

Expanding Coherent Array Processing to Larger Apertures Using Empirical Matched Field Processing

**Frode Ringdal
Steven J. Gibbons
David B. Harris
Tormod Kværna**

**NORSAR
PO Box 53
N-2027 Kjeller, Norway**

Final Report

17 Dec 2010

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION IS UNLIMITED.



**AIR FORCE RESEARCH LABORATORY
Space Vehicles Directorate
29 Randolph Rd
AIR FORCE MATERIEL COMMAND
HANSCOM AFB, MA 01731-3010**

NOTICES

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report was cleared for public release and is available to the general public, including foreign nationals. Qualified requestors may obtain copies of this report from the Defense Technical Information Center (DTIC) (<http://www.dtic.mil>). All others should apply to the National Technical Information Service.

AFRL-RV-HA-TR-2011-1001 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//signature//

ROBERT J. RAISTRICK
Contract Manager

//signature//

JILL PARKER, Maj, USAF, Acting Chief
Battlespace Surveillance Innovation Center

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) 17-12-2010		2. REPORT TYPE Final Report		3. DATES COVERED (From - To) 20-11-2007 to 19-11-2010	
4. TITLE AND SUBTITLE Expanding Coherent Array Processing to Larger Apertures Using Empirical Matched Field Processing				5a. CONTRACT NUMBER FA8718-08-C-0007	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER 62601F	
6. AUTHOR(S) F. Ringdal, S. Gibbons, D.B. Harris and T. Kværna				5d. PROJECT NUMBER 1010	
				5e. TASK NUMBER SM	
				5f. WORK UNIT NUMBER A1	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) NORSAR Gunnar Randers vei 15 N-2027 Kjeller, Norway				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Research Laboratory 29 Randolph Rd. Hanscom AFB, MA 01731-3010				10. SPONSOR/MONITOR'S ACRONYM(S) AFRL/RVBYE	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S) AFRL-RV-HA-TR-2011-1001	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited.					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT This final report summarizes the results of a three-year collaborative project undertaken by NORSAR, Lawrence Livermore National Laboratory, and Deschutes Signal Processing to explore the applicability of empirical matched field processing (EMFP) to the detection and classification of seismic signals over increasing sensor apertures. EMFP uses the observations of historical events to calibrate the amplitude and phase structure of an incident wavefield over a given sensor configuration for particular repeating sources. We have demonstrated an excellent source classification both for mines on the Kola Peninsula, NW Russia and in Kazakhstan using calibrations constructed from multiple Ground Truth events. A multitaper procedure has been developed for evaluating covariance matrices over short and precisely defined time-series and, while there are advantages to characterizing source regions with ensemble covariance matrices, single-event calibrations appear to work well in many cases meaning that the method is applicable also to source regions with few observations. We have demonstrated single-phase EMFP to be a viable event detector. Given only single observations or rank-1 matched field statistics, it appears that an increase in the receiver aperture makes a calibration specific to a more limited source region. Higher rank matched field statistics may make broader regions of diffuse seismicity amenable to coherent processing over wider sensor apertures. Considering the aftershock sequence from the M=7.4 October 8, 2005, Kashmir event, we suggest a partially coherent procedure whereby 3-component matched field statistics from the individual stations of KNET can be stacked to form a robust detector of larger events from a broad source region.					
15. SUBJECT TERMS Seismic arrays, Empirical Matched Field Processing, Seismic detection, Seismic classification					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT SAR	18. NUMBER OF PAGES 89	19a. NAME OF RESPONSIBLE PERSON Robert Raistrick
a. REPORT UNC	b. ABSTRACT UNC	c. THIS PAGE UNC			19b. TELEPHONE NUMBER (include area code) 781-377-3726

Table of Contents

Summary	1
1. Introduction	3
2. Formulation	11
3. Using Empirical Matched Field Processing as a Source Classifier for Repeating Industrial Events: Case Study - The Kola Peninsula	15
4. Using Empirical Matched Field Processing as a Source Classifier for Repeating Industrial Events: Case Study - Central Kazakhstan	29
5. Matched Field Processing as a Tool for Detection	43
6. Improving Phase Identification in Regions of Diffuse Seismicity	53
7. Examination of an Aftershock Sequence using Empirical Matched Field Processing	65
8. Conclusions	77
References	79
Appendix A: Calculating Spectral Covariance Matrices Using the Multitaper Method	81
List of Symbols, Abbreviations, and Acronyms	89

Figures

1.1	Locations of mines in the Khibiny and Olenegorsk regions of the Kola Peninsula and geometry of the ARCES array	6
1.2	Classification of 549 events from mines on the Kola Peninsula using only the Pn arrival at ARCES	7
1.3	Schematic view of matched field procedures for multiple array processing of regions of diffuse seismicity	9
3.1	Locations of sites of repeating seismicity near the ARCES array	16
3.2	Slowness estimates for initial P-arrivals at ARCES for sites of repeating seismicity	18
3.3	Image of the Zapoljarni mines from Google Earth	19
3.4	Slowness estimates (fully automatic and fixed frequency band) for P-arrivals from events at the two Zapoljarni mines	20
3.5	Separation using Empirical Matched Field Processing between the ZP1 and ZP2 events in the time-period 2003-2004	23
3.6	Similarity matrices for calibration events from the ZP1 and ZP2 mines (2001-2002)	24
3.7	Ward linkage dendrograms for ZP1 and ZP2 calibration events	25
3.8	ARCES waveforms for the 2001-2002 ZP2 calibration events	26
3.9	Separation using Empirical Matched Field Processing between the ZP1 and ZP2 events in the time-period 2003-2004 with alternative empirical steering vectors for the ZP2 cluster	27
4.1	Arrays and seismicity in Kazakhstan	30
4.2	Location estimates for three mining clusters from Hartse et al. (2008)	31
4.3	Application of Google Earth for the identification of candidate sites for clusters	31
4.4	Location estimates in KNDC bulletin for correlating clusters of mining events	32
4.5	A three minute long data segment on the Makanchi array	33

4.6	Slowness vector estimates at MKAR for mining explosions in Central Kazakhstan	34
4.7	Geometry of the MKAR and ARCES arrays	35
4.8	Distance matrix for Pn-arrivals at MKAR for cluster 328	36
4.9	Ward dendrogram for MKAR Pn-arrivals from cluster 328	37
4.10	Reordered similarity matrix for cluster 328	38
4.11	Reordered matched field similarity matrix over all three clusters	39
4.12	Separation using EMPF between events from clusters 310, 328, and 344 using both Rank-1 and Rank-2 matched field statistics	40
5.1	Location of the Kara Zhyra mine in Eastern Kazakhstan with respect to MKAR and ZALV	44
5.2	Matched field statistics at ZALV and MKAR for Pn arrivals from an event assumed to originate close to the Kara Zhyra mine	45
5.3	Matched field statistics as a function of data-window positioning for two events from the Kara Zhyra mine recorded at Makanchi	47
5.4	Matched field statistics as a function of time of day from empirical steering vectors from a single event at the Kara Zhyra mine	49
6.1	Seismicity well observed on both MKAR and KKAR arrays	53
6.2	MKAR Pn matched field similarity matrix for selected region of diffuse seismicity	54
6.3	Location estimates in the KNDC and IDC reviewed event bulletins for the events displayed in Figure 6.2	55
6.4	Location estimates in the ISC bulletin and the current study for the events displayed in Figure 6.2	56
6.5	Waveforms at station KBL in Afghanistan for event 2008-035:20.12.12 (lat 36:504 N, 70:157 E, depth ~ 200km)	57
6.6	Variability of the waveforms over KNET from the event 2008-035:20.12.12 (lat 36:504 N, 70:157 E, depth ~ 200km).	58

6.7	KKAR Pn arrival phase and coherence co-array at 2 Hz for two similar events in a region of diffuse seismicity with predicted pattern	59
6.8	KKAR Pn arrival phase and coherence co-array at 4 Hz for two similar events in a region of diffuse seismicity with predicted pattern	60
6.9	Matched field statistic for theoretical and empirical steering vectors on the KKAR array (wide band)	60
6.10	Matched field statistic for theoretical and empirical steering vectors on the KKAR array (narrow band)	61
6.11	Matched field statistics from theoretical (blue) and empirical (red) steering vectors in narrow frequency bands	62
7.1	Location of the M=7.4 October 8, 2005, Kashmir event mainshock (large red circle) with respect to KKAR, KNET, and the NIL 3-C station	65
7.2	Simulated helicorder plot of the KK01 short period vertical sensor on October 8, 2005	66
7.3	KK01 waveform with wavefield attributes for detecting Pn phases from the source region	68
7.4	As for Figure 7.3 except that we consider the time-interval surrounding the main shock	70
7.5	Continuous traces of matched field statistics with ranks 1, 2, and 3 where the ensemble covariance matrix is generated from the 20 best matches with the data segment starting 2005-281:05.37.06.7	71
7.6	Waveforms over KNET (filtered) 2.0-8.0 Hz from an aftershock of the October 8, 2005, Kashmir event	72
7.7	3-component matched field detection statistic traces for the NIL and AAK stations, together with an (incoherent) beam of the corresponding traces across KNET	73
7.8	Comparison of coherent and incoherent matched field detection statistic beams for KNET	75

Tables

3.1	List of sites of repeating seismicity considered displayed in Figure 3.1 with backazimuth and distances from ARCES	16
3.2	List of Zapoljarni Ground Truth events 2001-2002 used for generating the covariance matrices from which the empirical steering vectors are derived	21
3.3	List of Zapoljarni Ground Truth events 2003-2004 used for evaluating the performance of empirical matched field processing for source identification	22

Acknowledgments

We thank the Kazakhstan National Data Center for permission to use array data from the Kazakh stations and to Terri Hauk at Lawrence Livermore National Laboratory for preparing the data.

Data from KNET, KBL, KURK, and NIL was obtained from the IRIS Data Management Center with data coming from the KN, IU, KZ, and II networks respectively.

We are grateful to Hans Hartse for providing event lists and results of correlation analysis for mining events in Kazakhstan.

All maps are produced using GMT software (Wessel & Smith, 1995).

SUMMARY

This final report summarizes the results of a three-year collaborative project undertaken by NORSAR, Lawrence Livermore National Laboratory, and Deschutes Signal Processing. The objective has been to explore the applicability of empirical matched field processing (EMFP) to the detection and classification of seismic signals over increasing sensor apertures. EMFP uses the observations of historical events to calibrate the amplitude and phase structure of an incident wavefield over a given sensor configuration for particular repeating sources. The matching statistics are calculated using empirical steering vectors such that the plane wavefront model, which breaks down over wide apertures, is avoided. Since EMFP is a narrow-band procedure, it is relatively insensitive to source-time history which is an advantage over correlation detectors for the characterization of mining seismicity. On the small aperture ARCES array, we have demonstrated an excellent source classification for mines on the Kola Peninsula using calibrations constructed from multiple Ground Truth events. This study has highlighted the need to perform cluster analysis on the calibration events prior to the construction of ensemble covariance matrices. A similar performance has been demonstrated on sparser, larger aperture arrays in Kazakhstan at frequencies at which classical f-k analysis demonstrably fails. In this study, the separation between different source regions was improved by using higher rank matched field statistics which mitigate the effects of variability within the individual source regions.

A multitaper procedure has been developed for evaluating covariance matrices over short and precisely defined time-series and, while there are advantages to characterizing source regions with ensemble covariance matrices, single-event calibrations appear to work well in many cases meaning that the method is applicable also to source regions with few observations. We advocate extending f-k analysis in pipeline operations to include both theoretical and empirical steering vectors to improve classification and parameter estimates for calibrated sources. We have demonstrated single-phase EMFP to be a viable event detector. Given only single observations or rank-1 matched field statistics, it appears that an increase in the receiver aperture makes a calibration specific to a more limited source region. Higher rank matched field statistics may make broader regions of diffuse seismicity amenable to coherent processing over wider sensor apertures. Considering the aftershock sequence from the M=7.4 October 8, 2005, Kashmir event, we suggest a partially coherent procedure whereby 3-component matched field statistics from the individual stations of KNET can be stacked to form a robust detector of larger events from a broad source region. Using the events detected, more sensitive but more source-specific fully-coherent classifiers can then be spawned.

1. INTRODUCTION

Nuclear explosion monitoring is undergoing a steady progression toward event detection, location and identification at lower magnitudes. As the threshold for detection and characterization has decreased, the distances at which monitoring takes place have decreased correspondingly from originally teleseismic, to regional (< 2000 km) and now to local (< 200 km) ranges. The consequent challenges to monitoring include far larger numbers of events to process and the need to interpret broader categories of events. The principal avenues to reducing detection and estimation thresholds are to increase the number of monitoring stations and to increase the signal bandwidth over which coherent processing algorithms can be applied. Once reduced thresholds are achieved, the next objective will be to interpret the flood of smaller events detected, especially in areas of high natural and man-made seismicity. A natural by-product of extending coherent processing to larger apertures will be a concomitant improvement in resolution. This effect will provide new tools for interpreting events by making it possible to assign them to very specific sources such as individual mines or aftershock sequences.

Current operational seismic array processing methods for detection (beamforming) and parameter estimation (frequency-wavenumber analysis) have changed little since their introduction in the 1960s and 1970s. These methods rely upon a plane-wave assumption for predicting the spatial amplitude and phase structure of the seismic waves incident upon an array aperture. Scattering and refraction in strongly heterogeneous seismic propagation media constrain the size of usable coherent processing apertures under the plane-wave assumption to a few wave-lengths. This constraint severely limits the spatial resolution of beamforming and frequency-wavenumber (FK) methods. In addition, the spatial correlation of ambient seismic noise constrains the minimum usable element separation. The combination of these constraints bounds the maximum number of usable sensors and places a fundamental limit on coherent processing gain through beamforming. Existing beamforming and FK estimation methods make some attempt to compensate for the non-ideal spatial structure of seismic waves. For example, it is common to apply an amplitude correction to estimated spatial covariance matrices when using the indirect approach to FK spectrum estimation (i.e. by estimating the the spatial covariance function first, then evaluating its Fourier transform to obtain a wavenumber spectrum). Each element of the matrix is normalized by the square root of the diagonal elements in the same row and column. In direct methods (e.g. Kværna & Ringdal, 1986), it is now standard to integrate the FK spectrum over a band of frequencies to stabilize the narrow-band estimates that individually have high variance. That variance is controlled, in part, by the very small time-bandwidth product of signals in short analysis windows. However, it is likely that variance in the FK estimate from frequency to frequency also is a function of wavefield scattering. Recent advances in standardization of analysis windows and frequency bands for particular source regions have produced dramatic reductions in the variance of azimuth/slowness estimates derived from FK spectra (Gibbons et al., 2005) at least in frequency bands (below 6-8 Hz) where the wavefield is approximately coherent across regional-array apertures. Finally, it is common to apply post-processing (i.e. post-FK) vector slowness corrections (Schweitzer, 2001) to

azimuth/slowness measurements in an attempt to remove frequency-dependent biases caused (presumably) by wavefield refraction.

In our view, such slowness corrections, while valuable, come too late in the processing sequence. Rather than to apply post-FK corrections to the slowness estimate derived from the strongly non-linear process of maximizing a sample FK spectrum, we believe it is more effective to apply pre-FK corrections to the signal observed across the array aperture. This approach has much in common with adaptive optics used in astronomy to compensate optical wavefield perturbations caused by heterogeneity and turbulence in the atmosphere. Recently, detection algorithms that exploit aperture-level calibrations have been developed that often substantially reduce detection thresholds in beamforming operations. Correlation detectors (Gibbons & Ringdal, 2006) and subspace detectors (Harris, 2006) use previously-observed waveforms from events at specific sources in sensitive algorithms to detect subsequent events at those same sources. These algorithms rely upon the fact that the temporal and spatial structures of signals from these sources often are repeatable; past events constitute a spatio-temporal calibration across an array or network of sensors for future events. Such detectors simultaneously detect, locate and identify events as being consistent with previous events at the same source location, with the same source time history and mechanism.

However, while the full exploitation of the spatial and temporal structure of the signal frequently leads to detectors one to two orders of magnitude more sensitive than simple power detectors (with conventional beamforming), such exploitation also presents two barriers to widespread application. Correlation methods generally have a source-region geographic footprint one to two wavelengths across (Harris, 1991), which can be ameliorated to some extent with the more general waveform representation of a subspace detector (Harris, 2006). Correlation methods correct incoherence across the receiver aperture, but trade it for incoherence across the source region; the reciprocity principle is at play in the correlation detection strategy. Correlation methods also are sensitive to source mechanism and time history, and can fail when these attributes differ from event to event.

The sensitivities of FK methods and conventional beamforming to the plane wave assumption and of correlation methods to source variability leads us to consider another processing strategy based on temporally-incoherent, but spatially coherent signal processing extended with subspace techniques. Our general approach is to adapt narrowband empirical matched field processing, originally conceived for underwater acoustic applications (Baggeroer et al., 1993; Fialkowski et al., 2000) to the seismic monitoring task. Our processing agenda consists of (1) breaking the observed signal into narrow bands, (2) using empirical matched field methods to combine the near-monochromatic resultant waveforms coherently across an aperture to achieve processing gain, and (3) integrating the resulting power incoherently over the narrow bands to achieve a wideband result. We anticipate that this approach will ameliorate variations in source time history from event to event that defeat waveform correlation methods. Matched field processing derives its name from the realization that conventional narrow-band beamforming is a spatial matched filtering operation for plane waves, and that more

general forms of spatial matched filtering are possible. In underwater acoustics, the velocity structure of the medium is far less variable than the seismic propagation medium and frequently known in detail. In such circumstances, it is possible to compute the detailed phase and amplitude structure of a narrowband wavefield incident on an array and apply the computed structure as the focusing kernel in a beamforming operation. This approach works well in the SOFAR channel waveguide and allows coherent processing across arrays to be expanded beyond the physical aperture limits implied by the assumption of a planar structure (locally the field is planar). A clever empirical approach to estimating the structure of the narrowband acoustic wavefield directly from observed wavefields was attempted (Fialkowski et al., 2000) but largely failed due to the spatial non-stationarity of the source.

Under a previous DoE/NNSA funded contract, "Integrated Seismic Event Detection and Location by Advanced Array Processing" (Kværna et al., 2007), it was demonstrated that empirical matched field processing (EMFP) was able to distinguish between the signals on the ARCES array from very closely-spaced quarry blasts with a very high classification success-rate. The procedure and results have subsequently been documented in greater depth by Harris & Kværna (2010). The locations of the sources investigated are displayed in Figure 1.1 in relation to the anticipated resolution of the array. It is clear, from theoretical plane wavefront considerations, that the array should not be able to resolve between the different source regions. This is confirmed in Figure 1.2, where it is also demonstrated that the matched field statistic based classification identifies the correct source mine in essentially all cases. A comparison with classification by standard waveform correlation methods demonstrates the superiority of EMFP of waveform dissimilarity from signal to signal (resulting in particular from complicated source-time functions) appear to be mitigated by the narrow frequency band nature of the matched field processing formulation.

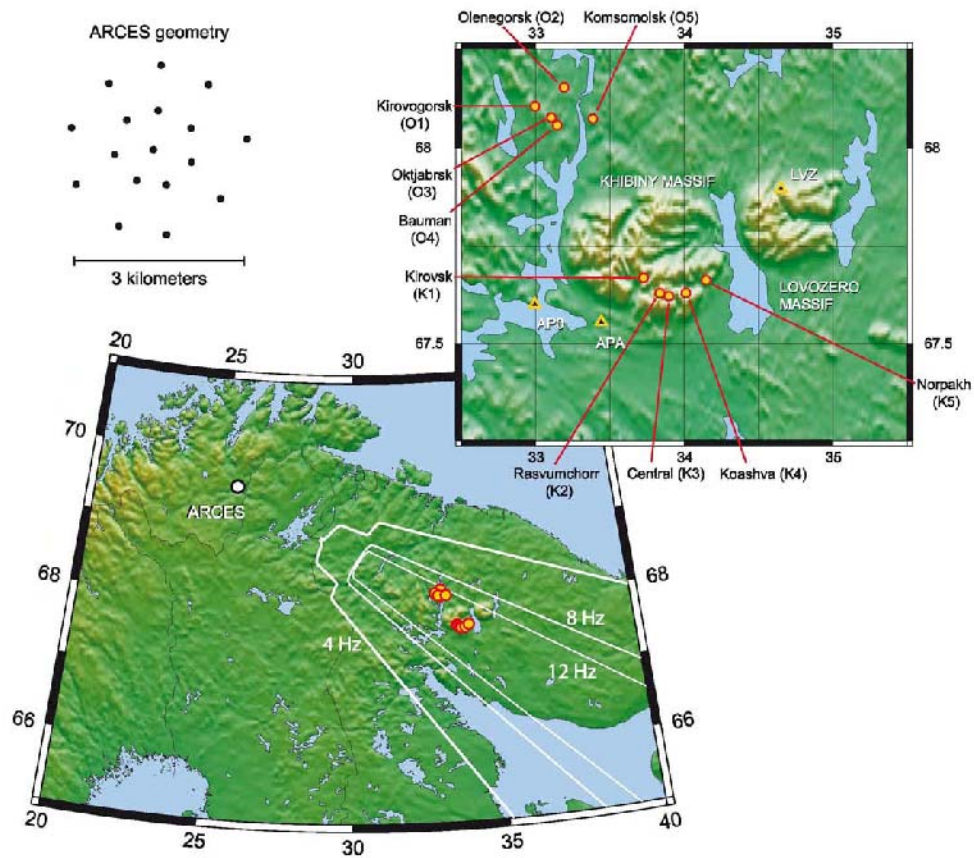


Figure 1.1. Locations of mines in the Khibiny and Olenegorsk regions of the Kola Peninsula together with the geometry of the ARCES array (central element, C-, and D-rings only) and projections of the half-power contours of the array response function at the frequencies displayed for a Pn beam steered towards the Rasvumchorr mine on the Khibiny Massif. Figure modified from Harris & Kværna (2010).

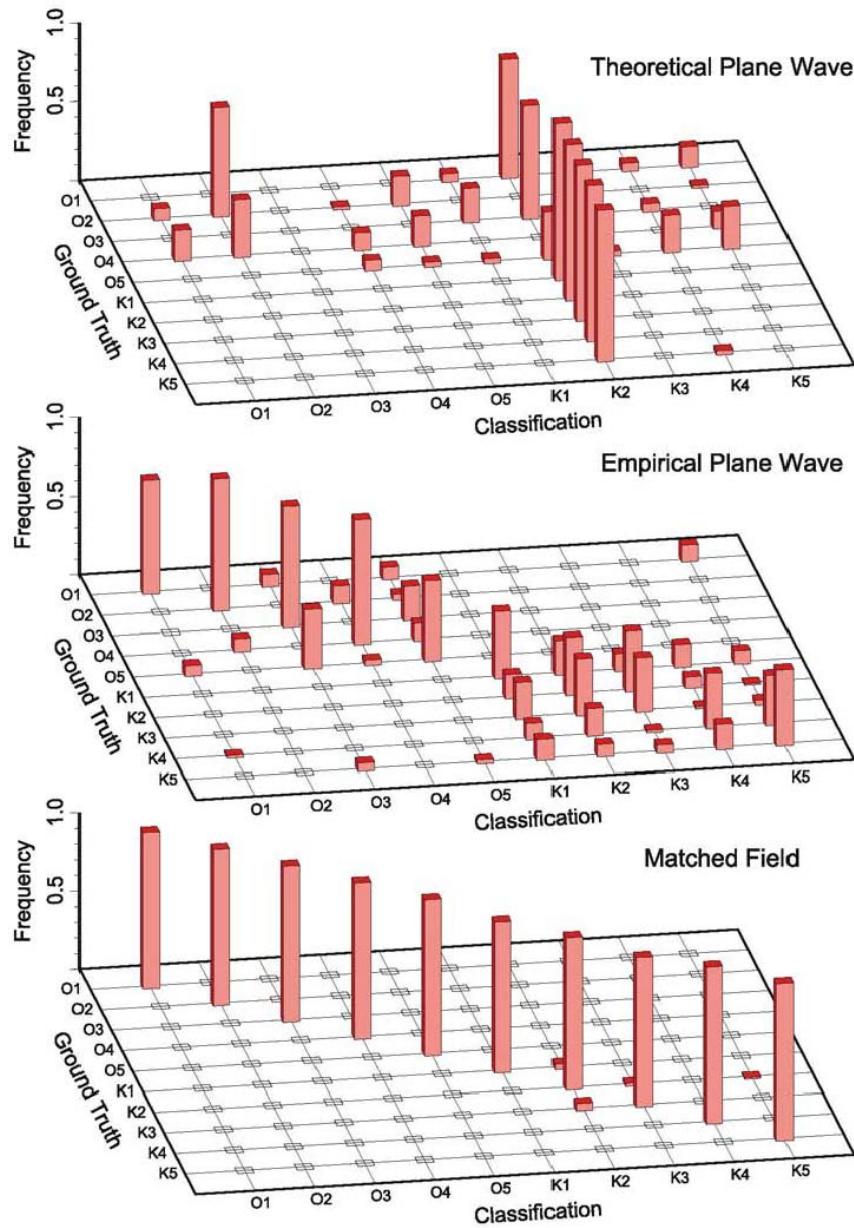


Figure 1.2. Classification of 549 events from mines on the Kola Peninsula using only the Pn arrival at ARCES. The theoretical plane-wave estimates (uncorrected f-k analysis, top panel) are unable to resolve the source mines. Applying slowness corrections (calibrated f-k analysis, middle panel) separates the two principal clusters well and shows some degree of success at separating the more sparse individual mines. The matched field method (EMFP, bottom panel) separates all mines with a 98.2 per cent success rate. Figure taken from Harris & Kværna (2010).

The only case attempted so far was that of compact source regions and a single small-aperture seismic array. 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.

Figure 1.3 provides a schematic representation of how EMFP may be expected to be applied to a band of diffuse seismicity. For each of several arrays, a calibration is built from existing observations of events from many overlapping regions. Chapter 2 defines the mathematical formalism for the procedures carried out during this study, and the five subsequent chapters each outline a case study to address a particular angle of enquiry. Chapter 3 describes a case study of using matched field processing on the ARCES array to classify events from two closely spaced mines near Zapoljarni on the Kola Peninsula. The key issue to be examined here is the calculation of the multi-event matched field steering vectors. Chapter 4 examines sets of industrial mining events in Central Kazakhstan recorded by the Makanchi array (MKAR). MKAR is larger and sparser than ARCES with the consequence that coherent estimation using the plane-wave model is unsuccessful at the high frequencies for which SNR is optimal. For this case, we also examine the use of higher order matched field statistics.

Chapter 5 expands the application of EMFP beyond the separation of alternative source hypotheses to the examination of matched field statistics evaluated on continuous incoming multichannel waveform data for event identification (detection). Issues that also need to be addressed here include the sensitivity to the setting of the data window, and a comparison of single-array and network processing. Chapter 6 examines the use of EMFP to event classification within a band of diffuse seismicity. In particular, we examine the comparison between empirical matched field and plane-wave model classification as a function of frequency. The concept of diffuse seismicity is widened to examine an extensive aftershock sequence from a large earthquake in Chapter 7, observed by both a medium aperture seismic array and a large network.

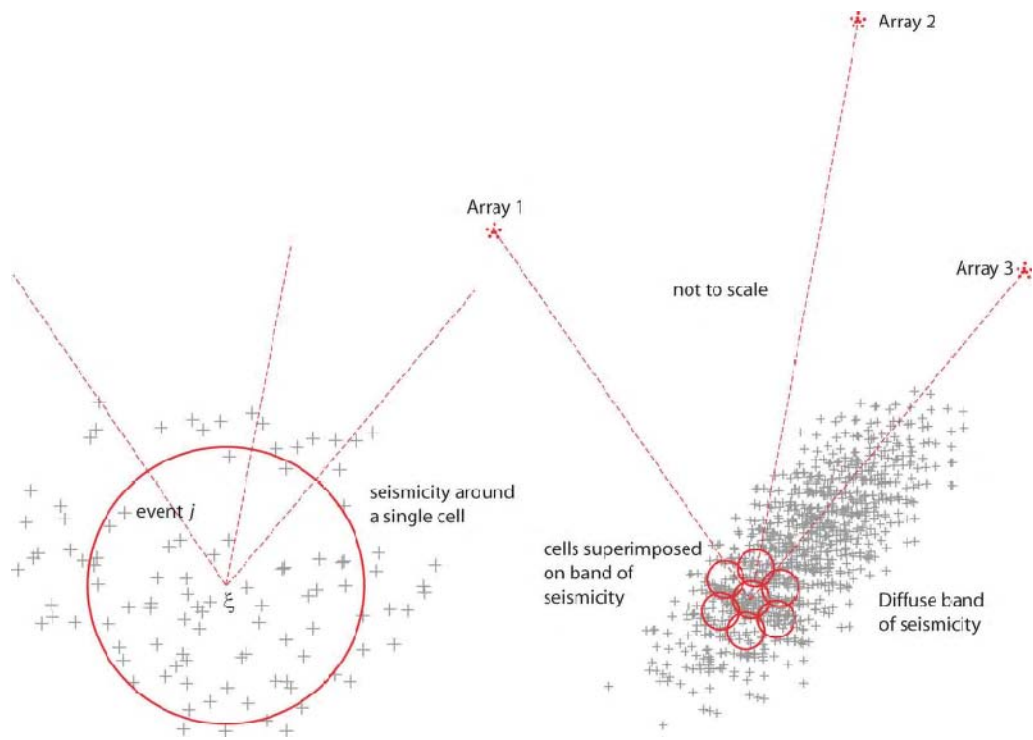


Figure 1.3. Schematic view of matched field procedures for multiple array processing of regions of diffuse seismicity.

2. FORMULATION

Throughout this study, we will consider exclusively seismic data from arrays and networks. If $\underline{x}_m$ denotes the coordinates of sensor m relative to some reference site, then we denote the time-series recorded at this site by

$$r_m(t) = r(t, \underline{x}_m) \quad (2.1)$$

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, $\underline{R}(\omega)$, of which the element R_{mn} is given by

$$R_{mn}(\omega) = \left(\int r(t, \underline{x}_m) e^{-i\omega t} dt \right) \left(\int r(t, \underline{x}_n) e^{i\omega t} dt \right) \quad (2.2)$$

Given the relative transience of seismic signals, it is understood that $\underline{R}(\omega)$ 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 $\underline{g}(\omega) = [s_1(\omega), \dots, s_N(\omega)]^T$ 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

$$\beta(\underline{g}(\omega)) = \frac{\underline{g}(\omega)^H \underline{R}(\omega) \underline{g}(\omega)}{\text{tr}[\underline{R}(\omega)]} \quad (2.3)$$

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

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:

$$\beta(\underline{g}([\omega_1, \omega_K])) = \sum_{k=1}^K a_k \frac{\underline{g}(\omega_k)^H \underline{R}(\omega_k) \underline{g}(\omega_k)}{\text{tr}[\underline{R}(\omega_k)]} \quad (2.4)$$

where $\underline{g}([\omega_1, \omega_K])$ 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:

$$\sum_{k=1}^K a_k = 1 \quad (2.5)$$

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

$$\hat{t}_m = \hat{t}_0 - \underline{s}^T \underline{x}_m \quad (2.6)$$

where $\hat{t}_0$ is the arrival time at the array reference site (with coordinate vector $\underline{x}_0$) and the slowness vector $\underline{s} = (s_x, s_y)$ is related to the backazimuth Θ and the apparent velocity v_{app} by

$$s_x = s \sin(\Theta), \quad s_y = s \cos(\Theta), \quad \text{and} \quad s = 1/v_{app} \quad (2.7)$$

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

$$\underline{s}(\omega, \underline{s}) = [e^{-i\omega \hat{t}_1}, \dots, e^{-i\omega \hat{t}_M}]^T \quad (2.8)$$

and the statistic evaluated over our K specified frequencies is

$$\hat{P}(\underline{s}([\omega_1; \omega_K], \underline{s})) = \sum_{k=1}^K \alpha_k \frac{\underline{s}(\omega_k, \underline{s})^H \underline{R}(\omega_k) \underline{s}(\omega_k, \underline{s})}{\text{tr}[\underline{R}(\omega_k)]} \quad (2.9)$$

Deviations of the wavefield from the predicted plane wavefront model will lead to a reduction in the value of the expression in Equation 2.9. The steering vector, $\underline{s}(\omega)$, which optimizes the value of the quadratic form in Equation 2.4 is the principal eigenvector of the matrix $\underline{R}(\omega)$. In empirical matched field processing (EMFP), 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 $\underline{R}(\omega, \alpha)$ is measured for an arrival from a source of interest (denoted by α) then the principal eigenvector $\underline{s}(\omega, \alpha)$ 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 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, $\underline{R}(\omega, \alpha)$, from the single arrival covariance matrices, $\underline{R}(\omega, \alpha_i)$, with

$$\underline{R}(\omega, \alpha) = \sum_{\alpha_i \in \alpha} \omega_i \underline{R}(\omega, \alpha_i) \quad (2.10)$$

where the ω_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.

Given that we may normalize any covariance matrix $\underline{R}(\omega)$, 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

$$\hat{P}(\underline{s}([\omega_1: \omega_R], \alpha)) = \sum_{k=1}^R a_k \underline{s}(\omega_k, \alpha)^H \underline{R}(\omega_k) \underline{s}(\omega_k, \alpha) \quad (2.11)$$

and

$$\hat{P}(\underline{s}([\omega_1: \omega_R], \alpha_i)) = \sum_{k=1}^R a_k \underline{s}(\omega_k, \alpha_i)^H \underline{R}(\omega_k) \underline{s}(\omega_k, \alpha_i) \quad (2.12)$$

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

Estimates of the spatial structure of a given wavefield which are constructed from multiple observations may result in high-rank covariance matrix estimates. In such cases, 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 single 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 $\underline{s}(\omega_k)$ then has to be a linear combination of the eigenvectors corresponding to the L largest eigenvectors of the template covariance matrix: $\alpha_1, \dots, \alpha_L$. If the α_i form the columns of the $N \times L$ steering matrix $\underline{E}$, with

$$\underline{s}(\omega_k) = \underline{E} \underline{v} \quad (2.13)$$

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

3. USING MATCHED FIELD PROCESSING AS A SOURCE CLASSIFIER FOR REPEATING INDUSTRIAL EVENTS: CASE STUDY - THE KOLA PENINSULA

While the intended focus region for this project is central Asia, it is highly beneficial to evaluate the methods proposed on datasets within a geographical region which has already been examined in great detail. Figure 3.1 displays the locations of sites of recurring anthropogenic seismicity close to the ARCES seismic array; the coordinates are provided, together with the distances and directions from ARCES in Table 3.1. A previous Department of Energy/NNSA sponsored collaboration between NORSAR and Lawrence Livermore National Laboratory (“Integrated Seismic Event Detection and Location by Advanced Array Processing”, contract number DE-FC52-03MA99517) addressed the degree to which repeating sources at regional distances from a small-aperture seismic array could be separated using more established forms of array processing (Kværna et al., 2007). Central to this project was a large database of mining Ground-Truth information for a number of sites on the Kola Peninsula of NW Russia, collected under a previous DoE-sponsored project (Harris et al., 2003). In addition, the use of multi-channel waveform correlation detectors (Gibbons & Ringdal, 2006) was able to expand greatly the set of source regions to which the study could be applied, providing “implicit Ground Truth information” for additional sites for which only a small number of confirmed events were known.

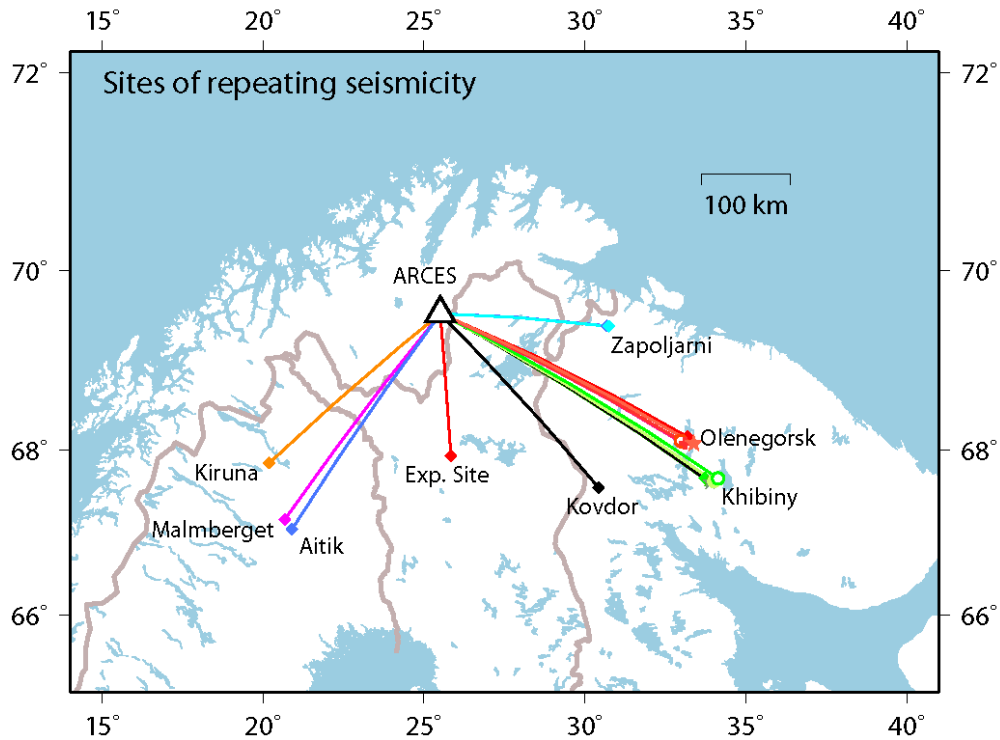


Figure 3.1. Locations of sites of repeating industrial or military seismicity within 500 km of the ARCES small-aperture seismic array in northern Norway. The Swedish underground mines at Kiruna and Malmberget are run by the LKAB company, and the copper quarry at Aitik is operated by the company Boliden. The site in northern Finland is used by the Finnish military to destroy outdated ammunition. The mining regions Zapoljarni, Olenegorsk, and Khibiny on the Kola Peninsula each contain several units within short distances of each other.

Table 3.1. List of sites of repeating seismicity considered displayed in Figure 3.1 with backazimuth and distances from ARCES

Identifier	Site Name	Latitude (degrees N)	Longitude (degrees E)	BAZ (ARCES)	Distance (km)	Description
ZP1	Zapadni (Zapoljarni)	69.404	30.682	91.7	203	Quarry
ZP2	Central (Zapoljarni)	69.397	30.742	91.8	205	Quarry
OL1	Olenegorsk (Olenegorsk)	68.154	33.192	112.8	346	Quarry
OL2	Kirovogorsk (Olenegorsk)	68.106	32.996	114.3	341	Quarry
OL3	Bauman (Olenegorsk)	68.057	33.145	114.5	349	Quarry
OL4	Oktjabrsk (Olenegorsk)	68.078	33.106	114.3	347	Quarry
OL5	Komsomolsk (Olenegorsk)	68.075	33.385	113.4	356	Quarry
KH1	Kirovsk (Khibiny)	67.670	33.729	118.0	393	Mine
KH3	Rasvumchorr (Khibiny)	67.631	33.835	118.0	400	Mine
KH4	Central (Khibiny)	67.624	33.896	118.0	403	Quarry
KH5	Koashva (Khibiny)	67.632	34.011	117.5	406	Quarry
KH6	Norpakh (Khibiny)	67.665	34.146	116.6	409	Quarry
KV1	Kovdor	67.557	30.425	135.4	298	Quarry
FES	Finnish Explosion Site	67.934	25.832	175.6	179	Military site
AIT	Aitik	67.060	20.900	216.7	335	Quarry
MAL	Malmberget	67.179	20.675	219.4	329	Mine
KIR	Kiruna	67.849	20.196	231.4	286	Mine

The work described by Kværna et al. (2007) confirmed that a large degree of variability in event location estimates was due to the measurement of backazimuth and slowness using broadband f-k analysis in variable frequency bands, as is generally the case for the routine detection and location procedures. Figure 3.2 displays slowness estimates for initial regional P-arrivals for selected events confirmed to have taken place at the sites listed in Table 3.1. The top panel in Figure 3.2 displays the slowness estimates obtained from the fully-automatic phase detection lists for ARCES. A spread of several degrees is observed for arrivals from all of the sites shown, in addition to a clear and persistent bias for a number of sites (in particular the Aitik quarry, shown in dark blue, and the Olenegorsk and Khibiny regions, shown in red and green).

Performing the f-k analysis in the fixed frequency band 2-4 Hz results in a significant improvement for the arrivals from many sites. Whereas the estimates for phases from the Khibiny and Olenegorsk regions are seen to overlap in the top panel, they form almost distinct populations in the lower panel, signaling an improvement in our ability to differentiate signals from the two regions using traditional array processing. The improvement is, however, not universal with the spread in slowness estimates at some sites being far larger for the 2-4 Hz band than for the variable bands; this is most likely the effect of a low Signal-to-Noise-Ratio (SNR) in this band. It has been demonstrated that many of the sites which perform poorly in the 2-4 Hz band perform far better in the 4-8 Hz band, and that in general no single frequency band provides optimal performance for all sites of interest (see, for example, Kværna et al., 2004). This is another motivation for studying the multiple narrow frequency band technique being investigated under the current contract.

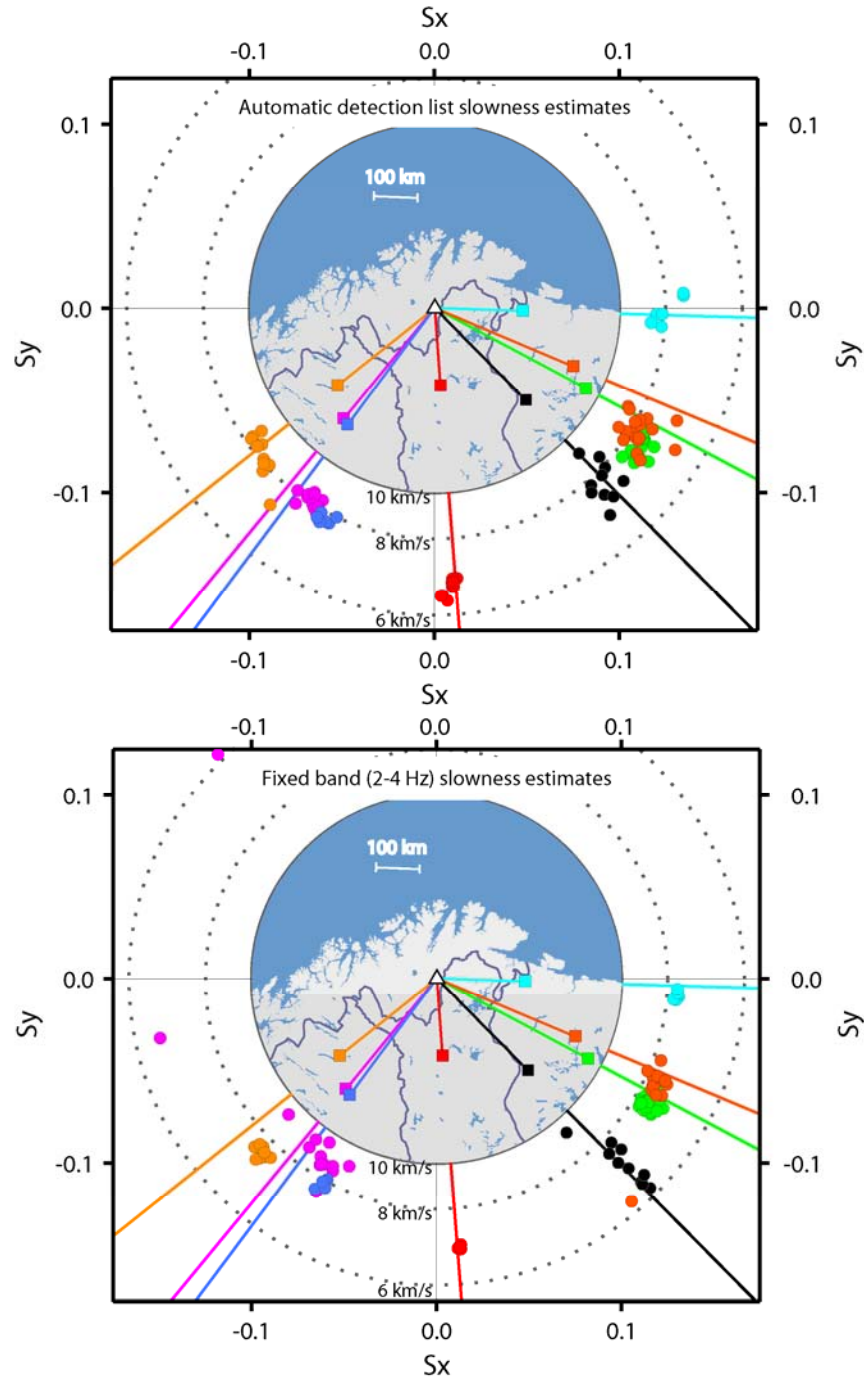


Figure 3.2: Slowness estimates for initial P-arrivals at ARCER for events confirmed to have taken place at the sites displayed in Figure 3.1 selected for analyst review at NORSAR. The upper panel displays the slowness vector indicated from the fully-automatic detection lists (f-k analysis performed in variable frequency bands) and the lower panel displays the results for the same set of arrivals using fixed-band re-estimation (2-4 Hz). The inset maps are for illustrative purposes, showing the true azimuthal lines from ARCER to each of the mines, also shown in Figure 3.1.



Figure 3.3: Relative locations (from Google Earth) of the Zapadny (ZP1) and Central (ZP2) mines close to the town of Zapoljarni on the Kola Peninsula in northwest Russia. The coordinates provided for ZP1 are 69.404°N, 30.682°E, and the coordinates provided for ZP2 are 69.397°N, 30.742°E. The distance between the ZP1 and ZP2 mine symbols is approximately 2.5 km.

There are two quarries close to the town of Zapoljarni for which the mining authorities provided the times of routine explosions to colleagues at the Kola Regional Seismological Center (KRSC) in Apatity. The Zapadny and Central mines (denoted ZP1 and ZP2 respectively, see Figure 3.3) are separated by only 2.5 kilometers, at a distance of approximately 205 km from ARCES. The geographical backazimuths from the ARCES array to the reference coordinates provided for the ZP1 and ZP2 mines are 91.7 and 91.8 degrees respectively. These sites are too close to each other for the ARCES array to be able to differentiate between the two sources by any conventional wavefield interpretation (Harris & Kværna, 2010). (It must also be assumed that the extent of each of the two mines is not insignificant compared with the distance between the reference coordinates provided, meaning that the actual distance between events occurring within the two mine complexes in many cases may be shorter than 2.5 km.)

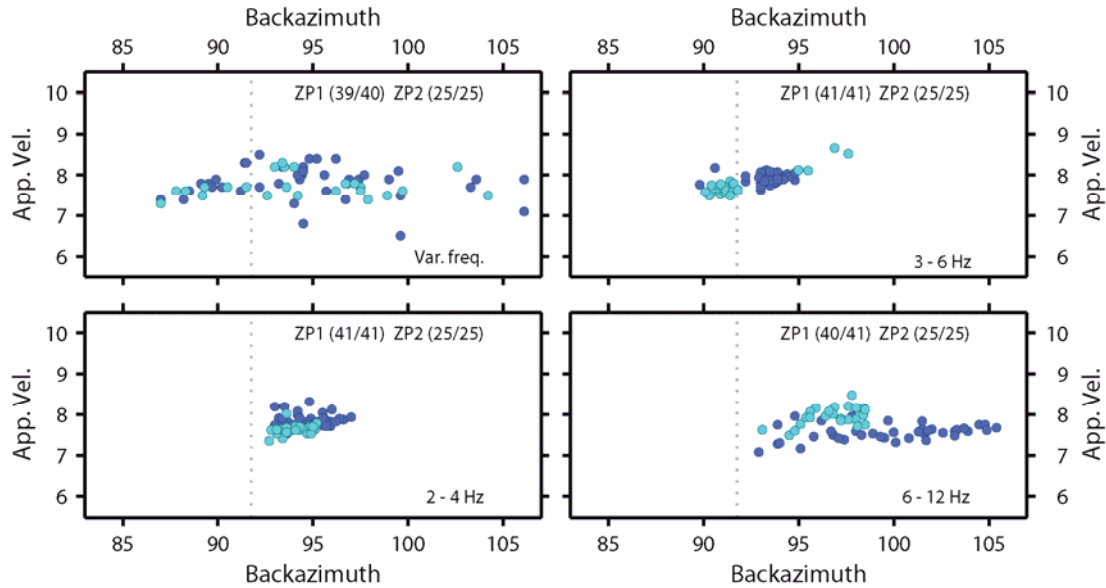


Figure 3.4: Slowness estimates for initial P-arrivals at ARCES from events at two mines ZP1 and ZP2 (see Figure 3.3) between October 2001 and September 2002. Estimates in the top left panel are from the fully automatic ARCES detection lists (variable frequency bands) and the remaining panels show estimates for the same arrivals processed in fixed frequency bands as indicated. The dotted vertical line at 91.7° backazimuth is the approximate great-circle backazimuth from ARCES. ZP1 and ZP2 symbols are dark and light blue respectively. A more complete analysis is provided in Gibbons et al. (2010).

Figure 3.4 displays slowness estimates for initial P-arrivals from confirmed events at the ZP1 and ZP2 mines over a period of one year. Origin times for the events are provided in Table 3.2. In the top left panel, it is clear that the azimuth and slowness estimates obtained for the two event populations in variable frequency bands occupy a broad region of parameter space. Given the coordinates of a parameter estimate for an arrival from an unknown source, it would not be possible to say to which population the phase is more likely to belong. For the three fixed frequency bands displayed (2-4 Hz, 3-6 Hz, and 6-12 Hz), not only is the variance of the two populations reduced, the two populations appear to form relatively distinct clusters in the velocity/azimuth parameter space displayed. While the clusters are not entirely distinct, such that one could not say without doubt to which population an unknown estimate belonged, the pattern for the two populations in each frequency band covers a reasonably characteristic region of parameter space. It is then clear that the inferred azimuth and slowness are affected by characteristics of the wavefield which are not modelled in the plane-wave formulation.

Table 3.2. List of Zapoljarni Ground Truth events 2001-2002 used for generating the covariance matrices from which the empirical steering vectors are derived

ZP1 : Zapadni		ZP2 : Central	
Evt. number	ARCES P time	Evt. number	ARCES P time
1	2001 283 11 26 24.613	1	2001 278 10 59 23.038
2	2001 292 11 05 03.913	2	2001 290 11 12 26.038
3	2001 299 11 01 24.363	3	2001 304 12 12 53.113
4	2001 306 12 10 32.763	4	2001 318 12 05 25.213
5	2001 313 12 11 51.738	5	2001 325 12 13 19.063
6	2001 320 12 41 13.613	6	2001 339 12 15 09.838
7	2001 327 12 49 57.413	7	2001 353 12 10 44.288
8	2001 332 12 21 03.413	8	2001 360 12 39 22.038
9	2001 341 12 02 11.963	9	2002 004 12 09 33.363
10	2001 355 12 11 35.738	10	2002 011 12 01 54.638
11	2001 362 12 05 53.763	11	2002 018 12 04 16.613
12	2002 016 13 01 38.863	12	2002 025 11 51 54.313
13	2002 023 12 39 22.238	13	2002 032 11 59 18.013
14	2002 030 12 21 01.063	14	2002 039 12 14 13.288
15	2002 037 12 03 32.913	15	2002 046 11 57 06.863
16	2002 044 12 06 10.588	16	2002 051 12 02 41.088
17	2002 053 13 02 38.988	17	2002 060 11 57 21.263
18	2002 053 13 04 33.263	18	2002 072 12 23 43.963
19	2002 058 11 54 57.513	19	2002 095 11 04 55.938
20	2002 072 12 24 31.688	20	2002 102 10 59 48.513
21	2002 074 12 19 32.613	21	2002 109 11 10 46.938
22	2002 079 12 31 14.663	22	2002 114 11 28 15.113
23	2002 088 12 15 10.188	23	2002 226 10 52 19.563
24	2002 093 11 02 24.038	24	2002 249 11 07 48.713
25	2002 107 11 03 24.538	25	2002 254 11 02 30.963
26	2002 114 11 35 10.438		
27	2002 128 11 00 18.913		
28	2002 137 11 05 15.613		
29	2002 144 11 00 16.613		
30	2002 165 11 15 03.613		
31	2002 172 11 05 02.863		
32	2002 179 11 05 12.238		
33	2002 191 11 34 36.988		
34	2002 198 11 06 13.338		
35	2002 205 11 06 25.988		
36	2002 219 11 18 00.788		
37	2002 228 11 04 05.988		
38	2002 233 11 00 41.738		
39	2002 242 11 06 29.788		
40	2002 263 11 12 23.913		
41	2002 270 10 55 59.113		

Table 3.3. List of Zapoljarni Ground Truth events 2003-2004 used for evaluating the performance of empirical matched field processing for source identification

ZP1 : Zapadni		ZP2 : Central	
Evt. number	ARCES P time	Evt. number	ARCES P time
1	2003 199 11 28 48.913	1	2003 183 10 58 53.488
2	2003 206 11 14 12.963	2	2003 185 11 03 30.963
3	2003 211 11 20 00.038	3	2003 192 10 54 32.913
4	2003 232 11 11 51.338	4	2003 197 11 38 56.363
5	2003 248 11 02 20.013	5	2003 218 11 01 11.788
6	2003 267 11 20 40.688	6	2003 225 11 35 19.713
7	2003 269 12 00 23.388	7	2003 234 11 39 50.063
8	2003 276 11 19 58.163	8	2003 241 11 08 50.788
9	2003 281 11 08 18.063	9	2003 255 11 25 50.588
10	2003 295 11 27 45.063	10	2003 260 11 11 45.163
11	2003 309 12 07 27.163	11	2003 262 11 17 50.563
12	2003 330 12 11 06.875	12	2003 274 11 11 58.038
13	2003 353 12 20 25.025	13	2003 290 11 14 39.138
14	2003 360 12 12 55.100	14	2003 297 11 10 51.638
15	2004 009 12 24 52.775	15	2003 302 12 09 44.463
16	2004 016 12 45 51.650	16	2003 316 12 43 00.238
17	2004 044 12 21 51.325	17	2003 323 12 32 26.138
18	2004 051 12 49 00.450	18	2003 325 12 16 27.988
19	2004 056 12 51 55.400	19	2003 332 12 05 55.800
		20	2003 337 12 29 41.850
		21	2003 344 12 53 03.950
		22	2003 351 12 38 35.575
		23	2003 358 12 32 51.250
		24	2003 364 12 07 49.150
		25	2004 014 12 14 30.025
		26	2004 021 12 39 00.825
		27	2004 028 12 15 16.100
		28	2004 035 12 50 02.300
		29	2004 042 12 14 55.925
		30	2004 049 12 06 22.800
		31	2004 058 12 52 57.425

Table 3.3 provides a list of Ground Truth events for the Zapoljarni mines in a time period distinct from that for the events listed in Table 3.2. We intend first to construct ensemble covariance matrices $\underline{R}_{ZP1}$ and $\underline{R}_{ZP2}$ (c.f. Equation 2.10) for the two mines ZP1 and ZP2 measured from initial P-arrivals for the events listed in Table 3.2. From the two covariance matrices we calculate the corresponding empirical steering vectors, $\underline{s}_{ZP1}$ and $\underline{s}_{ZP2}$. For each P-arrival from the events listed in Table 3.3, we measure the covariance matrix $\underline{R}$ and calculate two corresponding empirical matched field statistics:

$$\hat{P}(\underline{s}_{ZP1}) = \underline{s}_{ZP1}^H \underline{R} \underline{s}_{ZP1} \quad (3.1)$$

and

$$\hat{P}(\underline{s}_{ZP2}) = \underline{s}_{ZP2}^H \underline{R} \underline{s}_{ZP2} \quad (3.2)$$

(c.f. Equation 2.11). The aim is to investigate how successfully the events in Table 3.3 are classified based upon the size of the two values $\hat{P}(\underline{s}_{ZP1})$ and $\hat{P}(\underline{s}_{ZP2})$.

Template: 2001-2002 GT events
2.5-12.5 Hz
Events: 2003-2004 GT events

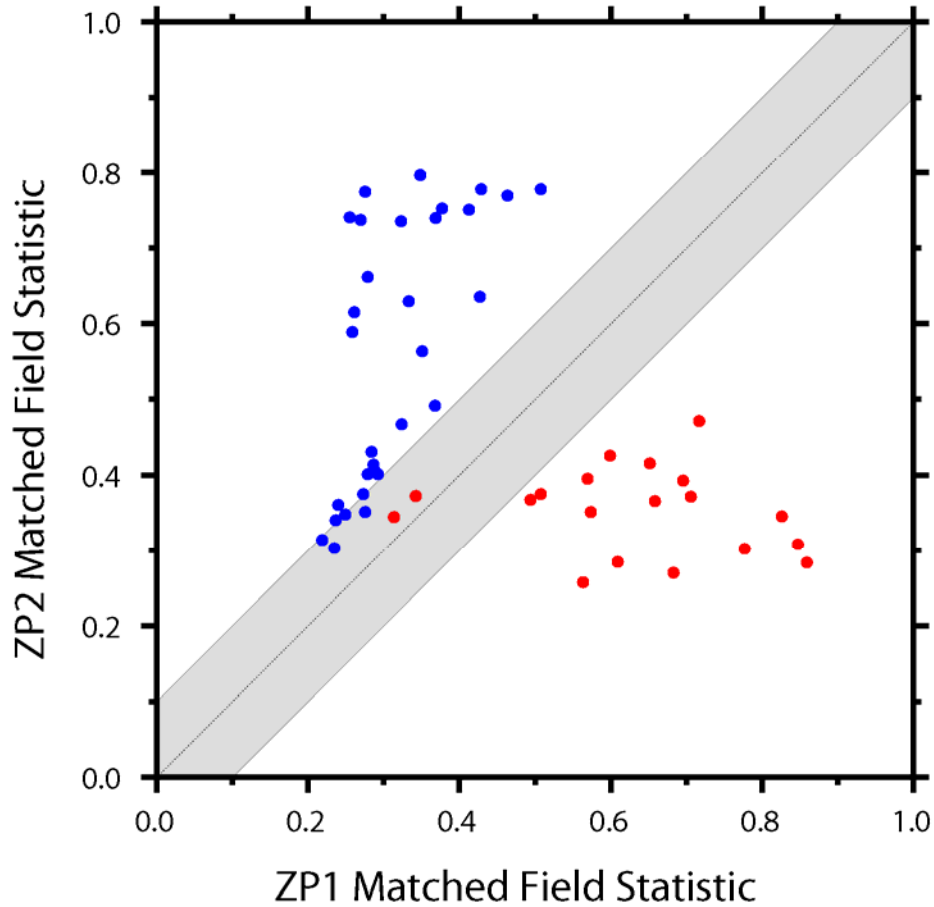


Figure 3.5: $\hat{P}(\underline{s}_{ZP1})$ versus $\hat{P}(\underline{s}_{ZP2})$ for events attributed to the ZP1 mine (red symbols) and the ZP2 mine (blue symbols) in the period 2003-2004 (see Table 3.3). The grey zone along the diagonal indicates a region where the difference between $\hat{P}(\underline{s}_{ZP1})$ and $\hat{P}(\underline{s}_{ZP2})$ is not great, indicating a less certain classification. The estimates are made in the broad frequency band 2-5 Hz to 12.5 Hz. Only the central element of ARCES, the C-ring and D-ring are used.

The values of $\hat{P}(\underline{s}_{ZP1})$ and $\hat{P}(\underline{s}_{ZP2})$ for the 2003 and 2004 Zapoljarni events are plotted in Figure 3.5 where the colour of the symbol indicates the mine the event is attributed to from the Ground Truth information. There is a clear separation in the $(\hat{P}(\underline{s}_{ZP1}) \text{ and } \hat{P}(\underline{s}_{ZP2}))$ parameter space between the two populations with all but two of the events identified as coming from the same source mine as the Ground Truth

attribution. There are however many symbols occupying the grey-shaded region suggesting that several of the classifications are associated with significant uncertainty.

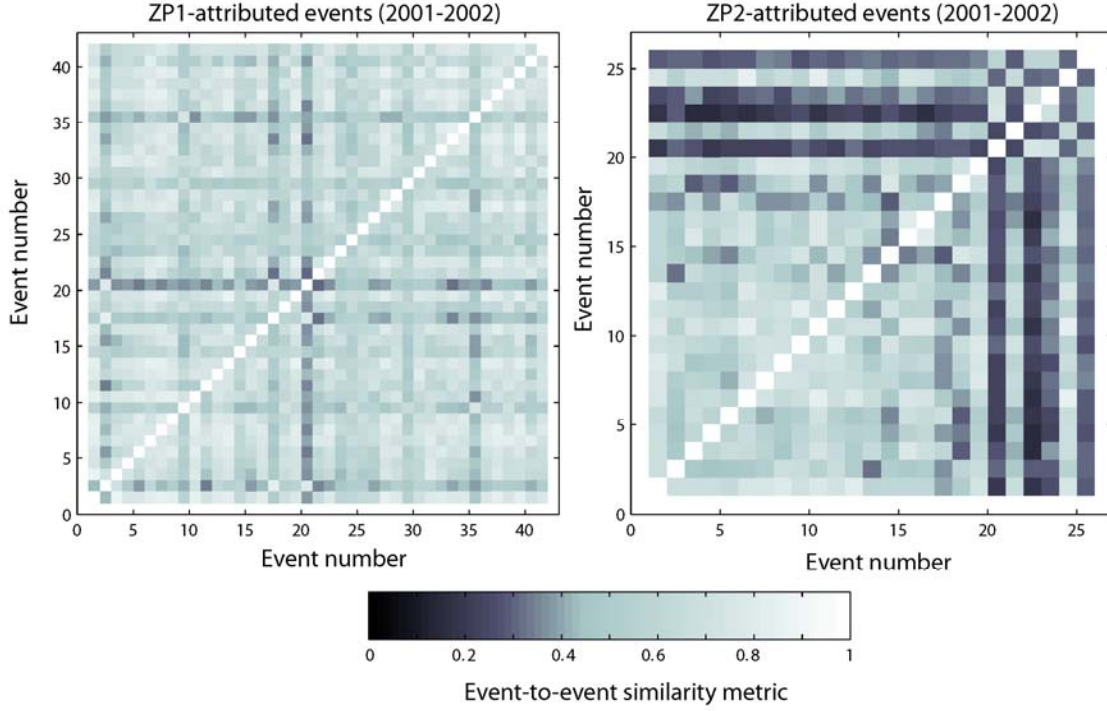


Figure 3.6: Similarity metrics (c.f. Equation 3.3) between pairs of events from the ZP1 and ZP2 calibration event populations in the time-period 2001-2002 (see Table 3.2).

We examine the two populations of events from which the empirical steering vectors were calculated (Table 3.2). For each event i , we calculate the covariance matrix $\underline{R}_i(\omega_k)$ for frequency band k and its fully normalized principal eigenvector $\underline{s}_{ik}$. Over the N_B frequency bands, the similarity metric

$$d_{ij} = \frac{1}{N_B} \sum_{k=1}^{N_B} |\underline{s}_{ik}^H \underline{s}_{jk}| \quad (3.3)$$

between events i and j measures the projection of one phase's matching field onto the other and is normalized to range between 0 (no projection) and 1 (identical matching fields). The d_{ij} are displayed for each event pair from the ZP1 and ZP2 calibration event populations in Figure 3.6.

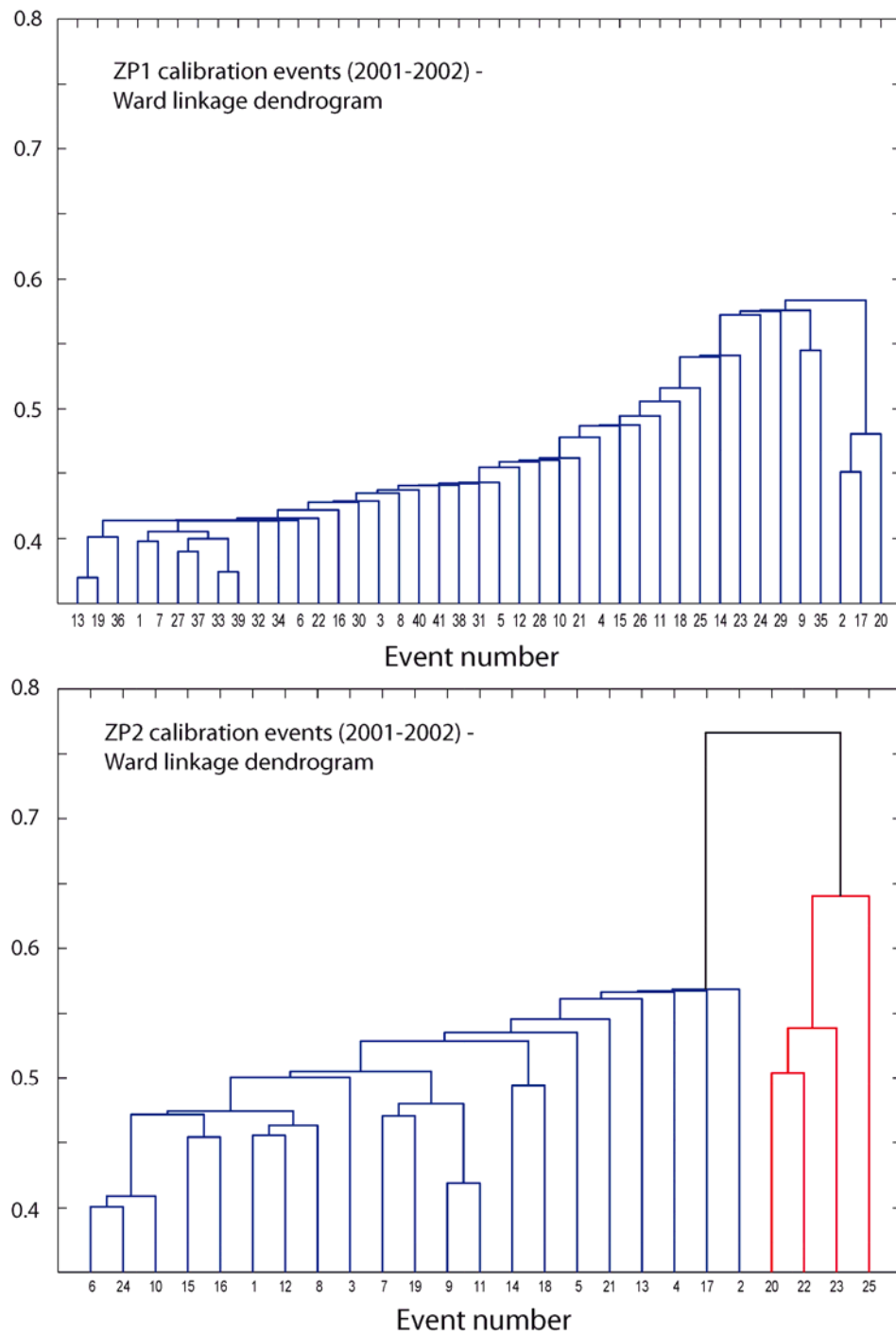


Figure 3.7: Ward linkage dendrograms for ZP1 and ZP2 calibration events as indicated. For the set of ZP2 events, two clusters are defined based upon a similarity threshold of 0.7.

The population of ZP1 calibration events (left panel, Figure 3.6) is considerably more homogeneous than the population of ZP2 calibration events (right panel, Figure 3.6). This is illustrated further by calculating dendrograms for the two mines (Figure 3.7) using the agglomerative clustering linkage algorithm of Ward (1963). Using a similarity level of 0.7 for defining event clusters, we find that the ZP2 calibration events can be subdivided into two groups. The smallest of these groups consists of only 4 events (numbers 20, 22, 23, and 25) displayed in red in Figure 3.8. It is also worth noting that these four events all occur after April 2002 (see Table 3.2).

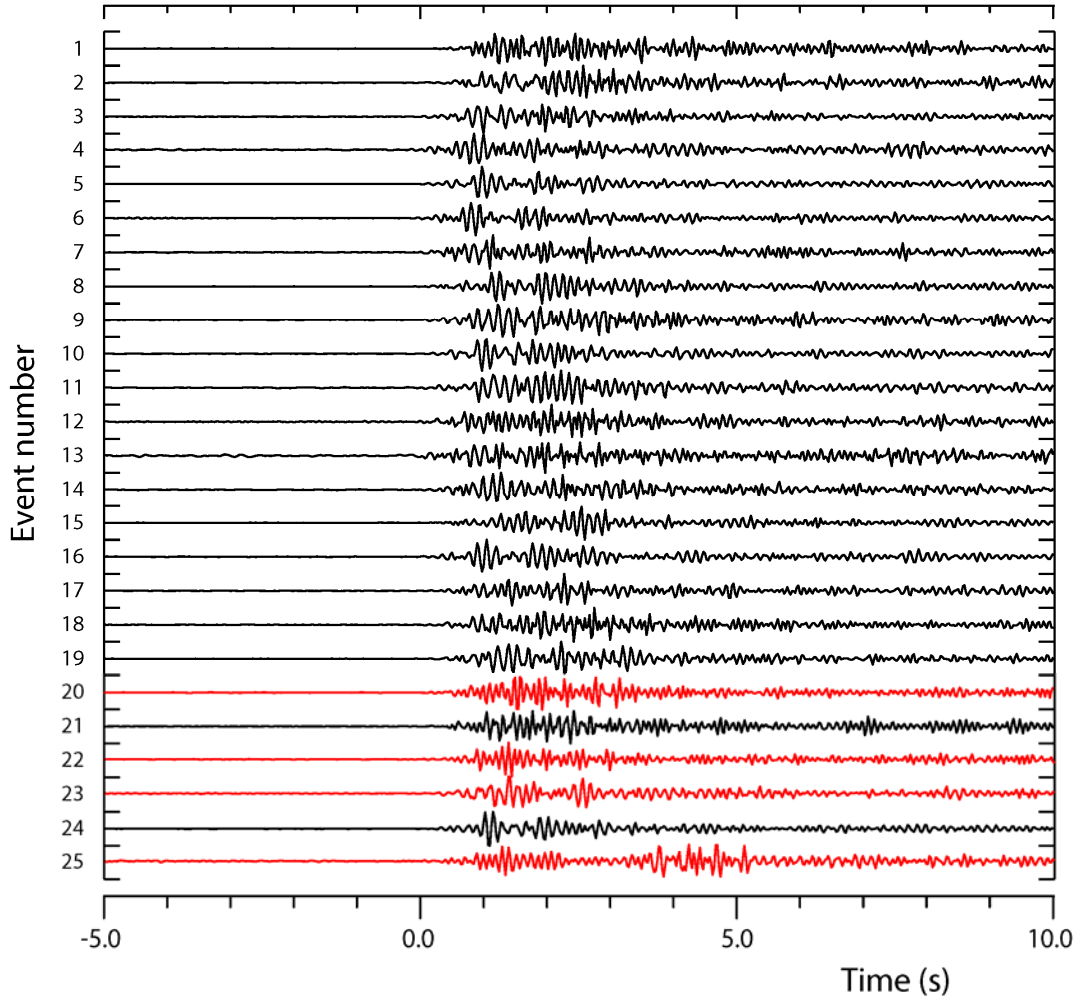


Figure 3.8: Waveforms (initial region P-wave arrivals) at the ARCES center instrument ARA0 sz for the 2001-2002 ZP2 calibration events listed in Table 3.2. The data are filtered in the passband 2.5 - 12.5 Hz, also used in the Matched-Field processing. The events of the split calibration datasets are shown by the black (ZP2 sub1) and the red (ZP2 sub2) traces, respectively.

We repeat the classification procedure displayed in Figure 3.5 except that, instead of a matching statistic for a single ZP2 empirical steering vector, we calculate a matched field statistic for two different steering vectors - one from each of the two clusters demonstrated - and take the ZP2 classifier as the maximum of the two values. The outcome is displayed in Figure 3.9. The separation between the populations of ZP1 and ZP2 events is far better than that displayed for the single ZP2 calibration (Figure 3.5). It is clear that the creation of an ensemble covariance matrix and subsequent empirical steering vector for the entire population of ZP2 events has actually degraded the classification performance. A far better performance is achieved when the ZP2 calibration events are separated on the basis of cluster analysis prior to the formation of covariance matrices and empirical steering vectors.

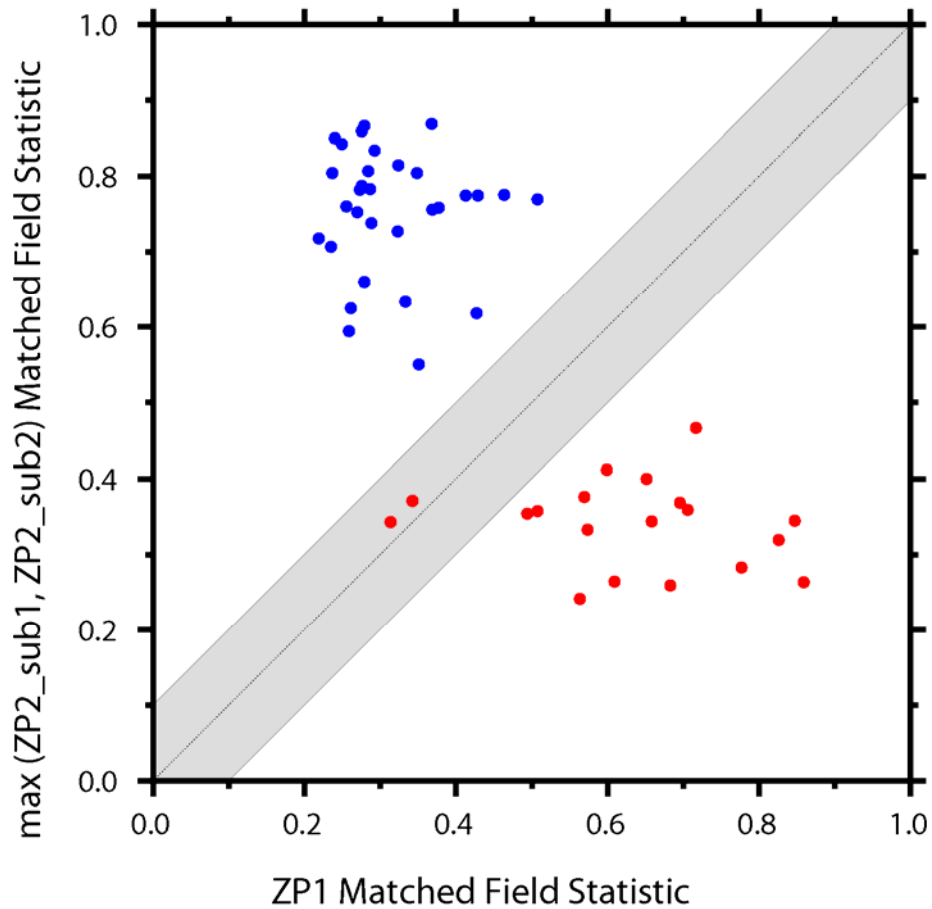


Figure 3.9: Classification of 2003 and 2004 events from ZP1 and ZP2 mines using empirical matched field processing. The procedure followed is the same as for Figure 3.5 except that the ZP2 matched field statistic, $\hat{P}(\mathbf{s}_{ZP2})$, is replaced by the maximum of the two values $\hat{P}(\mathbf{s}_{ZP2_sub1})$ and $\hat{P}(\mathbf{s}_{ZP2_sub2})$.

Summary

We have demonstrated the ability of empirical matched field processing to distinguish between two different sources of mining explosions separated by a distance of approximately 2 km using an array station at a distance of approximately 200 km. From theoretical considerations of a plane wavefront assumption, the ARCES array should not be able to differentiate between P-arrivals from the Zapadny (ZP1) and Central (ZP2) mines at Zapoljarni on the Kola Peninsula.

Broadband f-k analysis in fixed frequency bands does provide a quasi-distinction between confirmed events from the two mines, demonstrating first and foremost the significance of scattering in the arriving wavefront. We calculate ensemble covariance matrices for the two mines from confirmed calibration events over a 12 month period and obtain empirical steering vectors (the principal eigenvectors of the ensemble covariance matrices). We then attempt to classify a second set of Ground Truth events from the two mines, distinct from the calibration dataset, by calculating matched field statistics based upon the two empirical steering vectors. The resulting classification is very successful with only two events misclassified. It is noted however that, for many events, the correct mine is chosen over the incorrect mine quite marginally.

Cluster analysis is performed within the sets of calibration events for both mines, by considering a similarity metric between the principal eigenvectors from each single-event covariance matrix. The resulting dendrogram identifies four events from the ZP2 mine which appear quite different in nature to the other 21 training events. We calculate two separate ensemble covariance matrices for the two subgroups of ZP2 training events and re-classify subsequent events according to the maximum of three different matched field statistics: one for the ZP1 mine and two for the ZP2 mine. This new reclassification is greatly improved with a far greater separation between the two populations. Calculating an ensemble covariance matrix is likely to reduce the variability in the estimates for the phase and amplitude relations for a given source region, and capture generic information about the source to receiver Green's function which may evade observation given only a single event. However, we have demonstrated that calculating an ensemble covariance matrix from a group of calibration events without examining the associated similarity matrix can significantly reduce the matched field classification performance. We therefore advocate performing cluster analysis on all calibration events prior to calculating ensemble covariance matrices.

4. USING MATCHED FIELD PROCESSING AS A SOURCE CLASSIFIER FOR REPEATING INDUSTRIAL EVENTS: CASE STUDY - CENTRAL KAZAKHSTAN

The target region of the work carried out in this contract is central Asia, with a focus on Kazakhstan and the surrounding region. There are four 9-element seismic arrays within Kazakhstan, ABKAR, BVAR, KKAR, and MKAR, each with an aperture of approximately 4 km. These so-called hybrid arrays are based upon the same principles as ARCES (concentric rings with increasing radius) but, due to the combined effect of a larger aperture and fewer sensors, lack the short intersite separations which enable ARCES and similar small aperture arrays to process high frequency regional signals coherently. Also in the region is the Kurchatov cross-array (20 sites, aperture ~ 20 km), the KNET telemetry network in Kirgizstan (10 sites, aperture ~ 250 km), and a number of additional three-component stations. The seismic stations in the region comprise a broad spectrum of apertures over which empirical matched field processing can be evaluated and compared with existing procedures. The region is also of interest due to the high volume of seismicity, both anthropogenic and natural. Figure 4.1 displays the locations of the four 9-element Kazakh arrays together with a summary of reported seismicity and locations of known mines.

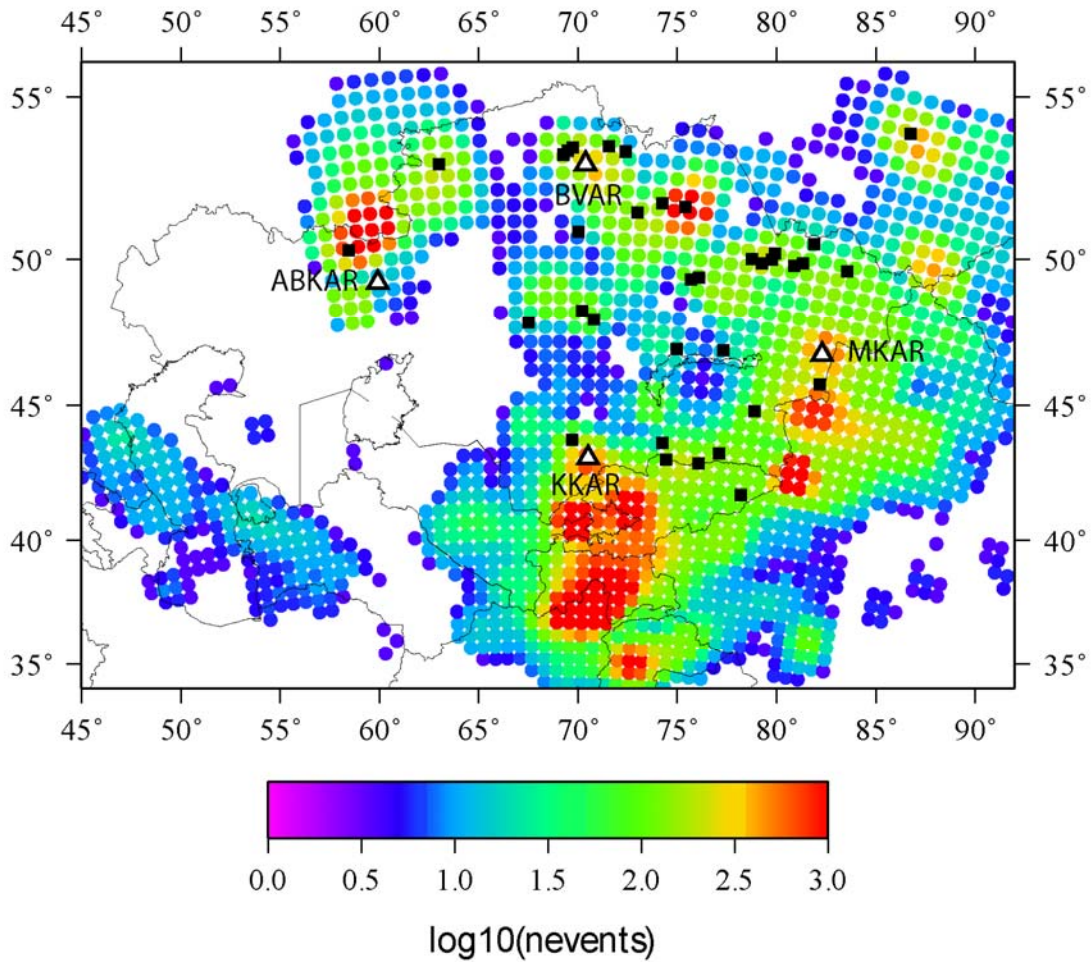


Figure 4.1: Locations of the four arrays ABKAR, BVAR, MKAR, and KKAR in Kazakhstan together with an illustration of the density of seismic events from the KNDC bulletin from 2005-001 to 2008-269. The “event bins” are coloured according to the number of events which have at least one defining P-phase from one of the four arrays shown. A total of 54443 events are contained within the bulletin during the time-period specified. The black symbols indicate the locations of known mines.

Unlike in the European Arctic, where we have received exceptional access to mining Ground Truth data, we have few instances of confirmed explosions at given sites. However, significant efforts have been made to classify sources of mining seismicity using waveform correlation methods in conjunction with satellite imagery (e.g. Hartse et al., 2008; MacCarthy et al., 2008). Figure 4.2 is taken from Hartse et al. (2008) and displays the locations of mines to which sets of mining blasts have been attributed. Figure 4.3 displays a satellite picture of a mine in one of the regions identified.

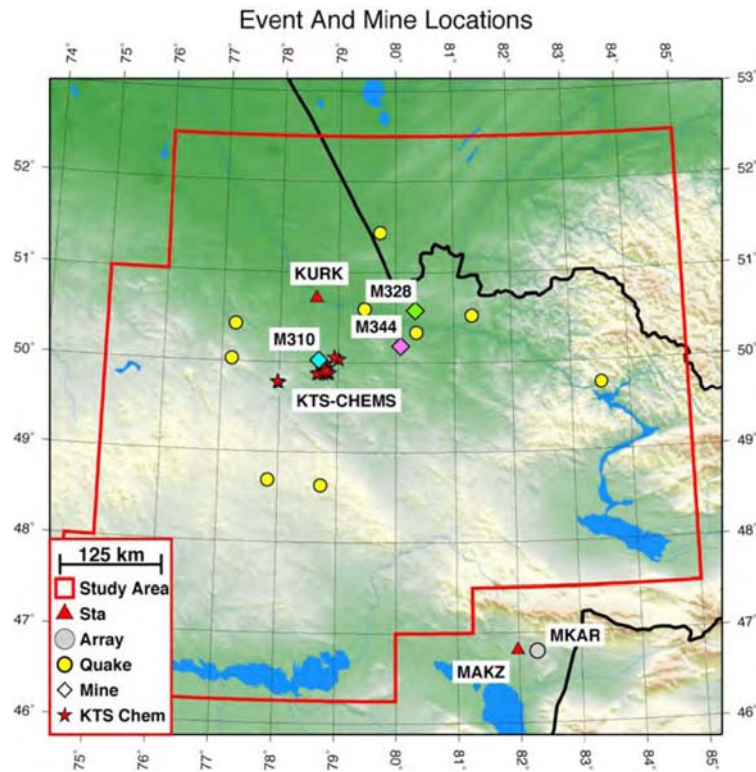


Figure 4.2: Location estimates for three mining clusters in eastern Kazakhstan obtained from correlation analysis by Hartse et al. (2008) and MacCarthy et al. (2008). Figure taken from Hartse et al. (2008).

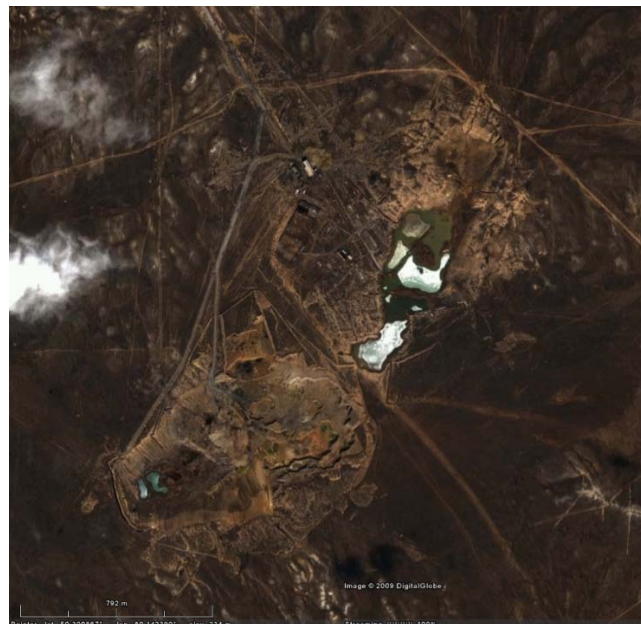


Figure 4.3: Application of Google Earth for the identification of candidate sites for clusters. This mine is in the vicinity of the location estimates for clusters 328 and 344 (c.f. Hartse et al., 2008).

Figure 4.4 illustrates the location estimates from the Kazakhstan National Data Center (KNDC) reviewed event bulletin (<http://www.kndc.kz/eng/index.php?p=0&f=data.html>) which correspond to events which Hartse et al. (2008) attributes to the clusters labelled 310, 328, and 344. The spread in location estimates emphasizes the need to apply additional processing methods such as EMFP. The waveform similarity constrains the source locations to be separated by no more than at most a few wavelengths at the dominant frequency; at most a few kilometers. The elongated uncertainty ellipses associated with the locations displayed in Figure 4.4 have dimensions of the order 100 km. Gibbons et al. (2010) attribute part of the considerable spread in automatic event location estimates for industrial events in the European Arctic to the practice of measuring slowness and azimuth at array stations in variable frequency bands. The same would apply to central Asia, although we are in addition restricted by the array geometries to processing in a lower frequency bands.

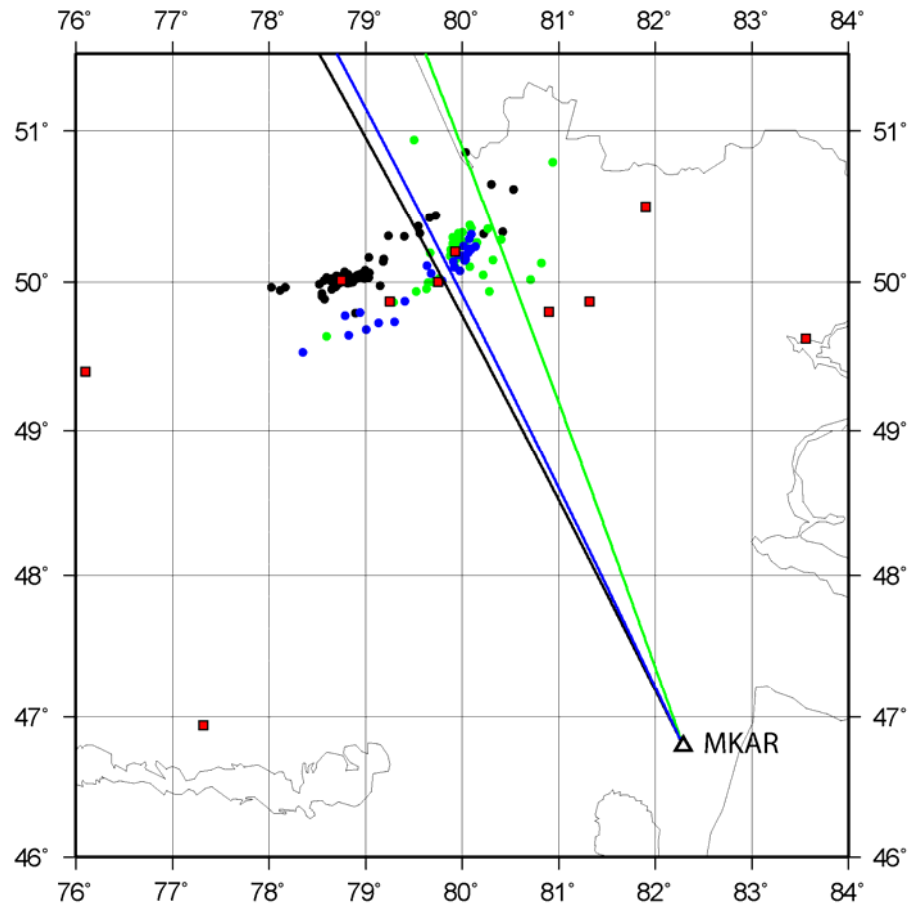


Figure 4.4: Location estimates for events from clusters 310 (black), 328 (green), and 344 (blue) from Hartse et al. (2008) as provided in the KNDC reviewed bulletin. Red symbols indicate the locations of known mines. The black, blue, and green lines give the median backazimuth estimates for the respective populations for the Pn phase arrivals at the MKAR array, whereby the estimate is made in the fixed frequency band 2.0 - 4.0 Hz. Clusters 310, 328, and 344 contain 65, 40, and 27 events respectively.

To illustrate, Figure 4.5 displays the signal from one of the industrial events classified by (Hartse et al., 2008) recorded on the Makanchi array (MKAR) at a distance of approximately 400 km. The waveform is filtered in a cascade of increasing frequency bands. While the secondary and Lg phases are best observed below 4 Hz, the SNR for the Pn phase (typically the best for parameter estimates: c.f. Figure 6 of Gibbons et al., 2010) is very poor below 4 Hz and increases steadily up to the cut-off frequency of the sensor. Figure 4.6 displays the consequences of the signal spectrum and array geometry for Pn slowness estimates using classical f-k analysis in two fixed frequency bands. In the upper row (2-4 Hz), the coherence over the array is relatively good while the SNR is poor. A majority of slowness estimates (top left panel) are in the correct region of slowness space, although the number of estimates which are not classified even qualitatively correct is significant. In the lower row (5-10 Hz), the SNR is high but the coherence between sensors is diminished, resulting in an even greater spread and poorer classification success. In neither frequency band is there any resolution in slowness space between Pn arrivals from the three different clusters. Figure 4.7 displays the geometry of MKAR together with the layout of ARCES for comparison.

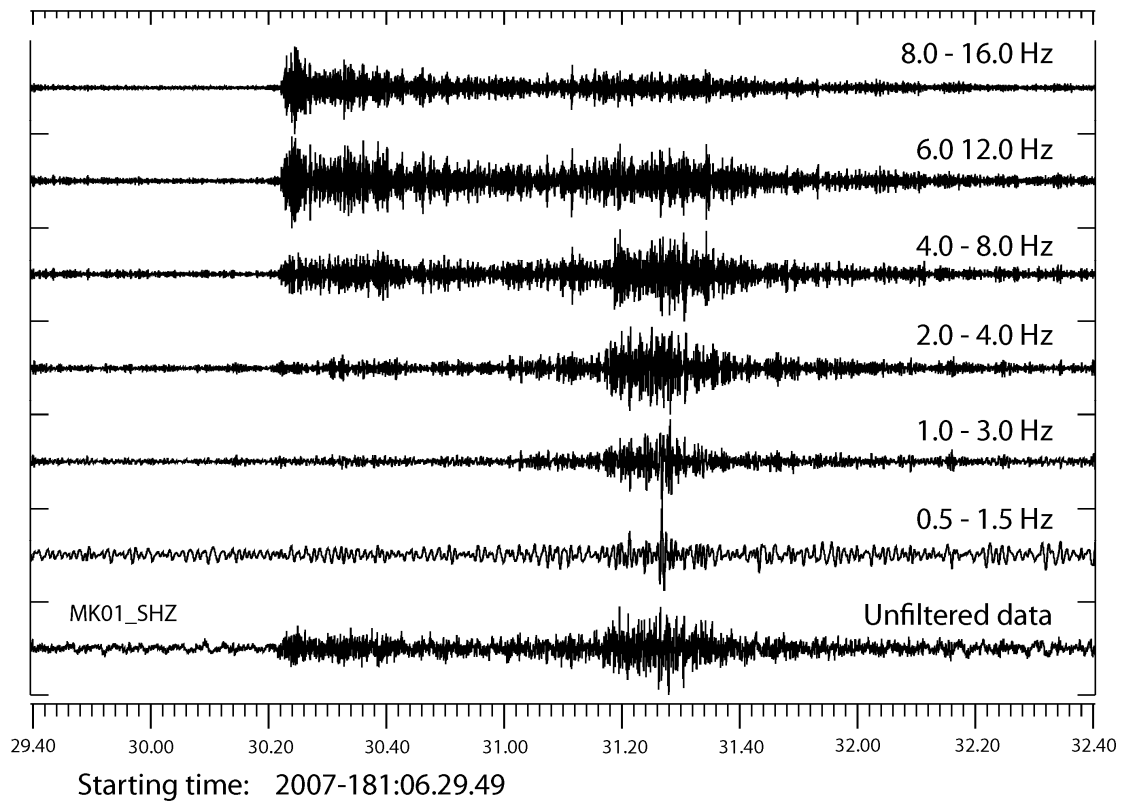


Figure 4.5: A three minute long data segment with signals from an industrial event recorded on the MK01 SHZ channel of the Makanchi array bandpass filtered in various frequency bands as indicated.

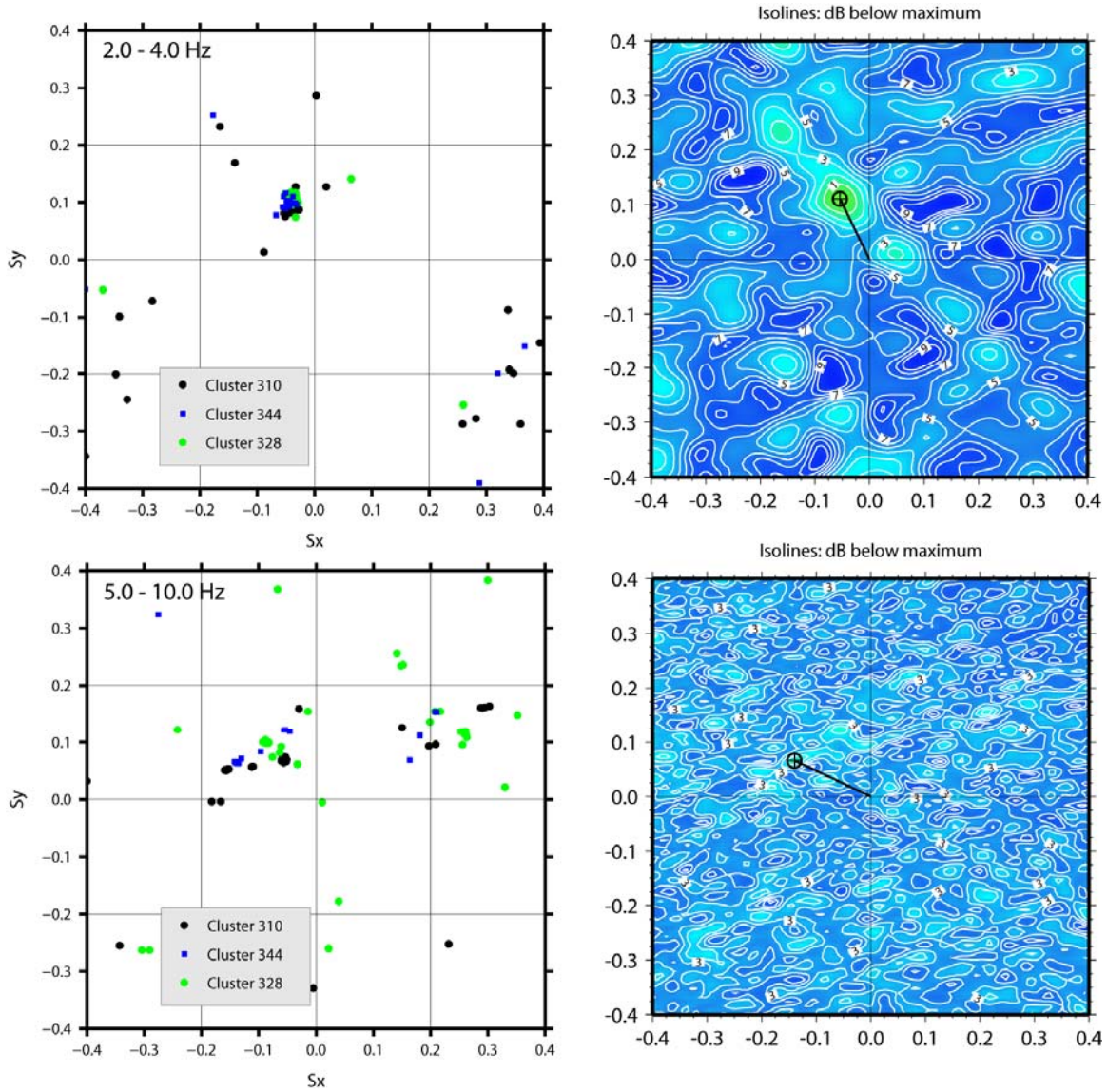


Figure 4.6: The left hand panels show scatter-plots of slowness vector estimates in the frequency bands indicated for initial Pn arrivals at Makanchi for events in clusters 310, 328, and 344 in Hartse et al. (2008). The anticipated backazimuth for all events is close to 340 degrees and the apparent velocity should be close to 8 km/s. The f-k grids to the right display the relative beam-power as a function of the horizontal slowness for the arrival at MKAR at a time 2007-181:06.30.31.425 (see Figure 4.5).

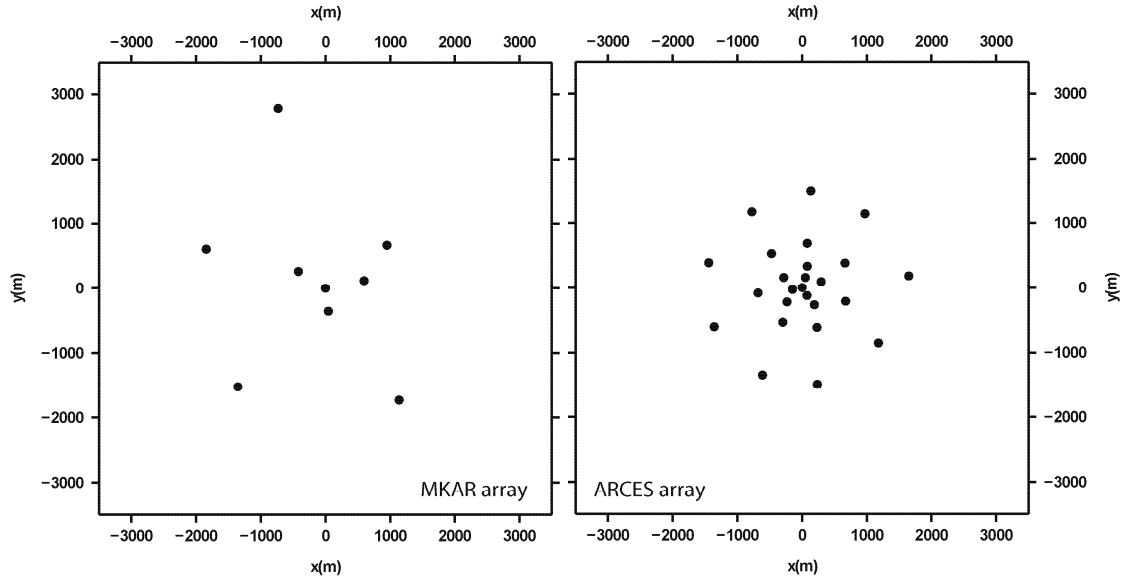


Figure 4.7: Geometries of the Makanchi array (MKAR) in Kazakhstan and, for comparison, the ARCES array in Norway. While MKAR is based on the same design concept as ARCES (concentric rings at increasing radius, each with a greater number of sensors) the lack of small intersite distances means that the high frequency signals (i.e. those relevant to low yield explosions at regional distances) are incoherent between sensors.

A similar procedure to that followed in the previous chapter was attempted to use empirical matched field processing to separate and classify the sources. A frequency band between 5.9375 and 10.0 Hz was selected, primarily to illustrate the performance of EMFP in a frequency band for which conventional frequency-wavenumber analysis demonstrably fails (c.f. Figure 4.6). Figure 4.8 displays the wideband similarity metric (c.f. Equation 3.3) between the empirical steering vectors for Pn arrivals at MKAR for each of the 40 events in cluster 328. As performed previously, a Ward dendrogram was constructed (Figure 4.9) and the resulting reordering (Figure 4.10) shows clear clustering within the 40 events.

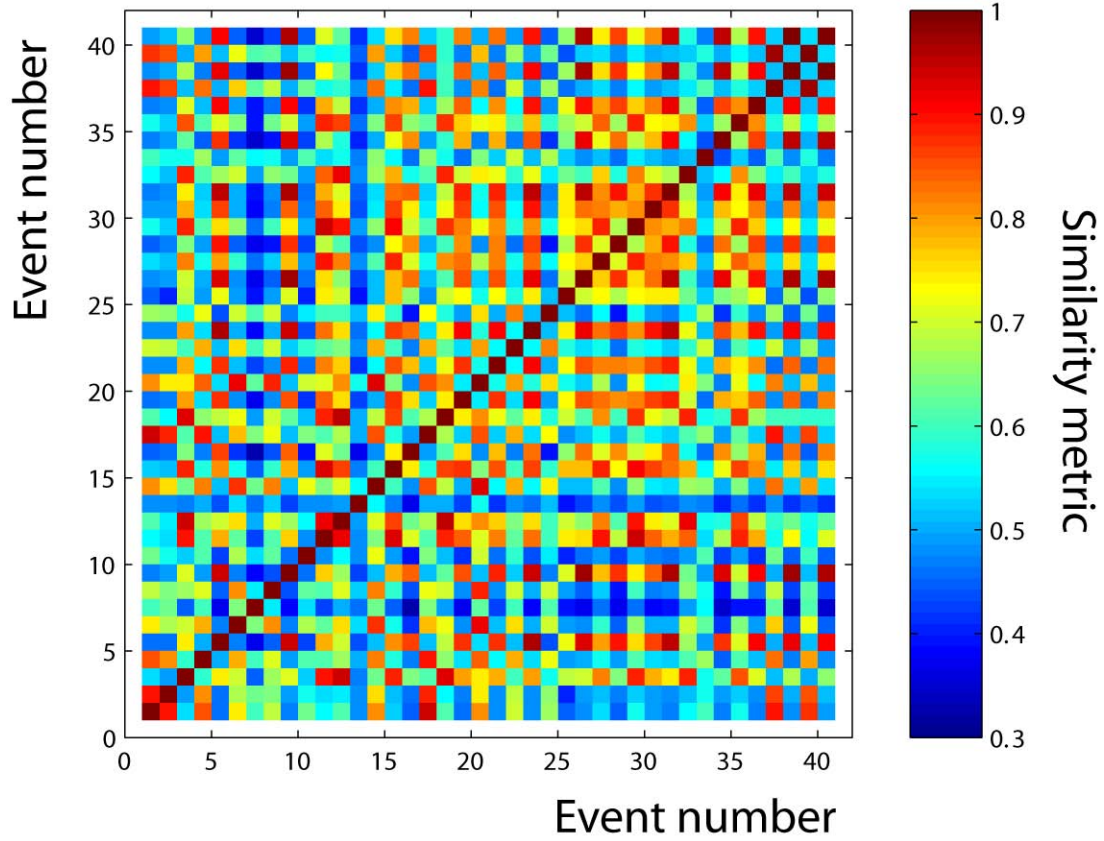


Figure 4.8: Distance matrix between the principal eigenvectors from the covariance matrices for Pn-arrivals at MKAR for each of the 40 events in cluster 328 of Hartse et al. (2008). The covariance matrix is formed using a data window of 9 seconds from all short-period vertical channels of the MKAR array, within the frequency band 5.9375 Hz to 10 Hz.

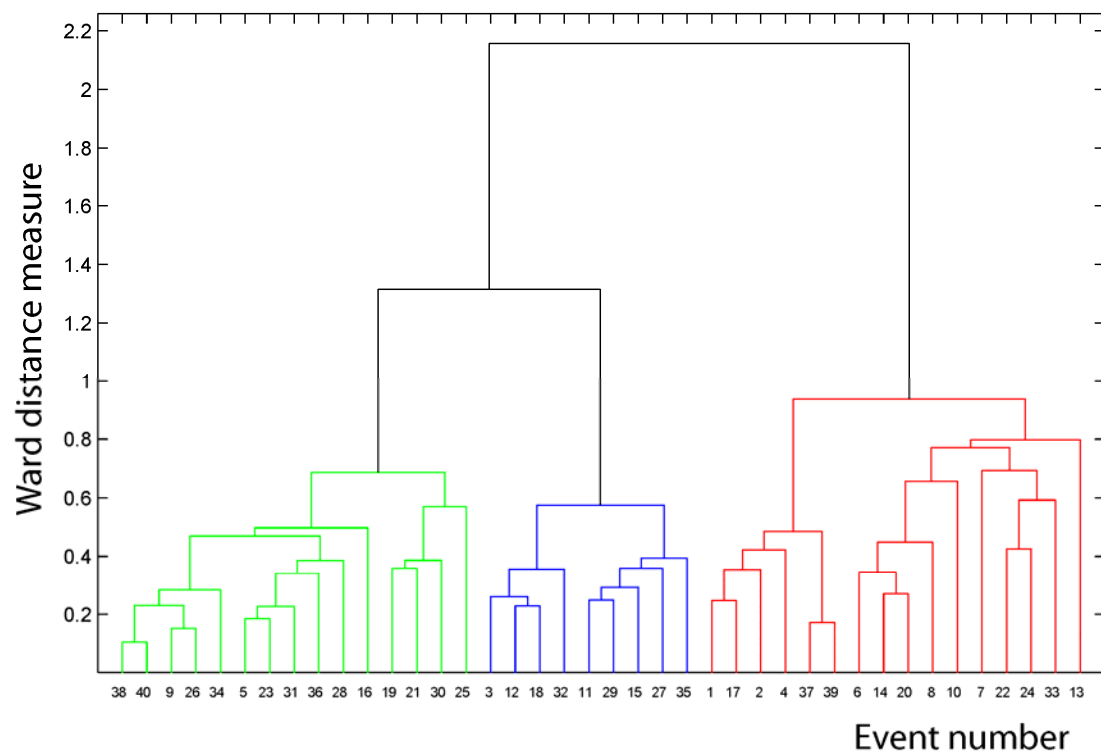


Figure 4.9: Ward dendrogram for MKAR Pn-arrivals from cluster 328 for the values displayed in Figure 4.8.

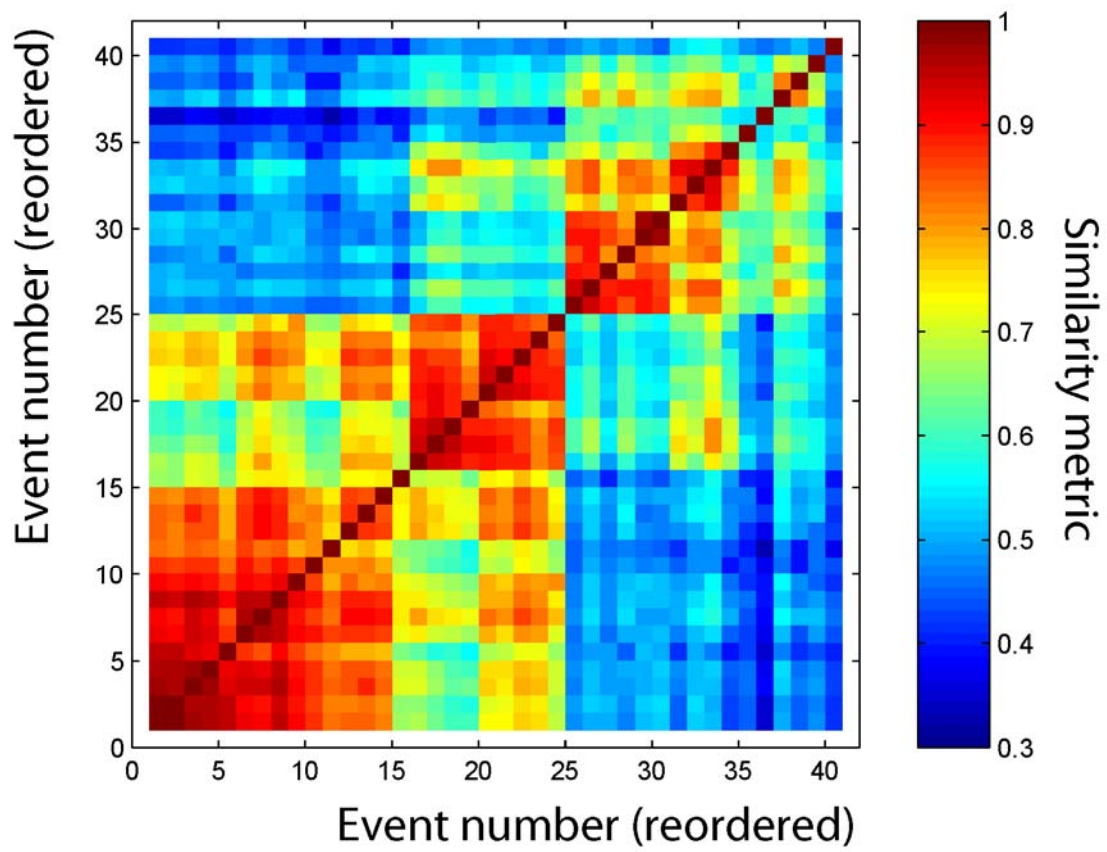


Figure 4.10: Similarity matrix displayed in Figure 4.8 with elements reordered according to the clustering displayed in the dendrogram in Figure 4.9.

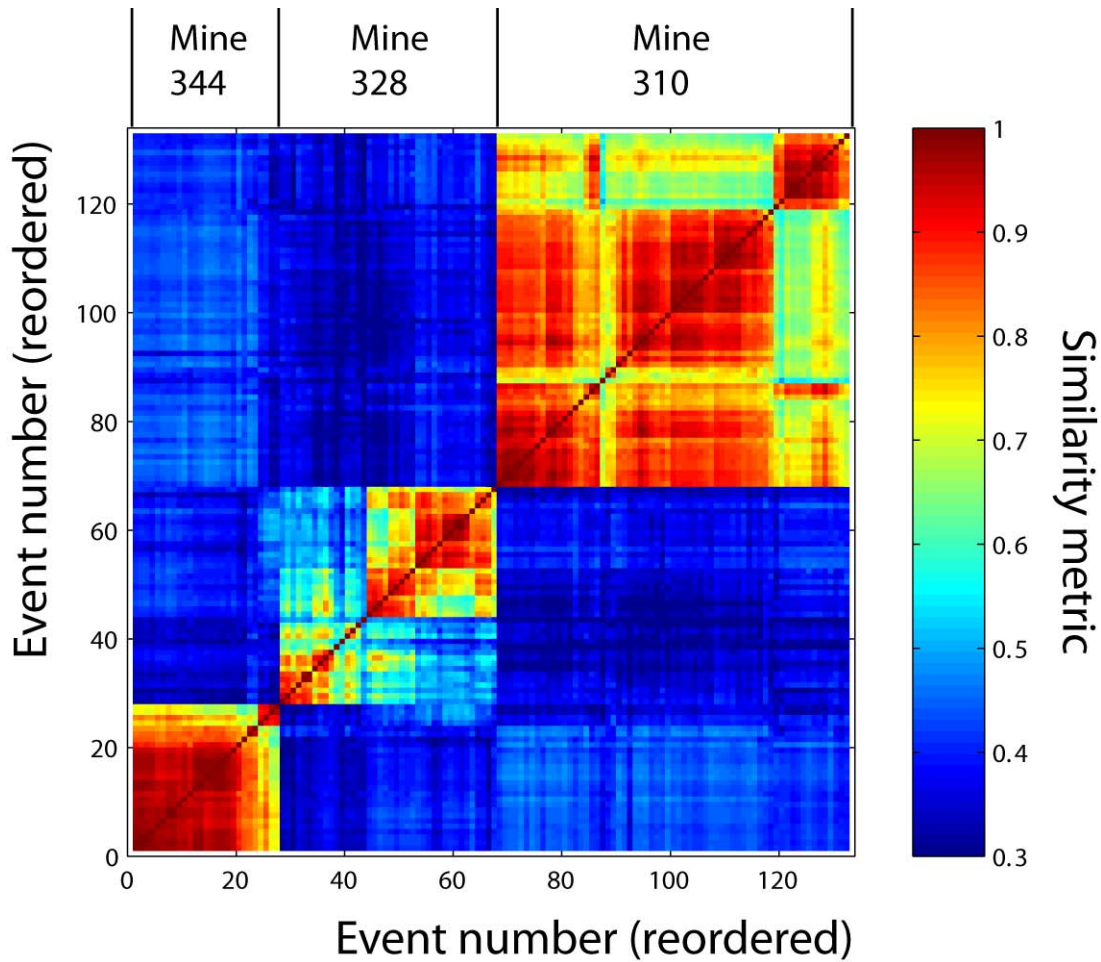


Figure 4.11: Reordered matched field similarity matrix over all three clusters.

The distance metric between empirical steering vectors was calculated for the entire set of MKAR Pn arrivals from clusters 310, 328, and 344, and the subsequent cluster analysis performed. The reordered similarity matrix is displayed in Figure 4.11, from which it is clear that the empirical steering vectors from arrivals from each of the three primary clusters have a far greater projection onto empirical steering vectors from the same cluster than onto empirical steering vectors from the other clusters. As in the previous chapter, an ensemble covariance matrix was calculated from each of the three main mining clusters together with the corresponding empirical steering vectors. For each Pn observation, three matched field statistics were calculated: one each of clusters 310, 328, and 344 (upper row of Figure 4.12). The symbols representing the statistics from each of the three clusters appear to be well separated in this 3-dimensional parameter space. The procedure described at the end of Chapter 2 was followed for the calculation of rank-2 matched field statistics for each of the three clusters. The rank-2 statistics (displayed in the lower row of Figure 4.12) are arguably better separated than the rank-1 statistics with each of the clusters more compact in space.

The improvement in the visual separation between clusters from Figure 4.6 (lower left panel) to Figure 4.12 is remarkable given that they use essentially the same data segments, from the same sensors, and in the same frequency band. The difference is that the separation in Figure 4.6 relies on the match between an observed and a theoretical wavefield, and the separation in Figure 4.12 relies on the match between an observed wavefield and previous observations of well-characterized wavefields.

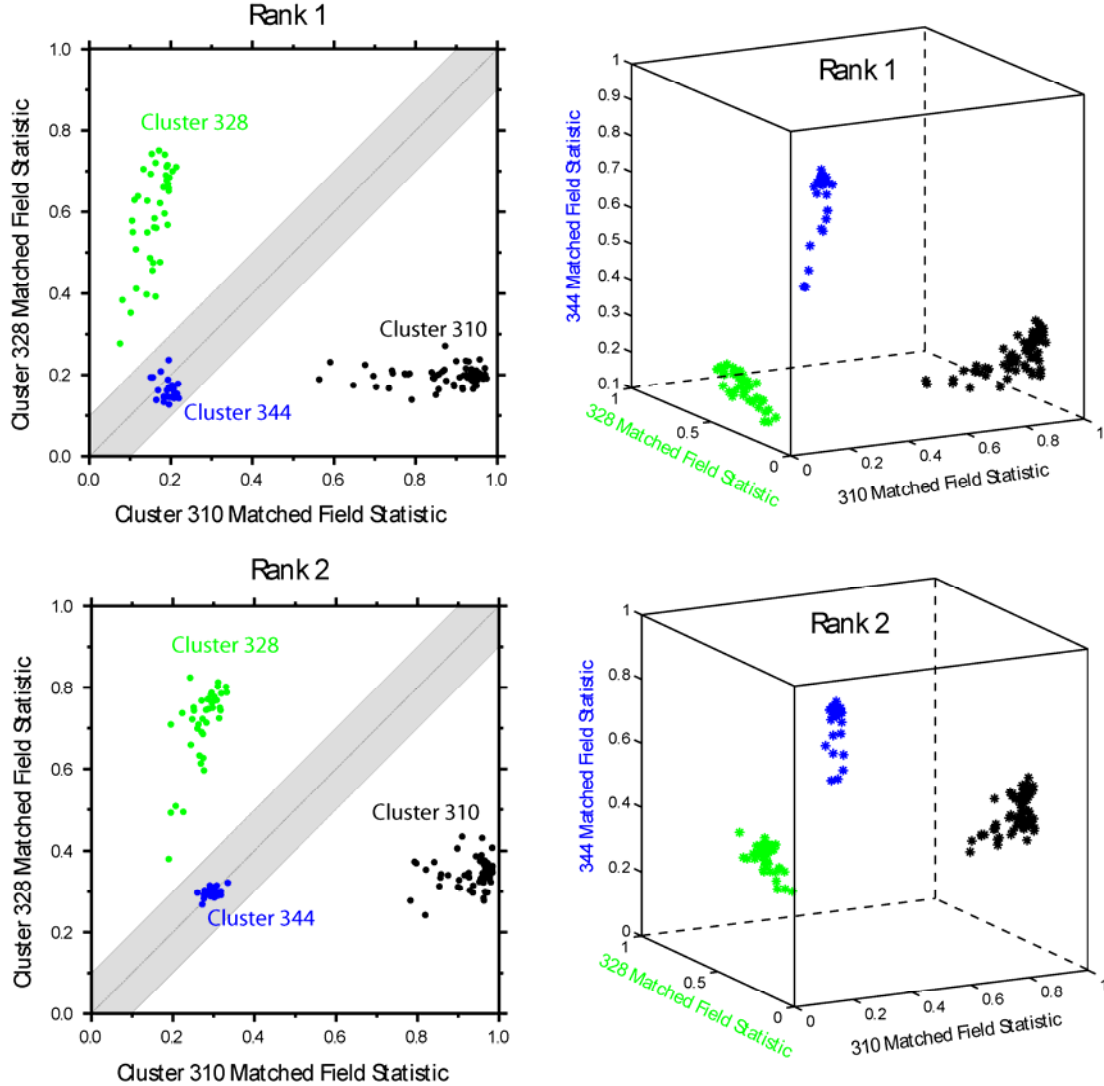


Figure 4.12: Separation using EMPF between events from clusters 310, 328, and 344 using both Rank-1 and Rank-2 matched field statistics.

It is noted that despite the indicated subgroups of events within cluster 328, only single sets of empirical steering vectors were calculated for each of the three major clusters. The similarity metric between the major clusters indicated in Figure 4.11 indicates that the separation between the three primary clusters is so great that a reduction

in performance due to the formation of ensemble covariance matrices from relatively diverse sources is not a problem. It appears that the examination of the rank-2 detection statistic appears to mitigate the spread of the populations in the 3-dimensional parameter space. If a more source-specific classification were necessary, a breakdown of the training events into sub-clusters as was performed in the previous chapter may pay dividends.

Summary

We have demonstrated that the empirical matched field source classification has comfortably distinguished signals from mining blasts in three different source regions in Kazakhstan using only the Pn arrivals at the MKAR array at a distance of approximately 400 km. The calculations were performed in a frequency band from 6 to 10 Hz, in which it has been demonstrated that classical f-k analysis on MKAR is unable even to provide qualitatively reliable estimates of the slowness vector. Performing cluster analysis on the empirical steering vectors from individual events provides a good indication of how reliable source identification is likely to be.

We have examined both rank-1 and rank-2 matched field statistics and it appears that the classification is improved for the rank-2 case. This provides a possible alternative strategy for classification of groups containing inherent dissimilarities, possibly as a result of an expanded source region. In the previous chapter, a successful characterization of a source region with a heterogeneous set of calibration events was only possible by performing cluster analysis on the individual set of calibration events and splitting these events into distinct groups. We have demonstrated that the separation between different source regions is also improved by considering covariance matrices calculated from all calibration events and examining higher order matched field statistics. This may remove the need to split the calibration events a priori since the different wavefield characteristics will be represented in different eigenvectors of the covariance matrix.

5. MATCHED FIELD PROCESSING AS A TOOL FOR DETECTION

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 & 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 presented in the previous sections has demonstrated that EMFP can provide excellent source classification for ripple-fired events, despite significant waveform diversity (see also Harris & Kværna, 2010).

Figure 5.1 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 (c.f. Figure 4.6). 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 (see Appendix A). 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.

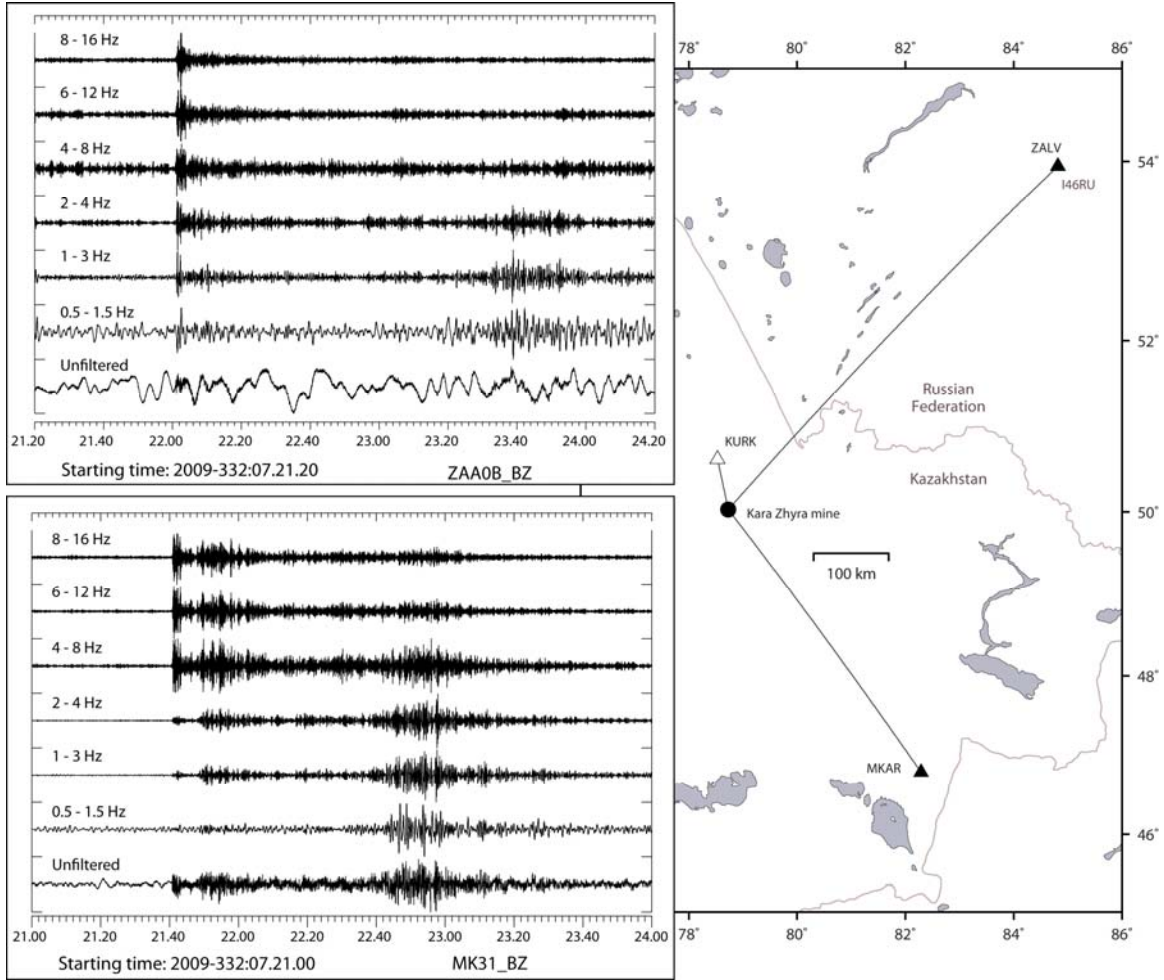


Figure 5.1: Location of the Kara Zhyra mine in Eastern Kazakhstan with respect to the MKAR and ZALV arrays (at 445 km and 600 km respectively). Waveforms from the vertical component broadband sensor at both arrays are shown at the times and in the frequency bands indicated.

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 5.2. The theoretical plane-wave, single observation empirical, and ensemble matched field statistics are as given in equations 2.9, 2.12, and 2.11 respectively. The upper panels of Figure 5.2 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 5.2 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.

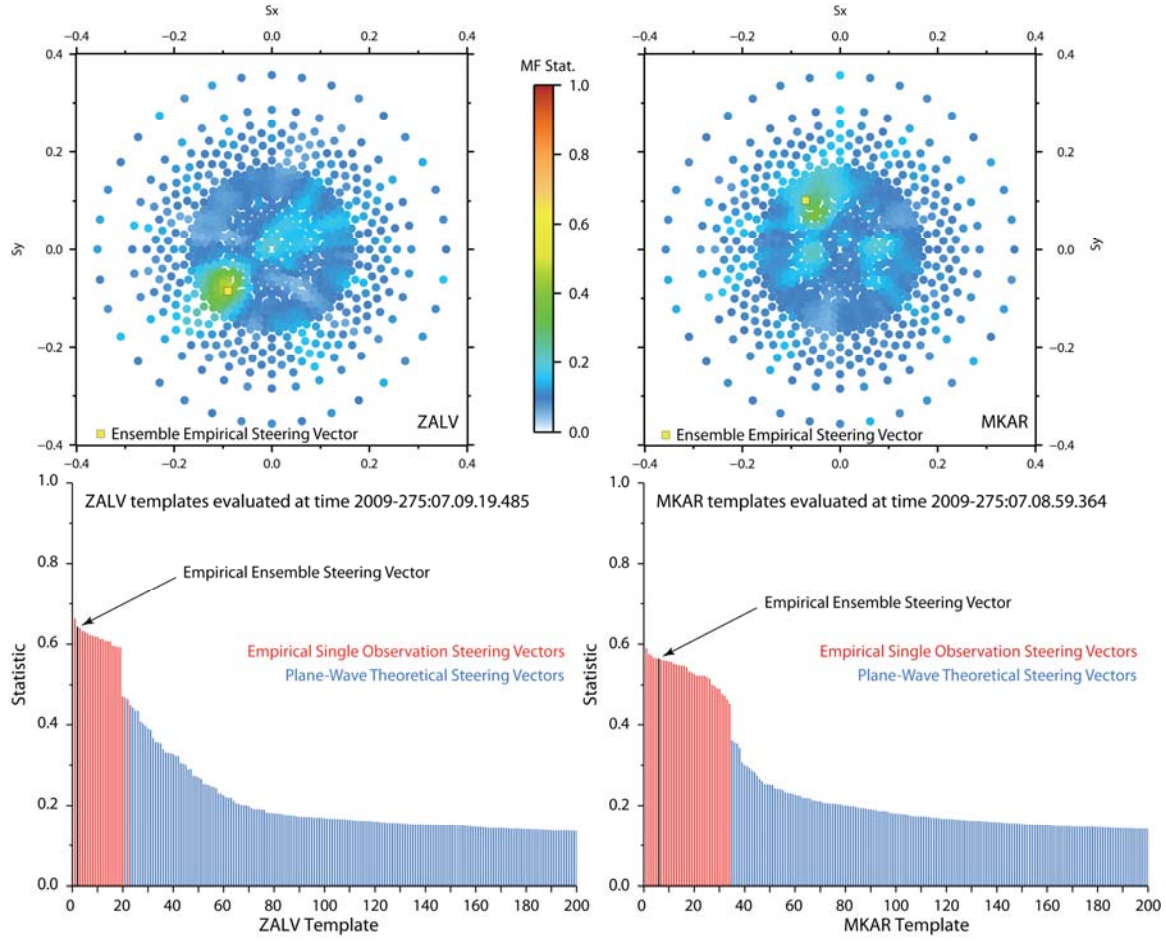


Figure 5.2: Matched field statistics for the specified times at ZALV and MKAR for Pn arrivals from an event assumed to originate close to the Kara Zhyra mine in Eastern Kazakhstan. Both theoretical plane wave steering vectors (blue), single observation empirical matched field steering vectors (red) and ensemble matched steering vectors (black) are used. The bars in the lower panels are rearranged in descending order of the matched field statistic.

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, labeled 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 5.2 is that

essentially every one of the single- observation empirical matched field steering vectors performs significantly better than the best theoretical steering vector. EMFP may 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.

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.

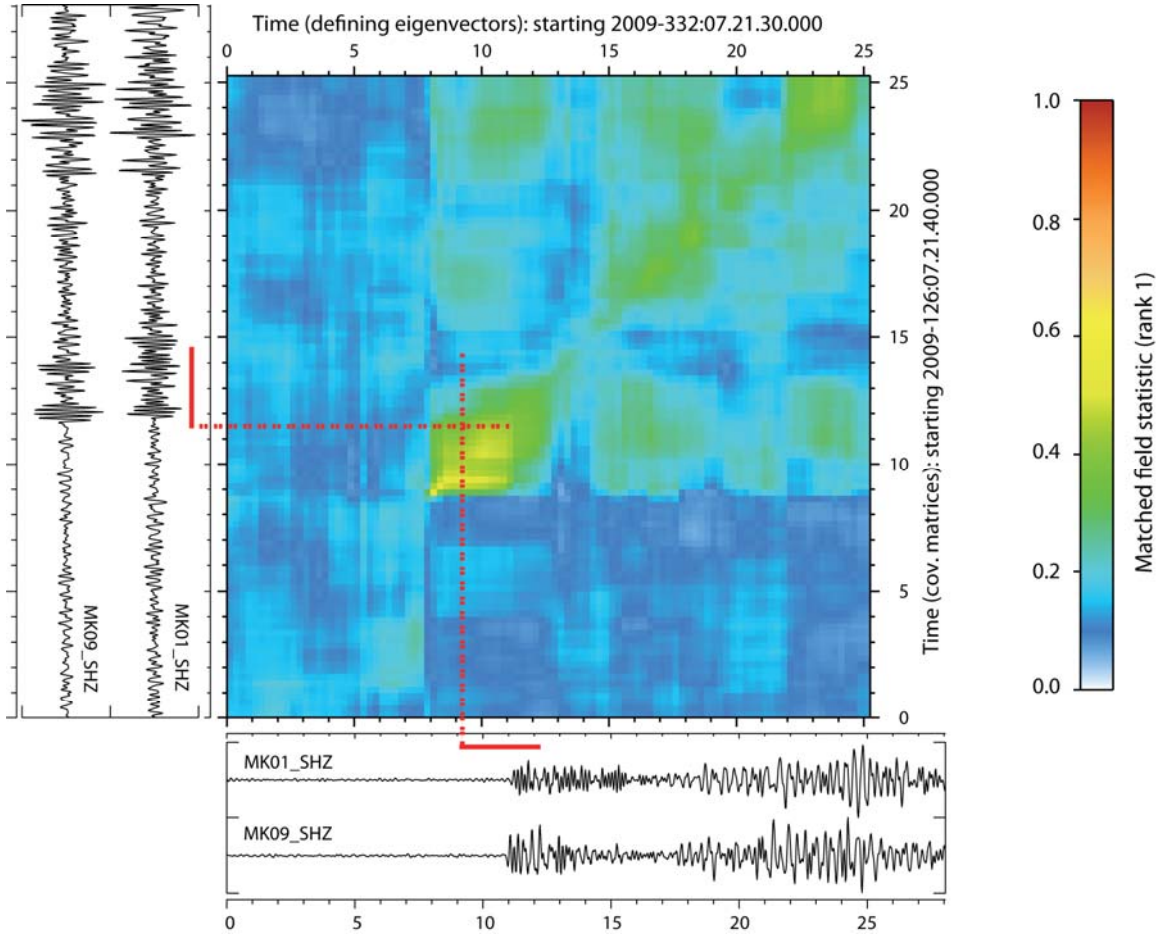


Figure 5.3: Matched field statistics as a function of data-window positioning for two events from the Kara Zhyra mine recorded at Makanchi. The x-axis of the grid indicates the starting time of the 3 second (121 sample) long data-window from which the (single event, single phase) empirical steering vector 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.

Figure 5.3 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 5 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 5.3 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 significant 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.

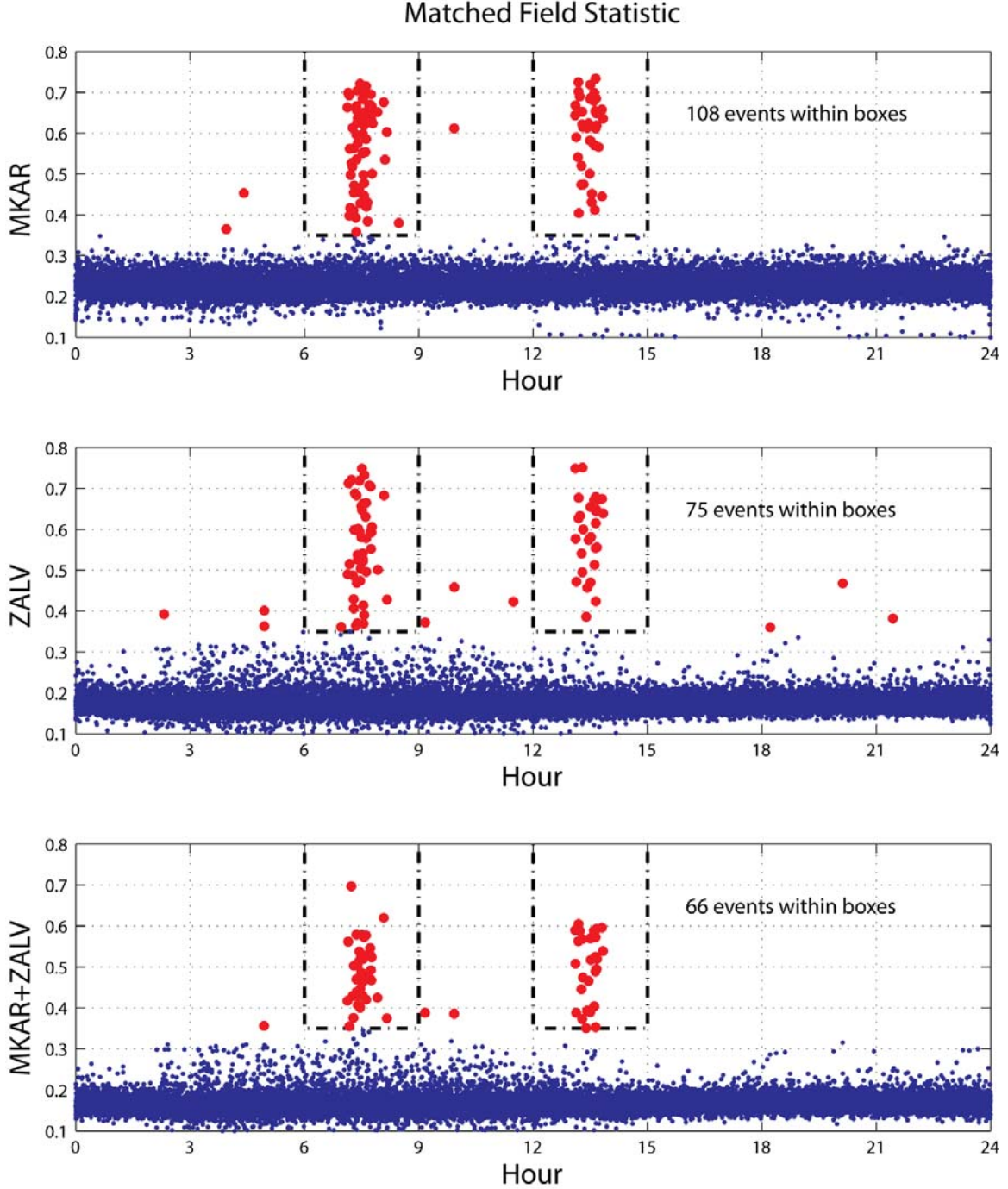


Figure 5.4: Matched field statistics as a function of time of day from empirical steering vectors from a single event at the Kara Zhyra mine for the array configurations indicated from January 1, 2009, to December 31, 2009. Each point indicates the maximum value obtained in a 20 minute segment (and so multiple events within the same 20 minute 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 5.4 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 for 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 covariance matrices for the October 28, 2009, event. 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.

Summary

We have demonstrated that, while it is beneficial to create ensemble covariance matrices for characterizing the spatial structure over a sensor array of signals from a given source region, empirical steering vectors generated from single observation may provide a good matching field representation. In particular, for the example of Pn arrivals at MKAR and ZALV from events at the Kara Zhyra mine in Kazakhstan, every one of the single event empirical steering vectors outperformed the best theoretical steering vector.

We note that this form of matched field processing, where the spectral covariance matrix is evaluated using the multitaper method on a short data segment, is simply an extension of classical f-k analysis and we advocate the inclusion of empirical steering vectors in pipeline detection recipes for all well-constrained seismic sources. This would eliminate the need to apply SASCs (Slowness and Azimuth Station Corrections) and improve the performance of phase association algorithms due to the improved parameter estimates provided.

We have assessed the applicability of single phase matched field processing as a source-specific detector. Firstly, we have examined the sensitivity of the matched field statistic to the precise definition of the data-window. We found that, for a 3 second long data segment, evaluation of the spatial covariance matrix every second is sufficient for calculating a representative matched field statistic trace. We performed a search on data

from the MKAR and ZALV arrays over one calendar year for events at the Kara Zhyra mine using only a single event Pn phase as a template. Due to the time-of-day of the detections made, we assume the false alarm rate to be quite low. A virtual wide-aperture network comprising both MKAR and ZALV, with time-shifts applied to compensate for differences in travel-time, performs reasonably well but misses a number of events that are detected using the two arrays separately. We speculate that the increased sensor separation may make the detector more sensitive to the precise source location. This is to say that increasing the receiver aperture has reduced the size of the source region to which the matched field is applicable.

6. IMPROVING PHASE IDENTIFICATION IN REGIONS OF DIFFUSE SEISMICITY

In the previous sections, it has been demonstrated that EMFP can be used as a means for both classification and detection of industrial seismicity in Central Kazakhstan. This has been demonstrated on medium-aperture seismic arrays in frequency bands at which classical f-k analysis demonstrably fails. An additional aim of the current project is to extend the notion of matched field processing to bands of diffuse seismicity. Figure 6.1 displays density of events in the KNDC seismic bulletin which are recorded with a good signal-to-noise ratio (SNR) on both of the KKAR and MKAR arrays. Events within the red ring displayed were selected for further investigation if the SNR for the Pn phase exceeded 10 at both MKAR and KKAR, and if data existed for all nine short-period vertical channels of both arrays. 267 events fulfilled these criteria.

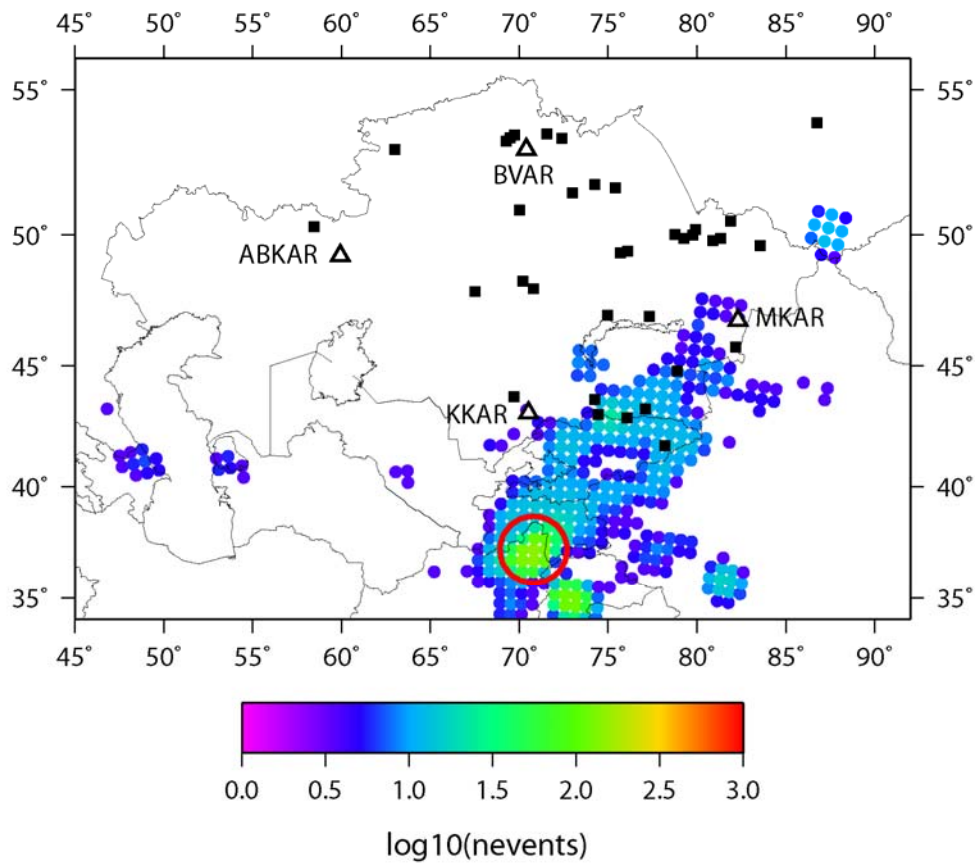


Figure 6.1: Events from the KNDC reviewed event bulletin for which a regional P-phase with an SNR equal to or greater than 10 was observed at both the KKAR and MKAR arrays. The colour indicates $\log_{10}$ of the number of events within each bin of radius ~ 83 km for the time period 2005 to 2008. The red circle indicates a region selected for the study of diffuse seismicity within Afghanistan and Tadjikistan. The radius of the search window is 167 km, centered on 37.2 N, 70.8 E.

A spectral covariance matrix was calculated for the Pn arrival at MKAR for each of the 267 events and a the principal eigenvector was extracted. A pairwise similarity matrix and corresponding Ward dendrogram was calculated and the reordered similarity matrix is displayed in Figure 6.2. It is clear that the set of 267 events contains a large number whose signals do not match the signals from any other events, although a small number of clusters are observed.

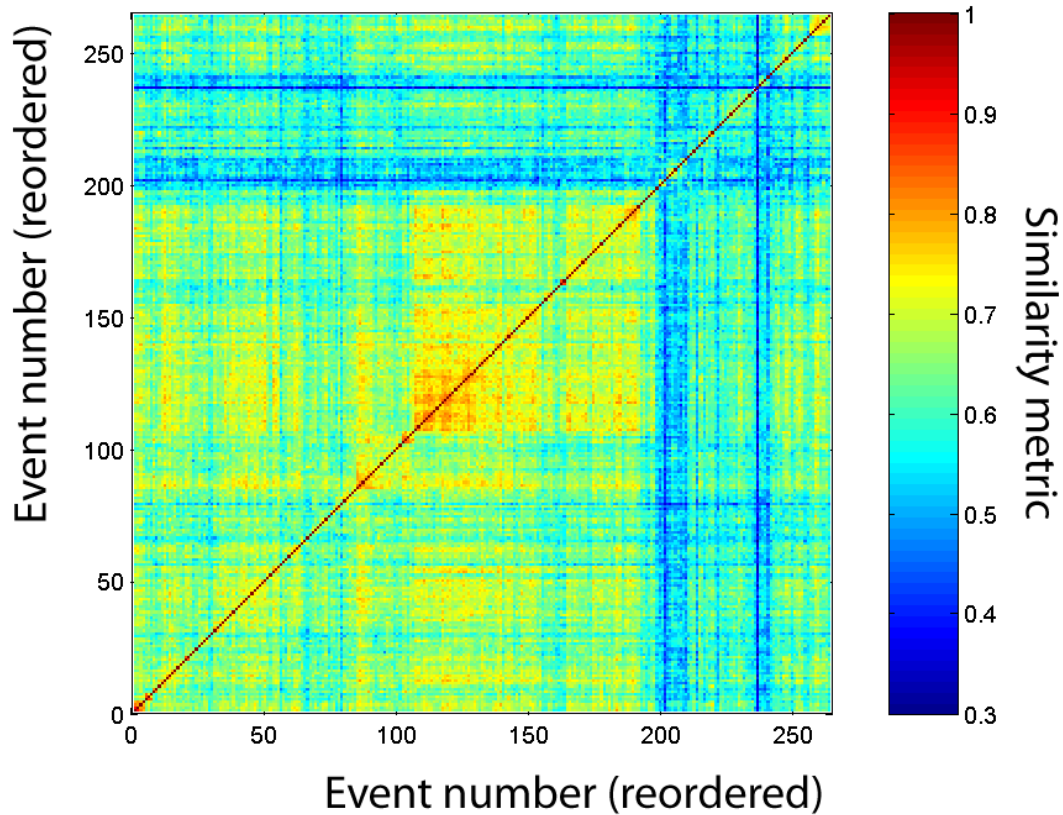


Figure 6.2: Similarity matrix for the Pn observations at MKAR for the region of diffuse seismicity labelled in Figure 6.1.

A significant number of the larger events were also found in the Reviewed Event Bulletin (REB) of the International Data Center (IDC) of the Comprehensive nuclear Test-Ban-Treaty Organization (CTBTO); these are displayed in the right hand panel of Figure 6.3. A comparison of the left and right panels in Figure 6.3 highlights an immediate problem; there is a disagreement in event locations of the order 100 km. Both sets of solutions have the disadvantage that the observations are predominantly to the North, meaning an almost direct tradeoff between event depth and latitude. The KNDC solutions use exclusively stations in Kazakhstan and Kirgizstan (usually only AAK is used) and the REB solutions have the disadvantage of not being able to exploit the signals on KKAR, the most sensitive of these stations for this target region.

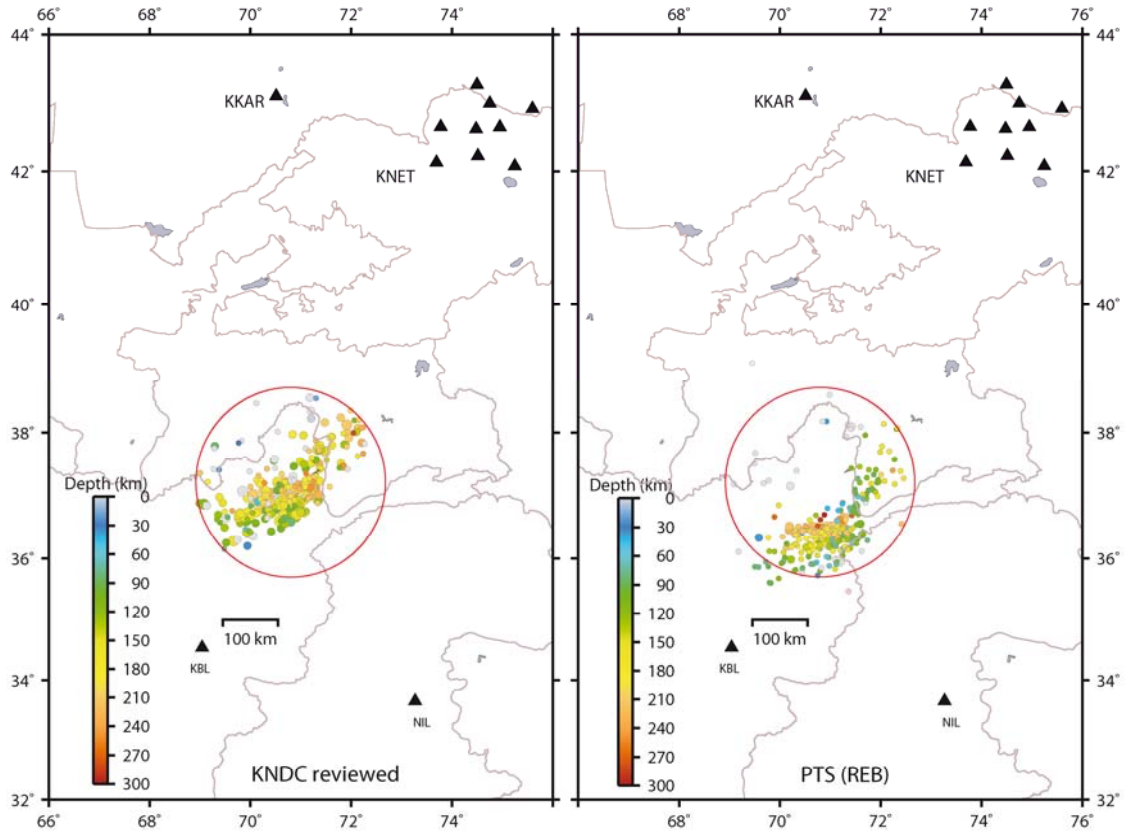


Figure 6.3: Event location estimates provided in the reviewed event bulletin of the KNDC (left) and in the REB of the International Data Center (right) for the 267 events whose similarity metrics are displayed in Figure 6.2.

It was concluded that there was significant uncertainty attached to the location estimates for these events (the uncertainty in the depth estimates can result in comparable lateral variability, particularly for small events with particularly unfavorable station distributions). It was decided that for any matched field study of this seismicity of KNET, a preliminary set of arrival times was required for each of the phases to be observed. None of the available bulletins contain readings for all of the KNET stations and it was decided to re-examine arrival times for all of the selected events for all available stations at regional distances, in addition to a number of IMS array stations at teleseismic distances. This would serve the dual purposes of creating an arrival time database for the generation of matched field spectral covariance matrices, and also the generation of new location estimates. The new location estimates are displayed in the right hand panel of Figure 6.4, together with the locations from the ISC catalog to the left.

From considerations of station coverage, we considered only events which were recorded to the south by either NIL (Nilore, Pakistan) or KBL (Kabul, Afghanistan; data available from IRIS since January 31, 2007). Figure 6.5 displays KBL waveforms for an event on February 4, 2008, with P and S arrivals clearly visible on the vertical and

transverse components respectively. Figure 6.6 gives an impression of how variable the waveforms from these events over KNET.

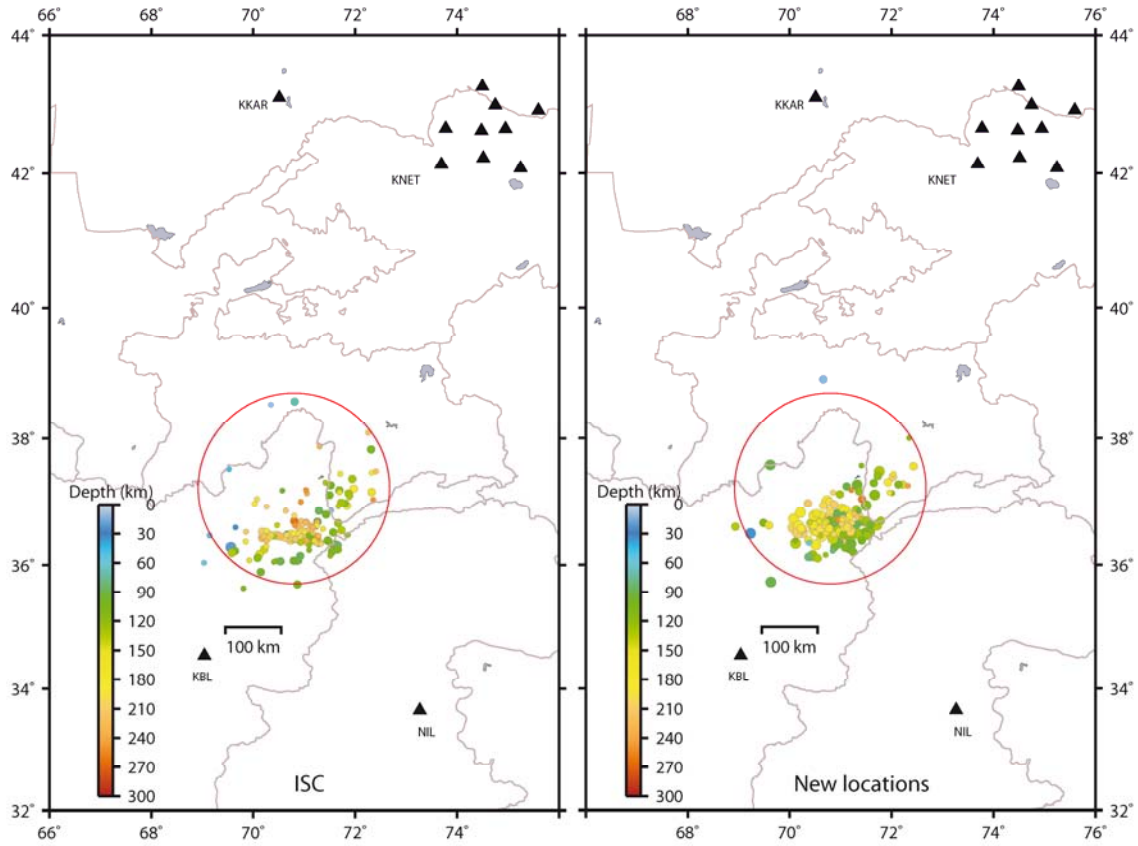


Figure 6.4: Event location estimates provided in the ISC bulletin (left) together with new solutions (right) for the 267 events whose similarity metrics are displayed in Figure 6.2. Note that only a subset of 155 of these events were available in the ISC bulletin at the time this work was performed.

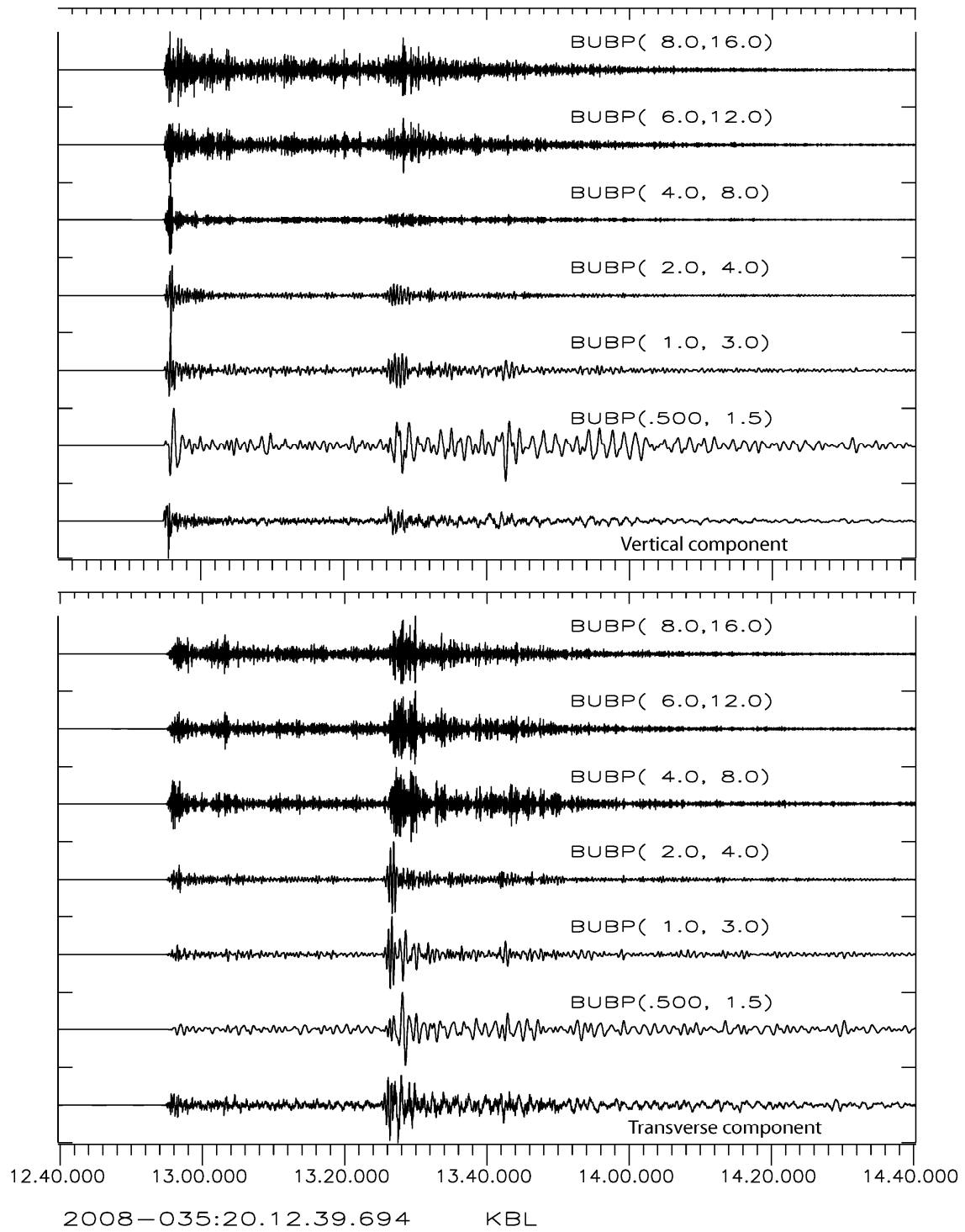


Figure 6.5: Waveforms at station KBL in Afghanistan for event 2008-035:20.12.12 (lat 36:504 N, 70:157 E, depth ~ 200km).

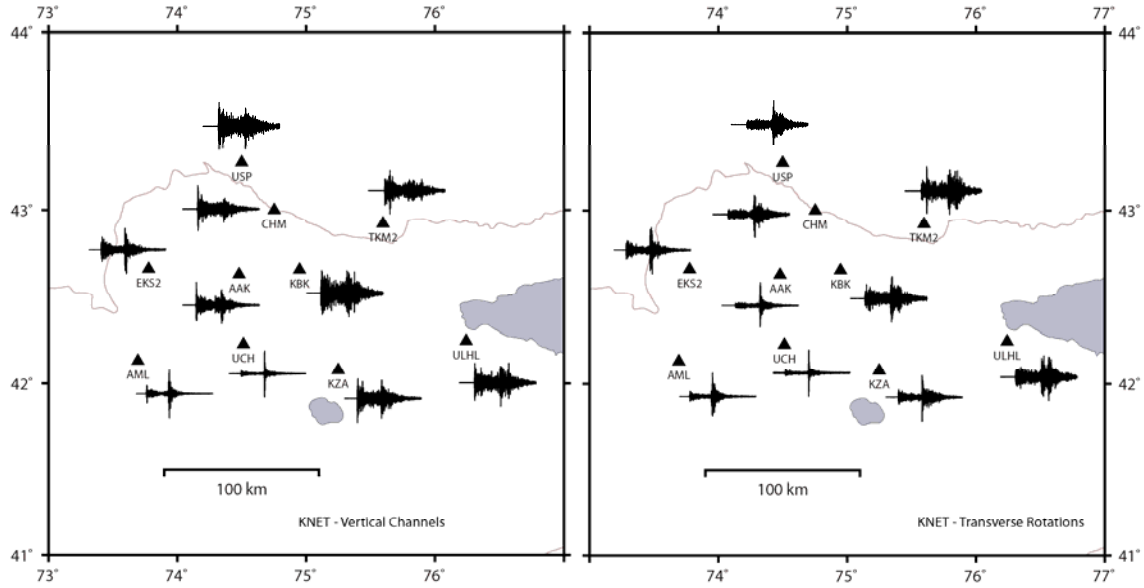


Figure 6.6: Variability of the waveforms over KNET from the event 2008-035:20.12.12 (lat 36:504 N, 70:157 E, depth $\sim$ 200km). All waveforms are bandpass filtered 2-8 Hz.

A number of events were identified whose principal eigenvectors resulted in good matches (Figure 6.2). Figure 6.7 displays the structure of the spectral covariance matrices at 2 Hz for the KKAR Pn arrival from two such events, together with the anticipated plane wavefront pattern. Each symbol indicates the phase difference (color) and the coherence (size) between the signals at two different sites on the array, and is plotted at coordinates indicating the vector distance between the two stations. The plane wavefront model appears to be reasonable although, already at 2 Hz, the coherence diminishes rapidly with intersite distance.

At 4 Hz (Figure 6.8), the differences between the measured and theoretical covariance matrices are far greater. However, the measured covariance matrices from the two different events are remarkably similar both for 2 Hz and 4 Hz.

In Figure 6.9, we evaluate a spectral covariance matrix for a third Pn arrival at KKAR and compare the matched field statistics (2-8 Hz band) obtained for a deployment of theoretical steering vectors and for the empirical steering vectors from the November 18, 2005, and February 4, 2008, events. As for the industrial Kara Zhyra events in the previous section, the empirical steering vectors provide a stronger match with the new event than any of the theoretical slowness vectors (Figure 6.9). We have so far only considered the wide band matched field statistics, where we have summed incoherently across all of the narrow frequency bands. Figure 6.10 displays the matched field statistics for the theoretical and empirical steering vectors as a function of horizontal slowness for a number of frequency bands as indicated. We note first how the backazimuth and velocity inferred from f-k analysis differs from 2 to 3 Hz. At 2 Hz, the maxima in slowness space coincides almost with the empirical slowness vector, whereas at 3 Hz, a

slightly different backazimuth and a far higher apparent velocity are inferred. In both cases, the match with the best empirical steering vector is higher than the maximum of the theoretical slowness vectors. As the frequency increases, the improvement of the empirical steering vectors over the theoretical steering vectors increases (see Figure 6.11). For some frequencies, the performance of one of the empirical steering vectors is poorer than a number of theoretical steering vectors. The reason for this is not clear; it may be an SNR issue or it may simply indicate a different source location or mechanism. If it is for the latter reason, it is somewhat surprising that the performance of this particular vector improves again at even higher frequencies.

Similar results are observed for many of the clusters apparent in Figure 6.2; that matched field statistics evaluated using empirical steering vectors from other events provide a better wavefield characterization than any plane-wave steering vector. However, for events in Figure 6.2 which did not show significant similarity with other events, there was unsurprisingly no improvement over the theoretical steering vectors.

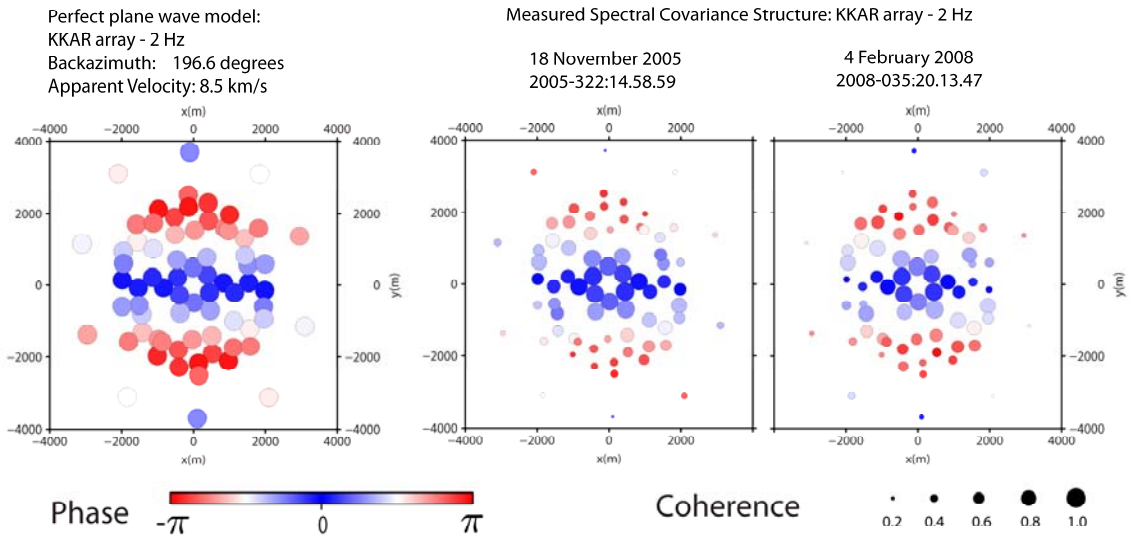


Figure 6.7: KKAR Pn arrival phase and coherence co-array at 2 Hz for two similar events in a region of diffuse seismicity together with predicted pattern.

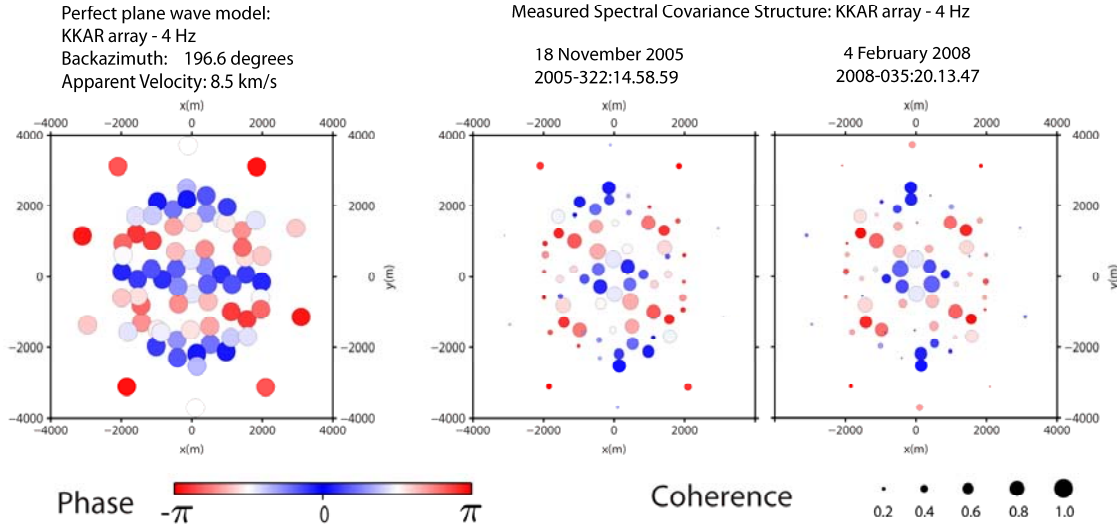


Figure 6.8: KKAR Pn arrival phase and coherence co-array at 4 Hz for two similar events in a region of diffuse seismicity together with predicted pattern.

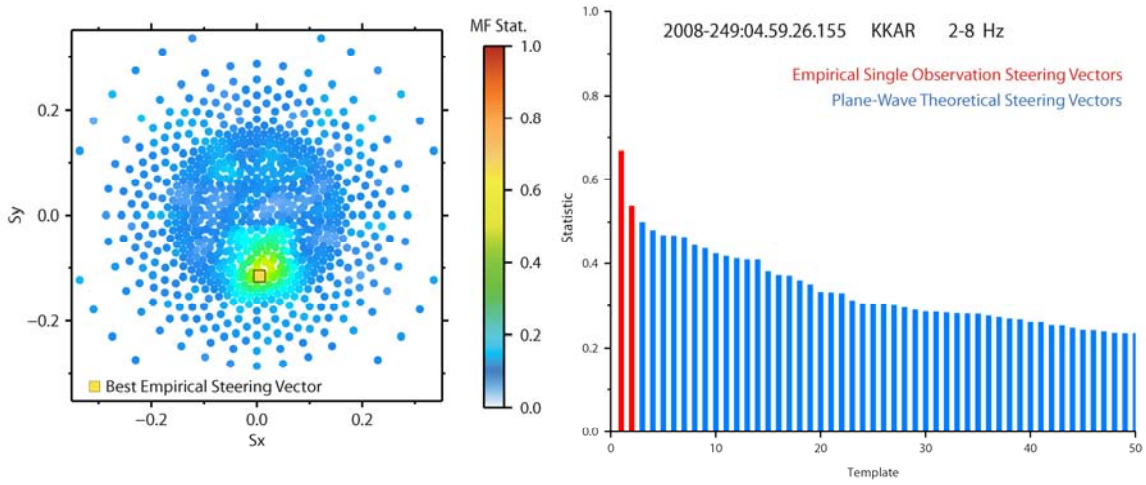


Figure 6.9: (left) Matched field statistic for theoretical (circles) and empirical (square) steering vectors on the KKAR array in the 2-8 Hz frequency band, plotted as a function of slowness vector (the slowness vector for the empirical matched field statistics is the theoretical one). The data covariance matrix is evaluated at a time 2008-249:04.59.26.155 and two empirical steering vectors are evaluated at times 2005-135:00.31.03.025 and 2006-014:09.00.25.475. The colour of the square indicates the maximum of the matched field statistics from the two empirical steering vectors. (right) 50 greatest values of the matched field statistic arranged in descending order with empirical and theoretical values colored red and blue respectively.

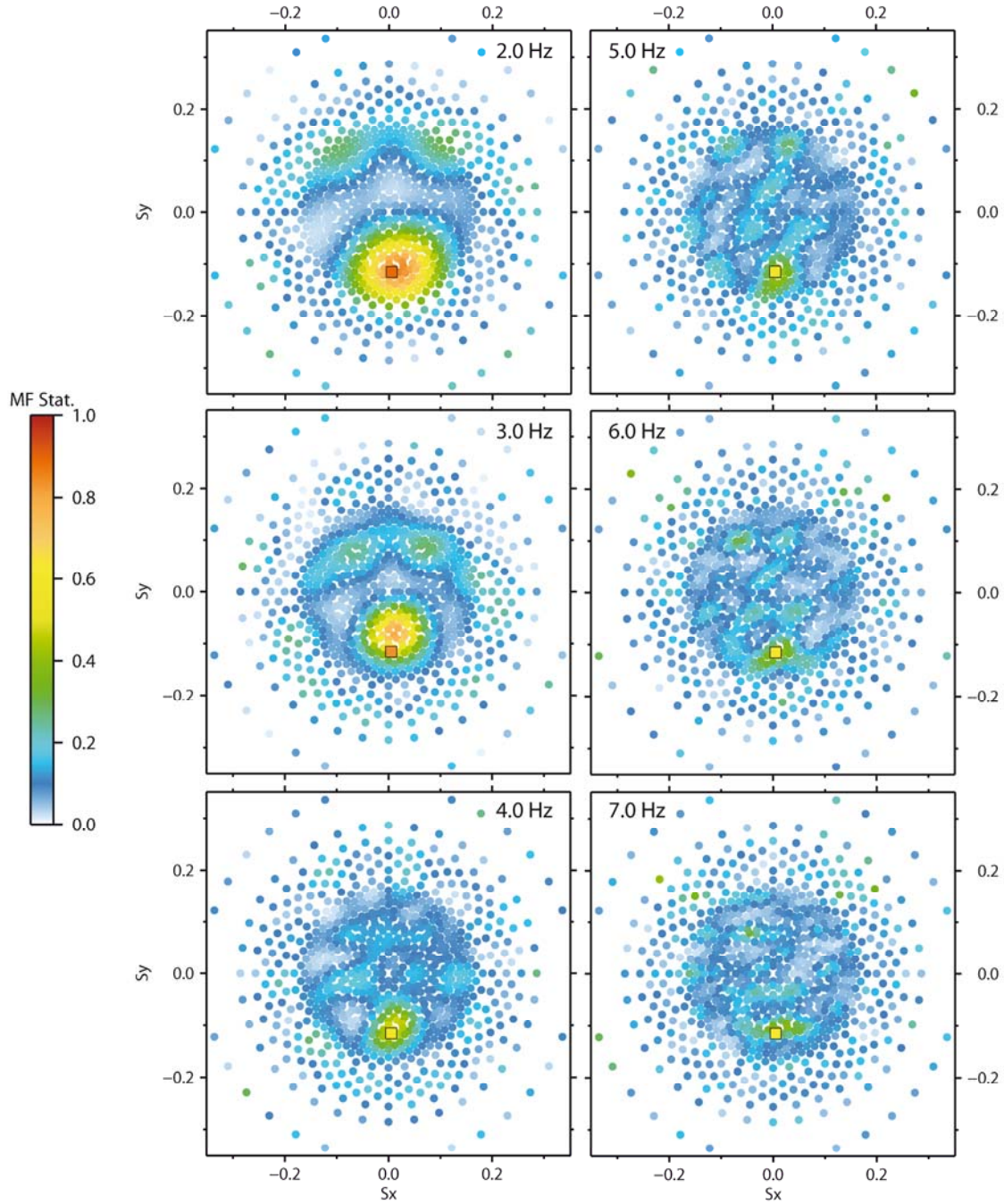


Figure 6.10: Matched field statistic for theoretical (circles) and empirical (square) steering vectors on the KKAR array in narrow frequency bands as indicated. The data covariance matrix is evaluated at a time 2008-249:04.59.26.155 and two empirical steering vectors are evaluated at times 2005-135:00.31.03.025 and 2006-014:09.00.25.475. The color of the square indicates the maximum of the matched field statistics from the two empirical steering vectors.

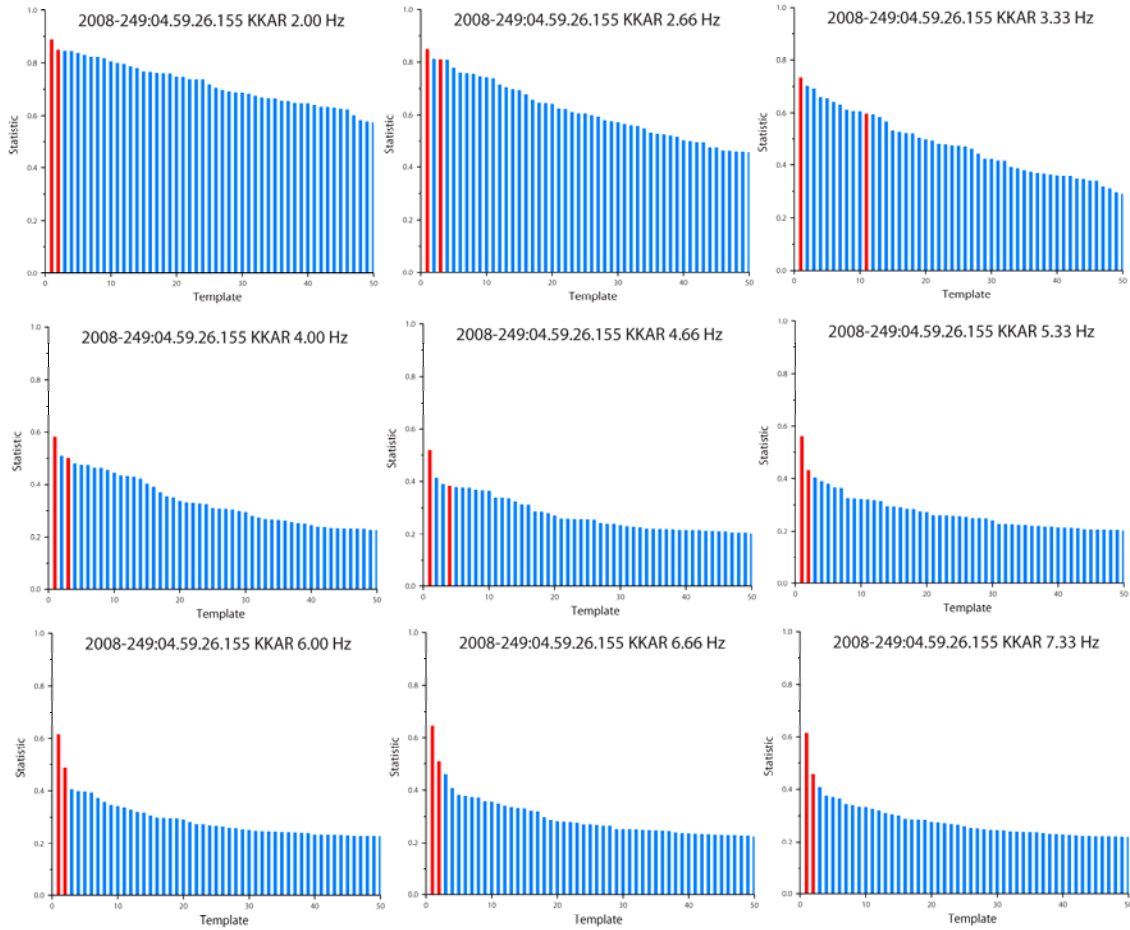


Figure 6.11: Matched field statistics from theoretical (blue) and empirical (red) steering vectors in narrow frequency bands as indicated arranged in descending order. All parameters are as in Figure 6.10.

Summary

We selected a region to the south of Kazakhstan containing significant natural (or diffuse) seismicity. The events were initially selected from the reviewed event bulletin of the Kazakhstan National Data Center, although subsequent investigation showed these solutions to be significantly different to those in the IDC Reviewed Event Bulletin and in the ISC catalog. A number of events were selected and relocated using manual picks from all openly available data. Spectral covariance matrices were calculated for KKAR Pn arrivals for the selected events and a number of clusters emerged. For clustering events, we have examined the form of the covariance matrices and found great similarity between some events and demonstrated that empirical steering vectors can provide better characterization than theoretical steering vectors, in particular at higher frequencies (where the best SNR is likely to be for lower magnitude events).

Due to the very poor constraints on the locations of these events, questions regarding the spatial footprint of the empirical matched field calibrations cannot be answered. Addressing these questions requires high-quality observations at far closer stations than those available to us for this case study.

7. EXAMINING AN AFTERSHOCK SEQUENCE USING EMPIRICAL MATCHED FIELD PROCESSING

A form of diffuse seismicity which it is necessary to characterize reliably, completely, and rapidly is seismicity in the immediate aftermath of a large earthquake. Many thousands of events can occur shortly after, and close to the source of, a large earthquake and can overwhelm data processing and analyst resources for a long time. Figure 7.1 shows the location of the M=7.4 October 8, 2005, Kashmir event relative to the 3-component station NIL, the medium aperture array KKAR, and the KNET network. Figure 7.2 displays the first 20 hours of the Kashmir event sequence recorded on the instrument KK01 of the KKAR array.

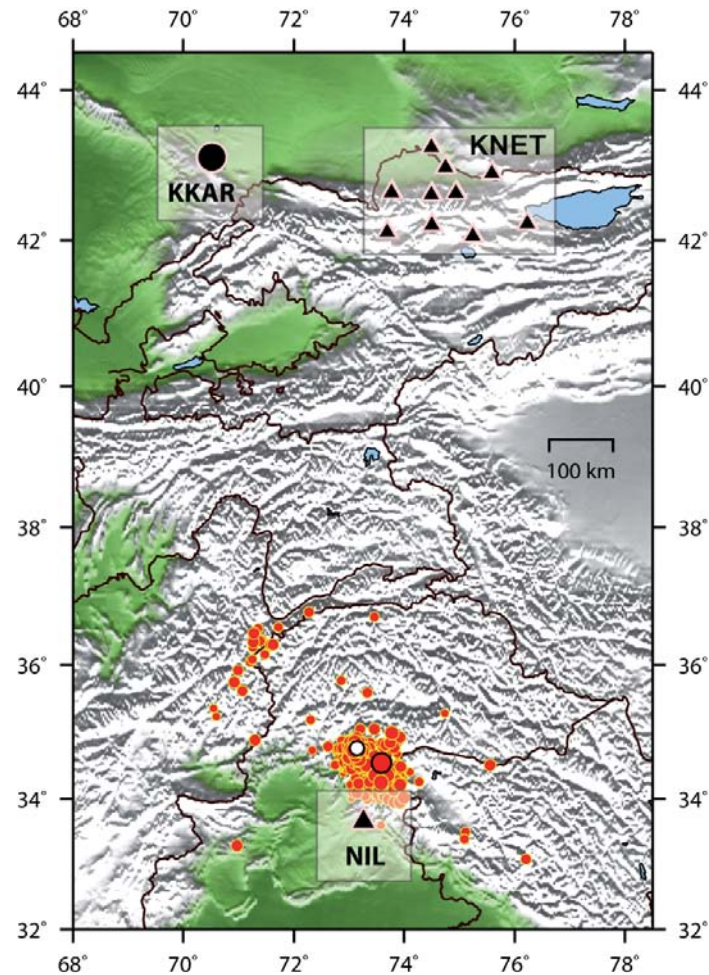


Figure 7.1: Location of the M=7.4 October 8, 2005, Kashmir event mainshock (large red circle) with respect to KKAR, KNET, and the NIL 3-C station. Additional seismic event locations of events within the ISC bulletin that fall within 300 km of the epicenter of the main shock between October 8, 2005, and December 31, 2005. The white symbol gives the ISC location of an M 5 aftershock with an origin time of approximately 2005-281:05.34.50.

The vast number of arrivals at all stations in a network challenge traditional association algorithms, with many incorrect associations filling the automatic bulletins and many true events being missed. An ideal solution would be to extract waveforms from the master event and perform a correlation operation and simply pick out aftershocks from the peaks in the matched filter detection statistic. Unfortunately, this procedure performs very poorly for very large events given the far greater dimensions of the source region, very different spectral composition of mainshock and aftershocks, and differing source mechanisms. Setting up correlation detectors using the signals of aftershocks which are well-separated from other events can frequently identify a number of events within a very limited footprint of the selected master event. However, due to the extended source regions that apply, only a very small proportion of the seismicity is likely to be covered using a single template. Multiple templates are required (correlation and subspace) and autonomously administering their deployment and interpretation in real-time is a great challenge.

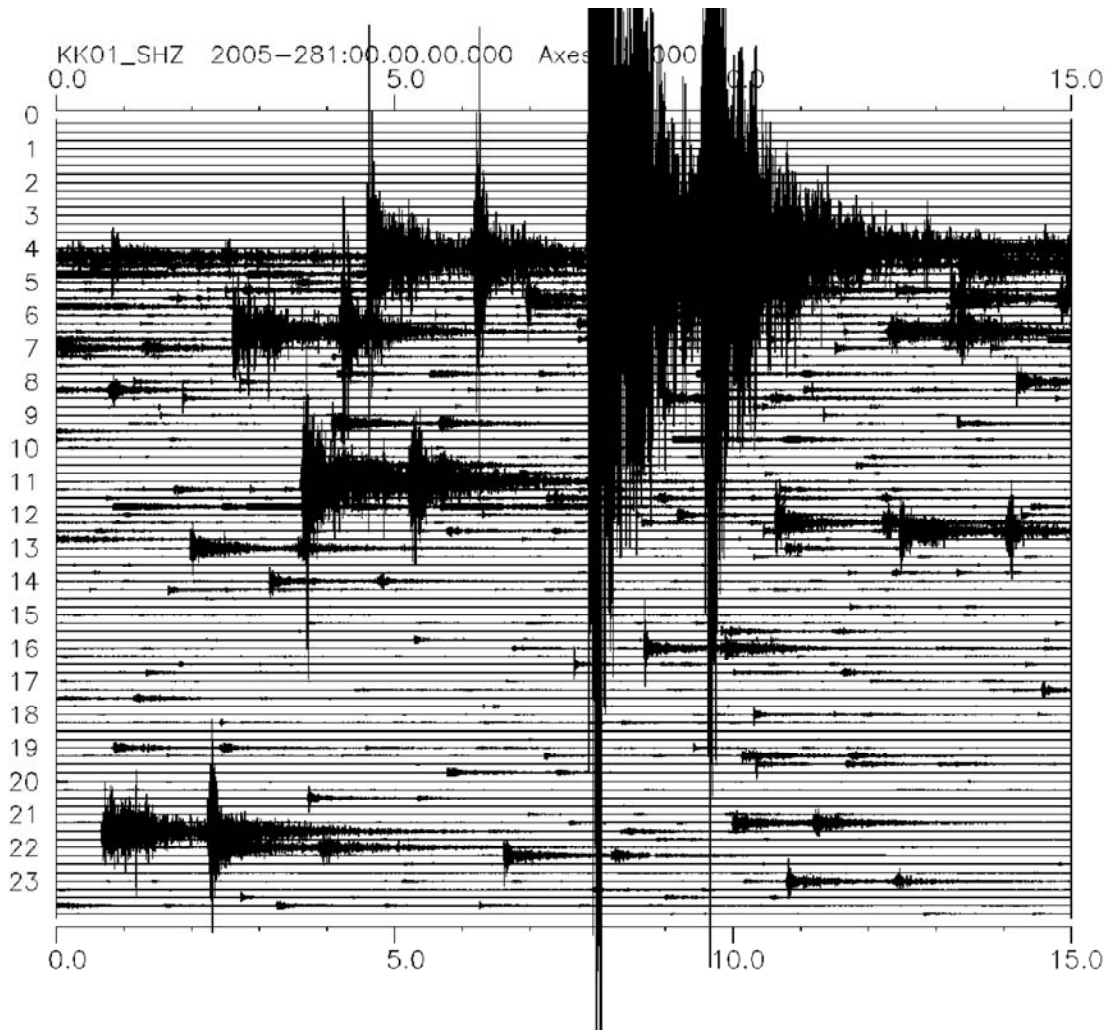


Figure 7.2: Simulated helicorder plot of the KK01 short period vertical sensor on October 8, 2005.

Figure 7.3 displays a 30 minute segment of KKAR data (trace 1 from top) surrounding one aftershock (indicated by an arrow) together with a number of waveform attributes which could be used to identify events in the sequence. Trace 2 from the top shows the relative beam power from a continuous f-k analysis in the 2-4 Hz frequency band. A high beam power will occur whenever a coherent wavefront passes over the array and so we need additional criteria to determine whether or not the detection corresponds to a specified arrival from the source region of interest. In practice, this will mean defining a region of slowness space within which the maximum beam power is to be found. As displayed in Figure 6.10, the optimal slowness vector can vary in slowness space as a function of frequency and become poorly defined due to incoherence at higher frequencies. Peaks corresponding to anticipated Pn arrivals from the aftershock region are colored red.

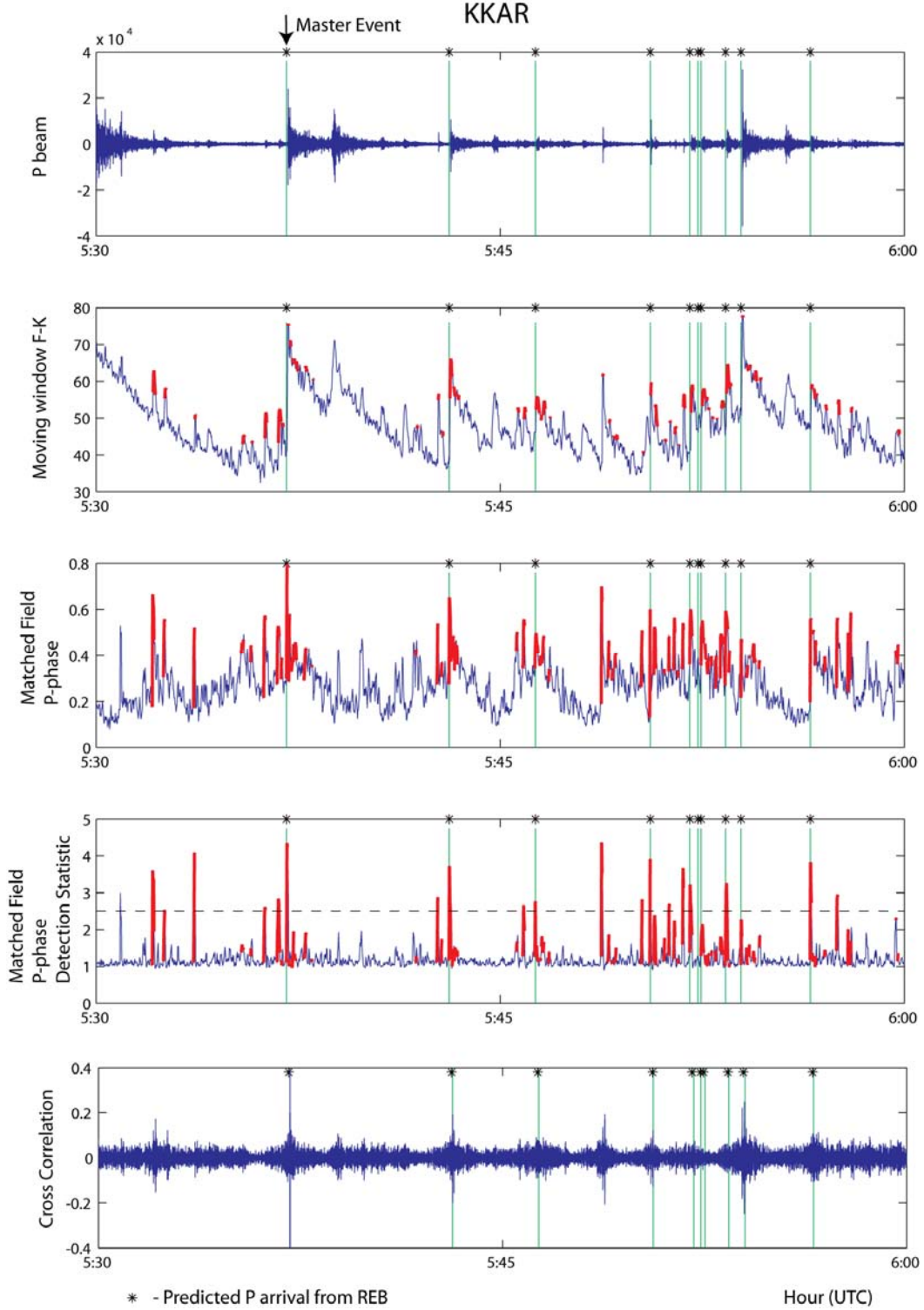


Figure 7.3: KK01 waveform (trace 1) together with 2-4 Hz f-k beam power (trace 2: red indicates slowness vector consistent with Pn phase from source region), scalar matched field statistic from the 2005-281:05.34.50 Pn arrival (trace 3), the weighted vector gradient (trace 4, c.f. Gibbons et al., 2008), and the correlation coefficient beam from a template from the 2005-281:05.34.50 event (trace 5).

Trace 3 from the top in Figure 7.3 displays the value of the matched field statistic where the empirical steering vector is generated from a single covariance matrix at a time 2005-281:05.37.06.7. This scalar peaks at the onset of each phase before trailing off into the coda. A transformation of the vector of narrow-band matched field statistics to a single "scaled gradient" scalar function (as described by Gibbons et al., 2008) was calculated in order to identify significant peaks. This function is displayed in Trace 4 from the top. In the lowest trace, the continuous correlation coefficient from the full waveform of this aftershock is displayed. It is clear that many more events are identified by the single-phase matched field statistic than from the correlation detector. Figure 7.4 shows a corresponding plot from a window containing the signal from the mainshock and the start of the aftershock sequence.

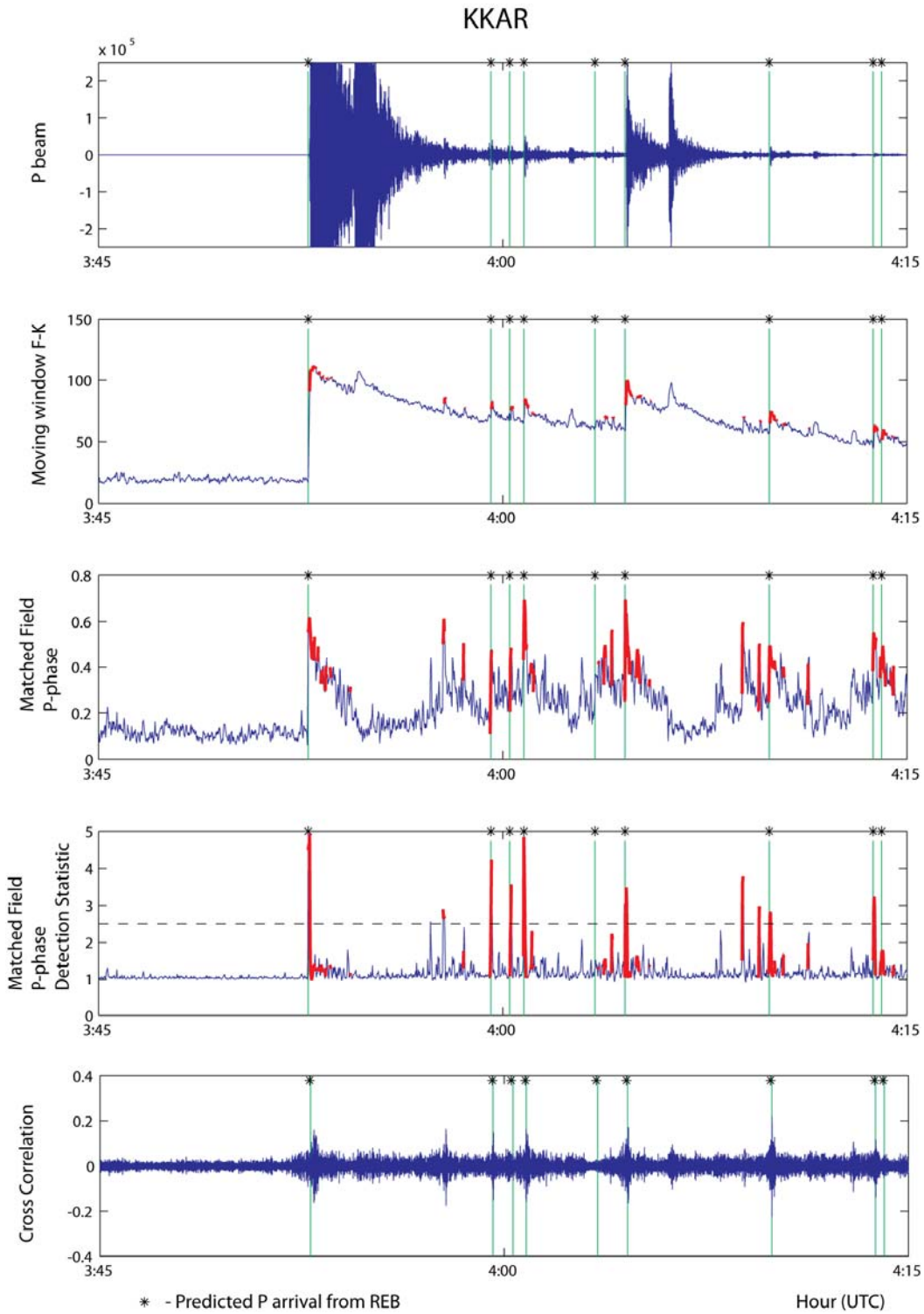


Figure 7.4: As for Figure 7.3 except that we consider the time-interval surrounding the main shock.

The KKAR Pn matched field statistic from the 2005-281:05.37.06.7 signal was calculated on continuous data for October 8, 2005, and the 20 maximum values throughout the day were examined and an ensemble covariance matrix from these events was generated. Figure 7.5 shows the rank-1, rank-2, and rank-3 matched field statistic traces for a short window surrounding a later Pn arrival. The values for the higher rank detection statistics are greater at the signal onset, but it is clear that the background level also increases significantly. This may be a result of the definition of ensemble covariance matrix which, designed to contain quite similar signals, may not contain significant structure in the higher order eigenvectors. Constructing higher rank matched field statistics from the resulting steering matrix may simply allow for a better match with noise segments. The effectiveness and performance of higher-rank matched field statistics and multiple rank-1 matched field statistics requires a more extensive investigation than can be provided here.

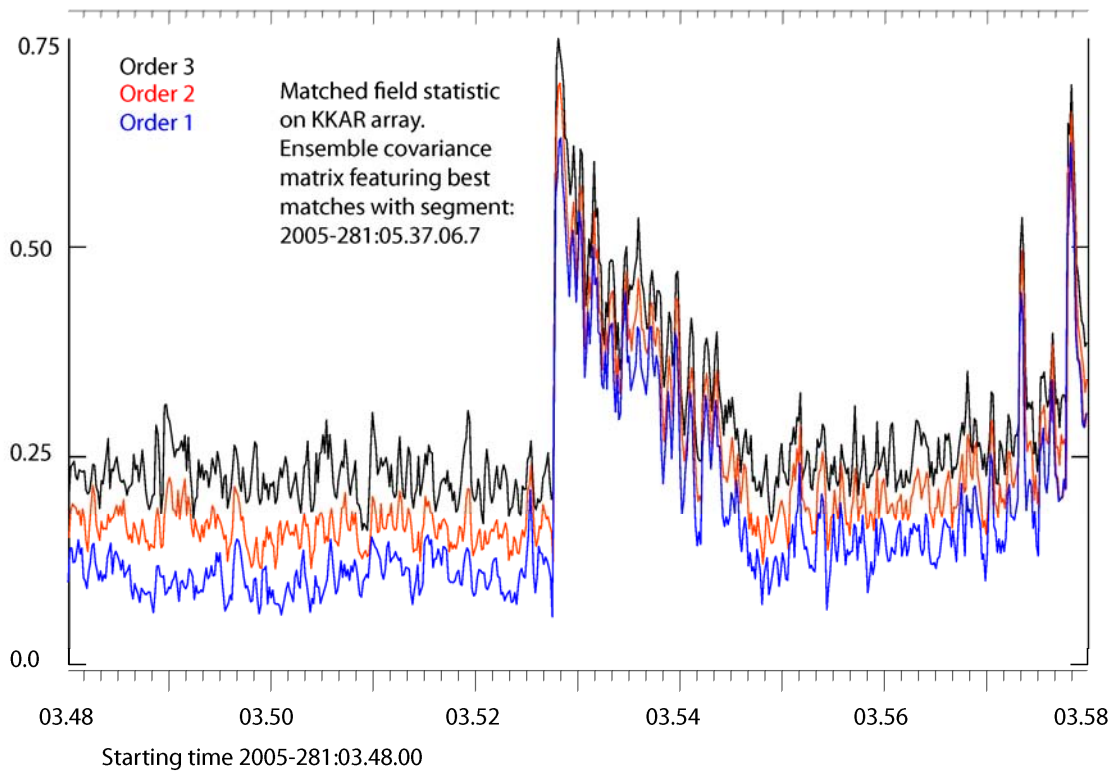


Figure 7.5: Continuous traces of matched field statistics with ranks 1, 2, and 3 where the ensemble covariance matrix is generated from the 20 best matches with the data segment starting 2005-281:05.37.06.7.

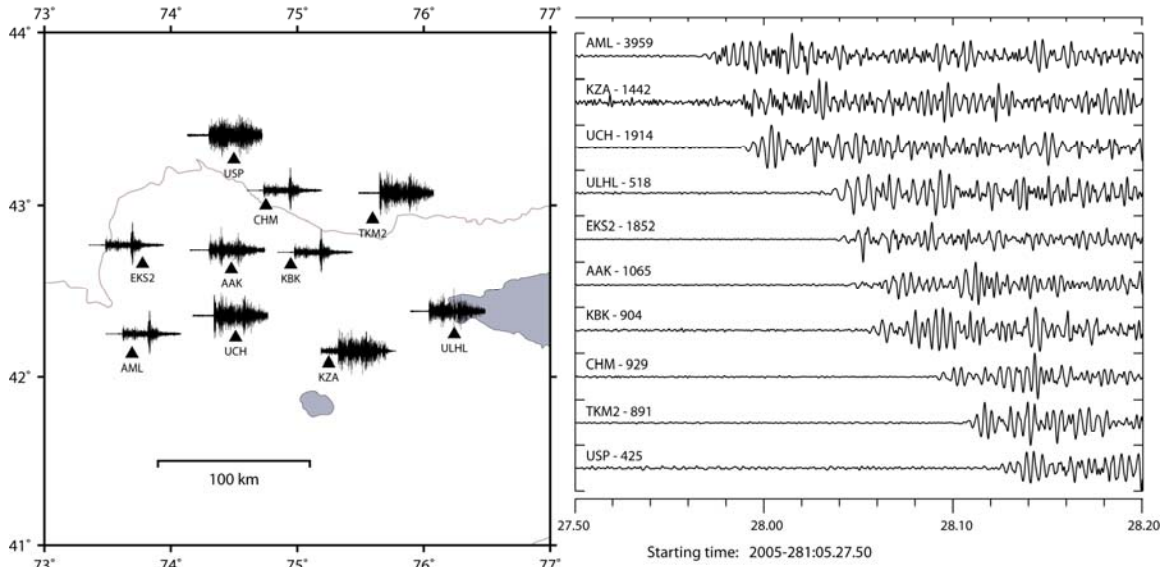


Figure 7.6: Waveforms over KNET (filtered) 2.0-8.0 Hz from an aftershock of the October 8, 2005, Kashmir event. (left) a four minute segment starting at a time 2005-281:05.27.03 and (right) a 30 second segment surrounding the Pn arrival illustrating the time delays across the network.

Figure 7.6 displays the waveforms from one aftershock over KNET showing how staggering of the waveforms is necessary prior to the calculation of a covariance matrix. On any 3-component station, it is possible to construct a 3 by 3 covariance matrix between the different sensors. (The relative sizes of the eigenvalues indicate the degree of polarization and the principal eigenvector describes the predominant particle motion.) It is therefore possible to calculate a 3-C matched field statistic for each station (see Figure 7.7). Trace 2 from the top displays the 3-C statistic for the NIL station and Trace 4 from the top displays the noisier trace for the more distant AAK station within KNET, with all traces being approximately aligned to account for the different travel-times from the source. The third trace from the top displays the (incoherent) sum of the 3-C matched field detection statistics stacked over all the stations of KNET with the appropriate time-delays (c.f. Figure 7.6). This incoherent matched field statistic beam over KNET displays a remarkably low variability of the background level and many more peaks can be identified on the beam than on the single channel AAK trace. The close correspondence between the peaks on the KNET beam and the NIL trace indicate that these arrivals are from events which are not located very far from the location of the master event.

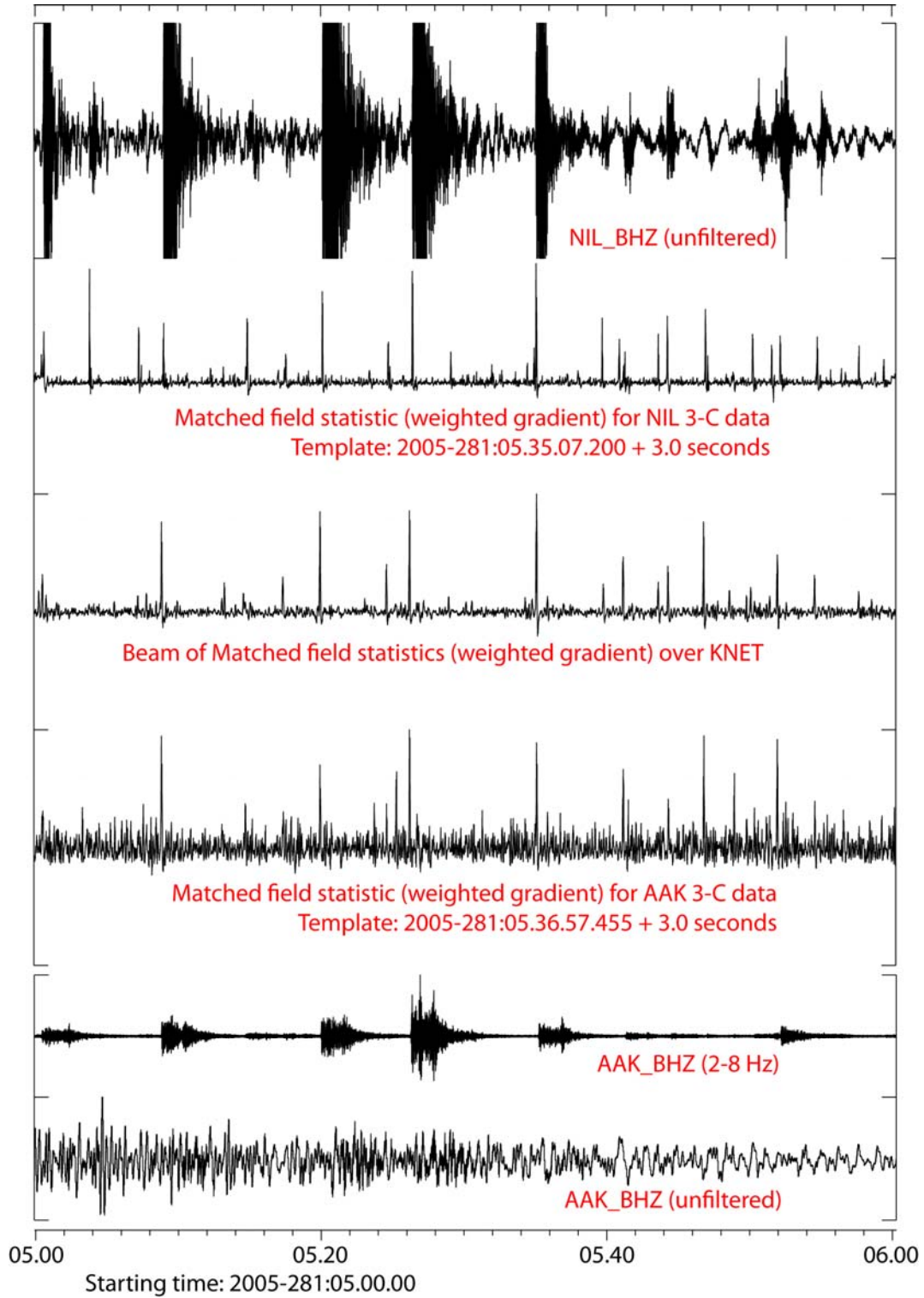


Figure 7.7: 3-component matched field detection statistic traces for the NIL and AAK stations, together with an (incoherent) beam of the corresponding traces across KNET.

Figure 7.8 displays these two traces together with a fully coherent matched field detection statistic for KNET. In the left of the two red boxes are three peaks on the fully coherent trace of which only the first is significant on the incoherent stack. The second peak on trace 5 is not recorded significantly on trace 4, but is present on the NIL trace, indicating that this is likely to be a genuine detection of a weak event. In the right of the two red boxes, a far greater peak is observed on the incoherent matched field statistic beam than on the coherent detection statistic. This is presumably an event which is somewhat further away from the source region which has led to a degradation of the coherent matched field statistic over the large network. It is however well detected by the partially coherent procedure and this detection could be used to generate a new coherent template to provide a more sensitive detector in the immediate vicinity of the new event.

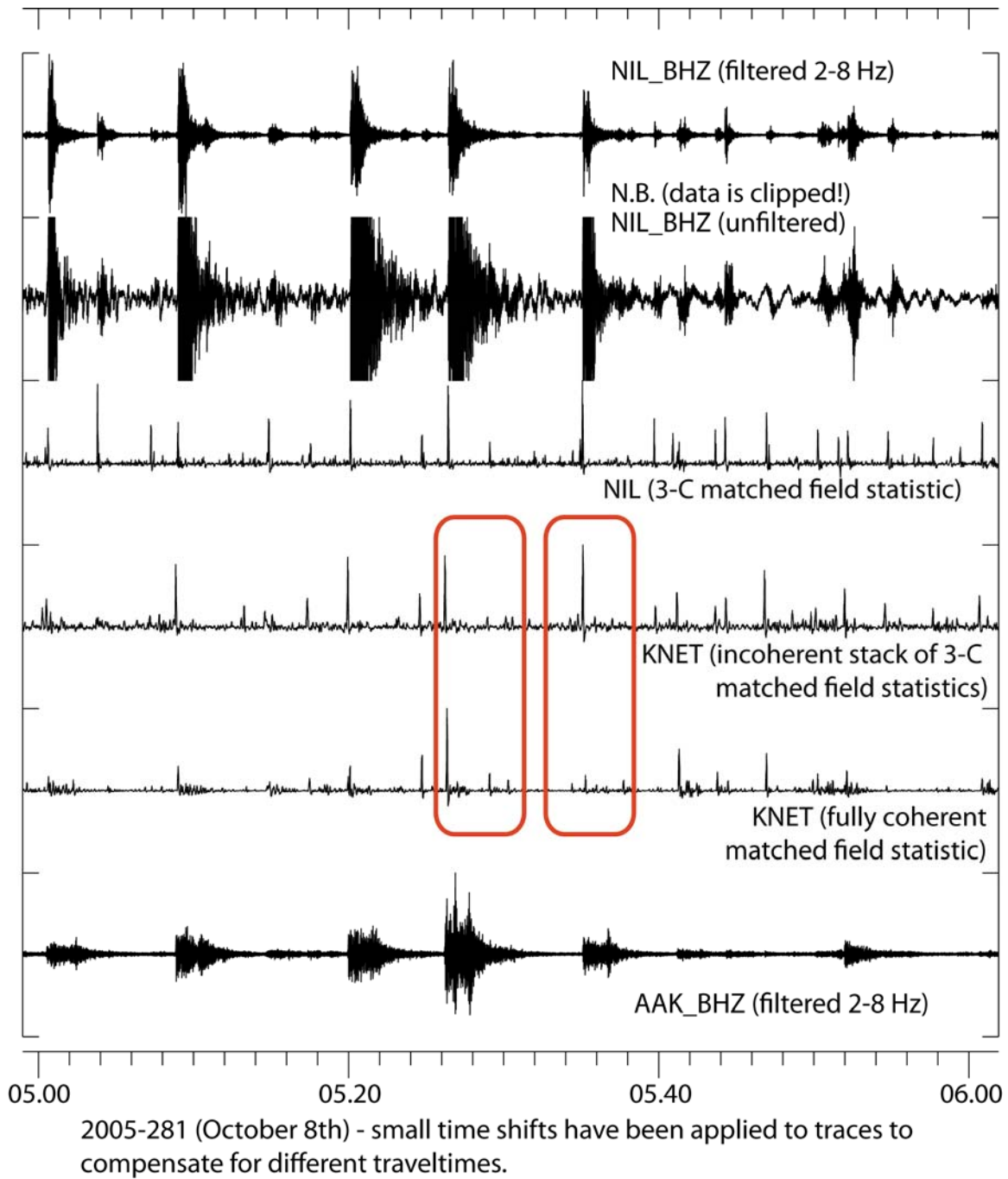


Figure 7.8: Comparison of coherent and incoherent matched field detection statistic beams for KNET for an aftershock of the October 8, 2005, Kashmir earthquake together with reference waveform traces from AAK (a station in KNET) and NIL (approximately 100 km from the epicenter).

Summary

Considering the extensive aftershock of the $M=7.4$ October 8, 2005, Kashmir earthquake, we have demonstrated that using a single Pn arrival on the KKAR array provides a detection statistic for identifying events from that source region which appears to outperform both the correlation detector and the standard pipeline f-k analysis. The f-k analysis is compromised by the limited frequency range over which coherent processing is possible and by the variability of the slowness vector as a function of frequency. The correlation detector is compromised by the very small geographic footprint of the target region and of the fact that a long wavetrain is required, causing possible interference with events closely spaced in time.

The empirical matched field processing method acts on a short data segment and yet is calibrated precisely to the spatial structure of the wavefield over the array aperture.

Monitoring the same sequence over the large aperture KNET network requires time-shifts to be applied because of the significant differences in travel-times to the different sensors. The fully coherent matched field statistic over the network is specific to a smaller source region than on the medium aperture KKAR array since differences between hypocenters will lead to greater phase shifts over the sensor aperture. However, a partially coherent procedure, whereby a 3-component matched field statistic is evaluated for each station of the network and then summed using appropriate time-delays, appears to provide a very robust detection statistic which covers a much broader geographical footprint. This suggests an operational pipeline for classifying the aftershock sequence whereby the partially coherent procedure identifies template events from distinct clusters, and numerous fully coherent procedures form more sensitive but more source-specific classifiers for each target location.

8. CONCLUSIONS

The overall objective of this three-year study has been to examine the limits of coherent processing on seismic sensor configurations of different apertures using empirical matched field processing (EMFP). EMFP is a narrowband technique which makes the detection and classification of seismic phases relatively insensitive to differences in source-time function. The technique is empirical, exploiting observations of historical events to calibrate the amplitude and phase structure of an incident wavefield over a given sensor configuration for particular repeating sources. Its applicability to wider sensor apertures than can process signals using the classical plane-wavefront formulation stems from the fact that matching with the incoming wavefield is done using an empirical description of the anticipated wavefield and not a theoretical one.

We have conducted a number of case-studies within the current contract:

Classification of quarry explosions on the Kola Peninsula

We have demonstrated the ability to distinguish between events from two mines on the Kola Peninsula using only the ARCES array at a distance of approximately 200 km. The mines are more closely spaced than can theoretically be resolved using the array. This study highlighted the importance of examining the similarity of the single-event empirical steering vectors prior to the construction of ensemble covariance matrices. Empirical steering vectors calculated from ensemble covariance matrices constructed from events that are too dissimilar will lead to a reduction in performance.

Classification of mining explosions in central Kazakhstan

EMFP also resolved a number of mining clusters in central Kazakhstan using only Pn arrivals on the MKAR array at a distance of approximately 400 km. MKAR is sparser than ARCES and it can be demonstrated that above 4 Hz (where the SNR is optimal) coherent f-k analysis rarely results in qualitatively correct slowness estimates. Matched field processing in the 6-10 Hz frequency band provided an excellent classification of the sources. We illustrate the fact that use of higher order matched field statistics can improve separation between clusters in cases where there is considerable variation within one or more of the clusters.

Events at the Kara Zhyra mine

We developed a routine for calculating covariance matrices on isolated short data segments using the multitaper method. We have demonstrated that, while it is often desirable to construct empirical steering vectors from ensemble covariance matrices calculated from multiple events, single-event calibrations frequently perform well making EMFP applicable to sites from which only a single event has been observed. We have

demonstrated that single-phase EMFP can constitute a working source-specific detector with a low false alarm rate.

Application of EMFP to diffuse seismicity

We identified a region in central Asia with diffuse seismicity and calculated empirical steering vectors from Pn arrivals at the MKAR array. Cluster analysis on the resulting similarity matrix indicated that many events resulted in a similar wavefield over MKAR which differed significantly from the theoretical plane wavefront. It was demonstrated that the improvement that EMFP provides in wavefield characterization over the theoretical wavefield description became greater as the frequency increased. The small number of events identified in each cluster, together with the significant location uncertainty, precluded the possibility of constructing covariance matrices from large ensembles of events from a broad source region, from which higher order matched field statistics could be calculated. We advocate investigating this question further in a region of diffuse seismicity for which event locations are better constrained.

Application of EMFP to an aftershock sequence

We considered the extensive aftershock sequence to the $M=7.4$ October 8, 2005, Kashmir earthquake and demonstrated that an empirical steering vector calculated from a Pn arrival at the KKAR array from a single aftershock provided a characterization of the sequence which appeared to perform favorably compared with classical continuous f-k processing and full waveform correlation. As the receiver aperture increases, we anticipate the source region over which the matched field calibration is specific to decrease. A single-event matched field detector on, e.g. KNET, is therefore likely to detect events over a much smaller source region than, for example, KKAR. However, a partially coherent system where the 3-component matched field statistics are beamformed over the stations of KNET appears to form a very robust detector for larger magnitude events from a broader source region. This opens the possibility for a matched field characterization of the aftershock sequence using the partially coherent procedure for identifying master events, and the fully coherent procedure for detecting smaller events close to the master events.

REFERENCES

- Baggeroer, A.B., Kuperman, W.A. & Mikhalevsky, P.N. (1993). An overview of matched field methods in ocean acoustics. *IEEE Journal of Oceanic Engineering*, **18**, 401-424.
- Fialkowski, L.T., Collins, M.D., Kuperman, W.A., Perkins, J.S., Kelly, L.J., Larsson, A., Fawcett, J.A. & Hall, L.H. (2000). Matched-field processing using measured replica fields. *J. Acoust. Soc. Am.*, **107**, 739-746..
- Gibbons, S.J. & Ringdal, F. (2006). The detection of low magnitude seismic events using array-based waveform correlation. *Geophys. J. Int.*, **165**, 149–166.
- Gibbons, S.J., Kverna, T. & Ringdal, F. (2005). Monitoring of seismic events from a specific source region using a single regional array: a case study. *J. Seismol.*, **9**, 277-294.
- Gibbons, S.J., Kverna, T. & Ringdal, F. (2010). Considerations in phase estimation and event location using small-aperture regional seismic arrays. *Pure appl. geophys.*, **167**, 381–399.
- Gibbons, S.J., Ringdal, F. & Kverna, T. (2008). Detection and characterization of seismic phases using continuous spectral estimation on incoherent and partially coherent arrays. *Geophys. J. Int.*, **172**, 405-421.
- Gubbins, D. (2004). Time Series Analysis and Inverse Theory for Geophysicists. Cambridge University Press, Cambridge, UK.
- Harris, D.B. (1991). A waveform correlation method for identifying quarry explosions. *Bull. Seism. Soc. Am.*, **81**, 2395-2418.
- Harris, D.B. (2006). Subspace Detectors: Theory. *Tech. Rep. UCRL-TR-222758*, Lawrence Livermore National Laboratory.
- Harris, D.B. & Kverna, T. (2010). Superresolution with seismic arrays using empirical matched field processing. *Geophys. J. Int.*, **182**, 1455–1477.
- Harris, D.B., Ringdal, F., Kremenetskaya, E.O., Mykkeltveit, S., Schweitzer, J., Hauk, T.F., Asming, V.E., Rock, D.W. & Lewis, J.P. (2003). Ground-Truth Collection for Mining Explosions in Northern Fennoscandia and Russia. In *Proceedings of the 25th Seismic Research Review, Tucson, Arizona, September 23-25, 2003. Nuclear Explosion Monitoring Building the Knowledge Base, LA-UR-03-6029*, 54–63.
- Harris, F.J. (1978). On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform. *Proc. IEEE*, **66**, 51–83.
- Hartse, H.E., Randall, G.E. & Arrowsmith, S.J. (2008). Regional Event Identification Research in Asia. In *Proceedings of the 30th Monitoring Research Review, Portsmouth, Virginia, September 23-25, 2008. Ground-Based Nuclear Explosion Monitoring Technologies, LA-UR-08-05261*, 615–624.

- Kværna, T. & Ringdal, F. (1986). Stability of various f-k estimation techniques. *NORSAR Scientific Report: Semiannual Technical Summary No. 1 - 1986/1987*, NORSAR, Kjeller, Norway, pp. 29-40.
- Kværna, T., Gibbons, S.J., Ringdal, F. & Harris, D.B. (2004). Integrated Seismic Event Detection and Location by Advanced Array Processing. In *Proceedings of the 26th Seismic Research Review, Orlando, Florida, September 21-23, 2004. Trends in Nuclear Explosion Monitoring. Los Alamos National Laboratory, LA-UR-04-5801*
- Kværna, T., Gibbons, S.J., Ringdal, F. & Harris, D.B. (2007). Integrated Seismic Event Detection and Location by Advanced Array Processing. Tech. Rep. UCRL-SR-228092, Lawrence Livermore National Laboratory, Livermore, CA, United States.
- Lees, J.M. & Park, J. (1995). Multiple-taper spectral analysis: a stand-alone C-subroutine. *Comput. Geosci.*, **21**, 199–236.
- MacCarthy, J., Hartse, H., Greene, M. & Rowe, C. (2008). Using Waveform Cross-correlation and Satellite Imagery to Identify Repeating Mine Blasts in Eastern Kazakhstan. *Seism. Res. Lett.*, **79**, 393–399.
- Prieto, G.A., Parker, R.L. & Vernon, F.L. (2009). A Fortran 90 library for multitaper spectrum analysis. *Computers and Geosciences*, **35**, 1701–1710.
- Schweitzer, J. (2001). Slowness Corrections - One Way to Improve IDC Products. *Pure appl. geophys.*, **158**, 375-396.
- Thomson, D.J. (1982). Spectrum Estimation and Harmonic Analysis. *Proc. IEEE*, **70**, 1055–1096.
- Ward, J.H. (1963). Hierarchical grouping to optimize an objective function. *J. Stat. Soc. Am.*, **58**, 236–244.
- Wessel, P. & Smith, W.H.F. (1995). New version of the Generic Mapping Tools. *EOS Trans. Am. Geophys. Union*, **76**, 329.

APPENDIX A: CALCULATING SPECTRAL COVARIANCE MATRICES USING THE MULTITAPER METHOD

Harris & Kværna (2010) provide a detailed description of a procedure for obtaining a narrow-band representation of an observed wavefield from which the covariance matrices can be obtained. The clustering and classification examples in Chapters 3 and 4 use this method exclusively to calculate the spectral covariance matrices.

Seismic signals are transient and it may be desirable to estimate the spectral covariance matrix over a very short and precisely defined time-window. There may be a very limited number of samples over which signal coherence between two sensors is observed, or we may want to calculate a spatial covariance matrix for one phase while avoiding a different phase arriving shortly before or after. Another motivation for wanting to examine the spatial covariance matrix over a precisely defined set of samples is to be able to provide a direct comparison between empirical matched field estimates and existing classical f-k analysis estimates on identical waveform segments.

The problem of how to estimate the spectrum of a time-limited signal is a nontrivial issue (see for example Gubbins, 2004). The sampling interval, Δt , limits the maximum meaningful frequency to the Nyquist frequency, $f_N = 1/(2\Delta t)$, and the length of the time series, T , limits the frequency spacing, Δf , to $\Delta f = 1/T$. In addition, spectral estimates from short data segments may be heavily biased due to spectral leakage whereby the cut-off in the time-domain forces energy at a given frequency into side lobes centered far from the true frequency. Tapering the data (see for example Harris, 1978) mitigates the spectral leakage problem at the expense of down-weighting valuable data samples and (in choosing the length and shape of tapering functions) introducing an arbitrary form of parametrization to the estimate. Stable, unbiased, and high-resolution spectral estimates are provided using the multitaper methods devised by Thomson (1982) with computational algorithms available through sources such as the commercial platform MATLAB and the published source code of Lees & Park (1995).

The freely available source code for multitaper spectral analysis was expanded significantly by Prieto et al. (2009) who also provide computational routines for the calculation of coherence for multivariate problems. The source code, additional documentation, and examples are provided at

<http://wwwprof.uniandes.edu.co/~gprieto/software/mwlib.html>

and the following piece of FORTRAN 90 code is a modification of the `mt_cohe` routine from Prieto et al. (2009) adapted to calculate the spectral covariance matrix from N-channel data.

```

! "n_mt_cohe.f90"
! Steven J. Gibbons Gibbons
! Mon Oct 19 09:37:49 CEST 2009
!
! Extends the multitaper coherence routine (mtcohe) of
! German Prieto to N-channels ...
!
! The original routines are described in
! "A Fortran 90 library for multitaper spectrum analysis"
! by G. A. Prieto, R. L. Parker, and F. L. Vernon III
! Computers and Geosciences, vol 35, 2009, pp. 1701-1710.
!
! nc is the number of channels to be examined ...
! each of the time-series contains npts samples.
!
! All the time series data is input via a real*4 array rwf and
! the locations of elements are specified by the integer array
! of pointers iptar.
!
! channel ic is contained in the elements
!
! rwf( iptar(ic)      )
! rwf( iptar(ic) + 1  )
! rwf( iptar(ic) + 2  )
!   :
!   :
! rwf( iptar(ic) + npts - 1 )
!
! rwf has dimension dimrwf
!
! If iptar(ic).eq.-1 then the channel is absent and we set
! the corresponding parts of the covariance matrix to the
! identity matrix ...
!
! A starting time channel ic is stored in element ic of the
! real*8 array dtimrel - these can be epoch times with respect
! to an arbitrary common time, since only the difference
! between the different dtimrel is referred to.
! This means that the different time-series can be staggered,
! and that the starting times are accounted for in the
! calculation of the phase differences.
! However, the sampling time interval has to be dt for all
! channels.
!
! nf, the number of frequency bins should be set to (npts/2 + 1)
!
! The output is in 3 real*4 arrays:
!
! rcmpcov( nc*(nc+1)*nf )
! rcohe(  nc*(nc+1)*nf/2 )
! rphas(  nc*(nc+1)*nf/2 )
!
! Each array contains the upper triangular part of the
! nc*nc matrices rcmpcov - the covariance matrix (complex

```

```

! numbers stored real, imag - hence twice the length)
! rcohe - the coherence - see Prieto et al. (2009) and
! rphas (both rcohe and rphas are real)
!
! The elements corresponding to frequency band 1 are stored
! first, followed by the elements corresponding to
! frequency band 2 etc.
!
! Within each frequency band, ifreq, we loop
!
! indexR = nc*(nc+1)/2*(ifreq-1)
! indexC = nc*(nc+1)*(ifreq-1)
! do i = 1, nc
!   do j = i, nc
!     Rij_real = rcmpcov( indexC+1 )
!     Rij_imag = rcmpcov( indexC+2 )
!     indexC = indexC + 2
!     COHij = rcohe( indexR+1 )
!     PHAij = rphas( indexR+1 )
!     indexR = indexR + 1
!   enddo
! enddo
!
subroutine n_mt_cohe( npts, nc, dt, dimrwf, rwf, &
                    iptar, tbp, kspec, &
                    nf, freqinc, rcmpcov, rcohe, &
                    rphas, iadapt, ierr, dtimrel )
!
use spectra
!
implicit none
!
! INPUT
!
!   npts      integer number of points in time series
!   nc        integer number of channels to be considered
!   dt        real, sampling rate of time series
!   dimrwf    integer dimension of the data array
!   rwf(dimrwf) real, data array
!   iptar(nc) integer, array of pointers
!   tbp       real, time-bandwidth product
!   kspec     integer, the number of tapers to use
!   nf        integer, number of freq points in spectrum
!   freqinc   real, the step in frequency
!   rcmpcov   real, the complex covariance matrix
!             in upper-right storage.
!             dimension ( nc*(nc+1)*nf )
!   rcohe     real, coherence values
!             dimension ( nc*(nc+1)*nf/2 )
!   rphas     real, phase differences
!             dimension ( nc*(nc+1)*nf/2 )
!   iadapt    integer 0 - adaptive, 1 - constant weights
!             default adapt = 1
!   ierr      integer: error flag. 0 good - other numbers

```

```

!           less good.
!   dtimrel    double precision, relative times of the traces
!           dimension ( nc )
!           Only the differences between these times
!           are considered ...
!
integer, intent(in) :: npts, nc, dimrwf, kspec, nf, iadapt
integer, intent(out) :: ierr
integer, dimension(nc), intent(in) :: iptar
real(4), intent(in) :: dt, tbp
real(4), intent(out) :: freqinc
real(4), dimension( dimrwf ), intent(in) :: rwf
real(4), dimension( nc*(nc+1)*nf ), intent(out) :: rcmpcov
real(4), dimension( nc*(nc+1)*nf/2 ), intent(out) :: rcohe
real(4), dimension( nc*(nc+1)*nf/2 ), intent(out) :: rphas
real(8), dimension( nc ), intent(in) :: dtimrel
!
! Local arrays and other variables ...
!
integer, dimension(nc) :: iptar2
real(4), dimension(nf) :: f, si, wt_scale
real(4), dimension(nf,kspec) :: wt_i
real(4), dimension(nf,kspec,nc) :: wtarr
real(4), dimension(nf,nc) :: sarr
real(4), dimension(npts) :: xi
!
complex(4), dimension(npts,kspec) :: yk_i
complex(4), dimension(npts,kspec,nc) :: ykarr
!
! Coherence freq matrices ...
!
complex(4), dimension(nf,kspec) :: dyk_i
complex(4), dimension(nf,kspec,nc) :: dykarr
!
real(4), dimension(nf) :: cohe, phase
complex(4), dimension(nf) :: spec
complex(4) :: cphass
!
real(4), parameter :: dtol = 0.000001
real(8), parameter :: dpi = 3.14159265358979312d0
real(8) :: dfreq, delfrq, dtd2pi
real(8) :: dphimp, dtdiff
real(8) :: drealtv, dimagv
!
! Temporary variables ...
!
integer :: ic, ip_first, ip_last, i, iad, ispec, ifreq
integer :: ipair, ic1, ic2, npairs, npairs2, init, idest
integer :: iptcov_real, iptcov_imag, iptcoh, iptpha
real(4) :: sumsqr
!
ierr = 0
!
iptar2 = iptar

```

```

!
if ( npts.lt.4 ) then
  write (6,*) 'Subroutine n_mt_cohe: Error ... '
  write (6,*) ' npts = ', npts
  ierr = 1
  return
endif
if ( nf.lt.4 ) then
  write (6,*) 'Subroutine n_mt_cohe: Error ... '
  write (6,*) ' nf = ', nf
  ierr = 1
  return
endif
if ( nc.lt.2 ) then
  write (6,*) 'Subroutine n_mt_cohe: Error ... '
  write (6,*) ' nc = ', nc
  ierr = 1
  return
endif
if ( dt.lt. dtol ) then
  write (6,*) 'Subroutine n_mt_cohe: Error ... '
  write (6,*) ' dt = ', dt
  ierr = 1
  return
endif
if ( iadapt.ne.0 .and. iadapt.ne.1 ) then
  write (6,*) 'Subroutine n_mt_cohe: Error ... '
  write (6,*) ' iadapt = ', iadapt
  ierr = 1
  return
endif
iad = iadapt
npairs = nc*(nc+1)
npairs2 = npairs/2
!
! The dimensions of the arrays appear to be sensible
! so we can then zero all of our output vectors ...
!
do i = 1, npairs*nf
  rcmpcov( i ) = 0.0
enddo
do i = 1, npairs2*nf
  rcohe( i ) = 0.0
enddo
do i = 1, npairs2*nf
  rphas( i ) = 0.0
enddo
!
! Need to check on the indices and the array dimensions ...
! Remember that all ipt2( ic ) are set to -1 if the channel
! is not found ...
!
do ic = 1, nc
  ip_first = ipt2( ic )

```

```

if ( ip_first.ne.-1 ) then
  ip_last = ip_first + npts - 1
  if ( (ip_first.lt.1) .or. &
      (ip_last.gt.dimrwf) ) then
    write(6,*) 'Subroutine n_mt_cohe: Error ... '
    write(6,*) ' Channel ', ic
    write(6,*) ' PTR( ', ic, ' ) = ', ip_first
    write(6,*) ' npts      = ', npts
    write(6,*) ' dimrwf    = ', dimrwf
    ierr = 1
    return
  endif
! Nasty things will happen if the channel is zero ...
! If we detect a zero channel, we will set ipta2( ic )
! to -1 to prevent the channel being looked at ...
  sumsqr = 0.0
  do i = ip_first, ip_last
    sumsqr = sumsqr + rwf( i )**2
  enddo
  if ( sumsqr.lt. dtol ) then
    ipta2( ic ) = -1
  endif
endif
enddo
!
! Now loop around the channels and calculate the spectra ...
!
do ic = 1, nc
  ip_first = ipta2( ic )
  if ( ip_first.ne.-1 ) then
    ip_last = ip_first + npts - 1
    idest = 1
    do i = ip_first, ip_last
      xi( idest ) = rwf( i )
      idest = idest + 1
    enddo
    call mtspec( npts, dt, xi, tbp, kspec, nf, f, &
                si, yk=yk_i, wt=wt_i, adapt=iad )
    wtarr( :, :, ic ) = wt_i( :, : )
    ykarr( :, :, ic ) = yk_i( :, : )
    freqinc      = f( 2 )
    delfrq       = dble( freqinc )
  endif
enddo
!
! Use the minimum weights -
!
init = 0
do ic = 1, nc
  if ( ipta2( ic ).ne.-1 ) then
    if ( init.eq.0 ) then
      wt_i = wtarr( :, :, ic )
      init = 1
    else

```

```

        wt_i = min( wt_i, wtarr( :, :, ic ) )
    endif
endif
enddo
!
wt_scale = sum( wt_i**2, dim=2 ) ! Scale weights to keep power ...
do ic = 1, nc
    if ( ipta2( ic ).ne.-1 ) then
        do ispec = 1, kspec
            wtarr( :, ispec, ic ) = wtarr( :, ispec, ic )/sqrt(wt_scale)
        enddo
    endif
enddo
!
! Now calculate the complex spectra ...
!
do ic = 1, nc
    if ( ipta2( ic ).ne.-1 ) then
        do ifreq = 1, nf
            do ispec = 1, kspec
                dyk_i(ifreq,ispec) = &
                    wtarr( ifreq, ispec, ic )*ykarr( ifreq, ispec, ic )
            enddo
        enddo
        si = sum( abs(dyk_i)**2, dim=2 )
        dykarr(:, :, ic) = dyk_i(:, :)
        sarr( :, ic ) = si( : )
    endif
enddo
!
! Now loop around all pairs and calculate the cross-spectra ...
!
ipair = 0
do ic1 = 1, nc
    do ic2 = ic1, nc
        ipair = ipair + 1
        if ( ipta2(ic1).ne.-1 .and. ipta2(ic2).ne.-1 ) then
            dtdiff = dtimrel( ic1 ) - dtimrel( ic2 )
            dtd2pi = 2.0d0*dpi*dtdiff
            dfreq = 0.0d0
            do ifreq = 1, nf
                dphimp = dtd2pi*dfreq
                drealv = dcos( dphimp )
                dimagv = dsin( dphimp )
                cphass = cmplx( real(drealv), real(dimagv) )
                spec( ifreq ) = &
                    sum( dykarr(ifreq,:,ic2) * conjg( dykarr(ifreq,:,ic1) ) )
                spec( ifreq ) = spec( ifreq )*cphass
                cohe( ifreq ) = &
                    (abs(spec(ifreq)))**2/(sarr(ifreq,ic1)*sarr(ifreq,ic2))
                phase( ifreq ) = &
                    atan2( aimag( spec(ifreq) ), real( spec(ifreq) ) )
                dfreq = dfreq + delfrq
            enddo
        enddo
    enddo
enddo

```

```

!      !
!      ! OK - we have calculated the complex values ...
!      ! Now we need to put them into our output arrays in the
!      ! right places ...
!      !
      iptcov_imag = 2*ipair
      iptcov_real = iptcov_imag - 1
      iptcoh      = ipair
      iptpha      = ipair
      do ifreq = 1, nf
        rcmpcov( iptcov_real ) = real( spec(ifreq) )
        rcmpcov( iptcov_imag ) = aimag( spec(ifreq) )
        rcohe( iptcoh )      = cohe( ifreq )
        rphas( iptpha )      = phase( ifreq )
        iptcov_real = iptcov_real + npairs
        iptcov_imag = iptcov_imag + npairs
        iptcoh      = iptcoh      + npairs2
        iptpha      = iptpha      + npairs2
      enddo
!      !
      endif
    enddo
  enddo
!
  return
end
!

```

The routine was tested initially by reproducing the matched field classification of the mining events previously evaluated using the method described by Harris & Kværna (2010) and references therein. When generating covariance matrices from which empirical steering vectors were to be derived, it became routine practice to calculate several estimates in overlapping time-windows by shifting the start of the data window along by a single sample, and then using a mean of the resulting normalized covariance matrices. For example, if 121 samples (3 seconds on 40 Hz data) were to be used for a covariance matrix estimate, we would frequently calculate 20 estimates on consecutive, overlapping data segments, requiring a total of 141 samples (or 3.5 seconds).

List of Symbols, Abbreviations, and Acronyms

AFRL	Air Force Research Laboratory
EMFP	Empirical Matched Field Processing
SASC	Slowness and Azimuth Station Correction