AUTOMATIC INTERPRETATION OF REGIONAL SEISMIC SIGNALS USING THE CUSUM-SA ALGORITHMS

Zoltan A. Der and Matthew W. McGarvey, Ensco Inc

Robert H. Shumway, University of California at Davis

Sponsor; Defense Threat Reduction Agency
Contract DSWA01-98-C-0155

ABSTRACT

Automatic onset estimation is a desirable goal facilitating the rapid location and discrimination of seismic events, automation also relieves the human operator from onerous repetitive tasks. At regional distances onsets of seismic phases are associated with gradual, rather than sudden, changes in the statistical properties of the seismic time series, i.e. changes in autoregressive models and mean square amplitudes. The work presented combines two aspects of the onset time estimation, the detection of the onset as a change in the trend of the cumulative sum (CUSUM) of a suitable test statistic, and a randomized search for its location, simulated annealing (SA). The latter is applied repeatedly thus resulting in multiple onset time estimates. These typically form clusters around the arrival times of the know seismic phases, but there are some scattered values as well that have to be discarded. The uncertainties in the onset times estimated can be quantified in terms of the spread in the values of the best defined clusters. Various standard cluster analysis methods can be applied to define individual clusters.

During past work we have used cumulative sums of the absolute amplitudes of the trace amplitudes after suitable prefiltering. We have found that for determining Pn onset times the method performed comparable to a human analyst. We have also been successful in picking later regional arrivals, even though many of these were emergent and poorly defined. In the followup work we have changed to the iterative cumulative sum of squares (ICSS), developed by Inclán and Tiao (1996), and applied to suitably prewhitened time series. This method has a better theoretical foundation because it can be justified on the basis of a Gaussian time series model. Our tests showed that it performs comparably to the previously described approach. We are extending this approach to the CUSUM of other time-variable statistics of single and multiple (three-component) time series. The system is presently developed using MATLAB, utilizing its graphical user interface capabilities.

Key Words: onset times, location, discrimination

OBJECTIVE

The objective of this work is to develop automated methods for interpreting regional seismograms.

INTRODUCTION

Automatic phase arrival time estimation is of considerable interest because of the need for rapid location and identification of numerous seismic events by networks monitoring natural and man-made seismic activity. Times of seismic 'phase' arrivals can be defined as time instants where some visible characteristic, such as amplitude, frequency content or wave polarization changes in some recording. Typically, regional arrivals are high frequency, broadband, emergent wave groups containing numerous cycles. Later arrivals generally have no clear, impulsive waveforms and are preceded by the codas of earlier ones and their onset times can only be defined to within a few cycles. Besides locating events from multiple Pn times, onset time estimation is useful in facilitating location using multiple arrivals in the same seismogram and in automated application time and frequency domain discriminants.

GENERAL STATISTICAL BACKGROUND.

There are numerous statistical approaches in the literature to determining the arrival time of 'phases'. Onsets of packets of seismic energy associated with various paths, i.e. seismic phases, are usually associated with sudden changes in amplitudes, spectra, polarization and slowness. Generally it is assumed that it is known that the change occurs somewhere in a predefined time interval and that the statistical model of the time series changes abruptly. It is usually assumed that the time series are stationary before and after such a change and that the statistical distributions are multivariate Gaussian (e.g. Basseville and Nikiforov 1993). A popular approach is to test for changes in single- or multi-channel autoregressive models to the data of the form

$$\mathbf{x}(t) = \sum_{k=1}^{p} \mathbf{A}_k \, \mathbf{x}(t-k) + \mathbf{e}(t)$$

where $\mathbf{x}$(t) is the multidimensional time series, the constants in matrix $\mathbf{A}_k$ are regression coefficients and $\mathbf{e}$(t) is a stationary noise vector process. The number of channels p in various applications may be one (a single seismic channel), three (three component station) or many (a seismic array). In this process the autoregressive parameters $\mathbf{A}_k$ will have to be estimated (Pisarenko et al 1987). The determination of the onset times is based on residuals $\mathbf{r}$ of the autoregressive models computed by using the estimated autoregressive parameters

$$\mathbf{r}(t) = \mathbf{x}(t) - \sum_{k=1}^{p} \overline{\mathbf{A}}_k \, \mathbf{x}(t-k)$$

Depending on the autoregressive models applied one may test for spectral and amplitude changes (single channnel model), slowness changes (multichannel model applied to array sensor outputs) or polarization (three-channel model) as described by Kushnir (1996). The model fitting can be performed for each guess of the onset time (Taylor et al 1992, Takanami 1991) or the fit to two fixed models fitted at each side of the possible arrival times may be used (Kvaerna 1966a,b). The minimum of the summed residuals occurs when two models are fitted such that the boundary of the fitting regions correspond to the onset time. In any case the estimation process decreases the degrees of freedom in the process, thus decreasing the stability and its sensitivity. Common to all these methods is the assumption that the time interval where the phase onset occurs is known and there are sufficient segments of the time series preceding and following the onset for effect the estimation models.

The approach described above is quite general and is applicable to many kind of problems involving changes in time series. Regardless what kind of test statistic is used for detecting changes a steplike change can be detected easier if a cumulative sum (CUSUM) of it is computed (Der and Shumway 1999, Iclan and Tiao 1994, Basseville and Nikiforov 1993, Shumway 1998), after the application of appropriate prefilters. After imposing a linear trend to the CUSUM, a repetitive onset time estimation is performed, locating the minima of the resulting function, by simulated annealing. The clustering in the multiple onset time estimates provides a means to assess the reliability of the process and the significance of each major arrival. In practical tests the method has outperformed those based testing for changes in autoregressive models.

The approaches based on the cumuative sum statistics are simple to apply and can be automated for applications in a CTBT monitoring environment. Inclan and Tiao (1996) developed a properly centered and normalized CUSUM test statistic and tabulated critical points for testing of significance. Furthermore, they developed an Iterative Cumulative Sum of Squares (ICSS) algorithm that allows sequential identification of multiple changepoints in a white noise series. In this paper we have tested their approach on a set of regional seismic data.

Suppose first that a single channel time series xt, t=1,2,….n is observed and we have a possible changepoint that is to be detected using a CUSUM type statistic. Assume that the time series xt has been prewhitened and has a zero mean, so that the change in regime can be modeled simply by a change in the variance of the white noise process. Inclan and Tiao (1996) proposed using the centered and scaled cumulative sum of the squared amplitudes. First define the sum of squares function over the interval [t1 t2] of length T=t2-t1+1 points as

$$S(t_1, t_2) = \sum_{t=t_1}^{t=t_2} x_t^2$$

The scaled and normalized CUSUM statistic over the interval $[t_1\ t_2]$ at the point $t_1<t<t_2$ is defined as

$$D(t_1, t) = \sqrt{T/2}\left|\frac{S(t_1, t)}{S(t_1, t_2)} - \frac{t}{T}\right|$$

The F statistic for testing for a change of variance at time t is

$$F_{T-t,t} = \left(\left(C_T - C_t\right)/(T-t)\right)/\left(C_t/t\right)$$

This is, of course the standard power detector. With no change in variance in the time interval $t1<t<t2$ $D(t_1,t)$ is a monotone function if t. If the variance increases $D(t_1,t)$ will have a maximum at the change point.

If we assume that the x_i are normally distributed with mean 0 and variances σ then we can obtain the likelihoods for testing one change against no change and let $NT=1$ represent one change. The likelihood for $NT=0$ is

$$l\left(N_T=0;x\right) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log\left[S(t_1,T)/T\right] - \frac{T}{2}$$

The likelihood function for NT=1, change at point t is

$$l\left(t, N_T=1;x\right) = -\frac{T}{2}\log(2\pi) - \frac{t}{2}\log\left[S(t_1,t)/t\right] - \frac{T-t}{2}\log\left[\frac{S(t_1,T)-S(t_1,t)}{T-t}\right] - \frac{T}{2}$$

The best estimate of the change point is where the likelihood ratio is maximized

$$LR_{0,1} = -\max_t\left\{-\frac{t}{2}\log\left[1+\frac{T}{2}D(t_1,t)\right] - \frac{T-t}{2}\log\left[1-\frac{T}{T-t}D(t_1,t)\right]\right\}$$

The function puts more weight on the middle of the time series. This is not a serious disadvantage since prior to applying the algorithm one has a fairly good idea where the main regional phases are (from standard detection algorithms and F-K analyses).

The iterative version of the CUSUM algorithm uses a repetitive application of the approach described above for subintervals of the first time interval. The method applies the following steps;

1) Calculate D(t1,tn) from the start to the endpoint of the initial time segment
2) Search for a significant maximum as defined by equation (3). If one is found at t2 then
3) Search the time interval $t1<t<t2$ for another significant maximum
4) Similarly search the $t2<t<tn$ intervals for a significant maximum
5) Continue the same procedure in other subinterval until no more significant maxima are found.

PRE-FILTERING ISSUES.

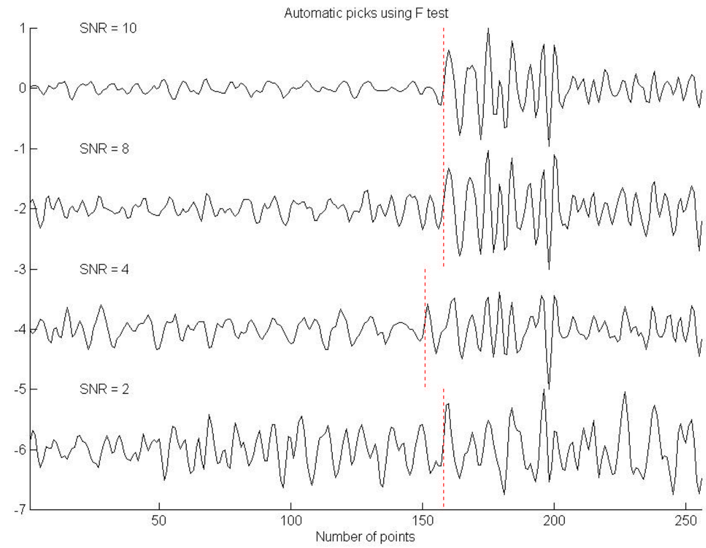
Subtle changes in the frequency content in noisy records often give clue to a human indicating the onset of a seismic phase arrival. On the other hand, when such human capabilities are needed it simply indicates a failure of applying appropriate frequency domain prefiltering that could produce an enhanced amplitude contrast between the noise and the arrival. Appropriate prefiltering in frequency is a prerequisite for further onset time determination regardless of the method (Kvaerna 1996a,b). Various schemes for optimum pre-filtering were suggested and most of these seem to work well in practice and the exact nature of filter is not critical as long as they are reasonably close to the optimum S/N^2 shape (averages of signal amplitude spectrum divided by the squared noise amplitude spectrum). Minimum phase filters based on Kvaerna 's definition of "usable bandwidth" (Kvaerna 1995, 1996a), those shaped according to S/N^2 or noise-adaptive predictive filtering all enhance the amplitude contrast between the noise background preceding the onset of the signal. Empirical fixed 'optimum' filters defined for a given region being monitored or path type seem to work as well. Such filters for regional signals will reduce the amplitude of the low frequency noise below 2 Hz, emphasize the band where the S/N is nearly constant, in the 3-10 Hz range, and reduce the low S/N portions of the spectra at very high frequencies. This enhancement of the signal amplitude happens at the expense of decreasing the visible frequency contrast between the signal and noise. This happens because both regional seismic signals and noise have similar spectral shapes in the 3-10 Hz band, they both fall off with frequency.

The CUSUM-based onset time estimation

The CUSUM algorithms described in Basseville and Nikiforov (1993) were designed for pinpointing the time of a change in a system and have their primary applications in quality control and machine diagnostics. The basic idea is detecting *changes in the trends* of the cumulative sum of some suitable statistic that abruptly changes with time as the properties of the time series changeIt is much easier to see and quantify a change in a trend, than to pinpoint the exact time of the first point where the change occurred (such as picking the first large value of a trace processed by a unit-distance predictive filter). The CUSUM-based methods have indeed been used for determining P onset times in the past (Nikiforov and Tikhonov 1986, Nikiforov et al 1989). In this paper we rely on the approach of Inclan and Tiao (1996).

F TESTS.

F tests are standard tools for verifying the arrival of a new package of energy from the source, i. e. a phase arrival. In this work we pick the arrival time from the modified CUSUM D(t) and test whether the variances on both sides of this arrival are indeed different. **Figure 1** shows two examples of this approach as applied to two Kola peninsula events observed at ARCESS. In these examples amplified noise samples were superposed on otherwise clean signals to simulate various S/N ratios. In making F tests appropriate adjustments must be made to account for the effective bandwidths of the signals.



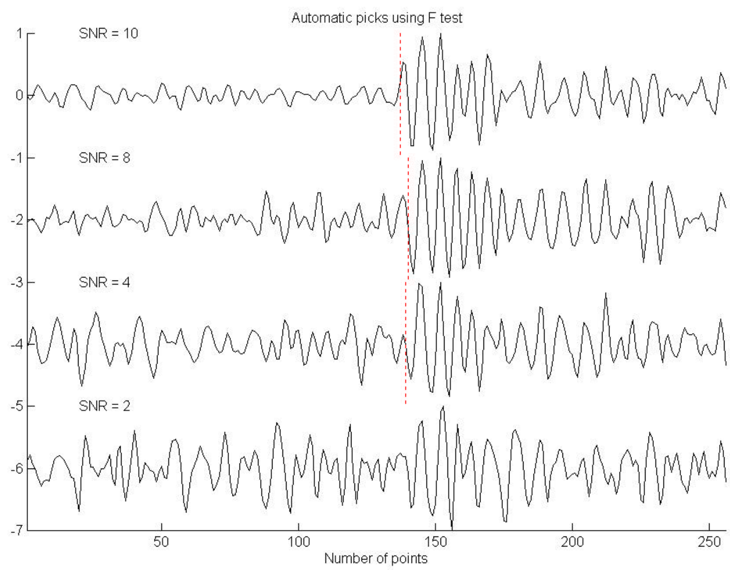


Figure 1. Two examples of F detection of signals under various noise conditions. Amplified noise samples were superposed on otherwise clean signals to simulate various S/N ratios. The F test was applied to the two sides of the window separated by the minimum of the normalized CUSUM statistic D(t). The picks are marked with vertical lines. Only one phase arrival with low S/N was missed.

SIMULATED ANNEALING.

Instead of finding an absolute minimum in the CUSUM-linear trend sum the minimum can also be located by randomized search methods. In the work we have used the method of simulated annealing (SA). SA was designed to find global minima of irregular functions where many local minima may exist. It tends to disregard minor local minima and converge to the lowest points. It uses the randomized Metropolis search algorithm which is based on a thermodynamic analogy (Press et al 1986). Initially it allows the search using large steps in the independent variables which may even be associated with increased values of the function. This allows the solution to “jump out” from local minima and resume search for other minima. As “cooling” occurs such steps are accepted less and less and finally the solution will settle in broad global minima. Repeated application of SA with random starting points will give rise to *populations* of onset time estimates that can be used for evaluation of the efficiency and accuracy of the method.

CLUSTER ANALYSIS.

The multiple application of SA searches automatically provides means for assessing the stability and accuracy of the onset estimates derived from the SA method. Starting out with numerous randomly chosen positions for the onset time within a search window these will converge into positions of prominent CUSUM minima and form a varying number of tight clusters. Besides these there will be hopefully much fewer scattered ‘solutions’ that are obviously spurious and thus must be discarded. This approach was not pursued much in our recent work but it has a considerable potential (Der and Shumway 1999).

PRACTICAL EXAMPLES OF THE APPLICATIONS OF THE METHODOLOGY.

PERFORMANCE OF THE CUSUM PROCEDURE FOR ESTIMATING PN ONSET TIMES.

In the following we show evaluations of the performance of the CUSUM minimum and the combined CUSUM-SA procedure for picking onset times of first-arriving Pn phases. The evaluation is based on comparing the performance of **a)** human analysts **b)** picking the absolute minimum of normalized CUSUM and keeping the arrival if the F tests is passed **c)** picking the median of multiple picks using SA on the CUSUM and taking the median of these, but discarding the result if the SA arrivals are highly scattered.

Prefiltering preceded all the data processing. The two kinds of prefilters applied to the seismograms were 2-7 Hz Butterworth band-pass filters and minimum phase filters designed by taking the S/N^2 spectral ratio such that the maximum was set at unity and cutoffs were placed at the values at 0.24. The latter are similar to the filters that define “useful bandwidth”. We have seen little difference in the performance of these filters in accordance with the comments made by Kvaerna (1996a). The events used had originally very high S/N ratios, especially on the prefiltered traces. To provide a “true” onset time the practically noise-free original trace was picked by the analyst after he has gone through all the noisy examples presented to him. In order to construct noisy data we have fitted a 15-th order AR model to the noise prior to the signal arrival and this model was used to construct independent noise samples by filtering Gaussian white noise. The noise sampled were added to the signal (mean removed) with various signal-to-noise ratios. The differences between this value and the other onset time estimates were plotted for all the three methods against the logarithms of SNR values, with the random time shifts corrected for, of course. This is the same kind of evaluation method as the one used by Kvaerna (1996a,b) and Yokota et al (1981) and Maeda (1985).Examples of results from these procedures are shown in.

The surprising finding of these tests is that most automatic onset times at SNR levels above 2 were comparable in quality to that of the analyst, although the relative
performance of these methods varies from event to event.

APPLICATION OF THE ITERATIVE VERSION OF NORMALIZED CUSUM METHOD TO SEGMENT COMPLETE REGIONAL SEISMOGRAMS.

The iterative method of Inclan and Tiao (1996) was applied to regional seismograms. Typically, three iterations were used. During the first iteration, the largest arrival was commonly picked. In the succeeding iterations the smaller arrivals were identified

Figure 4 and 5 show examples of the onset time determinations by the automatic algorithm described above as applied to the events from the Ground Truth Data Base assembled by Multimax.

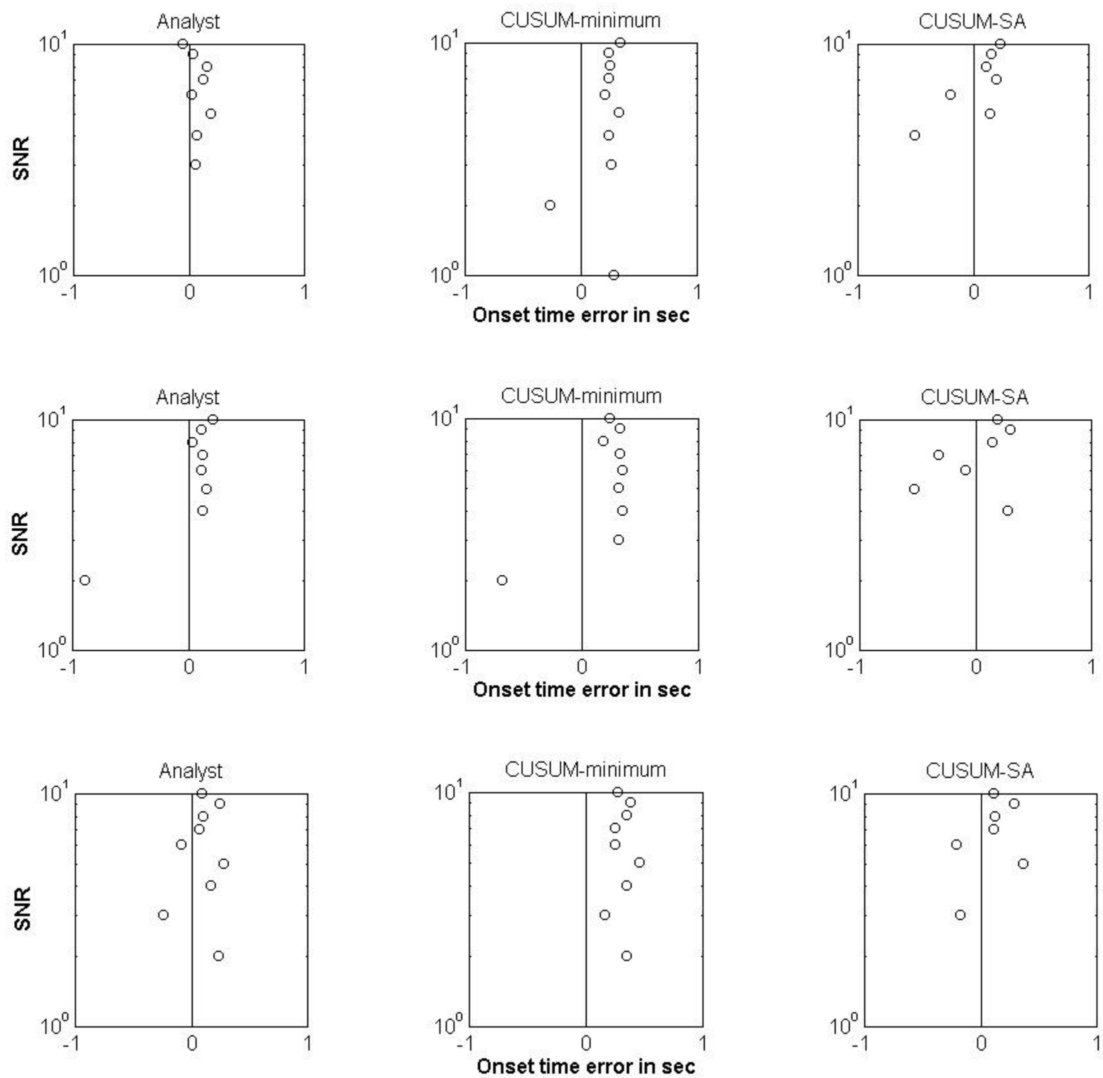


Figure 2. Comparison the Pn onset picks with various S/N ratios. The picks on the left were made by an analyst, those in the middle were CUSUM minimum picks accepted only if the F-test was passed. Those on the right were made by a CUSUM-SA combination, and were accepted only if _ of the trials were less than .5 sec near the median. Even though the analyst has the best performance, the second method come fairly close. Automatic picks tend to be somewhat later than the analyst. These picks were made for the events K2054, K2066 and K2110 listed by Der and Shumway (1999).

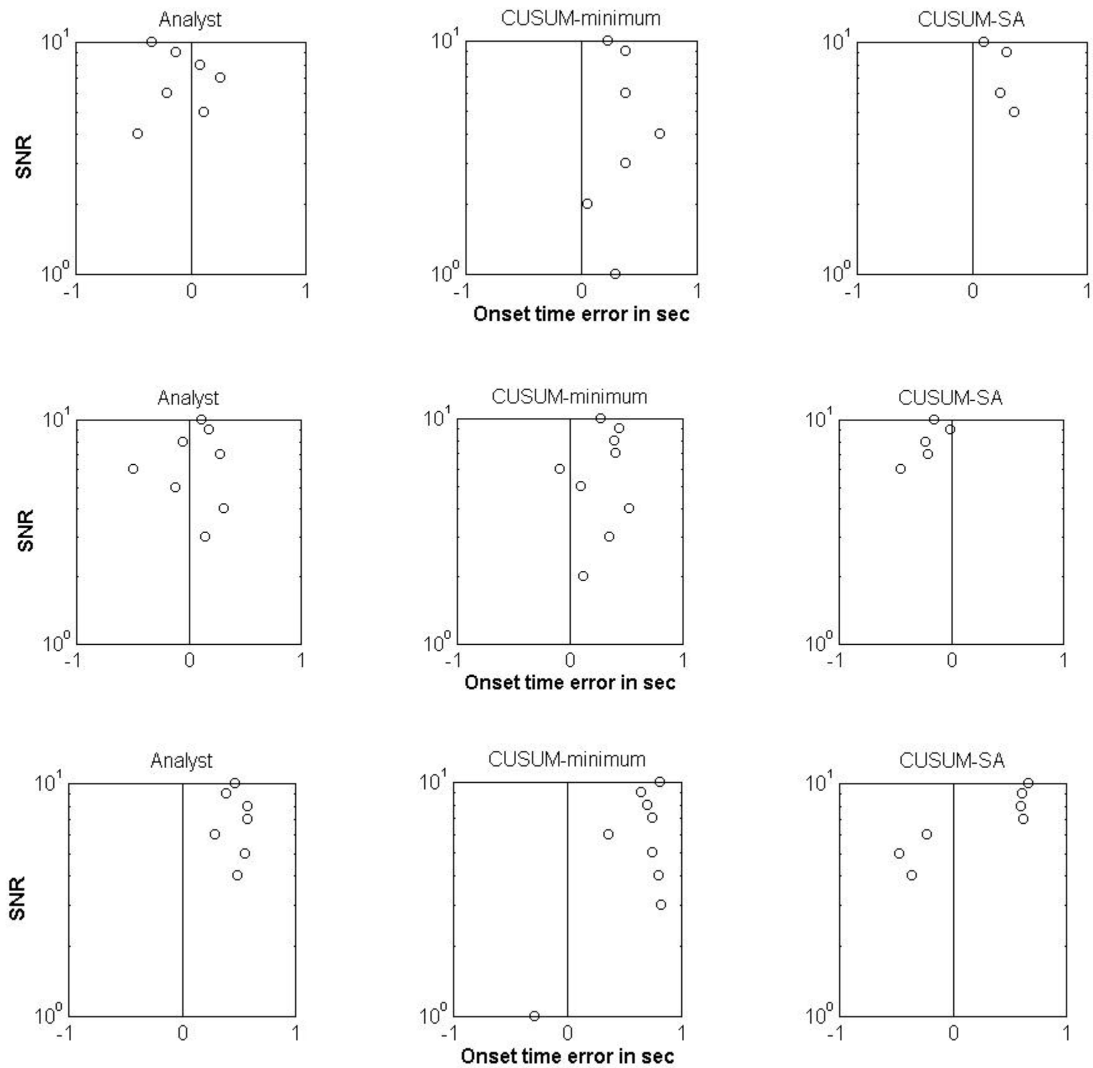
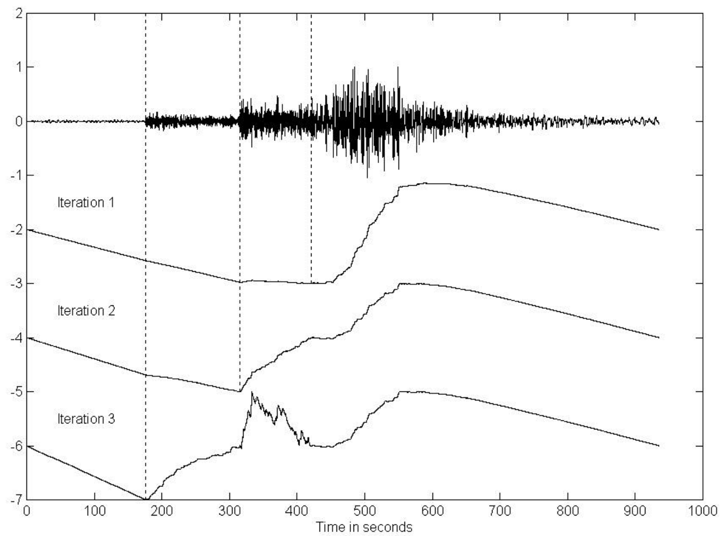


Figure 3. Comparison the Pn onset picks with various S/N ratios. The picks on the left were made by an analyst, those in the middle were CUSUM minimum picks accepted only if the F-test was passed. Those on the right were made by a CUSUM-SA combination, and were accepted only if _ of the trials were less than .5 sec near the median. Even though the analyst has the best performance, the second method come fairly close. Automatic picks tend to be somewhat later than the analyst. These picks were made for the events K2285, K4130 and K5040 listed by Der and Shumway (1999).



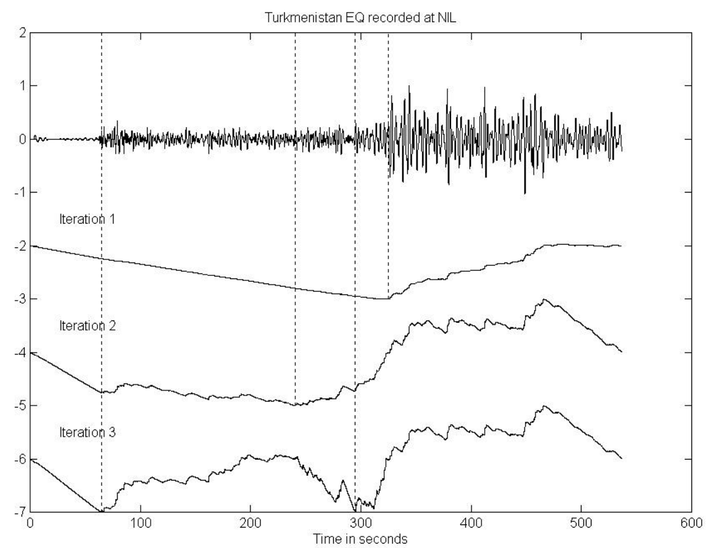


Figure 4. Examples of seismogram segmentation using the iterative CUSUM procedure of Inclan and Tiao (1996). In these figures we plot the raw trace (top) followed by the segmented CUSUM functions between the previously found arrivals. The final onset time picks are the vertical thin lines. Minima that failed the F tests were discarded. The figure shows the results for the stations KIV and NIL recording a Turkmenistan event.

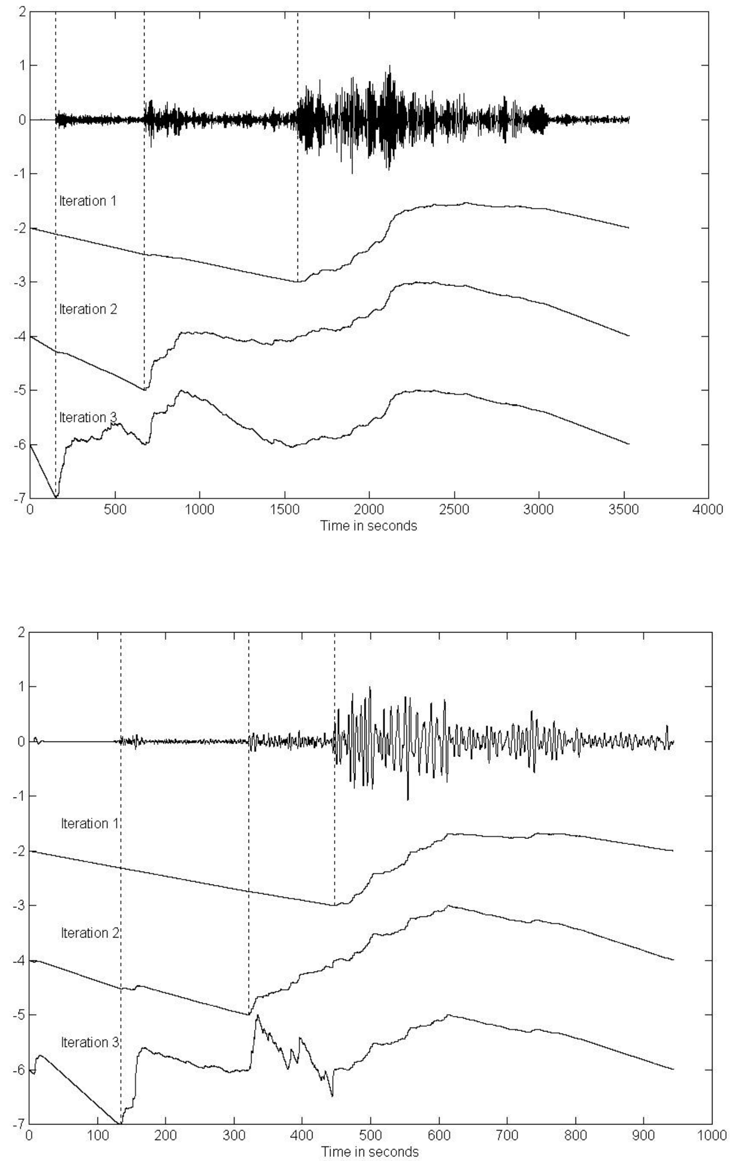


Figure 5. Examples of seismogram segmentation using the iterative CUSUM procedure of Inclan and Tiao (1996). In these figures we plot the raw trace (top) followed by the segmented CUSUM functions between the previously found arrivals. The final onset time picks are the vertical thin lines. Minima that failed the F tests were discarded. The figure shows the results for the stations DBIC and RAYN recording a Northern Iran event.

CONCLUSIONS AND FUTURE PLANS.

CUSUM-based methods seem to be quite suitable for processing regional seismograms since these consist of long wavetrains and have emergent phase onsets. Since CUSUM methods emphasize changes in the properties of signals over several cycles this kinds of methods can be used to segment regional seismograms. CUSUM-based methods to pick seismic phase onset times can also be developed based on a variety of statistics that are diagnostic of polarization, slowness, and spectral changes (Jurkevics 1988, Der et al 1993). Other candidates may include instantaneous relative phase differences among components, adaptive slowness estimates or their combination.

A completely new application of these methods in a CTBT could be the timing of T phase, oceanic seismic and infrasonic wave groups. Upon visual inspection signals of these types do not appear much different in nature than regional seismic signals and the methodology clearly could be used for them. Often such signals contain contributions from multiple raypaths through the atmosphere and the oceans.

REFERENCES

Basseville, M. and I.V. Nikiforov (1993), Detection of Abrupt Changes: Theory and Application. Prentice Hall Information and System Science Series, Prentice Hall, Englewood Cliffs, NJ.

Der, Z.A., Baumgardt,D.R. and R.H. Shumway (1993), The nature of particle motion in regional seismograms and its utilization for phase identification. Geophys. J. Int., 115, 1012-1024.

Der, Z. A. and R. H. Shumway (1999), Phase onset time estimation at regional distances using the CUSUM-SA algorithm. Phys. Earth and Planet. Int., 113,227-246..

Inclán, C., and G. C. Tiao (1994), Use of cumulative sums of squares for retrospective detection in the changes of variance. J. Amer. Statist. Assoc., 89,, 913-923.

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

Kushnir, A. F. (1996), Algorithms for adaptive statistical processing of seismic array data. In ‘Monitoring a Comprehensive Test Ban treaty’, E. S. Husebye and A. M. Dainty Eds., Kluwer Academic Publishers.

Kvaerna, T. (1995), Automatic onset time estimation based on autoregressive processing. Semiannual Technical Summary, 1 April-30 September 1995. NORSAR, Sci. Report No. 1-95/96, Kjeller, Norway.

Kvaerna, T. (1996a), Quality assessment of automatic onset times estimated by the autoregressive method. Semiannual Technical Summary, 1 April-30 September 1995. NORSAR, Sci. Report No. 1-95/96, Kjeller, Norway.

Kvaerna, T. (1996b), Time shifts of phase onsets caused by SNR variations, Semiannual Technical Summary, 1 October 1995-31 March 1996. NORSAR, Sci. Report No. 2-95/96, Kjeller, Norway.

Maeda, N. (1985), A method for reading and checking phase time in auto-processing ssystem of seismic wave data (in Japanese with English abstract), J. Seismol. Soc. Jpn., 38, 365-379.

Nikiforov, L.V. and Tikhonov, I. N. (1986), Application of change detection theory to seismic signal processing. In “Detection of Abrupt Changes in Signals and Dynamical Systems “ M. Basseville and A. Benveniste Editors. Lecture Notes in Control and Information Sciences, LNCIS 77, Springer, New York.

Nikiforov, L.V., Tikhonov, I.N., and Mikhailova, T.G. (1989), Automatic on-line oricessing of seismic data- Theory and applications. Far Eastern Dept. of USSR Academy of Sciences, Vladivostok, USSR (In Russian).

Press, W.H., Flannery, B.P., Teukolsky, S.A. and W.T. Vetterling (1986), Numerical Recipes: The Art of Scientific Computing. Cambridge University Press.

Pisarenko, V. F., Kushnir, A. F. and Savinn, I. V. (1987), Statistical adaptive algorithms for estimation of onset moments of seismic phases. Phys. Earth. Planet Int., 47, 4-10.

Shumway, R. H. (1998), An iterative cumulative sum of squares algorithm for phase onset estimation. ENSCO Technical Memo.

Takanami, T. (1991), A study of detection and extraction methods for earthquake waves by autoregressive models. J. Fac. Sci. Hokkaido, U., Ser VII, 9, 67-196.

Taylor, D. W. A., Ghalib, H. A. A. and Kimmel, R. H. (1992), Autoregressive analysis for seismic signal detection and onset time estimation. ENSCO Technical Report DCS-92-90.

Yokota, T., Zhou, S., Mizoue, M., and Nakamura, I. (1981), An automatic measurement of arrival time of seismic waves and its application to an on-line processing system (in Japanese with English abstract), Bull. Earthquake. Res. Inst. Univ. Tokyo, 55, 449-484.