ADAPTING PIPELINE ARCHITECTURES TO TRACK DEVELOPING AFTERSHOCK SEQUENCES AND RECURRENT EXPLOSIONS

Tormod Kværna[1], Steven J. Gibbons[1], David B. Harris[2], and Douglas A. Dodge[3]

NORSAR[1], Deschutes Signal Processing, LLC[2], and Lawrence Livermore National Laboratory[3]

Sponsored by the Air Force Research Laboratory and the National Nuclear Security Administration
Award Nos. FA9453-11-C-0230[1,2] and LL11-BAA11-26-NDD03[3]
Proposal No. BAA11-26

ABSTRACT

Pattern detectors (e.g., correlation, subspace, and matched field detectors) fuse the signal detection and source identification processes into a single operation. The organization of repeating waveforms for efficient analyst interpretation may result in significant gains in productivity when analyzing extensive aftershock sequences and explosions from repeating sources. Under current practice, pattern detectors run entirely independently of the pipeline signal detectors and the preparation and supervision of pattern detectors is relatively labor-intensive. It is the aim of this two-year study to investigate algorithms for adapting processing pipelines to create and supervise pattern detectors semi-automatically for incoming multi-channel data streams.

A functional model of an operational detection pipeline will be constructed with extensions that create and manage pattern detectors under a variety of spawning policies. The system will be tested on the October 2005 Kashmir earthquake and aftershock sequence using a network of observing seismic arrays and 3-component stations. This example is representative of challenging aftershock sequences given its vast number of events, relatively large source region, and absence of high-quality observations at close epicentral distances. Seismic arrays at both regional and teleseismic distances record thousands of detections of similar signals, of which only limited numbers match sufficiently well that they can be attributed automatically to given families of events. We will investigate whether it is most effective to limit pattern detectors to single arrays, or extend to coherent operations over a larger network. Studies will address whether templates can and should be updated as observations accumulate. It is essential that events of monitoring interest are not screened out by this autonomous system and this will be tested by superimposing signals of monitoring significance into the data stream for the Kashmir aftershock sequence.

A serious complication in the evaluation of the performance for the Kashmir sequence is the poor quality of the available bulletins for these events. The relative location of events is much poorer than it need be due to the fact that many high quality arrival time picks from stations at regional distances (in particular NIL at ~1 degree and the stations of KNET at ~8 degrees) are not used. Event location using all the available regional picks needs to correct for the anomalously slow velocities for many of the paths concerned. Station calibrations exist for some, but not all, of the observed phases and so an iterative re-location procedure was carried out whereby, at each iteration, a time-correction term for a given phase from a specified source region was estimated from the mean travel-time residual from the previous iteration. The resulting event location estimates cluster far better than in any of the available bulletins indicating improved relative event locations, from which the significance and consistency of groupings of similar waveforms can be more easily assessed.

The correlation-coefficient traces for templates of events in the Kashmir sequence display a high variability with a modulation which can be ascribed to the contrast in amplitudes of the incoming data. Many significant local maxima can be demonstrated to result from the presence of impulsive high amplitude signals in the data stream. Analysis of the alignment of single-channel detection statistic traces reveals many such maxima to be false alarms, although their presence makes the setting of robust detection thresholds difficult. The ratio-to-moving-average (RMA) method is frequently used in diverse fields for the seasonal adjustment of time-series. Replacing seismograms with the corresponding RMA traces prior to the correlation procedure results in detection statistic traces with a far lower variability. This reduces the number of preliminary detections and facilitates the setting of lower, more aggressive, detection thresholds.

OBJECTIVES

This two-year study will investigate the adaptation of processing pipelines to create pattern detectors (i.e., correlation, subspace and matched field detectors) that discover and organize repeating waveforms in data streams from a network of seismic arrays. The monitoring applications of this technology will include real-time responses to major developing aftershock sequences to ameliorate analyst overload, and autonomous discovery of repetitive explosions.

A functioning model of the detection stage of a pipeline implementing conventional beam recipes will be constructed, but extended to create and manage pattern detectors under a variety of spawning policies. This system will be used to test a number of strategies for discovering repeating waveform patterns and organizing detected occurrences for efficient interpretation by analysts. The system will be tested using the four regional Kazakhstan arrays as a network observing the 2005 Kashmir earthquake aftershock sequence.

It will be investigated as to whether pattern detector waveform templates should be limited to individual arrays or extended to coherent operation across the network, and whether templates can be improved as observations accumulate. An autonomous supervisory function will be introduced that keeps track of detector performance, and culls, updates or merges detectors to improve overall system performance. This includes periodic reprocessing of the data stream with the suite of maturing pattern detectors, to be conducted as a parallel operation so as not to slow the main detection process.

Alternative pattern detector spawning policies will be examined, one with new detectors created only from special analyst-designated primary detectors and another with spawning from all of the conventional STA/LTA or F detectors implemented on recipe beams. System performance will be graded, with the ultimate metric being a measure of the consolidation of detections into efficiently-interpreted families. This includes checks to ensure that this autonomous system does not screen events of interest from analyst evaluation, by superimposing waveforms from other events among the Kashmir aftershocks.

RESEARCH ACCOMPLISHED

Introduction

Despite years of research and development on seismic network operations and decades of practical experience with processing pipelines, aftershock sequences associated with large earthquakes (magnitude 6 and above) still overwhelm analytical resources. Very large events such as the 2004 Sumatra-Andaman Islands earthquake, the 2008 Sichuan earthquake, and the 2011 Japan earthquake flood seismic networks with thousands of aftershocks in the weeks that follow the main shock. Aftershocks may be distributed over large geographic regions. One strategy for coping with such challenges exploits the fact that many of the aftershocks produce highly similar waveforms. It may be possible to group events on the basis of waveform similarity and reduce the load of analytical work by limiting detailed interpretation to a smaller number of events that represent the range of waveforms produced by a major sequence or swarm. To be useful, a system built to realize this strategy must discover and associate repeating waveform occurrences in real time as the sequence unfolds.

Correlation, subspace and matched field detectors (here collectively referred to as pattern detectors) provide efficient means to detect and associate occurrences of specific repeating waveforms, but are labor-intensive to develop and deploy in most implementations. Recently, Harris and Dodge (2011) developed a prototype system to automate creation of correlation and subspace detectors that organize waveforms from a developing aftershock sequence. This system was tested on a data stream from a single array (NVAR) observing, at a range of 390 kilometers, the aftershock sequence of a moderately large earthquake (2003 San Simeon, mb 6.5). We found that this system had the potential to reduce analyst workload anywhere from 40% to 70% while simultaneously increasing the completeness of a catalog of events (even above magnitude 3) formed from this one array. The system was a relatively simple prototype operating with a single beam and STA/LTA detector directed at P phases emanating from the San Simeon region. This fixed detector produced waveform templates used to "spawn" correlation detectors in real time. These correlation detectors in turn detected and associated later occurrences of events with similar waveforms. They produced often large groups of related events with the potential to be interpreted efficiently by an analyst performing post-detection review.

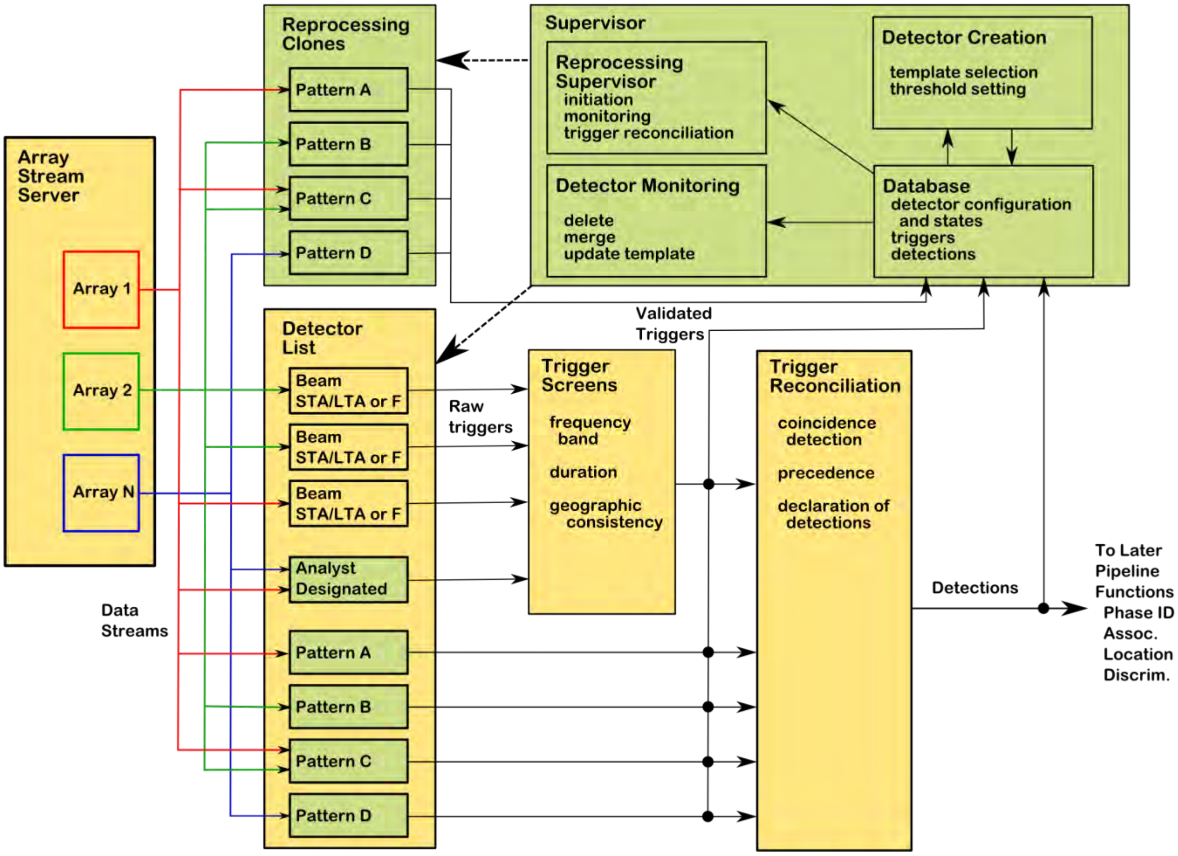

Figure 1. Block diagram of the detection processor proposed to test pattern detector creation and management strategies. The boxes in yellow indicate functions approximately shared with a conventional pipeline detection processor. Those in green represent entirely new functions. Note that some detectors may process data from more than one array, possibly coherently.

Figure 1 illustrates a proposed extension of the framework of Harris and Dodge (2011), progressing to a stream of multiple arrays. The purpose of this project is to examine the degree to which such a system can be scaled to a network of arrays, and investigate how data from the different arrays contribute to the system performance. The intention is to inform possible designs of new pipeline architectures that may create pattern detectors while operating conventional beam recipes, or to inform decisions to retrofit existing pipelines for the same purpose.

Finally, there is another motivation for building a system to discover repeating sources: detection of groups of explosions. Among other purposes, these find use in calibrating and testing discriminants. In some regions of the world, mining explosions dominate seismic detections. In such cases, automatically-created pattern detectors may themselves perform the discrimination function.

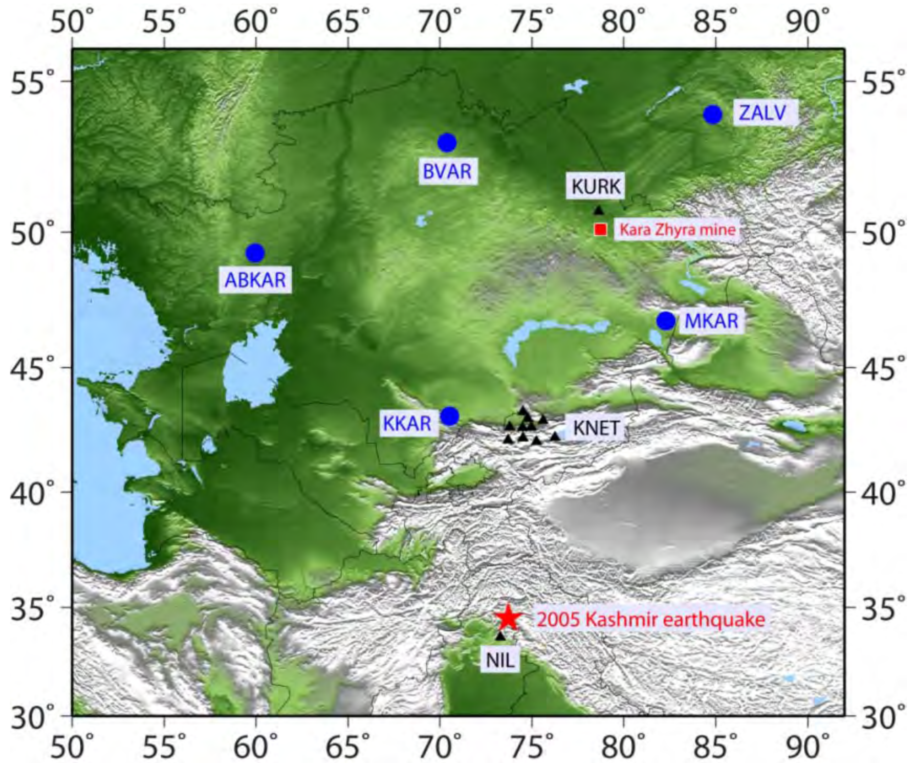

Figure 2. Location of 9-element arrays (blue circles) and 3-component stations (black triangles) relative to the site of the October 8, 2005 Kashmir earthquake. (Note that at the time of the earthquake, only a single 3-component station, ZAL, was in place at Zalesovo.)

The October 8, 2005, Kashmir Earthquake and Aftershock Sequence

We plan to evaluate the system on the aftershock sequence to the magnitude 7.6 Kashmir earthquake of October 8, 2005. This event was followed by hundreds of aftershocks in a relatively short time interval and represents a good example of a sequence which has the capability to overwhelm analyst capacity. We plan to run tests of the detector pipeline on a modest network of arrays (see Figure 2) with detector spawning likely to be restricted to regional waveform observations.

Figure 3 displays different processing attributes for a 30-minute data window from two of the arrays displayed in Figure 2: KKAR (990 km from the ISC location estimate for the main event) and MKAR (at 1540 km). Under the waveforms in Figure 3 are traces displaying the relative power for continuously evaluated f-k estimates, with a red color applied at the times of estimates that are consistent with initial P-arrivals from the appropriate directions. There is clearly a significant difference between the two arrays in the detection capability for events from this region. First P-arrivals at KKAR are frequently impulsive, providing a high signal-to-noise ratio (SNR). Emergent P-arrivals at MKAR, coupled with a long wavetrain, make this station a relatively poor array with which to observe the sequence, despite it being the closest International Monitoring System (IMS) array operational at the time. (The SNR for initial P-phases is far higher for many other IMS arrays at teleseismic distances.)

The difference in the performance between the two arrays raises an additional question. When is it optimal to combine waveforms from multiple arrays in a coherent detection process, and when is it preferable to consider arrays individually? In contrast to most of the classical array-processing methods, many of the pattern detectors can be applied over essentially arbitrary apertures. If an event is co-located exactly with a master event, then the only limitation to its detectability using a correlation or matched field detector is the signal-to-noise ratio. A finite hypocentral distance will be far more significant when observing over a larger aperture network than on a small aperture array, since the relative differences in travel-times to the different sensors will be larger. This may or may not be able to be compensated for by applying time-shifts; the issue becomes one of waveform similarity.

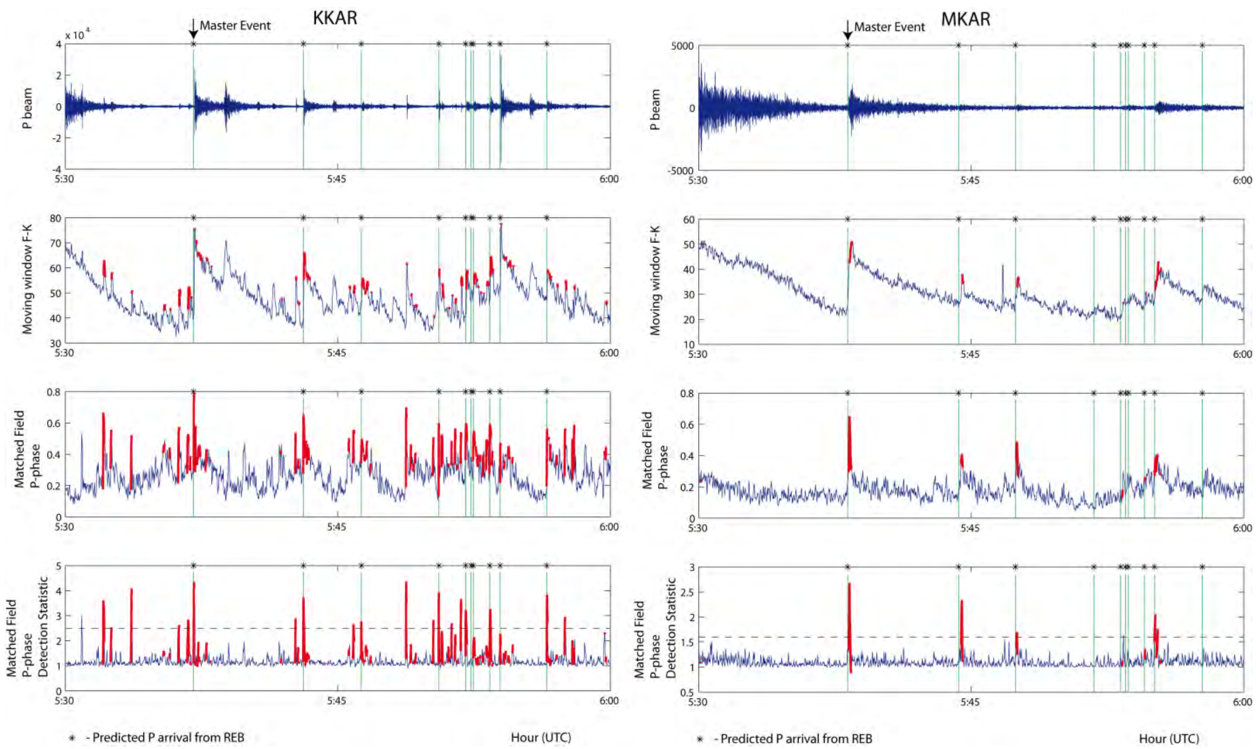

Figure 3. Performance of continuous classical f-k analysis and empirical matched field detection statistics for a 30-minute data segment from two arrays at regional distances from the October 8, 2005 Kashmir earthquake (KKAR at ~9 degrees and MKAR at ~14 degrees). The vertical green bars indicate predicted P-wave arrivals from events in the Reviewed Event Bulletin (REB) of the International Data Center (IDC) in Vienna. Red bars indicate times at which the slowness vector generated by classical f-k analysis was consistent with a P-phase from the appropriate direction.

It is clear from Figure 3 that P-phases at KKAR arrive in far quicker succession than the length of the wavetrain for each event. This presents a challenge for correlation detectors given that no single template should include the signals from more than one event, and that detectability will decrease due to the overlapping waveforms. This suggests a possible advantage of the single-phase matched field detector (Harris and Kværna, 2010) which forms a narrow frequency band representation of a short data segment. Whereas the correlator compares the temporal structure of the wavefield recorded at each sensor, the matched field detector compares the spatial structure of the wavefield over multiple sensors. In this respect, it is analogous to the classical f-k analysis except for the fact that the templates (or steering vectors) are empirical (based upon the observation of previous signals) rather than theoretical (based upon a plane-wave propagation model) and so encode implicitly the scattering and diffraction effects which can make classical f-k analysis perform so poorly, particularly at high frequencies.

In Figure 3, a master event is chosen and the spectral covariance matrices are estimated for the initial P-arrivals at the two arrays. For each covariance matrix, the principal eigenvector constitutes an empirical steering vector for that observation and the third trace down in each plot displays the matched field statistic for this steering vector evaluated continuously for the incoming data stream. There is a far greater contrast between peak values and background level for the matched field statistic than for the moving window f-k. The lowermost row in Figure 3 displays a derived detection statistic based upon the vector differencing scheme of Gibbons et al. (2008) which appears to be a good candidate statistic for the identification of new arrivals. It can be anticipated that a new matched field template be generated for every detected phase, covering the entire source region.

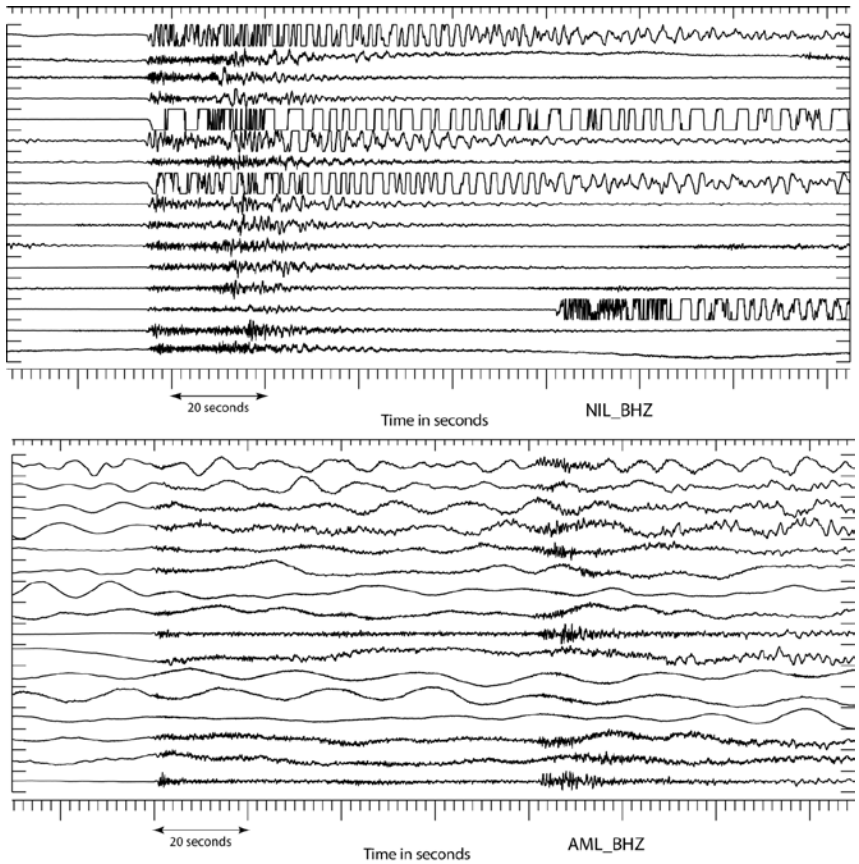

Figure 4. Unfiltered waveforms from NIL (Nilore, Pakistan: distance ~1 degree) and the AML station of KNET (distance ~9 degrees) for a selection of aftershocks from the October 8, 2005 Kashmir earthquake. Note that the dynamic range is exceeded for many of the signals at NIL.

Event Relocation for the October 8, 2005 Kashmir Aftershock Sequence

To be able to evaluate the performance of the pattern detectors and subsequent clustering techniques, we need the best possible event location estimates. The REB (Reviewed Event Bulletin of the IDC, Vienna), the USGS bulletin (http://earthquake.usgs.gov/earthquakes/eqarchives/epic/), and the bulletin of the International Seismological Center (ISC) (http://www.isc.ac.uk/) all cover a great many events in the sequence but all display rather large scatter and considerable inconsistencies. Common to all bulletins is the absence of arrival times at the NIL station and limited numbers of picks from KNET (see Figure 4). The REB fixes the depth of most events to 0 km whereas the USGS fix the depth of most events to 10 km. The ISC solutions (see left hand panel of Figure 5, for a limited set of events) display a wide range of depth estimates which is likely to contribute considerable scatter to the epicenter estimates. Using a partially automated procedure, arrival times were re-estimated at all available stations and a relocation procedure initiated. While some station corrections exist (e.g., Murphy et al., 2005), there are many paths without travel-time corrections and an iterative procedure has been initiated whereby a phase-specific travel-time correction for a given source region is estimated from the spread of travel-time residuals for that phase from the previous iteration. Preliminary results for 160 events are displayed in Figure 5, and the apparently improved consistency of depth estimates and lineation of epicenters gives reason to believe that the relative location estimates will improve. So far, only a constant station correction has been applied which is clearly not appropriate for the closest stations over the extended source region. We aim to produce an event list for the full aftershock sequence, with improved relative location and uncertainty estimates, using station corrections which are permitted to vary over the region displayed.

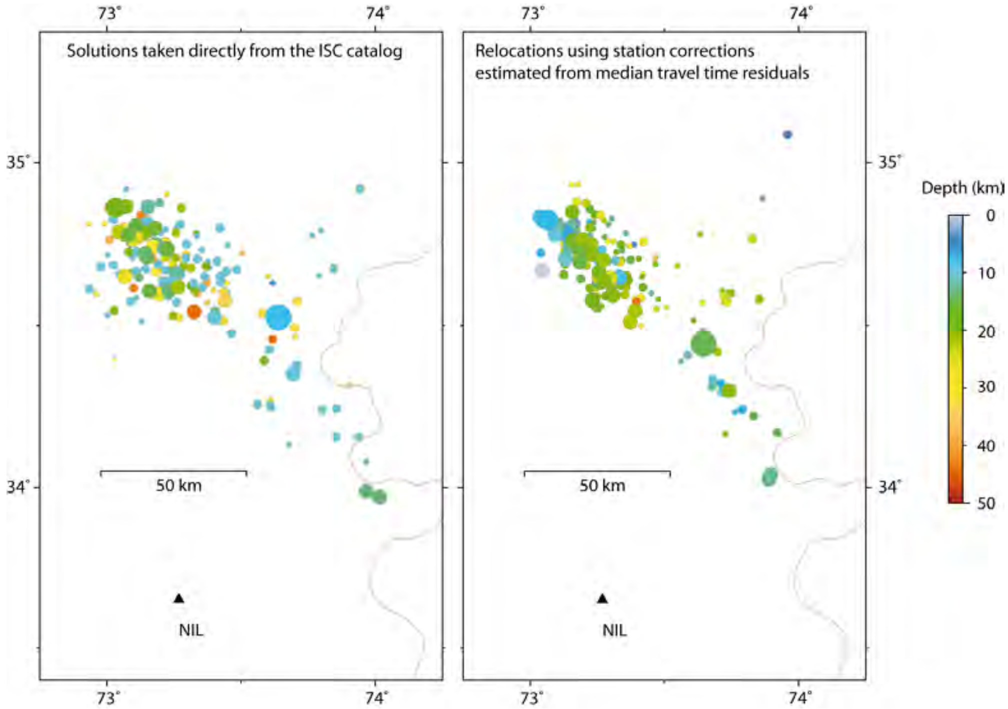

Figure 5. Preliminary event relocation estimates for 160 aftershocks in the October 8, 2005, Kashmir sequence. The left hand panel displays initial location estimates taken directly from the ISC catalog (http://www.isc.ac.uk/) and the right panel displays relocations using HYPOSAT (Schweitzer, 2001) where a station correction has been applied for each phase, estimated as the median travel-time residual relative to the ak135 model (Kennett et al., 1995) for the locations in the left hand panel.

Ratio-to-Moving-Average Seismograms: A strategy to Improve Correlation Detector Performance

The classical waveform correlation detector takes a template signal from an event of interest and calculates a correlation coefficient trace, sample-by-sample, against the incoming data. A significantly high value of the correlation coefficient indicates that a new occurrence of a similar signal, corresponding almost ripple-for-ripple to the template waveform, has been identified. What constitutes significance will depend upon the variability in the correlation coefficient trace that can be expected when correlating the waveform template against background noise and unrelated signals. The greater the complexity or time-bandwidth product of the waveform template, the lower the “background level” of the correlation coefficient trace is likely to be. This will allow us to set a more aggressive (lower) detection threshold without being inundated with false alarms. To maximize the signal complexity, it is usual to take the longest possible part of the wavetrain, bandpass filtered in the widest frequency band possible without compromising the signal to noise ratio. As discussed previously, in rapidly evolving aftershock sequences, it is frequently the case that phases from new events arrive at a sensor before the coda from the previous event has diminished (a clear example of this is seen for the NIL station in Figure 4). In such cases it may be necessary to limit the length of templates taken for some events, a procedure likely to require significant analyst supervision.

In preliminary correlation runs on the Kashmir sequence, an even more significant problem became apparent. The continuous correlation coefficient traces displayed a very high variability with frequent peaked local maxima which resulted in very high numbers of detections. When performing array-based waveform correlation on a medium-aperture array, such as KKAR, we are able to screen out many false alarms by performing f-k analysis on short segments of the single channel correlation coefficient traces (Gibbons and Ringdal, 2006). This procedure indicated that indeed many of the Kashmir correlation detections were invalid. While the majority of detections did correspond to arrivals from events in the sequence, if the f-k post-processing demonstrates that a detected signal arrives from a measurably different direction to the master event, such detections can be eliminated immediately as false alarms. The rapid modulation of the correlation coefficient traces results from relatively abrupt changes in the waveform amplitudes and we speculate that a procedure which reduces the variability in the dynamic range of the waveforms, without destroying the characteristic "fingerprint," may reduce the variability of the detection statistic traces without the detectors becoming any less sensitive.

We note the common practice in the correlation of long segments of ambient noise to reduce seismograms to one-bit traces prior to the correlation, in order to reduce the susceptibility to high amplitude transients (e.g., Larose et al., 2004). However, we are not at liberty to lose this degree of detail from our waveforms given the relatively short durations of the signals in question. We follow instead the procedure illustrated in Figure 6 whereby the bandpass filtered waveforms are divided by their corresponding short term average traces to result in what we denote RMA seismograms. Over short intervals of the order of a few seconds, these traces are almost indistinguishable from the original filtered seismograms. Over longer intervals, the traces appear almost featureless except for the occasional "spike" where an impulsive signal (of shorter duration than the STA length) exceeds the background level significantly (as displayed for the black trace in Figure 6).

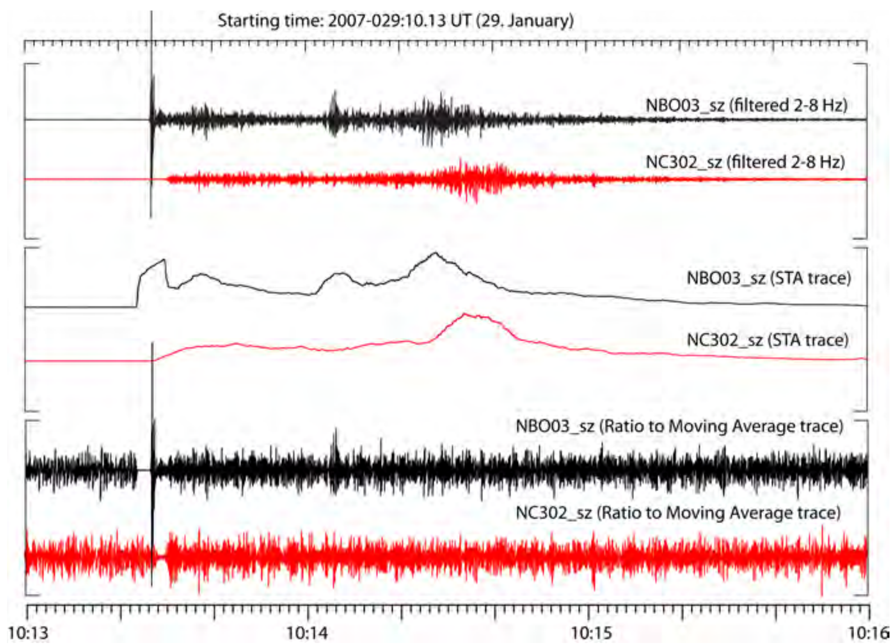

Figure 6. Generation of Ratio-to-Moving-Average (RMA) seismograms. The uppermost two traces display bandpass filtered waveforms on two sites of the large-aperture NORSAR array from a single earthquake off the western coast of Norway. The central two traces display STA traces constructed with a 0.025 second sampling interval where each sample gives a mean of absolute values over a 6 second interval. The lowermost two traces display the RMA traces: the bandpass filtered waveforms divided by the STA traces.

Figure 7 shows the filtered waveforms on the KKAR array for a one hour segment some 15 hours after the start of the aftershock sequence. There is a very large dynamic range, given frequent arrivals with amplitudes covering several orders of magnitude above the background noise. The single channel correlation coefficient trace (CC-trace) (third from top) generated from the template shown displays great variability with multiple peaks (detection candidates). The array CC-trace (fourth from top) has reduced the variability under the stacking operation, but appears problematic with respect to the setting of a detection threshold. The lowermost 3 traces of Figure 7 show an

RMA seismogram, the corresponding single-channel CC-trace, and the array CC-trace respectively. All of these traces have a fairly constant background level and it becomes far clearer where a detection threshold should be set in order that it be exceeded on only very few occasions. This simple transformation of the waveforms, prior to carrying out the correlation procedure, is likely to reduce the number of spurious correlation detections in an extensive aftershock sequence and may reduce significantly the danger of incorrect event associations.

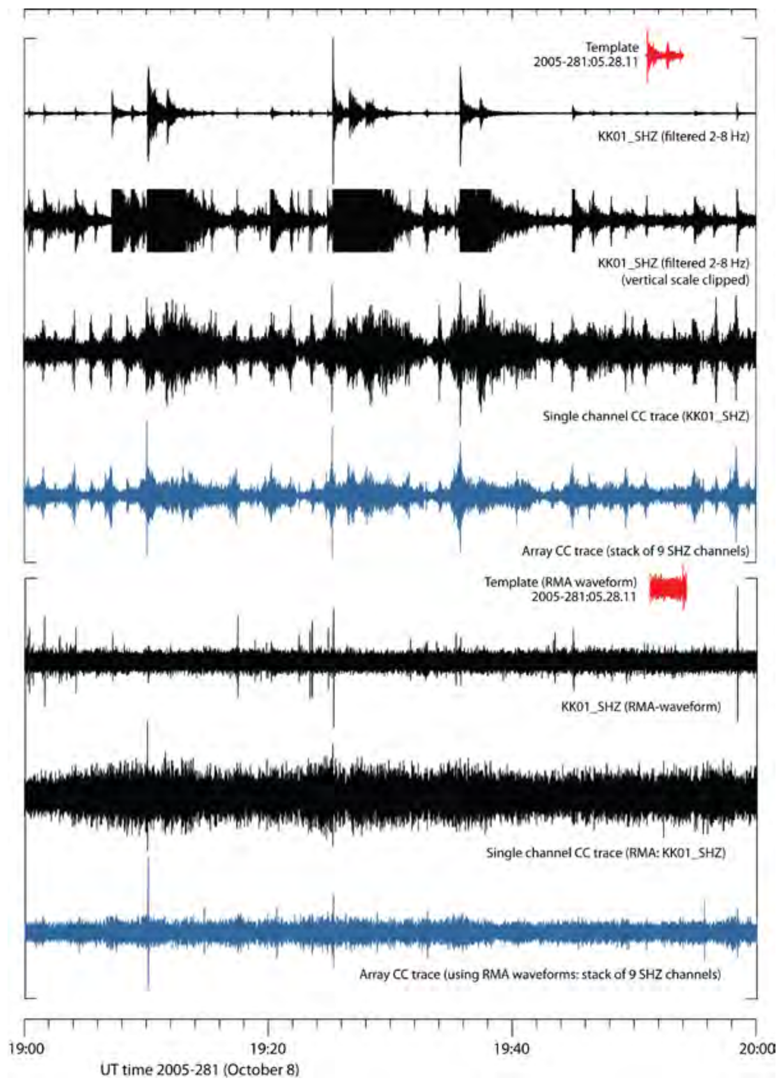

Figure 7. Performance of a multi-channel correlation detector on the 9 element KKAR array in Kazakhstan (distance ~9 degrees) for a one hour interval starting 19:00 UT on October 8, 2005, where the template is an aftershock earlier the same day. The upper panel shows the waveforms and corresponding detection statistic traces when the only preprocessing is bandpass filtering. The lower panel shows the corresponding traces when the filtering is followed by a calculation of the RMA for both template and incoming data. In this segment, only a single detection shortly after 19:10 UT can be unequivocally associated with an event located in the immediate vicinity of the master event. When the correlation procedure is applied to the filtered waveforms alone, this detection is not significantly above the background noise for the single channel case. Note the reduced dynamic range for the detection statistic traces as well as the waveforms for the RMA case in the lower panel.

CONCLUSIONS AND RECOMMENDATIONS

We have proposed a framework for adapting processing pipelines to incorporate and generate correlation, subspace, and matched field detectors that identify and exploit repeating waveforms in the incoming data stream. The purpose is to mitigate analyst overload under developing, intensive, aftershock sequences and to classify and track repeating explosions. We will investigate strategies for the spawning of new pattern detectors and for supervising the performance of different elements in the pipeline architecture.

We focus on the October 8, 2005, Kashmir earthquake and subsequent aftershock sequence observed primarily by the network of small to medium aperture arrays in Kazakhstan. The relatively poor quality of the available event bulletins makes evaluation of results difficult, particularly for the clustering algorithms. We have begun a systematic relocation of events in the sequence using a complete repicking of arrival times, particularly for phases at regional distances. An iterative procedure for event relocation has been started whereby station corrections are estimated from the median time-residuals from the previous iteration.

It was noted when applying waveform correlation methods to the 2005 Kashmir aftershock sequence that robust detection thresholds were difficult to set and that numerous spurious detections occurred. This is mainly due to the modulation and high variability of the cc-traces caused by the large dynamic range of the waveforms themselves. To mitigate this problem, we perform a transform of the filtered waveforms whereby each sample is divided by value at the same sample on the corresponding STA trace. The resulting RMA seismograms retain the detail of the original seismograms over short time-windows but have a far smaller dynamic range. The resulting CC-traces have a far lower variance, with fewer false alarms, and allow the setting of a lower detection threshold.

REFERENCES

Gibbons, S. J., and F. Ringdal (2006). The detection of low magnitude seismic events using array-based waveform correlation, *Geophys. J. Int*., 165: 149–165.

Gibbons, S. J., F. Ringdal, and T. Kværna (2008). Detection and characterization of seismic phases using continuous spectral estimation on incoherent and partially coherent arrays, *Geophys. J. Int*., 172: 405–421.

Harris, D. B. and D. Dodge (2011). An Autonomous System for Grouping Events in a Developing Aftershock Sequence. *Bull. Seismol. Soc. Am.*, 101: 763–774.

Harris, D. B. and T. Kværna (2010). Superresolution with seismic arrays using empirical matched field processing. *Geophys. J. Int.*, 182: 1455–1477.

Kennett, B. L. N., E. R. Engdahl, and R. Buland (1995). Constraints on seismic velocities in the Earth from travel times. *Geophys. J. Int.*, 122: 108–124.

Larose, E., A. Derode, M. Campillo, and M. Fink (2004). Imaging from one-bit correlations of wideband diffuse wave fields, *J. Appl. Phys*., 95: 8393–8399.

Murphy, J. R., W. Rodi, M. Johnson, D. D. Sultanov, T. J. Bennett, N. Toksöz, V. Ovtchinnikov, B. W. Barker, D. T. Reiter, A. C. Rosca, and Y. Shchukin (2005). Calibration of International Monitoring System (IMS) Stations in Central and Eastern Asia for improved seismic event location, *Bull. Seismol. Soc. Am.*, 95: 1535–1560.

Schweitzer, J. (2001). HYPOSAT - An Enhanced Routine to Locate Seismic Events. *Pure. Appl. Geophys.*, 158: 277–289.