EXPANDING COHERENT ARRAY PROCESSING TO LARGER APERTURES USING EMPIRICAL MATCHED FIELD PROCESSING

Frode Ringdal[1], David B. Harris[2,3], Tormod Kværna[1], Steven J. Gibbons[1]

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

Sponsored by the Air Force Research Laboratory and the National Nuclear Security Administration
Contract Nos. FA8718-08-C-0007[1] and DE-AC52-07NA27344[2,3]
Proposal No. BAA08-39

ABSTRACT

We have adapted matched field processing—a method developed in underwater acoustics to detect and locate targets— to classify transient seismic signals arising from mining explosions. Matched field processing, as we apply it, is an empirical technique, using observations of historic events to calibrate the amplitude and phase structure of wavefields incident upon an array aperture for particular repeating sources. The objective of this project is to determine how broadly applicable the method is and to understand the phenomena that control its performance.

We obtained our original results in distinguishing events from ten mines in the Khibiny and Olenegorsk mining districts of the Kola Peninsula, for which we had exceptional ground truth information. In a cross-validation test, some 98.2% of 549 explosions were correctly classified by originating mine using just the Pn observations (2.5-12.5 Hz) on the ARCES array at ranges from 350 – 410 kilometers. These results were achieved despite the fact that the mines are as closely spaced as 3 kilometers. Such classification performance is significantly better than predicted by the Rayleigh limit. Scattering phenomena account for the increased resolution, as we make clear in an analysis of the information carrying capacity of Pn under two alternative propagation scenarios: free-space propagation and propagation with realistic (actually measured) spatial covariance structure. The increase in information capacity over a wide band is captured by the matched field calibrations and used to separate explosions from very closely-spaced sources. In part, the improvement occurs because the calibrations enable coherent processing at frequencies above those normally considered coherent.

We are investigating whether similar results can be expected in different regions, with apertures of increasing scale and for diffuse seismicity. We verified similar performance with the closely-spaced Zapolyarni mines, though discovered that it may be necessary to divide event populations from a single mine into identifiable subpopulations. For this purpose, we perform cluster analysis using matched field statistics calculated on pairs of individual events as a distance metric. In our initial work, calibrations were derived from ensembles of events ranging in number to more than 100. We are considering the performance now of matched field calibrations derived with many fewer events (even, as mentioned, individual events). Since these are high-variance estimates, we are testing the use of cross-channel, multitaper, spectral estimation methods to reduce the variance of calibrations and detection statistics derived from single-event observations.

To test the applicability of the technique in a different tectonic region, we have obtained four years of continuous data from 4 Kazakh arrays and are extracting large numbers of event segments. Our initial results using 132 mining explosions recorded by the Makanchi array are similar to those obtained in the European Arctic. Matched field processing clearly separates the explosions from three closely-spaced mines located approximately 400 kilometers from the array, again using waveforms in a band (6-10 Hz) normally considered incoherent for this array. Having reproduced ARCES-type performance with another small aperture array, we have two additional objectives for matched field processing. We will attempt to extend matched field processing to larger apertures: a 200 km aperture (the KNET) and, if data permit, to an aperture comprised of several Kazakh arrays. We also will investigate the potential of developing matched field processing to roughly locate and classify natural seismicity, which is more diffuse than the concentrated sources of mining explosions that we have investigated to date.

OBJECTIVE

The objectives of this project are

- To investigate the limits of empirical matched field processing and other coherent array detection and parameter estimation methods as receiver aperture size increases from a few kilometers to many hundreds of kilometers.
- To investigate techniques for extending the geographical source-region footprint over which empirical matched field processing and other coherent calibrated methods apply.

We will begin by reanalyzing data from the European Arctic in order to reconfirm the potential of empirical matched field processing that has been indicated by our work under a previous contract. We then will proceed to study the central Asia region to assure programmatic relevance and to exploit the large belts of natural and man-made seismicity to test the applicability of the technique to diffuse seismicity.

RESEARCH ACCOMPLISHED

The effort was focused initially on examining the degree to which empirical matched field processing (EMPF) can classify signals recorded on the ARCES array in northern Norway from closely spaced mining explosions. There were many good reasons for testing the new algorithms on this data set. Through close collaboration with scientists at the Kola Regional Seismological Center in Apatity, and with the operators of numerous mines on the Kola Peninsula, a superb database of Ground Truth information has been accumulated (Harris et al., 2003), consisting of many hundreds of events from a period of several years. In addition, the parameter estimation capabilities of the ARCES array for such signals have been explored in depth. The ability of the array to resolve source location using the classical array-processing paradigm is relatively well understood, providing an excellent basis for comparison with the new technology. The panel to the left of Figure 1 shows the location of the sources with respect to ARCES.

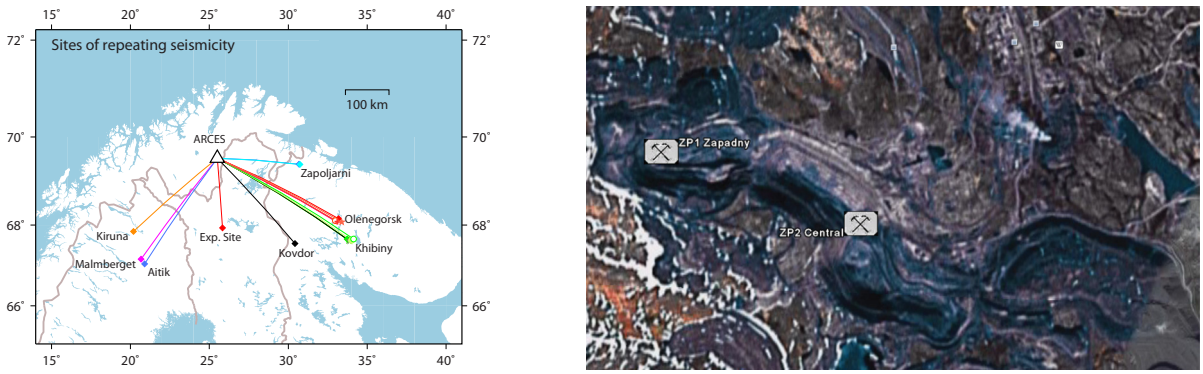

Figure 1. Locations of sites of industrial and repeating seismicity in relation to the ARCES array (left) and Google Earth image of the Zapoljarni open cast mining region with the ZP1 (Zapadny) and ZP2 (Central) mines labelled. The mines are separated by approximately 2.5 km and are approximately 200 km from ARCES.

The new data from Kazakhstan and Kyrgyzstan has allowed the application of the same techniques to a different tectonic setting, of programmatic relevance, with high levels of natural and artificial seismicity. Three different apertures of sensor configuration, with which the ability to expand Coherent Array Processing to greater sensor separations can be examined, are present. At the smallest scale, the four single arrays in Kazakhstan can be processed in the same way as ARCES. These arrays are however somewhat larger and far more sparse than ARCES which will have consequences for waveform semblance, in particular at the higher frequencies. At an intermediate scale, the stations of KNET constitute a network with an aperture of the order 200 km and a relatively broad spectrum of sensor separations. The largest aperture to be considered is that formed by the four Kazakh arrays as a single network.

Ground Truth of the kind which we have exploited in our examination of industrial seismicity on the Kola Peninsula is not yet available for Kazakhstan. This will complicate the evaluation of the event classification attempted under the current contract. We have however, been assisted greatly by access to classification results obtained by MacCarthy et

al. (2008) and Hartse et al. (2008). These studies are aimed at event screening and source identification using, in particular, waveform similarity techniques and satellite imagery. Their original source of information is the bulletin of the Kazakh National Data Center (KNDC).

Detected seismic phases are usually characterized by the parameters v_{app} (apparent velocity) and Θ (backazimuth) for the purpose of phase identification and association, and subsequent event location. In the estimation process, these parameters are coded in the slowness vector, $\boldsymbol{s}$, defined by

$$\boldsymbol{s} = (s_x, s_y) \quad \text{where} \quad s_x = s\sin(\Theta)\ , \quad s_y = s\cos(\Theta) \quad \text{and} \quad s = 1/v_{\mathrm{app}}\ . \tag{1}$$

If $\boldsymbol{x}_j$ denotes the coordinates of sensor *j*, relative to the array reference site, then we denote the time-series recorded at this site by $r(t, \boldsymbol{x}_j)$.

For a given frequency ω, the steering vector $\boldsymbol{\varepsilon}(\omega, \boldsymbol{s}) = \left[e^{-i\omega \boldsymbol{s}.\boldsymbol{x}_1} \quad \ldots \quad e^{-i\omega \boldsymbol{s}.\boldsymbol{x}_N}\right]^T$ allows a measurement of the energy incident on the array consistent with the plane wavefront hypothesis, $\boldsymbol{s}$, using (e.g. Capon, 1969)

$$P(\omega, \boldsymbol{s}) = \left| \int \sum_{j=1}^{N} r(t, \boldsymbol{x}_j) e^{i\omega(t - \boldsymbol{s}.\boldsymbol{x}_j)} \right|^2 = \boldsymbol{\varepsilon}(\omega, \boldsymbol{s})^H \boldsymbol{R}(\omega) \boldsymbol{\varepsilon}(\omega, \boldsymbol{s}) \tag{2}$$

where *N* denotes the number of sensors in the array, *H* denotes the Hermitian transpose, and the elements of the spatial covariance matrix, $\boldsymbol{R}(\omega)$, are given by

$$R_{mn}(\omega) = \left(\int r(t, \boldsymbol{x}_m) e^{-i\omega t} dt \right) \left(\int r(t, \boldsymbol{x}_n) e^{i\omega t} dt \right). \tag{3}$$

When estimating the slowness vector, $\boldsymbol{s}$, which best fits the incoming wavefront, it is typical to evaluate a normalized relative beam-power function of the form

$$\hat{P}(\boldsymbol{s}) = \frac{\sum_j \boldsymbol{\varepsilon}(\omega_j, \boldsymbol{s})^H \boldsymbol{R}(\omega_j) \boldsymbol{\varepsilon}(\omega_j, \boldsymbol{s})}{\sum_j tr\left\{\boldsymbol{R}(\omega_j)\right\}} \tag{4}$$

over a grid of slowness space and summed (incoherently) over a wide range of frequencies *j*.

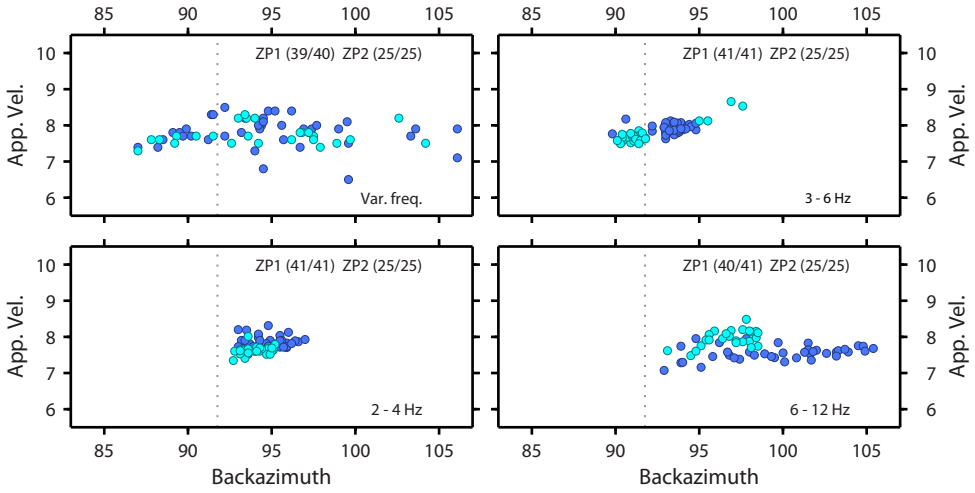

Figure 2. Slowness estimates at ARCES using broadband f-k analysis for first regional P-phases from the ZP1 (dark blue) and ZP2 (light blue) mines. The true backazimuth is approximately 91.8 degrees for both mines as indicated by the vertical dashed lines. The top left panel displays estimates taken from the fully automatic ARCES detection lists and the remaining panels are reprocessed in the fixed frequency bands as indicated. The 41 events from the ZP1 mine and the 25 events from the ZP2 mine all took place between October 2001 and September 2002.

The right panel of Figure 1 shows a satellite image of two mines, denoted ZP1 and ZP2, which are separated by less than 3 km. Figure 2 shows four panels of slowness estimates for initial P-arrivals from these two mines over a 12 month period from October 2001 to September 2002.

The top left panel (in which measurements are taken from the fully automatic detection lists) shows a very high spread in the measured backazimuth and no separation between the two populations. (Under standard array processing considerations, we would not anticipate being able to resolve these two nearby sources with a single seismic array at this distance.) The frequency band selected in the automatic detection lists is determined by an automatic algorithm on a case-by-case basis to try to optimize the signal-to-noise ratio (SNR). The high variability of the slowness estimates has significant consequences for the accuracy of fully-automatic event bulletins (e.g., Ringdal and Kværna, 1989) and all undertakings which can result in more stable parameter estimates should be investigated.

The three remaining panels are interesting as they demonstrate (a) that the stability is very much improved by selecting a fixed frequency band, (b) that the bias and spread of the slowness estimates are very different for the different frequency bands, and (c) that there is a significant separation in parameter space for the two different mines. Whilst this separation is evident visually, it would be very difficult to identify the source confidently just given the parameter estimates for the different frequency bands. It is the intention that, in bypassing the plane-wavefront parametrization, that Empirical Matched Field Processing will be able to exploit the characteristic source and path effects responsible for the patterns shown in order to identify a likely source.

If we calculate spatial covariance matrices, $\boldsymbol{R}(\omega)$, for seismic arrivals known to have originated from a given source, we can form an ensemble covariance matrix $\boldsymbol{R}_\alpha(\omega)$ from a specified population α of events. If we replace the theoretical plane wavefront steering vectors $\boldsymbol{\varepsilon}(\omega, \boldsymbol{s})$ in Equation 4 with the principal eigenvectors of $\boldsymbol{R}_\alpha(\omega)$, we can obtain an equivalent normalized measure of the energy incident on the array for the "α" hypothesis:

$$\hat{P}(\alpha) = \frac{\sum_j \boldsymbol{\varepsilon}(\omega_j, \alpha)^H \boldsymbol{R}(\omega_j) \boldsymbol{\varepsilon}(\omega_j, \alpha)}{\sum_j tr\{\boldsymbol{R}(\omega_j)\}} \qquad (5)$$

Ensemble spatial covariance matrices $\boldsymbol{R}_\alpha(\omega)$ were calculated for the two populations of events displayed in Figure 2 and the "matched field statistic" specified in Equation 5 was evaluated for both these empirical steering vectors for each event known to have taken place from the same two mines during the period 2003-2004 (see Figure 3).

Whilst the separation of the ZP1 and ZP2 populations against the calibrated empirical steering vectors is reasonable, many symbols lie in a zone of high uncertainty (shaded grey) where the match of the measured wavefield is quite similar for both of the source hypotheses. Two events from the ZP1 mine have been misclassified as belonging to the ZP2 cluster (with a marginal difference in detection statistics), although we do not rule out errors in the Ground Truth collection.
An ideal classification would see all of the dark blue symbols in the lower right corner and all the pale blue symbols in the upper left corner.

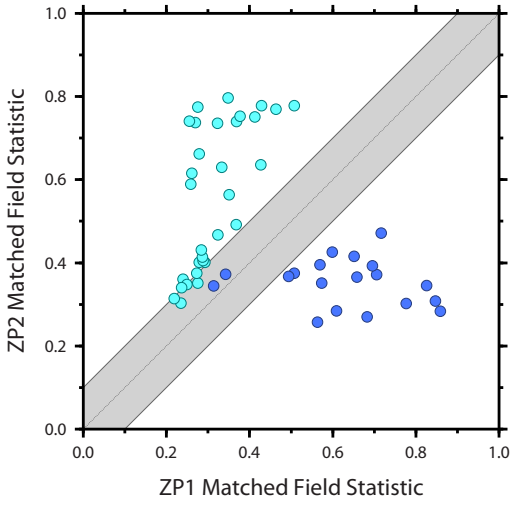

Figure 3. Matched field statistics as defined in Equation 5 for empirical steering vectors calibrated from the 2001-2002 populations for the ZP1 and ZP2 mines on the x and y axes respectively. Dark and light blue symbols indicate that the spatial covariance matrices were calculated from signals from ZP1 and ZP2 events respectively during the period 2003-2004. Frequency band used: 2.5 - 12.5 Hz.

There is a clear need to understand why the classification performance is as shown in Figure 3 and investigate which, if any, improvements are possible. The ability to attribute candidate signals to one or other of the sources will depend upon how well the principal eigenvector of the ensemble spatial covariance matrix for each cluster represents the spatial covariance observed for individual events associated with that source, and the degree to which the empirical steering vectors from the different sources differ from each other. To this end, we perform cluster analysis upon empirical steering vectors calculated from individual events. Specifically, we calculate a similarity measurement as defined in Equation 6 between the principal eigenvectors of the spatial covariance matrices between any two distinct data segments. Note that the measure is an average value taken over a wide range of frequency bands. Figures 4 and 5 show cross-matrices and corresponding single-linkage dendrograms for the ZP1 and ZP2 event sets for 2001-2002.

$$d_{\alpha\beta} = \frac{1}{N_{\mathrm{B}}} \sum_{j=1}^{N_{\mathrm{B}}} \left| \boldsymbol{\varepsilon}(\omega_j, \alpha)^H \boldsymbol{\varepsilon}(\omega_j, \beta) \right| \tag{6}$$

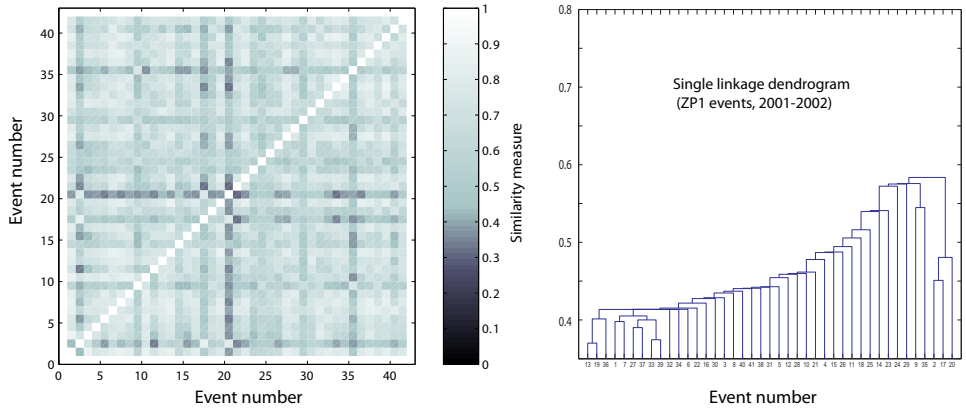

Figure 4. Similarity matrix and corresponding single-linkage dendrogram for events from the ZP1 mine 2001-2002.

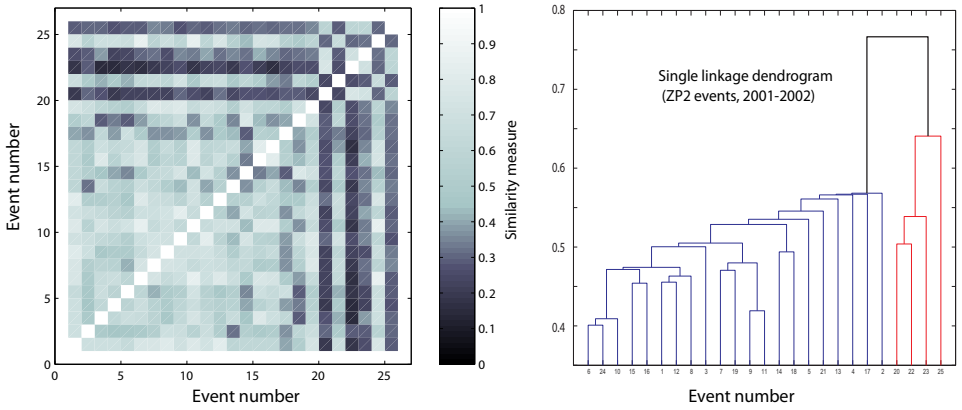

Figure 5. Similarity matrix and corresponding single-linkage dendrogram for events from the ZP2 mine 2001-2002.

It is clear from Figure 5 that the ensemble of events selected to represent the ZP2 mine consists of two distinct clusters according to the similarity metric defined in Equation 6. We then repeat the classification exercise displayed in Figure 3 but replacing the ZP2 matched field statistic with the maximum of two different statistics, calculated from the ensemble covariance matrices from each of the two clusters evident in Figure 5. It is clear in the right hand panel of Figure 6 that the classification is far better than that displayed in Figure 3 where only a single empirical steering vector was available to represent the whole ZP2 source region. Almost no events appear in the grey shaded region when the two steering vector representation was used for the ZP2 site. The same two events in the ZP1 population remain misclassified, and there is evidence in Figure 4 for (at least) two events which appear to match quite poorly.

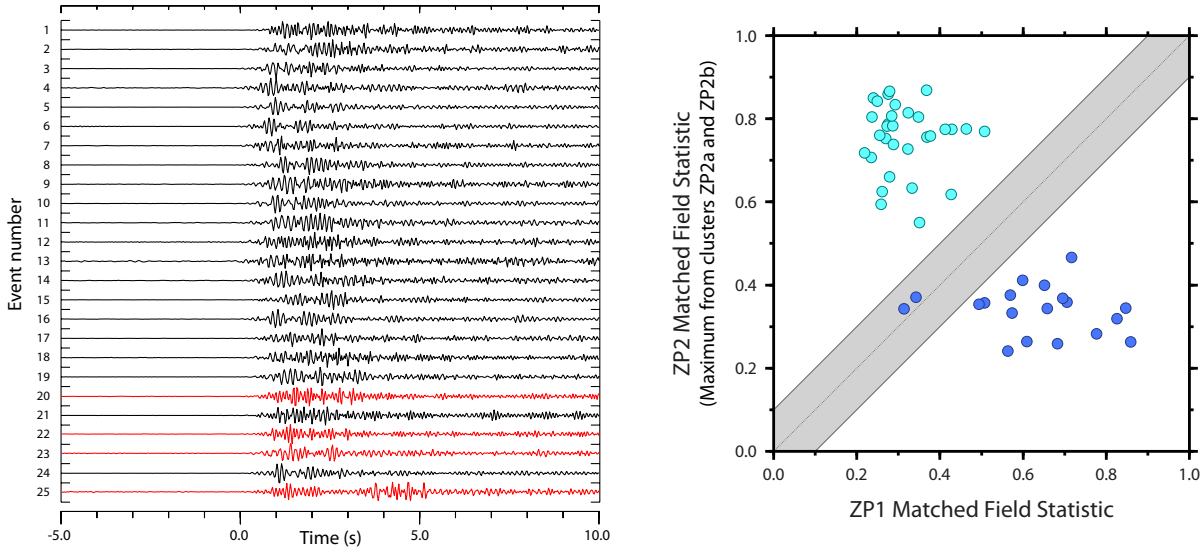

Figure 6. Waveforms from the ZP2 events (left) with traces colored according to the classification displayed in Figure 5, and Matched Field Statistics (right) with ZP2 events represented by two different empirical steering vectors. The empirical steering vectors for both mine clusters are calculated from the 2001-2002 event pools and the dark and light symbols represent ZP1 and ZP2 events respectively from 2003-2004. Frequency band used: 2.5 - 12.5 Hz.

We conclude that cluster analysis is necessary as a prerequisite to the calculation of ensemble covariance matrices in order that the validity of the selected cluster can be determined. If a cluster clearly separates into two populations, it is likely that matched field statistics calculated using an ensemble spatial covariance matrix generated from the full population are not going to provide a good classification. Events which are spurious and which clearly do not belong to the cluster will be identified using the procedure described here.

The scattering effects which are so problematic for coherent array processing of high frequency phases on ARCES are even more so for the arrays in Kazakhstan due to the sparse geometries (see Figure 7). Whilst the problem of low SNR at low frequencies is less acute than at ARCES, the SNR still improves greatly at higher frequencies (see Figure 8). Estimating the slowness using broadband f-k analysis is a trade-off between SNR and signal coherence.

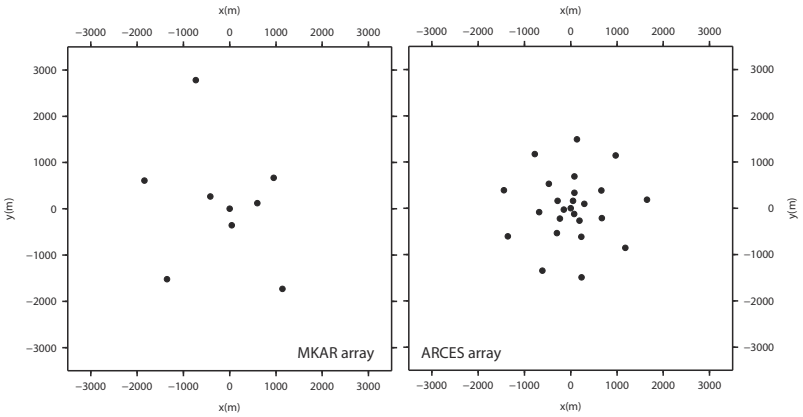

Figure 7. Comparison of the configurations of the MKAR and ARCES arrays. Whilst both arrays have the same fundamental design (of concentric rings with increasing numbers of sensors), the sparsity of the MKAR array makes processing of regional phases very difficult. The inner-most sensors of ARCES which are so crucial to the detection and estimation of high frequency phases are missing from MKAR.

The data segment in Figure 8 is the result of a mining explosion on June 30, 2007, at a distance of approximately 400 km from the MKAR array. This event belongs to the cluster labelled 344 by Hartse et al. (2008).
The SNR for the Pn arrival is high, but only for frequencies above 4 Hz. This is unfortunate from the point of view of phase estimation using traditional array processing since (see Figure 9) almost all slowness estimates above 4 Hz are qualitatively incorrect due to scattering and incoherence together with sidelobes in the array response function for these frequencies. Many of the estimates below 4 Hz are qualitatively misleading due to a low SNR.

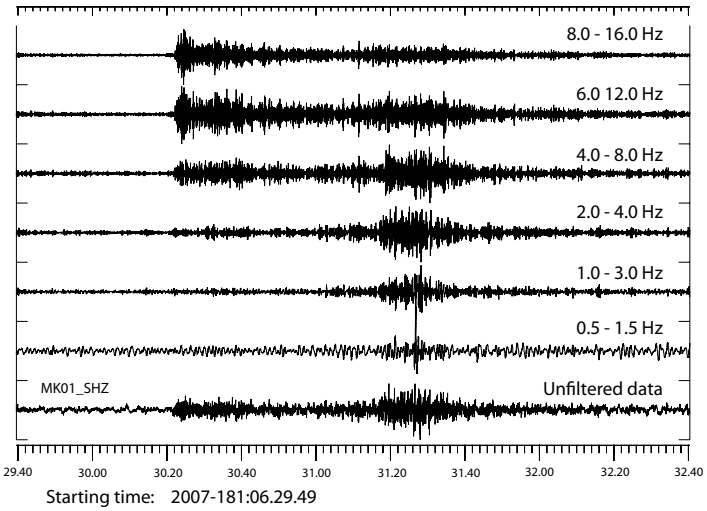

Figure 8. A three-minute long data segment on the single channel MK01_SHZ filtered in various frequency bands as indicated. See above text for details.

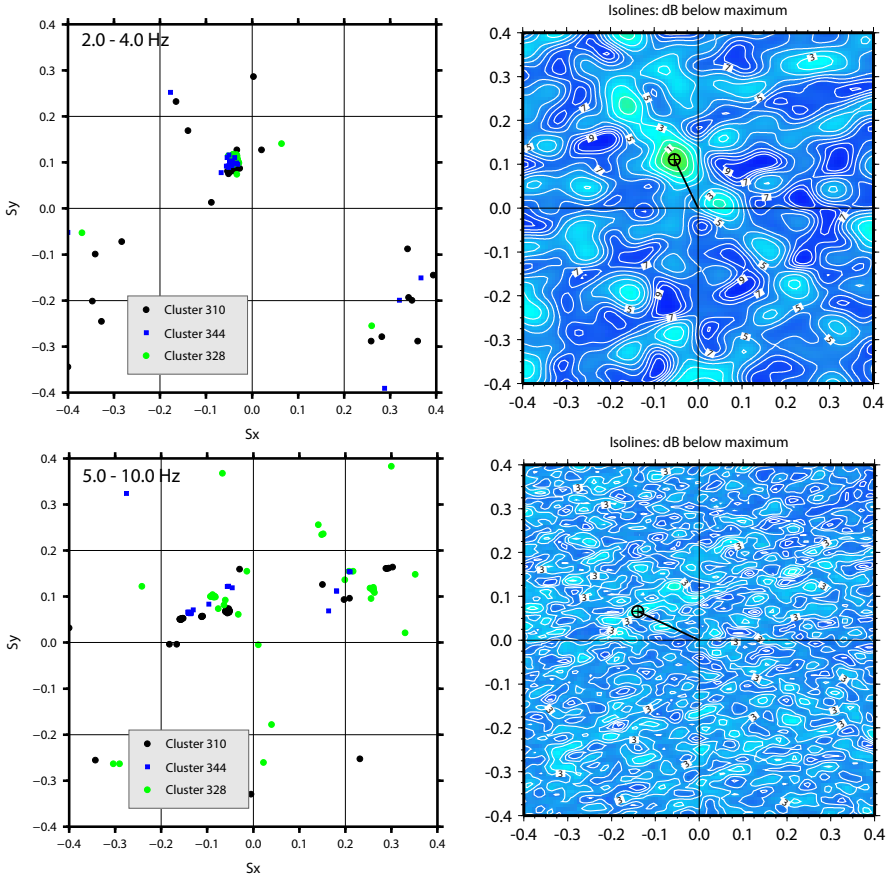

Figure 9. The left hand panels show scatter-plots of slowness estimates in the frequency bands indicated for initial Pn arrivals at MKAR for events in clusters 310, 328, and 344 in Hartse et al. (2008). The anticipated backazimuth for all events is close to 340^{O} and the apparent velocity should be close to 8 kms^{-1}. The f-k grids to the right display the relative beam-power as a function of slowness for the arrival at MKAR at a time 2007-181:06.30.31.425 (see Figure 8).

Figure 9 indicates at a glance that an attempt at source classification based upon direction estimates from traditional plane wavefront array processing are not likely to succeed given the abysmal coherence on the MKAR array for high frequency regional phases. We attempt a classification using empirical matched field processing; given that the clusters were derived by Hartse et al. (2008) and MacCarthy et al. (2007) using waveform correlation methods, some degree of success is anticipated. Cluster analysis was performed on the empirical steering vectors for each of the individual events prior to an attempt at classification using ensemble covariance matrices. The results are displayed in Figure 10. From this plot, it is clear that multiple empirical steering vectors are necessary to characterize the wavefields recorded by the range of events from each of these clusters.

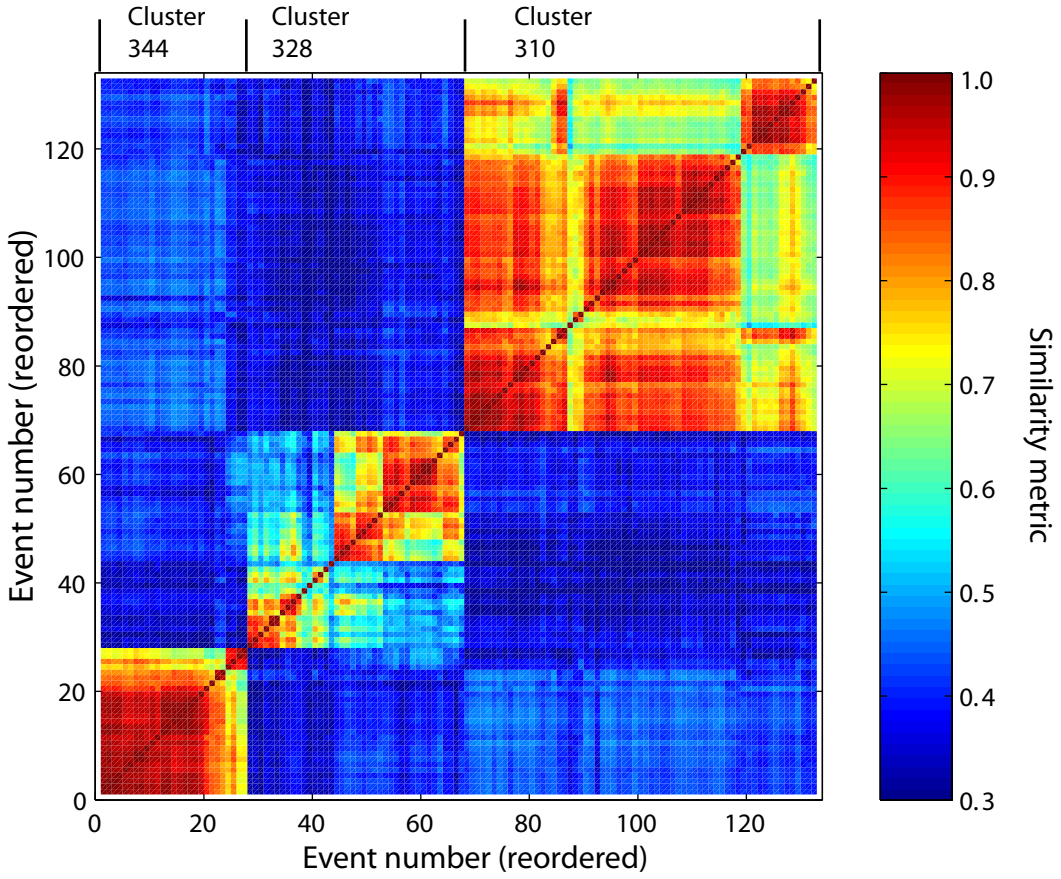

Figure 10. Matched field similarity matrix compiled over three distinct clusters (310, 328, and 344) from the study of Hartse et al. (2008). A number of distinct subpopulations are revealed from the cluster analysis of the empirical steering vectors for the Pn arrival at the MKAR array (especially for cluster 328) and these need to be taken into consideration prior to matched field classification using ensemble spatial covariance matrices.

Figure 11 shows that the matched field classification has attributed each of the individual events to the correct source. We have therefore demonstrated that the spatial covariance matrices provide a coherent array-processing technique which works at high frequencies over an aperture for which simply delay-and-stack beamforming and broadband f-k analysis fail.

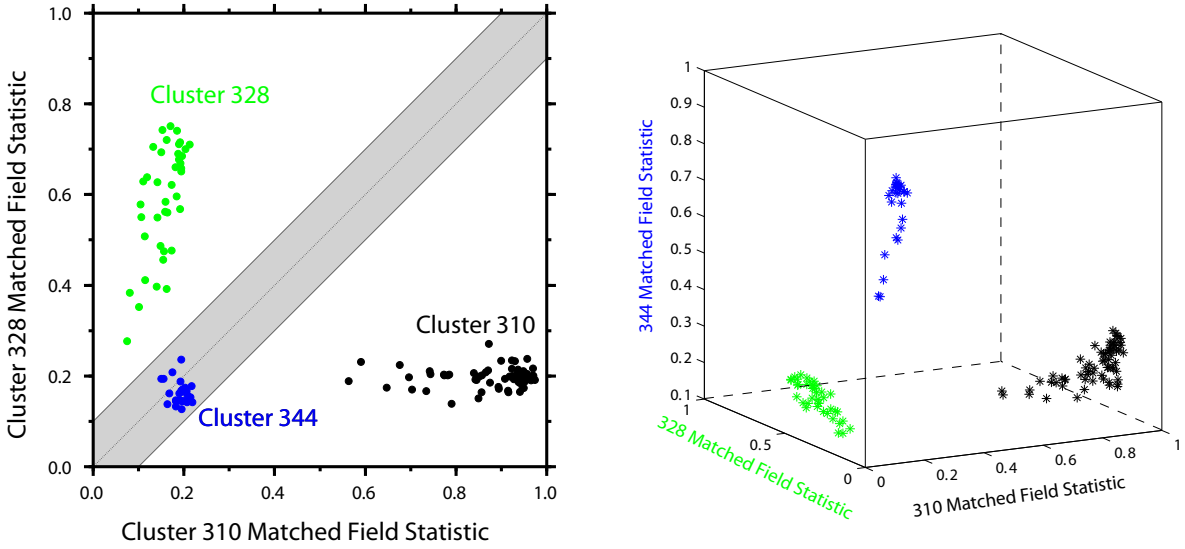

Figure 11. Classification results for clusters 310, 328, and 344, using multiple ensemble covariance matrices for each of the subpopulations displayed in Figure 11.

CONCLUSIONS AND RECOMMENDATIONS

We have demonstrated that empirical matched field processing is able to classify, with a high degree of confidence, events from sources of repeating seismicity even for cases where classical waveform correlation methods become problematic due to differences in the source time function. This has also been demonstrated for an array with an aperture over which incoherence defeats traditional methods of array processing.

One of the most important aspects of the procedure that has come to light during this reporting period is the necessity of performing cluster analysis on the empirical steering vectors prior to the building of ensemble covariance matrices. If an attempt is made to use an ensemble covariance matrix constructed from significantly differing subpopulations, this will result in empirical steering vectors which represent the entire population relatively poorly.

It is of great interest as to how many events are required to characterize a specified source region adequately; there are clearly circumstances where one would hope to achieve a reasonable performance with only a single observation. Part of the motivation for the construction of the ensemble covariance matrices is to reduce the variance in the narrow-band spatial covariance matrices (Equation 3). Multitaper methods (Thomson, 1982) produce spectral estimates with low variance and the recently published codes of Prieto et al. (2009) include routines for robust and stable multitaper coherence estimation. We have implemented these routines into our codes and are currently investigating the variability of the estimates of the spatial covariance matrices.

ACKNOWLEDGEMENTS

We are grateful to the National Data Center in Kazakhstan for permission to use the data from the Kazakh seismic arrays. We thank Hans Hartse of Los Alamos National Laboratory for kindly providing the results of waveform cross-correlation and cluster analysis for mining events in Kazakhstan. We are grateful to Prof. Germán Prieto of Universidad de los Andes, Bogotá, for making available the source code and documentation for the Multitaper Spectral Analysis code. We thank the IRIS DMC and the KN network for making continuous data from the KNET stations available. All maps produced using GMT software (Wessel and Smith, 1995).

REFERENCES

Capon, J. (1969), High-resolution frequency-wavenumber spectrum analysis, in *Proc. IEEE* 57:1408–1418.

Harris, D. B., F. Ringdal, E. O. Kremenetskaya, S. Mykkeltveit, J. Schweitzer, T. F. Hauk, V. E. Asming, D. W. Rock, and J. P. Lewis (2003). Ground-truth collection for mining explosions in Northern Fennoscandia and Russia, in *Proceedings of the 25th Seismic Research Review - Nuclear Explosion Monitoring: Building the Knowledge Base*, LA-UR-03-6029, Vol. 1, pp. 54–62.

Hartse, H. E., G. E. Randall, and S. J. Arrowsmith (2008). Regional event identification research in Asia, in *Proceedings of the 30th Monitoring Research Review, Ground-Based Nuclear Explosion Monitoring Technologies*, LA-UR-08-05261, Vol.1, pp. 615–624.

MacCarthy, J. K., H. Hartse, M. Greene, and C. Rowe (2008). Using waveform cross-correlation and satellite imagery to identify repeating mine blasts in Eastern Kazakhstan, in *Seism. Res. Lett.*, 79:393–399.

Prieto, G. A., R. L. Parker, and F. L. Vernon (2009). A Fortran 90 library for multitaper spectrum analysis, in *Computers and Geosciences (in press: corrected proof online)*.

Ringdal, F. and T. Kværna (1989). A multi-channel processing approach to real time network detection, phase association, and threshold monitoring, in *Bull. Seism. Soc. Am.* 79:1927–1940.

Thomson, D. J. (1982). Spectrum estimation and harmonic analysis, in *Proc. IEEE* 70:1055–1096.

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