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
Award Nos. FA8718-08-C-0007[1,3] and DE-FC52-06NA27324[2]
Proposal No. BAA08-39

ABSTRACT

Matched field processing, a method developed in underwater acoustics to detect and locate targets, has been adapted to classify transient seismic signals arising from mining explosions and other sources of recurring seismicity. Empirical matched field processing (EMFP) uses observations of historical 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. EMFP has been evaluated using seismic arrays and networks in Central Asia which comprise a wide spectrum of sensor geometries. These range from medium aperture arrays (e.g., MKAR and BVAR), to teleseismic arrays and networks (e.g., KNET and KURK), through to very large aperture virtual networks consisting of arrays and single-site stations separated by distances of several hundreds of kilometers. The region studied is also of interest given the large number of sources of repeating industrial events and regions of diffuse seismicity.

Multitaper methods have been used to calculate narrow frequency-band spectral covariance matrices for short data segments. Windows of the order 3 seconds are typically used in classical f-k analysis, and calculating the empirical steering vectors from data segments of this length allows a direct comparison between matching statistics from theoretical and empirical steering vectors. Using first P-arrivals from events at the Kara Zhyra mine near Semipalatinsk, observed on the MKAR and ZALV arrays, we have demonstrated that the empirical matched field statistics are significantly greater than those obtained for plane-wavefront models. The gain is greatest for high frequencies, at which coherent processing is difficult even on these relatively small aperture arrays. We have also demonstrated that the performance using empirical steering vectors calibrated from single events is frequently comparable to that possible using vectors calibrated from ensembles of many events. This indicates that the method is applicable to sources of interest from which only single events have been observed.

The sensitivity of the spectral covariance estimates to the exact positioning of the data-window has been investigated. For 3-second windows surrounding P-arrivals, the resulting matched field statistics appear to be almost unaffected by shifts of over a second. This has the consequences that the method can be applied to signals with fairly uncertain onset times, and that a single-phase matched field detector can be formulated where the covariance estimates on the incoming data need not be evaluated more often than once per second. Detection runs have been performed for Kara Zhyra events using only 3-second P-phase matched field templates on MKAR, ZALV, and a wide-aperture network comprising both arrays (where time-shifts are applied prior to calculating the spectral covariance matrix). In all three cases, values of the matched field statistic exceeding the background level significantly were only recorded on a very limited number of occasions. All were at characteristic times of day, and all were accompanied by additional evidence, both from correlation detectors and signals at a closer station (KURK), that the detected signal did indeed come from the target source region. A P-phase matched field detector was run using templates from single after-shocks of the M=7.4 October 8, 2005, Kashmir earthquake for both KNET (with staggered time-windows) and the KKAR array. Significant values of the matched field statistic were obtained for both sensor geometries on multiple occasions, at times consistent with events from the same source region. The source region within which events can be associated using EMFP is likely to decrease as the sensor network aperture increases. Consequently, an unequivocal matched field identification over a very large aperture sensor array is likely to attribute a signal to a specific source region with a greater degree of confidence than when a small aperture array is used.

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 began 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 proceeded 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

Throughout this study, we consider exclusively seismic data from arrays and networks. If $\underline{\boldsymbol{x}}_m$ denotes the coordinates of sensor *m* relative to some reference site, then we denote the time-series recorded at this site by $r_m(t) = r(t, \underline{\boldsymbol{x}}_m)$. If *N* denotes the number of sensors in our array then, for a given frequency ω, one can form the $N \times N$ spatial covariance matrix, $\underline{\boldsymbol{R}}(\omega)$, of which the element R_{nm} is given by

$$R_{mn}(\omega) = \left(\int r(t, \underline{\boldsymbol{x}}_m) e^{-i\omega t} dt \right) \left(\int r(t, \underline{\boldsymbol{x}}_n) e^{i\omega t} dt \right) \tag{1}$$

Given the relative transience of seismic signals, it is understood that $\underline{\boldsymbol{R}}(\omega)$ should be calculated on a relatively short data segment (at most a few seconds for high frequency studies of regional events).

The complex *N*-vector $\underline{\boldsymbol{\varepsilon}}(\omega, \underline{\boldsymbol{s}}) = \left[e^{-i\omega \underline{\boldsymbol{s}} \cdot \underline{\boldsymbol{x}}_1}, \quad \cdots, \quad e^{-i\omega \underline{\boldsymbol{s}} \cdot \underline{\boldsymbol{x}}_N} \right]^T$ describes a wavefield over the array at the frequency ω with the ε_m defining the time-delays (phase shifts) and amplitudes for the waveforms at the *N* sites. If *H* denotes the Hermitian transpose, then a scalar of the form

$$\hat{P}(\underline{\boldsymbol{\varepsilon}}(\omega)) = \frac{\underline{\boldsymbol{\varepsilon}}(\omega)^H \underline{\boldsymbol{R}}(\omega) \underline{\boldsymbol{\varepsilon}}(\omega)}{\mathrm{tr}\left\{ \underline{\boldsymbol{R}}(\omega) \right\}} \tag{2}$$

provides a measure of how consistent the data (from which $\underline{\boldsymbol{R}}(\omega)$ is calculated) are with the wavefield hypothesis defined by $\underline{\boldsymbol{\varepsilon}}(\omega, \underline{\boldsymbol{s}})$.

Given the broadband nature of the majority of the seismic signals examined, it is usual to consider a wide-band statistic, summed incoherently over a range of *K* frequency bands:

$$\hat{P}(\underline{\boldsymbol{\varepsilon}}([\omega_1 : \omega_K])) = \sum_{k=1}^{K} a_k \frac{\underline{\boldsymbol{\varepsilon}}(\omega_k)^H \underline{\boldsymbol{R}}(\omega_k) \underline{\boldsymbol{\varepsilon}}(\omega_k)}{\mathrm{tr}\left\{ \underline{\boldsymbol{R}}(\omega_k) \right\}} \tag{3}$$

where $\underline{\boldsymbol{\varepsilon}}(\omega_1 : \omega_K)$ comprises *K* complex vectors of length *N*: the so-called steering vectors for each of the *K* frequencies. The coefficients a_k are weights which ensure normalization of the broadband statistic:

$$\sum_{k=1}^{K} a_k = 1. \tag{4}$$

The most obvious solution is to set $a_k = 1/K$ (for all *k*) for a simple mean, although non-uniform coefficients may be preferable due to signal-to-noise considerations or stability.

In the most straightforward wavefield parametrization, the plane wavefront assumption, we assume that the waveforms on all sensors are identical except for the time-delay and that the arrival time at sensor *m* is specified by

$$t_m = t_0 - \underline{\boldsymbol{s}} \cdot \underline{\boldsymbol{x}}_m \tag{5}$$

where t_0 is the arrival time at the array reference site (with coordinate vector $\underline{\boldsymbol{x}}_0$) and the slowness vector

$\underline{\boldsymbol{s}} = (s_x, s_y)$ is related to the backazimuth Θ and the apparent velocity v_{app} by

$$s_x = s\sin(\Theta), \quad s_y = s\cos(\Theta), \quad \text{and} \quad s = 1/v_{\text{app}}. \tag{6}$$

In this case, our theoretical steering vector $\underline{\boldsymbol{\varepsilon}}$ is given by

$$\underline{\boldsymbol{\varepsilon}}(\omega, \underline{\boldsymbol{s}}) = \left[e^{-i\omega\underline{\boldsymbol{s}}\cdot\underline{\boldsymbol{x}}_1}, \quad \cdots, \quad e^{-i\omega\underline{\boldsymbol{s}}\cdot\underline{\boldsymbol{x}}_N}\right]^T \tag{7}$$

and the statistic evaluated over our *K* specified frequencies is

$$\hat{P}(\ \underline{\boldsymbol{\varepsilon}}([\omega_1 : \omega_K], \underline{\boldsymbol{s}})\ ) = \sum_{k=1}^{K} a_k \frac{\underline{\boldsymbol{\varepsilon}}(\omega_k, \underline{\boldsymbol{s}})^H \underline{\boldsymbol{R}}(\omega_k)\underline{\boldsymbol{\varepsilon}}(\omega_k, \underline{\boldsymbol{s}})}{\text{tr}\left\{\underline{\boldsymbol{R}}(\omega_k)\right\}} \tag{8}$$

Deviations of the wavefield from the predicted plane wavefront model will lead to a reduction in the value of the expression in Equation (8). The steering vector, $\underline{\boldsymbol{\varepsilon}}(\omega)$, which optimizes the value of the quadratic form in Equation (3) is the principal eigenvector of the matrix $\underline{\boldsymbol{R}}(\omega)$. In empirical matched field processing (EMPF), it is anticipated that the spatial structure of an incoming wavefront over a sensor array from a source of repeating seismicity will be approximately the same for each event. Therefore, if a covariance matrix $\underline{\boldsymbol{R}}(\omega, \alpha)$ is measured for an arrival from a source of interest (denoted by α) then the principal eigenvector $\underline{\boldsymbol{\varepsilon}}(\omega, \alpha)$ is likely to constitute an empirical steering vector which will provide a better match than the closest theoretical (plane-wave) steering vector, given that differences in amplitude and deviations in phase-shifts resulting from diffraction and scattering will be accounted for in the spectral covariance matrix estimate.

There may be many occurrences, α_i, of a given arrival from subsequent events at the site of interest. In this case, it may be desirable to use many different observations on which to base our EMFP spatial template. We could form an ensemble covariance matrix, $\underline{\boldsymbol{R}}(\omega, \boldsymbol{\alpha})$, from the single arrival covariance matrices, $\underline{\boldsymbol{R}}(\omega, \alpha_i)$, with

$$\underline{\boldsymbol{R}}(\omega, \boldsymbol{\alpha}) = \sum_{\alpha_i \in \boldsymbol{\alpha}} w_i \underline{\boldsymbol{R}}(\omega, \alpha_i) \tag{9}$$

where the w_i are weights which define the contribution for each of the single arrival covariance matrices. Forming the ensemble covariance matrix may provide a more stable estimator of the optimal empirical steering vector for the arrival by reducing the variability in the single arrival matrices. On the other hand, there is a danger that the ensemble covariance matrix will be degraded by the inclusion of arrivals from too great a diversity of sources (for example, a larger source region than was at first anticipated). In such a case it may be better to consider multiple steering vectors, each calculated from a single arrival. Ringdal et al. (2009) show an example of how source classification performance improved significantly by constructing two separate ensemble estimates from the individual observations. Cluster analysis on the single arrival covariance matrices is necessary before forming the covariance matrix in Equation 9.

Given that we may normalize any covariance matrix $\underline{\boldsymbol{R}}(\omega)$, to have unit trace (either prior to or after the forming of an ensemble covariance matrix), we can express our ensemble and single observation matched field estimators for the source region α as

$$\hat{P}(\ \underline{\boldsymbol{\varepsilon}}([\omega_1 : \omega_K], \boldsymbol{\alpha})\ ) = \sum_{k=1}^{K} a_k \underline{\boldsymbol{\varepsilon}}(\omega_k, \boldsymbol{\alpha})^H \underline{\boldsymbol{R}}(\omega_k)\underline{\boldsymbol{\varepsilon}}(\omega_k, \boldsymbol{\alpha}) \tag{10}$$

and

$$\hat{P}(\ \underline{\boldsymbol{\varepsilon}}([\omega_1 : \omega_K], \alpha_i)\ ) = \sum_{k=1}^{K} a_k \underline{\boldsymbol{\varepsilon}}(\omega_k, \alpha_i)^H \underline{\boldsymbol{R}}(\omega_k)\underline{\boldsymbol{\varepsilon}}(\omega_k, \alpha_i) \tag{11}$$

respectively. Here, the steering vectors $\underline{\boldsymbol{\varepsilon}}(\omega_k, \boldsymbol{\alpha})$ and $\underline{\boldsymbol{\varepsilon}}(\omega_k, \alpha_i)$ are respectively the principal eigenvectors of ensemble covariance matrices, $\underline{\boldsymbol{R}}(\omega_k, \boldsymbol{\alpha})$, and single arrival covariance matrices, $\underline{\boldsymbol{R}}(\omega, \alpha_i)$.

If the eigenspectrum of the covariance matrix is dominated by a single eigenvalue, then it is very likely that the principal eigenvector will provide a good representation of an arrival. If the wavefield at a given time is not well represented by a simple wavefront, the eigenspectrum is not likely to be dominated by one eigenvalue and a subspace approach may be more appropriate where we consider a signal-space spanned by more than one eigenvector. Our single-band steering vector $\underline{\boldsymbol{\varepsilon}}(\omega_k)$ then has to be a linear combination of the eigenvectors corresponding to the *L* largest eigenvectors of the template covariance matrix: $\underline{\boldsymbol{e}}_1, \cdots, \underline{\boldsymbol{e}}_L$. If the $\underline{\boldsymbol{e}}_i$ form the columns of the $N \times L$ steering

matrix $\underline{\underline{\boldsymbol{E}}}$, with

$$\underline{\varepsilon}(\omega_k) = \underline{\underline{\boldsymbol{E}}}\ \underline{\boldsymbol{v}} \tag{12}$$

then the vector of coefficients, $\underline{\boldsymbol{v}} = [v_1, \cdots, v_L]^T$, subject to $\underline{\boldsymbol{v}}^H \underline{\boldsymbol{v}} = 1$, is the principal eigenvector of the $L \times L$ matrix $\underline{\underline{\boldsymbol{E}}}^H \underline{\underline{\boldsymbol{R}}}\ \underline{\underline{\boldsymbol{E}}}$. L is referred to as the rank of the matched field statistic.

One of the most appealing features of empirical matched field processing (EMFP) is the expectation that, due to the narrow-band nature, the method will be far less sensitive to differences in the source-time function. Ripple-firing practices can cause significant problems for the classical correlation detectors (e.g., Gibbons and Ringdal, 2006) due to significant waveform variation from event to event. The spectral covariance matrices encode the spatial structure of the wavefield over the receiver aperture far more than the temporal structure and work on quarries on the Kola Peninsula (Harris and Kværna, 2010, Ringdal et al., 2009) has demonstrated that EMFP can provide excellent source classification for ripple-fired events, despite significant waveform diversity.

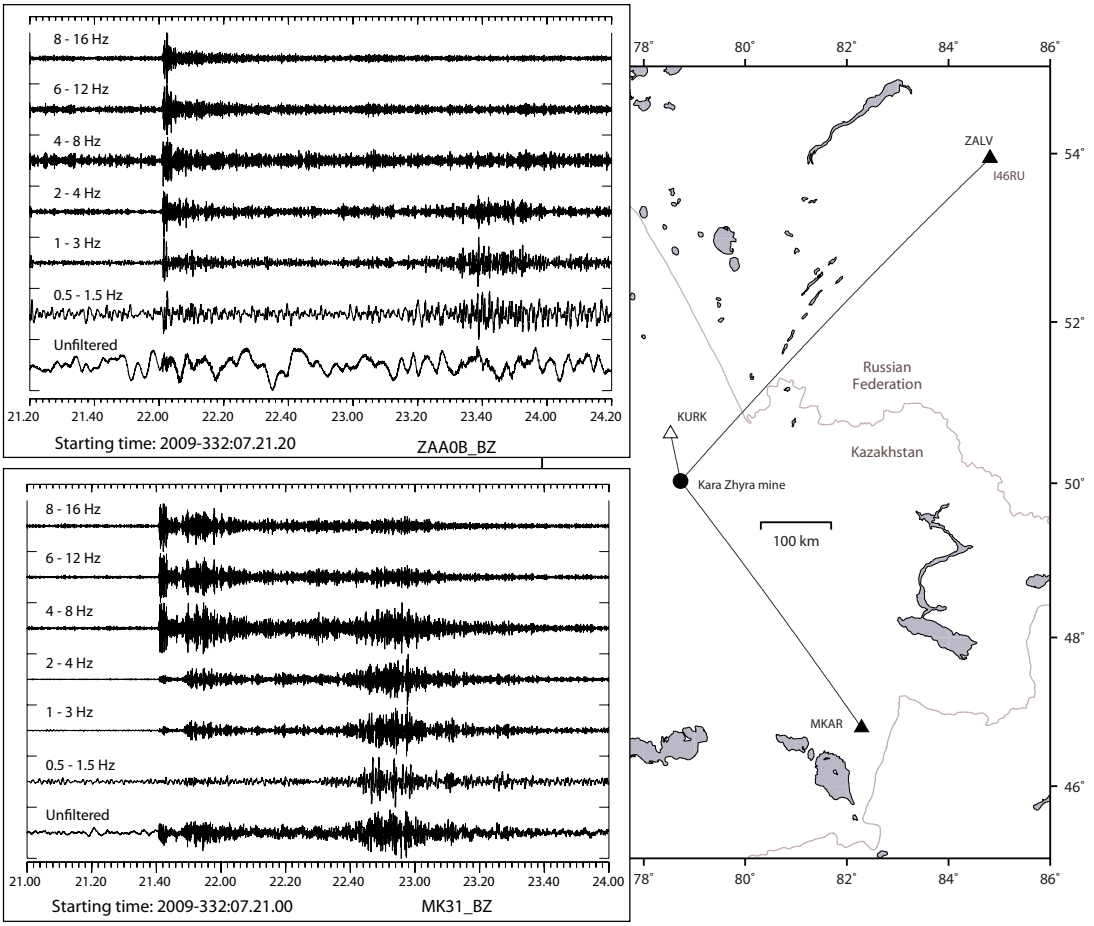

Figure 1. Location of the Kara Zhyra mine in Eastern Kazakhstan with respect to the MKAR and ZALV arrays (at 445 km and 600 km respectively). Waveforms from the vertical component broadband sensor at both arrays are shown at the times and in the frequency bands indicated.

Figure 1 displays the locations of two primary seismic arrays in central Asia in relation to a source of repeating seismic events. Both arrays consist of 9 sites over an aperture of ~5 km over which coherent processing of higher frequencies (> 4 Hz) demonstrably fails (Ringdal et al., 2009). The wavefield characteristics vary greatly between the two arrays. The only impulsive arrival at ZALV is Pn; Lg is highly emergent and there is no clear Sn onset. At MKAR, there are clear Pn and Pg onsets and high amplitude Lg. Pn is a very high frequency arrival and is often estimated poorly due to a low SNR below 4 Hz. The Pg arrival has far more low frequency energy. An event on October 28, 2009, was confirmed to have occurred at the mine and an extensive search using multiple correlation detectors on all available

arrays, followed by careful analysis of the signals at the closest station (KURK), provided a large database of events which were likely to have come from that source. For each event, a spectral covariance matrix and corresponding eigenvectors were evaluated for analyst-picked Pn arrivals at both the MKAR and ZALV arrays. We chose a window length of 121 samples (3.0 seconds) and 19 discrete frequencies from 2.0 to 8.0 Hz. The multitaper (Thomson, 1982) software described by Prieto et al. (2009) was used to construct the covariance matrices. An ensemble covariance matrix was also constructed for each of the two arrays, taking care to account properly for missing or corrupted data channels for a number of events.

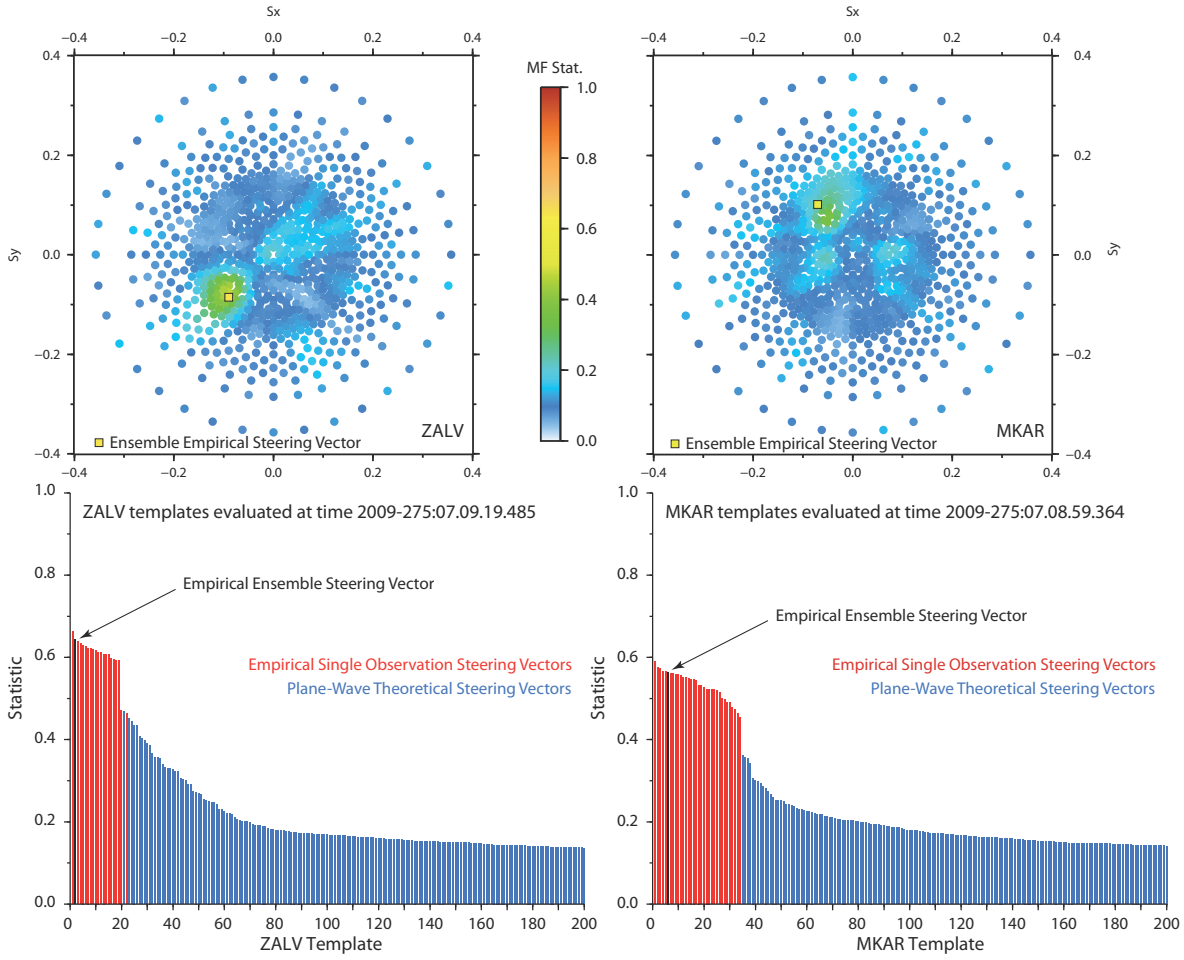

Figure 2. Matched field statistics for the specified times at ZALV and MKAR for Pn arrivals from an event assumed to originate close to the Kara Zhyra mine in Eastern Kazakhstan. Both theoretical plane wave steering vectors (blue), single observation empirical matched field steering vectors (red) and ensemble matched steering vectors (black) are used. The bars in the lower panels are rearranged in descending order of the matched field statistic.

For a number of candidate arrivals, the matched field statistic (an arithmetic mean across all frequency bands) was evaluated against steering vectors as displayed in Figure 2. The theoretical plane-wave, single observation empirical, and ensemble matched field statistics are as given in equations 8, 11, and 10 respectively. The upper panels of Figure 2 show the deployment in slowness space of the theoretical plane-wave steering vectors. Empirical steering vectors were only calculated for phase arrivals for which a clear onset could be picked by an analyst and the calculations displayed in Figure 2 are limited to events where the SNR on the array beam exceeds 20. In almost all cases, the matched field statistics for the empirical steering vectors (red/black) is considerably higher than that for the theoretical steering vectors (blue). The improvement is greater for MKAR than for ZALV, probably due to the high frequency content of the MKAR signals (for which the theoretical steering vectors perform poorly). The plot of matched field statistic as a function of slowness indicates local maxima close to the theoretical slowness vector; more clearly defined for ZALV than for MKAR over this frequency range.

The slowness plots also suggest a strategy for the calculation of f-k spectra in routine processing. A matched field statistic should be evaluated over a dense slowness grid (as is typically done today) but, in addition, a number of empirical steering vectors for repeatedly observed arrivals should augment the slowness space, labelled with the corresponding theoretical slowness vectors. Should an empirical steering vector attain a higher matched field statistic than any of the theoretical steering vectors, the corresponding theoretical slowness vector should be returned. Given a suitably dense coverage of empirical slowness vectors, this could remove the need to apply Slowness and Azimuth Station Corrections (SASCs). A final observation from Figure 2 is that essentially every one of the single-observation empirical matched field steering vectors performs significantly better than the best theoretical steering vector. EMFP can therefore be expected to perform well even for situations in which only a single observation is available. The evaluation of the matched field statistic for a given steering vector against the current covariance matrix is very rapid and it could be envisaged that huge numbers of steering vectors could be processed for each covariance matrix without the computational demands becoming prohibitive. The ensemble covariance matrices in these cases do perform well and it may be advisable to use empirical steering vectors both from ensemble estimates and from single observations.

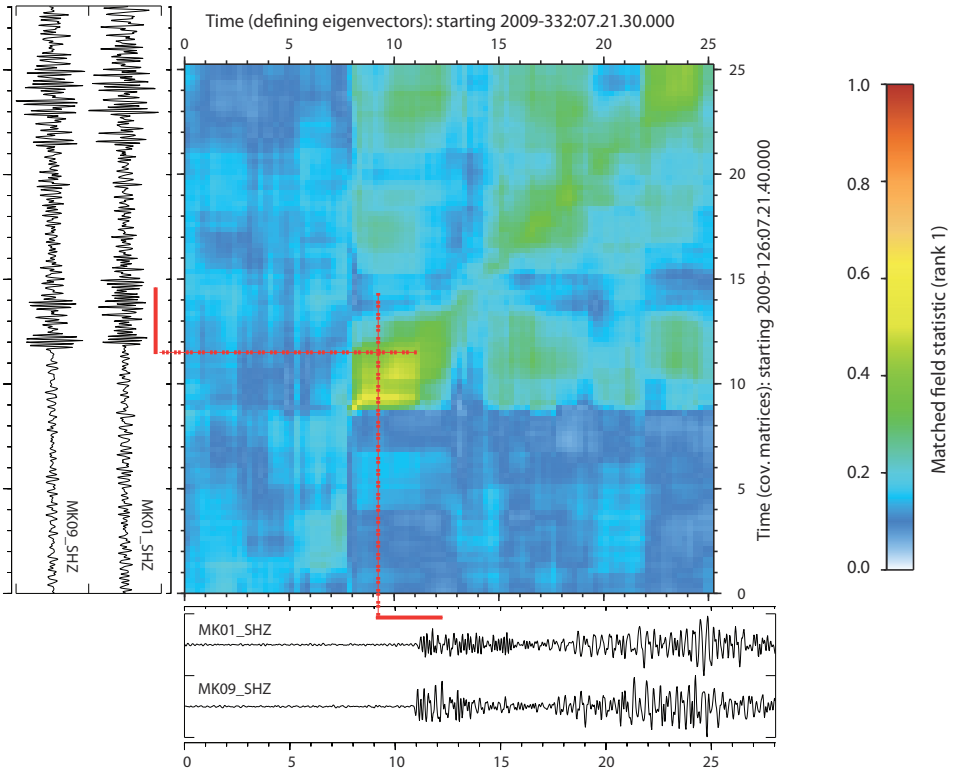

Figure 3. Matched field statistics as a function of data-window positioning for two events from the Kara Zhyra mine recorded at Makanchi. The x-axis of the grid indicates the starting time of the 3 second (121 sample) long data-window from which the (single event, single phase) empirical steering vector is calculated. The colors of the pixels for that value of x indicate the values of the matched field statistic evaluated against a data-window starting at the time indicated by the y coordinate. For example, the intersection of the dashed red lines indicates the value of the rank-1 statistic obtained using the spectral covariance matrix evaluated for the data segment covered by the vertical solid red bar against an empirical steering vector calculated from the template indicated by the horizontal red bar.

All of the matched field statistics presented so far have been evaluated over data segments carefully selected by an analyst. How sensitive are the results to the positioning of the data window? The three-second window chosen here is typical of the classical f-k analysis performed on small and medium aperture arrays and, given the relative transience of seismic phases, a significant increase in the data window length is probably not advisable. The uncertainties associated with arrival time estimates can be significant, especially for low SNR signals, and the method may be of limited applicability in automated processes if the matched field statistic varies greatly within the uncertainty of the phase arrival estimate.

Figure 3 displays the matched field statistic as a function both of time of the template window (for the calculation of the empirical steering vector) and of the time of the data window. The transience of the statistic is evident and it is clear that there is a modulation as the statistic is evaluated over time-intervals with different degrees of coherence. (For example, the matched field statistic for the Pg phase of event 2 using the Pn empirical steering vector for event 1 is considerably higher than for the coda immediately following the Pn arrival.) However, it appears that the matched field statistic is significantly greater than the background level for a duration long enough to cover the uncertainty surrounding the arrival time estimate. The variability will clearly change from arrival to arrival, and from array to array, although the results displayed here appear to be fairly representative of a wide range of phases observed.

The time dependence indicated in Figure 3 suggests that, for a continuous evaluation of a matched field statistic against a set of reference steering vectors, an evaluation of the data covariance matrix every second is probably sufficient. A more frequent evaluation of the spectral covariance matrix is likely to lead to a significantly higher computational expense without a signficant gain in performance. This provides for the possibility of a single-phase matched field detector to find the occurrences of incoming wavefields resembling a previously observed template. This is analogous to the correlation detectors except for that we are seeking primarily a spatial wavefield structure over the array and not necessarily a temporal wavefield structure.

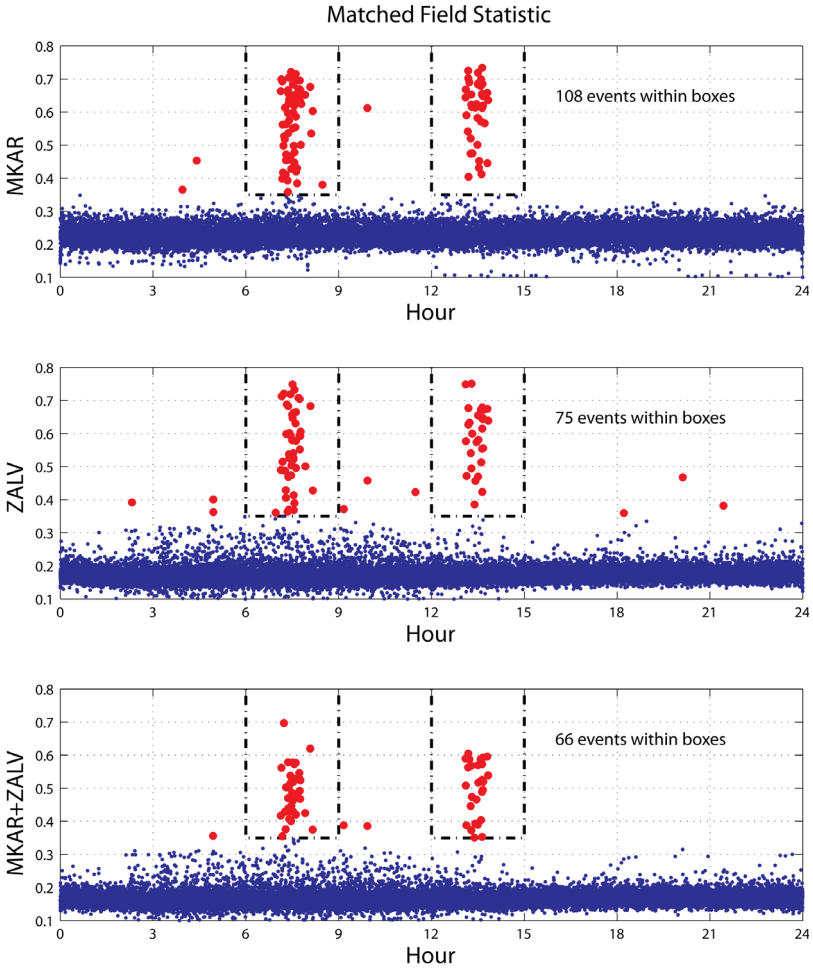

Figure 4. Matched field statistics from empirical steering vectors from a single Kara Zhyra event as function of time of day from the array configurations as displayed from January 1, 2009, to December 31, 2009. Each point indicates the maximum value obtained in a 20 minute segment (and so multiple events within the same data window would not be resolved). Values greater than 0.35 are highlighted. The covariance matrices for the two-array matched field calculation in the lowermost panel are calculated using a 20.0 second delay imposed on the ZALV waveforms relative to the MKAR waveforms.

Figure 4 shows the results of using the single-phase EMFP detector to identify Pn arrivals from Kara Zhyra events, at both MKAR, ZALV, and a virtual network consisting of the two arrays with the ZALV waveforms moved forward by 20 seconds to compensate the longer travel-time. The matched field statistic was evaluated every second for the calendar year 2009, for the three array configurations, using the empirical steering vectors from the ensemble covariance matrices. Every point shown in the plot is the maximum value obtained within predefined 20 minute segments. From the time-of-day plots, it does not appear that many triggers have occurred outside of two very narrow time-zones in the day, indicating that these matched field statistics may constitute a detector with a low false alarm rate. The virtual wide-aperture array works quite well (despite capturing slightly fewer events than the single array processes). The distribution of points for the two-array system follows very closely the results of the ZALV array which may indicate that, at some times, the covariance matrix is likely to be dominated by the contributions from channel pairs within one of the arrays. It is also possible that the single arrays cannot resolve between events from distant extremes of an extended source region since the relative phase-shifts are very small, whereas the two-array system is sensitive to far smaller changes of source region. This is identical to what is observed using the multi-channel waveform correlation detector. On a small-aperture array, detections of some significance can occur when an unrelated wavefront approaches from a similar direction to the master event. On a large array or network, it is almost impossible to get a trigger at all sites simultaneously and the detector responds to events from a much smaller source region.

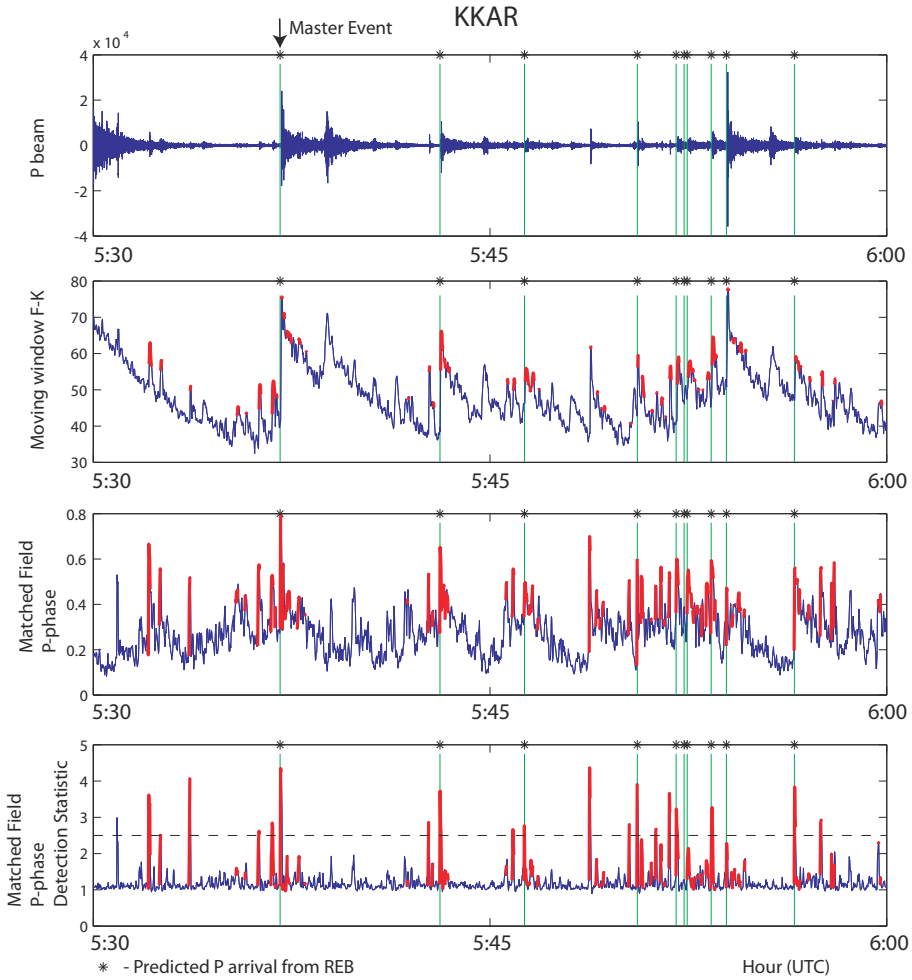

Figure 5. Panels displaying different processing attributes for a 30 minute long window of data from the KKAR array on October 8, 2005. Red parts of the traces indicate that the slowness vector from moving window F-K analysis is consistent with the P-phase of the master event. The lowermost trace shows the mean values over the narrow frequency bands where the matched field statistics for the different bands have been transformed using the procedure of Gibbons et al. (2008).

Figure 5 shows matched field statistics as a function of time for a 30 minute segment of data from the KKAR array in the aftermath of the M=7.4 Kashmir event of October 8, 2005, (at a distance of approximately 1000 km). Using a Pn arrival for a single aftershock to develop an empirical steering vector, the continuous matched field statistic attains a peak close to the first Pn arrival for each aftershock from this extended source region, diminishing throughout the coda. The transformed narrowband matched field trace appears to be a much more easily interpreted detection statistic than the simple matched field statistics (third trace) and this method may help greatly in the rapid identification of first arrivals in an extensive aftershock sequence. Performing the same calculation over the large aperture KNET network (but with time-shifts of up to several seconds provided a priori between the different stations) provided clear detections for a number of events, although far fewer than at the medium aperture KKAR array. This again suggests that the larger aperture network provides a sensitive coherent detector for events from a more limited source region.

CONCLUSIONS AND RECOMMENDATIONS

We have demonstrated that EMFP provides an effective means of identifying signals from sources of repeating seismicity. The matched field statistics obtained using empirical steering vectors, derived from previous observations of arrivals from the target site, are significantly higher than those obtained using the optimal theoretical steering vectors since deviations from the plane wavefront model are accounted for in the formulation. The improvement is greatest for higher frequencies and for larger aperture arrays over which classical array processing is made difficult or impossible due to waveform dissimilarity between the different sites.

If the theoretical steering vectors provided for scans of slowness-space in classical f-k analysis are augmented by empirical steering vectors from sources of repeating seismicity, this could provide “instant identification” of phases from known sources and may also circumvent the need for applying Slowness and Azimuth Station Corrections (SASCs).

The multitaper methods for spectral estimation appear to provide very stable estimates of the narrow-band phase and amplitude differences between different channels when only short duration waveform segments are to be used.

EMFP is applicable to sites of interest for which only a single observation is available. It is clearly an advantage to have numerous observations for estimation of the variability and a reduction of the “noise” in the matched field templates. However, using empirical steering vectors calculated from a single observation appears to give significantly better performance than the best theoretical steering vector.

The matched field statistics obtained for a typical 3-second long data segment appear to be sufficiently insensitive to the exact starting time of the selected data-windows that uncertainty in arrival time picking should not usually be an issue. In particular, for the application of EMFP as a continuous signal detector, it has been deemed that estimating a spectral covariance matrix every second on the incoming data is probably sufficient. We have demonstrated a successful application of EMFP for detecting occurrences of routine mining blasts at the Kara Zhyra mine in Eastern Kazakhstan on three different array configurations: the MKAR and ZALV arrays, and a virtual wide-aperture network comprising both arrays.

EMFP can be extended to far larger sensor apertures than are amenable to classical coherent array processing, although increasing the size of the receiver aperture reduces the size of the source region which can be monitored effectively. This has both advantages and disadvantages; a large network can provide a highly sensitive monitoring capability for a limited source region. A medium aperture array is likely to provide coverage of a greater source region with a somewhat higher false alarm rate if only a very limited source region is to be observed.

We have demonstrated that a transformation applied to continuous matched field statistic traces for narrow frequency bands can provide a very effective identification of arrivals in extensive aftershock sequences. This may contribute to a dramatic improvement of how efficiently such aftershock sequences can be classified.

ACKNOWLEDGEMENTS

We are grateful to the National Data Center in Kazakhstan for permission to use the data from the Kazakh seismic arrays. 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

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

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

Harris, D. B. and Kværna, T. (2010). Superresolution with Seismic Arrays using Empirical Matched Field Processing, *Geophys. J. Int.* (in press).

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*, Los Alamos National Laboratory, LA-UR-03-6029, Vol. 1, pp. 54–62.

Hartse, H. E., Randall, G. E., and Arrowsmith, S. J. (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., Hartse, H., Greene, M., and Rowe, C. (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., Parker, R. L. and Vernon, F. L. (2009). A Fortran 90 library for multitaper spectrum analysis, in *Computers and Geosciences,* 35: 1701–1710.

Ringdal, F., Harris, D. B., Kværna, T., and Gibbons, S. J. (2009). Expanding Coherent Array Processing to Larger Apertures Using Empirical Matched Field Processing, in *Proceedings of the 2009 Monitoring Research Review, Ground-Based Nuclear Explosion Monitoring Technologies*, LA-UR-09-05276, Vol. 1, pp. 379–388.

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.