Automatic time-picking of first arrivals on large seismic datasets

Joe Wong

ABSTRACT

Determining first-arrival times on seismograms in high-volume 3D datasets with as many as one hundred million seismograms requires fast automatic picking procedures. I present a procedure based on a modified energy ratio (MER) attribute that is effective for identifying the first-break times of arrivals with signal-to-noise ratios (SNRs) greater than 10dB. For arrivals with lower SNR values, the procedure exploits spatial coherence of events on common-source gathers to obtain reasonably reliable picked times. The report shows examples of first-arrival picking on common source gathers from a 3D survey.

INTRODUCTION

First-arrival picking on seismic survey data has been well-studied (Willis and Toksoz., 1983; Coppens, 1985; Boschetti et al., 1996, Chen and Stewart, 2005). This fundamental task is ubiquitous in seismic processing, and in particular it provides the arrival times required by refraction analysis for determining the velocities and thicknesses of near-surface low-velocity zones. Similarly, first-arrival times from full-waveform sonic well logs and VSP seismograms are needed to calculate interval velocities in a well. Arrival times are also required for locating focal points of global earthquakes and microseisms caused by hydraulic fracturing.

A standard method widely used for identifying valid events and picking arrival times is based on long-term average to short-term average (STA/LTA) ratios of seismogram energies. Han et al. (2009) proposed the modified energy ratio (MER) attribute as an alternative to the STA/LTA ratio. Expressions for the LTA/STA ratio and the MER attribute are found in the Appendix. Han (2010) compared the effectiveness of techniques based on the two types of ratios for arrival time picking on high-noise microseismic data. Testing on both field and synthetic microseismograms indicated that STA/LTA ratios and MER attributes are equally effective in picking first-break times on traces with good signal-to-noise (SNR) levels. However, when the traces are noisy, MER time-picking yielded more consistent results, and was significantly faster.

For large datasets produced by high-resolution 3D surveys consisting of hundreds of millions or even billions individual seismograms, it is impossible to pick first-arrival times manually. Picking must be done automatically by computer software. Quality control (QC) should be applied to time picks that number in the tens or hundreds of millions, regardless of whether they are made by humans or by machines. Inevitably, the quality control of such huge numbers must rely on statistics and the theory of population sampling, much like what is done in mass-production QC. Since the MER-based automatic picks are reliable for signal-to-noise ratios (SNR) of 10dB or greater, the statistical sampling can be restricted to comparisons of automatic to manual picks from seismograms for which the first arrivals have SNR less than 10dB.

Because speed is a prime consideration when picking hundreds of millions of first arrival time, it is imperative that the computer-based decisions used in automatic picking be made as fast as possible. This means that simple calculations with few floating point operations must be given precedence over more involved calculations. Boschetti et al. (1996) presented a method based on fractal mathematics. Baziw (2002) and Qiao and Bancroft (2010) used Kalman filtering with wavelet transforms in their procedures. These methods have the potential for giving more accurate estimates of first break-times in the presence of noise, but they involve complex calculations that would slow down automatic picking of high-volume datasets. A good strategy may be to use a fast technique based on the simple MER attribute for those traces with high SNR, and resort to the slower techniques for the low-quality seismograms.

METHOD

Trace pre-conditioning

To minimize the effect of high-amplitude surface wave arrivals on the picking of first arrivals, we apply an appropriate bandpass filter to each trace. The trace is then normalized by its maximum peak-to-trough, and the normalized amplitudes are divided by the time indices to reduce late-arriving surface wave amplitudes. To minimize the effects of noise prior to the first arrival, the trace is tapered up from zero at the beginning of the trace to the normalized values at time P_D, where P_D is the dominant period of the arrival. MER attributes are then calculated for every point the trace, and we estimate the first-break time on the trace by picking the time of the maximum MER value. The uncertainty in the time-pick for a low-noise first arrival is in the range $\pm P_D/8$; the uncertainty for high-noise ratios is higher, or the pick could be totally false.

Summary of the MER-based time-picking procedure

1. Kill bad traces in CSGs; apply bandpass filtering (20-40-100-200) Hz to reduce surface wave amplitudes; normalize each trace by its maximum amplitude.

2. Divide trace amplitudes by the time indices; taper the beginning of each trace.

3. Calculate MER attributes for each trace; make an initial estimate of estimate first-break times for all traces in the CSG by finding the maximum MER values.

4. Use the initial time picks to find parameters for a best-fit hyperbola in space-time coordinates; use the hyperbolic trajectory to define a retention corridor for the common source gather.

5. Mute all signals outside the retention corridor. Reject all initial time picks outside the corridor; retain initial time picks inside the corridor that have high SNRs.

6. Shift and flatten the muted traces with the times on the upper boundary of the retention corridor can be set to zero. This creates a CSG of reduced seismograms with a new time origin.

7. Re-calculate MER attributes for the muted CSG, and re-pick the first-arrival times.

8. For traces with low SNR and poor coherence of the picked first-break times, try picking the first trough or peak of the first arrival.

9. Apply smoothing to selected picked times, if desired. Interpolate between the good picks to fill in gaps caused by killed or high-noise traces, if desired

MER PICKING ON A COMMON-SOURCE GATHER

The remainder of this report demonstrates MER picking of first-arrival times using real field data. The seismograms on Figure 1 were acquired as part of a large land 3D survey using Vibroseis acquisition. Bandpass filtering has been applied to attenuate the amplitudes of strong surface waves. Reflections from subsurface geological boundaries are visible, but this report is concerned with the first arrivals only.

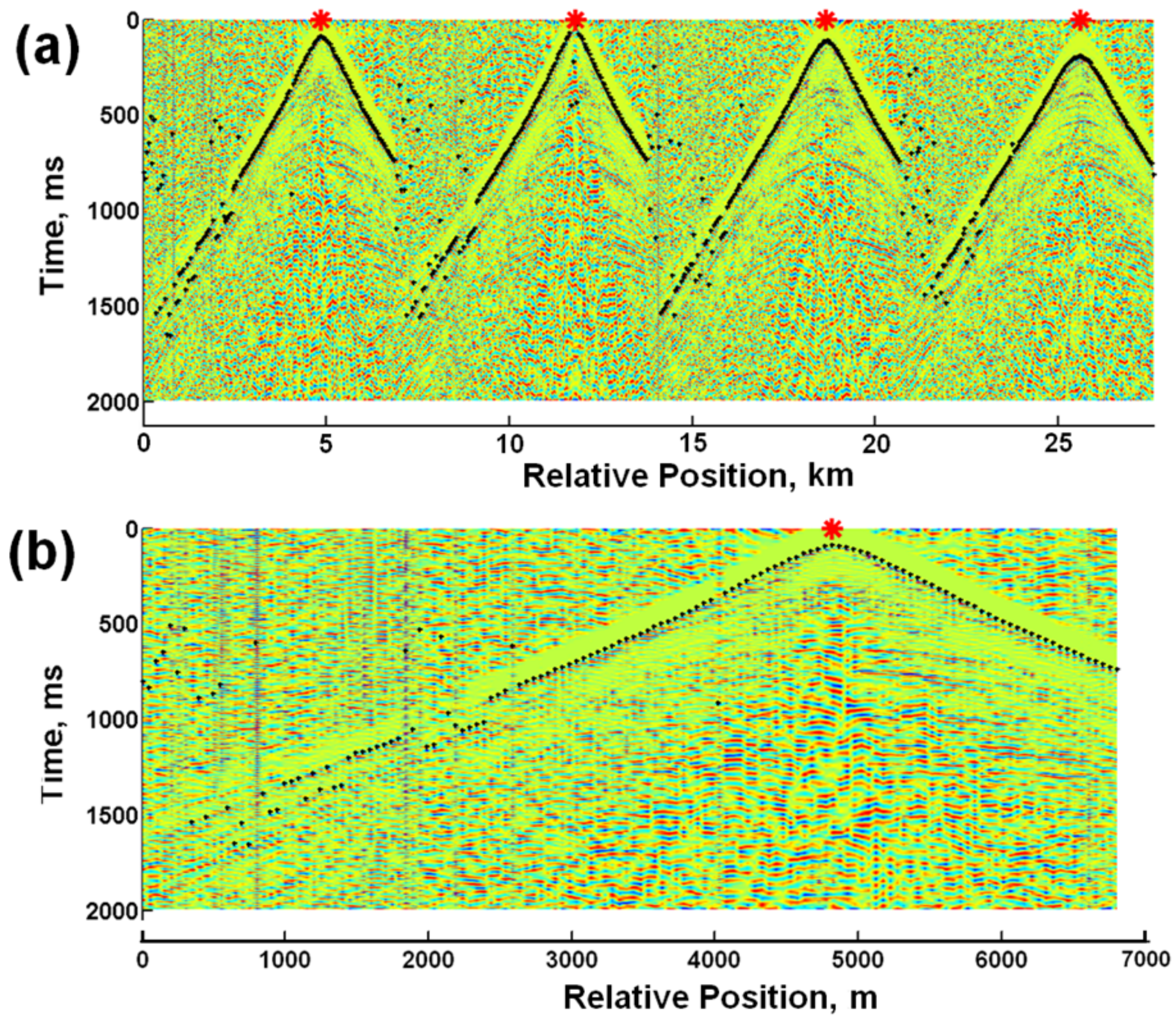

FIG 1. (a) AGC plot of four common-source gathers. (b) Expanded view of the left-hand-side CSG of Figure 1(a). Black crosses are initial MER time picks. The red stars indicate the positions of the sources.

Initial picking first-arrival times

The MER-based time-picking procedure was employed to obtain the initial first-arrival times of the seismograms on Figure 1. Although Figure 1 is plotted with AGC applied, the time picking was done on normalized traces. The initial picks appear to be very good estimate of first-arrival times for source-offset distance up to about 2km. They become erratic and less reliable when SNR values become low at larger offset distance. The loss of arrival amplitudes at large offsets mostly is due to poor and inconsistent coupling of geophones to the ground. Automatic picking of first-arrival times using the MER attribute is straightforward for most traces, but extra effort needs to be applied to improve the automatic picks for the high-noise traces.

On Figure 2 we have re-plotted the CSG of Figure 1(b) with each trace normalized by its maximum value. We estimated that the duration of the first arrivals on the seismograms to be about 30ms to 50ms, so we set the length of MER energy-collection window at 20ms. Since the sampling interval is 2ms, the index length L in Equation A1 is 10.

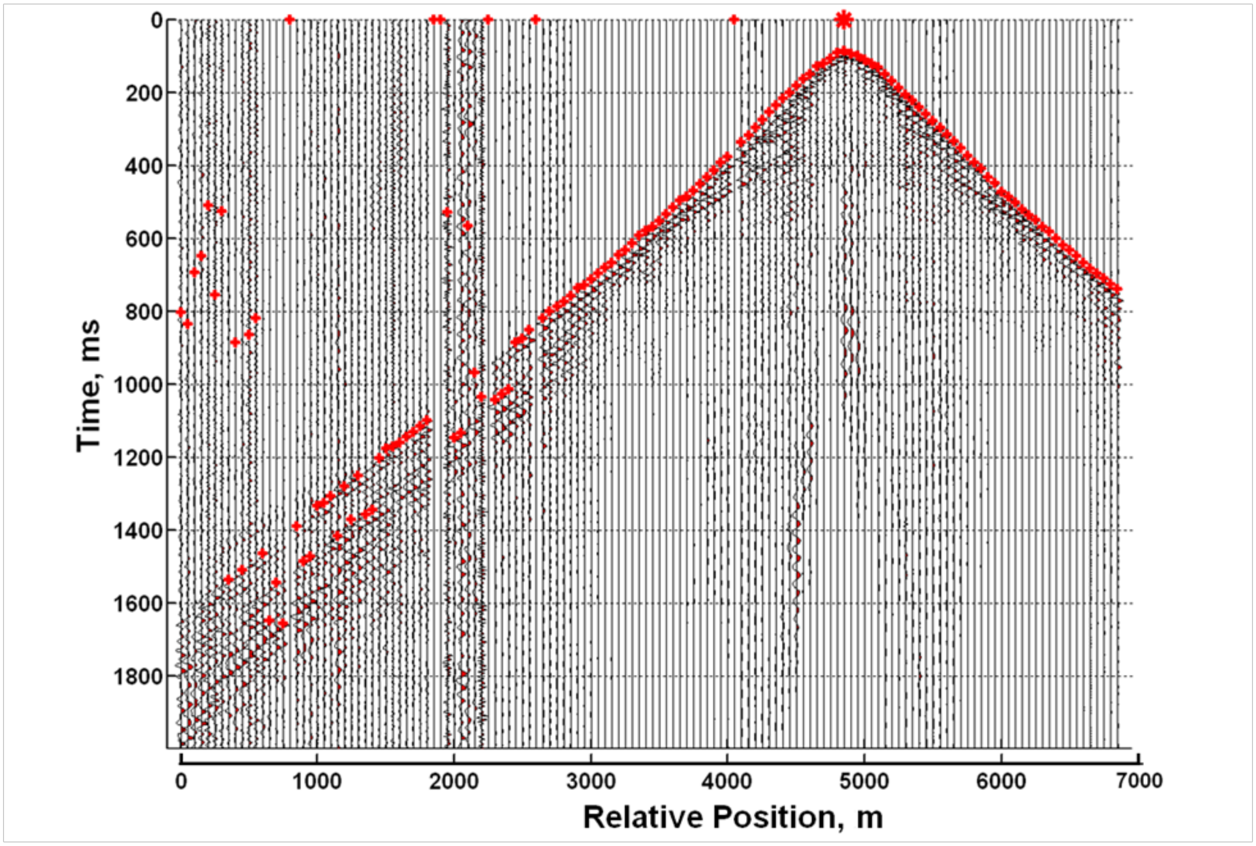

FIG 2. Common-source gather of normalized seismograms. All initial time picks are plotted as red crosses regardless of the SNR values associated with the picks. Outlier picks are tied to killed traces and traces with low SNR traces.

The first-break times obtained by an initial application of the MER-based picking procedure are plotted as red crosses. With the displayed time scale, the MER picked times appear to be excellent estimates of the first arrival times except for receivers far from the sources. The poor picks are tied to traces with low signal-to-noise ratios (SNRs). For this report, SNR is defined in decibels by the expression

$$SNR(i) = 20\log(\sum_{j=i}^{i+L} grm(j)^2 \, / \, \sum_{j=i-L}^{i} grm(j)^2)) \,, \tag{1}$$

where $SNR(i)$ is the signal-to-noise ratio at trace index i and $grm(j)$ are trace amplitudes. The denominator and numerator are the energies in windows of length L preceding and following the test point i. Comparison of Equation 1 with Equation A1 in the Appendix shows the SNR at a test point is exactly the same as the energy ratio, except that it is expressed in decibels. The SNR value associated with every estimated first-break time should be kept for use in quality control: the higher the SNR value, the more reliable the estimated first-arrival time will be.

Defining a retention corridor

The first-pass time picks shown on Figure 2 include quite a few outliers. Arbitrarily, we could reject any time pick with an SNR value of less than 6dB as being bad picks. If we wish to get better estimates for many of the picks with low SNR values, we can do a second iteration of MER picking. The first arrivals on the gather roughly follow a normal moveout (NMO) or hyperbolic trajectory. Using the pick times with good SNR values, we estimate a velocity v_0 and a time t_0 to find a suitable hyperbolic trajectory to fit the picked times using the standard equation for calculating NMO times:

$$t_r = \sqrt{(t_0^2 + [(x_r - x_s)^2 + (y_r - y_s)^2]/v_0^2} \ , \qquad (2)$$

where t_0 and v_0 are time and velocity parameters determined from the time picks, and t_r is the hyperbolic time of a receiver at position (x_r, y_r) separated from the source at position (x_s, y_s) .

We require values for the parameters t_0 and v_0 in order to make the calculations. These can be estimated easily from the initial picked times since most of these times already fall on a nearly hyperbolic path. One way of finding good estimates for t_0 and v_0 is to fit the initial picked times with high SNR values to Equation 2 by nonlinear optimization. An easier approach is to square both sides of Equation 2 to get

$$t_r^2 = (t_0^2 + [(x_r - x_s)^2 + (y_r - y_s)^2] * u_0^2 \ . \qquad (3)$$

Equation 3 is linear in the unknown parameters t_0^2 and u_0^2 ($u_0 = 1/v_0$ is slowness), and we can use linear rather than nonlinear inversion to find values that fit the squared picked times t_r^2 optimally.

Once a suitable hyperbola is found, we shift it up so that it lies entirely above all the first arrivals on the CSG. We shift it down to create a second hyperbola that lies entirely below all the first arrivals on the CSG. The two hyperbolas define a retention corridor containing the first arrivals. All trace values and all initial picked times outside the corridor are muted. Initial picked times inside the corridor but with SNRs below a cutoff value are rejected. The initial picked times inside the corridor with SNRs above the cutoff value are retained since they are considered to be good picks. Their retention adds to the redundancy of picked first-arrival times for the CSG. Redundancy of data may be important for subsequent statistics-based QC analysis on sampled populations.

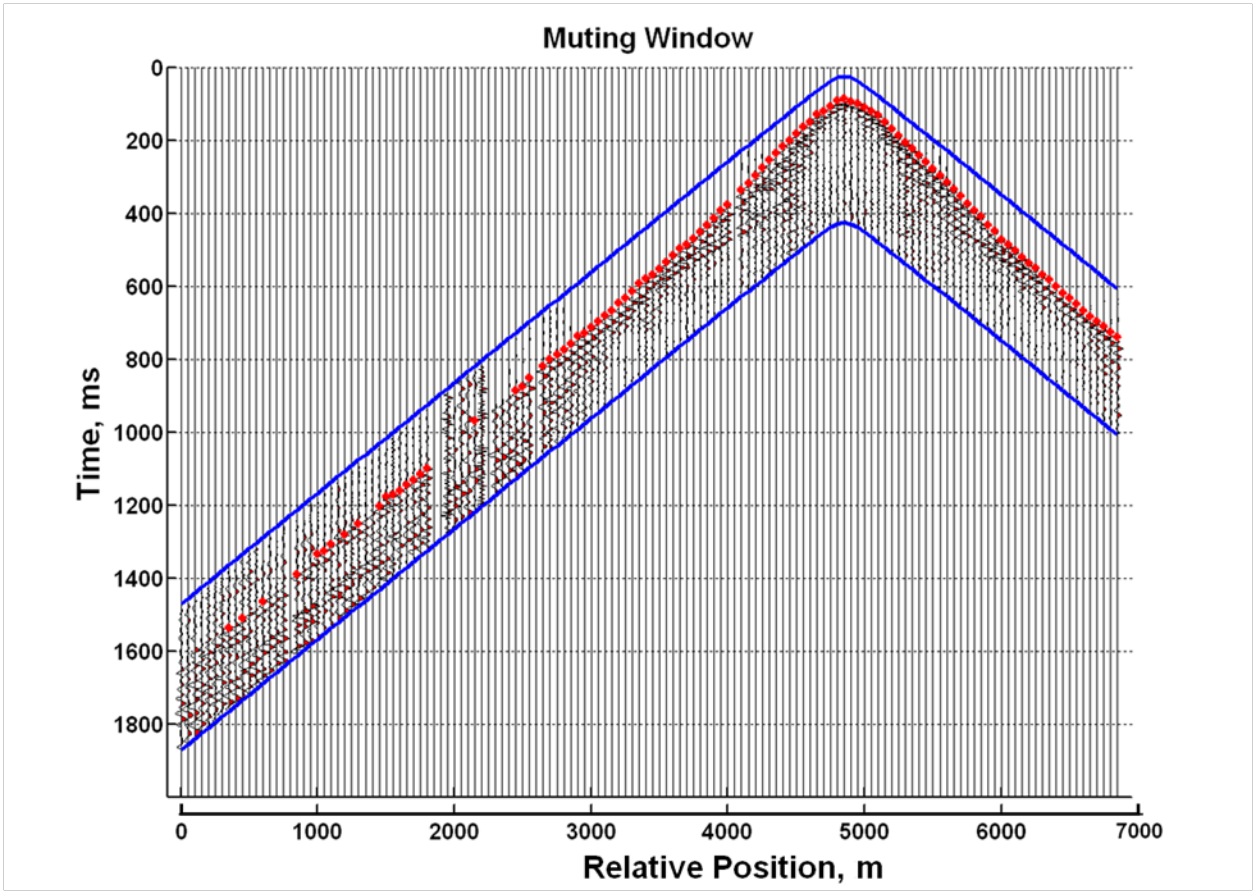

FIG 3. Trace values and initial time picks (red crosses) inside the retention corridor defined by the hyperbolic trajectories plotted in blue.

For our example CSG, the parameters t_0 and v_0 were found from the initial time picks to be 25ms and 2800m/s respectively. On Figure 3, the hyperbolic trajectory calculated with these values in Equation 3 is plotted, and then plotted again displaced downwards by 400ms. The two hyperbolas define a retention corridor of width 400ms. It is important that they bracket short portions of the seismograms near the expected first arrivals. All seismogram values not between the two plotted trajectories are muted. All time picks outside the retention corridor are discarded, thereby eliminating most of the outlier picks at far offsets. We retain the picked values between the trajectories with SNRs greater than the cutoff value as potentially good picks.

On the time scale at which Figure 3 is displayed, it is impossible to see any detail about how closely the picked times approach the first breaks on the seismograms. In order to see more detail, we generate a reduced-seismogram gather by shifting all the traces upwards so that zero time is set at the upper hyperbolic trajectory. All picked times are also corrected by the shift times to conform to the new time axis origin. The reduced seismograms and corrected time picks are plotted with an expanded time scale on Figure 4. On this figure, we can confirm that most of the initial picked times are very good estimates of the first breaks. Also, we now can see clearly which initial time picks are problematic.

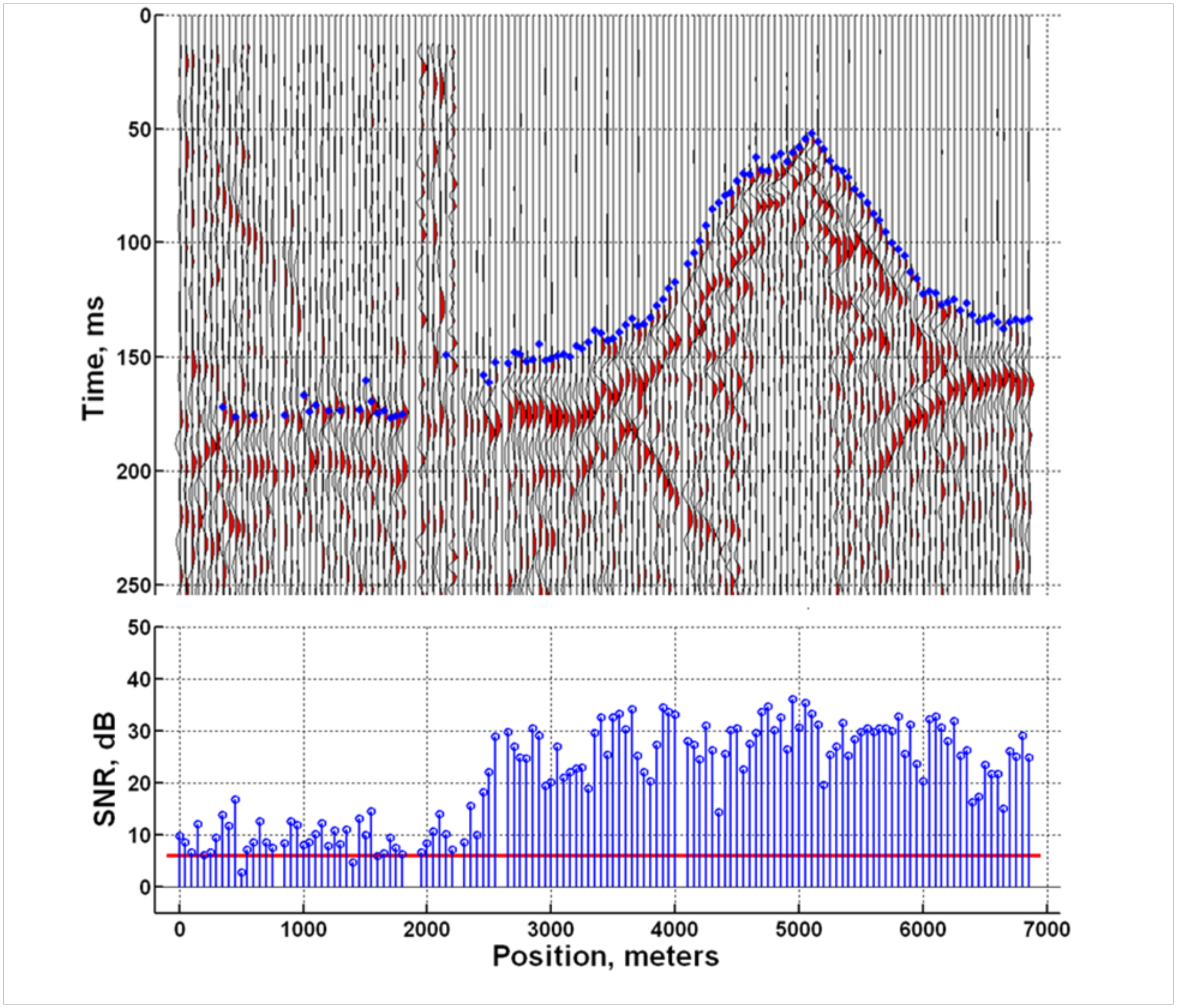

FIG 4. Reduced seismograms with shifted initial time picks displayed as blue crosses. The gaps in the plot of SNRs associated with automatic picks are due to killed traces. The red line is drawn for the SNR limit of 6dB.

Also plotted on Figure 4 at the bottom are the SNR values associated with the initial time picks. For this example, all initial time picks outside retention corridor, or have SNR values less than the cutoff value of 6dB, have been rejected and not plotted. Gaps along the position axis of the SNR plot indicate the position of killed traces. Many of the seismograms positioned between 0m and 2800m are quite noisy, and could have been killed at the very beginning. We have retained them to show that their inclusion does not affect MER automatic time-picking; the picks on such traces will have very low SNRs and can be rejected on that basis. At the cutoff value of 6dB, the ratio of signal energy to noise energy is 2, which means that average signal amplitudes is only about 1.4 times the average noise amplitudes. This actually is quite small, but setting the cutoff limit to a higher value (say, 10dB) would eliminate too many potentially good picks.

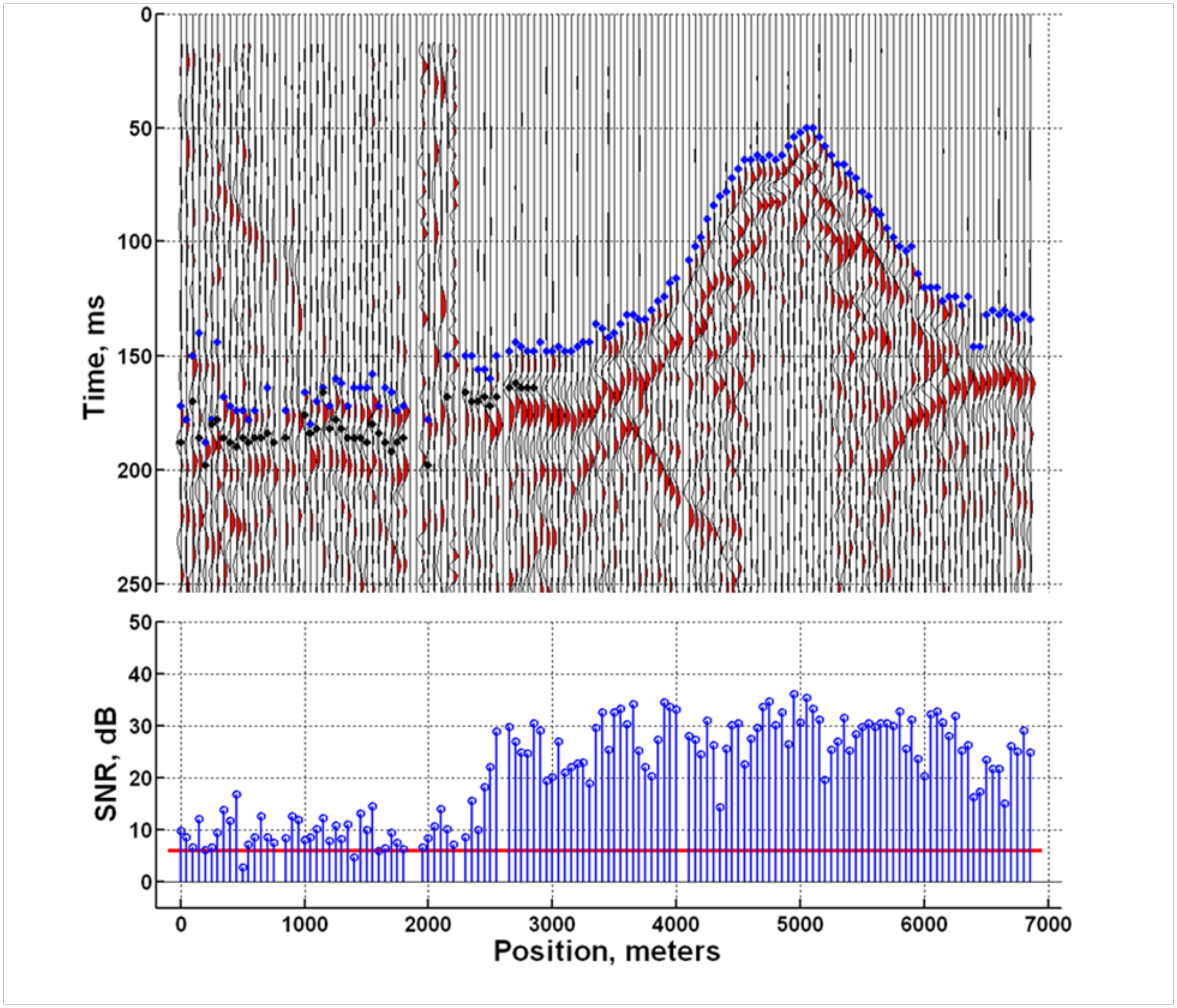

FIG 5. Reduced seismograms. Blue crosses are first-arrival picks made on the reduced seismograms. Black crosses are picks made on the first negative troughs following the first-break times for traces with low SNR (<=10db). Note the two outlier picks at 6400m and 6450m. Here, the MER attribute has skipped the first arrival and picked a later, more energetic event.

Second iteration of picking first-arrival times

We treat the reduced seismograms as a new dataset, and perform a second round of MER picking to obtain updated picks of first-arrival times. The reduced seismograms and the updated picks are shown on Figure 5. The second round of MER picking has increased the number of potentially good picks at positions between 0m and 2700m. To build confidence in the estimated first-break times, and possibly improve these estimates, the automatic procedure searches for the times of the first minimum and maximum trace amplitudes following the first-break picks. These trough or peak times often are more easily discerned than are energy onset times. If the picked first-break times are reliable, the trough and peak times should be later by one quarter to three quarters of the dominant periods. For our example CSG, trough times were picked and retained for positions between 0 and 2800m. They are plotted as black crosses on Figure 5.

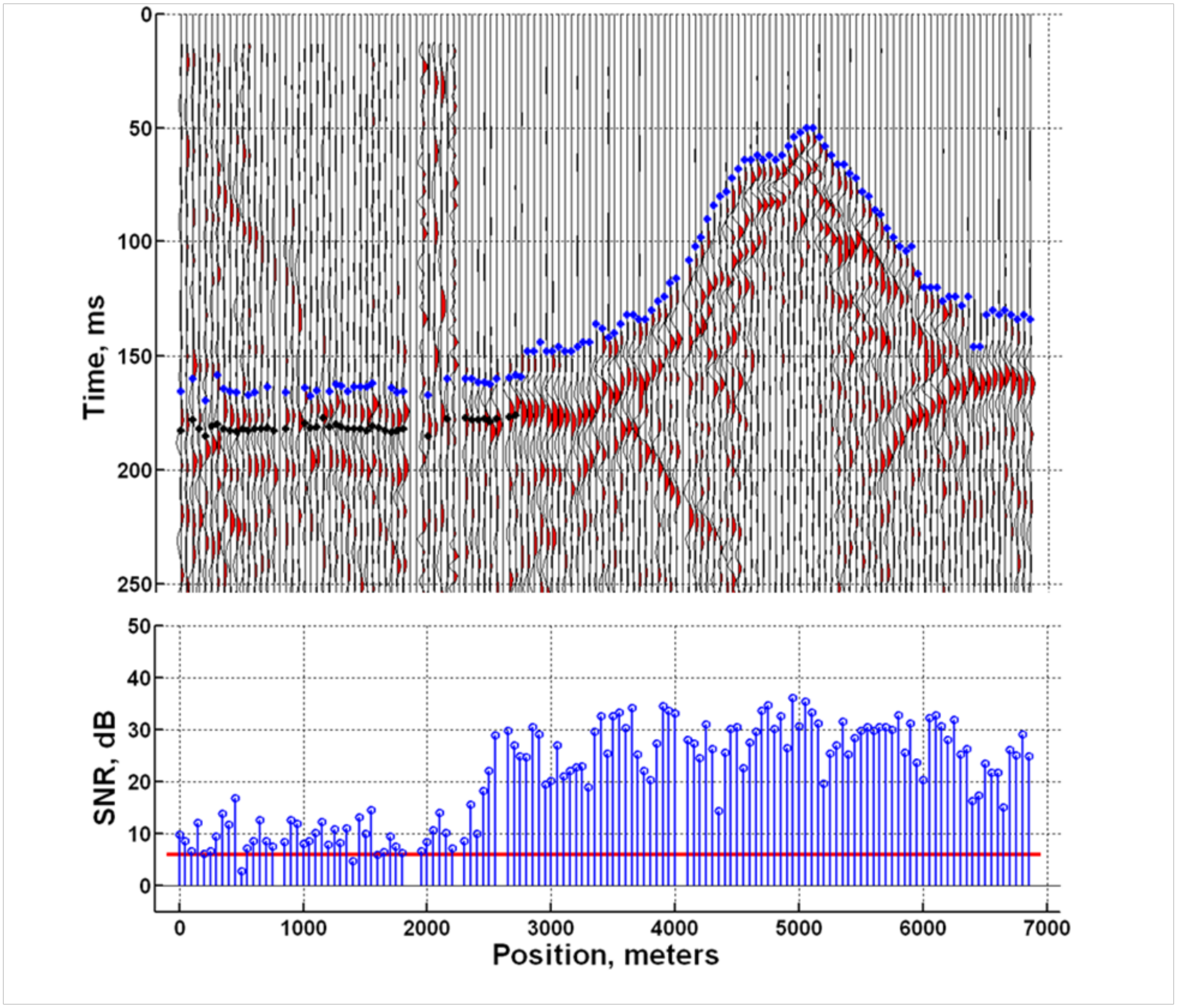

FIG 6. Reduced second-pass time picks after smoothing to reduce deviations from the mean.

The updated first-break times and the trough times for positions 0m to 2800m show some scatter on the plot of Figure 5. They arguably are the best estimates that can be made automatically by computer software, and in all likelihood they and the attendant scatter should be accepted as is without question. However, there is a philosophy that regards mean values, or smoothed values that follow a trend with less scatter when no scatter is expected, to be closer to the "truth". This philosophy leads to sophisticated error analyses based on statistical principles. Such an analysis is beyond the scope of this report, but we have calculated the means and standard deviation values of the picked first-break and trough times for positions 0m to 2800m with SNRs above the cutoff value of 6dB. We have smoothed these picked times, and display the smoothed values on Figure 5.

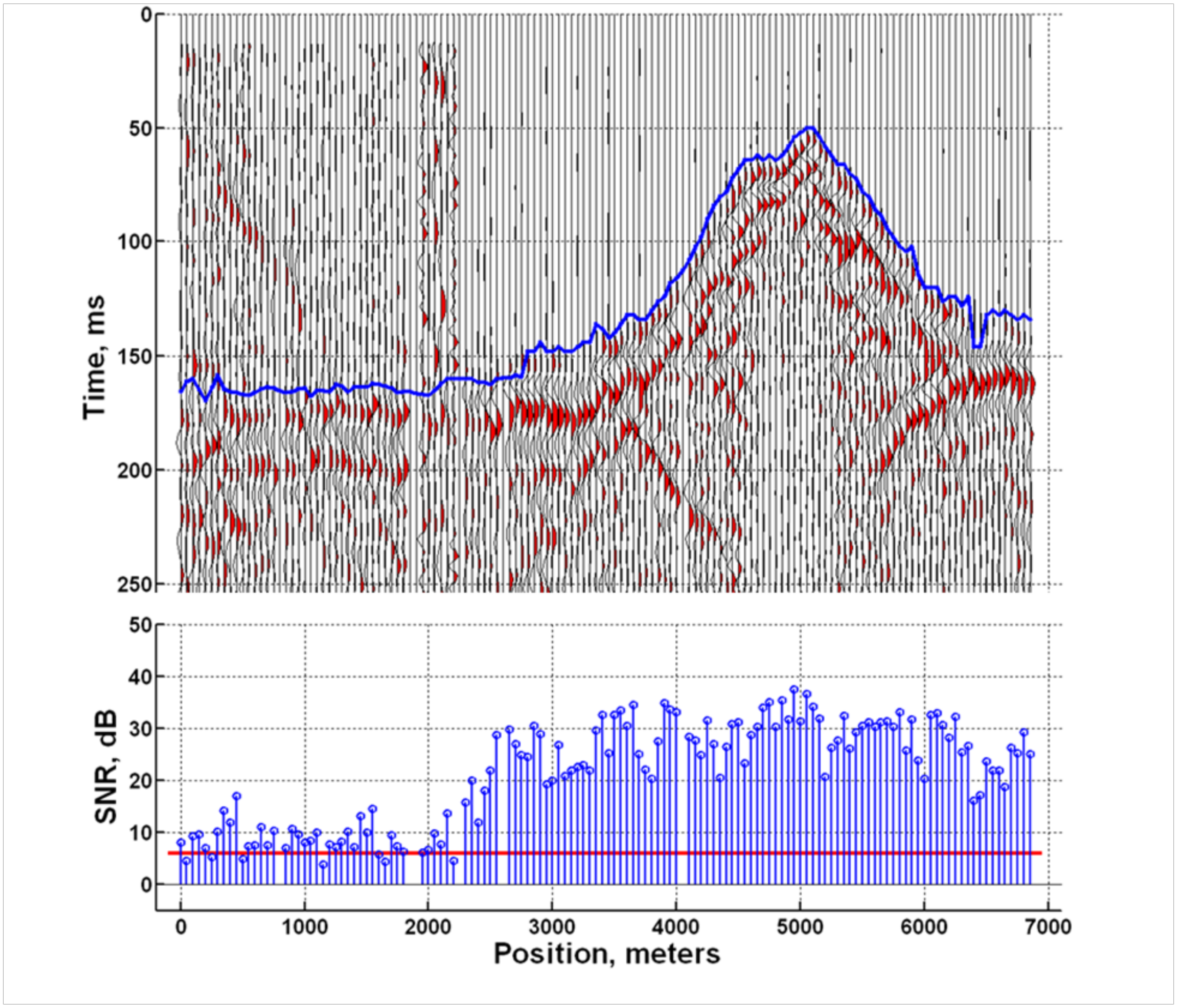

FIG. 7. Interpolated reduced first-arrival times plotted as a solid blue line. Interpolation has filled in the gaps caused by killed traces and extremely noisy traces.

We interpolated the smoothed picked times to fill in the gaps at the positions of killed traces and of traces for which pick times have been rejected because of low SNR. The interpolated first arrival times are plotted on Figure 6. For some people, an interpolated value is better than no value at all. For other people, interpolated values are "created" data because they are not directly measured. I have provided the interpolated values, and the reader can decide on the desirability of including them in any further analysis.

Possible use of coherency

On the plots of reduced seismograms, a weak linear event trending down to the right appears at reduced time of 40ms to 150ms. It shows up very clearly on the AGC plot of Figure 8, and it caused by refraction from a deep, high-velocity geological boundary. Even though it is easily recognized by the human eye, its amplitudes are not high enough above background noise for the energy ratio attributes to be able to identify it. Automatic recognition by computers will need to rely on other considerations as well. Despite the low SNR of the data shown on Figure 8, the coherence of the event in time and space is obvious, and the attribute most likely to succeed is a dip-focused measure of coherency (Coppens, 1985; Tomada, 1956). Using coherency to analyze the seismograms shown on Figure 8 is beyond the scope of the present report, but will be done in future work.

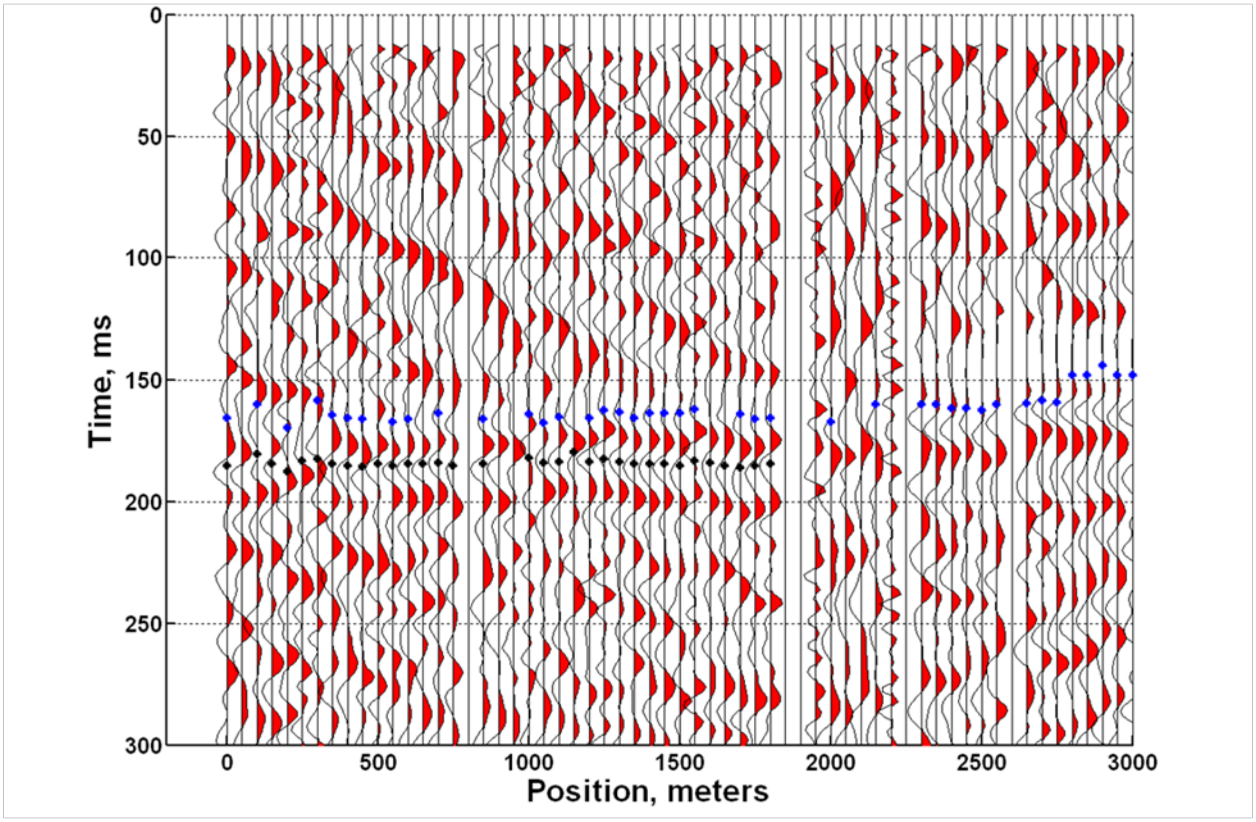

FIG. 8. AGC plot of reduced seismograms at positions 0m to 3000m. Note the clear linear trend of the traces in the upper part of the figure.

CONCLUSION

I have devised an automatic scheme for picking first-arrival times common source gathers. The scheme is based on the modified energy ratio (MER) attribute. This attribute is simple and so is fast to calculate. This fact makes it attractive for use in situation where first-arrival picking on hundreds of millions of seismograms must be done.

The procedure has been tested only on one CSG with 138 traces, but since it operates on a single trace at a time, it can be applied repeatedly for as many CSGs as are available. Other than the calculation of a signal-to-noise ratio as a figure of merit for each pick and the determination of whether or not the time pick fall within a retention corridor shaped according to the NMO hyperbola for the particular CSG, the scheme does not provide any quality control. However, external quality control using the theory of population sampling and statistical analysis can be applied to the potentially huge number of first break times generated by the picking procedure for high-volume 3D seismic datasets.

ACKNOWLEDGEMENTS

We thank the Natural Sciences and Engineering Research Council of Canada (NSERC) and the industrial sponsors of CREWES for supporting this research. Devon energy Corporation provide the field data used in this report.

REFERENCES

Baziw, E., 2002, Application of Kalman filtering techniques for microseismic event detection: Pure and Applied Geophysics, **159**, 449-471.

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

Coppens, F., 1985, First arrival picking on common-offset trace collection for automatic estimation of static corrections: Geophysical Prospecting, **33**, 1212-1231.

Chen, Z. and Stewart, R., 2005, Multi-window algorithm for detecting seismic first arrivals: Abstracts, CSEG 2005 National Convention, 355-358.

Han, l., Wong, J., and Bancroft, J., 2009, Time picking and random noise reduction on microseismic data: CREWES Research Reports, **20**, 30.1-30.13.

Han, L., 2010, Microseismic monitoring and hypocenter location: M.Sc. Thesis, University of Calgary.

Munro, K.A., 2004, Automatic event detection and picking P-wave arrivals: CREWES Research Report **18**, 12.1-12.10.

Qiao, B., and Bancroft, 2010, Picking microseismic first arrival times by Kalman filters and wavelet transforms: CREWES Research Report, **22**, 68.1-68.10.

Saari, J., 1991, Automated phase picker and source location algorithms for local distances using a single three-component seismic station, Tectonophysics, **189**, 307-315.

Tomada, Y., 1956, A simple method for calculating the correlation coefficients: Journal of the Physics of the Earth, **4**, 67-70.

Withers, M, Aster, R., Young, C., Beiriger, J., Harris, M., Moore, S., and Trujillo, J., 1998, A comparison of selected trigger algorithms for automated global seismic phase and event detection, Bulletin of the Seismological Society of America, 88, 95-106.

Willis, M.E., and Toksδz, M.N., 1981, Automatic P and S velocity determination from full waveform digital acoustic logs full waveform logging: Geophysics, **48**, 1631-1644.

APPENDIX: MODIFIED ENERGY RATIO METHODS

Techniques for time-picking of first-arrival seismic events have employed two common energy-ratio attributes: the short-term-average to long-term average ratio (Coppens, 1985; Saari, 1991; Withers et al., 1998), and the modified energy ratio (Han et al., 2009; Han, 2010). These ratios are visually defined on Figure A1. The energy ratio at a test point *i* on a digitized seismogram is

$$er(i) = (\sum_{j=i}^{i+L} grm(j)^2 \,/\, \sum_{j=i-L}^{i} grm(j)^2 \,, \qquad \text{(A1)}$$

where $grm(j)$ are trace amplitudes. The modified energy ratio is

$$mer(i) = [\, abs\big(grm(i)\big) * er(i) \,]^3 \,. \qquad \text{(A2)}$$

The exponent 3 in Equation A2 is not required for first-arrival picking. It is included to give plots of the MER attribute a much sharper, spike-like appearance at the first-break time (see Figure A2(b).

The window lengths in the formulas for accumulating energy have a significant effect on time-picking. For clean seismograms, the short window lengths L and *L1* can be set to as low as one half of the duration for the arrival wavelet but no more than one full duration. The duration is closely linked to the observed dominant period in the arrival. For seismograms with significant noise, the window lengths may need to be set to longer times. The tradeoff is that longer window lengths give more stable values of the attributes in the presence of noise, but increases the error in locating the first-break time. This error is always towards later times. For the clean trace, both ratios locate first break very well. The peaky nature of the MER attribute makes it easier than the STA/LTA ratio to use in picking first-arrival times.

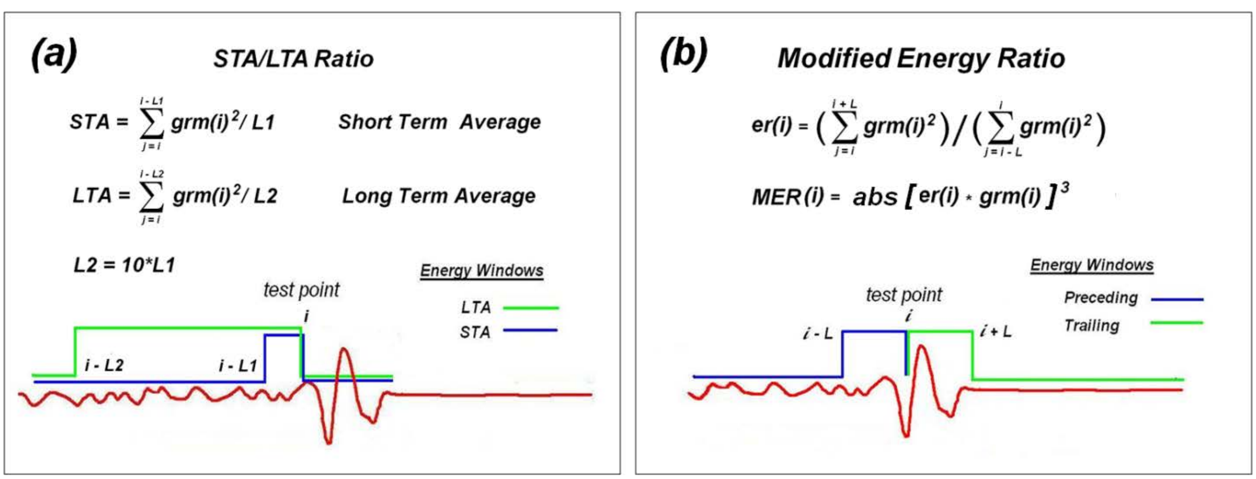

FIG. A1: (a) Definition of the STA/LTA ratio at a test point *i*. The length *L1* of the STA energy collection window should be one-half to one times the duration of the seismic arrival. The length *L2* of the LTA energy collection window should be five to ten times longer. (b) Definition of the MER attribute. The preceding and trailing energy collection windows have equal lengths located at a test point *i*. The window length *L* should be one-half to one times the duration of the seismic arrival.

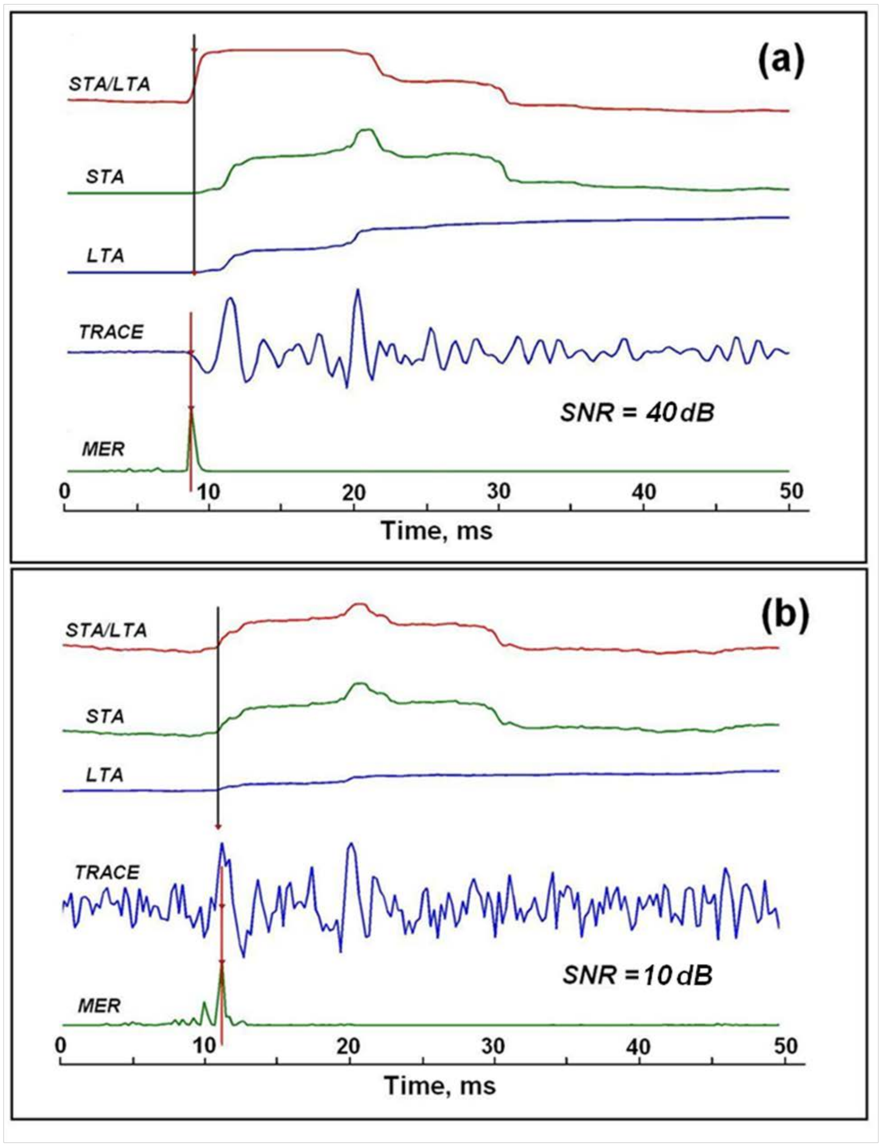

FIG. A2: Performance of the STA/LTA ratio and MER attribute in locating the onset of energy for a (a) clean and (b) a noisy seismic trace. For the clean trace, both ratios locate the first break very well. For the noisy trace, the estimated first-arrival times are shifted toward the largest amplitude in the seismogram.

Figure A2 compares the behaviour of the two attribute for a clean seismic trace and a noisy seismic trace. For the clean trace on Figure A2(a) with high signal-to-noise ratio, the first break time of the first arrival is nearly coincident with the maximum slope of the short-term-average to long-term average (LTA/STA) ratio, or with the maximum peak value of the modified energy ratio (MER). In this case, both attributes locate the first-break time accurately. For the noisy seismogram of Figure A2(b) with low signal-to-noise ratio, neither ratio is effective for identifying the first-break time accurately; the picked first-arrival times are both displaced towards the maximum amplitude for the arrival. For gathers of noisy seismograms on which the wavelets associated with the first-arrival are expected to resemble each other from trace to trace, both attributes can be used in an iterative fashion with semblance or coherence calculations to improve the first-break times.

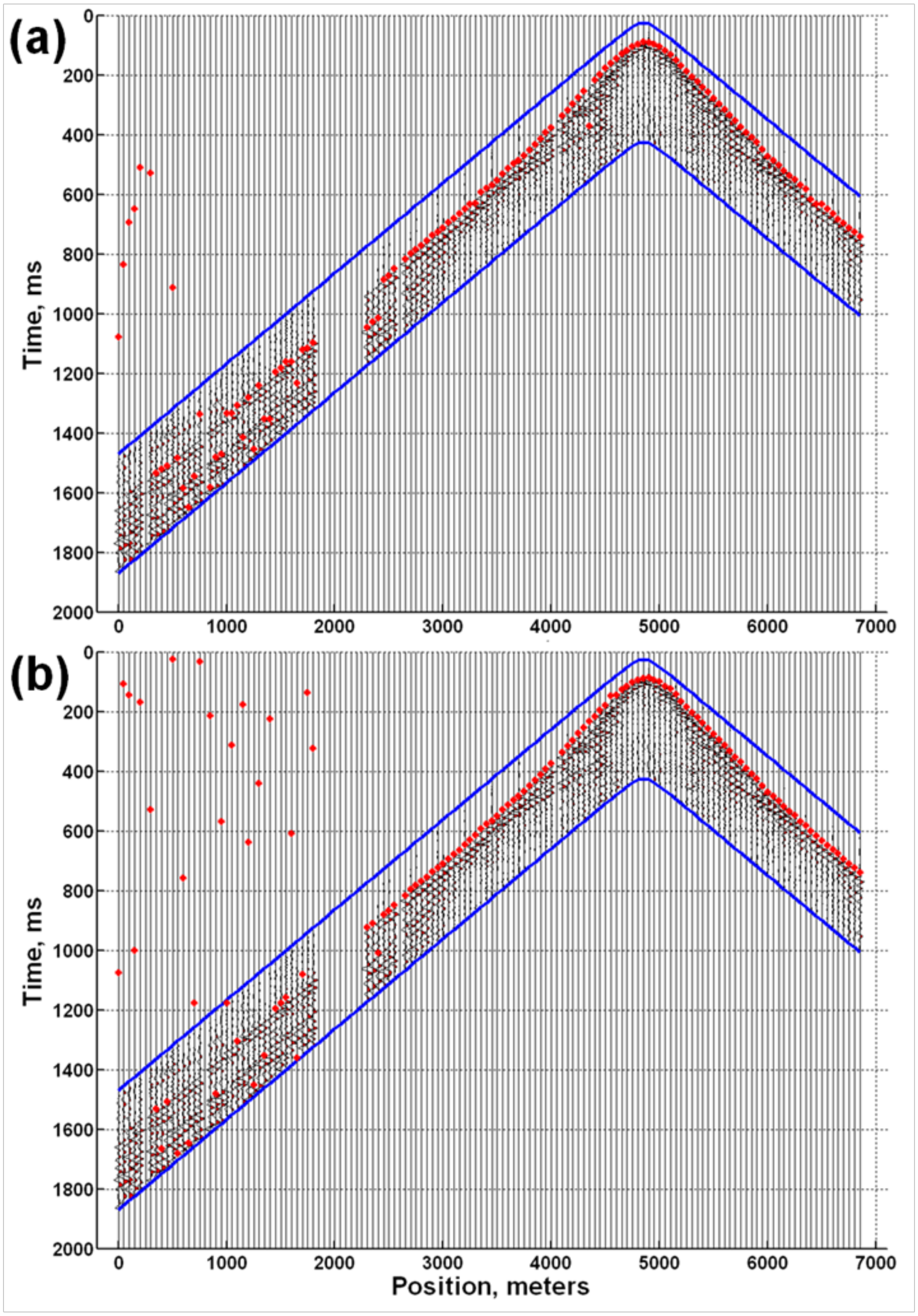

FIG. A3: Automatic time picks produced (a) using the MER attribute, and (b) using the unmodified energy ratio. For traces with low SNR, the MER attribute produces fewer false picks before the true first breaks.

Coppens (1985) used the unmodified energy ratio (Equation A1) in his time-picking algorithm. We have found that the addition of the factor $abs(grm(i))$ to modify the energy ratio (Equation A2) results in fewer false picks when the noise before the first arrival is high. This is indicated by the examples shown on Figure A3. At far receiver-source offsets, where SNR levels are less than 10dB, many more false picks, occurring much earlier than the true first-arrival times, are generated by picking the maximum of the energy ratio rather than the MER attribute.