NORSAR Scientific Report No. 1-2012

Semiannual Technical Summary

1 October 2011 – 30 June 2012

Svein Mykkeltveit (Ed.)

Kjeller, August 2012

6 Technical Reports / Papers Published

6.1 Adapting Pipeline Architectures to Track Developing Aftershock Sequences and Recurrent Explosions (*Paper Presented at the 2012 Monitoring Research Review*)

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 is being constructed with extensions that create and manage pattern detectors under a variety of spawning policies. The system is being tested on two aftershock sequences: that for the 8 October 2005, M=7.6, Kashmir earthquake and that for the 23 October 2011, M=7.1, Eastern Turkey event. Both cases are representative of challenging aftershock sequences given the vast numbers of events and relatively large source regions.

Pattern detectors that are coherent over multiple arrays and 3-component stations can constitute exquisitely sensitive detectors that increase the detection capability greatly for events in the immediate geographical vicinity of the master events. An alternative strategy would be to operate pattern detectors coherently over single arrays or other limited subsets of sensors, and combine the results incoherently across the complete network. This alternative strategy may allow a greater geographical region to be covered by given templates. The merits and limitations of the two strategies are being investigated for a range of different case studies. For correlation detectors on single arrays the validity of detections can be assessed by performing f-k analysis on single-channel detection statistic traces, eliminating enormous numbers of false alarms and allowing a significant reduction in the detection threshold.

Policies for triggering and spawning of correlation detectors are being studied extensively. The simplest trigger is using an STA/LTA detector on an array beam steered towards the slowness of anticipated first P-wave arrival from the source region considered, with the classification of the detected phase being confirmed using classical f-k analysis. This strategy is reinforced significantly when considering multiple observing stations. Considering only detections which are confirmed on multiple stations, with limits on the time-delays determined by the dimensions of the source region, will lead to fewer detectors generated by false triggers.

The most promising triggering algorithm considered so far is the single phase empirical matched field detector (EMFD). This narrow frequency band approach mitigates the effects of, and indeed exploits, the scattering which frequently confounds classical array processing. Correlation detectors using a waveform template from the main event are frequently very unsuccessful at detecting large number of aftershocks, since large spectral differences between the main shock and aftershocks may lead to

significantly different waveforms. The empirical matched field processing recognizes the characteristic spatial structure of the incoming wavefronts over the array aperture in each of many narrow frequency bands and this appears to be a more stable characteristic of a given source region than the temporal structure of the waveforms on each sensor. For both the Turkey and Kashmir sequences, the EMFD readily detects many very low signal-to-noise ratio (SNR) signals which are all confirmed by stations in the vicinity of the source region to correspond to real events in the sequences.

6.1.1 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. For additional system testing, we also plan to analyze signals from the October 2011 Van sequence in Turkey recorded at the Kazakhstan arrays. For this sequence a bulletin of quite accurate event locations and magnitudes, provided by the local Turkish network, form a good reference for evaluating the pipeline performance.

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.

6.1.2 Research accomplished

The Framework

The detection framework that we are building (Figure 6.1.1) models the detection front end of many pipelines that process array data, using an object-oriented architecture to allow different types of detectors to be added to the system dynamically and with any number of instances. The heart of the system is a list of detectors that can hold traditional beamformers with STA/LTA detectors and also several types of pattern-matching processors: correlation, subspace and matched-field detectors. The idea behind the system is to allow traditional power detectors (and one type of targeted pattern detector) to spawn new pattern detectors for specific aftershock families as they are detected and add them to the pipeline in real time. These pattern detectors are intended to sweep up some fraction of the aftershocks into groups for efficient interpretation later in the pipeline.

The system supports or will support other innovations, such as the ability to construct pattern detectors that span more than one array in the network and a capability to reprocess older data in a developing aftershock sequence with detectors created late in the sequence. Coherent pattern detectors with a larger network footprint may obviate some association problems that lead to incorrect event formation in automatic associators. They also should build event clusters that may be assumed to be families with a very high degree of confidence, which may lead to strategies for efficient use of analyst resources. The reprocessing capability (complete as of this writing) allows pattern detectors formed late in an aftershock sequence to sweep up similar events early in the sequence. This function may prove useful if analysts get behind by reducing the backlog of events to be reviewed.

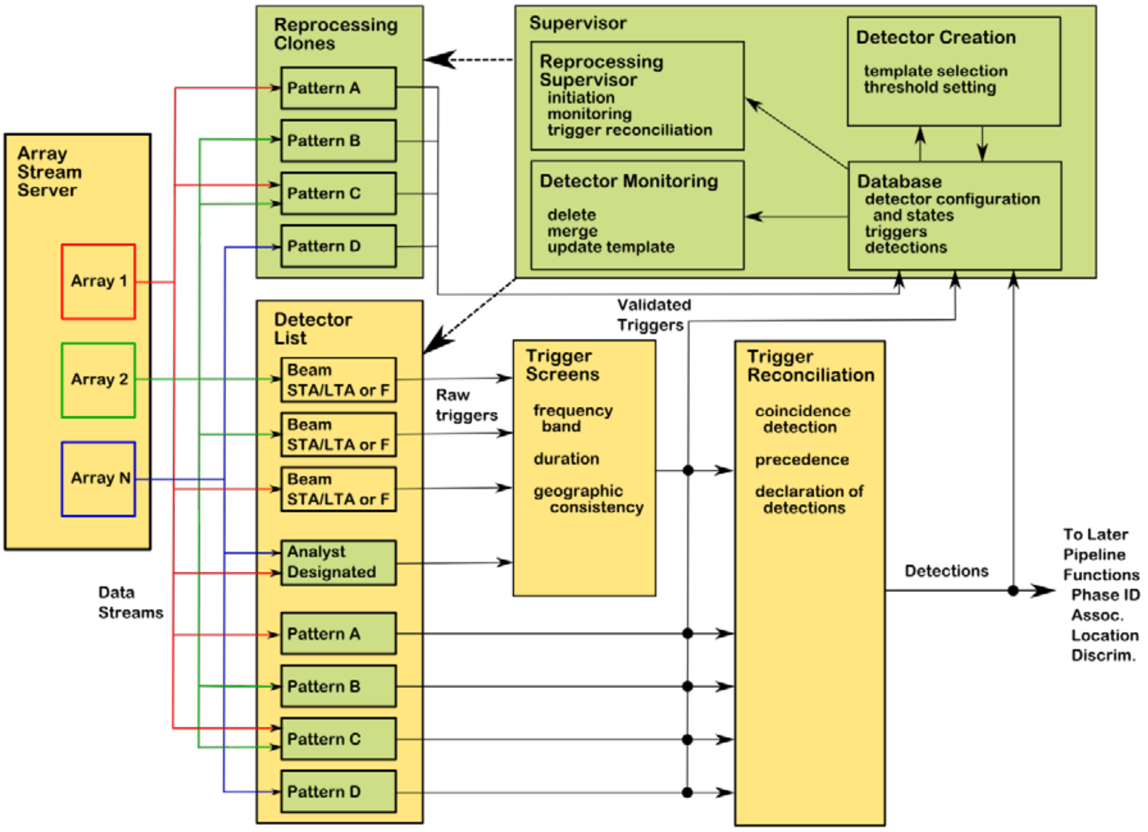

Fig. 6.1.1 Block diagram of the detection processor being developed to test pattern detector creation and management strategies. The boxes in yellow indicate functions approximately shared with a conventional pipeline detection processor. Those in green represent entirely new functions. Note that some detectors may process data from more than one array, possibly coherently.

To simplify the construction of detectors that span multiple arrays, we have completed an Array Stream Server function that acquires continuous stream data from different types of sources (flat files, database or conceivably real-time streams) and, so to speak, puts it on a common footing. By that we mean it resamples data to a common sampling rate and interpolates all samples to fall on the same time instants. This function simplifies selection of data from disparate stations to be combined (possibly coherently) in a single detector.

The other innovation is a supervisory function that creates detectors and monitors their performance. It incorporates an extensive database archiving the configuration of all detectors either designated by the system operators or created autonomously by the framework, all triggers produced by the detectors and the detections that emerge from a process of reconciliation and ranking of nearly simultaneous triggers. The archive supports tests of policies for updating (and possibly retiring and combining) detectors, as well as extensive post-operation analysis of overall system behavior and the performance of individual detectors. We are contemplating an ability to track and adapt the templates of detectors with new observations as they occur.

The kind of tests that we contemplate performing with this framework might be appreciated from an example. We would like to investigate how a subspace detector might be constructed automatically with a template that spans several arrays. One question is how to choose the receiver aperture of the template, i.e., which arrays to combine. Several possible strategies emerge. In a conservative approach, we might allow the system to operate for a time on a developing aftershock sequence to see if detectors are created among the arrays that have many triggers in common (discovered by examining whether patterns emerge in the relative timing of triggers made by the detectors). The absolute trigger times and the pattern of observed relative arrival times then could be used to extract waveforms for a correlation template. Alternatively, with a single event detected and under the assumption that we know the approximate location of the aftershock sequence, we could use a trigger time from a single station to predict arrival times at other stations for waveform extraction. We could operate the detector for a while to see whether it performed as well (had as many triggers) as a single-array correlation detector.

Because the pattern detectors we are using have efficient implementations, it may make sense to try fairly liberal spawning policies (such as the one just described), operate a fairly large number (thousands) of detectors, measure their performance and prune off the ones that don’t perform well. In this view the system could come to implement a kind of natural selection for empirical detectors.

Single Phase Empirical Matched Field Processing for Spawning Detectors

A pipeline which generates pattern detectors autonomously needs sensitive but robust criteria for spawning new detectors. Harris and Dodge (2011) classified an extensive aftershock sequence effectively by using triggers on an STA/LTA trace for the beam on a single array steered optimally to detect the initial P-arrivals from events in the source region. An ideal strategy for triggering the generation of a new pattern detector would be a correlation detector which simply took a waveform template from the main event and declared a detection on each occasion that the correlation coefficient trace attained a significant value. In practice, for large earthquakes, the contrast in source dimensions, magnitude, and consequent form of the signals usually make the mainshock signals very poor waveform templates for identifying aftershocks. This is demonstrated in the lowermost trace of Figure 6.1.2 where a 15 second long template for a teleseismic P-arrival from a magnitude 7.1

earthquake essentially fails to detect any of the aftershocks convincingly in the window displayed. The situation is improved considerably by considering the zero-moveout correlation stack on the array (trace 4) and it is possible that the noise floor could be lowered further by a flattening of the amplitudes in the incoming data stream (Gibbons et al., 2012).

Harris and Kværna (2010) present an application of empirical matched field processing (EMFP) on array signals to identifying the source of mining blasts. EMFP is a narrow frequency band procedure which matches the spatial structure of the incoming wavefield over the set of sensors, rather than the temporal structure of each time-series. EMFP outperforms waveform correlation for most mining sources since the ripple-fired nature of the shots makes the wavetrains from different blasting sequences very dissimilar, whereas the narrow band representations of the waveforms are relatively insensitive to the events' source-time functions and are highly characteristic for a given source region. EMFP also outperforms classical plane wavefront array methods since it takes templates from existing records from the relevant source region, producing pattern detectors which are uniquely calibrated for each source and account for the deviations that heterogeneous Earth structure imposes upon the arriving wavefront.

Wavefronts propagating to a distant array from an aftershock sequence will propagate along a very similar path. Therefore, the characteristic phase and amplitude relations between signals on the different sensors of the array are likely to be very similar from event to event and will comprise a fairly characteristic spatial seismic fingerprint for the sequence. Correlators ideally use the full wavetrain to maximize the signal's time-bandwidth product, and this can be problematic in rapidly unfolding sequences of events when the delay between events is often shorter than the wavetrains. Harris and Kværna (2010) demonstrated that a short window surrounding the initial P-arrival was a highly effective "information carrier" meaning only a short data segment may be required for each event. This will mitigate problems due to overlapping wavetrains. A further possible advantage of EMFP over waveform correlation in relating a large main event to subsequent smaller aftershocks results from the narrowband nature of the procedure. Correlation between the signals generated by the main event and aftershocks will usually be diminished by the significant spectral differences. In EMFP, in which the template is defined by the spatial structure of the wavefield in each narrow frequency band, the spectral content of each incoming wavefront is likely to be less significant.

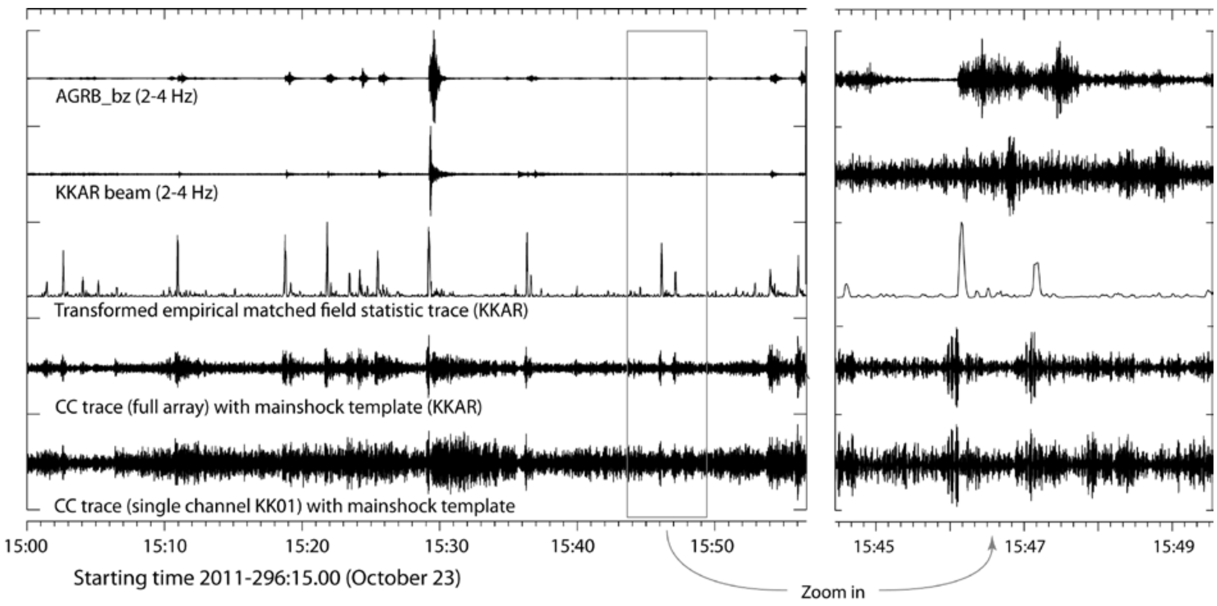

Fig. 6.1.2 Comparison of empirical matched field and correlation detectors on the KKAR array where both the empirical steering vectors and waveform template are taken from the initial P-arrival at time 2011-296:10.46.04 from the October 23, M=7.1 Van earthquake, Turkey (38.733°N, 43.483°E). The covariance matrix used to calculate the empirical steering vector was calculated using a data segment with 256 samples (6.4 seconds) and 26 narrow frequency bands between 1.09 and 5.00 Hz were used. The waveform correlation calculations were performed with a 15 second waveform template filtered 1-5 Hz. Data from the AGRB station (approximately 1 degree from the mainshock) is delayed by 266 seconds to provide an optimal alignment of initial P arrivals with the P phases at KKAR.

A matched field statistic trace can be evaluated for consecutive windows of incoming data, in a given narrow frequency band, for a given template or empirical steering vector. The third trace in Figure 6.1.2 displays the mean over 26 narrow frequency bands of transformed matched field statistic traces, where the transformation (described in Gibbons et al., 2008) results in local maxima at times characterized by an increase. There are numerous clear peaks in this function which can all, on closer inspection, be associated with events visible in data from far closer stations. The zoom panel makes it clear that the matched field statistic, measuring the spatial characteristics of the wavefield over the array, provides a significant improvement in SNR over an optimal beam steered with the appropriate time-delays. Times of arrival are clearer on the matched field traces than the correlation traces and this may be particularly useful if valid event hypotheses are to be validated by observations over multiple arrays.

STA/LTA detectors on a beam return many detections that are unrelated to the source region of interest (a strong signal from a completely different direction is likely to result in an increase of energy on all beams). The same may occur for the matched field detector. Just because sensitivity is optimal for the direction of the template steering vectors, this does not mean that the response to wavefronts arriving from other directions will be negligible. Gibbons and Ringdal (2006) demonstrate how vast numbers of false alarms using correlation detectors on arrays can be eliminated fully automatically by performing f-k analysis on the correlation coefficient traces, essentially testing whether or not directions other than the one corresponding to the master event are preferred. We here demonstrate an equivalent procedure for matched field detections.

Figure 6.1.3 shows f-k analysis in two different narrow frequency bands, at 2 Hz and 4 Hz, for the Pn phases for the main October 8, 2005, Kashmir event (left) and a selected aftershock (center). The similarity between the f-k grids strengthens the claim that two earthquakes in the same sequence (in this case separated by over two units of magnitude) will result in wavefronts with similar phase and amplitude relations between sensors in the various narrow bands. As anticipated, the relative beam power at 4 Hz is less than that at 2 Hz with greater sidelobes resulting from aliasing and incoherence. The differences between the 2 Hz and the 4 Hz slowness grids provide some of the motivation for using EMFP as opposed to the simple plane-wave model.

If a detection is made using a given empirical steering vector, for example from a “mainshock”, then we can scan slowness space in the same way that we do in classical f-k analysis, only that the phase shifts for the theoretical plane-waves are superimposed onto the phase shifts specified by the empirical steering vector.

In the right panels of Figure 6.1.3, we map out the f-k spectrum at the time of the Pn arrival at KKAR from the aftershock, relative to the empirical steering vector obtained from the Pn arrival from the mainshock. At both 2 Hz and 4 Hz, the maximum relative beam power is obtained close to the zero slowness vector, indicating that the detected phase is very likely to be from the same direction as the master event phase.

These results suggest that a single-phase empirical matched field primary detector designed from the main or early events of an on-going earthquake sequence would provide a robust and sensitive tool for spawning new pattern detectors. When used in combination with the described screening procedure, this primary detector can operate with a low false-alarm at a low detection threshold.

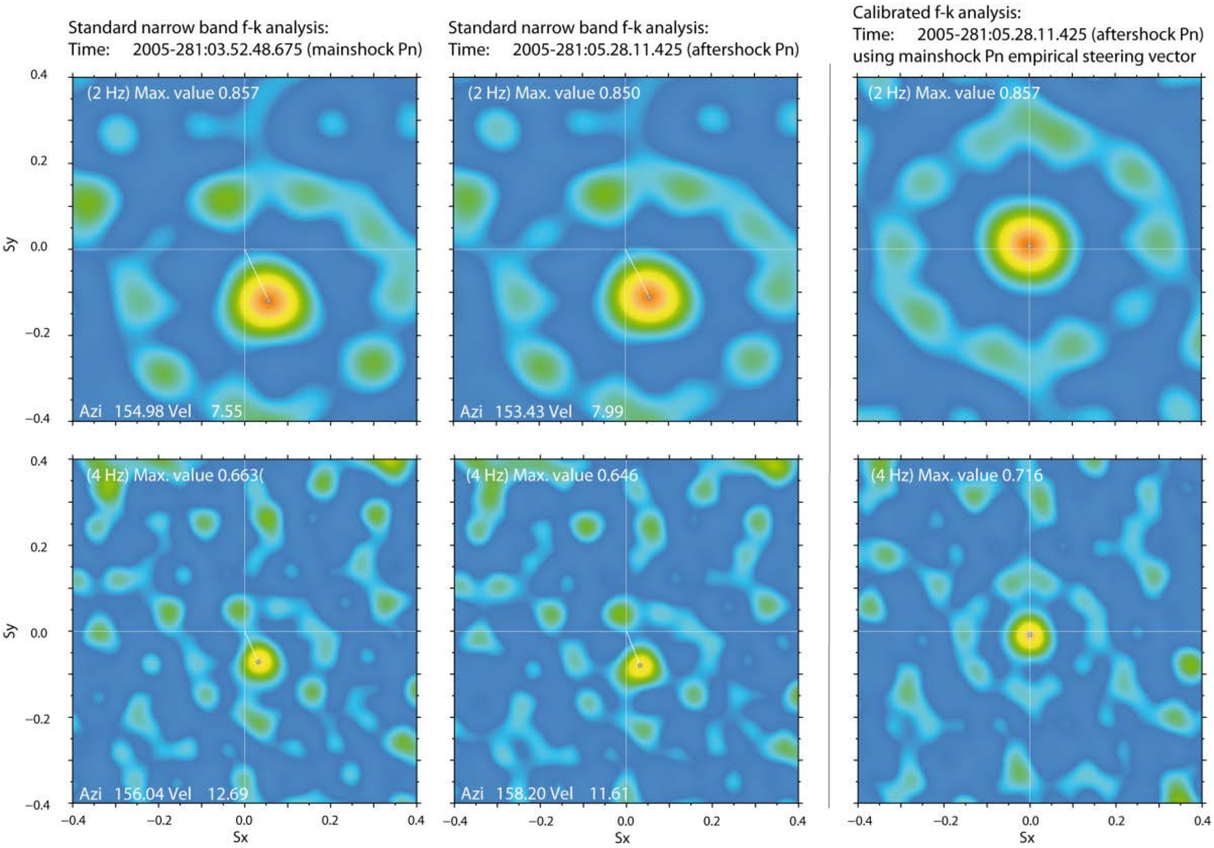

Fig. 6.1.3 Narrow band f-k analysis performed at the times indicated at 2.0 Hz and 4.0 Hz. The procedure employs the multitaper subroutines of Prieto et al. (2009). In the panels on the right, the phase shifts in the empirical steering vector are imposed and the zero slowness vector indicates that the wavefront observed comes from the same direction as the master signal.

The framework operating on aftershocks of the 2005 Kashmir earthquake

Our eventual goal is to enable the framework automatically to create detectors that operate on individual arrays or combinations of arrays. While an algorithm for creating (spawning) individual-array detectors is in place, we are experimenting with procedures for spawning detectors that operate across a network of arrays. Subspace detectors with a wider geographic footprint (on the receiver end) will have advantages in defining groups of events that are reliably related (families) and may obviate association problems among multiple stations. We are emulating manually parts of the eventual automatic process with tests on a ten-day period (2005:281-290) with data from the Kazakh arrays.

In our initial test, the spawning detectors were power detectors (STA/LTA) operating on beams directed at the backazimuths and slownesses of the initial P phase of the Kashmir mainshock (determined by FK analysis) for stations KKAR and ABKAR. KKAR is 10.5 degrees from the mainshock and ABKAR is 19.4 degrees distant. After some experimentation, we found that a two-pass process works well to define initial groups of correlated events. For example, we performed an initial run of the framework in which the spawned correlators were given a detection threshold of 0.6 (KKAR) and 0.4 (ABKAR) and template lengths of 160 and 310 seconds respectively. Following the initial pass all detectors with fewer than 4 detections (at KKAR) and 2 detections (at ABKAR) were removed from the system, and the framework was run again (with no spawning) with the surviving correlation detectors operating at lower thresholds (0.1).

Table 6.1.1 summarizes the number of detections made in the second pass. Statistics are shown only for the detectors that made 4 or more detections on the second pass. We emphasize that the two arrays were treated independently until this point.

Table 6.1.1 Numbers of detections made by automatically-created correlation detectors on a second pass.

KKAR Detectors		ABKAR Detectors	
DETECTORID	Detection Count	DETECTORID	Detection Count
52037	631	54422	467
52204	67	54470	14
52150	61	54902	5
52583	44	54442	5
52130	26	54610	5
52181	26	54823	4
52085	21	54512	4
52182	19	54558	4
52090	16		
52349	14		
52264	14		
52424	13		
52101	11		
52128	11		
52135	7		
52199	7		
52139	6		
52369	5		
52681	4		

Figure 6.1.4 displays seismograms from the most prolific detectors (52037 at KKAR and 54422 at ABKAR). Although most detail is not visible at the scale of these plots, nonetheless, the arrivals of P and S waves are clear with S-P times of about 110 seconds at KKAR and 210 seconds at ABKAR.

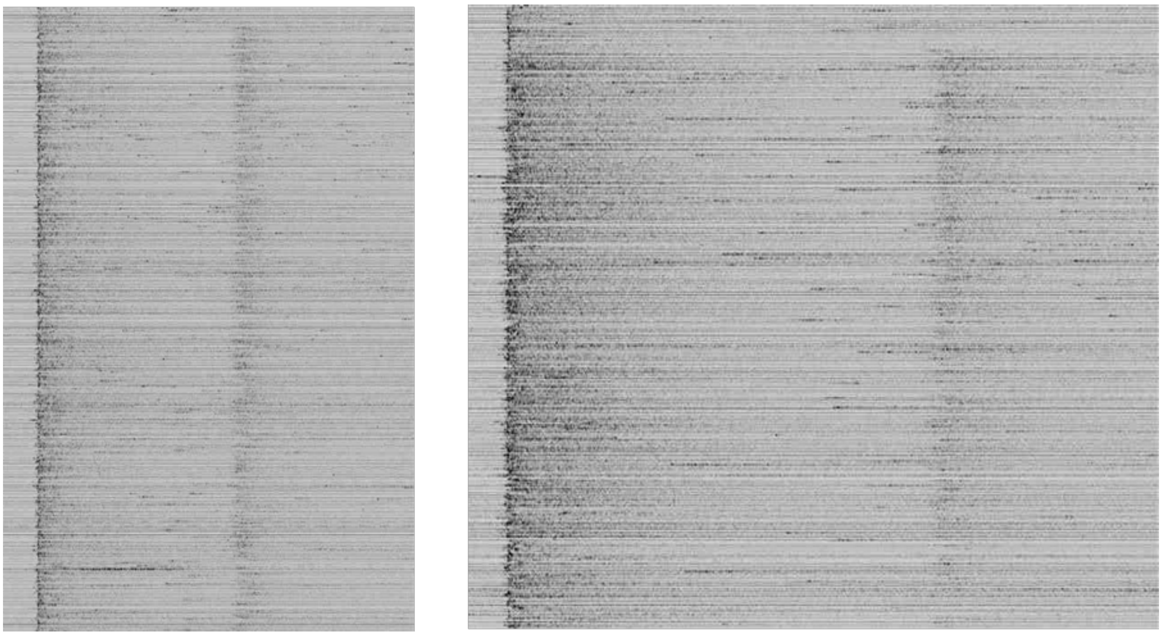

Fig. 6.1.4 Aftershocks of the 2005 Kashmir event recorded at KKAR (left) and ABKAR (right). The plot at left displays 631 events found by automatically-created detector number 52037 (220 second window of data) and the plot at right displays 467 events found by detector 54422 in a window 320 seconds long.

Our next step was to determine which detectors found events in common. For this purpose, we searched for events with common offsets in trigger times between the arrays, based upon observed P travel times (143 seconds for KKAR and 260 seconds for ABKAR; 117 second offset) on a sample of 10 detected aftershocks. Table 6.1.2 shows the result for all pairs of detectors that had more than 4 detections in common (in fact, perhaps twice as many detector pairs had at least one detection in common).

To create a trial joint detector, we used the 172 common detections between KKAR detector 52037 and ABKAR detector 54422. The template was 380 seconds long and included the 9 KKAR SHZ channels and the 9 ABKAR SHZ channels. A subspace detector was built with an energy capture threshold of 0.9, resulting in a very high rank (128) detector. Electing caution in our first attempt, we reprocessed that combined data stream of the two arrays for the ten days 2005:281 – 290 with a conservative detection threshold of 0.2. The detector collected 507 detections, compared with 631 for detector 52037 operating at KKAR alone, with few or no false triggers. Figure 6.1.5 displays plots of the extracted detections first at KK01 and then at ABK01.

Table 6.1.2 Numbers of detections found in common between pairs of independently-created detectors at stations KKAR and ABKAR.

KKAR Detector ID	ABKAR Detector ID	Common detections
52037	54422	172
52204	54422	22
52150	54422	14
52182	54422	11
52090	54422	10
52583	54422	9
52181	54422	8
52085	54422	6
52424	54422	5
52037	54610	5
52349	54422	4
52101	54422	4
52128	54422	4
52369	54422	4

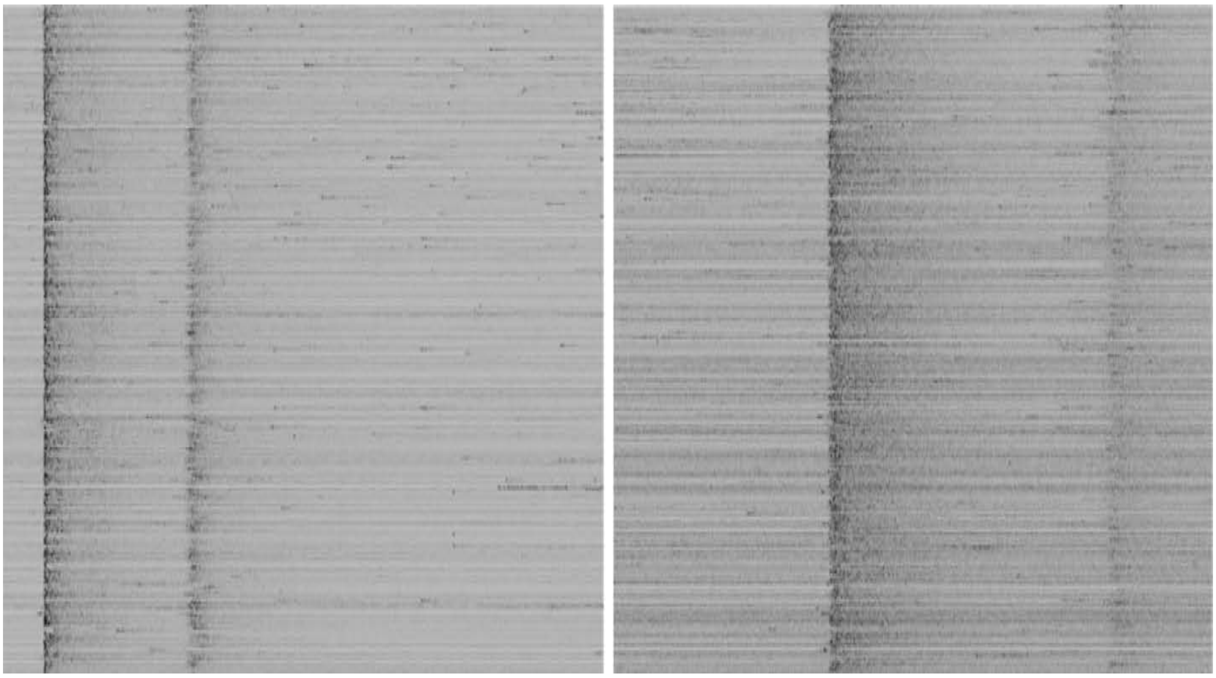

Fig. 6.1.5 Plots of 507 events at station KK01 (left) and ABK01 (right) detected by the joint subspace detector of rank 128.

6.1.3 Conclusions and recommendations

Correlation and subspace detectors are the principal methods for event detection and grouping. However, correlators using templates from extremely large earthquakes are often ineffective at spawning new pattern detectors for classifying large numbers of far smaller aftershocks. This is due to waveform dissimilarity resulting from the disparity of event magnitudes and source dimensions. While STA/LTA detectors on array beams steered towards the source region of interest, followed by f-k analysis for verification of direction and slowness, constitute an intuitive triggering algorithm for detector spawning, we argue that single-phase EMFP is both a sensitive and robust method for triggering new pattern detectors. EMFP operates on short data segments, which mitigates the problems associated with overlapping signals from consecutive events in a sequence. EMFP is a narrowband procedure, measuring the spatial structure of an incoming wavefront over an array of sensors, and may be less susceptible to differences in the spectral content between the signals from different events. We have demonstrated clear detections of confirmed aftershocks using EMFP for which the signal-to-noise ratio on the beam itself is so low that detection using an STA/LTA detector would not be feasible. It is also noted that false alarms from the EMFP detector are readily screened out by scanning the slowness space relative to the imposed template empirical steering vector.

Our initial trial of a detector operating across two arrays (KKAR and ABKAR) produced a very high-dimension (128) subspace detector that swept up a large number (507) of aftershocks from the 2005 Kashmir sequence. Because the dimension of the detector was so large, we elected to use the detector with a relatively high correlation power threshold (0.2, comparable to 0.45 linear correlation). Probably because the template was so large (TB > 10,000) and encoded very precise and large time delays across the combined aperture of the two arrays, few or no false alarms were detected. This fact suggests that the threshold could be reduced substantially to allow an even larger collection of aftershocks to be swept up in a correlated group. We plan to experiment with lower thresholds and to build joint detectors from additional pairs independently-defined single-station event clusters. We intend to automate the process (manual to this point) of forming joint detectors, experimenting with several different policies for initiating joint detectors.

Acknowledgements

We are grateful to the Kazakhstan National Data Center for permission to use the data from the KKAR array. Data from the AGRB station in Turkey was obtained through ORFEUS and is courtesy of the Kandilli Observatory Digital Broadband Seismic Network.

This research has been supported by the US Air Force Research Laboratory and the US National Nuclear Security Administration, award nos. FA9453-11-C-0230[1,2] and DE-AC52-07NA27344 [3]

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]

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.

Gibbons, S. J., F. Ringdal and T. Kværna (2012): Ratio-to-moving-average seismograms: a strategy for improving correlation detector performance, Geophys. J. Int. 190, 511–521.

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. Kvaerna (2010): Superresolution with seismic arrays using empirical matched field processing. Geophys. J. Int. 182, 1455–1477.

Prieto, G. A., R.L. Parker and F.L. Vernon (2009): A Fortran 90 library for multitaper spectrum analysis. Computers and Geosciences 35, 1701–1710.