A better automatic body-wave picker with broad applicability.

Frédérick Massin* and Alison Malcolm, Memorial University

SUMMARY

For robust earthquake analysis, we need efficient and reliable automatic body-wave recognition methods. To do this, we combine the advantages of standard methods in an innovative and generalized approach. Using the component energy correlation method, we demonstrate the mathematical and practical advantages of the correlation operator and apply this operator to the $\bar{S}T/\bar{L}T$ and $\bar{R}P/\bar{L}P$ methods. We also implement multi-scale versions of these methods to reduce the dependence on user-defined time-scale parameters. We compare our results systematically to different methods, propose an optimal approach and demonstrate its reliability.

INTRODUCTION

Whether analyzing global tectonic events or micro-earthquakes (natural or from human activity), we require estimates of body-wave arrival times from seismic records. The constant sensitivity and precision of automatic body-wave recognition methods are the basis of a robust earthquake analysis. Although body-wave arrival detection is done by automatic systems, these systems need configuration and review by an operator. Our study proposes a renewed generalized approach to reduce configuration, and to increase the sensitivity and precision of automatic systems resulting in a better earthquake analysis.

The evolution of automatic body-wave recognition methods has followed two main directions. The methods in the first branch are based on characteristic functions (Cf), which give a proxy for the continuous probability of a body-wave arrival from the continuous seismic signal, which is given as input. Such methods are called uninformed because they are based on continuous seismological data, i.e. they have an uninformed prior. They often rely on variants of the Short Term - Long Term averages ratio algorithm ($\bar{S}T/\bar{L}T$ is used in Earthworm and SeisComP3; Allen, 1982; Baer and Kradolfer, 1987; Lomax et al., 2012). In this method, the ratio of two running means (one short and one long time scale) increases with amplitude changes. By definition, the long time scale defines the lower bound on the period of detected body-waves with the short time scale defining the period of best sensitivity. Similar to $\bar{S}T/\bar{L}T$, other uninformed methods are based on an operator (e.g. ratio) by which multiple data streams (e.g. short and long term running means) are combined into one stream, a process we refer to as multiplexing. The resulting Cf can be used for probabilistic earthquake source scanning (Kao and Shan, 2004) or to extract body-wave information (arrival time, amplitude, frequency) for micro-earthquake location (in industrial environment Chen, 2006; Wong et al., 2009) or further analysis.

In addition to amplitude changes, body-wave arrivals are also associated with frequency and polarization changes that may

not be detected by $\bar{S}T/\bar{L}T$. Informed methods can help to detect these events. These methods aim to improve arrival time data by increasing their time precision and the number of detected arrivals, using pre-existing and potentially incomplete arrival time datasets from uninformed methods as prior information. Significant work has been done with informed methods (using autoregression, Wenner and match filters, see Song et al., 2010, for a application to hydrofracture induced microseismic) and open-source standard libraries (obspy, Megies et al., 2011) are now available to combine their advantages. The advantages of informed methods include sub-sample arrival time precision and sensitivity on data with signal to noise ratio below one (with match filters, Chamberlain et al., 2014). However, informed methods require pre-processing, advanced configuration, and are difficult to use for real time applications (as is necessary in industrial or hazard monitoring).

With further work, uninformed methods might be improved for reliable and standard body-wave arrival analysis. Along these lines, we combine the advantages of several Cf from uninformed methods to minimize the necessary configuration parameters, and to improve the sensitivity and precision of automatic arrival time picking. We first investigate the interesting advantages of the component energy correlation method (Nagano et al., 1989), which we call $C\star_{rms}$. To get the same advantages as $C\star_{rms}$, the correlation operator is then generalized for Cf based on multiplexing. Finally, we follow pre-existing work for multi-scaling with multiple Cf to reduce the dependence on time-scale parameters. We compare our results with three Cf ($\bar{S}T/\bar{L}T$, $C\star_{rms}$ and $\bar{R}P/\bar{L}P$, Zahradník et al., 2015) and propose an optimal approach.

METHOD

Component energy correlation method

P-waves produce displacements primarily along the propagation direction, while S-waves produce displacements primarily perpendicular to the propagation direction. Seismic noise by contrast, is a stochastic signal (Wentzell, 1981) and as such, exists on all polarizations. More specifically, the energy dissipation through time of stochastic noise is correlated between all polarization angles. P and S-wave arrivals temporarily decrease the cross-channel energy correlation, a measure we can exploit to detect their arrivals. This component energy correlation method was introduced by (Nagano et al., 1989, $C\star_{rms}$); the estimation of P and S wave arrival times and their uncertainties is described in detail in (Zhizhin et al., 2006; Massin et al., 2009). To account for body-waves incoming along the bisector of the sensor orientation in down-hole acquisition environments, we can use a second instance of $C\star_{rms}$ with corresponding rotated components. $C\star_{rms}$ is the product of $R_{ZE(t,T_C)}$ and $R_{ZN(t,T_C)}$, the correlations between the energy dissipation

SEG International Exposition and 86th Annual Meeting

of the vertical and the two horizontal components given by:

$$R_{j,k(t,T_C)} = \frac{\sum_{i=t}^{t+T_C} E_j(i)E_k(i)}{\sum_{i=t}^{t+T_C} E_j^2(i) \sum_{i=t}^{t+T_C} E_k^2(i)}. \quad (1)$$

The energy dissipation $E_k(t)$ on $k(t)$, a given component of the seismic signal as a function of time t, is defined as a sliding *rms* (root mean square, $\sum_{i=t-T}^{t} k(i)^2$). $C\star_{rms}$ is used as a Cf of a P-wave, $C\star_{rms(E)}$ (function of $R_{EZ(t,T_C)}$) and $C\star_{rms(N)}$ (function of $R_{NZ(t,T_C)}$) are used for S-waves. The correlation coefficient can be used as a probability of arrival time estimate between zero and one (one being worst). The correlation window (T_C) defines the dominant period of the P-waves with best sensitivity (Nagano et al., 1989). The energy dissipation window (T) can be chosen arbitrarily as twice T_C and defines the lowest frequency of sensitivity (Zhizhin et al., 2006).

$C\star_{rms}$ behaves as a multiplexor in which the three directional channels of a seismometer are combined into one Cf. By this design, it is sensitive to polarization changes, a physical property specific to body-waves.

Generalized correlation method

There are several Cf based on the ratio operator. The $\bar{ST}/\bar{LT}$ uses the averaged absolute value of the signal (Allen, 1982) or its envelope (Baer and Kradolfer, 1987). The $\bar{RP}/\bar{LP}$ uses the summed absolute values of the signal in the right and left part of a moving time window (Chen, 2006; Wong et al., 2009; Zahradník et al., 2015). In these Cf, the ratio operator could be substituted by the correlation operator with two advantages: 1) the correlation outputs normalized values, whereas the $\bar{RP}/\bar{LP}$ method gives values dependent on the input data, 2) correlation is a robust approximation of deconvolution.

An operator that estimates normalized probability allows the use of a quantitative triggering threshold and the comparisons of multiple properties that might have different characteristics. The second advantage is based on the mathematical robustness of the correlation operator. The deconvolution of $g(t)$ from $f(t)$ is written in the frequency domain as:

$$\frac{\mathscr{F}\{f\}}{\mathscr{F}\{g\}} = \frac{\mathscr{F}\{f\}\mathscr{F}\{g\}^*}{\mathscr{F}\{g\}\mathscr{F}\{g\}^*}, \quad (2)$$

where $\mathscr{F}$ denotes the Fourier transform, and an * indicates the complex conjugate. Noting that Equation 2 is the Least-Squares inverse of the convolution operator, $\mathscr{F}\{f\}\mathscr{F}\{g\}^*$ is the adjoint of the convolution. In the time domain, the deconvolution of g from f is thus approximated by the cross-correlation of g and f ($\mathscr{F}\{f\star g\} = \mathscr{F}\{f\}\mathscr{F}\{g\}^*$)

The correlation produces a finite, normalized probability; allowing us to combine several correlation-based Cf for improved sensitivity. For example, combining $\bar{ST}\star\bar{LT}$ and $C\star_{rms}$ will give a detector that is sensitive to both amplitude and polarization changes.

Multi-scaling

To adapt to multiple earthquake magnitudes and distances, we use a multi-scale calculation of the Cf following (Lomax et al., 2012). The signal is first decomposed using high pass filters

which are then used for running mean and *rms* calculations. Our formulation of multi-scaled Cf (^MCf) is the product of a range of Cf based on a time scale T as defined in Equation 3:

$$^MCf = Cf_{(t,T_0)}.Cf_{(t,T_1)}...Cf_{(t,T_n)}. \quad (3)$$

Each $Cf_{(t,T)}$ is calculated with the running mean ($\sum_{i=t-T}^{t} \frac{|k(i)|}{T}$) or *rms* ($\sum_{i=t-T}^{t} k^2(i)$) for a frequency band $[\frac{1}{T_n}, \frac{1}{T_{n+1}}]$ using the output of a high pass filter with corner frequency $\frac{1}{T_{n+1}}$. We take advantage of the running mean filter capabilities in the time domain for reducing low frequency signals while retaining a sharp step response (Smith, 1997).
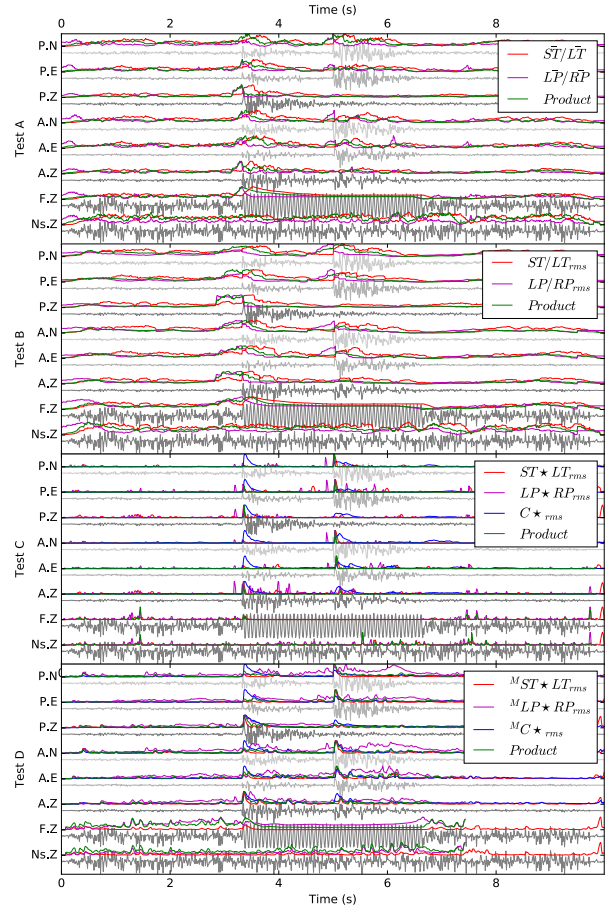

Figure 1: Artificial data tests. Gray: artificial noisy data added to three perturbations each simulating one property of P and S waves. Dark (light) gray: vertical (horizontal) component of ground displacements. Ns.Z : noise only (vertical component of a 3-dimensional random motion) ; F.Z : frequency changes (vertical component); A.(ZEN): amplitude changes (three components) ; P.(ZEN): vertically polarized motion added at 10/3 s and horizontally polarized motion added at 5s (three components). Characteristic functions are represented by overlying colored lines.

Using correlation and multi-scaling, we can combine methods and frequency bands for a robust general detector of body-wave arrivals. The threshold over which a body-wave arrival might be considered is the only required parameter. In the next

section, we show the response of $C\star_{rms}$, generalized correlation methods ($\bar{ST}\star\bar{LT}$ and $\bar{LP}\star\bar{RP}$) and multi-scaling with artificial and real data.

APPLICATIONS

Artificial tests

We use what we call artificial data (because it does not aim to simulate a realistic Earth response) to study individually three properties of body-waves (changes in frequency, amplitude, and polarization) in the presence of seismic noise. Our first key result illustrates that $C\star_{rms}$ has the best sensitivity of the standard approaches, characterized by narrower probability peaks at the wave arrival time. Our second key result shows that correlation-based methods have the best sensitivity to pure frequency changes. Amongst multi-scales approaches, our last key result is that $^{M}C\star_{rms}$ has a similar response to $C\star_{rms}$.

The response of $\bar{ST}/\bar{LT}$ and $\bar{RP}/\bar{LP}$, shown by test A in Figure 1 to all three properties (described in the caption) have similar precision and sensitivity. Using test A and B, we compare the ratio methods based on running mean to the same methods based on *rms*. Although different in details, the responses of ST/LT_{rms} and RP/LP_{rms} in test B are globally similar to $\bar{ST}/\bar{LT}$ and $\bar{RP}/\bar{LP}$. The probability peaks of RP/LP_{rms} are clearer that the other ratio methods. The onset of ST/LT methods are consistent with the location of property changes as is the local maximum for RP/LP methods. Using test B and C, we compare the ratio and the correlation methods using *rms* (Cf_{rms}). The correlation methods have a lower noise level and an improved sensitivity to pure frequency changes, however there is a loss of precision in $LP\star RP_{rms}$. $C\star_{rms}$ has the most impulsive response to amplitude and polarization changes compared to all tested correlation or ratio methods.

The comparison of multi-scale implementations in test D is based on correlation methods and *rms* ($^{M}Cf_{rms}$). The multiple inaccurate peaks of $^{M}LP\star RP_{rms}$ make it overall the least reliable of the tested methods. The noise level and the peak narrowness of $^{M}ST\star LT_{rms}$ and $^{M}C\star_{rms}$ are respectively the best of all implementations of these two methods. $^{M}ST\star LT_{rms}$ has the advantage of being sensitive to all tested properties of body-waves, while $^{M}C\star_{rms}$ is the most sensitive to amplitude and polarization changes amongst the tested multi-scale approaches.

Data test

The use of real data allows visual comparison of manual picking data with automatic methods. We use a set of micro-earthquakes over a range of magnitudes and distances (Figure 2, 3, 4 chosen to be as consistent as possible with properties of industrial experiments). At first glance, this test shows good agreement of the Cf onsets with manual picks. However the data signal to noise ratio affects the response of Cf to body-wave arrivals, particularly for $^{M}ST\star LT_{rms}$.

The consistency between the onsets of tested Cf and the manual P-wave arrival picks is significant through the entire distance range (Figure 2 and Figure 3). Before the arrival, the Cf noise level are stable enough to distinguish a clear onset.
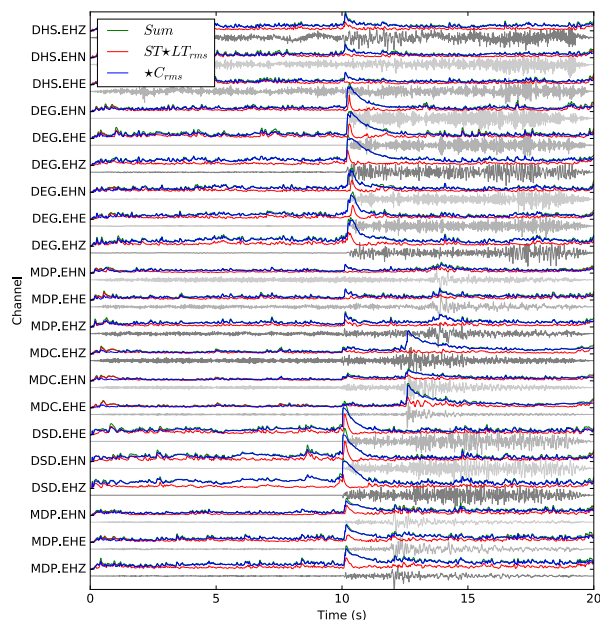

Figure 2: Real data tests for P-waves (S-waves in Figure 3). Gray: data recorded from M2 to M3 earthquakes with increasing distance from 3 to 60 km from bottom to top. Data are normalized by their respective maximum absolute values. The manual P-wave arrival time is centered at 10 s. Various instrument types and signal to noise ratios are shown. Dark (light) gray: vertical (horizontal) component of ground displacements. Characteristic functions are represented by overlying colored lines with an exponential exaggeration of 0.5 (low values are the most exaggerated).
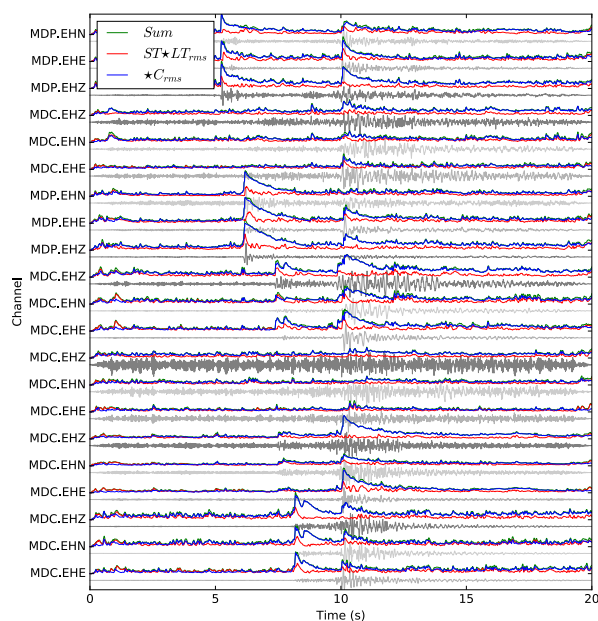

Figure 3: Identical to Figure 2 for S-waves.

It is important to note that the onset of local maximum in the Cf peak is correlated with the body-wave arrival, not the maximum itself. These observations are clearer for $^{M}C\star_{rms}$ than for $^{M}ST\star LT_{rms}$. For $^{M}C\star_{rms}$, the maximum level of noise is below 0.2, which could be used as a general threshold to detect probability peaks associated with body-wave arrivals.

The data noise level is the most important factor affecting the results, but it is not explained by body wave attenuation as function of distance. The worst results are not obtained with data from further distances but rather with exceptionally noisy data in the mid-distance range. The most important effect of distance is the appearance of multiple sub-peaks in the vicinity of the local maximum, reducing the precision of the P-wave arrival pick.
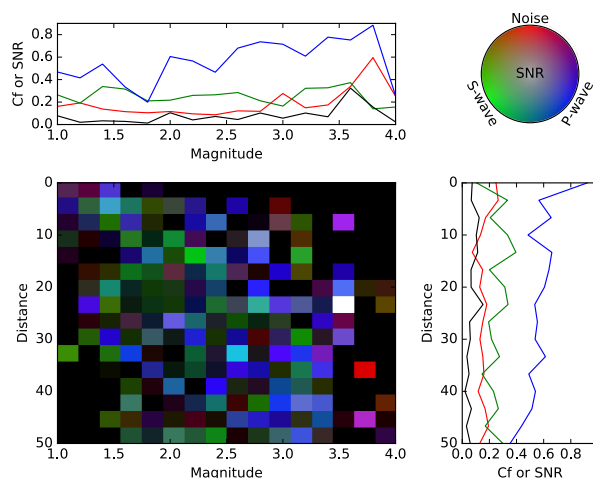

Figure 4: Cf response as a function of magnitude and distance with associated averaged probabilities and colormap. Dark: low Cf values and signal to noise ratio (SNR). White: low Cf values and high SNR. Blue (green): high Cf at P-wave (S) arrival compare to noise level. Red cells: low Cf at arrival.

DISCUSSION

We investigated several Cf formulations in order to find the most reliable way to estimate body-wave arrival times automatically. The methods based on *rms* calculations, polarization changes, and correlation showed the best sensitivities to body-wave arrivals and they can all be implemented in a multi-scale approach for application to a wide range of distances and magnitudes.

The worst responses are obtained with noisy data independent of distance. At constant magnitude below 60 km, data quality from station settings and site effects seem to impact the results more than attenuation or source effects. Data quality appear as a major limitation in our study. $^{M}ST\star LT_{rms}$ appears to be less reliable than $^{M}C\star_{rms}$. The onsets of $^{M}ST\star LT_{rms}$ are significantly less consistent with manual picks and its probability peaks have significantly smaller amplitudes than $^{M}C\star_{rms}$. In the absence of three-component data, $^{M}C\star_{rms}$ cannot be applied and the sensitivity and precision of $^{M}ST\star LT_{rms}$ are fundamental limitations.

The consistency of ratio methods using running mean or *rms*, and the advantages of RP/LP_{rms} have already been shown for industrial applications (Chen, 2006; Wong et al., 2009) or tectonic events (Zahradník et al., 2015). In our broader study, the best responses are obtained with the methods based on *rms* calculations, polarization changes and correlation. This is consistent with the results of previous authors (Nagano et al., 1989; Zhizhin et al., 2006; Massin et al., 2009) but our study goes further. $^{M}C\star_{rms}$ is our best multi-scale implementation and can estimate body-wave arrivals with broadband data without pre-filtering. This method is based on recursive algorithms and does not require high-order statistics. The optimal use of $^{M}C\star_{rms}$ requires three-dimensional data and spreads over all tested distances and magnitudes (M1 to M4, Figure 4).

The arrival time of a body-wave could be approximated with the onset of the probability peak from a reliable Cf. To get further, our method has to be coupled with an onset recognition algorithm. Only local maximums over the threshold of 0.2 (as defined for $^{M}C\star_{rms}$) should be considered. Theoretically, the minimal time span allowed between two peaks is dictated by the period of the first detected signal. In practice multiple peaks between subsequent P and S arrivals might be an issue for earthquake location (there are no such limitations with shift and stack methods, Kao and Shan, 2004; Zahradník et al., 2015). Time derivative and Kurtosis have been shown as reliable onset indicators for seismic wave detection (Baer and Kradolfer, 1987; Lomax et al., 2012); their maximum in the vicinity of the probability peak should give best estimates of arrival times.

CONCLUSION

We combined the advantages of correlation and multi-scaling with the detection of amplitude and polarization change into a single Cf. We obtained a body-wave recognition tool with broad applicability from industrial surveys to earthquake monitoring. We defined an optimal calculation scheme and the related threshold that can then be used for event detection with our approach. We discussed its limitations, which are related to data quality and to the need for 3-dimensional seismic records.

ACKNOWLEDGMENTS

EDITED REFERENCES

Note: This reference list is a copyedited version of the reference list submitted by the author. Reference lists for the 2016
SEG Technical Program Expanded Abstracts have been copyedited so that references provided with the online
metadata for each paper will achieve a high degree of linking to cited sources that appear on the Web.

REFERENCES

Allen, R. V., 1982, Automatic phase pickers: Their present use and future prospects: Bulletin of the
Seismological Society of America, **72**, S225–S242.

Baer, M., and U. Kradolfer, 1987, An automatic phase picker for local and teleseismic events: Bulletin of
the Geological Society of America, **77**, 1437–1445.

Chamberlain, C. J., D. R. Shelly, J. Townend, and T. A. Stern, 2014, Low-frequency earthquakes reveal
punctuated slow slip on the deep extent of the Alpine Fault, New Zealand: Geochemistry,
Geophysics, Geosystems, **15**, 2984–2999, http://dx.doi.org/.1002/2014GC005436.

Chen, Z., 2006, A multi-window algorithm for real-time automatic detection and picking of P-phases of
microseismic events: CREWES Research Report, **18**.

Kao, H., and S.-J. Shan, 2004, The Source-Scanning Algorithm: Mapping the distribution of seismic
sources in time and space: Geophysical Journal International, **157**, 589–594,
http://dx.doi.org/10.1111/j.1365-246X.2004.02276.x.

Lomax, A., C. Satriano, and M. Vassallo, 2012, Automatic picker developments and optimization:
FilterPicker — A robust, broadband picker for real-time seismic monitoring and earthquake early
warning: Seismological Research Letters, **83**, 531–540,
http://dx.doi.org/10.1785/gssrl.83.3.531.

Massin, F., V. Ferrazzini, P. Bachèlery, and Z. Duputel, 2009, A real time process for detection,
clustering, and relocation of volcano-tectonic events at Piton de La Fournaise volcano: Presented
at the American Geophysical Union, Fall Meeting.

Megies, T., M. Beyreuther, and R. Barsch, 2011, ObsPy — What can it do for data centers and
observatories? Annals of Geophysics, **54**, 47–58.

Nagano, K., H. Niitsuma, and N. Chubachi, 1989, Automatic algorithm for triaxial hodogram source
location in downhole acoustic emission measurement: Geophysics, **54**, 508–513,
http://dx.doi.org/10.1190/1.1442677.

Smith, S. W., 1997, Moving average filters, *in* The scientist and engineer's guide to digital signal
processing, 277–284.

Song, F., H. S. Kuleli, M. N. Toksöz, E. Ay, and H. Zhang, 2010, An improved method for hydrofracture-
induced microseismic event detection and phase picking: Geophysics, **75**, no. 6, A47–A52,
http://dx.doi.org/10.1190/1.3484716.

Wentzell, A. D., 1981, A course in the theory of stochastic processes: McGraw-Hill International.
Advanced book program: McGraw-Hill.

Wong, J., L. Han, J. Bancroft, and R. Stewart, 2009, Automatic time-picking of first arrivals on noisy
microseismic data: Presented at the CSEG Microseismic Workshop, CREWS, University of
Calgary.

Zahradník, J., J. Janský, and V. Plicka, 2015, Analysis of the source scanning algorithm with a new P-
wave picker: Journal of Seismology, **19**, 423–441.

Zhizhin, M. N., D. Rouland, J. Bonnin, A. D. Gvishiani, and A. Burtsev, 2006, Rapid estimation of
earthquake source parameters from pattern analysis of waveforms recorded at a single three-
component Broadband Station, Port Vila, Vanuatu: Bulletin of the Seismological Society of
America, **96**, 2329–2347.