SAND2011-9152 C

Processing the Tohoku Earthquake using Waveform Correlation

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

National Nuclear Security Administration

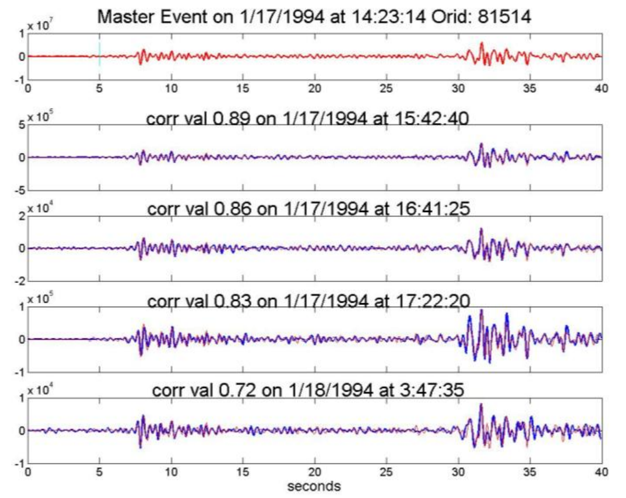
MOTIVATION AND OBJECTIVES

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, each event in these swarms/sequences must be treated as a possible nuclear test until it can be proven otherwise. 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) and applied it to several large aftershock sequences where we found that it detected 47% - 92% of the cataloged 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 log the match. 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.



Event on day 1

Four events in the next two days correlated well with the first event.

All were later localized to 34.3 lat, -118.4 lon

Figure 1: A typical family found by the WCD. The master waveform is shown in red, and the found matches are shown in blue. This figure shows data from the 1994 Northridge earthquake aftershock sequence.

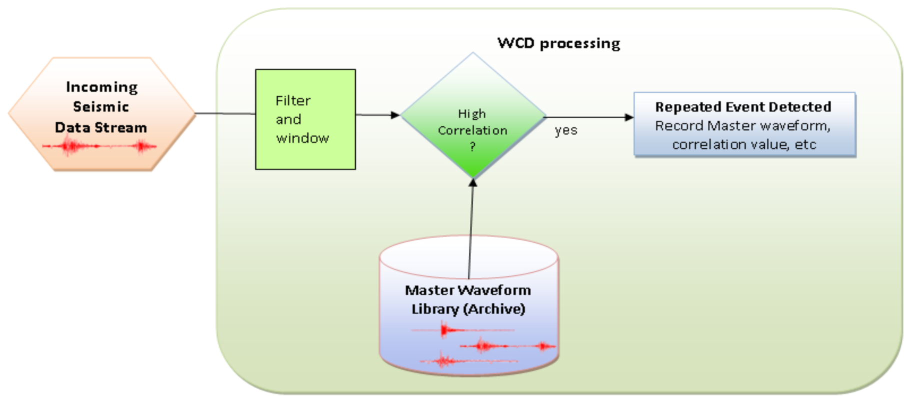


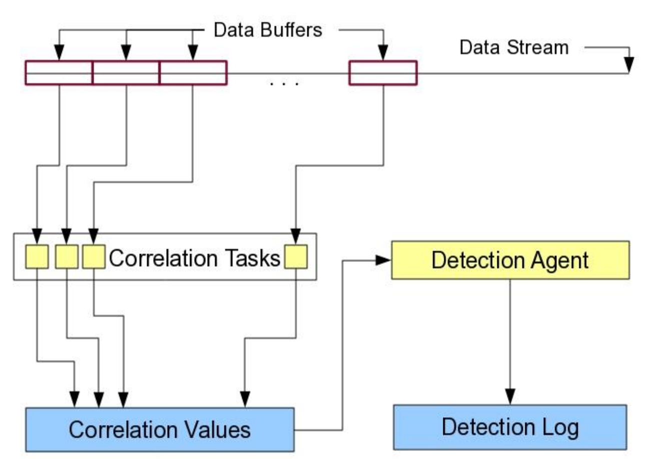
Figure2: Our WCD flow. The incoming raw data stream is filtered, windowed, and then correlated with each waveform in the Master Waveform Library. If a correlation is above the threshold, then we say a match is found, and record information such as the start time of the data segment, the correlation strength(s), 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 for the WC Detector 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 because the time required by the correlation tasks will dominate the overall processing time.

Data Stream Parallelism



PERFORMANCE

Our current processing time is:

9 sec per day, library event, Hz, core.

Processing Time	Library events	Days	Sampling Rate	Cores
23 seconds	1	1	20 Hz	8
5 minutes	1	14	20 Hz	8
108 minutes	20	14	20 Hz	8
9 hours	100	14	20 Hz	8

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

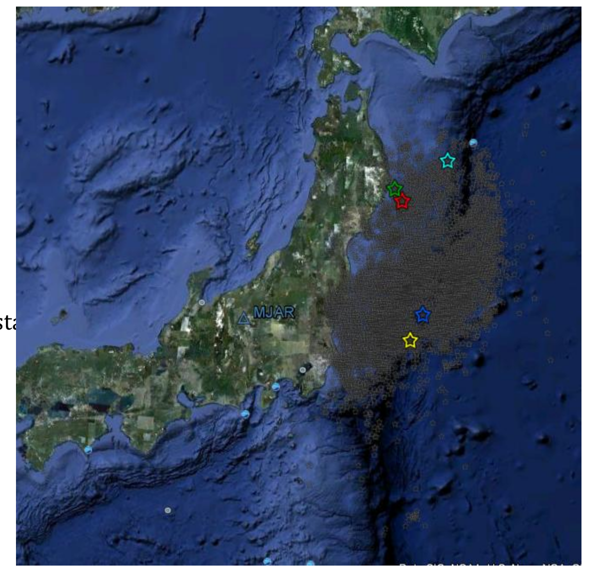
•Aftershocks used in our study were limited to a lat-lon box of 32-42°N and 139-146°E; the diameter of the cluster is approximately 745 km.

•The time period used in our study was March 11 6:00 UTC to March 30 02:00 2011; there were 1013 aftershocks in the IDC-REB catalog for this period.

•Templates were obtained started March 11. The entire datastream was processed from March 17 – 30.

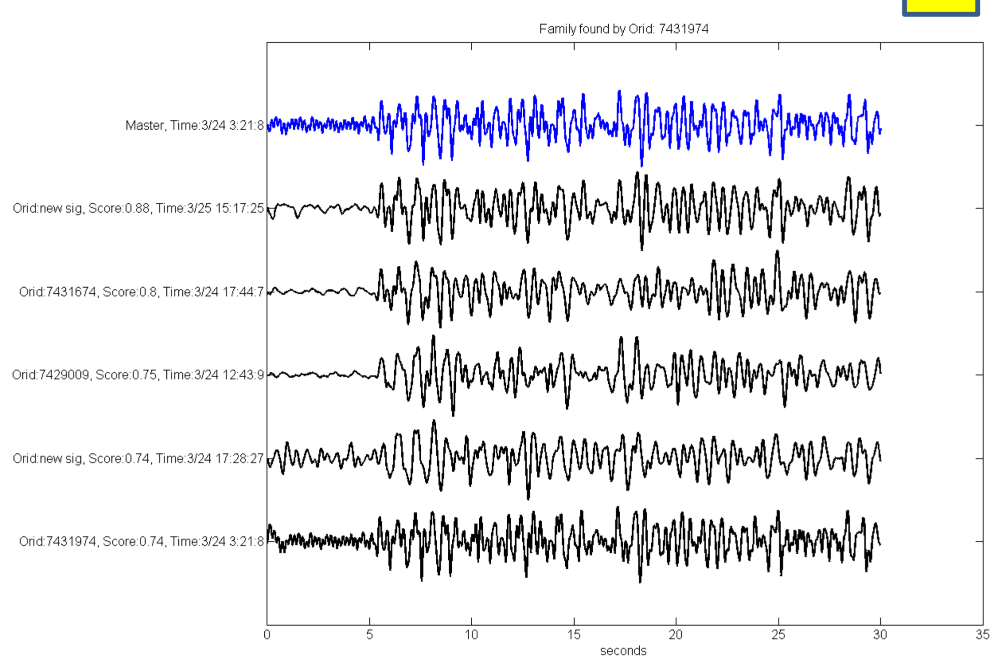
•We retrieved data from array MJAR, 415 km away from the main shock.

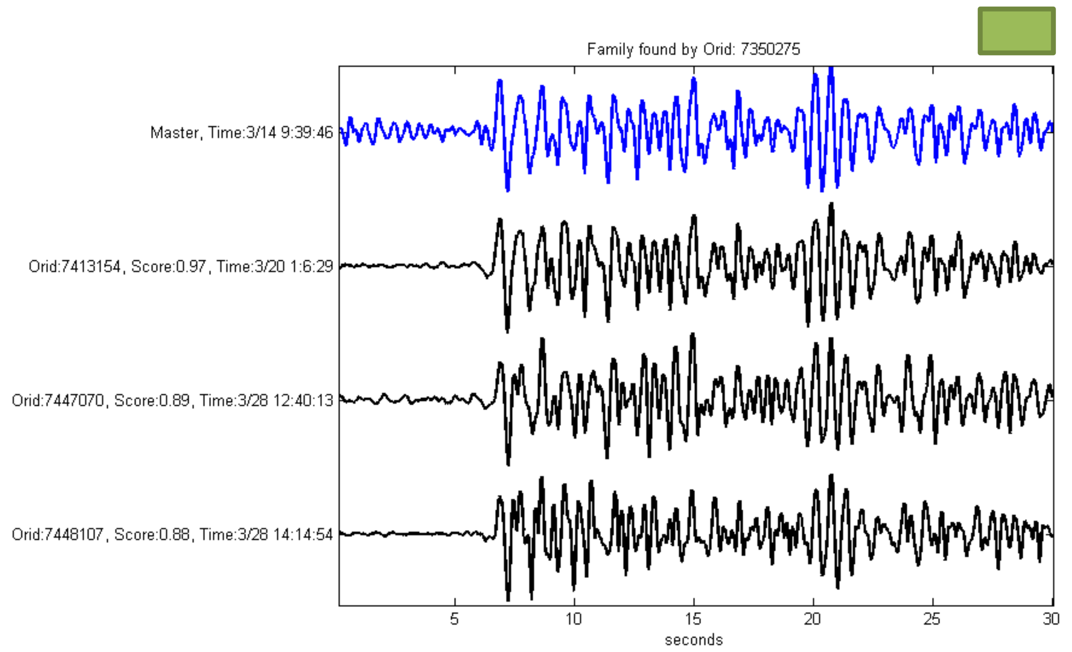
Figure 4. Swarm and st

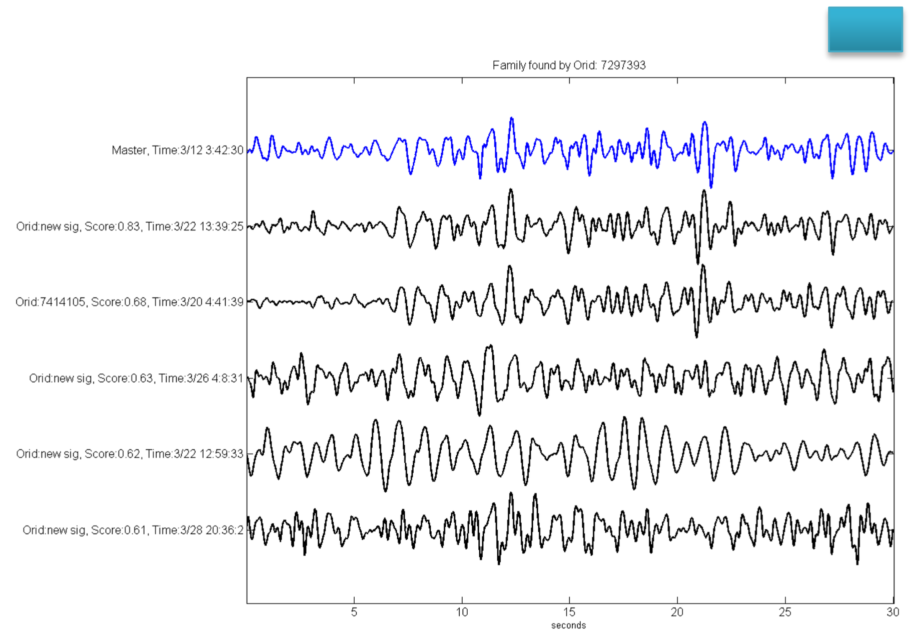


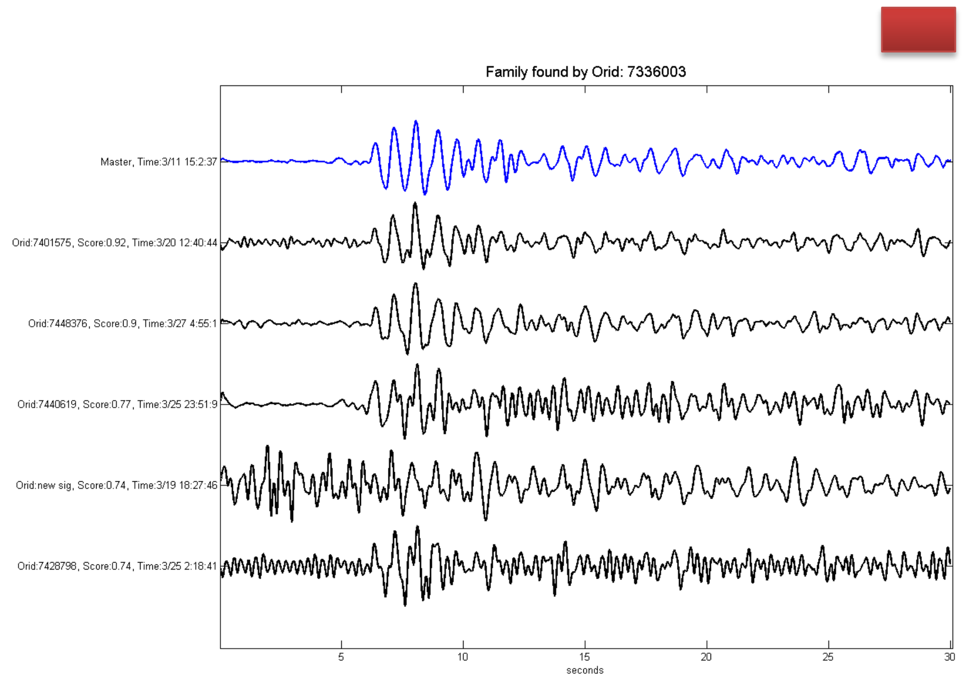
RESULTS

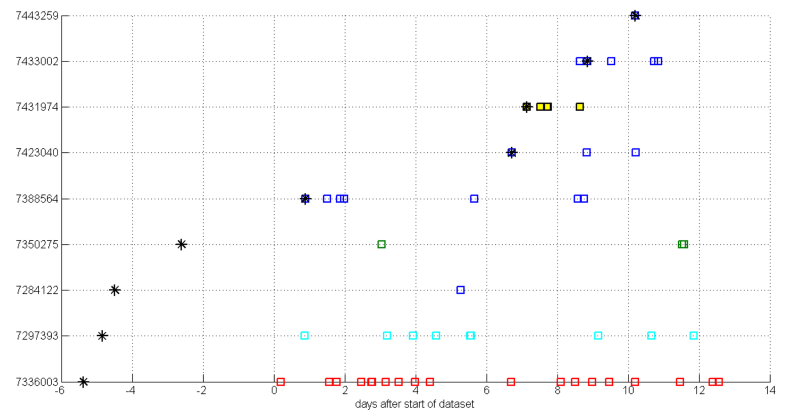
Typical families are plotted below. We found 51 events using 9 templates. Matches were often found days or weeks after the template event.



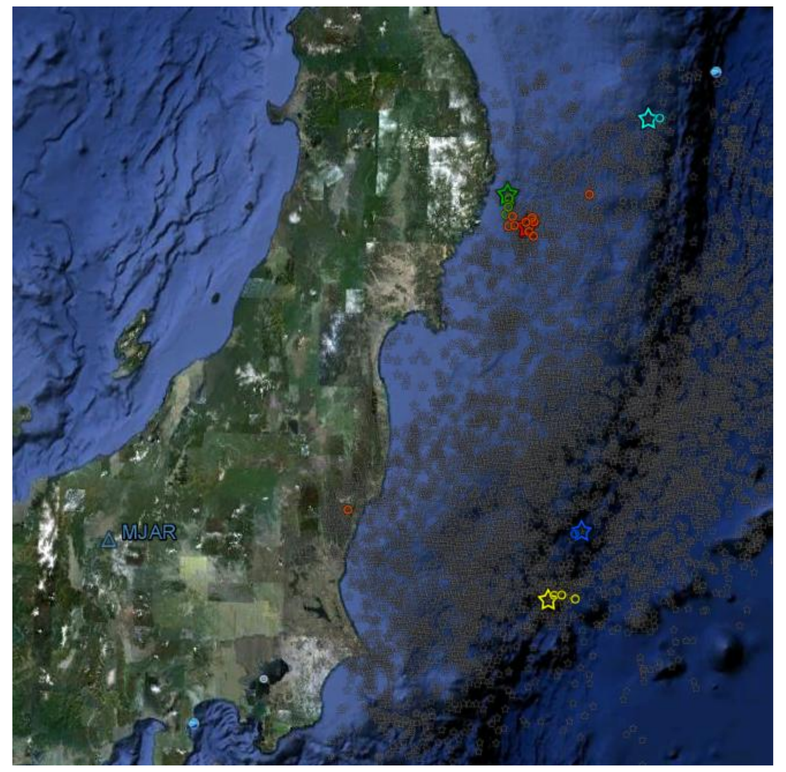








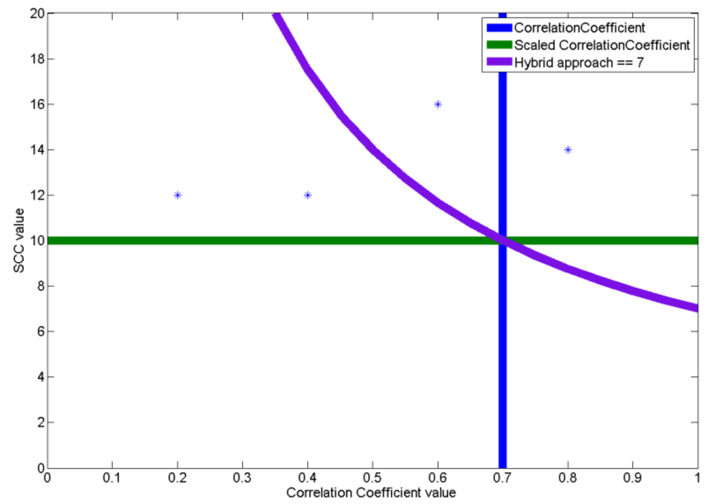
We can plot the templates and their matches in time, to see the time distribution of a family.



Families are plotted to show the spatial distribution of a family.

CORRELATION THRESHOLD METHOD

There are two methods currently commonly used for determining correlation: the correlation coefficient, and the scaled correlation coefficient (where STA/LTA is applied to correlation values). We introduce a third, HYBRID approach. This lowers the false alarm rate compared to the CSS method, yet allows more subtlety in recognizing a significant match than the correlation coefficient method.



ARRAY PROCESSING

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.

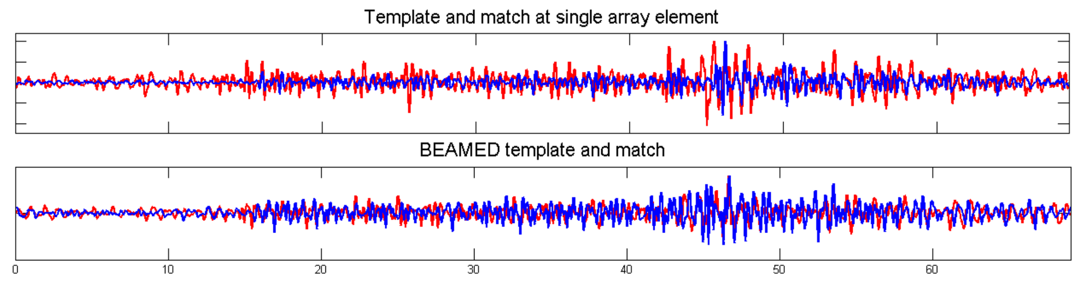


Figure 4: The advantages of using arrays vs single element stations is demonstrated by plotting template and match waveforms from just one element of the MJAR array, and from the array after beaming. The beamed signals correlate much more strongly, and are significantly less noisy.

FUTURE WORK

Parameter testing: We have developed a distributed system which allows us to very quickly process large datasets. Now we can turn our attention to evaluating the effect of various parameters on the results. In particular: optimal Threshold Method, and Window Length

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.

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 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 Development**