EXPLORING THE LIMITS OF WAVEFORM CORRELATION EVENT DETECTION AS APPLIED TO THE 1994 NORTHRIDGE EARTHQUAKE AFTERSHOCK SEQUENCE

Dorthe B. Carr, Megan E. Resor, Christopher J. Young, and Michael J. Procopio

Sandia National Laboratories

Sponsored by the National Nuclear Security Administration

Award No. DE-AC04-94AL85000/SL09-WFCorr-NDD02

ABSTRACT

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. Such occurrences can provide a large amount of useful information to seismologists. For those who monitor seismic events for possible nuclear explosions, however, these swarms/sequences are a nuisance. In an explosion monitoring system, each event must be treated as a possible nuclear test until it can be proven, to a high degree of confidence, not to be.

Correlation provides a measure of similarity between two waveforms, scaled from 0 (no similarity) to 1 (identical). Seismic events recorded by the same station which correlate well almost certainly have a similar location and source type, so clusters of events within a swarm can quickly be identified as earthquakes. We have developed a number of tools that can be used to exploit the high degree of waveform similarity expected to be associated with swarms/sequences: Dendro Tool (Merchant, 2007), the Waveform Correlation Detector (Resor et al., 2008) and the Self Scanner (Resor et al., 2009). Dendro Tool measures correlations between known events that have been detected by other means. The Waveform Correlation Detector creates its own event library of families that have correlated signals. The Self Scanner is used to establish the overall amount of correlation within a data steam and does not require an event library. All three techniques together provide an opportunity to study the similarities of events in an aftershock sequence in different ways.

We chose the well-studied aftershock sequence from the 1994 Northridge earthquake and used the three different methods to comprehensively characterize the waveform similarity available in a major aftershock sequence. Two different catalogs – the Earthquake Data Report (EDR) and the Southern California Earthquake Center (SCEC) catalog—provide ground truth for event locations and magnitudes. Station PAS from the Terrascope seismic network was the closest station to the Northridge data at a distance of approximately 33 kilometers from the main shock. We clustered events on PAS using all three methods—Dendro Tool, Waveform Correlation Detector and the Self Scanner. Most known events clustered in a similar manner, but with both the Waveform Correlation Detector and Self Scanner, we found significant numbers of new events that were not in either the EDR or SCEC catalog.

Because this sequence occurred in Southern California, there were many recordings of the aftershocks at a variety of azimuths and distances. Using Dendro Tool, we clustered known events occurring in the first 5 hours of the aftershock sequence on 13 stations from the Northern California Seismic Network (NCSN), along a transect of nearly constant azimuth from the Northridge events (338–347 degrees). Distances ranged from 238 to 960 kilometers away. Due to attenuation of the signal, the number of clusters with correlation values over 0.70 decreased from 7 clusters at the station 238 km away to only 1 cluster at the three stations over 850 kilometers away. In addition, the correlation values for the clusters decreased with distance as well. The Waveform Correlation Detector detected events that were not in either the EDR or SCEC bulletins and grouped them into families with known events. For non-catalog events, we estimate magnitudes based on scaling with events in the family. We found that we were getting good correlations for events where the amplitudes were 1000 times lower than the master event. This suggests that the Waveform Correlation Detector can be used to create a more complete catalog, and establish detailed clusters that can help better characterize the total seismicity in aftershock sequences.

OBJECTIVES

Previous work has established, to a high degree of certainty, that events with high correlation values are from the same location (Israelsson, 1990). The Northridge Earthquake of 1994 provides a wealth of data with which to study the effectiveness of correlation-based detection and classification schemes. We compared the results obtained from three correlation-based analysis and detection methods developed by Sandia National Laboratories. In addition, we studied how the output of correlation-based algorithms changed with regards to the distance between earthquake and station.

RESEARCH ACCOMPLISHED

Analysis was performed using three analysis tools developed and implemented by Sandia: Dendro Tool, Waveform Correlation Detector, and Self Scanner. All three tools implement correlation based algorithms to find similar events. They differ primarily in whether they are used on cataloged origins or raw data and in their post processing techniques. The three methods are summarized below (Table 1); detailed descriptions follow.

Table 1. Summary of the features of the three correlation tools used in the study.

	Dendro Tool	Waveform Correlation Detector	Self Scanner
Type of event/data	Cataloged origins	Compares cataloged origins with raw data	Raw data
Function	Measures similarity of cataloged origins	Finds all signals similar to the cataloged origins	Finds all similar signals in raw data
Clustering Actions	Several cluster options; shows dendrogram of clustering	No true clustering; creates families of events related to cataloged origins	Clusters results into families of similar signals
Output plots	Graphical interface shows connectedness of all origins	Plots show catalog origins and all matched signals found in raw data	Plots show all signals in an event family
Speed	Fastest – nearly instantaneous	Faster - ~1 week of data processed overnight	Fast - ~3 days of data processed overnight

Dendro Tool

The Dendro Tool (Merchant, 2007) allows analysts to quickly and easily determine the similarity between seismic origins by computing the cross-correlation values of the origin waveforms. Origins can then be categorized into clusters of similar events. This analysis technique can be used to characterize historical archives of seismic events in order to determine many of the unique sources that are present. In addition, the source of any new events can be quickly identified simply by comparing the new origin to the historical set.

The primary interface of the tool is the dendrogram window which plots the correlation relationship between all the origins. Selecting a correlation threshold value separates the waveforms into clusters. A detailed discussion of the clustering process can be found in Merchant (2007).

Waveform Correlation Detector

The Waveform Correlation Detector (Resor et al., 2008) was developed to simulate a real-time system where incoming raw data is compared to cataloged origins to try to screen out similar events. Catalog origins can include historical data and events from earlier in an aftershock/swarm sequence. The Waveform Correlation Detector is intended for use during an aftershock/swarm sequence to aid analysts by allowing them to prioritize signals with new locations and/or source types (possible explosions) over signals with a high degree of similarity to a cataloged origin known to be an earthquake, and therefore almost certainly an aftershock. Motivation for the tool was to provide the following benefits during a swarm/sequence scenario:

- Better use of analyst resources
- Faster processing
- Reduction of detection threshold
- More accurate phase picks.

A flow chart of the algorithm and how it compares with the traditional pipeline is shown in Figure 1.

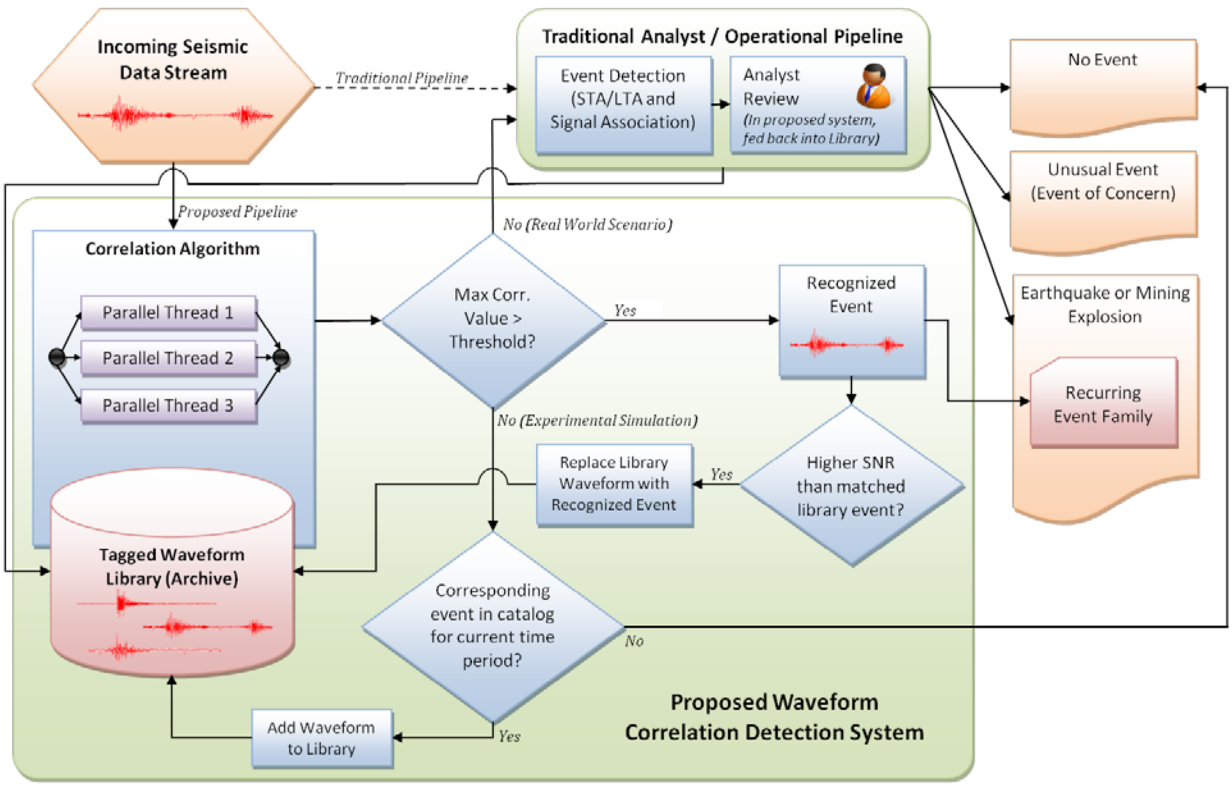


Figure 1. Waveform Correlation Detector algorithm flow chart.

In a traditional monitoring pipeline, an incoming data stream is passed through an event detector, followed by analyst review to categorize the event. The Waveform Correlation Detector is designed to be inserted before the traditional pipeline, reducing the analyst workload by screening out events similar to those which have been seen and previously categorized.

Our system compares the incoming data stream to previously identified origins held in a "library" or archive of tagged events. If the data stream and a particular library entry have a correlation value above a threshold, then we declare a recognized similar event. During a swarm, the library will grow as previously unseen waveforms, detected using traditional detection methods and verified by an analyst, are added to the library. In our experiments, the library is initially empty, though in practice the library could be pre-populated with historical data. The Waveform Correlation Detector outputs all the signals in the raw data found to correlate well with each cataloged origin (library master origin).

Self Scanner

The Self Scanner was developed to find all similar signals within a raw data stream (Resor et al., 2009). Each segment of data is correlated with every other segment. The correlation values are then run through a clustering algorithm which outputs families of similar waveforms. Figure 2 shows the self scanning correlation algorithm in pseudo-code. In this manner, every segment of raw data is correlated with every other segment.

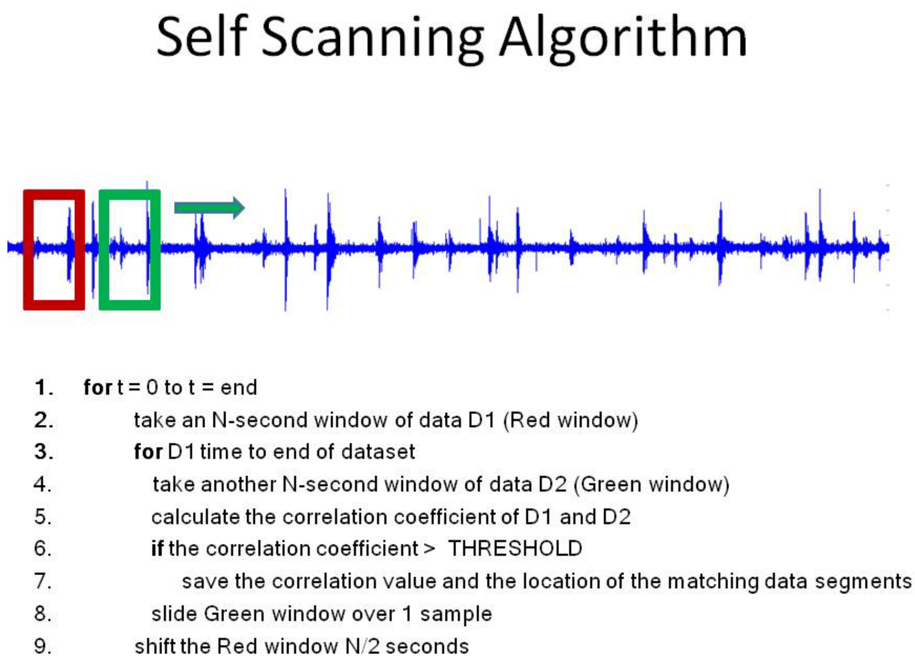


Figure 2.Self Scanning Algorithm.

The clustering routine is run after performing the self scanning algorithm. The algorithm links events with high correlation values to create families of similar signals.

Data

The Northridge Earthquake occurred at local time 4:30 am on January 17, 1994 and had a moment magnitude (Mw) of 6.7. This earthquake produced the strongest ground motions ever instrumentally recorded in an urban setting in North America at that time. The location of the earthquake was 34.213°N, 118.547°W and at a depth of 18.4 km according to the EDR catalog from the United States Geological Survey (USGS). There were numerous aftershocks from this major earthquake. In a period of 7 days, January 17-23, 1994, the EDR recorded 457 origins with m_b magnitudes between 2.9 and 6.4 in the lat-lon box of 33.9°-34.6°N and 118.8°-118.2°W (Figure 3). In that same time period and lat-lon box, the SCEC catalog recorded 2395 origins. Local magnitudes from the SCEC catalog ranged from 1.5 to 5.9.

We collected waveforms from the Incorporated Research Institutions for Seismology (IRIS) on station PAS for the time period January 17-23. Station PAS is only 35 km from the main shock, and has good signal to noise. Because the Northridge Earthquake occurred in Southern California, it provides a good opportunity to get waveform data from a large number of stations in the area to determine how correlation is affected by the distance from the station to the aftershock swarm. We collected waveforms from the Northern California Earthquake Data Center (NCEDC) and Northern California Seismic Network of the USGS for 13 stations that were at various distances from the main shock, but all along a transect of nearly constant azimuth of 338 to 347 degrees. Figure 3 shows a map of where the stations are located with relation to the Northridge Earthquake aftershocks.

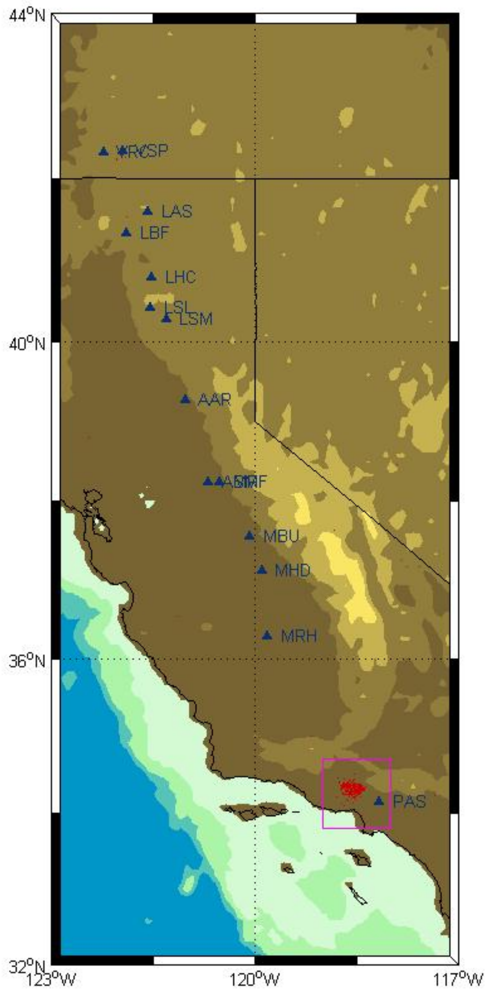


Figure 3. The Northridge Earthquake aftershock sequences (pink box) and stations used in the study.

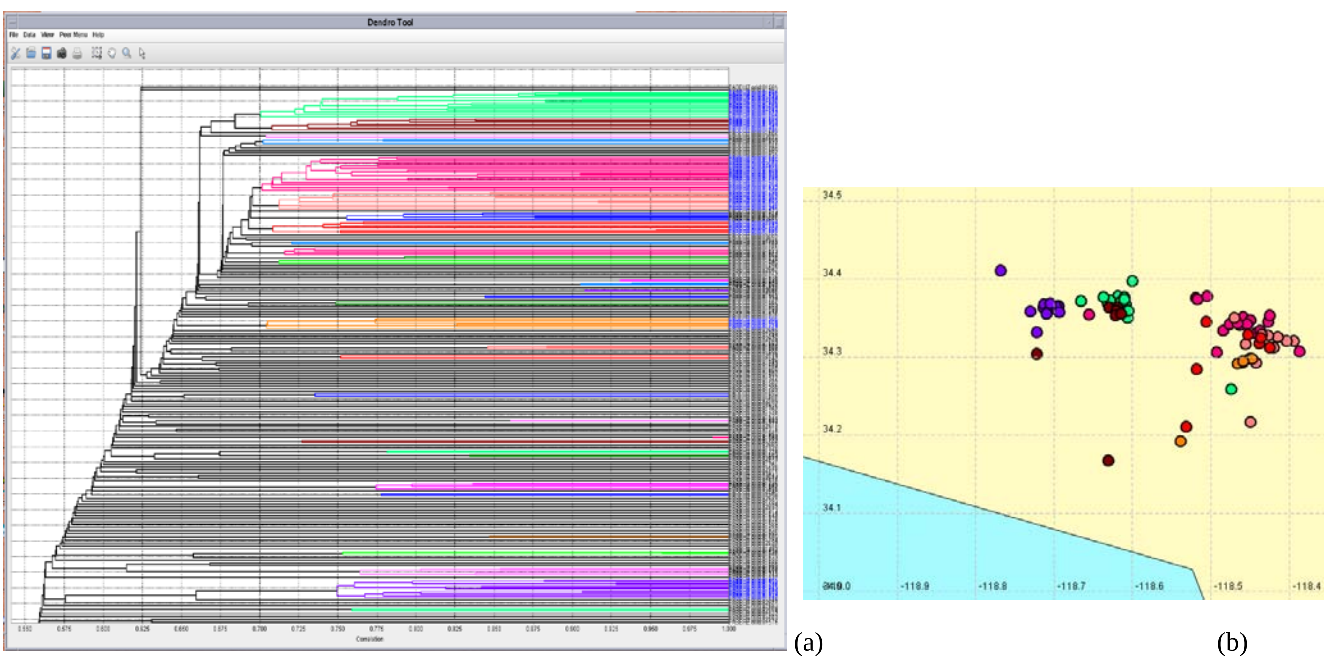
Comparison of Waveform Correlation Tools

We began with Dendro Tool to cluster the 412 origins found in the EDR catalog in the chosen time period and lat-lon box. Filter and window parameters are given in Table 2. One hundred eighty of the 412 origins (44%) were categorized into 50 clusters of similar signals using a correlation threshold of 0.7. The largest cluster had 19 origins and there were 33 clusters that had only 2 origins. Figure 4a shows part of the dendrogram that was created using Dendro Tool on this data and Figure 4b is a map showing the locations of the 7 clusters with more than 5 origins. The color of the origins on the map correlates to the color of the cluster on the dendrogram. The seven clusters separate fairly well, but there are some origins that may not have an accurate location.

Next we applied the Waveform Correlation Detector using origins from the EDR catalog as master origins and the parameters given in Table 2. There are 326 master origins and signals in the data stream are matched using a correlation value of 0.7. The signals that are matched are either known origins from the EDR catalog or unknown signals. The biggest event family had 16 signals that matched the master origin (17 signals total) and there were 183 master origins (56%) that matched no other signals in the data. Figure 5 shows events from one of the event families created by running the Waveform Correlation Detector on the Northridge data. The master origin for event family #20 is orid 81501. Signals that match this waveform are orids 81538, 81540 and 81544, plus a new signal that occurred at 1994017 15:46:23 that was not in the EDR origin table. In a pipeline scenario, an analyst would not have had to look at the signals that matched orid 81501. These signals would have been considered as matches to an

Table 2. Input parameters for the correlation tools.

	Dendro Tool	Waveform Correlation Detector	Self Scanner
Time	1994017 12:30 to 1994021 14:35	1994017 12:30 to 1994021 14:35	1994017 12:30 to 1994021 14:35
Latitude Span	33.9°N to 34.6°N	33.9°N to 34.6°N	Not needed
Longitude Span	118.8°W to 118.2°W	118.8°W to 118.2°W	Not needed
Catalog	Earthquake Data Report	Earthquake Data Report	Not needed
Station/Channel	PAS, BHZ	PAS, BHZ	PAS, BHZ
Window length	55 seconds, starting 5 seconds before P	40 seconds starting 5 seconds before P	36 seconds
Correlation Threshold Value	0.7	0.7	0.7
Filter	0.8-3.5 Butterworth, order 3, filter before windowing	0.8-3.5 Butterworth, order 6	0.8-3.5 Butterworth, order 3, forward and backwards



(a) (b)

Figure 4 (a)Dendrogram showing the some of the clusters formed on the Northridge data from station PAS and (b) Locations of the seven dendrogram clusters that had more than 5 origins. The color of the origins matches the colors of the clusters in 4(a).

already analyzed event. The signal found that didn't match a known origin indicates that using the Waveform Correlation Detector can find additional aftershock data without increasing analyst burden.

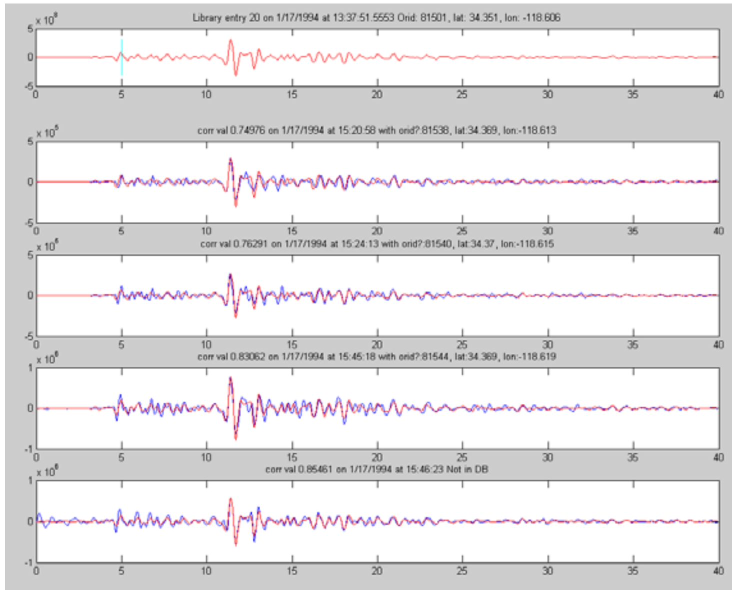


Figure 5. Waveform Correlation Detector output plot for event family #20 (5 of the 17 signals in library).

The Self Scanner did not use any origin information. Correlations were made with signals that were found as the self scanner went through the data and clustered signals based on correlation values over 0.7. Over 950 families were created with the Self Scanner. Most of the families had only 2 signals, but the top 10 families had between 63 and 202 signals. Figure 6 shows the first few signals in a family that was composed of 92 signals, designated Family 4. The signals from Family 4 are the same signals that were in the Waveform Correlation Detector event Family 20 (Figure 5). The signals at times 13:37:46, 15:20:48, 15:24:03 and 15:45:10 are the signals from orids 81501, 81538, 81540 and 81544 respectively. Additionally, the signal at time 15:46:17 matches up to the signal timed at 15:46:23

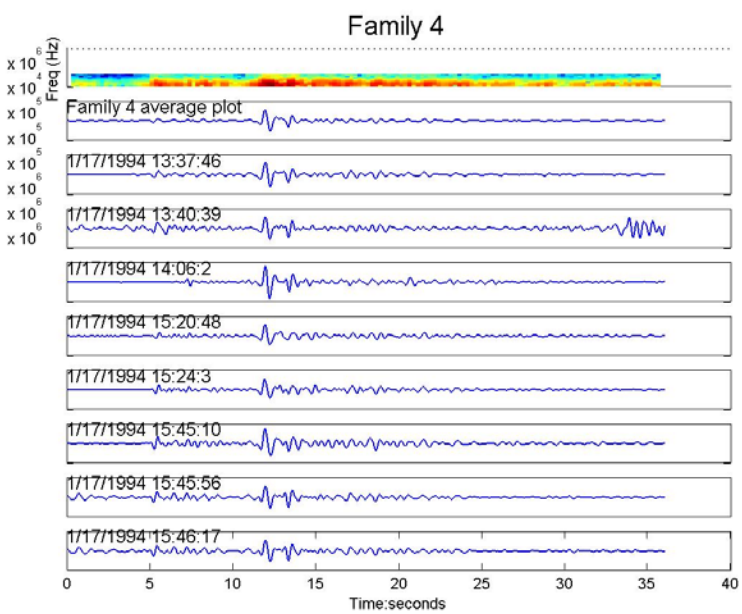


Figure 6. Self Scanner output plot for Family #4 (8 of the 92 signals).

in Figure 5. The Self Scanner found some additional signals as well (13:40:39 and 14:06:02). These signals are small and have low correlation values (0.71, 0.73) with the first signal at 13:37:46. The Self Scanner also extends the known families found with Dendro Tool, finding smaller events that make the family more complete.

To get the best idea of the differences seen with the three different correlation tools, we looked at the results seen in Figures 5 and 6 from the Waveform Correlation Detector and Self Scanner to see how those tools correlated the events in the Dendro cluster (cluster 1, green cluster on Figure 4). Cluster 1 consisted of 16 origins that spanned the time from 1994017 13:37 through 1994021 11:31. These events are listed in Table 3. Fifteen of the 16 origins had locations that fit into the lat-lon box of 34.35-34.40°N latitude 118.66-118.60°W latitude. One origin, orid 81543 was outside this box at a location of 34.26°N, 118.47°W. We looked at the output from the Waveform Correlation Detector and Self Scanner to see how the origins in Dendro Cluster #1 were matched/clustered with these tools (Table 3). The origins fell into four different event families with the Waveform Correlation Detector, Families 20, 91, 94, and 279. Eleven of the 16 origins in the Dendro Cluster were found in the Self Scanner Family 4. The other origins did not have correlation values with the other signals in Family 4 that were greater than 0.7. Two of the origins, orid 81598 and 81601, correlated well with each other (correlation value = 0.83), but not with any other of the 16 origins in Dendro Cluster 1. The Waveform Correlation Detector and Self Scanner subdivided the Dendro Cluster in an identical manner. The fact that Dendro put all the events in one cluster is due to slight differences in the algorithms handling of lower correlated waveforms.

To compare the correlation algorithms in the three different tools, we compared the correlation value between the same origins with each tool where it was possible. The correlation value between origins 81501 and 81548 can be found looking at all three methods. The value is 0.891 for Dendro Tool, 0.893 for the Waveform Correlation Detector, and 0.89 for the Self Scanner. Additionally, the origins 81598 and 81601 both have correlation values of

Table 3. Results from the three tools using Dendro Cluster #1.

Cluster origins	Lat (°N)	Long (°W)	Waveform Correlation Detector Family	Self Scanner Family
81501	34.351	118.606	20 – master orid	4
81538	34.369	118.613	20	4
81540	34.37	118.615	20	4
81543	34.259	118.474	-	-
81544	34.369	118.619	20	4
81548	34.374	118.622	20	4
81554	34.398	118.6	20	4
81598	34.373	118.629	91 – master orid	103
81601	34.37	118.632	94 – master orid	103
81812	34.36	118.605	20	4
81930	34.378	118.618	20	4
81944	34.377	118.61	20	4
81950	34.372	118.666	279 – master orid	-
81965	34.374	118.61	-	4
82102	34.377	118.637	-	4
82106	34.367	118.608	-	-

0.83 in Dendro Tool and the Self Scanner; although they were found not to match with the Waveform Correlation Detector since they are in different families.

In addition to the known origins, the Waveform Correlation Detector and Self Scanner also matched signals that were not known origins in the EDR catalog. There were 13 new signals that matched the origins in the 4 Waveform Correlation Detector families. Six of those signals matched up to origins in the SCEC catalog. The locations of the SCEC origins all fit within the lat-lon box of 34.35-34.40°N and 118.66-118.60°W, except for SCEC orid 3143149 which had a location or 34.304° N, 118.447° W. Seven of the new signals that were in the Waveform Correlation Detector families were also found with the Self Scanner and clustered into Family #4.

Waveform Correlation and Magnitude

Changes in the size of events imply changes in the source mechanism (e.g., corner frequency) which might affect correlation between nearby events. To investigate this, we examined the range in magnitude within the clusters/families of origins found by our processing. Unfortunately, the EDR does not always provide magnitude estimates for their origins. The SCEC catalog provides magnitudes, but they are one of three types – local, coda or hand determined, so it is difficult to do direct comparisons. For the known SCEC origins that were grouped into the four Waveform Correlation Detection families 20, 91, 94, and 279, the origins with local magnitudes range from 2.43 to 5.07, a difference of 2.7 magnitude units.

Because some of the signals that were matched with the Waveform Correlation Detector were not origins in the SCEC catalog, we calculated our own magnitude range within families based on log of amplitude of the Lg phase and found ranges of up to 3.5. We believe that the Waveform Correlation Detector can find and match signals that are more than 2 magnitude units apart in size.

Waveform Correlation and Distance

Both attenuation of the signal and the signal-to-noise ratio for a signal can have impacts on the correlation values. To study the relationship between correlation/clustering and distance, we correlated 5 hours of data, starting at 1994017 12:00, on 13 different stations that were between 238 to 961 km from the Northridge swarm (Figure 3, Table 4). We used Dendro Tool to do the correlation, since we chose to base our analysis on catalog origins only. For each station we counted the number of clusters and determined how many orids were in each cluster. Table 4 shows how many clusters were created with Dendro Tool at each station, and the number of origins in the cluster with the most events.

Table 4. Station information for stations used in the distance study.

Station	MRH	MHD	MBU	MRF	ASM	AAR	LSM	LSL	LHC	LBF	LAS	VSP	VRC
Distance (km)	238	348	399	484	549	607	720	743	780	849	866	954	961
Azimuth (°)	347	339	340	338	340	339	341	340	341	340	343	342	342
# of clusters	8	5	5	2	4	3	2	2	2	3	2	2	2
Max # of origins	4	4	3	2	3	2	2	2	2	2	2	2	2

As expected, the number of clusters formed at a station decreased with distance. Origins that are seen well and without filtering at stations within 400 km of the aftershock sequence cannot be seen, even with filtering, at the stations over 800 km away from the aftershock sequence. One cluster was seen at all the stations except for the station closest to the source, MRH. The two origins in this cluster are 81481 and 81488 which have local magnitudes of 4.89 and 4.62 respectively, two of the largest events in the swarm. According to the SCEC catalog, there are 3 other events that occur within two minutes of origin 81481 and there is an overlap of all the different signals for these 4 events in the time window being used to do the correlation of orid 81481 on station MRH. Therefore, this signal does not correlate well to orid 81488. Since the other stations are farther away, the overlap of signals does not impact the correlation as much, although the correlation values for these two orids at stations MHD and MRF are lower than the correlation values at the stations between 450 and 900 km.

Our results suggest that attenuation of the signals does play a role in how well a station can correlate signals seen from an aftershock sequence. In order to match and correlate signals from smaller events, the station used to do the correlation must be within 200–400 km from the signals.

CONCLUSIONS AND RECOMMENDATIONS

The three different methods we have developed for doing waveform correlation work similarly and provide good information about how similar events from an aftershock sequence or swarm are related. The Dendro Tool can be used for historical archive data where good origin information exists. The Waveform Correlation Detector can be used in a real-time system, detecting and matching signals to known events in the library. The Self Scanner works on "raw" data where there is either sketchy or no origin information. A combination of all three tools can be valuable when doing in-depth studies of aftershocks or swarms. Smaller signals that may or may not have origin information can be found and matched/correlated to known origins, even those that are several magnitude units greater.

REFERENCES

Israelsson, H. (1990). Correlation of waveforms from closely-spaced regional events, *Bull. Seism. Soc. Am*. 80: pp 2177–2193

Merchant, B. J. (2007). The GNEMRE Dendro Tool, SAND Report #2007-6439. Sandia National Laboratories, Albuquerque, NM, October 2007 (Unclassified).

Resor, M. E, D. B. Carr, M. J. Procopio, C. J. Young, and C. Rowe (2009). Self Scanning Event Detection in Volcanic Swarms [abstract], *Seismological Research Letters*, 80: 340.

Resor, M. E., M. J. Procopio, C. J. Young and D. B. Carr (2008). Processing Aftershock Sequences Using Waveform Correlation [abstract], *AGU Fall Meeting Abstracts*, 1802.