AFRL-RV-PS-
TP-2017-0006

COMPREHENSIVE SEISMIC DETECTION AND ESTIMATION USING MATCHED FIELD PROCESSING
Annual Report

Steven J. Gibbons, et al.

NORSAR
PO Box 53
Gunnar Randers vei 15
N-2027 Kjeller, Norway

08 November 2016

Technical Report

AIR FORCE RESEARCH LABORATORY
Space Vehicles Directorate
3550 Aberdeen Ave SE
AIR FORCE MATERIEL COMMAND
KIRTLAND AIR FORCE BASE, NM 87117-5776

DTIC COPY

NOTICE AND SIGNATURE PAGE

AFRL-RV-PS-TP-2017-0006 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//SIGNED//	//SIGNED//
Dr. Frederick Schult Program Manager, AFRL/RVBYE	Dr. Thomas R. Caudill, Acting Chief AFRL Battlespace Environment Division

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
08-11-2016	Technical Report	08 Oct 2015 – 08 Oct 2016

4. TITLE AND SUBTITLE

COMPREHENSIVE SEISMIC DETECTION AND ESTIMATION USING MATCHED FIELD PROCESSING
Annual Report

5a. CONTRACT NUMBER
FA9453-16-C-0006

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
62601F

6. AUTHOR(S)
Steven J. Gibbons, Tormod Kværna, David B. Harris[1], Douglas A. Dodge[2]

5d. PROJECT NUMBER
1010

5e. TASK NUMBER
PPM00027855

5f. WORK UNIT NUMBER
EF128707

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

NORSAR
PO Box 53
Gunnar Randers vei 15
N-2027 Kjeller, Norway

[1]Deschutes Signal Processing, LLC
81211 E. Wapinitia Road
Maupin, OR 97037

[2]Lawrence Livermore National Laboratory
7000 East Avenue
Livermore, CA 94550

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory
Space Vehicles Directorate
3550 Aberdeen Avenue SE
Kirtland AFB. NM 87117-5776

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/RVBYE

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-RV-PS-TP-2017-0006

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited. (AFMC-2018-0112 dtd 20 Jun 2018)

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Empirical Matched Field Processing (EMFP) is a narrow frequency band method for signal detection and estimation which recognizes a characteristic phase and amplitude structure over an array of sensors. EMFP can provide an enhanced capability for detecting and identifying signals of special interest from a target source region and we here explore the extent to which it can constitute a comprehensive front-end detector for general seismic signals. On small aperture seismic arrays, we demonstrate that a relatively small number of matched field templates are required to cover the parameter space of anticipated signals. In addition, estimates of apparent velocity and backazimuth made directly from the empirical steering vectors are likely to be far more accurate than those made from classical array processing since the bias resulting from path and near-receiver characteristics are corrected for empirically. On larger aperture arrays, significantly more templates are required to provide the same coverage of the anticipated parameter space. On seismic arrays for which signal coherence in the frequency band of interest is poor, a given matched field detector will provide a very signal-specific detector and vast numbers of templates would be required for a comprehensive detector. We demonstrate how classical slowness estimates for high frequency signals of interest can be enhanced by considering only selected sensor pairs and recommend similar approaches for mitigating signal incoherence for general matched field detectors.

15. SUBJECT TERMS
nuclear explosion monitoring, signal detections seismic monitoring, array processing

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Dr. Frederick Schult
Unclassified	Unclassified	Unclassified	Unlimited	58	**19b. TELEPHONE NUMBER** *(include area code)*

This page is intentionally left blank.

Table of Contents

List of Figures

1. SUMMARY

Empirical Matched Field Processing (EMFP) is a narrow frequency band method for signal detection and estimation which recognizes a characteristic phase and amplitude structure over an array of sensors. EMFP can provide an enhanced capability for detecting and identifying signals of special interest from a target source region and we here explore the extent to which it can constitute a comprehensive front-end detector for general seismic signals. On small aperture seismic arrays, we demonstrate that a relatively small number of matched field templates are required to cover the parameter space of anticipated signals. In addition, estimates of apparent velocity and backazimuth made directly from the empirical steering vectors are likely to be far more accurate than those made from classical array processing since the bias resulting from path and near-receiver characteristics are corrected for empirically. On larger aperture arrays, significantly more templates are required to provide the same coverage of the anticipated parameter space. On seismic arrays for which signal coherence in the frequency band of interest is poor, a given matched field detector will provide a very signal-specific detector and vast numbers of templates would be required for a comprehensive detector. We demonstrate how classical slowness estimates for high frequency signals of interest can be enhanced by considering only selected sensor pairs and recommend similar approaches for mitigating signal incoherence for general matched field detectors.

2. INTRODUCTION

Methods for detecting seismic signals can be classified in many ways. One classification is whether or not we know what the signal looks like a priori. Given only a single channel, a signal which we do not know the shape of is detected using a power detector - resulting in a trigger when a sudden increase in amplitudes on a time-series is encountered (see e.g. Withers et al., 1998, for an overview). If we do know the shape of an anticipated signal beforehand, we can use correlation or matched filter techniques (e.g. Gibbons and Ringdal, 2006), or more general subspace detectors (e.g. Harris, 2006; Barrett and Beroza, 2014), which recognize the ripple-for-ripple similarity between the incoming waveform and a template signal.

When we have an array of sensors (see Figure 1) the distinction is more subtle. The spatial distribution of sensors means that, although we don't know the shape of the actual ripples beforehand, we are looking for patterns of anticipated time-delays or phase shifts between the different sites. Even the delay-and-stack methods of detecting general signals on a seismic array are therefore pattern detectors to a certain degree.

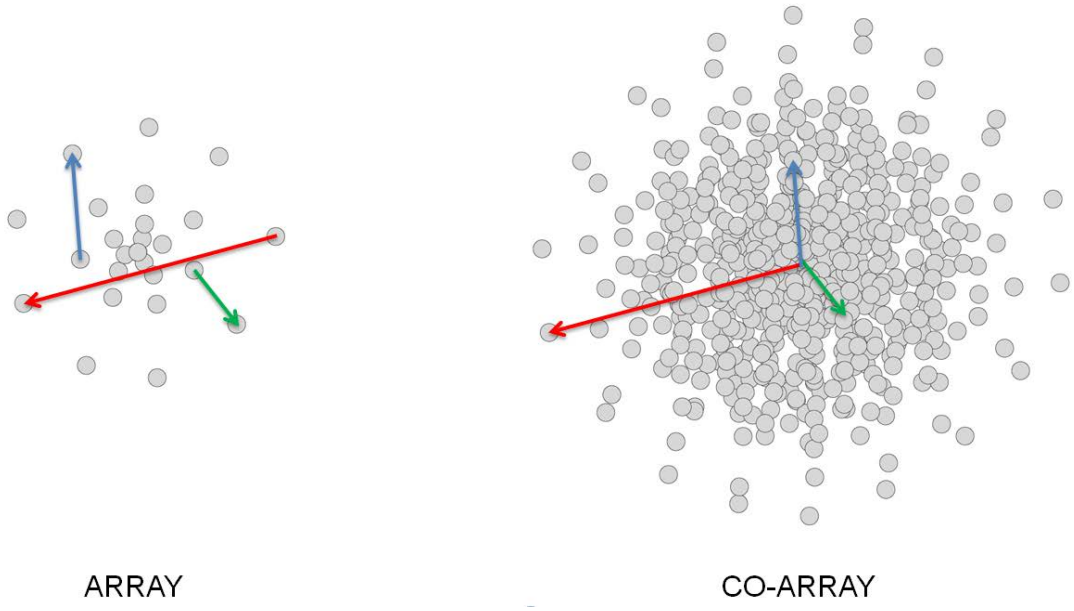

Figure 1: The relationship between the geometry of a seismic array and its co-array. ***The co-array consists of the set of vector differences between each location in the array.***

The co-array is a useful construct since it allows us to visualize the covariance matrix evaluated on a short time-window of data on a seismic array. A typical example is shown in Figure 2, both for an ideal or theoretical signal and for a real-world signal coming from the same direction. For a theoretical plane wavefront (left), the signals across the array aperture are perfectly coherent and the phase-shifts conform exactly to a set of time-

delays defined by a steady progression of a plane wavefront over the sensors. On real world seismic arrays (right), the situation is more complicated.

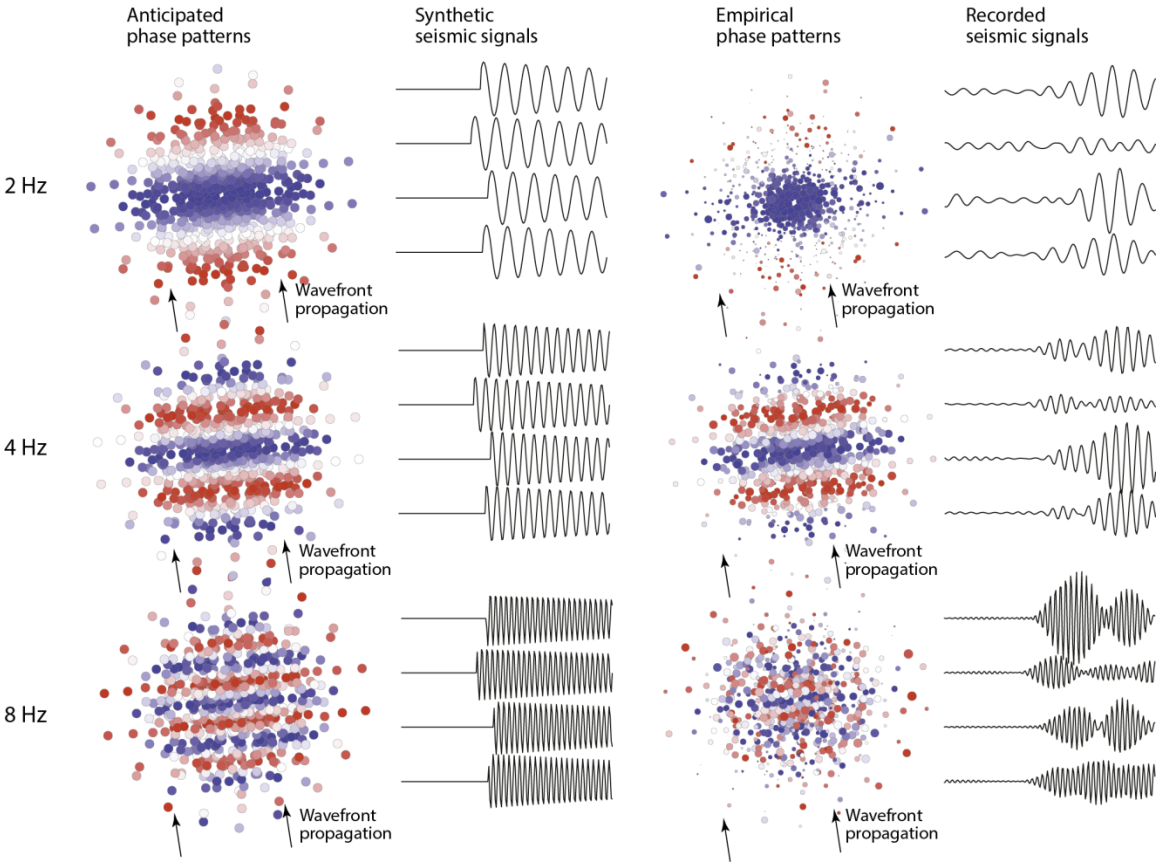

Figure 2: Illustration of phase shift and coherence patterns over a seismic array for an incoming wavefront for a theoretical plane wave (left) and real data (right). ***The signal is broken down into many very narrow frequency bands (of which 3 are displayed here). The colors of the points of the co-array represent the phase-shift for that particular frequency band between the corresponding two sensors, and the size of the symbols indicates the coherence – a measure of the similarity between the narrow-band signals on the two sensors. A given set of symbols is a graphical representation of the covariance matrix evaluated on the incoming wavefront over an array over a short time interval.***

At 4 Hz, the empirical and theoretical covariance matrices look very similar, with a banded structure of phase shifts. However, we note that the symbols in the center of the co-array are larger than those at the periphery – meaning that the signals on closely-spaced sensors are more similar than those on more distant sensors. At 8 Hz, for which the seismic wavelengths are shorter and the wavefield more sensitive to small-scale heterogeneities in the real earth, the pattern becomes more fragmented and the signal coherence is diminished for all sensor separations. The difference between the theoretical and empirical covariance matrices is also large at 2 Hz. This is most likely the result of low signal-to-noise ratio (SNR). A signal at 2 Hz would likely be highly coherent over this array aperture, but at 2 Hz the waveforms are dominated by a diffuse noise field and hence the coherence is low.

Matched field processing involves comparing the phase, coherence and amplitude relations between sensors of an array or network with those anticipated from a given signal. The idea is displayed in Figure 3. In practice, a statistic is calculated from the covariance matrix of the incoming wavefront and a prescribed steering vector (this is a quadratic form). A theoretical steering vector is constructed from a set of modeled time-delays and an empirical steering vector is the principal eigenvector of the covariance matrix evaluated for a previously observed signal. Matched field processing (MFP) using empirical steering vectors is referred to as empirical matched field processing (EMFP).

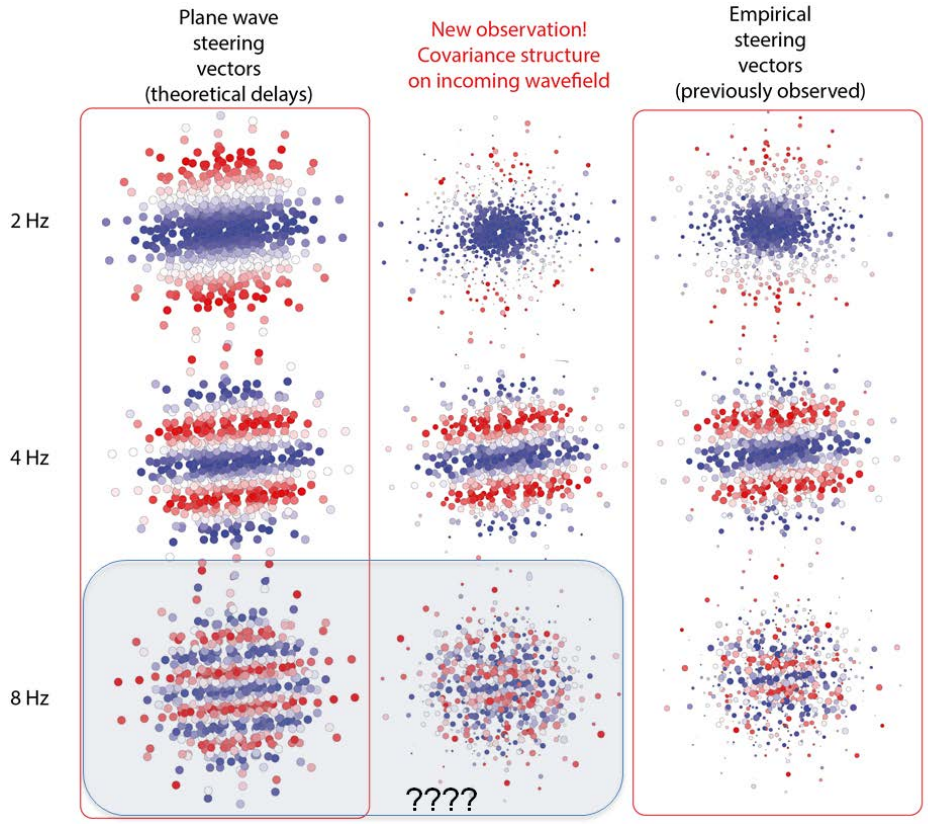

Figure 3: A graphical display of the comparison between the covariance matrix of an unknown wavefront with two possible phase and coherence patterns – one theoretical and one empirical.

EMFP was demonstrated by Harris and Kværna (2010) to provide an excellent classifier of seismic signals from very closely spaced active mines. The small differences in the paths taken from the different sites (on the Kola Peninsula in northwestern Russia) to the ARCES seismic array (at a distance of approximately 440 km) were sufficient to provide differences in the signal covariance matrices that were sufficiently characteristic that the small differences in the phase shifts observed could actually distinguish between blasts from the different mines. With only a few km between each mine, the sources are far

closer to each other than a seismic array at this distance should be able to resolve using theoretical considerations. EMFP is actually exploiting the diffraction of the seismic wavefront which degrades the performance of theoretical frequency wavenumber (or f-k) analysis.

The EMFP classification was also demonstrated to provide far superior identification of the sources than waveform correlation (Harris and Kværna, 2010). The reason for this is that the seismic signals from the mining explosions differ strongly temporally from one event to the next. The sources are ripple-fired shots with many temporally spaced charges detonating within fractions of a second of each other. This results in a very different sequence of ripples, depending upon the orientation and source time-function of each event. EMFP however breaks the signal down into narrow frequency bands such that temporal differences are mitigated; the phase shifts between each sensor of the array are still characteristic for the appropriate path.

EMFP has been shown to be an effective signal classifier, but how do we perform continuous detection? Figure 4 shows a sequence of signals from aftershocks of the 8 October 2005 Kashmir earthquake. For an effective characterization of the sequence we would like a detector which provided a trigger for every event coming from the aftershock region but not in the absence of such an event. Waveform correlation has been demonstrated by Slinkard et al. (2013) to be a poor classifier of such sequences with very few pairs of events with sufficiently similar signals. In particular, the differing in source dimensions between the main event (M 7.6) and the aftershocks (typically below magnitude 5.5) mean that the signal from the main shock is a poor representation of the Greens function for any given aftershock. However, for a seismic array at some distance from the source region, the phase shifts between sensors for incoming wavefronts from different events in the sequence are likely to show significant similarity.

Figure 4 demonstrates how EMFP for a single phase arrival can be used to characterize the aftershock sequence from a large earthquake observed at a single seismic array. A phase/amplitude template (characterized by the steering vectors, $\boldsymbol{\varepsilon}$) is obtained for the initial P-wave arrival from the main event crossing the array. A continuous comparison is made between these template phase shifts and the phase shifts inherent in the incoming wavefield measured over the array at any given time. For each time window, a statistic indicating the degree of likeness is obtained for each narrow frequency band resulting in a time-frequency display that is reminiscent of a spectrogram but where, instead of energy in the signal, we measure the goodness of fit between the instantaneous wavefield over the array and our template wavefront.

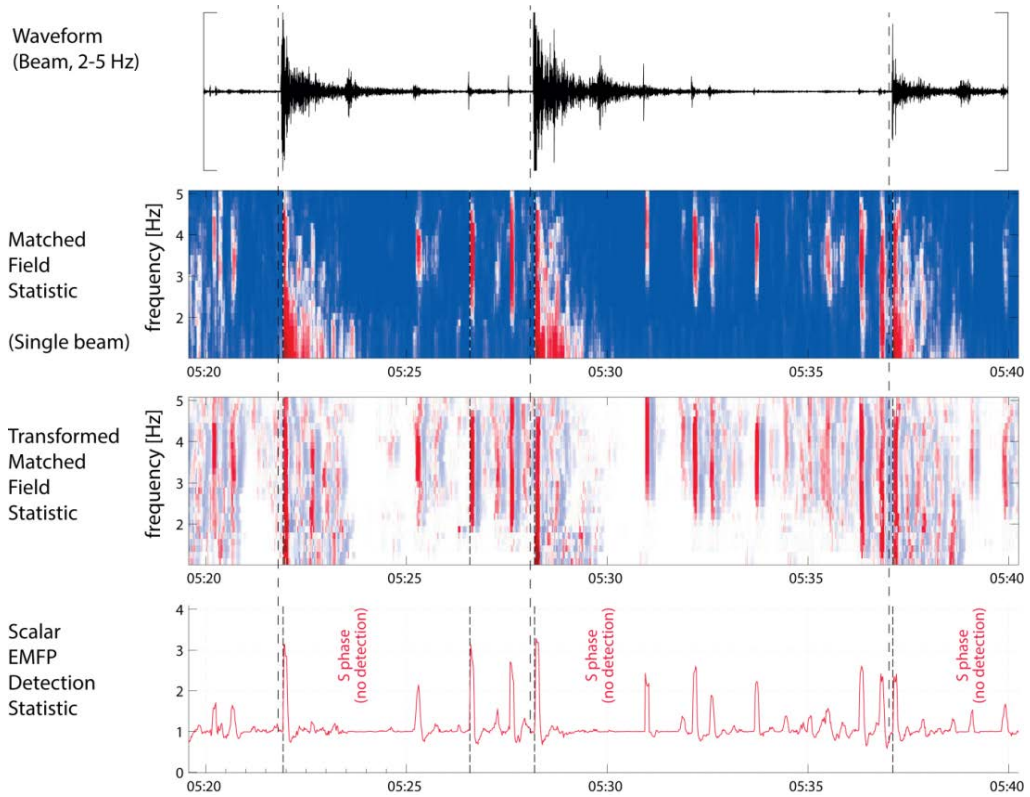

Figure 4. An illustration of the procedure to generate a detection statistic trace for a single matched field "beam" under an ongoing aftershock sequence, based on a single P-phase. ***The covariance matrix is calculated across all sensors of the array, in many narrow frequency bands, for the initial P-phase from a large earthquake, and empirical steering vectors are obtained as the principal eigenvectors of these matrices. A continuous matched field statistic trace is generated for each narrow frequency band using these steering vectors, and transformations of these single-frequency traces are converted to a scalar detection statistic trace by taking averages across selected frequency ranges. New occurrences of the template phase structure over the array appear as vertical bars in the "pseudo-spectrograms" and then as local peaks in the scalar trace. The waveforms shown are from the KKAR array in Kazakhstan on October 8, 2005. The main event is the magnitude 7.6 Kashmir earthquake at a distance of approximately 9 degrees from the station.***

The properties of this "pseudo-spectrogram" vary from arrival to arrival. For the largest aftershocks in the sequence, the likeness is strong for all frequencies and the wavefield structure persists for a long duration into the coda. For smaller events, the likeness is only significant for higher frequencies for which the SNR is high. A differencing transformation (Gibbons et al., 2008) is applied for each frequency band that replaces the matched field statistic with a derivative function which is maximized at the phase arrival time, giving the impression of vertical bars in the time-frequency plot. A transformation across frequency bands (most simply a mean value) then generates a scalar trace on which significant local maxima occur at the times of P-wave arrivals from the different aftershocks. It is important to note that unrelated energy arrivals (e.g. the S waves) do not result in detections, whereas all P-arrivals from a source region with an aperture of the order ~70 km do register a trigger.

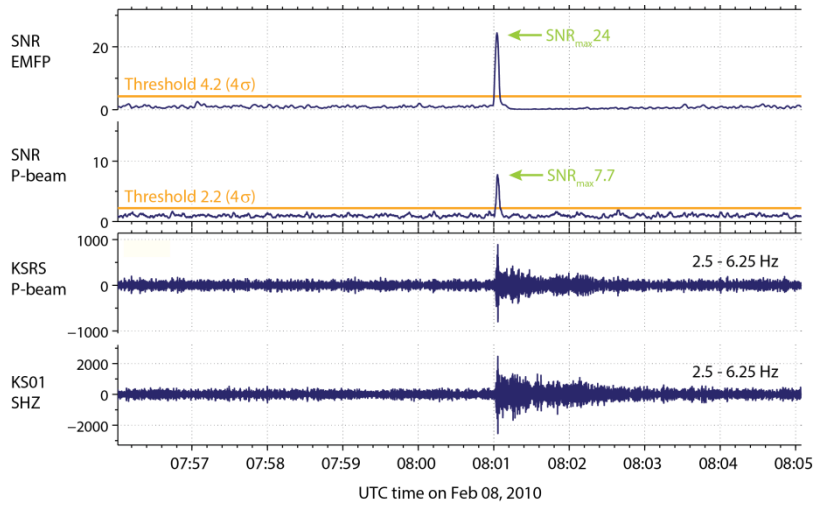

Figure 5. Matched field detection of the signal from a DPRK nuclear test I. ***The signal from the February 12, 2013, DPRK nuclear test, recorded on the KSRS array, is submerged, scaled by a factor of 0.0125, into background noise on the same channels at the time shown on February 8, 2010. A beam is formed with apparent velocity 8.5 km/s and backazimuth of 15 degrees and an SNR (STA/LTA) trace is formed. At the top is the SNR trace of the empirical MFP scalar detection statistic, generated in the manner displayed in Figure 1, using the Pn phase of the signal from the May 25, 2009, DPRK nuclear test as the matched field template. Thresholds for the different SNR traces are determined from statistics of the (log) background variability of these traces over long time-windows.***

To demonstrate that the MFP detector really does perform robustly below the detection threshold for classical (beam/STA:LTA) detectors, we submerge copies of the 2013 DPRK nuclear test signal, recorded on the KSRS array, with different scaling factors into background noise on the array and attempt to detect the initial P-phase using a matched field template obtained from the 2009 nuclear test signal. Figure 5 shows a case where the 2013 signal is submerged into background noise with a 1/80 scaling factor. We form a classical beam SNR and also an SNR trace on the output from the matched field detector as displayed in Figure 4. Thresholds for the two traces are obtained by examining the statistical variability of the traces over long time-durations. A clear detection is made for the signal on both SNR traces.

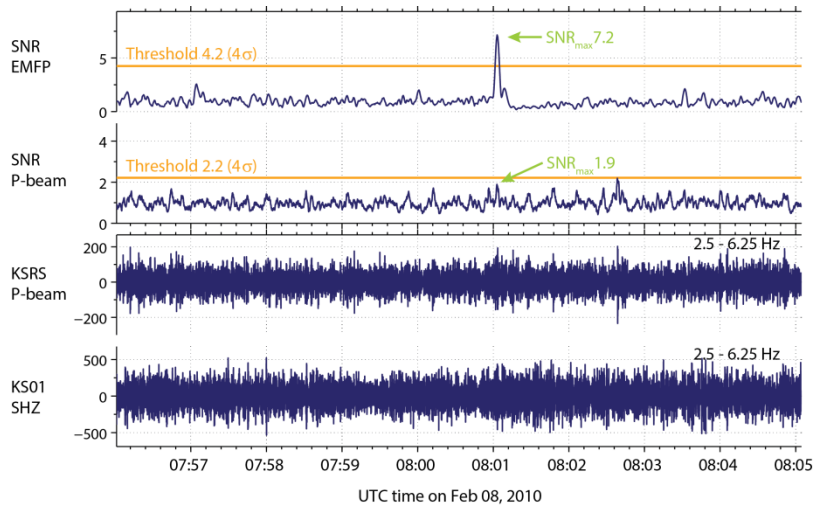

Figure 6. Matched field detection of the signal from a DPRK nuclear test II. ***Exactly the same as for Figure 5 except that the scaling factor now applied to the signal from the 2013 nuclear test is now 0.0025 (i.e. a factor of 5 lower than that displayed in Figure 5). The submerged signal is no longer observed over the background noise, even on the beam, and the SNR calculated from the beam fails to exceed the empirically determined threshold. The SNR of the empirical MFP trace does exceed the statistically determined detection threshold indicating that, despite the fact that no significant increase in signal amplitudes is observed, the spatial structure of the wavefield immediately after the signal arrival fits the matched field template significantly better than the spatial structure of the wavefield immediately before the signal arrival.***

The same is done in Figure 6, the only difference being that the scaling factor now applied is 1/400, taking the signal amplitude right down to the noise level. No detection is made on the classical beam SNR trace. The SNR trace derived from the MFP statistics still exceeds the threshold comfortably at the time of the signal onset. A detection statistic that measures the rapid increase in the similarity between the phase structure of the wavetrain and the target phase structure is significant at the time of the signal arrival, whereas a detection statistic that measures an increase in waveform amplitude is not significant at this time.

The system displayed in Figure 6 has been optimized using an empirical template. In the absence of previous observations, a narrow band theoretical matched field template can be used that may still exceed the detection capabilities of beam/SNR-based algorithms if the increase in spatial coherence of the wavefield with the specified phase shifts is more significant than the increase in waveform amplitude. Such templates will usually consist of plane-wave phase shifts, although we may have additional models for wave propagation close to the array which may predict deviations from the simplest plane-wave models. Figure 7 illustrates the detection of the 2013 DPRK nuclear test signal at KSRS using plane-wave matched field detectors and empirical matched field detectors with the templates taken from the events in 2006 (distance ~2km) and 2009 (distance < 500 m). In the lower frequency band, there is little difference in the performance between the 3 detectors. At higher frequencies, both empirical templates outperform the plane-

wave MFP detector with the template from the closer 2009 event outperforming the template from the 2006 event.

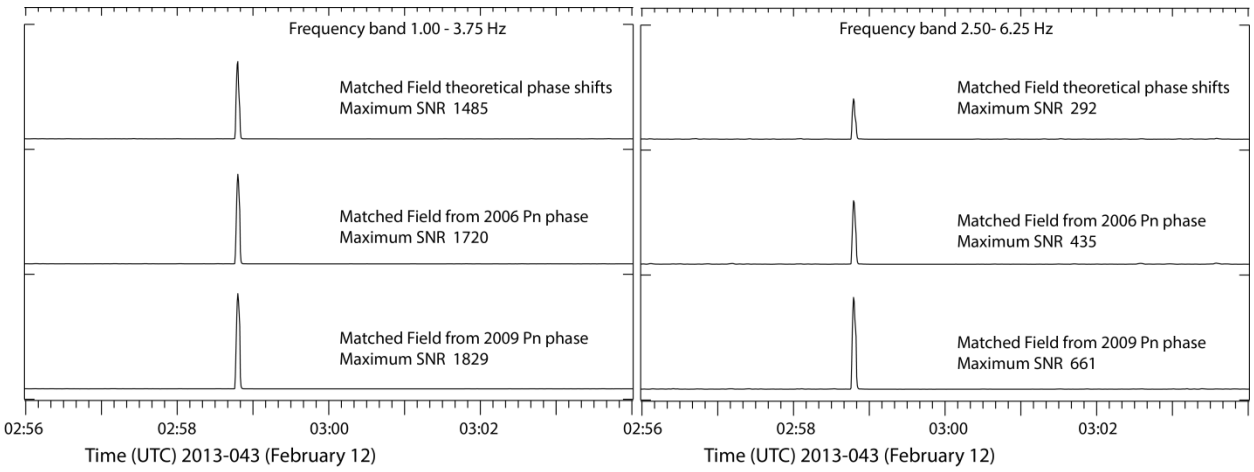

Figure 7. Detection of the signal from the 2013 DPRK nuclear test on the KSRS array using matched field templates as indicated. ***For both frequency bands, the empirical matched field templates perform better than the plane-wave template (top) with the template from the closer 2009 event signal performing better than the template from the 2006 event signal. However, the difference is more pronounced at the higher frequencies indicating that the empirical MFP detector exploits significantly more of the critical high frequency energy than the plane wave matched field detector.***

We have demonstrated that a specially targeted matched field beam will detect events in that given direction effectively. However, the purpose of the current contract is to explore the extent to which EMFP can be extended to cover the spectrum of detectors from the highly source-specific spot detectors to the general detectors for unknown signals. Which considerations are most important when designing generalized detectors using the principles of MFP? Figure 8 depicts deployments of matched field "beams" for seismic arrays covering the theoretical slowness space and augmented by source-specific, empirical, matched field detectors. A general beam deployment has been augmented by a few specially chosen site-specific matched field templates, looking for e.g. Pn, Pg, Sn, and Lg phases from regional events at a particular site. In addition, a few other templates aimed at targets at teleseismic distances are deployed. Other templates may be added for spotlight detectors on target sites, or even for repeating sites that generate signals that require safe classification for screening.

In principle, every well-classified detected phase can result in a new matched field detector. In practice, this could escalate rapidly to a deployment containing an intractable number of empirical MFP detectors. Repeating sources such as mines may be adequately represented with one or a few matched field templates; other very limited regions of slowness space may correspond to quite broad geographical regions and may need very dense coverage of matched field detectors to provide optimal detection and parameter estimation.

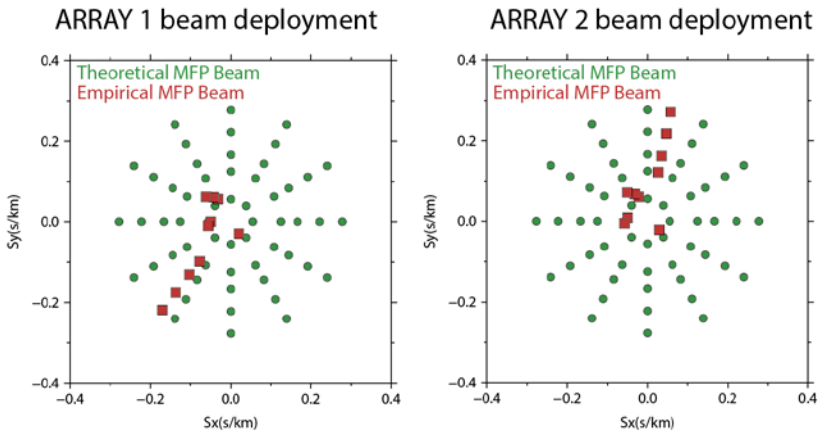

Figure 8. Schematic representation of a matched field beam deployment on two arrays both with general beams and empirical beams aimed at specific wavefronts. ***After the deployment of a new seismic array, an initial beam deployment would consist only of the (green) matched field "beams" with the theoretical phase shifts. As events are recorded and located, the (red) empirical matched field beams can be deployed in addition to detect new occurrences of previously observed phases.***

Following signal detection, an estimate must be made of parameters which describe the propagation of a wavefront, in order that the signal's origin can best be estimated. Under existing procedures, slowness space is scanned using plane-wave phase shifts which frequently results in bias and waveform misalignment (hence the need for SASCs – Slowness and Azimuth Station Corrections). For example, the top panels of Figure 9 show f-k analysis for the P-phase of an aftershock from the October 8, 2005, Kashmir earthquake recorded at KKAR, performed at 2 Hz and 4 Hz, together with the theoretical slowness vector indicated by a red arrow. The estimate at 2 Hz is close to the expectation whereas the estimate at 4 Hz gives a qualitatively different estimate (a teleseismic apparent velocity). The frequency dependency of plane-wave slowness estimates is one reason for the variability of broadband f-k estimates, which in turn requires large tolerance bounds in phase association algorithms.

A significant part of the project will investigate how EMFP can result in calibrated parameter estimates for placing the direction of signal in the classical slowness space. However, we will also investigate the potential for mapping signals directly into the geographical space of the source regions: a procedure which may change significantly the nature of standard seismic processing pipelines.

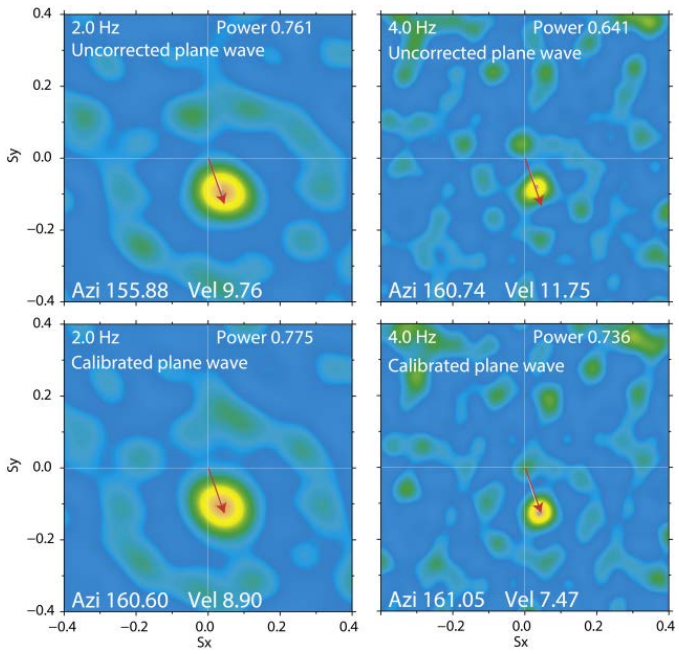

Figure 9. Demonstration of how empirically calibrated f-k analysis results in an increased array gain (in particular at higher frequencies) and parameter estimates that are more consistent between the different frequency bands. ***All of the above slowness grids are evaluated at the time of a P-arrival at the KKAR array from an aftershock of the 8 October 2005 Kashmir earthquake. The "calibrated" slowness grids are evaluated using imposed phase shifts calculated from the P-arrival from the main shock which is attributed an apparent velocity of 8 km/s and a backazimuth of 160 degrees. The aftershock is approximately 50 km to the North West of the main shock from which the template is calculated. The array is at a distance of approximately 9 degrees from the source region.***

Most of the calculations performed in this project will be at frequencies above 1 Hz since this is the band in which energy is primarily detected for body waves from low magnitude seismic events of monitoring interest. However, we wish also to examine the physical basis for the observed variability in matched field templates. Which factors of the propagation medium are most significant in determining the form of the received signal? We therefore wish to study variability of matched field templates in a frequency band that allows comparison with results obtained through model calculations. With anything but the largest supercomputers, the calculation of synthetics requires that the comparison be carried out at long periods. Consequently, we intend to investigate the variation of matched field templates for long-period surface waves.

3. TECHNICAL APPROACH

We seek to understand how combinations of theoretical and empirical templates will cover the parameter space that describes the set of incoming wavefronts, anticipated at a given station, that are of interest for monitoring purposes. This is to say that we want to be able to detect all signals of interest and classify them correctly, using a Matched Field method of the type described by Harris and Kværna (2010). How do we implement this in operational systems? An interesting test-case here is seismicity in the European Arctic detected and classified using the SPITS array. The case is interesting because this small aperture array (diameter approximately 1 km) is critical for monitoring needs (due in particular to its proximity to Novaya Zemlya) and produces slowness and backazimuth estimates that require calibration (see Gibbons et al., 2011).

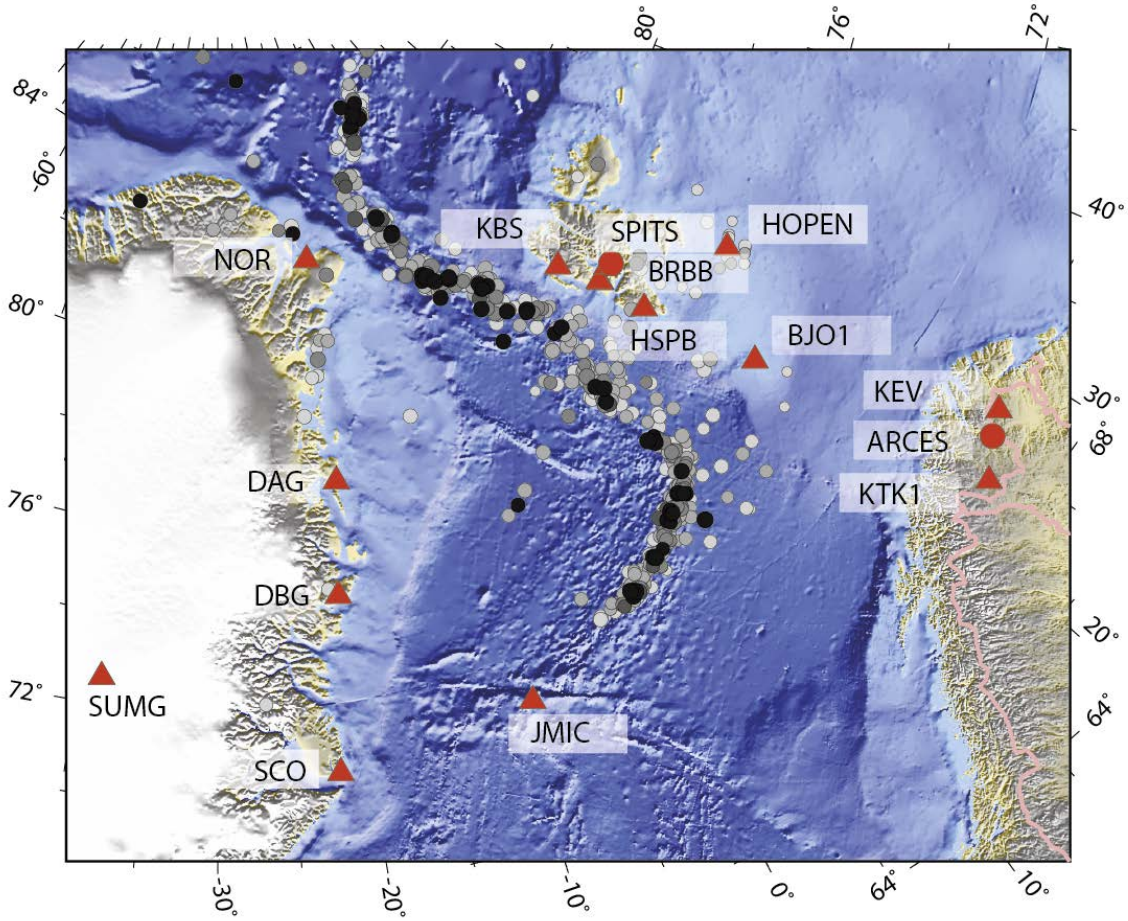

Figure 10. The northern mid-Atlantic ridge. ***Events north of 72 degrees from Engen et al. (2003) together with the stations in the current network which are most sensitive to low-magnitude events on the spreading ridge. The events are marked with circles with size proportional to magnitude and with the shade a function of the number of associated phase readings (essentially all teleseismic). The palest symbols have as few as 10 defining phases and the darkest have over 100.***

Figure 10 displays the location estimates for seismicity between 1955 and 1999 north of 72 degrees as compiled by Engen et al. (2003). Since this time, the highly sensitive

SPITS array and additional 3-component stations on Greenland and in the European Arctic have resulted in a significant improvement in the detection and location capability for events in the region. However, most of the available event location estimates (e.g. the routine solutions calculated by NORSAR, and those in the Reviewed Event Bulletin – REB - of the International Data Center in Vienna) use backazimuth measured at SPITS as a location constraint. This disqualifies such solutions from being used to calibrate parameter estimation using the SPITS array.

It was necessary to obtain a set of new and reliable event location estimates that were independent of the parameter estimates obtained from SPITS. The Bayesian hierarchical multiple event location algorithm Bayesloc (https://www-gs.llnl.gov/about/nuclear-threat-reduction/nuclear-explosion-monitoring/bayesloc: see Myers et al., 2007, Myers et al., 2009) was chosen for this purpose. The seismicity covers a range of magnitudes with many significant events with magnitudes greater than 5 that are recorded well teleseismically, in addition to many events of magnitude 3 and below which are recorded only by a subset of the stations displayed in Figure 10. The larger events which are very well constrained using teleseismic phases have far smaller location uncertainties. These uncertainties are reflected in the event priors and this is likely to provide additional constraints on the "regional only" events when they are relocated in the multiple event procedure.

Over 15 years of continuous data was obtained for the stations on the Greenland side of the ridge. Very few of the arrivals on stations such as NOR and DBG have been used in location estimates found in, for example, the bulletin of the International Seismological Center, and it is only very recently that these stations have been used in the routine NORSAR event location estimates.

The preliminary set of relocations using Bayesloc is displayed in Figure 11. These solutions use regional P and S arrivals (largely Pn and Sn) from the stations displayed in Figure 10, in addition to teleseismic P readings taken from the Reviewed Event Bulletin. Note that the initial event location estimates were provided by the HYPOSAT program (Schweitzer, 2001a). The HYPOSAT program is useful in many respects as a single event algorithm, providing very robust preliminary location estimates. Phases with anomalously high time-residuals, for example, are excluded from the inversion such that the arrival input to Bayesloc is already screened for obvious outliers. HYPOSAT in addition provides an initial phase label for each arrival (Pg, Pb, Pn, etc.) even if we only specify "P1" (for first arriving P-phase) – and, once in Bayesloc, the probabilistic phase labelling approach of Myers et al. (2009) is applied.

To give an impression of the improvement that the Bayesloc solutions provide over previously available event bulletins, we compare the solutions displayed in Figure 11 with three available bulletins in Figure 12 a), b), and c) – and with the input solutions from the HYPOSAT algorithm in Figure 12 d). Only panel (d) of Figure 12 shows exactly the same events as displayed in Figure 11. The events displayed in panels a), b), and c) have significant overlap with the events displayed in Figure 11. The ISC solutions for example are only available two or so years before the present. The REB solutions are

limited to events recorded by more than one International Monitoring System (IMS) station; there are many events displayed in Figure 12 for which SPITS is the only IMS station that recorded the event well enough for inclusion.

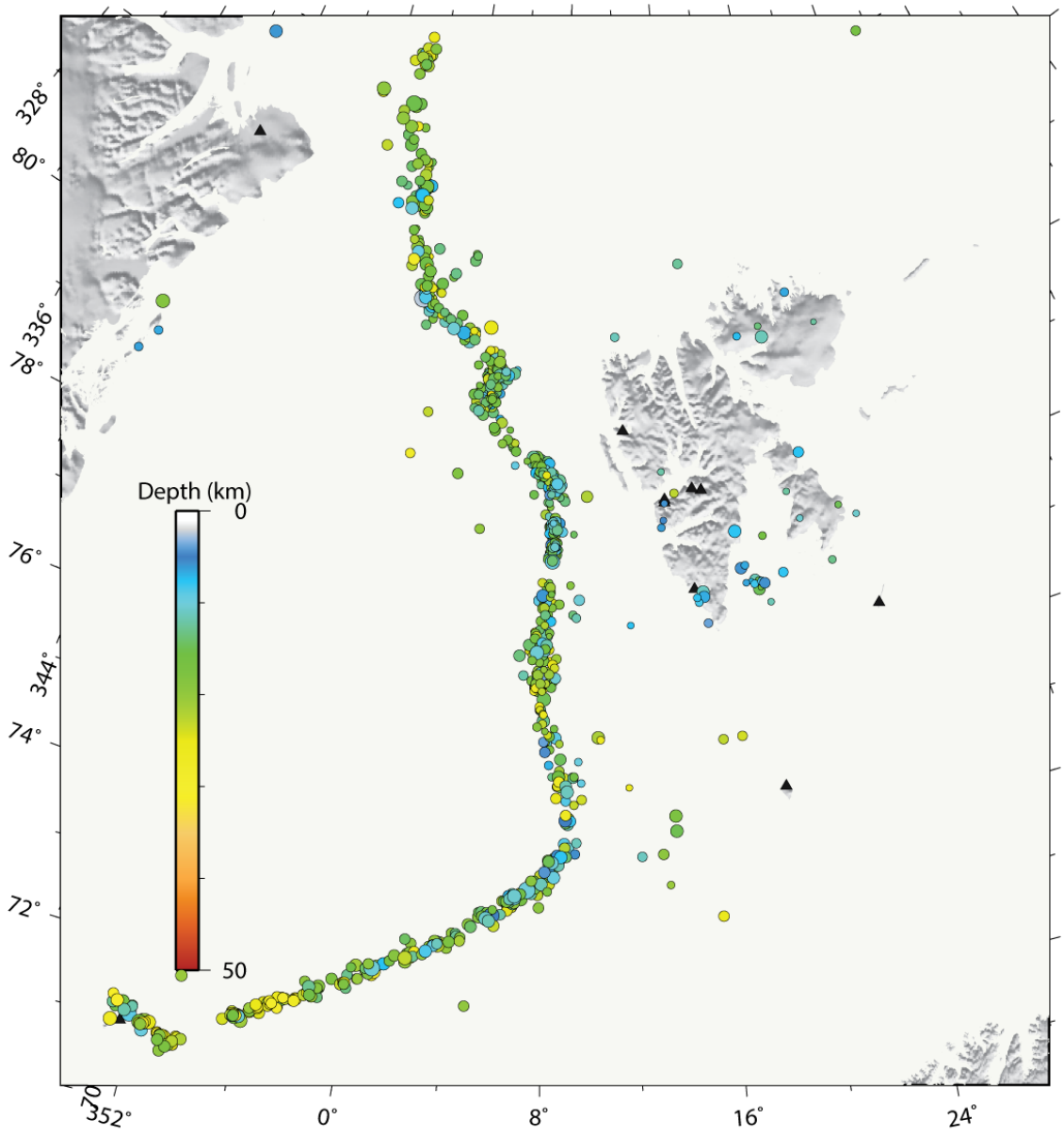

Figure 11. Location estimates for 971 events using the Bayesloc program and both regional and teleseismic phases.

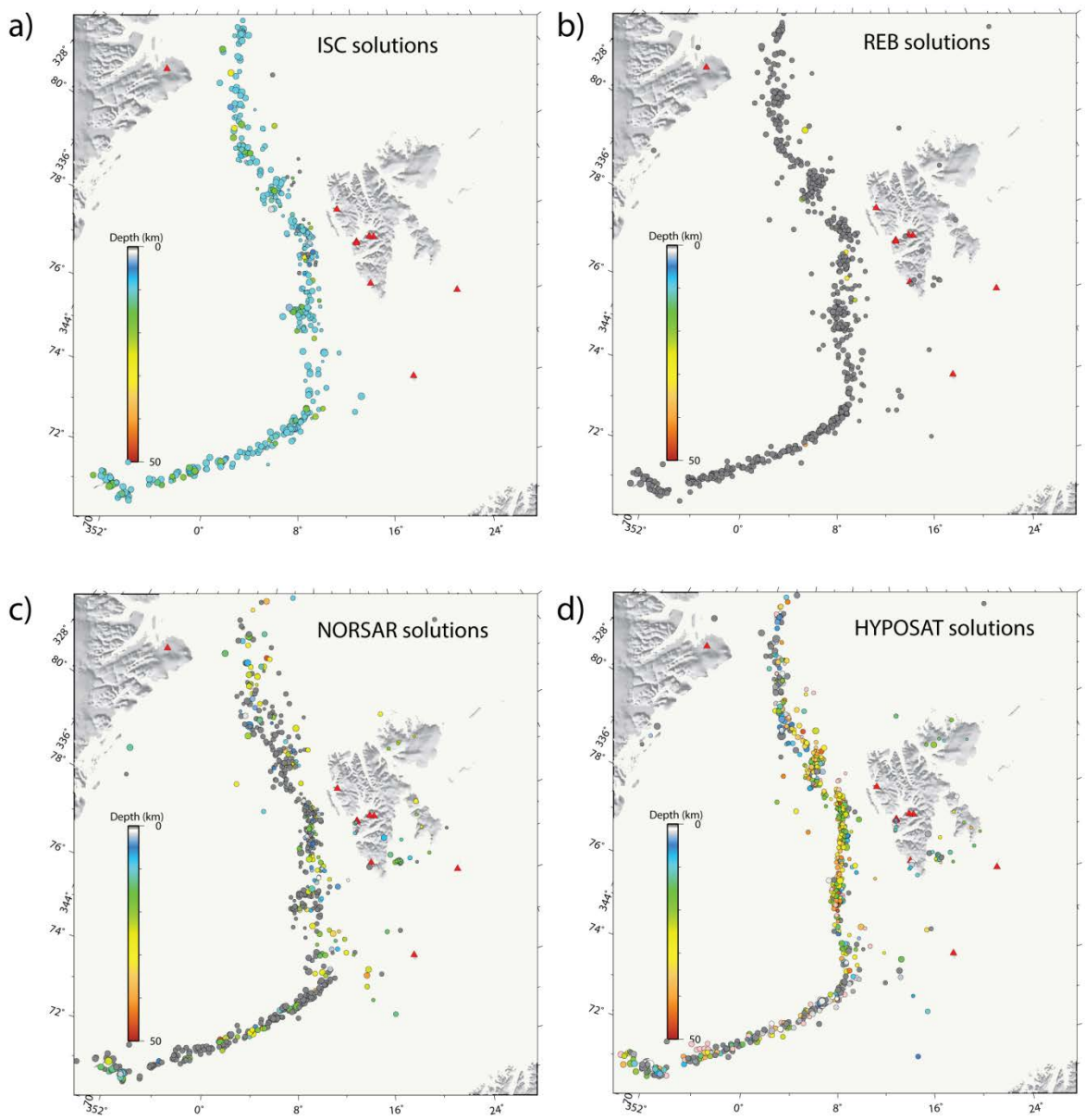

Figure 12. Comparison of event location estimates. ***(a) Events in the period 1998-2011 from the bulletin of the International Seismological Center (www.isc.ac.uk), (b) Events in the period 2000-2015 from the Reviewed Event Bulletin of the International Data Center for the Comprehensive Nuclear-Test-Ban Treaty, (c) Events in the period 1998-2015 from the routine event location carried out at NORSAR (http://www.norsardata.no/NDC/bulletins/regional/), and (d) the events displayed in Figure 11, located one by one using the HYPOSAT program.***

A useful insight into the properties of the Bayesloc solutions is provided by examining the residuals associated with one given phase for each event (Figure 13). For a single 1-d velocity model, we display the time residual for the first regional P-arrival at SPITS for the multiple event location estimates (Figure 13a) and for the single event locations (Figure 13b). The residual for the multiple event locations varies exceptionally smoothly as a function of source location, whereas it varies significantly with location for the single event location estimates. Like most single event location algorithms, HYPOSAT

seeks a solution which minimizes a measure of the total time residual vector. Bayesloc, given the additional constraints provided by the better observed events, favors those solutions with residuals which are consistent with the residuals of neighboring events. The procedure was repeated for several different velocity models and the Bayesloc location estimates were found to vary significantly less than the single event algorithm location estimates; only the residual distributions looked different. The significance of this is that, given a continuous band of seismicity containing many events with exceptionally good teleseismic location constraints, Bayesloc can provide quite robust event location estimates that are relatively insensitive to the exact structure of the applied regional velocity model. The spatial statistics of the regional phase residuals provides implicit correction terms for the applied velocity model.

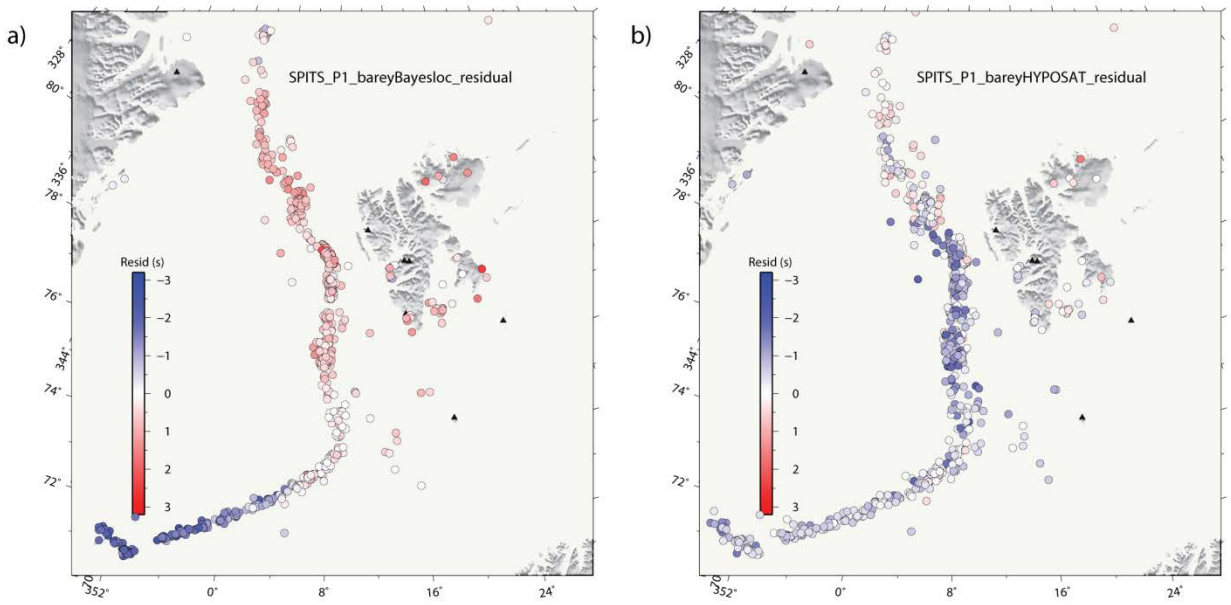

Figure 13. Consistency of time-residuals for Bayesloc origin estimates. ***Time residual in seconds for the initial P-arrival at SPITS relative to the BAREY velocity model (see Hicks et al., 2004) for (a) the mean solutions provided by Bayesloc and (b) the preliminary single event locations provided by HYPOSAT.***

Finally, in Figure 14, we demonstrate how insensitive to the applied regional velocity model the Bayesloc solutions are. The different velocity models for the crust and uppermost mantle result in different regionalized correction terms but the hypocenters are affected little.

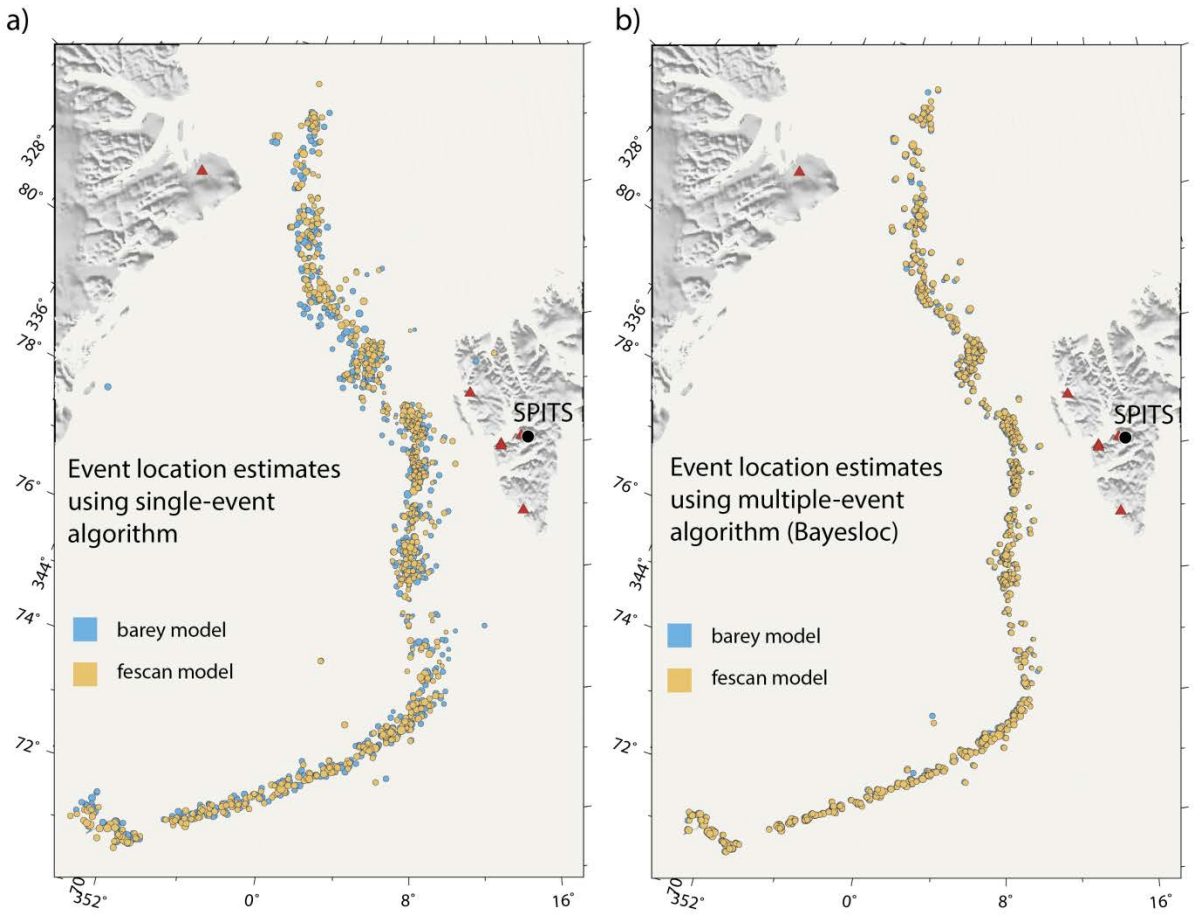

Figure 14: Robustness of Bayesloc event locations to choice of regional velocity model. ***Event location estimates for earthquakes on the mid-Atlantic ridge between 71 and 84 degrees North using (a) individually located events and (b) the Bayesloc routine. Two different velocity models, fescan (Mykkeltveit and Ringdal, 1981) and barey (Hicks et al., 2004) have been used. The sets of events displayed are identical for both panels as are the sets of arrival times used.***

Given our set of Bayesloc solutions – not explicitly constrained by slowness parameter estimates measured on the SPITS array – we can examine how the array backazimuth estimate varies with the geographical backazimuth (Figure 15). As indicated by Gibbons et al. (2011), there is an enormous variation of the apparent velocity indicated for different source regions. For events to the South of the SPITS array, regional P-phases can generate apparent velocity estimates indicative of regional S-phases. For events to the north of SPITS, regional P-phases generate apparent velocity estimates indicative of teleseismic P-phases. In addition, and of possibly greater consequence for event location using slowness estimates as a constraint, the apparent backazimuth deviates from the true backazimuth in a significant and almost sinusoidal manner. Existing correction procedures involve mapping a slowness vector using a look-up table from a set of observations to the likely real-world velocity and azimuth (e.g. Bondar et al., 1999,

Schweitzer, 2001b). In Figure 16 we display how we can perform this correction empirically using matched field processing at an earlier stage.

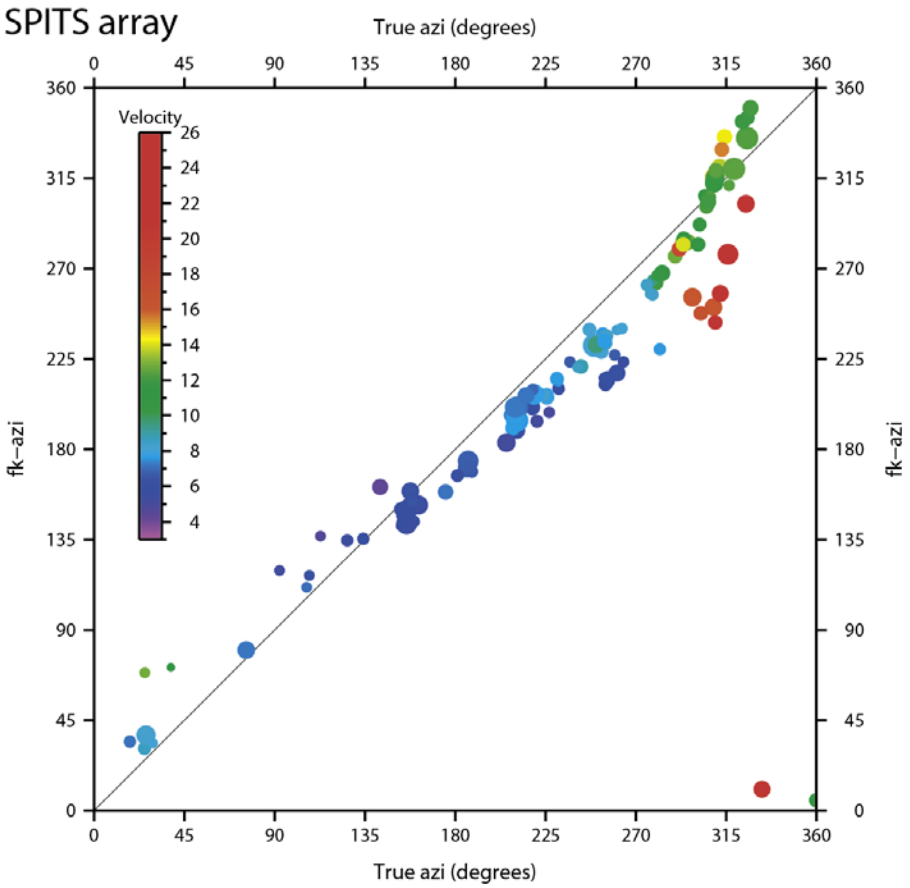

Figure 15. Deviation in backazimuth estimates using classical plane-wave models at SPITS. ***Variation of the parameter estimates obtained using broadband f-k analysis for the first P-arrival at SPITS as a function of geographical backazimuth obtained using a subset of the Bayesloc locations displayed in Figure 11.***

Figure 16 displays how – using the Bayesloc event location estimates – we are able to obtain an empirical parameter estimate using matched field processing in the context of previously observed events. We know from Figure 15 (and previous studies) that a backazimuth correction is likely to be needed and we demonstrate here how matched field calculations can provide an improved parameter estimate which is implicitly calibrated from previous observations. We demonstrate also in Figure 21 how the EMFP estimate maps the direction of the seismic signal back into geographical parameter space and not just slowness space.

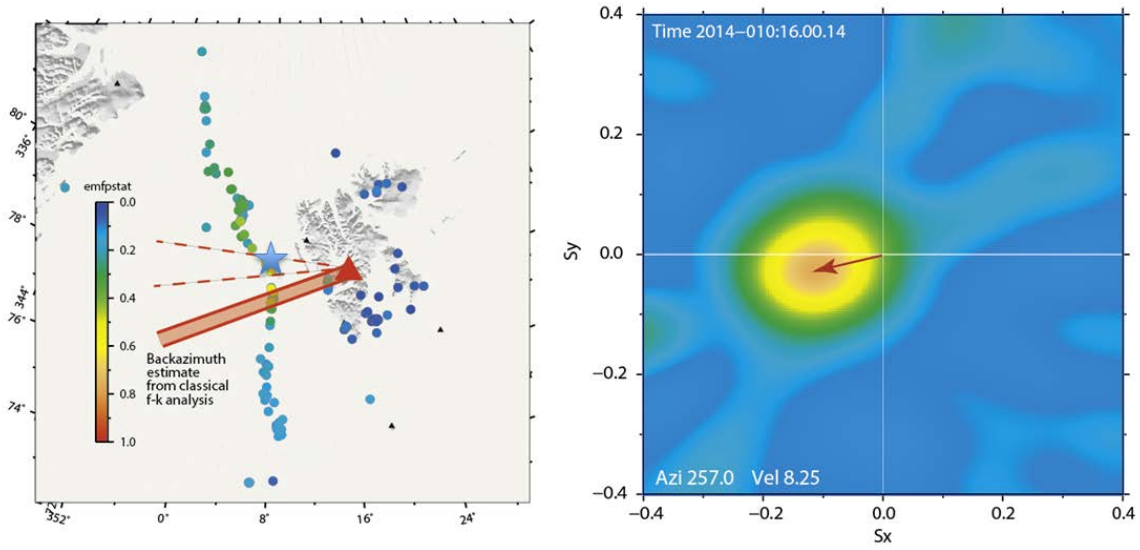

Figure 16. Display of potential utility of empirical matched field processing for providing an empirically corrected direction estimate for a detected phase arrival at SPITS. ***In the right hand panel we display the slowness plot obtained using conventional and uncalibrated f-k analysis for an arrival on January 10, 2014. In the left hand panel, we see the geographic implication of this slowness estimate for the event location, but also the matched field statistic (see Harris and Kværna, 2010) calculated between the detected phase and the empirical steering vectors calibrated from the P-arrivals from the events displayed. The blue star indicates the independent location estimate for the detected event – not used for calibration purposes.***

The European Arctic study is likely to indicate the performance of EMFP on small aperture seismic arrays optimized for processing high-frequency regional seismic arrivals. To extend the study to arguably more challenging arrays, we consider also a part of the world which is both of programmatic relevance and monitored by several seismic arrays with very different properties. The region in question is that within regional distances of the MJAR, KSRS, and USRK arrays of the International Monitoring System (IMS) and this includes the Korean peninsula, Japan, and regions of western Russia and China (Figure 17).

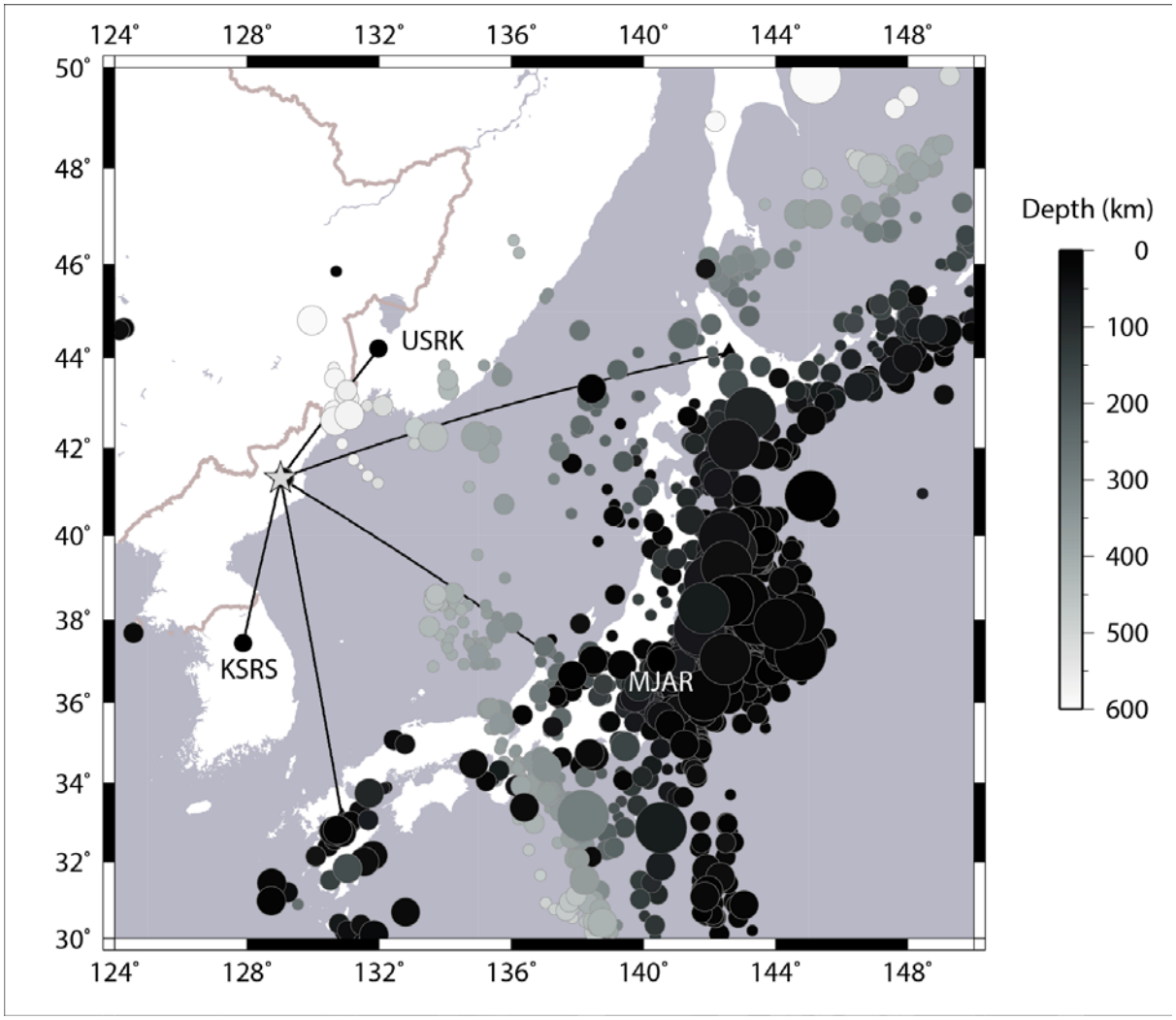

Figure 17. Study region in the Far East for the USRK, KSRS and MJAR seismic arrays. ***Event locations were obtained using Bayesloc (Myers et al., 2007, 2009) based upon events in the Reviewed Event Bulletin (REB) of the International Data Center (IDC) of the Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO). The criterion for events (between 2008 and 2016) was that an SNR > 10 was recorded for the P phases on all of the three arrays displayed. The star indicates the location of the North Korean nuclear test site.***

A vast amount of natural seismicity is recorded in this region, in addition to significant industrial seismicity and the North Korean nuclear test site. The geometries of these three arrays are displayed in Figure 18. From the geometry alone, and from coherence analysis of real world signals (Gibbons and Ringdal, 2012), we anticipate significantly different performance from the matched field detectors.

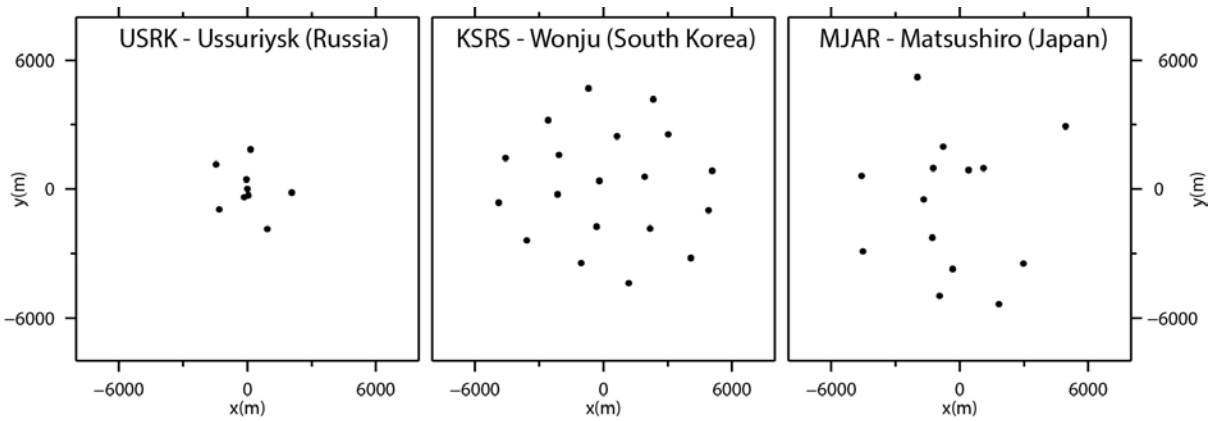

Figure 18. Geometries of the USRK, KSRS, and MJAR seismic arrays. ***In addition to the differences resulting from the array geometry, there is a significant difference between the degree of waveform similarity observed between sensors of comparable spacing. The KSRS and MJAR arrays have a similar aperture but, due to geologic and topographic factors, wavefronts crossing the MJAR array are far less coherent than those crossing KSRS. Modified from Gibbons and Ringdal (2012).***

Our investigation of the physical basis for variability in the matched field templates takes us to lower frequencies than are well-resolved on the arrays in the European Arctic and in Far East Asia. The Warramunga array (WRA) in Australia's Northern Territory has a larger aperture than any of the arrays displayed in Figure 18 and observes mid-size to large earthquakes over a wide range of backazimuth. The arc defined by subduction zones from Sumatra through and beyond Indonesia, north of Australia, is highly productive of earthquakes that reliably produce observable surface waves. A search for shallow (depth < 50 km) events in this arc using the NEIC catalog turned up 606 events in the magnitude range 5.5 – 6.5 during the time interval January 2005 – July 2016. The event locations are indicated in Figure 19. Waveform recordings for an event occurring in 2005 on day 155 are shown in Figure 20; clear surface waves are visible.

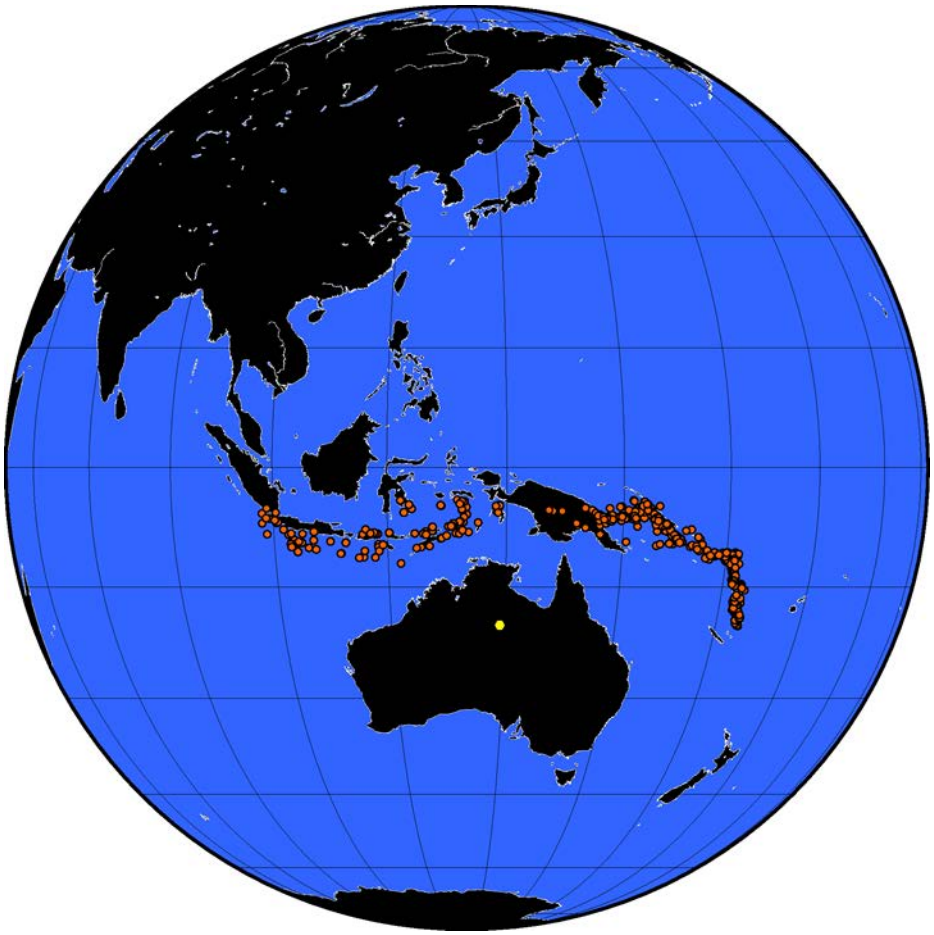

Figure 19. Locations of 606 NEIC catalog events in the magnitude range 5.5 - 6.5 during the interval from January, 2005 through July 2016. ***The event locations are shown with orange circles and the location of the observing array at Warramunga is shown as a yellow hexagon.***

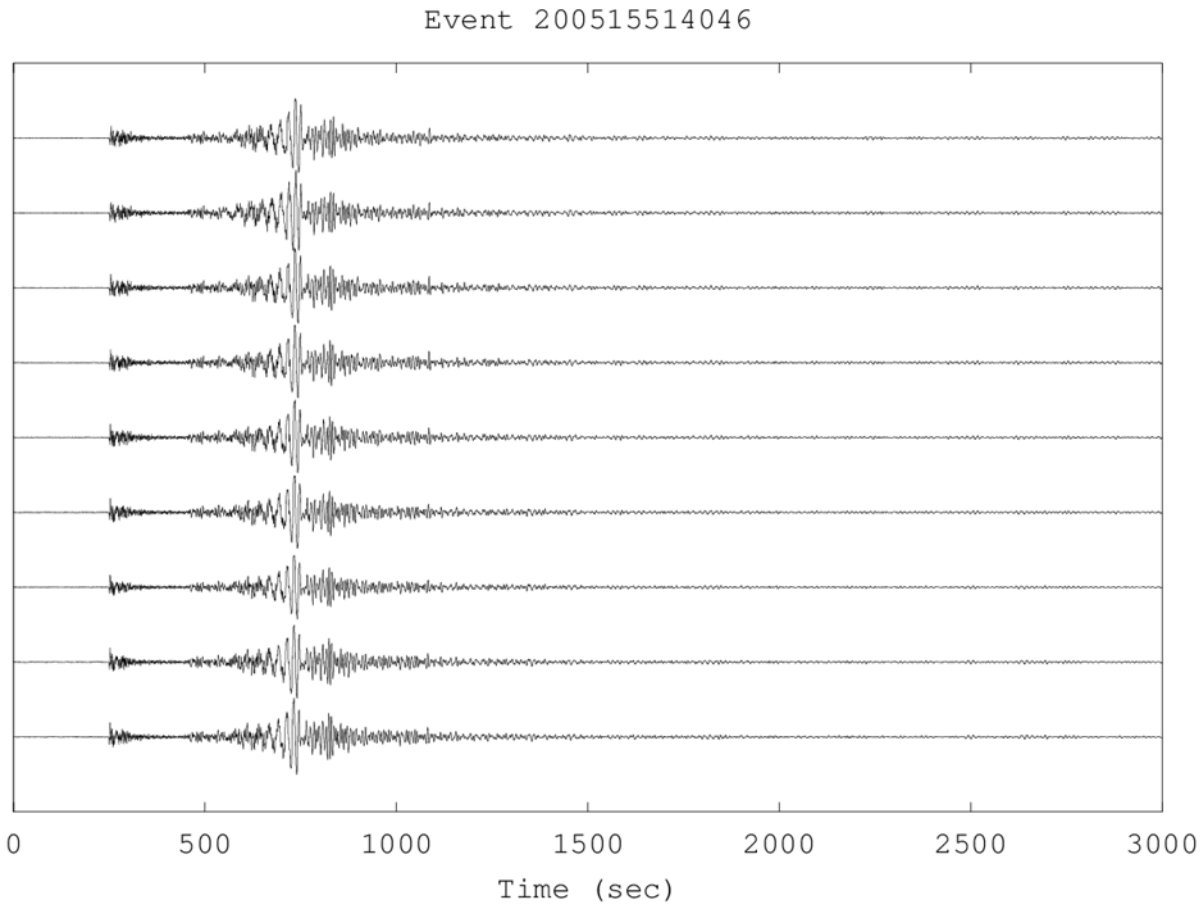

Figure 20. Waveform data for an event in 2005 on day 155 (B line of the Warramunga array only).
Surface waves are well recorded at Warramunga for the events shown in Figure 19.

4. RESULTS AND DISCUSSION

The question of the number of matched field beams required to cover seismicity in a given geographical region can be addressed by calculating EMFP statistics for the covariance matrix evaluated for a given arrival against empirical steering vectors calculated for vast numbers of seismic events covering the region in question.

Figure 21 displays matched field statistics evaluated for a given arrival at SPITS as a function of the location of the event used to generate the empirical steering vectors. The black line in Figure 21 indicates the backazimuth measured for the arrival using classical array processing for the SPITS Pn arrival and the red line indicates the direction towards the geographical locations of those events whose empirical steering vectors generated the highest matched field statistics. Given the Bayesloc location estimate of the event in question, we see that the empirical backazimuth estimate provides a far more accurate estimate than the plane-wave estimate generated from the theoretical steering vectors. We note however that, in the direction indicated, there are two "crossings" of the mid-Atlantic ridge – one at a distance of about 300 km and one at a distance of about 900 km. The matched field statistics obtained from the single Pn phase, in this direction, are unable to differentiate between the two regions. This is to say that, without additional information, we are unable to map this arrival directly in geographical space.

Figure 22 shows the equivalent plot for the Pn phase at the ARCES seismic array from the same event evaluated against empirical steering vectors calculated for the initial ARCES Pn arrivals for the same set of events. Again, we land far closer to the correct direction than the backazimuth estimated from classical f-k analysis at the array. What is demonstrated in both of these figures is that we avoid the bias in backazimuth obtained when using classical f-k analysis with a plane wavefront assumption (the black line) and that the empirical matched field statistic varies quite smoothly as a function of geographical source location. However, in this plot, we have evaluated over 1000 matched field statistics, many of which are almost identical. We wish to use EMFP as a detection algorithm and so we need to devise a manageable "beam deployment" which covers the parameter space adequately but without excessive duplication.

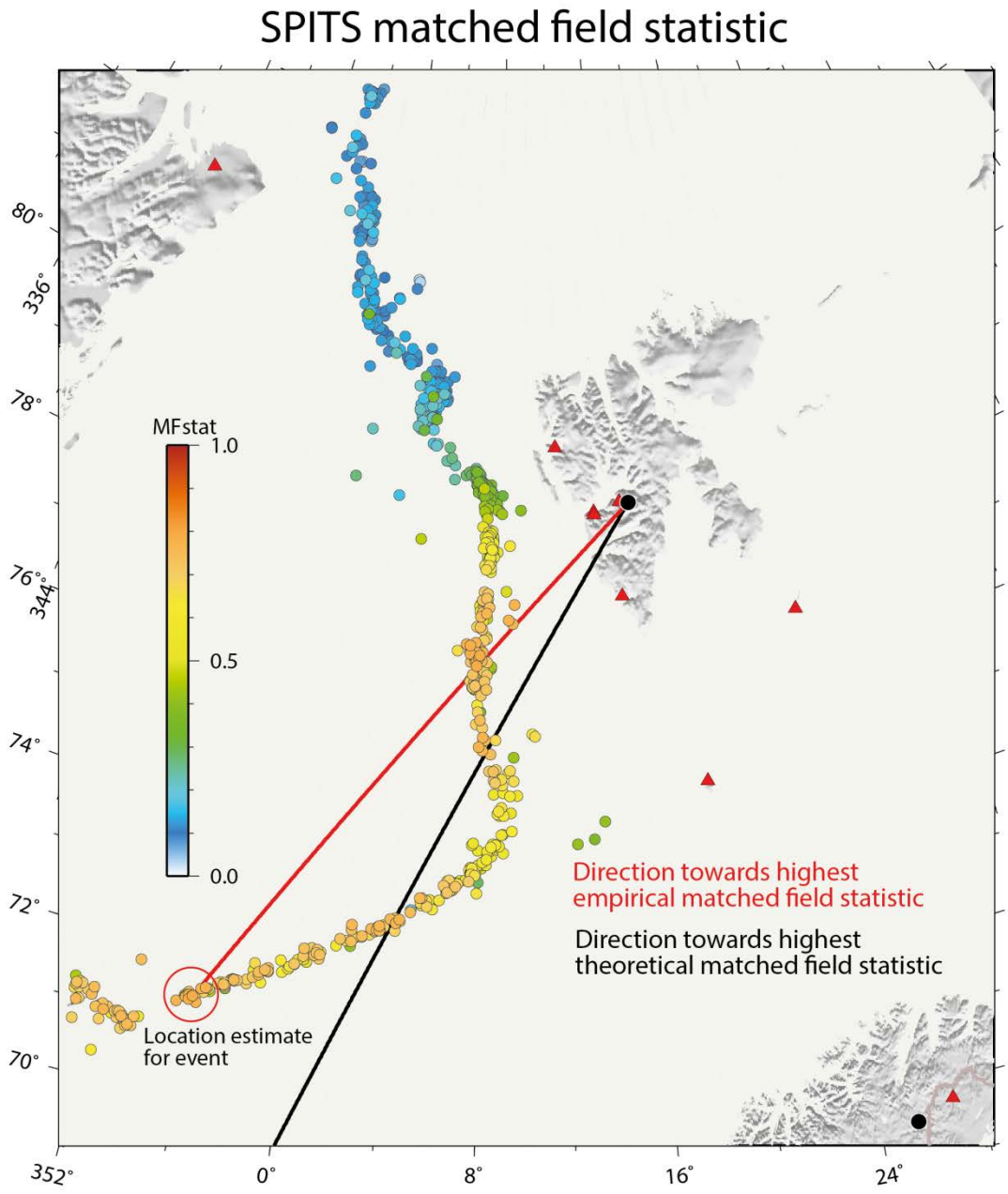

Figure 21: Geographical matched field direction estimates on the SPITS array. ***Matched field statistic evaluated on the SPITS array at a time 2015-276:01.48.27.425 as a function of the location of the template events together with the best estimate of the true location of the event at 2015-276:01.46.24. The backazimuth as indicated by the best theoretical (plane wave) matched field statistic is displayed.***

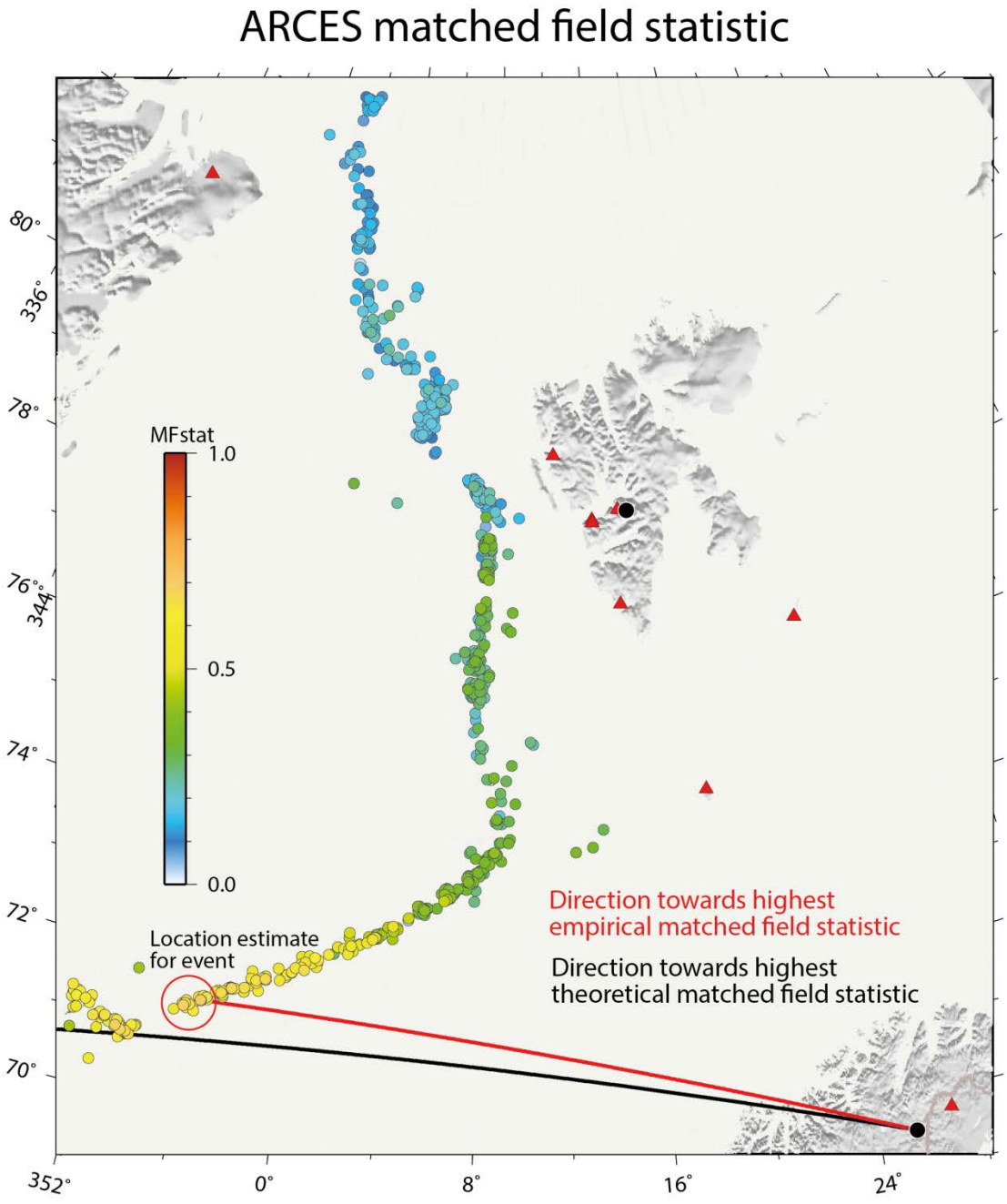

Figure 22: Geographical matched field direction estimates on the ARCES array. ***Matched field statistic evaluated on the ARCES array at a time 2015-276:01.48.44 as a function of the location of the template events together with the best estimate of the true location of the event at 2015-276:01.46.24. The backazimuth as indicated by the best theoretical (plane wave) matched field statistic is displayed.***

We wish to divide the seismicity into zones over which we can calculate “ensemble covariance matrices” (see Harris and Kværna, 2010). This is to say that we want to find the most representative and stable estimates possible of the phase and amplitude relationships between the narrowband signals on the different sensors of the SPITS array for events in a given source region. A covariance estimate taken from a single event may be biased due to low signal-to-noise ratio or, for example, a timing anomaly. If we can take the covariance matrices for many very closely spaced events and perform a careful averaging procedure for the phase-differences between each pair of sensors, we will hopefully be able to avoid any corruption of the ensemble covariance matrix for that region resulting from a data anomaly.

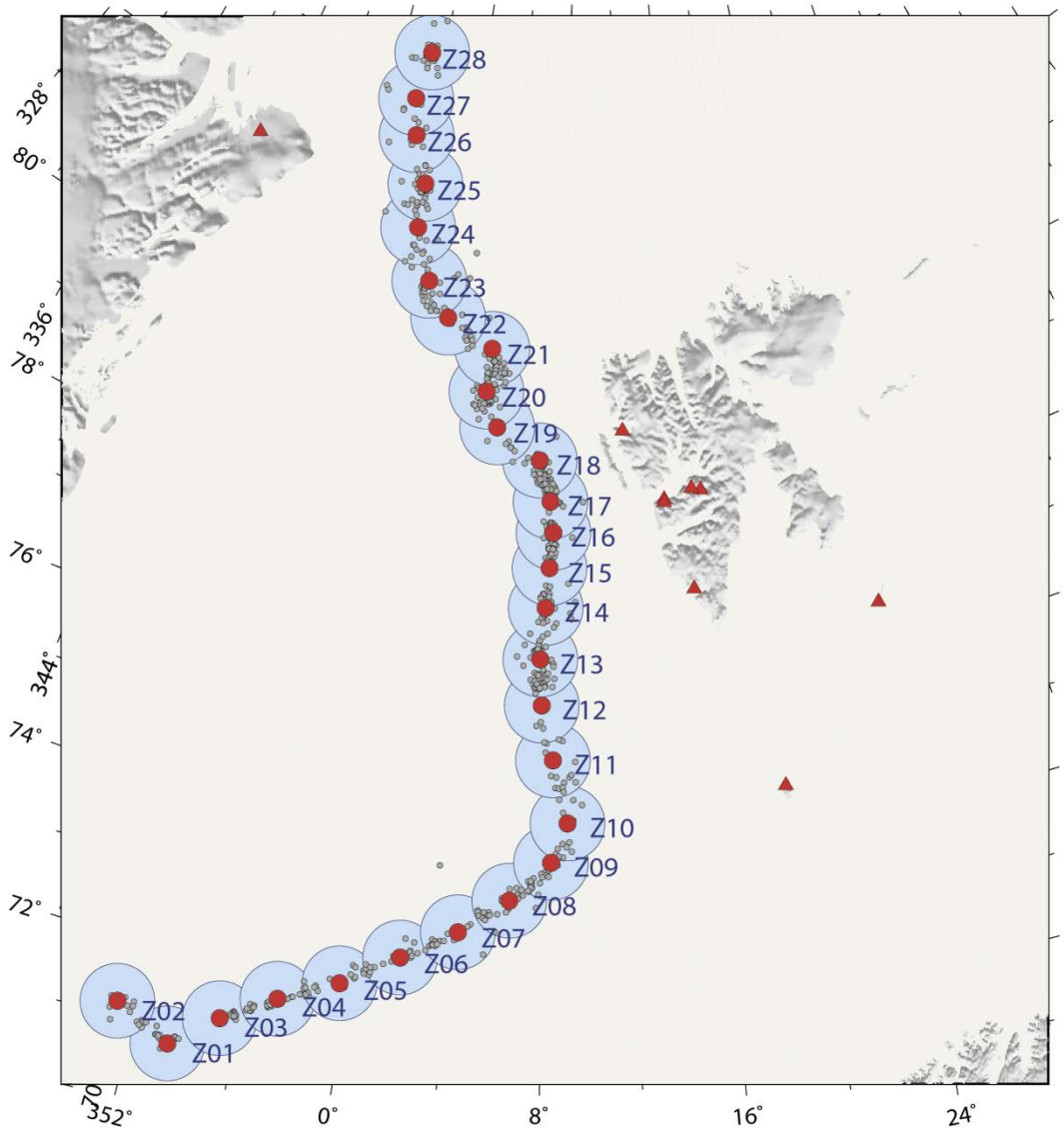

Figure 23: Provisional zone map for events along the ridge between 71 and 84 degrees north. ***The zones have radius of 70 km and overlap.***

Figure 23 displays a trial zonation of the seismicity along the mid-Atlantic ridge. The spacing was chosen based upon considerations of geographical distance between events.

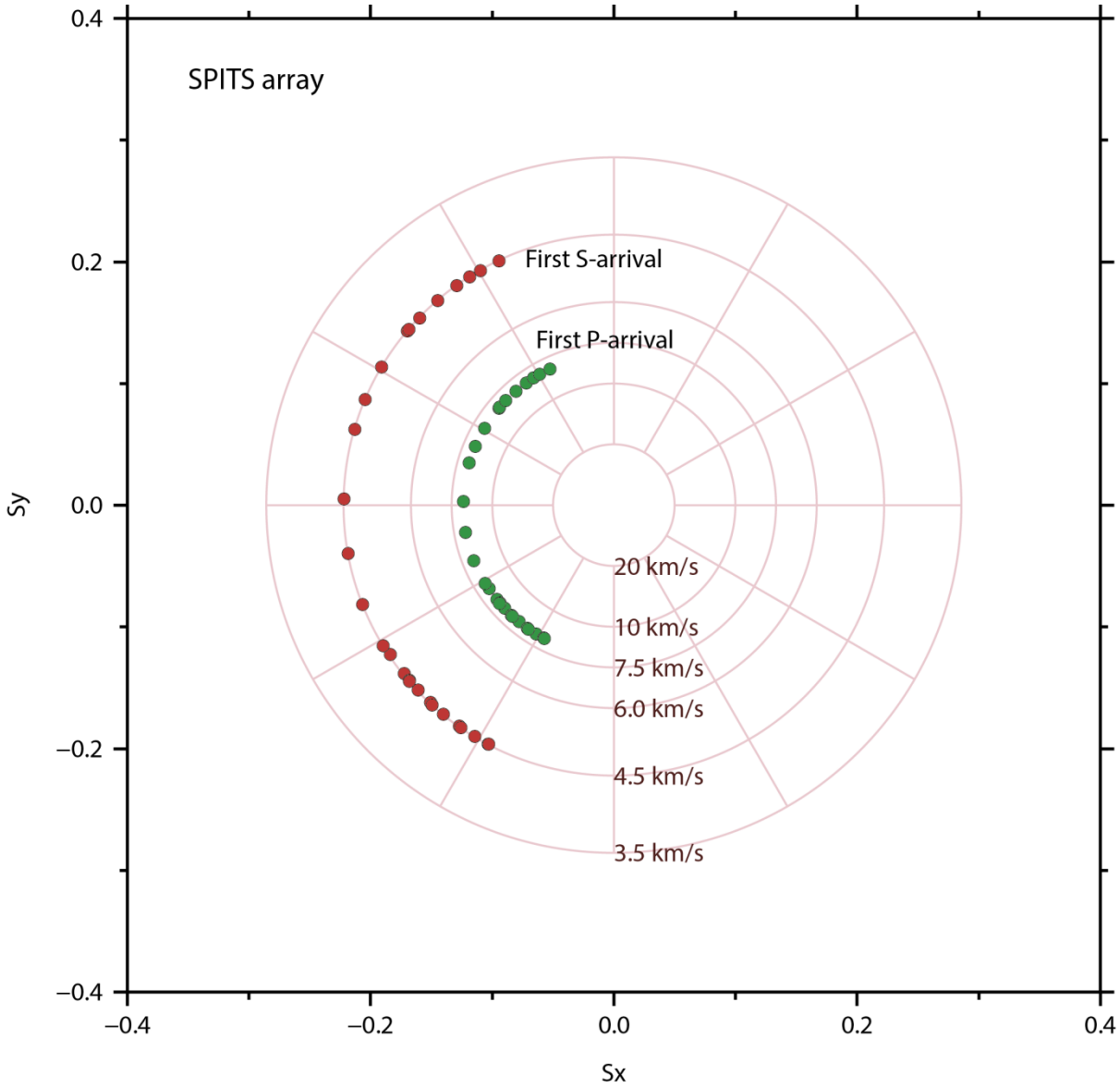

Figure 24: Beam deployment in slowness space. ***For each of the zones displayed in Figure 23, the theoretical slowness of the first P and the first S arrival is calculated for the central point.***

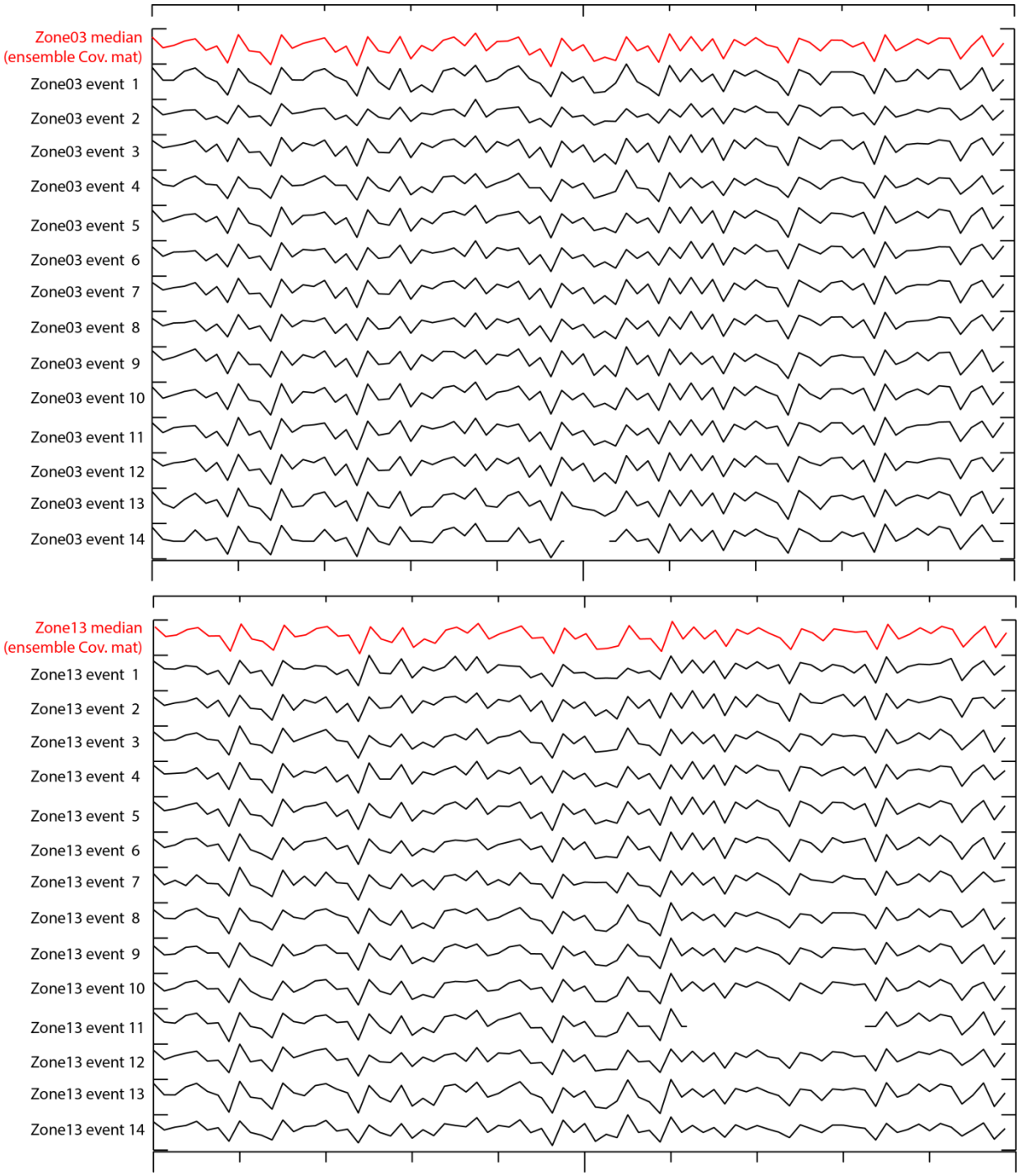

Figure 25: The construction of ensemble covariance matrices for first P-arrivals at SPITS for events from two of the clusters, obtained by calculating median values of each real and each imaginary value over all events within one of the given zones. ***Zones Z03 and Z13 are displayed which have a very similar backazimuth from the SPITS array. The gaps are zeros resulting from an absence of data from specific channels at a given time and have little effect on the median ensemble averages.***

One of the questions that we wish to address is the extent to which we should organize our signal classification by the source region, as opposed to the classical slowness and backazimuth parameter space which is usually used to classify incoming arrivals on

seismic arrays. The study designed here should be able to test whether or not there is an advantage in processing primarily using empirical templates classified in geographical space, and then returning the corresponding theoretical slowness and backazimuth parameters (if needed) to the phase association algorithm. For example, Zones Z03 and Z13 in Figure 23 have essentially the same backazimuth and slowness, despite being in two very distinct geographical regions. Scattering of the wavefield along the path may result in quite distinct patterns in the phase relationships which allow for a resolution greater than the theoretical capability of the array (c.f. Harris and Kværna, 2010).

The covariance matrices themselves are calculated by taking a short time-window of the array data surrounding the initial P-arrival and calculating the coherence and phase shift between each pair of channels using the multitaper coherence routines provided by Prieto et al. (2009). This provides, for an N-channel datastream, for every one of the NB narrow frequency bands, N*(N-1)/2 complex phase shifts. So while each covariance matrix consists of NB*N*(N-1)/2 complex numbers, we display in Figure 25 a short segment of the corresponding NB*N*(N-1) real numbers simply arranged sequentially. There is no physical interpretation of the lines displayed in Figure 25 other than that the relative heights of the various real and imaginary parts should share characteristics from event to event. By constructing the ensemble covariance matrices by taking medians of each element for each event in the cluster, we obtain an apparently stable estimate which is not greatly influenced by outliers in the population.

The SPITS array was upgraded in 2004 from being a vertical-only array (with a single 3-component instrument) to having 6 out of 9 sites equipped with 3-component instruments (Gibbons et al., 2011). For the current study, we compiled a set of template events which included events prior to 2004 and so, for simplicity, only the 9 vertical sensors have initially been used. Gibbons et al. (2011) demonstrated that the horizontal components not only had far greater amplitudes than the vertical components for regional S-phases, but also that the coherence between sensors was often far greater for the horizontal ground motion. A final matched field beam deployment for the SPITS array will inevitably have to exploit both vertical and horizontal ground motion for optimal detection of all phases of interest, and a detection recipe will need to take into account that different sets of channels may be required for the optimal detection of different types of signals.

For each of the zones of seismicity displayed in Figure 23, we calculate empirical steering vectors as the principal eigenvectors of the ensemble covariance matrices calculated. Figure 26 displays the matched field statistics calculated as a function of time and frequency for six of these first P-arrival empirical templates, evaluated over a short time-window (slightly in excess of 5 minutes) on August 9, 2015. This data segment includes the signal from a magnitude 3.4 event from the North West along the spreading ridge. These six steering vectors constitute part of the empirical “beam deployment” displayed in Figures 23 and 24 with the color of each pixel indicating the degree to which the pattern of phase-shifts at the time and frequency indicated match the phase-shifts specified in the empirical steering vector. As we would hope, at the time of the P-arrival from our signal, a strong matching is observed across all of the narrow frequency bands

for a number of our matched field detector traces. The values obtained are highest for the empirical vectors from zones 21 to 24 which all correspond to a similar backazimuth from the SPITS array (~325 degrees). As the direction of the events from which the empirical steering vectors are calculated deviates from this backazimuth, the match between the steering vector phase shifts and the phase shifts in the incoming data diminishes.

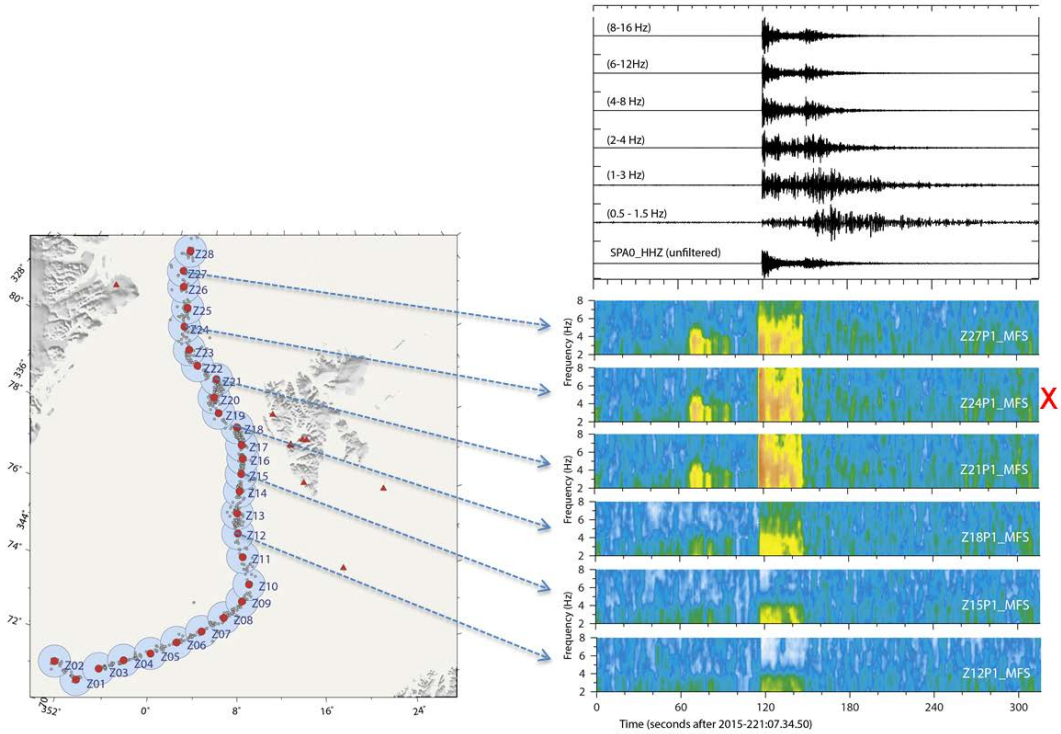

Figure 26: Time-frequency displays of matched field statistics for "beams" based upon the empirical steering vectors calibrated for first P-arrivals for ensembles for the zones of seismicity indicated. ***The waveform for the channel SPA0_HHZ is displayed for reference along the same time-scale bandpass filtered in the bands as indicated. Note that approximately 50 seconds before the main signal, there is significant peak in the matched field statistics, especially at lower frequencies. This is a low amplitude signal with a similar backazimuth and apparent velocity.***

The challenge of controlling operational detectors based upon these principles is to be able to declare detections based upon the information contained in such time-frequency plots, together with absolute measures of the waveform amplitudes. As discussed by Gibbons et al. (2011), the detection of events from this source region at the SPITS array has always been challenging since the S-phases have typically been quite incoherent on the vertical components and f-k analysis always results in typical P-wave velocities, making it very difficult for phase association algorithms to interpret the detections correctly. This is illustrated in Figure 27.

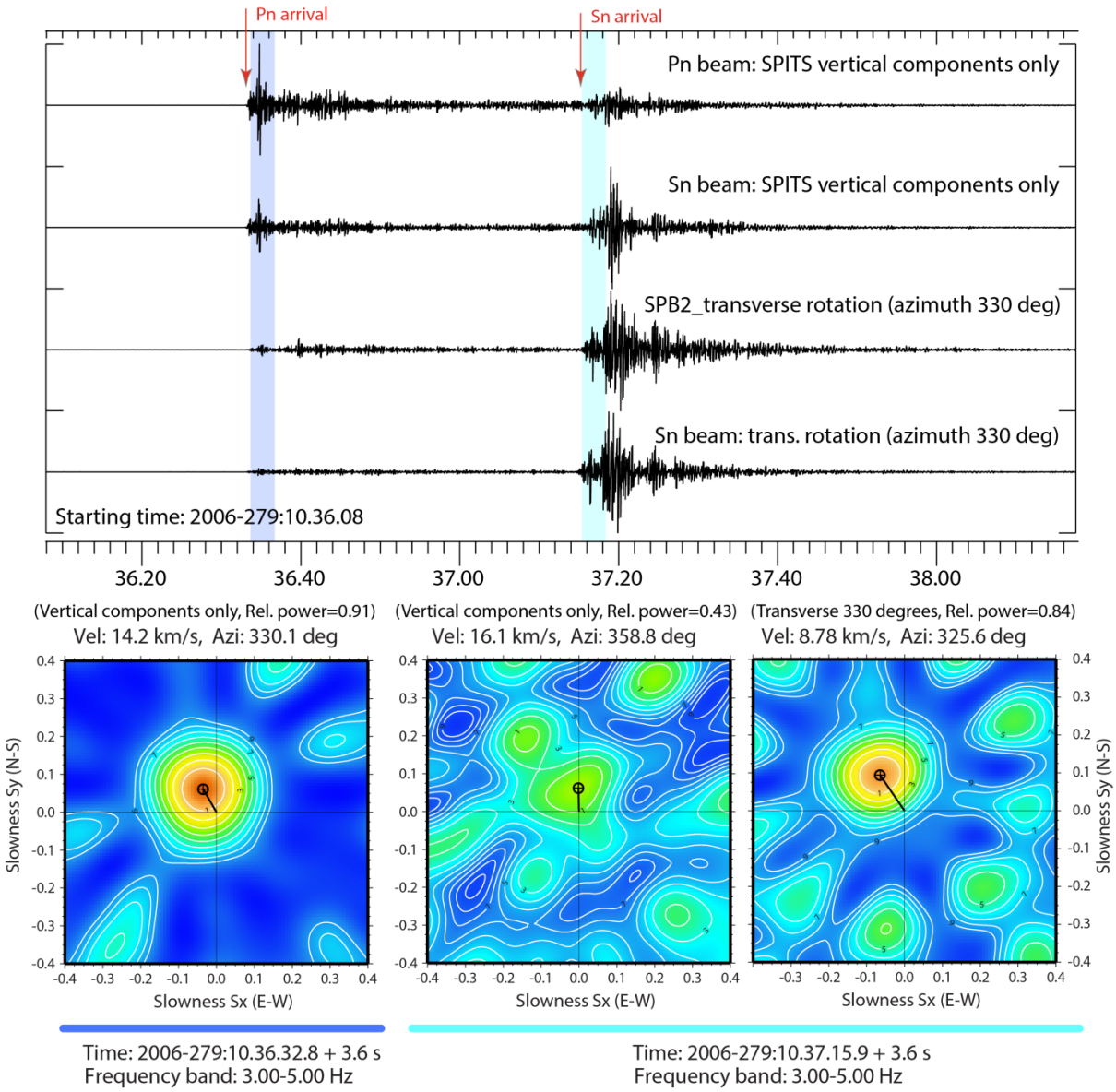

Figure 27: Plane-wave f-k analysis for Pn and Sn phases for an event on the ridge to the North West of the SPITS array. ***The slowness estimate for the Pn phase is evaluated only on the vertical traces whereas the slowness estimate for the Sn phase is evaluated both on the vertical and horizontal components. The coherence of the Sn phase on the horizontal traces is far greater than that for the Sn phase on the vertical traces. (Reproduced from Gibbons et al., 2011).***

The time-frequency plots in Figure 26 show that throughout the P-coda of the regional event displayed, there is a good match between the phase structure of the coda wavetrain and the first P-arrival template. On the arrival of the Sn phase, this phase structure completely disappears and we return to the "background level" of these "P-beam" detection statistics. The corresponding plots for the Sn matched field beams do not show a convincing correspondence with the template at the time of the Sn arrival. This result is a disappointment given an expectation that representative matched field templates could be calculated for all phases from ensemble covariance matrices evaluated over a broad

region. It works well for the P-arrival but poorly for the S-arrival on the vertical sensors for events in this source region. It is to be expected from Figure 27 that an S-phase detector would work far better on the horizontal sensors given the improved signal coherence.

The phase structure of coda waves on the SPITS array for regional events was studied by Schissele and Schweitzer (2004). They demonstrated that for some paths (for example from Novaya Zemlya to Spitsbergen) the coda waves were highly coherent and indicated the correct backazimuth and characteristic P- and S-wave velocities, whereas for other paths (for example from Mohns Ridge to Spitsbergen) there was little stability in the coda backazimuth estimates and that little evidence of S-wave apparent velocities could be detected. The work of Schissele and Schweitzer (2004) has significant implications for the likely effectiveness of a time-frequency matched field detector beam recipe. It is quite likely that a full wavetrain analysis, rather than a first-arrival-only detector is likely to provide the information necessary to be able to make a signal classification. The sudden disappearance of the P-wave-consistent coda wave energy, combined with an increase in energy and decrease in coherence on the vertical traces, could in fact be exploited as a wavefield characteristic used to identify the most likely source of a signal.

Schissele and Schweitzer (2004) were limited to the vertical component array, but we can now extend our analysis to the 3-component subarray at SPITS. Figure 28 shows f-k analysis of the coda waves for the October 11, 2010, M = 4.4 Novaya Zemlya earthquake on three different components at SPITS. Unlike the ridge event example displayed in Figure 27, the S-wave energy is highly coherent on the vertical sensors. The coherence between the horizontal sensors is even greater following the Sn arrival, but it is also worthy of note that the coherent energy in the P-wave coda for this event is present on all components - but propagating with different apparent velocities.

In summary, we have made significant progress in evaluating a matched field frontline pipeline detector for the SPITS array. We have been able to calibrate empirical steering vectors for the detection of initial P-wave arrivals from events on the ridge. It is likely that for robust detection and identification of signals – in particular S-phases - we ought to consider all components and examine the full wavetrain characteristics and not just the initial signal onset.

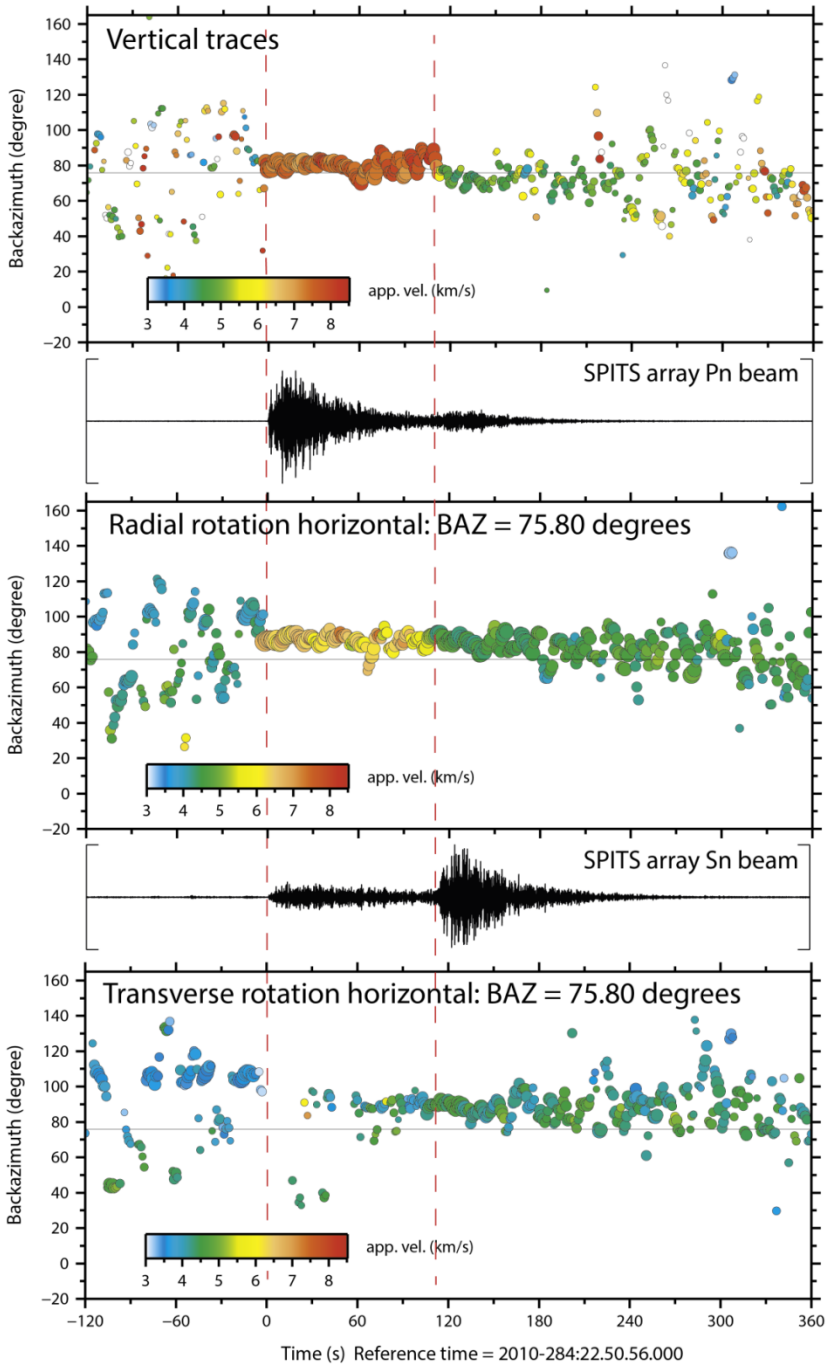

Figure 28: Evaluation of slowness vectors on the vertical, radial and transverse components at the SPITS array for the October 11, 2010, Novaya Zemlya earthquake. ***An evaluation is made every second using a 4-second long time-window with a 2-8 Hz broad frequency band used. The size of the symbols is proportional to the relative power in the f-k beam (essentially a measure of coherence).***

Empirical steering vectors were calculated for each arrival present in the NORSAR waveform archive for all of the events displayed in Figure 17 and, for a number of phase arrivals, we evaluated the correspondence with the complete set of empirical steering vectors. An example, for an earthquake south of Japan, is displayed for the three arrays in figures 29, 30, and 31.

The results for the USRK array, for an earthquake recorded on January 2, 2009, are displayed in Figure 29. The covariance matrix was evaluated over the nine vertical channels of the array for a window with length 256 samples (6.4 seconds) and frequency bands between 1.09 Hz and 5.00 Hz were considered. As for the SPITS examples described in the previous reporting period, the highest values of the matched field statistics are close to the great circle path linking the event hypocenter and the station. As we move away from this line, the statistic diminishes indicating a decrease in the similarity in the observed phase shifts. There is clearly also a distance effect with little similarity observed between the phase shifts observed for the closest events to the station and the phase shifts observed in the signal from the January 2 event. One explanation for this is likely to be that closest events are very deep, such that the arriving wavefronts are likely to reach the array at a very steep angle of inclination.

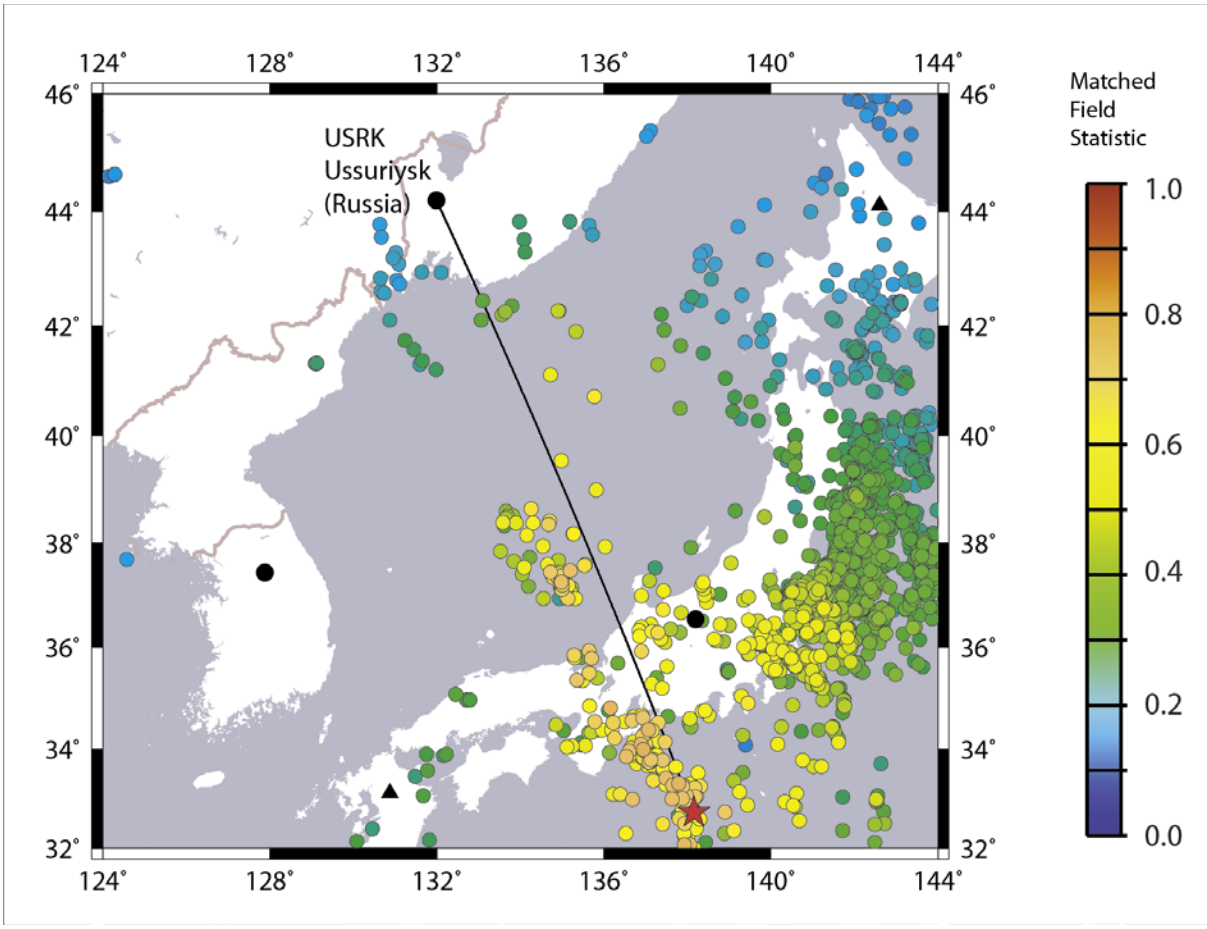

Figure 29. Geographical matched field direction estimates on the USRK array. ***Empirical matched field statistics calculated between the covariance matrix evaluated for the regional P arrival at the USRK array at a time 2009-002:07.45.38.850 and the steering vectors calculated from the covariance matrices evaluated for the first P arrival at USRK for each of the events indicated. The red star denotes the location of the earthquake with origin time 2009-002:07.42.51.34 and epicenter (33.72, 138.15).***

The corresponding image for the KSRS array is displayed in Figure 30. The matched field statistics evaluated for this larger aperture array are significantly lower in absolute terms and only exceed the background level in a rather narrow range of backazimuth from the array. It is noted however that the "background level" for this 10km plus aperture array is significantly lower than for the far smaller aperture USRK array. So, while the colors on figures 29 and 30 indicate a weaker correspondence for KSRS than for USRK, the significance of a given value is greater for the larger aperture array.

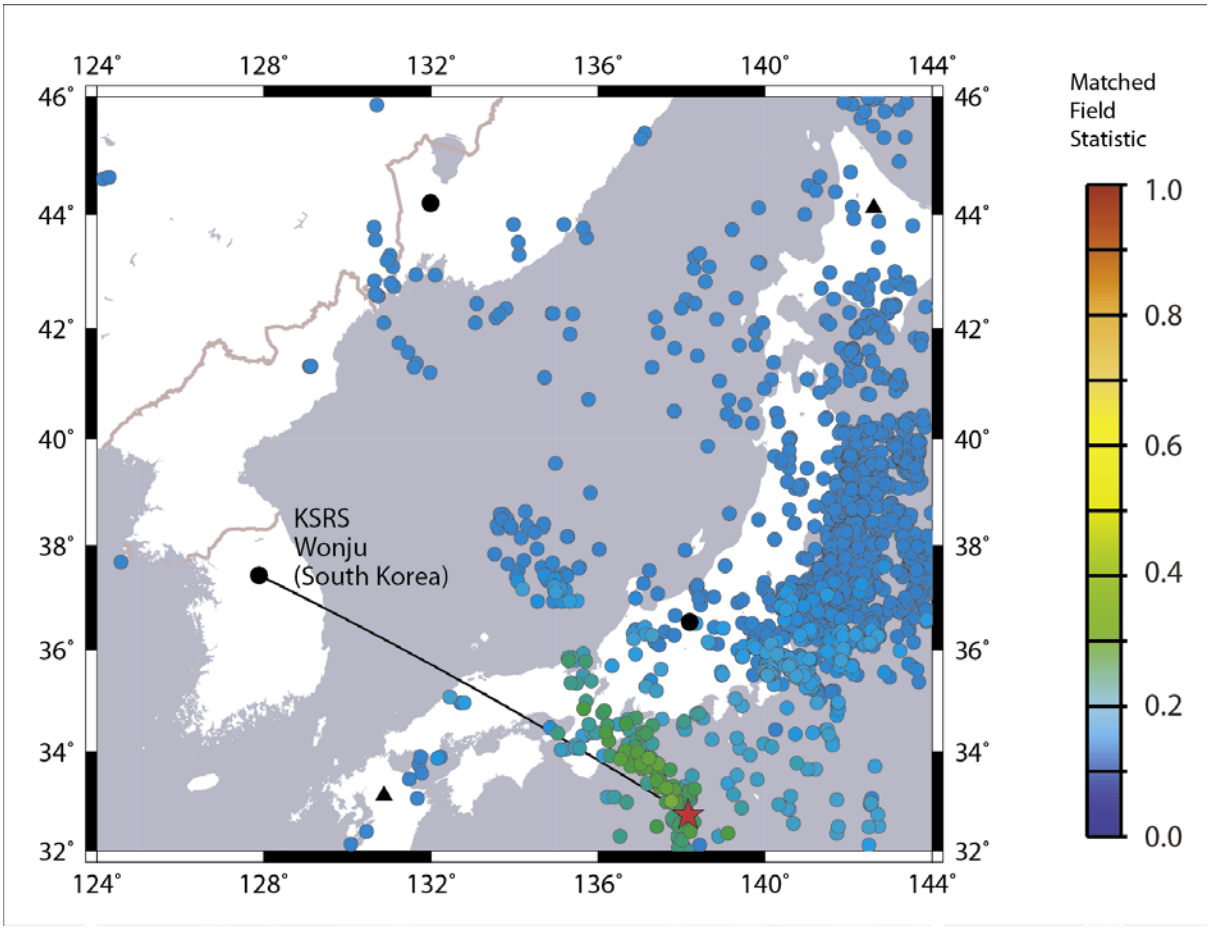

Figure 30. Geographical matched field direction estimates on the KSRS array. ***Empirical matched field statistics calculated between the covariance matrix evaluated for the regional P arrival at the KSRS array at a time 2009-002:07.45.07.350 and the steering vectors calculated from the covariance matrices evaluated for the first P arrival at KSRS for each of the events indicated. The red star denotes the location of the earthquake with origin time 2009-002:07.42.51.34 and epicenter (33.72, 138.15).***

It comes as little surprise that the geographical response of the MJAR array is even narrower, with values of the matched field statistic only exceeding the background level in a very limited region surrounding the estimated event epicenter. Observing many plots similar to those displayed in figures 29, 30, and 31 we conclude that the empirical matched field statistics do indeed have a footprint which maps back the geographical source region with a pattern which is consistent with expectations. With a number of observing phases which are associated correctly, it is likely that a good preliminary estimate of the source region can be obtained from a numerical evaluation of matched field statistics even without explicit reference to arrival times. This alludes to a possible algorithm for evaluating source hypotheses.

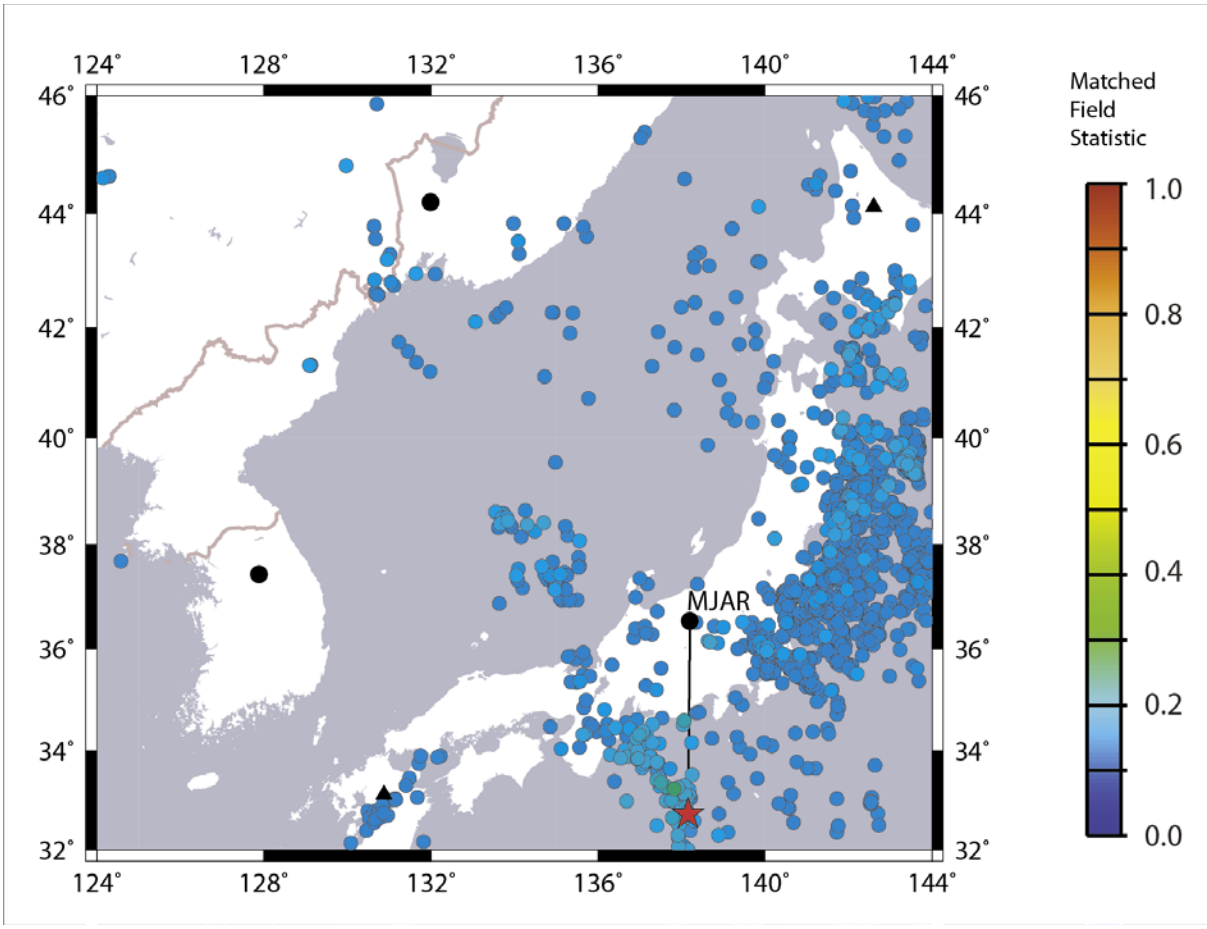

Figure 31. Geographical matched field direction estimates on the MJAR array. ***Empirical matched field statistics calculated between the covariance matrix evaluated for the regional P arrival at the MJAR array at a time 2009-002:07.43.56.675 and the steering vectors calculated from the covariance matrices evaluated for the first P arrival at MJAR for each of the events indicated. The red star denotes the location of the earthquake with origin time 2009-002:07.42.51.34 and epicenter (33.72, 138.15).***

We have highlighted the specific dependence of the geographical resolution of the matched field statistic on the array aperture and coherence characteristics of the relevant seismic waves on the sensors. This is to say that, like in classical beamforming, a small aperture array (such as USRK or even more so SPITS) will be a relatively blunt instrument in location estimation with a high response over a wide parameter space. A large or sparse array (such as MJAR), for signals in a similar frequency band, will act more as a spotlight or pattern detector and have a highly source-specific response. This will require a far denser deployment of matched field beams to be able to cover the appropriate parameter space.

We have begun to examine ways of mitigating this, in order to provide more robust and less source-specific detectors for arrays which suffer from lower coherence. Approaches include the formation of ensemble covariance matrices for multiple events in a given region (as was demonstrated in the previous reporting period for the SPITS array). This will reduce the influence of cross-terms which are highly source-specific and increase the

relative influence of those characteristics which are more consistent over a wider source region. In the absence of numerous calibration events, an explicit removal of cross-terms from sensors with typically low coherence (usually those at greatest distance from each other) may also facilitate more robust detectors from given templates.

In Figure 32, we revisit our definition of the co-array to consider only pairs of sensors with a spatial separation below a specified distance threshold. Performing f-k analysis for high frequency signals on large aperture arrays has long been a problem for standard seismic processing pipelines (see Gibbons et al., 2008, and references therein). The diminishing waveform similarity over large inter-site distances can result in spurious or unstable estimates of direction for some seismic arrivals.

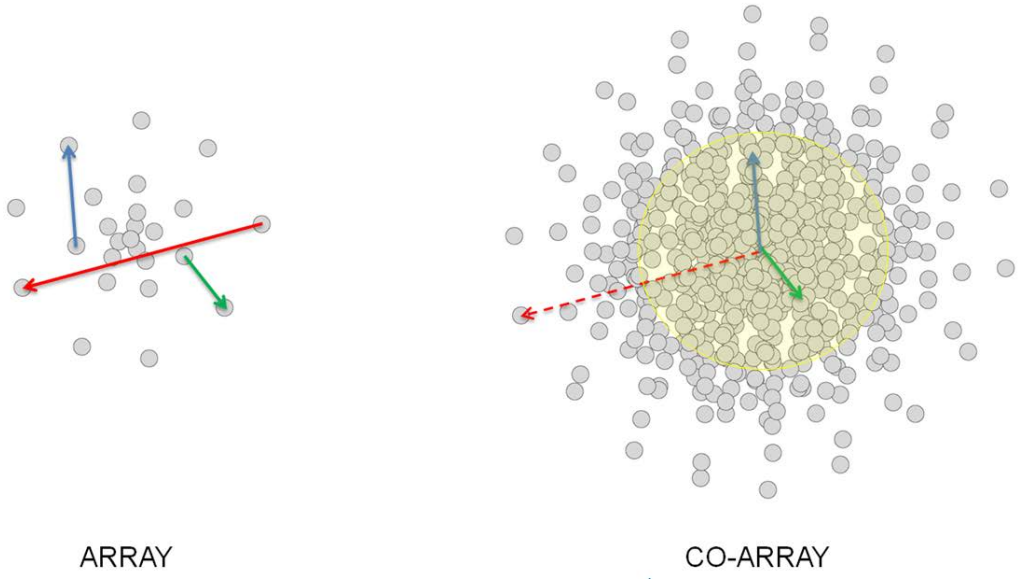

Figure 32. Limiting the covariance matrix to contain only cross-terms between sensors with limited spatial separation. ***All sensor pairs for which the distance exceeds the radius of the yellow circle are not included.***

A typical example of a problematic arrival on a large aperture seismic array is displayed in Figure 33. This a high frequency regional Pn signal recorded on an array with an aperture exceeding 25 km. The f-k spectrum (left) appears to have a maximum with a slowness vector which is consistent with the anticipated direction of arrival, but the image is speckled and the contours of beam-gain follow an irregular pattern. A possible consequence of this is that with slightly different processing parameters, or a somewhat lower SNR, it is likely that a qualitatively different slowness estimate may have been made.

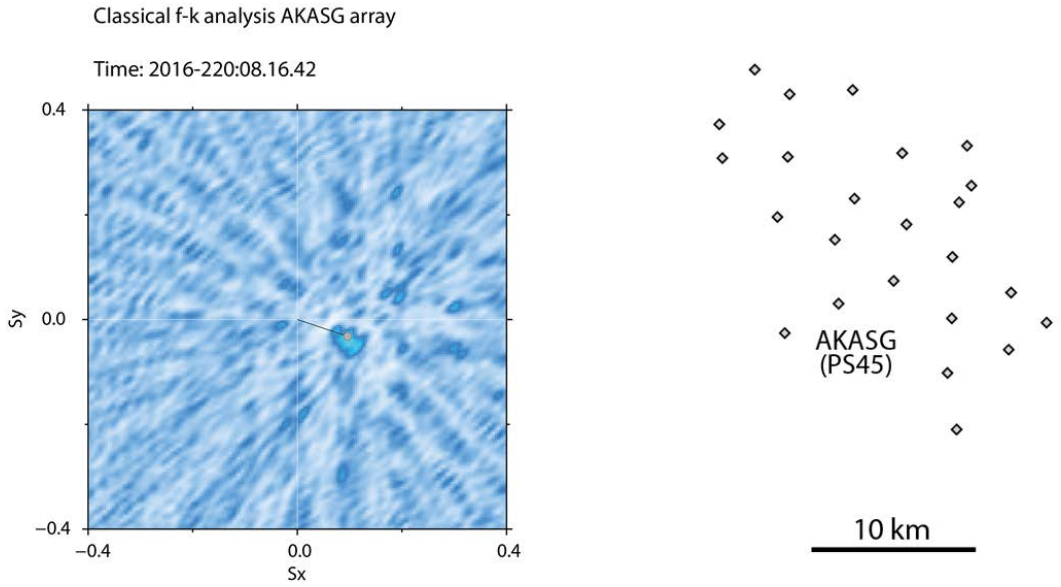

Figure 33. "High-risk" direction estimate of a high frequency signal on a large aperture seismic array (Ukraine example). ***(left) Slowness estimate for the Pn arrival at the AKASG array (Malin, Ukraine) for an regional earthquake in eastern Ukraine on 7 August 2016. (right) geometry of the AKASG array.***

If we were instead to calculate correlation traces between sensors, attribute the coordinates of each trace to the corresponding elements of the co-array, and then perform f-k analysis (c.f. Gibbons et al., 2015) then we should get a similar direction estimate to that obtained when performing f-k analysis on the original data. However, we are now at liberty to remove traces from this "virtual array". It is likely that the correlation traces calculated between closely spaced sensors are likely to have clearer and better defined maxima than those between more distant sensors, and we can simply exclude those sensor pairs with an inter-site distance exceeding a pre-specified threshold. Note that there are clear analogies with the PMCC method (Cansi, 1995) which uses only selected "triads" of sensors and need not exploit all possible triads. The resulting f-k grid is displayed in the right panel of Figure 34 and we see a clear well-defined maximum. We note that this slowness estimate still exploits all of the waveforms on the array; no raw data is discarded.

Figure 35 displays the slowness estimate for the P-arrival at the NORSAR array from the 9 October 2006 declared nuclear test in North Korea. Almost all of the energy in this signal is in the 2-4 Hz frequency band (Gibbons et al., 2008) and the corresponding f-k grid from the same frequency band has multiple local maxima. If we repeat the same procedure applied to the AKASG array case, we again see that a single well-defined maximum with symmetrical contours of beam-gain is now obtained.

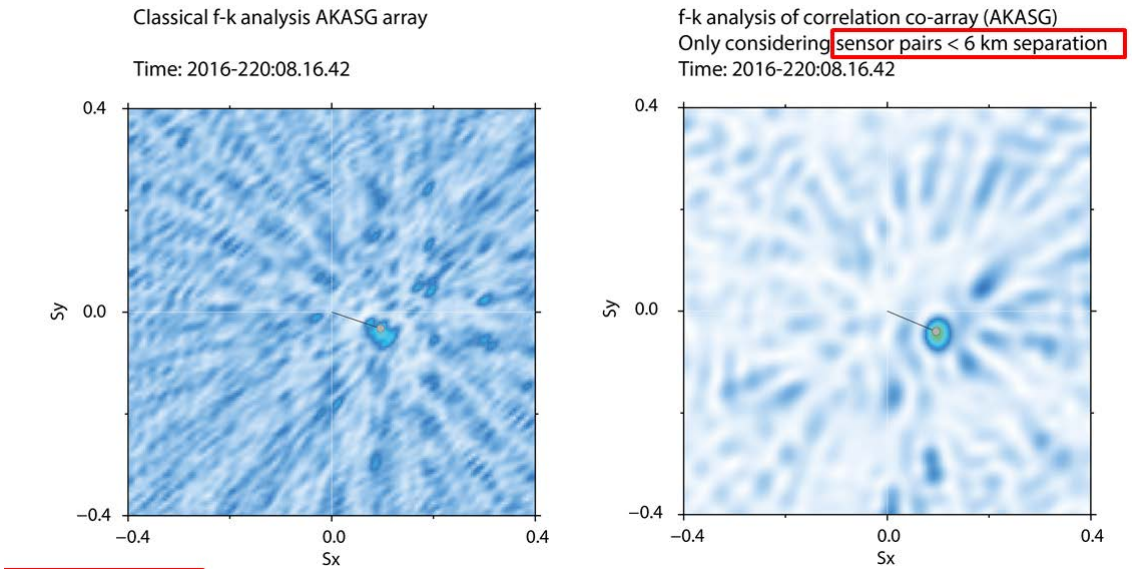

Figure 34. Improvement of direction estimate for the Ukraine example using f-k analysis on limited sensor pair correlations. ***(left) f-k plot from Figure 33 and (right) f-k analysis of correlation co-array containing only sensor pairs with separation below 6 km.***

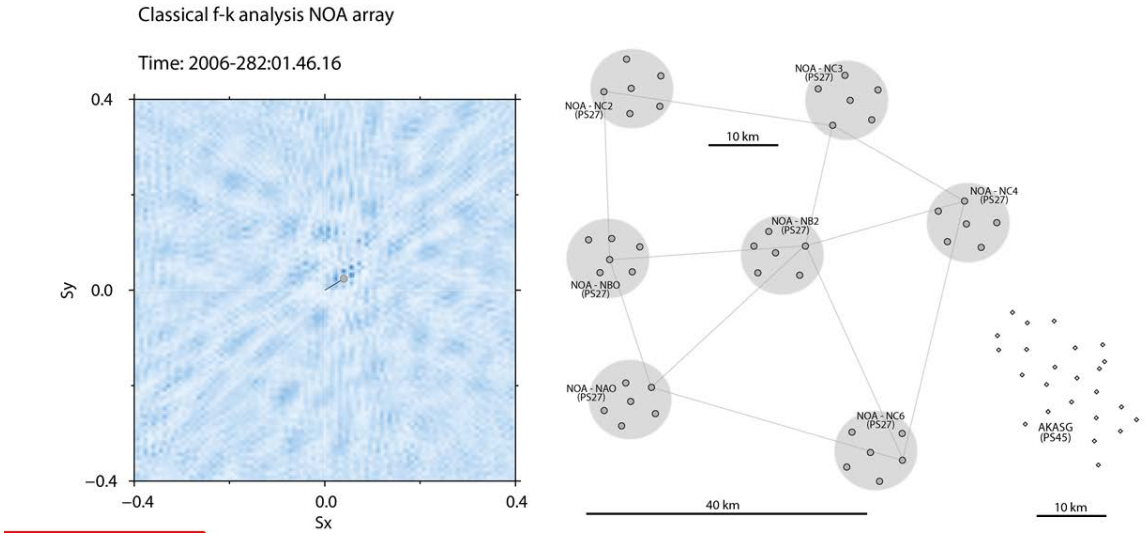

Figure 35. "High-risk" direction estimate of a high frequency signal on a large aperture seismic array (North Korea example). ***(left) f-k estimate for the slowness of the teleseismic P-arrival on the NORSAR array from the October 9, 2006, nuclear test in North Korea. (right) geometry of the NORSAR array with that of AKASG to the same scale.***

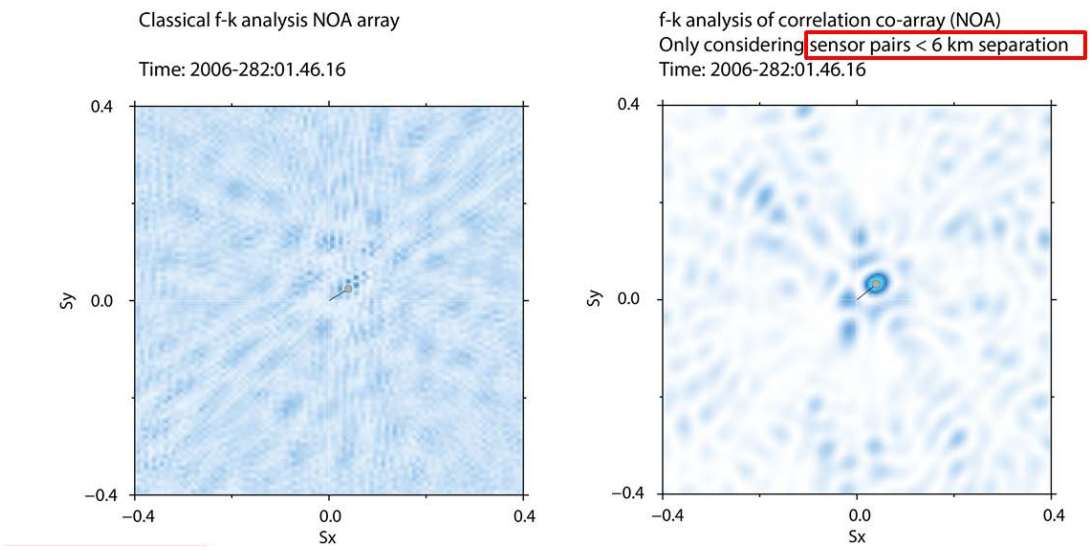

Figure 36. Improvement of direction estimate for the North Korea example using f-k analysis on limited sensor pair correlations. ***(left) f-k plot from Figure 35 and (right) f-k analysis of correlation co-array containing only sensor pairs with separation below 6 km.***

In each of these examples, we are essentially forcing a large aperture seismic array to behave as a small aperture array by considering only a limited set of inter-sensor correlations. In the final year of the project we will explore strategies for applying the same principles to the matched field processing algorithms in order to extend the geographical footprint of the target. The simplest step in this direction is simply to zero those elements of the covariance matrix which correspond to channel pairs that we would prefer to ignore. However, if this results in a covariance matrix which is block-diagonal (as would be the case with the NORSAR array) then more complicated procedures, such as calculating the full correlation matrix at each time-step, may be necessary.

For the long-period set of signals observed at Warramunga, we see also characteristic geographical trends. After screening the 606 events displayed in Figure 19 for those with complete recordings along the two arms (WR1-9, WB1-9) of the Warramunga array, 378 events were available for further analysis. Figure 37 shows a measure of similarity between empirical matched field templates for 17-second Rayleigh waves constructed from observations at Warramunga for the 378 events. The template for a reference event in the Far East is used as the basis for comparing all other templates in the dataset. The similarity measure is the norm of the inner product between templates, and since all templates are normalized to have unit energy, these norms range between 0 and 1. In the figure, the symbols indicate the location of the events for which templates are being compared to the reference, with symbol color encoding the similarity measure. Generally, the similarity measures have high values in the east, with values declining the further events are to the west. However, there are significant variations in the trend, which probably are associated with data problems, particularly high-amplitude dropouts, but also superimposed events.

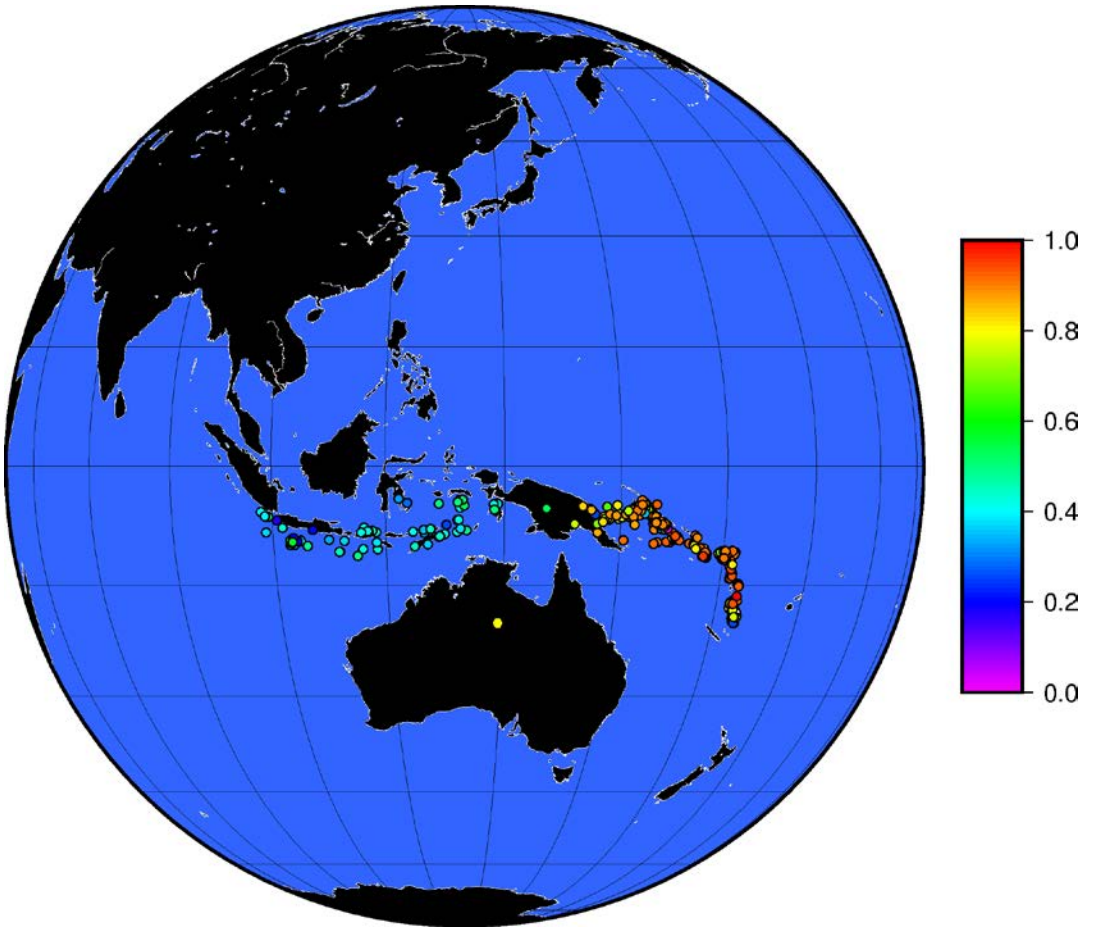

Figure 37. Map showing the size of the inner product (rendered with color) of matched field templates against a template for a reference event at the eastern end of the arc. ***An inner product of 1 indicates that two (normalized) templates are identical. Generally the pattern shows high similarity between templates in the east, with similarity decreasing to the west. However, there is significant variation in template similarity that probably is caused by dropouts and other data problems, indicating the need for additional screening.***

A similar (but reversed) pattern emerges when the reference event is chosen as one of the far western events (Figure 38). The concern with the variations that we see from the trend is that the variations may not represent a structural signal, but rather data problems.

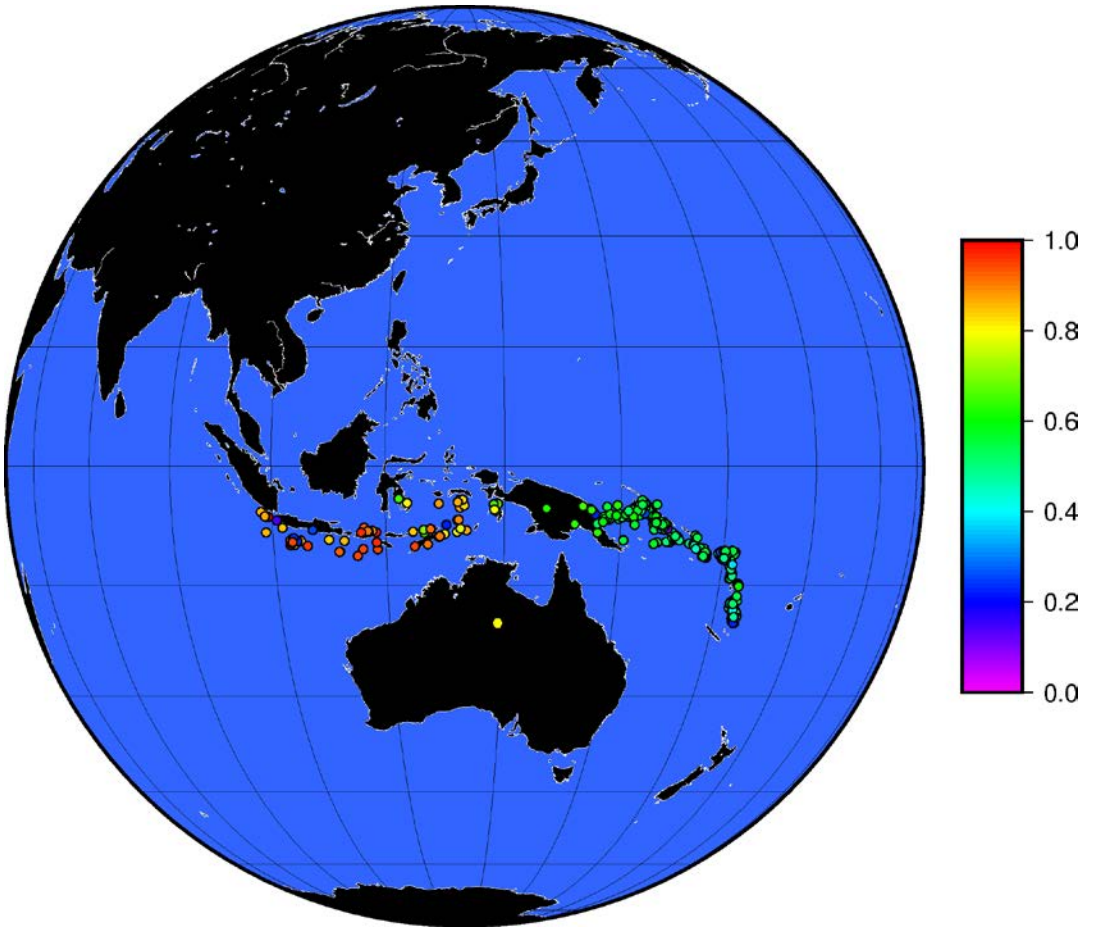

Figure 38. A map of Rayleigh wave template similarity for a reference event in the far west of the arc.

5. CONCLUSIONS

Empirical Matched Field Processing (EMFP) is a narrow frequency band method for signal detection and estimation which recognizes a characteristic phase and amplitude structure over an array of sensors. It has already been demonstrated that EMFP is an excellent "pattern detector" for recognizing new occurrences of previously observed signals. It is also demonstrated that the classical array processing tools for the detection of "unknown signals" in front-end seismic pipelines are in fact generalizations of EMFP. In this project, we explore the extent to which EMFP can provide a comprehensive detection and estimation system both for targeted and general seismic signals.

We seek

- completeness of signal detection
  (i.e. no clear signals from any directions shall be missed)
- high sensitivity for signals of special interest
  (signals of monitoring interest shall be detected well below the noise level)
- accuracy of parameter estimates
  (estimates of signal attributes shall provide accurate information about the likely source of the signal - bias due to path and receiver effects shall be minimal).

A preliminary study targeted the European Arctic. The detectability of small magnitude seismic events in this region of monitoring interest is governed by two small-aperture seismic arrays (SPITS and ARCES) and parameter estimates at both of these arrays are known to be susceptible to bias due to near-array structure.

An accurate relocation of seismic events along the northern mid-Atlantic ridge was undertaken and, using the Bayesloc program with both regional and teleseismic signals, we obtained a set of reference events for which the location estimates are robust to small changes to the regional velocity model employed. Matched Field templates were constructed for first arrivals at both SPITS and ARCES for each of many zones covering the relocated seismicity. A careful median-stack procedure is used to construct the so-called ensemble covariance matrices; the procedure makes the templates insensitive to missing, noisy, or faulty data on individual channels. Analysis of signal detection and estimation algorithms indicates that one template (or "matched field beam") for each one of these zones is sufficient to cover the parameter space of anticipated signals and the variability of the matched field template is very smooth as a function of the master event location. Parameter estimates made on SPITS and ARCES using the empirical matched field templates show far less bias than the plane-wavefront estimates used in the current processing pipelines.

We examine the covariance structure of the full wavetrain for events in the European Arctic observed at regional distances and anticipate that optimal monitoring of such signals will require consideration of coda in addition to the first arrival.

To explore the applicability of EMFP to different array geometries and tectonic settings, we performed a study in the Far East (Japan, the Korean Peninsula and surrounding areas). A dataset of over 2500 seismic events was obtained for the region with each event providing clear Pn arrivals for each of the 3 arrays USRK (Russian Federation), KSRS (Republic of Korea), and MJAR (Japan). USRK consists of 9 elements over an aperture of 3.8 km and P-wave signals up to 4 Hz are typically acceptably coherent over the array. KSRS has an aperture of 10.1 km but also supports coherent array processing for signals up to approximately 3 Hz. MJAR has an aperture of 11.3 km but, due to mountainous terrain and complex near-receiver geology, typically fails for coherent signal processing above 1 Hz. For USRK, the matched field templates covered a wide region and varied smoothly as for the SPITS and ARCES examples. For KSRS, the resolution is far sharper due to the larger aperture and it is clear that more matched field beams will be required to provide detection completeness in the same frequency bands. For MJAR, signal scattering results in very small geographical coverage for each matched field template; EMFP detectors on MJAR will be essentially spotlight detectors.

For the matched field detectors to provide more general coverage across geometries of existing IMS-type arrays, steps need to be made to mitigate the effects of incoherence. We demonstrate a significant improvement in f-k estimates for high frequency signals on large aperture seismic arrays by performing correlation on selected pairs of sensors prior to classical f-k analysis on the resulting reduced co-array. This is shown for a regional Pn phase on the AKASG array in Ukraine (aperture 27.6 km) and a high frequency teleseismic signal on the NOA array in Norway (aperture 77.5 km), both processed in the 2-4 Hz frequency band and considering only sensor pairs with inter-receiver distances below 6 km. We recommend that similar steps are taken in the formulation of the matched field detection and estimation algorithms.

REFERENCES

Barrett, S. A. and Beroza, G. C. (2014), An Empirical Approach to Subspace Detection, Seismological Research Letters, **85**, pp. 594-600, doi:10.1785/0220130152.

Bondar, I., North, R. G., and Beall, G. (1999), Teleseismic slowness-azimuth station corrections for the International Monitoring System Seismic Network, *Bulletin of the Seismological Society of America*, **89**, pp. 989-1003.

Cansi, Y. (1995), An automatic seismic event processing for detection and location: The P.M.C.C. Method, Geophysical Research Letters, **22**, pp. 1021-1024, doi:10.1029/95gl00468.

Engen, Ø., Eldholm, O., and Bungum, H. (2003), The Arctic plate boundary, *Journal of Geophysical Research - Solid Earth*, **108**, B2. 2075, doi:10.1029/2002jb001809.

Gibbons, S. J. and Ringdal. F. (2006), The detection of low magnitude seismic events using array-based waveform correlation, Geophysical Journal International, **165**, pp. 149-166, doi:10.1111/j.1365-246x.2006.02865.x.

Gibbons, S. J., Ringdal, F., and Kværna, T. (2008), Detection and characterization of seismic phases using continuous spectral estimation on incoherent and partially coherent arrays, *Geophysical Journal International*, **172**, pp. 405-421, doi:10.1111/j.1365-246x.2007.03650.x.

Gibbons, S. J. and Ringdal, F. (2012), Seismic Monitoring of the North Korea Nuclear Test Site Using a Multichannel Correlation Detector, IEEE Transactions of Geoscience and Remote Sensing 50, pp. 1897-1909, doi:10.1109/tgrs.2011.2170429.

Gibbons, S. J., Schweitzer, J., Ringdal, F., Kværna, T., Mykkeltveit, S., and Paulsen, B. (2011), Improvements to Seismic Monitoring of the European Arctic Using Three-Component Array Processing at SPITS, *Bulletin of the Seismological Society of America* **101**, pp. 2737-2754, doi:10.1785/0120110109.

Gibbons, S. J., Asming, V., Eliasson, L., Fedorov, A., Fyen, J., Kero, J., Kozlovskaya, E., Kværna, T., Liszka, L., Näsholm, S. P., Raita, T., Roth, M., Tiira, T., and Vinogradov, Y. (2015), The European Arctic: A Laboratory for Seismoacoustic Studies, *Seismological Research Letters*, **86**, pp. 917-928, doi:10.1785/0220140230.

Harris, D. B. (2006), Subspace Detectors: Theory, UCRL-TR-222758, Lawrence Livermore National Laboratory, 700 East Avenue, Livermore, CA.

Harris, D. B. and Kværna, T. (2010), Superresolution with seismic arrays using empirical matched field processing, *Geophysical Journal International*, **182**, pp. 1455-1477, doi:10.1111/j.1365-246x.2010.04684.x.

Hicks, E. C., Kværna, T., Mykkeltveit, S., Schweitzer, J., and Ringdal, F. (2004), Travel-times and Attenuation Relations for Regional Phases in the Barents Sea Region, *Pure and Applied Geophysics*, **161**, pp. 1-19, doi:10.1007/s00024-003-2437-6.

Myers, S. C., Johannesson, G., and Hanley, W. (2007), A Bayesian hierarchical method for multiple-event seismic location, *Geophysical Journal International*, **171**, pp. 1049-1063, doi:10.1111/j.1365-246x.2007.03555.x.

Myers, S. C., Johannesson, G., and Hanley, W. (2009), Incorporation of probabilistic seismic phase labels into a Bayesian multiple-event seismic locator, *Geophysical Journal International*, **177**, pp. 277-289, doi:10.1111/j.1365-246x.2008.04070.x .

Mykkeltveit, S. and Ringdal, F. (1981), Phase identification and event location at regional distances using small-aperture array data, in "Identification of seismic sources - Earthquake or underground explosions" (ed.s Husebye, E. S. and Mykkeltveit, S.), Reidel Publishing Company, pp. 467-481.

Prieto, G. A., Parker, R. L., and Vernon, F. L. (2009), A Fortran 90 library for multitaper spectrum analysis, Computers & Geosciences, 35, pp. 1701-1710, doi:10.1016/j.cageo.2008.06.007.

Schissele, E. and Schweitzer, J. (2004), Study of regional variations of the coda characteristics in the Barents Sea using small-aperture arrays, Semiannual Technical Summary Report 1-2004, NORSAR, PO Box 53, Gunnar Randers vei 15, N-2027 Kjeller, Norway, February 2004, pp. 70-80.

Schweitzer, J. (2001a), HYPOSAT – An Enhanced Routine to Locate Seismic Events, Pure and Applied Geophysics, **158**, pp. 277-289, doi:10.1007/pl00001160.

Schweitzer, J. (2001b), Slowness Corrections — One Way to Improve IDC Products, Pure and Applied Geophysics, **158**, pp. 375-396, doi:10.1007/pl00001165.

Slinkard, M. E., Carr, D. B., and Young, C. J. (2013), Applying Waveform Correlation to Three Aftershock Sequences, Bulletin of the Seismological Society of America, **103**, pp. 675-693, doi:10.1785/0120120058.

Withers, M., Aster, R., Young, C., Beiriger, J., Harris, M., Moore, S., and Trujillo, J. (1998), A comparison of select trigger algorithms for automated global seismic phase and event detection, Bulletin of the Seismological Society of America, **88**, pp. 95-106.

LIST OF SYMBOLS, ABBREVIATIONS, AND ACRONYMS

EMFP	Empirical Matched Field Processing
MFP	Matched Field Processing
NEIC	National Earthquake Information Center
SNR	Signal to Noise Ratio

DISTRIBUTION LIST

DTIC/OCP
8725 John J. Kingman Rd, Suite 0944
Ft Belvoir, VA 22060-6218 1 cy

AFRL/RVIL
Kirtland AFB, NM 87117-5776 1 cy

Official Record Copy
AFRL/RVBYE/Dr. Frederick Schult 1 cy