



LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

NORSAR Final Scientific Report Adaptive Waveform Correlation Detectors for Arrays: Algorithms for Autonomous Calibration

S. J. Gibbons, F. Ringdal, D. B. Harris

April 22, 2009

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.



Final Scientific Report

Contract Number: DE-FC52-05NA26604

Adaptive Waveform Correlation Detectors for Arrays: Algorithms for Autonomous Calibration

submitted by

NORSAR
Kjeller, Norway

15 April 2009

Principal Investigator

Dr. Frode Ringdal
NORSAR, P.O. Box 53
N-2027 Kjeller, Norway

Report prepared by

Steven J. Gibbons

Frode Ringdal

David B. Harris

Executive Summary

Correlation detection is a relatively new approach in seismology that offers significant advantages in increased sensitivity and event screening over standard energy detection algorithms. The basic concept is that a representative event waveform is used as a template (i.e. matched filter) that is correlated against a continuous, possibly multichannel, data stream to detect new occurrences of that same signal. These algorithms are therefore effective at detecting repeating events, such as explosions and aftershocks at a specific location.

This final report summarizes the results of a three-year cooperative project undertaken by NORSAR and Lawrence Livermore National Laboratory. The overall objective has been to develop and test a new advanced, automatic approach to seismic detection using waveform correlation. The principal goal is to develop an adaptive processing algorithm. By this we mean that the detector is initiated using a basic set of reference (“master”) events to be used in the correlation process, and then an automatic algorithm is applied successively to provide improved performance by extending the set of master events selectively and strategically. These additional master events are generated by an independent, conventional detection system. A periodic analyst review will then be applied to verify the performance and, if necessary, adjust and consolidate the master event set.

A primary focus of this project has been the application of waveform correlation techniques to seismic arrays. The basic procedure is to perform correlation on the individual channels, and then stack the correlation traces using zero-delay beamforming. Array methods such

as frequency-wavenumber analysis can be applied to this set of correlation traces to help guarantee the validity of detections and lower the detection threshold.

In principle, the deployment of correlation detectors against seismically active regions could involve very large numbers of very specific detectors. To meet this challenge, we have examined two strategies:

- use of *subspace detectors*, a multi-dimensional extension of correlators, which allow representation and detection of signals exhibiting some degree of variation
- autonomous calibration of many subspace and correlation detectors in an adaptive detection framework, subject to analyst review.

Because correlation detectors are relatively new to seismology, a significant amount of research on how to tune these detectors has been needed to address later calibration efforts that will arise as they are adopted for operational use.

We have approached these challenges by carrying out a number of case studies, encompassing various monitoring scenarios such as earthquake aftershock sequences and swarms, recurring mining explosions, other types of explosions, and rockbursts. We have studied several different geographical regions (the European Arctic, Central Asia, and the western United States). We have drawn on available Ground Truth data in assessing the results of the various processing schemes. In all cases, we have benefited from the high-quality seismic arrays or networks available in these regions, and we have thus been able to evaluate the performance of array-based correlation processing under a variety of conditions.

The main results of the project are summarized as follows:

1. Array-based waveform correlation has been demonstrated to lower significantly detection thresholds in comparison with standard

single-channel waveform correlation. The scale of improvement is a function of the array geometry, the number of sensors and the selection of time window, but is typically of the order of 0.5 to 1.0 magnitude units for the cases studied here.

2. Frequency-wavenumber analysis of the correlation traces on a small-aperture array provides an effective method for screening out a certain category of false alarms, and can therefore be used to improve detector sensitivity by lowering the threshold for automatic array detection.
3. We have developed and tested a framework for autonomous correlation detection. The framework comprises a set of conventional (STA/LTA) detectors on a collection of array beams, augmented by correlation and subspace detectors. The detectors are applied in parallel. Triggers occur on both conventional and correlation detectors and we apply rules to determine when new correlation detectors are spawned, and when multiple correlation detectors are combined into one or several subspace detectors. Periodically the system is halted for an analyst-supervised recalibration of the correlation detectors.
4. We have experimentally tested the processing of a large aftershock sequence using various transformations of the signals from the main event to construct what we denote an *incoherent correlation detector* or *characteristic function correlation detector*. The prototype detector is not sensitive to the same degree as the standard correlation detector, but may trigger on more events from a broader target region, resulting in a more complete characterization of the aftershock sequence. The intention is that such a detector will identify potential master events within the region of interest whilst being relatively insensitive to signals from unrelated sources.
5. An important application of multi-channel waveform correlation is to identify instances of erroneous instrumental timing. We

provide examples of detected timing errors on single channels of an array, as well as on a remote three-component station.

While we consider that we have achieved the main goals set out for the project, it is clear that many issues need to be explored further. The framework for autonomous correlation detection needs to be further elaborated and tested, as do the ideas presented in this report for increasing the size of the “footprint” of the correlation detectors. Cluster analysis has proved very useful in designing and applying correlation detectors, and should be further explored for this purpose. Finally, the remarkable performance that has recently been demonstrated for the empirical matched field technique to classify mining events suggests that this technique should be applied in conjunction with the correlation methods, subspace methods, and classical detectors to obtain an optimized approach to detection and location using arrays or networks of stations.

Acknowledgements

We are grateful to Doug Dodge for software development related to the adaptive framework for correlation and subspace detectors. We thank Tim Paik for work on developing software for the efficient implementation of subspace detectors.

We are grateful to Mathilde Bøttger Sørensen and Jens Havskov from the Department of Earth Sciences at the University of Bergen (UiB), Norway for providing data from the Norwegian National Seismic Network (NNSN). The NNSN is operated by the Department of Earth Science, University of Bergen, Norway and supported by Oljeindustriens Landsforening (OLF) and the Faculty of Mathematics and Natural Sciences, University of Bergen.

We are grateful to the operators of the Kyrgyz Seismic Telemetry Network (Kyrgyz Institute of Seismology, IVTAN/KIS, network code KN) and the IRIS data exchange center for making this data available.

We thank Dr. Oliver Ritzmann of the University of Oslo for providing the coordinates and origin times of the shots in the PETROBAR seismic profiles. We are grateful to Berit Paulsen of NORSAR for locating aftershocks of the February 21, 2008, Svalbard earthquake.

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

Contents

1	Introduction	1
2	Detection Thresholds for Multichannel Correlation Detectors	9
2.1	A sequence of earthquakes in northern Norway	9
2.2	Setting the detection threshold	12
2.3	Improvement in detectability over standard STA/LTA detectors .	21
2.4	Relative location of events	23
2.5	Summary	25
3	Performance of Correlation Detectors and Subspace Detectors	26
3.1	Introduction	26
3.2	The Coso Sequence in Eastern California	26
3.3	Summary	33
4	Application of array-based waveform correlation detectors to events with complicated source-time functions	34
5	Control of Instrumental Timing Using Correlation Detectors	40
5.1	Instrumental Timing: Introduction	40
5.2	Instrumental Timing: Examples	42
5.2.1	Timing Example: Rana earthquake swarm	42
5.2.2	Timing Example: The KBS station	44
5.3	Instrumental Timing: Summary	49
6	Size of the Correlation Footprint	51
6.1	Correlation Footprint: Introduction	51

6.2	A Reciprocity Argument	52
6.3	An Example from a Marine Seismic Profile	54
6.4	Correlation Footprint: Summary	57
7	Characteristic Function Correlation Detectors	59
7.1	Introduction	59
7.2	The February 21, 2008, Svalbard earthquake and aftershock sequence	60
7.3	Characteristic Function Correlation Detectors: Summary	66
8	Autonomous Correlation Detection Frameworks	67
8.1	Motivation for Developing an Autonomous Detection Framework .	67
8.2	Proposed Solution	68
8.3	Autonomous Framework: Implementation	71
8.4	Autonomous Framework: Test Cases	75
8.4.1	Test Case: Mt. St. Helens Sequence	75
8.4.2	Test Case: Orinda Sequence	78
8.4.3	Final test: 2003 San Simeon Earthquake	81
8.5	Autonomous Framework: Summary	90
9	Conclusions	92
	References	99

List of Figures

1.1	Detection using waveform correlation on the NORES array of a high frequency signal buried deep in the background noise	3
1.2	Location and waveforms from the M=7.7 October 8, 2005, Pakistan Earthquake.	4
1.3	A detection by correlation over KNET using an aftershock of the October 8, 2005, Pakistan event as a master event.	6
2.1	Approximate location of the Rana earthquake sequence relative to the four Fennoscandian IMS array stations and the Norwegian National Seismic Network	11
2.2	Selection of a detection threshold for a correlation detector on the NORSAR array	13
2.3	Preliminary detections between Jan 1, 2005, and Dec 31, 2005	15
2.4	Detection of an $m_b = 1.4$ aftershock to an m_b 2.4 earthquake on April 28, 2005	19
2.5	Recordings of the April 28, 2005, Rana sequence recorded at the STOK station at Stokkvågen.	20
2.6	Magnitude estimates using independent amplitude measurements on the STOK and STOK1 stations and derived estimates from the NORSAR array	22
2.7	Relative location estimates obtained for 8 events in the Rana region of northern Norway using the Lg double-difference formulation of Schaff & Richards (2004)	24
3.1	Location of the NVAR array in western Nevada and the 2001 Coso sequence in south eastern California.	28

LIST OF FIGURES

3.2	Screen shots from a new tool showing several of the steps in constructing a subspace detector for the Coso sequence with data from the NVAR array.	29
3.3	Detail maps showing the spatial distribution of training events (green), detected events (red) and undetected events (grey)	31
3.4	Histograms of detected events in this sequence in black for the subspace detector (left) and the correlator (right)	32
4.1	ARCES waveforms showing seismic and acoustic signals from presumed military explosions near the northern coast of the Kola Peninsula	36
4.2	Automatic (GBF) location estimates for the events displayed in Figure 4.1 (left) and for events known to come from the mines at Zapoljarni (right).	37
4.3	Detection using waveform correlation of an event on the Kola Peninsula	38
5.1	Detecting and Measuring Timing Anomalies Using Waveform Correlation	41
5.2	Misalignment of single channel cross-correlation traces on the FINES array	43
5.3	Location of the March 5, 2006, Novaya Zemlya event in relation to the stations SPITS, ARCES, and KBS.	45
5.4	Signals from the March 5, 2006, Novaya Zemlya event on the KBS station.	46
5.5	Detection of the first identified Barentsburg rockburst following the March 5, 2006, Novaya Zemlya event	47
5.6	Measurement of the timing error on KBS from a sequence of cross-correlations of Barentsburg events	48
6.1	Geometry of the ARCES and detection by correlation of signals on single sensors using the signal on the central ARA0 element as a waveform template	53
6.2	Locations of shots in the PETROBAR 1 marine seismic profile . .	55

LIST OF FIGURES

6.3	Waveforms and cross-correlation traces with a template from the shot 121 signal	56
7.1	Location of the February 21 Svalbard event in relation to the SPITS and ARCES arrays and the 3-component stations HSP and KBS, at Hornsund and Ny Ålesund respectively	61
7.2	Mock helicorder plot of data from SPA1_BHZ, 23 February, 2008 .	62
7.3	Construction of an “incoherent matched filter detector” for ARCES from the February 21, 2008, Storfjorden main event	64
7.4	Detections of aftershocks using coherent and incoherent correlation detectors	65
8.1	Autonomous Detection Framework	70
8.2	Diagram of the final system configuration.	72
8.3	Twenty-five hours of data from station STD on the northwest flank of Mt. St. Helens	75
8.4	Detail of an hour of the Mt. St. Helens data	76
8.5	Waveforms from twenty-one events in one representative cluster after the reclustering operation for the Orinda Sequence	80
8.6	San Simeon main shock and aftershocks in relation to the NVAR array	82
8.7	A map of the ANSS catalog events in the vicinity of the San Simeon main shock (grey) and the events detected automatically by our detection framework (red).	84
8.8	Histograms of the ANSS catalog events against magnitude (grey) and the events detected by our system (red).	85
8.9	Example waveforms of detections associated with one detector. .	87
8.10	Two of the larger clusters of San Simeon aftershocks.	88
8.11	Two of the larger clusters of detections unrelated to the San Simeon sequence.	89

Chapter 1

Introduction

Waveform correlation detectors compare a signal template with successive windows of a continuous data stream and report a detection when the correlation coefficient, or some comparable detection statistic, exceeds a given threshold. Since these methods exploit characteristic details of the full waveform, they provide exquisitely sensitive detectors with far lower detection thresholds than typical STA/LTA algorithms. The drawback is that the form of the sought after signal needs to be known quite accurately a priori which limits such methods to instances of seismicity whereby a very similar signal has already been observed by every station used. Such instances include earthquake swarms, aftershock sequences, repeating industrial seismicity, and many other forms of controlled explosions. The reduction in the detection threshold is even greater when the techniques are applied to multiple channels since stacking can be performed on the correlation coefficient traces with a significant array-gain. The time-difference between a signal from one event and the signal from a subsequent co-located event will be identical for all stations in an arbitrarily spaced network and this means that the correlation coefficient traces can be stacked coherently even when there is little or no similarity between the actual signals at the different sites.

An example of the detection by waveform correlation using data from a seismic array is displayed in Figure 1.1. A signal with high signal-to-noise ratio (SNR) was generated by the detonation of 10000 kg TNT equivalent in a granite chamber. This waveform was used as a template to detect a signal from the detonation of 500 kg TNT in the same chamber. No signal from this smaller event can be

observed in the waveform data, and any signal present is buried so deep in the background noise that even waveform correlation on the single channels is not sufficient to detect it. The stacking procedure reduces all the unrelated peaks and troughs through cancellation of the correlation coefficient traces at all time except one: the moment marking the start of the sought after signal from the explosion. This situation is ideal for the use of the matched filter detector; the source is simple, consisting of a single explosive force, and co-located exactly with the master event.

The majority of event detection scenarios within nuclear explosion monitoring, and detection seismology in general, will be far less favourable. The Green's function for a given seismic source varies rapidly as a function of source location and a given waveform template is likely to have a very limited region of applicability (*correlation footprint*). Variable source-time functions, for example due to ripple-fired explosive sequences, are likely to result in highly diverse waveforms which may not correlate well even if the events are essentially co-located.

Correlation detectors combine the operation of detection with that of source identification which makes them highly desirable as event classifiers. One of the greatest challenges for data centers is the occurrence of large earthquakes followed by extensive aftershock sequences. The manual review of up to many thousands of resulting events can cause long delays in the production of seismic bulletins and divert analyst time away from potential events of monitoring interest. An example of such an event is the devastating October 8, 2005, Pakistan earthquake (Figure 1.2). This event was followed by many hundreds of aftershocks covering a wide geographical region and a wide spectrum of magnitudes. An ultimate aim is for correlation detectors to classify this event hierarchy semi-automatically giving, with a high level of confidence, time, coordinates, and magnitude, of each aftershock. Events not related to this tectonic activity could then be identified rapidly and given an appropriate priority in the event screening process.

The signal from the main October 8 event was used as a template to detect aftershocks in a correlation detection procedure, but found essentially nothing. It is likely that the magnitude of the event and the physical dimensions of the source rupture were so large that the resulting waveforms did not provide a good

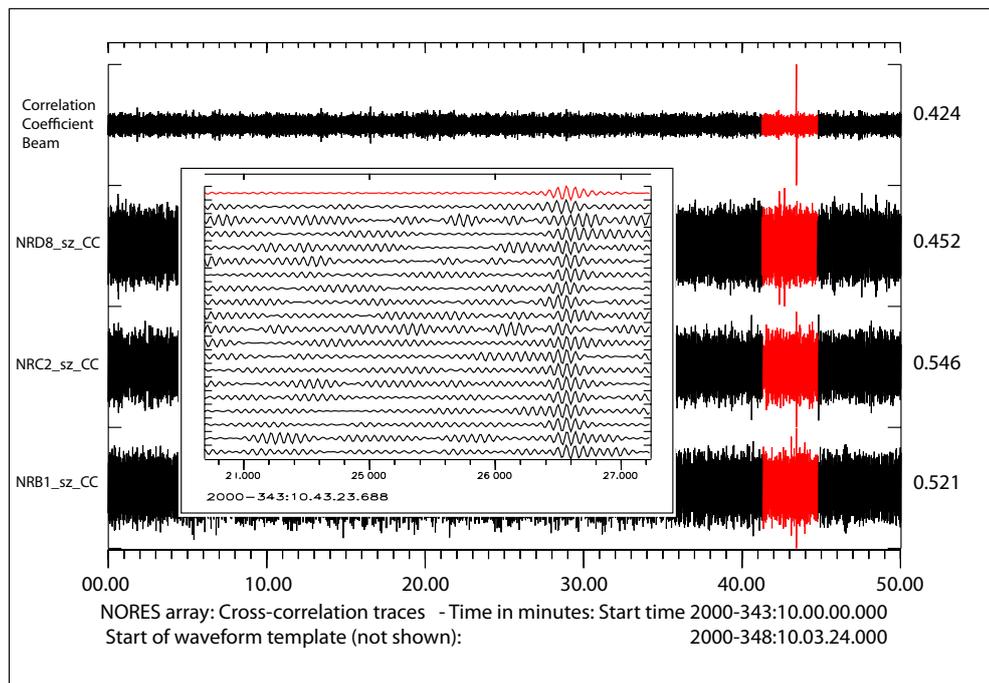


Figure 1.1: Detection using waveform correlation on the NORES array of the signal from a 500 kg TNT explosion in a non-spherical cavity, using the high-SNR signal from a 10000 kg TNT explosion in the same cavity as a master event. Note that the signal is not detectable on a single channel (the waveform correlation traces frequently do not even show a local maximum at the correct time). However, the zero-delay stacking results in a clear detection on the correlation coefficient beam. The events are described in [Stevens *et al.* \(2006\)](#) together with waveforms and waveform spectra for the larger events.

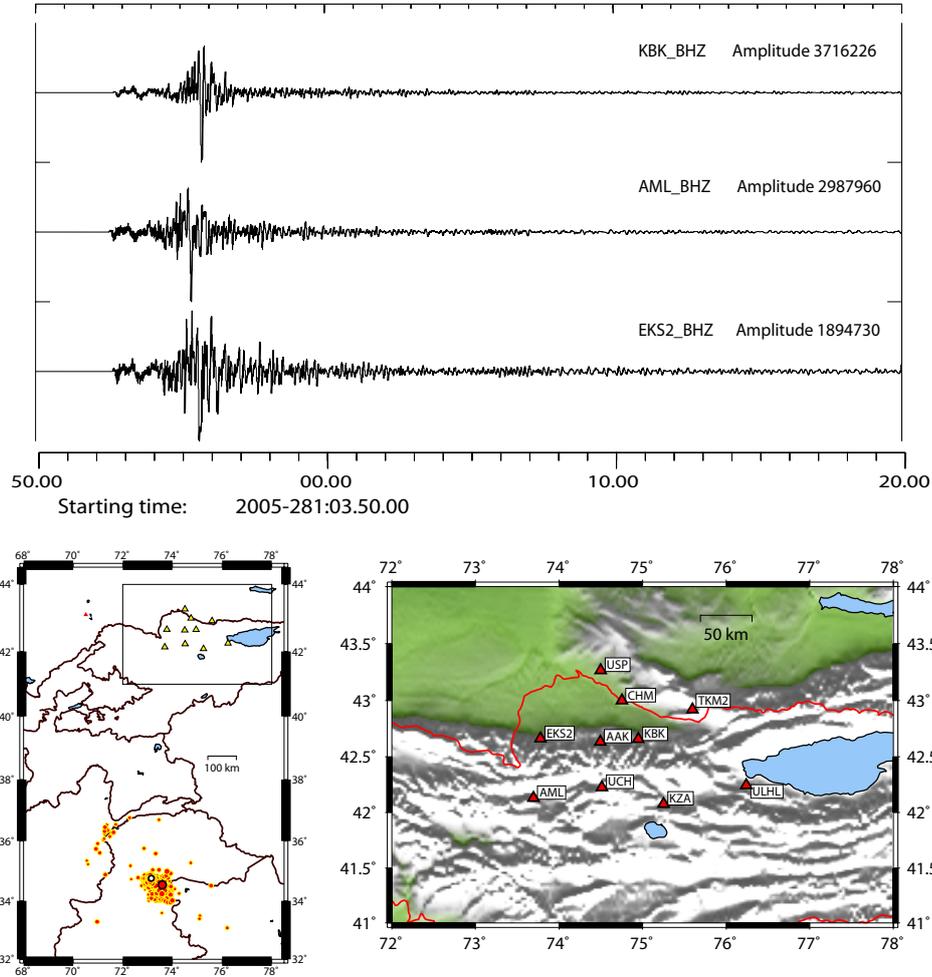


Figure 1.2: Location of the $M=7.7$ October 8, 2005, Pakistan earthquake together with locations from the USGS/NEIC database of subsequent events within 300 km of the main shock epicenter (between October 8, 2005, and December 31, 2005). The waveforms shown are recorded on selected instruments of the KNET network shown in relation to the earthquakes in the lower left panel and in more detail in the lower right panel.

representation of the Green's functions for any particular source location. However, there are many instances where the signals from aftershocks have detected numerous subsequent events (see Figure 1.3) by the correlation procedure. This suggests that, within the bands of diffuse seismic activity, there are pockets of repeating seismicity which are well characterized by a single template waveform. Figure 1.3 is also a superb illustration of how coherent array processing can be achieved over a large seismic network with little or no similarity between the waveforms on the different sensors. It should also be noted that this "network correlation beam" is constructed from correlation coefficient traces from all three components at each site.

Under existing procedures, the setting up of new correlation detectors is labor-intensive and has to be done on an event-by-event basis. In an adaptive and autonomous framework, the detection process displayed in Figure 1.3 would have been generated automatically. The system requires a correlation-type detector for the characterization of each new occurrence of seismicity and, in the absence of a detector specific to the waveform displayed, this would have been spawned as a result of an STA/LTA detection of a hitherto unobserved signal.

Due to the relatively short time that waveform correlation methods have been in routine operation within detection seismology, there are many questions regarding the sensitivity and applicability of correlation detectors which have barely been addressed. Whilst the primary aim of this project has been the very issue of automation, we have dedicated a number of chapters to a number of related issues: all necessary to an understanding of the circumstances under which correlation detectors can be applied effectively.

Chapter 2 considers a sequence of earthquakes in northern Norway which are recorded by several IMS seismic arrays at distances over 600 km. We attempt to detect these by multichannel waveform correlation using the signals from the largest event (magnitude 3.5 on June 24, 2005) as waveform templates. The events are also recorded on network stations at far smaller epicentral distances (~ 15 km) which provides independent confirmation and relatively accurate estimates of event magnitude. The local recordings serve as a form of Ground Truth for assessing the performance of the correlation detectors on different arrays, at different distances, and with different template parameters.

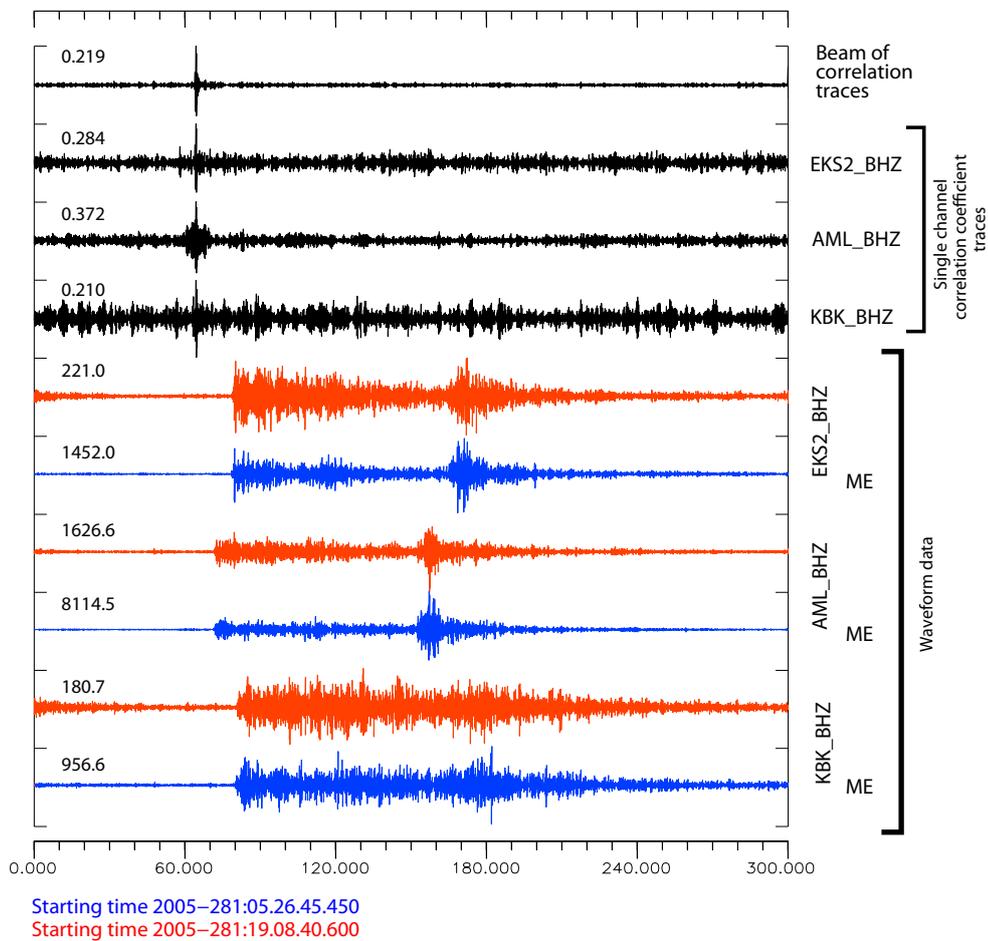


Figure 1.3: A detection by correlation over KNET using an aftershock of the October 8, 2005, Pakistan event as a master event. All waveforms are filtered between 2.5 and 8.0 Hz and the waveform template has a length of 180.0 seconds beginning at a time 2005-281:05.27.50.000. The USGS/NEIC event bulletin lists the master event as origin time 2005-281:05.26.05.120, latitude 34.760°N, longitude 73.150°E, depth 10.0 km and magnitude 5.5. The same bulletin lists the detected event with origin time 2005-281:19.08.00.490, latitude 34.800°N, longitude 73.170°E, depth 10.0 km and magnitude 4.9.

Chapter 3 examines an earthquake cluster in eastern California which is observed by the NVAR array at a distance of ~ 270 km. For these events, we have an excellent double-difference earthquake catalog in addition to recordings from a local station at a distance of only a few kilometers. We apply a correlation detector on a single master event and also a multi-dimensional correlation or subspace detector constructed using the waveforms of several master events. The performance of both types of correlation detector is evaluated against the earthquake bulletin.

In chapter 4 we examine briefly a sequence of near-surface explosions on the Kola Peninsula in north west Russia at a distance of some 250 km from the ARCES array. The correlation coefficients obtained between events (which are presumed to originate from the same source region) are not high. It is assumed that the differences between the waveforms are due primarily to different source-time functions for the different events. Examining the alignment of the single channel cross-correlation traces, in addition to the values attained, suggests that array-based waveform correlation can detect large numbers of these events and still with a very low false alarm rate.

Chapter 5 examines cases of demonstrably erroneous instrumental timing and describes how the operational use of correlation detectors is able to detect, and in some cases correct, the timing error.

Chapter 6 examines briefly the question of how far apart two events may be such that one can still be detected using the signals from the other as a waveform template. Due to the number of issues involved (source type, event magnitude, tectonic setting, frequency content of signals etc.) this question alone could be the subject of an entire research project. However, with two case studies using high-frequency signals from well-constrained regional events, we are able to obtain presumably quite generic results. Firstly, we invoke an argument of reciprocity and examine to what degree the signals from sites in a seismic array can be used to detect the signals on other sites, at varying distances, using a one-channel correlation detector. Secondly, we examine signals from a marine seismic profile recorded on the ARCES array to assess the detectability using waveform correlation on one and many channels as a function of the distance between events.

Chapter 7 examines to what extent we can expand the correlation footprint by considering characteristic transformations of waveform data on arrays, rather than the raw data itself, as input to a matched filter detector. The motivation comes from extensive earthquake aftershock sequences where the waveform from the main event fails to correlate with the signals from aftershocks, and therefore does not constitute a useful correlation detector template. It is suggested that the signal envelopes from different bandpass filterings of the waveforms, together with semblance measures for different beam-steering parameters, constitute more representative characteristics of the master event than the waveforms themselves and that these so-called incoherent correlation detectors may detect occurrences of hitherto unobserved signals which can subsequently be added to a pool of correlation detectors.

Finally, chapter 8 discusses the development of an adaptive framework in which correlation detectors can be generated as required to characterize the observed seismicity. Three case-studies are selected. The first considers data from a single station close to Mt. St. Helens. This is a straightforward example where many events, quite well separated in time, are easily detected using traditional STA/LTA type detectors. The second case-study considers an aftershock sequence near Orinda in southern California, observed at close range. The final test case considers an extensive aftershock sequence from the December 22, 2003, San Simeon earthquake, also in California. The test cases increase in complexity, both in terms of the seismicity examined and of the techniques required to carry out effective monitoring. In the first case, we consider almost repeating seismic events confined to a very limited geographical region and monitored by a single seismic sensor. In the final test case, we consider an aftershock sequence covering a relatively wide geographical region and which is monitored using array processing at regional distances.

Chapter 2

Detection Thresholds for Multichannel Correlation Detectors

2.1 A sequence of earthquakes in northern Norway

The Rana region of Norway is the site of constant intraplate seismicity (Hicks *et al.*, 2000) and, on June 24 2005, was the site of an $m_b = 3.5$ event which was well recorded at all the Fennoscandian International Monitoring System (IMS) array stations in addition to the National Seismic Networks of Norway and Finland (Figure 2.1). An $m_b = 2.4$ event on April 28, 2005, had been located to almost the same location and subsequent analysis of the waveforms recorded at regional distances indicated very high correlation coefficients between the two events at frequencies up to 10 Hz. The high waveform similarity indicates that the spatial separation between the events is very small (e.g. Geller & Mueller, 1980) and waveform templates were extracted from the June 24 event to attempt to detect occurrences of smaller events from this location which were not detected using traditional array processing. We focused upon the NORSAR array at a distance of approximately 600 km from the earthquake epicenters. The large inter-site distances on this teleseismic array preclude the effective processing of high-frequency

2.1 A sequence of earthquakes in northern Norway

regional phases using traditional array methods for small-aperture arrays due to the lack of waveform similarity between sites. However, as was demonstrated by [Gibbons & Ringdal \(2006\)](#), the correlation coefficient channels are coherent over arbitrarily spaced networks and arrays even when the waveforms are not (the condition of waveform similarity between sites is replaced by a condition of waveform similarity between events).

2.1 A sequence of earthquakes in northern Norway

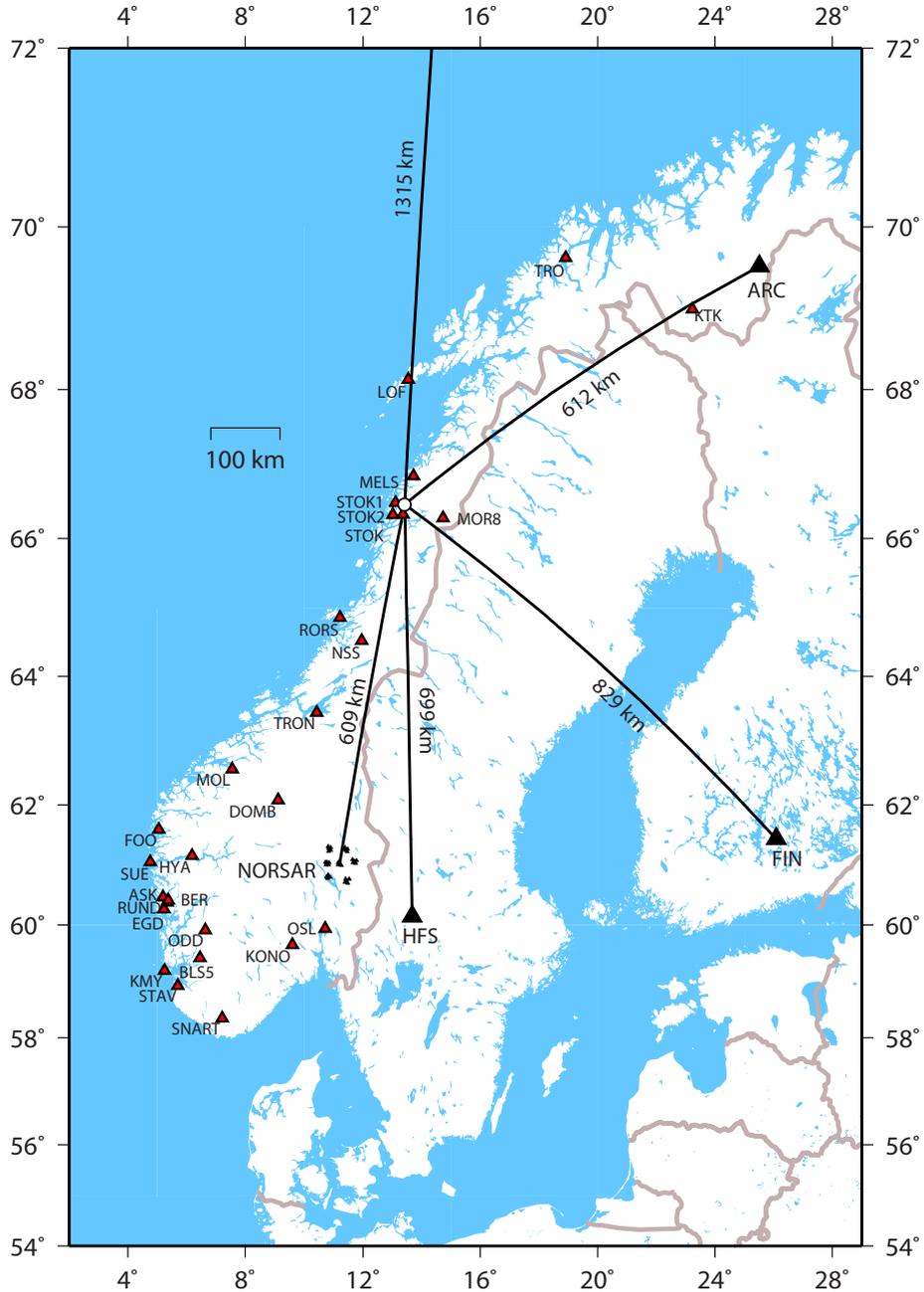


Figure 2.1: Approximate location of the Rana earthquake sequence relative to the four Fennoscandian IMS array stations (black symbols) and the seismic stations of the Norwegian National Seismic Network (NNSN: red triangles). The line leading Northwards leads towards the SPITS IMS seismic array on the island of Spitsbergen.

2.2 Setting the detection threshold

As for any detector, we need to declare the circumstances under which a detection is defined. For the correlation detector, we seek significant values of the normalized correlation coefficient or, in the array situation, the array correlation beam. Gibbons & Ringdal (2006) defined a *scaled correlation coefficient* which, analogously to an SNR, indicates the ratio between the correlation coefficient at a given time and the background level at surrounding times. (A similar quantity has subsequently been demonstrated by Schaff, 2008, to provide a very stable detection statistic for correlation detectors on 3-component stations.) For the current investigation, the values of the (unscaled) correlation beam were examined for many different data segments, of which three are displayed in Figure 2.2. The uppermost panels in Figure 2.2 correspond to a data segment in which no signals are observed. The correlation coefficient for a single channel does not exceed ± 0.14 and the zero-delay stacking over the 42 sites of the NORSAR array reduces the variability of the correlation beam to within ± 0.02 over the 10-minute interval. The linearity of the quantile-quantile plot (upper right panel in Figure 2.2) indicates that the values of the correlation beam are almost perfectly normally distributed for this interval. The normal distribution with this standard deviation appears to be quite typical for data segments without detections. The presence of any seismic signal appears to complicate matters and an example featuring a completely unrelated regional signal is displayed in the central panels of Figure 2.2. The foreign signal leads to a modulation of the correlation coefficient traces and the quantile-quantile plot indicates a departure from the normal distribution and an increase in the extreme values observed. Similar plots were observed for large numbers of different signals; the highest values of the correlation beam and the greatest departures from a normal distribution observed were for regional signals from the Rana region.

For all the data segments examined, with the exception of the times of events known to correlate well with the waveform template, the correlation beam was never observed to exceed a value of 0.03. The signal and correlation coefficients corresponding to the time of the April 28, 2005, earthquake are displayed in the lowermost panels of Figure 2.2. The difference between the single channel

2.2 Setting the detection threshold

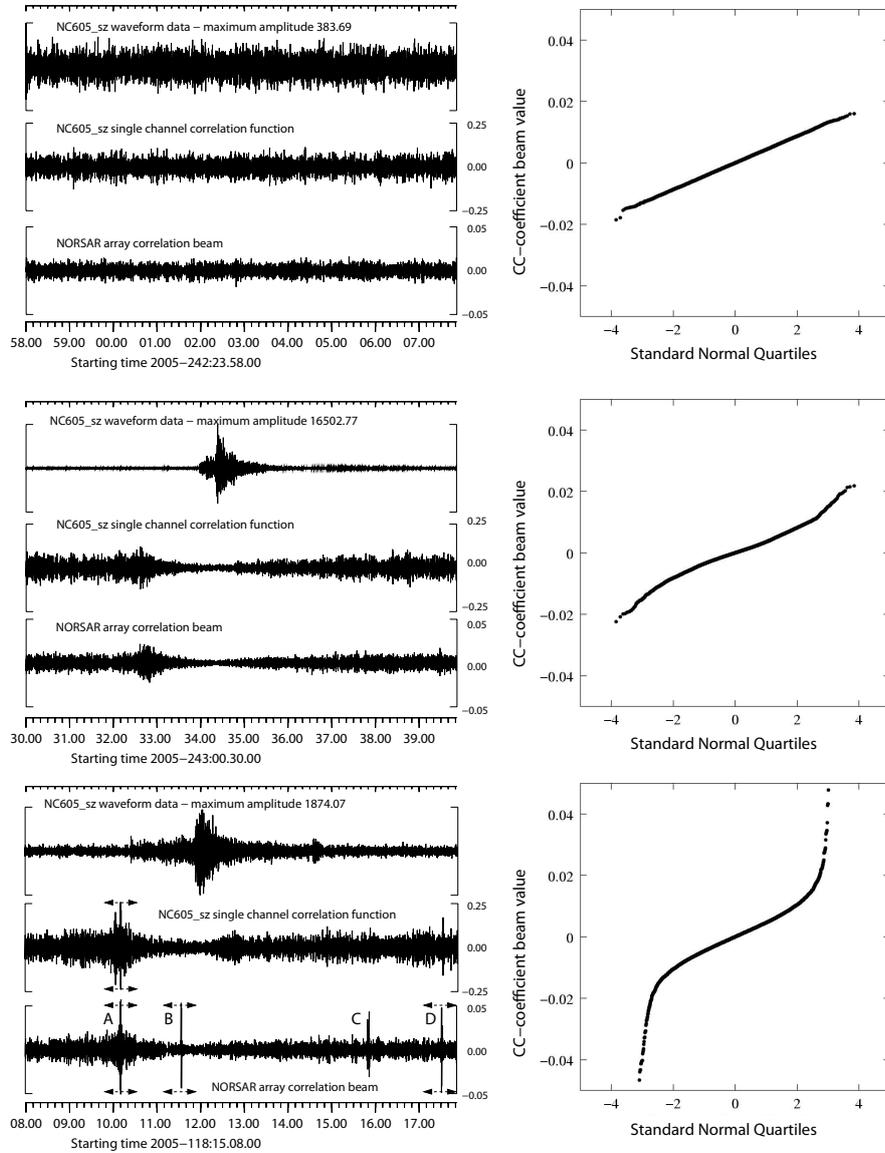


Figure 2.2: Selection of a detection threshold for a correlation detector on the NORSAR array where the template consists of 120.0 second long data segments of waveforms, bandpass filtered between 2.0 and 8.0 Hz, from the June 24 main event. The left panel in each row displays a ten-minute long filtered data segment on a single channel together with the corresponding correlation coefficient channel and the correlation beam. The right panel in each row shows the values of the correlation beam reordered and plotted against the standard Normal Quantiles. The correlation traces in the bottom row are clipped as indicated.

2.2 Setting the detection threshold

and full-array correlation coefficient traces is very clear with several clear peaks observed in the array beam which do not rise above the background levels in the single channel case. The nominal threshold value of 0.03 is exceeded at the times indicated by A, B, C, and D on the lowermost trace. On closer inspection, the correlation peak at time C actually consists of two distinct maxima separated by approximately 2.5 seconds. Whilst we expect 0.03 to constitute a robust detection threshold, we report initially all occasions on which a value of 0.025 is exceeded and, in addition, all occasions on which the scaled correlation coefficient (as defined by [Gibbons & Ringdal, 2006](#)) exceeds a value 6.0. On each occasion on which a local maximum is identified in the correlation beam, a detection reduction rule prevents any subsequent detection being made within 3.0 seconds of this time.

Figure 2.3 displays the fully normalized array coefficient for each of the 554 occasions during 2005 for which these provisional detection criteria were met. The value unity is obtained exactly once: for the time interval corresponding to the master waveform template. It is clear from the distribution of points in Figure 2.3 that any reduction in the detection threshold below the provisional value of 0.03 suggested by examination of the quantile-quantile plots in Figure 2.2 would result in a large increase in the number of detections. Since tests involving unrelated signals frequently resulted in coefficients exceeding 0.020, it is clear that a robust threshold must be set higher. On the basis of Figure 2.2 and Figure 2.3, it was deemed that 0.03 appeared to constitute a sensible threshold and all provisional detections where the correlation beam did not exceed 0.03 were discarded. This threshold was exceeded on 32 occasions during 2005, almost none of which corresponded to times when a signal detectable by traditional energy detectors was observed on the NORSAR array. These instances are listed in Table 2.1.

The first observation is that many of these detections are in quick succession of each other. Figure 2.4 displays the signals from the June 24, 2005, master event together with detected signals on April 28. Fortunately, in this study, the validity of correlation detections at the distant array stations can be verified using data from a 3-component station (STOK) located only some 15 kilometers away from the epicenter. Figure 2.5 confirms that each of the detections on the NORSAR

2.2 Setting the detection threshold

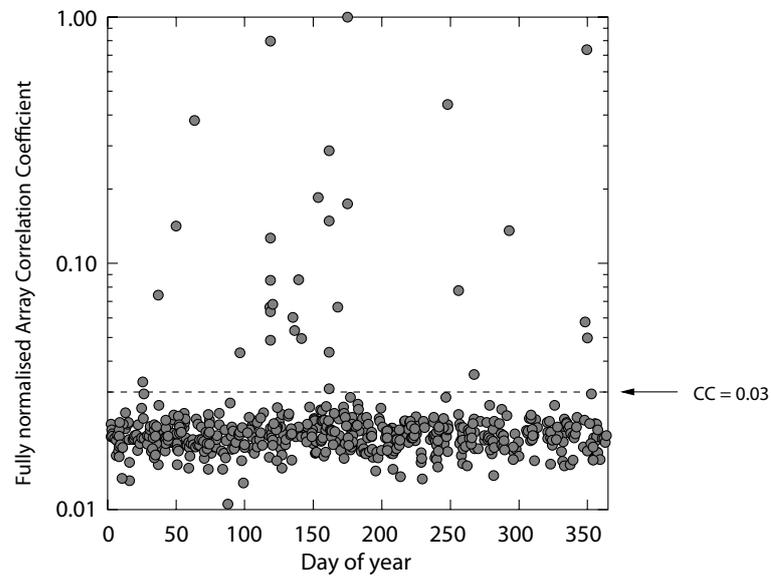


Figure 2.3: Preliminary detections between Jan 1, 2005, and Dec 31, 2005, for a matched filter detector on the NORSAR array where the template consists of 120.0 second long data segments of waveforms, bandpass filtered between 2.0 and 8.0 Hz, from the June 24 main event. Each detection was triggered by either an array correlation coefficient exceeding 0.025 or a scaled array coefficient exceeding 6.0.

2.2 Setting the detection threshold

Table 2.1: NORSAR correlation detections (see Figure 2.3) for which the array correlation coefficient exceeded 0.03. C.C. coef. is the fully-normalized array correlation coefficient and the scaling factor, α , is the scalar multiple of the master waveform which minimizes the residual when the detected waveform is subtracted (see Gibbons & Ringdal, 2006). The subsequent magnitude estimate is given by $M = \log_{10}(\alpha) + 3.5$ (where the value 3.5 is the magnitude attributed to the master event). The reference time, t_R , used is 2005-175:04.25.41.000.

Det.	Date	Julian time	C.C. coef.	Scaling factor.	mag.
1	Jan 25	025:16.47.29.719	0.0330	0.000037	-0.93
2	Feb 6	037:02.13.07.446	0.0740	0.000770	0.39
3	Feb 19	050:04.03.00.086	0.1411	0.001730	0.74
4	Mar 4	063:10.32.20.833	0.3796	0.007940	1.40
5	Apr 6	096:10.54.57.970	0.0432	0.000543	0.23
6	Apr 28	118:10.48.42.936	0.0662	0.000901	0.45
7	Apr 28	118:15.08.57.788	0.7970	0.044560	2.15
8	Apr 28	118:15.10.21.033	0.0854	0.004700	1.17
9	Apr 28	118:15.14.38.837	0.0487	0.000348	0.04
10	Apr 28	118:15.16.18.097	0.1267	0.001410	0.65
11	Apr 28	118:15.50.02.263	0.0636	0.000672	0.33
12	Apr 30	120:12.41.24.221	0.0682	0.000687	0.34
13	May 15	135:03.31.10.775	0.0602	0.000461	0.16
14	May 16	136:07.00.16.105	0.0534	0.000562	0.25
15	May 19	139:03.58.24.572	0.0859	0.000795	0.40
16	May 21	141:11.28.53.963	0.0493	0.000670	0.33

2.2 Setting the detection threshold

Table 2.1: continued.

Det.	Date	Julian time	C.C. coef.	Scaling factor.	mag.
17	Jun 2	153:14.07.49.892	0.1846	0.002650	0.92
18	Jun 10	161:15.39.30.817	0.2854	0.004390	1.14
19	Jun 10	161:16.25.34.695	0.1486	0.002140	0.83
20	Jun 10	161:16.39.01.043	0.0435	0.000481	0.18
21	Jun 10	161:17.46.26.336	0.0309	0.000250	-0.10
22	Jun 17	168:00.50.55.884	0.0664	0.000645	0.31
23	Jun 24	175:04.25.41.000	1.0000	1.000000	3.50
24	Jun 24	175:05.02.16.254	0.1741	0.002790	0.95
25	Sep 5	248:04.58.22.973	0.4419	0.007200	1.36
26	Sep 12	255:23.49.03.892	0.0774	0.000747	0.37
27	Sep 24	267:09.56.29.013	0.0354	0.000360	0.06
28	Oct 19	292:19.40.56.155	0.1355	0.001330	0.62
29	Dec 14	348:05.53.06.775	0.0577	0.000663	0.32
30	Dec 15	349:16.47.13.008	0.7363	0.165670	2.72
31	Dec 15	349:16.47.57.783	0.0497	0.010390	1.52
32	Dec 31	365:09.54.11.126	0.0443	0.000422	0.13

2.2 Setting the detection threshold

array corresponds to a small aftershock recorded by the STOK station. In the lower right panel of Figure 2.5, we even see that the small double correlation detection (labelled C in Figure 2.2) does indeed correspond to two distinct events which can just be resolved by the local station.

2.2 Setting the detection threshold

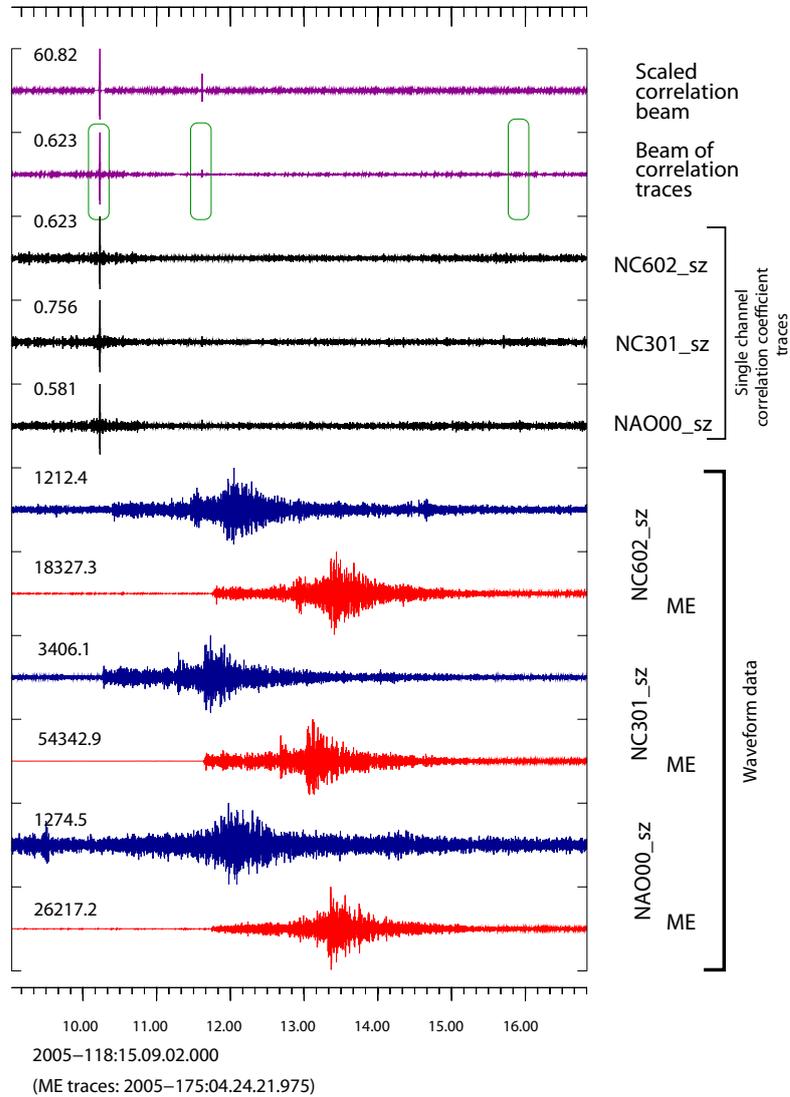


Figure 2.4: Detection of an $m_b = 1.4$ aftershock to an $m_b = 2.4$ earthquake on April 28, 2005, in the Rana region of Northern Norway using multichannel waveform correlation on the large aperture NORSAR array at a distance of over 600 km. The master event signals (red traces) result from a magnitude 3.5 earthquake on June 24, 2005, in the same region, and are aligned with the April 28 data (blue traces) according to the first aftershock detection. Only the main shock can be detected on the single channel correlation coefficient trace; the numerous aftershocks are only detected using array processing of the correlation coefficient channels. The green rings mark detections on the array correlation beam. The third detection comprises two distinct correlation maxima, both close to the detection threshold.

2.2 Setting the detection threshold

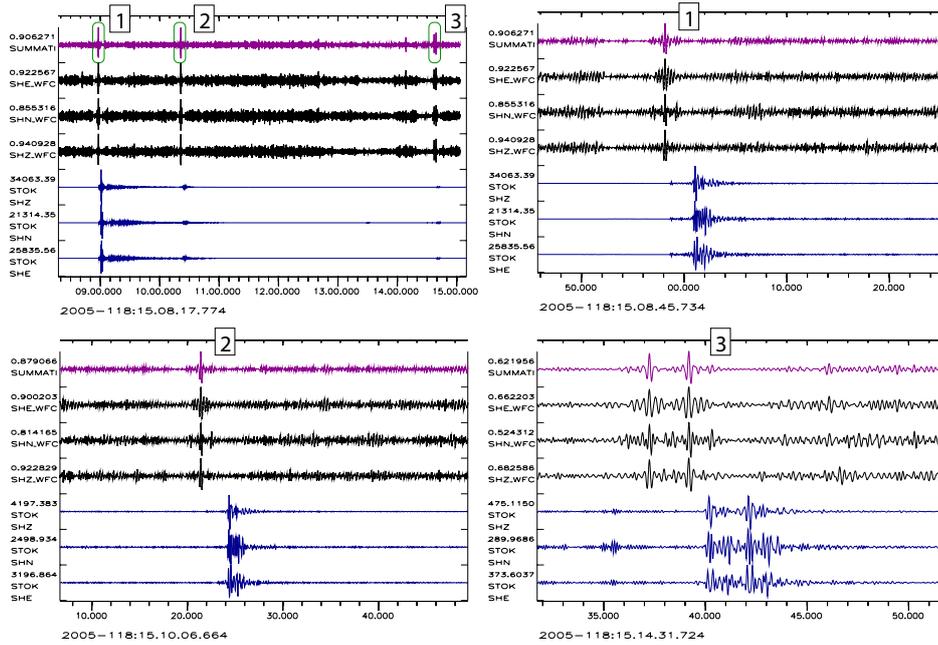


Figure 2.5: Part of the April 28, 2005, Rana earthquake sequence as recorded by the Stokkvågen 3-component station of the NNSN at a distance of approximately 15 km from the earthquake epicenter. The top left panel shows a seven minute long time window which (correcting for the difference in traveltimes) corresponds to the interval shown in Figure 2.4 for the NORSAR array. The remaining panels provide a close-up view at the times indicated. In each panel, the blue traces are STOK data from April 28, bandpass filtered between 2.5 and 8.0 Hz, the black traces show the componentwise correlation coefficients with filtered STOK data from the June 24, 2005, event (waveforms not shown), and the magenta trace indicates the mean of the 3 single component correlation traces.

2.3 Improvement in detectability over standard STA/LTA detectors

Using data from the STOK station, we confirmed that all detections in Table 2.1 except the first corresponded to real and verifiable events in the Rana region. Detections 7, 23 and 30 clearly correspond to the April 28, June 24 and December 15 events published in the reviewed regional NORSAR event bulletin. We can estimate event magnitudes for the correlated events by assigning a magnitude of 3.5 to the master event and estimating magnitudes for the remaining event using

$$M = \log_{10}(\alpha) + 3.5$$

where α is least squares scaling factor described by [Gibbons & Ringdal \(2006\)](#). The magnitude estimates of 2.16 and 2.73 for the April 28 and December 15 events are slightly lower than (albeit of the same order as) the respective GBF magnitude estimates of 2.51 and 3.03.

Due to the high SNR of the signals at the local STOK station, we can estimate the event magnitudes directly by comparing maximum waveform amplitudes. The event magnitudes estimated from the NORSAR detections are plotted as a function of the event magnitudes estimated from STOK data in Figure 2.6. We also show independent estimates of the magnitudes inferred from maximum waveform amplitudes at the STOK1 station, which was deployed in June 2005. The NORSAR estimates have a tendency to underestimate the magnitude, but show a reasonable linear relationship with the magnitude estimates inferred from the local station. We conclude that using the signal from a magnitude 3.5 event as a master waveform has allowed us to detect almost co-located events down to magnitude 0.5 on array data from stations at distances of over 600 km. For the ARCES array, this represents an improvement of approximately one unit of magnitude over the detection threshold possible using standard energy detectors. The improvement for the large-aperture NORSAR array is even greater (approximately 1.5 units) since incoherence of the high-frequency regional signals essentially precludes coherent array processing for these events.

2.3 Improvement in detectability over standard STA/LTA detectors

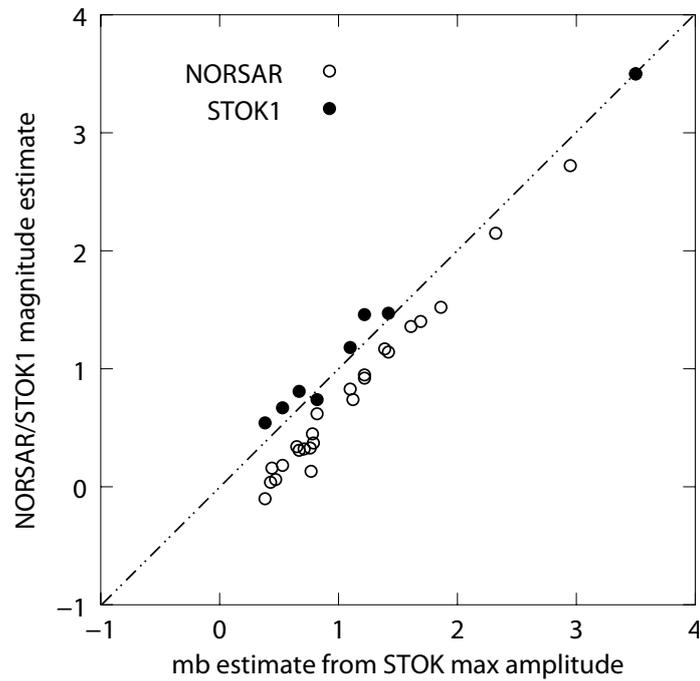


Figure 2.6: Comparison of magnitudes estimated using maximum amplitudes at the STOK station and those obtained using independent measurements on amplitudes at the temporary STOK1 station and correlation and least squares inversion at the NORSAR array. The magnitude of the main June 24 event is fixed to 3.50.

2.4 Relative location of events

Central to the theme of detection through waveform similarity, is the question of spatial separation of events and the ability to provide accurate relative locations (see [Richards *et al.*, 2006](#)). The stacking of correlation coefficient traces over a seismic array lowers the detection threshold and, for the same reasons, provides improved cross-correlation time estimates, especially in cases where a single station observation would be insufficient. If an event cluster is recorded by a network of seismic arrays, there may be a possibility that events of lower magnitude may be included in inversions for relative location estimates using double-difference type algorithms. The events described in [Gibbons *et al.* \(2007a\)](#) were recorded by both array stations at regional distances and 3-component stations at local distances. The local network recordings were unfortunately too incomplete for a full double-difference relative relocation. However, a stacking of correlation coefficient traces for the Lg-phase over each of the seismic arrays, together with an application of the [Schaff & Richards \(2004\)](#) DD-algorithm, allowed for a stable solution (Figure 2.7) - consistent with the existing local recordings - despite the low SNR of the events at regional distances.

2.4 Relative location of events

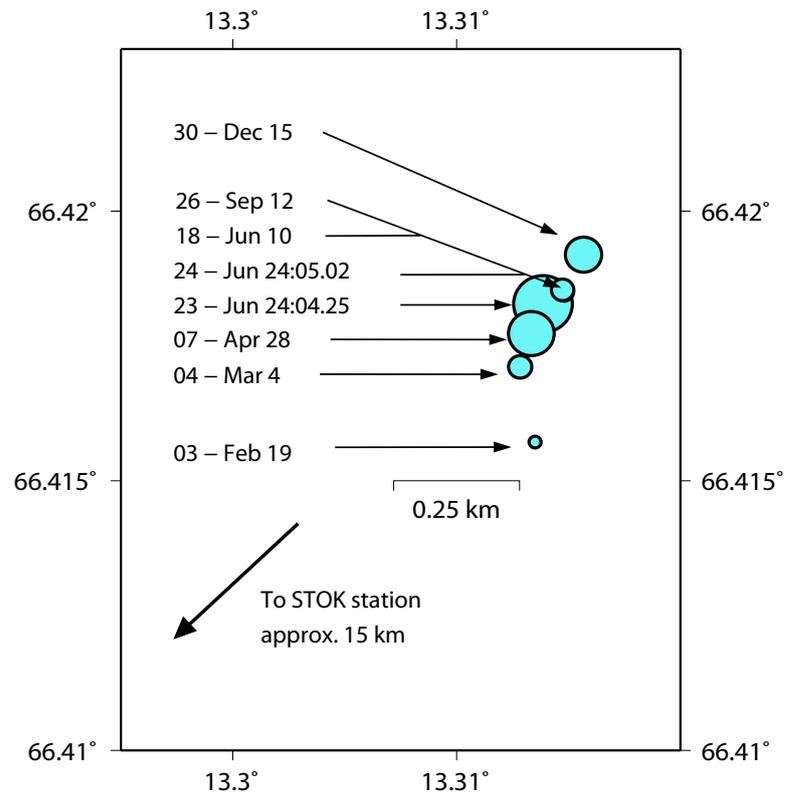


Figure 2.7: Relative location estimates obtained for 8 events in the Rana region of northern Norway using the Lg double-difference formulation of [Schaff & Richards \(2004\)](#) based upon delay times indicated by correlation coefficient beams at four seismic arrays at distances over 600 km. See [Gibbons *et al.* \(2007a\)](#) for details.

2.5 Summary

A fortuitous sequence of earthquakes in the Rana region of northern Norway has allowed us to examine the potential of detection at regional distances using multi-channel waveform correlation. The sequence was useful for a number of reasons. Firstly, a 3-component station of the Norwegian National Seismic Network was located only 15 km from the events, providing confirmation of the occurrence of detected events together with an independent estimate of event magnitude. Secondly, the sequence was essentially equidistant from the very large aperture NORSAR array and the small-aperture regional ARCES array. Based upon correlation detections at the arrays, and observations at local stations, we conclude that a master event of magnitude 3.5 has allowed us to detect aftershocks down to magnitude 0.5 at distances of over 600 km.

The detection capability for the correlation detector was approximately the same for the ARCES and NORSAR arrays. The detection threshold using standard procedures is considerably higher for the NORSAR array than for ARCES due to waveform incoherence over the larger inter-site distances. Since waveform dissimilarity is not an issue for correlation detectors, we point out that the improvement in detectability is even greater for large arrays or seismic networks.

It should be noted that the signals from many events are hidden within the coda of preceding events on the recordings made at regional distances. The local network stations confirm these events to be distinct earthquakes. There is no evidence from the local stations of any aftershocks within ten minutes following the main June 24, 2005, master event and we are quite confident that the waveform template for this event is free of interfering signals from this region. The ability of the correlation detectors at regional distances to detect these *hidden* events is quite impressive, as it the ability to resolve two events with less than 3 seconds separation.

This investigation is described fully in [Gibbons *et al.* \(2007a\)](#).

Chapter 3

Performance of Correlation Detectors and Subspace Detectors

3.1 Introduction

Whilst the event sequence described in chapter 2 provided us with a good overview of the improvement in detection capability which could be attained using waveform correlation methods, it must be remembered that this sequence is an almost ideal situation for the matched filter procedure and that many, if not most, occurrences of seismicity will provide additional challenges in terms of waveform diversity.

3.2 The Coso Sequence in Eastern California

Eastern California was selected as a quite different tectonic setting in which to evaluate correlation detector performance. Studies in this region benefit from the double-difference catalogue produced by [Hauksson *et al.* \(2003\)](#), which provides good relative event locations important to understanding how event waveform correlations are influenced by hypocenter proximity. The location of one cluster of seismic events, approximately 270 kilometers south of the NVAR array (Mina,

3.2 The Coso Sequence in Eastern California

Nevada), is displayed in Figure 3.1. This cluster of 406 events is part of a larger sequence just north of the geothermal region in Coso, southeastern California. The events in the smaller cluster occurred in 2001 (Julian days 195-197) and had reported magnitudes ranging from 0.4 to 3.9. A local broadband station, JRC, within 5 kilometers of the cluster provides further ground truth information on the sequence.

Examination of the waveforms from these events made it clear that even this relatively limited cluster would not be able to be characterized by a single waveform template in the way that the Rana sequence in the previous chapter could. One option is to initiate a number of distinct correlation detectors, each using a waveform from a different master event. Another is to construct a multi-dimensional correlation detector, a so-called subspace detector (Harris, 1989, 2006; Harris & Paik, 2006).

A set of 29 events with magnitudes above 2.8 were used to characterize the events in the cluster - these are shown in Figure 3.1 as green crosses. They were drawn from a longer time interval (days 195-211), but included many of the larger events in the detection interval. Waveforms from the 29 events were extracted from the NVAR data stream; a sample of these, recorded at the array site NV01, are displayed in Figure 3.2. The figure shows screenshots from a new tool written to select and align waveform data, and compute subspace representations with user-selected parameters (filter band, representation dimension, etc.). Besides a waveform display, used to screen bad events and channels, the tool has a correlator that aligns the waveforms, and a dendrogram display allowing the user to set clustering thresholds and select clusters. It also has a series of subspace panels that display the fidelity of representation (energy capture, shown in the figure) of waveforms and the probability of detection (at a fixed, selectable false alarm probability) both as a function of the subspace dimension. These analytical tools allow the user to select a subspace (correlator operator) dimension. The tool finally allows the user to write a detector definition file which can be used by a separate code to run the detector against a continuous (multichannel) data stream. In this instance, an 8-dimensional detector was constructed and applied against the 9 channels of continuous NVAR data for the 3-day interval 2001:195-197.

3.2 The Coso Sequence in Eastern California

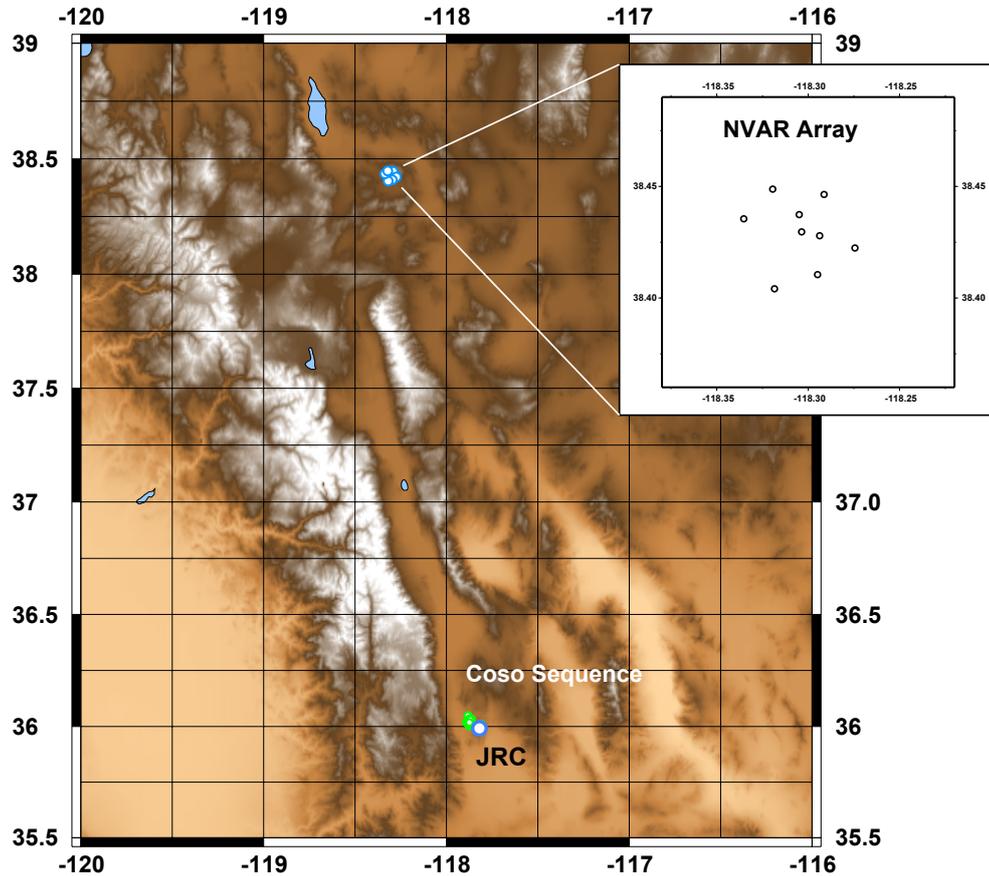


Figure 3.1: Nine elements of the array were used to construct and apply detectors. The other elements of the array had significant data gaps or poor signal to noise ratio. The Coso sequence and two others in eastern California were chosen for study because of excellent catalog information. The Hauksson catalog (Hauksson *et al.*, 2003) provides double-difference locations for events in the Coso sequence. While the overall sequence consisted of more the 3500 events (as reported in the catalog), 406 occurring in a short time interval 2001: 195-197 and in a fairly compact region were selected for study. Reported catalog magnitudes ranged from 0.4 to 3.9 for these events.

3.2 The Coso Sequence in Eastern California

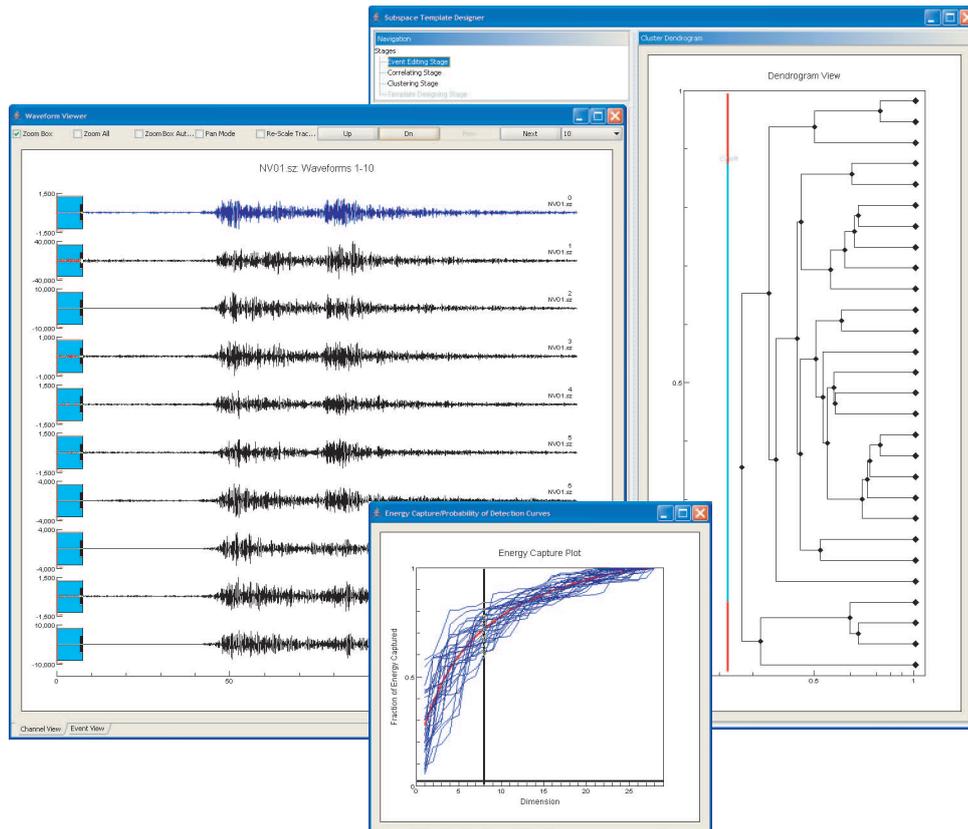


Figure 3.2: Screen shots from a new tool showing several of the steps in constructing a subspace detector for the Coso sequence with data from the NVAR array. A waveform display tool (at left) allows the user to screen bad events and bad channels. A correlator aligns the waveforms and provides measurements for building a dendrogram display (at right). The dendrogram allows the user to define and select event clusters. A series of subspace panels (one shown at center: energy capture as a function of subspace dimension) allow the user to select a window on the aligned waveforms and choose a subspace dimension.

3.2 The Coso Sequence in Eastern California

The results of the detection operation are shown in Figures 3.3 and 3.4. Figure 3.3 provides maps of the 406 events comprising the cluster in question. The green crosses indicate the reference events, the red circles indicate detected events and the grey circles indicate undetected events. Results for the 8-dimension multichannel subspace detector are shown at the top: 120 events were detected after adjusting the detection threshold (to 0.03) to allow only 2 presumed false alarms (defined as events not reconcilable against the Hauksson catalog). To the left in Figure 3.4 is a histogram (in black) of the 120 detections reconciled against the catalog as a function of catalog magnitude. Behind the histogram of detections is another histogram (in red) of all the catalog events. Comparison of the two histograms makes it clear that the subspace detector missed at least half the events between magnitudes 1.5 and 2.2, and two magnitude 3.2 events (these were in coda of still larger events). For comparison, a multichannel correlation detector was developed using one event (shown as the green cross in the bottom of Figure 3.3) and a similar detection operation was performed. Adjusting the detection threshold (to 0.0125) to allow the same number (2) of unreconciled triggers as in the subspace detector case, this detector found 95 events distributed about as widely as for the subspace detector. The correlation detector threshold was substantially lower than the subspace detector because of the fact that the 8 basis functions of the subspace detector significantly raise the noise floor of the detection statistic (while simultaneously raising the values of the statistic for actual events). We can only speculate about the reasons for the relatively poor performance of both detector types in this case (as opposed to almost complete detections in other instances). From cursory examinations of the three-component signals observed at the local station JRC, it does appear that the mechanisms of the events not detected are significantly different from the larger events used for design. One hypothesis consistent with the observations is that the larger events have principal axis orientations aligned with the dominant stress field in the region, and thus have relatively uniform mechanisms. This hypothesis would explain neatly why the correlation detector performed nearly as well as the subspace detector designed from 29 events: the additional 28 events would add relatively little information to the representation relevant to the smaller events. In this view, the smaller events might have relatively random orientations.

3.2 The Coso Sequence in Eastern California

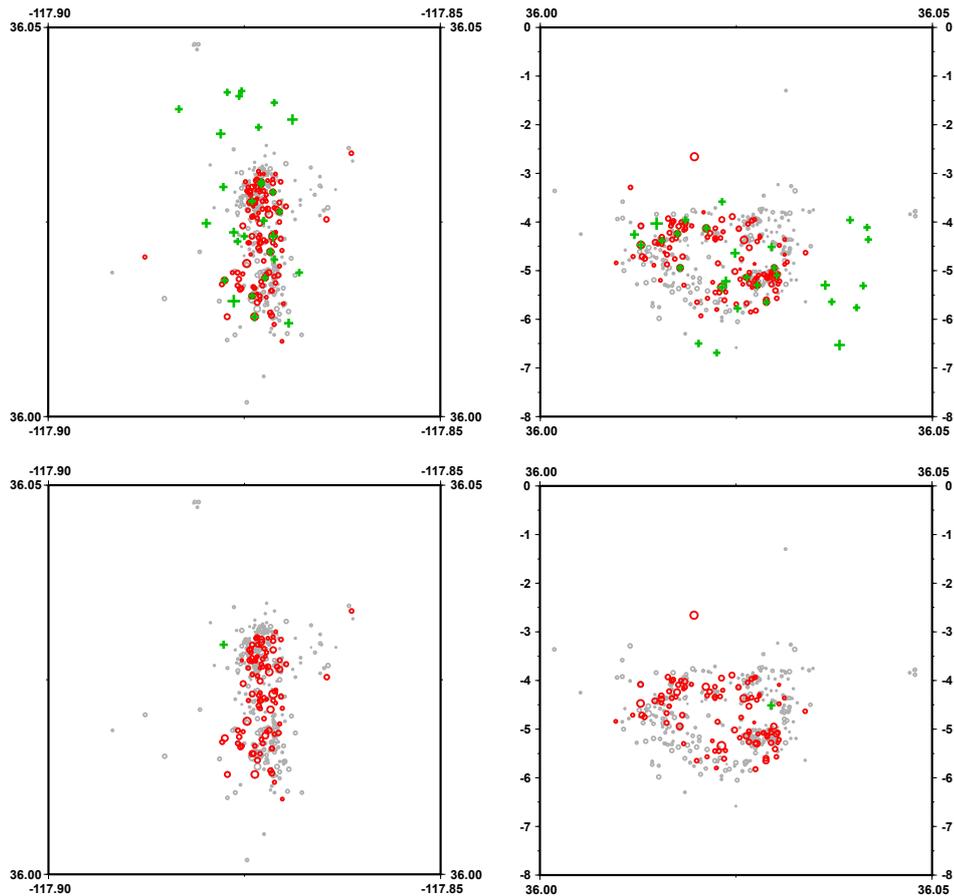


Figure 3.3: Detail maps showing the spatial distribution of training events (green), detected events (red) and undetected events (grey). Results for the 8-dimension subspace detector are on top and for a correlation detector are on the bottom. The maps show the distribution of events in latitude and longitude (left) and latitude and depth (right). Symbol size is scaled to event magnitude.

3.2 The Coso Sequence in Eastern California

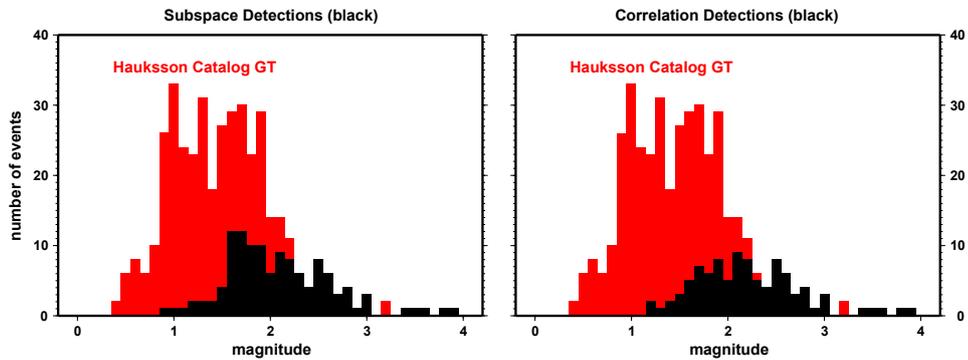


Figure 3.4: Histograms of detected events in this sequence in black for the subspace detector (left) and the correlator (right). The histogram for the Hauksson catalog is shown in red behind the detector histograms. Waveform representation is a significant challenge for this sequence.

3.3 Summary

The greatest drawback of matched filter detectors is the requirement that we have a representative waveform to use as a detection template. Whilst we are encountering surprisingly many situations in which correlation detectors are highly effective, we have also identified a number of situations where very closely spaced events display a very high degree of waveform dissimilarity and where correlation detectors subsequently miss a large number of events. In a case study in the western United States, using NVAR array data from the Coso earthquake sequence from 2001 (at a distance of 270 km), multidimensional subspace detectors were found to perform only modestly better than a correlation detector. Large numbers of events went undetected by either correlation or subspace detectors, suggesting that the smaller magnitude events in this sequence may display greater diversity in mechanism than the larger events used to design the detection templates.

Chapter 4

Application of array-based waveform correlation detectors to events with complicated source-time functions

In March 2005, Norwegian authorities alerted NORSAR to concerns of citizens living in the far north of Norway who had heard loud noises from an unknown source. From an association of reported times with seismic signals, it was concluded that numerous events had occurred on the Kola Peninsula of NW Russia (Figure 4.2); sound waves coming from the appropriate directions were also observed on the seismic sensors of the ARCES array and on the Apatity seismic array and microbarograph subarray. The automatic location estimates (Generalized Beamforming, or, GBF, Ringdal & Kværna, 1989) showed considerable variation (Figure 4.2, left panel) and displayed a very similar pattern to the distributions of GBF event location estimates obtained for verified mining events at the Zapoljarni mines at a distance of approximately 200 km (Figure 4.2, right panel). Analyst reviewed solutions (including some which applied direction estimates from the infrasound phases) appeared to result in quite consistent location estimates close to the circle in Figure 4.2. The reasons for the large spread in automatic event locations are well understood; slowness estimates obtained using broadband f-k analysis are made using variable frequency bands (see Gibbons

et al., 2009), and multiple detections are made for both P-type and S-type regional phases in each signal with highly differing patterns of occurrence. This indicates multiple firing sequences. The automatic phase association algorithm is unable to identify which bursts of energy correspond to which shots and different hypothetical source locations will provide arbitrarily better or worse fits to phase detection sequences. A closer examination of the waveforms displayed in Figure 4.1 reveals great heterogeneity within the set of waveforms observed for this particular event sequence.

Our aim is to identify similar and subsequent events. The fully automatic event lists are clearly little help in this aim without the introduction of much manual analysis. The Zapoljarni mines on the Kola Peninsula frequently result in seismic signals whose automatic event location estimates overlap with the presumed explosion site. Waveform correlation is an appealing method, but low values of the correlation coefficients resulting from comparing the seven traces in Figure 4.1 make it clear that a detection threshold for a correlation detector would have to be very low and that we might have to accept a very high false alarm rate. One observation from the mutual correlations between these seven events is that despite relatively low values of the Array Correlation Coefficient Beam (ACCB), performing f-k analysis on the single channel correlation coefficient traces (c.f. Gibbons & Ringdal, 2006) resulted in very well defined slowness vectors (i.e. indicating good alignment of the correlation coefficient traces).

With the detection threshold set necessarily low, between January 1, 2002, and December 31, 2005, a total of 17485 detections were made based upon the array correlation beam (ACCB) value alone. The vast majority of these were demonstrably false alarms by inspection and other selection criteria were deemed necessary. Criteria chosen include the ratio of ACCB maximum to the standard deviation, the apparent slowness, of the correlation coefficient traces, and the coherence of the correlation coefficient traces. This multi-variate selection condition reduced the number of detections to 243. Of these detections, 220 corresponded to the times detected events on the GBF which appeared to originate from a similar source region and which often showed similar properties. Of special interest is the appearance of sound waves on the ARCES array some 12 to 14 minutes after

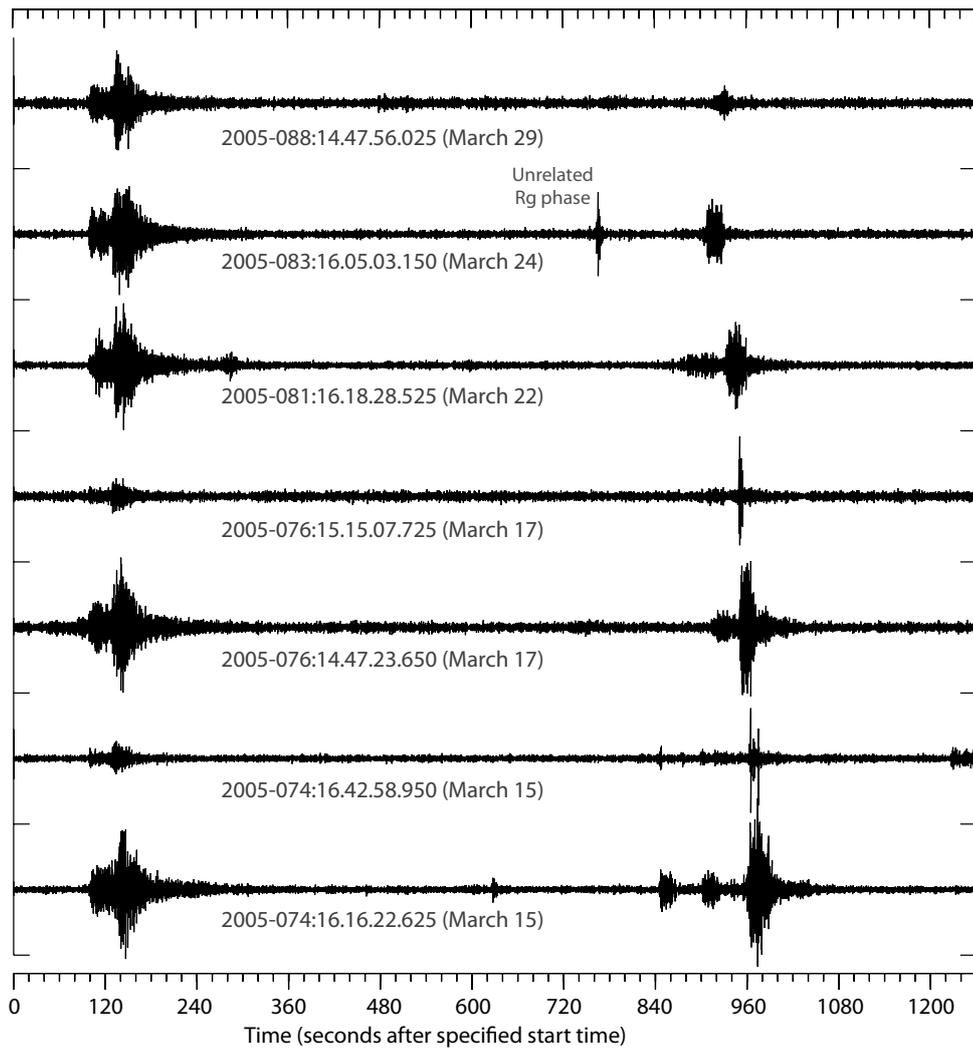


Figure 4.1: ARCES waveforms showing seismic and acoustic signals from presumed military explosions near the northern coast of the Kola Peninsula.

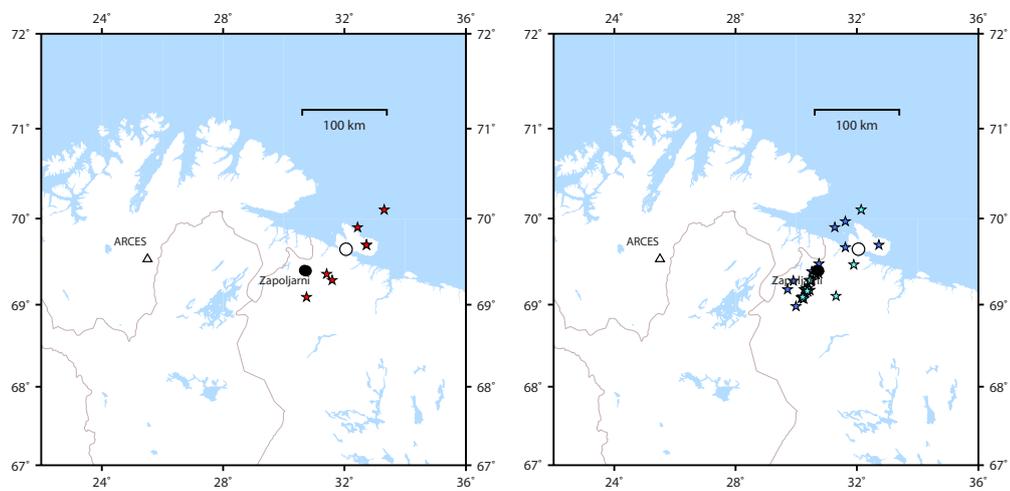


Figure 4.2: Automatic (GBF) location estimates for the events displayed in Figure 4.1 (left) and for events known to come from the mines at Zapoljarni (right). The two different shades of blue in the right hand panel distinguish between events from two nearby mines. The open circle indicates the location estimates obtained by Ringdal *et al.* (2005) based only upon backazimuth estimates for the infrasound phases observed on the ARCES seismic array and the microbarograph array at Apatity.

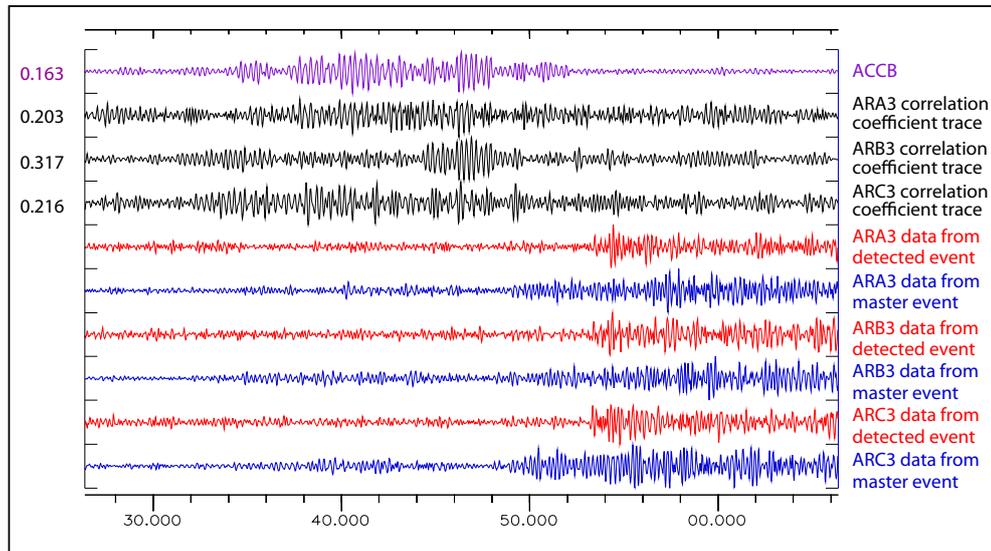
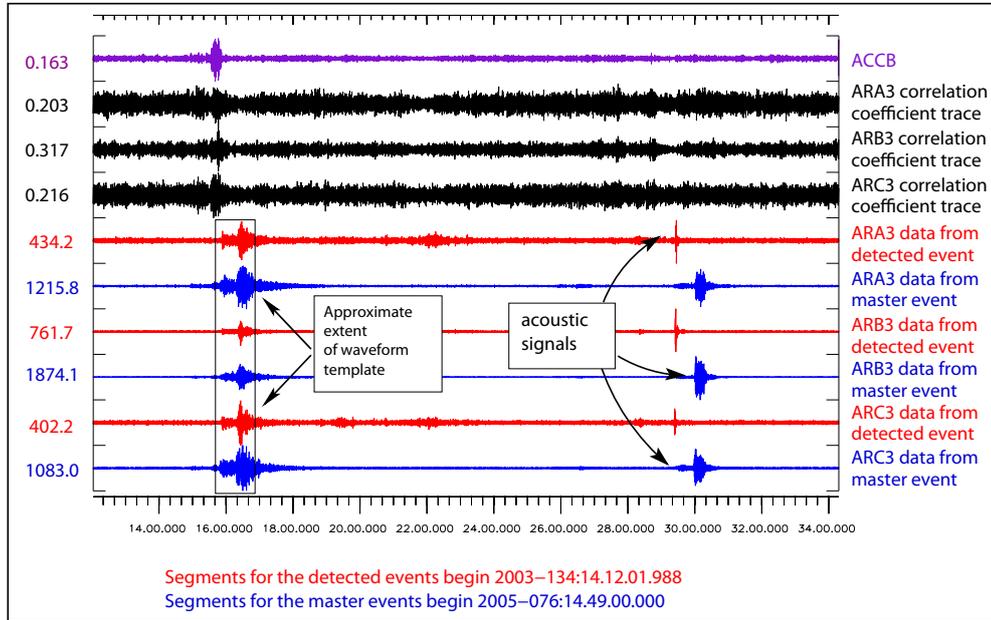


Figure 4.3: A typical correlation detection on the ARCES array using a template from a Kola Peninsula event on March 17, 2005. The seismic signals from both master and detected events are followed by infrasound arrivals. Note the long duration of the semblance of the correlation coefficient traces and the absence of well-defined peak in the ACCB (Array Cross-Correlation Coefficient trace Beam).

the explosions, which supports the hypothesis that the events are of a similar nature and almost certainly near-surface explosions. An example of such a detected event is displayed in Figure 4.3, displaying the somewhat dissimilar waveforms and the ACCB with no clearly defined maximum. Three of the 243 detections did not correspond to automatic GBF event hypotheses, although careful manual VESPAgram analysis indicated evidence of weak P- and S- type regional phases coming from the appropriate direction. The remaining events were clearly false alarms and resulted from the occurrence of very high amplitude, short-duration, Rg-type phases arriving from approximately the same backazimuth. Whilst the actual events are somewhat poorly constrained (i.e. there is no Ground Truth, and no independent confirmation of the location of the explosion site or sites) the method we have applied has demonstrated that correlation detectors which apply additional constraints (primarily on the alignment of the single channel correlation coefficient traces) have been able to produce an extensive list of very likely candidate events with very few obvious false alarms. It is worth noting that there are no coincidental correlation detections with signals from Zapoljarni mining events. Any other existing procedure to identify that number of candidate events would almost certainly also result in many false alarms and much additional analyst time.

Chapter 5

Control of Instrumental Timing Using Correlation Detectors

5.1 Instrumental Timing: Introduction

[Gibbons & Ringdal \(2006\)](#) pointed out that if two seismic events are co-located, since the travel time to any given station is identical for both events, the time separating the start of the waveform template for the first event and the maximum of the correlation coefficient channel (whereby the second event is detected) should be identical for all stations. This is the basis by which we can perform coherent beamforming of correlation coefficient traces over sparse networks, even when the actual waveforms show no similarity whatsoever. If this is not the case (and the difference cannot be ascribed to waveform dissimilarity - whether due to differences in the seismic sources or to a low SNR) then we have to conclude that there is an inconsistency in instrumental timing at one (or both) of the stations at the time of one (or both) of the events. The principle is illustrated in [Figure 5.1](#).

In [section 5.2](#) we examine two cases of erroneous instrumental timing encountered during this contract. The first ([5.2.1](#)) deals with numerous examples of timing anomalies revealed on IMS seismic arrays whilst monitoring a swarm of earthquakes in the Rana region of Norway during 2005 ([Gibbons *et al.*, 2007a](#)). The second ([5.2.2](#)) deals with a case where erroneous timing on a single 3-component station (KBS) was both detected and measured using time-delays measured on

5.1 Instrumental Timing: Introduction

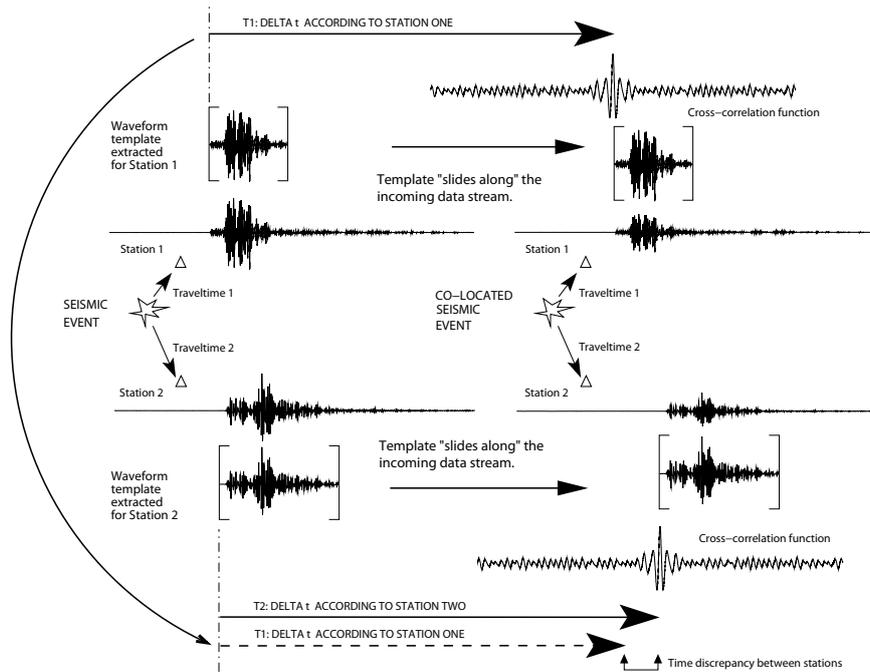


Figure 5.1: A schematic illustration of how two successive events from almost identical seismic sources can be exploited to reveal anomalies in the timing at a given station. Assuming that no measurable changes occur to the velocity structure between source and receivers, seismic waves from two co-located events will take the same length of time to reach any given sensor. The cross-correlation function for a given signal at a given station measures how similar the subsequent portion of the seismogram is to the waveform template. The time separating the start of the template and the maximum of the cross-correlation function should equal the time separating the two event origin times for all stations. Any discrepancy in the separation times measured at two different stations, which is not attributable to source differences or a poor SNR, must be the result of a timing anomaly at one, or both, of the instruments.

two different stations between repeating mining-induced events at the Barentsburg coal mine in February and March 2006.

5.2 Instrumental Timing: Examples

5.2.1 Timing Example: Rana earthquake swarm

The earthquake swarm in the Rana region of northern Norway described by [Gibbons *et al.* \(2007a\)](#) was fortuitous in many ways. Not only did it contain events covering a wide range of magnitudes, but it was recorded to varying degrees by at least five different IMS seismic array stations (NORSAR, ARCES, HFS, FINES, and SPITS). [Gibbons & Ringdal \(2006\)](#) point out that the validity of correlation detections on arrays can be supported by examining the alignment of the cross-correlation traces, at the times of local maxima on the correlation beam, using f-k analysis. A non-zero slowness vector indicates that the correlation detection is probably the result of a coincidental similarity between two wavefronts arriving from slightly different directions. However, even if the slowness vector measurement is satisfactory, there can be single-channel cross-correlation traces which are not aligned with the others which, for the reasons illustrated in [Figure 5.1](#), indicate a timing aberration. [Figure 12](#) of [Gibbons *et al.* \(2007a\)](#) illustrates two such cases, one on the NORSAR array and one on the SPITS array.

[Figure 5.2](#) shows another example, this time on the FINES array. In each of the examples shown, the values of the cross-correlation coefficient traces were high for each single channel meaning that the individual time-delays could be measured quite accurately using the procedure of [VanDecar & Crosson \(1990\)](#). For more marginal detections, it may not be possible to obtain accurate time-delay measurements.

The cases of anomalous timing are much more easily seen in the misalignment of the cross-correlation traces than, for example, the misalignment of waveforms at the times of strong teleseismic signals. This is because all time-delays, other than those resulting from timing errors, are already accounted for and require no theoretical model. However, we recommend that for every detection which passes

5.2 Instrumental Timing: Examples

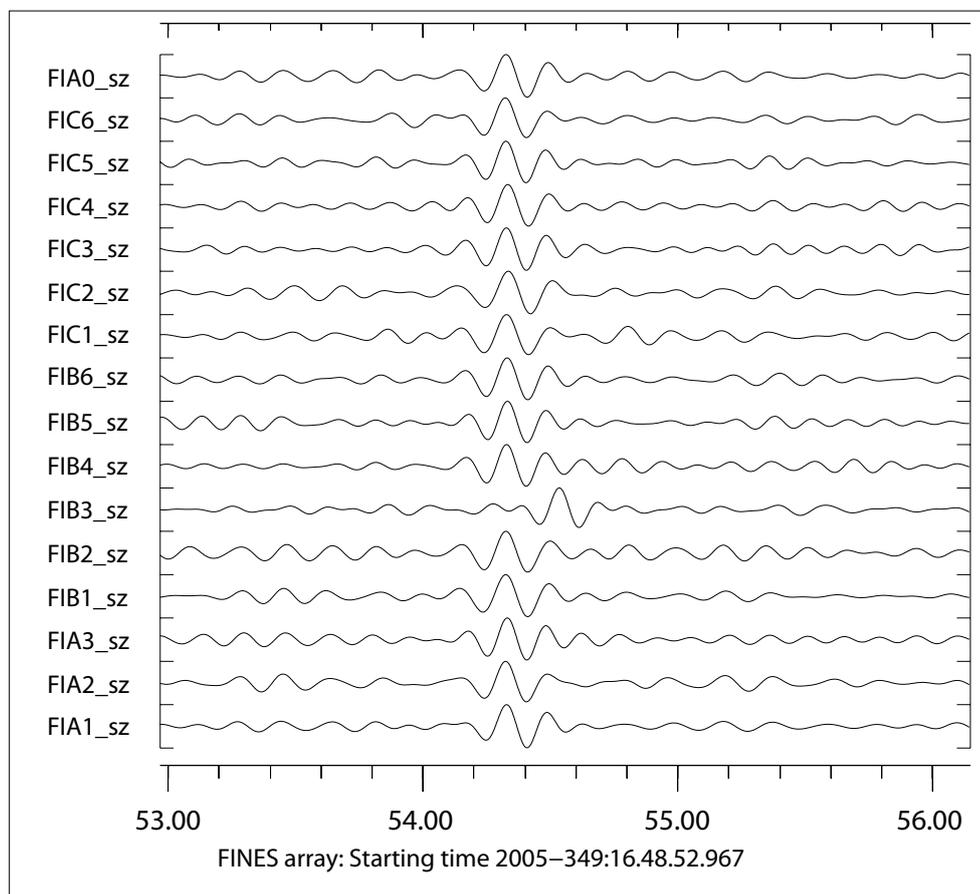


Figure 5.2: Misalignment of cross-correlation traces on the FINES array. Given a detection on a stack of cross-correlation traces, the least squares inversion of [VanDecar & Crosson \(1990\)](#) can be used to measure accurately the time-delays between the correlation maxima for the single channels. The FIB3.sz site is clearly not synchronized with the other elements although, since GPS-lock for this site was on for neither the master event nor the detected signals, it is impossible to calculate a correction from this data alone. The master signal template is a 20 second-long data segment beginning at a time 2005-175:04.27.22.35000 (bandpass filtered between 2.5 and 8.0 hz). The maximum of the correlation beam occurs at a time 2005-349:16.48.54.32750. Details of the events are provided in [Gibbons *et al.* \(2007a\)](#).

the slowness vector test described by [Gibbons & Ringdal \(2006\)](#), a calculation of the individual time-delays be attempted in order to identify clear anomalies.

5.2.2 Timing Example: The KBS station

An event in the vicinity of Novaya Zemlya on March 5, 2006, was well-recorded by the SPITS and ARCES arrays and also by the broadband station KBS on Spitsbergen (Figure 5.3). The arrival times for the Pn and Sn phases at the KBS station (Figure 5.4) could not be reconciled with those at the array stations and the residuals obtained by various attempts to locate the event indicated that a consistent offset in the timing at KBS was to blame.

The operators of the station confirmed that a technical fault had occurred with the station on February 17, 2006, which was repaired on March 22, 2006. Continuous real-time correlation detectors have been run on SPITS data at NOR-SAR for some time to detect mining-induced seismic events at the Barentsburg coal mine, the source of many almost-repeating seismic signals. Many of the Barentsburg events were recorded by both the SPITS array and the KBS station (see Figure 5.5). If t always denotes a UT time, then we can define a correction function $C_{\text{KBS}}(t)$ that allows the apparent time according to the KBS station to be calculated using

$$t_{\text{KBS}}^{\text{app}} = t - C_{\text{KBS}}(t). \quad (5.1)$$

The time separating the origin times of the two events is equal to the time separating the start of the waveform template and the maximum correlation coefficient for all stations.

Assuming that the SPITS array recorded both master and detected events with the correct time, and that the KBS recorded the master event with the correct time, we can calculate $C_{\text{KBS}}(t)$ using

$$C_{\text{KBS}}(t) = (t_{\text{KBS}}^{\text{ccm}} - t_{\text{KBS}}^{\text{wft}}) - (t_{\text{SPI}}^{\text{ccm}} - t_{\text{SPI}}^{\text{wft}}) \quad (5.2)$$

where $t_{\text{x}}^{\text{wft}}$ is the start of the waveform template for station “x” and $t_{\text{x}}^{\text{ccm}}$ is the apparent time of the correlation coefficient maximum for station “x”. Measurements of correlation maxima at both stations for many events before, during, and following the period affected by the technical fault allowed us to measure

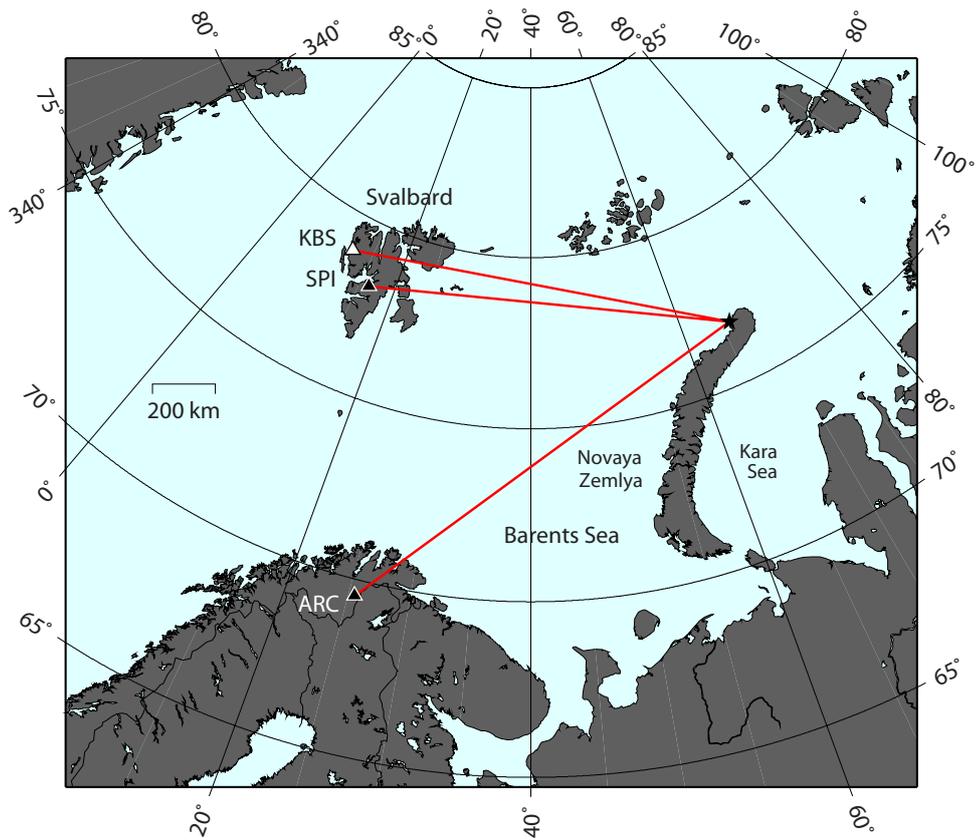


Figure 5.3: Automatic location estimate (using the GBF algorithm, Ringdal & Kväerna, 1989) of the March 5, 2006, Novaya Zemlya event. Also shown are the IMS seismic arrays ARCES and SPITS, and the IRIS/GEOFON/AWI 3-component station KBS at Ny Ålesund, Kings Bay, on Spitsbergen.

5.2 Instrumental Timing: Examples

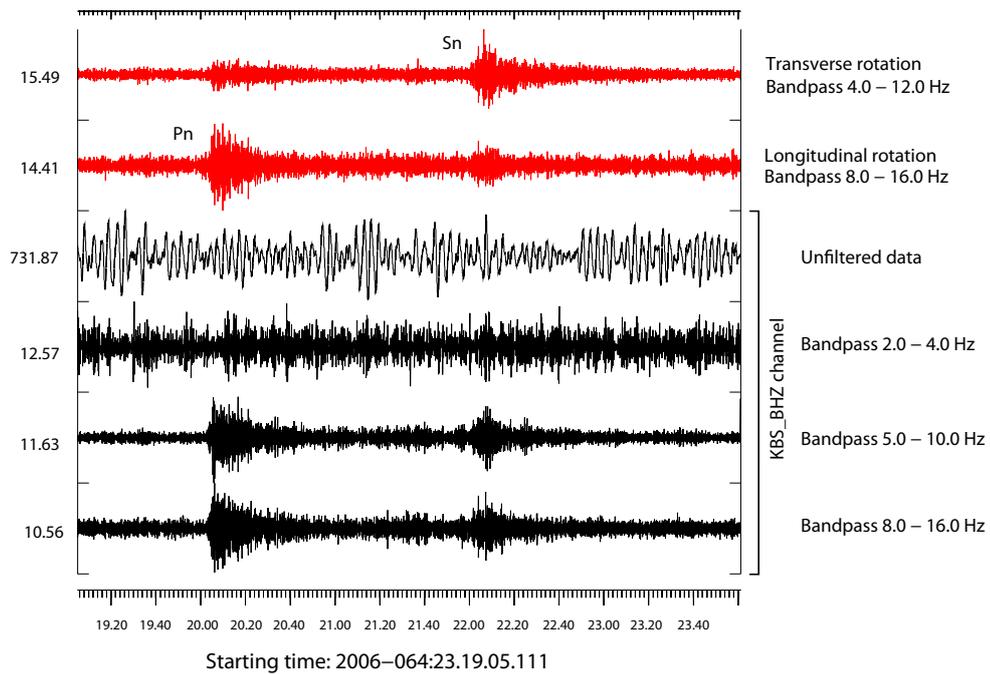


Figure 5.4: Signals from the March 5, 2006, Novaya Zemlya event as indicated on the KBS station. Due to an instrumental technical fault, the time-stamp for the KBS station could not be relied upon and attempts to locate the Novaya Zemlya event using phase determinations from ARCES, SPITS, and KBS demonstrate that the time-stamp at KBS must be too early by several seconds.

5.2 Instrumental Timing: Examples

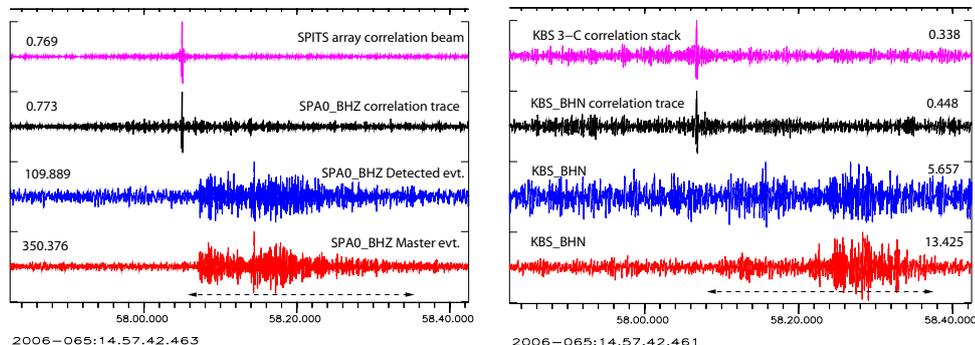


Figure 5.5: Detection by waveform correlation of the first identified rockburst at the Barentsburg coal mine, Spitsbergen, following the March 5, 2006, Novaya Zemlya event using the SPITS array (distance 50 km) and the KBS 3-component station (distance 120 km). The master event used is the first identified Barentsburg event following the repair of the KBS station on March 22, 2006. All data are bandpass filtered between 3.0 and 6.0 Hz and a 30.0 second waveform template is extracted for all available channels at both stations beginning at the estimated onset time of the first P-arrival. The SPITS waveform template begins at a time 2006-081:23.25.07.15000 and the interpolated correlation coefficient maximum at SPITS occurs at a time 2006-065:14.58.05.00229. The KBS waveform template begins at a time 2006-081:23.25.17.00760 and the interpolated correlation coefficient maximum at KBS occurs at a time 2006-065:14.58.06.78268.

that, at the time of the Novaya Zemlya event, the time stamp at the KBS station was approximately 8.07 seconds earlier than the actual UTC time (Figure 5.6). With the corresponding correction applied to the arrival time estimates, a well-constrained location is obtained for the Novaya Zemlya event including the KBS phase determinations. This study is described in full in [Gibbons \(2006\)](#).

5.2 Instrumental Timing: Examples

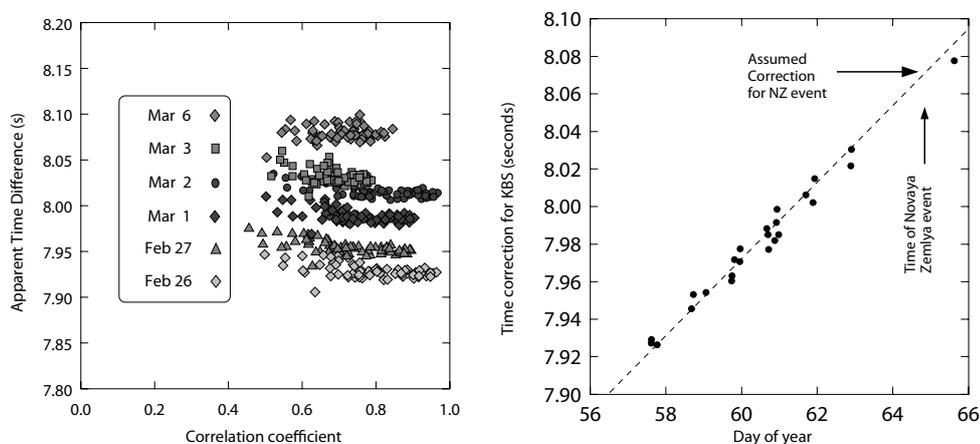


Figure 5.6: (Left) Variability of the time-correction function $C_{\text{KBS}}(t)$ for a number of detected Barentsburg events. Each point indicates the time-difference calculated for the indicated event using a certain master event with the SPITS array correlation coefficient displayed on the x-axis. All master events are taken from 22 March, 2006, or later, after the KBS station was repaired. (Right) $C_{\text{KBS}}(t)$ estimates and best fit line for the interval February 25, 2006, to March 6, 2006. Reading from this line indicates that approximately 8.072 seconds must be added to the KBS time-stamp to give the correct time at the time of the Novaya Zemlya event.

5.3 Instrumental Timing: Summary

Given two seismic events which are approximately co-located, the time separating corresponding parts of the two wavetrains generated is identical for any observing station. This makes errors in instrumental timing relatively easy to detect when performing studies of event detection using multichannel waveform correlation.

When running a correlation detector on a small or medium aperture array, the validity of detections is frequently assessed by performing f-k analysis on the set of single channel cross-correlation traces (Gibbons & Ringdal, 2006). If the cross-correlations on the single channels are sufficiently well-defined, the inversion procedure of VanDecar & Crosson (1990) can be applied to measure accurately the corresponding time-delays. If the absolute time-delay obtained for any channel is clearly non-zero, it is essentially guaranteed that the instrumental timing has been erroneous. For the Rana earthquake swarm in northern Norway in 2005, instances of erroneous timing were observed on three IMS arrays: NOA, SPITS (see Gibbons *et al.*, 2007a), and FINES (Figure 5.2).

Timing errors on remote stations can be detected and sometimes measured by comparing time-differences between almost repeating seismic events at two or more stations. Here, we considered the IRIS 3-component station KBS on the island of Spitsbergen in the European Arctic. The presence of a timing error was identified when attempting to use phase readings from this station to locate an event close to Novaya Zemlya in March 2006. At the time, almost identical signals were being generated by mining-induced seismicity at the Barentsburg coal mine, recorded by both KBS and the SPITS array. An extrapolation of measurements from numerous events allowed us to determine that the time indicated by the clock at KBS was approximately 8.07 seconds earlier than real-time when the Novaya Zemlya event occurred. A complete description of this study is found in Gibbons (2006).

The method of Koch & Stammer (2003) for detecting timing aberrations on a single seismic array has the advantage that it does not depend upon fortuitous sequences of repeating seismic events. It does however have the disadvantage that all stations considered are sufficiently close for the microseismic background noise to be coherent between sensors. The method outlined in Gibbons (2006) can

5.3 Instrumental Timing: Summary

be used on a network of arbitrary dimensions, provided that all stations record occurrences of the same almost-repeating seismic source.

Rubin (2002) describes how erroneous instrumental timing can be identified and measured from double-difference earthquake catalogs, although it is implicit also in this method that recurring seismicity is recorded by multiple stations. More recently, Sens-Schönfelder (2008) has outlined a method of measuring instrumental timing errors in the process of retrieving Green's Functions from pairwise correlations of ambient noise. This method is promising given that it does not require the recording of repeating events. It does, however, require the recording of very long segments of data and a more complicated retrieval process. The information required for the verification of consistent timing described in this chapter is part of the standard output of the correlation detection procedure and can be exploited at little additional cost to provide a timing consistency check across the network employed. This study indicates that verification of timing consistency alone is an incentive to identify more sources of almost repeating seismicity.

Chapter 6

Size of the Correlation Footprint

6.1 Introduction

When assessing the applicability of waveform correlation detectors, it is of paramount importance to be able to estimate how far from the master event a subsequent event can be such that it can still be detected using the available signal template. We use the term *correlation footprint* to describe the region surrounding the master event location in which seismic events can be detected using this template. It is important to note that we are assuming the “best case scenario” for detectability, i.e. that the events have similar source mechanisms and similar source-time functions. Two seismic events which are essentially co-located may not correlate if these properties differ significantly.

The coherency of seismic signals over receiver arrays has been studied at great lengths since the signal semblance at various inter-site distances dictates how useful a given design of seismic array will be for detecting and estimating different classes of seismic signals. For example, a teleseismic signal dominated by low frequency energy can be highly coherent over a network with an aperture of many tens of kilometers, whereas higher frequency signals can be quite incoherent if the instruments are separated by more than a few hundred meters. Similar arguments are likely to apply to the source region and the size of the correlation footprint is likely also to be a function of the dominant frequency of the signal. [Geller & Mueller \(1980\)](#) suggest that well-correlating events are likely to be separated by no more than a quarter wavelength at the dominant wavelength. To examine

the extent of the correlation footprint as a function of frequency band and signal duration, we would ideally have records of many hundreds of seismic events with precisely known locations over a very wide range of inter-event distances. The locations of earthquakes are seldom sufficiently well known, and it is expensive and logistically difficult to produce an event array of explosions from which a sufficient number of observations can be made. Mining explosions are often suitable for such studies (Harris, 1991) although these events are often associated with very differing and complicated source-time functions.

In this chapter, we discuss briefly two procedures which aim to obtain representative estimates of how far apart two events may be separated such that one can still detect one using a waveform template from the other. Section 6.2 examines the detectability on single sensors by using a reciprocal argument, considering waveforms from a single event recorded over a small-aperture array. Section 6.3 examines differences between the detectability at different event separations on single channels and the detectability at different event separations using stacked traces over an array. For this case study, we examine the recording on a small-aperture array of a marine seismic profile in which events are separated by approximately 200 meters.

6.2 Investigating Correlation Distance Using a Reciprocity Argument on a Small Aperture Seismic Array

It can be informative to consider the reciprocal case, where the signal from a single event is recorded on closely spaced instruments of an array (an approach also taken by, for example, Menke, 2001). Figure 6.1 shows the signal recorded at the central element (ARA0) of the ARCES array from a surface explosion at a site in Finland at a distance of approximately 178 km (details of these events are provided by Gibbons *et al.*, 2007b). At the sites in the A-ring (see Figure 6.1), with inter-site distances of the order 150 m, a clear correlation peak indicates that the signals at these sites can easily be detected by a matched filter using the ARA0 signal as the template. Under our supposition of reciprocity, we assume

6.2 A Reciprocity Argument

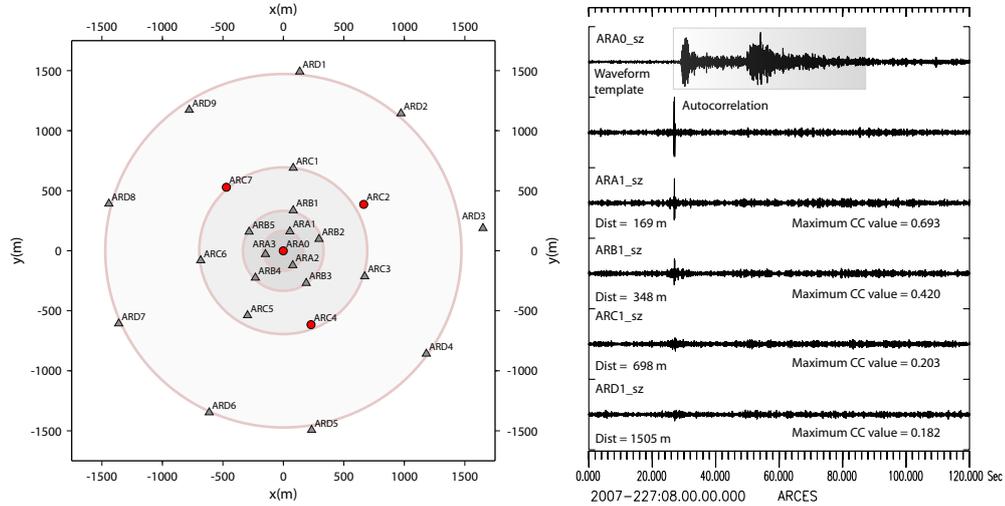


Figure 6.1: (Left) Geometry of the ARCES seismic array. (Right) An attempt to detect, using single-channel waveform correlation, the seismic signal from a given event on various vertical sensors of the ARCES array using the signal recorded at the central ARA0 site as the waveform template. The length of the template is 60.0 seconds and all waveforms are filtered 2.0 - 8.0 Hz.

that the signal from an event 150 m from the master event could probably be detected using this signal template. Figure 6.1 shows that the signal at the ARB1 site (a distance of 350 m) is also detected clearly using the ARA0 template, albeit with a lower value of the correlation coefficient, whereas the signals at the outer-ring sites (distances over 500 m), are too dissimilar for a matched filter detection. Similar calculations were repeated using the signals from all of the other elements as templates, and hence covering a very large range of inter-site distances. This exercise confirms an approximately exponential decrease in the coherence between waveforms with inter-site distance (see, for example, [Menke *et al.*, 1990](#)). For this regional event, with a 60 second long template filtered from 2 to 8 Hz, the signal at a site at ARCES can be detected by correlation using the signal from a different site provided the inter-site distance is less than about 500 m. This is consistent with the quarter wavelength argument of [Geller & Mueller \(1980\)](#).

6.3 Examining the Distance of Detectability With Waveform Correlation Using a Marine Seismic Profile

In August 2007, a series of marine seismic profiles was shot in the Barents Sea under the name PETROBAR. One of these profiles is displayed in Figure 6.2 and was of great interest because most of the shots were reasonably well recorded by the ARCES array at a distance of less than 250 km. Unusually for such marine profiles, there are clear S-phases detected at the array. The signals are best observed in the 4-8 Hz frequency band and waveforms filtered in this range are displayed for a segment of a little over 26 minutes in Figure 6.3.

The distance between the events is approximately 200 m, with an approximate 90 second delay between each shot. For the distance between the profile and ARCES, this allows a 60 second template to be extracted for each event, containing both P- and S-phases and coda, which is not contaminated with the signal from any other shots. Figure 6.3 also displays the single-channel cross-correlation traces calculated from the template as indicated. There is some variation between the different ARCES channels but the primary observation is that, with only a single channel, the signal from one event can be used to as a template to detect events up to 600 m away. This is consistent with the results from the reciprocity experiment described in section 6.2. This calculation was repeated using every single event as a master event and the pattern remained largely unchanged. Whilst signals with an exceptionally low SNR usually resulted in a poorer detection capability, there was no obvious connection between the SNR of the master event signal and how many adjacent shots could be detected using the matched filter detector (Shelly *et al.*, 2007, describe a correlation detection case study whereby the master event signals almost inevitably suffer from a low SNR)

The most important observation is that when the cross-correlation traces are stacked, the signals from events at several kilometers distance are detected from the given template. This result is non-intuitive and provides considerable motivation for the use of array-based waveform correlation detectors as opposed to detectors using only single stations. The preliminary investigations of Gibbons

6.3 An Example from a Marine Seismic Profile

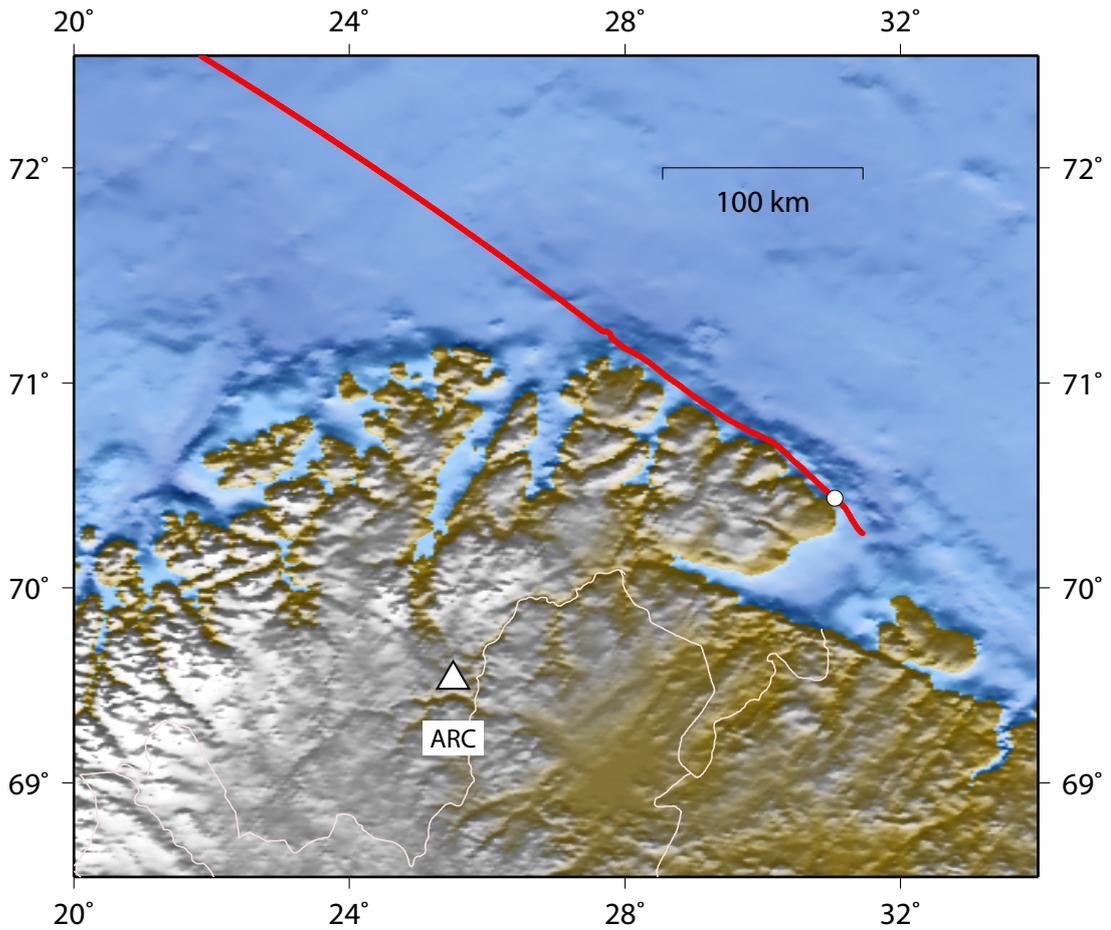


Figure 6.2: Locations of shots in the PETROBAR 1 marine seismic profile in relation to the ARCES regional seismic array. The white circle shows the location of shot 121 with coordinates $70.444386^{\circ}\text{N}$, $31.053447^{\circ}\text{E}$ and origin time 2007-221:03.02.32.523. For this part of the profile, shots are separated by approximately 200 m.

6.3 An Example from a Marine Seismic Profile

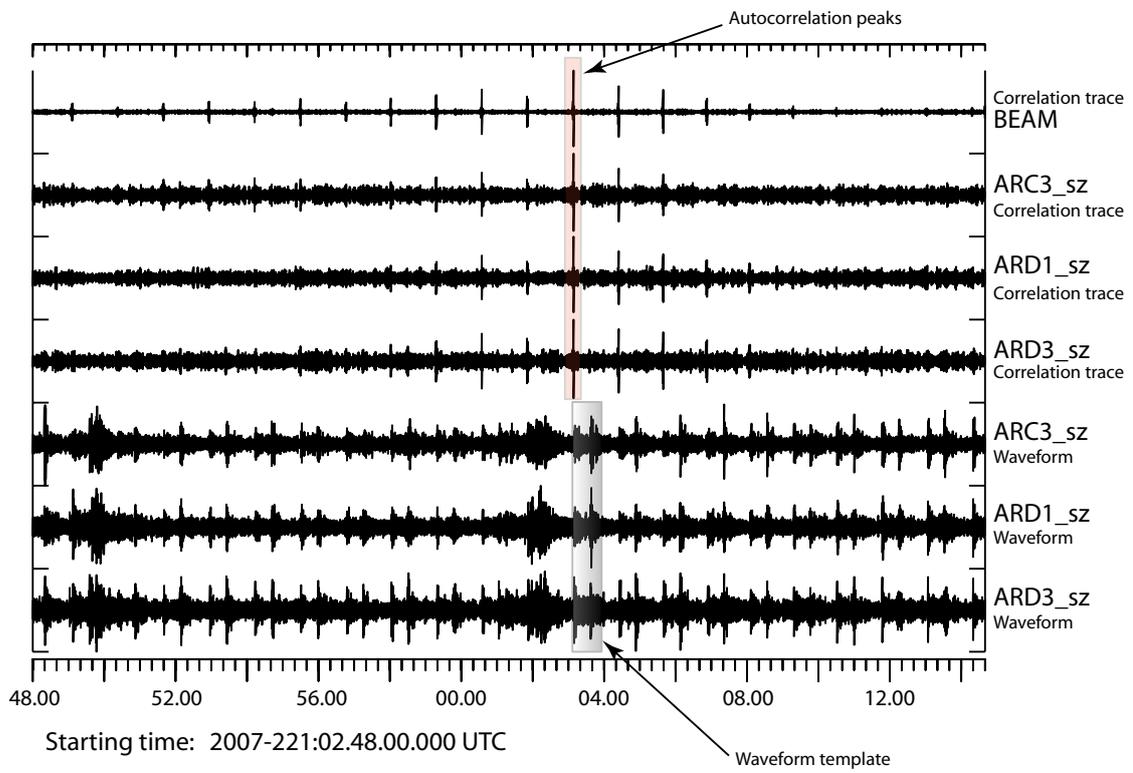


Figure 6.3: Waveforms and cross-correlation traces with a template from the shot 121 signal.

& Ringdal (2006) and Gibbons *et al.* (2007a) went only as far as to suggest that correlation trace stacking should improve the detectability of a co-located event with a lower SNR. That network stacking should increase the size of the correlation footprint is, however, consistent with an observation of Harris (1991) that statistically significant correlation coefficients were obtained between events at greater distances when a full-array was used than when only a single channel was used.

6.4 Summary

Firstly, it has been demonstrated that the Geller and Mueller quarter wavelength argument is probably reasonable for the size of the correlation footprint for single channel matched filter detectors. Stacking the correlation coefficient traces over an array or network results not only in an improvement in detectability for weaker co-located events, but also an expansion of the size of the correlation footprint. The stacking operation presses down the correlation detection statistic for the background noise faster than that for nearby events resulting in significant values of the detection statistic at greater distances when the full array is used. For the regional signals in the PETROBAR 1 profile displayed here, the correlation distance for a single channel was of the order 500-600 m. When the full ARCES array was employed, this distance was increased to up to several kilometers.

The correlation distance will vary greatly as a function of the source characteristics (see in particular chapter 3) as well as the dominant wavelength. In addition, the homogeneity of the surrounding geological structures will need to be considered and evaluated on a case-by-case basis (Nakahara, 2004).

The level of waveform similarity required to confirm that two seismic events occurred within a specified distance (within, for example, a quarter wavelength) is very high. The observational and numerical investigation of Baisch *et al.* (2008) operates with a correlation threshold value of 0.95. Such high values are almost never observed in the practical situations we have encountered during the studies described within this report. It should be pointed out that the array-stacking of single channel correlation coefficient traces, followed by the verification of their zero-delay alignment, allows for very low values of the correlation coefficient to

6.4 Correlation Footprint: Summary

constitute significant correlation detections which represent strong evidence for almost-repeating seismicity. If a sufficient number of such detections can be observed from different directions, arguments related to the relative timing of the detection statistic maxima are likely to provide a far more convincing case for co-location of events than the values of the correlation coefficients themselves (see [Gibbons & Ringdal, 2005](#)).

Chapter 7

Characteristic Function Correlation Detectors

7.1 Introduction

The classification of extensive aftershock sequences from large earthquakes is one of the most challenging problems faced by data centers engaged in the monitoring of nuclear explosions. The sheer number of events requiring manual analysis and relocation can lead to significant backlogs in the compilation of event bulletins and the subsequent screening of unambiguous earthquakes. The identification and association of aftershocks should ideally be performed with as high a degree of automation as possible such that limited analyst resources are only necessary to provide checks on the results of the automatic processing.

In a semi-autonomous system for the classification of aftershock sequences, we have to run simultaneously correlation detectors (for finding new occurrences of signals already observed) and traditional power detectors (for occurrences of signals which have not been observed previously). In most cases, we will only want to consider seismicity originating from a rather limited source region and we therefore need a signal detection system which is relatively insensitive to signals from completely unrelated sources.

In this chapter, we consider a modification to the standard multichannel matched filter detector which, rather than correlating the waveforms themselves, correlates quantifiable properties of the wavefield which may be characteristic of

7.2 The February 21, 2008, Svalbard earthquake and aftershock sequence

the source region of interest. Some attention has been given to processes which correlate waveform envelopes (e.g. [Withers *et al.*, 1999](#)) although such systems may be of limited application in cases where the events are observed by relatively few stations.

7.2 The February 21, 2008, Svalbard earthquake and aftershock sequence

On February 21, 2008, a magnitude 6 earthquake occurred in the Storfjorden region close to Svalbard (Figure [7.1](#)). Whilst not large in a global perspective, this event illustrates beautifully the difficult task of aftershock identification. The aftershock sequence consisted of many hundreds of events within the first two days and a very much increased level of seismicity for several months afterwards. The simulated helicorder plot from the SPITS array (Figure [7.2](#)) illustrates the largest events for February 23, 2008. The dynamic range at the SPITS array saturated under the main event and the data is clipped on all channels. The signal at the ARCES array was not clipped and a template was extracted rapidly in an attempt to detect aftershocks using a correlation detector. It became evident that very few of the numerous aftershocks were being detected using this template, with essentially no convincing detections. Following the analyst review of the aftershocks in the first few days following the main event, matched filter detectors were initiated for a large number of master events. Many of these detectors were successful in detecting large numbers of other events and it became clear that a cluster analysis of the sequence would be possible using associations indicated by high values of the correlation coefficient.

In performing this correlation analysis, we have however been completely dependent upon a bulletin of manually located events; our goal is to classify the sequence automatically. The preliminary automatic event bulletin (GBF, [Ringdal & Kværna, 1989](#)) upon which the reviewed bulletin is based is not accurate enough for an unambiguous association of events; errors in origin time and hypocentral location are too large even for the setting of time-windows for waveform correlation analysis. The fixed-window template method of [Gibbons *et al.* \(2005\)](#) is also

7.2 The February 21, 2008, Svalbard earthquake and aftershock sequence

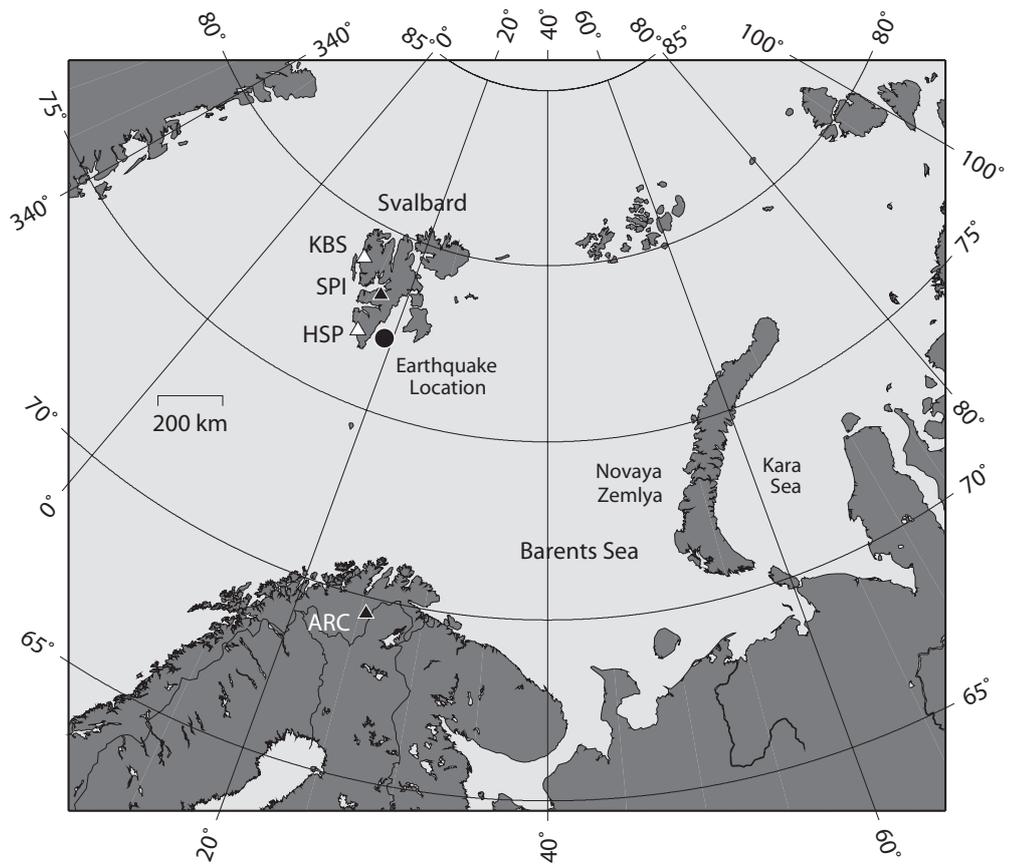


Figure 7.1: Location of the February 21 Svalbard event in relation to the SPITS and ARCES arrays and the 3-component stations HSP and KBS, at Hornsund and Ny Ålesund respectively.

7.2 The February 21, 2008, Svalbard earthquake and aftershock sequence

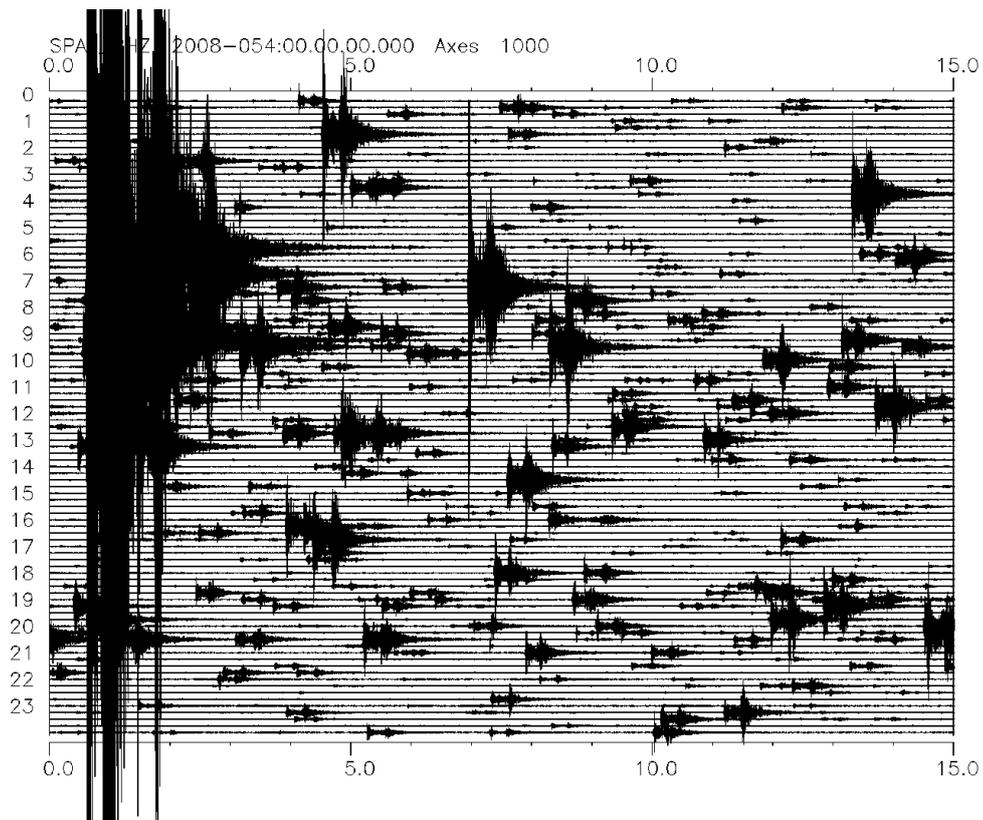


Figure 7.2: Mock helicorder plot of data from the SPA1_BHZ element of the SPITS array, 23 February, 2008. All data is bandpass-filtered between 2.0 and 8.0 Hz. Almost every signal on this plot is an aftershock of the February 21 M=6.2 Svalbard event.

7.2 The February 21, 2008, Svalbard earthquake and aftershock sequence

not applicable. This algorithm is designed for very small source regions (maximum length-scale of 2 or 3 km) for which numerous events have already been observed for the calibration of processing parameters. A preliminary examination of event location estimates for the Svalbard sequence indicates a source region aperture at least an order of magnitude larger than this. In addition, we would like to assume that our main shock has occurred in a location where no event has been observed previously meaning that no calibration data is available.

An alternative possibility is that of correlating characteristic functions of the waveforms against the corresponding transformations of the incoming data. Since we do not have “ripple for ripple” correspondence between the events we wish to associate, we need to correlate some coarser signature of the generated signals. Viable automatic detection and location schemes have been constructed which correlate waveform envelopes but we need to proceed with caution since the small aperture of our regional seismic array means that the waveform envelopes on the different channels show essentially the same features at the same times. The standard waveform correlation detector (a *coherent correlation detector*) exploits details in the full wavetrain which are specific to the given site. For this reason, one would rarely apply beamforming prior to correlation since this would lose much of the available information about the signals’ signature at each site; beamforming is applied after the correlation to much greater effect. However, since we wish to compare waveform envelopes, or some other signal attributes, (i.e. an *incoherent correlation detector*) we should apply every possible transformation to the waveforms from a given event which are likely to result in a shape specific to that source region. The construction of part of the *incoherent template* at ARCES for the Svalbard event is displayed in Figure 7.3.

The relationship between the three uppermost traces in Figure 7.4 suggests a system for the semi-autonomous readjustment of a waveform template pool for new master events. Signals from such events could be selected either for standard correlation detectors or, alternatively, as new templates to add to the basis for a signal subspace detector (Harris, 1989, 2006). A candidate template is any event which triggers a detection on the incoherent detector (probably verified by an analyst before accepting). The time of the corresponding event is searched for in the lists of correlation detections from the existing correlation detector

7.2 The February 21, 2008, Svalbard earthquake and aftershock sequence

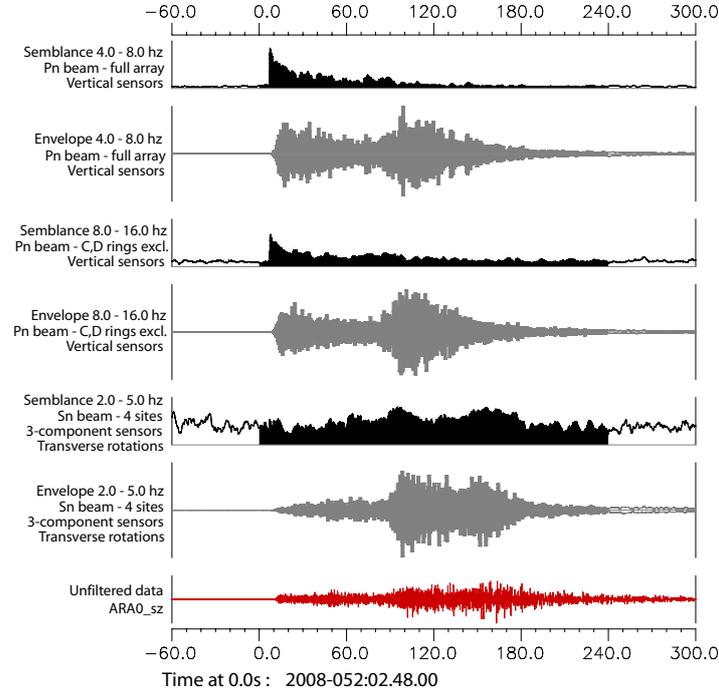


Figure 7.3: Construction of an *incoherent matched filter detector* from ARCES waveforms for the February 21, 2008, Storfjorden main event. The small aperture of the ARCES array means that the variability of waveform envelopes between sensors is minimal and there would therefore be very little advantage in stacking of correlation coefficient traces. However, the coherence of the waveforms can be exploited to form beams using time-delays corresponding to the slownesses of the various observed seismic phases. This emphasizes different sections of the seismograms at times characteristic of the source-receiver distance, although with far less sensitivity than that necessary for setting of the time-windows for f-k analysis prescribed in the algorithm of [Gibbons *et al.* \(2005\)](#). In addition to the envelopes of beams formed for different slowness vectors, different component rotations, and different frequency bands, we have the corresponding characteristic functions (semblance and the related F-statistic) which can also be correlated against the same functions evaluated for the incoming data stream. The semblance functions are greatest for the initial Pn-arrivals on the corresponding beams. The semblance function for the Sn beam on the transverse components suffers from the reduced number of instruments and non-optimal coherence of the Sn phase between these 4 widely spaced sensors.

7.2 The February 21, 2008, Svalbard earthquake and aftershock sequence

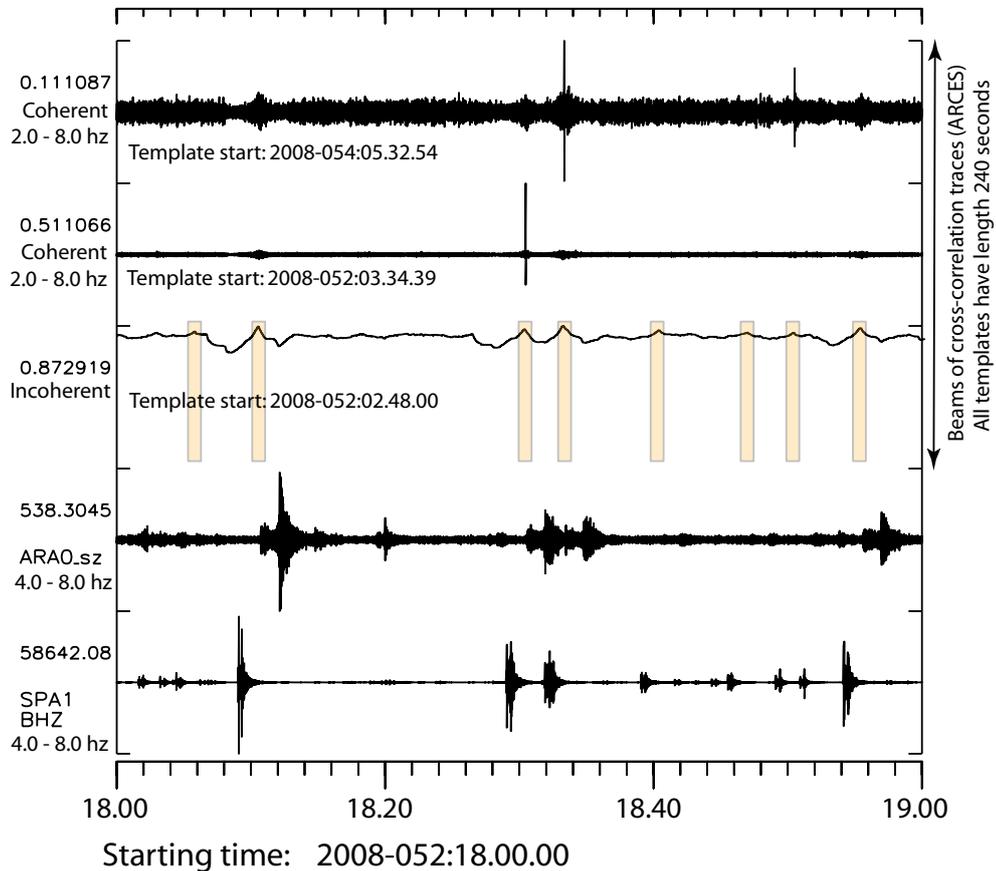


Figure 7.4: Detections of aftershocks using the incoherent template illustrated in Figure 7.3 and using two standard correlation detectors. The template for the top trace detects two events during this one-hour interval, the template for the second trace detects one (different) event, and the incoherent template takes all three events plus 5 additional events. Data from the SPITS array in the lowermost trace confirms that each of the incoherent correlation detections does indeed correspond to a Svalbard aftershock. Note that the signals at the SPITS array arrive approximately 90 seconds prior to the ARCES detections due to the travel-time differences. The Svalbard aftershocks visible in the data that are not detected by the incoherent correlator occur within close succession of other Svalbard events, and such rapid sequences cannot be resolved using the four-minute templates at ARCES.

7.3 Characteristic Function Correlation Detectors: Summary

pool. If an entry is found with a sufficiently high correlation coefficient, we can most likely ignore the new event since it seems that we already have a template which is likely to detect subsequent occurrences of this signal. If no entry in existing detection lists is found, or if there are detections which are marginal, it is probably sensible to include the new event as a waveform template in order to detect further occurrences of this signal, including those which are too weak for the incoherent method to detect.

7.3 Summary

An extensive aftershock sequence has been studied whereby the signal from the main event did not constitute a suitable template for the detection of subsequent seismic events. Using various transformations of the signal from the main event (e.g. beamforming, filtering, rotation) we can construct a template of characteristic functions for the extended source region which constitutes an *incoherent correlation detector*. This detector successfully identified large numbers of the aftershocks with a very low false alarm rate. It is likely that the simultaneous running of an incoherent detector together with numerous standard correlation detectors will allow a more complete characterization of the aftershock sequence than would be possible if only the traditional matched filter detectors were to be used.

Chapter 8

Autonomous Correlation Detection Frameworks

8.1 Motivation for Developing an Autonomous Detection Framework

Correlation detection is becoming a mainstream option for network operations due to its advantages in sensitivity and event screening. Correlation detectors wrap detection, location and event identification functions into a single operation, which potentially makes them effective at reducing the burden of analysts in network monitoring operations. Since such detectors both detect and classify, they support a strategy of reviewing repeating events from particular sources as aggregated groups rather than one event at a time. This feature may lead to significant efficiencies during aftershock sequences and swarms and in regions with large amounts of mining activity.

However, the development of correlation detectors is labor intensive as currently practiced. Typically, large numbers of events must be assembled from catalogs or by running power detectors over continuous data streams. Cross-correlations among the event waveforms are computed and events are grouped by a clustering algorithm, which brings together events with significant waveform similarity. Waveforms from the clustered events then are selected to define a correlation template, which can be applied as a matched filter to the data in

a continuous stream. Subspace detectors, which are higher-dimensional extensions of correlation detectors, require even more care in their construction. They require careful alignment of waveforms from multiple events in a cluster, construction of an orthonormal basis for the event waveforms and selection of an optimum basis dimension for signal representation. In return, they provide greater scope for signal representation.

As these detector design activities currently are manual or only semi-automated, it is not possible to keep up with the occurrence of swarm events or aftershock sequences. However, one of the most attractive potential applications of correlation detectors is as a real-time screen for the very large number of similar events that can overwhelm network operations.

It also is the case that a large number of sources surround many stations, requiring distinct, dedicated detectors. The number of detectors required is compounded by the fact that the detailed structure of signals may change over time for many of these sources, requiring detector updates.

Our solution to these problems is to automate correlation detector development to the greatest extent possible, eventually under analyst review. The objective in seeking automation is not to replace analysts, but rather to assist them by organizing detected events into categories prior to review. We anticipate that even a partial organization of events can substantially reduce the burden of event formation and review.

8.2 Proposed Solution

Our proposed solution from the outset of this project is shown in Figure 8.1. We intended this framework to be a research tool designed to explore strategies for autonomous detector development. The framework would support simultaneous operation of numerous instances of detectors of several different types. At the core of the framework would be a list of conventional detectors (e.g. STA/LTA processes operating on a deployment of array beams) and correlation detectors. The purpose of the conventional detectors was to provide an event pool from which candidate events defining new correlation detectors could be drawn. The conventional detectors would provide an insurance policy against contingencies

that new sources may develop, or existing sources may transform to produce unrecognizable signals. Based on past experience, the correlation detectors were anticipated to provide significantly more sensitive detection of events at their target sources. There may be a great number of these, corresponding to a possibly large number of distinct repeating sources (e.g. mines, aftershock loci or swarms). Generally, we intended the framework to operate in the following manner:

1. Conventional detectors and correlation/subspace detectors operate simultaneously on data streams from each station in the network. All detections are sent to a list for further processing. Detections declared by conventional detectors only would, in addition, be sent to a pool for investigation as possible defining events for new correlation detectors. Detections declared by correlation/subspace detectors only would be logged immediately to event catalogs and marked as new events at currently recognized sources. Detections declared by both would be treated as correlation/subspace detections.
2. Detections in the pool would be subjected to a variety of checks intended to screen out noise triggers and to identify interesting candidates for possible promotion as correlation master events. The principal check is to correlate the new waveform against all others in the pool to look for a correlating pair. This check may be incorporated into an incremental clustering operation triggered by the arrival of a new event in the pool. When enough events accrue to a cluster, a subspace detector can be constructed. Other checks (e.g. on signal duration or bandwidth) might be required to guard against glitches and dropouts. These tend to produce spurious clusters.
3. Add the new correlation/subspace detectors to the detector list and continue operation.
4. As an option, redesign correlation/subspace detectors each time they detect, using the newly-detected waveform to update the correlation template. For subspace detectors, an elegant update to the subspace basis is possible. For correlators, a new correlator may be added to the bundle representing the source, and other correlators, not having produced triggers for a while, could be dropped.

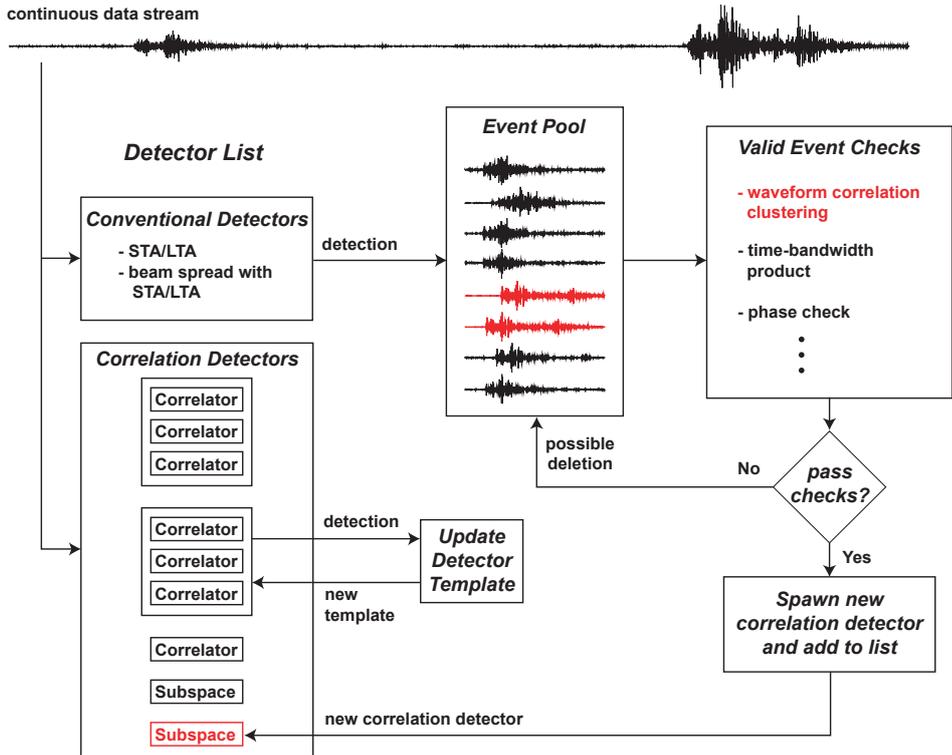


Figure 8.1: Detection framework proposed at the outset of this project. At the core of the framework is a list of detectors that operate simultaneously on a continuous data stream. Conventional power detectors are permanent members of the list and serve to detect new events for which no prior observations can serve as templates. These detectors were intended to populate an event pool which would be used to spawn new correlation detectors. Checks for event validity are intended to be sure that triggers on dropouts and noise bursts would not lead to a proliferation of “junk” correlators. The most significant test would be an automated clustering algorithm based on waveform correlation among all event pairs in the pool.

8.3 Autonomous Correlation Detection Framework: Implementation

We chose to implement the detection framework in Java, as (1) we had implementations of subspace detectors already in this language, (2) an object-oriented language facilitates construction of an adaptive framework, and (3) Java is platform independent.

At first, we implemented a simple STA/LTA detector on the continuous data stream from a single sensor. Our initial implementation had no screens or clustering operation, but simply promoted the waveform from every power detection as the template of a new correlation detector automatically. This is the simplest implementation imaginable, but allowed a quick look at performance and issues to be addressed. This system was applied to a relatively short (25 hour) sample of data from a station close to Mt. St. Helens, and another day long sample of a borehole station near a low-magnitude aftershock sequence near Orinda, California. These sequences were chosen for testing because they have many hundreds of events in a short period of time (thus testing could proceed without processing months of data; see Figure 8.3), and, at least in the case of the Mt. St. Helens data, the events are well separated in time (Figure 8.3), which minimizes complications with superimposed events corrupting correlation detector templates.

In these sequences, we learned that a policy of spawning correlation detectors directly from individual STA/LTA detections did not produce a huge number of correlation detectors (14 in the case of the Mt. St. Helens data with 409 detections, STEVE in the case of the Orinda sequence with STEVE detections). The system, as implemented, also was exceedingly fast: approximately 89,000 seconds of 100 sps, single channel Mt. St. Helens data were processed in 8.5 seconds. We are using an aggressive decimation scheme involving a single-sideband complex analytic representation of the data for efficient implementation of waveform correlations (Harris & Paik, 2006). Consequently there is not a huge penalty to pay for aggressively spawning correlation detectors - hundreds of detectors easily can be implemented in real time, even with array data.

We considered adapting our original concept to eliminate the detection pool, but institute a detection supervisor. Our revised concept was to spawn correla-

8.3 Autonomous Framework: Implementation

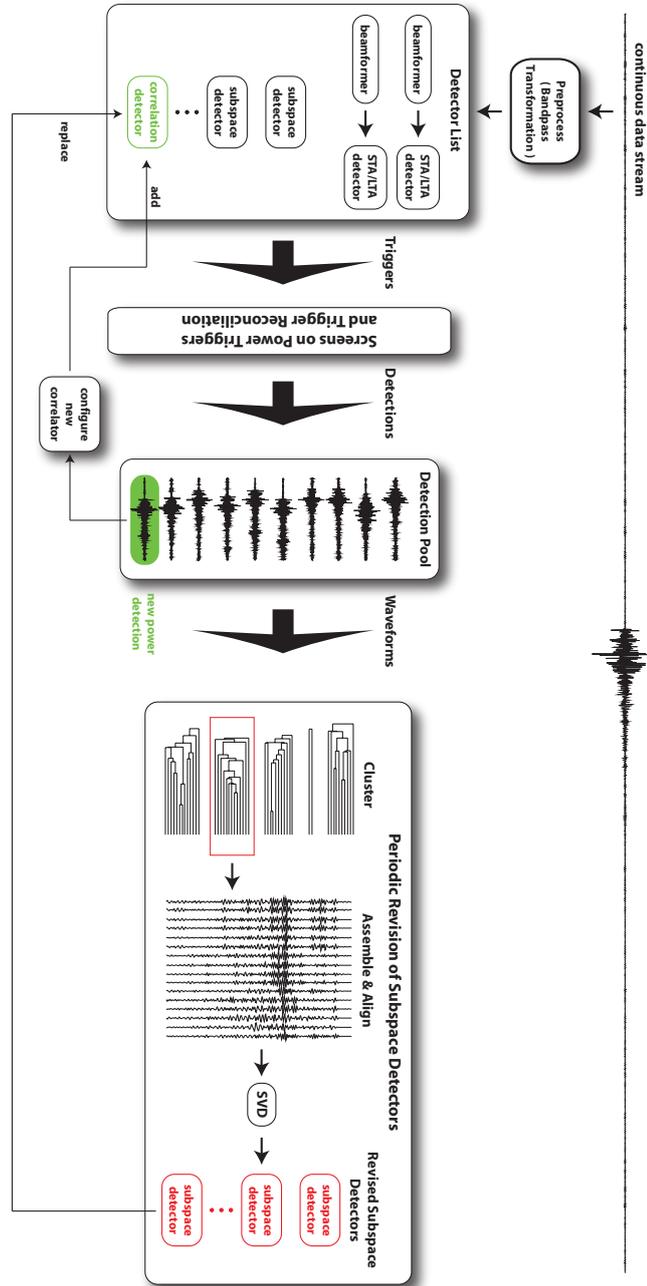


Figure 8.2: Diagram of the final system configuration.

8.3 Autonomous Framework: Implementation

tion detectors with every power detection, then monitor the performance of the detectors. We implemented a system to keep track of the number of triggers declared by each detector, and, in particular, the number of triggers declared in common by every pair of detectors. In the system, if two or more detectors of the same type declare simultaneous triggers, the detector with the largest detection statistic value has its trigger converted to a declared detection. Our intention was to retire detectors that triggered, but have no declared detections, and to consider aggregation of detectors that have many triggers in common. We learned from this exercise that many spawned detectors have triggers in common; in fact, groups of many detectors exist that have common triggers. We had intended to halt processing periodically to refine the list of detectors: detectors that had no declared detections might be retired (removed from the list) and detectors with common triggers might be merged and replaced by a single (ideally low-rank) subspace detector. After pruning and merging the detector list, processing would resume.

The existence of groups of many detectors with common triggers convinced us to modify our plan to monitor and trim the detectors. Since we need to assemble waveforms for a large number of detections from related detectors (i.e. with common triggers) in order to construct a subspace basis, we decided simply to assemble all detected waveforms periodically, perform all cross-correlations, cluster the detections on the basis of the correlations and create subspace detectors for each cluster. This change in strategy amounts to a complete overhaul of the set of correlation/subspace detectors periodically. It can be implemented on short lengths of data (we are working on individual aftershock sequences - 10 days of data in our current test). However, this policy cannot be implemented indefinitely as the number of detections will grow without bound. Consequently, we are considering more sophisticated strategies that would recluster events over some sliding window of recent history, perhaps extended to include events that have failed previous attempts at reclustering.

The system we finally built is displayed in Figure 8.2. The heart of the system is still a list of detectors with a set of fixed STA/LTA (power) detectors implemented on a collection of beams (to allow beam recipes to be implemented for an array). Correlation and subspace detectors are added to the list as described

8.3 Autonomous Framework: Implementation

below. The system acquires a block of array data from the continuous stream, preprocesses it (i.e. filters it into the desired frequency band and creates a single-sideband decimated signal). Each detector in the list is directed to process the block, i.e. to calculate a detection statistic from the block of data and examine the statistic for excursions above a predetermined threshold. A trigger is declared when the statistic exceeds the threshold. When two or more detectors produce triggers simultaneously, the triggers are compared, and only one is promoted as a detection subject to the following rules:

1. Triggers from the same type of detector (array power or subspace) are promoted or eliminated based on which has the largest detection statistic value. The trigger with the largest statistic is promoted.
2. Triggers from subspace detectors are always promoted over those from power detectors.

All detections are archived with information about the detector that originated them, the trigger time and the value of the detection statistic.

Detections from power detectors are assumed to be signals not yet seen: the system passes such detected waveforms through a series of screens (e.g. duration, bandwidth) in an attempt to eliminate spikes and other unwanted signal types. Waveforms that pass these tests are used to create correlation detectors which are added to the detector list. The system continues to on to the next block of data.

Periodically the system is halted to recalibrate the correlation detectors. The system halts when the number of newly detected events (since the last halt) exceeds some threshold. All of the events (from power and subspace detectors) are extracted from the archive (detection pool in Figure 8.2) and correlations are calculated between all event pairs. The correlation values are used to cluster the events and the correlation lags are used to align event waveforms from individual clusters. A subspace detector template is constructed from the aligned waveforms from each cluster, and the collection of subspace detectors so created replaces all of the correlation and subspace detectors in the detector list. At this point the system is restarted from the point where it left off in the stream.

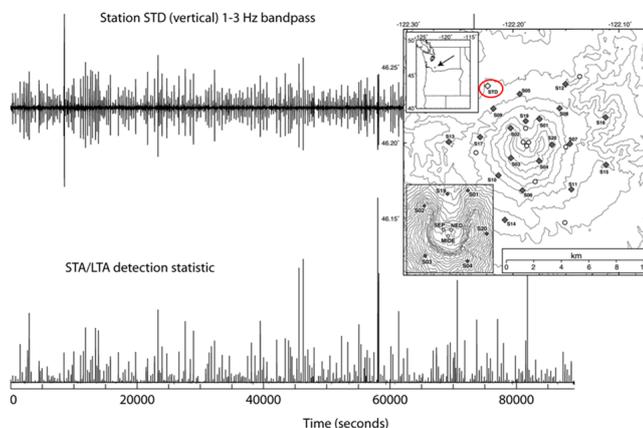


Figure 8.3: Twenty-five hours of data from station STD on the northwest flank of Mt. St. Helens, Washington state (map at upper right taken from [Waite *et al.*, 2007](#)). The very large number of events, apparent in the upper trace, is mirrored in the detection statistic from a simple STA/LTA detector, lower trace. Over 400 events occurred in 25 hours. See Figure 8.4.

8.4 Test Cases

8.4.1 Test Case: Mt. St. Helens Sequence

As mentioned earlier, we first tried a simple single-channel version of the system with a particularly favorable sequence of repeating events from Mt. St. Helens in Washington state. These are the so-called drumbeat events ([Waite *et al.*, 2007](#)) which are approximately periodic, very similar in size and usually well-separated in time. Figure 8.3 shows twenty-five hours of data from broadband station STD on the northern flank of the mountain. Detail of 60 minutes of data is shown in Figure 8.4, which clearly shows the events to be well-separated. This is a particularly benign case suitable for a first test of an automated system. Our system produced 14 detectors, capturing 409 events. We did not perform an exhaustive analysis of the completeness of the detections, but a quick calculation based on an examination of the one hour of data in Figure 8.4 suggests that about 400 events are expected.

In this test, the system simply spawned a correlation detector for every

8.4 Autonomous Framework: Test Cases

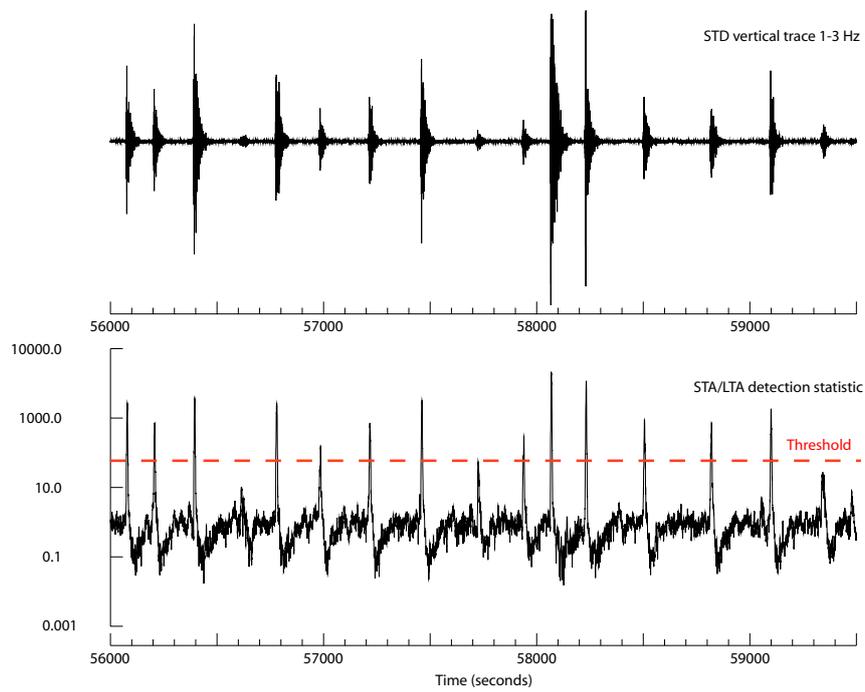


Figure 8.4: Detail of an hour of data from Figure 8.3. Note that the events are well separated, making this a particularly easy sequence to process, i.e. a good data set for a first trial. The output of the STA/LTA detector is shown in the lower panel (log scale).

8.4 Autonomous Framework: Test Cases

STA/LTA detection of a new signal pattern. We developed the system to compare triggers of the STA/LTA detector and the correlation detectors - if triggers from both algorithms occurred simultaneously, we assigned the detection to the correlation detector and suppressed creation of an additional correlator. This principal worked quite well to limit the otherwise unrestrained creation of correlation detectors. Only 14 patterns emerged, meaning that the vast majority of power detections were suppressed by the occurrence of a correlation detection.

8.4.2 Test Case: Orinda Sequence

We also tested the simple algorithm on an aftershock sequence near Orinda, California observed at close range (a few kilometers) with a borehole sensor. This was our first multichannel test, since we used a three-component sensor. This also was a trial of our first effort to keep track of the performance of the spawned correlation detectors. Our intention was to determine whether correlation detectors had a significant overlap in the events that they detected, providing incentive for combining individual correlation detectors perhaps into higher-rank subspace detectors.

We ran the detection framework on 24 hours of data from station BRIB (HLZ, HLN, HLE components). The system created 13 correlation detectors, though only 7 of these subsequently produced detections (Table 8.1). Among the active 7, it was common for a few dominant detectors to “steal” detections from the remaining detectors. We had a rule in place that correlation triggers would be promoted to detections only if they had no competing triggers - or if they had the largest value of the detection statistic. We intended to institute a supervisory system which would observe detector performance and periodically merge detectors with any triggers in common. Our intention was to pool the events detected by related detectors, create a subspace template from the pooled waveforms, then replace the pair of correlation detectors with a subspace detector.

We kept statistics on the numbers of triggers created simultaneously by all pairs of detectors (Table 8.2) in order to decide which detectors should be merged. To our surprise, there was considerable overlap among all of the detectors. Consequently, we decided that we needed to pool events from all of the detectors, cluster all events based on pairwise waveform correlations, then spawn subspace detectors from all clusters. We performed a reclustering operation following completion of the processing of 24 hours of data. The seven groups of events looked like the cluster in Figure 8.5, which suggests that events can indeed be interpreted in automatically defined groups to reduce the effort of analysts.

8.4 Autonomous Framework: Test Cases

Detector#	Type	# Triggers	# Detections
1	STA/LTA	283	13
2	Correlator	383	61
3	Correlator	469	158
4	Correlator	495	284
5	Correlator	433	205
6	Correlator	126	8
7	Correlator	71	2
8	Correlator	103	45
9	Correlator	1	0
10	Correlator	1	0
11	Correlator	1	0
12	Correlator	1	0
13	Correlator	1	0
14	Correlator	1	0
Total		2363	776

Table 8.1: Triggers and detections for the Orinda Sequence

	2	3	4	5	6	7	8
2	383	24	47	121	18	18	23
3		469	350	6	89	30	0
4			495	6	108	43	43
5				433	12	27	19
6					126	21	14
7						71	16
8							103

Table 8.2: Triggers common to detector pairs

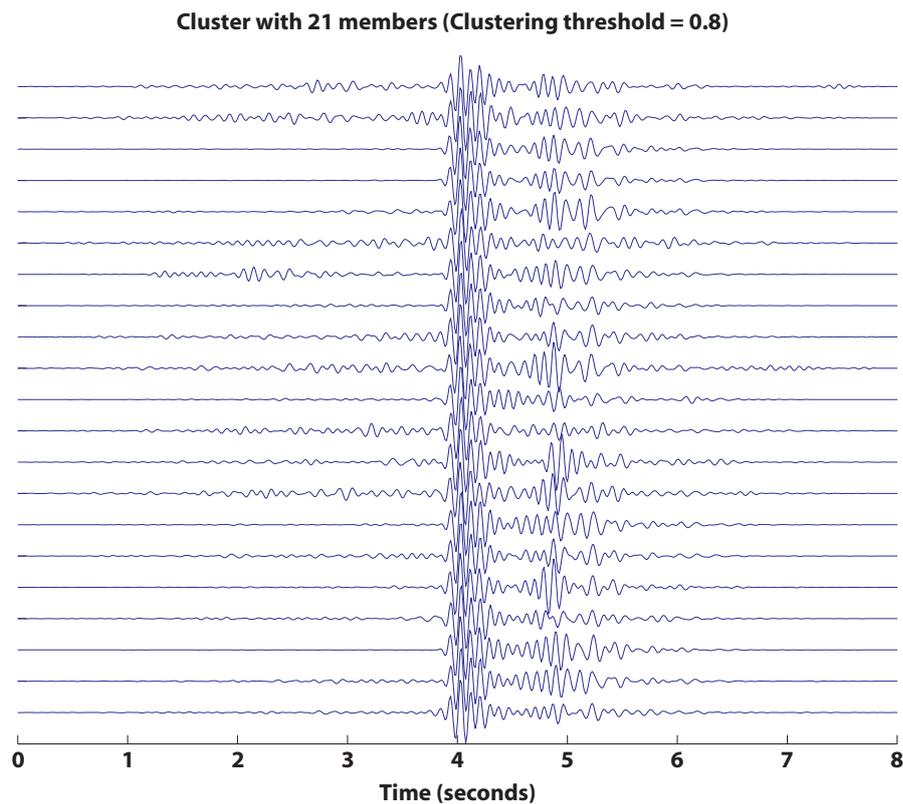


Figure 8.5: Waveforms from twenty-one events in one representative cluster after the reclustering operation for the Orinda Sequence. The point of this figure is that a large number of events can be organized by even a simple autonomous detection framework to speed the task of follow-on analyst interpretation and screening. Other clusters had more than 100 events. The events can be examined as a group rather than individually.

8.4.3 Final test: 2003 San Simeon Earthquake

As our main test of the system, we used the software to detect and cluster events occurring during the 2003 San Simeon earthquake (Figure 8.6). This was a moderately large event (mb 6.5) with thousands of aftershocks recorded by local networks in California. We acquired ten days of data (2003-356 - 2003-365) from the NVAR array for the test. NVAR is 390 kilometers from the main shock location. For ground truth information, we relied upon the Advanced National Seismic System (ANSS) composite catalog (NCEDC, 2009) which reported 1433 events in a 1x1 degree square around the main shock during the last ten days of 2003. For a suitable test, we required an event with large numbers of aftershocks in a short period of time and good Ground Truth information, observed by an array at regional distance with many high-SNR observations suitable for generating correlation templates. No examples of extensive aftershock sequences from Asia could be found for which the Ground Truth information was of comparable quality, and for which high quality array-data from regional distances was available.

Our objective in performing this test was not to reduce the detection threshold (as is a common objective with correlation detectors), but rather to group detected events automatically as an aid to analyst review. By this measure, a system is successful if it classifies automatically a large fraction of events as they occur or with periodic bulk processing during a sequence, without introducing large numbers of unwanted detections that are difficult to review. The system we implemented was able to process the 10 days of NVAR data (9 channels at 40 samples/second) in 15 to 20 minutes on a modern laptop computer. For ease in catalog reconciliation, and to increase the number of detections, we added a second processing step in which the 98 subspace detectors created during the first pass were held fixed and used to reprocess the entire ten days of data. This reprocessing step took an additional ten minutes. Since the system is so fast, periodic reprocessing of weeks of data is possible in a real-time monitoring operations.

We assume that, with such a system, analysts would not review events on an individual basis, but rather in groups determined automatically to be related on the basis of waveform similarity. Our results suggest that this approach is possible, and could form the basis of a change to current practice in network

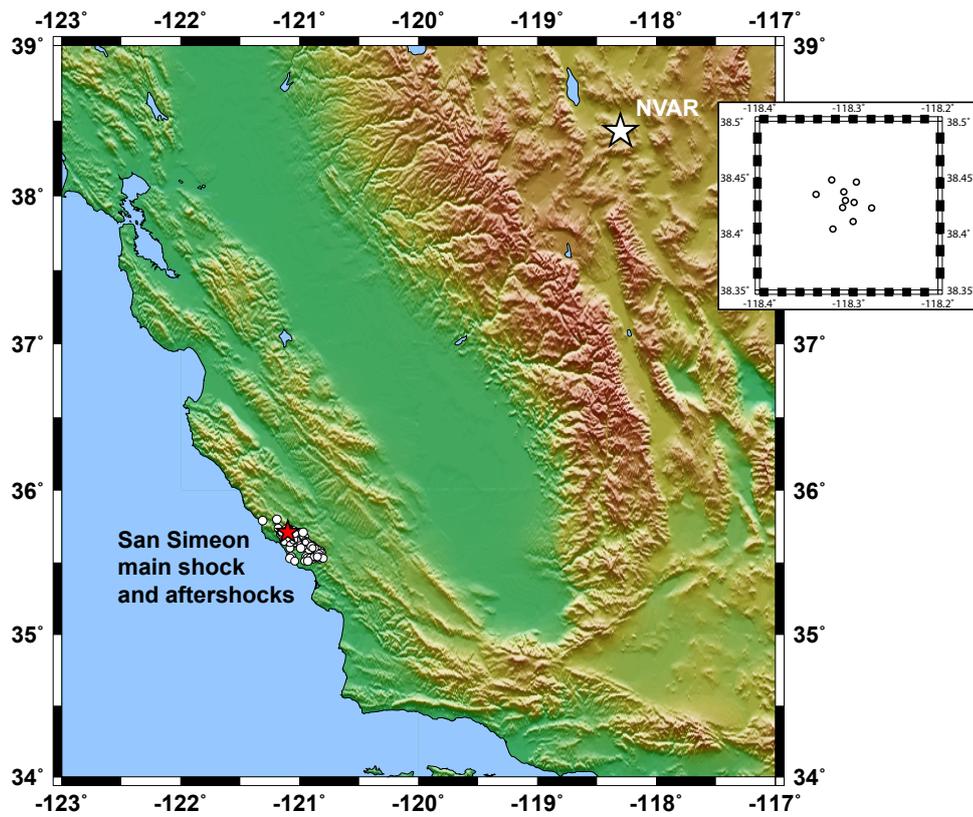


Figure 8.6: The main sequence of study consists of the magnitude 6.5 San Simeon earthquake and its aftershocks during the 10 day period Dec 22, 2003 - Dec 31, 2003. We acquired 10 days of data from the NVAR array to use in the test.

operations. We imagine a scenario where, immediately following the occurrence of a large event, an analyst directs the system to deploy a beam to the slowness of the main shock, anticipating aftershocks. The system then attempts to detect and cluster aftershocks for subsequent review. By reviewing events in groups, time and effort could be minimized, as only one or a few of the largest events in the cluster would require intensive interpretation, and the remainder could be treated as local related events of similar origin.

In our implementation, we used a standard beamformer with 9 elements of the NVAR array (2 of the 11 elements had very significant problems with dropouts). We used the great circle path backazimuth of 220 degrees and a velocity of 8 km/sec as the beamforming parameters. An STA/LTA detector was used on the beam to make power detections. The STA duration was 5 seconds, the LTA duration was 50 seconds, and we inserted a 5 second gap between the STA and LTA windows. The detection threshold was set at 5 (in amplitude, 25 in power) in order to obtain high-quality signals for correlation templates.

We set the system to halt and recalibrate every time 200 new detections were made. We distinguished between two classes of subspace detectors: first generation and second generation. The first generation detectors were those created directly from STA/LTA detections; they were correlators, i.e. subspace detectors of dimension one. Second generation detectors were those created from a recalibration operation and could be subspace detectors of higher rank. A database of detections was maintained with each detection “linked” to its originating detector. During a reclustering interlude, all new detections of any type and all detections from first generation detectors not previously reclustered were assembled into a pool. The pool events then were correlated pairwise and reclustered using a single link algorithm with a correlation clustering threshold of 0.5. Subspace detectors were developed for each identified cluster, using an energy capture metric to define the dimension of the subspace (Harris, 2006). Events used to construct the new second-generation detectors then were reassigned in the database to these new detectors. Detectors left with no detections following reassignment were removed from the system. Second generation detectors were assumed to be mature: their detections were not subject to reclustering.

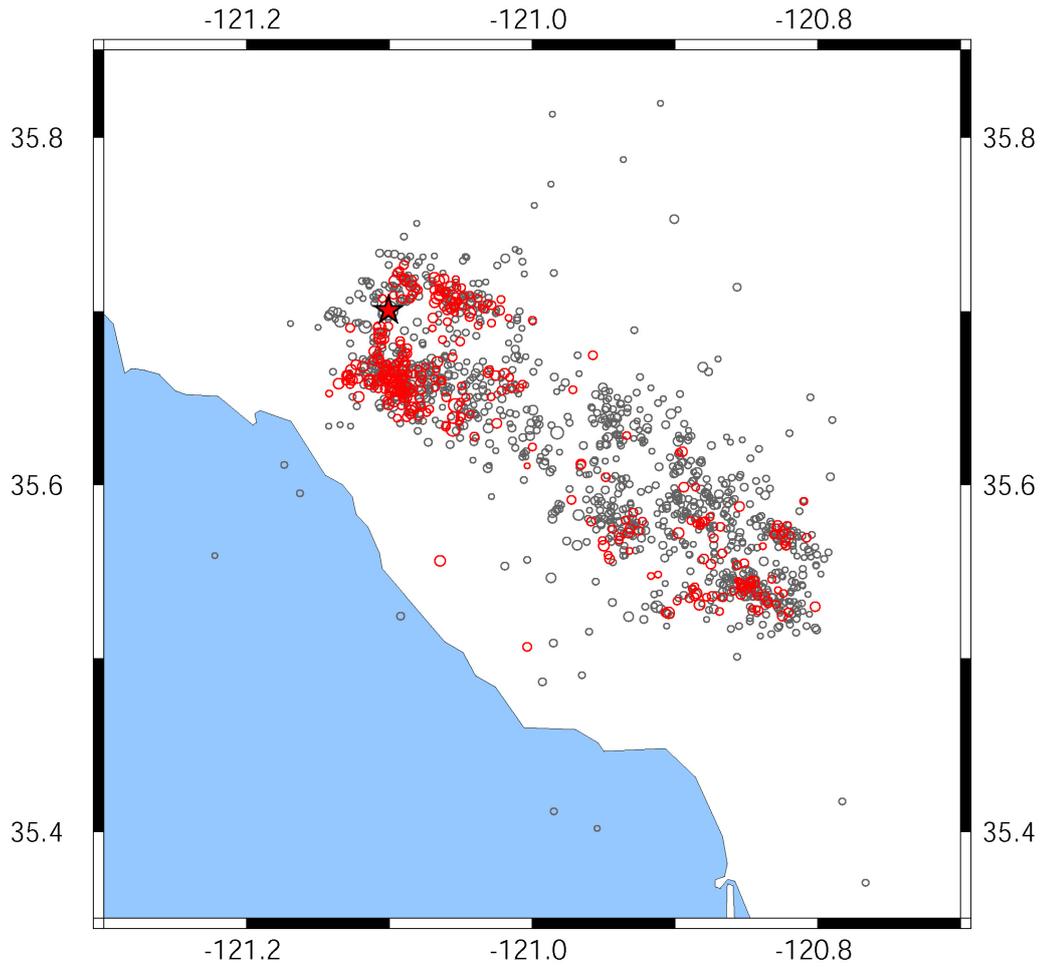


Figure 8.7: A map of the ANSS catalog events in the vicinity of the San Simeon main shock (grey) and the events detected automatically by our detection framework (red).

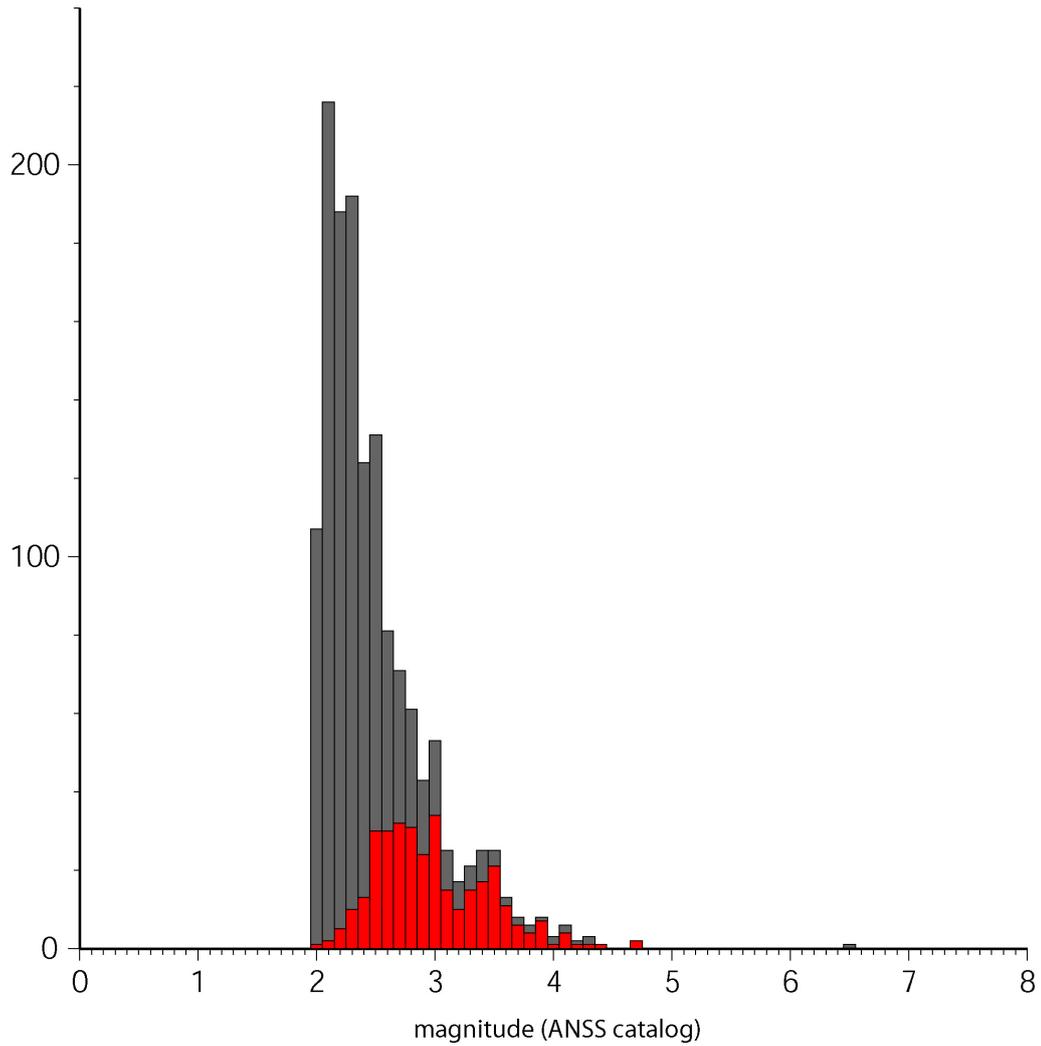


Figure 8.8: Histograms of the ANSS catalog events against magnitude (grey) and the events detected by our system (red). The system captured the majority of events above magnitude 2.8.

8.4 Autonomous Framework: Test Cases

This rather complex system was designed to control the number of detectors operating in the system. The number would tend to grow without bound without some process to identify and retire old detectors, replacing them with effectively merged detectors created as more events occur. Our hypothesis was that natural groupings among the aftershocks of the San Simeon sequence and other events observed by the array would be revealed only as events accumulate over time. In our view, the system needed to be flexible to allow detectors generated early in the process to be replaced as more events allowed better correlation and subspace templates to be developed.

Over the course of processing the 10 days of data, the STA/LTA algorithm spawned 124 correlation detectors directly. The system halted and reclustered three times, ultimately reducing the number of correlation/subspace detectors to 98 (in addition to the beamformer). As mentioned earlier, we reprocessed the data completely a second time (starting over from the beginning) using the detectors in existence at the end of the first pass. The system was cleared of all detection records and entirely new detections were declared during the second pass. This operation had the effect of putting detections from a single detector on a common relative timing basis, which greatly simplified reconciliation against the ANSS catalog. It also increased the number of correlation and subspace detections, since templates developed from high-SNR events occurring late in the 10 day period matched smaller events occurring earlier in the sequence that were overlooked in the first pass.

Overall, the second reprocessing step produced 702 detections, of which 360 could be reconciled against the ANSS catalog. The locations of the detected events are indicated as red symbols in Figure 8.7 against the background of grey symbols representing the ANSS catalog. Note that most of the larger events were captured by the system and that the detected events were distributed over most of the aftershock region.

Figure 8.8 compares histograms of ANSS catalog events and our detections as a function of event magnitude. These plots indicate that the system captured the majority of events above magnitude 2.8. We believe that a number of the higher magnitude events were missed due to the large duration of our long-term average in the beamforming detector. Many of these events occurred early in

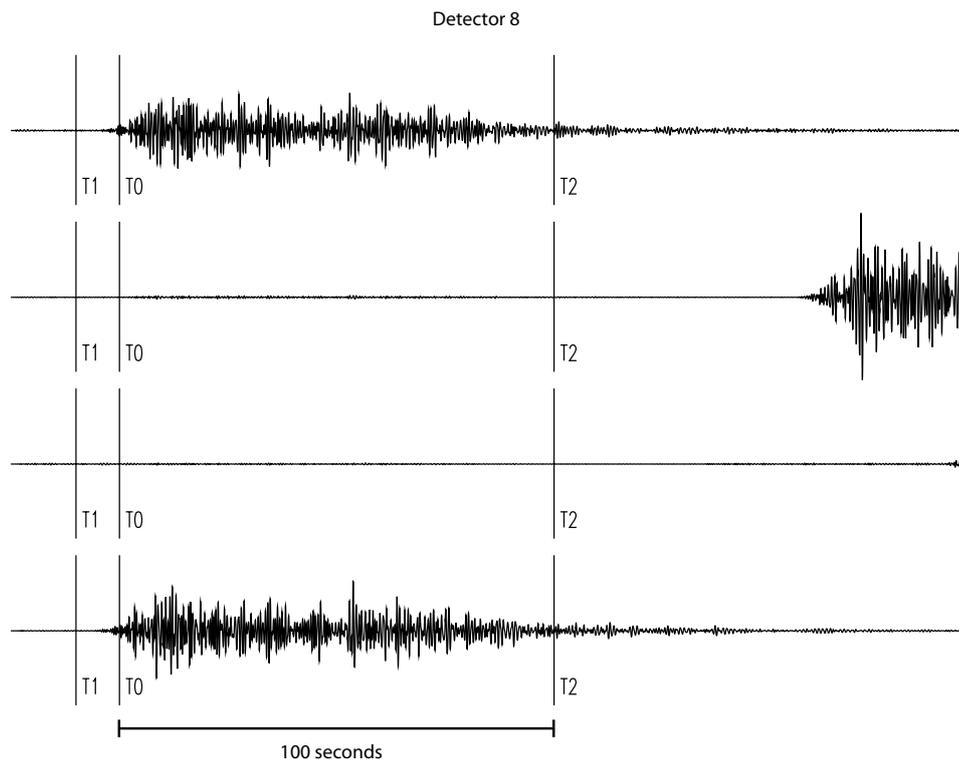


Figure 8.9: Example waveforms of detections associated with one detector (#8). The window used to create detectors is 110 seconds long and delimited by markers labeled T1 and T2. T0 is the nominal detection point (actually the point of detection by the STA/LTA algorithm). Common scaling of the traces causes the middle two events, which were smaller, to be invisible. The similarity of the traces allows multiple events to be interpreted simultaneously.

8.4 Autonomous Framework: Test Cases

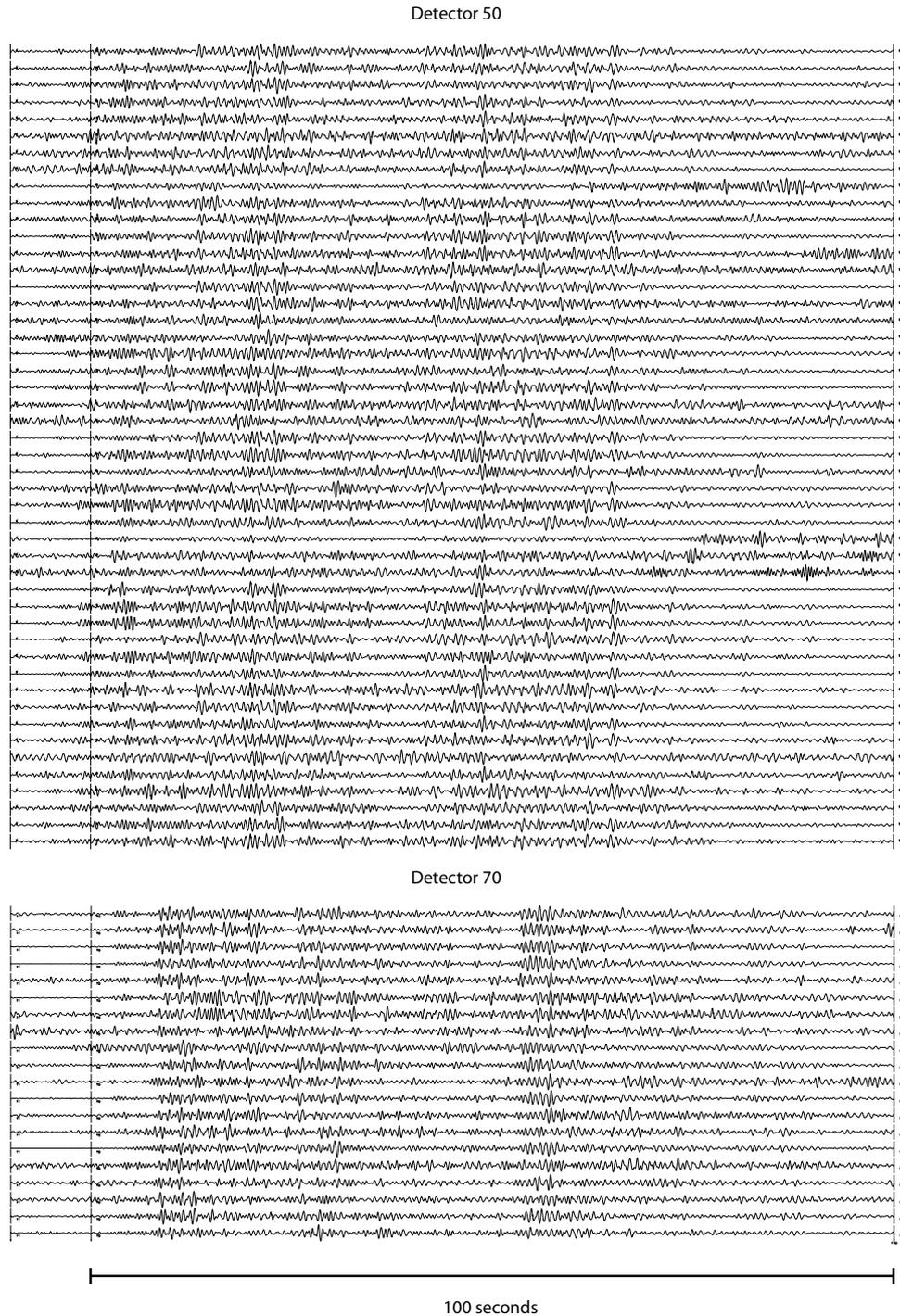


Figure 8.10: Two of the larger clusters of San Simeon aftershocks.

8.4 Autonomous Framework: Test Cases

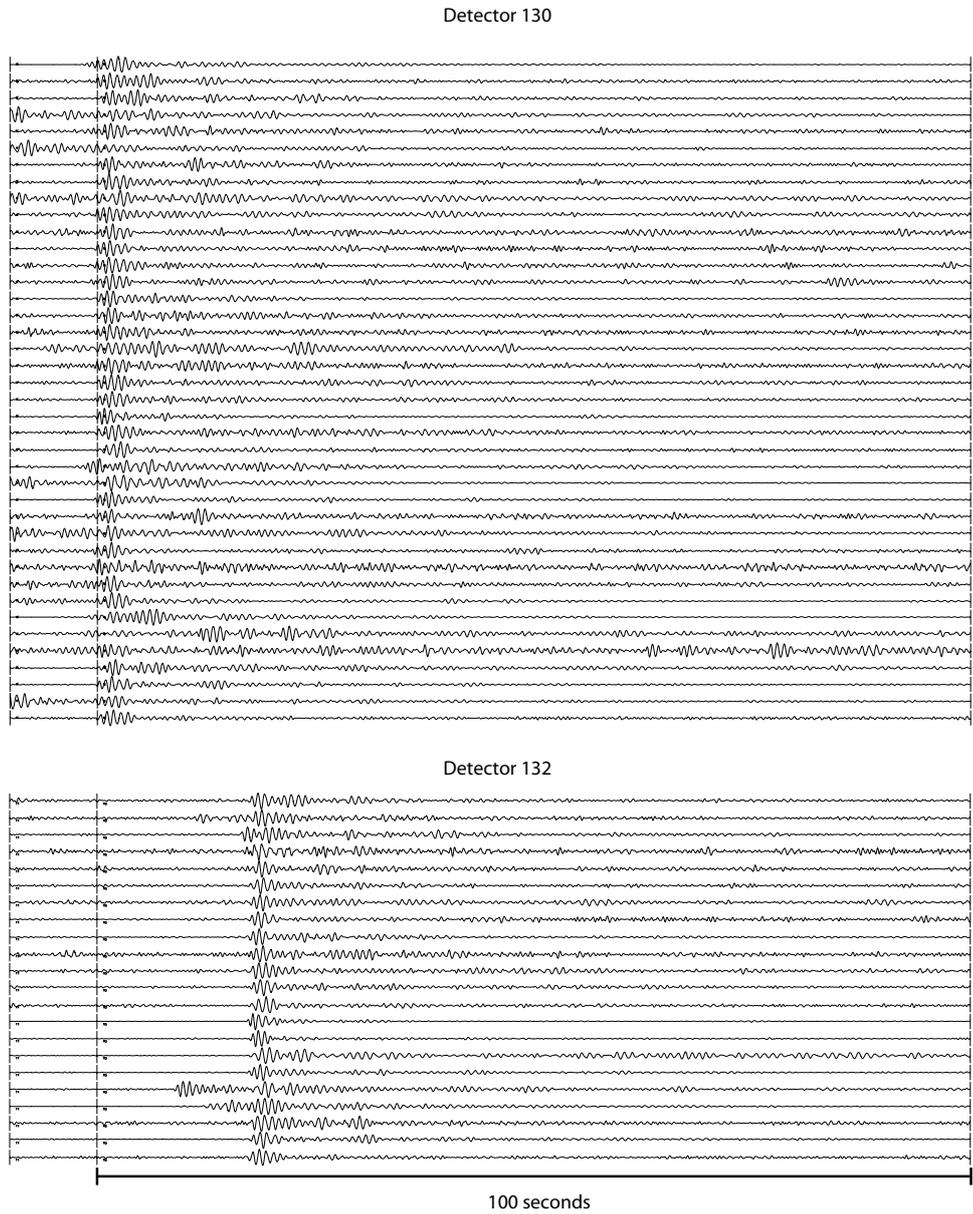


Figure 8.11: Two of the larger clusters of detections unrelated to the San Simeon sequence.

8.5 Autonomous Framework: Summary

the sequence when large numbers of events packed together, loading the LTA and reducing the STA/LTA ratio below the detection threshold. By missing the additional templates provided by these high magnitude events, we also lost many associated lower magnitude events that might otherwise have been detected.

The key issue is whether the events are clustered in some fashion that supports a reduction in analyst effort. Figure 8.9 shows four detections from one detector, and demonstrates that the waveforms can be aligned automatically for comparison based on waveform correlation measurements. Several of the clusters have a large number of events (Figure 8.10) which should aid rapid interpretation of the sequence.

An issue of major concern is whether automatically spawned detectors introduce a large number of spurious detections, which could overwhelm any advantage gained by organizing aftershocks into compact groups for interpretation. We note that the San Simeon aftershocks occurred in 75 clusters containing 360 detections. The other 342 detections were split among 23 clusters. These detections were predominantly legitimate seismic detections from sources other than the San Simeon sequence. The largest two clusters of these detections are shown in Figure 8.11. These events (in fact most of the other detections) appear to be very local signals that may have entered the system through a sidelobe of the array processor. They are relatively easy to distinguish and occur in large groups that would simplify their interpretation and dismissal.

8.5 Autonomous Correlation Detection Framework: Summary

Overall we believe this work demonstrates that self-calibrating, autonomous correlation detection frameworks are feasible. The system we describe is a pilot; it is clear that many improvements are possible. In particular, significantly more attention should be paid to the power detector used to originate waveform correlation templates. Our system missed a significant number of the higher magnitude events, in part due to inter-event interference early in the sequence. The loss of templates from the larger events has a cascading effect, potentially eliminating

8.5 Autonomous Framework: Summary

large numbers of lower magnitude detections. Window selection is another area of potential improvement. We used a particularly simple algorithm, defining the template window to start and end with fixed offsets from the STA/LTA detection point. This choice of window propagates thereafter into all generations of correlation detectors. A better approach might examine windows about the detection point for temporal coherence among events within a cluster to optimize window selection. The availability of many snapshots of the same signature could be used to advantage in defining the template window. Finally, though our system did not create large numbers of spurious detections which were difficult to dismiss, potential for mischief in automatic detector creation remains. One way to suppress this problem is to extend the construction of correlation detectors to networks of stations or arrays. A network-wide correlator may have the effect of suppressing interfering signal sources local to individual stations.

Chapter 9

Conclusions

The overall objective of this three-year study has been to develop and test a new advanced, automatic approach to seismic detection using waveform correlation, with special application to seismic arrays. The principal goal is to develop an adaptive processing algorithm. By this we mean that the detector is initiated using a basic set of reference (master) events to be used in the correlation process, and then an automatic algorithm is applied successively to provide improved performance by extending the set of master events selectively and strategically. These additional master events are to be generated by an independent, conventional detection system. A periodic analyst review would then be applied to verify the performance and, if necessary, adjust and consolidate the master event set.

We have examined the performance of waveform correlation detectors in a number of diverse applications within the field of detection seismology. Multi-channel systems have been studied in situations ranging from a single three-component station to small aperture arrays and large networks. In addition, we have examined the performance of multi-dimensional correlation detectors: so-called *subspace detectors*.

In chapter 2, we considered a sequence of natural earthquakes in northern Norway of which three (with magnitude greater than 2.0) were well-recorded by several IMS seismic arrays at distances over 600 km. Using the waveforms on the NORSAR array from the largest event (magnitude 3.5, on 2005 June 24) as a signal template, a multi-channel correlation detector was run on continuous array data and detected, with a low false alarm rate, 31 events between January

and December 2005. Local network stations within 15 kilometers of the event epicenters confirmed that each of these detections corresponded to actual events with magnitudes down to 0.5. Many of these events occurred in rapid succession of each other such that the signals at regional distances were hidden within the coda of preceding events. The improvement in event detectability, where applicable, using correlation detectors over conventional detectors is of the order one magnitude unit. We point out that the actual improvement is somewhat better for large aperture arrays or networks where coherent array processing cannot be used to improve the signal-to-noise ratio.

Chapter 3 considered a comparison between correlation and subspace detectors in a seismic sequence where waveform dissimilarity between events provides an evident challenge to the applicability of correlation methods. The correlation procedure was carried out on data from the NVAR array at a distance of approximately 270 km, and Ground Truth was again provided by data from a local station. The subspace detector in this case was found to perform only modestly better than the correlation detector selected. Large numbers of events went undetected by either correlation or subspace detectors, suggesting that the smaller magnitude events in this sequence may display greater diversity in mechanism than the larger events used to design the detection templates.

Chapter 4 provides a brief overview of the detection by correlation on ARCES array data of signals from near-surface explosions near the northern coast of the Kola Peninsula. Low waveform similarity between events would be assumed to preclude the use of a correlation detector on such a sequence. However, by considering the alignment of the single-site correlation coefficient traces (none showing very significant local maxima at any given time) and performing f-k analysis, we are able to detect a large number of these events with a very low false alarm rate. We conclude that waveform correlation on array data should not be ruled out as an effective means to detect events of interest provided that the events are essentially co-located and that the waveform dissimilarity is primarily the result of differing source-time functions.

Throughout this three year study, we have encountered instances of erroneous instrumental timing. We demonstrate two such cases in Chapter 5. The first concerns a timing error on an individual site within an array. A cross-correlation

and least squares inversion allows for very accurate “snapshot” estimates of timing errors when repeating seismic events are detected. The second concerns a timing error on a remote three-component station during a two-month period in 2006. A fortuitous sequence of rockbursts at a nearby coal mine, recorded by two different stations, could be exploited to measure, and hence correct, this drifting clock time.

Chapter 6 examines how far apart two seismic events may be such that one can still be detected using the signals from the other as a waveform template. Using an argument of reciprocity, we point out that correlations between the signals on different sites of a regional array indicate a correlation footprint of only a few hundred meters for a high frequency regional signal recorded on a single channel. A study of detection by correlation on a marine seismic profile suggests that this estimate is reasonable, and that the stacking procedure over all sensors of the array expands the correlation footprint to the order of 2 km.

Strategies for further expansion of the correlation footprint, and a controlled increase of the number of master events considered, are examined in Chapter 7. Instead of correlating waveform data, transformations of the raw data which provide characteristic properties of the wavefield may be correlated to form a less sensitive but more widely applicable form of matched filter detector. It is suggested that such a detector could be run in parallel with a sequence of classical correlation detectors in an adaptive framework.

In chapter 8, we consider experimental implementations of a framework for the autonomous calibration of correlation and subspace detectors. We consider three test-cases of increasing complexity and increasing relevance to strategies necessary in detection seismology.

1. *“Drumbeat” events near Mt. St. Helens*

This test case considers volcanic seismicity restricted to a very limited spatial extent in the vicinity of Mt. St. Helens. Monitored only by a single station, we tested an algorithm which simply spawned a correlation detector if an STA/LTA detection occurred for which the signal was not detected by the existing correlators. An apparently fairly exhaustive classification of

the seismicity was obtained using only 14 correlation detectors in the time interval studied.

2. *Aftershock sequence near Orinda, California*

In this single station study, it was found that very many events were shared by one or more of the initial correlators necessitating a recalibration of the detectors. Based upon a reclustering analysis, this resulted in the spawning of new sets of subspace detectors.

3. *Aftershock sequence from the December 22, 2003, San Simeon earthquake*

This aftershock sequence consisted of many thousands of events, covering a broad geographical footprint, and was observed by a seismic array at a regional distance. This makes the case study highly pertinent to many problematic sequences encountered in routine explosion monitoring. Despite a large number of subspace detectors covering a correspondingly large number of clusters of seismicity, a number of events were still missed. This was attributed, at least in part, to failures by the conventional detection system to generate suitable templates resulting in a system which was only partly able to characterize the observed seismicity.

The last of the test cases considered in particular emphasizes the need to coordinate the different procedures which need to be applied in our detection strategy. There are diminishing returns for operating a sophisticated event clustering procedure and hierarchy of subspace detectors if the auxiliary systems for conventional signal detection and parameter estimation are not optimal. A synergy exists between the different components of a detection and classification system and we need to consider, for example, how the results obtained from the correlation detectors and cluster analysis can be used to improve the beamforming and STA/LTA detection procedure. It may be the case that very non-optimal processing parameters are being applied for some target regions, and aligned signals from automatically classified clusters of events provide the ideal starting point for refining existing procedures.

References

- BAISCH, S., CERANNA, L. & HARJES, H.P. (2008). Earthquake Cluster: What Can We Learn From Waveform Similarity? *Bull. Seism. Soc. Am.*, **98**, 2806–2814. [57](#)
- GELLER, R.J. & MUELLER, C.S. (1980). Four Similar Earthquakes in Central California. *Geophys. Res. Lett.*, **7**, 821–824. [9](#), [51](#), [53](#)
- GIBBONS, S.J. (2006). On the Identification and Documentation of Timing Errors: An Example at the KBS Station, Spitsbergen. *Seism. Res. Lett.*, **77**, 559–571. [47](#), [49](#)
- GIBBONS, S.J. & RINGDAL, F. (2005). Detection of the aftershock from the 16 August 1997 Kara Sea event using waveform correlation. NOR SAR Scientific Report: Semiannual Technical Summary No. 2 - 2005, NOR SAR, Kjeller, Norway. [58](#)
- GIBBONS, S.J. & RINGDAL, F. (2006). The detection of low magnitude seismic events using array-based waveform correlation. *Geophys. J. Int.*, **165**, 149–166. [10](#), [12](#), [14](#), [16](#), [21](#), [35](#), [40](#), [42](#), [44](#), [49](#), [54](#)
- GIBBONS, S.J., KVÆRNA, T. & RINGDAL, F. (2005). Monitoring of seismic events from a specific source region using a single regional array: a case study. *J. Seismology*, **9**, 277–294. [60](#), [64](#)
- GIBBONS, S.J., BØTTGER SØRENSEN, M., HARRIS, D.B. & RINGDAL, F. (2007a). The detection and location of low magnitude earthquakes in northern

REFERENCES

- Norway using multi-channel waveform correlation at regional distances. *Phys. Earth Planet. Inter.*, **160**, 285–309. [23](#), [24](#), [25](#), [40](#), [42](#), [43](#), [49](#), [57](#)
- GIBBONS, S.J., RINGDAL, F. & KVÆRNA, T. (2007b). Joint seismic-infrasonic processing of recordings from a repeating source of atmospheric explosions. *J. Acoust. Soc. Am.*, **122**, EL158–EL164. [52](#)
- GIBBONS, S.J., KVÆRNA, T. & RINGDAL, F. (2009). Considerations in phase estimation and event location using small-aperture regional seismic arrays. *Pure appl. geophys.*, in press. [34](#)
- HARRIS, D.B. (1989). Characterizing Source Regions with Signal Subspace Methods: Theory and Computational Methods. Tech. Rep. UCID-21848, Lawrence Livermore National Laboratory. [27](#), [63](#)
- HARRIS, D.B. (1991). A waveform correlation method for identifying quarry explosions. *Bull. seism. Soc. Am.*, **81**, 2395–2418. [52](#), [57](#)
- HARRIS, D.B. (2006). Subspace Detectors: Theory. Tech. Rep. UCRL-TR-222758, Lawrence Livermore National Laboratory. [27](#), [63](#), [83](#)
- HARRIS, D.B. & PAIK, T. (2006). Subspace Detectors: Efficient Implementation. Tech. Rep. UCRL-TR-223177, Lawrence Livermore National Laboratory. [27](#), [71](#)
- HAUKSSON, E., CHI, W. & SHEARER, P. (2003). Comprehensive Waveform Cross-correlation of Southern California Seismograms: Part 1. Refined Hypocenters Obtained Using the Double-difference Method and Tectonic Implications. *AGU Fall Meeting Abstracts*, D325+. [26](#), [28](#)
- HICKS, E.C., BUNGUM, H. & LINDHOLM, C.D. (2000). Seismic activity, inferred crustal stresses and seismotectonics in the rana region, northern norway. *Quaternary Science Reviews*, **19**, 1423–1436. [9](#)
- KOCH, K. & STAMMLER, K. (2003). Detection and Elimination of Time Synchronization Problems for the GERESS Array by Correlating Microseismic Noise. *Seism. Res. Lett.*, **74**, 803–816. [49](#)

REFERENCES

- MENKE, W. (2001). Using Waveform Similarity to Constrain Earthquake Locations. *Bull. seism. Soc. Am.*, **89**, 1143–1146. [52](#)
- MENKE, W., LERNER-LAM, A.L., DUBENDORFF, B. & PACHECO, J. (1990). Polarization and coherence of 5 to 30 Hz seismic wave fields at a hard-rock site and their relevance to velocity heterogeneities in the crust. *Bull. seism. Soc. Am.*, **80**, 430–449. [53](#)
- NAKAHARA, H. (2004). Correlation distance of waveforms for closely located events - I. Implications of the heterogeneous structure around the source region of the 1995 Hyogo-Ken Nanbu, Japan, earthquake ($M_W = 6.9$). *Geophys. J. Int.*, **157**, 1255–1268. [57](#)
- NCEDC (2009). ANSS catalog website, www.ncedc.org/anss/. [81](#)
- RICHARDS, P.G., WALDHAUSER, F., SCHAFF, D. & KIM, W.Y. (2006). The Applicability of Modern Methods of Earthquake Location. *Pure appl. geophys.*, **163**, 351–372. [23](#)
- RINGDAL, F. & KVÆRNA, T. (1989). A multi-channel processing approach to real time network detection, phase association, and threshold monitoring. *Bull. Seism. Soc. Am.*, **79**, 1927–1940. [34](#), [45](#), [60](#)
- RINGDAL, F., GIBBONS, S.J., KVÆRNA, T., ASMING, V., VINOGRADOV, Y., MYKKELTVEIT, S. & SCHWEITZER, J. (2005). Research in Regional Seismic Monitoring. In *Proceedings of the 27th Seismic Research Review, Rancho Mirage, California, September 20-22, 2005. Ground-Based Nuclear Explosion Monitoring Technologies*, LA-UR-05-6407, 423–432. [37](#)
- RUBIN, A.M. (2002). Using Repeating Earthquakes to Correct High-Precision Earthquake Catalogs for Time-Dependent Station Delays. *Bull. seism. Soc. Am.*, **92**, 1647–1659. [50](#)
- SCHAFF, D.P. (2008). Semiempirical Statistics of Correlation-Detector Performance. *Bull. Seism. Soc. Am.*, **98**, 1495–1507. [12](#)

REFERENCES

- SCHAFF, D.P. & RICHARDS, P.G. (2004). *Lg*-Wave Cross Correlation and Double-Difference Location: Application to the 1999 Xiuyan, China, Sequence. *Bull. seism. Soc. Am.*, **94**, 867–879. [x](#), [23](#), [24](#)
- SENS-SCHÖNFELDER, C. (2008). Synchronizing seismic networks with ambient noise. *Geophysical Journal International*, **174**, 966–970. [50](#)
- SHELLY, D.P., BEROZA, G.C. & IDE, S. (2007). Non-volcanic tremor and low-frequency earthquake swarms. *Nature*, **446**, 305–307. [54](#)
- STEVENS, J.L., GIBBONS, S., RIMER, N., XU, H., LINDHOLM, C., RINGDAL, F., KVAERNA, T. & MURPHY, J.R. (2006). Analysis and simulation of chemical explosions in nonspherical cavities in granite. *J. Geophys. Res.*, **111**, B04306. [3](#)
- VANDECAR, J.C. & CROSSON, R.S. (1990). Determination of Teleseismic Relative Phase Arrival Times Using Multi-Channel Cross-Correlation and Least Squares. *Bull. seism. Soc. Am.*, **80**, 150–169. [42](#), [43](#), [49](#)
- WAITE, G.P., CHOUET, B.A. & DAWSON, P.B. (2007). Eruption dynamics at Mount St. Helens imaged from broadband seismic waveforms: Interaction of the shallow magmatic and hydrothermal systems. *J. Geophys. Res.*, **113**, B02305. [75](#)
- WESSEL, P. & SMITH, W.H.F. (1995). New version of the Generic Mapping Tools. *EOS Trans. Am. Geophys. Union*, **76**, 329. [vii](#)
- WITHERS, M., ASTER, R. & YOUNG, C. (1999). An Automated Local and Regional Seismic Event Detection and Location System Using Waveform Correlation. *Bull. Seism. Soc. Am.*, **89**, 657–669. [60](#)