A multi-window algorithm for real-time automatic detection and picking of P-phases of microseismic events

Zuolin Chen and Robert R. Stewart

ABSTRACT

There exist a variety of algorithms for the detection and picking of a seismic event in real-time seismic monitoring. However, all have some limitations. We propose and test a multi-window algorithm for the automatic detection and picking of impulsive P-phases of seismic events in low *SNR* (signal-to-noise ratio) environments. This method employs both the instantaneous and the averaged absolute amplitude of traces in several time windows before and after each time point (sample) as the characteristic functions. When the instantaneous absolute value of a characteristic function exceeds an automatically adjusted dynamic threshold, ratios based on the averages of the windows over time samples provide parameters to differentiate an expected event from unwanted noise. Examination of the algorithm by using synthetic and real data shows that the picking accuracy of impulsive first arrivals can be less than 1-2 samples even when the signal-to-noise ratio is lower than 3.

INTRODUCTION

Effective detection and accurate picking of the arrival time of the P-phase of a seismic event is of considerable importance for the automatic real-time monitoring and location of earthquakes from observation networks. Many algorithms for the detection and picking of the arrival time of a seismic event are carried out in the time-domain (e.g., Allen, 1978, 1982; Baer and Kradolfer, 1987; Dai and MacBeth, 1997; Der and Shumway, 1999; Kanasewich, 1981; Morita and Hamaguchi ,1984; Sleeman and Eck, 1999; Takanami and Kitagawa, 1988; Zhao and Takano 1999). Among the time-domain approaches, the short-term average over long-term average ratio (*STA*/*LTA*) automatic picker or/and its modified versions are most widely used, in which the absolute amplitude, the power, or the envelope of the seismic trace is usually selected as the characteristic function for the calculation of *STA* and *LTA* (Allen, 1978, 1982; Baer and Kradolfer, 1987; Kanasewich, 1981). The basic idea of the algorithm is that an event is considered detected when the *STA*/*LTA* ratio exceeds a pre-defined threshold. These approaches are fairly well suited for the arrival detection of seismic events, but have two inherent weaknesses: 1) possibly inaccurate arrival times due to the length of the *STA* window, and 2) difficulty in distinguishing events from high-amplitude noise. The remaining time-domain methods either use autoregressive methods (Morita and Hamaguchi, 1984; Takanami and Kitagawa, 1988), fractal-based algorithms (Boschetti et al., 1996), or neural networks (Dai and MacBeth, 1995; Zhao and Takano, 1999). However, most of the methods are based on the pre-condition that the events have been detected (by *STA/LTA*) and the approximate times of the main arrivals are already known.

In this paper, a multi-window algorithm is proposed, which effectively detects and picks the P-phase first arrivals at various levels of noise based on the instantaneous absolute values of the amplitude of seismic traces and their averages over three running time windows. This method can also be regarded as an improved *STA*/*LTA* algorithm

which combines the functions of both effective detection and accurate picking. In addition, a waveform correction is introduced to compensate for the time delay associated with the trigger threshold. This multi-window procedure resembles that of a human operator in the detection and picking of the onsets of seismic events, and is comparable in accuracy to the human operator.

METHODOLOGY

The multi-window automatic phase picker operates in the time-domain, and is appropriate for a single trace. To pick the first arrival of a P-phase, it includes procedures for defining time windows, metrics, corresponding thresholds and waveform correction for the acquisition of a more accurate first arrival time.

Time windows and metrics

The definitions of the multiple running time windows are mainly tailored to the purpose of measuring the averages of the signal-to-noise within certain time extents just before, after and after a delay of an instantaneous time point (sample). The averages of absolute values of a characteristic function within *BTA* (Before Term Average), *ATA* (After Term Average) and *DTA* (Delayed Term Average) windows are respectively defined as follows:

$$\overline{BTA(t)} = \sum_{i=1}^{m} | u(t-i) | / m \tag{1}$$

$$\overline{ATA(t)} = \sum_{j=1}^{n} | u(t+j) | / n \tag{2}$$

$$\overline{DTA(t)} = \sum_{k=1}^{q} | u(t+k+d) | / q \tag{3}$$

Here, $u(t)$ represents the amplitude or the square root of the power of seismic traces at a time point t; m, n, *and* q are lengths of the running time windows in samples respectively; d the time delay for the *DTA* window.

The instantaneous absolute value $|u(t)|$ is selected as the first metric $R_1(t)$ to judge the arrival of a possible high amplitude/power event or noise. Two other metrics, $R_2(t)$ and $R_3(t)$, are introduced based on the ratios of averages of *ATA* and *DTA* over *BTA*, and take the forms

$$R_2(t) = \overline{ATA(t)} / \overline{BTA(t)} \tag{4}$$

$$R_3(t) = \overline{DTA(t)} / \overline{BTA(t)} \,. \tag{5}$$

The functions of $R_2(t)$ and $R_3(t)$ are mainly designed to discriminate between high-amplitude short-duration and long-duration noise, respectively. The definitions for $R_2(t)$ and $R_3(t)$ are based on the following rationale:

When the first metric, $R_1(t)$, exceeds a pre-defined threshold $H_1(t)$ (see below) at a time point t, the second metric, $R_2(t)$, is used to separate high-amplitude but short-duration noise from a high-amplitude but long-duration event. Equation (4) shows that when the *ATA* window contains noise, and the length of the window is several times longer than the noise, the $R_2(t)$ value will only be weakly affected by the existence of the noise; in contrast, when the *ATA* window is occupied by a part or all of an event, the $R_2(t)$ value will be much amplified. Based on this feature of $R_2(t)$, short duration noise will be removed. However, when the duration of high-amplitude noise is comparable to the length of the *ATA* window, $R_2(t)$ will lose its ability to distinguish an event from noise. This difficulty can be addressed by moving the *ATA* window backward a certain time interval as a *DTA* window, yielding equation (5). An appropriate delay time will reduce the average of the *DTA* window and hence $R_3(t)$ to an expected level. In this way, false triggers caused by high-amplitude long-duration noise can be avoided.

Thresholds

Thresholds pertaining to the three metrics can be pre-defined based on the expected *SNR*. An optimal first threshold, $H_1(t)$, should be set larger than the fluctuations of most noise but lower than expected events. This demand can be satisfied by measuring the mean (E_m) and standard deviation (E_{sd}) of the envelope ($E(t)$) of the pre-existing noise level of a seismic trace within a *BTA* window. A convenient way to acquire the envelope $E(t)$ is using the absolute value of the Hilbert transform of the seismic trace. In practice, $H_1(t)$ is often shifted several samples backwards to avoid the unnecessarily early rising of $H_1(t)$ caused by the first arrival. Based on the factors, an instantaneous $H_1(t)$ is established as

$$H_1(t) = E_m(t-p) + \alpha E_{sd}(t-p) \cdot \tag{6}$$

Here, p is the number of shifted samples; α is a coefficient to adjust the height of the first threshold, and is often taken to be 3, whereby approximately 99% of the noise will not exceed the first threshold statistically. From equation (6), it is obvious that $H_1(t)$ is automatically adjusted with the variance of the level of the background noise.

Thresholds $H_2(t)$ and $H_3(t)$ corresponding to metrics $R_2(t)$ and $R_3(t)$ are generally set to be 3/4 of the level of the expected *SNR* of events. It is noted that extraordinarily high or low specification of $H_2(t)$ and $H_3(t)$ will either lead to failure of the detection of events, or the triggering by unwanted noise.

Waveform correction

As $H_1(t)$ is defined to be larger than most pre-existing noise levels, and further, the instantaneous absolute value at the trigger time point is higher than $H_1(t)$, the trigger time point must be somehow later than the real arrival time according to the configuration of the first arrival of an event. This belated arrival time can be compensated by introducing a concept of waveform correction. For an impulsive first arrival, this process is simply accomplished by using the height of the absolute value and the representative gradient at the trigger point (Figure 1).

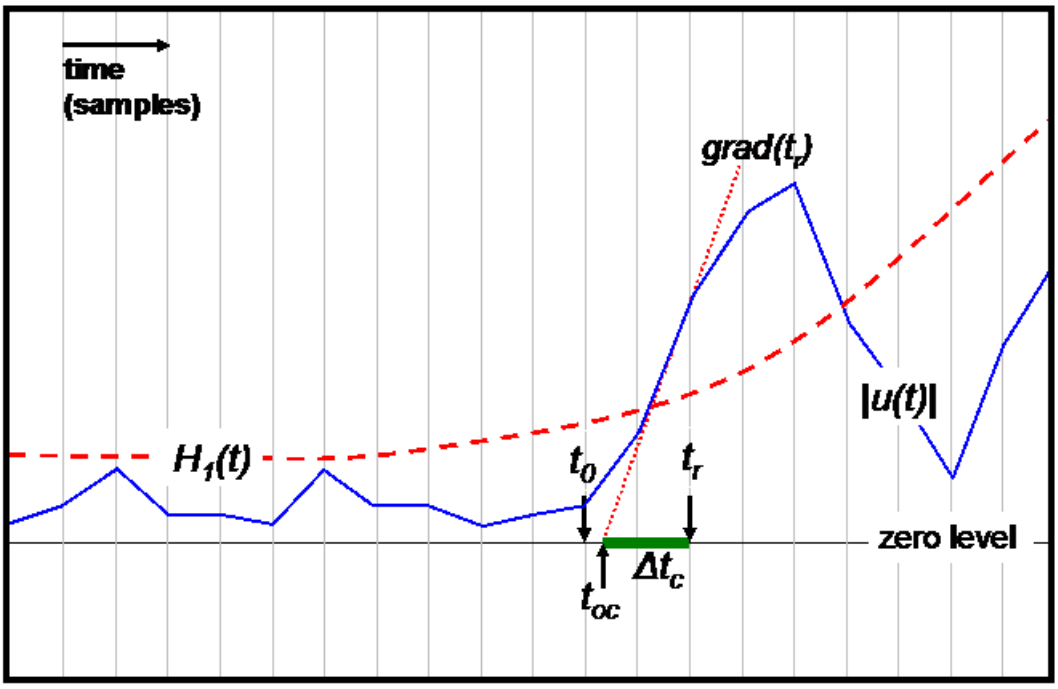

FIG. 1. Schematic diagram showing the items used in the waveform correction at trigger point (tr) for the corrected onset time (toc) of an event. The absolute value of seismic trace u(t) is drawn by a solid blue line; first threshold H1(t) is represented by a thick red dashed curve; the red dashed oblique line indicates the adapted gradient; the thick red bar marks the calculated amount of waveform correction (Δtc); the real onset time (to) is indicated by a black arrow. Vertical dotted lines mark the sample points at one sample interval.

SYNTHETIC AND REAL DATA

Both accuracy and reliability of the algorithm are examined by using synthetic and real seismic data respectively.

Picking accuracy

The performance in accuracy of the multi-window algorithm is first examined on a series of synthetic data examples with various *SNR* levels. Basically, the synthetic data are composed of three parts: computer-generated random noise with various gains, an impulsive high-amplitude short-duration noise and an impulsive exponential-decay sine wave seismic event with its first arrival at a fixed time point. Random noise still superimposes the event after the arrival of the sine waves.

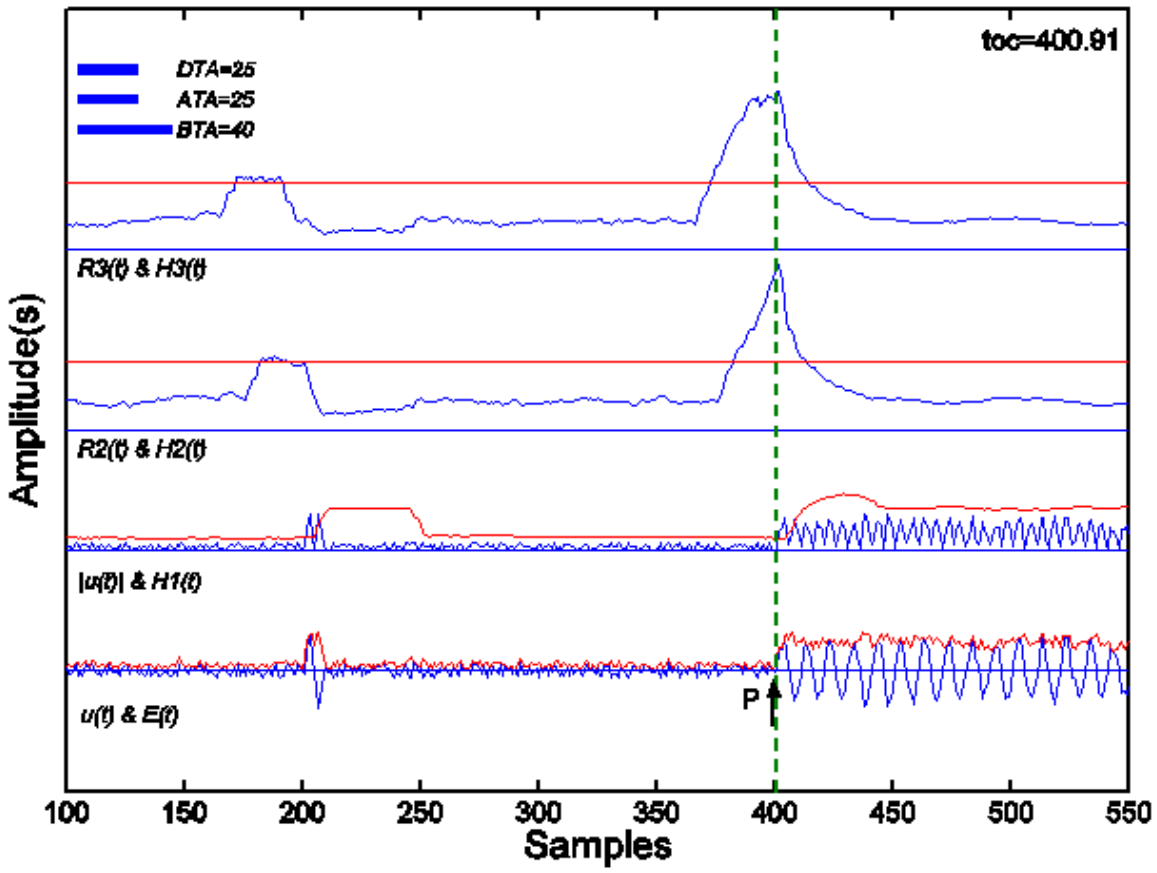

FIG. 2. Illustration of the picking processes of the first arrival of an event in a synthetic seismogram. Diagrams from bottom to top are synthetic seismogram (blue) and envelope (red); absolute value of seismic trace u(t) (blue) and threshold H1(t) (red); second metric R2(t) (blue) and threshold H2(t) (red); third metric R3(t) (blue) and threshold H3(t) (red). The solid blue line in each diagram represents the zero reference level. The lengths of running time windows are shown in the upper left corner. The expected SNR is set to be 3. The real arrival time of the assumed seismic event is set at time point 401. Vertical dashed green indicates the time of the corrected arrival time (t_{oc}), which is also written in the upper right corner. The delay of DTA window is set to 10 samples.

Figure 2 gives an example to show the multiple time windows, metrics and thresholds, and illustrates the picking accuracy of our algorithm on a synthetic seismic trace with a 20 Hz dominant frequency event. The maximum amplitudes of the event and a high-amplitude short-duration noise burst are assumed to be 1. The amplitude coefficient of random noise is assumed to be 0.25. It is noted that the three metrics of the event exceeded their corresponding pre-defined thresholds at sample 402, and the event is triggered there. The continuous procedure of waveform correction modifies the trigger point to the corrected arrival time of 400.91, which is merely 0.09 samples earlier than the real arrival time at time point 401.

The accuracy of our automatic P phase picker is also examined on some real local events recorded by a small six-station vertical-component seismic array deployed in southern Alberta, Canada (Bingham, 1996). Examination shows that the picking accuracy of real events with impulsive first arrivals can be stably picked at an accuracy of less than 1-2 samples compared to the pickings by a human operator (Figure 3).

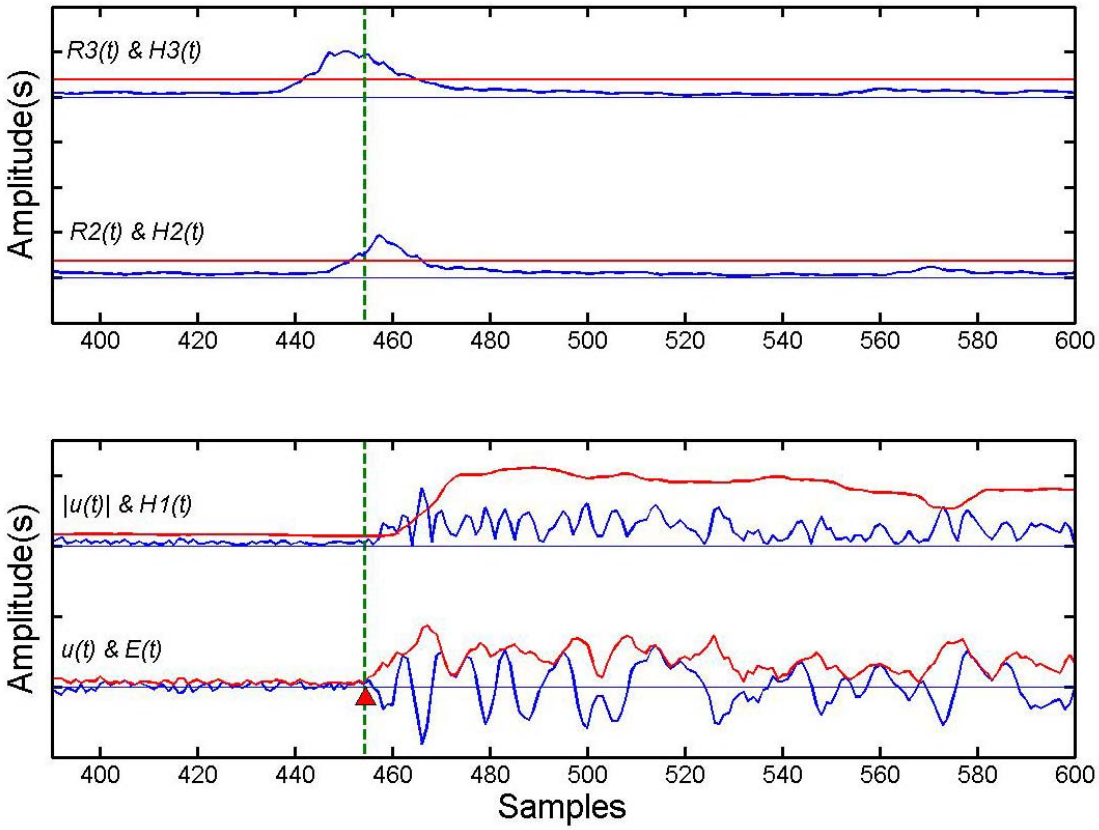

FIG. 3. Example of the picking processes of the first arrival of an event in a real seismic seismogram. Diagrams in the lower frame are seismogram(blue) and envelope(red), absolute value of seismic trace u(t) (blue) and threshold $H_1(t)$ (red) upwards. Diagrams in the upper frame are second metric $R_2(t)$ (blue) and threshold $H_2(t)$ (red), third metric $R_3(t)$ (blue) and threshold $H_3(t)$ (red) upwards. The solid blue line in each diagram represents the zero reference level. The lengths of running time windows are shown in the upper left corner. The vertical dashed green indicates the time of the corrected arrival time (t_{oc}); The red solid triangle is the first arrival picked by eye. The lengths of BTA, ATA and DTA are 40, 10 and 10 samples respectively; the delay of DTA window is set to be 10 samples.

Picking reliabilities

The reliability and accuracy of the algorithm is examined by the synthetic data in the case of various background noise levels. As the noise component of the synthetic data is computer-generated randomly, the time series used in the examinations is slightly different each time and can be regarded as new independent input. The examination of each noise levels is repeated 100 times to acquire a stable statistic of errors of detection between real and corrected arrival times.

Figure 4 shows the statistical results on the reliability and accuracy of our algorithm with the increase of the amplitude of random noise, where the maximum arrival time error ranges from -1.0 to +1.25 samples. When the coefficient of random noise was increased to 0.4, failures in picking of the first arrival times due to the existence of high-amplitude noise began to occur.

The scatter of the corrected arrival times increases with noise levels. This is considered to be caused by the disturbance of the waveform of first arrivals by the superimposed random noise. The disturbance of the waveform thus reduces the reliability of the two crucial factors of height and representative gradient adopted in the waveform correction at the trigger point. It is noted that the measurement of the reliability of the first arrival remains a difficult problem. *SNR* is normally used to measure the quality of a signal and can be improved in a single-component seismic trace by signal enhancement using, for example, bandpass filters or Wiener filters (Douglas, 1997). Also, pre-filtering of low-frequency noise and direct current in the recordings is crucial to remove their influences on the averages of the time windows.

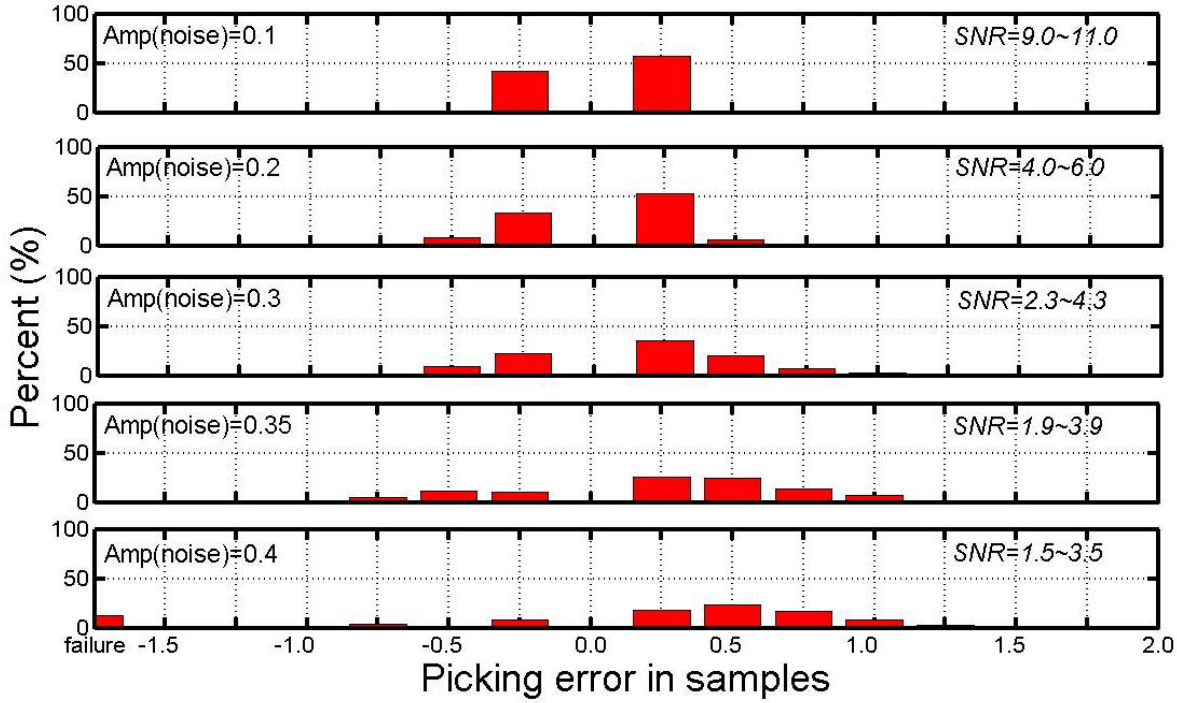

FIG. 4. Percentage of error distribution of corrected arrival times with various SNR levels. For each diagram, picking is repeated 100 times to get the stable statistics. Histograms indicate the percentage of events with corrected arrival time errors less than the indicated amounts in the horizontal axes. The SNR is estimated by equation SNR= Amp^{event} /Amp^{noise}, where Amp^{event} and Amp^{noise} represent the maximum amplitudes of event and random noise respectively. The lengths of the BTA, ATA and DTA windows are shown in the upper left corner. The time delay of the DTA window is set to 10 samples. $H_2(t)$ and $H_3(t)$ are set to be 0.75 times the expected SNR of 2.

CONCLUSIONS

An effective and accurate multi-window algorithm for automatically picking the arrival times of P-phases of impulsive local seismic events for a single trace is presented. The algorithm functions reliably when the *SNR* is higher than 3. The picking accuracy of the P-phase is less than 1 sample in most low *SNR* environments, which is comparable to that achieved by a human operator.

Variation of parameters, such as lengths of time windows, time delay interval and thresholds, can be tailored to a specific case. After the limitation of an expected *SNR*, the duration of the impulsive noise to be discarded, and the dominant frequency band of signal are indicated, all parameters can be defined directly or automatically. Due to this efficiency and simplicity of the method, our algorithm is appropriate and applicable in real-time automatic seismic monitoring and hypocenter location with only moderate computing capability.

ACKNOWLEDGEMENTS

The author thanks Rolf Maier, of CREWES, Univ. of Calgary for his helpful discussion. The author also thanks Dr. D. Bingham of Alberta Environmental Protection for providing the real seismic event data.

REFERENCES

Allen, R. V.,1978, Automatic earthquake recognition and timing from single trace: Bull. Seism. Soc. Am., **68**, 1521-1532.

Allen, R. V., 1982, Automatic phase-pickers: their present use and future prospects: Bull. Seism. Soc. Am., **72**, S225–S242.

Baer, M., and Kradolfer U., 1987, An automatic phase picker for local and teleseismic events: Bull. Seism. Soc. Am., **77**, 1437–1445.

Bingham, D. K., 1996, Seismic monitoring of Turtle Mountain: Internal report, 34 pp., Hydrology Branch, Alberta Environmental Protection, Edmonton, Alberta.

Boschetti, F., Dentith M. and List R., 1996, A fractal based algorithm for detecting first arrivals on seismic traces: Geophysics, **61**, 1095–1102.

Dai, H. and MacBeth C., 1997, The application of back-propagation neural network to automatic picking seismic arrivals from single-component recordings: J.Geophys.Res., **102**, 15105-15113.

Der, Z.A. and Shumway R. H., 1999, Phase arrival time estimation at regional distance using the CUSUM algorithm: Phys. Earth Planet. Inter., **113**, 227-246.

Douglas, A., 1997, Bandpass filtering to reduce noise on seismograms: is there a better way?: Bull. Seism. Soc. Am., **87**, 770–777.

Kanasewich, E.R., 1981, Time Sequence Analysis in Geophysics: Univ. of Alberta Press, Edmonton, Alberta, Canada.

Morita, Y. and Hamaguchi H., 1984, Automatic detection of onset time of seismic waves and its confidence interval using the autoregressive model fitting: Zisin **37**, 281–293.

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

Zhao, Y. and Takano K., 1999, An artificial neural network approach for broadband seismic phase picking: Bull. seism. Soc. Am., **89**, 670–680.