WCEDS: A Waveform Correlation Event Detection System

Christopher J. Young, Judy I. Beiriger, Julian R. Trujillo
Sandia National Laboratories
Mitchell M. Withers, Richard C. Aster
New Mexico Institute of Mining and Technology
Luciana Astiz and Peter M. Shearer
University of California, San Diego

work completed under DOE ST485D

ABSTRACT

We have developed a working prototype of a grid-based global event detection system based on waveform correlation. The algorithm comes from a long-period detector (Shearer, 1994) but we have recast it in a full matrix formulation which can reduce the number of multiplications needed by better than two orders of magnitude for realistic monitoring scenarios. The reduction is made possible by eliminating redundant multiplications in the original formulation. In the matrix formulation, all unique correlations for a given origin time are stored in a correlation matrix (C) which is formed by a full matrix product of a Master Image matrix (M) and a data matrix (D). The detector value at each grid point is calculated by following a different summation path through the correlation matrix. The Master Image is a critical component in the detection system because it determines how the data contribute to the detector output at each grid point. Master Images can be derived either empirically or synthetically. Ultimately we will use both types, but for our preliminary testing we have used synthetic Master Images because their influence on the detector is easier to understand.

We tested the system using the matrix formulation with continuous data from the IRIS (Incoroporate Research Institutes for Seismology) broadband global network to monitor a 2 degree evenly spaced surface grid with a time discretization of 1 sps; we successfully detected the largest event in a two hour segment from October 1993. Both space and time resolution results are encouraging: the output at the correct gridpoint was at least 33% larger than at adjacent grid points, and the output at the correct gridpoint at the correct origin time was more than 500% larger than the output at the same gridpoint immediately before or after. Analysis of the C matrix for the origin time of the event demonstrates that while the largest values are due to the expected correlations, there are many significant "false" correlations of observed phases with incorrect predicted phases. These false correlations dull the sensitivity of the detector and so must be dealt with if our system is to attain detection thresholds consistent with a Comprehensive Test Ban Treaty (CTBT). Current research efforts are focussing on this problem.

Keywords: waverform correlation, event detection, master image

INTRODUCTION

As the nuclear treaty verification community moves towards the goal of a Comprehensive Test Band Treaty (CTBT), the demands on event detection systems are changing and indications are that current systems will not be able to meet future needs without significant modifications. The next generation of detection systems must have performance which scales well with an increasing number of stations and should be able to take advantage of the new direction in high-performance computing, parallel processing. As part of the CTBT Research and Development program at Sandia Labs, we have been testing a new type of event detection system, the Waveform Correlation Event Detection System (WCEDS), which may be able to meet these requirements and which has other advantages as well. The technique uses waveform correlation and it has already been used to process 10 years of long-period data to search for "slow" or "silent" earthquakes (Shearer, 1994) and to detect and locate a Chinese nuclear test using broadband data (Wallace, 1994). Neither of these applications has established its true worth as a CTBT-quality event detection system, however.

WAVEFORM CORRELATION EVENT DETECTION

1. Qualitative Theory

The basic premise of this type of system is the idea that if data are sorted to the correct distances for a given event, any phases which are present in the waveforms should be obvious because of the moveouts.

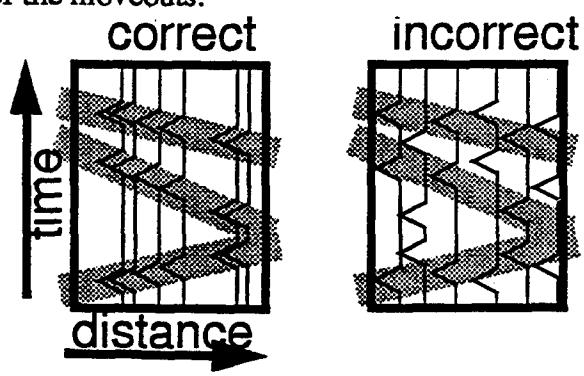

FIGURE 1. A simple example of correctly sorted and incorrectly sorted waveform profiles.

Clearly in the profile on the left the stations are plotted at the correct distances, while in the profile to the right they are not. This simple observation is the basis for the event detection system.

Making an event detection system out of this idea is straightforward. First we must select a network of stations appropriate to the region being monitored. Next we define a grid for the monitored region with a sufficient number of grid points to cover any areas where we think events might occur. Finally, we collect data from each station for the same specified time period (the length is determined by which phases we want to try to observe) and form profiles for each grid point by plotting the data at the appropriate epicentral distances for an event at the grid point. If any of the profiles show the expected moveouts for a seismic event, we declare that an event has occurred at that grid point at the current origin time. To run such a detector continuously through time, we simply repeat this procedure at some regular interval (every 2 minutes for the LP study of Shearer, 1994).

To form an automated system, we use waveform correlation to perform the moveout recognition. Consider two time series X(t) and Y(t). The the dot product of Y(t) with X(t) is:

(Equation 1)

$$Y(t) \bullet X(t) = \sum_{i=1}^{N} X_i Y_i$$

This is the zero time shift correlation of Y(t) with X(t). The greater the similarity, the greater the correlation will be.

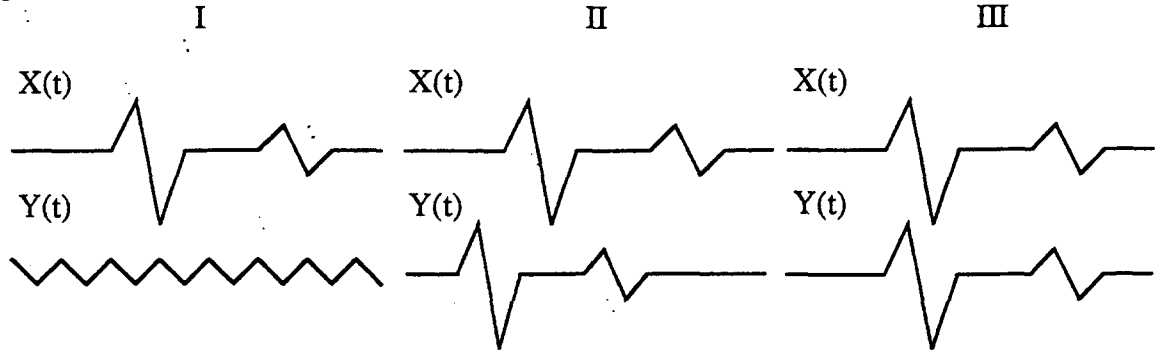

FIGURE 2. Examples of waveform pairs with poor, mediocre, and excellent correlations . Both waveform similarity and correct time alignment are important.

In case I, the waveforms are completely dissimilar and the dot product is small. In case II, the waveforms are similar but the time alignment is poor and the dot product is still small. Finally, in case III the waveforms are similar and they are time aligned so the dot product will be large. Thus, correlation can be used to identify similar waveforms.

For our detector, the data from each station are the waveforms which we wish to identify as valid or invalid with respect to a hypothesized event location (grid point) and origin time. Hence, the correlation waveform must be the expected waveform at the appropriate distance. High correlations (large dot products) will indicate that the observed waveform is consistent with an event at that grid point.

2. Master Images

The expected waveforms can either be derived synthetically or empirically. To derive them empirically, one can take the network data from a large number of events with reliable locations and bin and stack the data (see Shearer, 1991 for details). The correct waveforms emerge quite naturally as a result of averaging the many contributions, and if enough events from different locations are used, expected waveforms for all distances will be generated. If we consider all of the stacked waveforms together they form an image which looks like a plot of travel time curves. An example formed for this project is shown below.

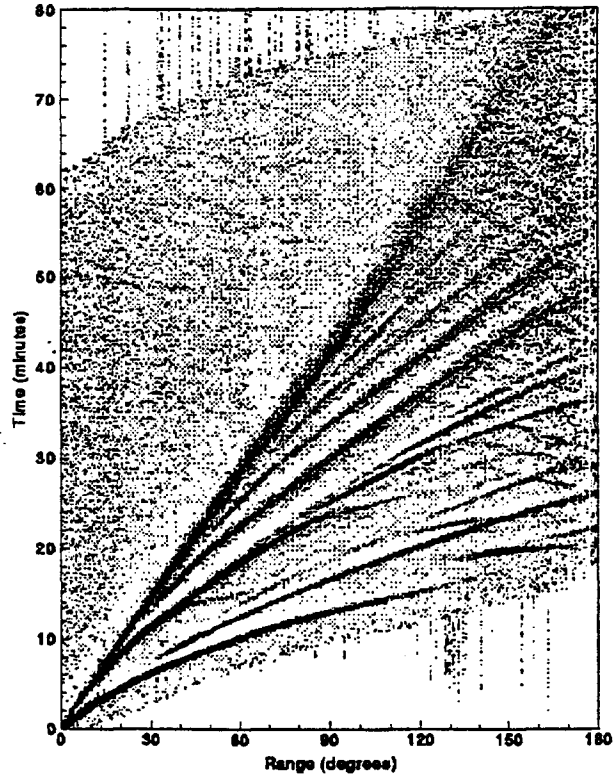

FIGURE 3. An empirical stack using the IRIS BB stations for all events with M_w > 5.8 and depth < 50 km.

We will refer to this type of image as a Master Image or MI. Empirically derived MIs have the advantage that the images do not require any phase picks and by definition must produce the true image (assuming that the event locations are accurate). The technique of producing an empirical MI is described in detail in Shearer, 1991. To form empirical MIs for WCEDS, we have gathered all of the IRIS BB data from 1988-1994 for shallow (d < 100 km) events with Mw > 5.8 and for deep (d > 100 km) events with Mw > 5.5. From this, we have formed images for d=0-50 km, 50-150 km, 150-350 km, 350-550 km, and 550+ km.

Unfortunately, the complexity of these images makes them ill-suited for our need to understand the detailed workings of WCEDS. For this reason, we have chosen to use synthetic MIs for our initial testing. Synthetic MIs offer the advantage of being able to control exactly which phases are contained in the image and their relative amplitudes. This information is important because it in turn determines which phases can contribute to an event detection and how they will be weighted. Our synthetic MIs were derived from the IASPEI 1991 travel time tables. Because travel time curves have no duration associated with their phases, the time series for each distance must be convolved with some response function(s) to yield waveforms with correlation windows of finite duration:

FIGURE 4. Transformation of a spike train-style waveform to a waveform that can be used for correlation by means of convolution with a simple response function.

3. Quantitative Theory

3.1 Original Formulation

In the original formulation of the detector (Shearer, 1994), the algorithm is essentially identical to what we have outlined above. For each grid point the data are sorted to the appropriate distance bins and the dot products of each bin with the master image bin are formed and summed to give the detector output at a given gridpoint at a given time:

(Equation 2)

$$E_{gridpoint} = \sum_{i=1}^{N_D} \sum_{j=1}^{N_T} M_{ij} D_{ij}$$

Where M_{ij} is the master image for distance bin i and time point j, D_{ij} is the data for distance bin i and time point j, N_D is the number of distance bins, and N_T is the number of time points to correlate. Thus, for each origin time to be monitored this double summation must be done for all grid points.

For a coarse grid the number of grid points is small (Shearer needed only 416 points for approximately 10 degree spacing on the Earth's surface) and this formulation is acceptable, but for a finer grid the method becomes inefficient (e.g. for our gridding scheme, 1 degree grid spacing for a single surface grid leads to about 41,000 points).

3.2 Matrix Formulation

The number of multiplications needed for the WCEDS processing can be reduced significantly (e.g. 2 orders of magnitude) by using a matrix formulation in which all dot products needed to monitor each grid point for a given origin time are calculated with one full matrix multiplication. The motivation for the matrix formulation is provided by the observation that in the original formulation many redundant dot products are calculated. Consider a given station at a distance of *r* degrees from the grid point being monitored. This same station will contribute exactly the same dot product to every other grid point the same distance away:

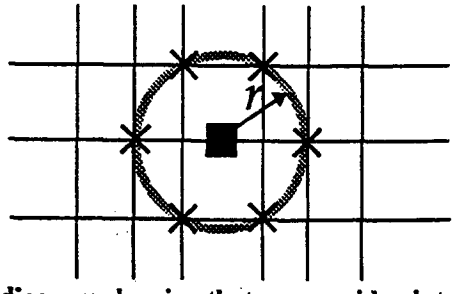

FIGURE 5. A simplified diagram showing that many grid points will lie at the same distance from each station. The shaded circle indicates equal distance from the station (solid square). Six grid points (indicated with Xs) lie at this distance from the station.

Another way to think of this is that each station can only form N_D unique dot products (i.e. it can be dotted once with each of the distance columns of the master image). This suggests a simple shortcut: form all of the possible dot products for each station immediately and then simply look up the dot products needed to monitor each grid point as they are needed. The only thing complicating this formulation is the binning of the station data in the original formulation, which (ironically) was done to limit the number of dot products. If we eliminate the distance binning, however, a much greater gain can be realized.

The matrix formulation is shown graphically below for N_S stations with a correlation window length of N_T points, and a distance discretization of 1 degree for clarity:

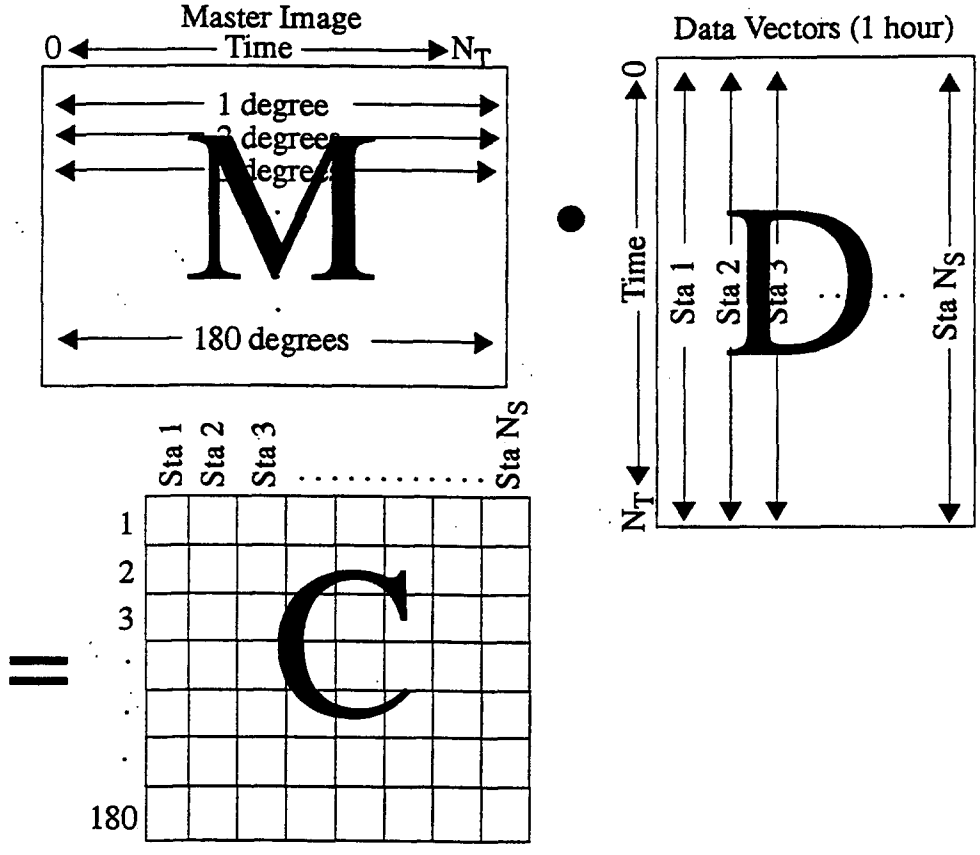

FIGURE 6. The matrix formulation of WCEDS. All of the information needed to monitor each grid point on the earth (contained in the correlation matrix C) can be obtained by forming the full matrix product of the Master Image matrix (M) with the data matrix (D).

Let us compare the number of multiplications needed to monitor a full grid at the surface using both the original and the matrix formulations for a realistic scenario. Let $N_S = 50$, $N_T = 3600$ (a 1 hour correlation window sampled at 1 sps), and using a 1 degree grid spacing (about 41,000 grid points):

- **original formulation**: 41,000 grid points x 50 stations x 3600 time points = 7.38 e09 multiplications. The true number is likely to be somewhat less because of the binning (for each grid point some of the stations will lie in the same distance bin).
- **matrix formulation**: 50 stations x 180 distances x 3600 time points = 3.24 e07 multiplications.

Thus we reduce the number of multiplications by a factor of 227 using the matrix formulation. Even if we assume that binning in the original formulation reduces the number of multiplications at each grid point by as much 50%, the reduction factor is still better than two orders of magnitude.

Once C has been calculated, event detection is simply a matter of referring back to the elements of C. E.g.:

grid point 1 --	**Station**	**Distance**	**C Matrix Element**
	#1	10.6 (11)	C(11,1)
	#2	88.7(89)	C(89,2)
	#3	23.4(23)	C(24,3)
	...	...	...

#50 16.7(17) C(17,50)

Correlation for grid point 1 = C(11,1) + C(89,2) + C(24,3) + ... + C(17,50)

The correlations for the rest of the grid points can be calculated in a similar manner.

3.2.1 Importance of Master Images

In general, the output of the detector at a given gridpoint is given by a particular summation path through the correlation matrix:

(Equation 3)

$$E_{gridpoint} = \sum_{j}^{N_S} C_{ij}$$

Where N_S is the number of stations and i = function(gridpoint, j). From Figure 6, we se that the C_{ij} are a weighted sum of the data D_{lj}:

(Equation 4)

$$C_{ij} = \sum_{l}^{N_T} M_{il} D_{lj}$$

Where N_T is the number of time points used in the dot products and M_{il} are the elements of the master image. Thus, the master image weights the contributions of the data to the C matrix and thereby to the detector output at each grid point.

4. Example Application

To test the WCEDS prototype, we used a two hour segment of data surrounding an m_b 6.2 event in southern Xinjiang Province, China which occurred at 08:42:32.7 GMT on October 2, 1993. Although there are four other events listed in the PDE for this interval, the Chinese earthquake is by far the largest. The output of our detector reflected this: the maximum value occurs at the grid point nearest to the PDE epicenter of the Chinese event at the given origin time (rounded to the nearest second).

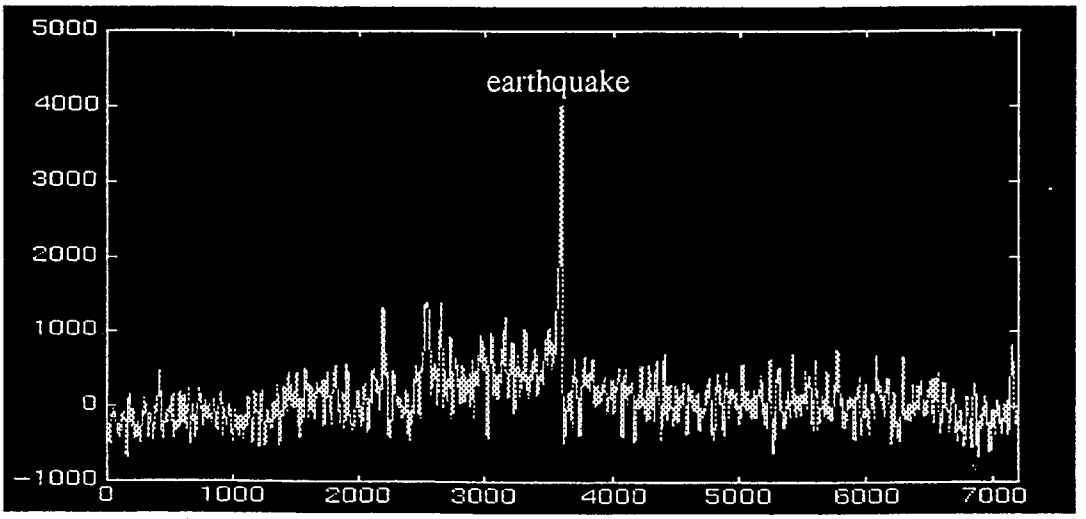

FIGURE 7. Detector output for the grid point nearest the Chinese earthquake for 1 hour before until 1 hour after the event.

Viewing the C matrix sorted for the epicentral grid point well before the event and at the time of the event sorted demonstrates that the correlations are occurring as theorized:

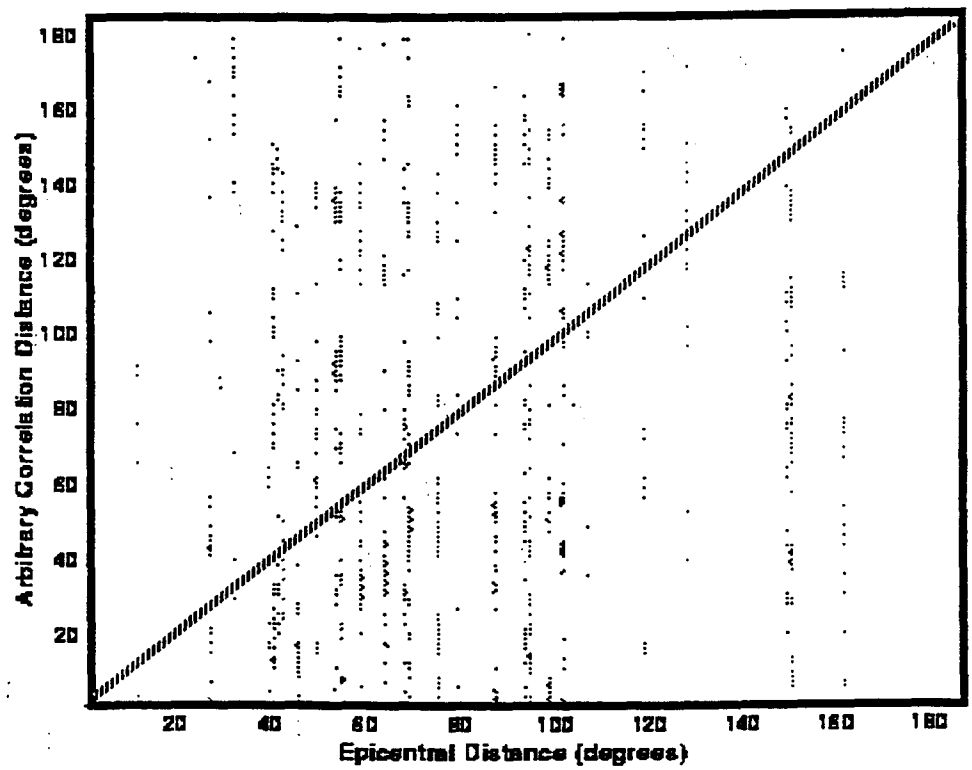

FIGURE 8. C matrix with for an origin time 1 hour prior to the event. The columns have been plotted at the correct distance for grid point nearest to the event location. The dashed line shows the expected positions for the correlations for the event.

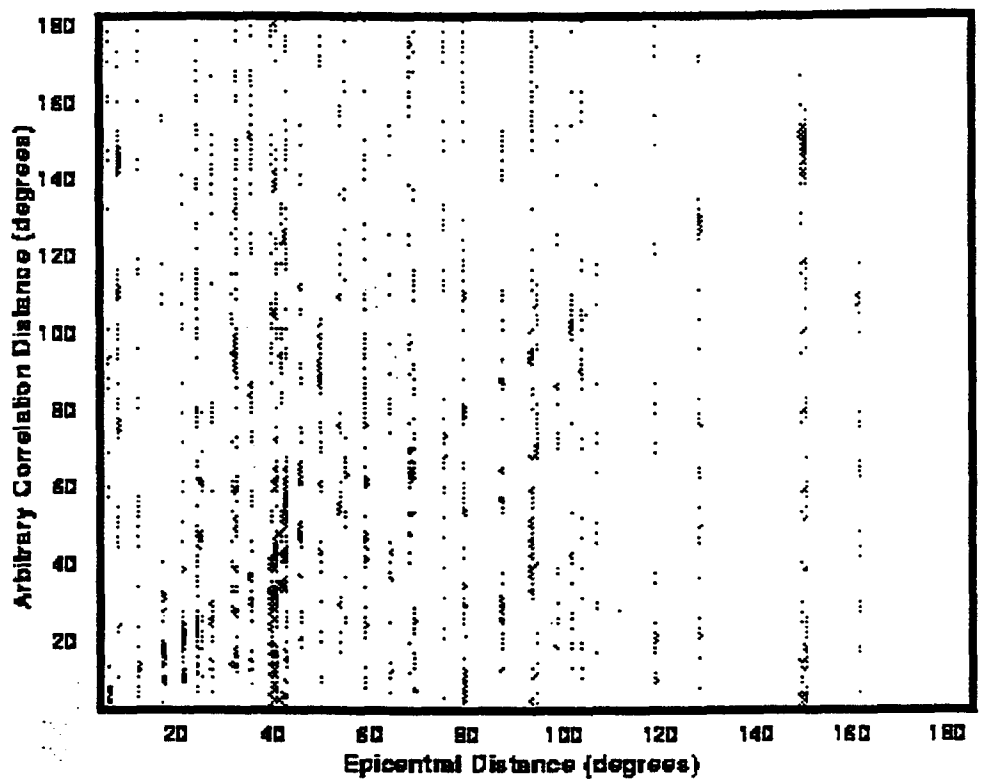

FIGURE 9. Same as Figure 7, but at the correct origin time.

FUTURE PLANS

If we examine the above figure carefully it can be seen that in addition to the expected line of correlations along the diagonal, there seem to be several other correlation lines:

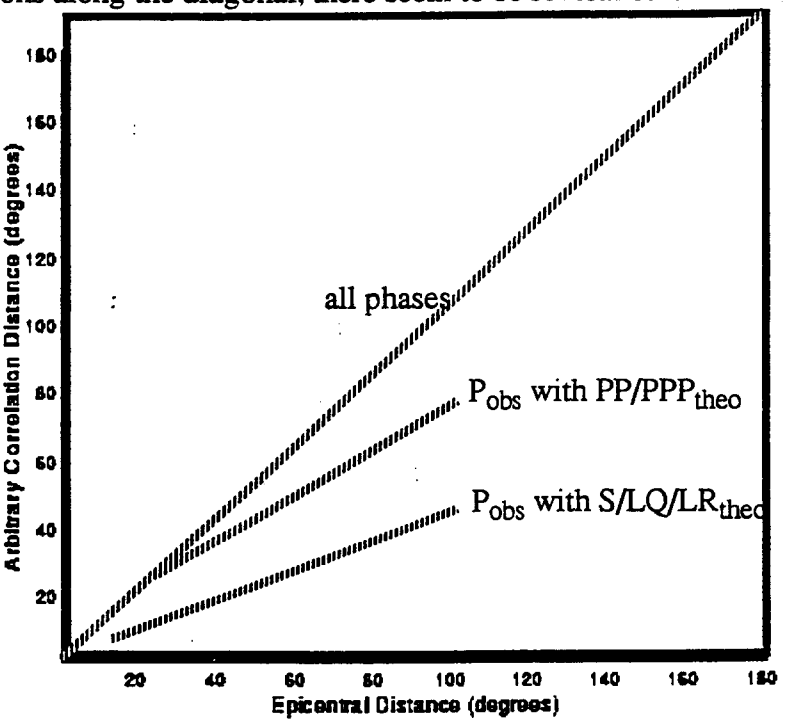

FIGURE 10. Correlation lines observed in Figure 9.

These other lines are the correlations of observed phases with the wrong theoretical phases. In particular, we can pick out P_{obs} with PP/PPP (notice that they begin to split near 40 degrees), and P_{obs} with S/surface waves. The "false" correlations do not contribute as much to the detector as the true correlations because fewer phases contribute:.

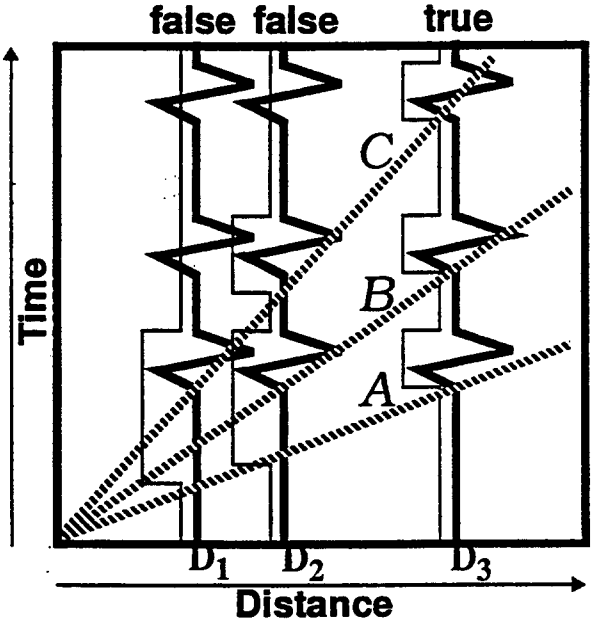

FIGURE 11. True and false correlations of a single observed waveform with theoretical (boxcar) correlation waveforms for a simple system with three phases: A, B, and C. Note that Correlation(D_3) > Correlation(D_2) > Correlation(D_1) because three phases contribute at D_3, two at D_2, and only one at D_1.

False correlations do, however, raise the background noise level of the detector because they randomly add to the detector output at grid points when no event is actually occurring. Thus, to lower the detection threshold of our system, we must find a means to eliminate the false correlations, or minimize their effects.

There are two obvious ways of eliminating the false correlations. The most obvious is to go back to the data, zero out the portions where signals were recorded which matched predicted phases for the event, and then recalculate the C matrix and rerun the detector. This method is not difficult to apply, but we would prefer to avoid it because of the recalculation of the C matrix, which we have found to take nearly 90% of the run time for the entire detector. A more computationally efficient method is not to recalculate the C matrix at all, but rather to determine where and when false correlations will occur and to use this information to screen the C matrix whenever running the detector. What we desire is an overlay for the C matrix which will "x" out the elements contaminated by false correlations: we will refer to this as the X matrix.

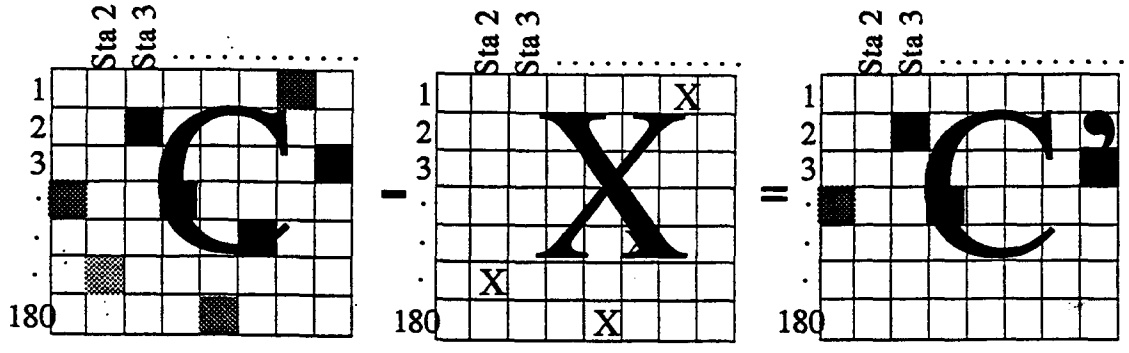

FIGURE 12. An example demonstrating the use of the X matrix to screen nuisance correlations from theC matrix: C - X = C'.

The implementation of the X matrix is one of our active areas of research.

CONCLUSIONS

We have developed a simplified prototype of a Waveform Correlation Event Detection System (WCEDS) which we have used to successfully detect the largest event in a two hour interval from October, 1993 using data from the IRIS broadband global network. The algorithm is based on a long-period detector developed by Shearer (1994) but we have recast it in a full matrix form which can reduce the number of multiplications needed by better than two order of magnitude. Even with the many simplifications which have been made in implementing the prototype, both the performance and the quality of output of the system show great promise. To make WCEDS a CTBT verification system however, will require significant progress in many areas.

REFERENCES

Bache, T. C., S. R. Bratt, J. Wang, R. M. Fung, C. Kobryn, and J. W. Given (1990). The intelligent monitoring system, *Bull. Seismol. Soc. Am.*, **80**, Part B, 1833-1851.

Shearer, P. M. (1994). Global seismic event detection using a matched filter on long-period seismograms, J. Geophys. Res., 99, 13713-13726.

Shearer, P. M. (1991). Imaging global body wave phases by stacking long-period seismograms, J. Geophys. Res., 96, 20353-20364.

Wallace, T. (1994). Chinese nuclear test - June, 1994, *IRIS Newsletter*, **2**, 6.