A MATCHED FIELD PROCESSING FRAMEWORK FOR COHERENT DETECTION OVER LOCAL AND REGIONAL NETWORKS

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. FA9453-10-C-0209[1,2] and LL10-BAA10-20-NDD03[3]
Proposal No. BAA10-20

ABSTRACT

The objective of this one-year feasibility study is to develop a data-adaptive matched field procedure to detect coherently and incoherently across networks of stations at local and regional distances. The detector will extend the current matched field processing approach, which classifies events on the basis of observations of first-arriving P-waves, to detection using the entire waveform. The framework is based upon a narrowband signal representation that we believe exposes invariant spatial and temporal correlation structure of network signals from repeating sources.

During the first months of the contract period, we have implemented a prototype matched field (MF) algorithm for detection processing using the full event waveforms. We have initially tested and validated the algorithm with regional events on the Kola Peninsula, for which good ground truth information is available and where signals from repeating sources are observed by network stations and arrays at distances of interest. We will continue testing of the prototype matched field detector on larger datasets from the Kola Peninsula and assess the performance against ground truth information and benchmarks set by correlation detectors and empirical matched field processing event classification studies. This algorithm will become an important part of a general detection framework to estimate autonomously the correlation structure of signals from repeating sources while processing continuous data streams. The framework will be tested for its ability to derive detectors adapted to the correlation structure of repeating sources using networks in Fennoscandia and the Korean peninsula.

In parallel, we have worked on the preparation of a database of continuous waveforms from the network of stations on and surrounding the Korean peninsula. We have obtained seismic event bulletins from KIGAM (the Korea Institute of Geology, Mining and Materials) in addition to the International Data Centre (IDC) Reviewed Event Bulletin, the preliminary determination of epicenters (PDE) bulletin from the USGS, and the bulletin of the International Seismological Center (ISC). However, it is almost guaranteed that the majority of small, repeating, industrial events will not be present in these event lists. We have therefore performed a single-array detection procedure on the KSRS teleseismic array in South Korea targeted at identifying regional events from repeating sources. The main purpose of this is to identify sources of repeating seismicity against which the new algorithms can be evaluated. The initial power detection procedure applies an incoherent detection scheme which calculates continuous spectral estimates of single channels and then forms beams of transformed spectrograms. This method has the advantage that it detects phases which are rich in high frequency energy, at which coherent processing on the KSRS array fails. Coherent processing in the 2-4 Hz band does usually provide fairly robust and accurate estimates of the slowness, even though the signal-to-noise ratio (SNR) in this frequency band is usually too low to provide a good detector with a low false alarm rate. This procedure of incoherent detection and coherent parameter estimation appears to provide an extensive bulletin of well-defined events at regional distances which fall into clusters. The detections which do belong to the same groupings are currently being clustered using the standard multi-channel correlation detectors, followed by f-k analysis post-processing to further validate the detections.

OBJECTIVES

The objective of this one-year feasibility study is to develop a data-adaptive matched field procedure to detect coherently and incoherently across networks of stations at local and regional distances:

- The detector will extend the current matched field processing approach, which classifies events on the basis of observations of first-arriving P-waves, to detection using the entire waveform.
- The framework will be based upon a narrowband signal representation that exposes invariant spatial and temporal correlation structure of network signals from repeating sources, even in the presence of source-time history variation. It is anticipated that familiar waveform correlation detectors and alternative temporally or spatially incoherent detectors will emerge as special cases of the proposed architecture.
- The specific form of the detector that emerges will depend upon the details of correlation structure present in the signal data. The architecture will be tested initially against events on the Kola Peninsula, for which good Ground Truth information is available and where signals from repeating sources are observed by network stations and arrays at distances of interest.

Once the basic detection algorithm has been developed and tested, a previously developed autonomous calibration framework will be adapted to drive the detector. The new detector will replace correlation and subspace detectors currently employed in that framework. As the framework operates and collects events from sources of repeating seismicity, the general network covariance will be estimated and used to update individual detectors to match the observed correlation characteristics of signals they are designed to detect. As more events are detected, the covariance estimates will be updated and used to adapt detectors to be more finely tuned to their targets. Once the framework is complete, it will be applied to the Korean Peninsula. The framework will be applied using data from several different time periods, mainly during time intervals when existing bulletins contain high numbers of seismic events in the regions of interest. Data intervals surrounding the known nuclear tests will be included.

RESEARCH ACCOMPLISHED

Types of sources and their bearing upon detector architecture

Correlation, subspace and matched field detectors offer advantages of great sensitivity at a low false alarm rate and the ability to wrap detection, association, location and identification functions into a single operation. These advantages are a consequence of the fact that the detectors are highly tuned to the characteristics of one particular source. The premise of this project is that repeated observations of waveforms from a given source can be used automatically to tune a detector to match the spatial and the statistical temporal structure of the waveforms generated by the source. Figure 1 illustrates two sources with substantially different temporal statistical structure. The top set of traces depicts waveforms from 10 consecutive explosions at a munitions demolition site in Finland observed at the ARCES array center element. The salient feature of these waveforms is that they are nearly identical. A waveform correlator (subspace detector of rank one) is the optimum detector for events of this source. The bottom set of traces shows waveforms from 10 consecutive explosions at the underground Kirovsk mine. The waveforms of this set vary considerably, so that the optimum detector is a higher-rank subspace detector or a matched field detector.

We seek a single model for both sets of observations, and, hence, a single framework for optimized detectors of both source types. Because the sources are approximately point sources, we assume that a fixed set of Greens functions describe propagation from each source to the observing array. The Greens functions are deterministic, but unknown, though subject to calibration. We assume the variation appears in the source time history and mechanisms of the sources, which we consider to be stochastic. In the case of the Finnish explosions, source variance is very low and the source is nearly deterministic. In the case of the Kirovsk explosions, source variance is substantially higher.

Matched field detectors are designed for sources like the Kirovsk mine, as they are insensitive to variations in source time history. They operate on a narrowband signal representation, where the bandwidth is chosen to be narrow enough to wash out the effects of the source forcing functions. For a sufficiently narrow band centered on frequency ω_i, the vector of signals $\mathbf{r}_i(t)$ observed across an array or network is, approximately:

$$\mathbf{r}_i(t) \approx F(\omega_i)\mathbf{g}(t,\omega_i) \tag{1}$$

This model assumes that the source forcing function $f(t)$ has a very short duration $\Delta t \ll {}^{1}/_{\Delta\omega}$, where $\Delta\omega$ is the width of the narrow bands. Here $\mathbf{g}(t,\omega_i)$ is the vector of Greens functions describing propagation to the collection of sensors from the source, also filtered into the band centered on ω_i, and $F(\omega_i)$ is the Fourier transform of $f(t)$. Matched field detectors operate by correlating a template with the signal in each frequency band, then combining the result over frequency bands. The statistical (covariance) structure of the source waveforms governs the combination over bands: incoherent in the case of a source like the Kirovsk mine, coherent in the case of a source like the Finnish demolition site.

The templates are obtained from an eigendecomposition of the covariance function for the observations:

$$\mathbf{C}_{\mathbf{ij}}(t_1,t_2) = E\{\mathbf{r}_i(t)\mathbf{r}_j^H(t)\} \approx \mathbf{g}(t,\omega_i)E\{F(\omega_i)F^*(\omega_j)\}\mathbf{g}^H(t,\omega_j) \tag{2}$$

The essential feature of this covariance function is the behavior of different frequency components for the two source types. For the Kirovsk source, we anticipate independence among frequency bands:

$$E\{F(\omega_i)F^*(\omega_j)\} = \delta_{ij}\, E\{|F(\omega_i)|^2\} \tag{3}$$

i.e., $E\{F(\omega_i)F^*(\omega_j)\} = 0$ whenever $i \neq j$. For the Finnish demolition source, we anticipate high correlation among different frequency bands, i.e. $E\{F(\omega_i)F^*(\omega_j)\} \neq 0$.

The form of the detector requires some explanation. First, we restructure the narrowband signals into a single vector:

$$\mathbf{r}(t) = \begin{bmatrix} \mathbf{r}_1(t) \\ \vdots \\ \mathbf{r}_i(t) \\ \vdots \end{bmatrix} \tag{4}$$

and the covariance function into a matrix:

$$\mathbf{C}(t_1,t_2) = \begin{bmatrix} \mathbf{C_{11}}(t_1,t_2) & \mathbf{C_{12}}(t_1,t_2) & \cdots \\ \mathbf{C_{21}}(t_1,t_2) & \mathbf{C_{22}}(t_1,t_2) & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \tag{5}$$

The covariance matrix has a Karhunen-Loeve representation:

$$\mathbf{C}(t_1,t_2) = \sum_p \lambda_p \boldsymbol{\phi}_p(t_1)\ \boldsymbol{\phi}_p^H(t_2) \tag{6}$$

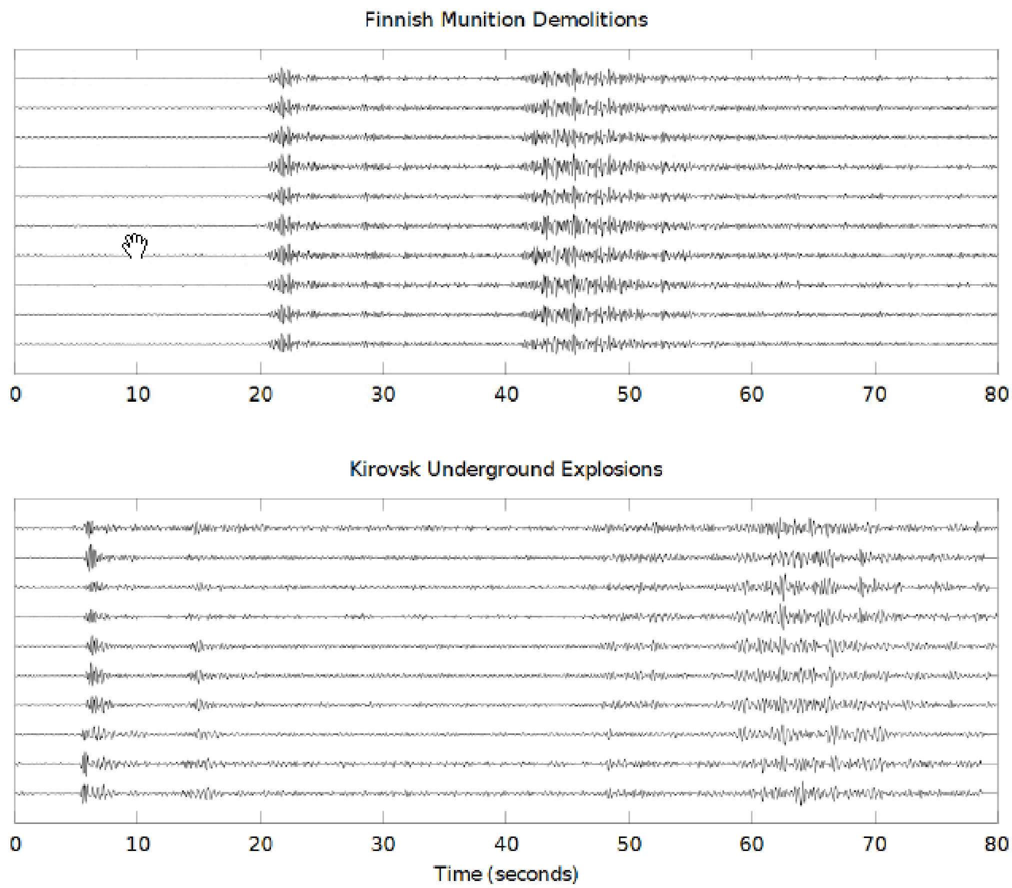

Figure 1. Representative waveforms from two different sources types: 10 consecutive explosions from a munitions demolition site in Finland and 10 consecutive explosions from the Kirovsk underground mine in the Khibiny Massif. Note the similarity of the munitions shots and the variation among the Kirovsk shots.

The detection statistic computed by the matched field detector is given by:

$$d(t) \; = \; \sum_{p} \left| \int \boldsymbol{\phi}_p^H(\tau)\mathbf{r}(t+\tau)d\tau \right|^{\mathbf{2}} \tag{7}$$

In the Kirovsk case, it is approximately true that any particular $\boldsymbol{\phi}_p$ will contain a contribution from only one frequency band. Consequently the contributions from different frequency bands will be summed incoherently. In the case of the Finnish demolition site, any particular $\boldsymbol{\phi}_p$ will contain contributions from many or all frequency bands and contributions from different frequency bands will be summed coherently in the detection statistic.

An automated system to design a detector appropriate to a particular source will begin by estimating the covariance matrix $\mathbf{C}(t_1, t_2)$ from many observations of waveforms from that source. The correlation kernel $\boldsymbol{\phi}_p$ will be estimated from the sample covariance matrix and will automatically assume the structure of a waveform correlator or an incoherent matched field detector depending upon the statistical characteristics of the source. Our intention is that, over time, as many observations are acquired, the detector will tune itself to the statistical characteristics of the source.

We plan to try two initialization strategies, starting with a single event observation for a particular source. The first strategy may begin with a correlation detector defined by the one available waveform and, by setting the detection threshold low, attempt to acquire more events to define the covariance matrix. This strategy obviously would work perfectly for the Finnish munitions explosions, but may fail for the Kirovsk mine shots with their considerable waveform variability. The second strategy will use the single observation to define the spatial structure of the signal across the array, but will force incoherence across frequency bands to suppress dependence on the source time history. This approach will provide a more general detector with the prospect of acquiring enough events to define the statistical structure of sources like the Kirovsk mine.

Implementing and testing a prototype Matched Field Detector

Priority during the first phase of this project has been given to implementation and testing of a prototype Matched Field (MF) detector. The prototype software is written in Matlab and provides a convenient tool for experiments. We have initially tested and validated the algorithm with regional events on the Kola Peninsula, for which good ground truth information is available and where signals from repeating sources are observed by network stations and arrays at distances of interest. In Figure 2 we show the result from processing of a 3-hour data interval using the new prototype MF detector software. Notice the peaks of the detection statistic associated with known blasts in the Kirovsk mine.

In order to facilitate further testing and experiments, we have implemented access to NORSAR's large database of continuous seismic data into the MF detector software. This is demonstrated in Figure 3 which shows results from processing of a recent 3-hour time interval from the ARCES array starting at 12 UTC on 10 April 2010. This first version of the MF detector software will be subjected to further validation and testing using ARCES recordings of mining events from the Kola Peninsula. In particular we will address aspects of the detector normalization, design of detector templates from event ensembles and threshold setting for triggering of events.

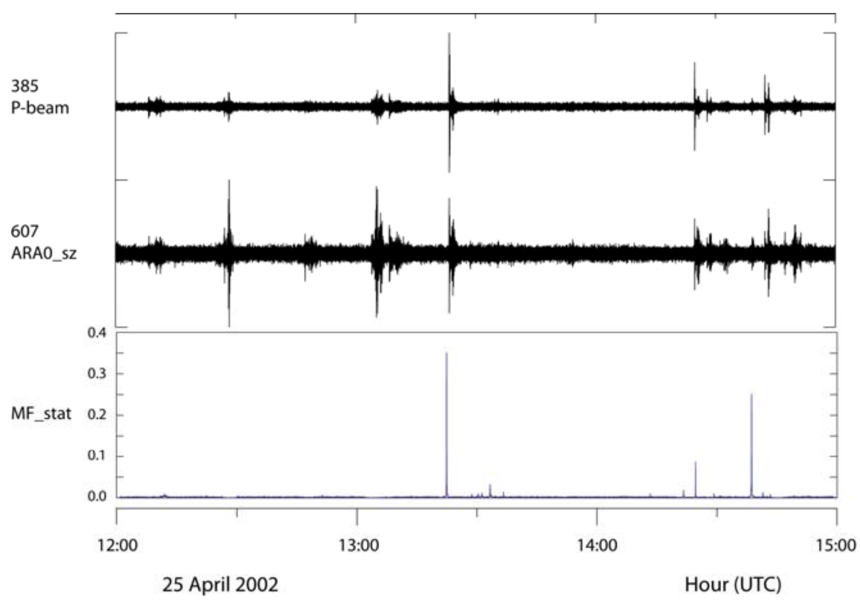

Figure 2. Result from the application of the prototype implementation of the full-waveform MF detector to three hours of data from the ARCES array on 25 April 2002. The detector was designed using observations of 114 explosions at the Kirovsk mine, which is one of the five closely spaced mines on the Khibiny Massif on the Kola Peninsula. The three largest peaks correspond to three Kirovsk explosions. The processing frequency band was 2.5 - 12.5 Hz, which the detector breaks into 33 narrow sub-bands.

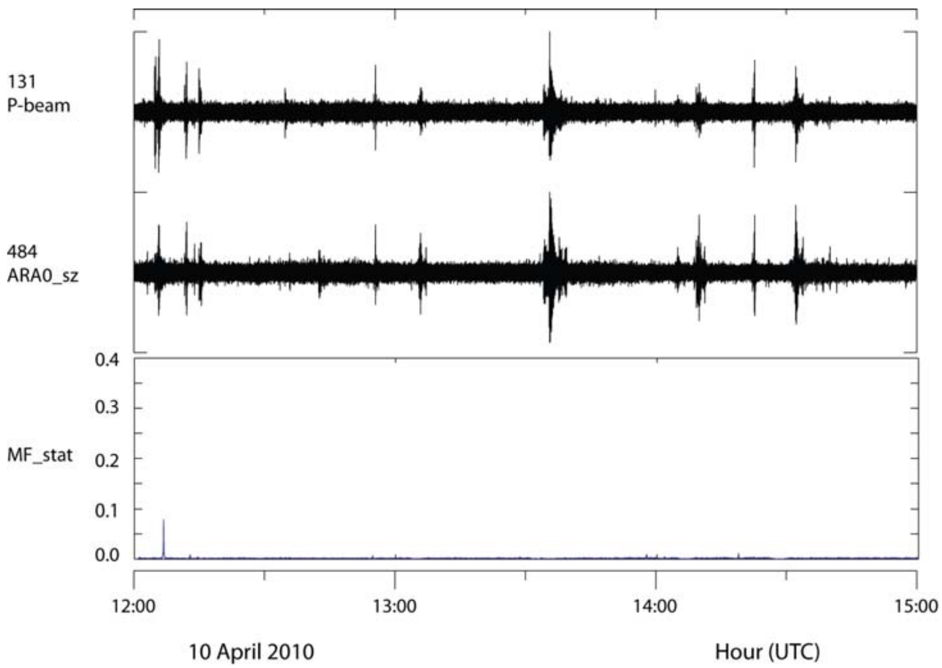

Figure 3. Result from the application of the new prototype MF detector to three hours of data from the ARCES array on 10 April 2010, again using the 114 explosions at the Kirovsk mine for designing of the detector. The peak just after 12 UTC corresponds to signals from a small event most likely located in one of the mines of Khibiny Massif. We do not currently have Ground Truth information for this event.

Preparation of a Database of Continuous Waveforms from the Network of Stations Surrounding North Korea

While the matched field detection architecture is currently being tested on datasets in the European Arctic, it is an important part of the current project to apply the algorithms to a region of programmatic interest: the Korean Peninsula. The closest two IRIS 3-component stations to the test site are INCN (Inchon, Republic of Korea, IRIS-USGS network) and MDJ (Mudanjiang, China, IC network). Data from these two stations have been archived back to January 1, 2006, although older archives may be accessed if this is deemed necessary. As the Norwegian NDC, NORSAR has access to data from the USRK, KSRS, and MJAR arrays from the IDC in Vienna. All current holdings from these stations have now been obtained. The geography of the stations in relation to the test site is displayed in Figure 4.

We have obtained seismic event bulletins from KIGAM in addition to the IDC Reviewed Event Bulletin, the PDE bulletin from the USGS, and the bulletin of the ISC. However, it is almost guaranteed that the majority of small, repeating, industrial events will not be present in these event lists.

We start by attempting to bootstrap lists of seismic events from repeating industrial sources.

Most of the signals from such events are characterized by high frequency regional phases. We begin with the KSRS array (since this has the longest history of continuous data), a teleseismic array over which signals are fairly incoherent above 4 Hz. This has the consequence that beamforming at the frequencies with the greatest SNR leads to substantial beam-loss, and that slowness estimates in these frequency bands are often qualitatively misleading. In order to detect the high frequency regional phases with a low false alarm rate, we use the spectrogram beamforming (incoherent) method described in Gibbons et al. (2008) - see Figure 5. Despite the typically low SNR in the 2-4 hz frequency band, the slowness estimates made in this band do appear to be fairly stable.

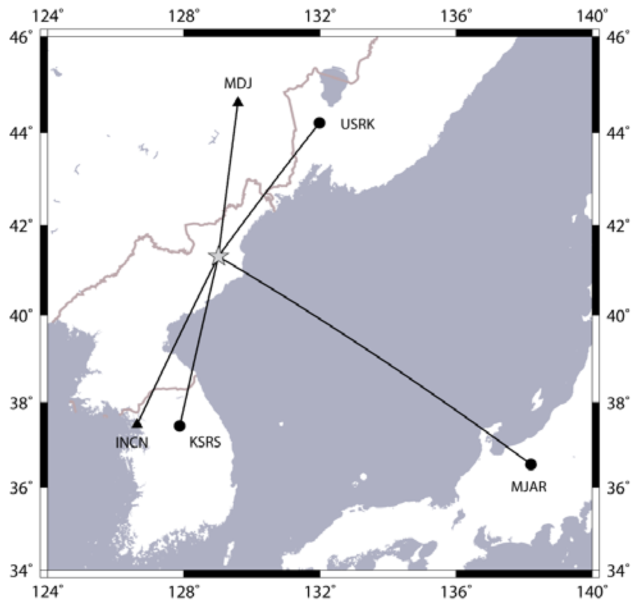

Figure 4.
Location of the IMS arrays KSRS (Wonju, South Korea), USRK (Ussirusk, Russian Federation), and MJAR (Matsushiro, Japan) and the IRIS 3-component stations INCN (IU network) and MDJ (IC network). The star denotes the location of the DPRK nuclear test site.

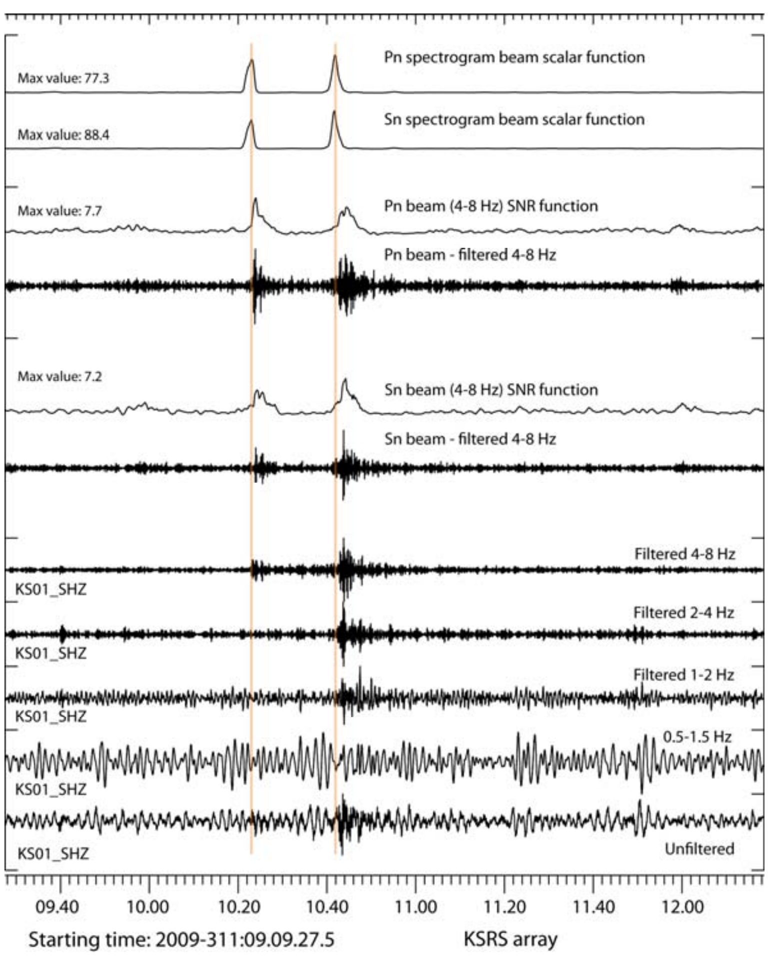

Figure 5.
An example of the detection of a P and S phase for an event within regional distances of the KSRS array. The only frequency band at which a reasonable SNR is attained is the 4-8 Hz band. The beamforming of transformed spectrograms described by Gibbons et al. (2008) provides a promising detection statistic for these phases, as shown in the upper two traces of the figure.

The detection lists obtained using this procedure are now complete from 2006 to the present day. A procedure of taking a single event, running a correlation detector, removing all of the power detections which are confidently associated with these correlation detections, and then spawning a new correlation detector from an event in the pool of remaining detections, is currently underway. We hope that, when completed, this will give quite a comprehensive list of sources of repeating seismicity within regional distances of KSRS. So far, a small number of clusters have been investigated using correlation detectors and four of these are displayed, together with the full set of automatic event locations, in Figure 6. In situations where multiple correlation detectors have common events, output from the different templates will be compared. This may indicate that events originate from a large source region and that several correlation templates are necessary to characterize the source. Such instances will be an interesting point of comparison for the new detection architecture.

Waveforms from events assigned to the four clusters displayed on the map are shown in Figure 7. It is clear that even these four clusters alone show significant variation in the both waveform amplitudes, SNR, and waveform semblance. Results of the detector and cluster analysis will be addressed in the next phase of the project.

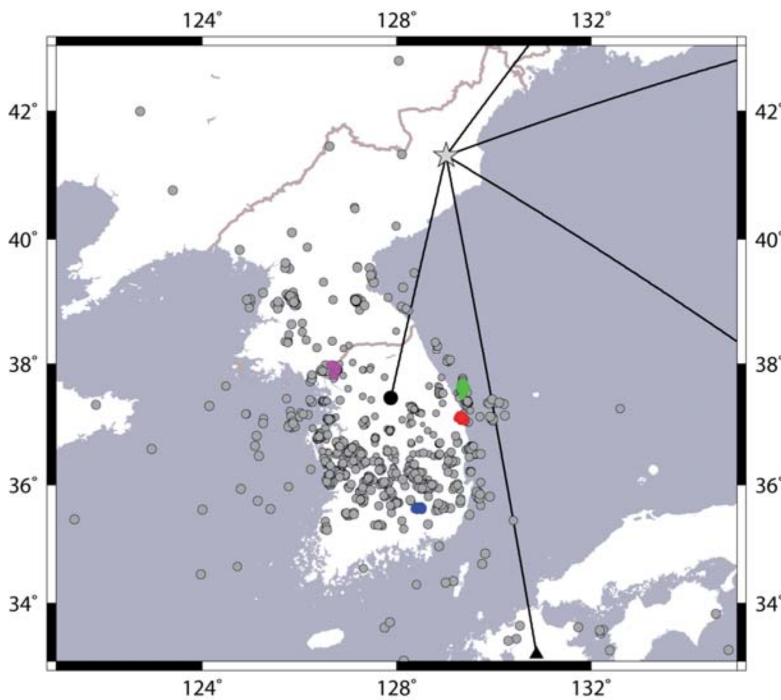

Figure 6. Fully automatic event location estimates using only regional P and S detections on the KSRS array (marked by the black circle at the center of the map). Given that the event locations at this stage are fully automatic, some of the events shown may be spurious associations. However, there are clearly many clusters of events and location estimates for four groups of events found to correlate are displayed in four different colors. The compactness of the automatic estimates within these event clusters gives some confidence about the robustness of the automatic phase detection and estimation procedure. The color-coded clusters on the map are associated with correlations with master events as follows:
red (2006-307:20.30.24, 37.100 N, 129.252 E), blue (2006-308:09.05.26, 35.610 N, 128.408 E),
green (2006-308:02.57.19, 37.593 N, 129.388 E), magenta (2006-310:03.07.22, 37.940 N, 126.794 E).

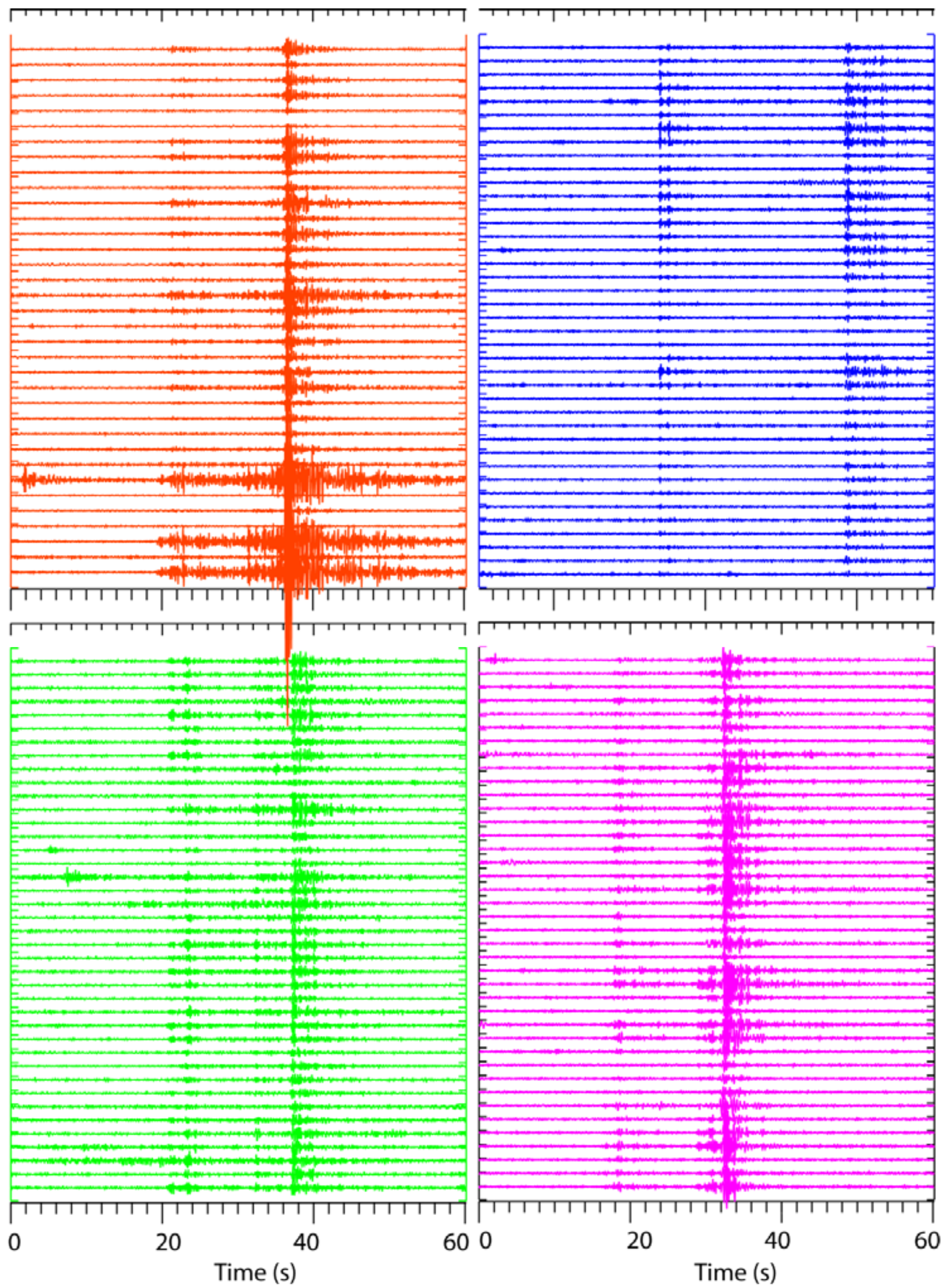

Figure 7. Waveforms from the KS01_SHZ instrument, aligned according to the times of best correlation coefficient for 40 events from each of the clusters displayed in Figure 6 in the appropriate color.

CONCLUSIONS AND RECOMMENDATIONS

In the development of the processing system, we plan to try two initialization strategies, starting with a single event observation for a particular source. The first strategy may begin with a correlation detector defined by the one available waveform and, by setting the detection threshold low, attempt to acquire more events to define the covariance matrix. The second strategy will use the single observation to define the spatial structure of the signal across the array, but will force incoherence across frequency bands to suppress dependence on the source time history. The applicability of these strategies is expected to depend on the degree of waveform variability of the events from the repeating source in question.

During the first few months of this project, we have achieved the following results:

We have evaluated a prototype matched field (MF) detector on ARCES data for well constrained mining events on the Kola Peninsula. The results are very promising, and the initial version of the MF detector will be subjected to further validation and testing using this data set. In particular we will address aspects of the detector normalization, design of detector templates from event ensembles and threshold setting for triggering of events.

We have collected seismic array and 3-component data for stations on or surrounding the Korean Peninsula. From the IDC of the CTBTO, we have obtained KSRS array data from November 2006 to the present day, USRK data from 2008 to the present day, and also MJAR data (Matsushiro array, Japan) from 2005 to the present day. The IRIS 3-component stations MDJ (IC network) and INCN (IU network) have been obtained from the IRIS Data Management Center for the years 2006 to the present day (older data is available in the archives and may be collected should this be deemed to be of interest).

We have performed a single-array detection procedure on the KSRS teleseismic array in South Korea targeted at identifying regional events from repeating sources. The main purpose of this is to identify sources of repeating seismicity against which the new algorithms can be evaluated. The initial power detection procedure applies the method of Gibbons et al. (2008)—an incoherent detection scheme which works by calculating continuous spectral estimates of single channels and then forming beams of transformed spectrograms. This method (which has been used in operational mode on the NORSAR array) has the advantage that it detects phases which are rich in high frequency energy, at which coherent processing on the array fails. Coherent processing in the 2-4 Hz band does usually provide fairly robust and accurate estimates of the slowness, even though the SNR in this frequency band is usually too low to provide a good detector with a low false alarm rate. This procedure of incoherent detection and coherent parameter estimation appears to provide an extensive bulletin of well-defined events at regional distances which fall into clusters. The detections which do belong to the same groupings are currently being clustered using the standard multi-channel correlation detectors with f-k analysis post-processing (Gibbons and Ringdal, 2006).

ACKNOWLEDGEMENTS

We are grateful to the IRIS DMC for data from the INCN and MDJ 3-component stations which belong to the networks IU and IC respectively. Array data was obtained from the IDC of the Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO) in Vienna, Austria. All maps have been produced using GMT software (Wessel and Smith, 1995).

REFERENCES

Harris, D. B. and T. Kvaerna (2010). Superresolution with seismic arrays using empirical matched field processing, accepted for publication, *Geophys. Jour. International.*

Harris, D. B. and D. Dodge (2010). An autonomous system for grouping events in a developing aftershock sequence, accepted for publication *Bull. Seismol. Soc. Am.*

Gibbons, S. J. and F. Ringdal (2006). The detection of low magnitude seismic events using array-based waveform correlation, *Geophys. J. Int*. 165: 149–166.

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.

Wessel, P. and Smith, W. H. F. (1995). New version of the generic mapping tools, *EOS Trans., Am. geophys. Union*., 76: 329.