SAND2011-2456 C

Towards an Automated Waveform Correlation Detector System

Megan Slinkard, Dorthe Carr, Stephen Heck, and Christopher Young | *Sandia National Laboratories**

MOTIVATION AND OBJECTIVES

Swarms of earthquakes and/or aftershock sequences can dramatically increase the level of seismicity in a region for a period of time lasting from days to months, depending on the swarm or sequence. For those who monitor seismic events for possible nuclear explosions, these swarms/sequences are a nuisance because each event must be treated as a possible nuclear test until it can be proven, to a high degree of confidence, not to be. Fortunately, swarms typically consist of groups of very similar looking waveforms, suggesting that they can be effectively processed using waveform correlation techniques, which have been shown to have excellent sensitivity and robustness.

We have designed a prototype Waveform Correlation Detector (WC Detector) which is used to simulate applying a waveform correlation based process to large aftershock sequences. Previous work has demonstrated that a high percentage of events in the sequences we studied could be detected and identified using the WC Detector. As we move toward an operational system, we must find ways to 1) recognize that a swarm has started, and 2) select appropriate parameters for use in running the WC Detector. This poster describes our techniques for automating and optimizing the setup of an operation system.

WAVEFORM CORRELATION PROCESSING OF CONTINUOUS WAVEFORMS

We developed the Waveform Correlation Detector (WCD) to simulate a real-time system where incoming raw data is compared to archived waveforms in order to screen out similar events. It is intended for use during an aftershock/swarm sequence to aid analysts by allowing them to quickly identify new events with a high degree of waveform similarity to a previous event already determined to be part of the aftershock sequence (Figure 1). Our system compares the incoming data stream to previously identified origins held in a "library" of master events. Our WCD flow is captured in the above flow chart (Figure 2). Our algorithm operates on a single station, during a prescribed time period. The incoming raw data stream is filtered, windowed, and then correlated with each waveform in the Master Waveform Library. If the data stream and a particular library entry have a correlation value above a threshold, then we declare a recognized similar event. Detected matches were identified as either a catalogued match or as a new (un-catalogued) event. The incoming data stream is then advanced one sample, and the process repeats.

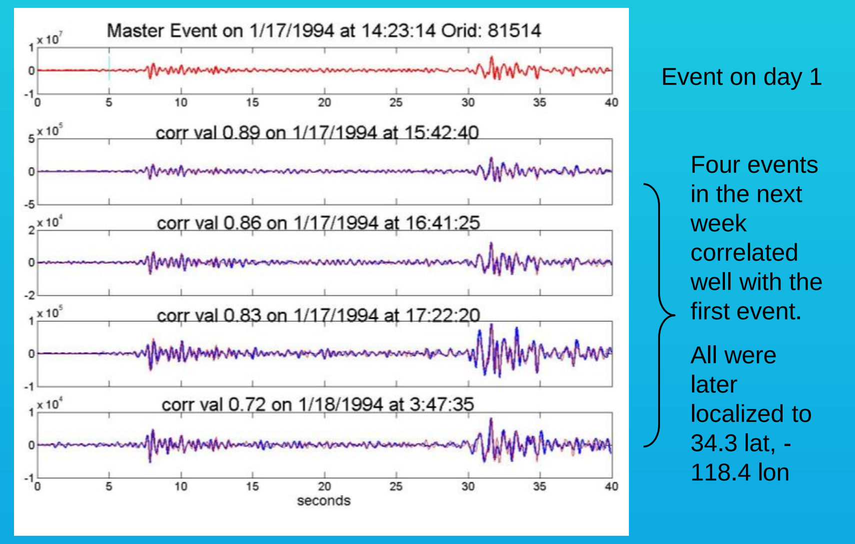


Figure 1: A typical family found by the WCD. The master waveform is shown in red, and the found matches are shown in blue.

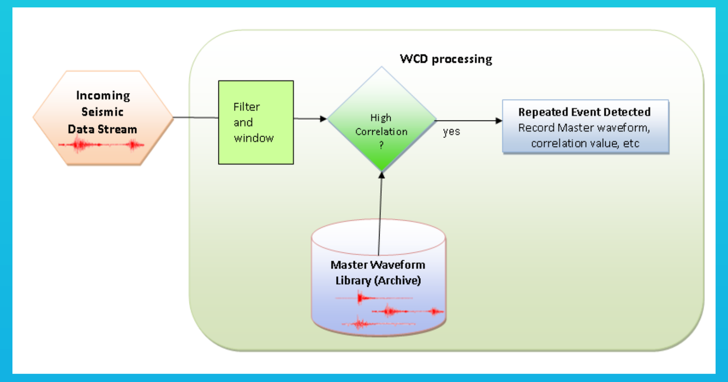


Figure2: Our WCD flow. Our algorithm operates on a single station, during a prescribed time period. The incoming raw data stream is filtered, windowed, and then correlated with each waveform in the Master Waveform Library. If a correlation is above the threshold, then we say a match is found, and record information such as the start time of the data segment, the correlation strength(s), and the master waveform(s) which found the match. The incoming data stream is then advanced one sample, and the process repeats.

Event	Station	Station distance	Window Length	Corr Threshold	# of catalog events seen at station	% catalog events belonging to a family	# of additional events identified
Northridge (M_w 6.7)	PAS	27km	40 sec	.5	412	**92%**	**942**
Wenchuan (M_w 7.9)	CD2	39km	40 sec	.5	262 (data dropouts)	**39%**	**300**
Pakistan (M_w 7.6)	NIL	99km	40 sec	.5	440	**78%**	**740**
Northridge (M_w 6.7)	MHD	348km	40 sec	.5	352	**59%**	**208**
Wenchuan (M_w 7.9)	XAN	621	90 sec	.5	752	**29%**	**218**
Pakistan (M_w 7.6)	AAK	907km	120 sec	.5	360	**24%**	**10**

Table 1: This table shows the effectiveness of the WC Detector *before* implementing the automated parameter selection describes in this poster. Results are taken from a paper under review. The success of our prototype encouraged us to continue to exploring optimized automation.

AUTOMATED PARAMETER SELECTION

To move towards an operational system, the system must automatically recognize that a swarm has started, set up a library of master events from a suitable region, pick parameters necessary for running the WC Detector, and start processing.

Based on our previous work, we determined that the following parameters had significant influence on the accuracy of results, and needed to be chosen with care: window length, filter band, and, most of all, correlation threshold. Our goal is twofold: first, to determine optimal parameters, and second, to develop ways of automating the software to determine these optimal parameters.

WINDOW LENGTH

Window length refers to the number of seconds of waveform captured for the master waveforms stored in the Master Waveform Library. We found that this parameter has a significant effect on the number and quality of matches found. Too short leads to false matches – S arrivals can correlated with a master waveform's P arrival, short snipets correlate when the overall envelops don't, etc. A longer than necessary window wastes processing time (calculating correlations is computationally expensive), and, increases the probability of new arrivals corrupting the signal. We found that a window length that includes the P arrival and beginning of the S arrival is the goldilocks length. This also helps improve accuracy, since event to station distance is reflected in the P-S separation.

Our approach: Given an station and an event region, we determine the difference between the theoretical P and S arrivals for historical events. We set the window length to 1.2 times the median P to S separation.

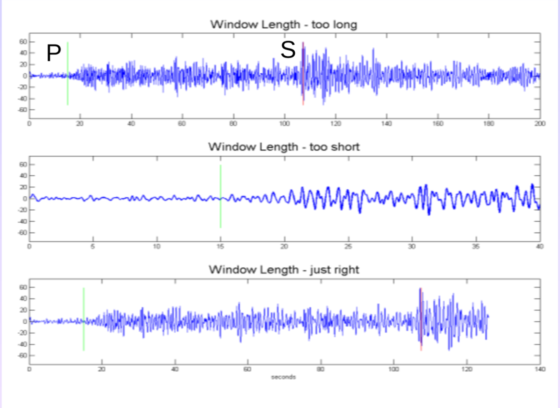


Figure 3: The same master waveform plotted at different window lengths. The green bar is the theoretical P arrival, the red the theoretical S arrival.

FILTER BAND

The optimal filter band depends on distance and path between the swarm and the station. We are still exploring how to determine an optimal filter band. Results shown on this poster were obtained using a 3rd order butterworth bandpass filter with cutoff frequencies .8 and 3.5 Hz.

REMOVING POOR DATA

As we lowered the correlation values used, we would occasionally declare a match which an analyst did not agree was a match. This almost always occurred when the incoming data window was very low amplitude except for an arrival at the end. We believe this is an artifact of using the normalized correlation coefficent. To discard these false matches, we check that the energy distribution in the data window is distributed appropriately for a seismic event. If the energy is disproportionately in the last half of the window, we discard it.

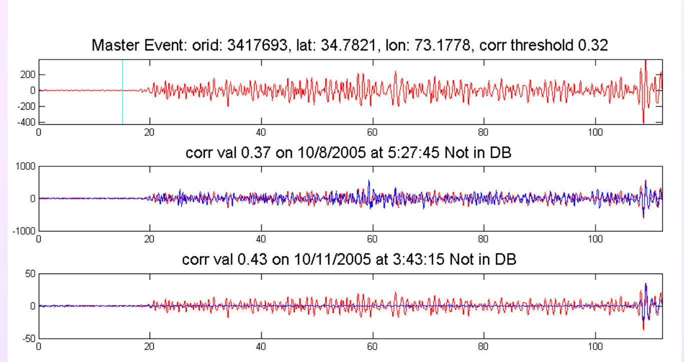


Figure 4. The top plot is the master waveform (red). The next two plots show waveforms (blue) that correlated with the master event (red).

SELECTING THE CORRELATION THRESHOLD

Selecting the correlation threshold is one of the most critical factors in the success and accuracy of the WC Detector. We wanted an objective method for automatically determining a suitable threshold for each master waveform. Calculating the threshold for a given probability of false alarm depends on the time-bandwidth product of the waveform'; thus, it depends on the window length and filter band chosen. Using Wiechieki-Vergara's technique for determining a threshold given a suitable Pe, we originally used station background noise. However, we felt the threshold returned was too low, and that the resultant families did not always look similar to the eye. We decided to instead treat distant events as noise; This method raised the correlation threshold and yielded resultant families that looked similar.

Method:

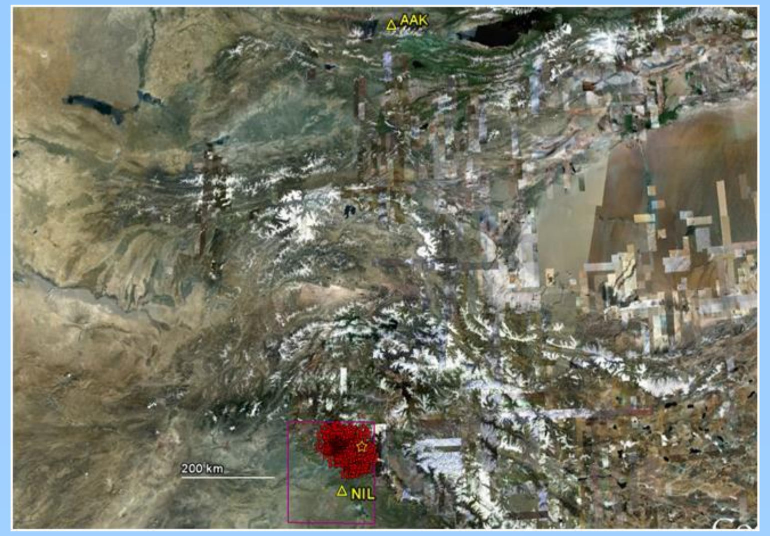
Assume that events more than 50 km distant from each other are effectively noise.

For each master waveform: Compare master to all other master events which are > 50 km away.

Use Wiechieki-Vergara's technique to figure out the threshold for a Pe which give 100 years between false matches. Take the mean of the results

For our dataset (described next), this returned correlation thresholds ranging from .26 to .39.

DATASET



To test our automated WC Detector we used the Pakistan (Kashmir) earthquake.
•Occurred on October 8, 2005 in northern Pakistan.
•Mw of 7.6.
•70 km northwest-trending thrust surface rupture (Kaneda, et al., 2008)
•Strongest earthquake in area for 100 years.
•More than 75% of the aftershocks occur in a cluster around 30 km southwest of the strike of the main rupture (Bendick, et al., 2007)
•Aftershocks used in our study were limited to a lat-lon box of 33-35°N and 72-74°E; the diameter of the cluster is approximately 150 km.
•According to the IDC-REB catalog there were 282 recorded events in the first 24 hours and 502 recorded events in the first 7 days.
•The time period used in our study was October 8 02:00 to October 13 02:00 2005; there were 462 aftershocks in the IDC-REB catalog for this period.
•We retrieved data from station AAK, 907 km away from the main shock.

RESULTS

The automated WC detector identified 47% of catalogued events as belonging to a family. It also found an additional 183 new, un-catalogued, signals.

Compared to our previous WC Detector results, the results from the automated WC Detector are an improvement both in % of events found and # new signals found. The addition of code to remove bad matches also improved the quality of the families.

Run	Station	Filterband	Window Length	Corr Threshold	# of catalog events seen at station	% catalog events belonging to a family	# of additional events identified
Short time window	AAK	.8-3.5Hz	40 sec	.5	360	**4%**	**3**
High correlation	AAK	.8 – 3.5 Hz	120 sec	.5	360	**24%**	**10**
Automated parameter selection	AAK	.8 – 3.5 Hz	112 sec (auto)	**.26 - .39 (auto)**	360	**47%**	**183**

Table2: Effectiveness of the WCD Detector on the Pakistan earthquake dataset, for 3 different combinations of parameters settings. We found our automated system performed substantially better than previous runs.

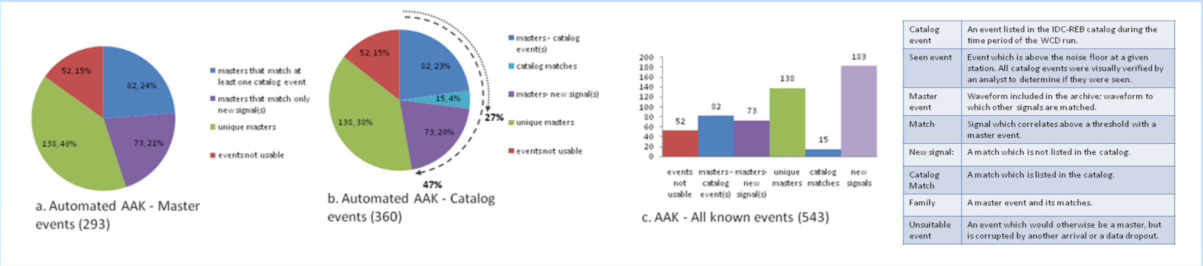


Figure 5. A detailed look at results from the automated WC Detector. A) shows the breakdown of events considered for the Master Waveform Library. B) Shows all cataloged events included in the study. C) shows the number of new, un-cataloged, signals found by the detector.

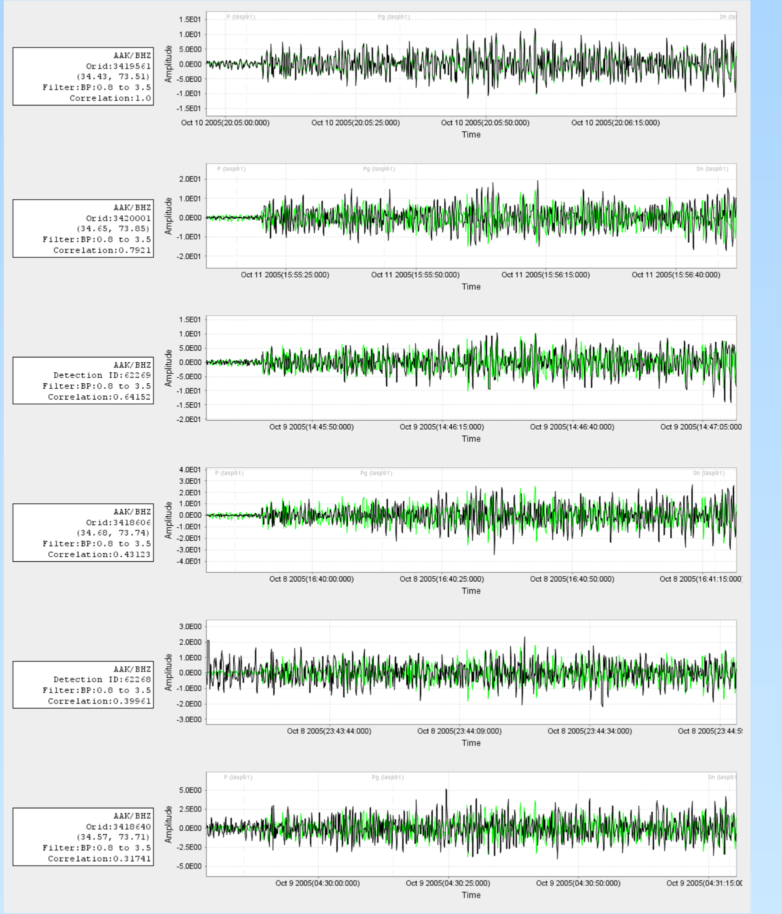


Figure 4. A typical family returned by the automated WC Detector. The top plot is the master waveform . Matches found by the detector are shown below; black is the match, green is the master shown for comparison. Some of the waveforms have an Orid associated with them, these are catalog matches; some have a Detection ID, these are new signals.

FUTURE WORK

Automated parameter selection is an important step toward an operation system. However, much additional work needs to be done before that vision can be a reality.

Integrating Waveform Correlation Results across a Network: Our work to date has focused on using waveform correlation on a station-by-station basis. For an operational system, waveform correlation must be used for a network of stations. In further research we plan to explore how to combine the results from multiple stations.

Multiple Family Correlations: In our current project we have found several cases where an incoming waveform correlates with more than one master event waveform. It is not clear how to resolve this ambiguity because the highest correlation match does not always match an analyst's choice for a match. This situation typically, but not always, occurs when the new event is located geographically between the two master events – close enough to both of them to correlate even though the masters aren't close to one another.

Integrating a WC Detector with traditional event detection and identification: An operational system would require integrating with the existing processing scheme. A WC Detector only finds repeated events; it will not replace traditional processing.

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

* {meslink, dbcarr, sheck, cjyoung}@sandia.gov // **Ground-Based Nuclear Explosion Monitoring Research and Engineering**