Processing the Tohoku Earthquake using Waveform Correlation

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

SAND2011-6613C

National Nuclear Security Administration

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.

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. We have applied our WC Detector to several large aftershock sequences, and have found it detected 47% - 92% of events. For this discussion, we chose to study the noteworthy Tohoku sequence.

WAVEFORM CORRELATION PROCESSING OF CONTINUOUS WAVEFORMS

We developed the Waveform Correlation Detector to simulate a real-time system where incoming raw data is compared to archived waveforms in order to identify similar events. The intended use is to aid analysts to quickly identify new events with a high degree of waveform similarity to previously seen events from an aftershock/smarm sequence (Figure 1). Our system compares the incoming data stream to the waveforms of previously identified events held in a "library" of master waveforms. The WC Detector flow is depicted in the flow chart (Figure 2). Our algorithm operates on an array, 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 are identified as either a cataloged match if they can be matched with an arrival from the IDC-DEB catalog, or as a new (un-cataloged) event. The incoming data stream is then advanced one sample and the process repeats.

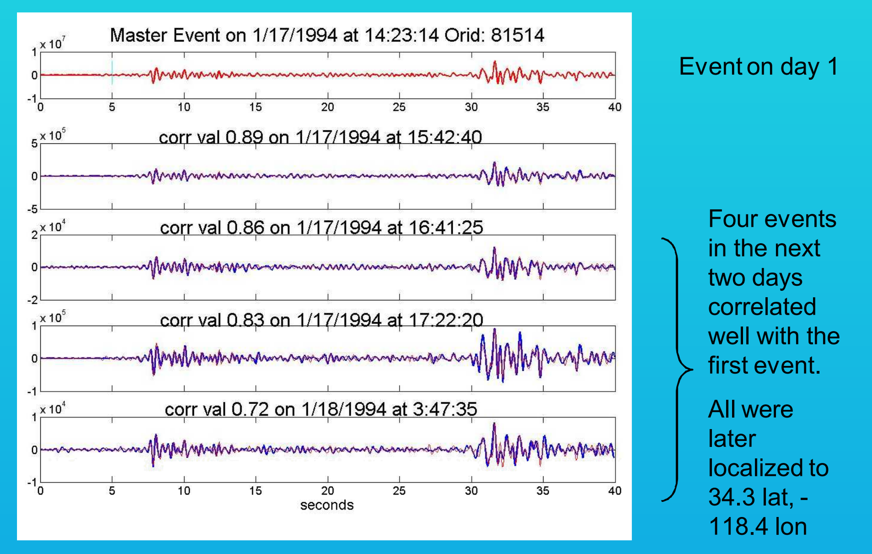


Figure 1: A typical family found by the WCD. The master waveform is shown in red, and the found matches are shown in blue.

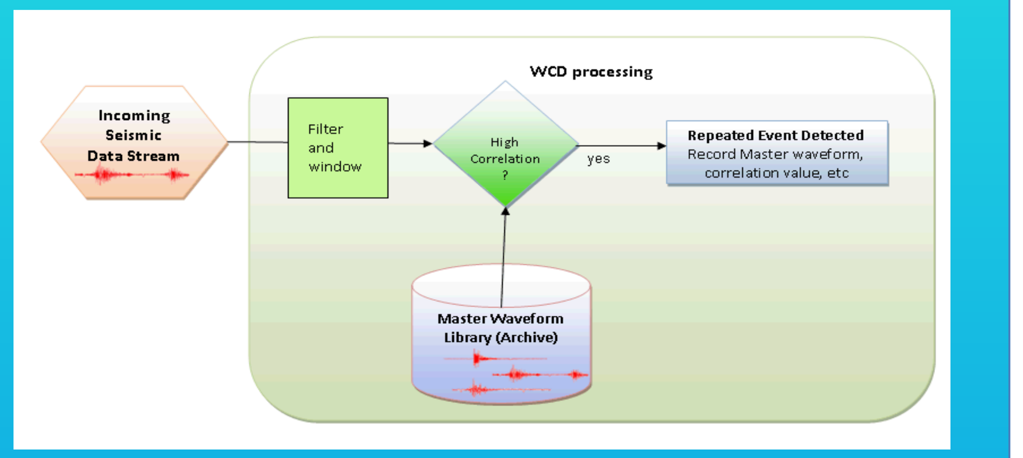
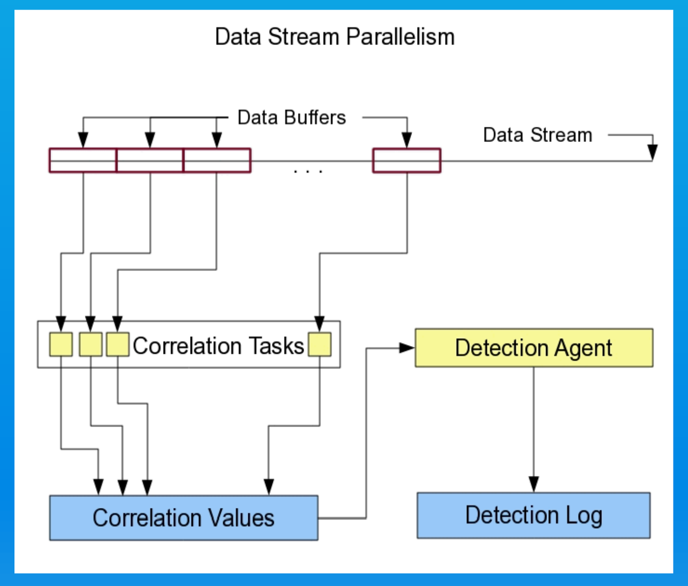


Figure2: Our WCD flow. The incoming raw data stream is filtered, windowed, and then correlated with each waveform in the Master Waveform Library. If a correlation is above the threshold, then we say a match is found, and record information such as the start time of the data segment, the correlation strength(s),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

DISTRIBUTED COMPUTING

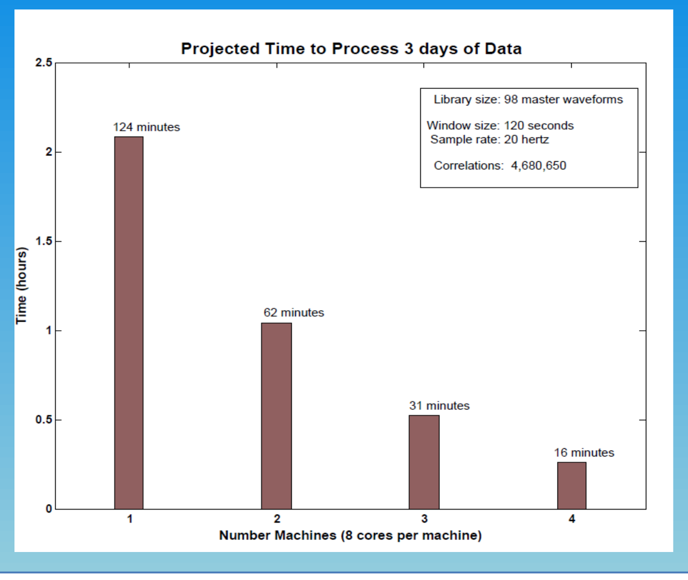
The biggest computational expense is calculating the correlation values of the data stream, but there is no need compute these correlation values sequentially. To take advantage of concurrency, we divide the data stream into a set of data buffers. These data buffers are correlated against the master waveform library concurrently and independently of each other via separate **Correlation Tasks**. The output of these **Correlation Tasks** is a series of correlation values. While the correlation tasks are running, the **Detection Agent** finds the maxima of the correlation data sequence. These maxima define our matches.

The **Detection Agent** processes the correlation values sequentially and thus represents the theoretical limit on how fast this system can run. However, finding the maxima in a data series is incredibly fast. It is doubtful the system will ever approach this theoretical limit as the correlation tasks dominate the computational time.



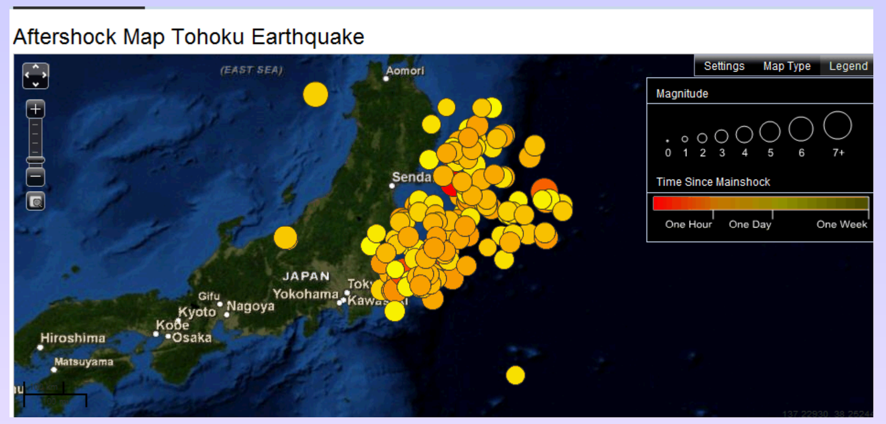
PERFORMANCE

Based on the average processing time of one machine with 8 cores, the chart to the right shows our projection of the amount of time required to process 3 days worth of data as we double the number of machines/cores. This chart represents an optimistic projection as it does not account for resource issues such as increased database accesses, network traffic, etc… that could arise when actually running multiple machines. However these are physical issues that can be solved with better hardware. The current architecture of the software will continue to scale well with increasing computational resources.



DATASET (Tohoku Aftershock Sequence)

- Mainshock occurred 5:46 UTC on March 11, 2011.
- Mw of 9.0 occurred at 38.322°N and 142.369°E
- Aftershocks used in our study were limited to a lat-lon box of 32-42°N and 139-146°E; the diameter of the cluster is approximately 745 km.
- The time period used in our study was March 11 6:00 UTC to March 13 02:00 2011; there were 1013 aftershocks in the IDC-REB catalog for this period.
- We retrieved data from array MJAR, 415 km away from the main shock.



AUTOMATED PARAMETER SELECTION

To move towards an operational system, the WC Detector must recognize that a swarm has started, and prepare itself for processing incoming data. Based on our previous work, we determined that the following parameters had significant influence on the accuracy of results: window length, filter band, correlation threshold. The most significant of these being the correlation threshold.

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. A window that is too short leads to false matches – S arrivals can correlate 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 and increases the probability of new arrivals corrupting the signal. We found that a window length that includes the P arrival and the beginning of the S arrival works well. This also helps improve accuracy, since event to station distance is reflected in the P-S separation.

Our approach: Given a station and an aftershock 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.

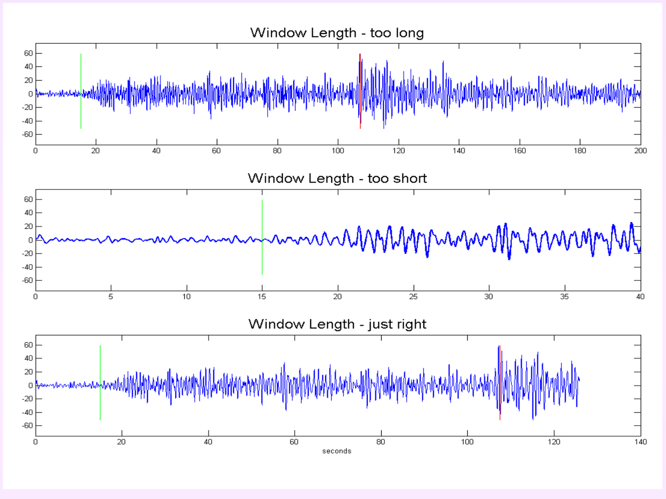


Figure 3: The same master waveform plotted at different window lengths. The green bar is the theoretical P arrival, the red the theoretical S arrival.

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 chose to implement the method described by Schaff (2008), which is similar to Gibbons and Ringdal (2006), because we wanted an objective method for automatically determining a suitable threshold for each master waveform. In Schaff's method, each array element is correlated against the template waveforms and the results are averaged. Then STA/LTA is used to identify when the correlation is at a local maxima, and declare a match. Using STA/LTA allows us to use a constant threshold regardless of the window length.

Our parameters: STA/LTA theshold = 5. STA length = 5 sec; LTA length = 20 sec.

** Gibbons and Ringdal (2006). The detection of low magnitude seismic events using array-based waveform correlation, Geophys. J. Int. 165, 149-166*
** Schaff (2008). Semiempirical Statistics of Correlation-Detector Performance, BSSA, 98, 1495-1507*

Performing waveform correlation on arrays allows for enhanced performance compared to using single element stations. Using multiple array elements ensures that directional information is factored into determining whether a match is declared. In addition, using multiple elements beats down the noise.

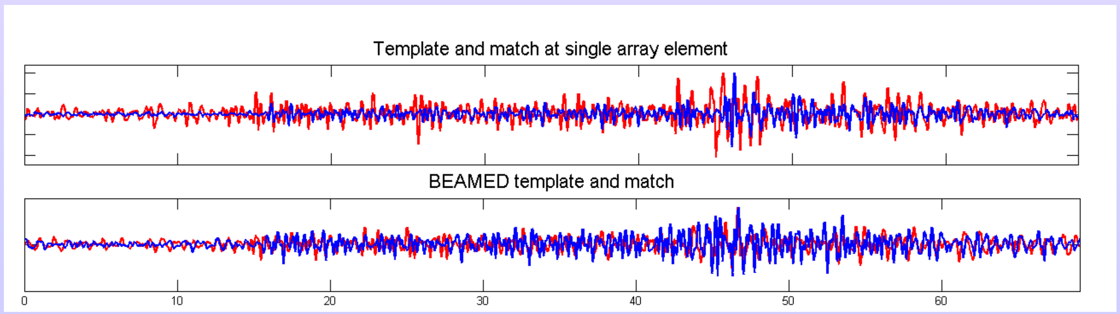


Figure 4: The advantages of using arrays vs single element stations is demonstrated by plotting template and match waveforms from just one element of the MJAR array, and from the array after beaming. The beamed signals correlate much more strongly, and are significantly less noisy.

RESULTS (PRELIMINARY)

Run	Station	Filter band	Array elements used	Window Length	Corr Threshold (STA/LTA)	# of catalog events seen at station	% catalog events belonging to a family	# of additional events identified
MJAR	MJAR	.8 - 5Hz	5	69 sec	5	1013	**85%**	**> 439**

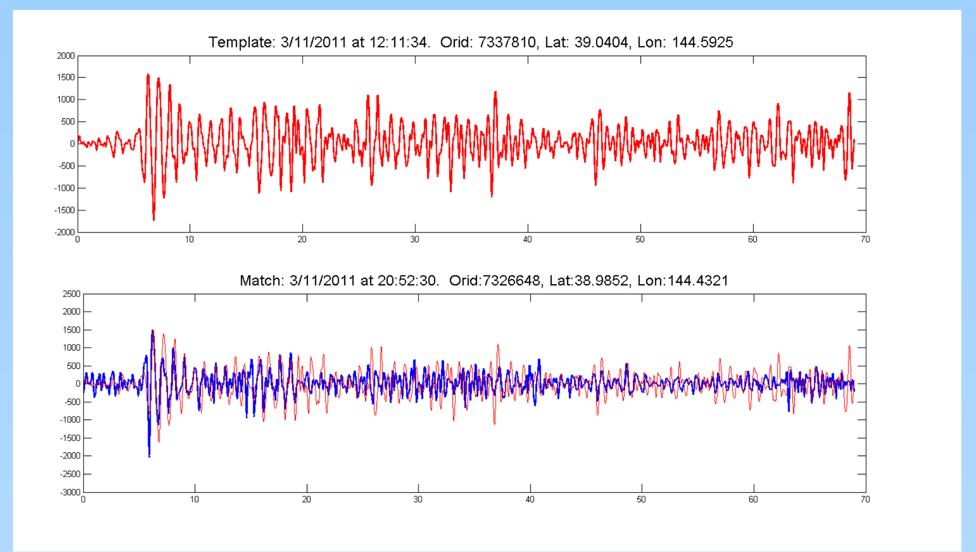


Figure 5: An event family, consisting of a template (red) and match (blue). In this case the match was an event in the LEB catalog, and we see that the template and match have nearly the same locations in the LEB catalog.

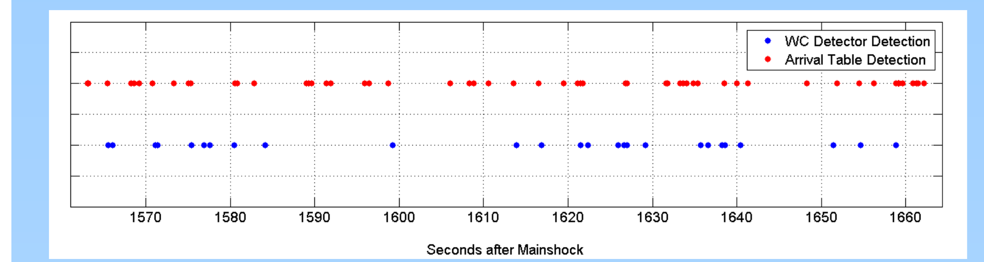
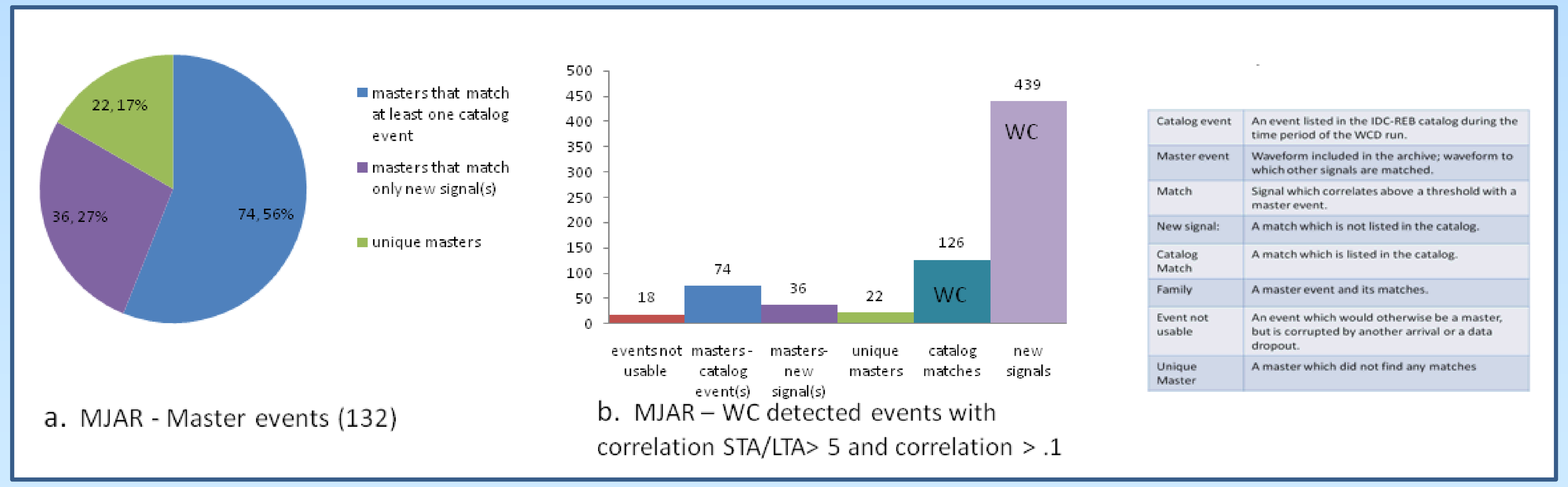


Figure 6: The WCDetector tended to find matches at times when the LEB arrival table had an arrival, as expected. Occasionally we detected "new signals" not in the arrival table. LEB arrivals *not* detected via WC were added to the library as new templates.



a. MJAR - Master events (132)

b. MJAR – WC detected events with correlation STA/LTA> 5 and correlation > .1

Catalog event	An event listed in the IDC-REB catalog during the time period of the WCD run.
Master event	Waveform included in the archive; waveform to which other signals are matched.
Match	Signal which correlates above a threshold with a master event.
New signal:	A match which is not listed in the catalog.
Catalog Match	A match which is listed in the catalog.
Family	A master event and its matches.
Event not usable	An event which would otherwise be a master, but is corrupted by another arrival or a data dropout.
Unique Master	A master which did not find any matches

FUTURE WORK

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

Automatic start-up: An operational system needs to automatically recognize that a swarm has started and set up the processing parameters.

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 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 finds only repeated events; it cannot replace traditional processing.

Sandia is a multi-program laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. **Unclassified Unlimited Release (UUR) SNL SAND # 2009-8141C**

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