

**ADAPTIVE WAVEFORM CORRELATION DETECTORS FOR ARRAYS: ALGORITHMS FOR
AUTONOMOUS CALIBRATION**

Frode Ringdal¹, Steven J. Gibbons¹, and David B. Harris²

NORSAR¹ and Lawrence Livermore National Laboratory²

Sponsored by National Nuclear Security Administration
Office of Nonproliferation Research and Development
Office of Defense Nuclear Nonproliferation

Contract Nos. DE-FC52-05NA26604¹ and W-7405-ENG-48²

ABSTRACT

Waveform correlation detectors compare a signal template with successive windows of a continuous data stream and report a detection when the correlation coefficient, or some comparable detection statistic, exceeds a given threshold. Since these methods exploit characteristic details of the full waveform, they provide exquisitely sensitive detectors with far lower detection thresholds than typical short-term average/long-term average (STA/LTA) algorithms. The drawback is that the form of the sought-after signal needs to be known quite accurately a priori, which limits such methods to instances of seismicity whereby a very similar signal has already been observed by every station used. Such instances include earthquake swarms, aftershock sequences, repeating industrial seismicity, and many other forms of controlled explosions. The reduction in the detection threshold is even greater when the techniques are applied to multiple channels since stacking can be performed on the correlation coefficient traces with a significant array-gain. A detected event that is co-located with the master event will record the same time-difference at every site in an arbitrarily spaced network which means that the correlation coefficient traces can be stacked coherently even when there is little or no similarity between the actual signals at the different sites.

In the first year of this three-year investigation, the emphasis was upon estimating the detection threshold reduction for a range of highly repeating seismic sources using arrays of different configurations and at different distances from the events examined. In the case studies pursued (induced seismicity in a coal mine, aftershock sequences in Fennoscandia, and the 2002 Coso sequence in Southern California) the master events were all of magnitude less than 4.5 and, in many cases, nearby instruments were able to confirm the similarity of waveforms at local distances which increased the likelihood of successful detection at regional distances. In the second year, we examined case studies aimed at investigating the applicability of matched filter detectors to more difficult sources. The first case is a sequence of military explosions on the Kola Peninsula in Russia at approximately 200 km from the ARCES seismic array. The difficulty here is the lack of waveform similarity from event to event, presumably due to large variation in the detonation procedure and source-time function, which results in quite low correlation coefficients. However, the alignment and coherence of the correlation coefficient traces between the different sites of the small-aperture array provides additional selection criteria which eliminate the vast majority of spurious event hypotheses. The second case study is the extensive aftershock sequence from the M=7.7 event 8 October 2005 in Pakistan. Many of these events were well in excess of magnitude 5 with correspondingly larger rupture sizes than had been considered previously. Driving multichannel correlation detectors at teleseismic distances using many of the larger events as master signals classifies rapidly subsequent events which are likely to be associated with the various templates. Many additional events were then successfully identified by correlating high-frequency regional waveforms over the large aperture K-NET network at distances of over 900 km.

The relative location of detected events is an important part of the event categorization process and we demonstrate how the horizontal double-difference relocation procedure of Schaff and Richards (2004) can be applied to events of smaller magnitude when the correlation trace stacking is applied over small to medium sized arrays at a range of different directions from the source. We also demonstrate cases of erroneous timing on International Monitoring System (IMS) arrays that have been measured accurately as a result of the correlation procedure for repeating seismic events.

OBJECTIVE

The overall objective of this proposed three-year study is to develop and test a new advanced, automatic approach to seismic detection using waveform correlation, with special application to seismic arrays. The principal goal is to develop an adaptive processing algorithm. By this we mean that the detector is initiated using a basic set of reference (“master”) events to be used in the correlation process, and then an automatic algorithm is applied successively to provide improved performance by extending the set of master events selectively and strategically. These additional master events are generated by an independent, conventional detection system. A periodic analyst review will then be applied to verify the performance and, if necessary, adjust and consolidate the master event set.

RESEARCH ACCOMPLISHED

In the first year of this project, the emphasis was on two major aspects of array-based waveform correlation detectors. Firstly was the issue of the reduction in detection thresholds possible using seismic arrays of different spatial configurations and at different epicentral distances. Secondly was the construction of subspace detectors (a generalization of the matched filter detector which accommodates a greater variation in the permissible waveforms) and a comparison between the success rate and false alarm rate of 1-dimensional and multidimensional subspace detectors.

In this second year, a number of case-studies have explored situations that are less than ideal for correlation detectors—primarily due to a demonstrable variation in the waveforms. We report on two such cases here. The first is a series of presumed multiple ammunition explosions which are challenging due to very heterogeneous source-time functions. The second is an extensive aftershock sequence from a large earthquake. Although all the scenarios studied in the first year were limited to small magnitudes (all less than 4.5 and mostly less than 3.5), this earthquake sequence consisted of events up to magnitude 7.7 covering a very large geographical area. Two further topics which are covered briefly: firstly, the rapid identification of erroneous instrumental timing over arrays through correlating long data segments from repeating seismic events and, secondly, relative relocation of seismic events.

Application of a Waveform Correlation Detector to a Sequence of Explosions with Differing Source-Time Functions

In March 2005, Norwegian authorities alerted NORSAR to concerns of citizens living in the far north of Norway who had experienced loud noise from an unknown source. From an association of reported times with seismic signals, it was concluded that numerous events had occurred on the Kola Peninsula of NW Russia (Figure 1); sound waves coming from the appropriate directions were also observed on the seismic sensors of the ARCES array and on the Apatity seismic array and microbarograph subarray (Ringdal et al., 2005). The automatic location estimates from generalized beamforming (GBF) (Ringdal and Kväerna, 1989) showed considerable variation (Figure 1, left panel) but analyst reviewed solutions (including some which applied direction estimates from the infrasound phases) appeared to result in quite consistent location estimates close to the circle in Figure 1. The primary reason for the large spread in automatic event locations is that multiple detections are made for both P-type and S-type regional phases in each signal with highly differing patterns of occurrence. This indicates multiple firing sequences. The automatic phase association algorithm is unable to identify which bursts of energy correspond to which shots and different hypothetical source locations will provide arbitrarily better or worse fits to phase detection sequences. The heterogeneity of the waveforms from this particular event sequence is clear from the traces in the right panel of Figure 1.

Our aim is to identify similar and subsequent events. The fully automatic event lists are clearly little help in this aim without the introduction of much manual analysis. The Zapoljarni mines on the Kola Peninsula frequently result in seismic signals whose automatic event location estimates overlap with the presumed explosion site. Waveform correlation is an appealing method, but low values of the correlation coefficients resulting from comparing the seven traces in Figure 1 make it clear that a detection threshold for a correlation detector would have to be very low and that we might have to accept a very high false alarm rate. One observation from the mutual correlations between these seven events is that despite relatively low values of the Array Correlation Coefficient Beam (ACCB), performing f-k analysis on the single channel correlation coefficient traces (e.g., Gibbons and Ringdal, 2006) resulted in very well defined slowness vectors (i.e., indicating good alignment of the correlation coefficient traces).

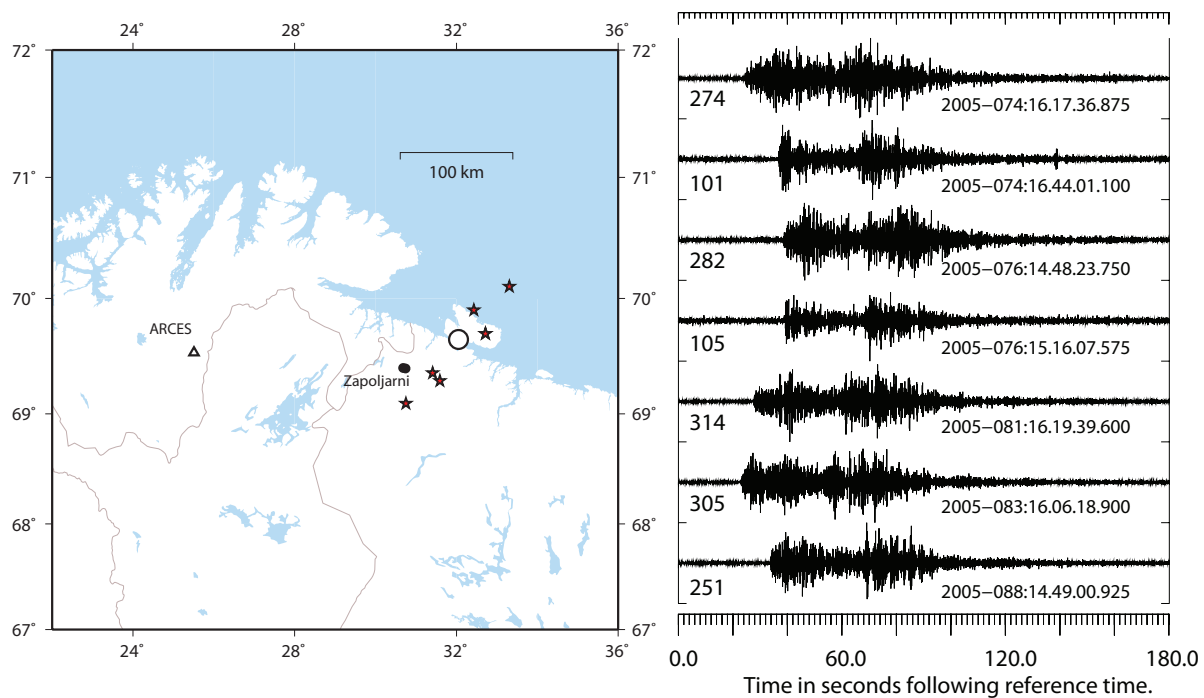


Figure 1. (left) Estimated location of the Kola Peninsula explosion site (open circle) in relation to the ARCES seismic array and the Zapoljarni mines (closed circles). The stars indicate fully automatic location estimates (GBF: Ringdal and Kværna, 1989) for seven events (two co-located) which NORSAR were made aware of due to reports of loud noises from residents in northern Norway. (right) ARA0_sz seismograms (bandpass filtered 3–8 Hz) for these events, each trace starting at the GBF origin time estimate as indicated. The number displayed is the maximum amplitude of the trace.

With the detection threshold set necessarily low, between January 1, 2002, and December 31, 2005, a total of 17,485 detections were made based upon the ACCB value alone. The vast majority of these were demonstrably false alarms by inspection and other selection criteria were deemed necessary: the ratio of ACCB maximum to the standard deviation, the apparent slowness of the correlation coefficient traces, and the coherence of the correlation coefficient traces. This multi-variate selection condition reduced the number of detections to 243. Of these detections, 220 corresponded to the times detected events on the GBF which appeared to originate from a similar source region and which often showed similar properties. Of special interest is the appearance of sound waves on the ARCES array some 12 to 14 minutes after the explosions, which supports the hypothesis that the events are of a similar nature. An example of such a detected event is displayed in Figure 2, displaying the somewhat dissimilar waveforms and the ACCB with no clearly defined maximum. Three of the 243 detections did not correspond to automatic GBF event hypotheses, although careful manual VESPAgram analysis indicated evidence of weak P- and S- type regional phases coming from the appropriate direction. The remaining events were clearly false alarms and resulted from the occurrence of very high amplitude, short-duration, Rg-type phases arriving from approximately the same backazimuth.

Whilst the actual events are somewhat poorly constrained (i.e., there is no Ground Truth, and no independent confirmation of the location of the explosion site or sites) the method we have applied has demonstrated that correlation detectors which apply additional constraints (primarily on the alignment of the single channel correlation coefficient traces) have been able to produce an extensive list of very likely candidate events with very few obvious false alarms. It is worth noting that there are no coincidental correlation detections with signals from Zapoljarni mining events. Any other existing procedure to identify that number of candidate events would almost certainly also result in many false alarms and much additional analyst time.

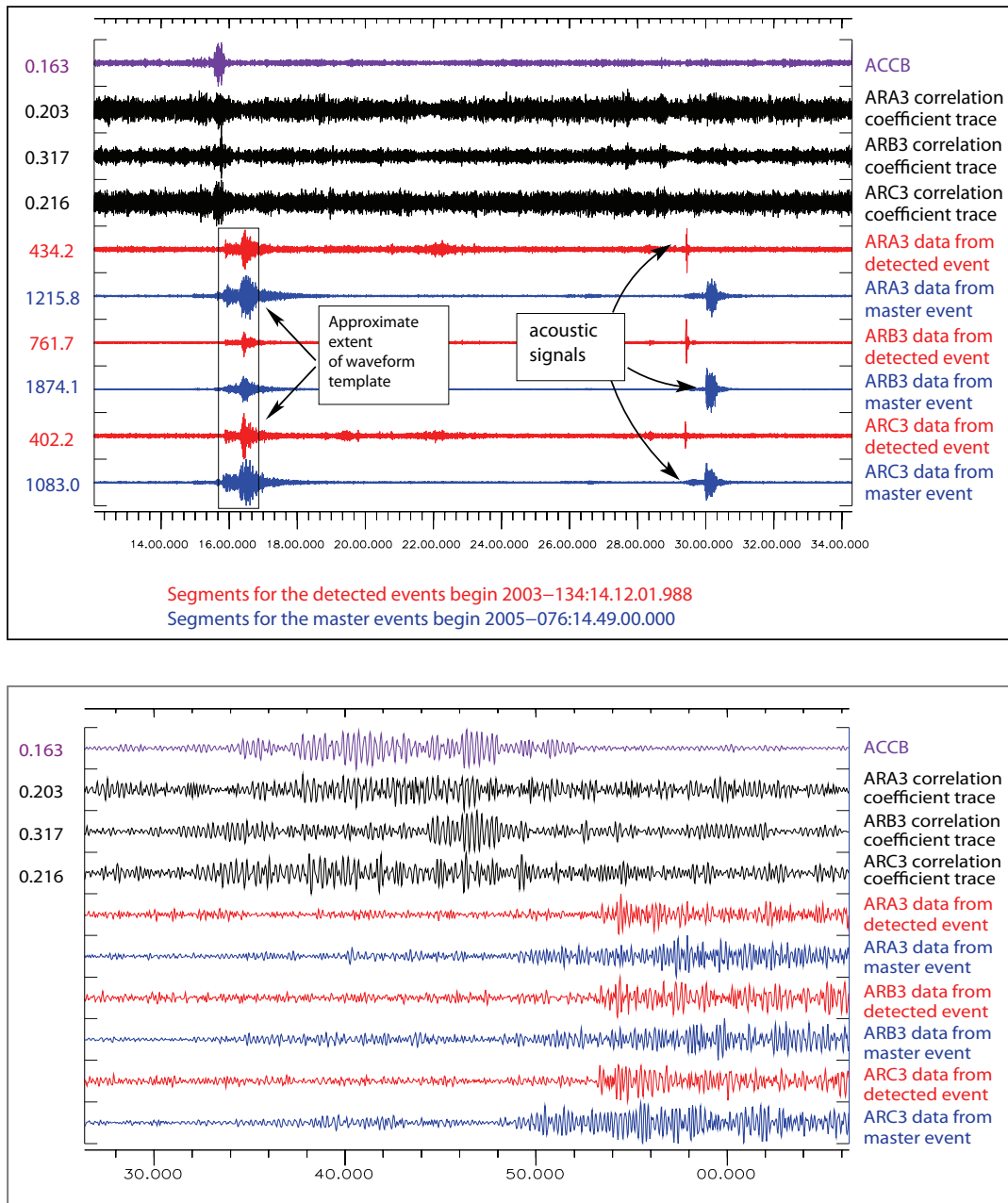


Figure 2. A typical correlation detection on the ARCES array using a template from a Kola Peninsula event on March 17, 2005. The seismic signals from both master and detected events are followed by infrasound arrivals. Note the long duration of the semblance of the correlation coefficient traces and the absence of well-defined peak in the ACCB.

Application of a Waveform Correlation Detector to an Extensive Aftershock Sequence using a Sparse Network

One kind of scenario in nuclear explosion monitoring where a large number of detection templates may be required to cover a wide source region is an active mine. The extent of the mine is likely to be larger than the “correlation footprint” of the event (Geller and Mueller, 1980), and a different template is likely to be required for many distinct regions of the mine. Autonomous calibration may be necessary given that as excavation continues into previously untouched rock masses, the distance of the new seismic sources from previously observed sources will increase and the correlation between the corresponding signals will decrease. The autonomous calibration algorithm will need to decide in which circumstances a new template should be added to (or removed from) the event pool.

A very different scenario is the occurrence of a large earthquake followed by an extensive series of aftershocks. In contrast to the gradual change to the continual low magnitude seismicity in the mine, a large earthquake is likely to occur in a region of a main fault zone where seismicity has not been observed for many years, if ever, and is likely to be followed by large numbers of seismic events covering a very large range of the magnitude spectrum, which warrant rapid and exhaustive categorization. The problems involved in applying correlation detectors are obvious. The main shock is likely to be in excess of magnitude 7 with a corresponding rupture length many times larger than the fundamental wavelengths of the aftershocks which are sought. The spectral content of the resulting waveform will also differ significantly from smaller aftershocks.

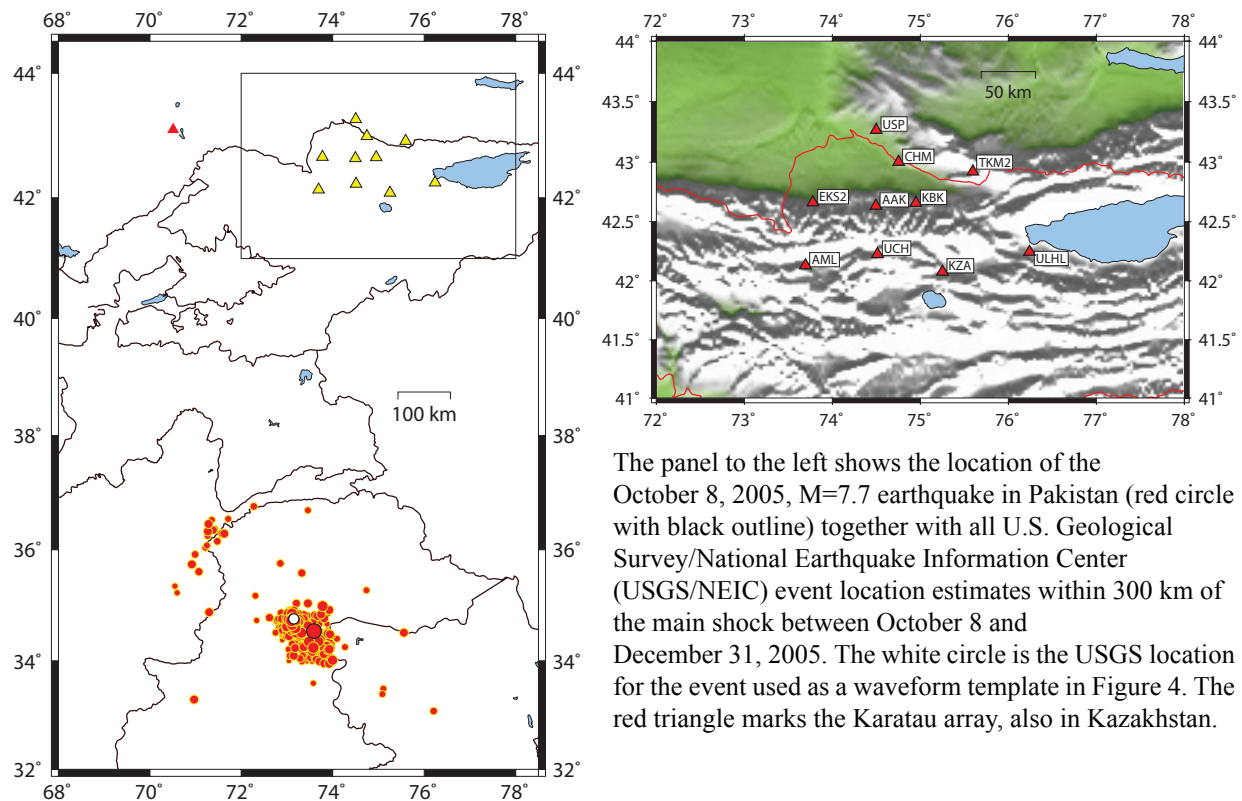


Figure 3. Locations of the October 8, 2005, Pakistan earthquake and related events together with stations of the K-NET network in Kyrgyzstan/Kazakhstan.

One such example is the $M=7.7$ earthquake which caused devastation in Pakistan on October 8, 2005. A template from the master event was extracted for the NORSAR and ARCES arrays at teleseismic distances and correlated experimentally against subsequent data. Not surprisingly, this resulted in few convincing detections (as characterized by sudden significant maxima on the ACCB) but rather a similarity of correlation traces over extended time segments. This is probably due to the fact that the wavefronts from subsequent events at these distances appears to the array to come from a similar direction, but that the actual waveforms (which are considerably more limited in frequency content than the regional signals displayed in Figure 2) do not show a convincing degree of similarity.

Data from the KNET stations (Figure 3) was obtained from the Incorporated Research Institutions for Seismology (IRIS) and similar correlation detectors were initiated. The master signal deriving from the main event was not useful for detecting aftershocks at these regional distances. However, waveform templates obtained from smaller aftershocks did result in many quite convincing detections on the KNET data. An example is shown in Figure 4. This is a larger aperture network than any that we have previously attempted the correlation coefficient beamforming on and the array gain is excellent with marginal peaks on the single channel traces summing to one well defined maximum and cancelling at other points in time. While the master event displayed in Figure 4 is significantly smaller than the main shock, it is also significantly larger than master events considered so far in this project. The recordings of these two events at teleseismic distances also show a degree of likeness but, as with the Russian explosions at ARCES in the previous section, are likely to require additional constraints (primarily regarding correlation trace alignment) in order to produce useful (low false-alarm-rate) detection lists.

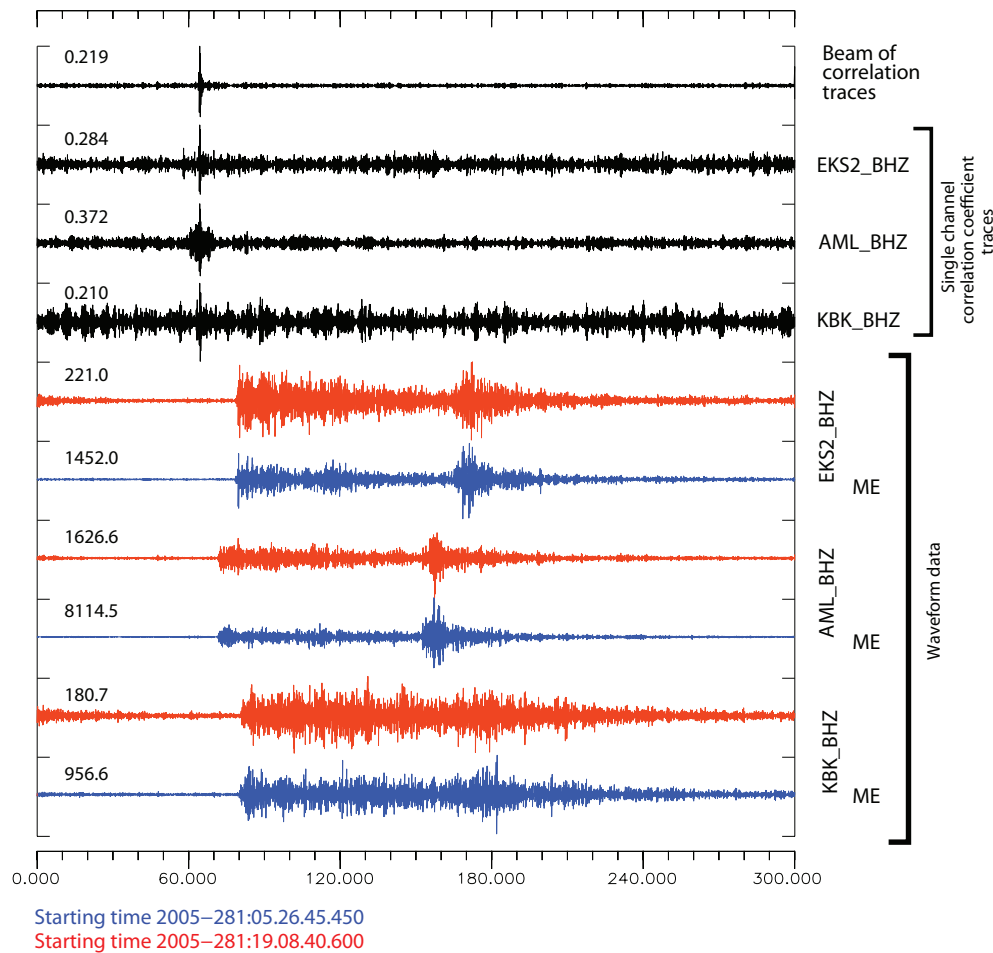
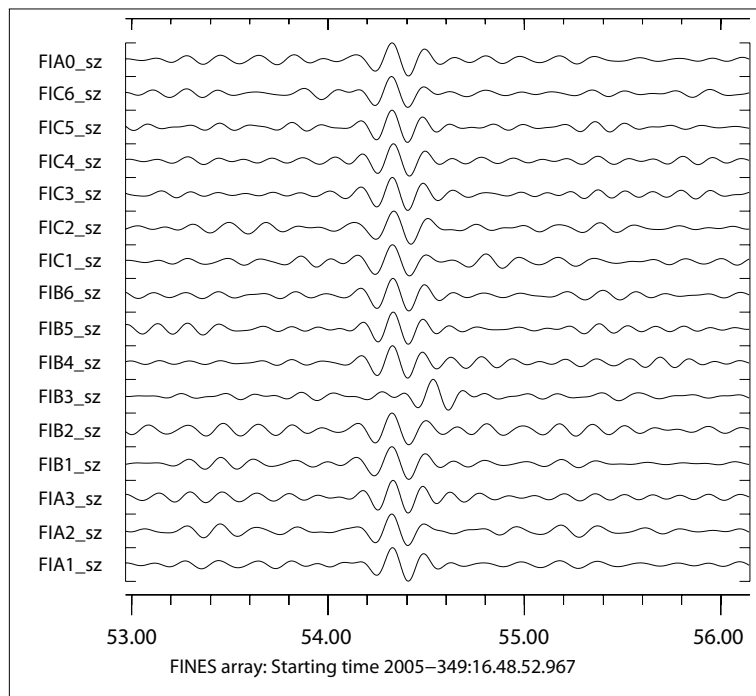


Figure 4. A positive correlation detection made over the KN-network. All waveforms are filtered between 2.5 and 8.0 Hz and the waveform template has a length of 180.0 seconds beginning at a time 2005-281:05.27.50.000. The USGS/NEIC event bulletin lists the master event as origin time 2005-281:05.26.05.120, latitude 34.760°N, longitude 73.150°E, depth 10.0 km and magnitude 5.5. The same bulletin lists the detected event with origin time 2005-281:19.08.00.490, latitude 34.800°N, longitude 73.170°E, depth 10.0 km and magnitude 4.9.

If a correlation-based rapid categorization scheme for major earthquakes and aftershock sequences is to be successful it is likely to have to consider a large number of events at a large range of magnitudes and using stations at different epicentral distances. The regional stations are likely to be necessary for identification and clustering analysis, and teleseismic stations at different azimuths might be useful in making preliminary, automatic judgments about whether events are likely to belong to the source region of interest.

The Identification of Erroneous Timing over Arrays from Correlation Detections

Rubin (2002) demonstrated how inconsistency between small errors inferred from high-accuracy cross-correlation delay-time measurements and large errors in relative event location estimates can be indicative of erroneous instrumental timing on one or more stations in a network. Given a single source of repeating seismic events, and two stations which record waveforms independently from all events, a misalignment of cross-correlation traces provides an immediate indication of an instrumental timing error at one, or both, of the stations. If such events occur sufficiently frequently over an extended period, well-determined time-dependent station corrections can be calculated using a limited number of “snapshot” measurements (Gibbons, 2006). Such methods are especially useful for sparse networks and 3-component stations since methods which exploit the coherence of microseismic noise (Koch and Stammer, 2003) are not applicable. However, even on small-aperture arrays, the alignment information for each channel can and should be applied at each opportunity to test consistency of timing across the network. An example of erroneous timing on the FINES array is displayed in Figure 5.



The panel to the left shows correlation coefficient traces for the 16 short-period vertical channels of the FINES array over a short time-window where the master signal template is a 20-second-long data segment beginning at a time 2005-175:04.27.22.35000 (bandpass filtered between 2.5 and 8.0 hz). The maximum of the correlation beam occurs at a time 2005-349:16.48.54.32750. Details of the events are provided in Gibbons et al. (2007). The multi-channel cross-correlation method of VanDecar and Crosson (1990) allows a very accurate measurement of the relative timing anomaly on the channel FIB3_sz. However, since GPS-lock for this site was on for neither the master event nor the detected signals, it is impossible to calculate a correction from this data alone. The FIB3 element has been non-operational since January 2006.

Figure 5. Rapid identification of a timing anomaly on the IMS primary array FINES from the misalignment of correlation coefficient traces.

Diagnostics regarding the alignment of the correlation coefficient traces are as meaningful and easy to implement over a sparse network (e.g., Figure 4) as over a small aperture array (e.g., Figure 5).

Horizontal Relative Event Locations using Stacked Cross-Correlation Traces

Central to the theme of detection through waveform similarity, is the question of spatial separation of events and the ability to provide accurate relative locations (Richards et al., 2006). The stacking of correlation coefficient traces over a seismic array lowers the detection threshold and, for the same reasons, provides improved cross-correlation time estimates, especially in cases where a single station observation would be insufficient. If an event cluster is recorded by a network of seismic arrays, there may be a possibility that events of lower magnitude may be included in inversions for relative location estimates using double-difference (DD) type algorithms. The events described in Gibbons et al. (2007) were recorded by both array stations at regional distances and 3-component stations at local distances. The local network recordings were unfortunately too incomplete for a full double-difference relative relocation. However, a stacking of correlation coefficient traces for the Lg-phase over each of the seismic arrays, together with an application of the Schaff and Richards (2004) DD-algorithm, allowed for a stable solution (Figure 6) - consistent with the existing local recordings - despite the low signal-to-noise ratio of the events at regional distances.

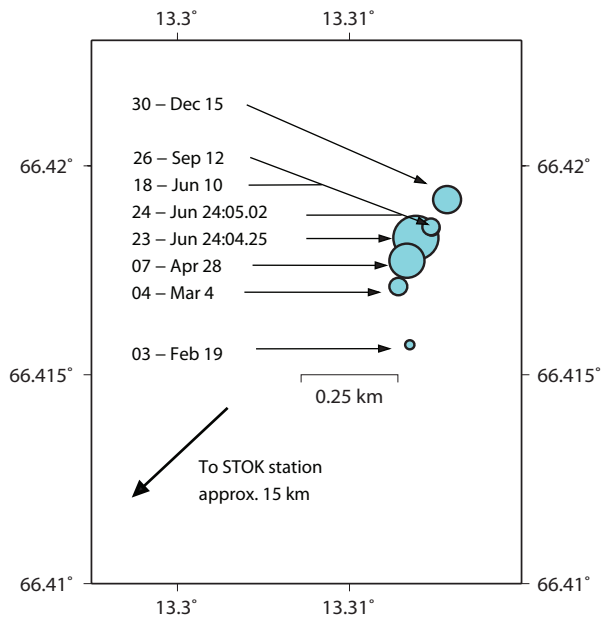


Figure 6. Relative location estimates obtained for 8 events in the Rana region of northern Norway using the Lg double-difference formulation of Schaff and Richards (2004) based upon delay times indicated by correlation coefficient beams at 4 seismic arrays at distances over 600 km. Details provided in Gibbons et al. (2007).

Gibbons et al. (2007) describe a sequence of events in the north of Norway between March and December 2005, ranging from magnitudes 0.5 to 3.5, which were all detected by correlation with a waveform template from the largest of the events on the large aperture NORSAR array at a distance of approximately 600 km. Observations from a local station (STOK at Stokkvågen) aligned by correlation of the large amplitude S-arrival indicate a small difference in S-P times (~0.05 secs) but with uncertainty due to the emergent and low-amplitude P-arrivals. However, a number of the events were also recorded by three other seismic arrays at regional distances at which the largest amplitude part of the wavetrain is the Lg-phase. Schaff and Richards (2004) observed that the Lg phase from many events in China was remarkably consistent in facilitating a double-difference horizontal relative location based upon differential delays for this phase from a number of regional stations. This procedure could be applied in this case to quite low magnitude events using time delays on the correlation stacks.

CONCLUSIONS AND RECOMMENDATIONS

We have explored the use of multi-channel waveform correlation detectors in a number of situations that present a number of challenges primarily due to waveform dissimilarity. Two case studies have been introduced in the current paper. The first case study is a sequence of presumed military explosions in northwestern Russia. Whilst the source location for these is poorly constrained, it is assumed that they take place within a reasonably compact region and that the primary reason for the waveform dissimilarity is differences in the source-time functions. This is supported by the sequence of multiple phase arrivals recorded by the ARCES array, which result in a characteristic large spread in automatic event location estimates determined by spurious phase associations. A template extracted from the simplest waveform was correlated against several years of ARCES data. Typical low correlation coefficient values between events that were evidently related resulted in an aggressively low detection threshold. Accepting triggers based upon the correlation beam value alone resulted in an excessive number of detections of which the vast majority were clearly spurious. Applying multi-variate selection criteria, with emphasis upon the alignment of the correlation coefficient traces at distinct sites, reduced this number dramatically. In a four year period, 243 detections passed the specified tests of which 220 corresponded with events which could be located manually close to the assumed

explosion site. All of the evident false alarms from this particular case-study were the result of high-amplitude local Rg phases from a similar direction.

The second case study is the extensive aftershock sequence from the magnitude 7.7 earthquake in Pakistan on October 8, 2005. Such sequences present a huge challenge to the nuclear explosion monitoring effort, due both to the increased noise level the global network is subjected to and the increased demand on analyst resources in locating and categorizing all events. A fully automatic event categorization based upon waveform similarity is therefore highly desirable, but is difficult due to the large magnitude of the main shock (large rupture length, source spectrum which will vary significantly from that of small aftershocks) and the wide geographical distribution of aftershocks (will clearly require a very large number of waveform templates to have a hope of covering with a correlation or subspace detector). A number of very clear correlation detections were made using data at teleseismic distances (the NORSAR and ARCES seismic arrays) and at regional distances (the KNET stations at a distance of approximately 900 km). In an example displayed, the master event has estimated magnitude of 5.5, substantially higher than any considered so far in this project (e.g., Gibbons and Ringdal, 2006; Stevens et al., 2006; Gibbons et al., 2007). It may be that a sequence of master events covering a cascade of different magnitudes may be necessary to bridge the spectral differences and geographical coverage involved. Application of the correlation trace beamforming method over the KNET stations represents a large increase in the array aperture that has been applied in the current project, and its success has been mirrored in recent work by other authors (e.g., Shelly et al., 2007).

Repeating seismic events facilitate “spot-checks” on instrumental timing, which can be applied to an array or network of arbitrary dimensions since coherence between sensors is not required. On small aperture arrays, where timing verification using coherent low-frequency microseisms is possible, correlation trace alignment testing can be used to validate and confirm results obtained using alternative methods. In many cases, the accuracy obtained measuring a moveout in a sharp correlation peak calculated from a long wavetrain might be higher than that possible by correlating transient coherent wavelets over pairs of channels. This is additional motivation for the identification of repeating seismic sources. In the current paper and in Gibbons et al. (2007) we demonstrate erroneous instrumental timing on three different IMS array stations.

The presence of networks of seismic arrays may facilitate the accurate relative relocation of seismic events of smaller magnitude than would be possible with networks of single stations. Provided that the interstation distance is short compared with the epicentral distance, the cross-correlation traces from several sensors can be stacked to measure a more stable time difference between two corresponding phases in a given direction from the source. If such measurements are available at enough sites, we can apply double-difference type techniques (Waldhauser and Ellsworth, 2000) to event relocation. We measured time-differences for the Lg-phase for earthquakes in the Rana region of northern Norway at four IMS seismic arrays and applied the horizontal DD-relocation procedure of Schaff and Richards (2004) to obtain a stable relative event location solution that is consistent with limited recordings at local distances.

ACKNOWLEDGEMENTS

We are grateful to the operators of the KN network (Kyrgyz Seismic Telemetry Network: Kyrgyz Institute of Seismology, IVTAN/KIS) and the IRIS data exchange center for making this data available.

REFERENCES

- Geller, R. J. and Mueller, C. S. (1980). Four Similar Earthquakes in Central California, in *Geophys. Res. Lett.* 7: 821–824.
- Gibbons, S. J. (2006), On the Identification and Documentation of Timing Errors: An Example at the KBS station, Spitsbergen, *Seism. Res. Lett.* 77: 559–571.

- Gibbons, S. J., Böttger Sørensen, M., Harris, D. B. and Ringdal, F. (2007). The detection and location of low magnitude earthquakes in northern Norway using multichannel waveform correlation at regional distances, in *Phys. Earth Planet. Int.* 160: 285–309.
- Gibbons, S. J. and Ringdal, F. (2006). The detection of low magnitude seismic events using array-based waveform correlation, in *Geophys. J. Int.* 165: 149–166.
- Koch, K. and Stammer, K. (2003). Detection and Elimination of Time Synchronization Problems for the GERESS Array by Correlating Microseismic Noise, *Seism. Res. Lett.* 74: 803–816.
- Ringdal, F. and Kværna, T. (1989). A multi-channel processing approach to real time network detection, phase association, and threshold monitoring, in *Bull. Seism. Soc. Am.* 79: 1927–1940.
- Ringdal, F., Gibbons, S. J., Kværna, T., Kremenetskaya, E., Asming, V., Vinogradov, V., Mykkeltveit, S. and Schweitzer, J. (2005). Research in regional seismic monitoring, in *Proceedings of the 27th Seismic Research Review: Ground-based Nuclear Explosion Monitoring Technologies*, LA-UR-05-6407, Vol. 1, pp 423–432.
- Richards, P. G., Waldhauser, F., Schaff, D. P., and Kim, W.-Y. (2006). The Applicability of Modern Methods of Earthquake Location, in *Pure. Appl. Geophys.* 163: 351–372.
- Rubin, A. M. (2002). Using Repeating Earthquakes to Correct High-Precision Earthquake Catalogs for Time-Dependent Station Delays, in *Bull. Seism. Soc. Am.* 92: 1647–1659.
- Schaff, D. P. and Richards, P. G. (2004). Lg-wave cross correlation and double-difference location: application to the 1999 Xiuyan, China, sequence, in *Bull. Seism. Soc. Am.* 94: 867–879.
- Shelly, D. R., Beroza, G. C., and Ide, S. (2007). Non-volcanic tremor and low-frequency earthquake swarms. *Nature* 446: 305–307.
- Stevens, J. L., Gibbons, S., Rimer, N., Xu, H., Lindholm, C., Ringdal, F., Kværna, T., Murphy, J. R. (2006). Analysis and simulation of chemical explosions in nonspherical cavities in granite, in *J. Geophys. Res.* 111: B04306, doi:10.1029/2005JB003768.
- VanDecar, J. C. and Crosson, R. S. (1990). Determination of teleseismic relative phase arrival times using multi-channel cross-correlation and least squares, in *Bull. Seism. Soc. Am.* 80: 150–169.
- Waldhauser, F. and Ellsworth, W. L. (2000). A double-difference earthquake location algorithm: method and application to the Northern Hayward Fault, California, in *Bull. Seism. Soc. Am.* 90: 1353–1368.