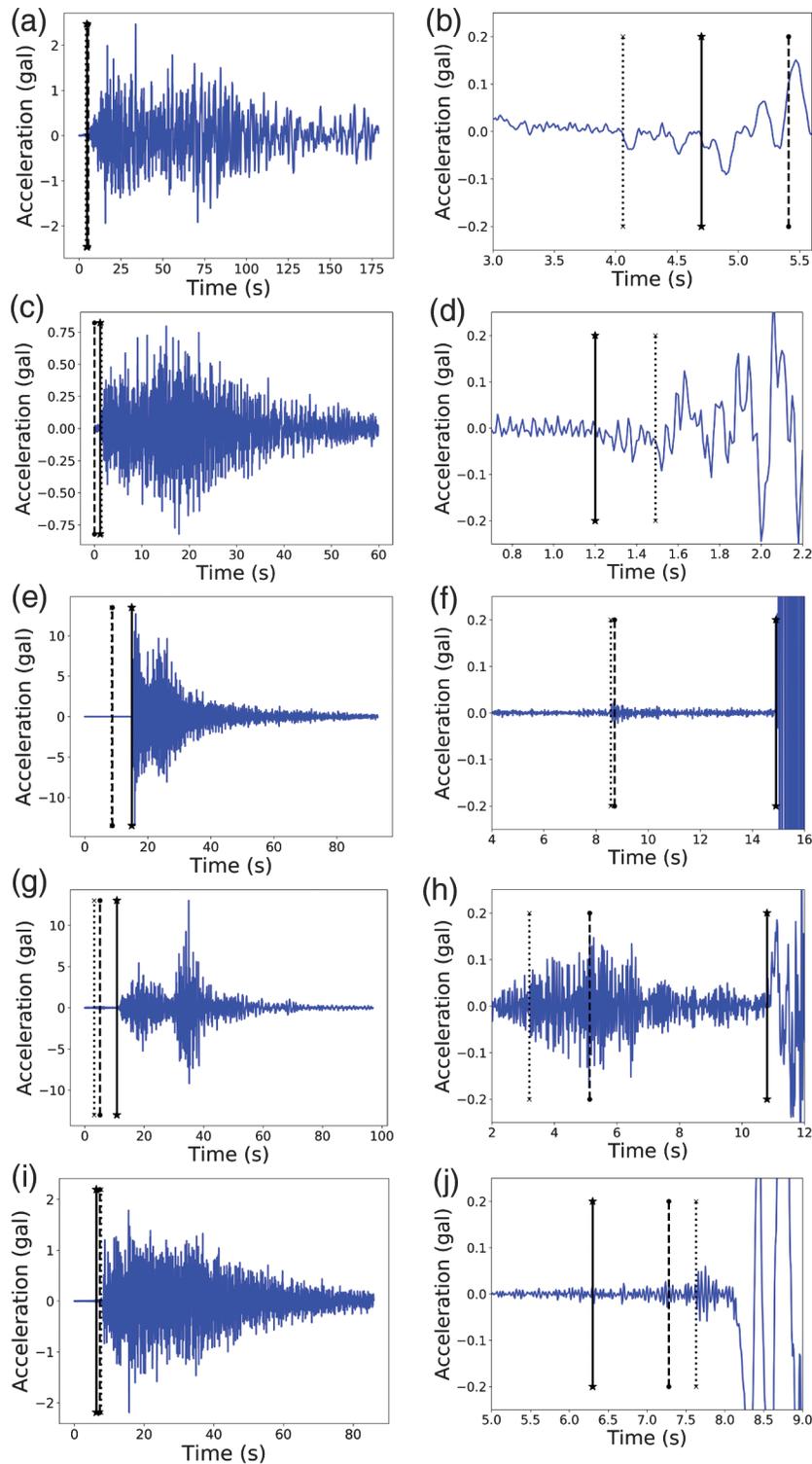


Figure 7. A gallery of traces with some failure cases. The solid line marked with stars shows the picked arrival time using our method, the long dashed line marked with circles shows that using the STA/LTA method, and the dotted line marked x symbols shows that using the AIC method. All the subfigures (b,d,f,h,j) on the right are the zoom-in of the left. (a) M_L 7.3, Noisy; (c) M_L 5.3, STA/LTA fallback case; (e) M_L 5.7, STA/LTA picks on noise; (g) M_L 6.2, AIC and STA/LTA early picks; and (i) M_L 5.8. The proposed method fails, early pick.

example, the window to add up, the filter used, and the allowed frequency band. The pick uncertainty is best represented by the standard deviation. And the bin width represents the resolution width, that is, the highest possible achievable precision. In Table 2, all the standard deviation is larger than the resolution. In terms of pick quality, we can separate them according to their SNR. The high-SNR group should be quality 0, the medium-SNR group should be quality 1, and the low-SNR group should be quality 2.

Tendency of Early Picks

Figures 4 and 5 show that the proposed method has a tendency to produce early picks, compared with STA/LTA and AIC. We think this phenomenon reflects the fact that there are many traces with emergent oscillation before P arrival and that can lead to early picks. When the long duration of emergent oscillation presented, it can result in a fail pick, like Figure 7i. Although it is possible to modify the thresholds to achieve a better result in long emergent oscillation cases, the trade-off is the possibility to deteriorate other picks.

Limitations

There are some limitations to our method. Upon an examination of each trace, we can observe that the presence of a triangle-shaped emergent signal introduces both bias and obscurity into the picks, both of which are difficult to handle in all three methods. Consequently, a further understanding of both the mechanisms and the waveforms of head triangle phenomena is necessary.

Another possible limitation is the presence of noisy spikes in the neighborhood of the P -wave arrival. Although our method can avoid the picking of spikes before the arrival time, it will produce unreliable results when such spikes occur within a head triangle or at times similar to the P -wave arrival time.

Finally, the proposed method needs more signals than STA/LTA. STA/LTA usually needs one or two seconds of P wave to trigger the pick. To have some apparent waves for analysis, the proposed method needs about 3–5 s of P -wave signals to have consistent results.

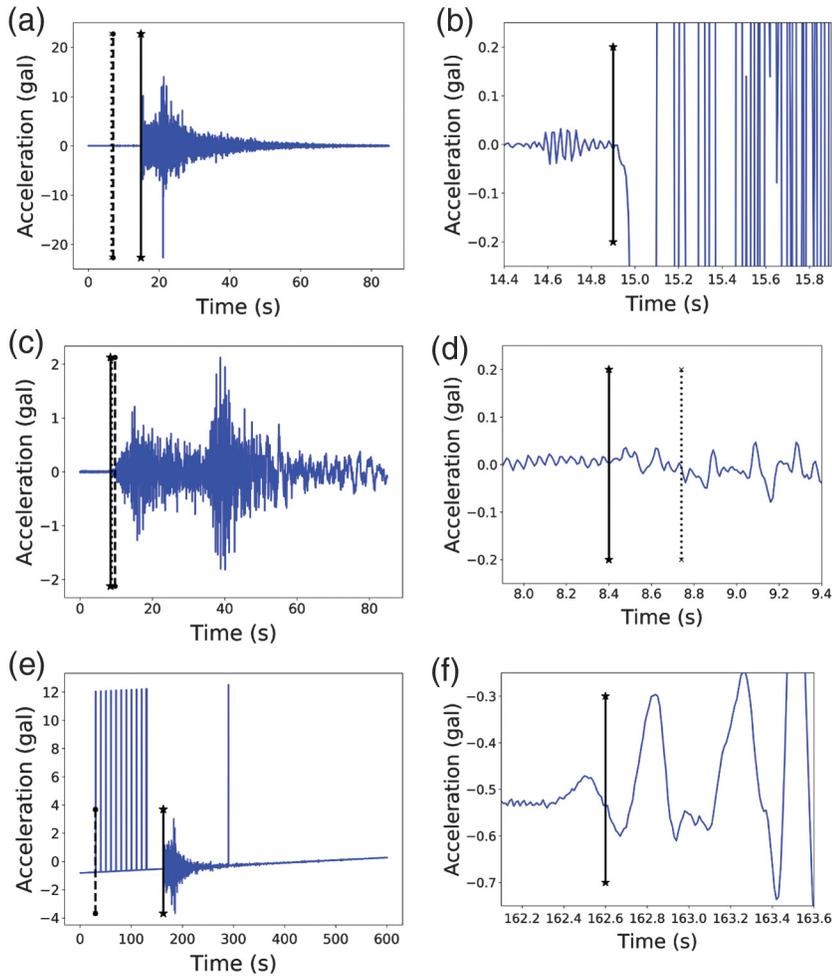


Figure 8. Three common problems in automatic picking. The solid line marked with stars shows the picked arrival time using our method, the long dashed line marked with circles shows that using the STA/LTA method, and the dotted line marked with x symbols shows that using the AIC method. (a) Picking in a noisy environment. SNR is 6.971, (c) triangle-shaped emergent arrival that affects the picking. SNR is 35.926, (e) picking with noisy spikes. SNR is 12.916, (b), (d), and (f) are the zoom-in of (a), (c), and (e), respectively.

Table 2

Statistical Analysis of Time Difference between the Pick of the Given Method and the Reference Pick at Different SNR Groups

Group	SNR _{dB}	Method	Mean (s)	Median (s)	St. dev. (s)
Total	All	This study	-0.024	-0.05	0.824
		STA/LTA	0.162	-0.1	4.023
		AIC	-0.297	0.0	2.896
High	> 60	This study	0.092	-0.02	0.417
		STA/LTA	0.403	0.075	2.121
		AIC	0.073	0.0	1.319
Medium	≥ 30, ≤ 60	This study	-0.088	-0.06	0.849
		STA/LTA	0.083	0.12	3.347
		AIC	-0.374	0.0	2.241
Low	< 30	This study	0.139	-0.03	0.905
		STA/LTA	0.3	0.08	6.587
		AIC	-0.256	0.0	5.081

SNR, signal-to-noise ratio; St. dev., standard deviation; STA/LTA, short-term average/long-term average; AIC, Akaike information criterion.

Conclusions

In this article, we demonstrate the possibility of transforming modified AMPD, an LES-based algorithm aiming to locate the peaks and valleys within a signal, into a robust arrival-time picking program. Numerical tests show that the PDF of the proposed technique consistently results in a smaller standard deviation than the STA/LTA and the AIC methods, especially in the low-SNR case. In addition, we observe the impact of a head triangle within a seismic trace and demonstrate how the presence of this phenomenon can substantially deteriorate the picking results regardless of the method used. The novel method proposed herein will be most suitable for near-real-time scenarios (after at least the P-wave arrival) or for postprocessing scenarios in relatively noisy environments.

Data and Resources

Waveform data from the K-Net seismic network were obtained from the National Research Institute for Earth Science and Disaster Resilience (NIED, <http://www.kyoshin.bosai.go.jp>, last accessed December 2018). Waveform data with noisy spikes were obtained from the National Center for Earthquake Engineering (<https://www.ncree.org>, last accessed December 2018).

Acknowledgments

T. C. H. thanks Li Zhao for his encouraging advice, which later led to this project. T. C. H. thanks Hsiang-Meng Yu for his helpful comments on handling noisy spikes using our method. The authors also thank the anonymous reviewers for their valuable comments to improve this work. Our work was supported by the Ministry of Science and Technology (MOST) of Taiwan under MOST 106-2116-M-002-019-MY3. Our work was also supported by the Research Center of Future Earth of the National Taiwan University (NTU) under Grant Number 107L901002. This work was also financially supported by the NTU Research Center for Future Earth from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan.

References

Ait Laasri, E. H., E. S. Akhouayri, D. Agliz, and A. Atmani (2013). Automatic detection and picking of P-wave arrival in locally stationary noise using cross-correlation, *Digit. Signal Process.* **26**, 87–100, doi: [10.1016/j.dsp.2013.12.009](https://doi.org/10.1016/j.dsp.2013.12.009).

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *Second International Symposium on Information Theory*, Budapest, Hungary, 267–281.

Allen, R. (1982). Automatic phase pickers: Their present use and future prospects, *Bull. Seismol. Soc. Am.* **72**, S225–S242.

- Allen, R. V. (1978). Automatic earthquake recognition and timing from single traces, *Bull. Seismol. Soc. Am.* **68**, 1521–1532.
- Baer, M., and U. Kradolfer (1987). An automatic phase picker for local and teleseismic events, *Bull. Seismol. Soc. Am.* **77**, 1437–1445.
- Bogiatzis, P., and M. Ishii (2015). Continuous wavelet decomposition algorithms for automatic detection of compressional- and shear-wave arrival times, *Bull. Seismol. Soc. Am.* **105**, no. 3, 1628–1641, doi: [10.1785/0120140267](https://doi.org/10.1785/0120140267).
- Jena, D. P., and S. N. Panigrahi (2015). Automatic gear and bearing fault localization using vibration and acoustic signals, *Appl. Acoust.* **98**, 20–33, doi: [10.1016/j.apacoust.2015.04.016](https://doi.org/10.1016/j.apacoust.2015.04.016).
- Küperkoch, L., T. Meier, J. Lee, W. Friederich, and 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, doi: [10.1111/j.1365-246X.2010.04570.x](https://doi.org/10.1111/j.1365-246X.2010.04570.x).
- Ross, Z. E., and Y. Ben-Zion (2014). Automatic picking of direct P, S seismic phases and fault zone head waves, *Geophys. J. Int.* **199**, no. 1, 368–381, doi: [10.1093/gji/ggu267](https://doi.org/10.1093/gji/ggu267).
- Scafidi, D., A. Vigano, G. Ferretti, and D. Spallarossa (2018). Robust picking and accurate location with RSNI – Picker₂: Real-time automatic monitoring of earthquakes and nontectonic events, *Seismol. Res. Lett.* **89**, no. 4, 1478–1487, doi: [10.1785/0220170206](https://doi.org/10.1785/0220170206).
- Scholkmann, F., J. Boss, and M. Wolf (2012). An efficient algorithm for automatic peak detection in noisy periodic and Quasi-periodic signals, *Algorithms* **5**, 588–603, doi: [10.3390/a5040588](https://doi.org/10.3390/a5040588).
- Sleeman, R., and T. van Eck (1999). Robust automatic P-phase picking: An on-line implementation in the analysis of broad-band seismogram recordings, *Phys. Earth Planet. In.* **113**, 265–275, doi: [10.1016/S0031-9201\(99\)00007-2](https://doi.org/10.1016/S0031-9201(99)00007-2).
- Wu, Y. M., and H. Kanamori (2005). Experiment on an onsite early warning method for the Taiwan early warning system, *Bull. Seismol. Soc. Am.* **95**, no. 1, 347–353, doi: [10.1785/0120040097](https://doi.org/10.1785/0120040097).
- Zhang, H., C. Thurber, and C. Rowe (2003). Automatic P-wave arrival detection and picking with multiscale wavelet analysis for single-component recordings, *Bull. Seismol. Soc. Am.* **93**, no. 5, 1904–1912, doi: [10.1785/0120020241](https://doi.org/10.1785/0120020241).
- Zhu, W., and G. C. Beroza (2018). *PhaseNet: A Deep-Neural-Network-Based Seismic Arrival Time Picking Method*, available at <http://arxiv.org/abs/1803.03211> (last accessed December 2018).

Department of Geosciences
National Taiwan University
No. 1, Section 4, Roosevelt Road
Taipei 10617, Taiwan
tingchunguang@gmail.com
drymwu@ntu.edu.tw

Manuscript received 27 April 2018;
Published Online 2 January 2019