Geophysical Prospecting

Full-waveform Based Microseismic Event Detection and Signal Enhancement: The Subspace Approach

Journal:	*Geophysical Prospecting*
Manuscript ID:	Draft
Manuscript Type:	Original Manuscripts
Date Submitted by the Author:	n/a
Complete List of Authors:	Song, Fuxian; Massachusetts Institute of Technology, Dept. of Earth, Atmospheric, and Planetary Sciences Warpinski, Norm; Pinnacle/Halliburton, Toksöz, Nafi; Massachusetts Institute of Technology, Dept. of Earth, Atmospheric, and Planetary Sciences Kuleli, Huseyin; Massachusetts Institute of Technology, Dept. of Earth, Atmospheric, and Planetary Sciences
Keyword:	Microseismic monitoring, Seismic processing

Full-waveform Based Microseismic Event Detection and Signal Enhancement: The Subspace Approach

Fuxian Song[1], Norm R. Warpinski[2], M. Nafi Toksöz[1], and H. Sadi Kuleli[1]

[1] Earth Resources Laboratory, Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts, U.S.A.
[2] Pinnacle, A Halliburton Service, Houston, Texas, USA.

E-mail: fxsong@mit.edu; norm.warpinski@halliburton.com; toksoz@mit.edu.

Prepared for

Geophysical Prospecting

ABSTRACT

Microseismic monitoring has proven to be an invaluable tool for optimizing hydraulic fracturing stimulations and monitoring reservoir changes. The signal to noise ratio (SNR) of the recorded microseismic data varies enormously from one dataset to another, and it can often be very low especially for surface monitoring scenarios. Moreover, the data are often contaminated by correlated noises such as borehole waves in the downhole monitoring case. These issues pose a significant challenge for microseismic event detection. On the other hand, in the downhole monitoring scenario, the location of microseismic events relies on the accurate polarization analysis of the often weak P-wave to determine the event azimuth. Therefore, enhancing the microseismic signal, especially the low SNR P-wave data, has become an important task. In this study, a statistical approach based on the binary hypothesis test is developed to detect the weak events embedded in high noise. The method constructs a vector space, known as the signal subspace, from previously detected events to represent similar, yet significantly variable microseismic signals from specific source regions. Empirical procedures are presented for building the signal subspace from clusters of events. The distribution of the detection statistics is analyzed to determine the parameters of the subspace detector including the signal subspace dimension and detection threshold. The effect of correlated noise is corrected in the statistical analysis. The subspace design and detection approach is illustrated on a dual-array hydrofracture monitoring dataset. The comparison between the subspace approach, array correlation method, and array short-time average/long-time average (STA/ LTA) detector is performed on the data from the far monitoring well. It is shown that, at the same expected false alarm rate, the subspace detector gives fewer false

alarms than the array STA/LTA detector and more event detections than the array correlation detector. The additionally detected events from the subspace detector are further validated using the data from the nearby monitoring well. The comparison demonstrates the potential benefit of using the subspace approach to improve the microseismic viewing distance. Following event detection, a novel method based on subspace projection is proposed to enhance weak microseismic signals. Examples on field data are presented indicating the effectiveness of this subspace-projection-based signal enhancement procedure.

Keywords: Microseismicity, Full waveform, Signal processing, Monitoring

INTRODUCTION

Microseismic monitoring has become a valuable tool for understanding physical processes in the subsurface. Besides its most common use in hydrofracture monitoring, it is also widely used for reservoir surveillance, geothermal studies, and monitoring of CO_2 sequestration (Phillips et al. 2002; Maxwell et al. 2004; Warpinski 2009).

The occurrence of microearthquakes follows a frequency-magnitude power law relation similar to tectonic earthquakes (Maxwell et al., 2006). The majority of microseismic events occur in the low magnitude range with a typical Richter magnitude M_L<-1. Moreover, the recorded microseismic waveforms are usually contaminated by the high amplitude noise. In downhole monitoring of hydraulic fracturing, as is the case for this study, the high amplitude noise may come from various sources, most notably from the borehole waves excited by pumps located at the surface. Therefore, the recorded microseismic data normally have a very low signal-to-noise ratio (SNR). This low SNR poses a great challenge in processing microseismic data and leads to two major consequences. Firstly, the accurate time picking of the P- and S-wave arrivals for individual events becomes a difficult task, which impacts the accuracy of the microearthquake location and, indeed, the success of the microseismic monitoring. Secondly, the low SNR values set a detection limit. As such, the minimum detectable event magnitude increases with increased distance from monitoring geophones due to the increased signal attenuation with distance. This causes the viewing-distance bias, which can be a significant issue when interpreting the completeness of the fracture geometry (Maxwell et al., 2010; Warpinski, 2009).

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

Known methods for automated microseismic event detection include short-time-average/long-time-average (STA/LTA) detectors and correlation-type detectors. The STA/LTA detector calculates the energy ratio of short-time window to long-time window and declares the appearance of seismic events when the ratio exceeds a threshold (Earle and Shearer, 1994). The correlation detector screens seismic events by calculating a correlation coefficient between the received signal and a template event known as the master event, assuming events that are to be detected have similar waveforms as the master event (Gibbons and Ringdal, 2006; Song et al., 2010).

Simple STA/LTA detectors are broadly applicable, but suffer from high false alarm rates when an aggressive threshold is set to detect smaller signals. Correlation detectors are highly sensitive, having high detection probability at low false alarm rates. However, they are applicable only to repetitive sources confined to very compact source regions.

Unlike the above two approaches, a detection method based on statistical hypothesis testing has been proposed to take into account the statistics of both signal and noise (Bose et al., 2009). In their detection algorithm, the microseismic event signal recorded at the downhole geophone array is assumed to be a scaled and delayed version of a common trace. The common trace is modeled as a deterministic Ricker wavelet signal convolved with a finite impulsive response (FIR) filter. The FIR filter is determined by maximizing the detection likelihood. Although the common trace can be adjusted from one detection window to another, there is only one microseismic signal template in each detection window; it therefore faces similar difficulties as correlation detectors. Moreover, this method cannot take advantage of previously detected events.

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

In order to overcome these limitations, we adapt the subspace detection method of Harris (2006) to replace the single matching template in a correlation detector with a suite of basis vectors (known as the signal subspace) that are combined linearly to match occurrences of variable signals from a specific source region. We extend the surface monitoring setup to the downhole monitoring configuration and consider correlated noises with different variances on different geophone channels. We introduce a systematic procedure to determine the parameters for the subspace detector.

The subspace design and detection approach is demonstrated on a dual-array hydrofracture monitoring dataset. We compare the subspace approach, array correlation method, and array short-time average/long-time average (STA/ LTA) detector using the data from the far monitoring well. The additionally detected events from the subspace detector are further validated using the data from the nearby monitoring well. The comparison illustrates the effectiveness of using the subspace approach to improve the detection capability. Furthermore, we develop a subspace projection approach to enhance the SNR of detected microseismic signals. Signal enhancement results on the field dataset are presented.

Exposition in this paper is necessarily mathematical. The number of symbols is sufficiently large that a table of symbols has been included (Table 1). To keep the number of symbols to a minimum, a few conventions have been adopted. First, the underlined lower-case symbol indicates a column vector, while a matrix is shown as the underlined upper-case symbol. Second, a symbol with a “hat” denotes the estimated value. When it refers to the embedding space dimension, the effect of correlated noise has been corrected.

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

METHODOLOGY

Subspace detector theory: statistical binary hypothesis testing

Current practice in seismic event detection is concentrated at the extremes of a spectrum of possibilities determined by the amount of information available about the temporal structure of signals to be detected. On one end, incoherent energy detectors, such as STA/LTA detectors, assume little knowledge of the underlying signals. On the other end, correlation detectors assume a completely known signal and coherently use the fine temporal and spatial structure of detected seismic signals to enhance the sensitivity (Harris, 1991). STA/LTA detectors are broadly applicable, but are insensitive to waveform information and thus have a high false alarm rate, while correlation detectors are sensitive to waveforms and have fewer false alarms, but are less flexible and are applicable only to repetitive sources. Therefore, the subspace detector was proposed to manage the tradeoff between sensitivity and flexibility (Scharf and Friedlander, 1994; Harris, 2006).

The basic idea of subspace detection is to replace the single matching template in the correlation detector with a suite of basis vectors, known as the signal subspace, that are combined linearly to match the occurrence of variable signals from a specific source region. The subspace detector is implemented as a binary hypothesis test on a window that slides along a continuous data stream.

The microseismic data recorded at multiple channels are multiplexed into a continuous stream according to the following equation:

$$x((n-1)\cdot N_c + i) = x_i(n), \quad (1)$$

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

wherein $x_i(n)$ is the n-th time sample from the i-th channel, where $n = 1,2,\ldots;\ i = 1,2,\ldots,N_c$. In the downhole setup, the three component (3C) data from all levels of geophones are multiplexed. The data within each window may be presented as

$$\underline{x}(n) = [x_1(n)\ x_2(n)\ \ldots\ x_{N_c}(n)\ x_1(n+1)\ x_2(n+1)\ldots\ x_{N_c}(n+N_T-1)\ ]^T. \quad (2)$$

$\underline{x}(n)$ is an N*1 vector with $N = N_c \times N_T$, where N_T is the temporal length of the window.

The subspace detection is posed as a binary hypothesis testing problem. Under the null hypothsis (H_0), the windowed data $\underline{x}(n)$ are assumed to consist of noise. In the alternative hypothsis (H_1), the data consist of both signal $\underline{s}$ and noise $\underline{\eta}$.

$$\begin{cases} \underline{x}(n) = \underline{\eta} & \text{under hypothesis } H_0 \\ \underline{x}(n) = \underline{s} + \underline{\eta} & \text{under hypothesis } H_1 \end{cases} \quad (3)$$

The noise $\underline{\eta}$ is assumed to be zero-mean Gaussian noise with an unknown variance σ^2. It is also assumed to be temporally and spatially uncorrelated. The signal $\underline{s}$ is assumed to be deterministic but dependent upon a vector of unknown parameters $\underline{a}$, and is expressed as an unknown linear combination of basis waveforms:

$$\underline{s} = \underline{U}\,\underline{a}\,, \quad (4)$$

where the N*d matrix $\underline{U}$ represents d unknown signal subspace bases. The signal subspace dimension d may take any value from 1 to N. Without losing generality, $\underline{U}$ can be made orthonormal:

$$\underline{U}^T\,\underline{U} = I_d. \quad (5)$$

Therefore, energy captured in the signal subspace may be simplified to

$$E_c = \underline{a}^T\underline{a}. \quad (6)$$

Under the above assumptions, the probability function for the recorded data is

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

$$p\left(\underline{x}(n)|H_0\right) = \left[\frac{1}{2\pi\sigma^2}\right]^{N/2} \exp\left(-\frac{1}{2\sigma^2}\underline{x}^T(n)\underline{x}(n)\right) \tag{7}$$

under the null hypothesis H_0 (no event present), and

$$p\left(\underline{x}(n)|H_1\right) = \left[\frac{1}{2\pi\sigma^2}\right]^{N/2} \exp\left(-\frac{1}{2\sigma^2}\left(\underline{x}(n) - \underline{Ua}\right)^T\left(\underline{x}(n) - \underline{Ua}\right)\right) \tag{8}$$

under the alternative hypothesis H_1 (event present). The generalized log likelihood ratio can be derived as (Van Trees, 1968),

$$l\left(\underline{x}(n)\right) = \ln\left(\Lambda\left(\underline{x}(n)\right)\right) = \ln\left[\frac{\max\limits_{\{\underline{a},\sigma\}} p\left(\underline{x}(n)|H_1\right)}{\max\limits_{\{\sigma\}} p\left(\underline{x}(n)|H_0\right)}\right] = -\frac{N}{2}\ln[1 - c(n)]\,. \tag{9}$$

The ratio of energy in the signal projected into the signal subspace $\underline{U}$ to the energy in the original data is represented by the quantity c(n), known as the subspace detection statistics,

$$c(n) = \frac{\underline{x}_p^T(n)\underline{x}_p(n)}{\underline{x}^T(n)\underline{x}(n)}. \tag{10}$$

The projected signal $\underline{x}_p(n)$ is the least-squares estimate of the signal $\underline{x}(n)$ in the detection window,

$$\underline{x}_p(n) = \underline{U}\underline{U}^T\underline{x}(n). \tag{11}$$

The subspace detection statistics c(n) is a positive quantity with values ranging between 0 and 1. The generalized likelihood ratio test detects an event of interest if the generalized log likelihood ratio exceeds a certain threshold α, that is

$$l\left(\underline{x}(n)\right) = -\frac{N}{2}\ln\left(1 - c(n)\right) > \alpha. \tag{12}$$

It means that an event is declared if the subspace detection statistics c(n) is larger than a detection threshold γ,

$$c(n) = \frac{\underline{x}_p^T(n)\underline{x}_p(n)}{\underline{x}^T(n)\underline{x}(n)} > \gamma. \tag{13}$$

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

Subspace detector implementation: subspace design

To perform subspace detection based on equations 11 and 13, the first step is to construct the signal subspace bases $\underline{U}$. Harris (2006) suggested a way to build the signal subspace from previously detected seismic events. The objective in constructing the signal subspace representation $\underline{U}$ is to obtain acceptably accurate orthogonal bases for seismic signals characteristic of events of interest in the target source region. A representation with a larger dimension provides a higher possibility of detecting weak events by capturing more of the energy of an incompletely known signal. However, a higher order dimension representation may also be expected to increase the false alarm rate by allowing the detector to match noise with great probability. Consequently, a parsimonious representation with an adequate signal energy capture is desired. Assuming that D previously detected events are selected to construct the signal subspace, for a given representation order of d, the signal subspace bases $\underline{U}$ should capture as much energy in the D design set events as possible.

For each event in the design set, the aligned channel-multiplexed data vector may be written as,

$$\underline{s}(n) = [s_1(n)\ s_2(n) \ldots\ s_{N_c}(n)\ s_1(n+1)\ s_2(n+1) \ldots s_{N_c}(n+N_T-1)]^T. \quad (14)$$

The design data matrix $\underline{S}$ is assembled with D channel-multiplexed column vectors, with each column representing one design set event,

$$\underline{S} = [\underline{s}^1(n)\underline{s}^2(n) \ldots \underline{s}^D(n)]. \quad (15)$$

To prevent large events in the design set from dominating the design data matrix, data from each event are normalized to have unit energy, that is,

$$\underline{s}^{iT}(n)\underline{s}^i(n) = 1, \qquad i = 1,2, \ldots, D. \quad (16)$$

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

The singular value decomposition (SVD) of the design data matrix $\underline{S}$ is

$$\underline{S} = \underline{W}\,\underline{\Sigma}\,\underline{V}^{T} = \underline{W}\,\underline{A}\,, \tag{17}$$

where $\underline{A} = \underline{\Sigma}\,\underline{V}^{T}$ is the representation coefficient matrix. According to Eckart and Young (1936), the best approximation to $\underline{S}$ in the least-squares sense for a given order d is the truncated SVD of the matrix to the rank d. Consequently,

$$\underline{S} = \underline{W}\,\underline{A} = [\underline{W}_{d} \quad \underline{W}_{D-d}] \begin{bmatrix} \underline{\Sigma}_{d} & \underline{0} \\ \underline{0} & \underline{\Sigma}_{D-d} \end{bmatrix} \begin{bmatrix} \underline{V}_{d}^{T} \\ \underline{V}_{D-d}^{T} \end{bmatrix} \sim \underline{W}_{d}\underline{\Sigma}_{d}\underline{V}_{d}^{T} = \underline{U}\,\underline{A}_{d} \tag{18}$$

$$\underline{U} = \underline{W}_{d} \tag{19}$$

$$\underline{A}_{d} = \underline{\Sigma}_{d}\underline{V}_{d}^{T} \tag{20}$$

The matrix of coefficients $\underline{A}_{d}$ provides an expression of the energy captured in the signal subspace $\underline{U}$ corresponding to the first d largest singular values for the D events in the design set. Consider, for example, the i-th design set event,

$$\underline{s}^{i} \sim \underline{U}\,\underline{a}_{d}^{i}, \tag{21}$$

where $\underline{a}_{d}^{i}$ is the i-th column of $\underline{A}_{d}$, the fractional energy capture for this event is

$$f_{c}^{i} = [\underline{a}_{d}^{iT}\underline{a}_{d}^{i}]/[\underline{s}^{iT}(n)\underline{s}^{i}(n)] = \underline{a}_{d}^{iT}\underline{a}_{d}^{i}. \tag{22}$$

The average fraction of energy captured for all D events in the design set may accordingly be calculated as:

$$\overline{f_{c}} = \frac{1}{D}\sum_{i=1}^{D} f_{c}^{i} = \mathrm{trace}(\underline{\Sigma}_{d}^{T}\underline{\Sigma}_{d})/\mathrm{trace}(\underline{\Sigma}^{T}\underline{\Sigma}). \tag{23}$$

The fractional energy capture for each of the D events and the average fraction of energy captured for all D events may be plotted as a function of the dimension of representation d, also referred to as the signal subspace dimension. Each fractional energy capture curve extends from 0 to 1, and increases with increased dimension of representation. When d reaches a certain value, the average fraction of energy captured

for all design events exceeds a predetermined threshold (e.g., 80%). This may be used as an aid to determining d. In the field study section, an example will be shown to determine the signal subspace dimension d by generating and processing plots, as described above.

The question remaining is to select the D design set events to represent the source region of interest. The single-link clustering algorithm is used to serve this purpose (Israelsson, 1990).

A template event library is built upon the previously identified events, for example based on the output of a STA/LTA detector with a conservative threshold. The multi-channel time series data are converted to a single channel multiplexed data vector according to equation 1. Pair-wise correlation coefficients for the template events in the library are then calculated from the single channel multiplexed data vector. Assuming that there are M events in the library, the event dissimilarity matrix $\underline{K}$ with a size of M*M is constructed using the following equation:

$$K_{p,q} = 1.001 - \lambda_{p,q}. \quad (24)$$

$K_{p,q}$ may be viewed as a measure of inter-event distance in waveform similarity space for events p and q, where $\lambda_{p,q}$ is the maximum waveform correlation between events p and q.

Next, the events are automatically clustered based on the dissimilarity matrix $\underline{K}$. Various clustering algorithms may be used in different embodiments. The choice of clustering algorithm depends on objectives and expectations in characterizing a source. In the context of hydraulic fracturing, the assumption is often that the source region to be characterized through representative waveforms may have significant variation in source mechanisms, some variation in source time function and location, or some combination of all these attributes. To this end, in this paper, an aggressive algorithm for linking

events into relatively extensive chains spanning the space of waveform variations, in the form of a single-link method is employed.

The single-link algorithm aggregates event clusters based solely on the single pair of events (one event in each of two clusters under consideration for merging) with a large waveform correlation (Israelsson, 1990). The hierarchical agglomerative clustering procedure is presented as a dendrogram, as will also be illustrated in the field study section. At each clustering step, the cophenetic correlation coefficient is calculated to measure how well the clustering models the actual event dissimilarity behavior, which is described in matrix K (Rowe et al., 2002). The cophenetic correlation coefficient may serve as an aid to select the design set events. Sudden decreases in the cophenetic correlation indicate that the cluster just formed has made the dendrogram less faithful to the data and thus may suggest that the clustering process should be terminated between this cluster and the previous one. The events that have been clustered up to that point are then automatically selected as the design set events. The design set is a set of events that are to be used to construct the signal subspace bases. Therefore, it is desirable for the design set to not only represent the actual inter-event correlation behavior described by the original dissimilarity matrix K, but also to comprise most of the larger events in the event library. To this end, the waveform root-mean-square amplitudes of the automatically selected events are checked to ensure that most of the larger events in the library are included. The details of the single-link clustering algorithm and design set event selection are further illustrated in Appendix A.

Subspace detector implementation: parameter determination and performance evaluation

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

Once the signal subspace $\underline{U}$ is constructed, in order to conduct subspace detection based on equations 11 and 13, it is necessary to determine the detection threshold γ. Harris (2006) studied the distribution of the subspace detection statistics c(n) and derived the threshold using the Neyman-Pearson criterion (Van Trees, 1968). Under this criterion, the subspace dimension d is firstly determined by maximizing the probability of detection P_D for a fixed false alarm rate P_F. The threshold γ is then derived from the false alarm rate using the following equation:

$$\begin{cases} 1 - F_{d,N-d}\left(\frac{\gamma}{1-\gamma}\frac{N-d}{d}\right) = P_F \\ 1 - F_{d,N-d}\left(\frac{\gamma}{1-\gamma}\frac{N-d}{d}, \bar{f}_c \cdot N \cdot SNR, \left(1-\bar{f}_c\right) \cdot N \cdot SNR\right) = P_D \end{cases} \quad (25)$$

where P_F is evaluated from the cumulative central F distribution $F_{d,N-d}(\cdot)$ with d and (N-d) degrees of freedom under null hypothesis H_0, while P_D is expressed in terms of the cumulative doubly non-central F distribution with the same degrees of freedom and a non-centrality parameters of $(\bar{f}_c \cdot N \cdot SNR)$ and $\left[\left(1-\bar{f}_c\right) \cdot N \cdot SNR\right]$ for the numerator and denominator term, respectively. $\bar{f}_c$ is the average fraction of energy captured for all D design set events, defined in equation 23. N denotes the embedding space dimension, out of which d is the signal subspace dimension, and (N-d) is the dimension of the orthogonal complement of the signal subspace. SNR is the signal-to-noise ratio in the detection window, defined as

$$SNR = E/\sigma^2/N, \quad (26)$$

where E is the signal energy over the detection window of length N, and σ^2 denotes the unknown Gaussian noise variance. The details of the derivation of Equation 25 are further presented in Appendix B.

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

When deriving equation 25, two implicit assumptions are made. Firstly, the Gaussian noise variance σ^2 is identical across all N_c channels. This is not generally true. As we will see in the following field study section, the noise variances on different channels are estimated from the pre-event noise and the microseismic data from different channels are normalized by their relative variances before multiplexing (equation 2). Secondly, the noise in the detection windows is assumed to be statistically uncorrelated. If the detection window is N samples long, the dimension of the embedding space, where the signal subspace resides, is also N. As Wiechecki-Vergara et al. (2001) point out, the effective dimension of the embedding space can be significantly lower than N if the data are filtered prior to detection. Even without filtering, noise is typically correlated and helps reduce the effective dimension of the embedding space. To correct for the influence of the correlated noise, according to Wiechecki-Vergara et al. (2001), the effective dimension of the embedding space $\hat{N}$ is related to the variance of the sample correlation coefficient $\hat{c}_{ij}$ between noise data $\underline{\eta}^j$ and event signal $\underline{s}^i$

$$\hat{c}_{ij} = \frac{\underline{s}^{iT}\underline{\eta}^j}{\sqrt{\left(\underline{s}^{iT}\underline{s}^i\right)\left(\underline{\eta}^{jT}\underline{\eta}^j\right)}}, \tag{27}$$

by:

$$\hat{N} = 1 + {}^{1}\!/_{\sigma_{\hat{c}}^2} \leq N\,. \tag{28}$$

Considering a decrease in the effective embedding space dimension resulted from the correlated noise and/or data pre-processing, equation 25 is rewritten as

$$\begin{cases} 1 - F_{d,\hat{N}-d}\left(\frac{\gamma}{1-\gamma}\frac{\hat{N}-d}{d}\right) = P_F \\ 1 - F_{d,\hat{N}-d}\left(\frac{\gamma}{1-\gamma}\frac{\hat{N}-d}{d}, \bar{f}_c \cdot N \cdot SNR, \left(1-\bar{f}_c\right) \cdot N \cdot SNR\right) = P_D \end{cases}. \tag{29}$$

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

Equation 29 gives the average probability of detection for the events in the design set assuming the design events are all equally likely. If the signals in the design set span the range of signals produced by the source of interest, the calculated average probability of detection P_D in equation 29 also indicates the detection probability for all possible events from this source region.

According to equation 29, assuming a given false alarm rate P_F, the detection threshold γ can be related to the signal subspace dimension d. Therefore, for the given false alarm rate P_F, the average probability of detection P_D over a SNR range of interest is a sole function of the signal subspace dimension d. A value for the signal subspace dimension d and the detection threshold γ are thus determined.

As illustrated in the previous section, the subspace dimension d can also be derived from the average fractional energy capture plots. We will discuss the d values determined by equation 29 and by the fractional energy capture plots in the field study section.

Note that when using equation 29 to determine d and γ, a known P_F is assumed. In this study, we will compare the subspace detection results with those from array correlation and STA/LTA detection. Therefore, we propose a method to determine P_F from the correlation detector and compare the three types of detectors under the same P_F. The array correlation detector on the channel-multiplexed data can be written as:

$$\hat{c} = \frac{\underline{s_m}^T \underline{x}}{\sqrt{\left(\underline{s_m}^T \underline{s_m}\right)\left(\underline{x}^T \underline{x}\right)}}, \tag{30}$$

where $\underline{s_m}$ is the correlation template, i.e. master event data, and $\underline{x}$ is the windowed data to be detected. Both of them have been band-pass filtered and, therefore, have a zero mean. A template event with a good SNR in both P- and S-waves is selected as the correlation

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

template. The correlation detection threshold ρ_c can be estimated from the histogram plot of template event-noise correlation and correlation between template events, which will be illustrated in the field study section. A comparison between equation 10 and equation 30 shows that the correlation coefficient $\hat{c}$ is equivalent to the square root of the subspace detection statistics c(n) with a signal subspace dimension of d = 1. Therefore, the correlation detector (equation 30) can be implemented as a subspace detector with d = 1,

$$c(n) = \frac{(\underline{s}_m{}^T \underline{x})^2}{(\underline{s}_m{}^T \underline{s}_m)(\underline{x}^T \underline{x})}. \quad (31)$$

In this paper, we define c(n) in equation 31 as the correlation detection statistics. The detection threshold associated with equation 31 is

$$\gamma_c = \rho_c^2. \quad (32)$$

According to equation 29, both detectors have a false alarm rate of

$$P_F = 1 - F_{1,\hat{N}-1}\left(\frac{\gamma_c}{1-\gamma_c}\frac{\hat{N}-1}{1}\right). \quad (33)$$

The array correlation algorithm claims an event if

$$c(n) > \gamma_c. \quad (34)$$

The array STA/LTA detector is employed on the channel multiplexed data as,

$$r(n) = \frac{[\underline{x}_{STA}^T(n)\ \underline{x}_{STA}(n)]/N_{STA}}{[\underline{x}_{LTA}^T(n)\ \underline{x}_{LTA}(n)]/N_{LTA}}, \quad (35)$$

and

$$\begin{cases} \underline{x}_{STA}(n) = [x_1(n)\ x_2(n)\ \dots\ x_{N_c}(n)\ x_1(n+1)\ x_2(n+1) \dots\ x_{N_c}(n+N_{STA}-1)]^T \\ \underline{x}_{LTA}(n) = [x_1(n-N_{LTA}) \dots\ x_{N_c}(n-N_{LTA}) \dots\ x_1(n-1)\ \dots\ x_{N_c}(n-1)]^T \end{cases}. \quad (36)$$

N_{STA} and N_{LTA} are the STA, LTA window lengths in samples, respectively. In the field study, typical values of 3 and 15 times the dominant period are selected as the STA and LTA window lengths (Song et al., 2010). Following the analysis in Appendix B, under

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

the null hypothesis H_0, the STA/LTA detection statistics $r(n)$ has a central F distribution with d and $(\widehat{N} - d)$ degrees of freedom. Therefore, at the same false alarm rate P_F as specified in equation 33, the array STA/LTA detection threshold γ_r can be calculated by

$$P_F = 1 - F_{\widehat{N}_{STA},\widehat{N}_{LTA}}(\gamma_r). \quad (37)$$

$\widehat{N}_{STA}$ and $\widehat{N}_{LTA}$ are the effective embedding space dimension of the STA and LTA window, respectively, which can be calculated using the method proposed in Wiechecki-Vergara et al. (2001). An event is declared from the array STA/LTA algorithm if

$$r(n) > \gamma_r . \quad (38)$$

The processing steps of subspace detection and signal enhancement can be summarized as follows:

(1) analyze the spectrum of the recorded data and determine the filter parameters;

(2) apply the band-pass filter; multiplex the filtered continuous data using only channels with good SNRs to form the channel-multiplexed data stream shown in equation 1;

(3) perform the initial detection on the channel-multiplexed data stream with STA/LTA algorithm according to equation 35;

(4) form the template event library and noise data library out of the initial detection; estimate the noise mean and variance on each geophone channel; normalize the filtered data in step 2 by noise variances and form the new channel-multiplexed data stream;

(5) calculate the waveform correlation pairwise for all M events in the template event library; carry out the single-link clustering algorithm on the template event correlation data and select D out of M template events to construct the design set; align the waveforms of D design set events and extract the temporal window

including both P- and S-waves that will be used to form the design data matrix S on equation 15 and define the signal subspace template;

(6) estimate the effective embedding space dimension $\hat{N}$ using equations 27 and 28; carry out the array correlation detection (equations 31, 34) on the new channel-multiplexed data stream and determine the correlation detection threshold γ_c and the corresponding false alarm rate P_F from equations 32 and 33;

(7) for the calculated false alarm rate P_F in step 6, determine the signal subspace dimension d by equation 29 and the fractional energy capture plots; calculate the subspace detection threshold γ by equation 29; construct the signal subspace $\underline{U}$ from the design data matrix S according to equations 18 and 19; calculate the array STA/LTA detection threshold γ_r by equation 37;

(8) conduct the subspace detection (equations 10, 13) and array STA/LTA (equations 35, 38) on the new channel-multiplexed data stream; compare the subspace detection, array STA/LTA and array correlation detection (step 6) results;

(9) for detected events, employ the subspace projection approach (equation 11) to enhance the weak microseismic signals.

FIELD STUDY

Field setup

A microseismic survey was conducted during the fracture stimulation of the Mesa Verde and Cameo Formations in the Mamm Creek Field of the Piceance Basin, Colorado at a depth approximately from 1310 m (4300 ft) to 1981 m (6500 ft) (Weijers et al., 2009). A total of 40 stages of hydraulic fracturing treatments in five wells were mapped. Figure

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

1 shows the located microseismic events from one stage stimulation in fracturing well 24D. The microseismic data were collected using two twelve-level, three-component (3C) geophone arrays deployed in two offset monitoring wells at a depth from 1905 m (6250 ft) to 2050 m (6726 ft). The monitoring well 13B is approximately 457 m (1500 ft) away from the fracturing well 24D, while the monitoring well 24C is closer to the fracturing well 24D, at a distance of about 117 m (386 ft). Therefore, on Figure 1, fewer events are observed from the far well 13B than the nearby well 24C. Two event clusters appear on Figure 1, with a cluster comprising the majority of the events located around well 24C and another minor cluster close to well 13B. Clearly, the microseismicity map is discontinuous and has a gap between the two clusters. This illustrates the footprint of a limited viewing distance from far well 13B, which hinders the interpretation of microseismic maps.

In this study, we apply the subspace approach to improve the detection capability for far well 13B. We also perform STA/LTA and array correlation algorithms on the data from well 13B and compare the detection results with those from the subspace detector. In the following section, we will follow the processing flow proposed in the methodology section and begin with data pre-processing and signal subspace construction. Next, we will determine the detection parameters and conduct the subspace, array STA/LTA and array correlation detections. After that, we will discuss the results from all 3 detectors. Finally, we will present the signal enhancement results using the subspace projection method.

Data pre-processing and signal subspace design

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

Figure 2 shows the 3C raw data for a typical microseismic event recorded by the 12 geophones in well 13B. It is clear that data from geophones 7-12 have higher SNRs. On Figure 2c, even high-amplitude S-waves are hard to see on geophones 1-6. Therefore, in this paper, only data from geophones 7-12 in well 13B are used in the following analysis.

Noise suppression is an essential step prior to detection. Figure 3 gives the spectrum analysis of the raw data for a typical event and noise file. A comparison of the average amplitude spectrum between noise and microseismic signals indicates a dominant signal frequency of [100, 400] Hz, which is depicted as the black square in Figure 3c. Thus, prior to detection, a band-pass filter of [100, 400] Hz was applied to the raw data to get an enhanced signal as shown in Figure 4. The filtered data in Figure 4b demonstrates an enhanced SNR over the raw data in Figure 4a. The band-pass filtered 3C continuous data from geophones 7-12 were next multiplexed to form a channel-multiplexed data stream according to equation 1.

In order to build the template event library for subspace construction, an initial detection was conducted on the channel-multiplexed data using the STA/LTA algorithm (equation 35). The detection results on a 30-minute continuous record are plotted in Figure 5b. At a conservative threshold of 30, 20 events with good waveforms are identified and shown as red stars in Figure 5b. Therefore, an initial template event library comprising M = 20 events was built. It is worth noting that several false alarms (fake events) appear on Figure 5b. This is due to the lack of sensitivity to waveforms, which is also the motivation to develop correlation and subspace algorithms.

As discussed in the methodology section, when developing the subspace detection theory, we assumed the noises on different channels are Gaussian distributed with a zero

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

mean and an identical variance. However, in practice, this assumption may not be true. For geophones at different depths and distances from the microseismic event, the noise variance may vary. Figure 6a-c shows the noise standard deviation of the identified 454 noise files from the initial detection on different geophones and components. It is clear that the noise standard deviation varies over different geophones and components. It is interesting to point out that the vertical component data, shown on Figure 6c, generally have larger standard deviations compared to horizontal components. This may be due to the poorer coupling of the vertical component geophone (Song and Toksöz, 2011). Figure 6d-f gives the noise mean across different geophones and components. It is presented as a multiple of the waveform absolute maximum value on the corresponding channel. It is observed on Figure 6d-f that the band-pass filtered noise data have negligible mean values. To comply with the noise model assumed by the subspace detector, the continuous data from different channels were normalized by their noise standard deviation values and multiplexed to form a new channel-multiplexed data stream. The new channel-multiplexed data have zero mean and identical variance. In the following section, we will apply array STA/LTA, correlation and subspace detection algorithms on the new channel-multiplexed data and compare their detection results.

Figure 7 plots the identified 20 template events from Figure 5 after the noise standard deviation normalization. On the common geophone plot of Figure 7a, significant waveform variations across events are observed. The motivation of subspace detection is to preserve these variations in the subspace representation and detect more events under the same false alarm rate. The template event data were windowed to include both P- and S-waves and the pair-wise correlation between template events was calculated. On the

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

single event plot of Figure 7b, coherent P- and S-wave arrivals across the geophone array are seen.

The pair-wise dissimilarity distances between template events were calculated from event correlation values according to equation 24. The template events were then clustered based on the pair-wise event dissimilarity distance K via the single-link algorithm. In the single-link algorithm, events are aggregated and aligned in a sequential manner (Israelsson, 1990). The dendrogram in Figure 8 shows the hierarchical clustering process. Events 11 and 19 have the smallest dissimilarity distance K (i.e. the largest correlation) and are clustered first. Next, the dissimilarity distances between the remaining 18 events and the newly formed event cluster (11, 19) are updated using the single-link algorithm (see Appendix A for details) and the clustering continues with the second smallest dissimilarity distance and forms a bigger cluster (11, 19, 5). The clustering goes on until all 20 events in the template event library have been clustered. The result is shown in Figure 8. At each clustering step, the cophenetic correlation coefficient (see Appendix A for details) is calculated to measure how well the clustering preserves the actual event dissimilarity behavior. A sudden decrease in the cophenetic correlation indicates the termination of the clustering and the formation of the design set. In this study, we also consider the event dissimilarity distance threshold when forming the design set. The dissimilarity distance threshold cannot be too small or too large, since a small threshold cannot preserve the waveform variations and a large threshold will lose the sensitivity to waveforms. Another consideration for the choice of design set events is to include as many large amplitude events as possible. Therefore, we select a

dissimilarity distance threshold of 0.6. It generates a design set of D = 12 events as shown in Figure 8.

As the clusters are aggregated, the waveforms from the design set events are also aligned to form the design data matrix $\underline{S}$ (see Appendix A for details on the waveform alignment). Proper alignment is crucial, as poor alignment will result in a subspace operator with a larger than necessary number of dimensions (Harris, 2006). Figure 9 shows the alignment result. Figure 9a gives the design set event waveforms before alignment, while Figure 9b presents the aligned waveforms. Several sub-clusters show up on Figure 9b, as also seen on Figure 8. Overall, the dominant phases such as the P- and S-waves are clearly aligned.

In order to construct the signal subspace operator $\underline{U}$ from the design data matrix $\underline{S}$ by equations 17 and 19, the signal subspace dimension d has to be determined. One way to determine d is to look at the fractional energy capture f_c^i for the D design set events as a function of d. Figure 10a shows the fractional energy captured for each design set event in blue. The fractional energy capture curves all begin at 0 and end at 1 and increase with increased dimension of representation (i.e. signal subspace dimension). The average fraction of energy captured for all design events $\bar{f}_c$ is plotted in red. When d reaches 4, the average fraction of energy captured for all design events exceeds 80%. This acts as an aid to determining d. Another way is to look at the increase in the average fractional energy capture $\Delta\,\overline{f_c}$ as a function of the increased subspace dimension d,

$$\Delta\,\overline{f_c} = \overline{f_c}(d+1) - \overline{f_c}(d)\ . \tag{39}$$

It is observed from Figure 10b that $\Delta\,\overline{f_c}$ decreases rapidly as d is increased, which indicates a marginal benefit in the signal energy capture at large d values. However, with

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

increased d, the noise energy capture also increases, which will lead to an increased number of false alarms. Therefore, an intermediate value of d = 4 is chosen as the signal subspace dimension. We will revisit this in the next section when we calculate the detection probability curve.

So far, the design data matrix $\underline{S}$ and the subspace dimension d have been derived. Finally, the signal subspace is built based upon equations 17 and 19.

Detection parameter estimation and performance evaluation

As discussed in the methodology section, in order to derive the detection threshold γ, it is necessary to determine the false alarm rate P_F first.

Figure 11a gives the histogram of the correlation between template event and noise, while Figure 11b presents the histogram of correlation values between template events. The effective dimension of the embedding space $\widehat{N}$ is related to the variance of the sample correlation coefficient $\hat{c}_{ij}$ under null hypothesis by equation 28. From Figure 11a, it is estimated that

$$\widehat{N} = 1 + {}^{1}/_{\sigma_{\hat{c}}^2} = 402.$$

According to Wiechecki-Vergara et al. (2001), the green line on Figure 11a shows the theoretical null probability density function of $\hat{c}_{ij}$,

$$f_R(r) = \frac{1}{B\left(\frac{1}{2}, \frac{\widehat{N}}{2}-1\right)}(1-r^2)^{\frac{\widehat{N}}{2}-2}, \qquad (40)$$

which fits well with the observed histogram.

To produce only 1 false alarm out of all 6240 correlation samples shown as the red line on Figure 11a, a correlation detection threshold of ρ_c= 0.385 is chosen. At this threshold plotted as the red line on Figure 11b, around 16% template events are detected

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

from the correlation detector. According to equation 33, the false alarm rate for the chosen correlation detection threshold is calculated as

$$P_F = 1 - F_{1,\hat{N}-1}\left(\frac{\rho_c^2}{1-\rho_c^2}\frac{\hat{N}-1}{1}\right) = 10^{-15}.$$

At this false alarm rate P_F, the threshold for correlation detection statistics in equation 31 and array STA/LTA detection statistics in equation 35 are given by,

$$\gamma_c = \rho_c^2 = 0.149,$$

$$\gamma_r = 3.989.$$

The threshold γ for the subspace detection statistics in equation 10 is determined by maximizing the detection probability P_D for the given false alarm rate $P_F = 10^{-15}$. Figure 12 illustrates the probability of detection P_D as a function of SNR. Twelve individual P_D curves appear in the figure, one for each possible signal subspace dimension d. It is clear that the P_D curve for d = 1, shown in yellow, does not reach 1 even at high SNRs such as 10 dB. This feature occurs because of the low signal energy capture for most events when the signal subspace consists of a single vector (see Figure 10a). This also justifies the disadvantage that comes with the correlation detector and the detection method proposed by Bose et al. (2009), both of which use only one basis vector to represent the signal subspace. On Figure 12, the P_D curve climbs quickly as d increases from 1 to 2, because the added dimension increases the signal energy capture significantly, as seen in Figure 10. The detection probability continues to improve until d increases to 4. By the time the subspace dimension grows to 4, the average fractional energy capture is above 0.8 for all design set events and good detection performance (P_D~1) is achieved for a SNR as low as -13.5 dB. Performance continues to improve, but marginally, until d reaches 8, beyond which the detection probability actually begins to decline. This is because the marginal

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

increase in signal energy capture afforded by additional increments to the signal subspace representation does not offset the increase in noise energy capture. The analysis suggests that a dimension of 4 would be a good choice for this example in terms of maximizing the probability of detection for a given false alarm rate. This is consistent with the fractional energy capture analysis in Figure 10. Given that d = 4, and $\widehat{N} = 402$, the subspace detection threshold γ is calculated from P_F using equation 29,

$$P_F = 1 - F_{d,\widehat{N}-d}\left(\frac{\gamma}{1-\gamma}\frac{\widehat{N}-d}{d}\right) = 10^{-15} \Rightarrow \gamma = 0.174.$$

Comparison of array STA/LTA, correlation and subspace detectors

The new channel-multiplexed data are formed by multiplexing the x, y, z component data from geophones 7-12 in well 13B, after noise standard deviation normalization. The array STA/LTA, correlation and subspace detectors are then applied to the new channel-multiplexed data. Figure 13 compares the detection results of the three detectors on the 30-minute continuous record. The false alarm rate has been set at $P_F = 10^{-15}$ for all three detectors. The detection thresholds are calculated from the previous section and indicated in Figure 13 with the black lines. The correlation detector has a lower threshold (0.149) than the subspace detector (0.174), since it has a single dimension and, thus, should match background noise less well. It is seen from the figure that the background level of the correlation statistics is lower than that of the subspace statistics commensurate with the calculated threshold.

At $P_F = 10^{-15}$, there are 604, 1571, and 2730 triggers on the 30-minute continuous record by array correlation, subspace, and STA/LTA detectors, respectively. It is difficult to analyze such a large number of triggers. We proceed with two alternative approaches.

First, we analyze a 1-minute portion (in this case, from 1100 seconds to 1160 seconds on Figure 13) of the data to compare the performance of three detectors under the same expected probability of false alarms. Table 2 presents the results. It is clear that at a constant false alarm rate $P_F = 10^{-15}$, the correlation detector gives the least amount of false alarms while STA/LTA detector generates the most false alarms. This is due to the fact that an incoherent detector, such as STA/LTA detector, has no sensitivity to waveforms and thus is prone to false alarms. The correlation detector, on the contrary, matches signals both temporally and spatially to enhance sensitivity and, therefore, has less chance to generate false alarms. It is interesting to point out that several design set events, shown as yellow crosses on Figure 13c, are missed by the correlation detector even at a low threshold of 0.149. However, they are detected as red crosses by both STA/LTA and subspace algorithms in Figure 13b and d. Figure 14 plots the four design set events missed by the correlation detector at a threshold of 0.149 (see Figure 13c). Substantial variations between the waveform of the four design set events on Figure 14a-d and the template event waveform on Figure 14e are observed. This demonstrates the inability of correlation detectors to capture waveform variations, which result from variations in the source mechanisms, locations, and source time functions.

Secondly, due to the double-precision limit of the machine used, it is hard to set a false alarm rate less than 10^{-15}. Instead, we look at a reduced number of triggers. Figure 15 presents the first 35 largest triggers from three detectors on the 30-min continuous record. On Figure 15b, the crosses represent the 21 events detected by the array STA/LTA algorithm with the minimum detected event plotted in red, while the false alarm with the largest STA/LTA detection statistics is shown as the green square. One

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

event, detected by the STA/LTA algorithm but missed by the subspace detector, is seen as the magenta cross on Figure 15b. Figure 16a plots the 3C waveforms of the minimum detected event, where coherent P- and S-waves are seen across the array. Figure 16b shows the STA/LTA event missed by the subspace detector. Figure 16c gives the false alarm with the largest STA/LTA statistics, where no coherent arrivals are observed.

Similarly, out of the largest 35 triggers by the correlation detector, 10 events are detected and plotted as crosses in Figure 15c. The minimum detected event and the correlation template event are shown as the red and magenta cross, respectively. Figure 17a and b shows the 3C waveforms of the minimum detected event and the template event. Coherent P- and S-waves appear on both figures. A good degree of waveform similarity is seen between the detected and template event especially for the dominant S-waves. Figure 17c shows the waveforms of the false alarm with the largest correlation statistics, plotted as the green square on Figure 15c. No coherent P-waves are seen. It is worth noting that dominant coherent energy at the moveout of S-waves is observed on Figure 17c, which justifies the appearance of the high correlation. Several factors could contribute to the missing P waves. Firstly, P-waves usually have small amplitudes and are embedded in high noise. Alternatively, the geophone array may be close to the nodal points of the P-wave radiation pattern. Either way, P-waves cannot be identified. However, the polarization information of P-waves is essential for locating microseismic events from downhole geophones (Warpinski et al., 2005; Rentsch et al., 2007). Therefore, for the purpose of location, we consider event triggers without P-waves as false alarms.

Among the 35 largest triggers by the subspace detector, 21 events are detected and plotted as crosses on Figure 15d. All 12 events, comprising the design set, are detected and shown as black crosses. The 9 additional events, denoted as magenta crosses, are also detected at the threshold of 0.619. Out of the 9 events, 2 events, labeled as '1' and '2' on Figure 15d, are missed by both the STA/LTA and correlation detectors. Figures 18a and 19a show the 3C waveform for events '1' and '2', respectively. Comparing Figures 18a and 19a with Figures 16a and 16b, it is clear that the STA/LTA algorithm allows more waveform variations than the subspace detector, but at the expense of increased false alarms. Moreover, the 10 events detected by the correlator are a subset of the subspace detections. Looking at the correlation and subspace statistics for the 10 correlation events, it is found that the subspace detector has an increased processing gain (larger statistics when events are present).

To study the 2 events from well 13B that are missed by both the STA/LTA and correlation detectors, we look at detections from nearby well 24C. By considering the possible arrival time difference on well 24C and 13B (for a well separation of 1100 ft, the maximum two-way travel time difference is around 0.3 seconds), and possible difference between STA/LTA picks and subspace picks (depending on the size of the detection window, in this study, around 0.5 seconds), we search the data from well 24C over an interval of +/- 1 second around the subspace picks. We look for the maximum STA/LTA peak within the searched time interval. Figures 18b and 19b give the 3C waveform plots associated with the maximum STA/LTA peak on nearby well 24C. Coherent S-waves are seen on Figure 18b, while P-waves are missing. As discussed before, for location purposes, we consider this event not detectable on well 24C. On the contrary, on Figure

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

19b, both coherent S- and P-waves are observed. We treat this event as a valid detection from well 24C. Although only the largest 35 triggers from this 30-min record are analyzed, we are able to detect one event (Figure 18a), which is not seen on nearby well 24C. This demonstrates the capability of the subspace algorithm in improving the detection capability of far well 13B.

As previously mentioned, limited by the double-precision machine, we could not perform a constant false alarm detection at $P_F < 10^{-15}$. However, the false alarm rate corresponding to a given detection threshold can still be calculated to an arbitrary precision by analytical approaches (Kendall and Stuart, 1979). The last column of Table 2 summarizes the detection results on the 30-min record for the largest 35 triggers. At a much lower expected false alarm rate, the STA/LTA detector ($P_F = 8 * 10^{-102}$) gives the same amount of false alarms as the subspace detector ($P_F = 4 * 10^{-82}$). This is consistent with the constant false alarm rate case shown in the middle column of Table 2, where, due to its limited sensitivity to waveforms, the STA/LTA detector generates many more false alarms than the subspace detector at the same false alarm rate. On the other hand, even at a larger allowed false alarm rate, the correlation detector ($P_F = 7 * 10^{-56}$) cannot detect as many events as the subspace detector ($P_F = 4 * 10^{-82}$). This indicates the inability of the correlation detector to capture the waveform variations.

Signal enhancement based on subspace projection

Once weak events are detected, the corresponding 3C data are projected into the signal subspace $\underline{U}$ according to equation 11 to obtain the enhanced signal $\underline{x}_p(n)$. As an example, Figure 20 shows the enhanced signal for two weak subspace events '1' and '2'. Compared to the original data on Figures 20a and 20c, the enhanced signals on Figures

20b and 20d have larger SNRs especially for the weak P-waves. For both events, the median value in the SNR gains from subspace projection, across six geophones and three components, is around 16 dB and 19 dB for P- and S-waves, respectively. The major contribution in the SNR increase comes from the pre-event noise suppression. Since noise mostly exists in the orthogonal complement of the signal subspace, the noise energy projected into the signal subspace is minimal.

CONCLUSIONS

In this paper, we introduced a full-waveform based event detection and signal enhancement approach for microseismic monitoring. The method constructs a vector space, known as the signal subspace, to represent variable microseismic signals from specific source regions. It models the signals to be detected as a linear combination of the orthogonal bases of the subspace. Unlike correlation detectors, the subspace approach is more broadly applicable. Furthermore, the subspace detector is sensitive to waveforms and, therefore, offers a lower probability of false alarms, compared to STA/LTA detectors.

A systematic procedure based on the statistical hypothesis testing theory was presented to build the signal subspace from previously detected events and determine the detection parameters. The subspace detector provides a way to manage the tradeoff between sensitivity and flexibility by adjusting the detection parameters such as the detection threshold and the subspace dimension, i.e. the number of bases used to present the signal subspace. The subspace design and detection approach was demonstrated on a dual-array hydrofracture monitoring dataset. The application of the subspace, STA/LTA,

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

and correlation detectors to the data from the far monitoring well was presented. It is found that, at the same false alarm rate, the subspace detector gives fewer false alarms than the array STA/LTA detector and more event detections than the array correlation detector. The additionally detected events from the far monitoring well by the subspace detector were compared with the detections from the nearby well. It was demonstrated that, with the subspace detector, we are able to detect additional events that are not seen on the nearby well. The limitation of the subspace detector is the complexity and relatively large computation cost in building the signal subspace. Fortunately, the signal subspace construction could be done off-line, which makes real-time subspace detection possible. Moreover, the template event library used to form the signal subspace could be dynamically updated as detection goes on. When the subspace detector is used as a post-processing tool, it would be more efficient to build the signal subspace from the spatial clusters of events.

The data of detected events were projected into the signal subspace to form the enhanced microseismic signals. It was shown that the SNR of detected weak microseismic events is improved after applying the subspace-projection-based signal enhancement procedure. By using both full waveforms from multiple events and the signal/noise statistics, the proposed subspace detection and signal enhancement approach is capable of handling strong noise and offers the potential for future application of hydrofracture monitoring with the treatment well, as the noise in the treatment well is much higher than the offset monitoring well.

ACKNOWLEDGMENTS

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

The authors would like to thank Halliburton for providing the data and for funding this research. We are grateful to Mr. Charlie Waltman from Halliburton and Dr. David Harris from Lawrence Livermore National Laboratory for their helpful discussions. We thank Halliburton and Bill Barrett Corporation for permission to publish this work.

APPENDIX A: DESIGN SET EVENT SELECTION AND WAVEFORM ALIGNMENT THROUGH THE SINGLE-LINK ALGORITHM

The single-link algorithm has been proposed for seismic event clustering and been used in the subspace algorithm (Israelsson, 1990, Harris, 2006). In this appendix, we review the steps of design set event selection and waveform alignment via the single-link algorithm.

The single-link clustering method begins by treating all events as individual clusters containing one event each. In each step of the clustering method, the minimum distance pair (i.e., largest correlation measurement) is selected and the two clusters (events), to which it corresponds, are merged. As two clusters are combined, the dissimilarity distances between the two clusters and any third remaining cluster are combined by selecting the smaller of the dissimilarity distance measurements to represent the inter-event distance of the new cluster with the third cluster. An updated dissimilarity matrix $\underline{K^g}$ is formed to reflect the inter-event distance changes caused by the clustering. This process of aggregation continues until a single cluster remains. The clustering results are summarized by a dendrogram, which shows the successive fusions of events. At each

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

clustering step, a cophenetic correlation coefficient (C_g) is calculated to measure how well the clustering models the actual similarity behavior, which is described in matrix $\underline{K}$. Assuming that there are M events in the template event library, the original dissimilarity matrix $\underline{K}$ has a size of M*M. The cophenetic correlation is computed as the correlation coefficient between $\underline{K}$ and $\underline{K^g}$, for successive steps g = 1, 2, … , M-1,

$$C_g = \frac{\sum_{q=1}^{M}\sum_{p=1}^{M} K_{p,q}K^g_{p,q}}{\left[\sum_{q=1}^{M}\sum_{p=1}^{M} K_{p,q}K_{p,q}\sum_{q=1}^{M}\sum_{p=1}^{M} K^g_{p,q}K^g_{p,q}\right]^{1/2}}. \quad \text{(A1)}$$

As clustering progresses, the correlation between the $\underline{K^g}$ matrix and the original $\underline{K}$ matrix will continue to decrease as the original entries are replaced with the dissimilarity distances calculated for the growing clusters. Overall, values of C_g will thus decline.

The design set is a set of events in the template event library that are to be used to construct the signal subspace bases. Therefore, it is desirable for the design set to comprise not only most of the larger template events, but also to represent the actual inter-event correlation behavior described by the original dissimilarity matrix $\underline{K}$. Therefore, a sudden decrease in C_g is used as an indicator to terminate clustering. Besides the cophenetic correlation criteria, in this paper the event dissimilarity distance threshold is also considered to ensure reasonable waveform variability when forming the design set.

The waveform alignment is done simultaneously with the design set event selection. The delays used for waveform alignment are calculated relative to the reference event, i.e., event 11 as shown in Figure 8. For each event that belongs to the left nodes of the dendrogram and is directly connected to the reference event (event 19, 5, 18, 2, 20, and 6 in Figure 8), the delay is the point in the cross-correlation function where the correlation between that event and event 11 is maximized. The rest of the design set events are

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

connected to the reference event through intermediate left node events. The delay of each of these events is calculated as the sum of all the delays on the connection path to the reference event. For example, the delay for event 16 is the sum of the delays from event pairs (16, 6) and (6, 11). Likewise, the delay for event 13 is the sum of the delays from event pairs (13, 12), (12, 2) and (2, 11). The waveform alignment results for all D=12 design set events after applying the delays are displayed in Figure 9.

APPENDIX B: DERIVATION OF EQUATION 25 VIA THE ANALYSIS OF THE DETECTION STATISTICS

In this appendix, we derive the subspace detection probability and false alarm rate from the analysis of the subspace detection statistics. According to Harris (2006), the subspace detection statistics $c(n)$ defined in equation 10 can be transformed into a F-distributed variable,

$$c'(n) = \frac{[\underline{x}_p^T(n)\underline{x}_p(n)]/\sigma^2/d}{[\underline{w}^T(n)\underline{w}(n)]/\sigma^2/(N-d)}, \tag{B1}$$

where $\underline{w}(n)$ is the projection of the detection data vector $\underline{x}(n)$ into the orthogonal complement to the subspace $\underline{U}$,

$$\underline{w}(n) = [I_N - \underline{U}\,\underline{U}^T]\,\underline{x}(n). \tag{B2}$$

Under null hypothesis H_0, $\underline{x}_p(n)$ and $\underline{w}(n)$ are two independent zero-mean Gaussian distributed variables with an identical variance of σ^2. Therefore, $[\underline{x}_p^T(n)\underline{x}_p(n)]/\sigma^2$ and $[\underline{w}^T(n)\underline{w}(n)]/\sigma^2$ are independent and chi-square distributed, with d and (N-d) degrees of freedom, respectively. Hence, $c'(n)$ in equation B1 follows the central F distribution under null hypothesis. From equation 13, the false alarm occurs when

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

$$c'(n) > \frac{\gamma}{1-\gamma}\frac{N-d}{d}, \tag{B3}$$

where γ is the threshold for c(n). Thus, the false alarm rate is calculated as

$$P_F = 1 - F_{d,N-d}\left(\frac{\gamma}{1-\gamma}\frac{N-d}{d}\right), \tag{B4}$$

where $F_{d,N-d}(\cdot)$ denotes the cumulative central F distribution with d and (N-d) degrees of freedom.

Similarly, under alternative hypothesis H_1, $\underline{x}_p(n)$ and $\underline{w}(n)$ are two independent Gaussian distributed variables with an identical variance of σ^2, but with non-zero mean values. Therefore, $\left[\underline{x}_p^T(n)\underline{x}_p(n)\right]/\sigma^2$ and $\left[\underline{w}^T(n)\underline{w}(n)\right]/\sigma^2$ are independent and noncentral chi-square distributed, with d and (N-d) degrees of freedom, respectively. Considering the fractional energy captured in the signal subspace $\underline{U}$, the noncentrality parameters of $\left[\underline{x}_p^T(n)\underline{x}_p(n)\right]/\sigma^2$ and $\left[\underline{w}^T(n)\underline{w}(n)\right]/\sigma^2$ are $\underline{a}^T\underline{a}/\sigma^2$ and $\left(1-\underline{a}^T\underline{a}\right)/\sigma^2$, respectively. Thus, $c'(n)$ in equation B1 follows the doubly noncentral F distribution under alternative hypothesis. An event is then detected according to equation B3. The detection probability is then derived as

$$P_D = 1 - F_{d,N-d}\left(\frac{\gamma}{1-\gamma}\frac{N-d}{d}, \underline{a}^T\underline{a}/\sigma^2, \left(1-\underline{a}^T\underline{a}\right)/\sigma^2\right). \tag{B5}$$

As discussed in the main text, if we assume 1) the signals in the design set span the range of signals produced by the source of interest, and 2) the design events are all equally likely, the noncentrality parameters $\underline{a}^T\underline{a}/\sigma^2$ and $\left(1-\underline{a}^T\underline{a}\right)/\sigma^2$ for any event can be replaced by the ratio of the average energy captured in the subspace and its orthogonal complement to the noise variance. This gives

$$P_D = 1 - F_{d,N-d}\left(\frac{\gamma}{1-\gamma}\frac{N-d}{d}, \bar{f}_c \cdot \frac{E}{\sigma^2}, \left(1-\bar{f}_c\right)\cdot\frac{E}{\sigma^2}\right). \tag{B6}$$

Substituting SNR from equation 26 into equation (B6) yields equation 25.

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

REFERENCES

Bose S., Valero H.P., Liu Q., Shenoy R.G. and Ounadjela A. 2009. An Automatic Procedure to Detect Microseismic Events Embedded in High Noise. 79th Annual International Meeting, SEG Expanded Abstracts, **28**, 1537-1541.

Earle P.S. and Shearer P.M. 1994. Characterization of global seismograms using an automatic-picking algorithm. *Bulletin of the Seismological Society of America*, **84**, 366-376.

Eckart C. and Young G. 1936. The approximation of one matrix by another of lower rank. *Psychometrika*, **1**, 211-218.

Gibbons S. J. and Ringdal F. 2006. The detection of low magnitude seismic events using array-based waveform correlation. *Geophysical Journal International*, **165**, 149-166.

Harris D.B. 1991. A waveform correlation method for identifying quarry explosions. *Bulletin of the Seismological Society of America*, **81**, 2395–2418.

Harris D.B. 2006. Subspace Detectors: Theory. *Lawrence Livermore National Laboratory Internal Report UCRL-TR-222758*, 46 pp.

Israelsson H. 1990. Correlation of waveforms from closely spaced regional events. *Bulletin of the Seismological Society of America*, **80**, 2177-2193.

Kendall M. and Stuart A. 1979. *The Advanced Theory of Statistics*, vol. 2, Charles Griffin, London.

Maxwell S.C., Underhill B., Bennett L., Woerpel C. and Martinez A. 2010. Key criteria for a successful microseismic project. *SPE Paper* 134695.

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

Maxwell S.C., Waltman C.K., Warpinski N.R., Mayerhofer M.J. and Boroumand N. 2006. Imaging Seismic Deformation Induced by Hydraulic Fracture Complexity. *SPE Paper* 102801.

Maxwell S.C., White D.J. and Fabriol H. 2004. Passive seismic imaging of CO2 sequestration at Weyburn. 74th Annual International Meeting, SEG, Expanded Abstracts, **23**, 568-571.

Phillips W.S., Rutledge J. T. and House L. 2002. Induced microearthquake patterns in hydrocarbon and geothermal reservoirs: six case studies. *Pure and Applied Geophysics* **159**, 345-369.

Rentsch S., Buske S., Lüth S. and Shapiro S.A. 2007. Fast location of seismicity: A migration-type approach with application to hydraulic-fracturing data. *Geophysics* **72**, no. 1, S33–S40.

Rowe C.A., Aster R.C., Borchers B. and Young C.J. 2002. An automatic, adaptive algorithm for refining phase picks in large seismic data sets. *Bulletin of the Seismological Society of America* **92**, 1660–1674.

Scharf L.L. and Friedlander B. 1994. Matched subspace detectors. *IEEE Transactions on Signal Processing* **42**, 2146–2157.

Song F., Sadi H.K., Toksöz M.N., Ay E. and Zhang H. 2010. An improved method for hydrofracture-induced microseismic event detection and phase picking. *Geophysics* **75**, A47–A52.

Song F. and Toksöz M.N. 2011. Full-waveform based complete moment tensor inversion and source parameter estimation from downhole microseismic data for hydrofracture monitoring. *Geophysics* **76**, WC101–WC114.

Van Trees H.L. 1968. *Detection, Estimation and Modulation Theory*, vol. 1. John Wiley and Sons, New York.

Warpinski N.R. 2009. Microseismic monitoring: inside and out. *Journal of Petroleum Technology* **61**, 80-85.

Warpinski N.R., Kramm R.C., Heinze J.R. and Waltman C.K. 2005. Comparison of Single- and Dual-Array Microseismic Mapping Techniques in the Barnett Shale. *SPE paper* 95568.

Weichecki-Vergara S., Gray H.L. and Woodward W.A. 2001. Statistical development in support of CTBT monitoring. *Tech. Rep. DTRA-TR-00-22*, Southern Methodist University, Dallas, Texas.

Weijers L., Kama Y., Shemeta J. and Cumella S. 2009. Bigger is better — Hydraulic fracturing in the Williams Fork Formation in the Piceance basin. *AAPG/Datapages, Search and Discovery* 110092.

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

FIGURE CAPTIONS

Figure 1. (a) Horizontal plane view of the microseismic event locations from one stage treatment plotted as black stars. The blue and black squares denote the monitoring wells 13B and 24C, respectively, while the fracturing well is shown as the red triangle. The origin (0, 0) corresponds to the wellhead location of well 13B. (b) The side view of the microseismic events. The blue squares and black squares represent the two twelve-level geophone arrays deployed in well 13B and 24C separately (from deep to shallow depths: geophone 1 to 12). The perforation locations are depicted as the red triangles in fracturing well 24D. Fewer events are detected on the far well 13B. Data from the far well 13B will be used in this study for subspace detection and signal enhancement.

Figure 2. The three-component raw data plot for a typical event recorded in the far well 13B: (a) x component, (b) y component, (c) z component.

Figure 3. (a) The raw x component data of a 0.5s event record from geophones 7-12 in well 13B. (b) The raw x component data of a 0.5s noise segment recorded by geophones 7-12 in well 13B. (c) Amplitude spectrum of the raw event and noise data in the panels (a) and (b), averaged over all 6 geophones. The black square demonstrates the dominant signal frequency range of [100, 400] Hz.

Figure 4. (a) The raw x component data of a 0.5s continuous record from geophones 7-12 in well 13B. (b) The [100, 400] Hz band-pass filtered result of the panel (a).

Figure 5. Array STA/LTA detection on a 30-min continuous record from far well 13B. a) The x component [100, 400] Hz band-pass filtered continuous data from one geophone in

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

well 13B. b) The STA/LTA detection results on the channel-multiplexed data. The x, y, z component data from geophones 7-12 are used in the STA/LTA detection. The template event library for the subspace detector, comprising the M = 20 identified events using a conservative STA/LTA threshold of 30, is plotted in red stars.

Figure 6. The standard deviation and mean of identified 454 noise data files across the six geophones (geophones 7-12 from well 13B). Left columns: noise standard deviation. a) x component. b) y component. c) z component. Right columns: noise mean as a multiple of its corresponding absolute maximum value. d) x component. e) y component. f) z component.

Figure 7. Waveform plot of the detected 20 template events (as described in Figure 5) after noise standard deviation normalization. a): Band-pass filtered unaligned waveforms of all 20 events from one geophone in well 13B. b): Band-pass filtered unaligned waveforms of one template event from all six geophones in well 13B (geophones 7-12).

Figure 8. Template event clustering and design set event selection through the dendrogram using the single-link algorithm. The red line shows the termination of clustering with a maximum event dissimilarity distance of 0.6, which gives a design set comprising D = 12 events (events 11 to 16).

Figure 9. The waveform alignment of design set events using the single-link algorithm. a) The unaligned z component waveform plot from one geophone. b) The waveform plot of panel a) after alignment.

Figure 10. a) Fractional energy captured f_c as a function of dimension of representation d (also known as the signal subspace dimension) for each design set event is plotted in blue, while the average fractional energy capture $\overline{f_c}$ for all D=12 design set events as a function of d is shown in the red curve. A threshold of at least 80% average fractional energy capture plotted as the vertical red line gives an optimal subspace dimension d = 4. The horizontal red line shows the theoretical detection threshold for the subspace detector with d = 4, and false alarm rate of $P_F = 10^{-15}$. b) The increase in the average fractional energy capture $\Delta\overline{f_c}$ as a function of an increased subspace dimension d.

Figure 11. a) The histogram of correlation values between template event and noise. b) The histogram of correlation values between template events.

Figure 12. The probability of detection as a function of the SNR at a fixed false alarm rate $P_F = 10^{-15}$. In this case, the detection probabilities are calculated as a function of SNR for subspace dimensions ranging from 1 to 12. The detection probability curve for the selected subspace detector with d = 4 is plotted in red, while the yellow and black curves demonstrate the detection probability curves for the subspace detector with d = 1 and d = 12, respectively.

Figure 13. The comparison of detection results on a 30-min continuous record in far well 13B at a fixed false alarm rate $P_F = 10^{-15}$. The new channel-multiplexed data, formed by the x, y, z component data from geophones 7-12 after noise standard deviation normalization, are used in the detection. a) The [100, 400] Hz band-pass filtered x

component data from one geophone in well 13B. b) The STA/LTA detection, c) the correlation detection, and d) the subspace detection (d=4) results on the new channel-multiplexed data. The threshold values at $P_F = 10^{-15}$, plotted as the black horizontal lines, are 3.989, 0.149, and 0.174 for the STA/LTA, correlation, and subspace detector, respectively. The four design set events missed by the correlation detector, but captured by STA/LTA and subspace detectors, are plotted as yellow and red crosses.

Figure 14. The band-pass filtered x component waveform plot. The dashed and solid black lines represent the P and S arrival picks on geophones 7-12 (geophone index: 1-6) in well 13B. a-d) The four design set events missed by the correlation detector, but captured by STA/LTA and subspace detectors at $P_F = 10^{-15}$. e) The correlation template event.

Figure 15. The comparison of the largest 35 triggers on a 30-min continuous record in far well 13B. The new channel-multiplexed data, formed by the x, y, z component data from geophones 7-12 after noise standard deviation normalization, are used in the detection. a) The [100, 400] Hz band-pass filtered x component data from one geophone in well 13B. b) The STA/LTA detector gives 21 events plotted as crosses, with the minimum detected event denoted as the red cross. The false alarm with the largest STA/LTA statistics is shown in the green square. One STA/LTA event missed by the subspace detector is plotted as the magenta cross. c) The correlation detector gives 10 events plotted as crosses, with the minimum detected event and correlation template event denoted as the red and magenta crosses, respectively. The false alarm with the largest correlation statistics is shown in the green square. d) The subspace detector with d=4 generates 21

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

events plotted as crosses, with 12 out of them being the design set events shown in black, and 9 additional detected events are plotted in magenta. Two events, detected by the subspace detector but missed by both STA/LTA and correlation detectors, are marked as 1, 2 on panel d).

Figure 16. The waveform plot of the band-pass filtered data (columns from left to right: x, y, z components). The dashed and solid black lines represent the P and S arrival picks on geophones 7-12 (geophone index: 1-6) in well 13B. a) The minimum detected event from the array STA/LTA detector (see the red cross on Figure 15b). b) The STA/LTA event missed by the subspace detector (see the magenta cross on Figure 15b). c) The false alarm with the largest STA/LTA statistics (see the green square on Figure 15b).

Figure 17. The waveform plot of the band-pass filtered data (columns from left to right: x, y, z components). The dashed and solid black lines represent the P and S arrival picks on geophones 7-12 (geophone index: 1-6) in well 13B. a) The minimum detected event from the array correlation detector (see the red cross on Figure 15c). b) The correlation template event of the array correlation detector (see the magenta cross on Figure 15c). c) The false alarm with the largest correlation statistics (see the green square on Figure 15c).

Figure 18. The three-component waveform plot of event 1 on Figure 15d, detected by the subspace detector, but missed by both array STA/LTA detector and array correlation detector (x in blue, y in red, z in black). The dashed and solid black lines represent the P and S arrival picks. a) The band-pass filtered data from geophones 7-12 (geophone index: 1-6) in the far well 13B. b) The corresponding detected waveforms from geophones 7-12

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

(geophone index: 1-6) in the nearby well 24C. The time difference between a) and b) is to account for the possible arrival time difference between the far well 13B and nearby well 24C.

Figure 19. The three-component waveform plot of event 2 on Figure 15d, detected by the subspace detector, but missed by both array STA/LTA detector and array correlation detector (x in blue, y in red, z in black). The dashed and solid black lines represent the P and S arrival picks. a) The band-pass filtered data from geophones 7-12 (geophone index: 1-6) in the far well 13B. b) The corresponding detected waveforms from geophones 7-12 (geophone index: 1-6) in the nearby well 24C. The time difference between a) and b) is to account for the possible arrival time difference between the far well 13B and nearby well 24C.

Figure 20. The subspace projection approach for microseismic signal enhancement. The waveform plot of the band-pass filtered (x, y, z) component data from geophones 7-12 (geophone index: 1-6) in the far well 13B (columns from left to right: x, y, z components). a) Data from the detected event 1 as shown in Figure 18, before signal enhancement. b) Data from the detected event 1 as shown in Figure 18, after signal enhancement. c) Data from the detected event 2 as shown in Figure 19, before signal enhancement. d) Data from the detected event 2 as shown in Figure 19, after signal enhancement.

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

TABLE HEADINGS

Table 1. Symbols.

Table 2. Summary of detections results on a 30-minute continuous record in far well 13B by the STA/LTA, correlation, and subspace detectors.

Table 1. Symbols.

H_0	The null hypothesis: event not present in the detection window
H_1	The alternative hypothesis: event present in the detection window
$x(n)$	The n-th sample in the channel multiplexed continuous data stream
$x_i(n)$	The n-th sample in the continuous data recorded by the i-th channel
$\underline{x}(n)$	The N*1 data vector in the subspace/correlation detection window starting at n-th time sample
$\underline{x}_{STA}(n)$	The channel multiplexed data vector in the STA window starting at n-th time sample
$\underline{x}_{LTA}(n)$	The channel multiplexed data vector in the LTA window ending at n-th time sample
N, N_T	The number of data samples, time samples in each subspace/correlation detection window
N_c	The number of recorded channels
$\widehat{N}$	The effective embedding space dimension of the subspace/correlation detection window
N_{STA}, N_{LTA}	The number of time samples in each STA, LTA window
$\widehat{N}_{STA}, \widehat{N}_{LTA}$	The effective embedding space dimension of the STA, LTA window
$\underline{s}, \underline{\eta}$	The signal and noise vector in the detection window
σ^2	The unknown noise variance
$\underline{U}$	The N*d matrix, comprising d signal subspace bases
$\underline{a}$	The d*1 coefficients, used to project the signal vector $\underline{s}$ into the signal subspace $\underline{U}$
$\underline{a}_d^i$	The d*1 coefficients, used to project the data vector $\underline{s}^i(n)$ from the i-th design set event into the signal subspace $\underline{U}$
E_c	The energy captured in the signal subspace after projection
$p(\cdot \mid H_0)$	The probability density function of the detection data under H_0
$p(\cdot \mid H_1)$	The probability density function of the detection data under H_1
$l\left(\underline{x}(n)\right)$	The generalized log likelihood ratio function of the detection data vector
γ	The subspace detection threshold, associated with the subspace detection statistics c(n) defined in equation 10
$\underline{x}_p(n)$	The projection of the detection data vector $\underline{x}(n)$ into the subspace $\underline{U}$
$\underline{s}^i(n)$	The N*1 normalized data vector from the i-th design set event
$\underline{S}$	The design data matrix, comprising D data vectors of design set events
$\underline{w}(n)$	The projection of the detection data vector $\underline{x}(n)$ into the orthogonal complement to the subspace $\underline{U}$
$\underline{W}, \underline{\Sigma}, \underline{V}$	The SVD of the N*D design data matrix $\underline{S}$
$\underline{A}$	The exact representation coefficient matrix of size N*D
$\underline{A}_d$	The approximate representation coefficient matrix of size d*D
D, d	The number of design set events, the signal subspace dimension
f_c^i	The fractional energy captured in $\underline{U}$ for the i-th design set event
$\overline{f_c}$	The average fractional energy captured in $\underline{U}$ for all D design set events

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

$\Delta\overline{f_c}$	The increase in average fractional energy capture
M	The number of template events
$\underline{K}$	The original template event dissimilarity distance matrix of size M*M
$\underline{K^g}$	The template event dissimilarity distance matrix of size M*M at clustering step g (g=1,2…, M-1)
$K_{p,q}$	The original dissimilarity distance between template event p and q
$K^g_{p,q}$	The dissimilarity distance between template event p and q at clustering step g (g=1,2…, M-1)
$\lambda_{p,q}$	The maximum waveform correlation between template event p and q
P_F	The probability of false alarms, i.e. the false alarm rate
P_D	The probability of detection
$\hat{c}_{ij}$	The sample correlation coefficient between noise data $\underline{\eta}^j$ and event signal $\underline{s}^i$
C_g	The cophenetic correlation coefficient between $\underline{K}$ and $\underline{K^g}$
$\sigma^2_{\hat{c}}$	The variance of the sample correlation between noise and event
$\underline{s_m}$	The N*1 correlation template vector, i.e. master event data vector
γ_c	The correlation detection threshold, associated with correlation detection statistics c(n) defined in equation 31
γ_r	The STA/LTA detection threshold, associated with STA/LTA detection statistics r(n) defined in equation 35
$f_R(\cdot)$	The probability density function of sample correlation coefficient under H_0
$F_{\cdot,\cdot}(\cdot)$	The cumulative distribution function of subspace/correlation detection statistics, could be central F distribution or doubly non-central F distribution

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

Table 2. Summary of detections results on a 30-minute continuous record in far well 13B by the STA/LTA, correlation, and subspace detectors.

Performance type / Type of detectors	Constant false alarm rate[1]	Constant # of triggers[2]
Array STA/LTA detector (# of detected events / # of false alarms) (expected false alarm rate P_F)	10 / 139 $P_F = 10^{-15}$	21 / 14 $P_F = 8 * 10^{-102}$
Array correlation detector (# of detected events / # of false alarms) (expected false alarm rate P_F)	5 / 9 $P_F = 10^{-15}$	10 / 25 $P_F = 7 * 10^{-56}$
Subspace detector (# of detected events / # of false alarms) (expected false alarm rate P_F)	6 / 27 $P_F = 10^{-15}$	21 / 14 $P_F = 4 * 10^{-82}$

Note 1: only the detection results from a 1-minute segment of the total 30-minute record are listed here under a constant false alarm rate $P_F = 10^{-15}$.

Note 2: the largest 35 triggers of the detection results from the total 30-minute record are analyzed and listed here.

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

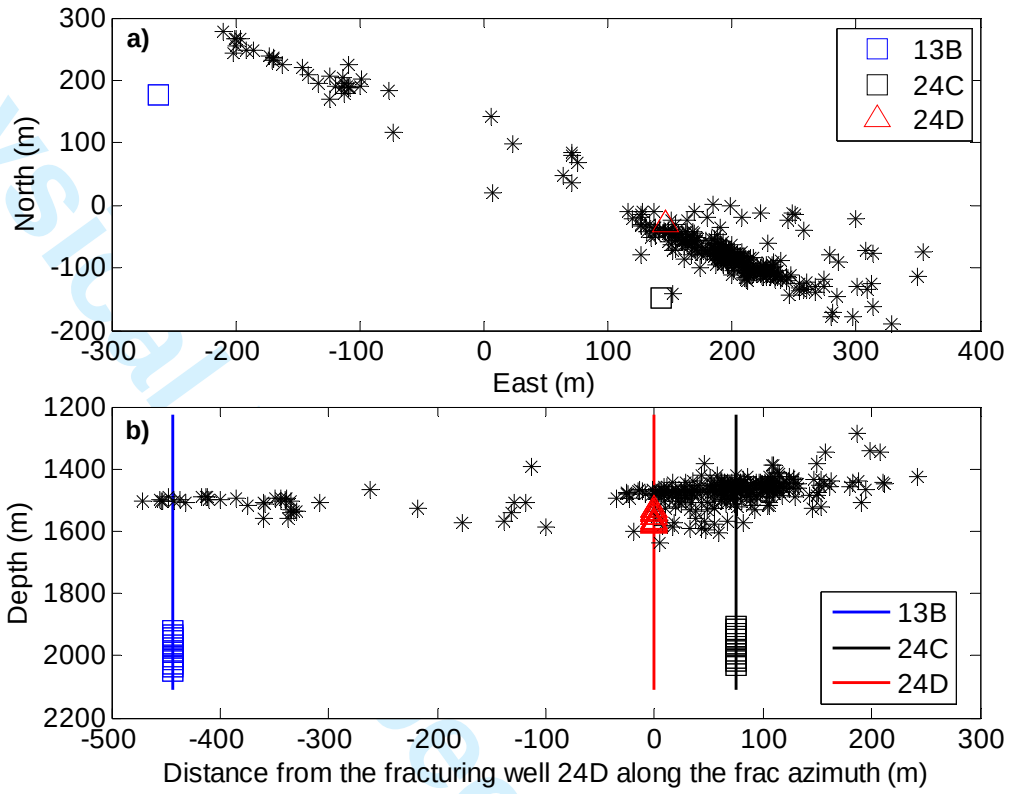

Figure 1. (a) Horizontal plane view of the microseismic event locations from one stage treatment plotted as black stars. The blue and black squares denote the monitoring wells 13B and 24C, respectively, while the fracturing well is shown as the red triangle. The origin (0, 0) corresponds to the wellhead location of well 13B. (b) The side view of the microseismic events. The blue squares and black squares represent the two twelve-level geophone arrays deployed in well 13B and 24C separately (from deep to shallow depths: geophone 1 to 12). The perforation locations are depicted as the red triangles in fracturing well 24D. Fewer events are detected on the far well 13B. Data from the far well 13B will be used in this study for subspace detection and signal enhancement.

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

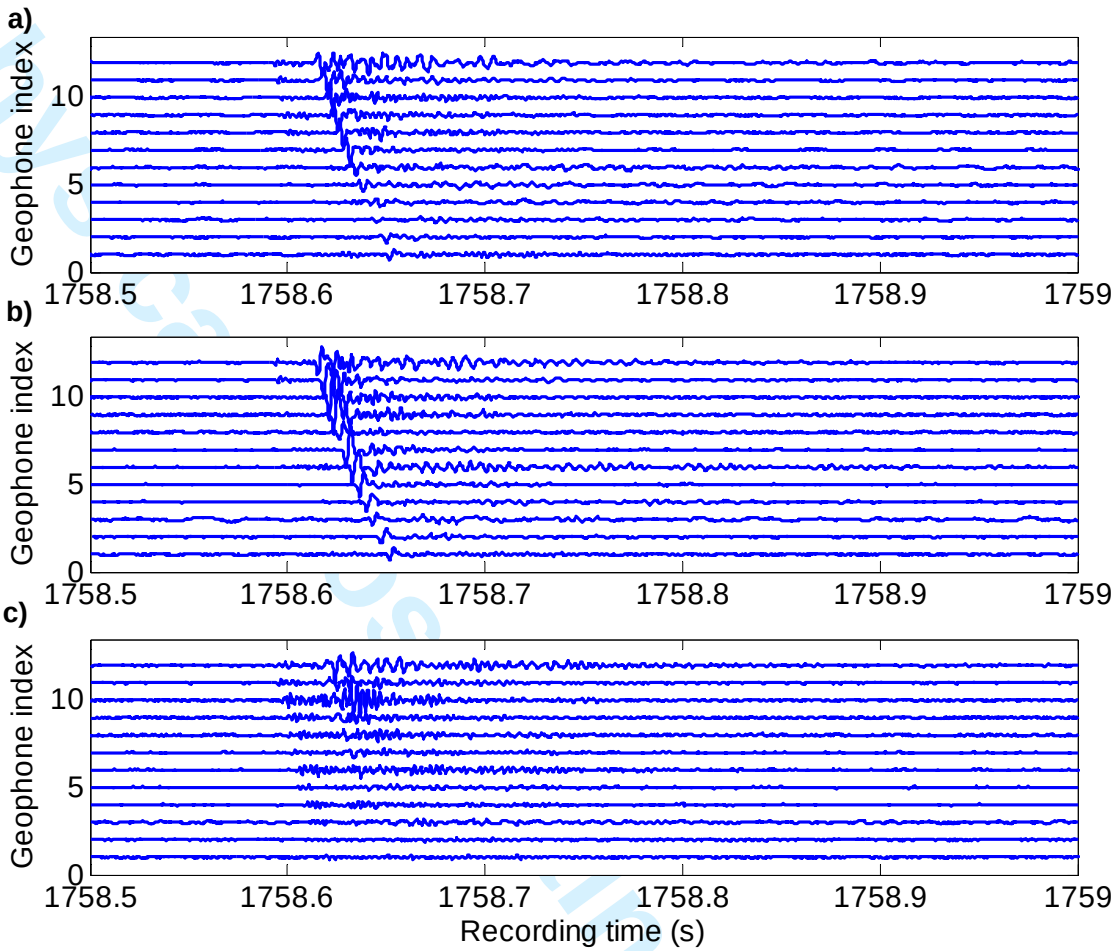

Figure 2. The three-component raw data plot for a typical event recorded in the far well 13B: (a) x component, (b) y component, (c) z component.

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

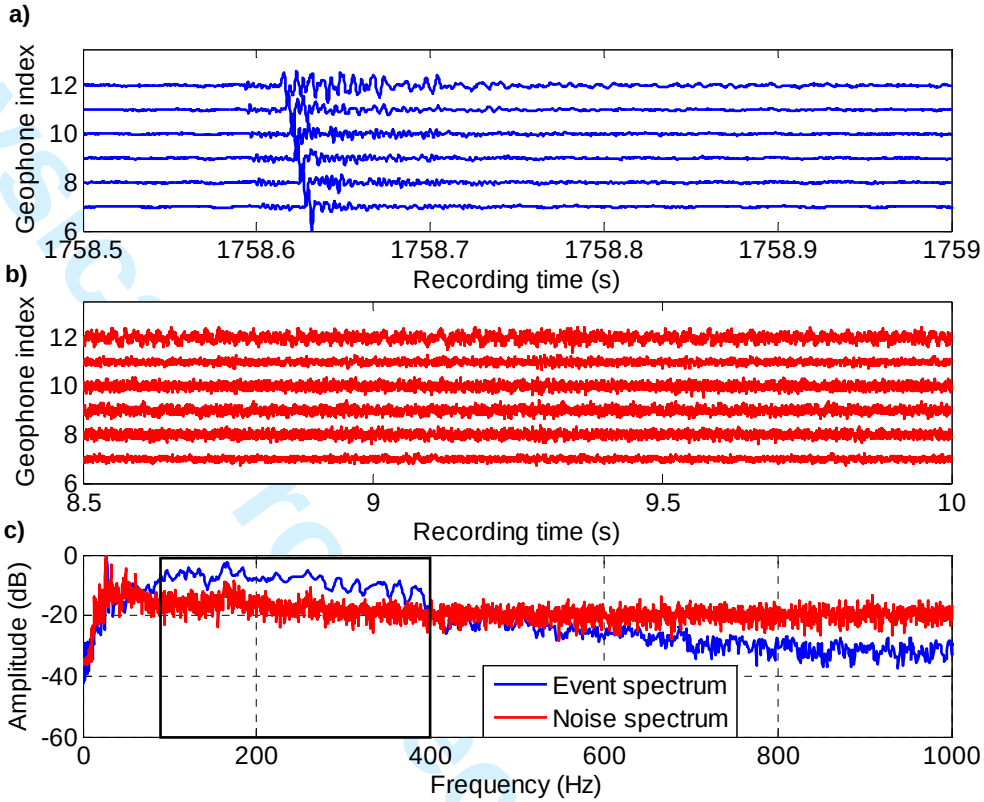

Figure 3. (a) The raw x component data of a 0.5s event record from geophones 7-12 in well 13B. (b) The raw x component data of a 0.5s noise segment recorded by geophones 7-12 in well 13B. (c) Amplitude spectrum of the raw event and noise data in the panels (a) and (b), averaged over all 6 geophones. The black square demonstrates the dominant signal frequency range of [100, 400] Hz.

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

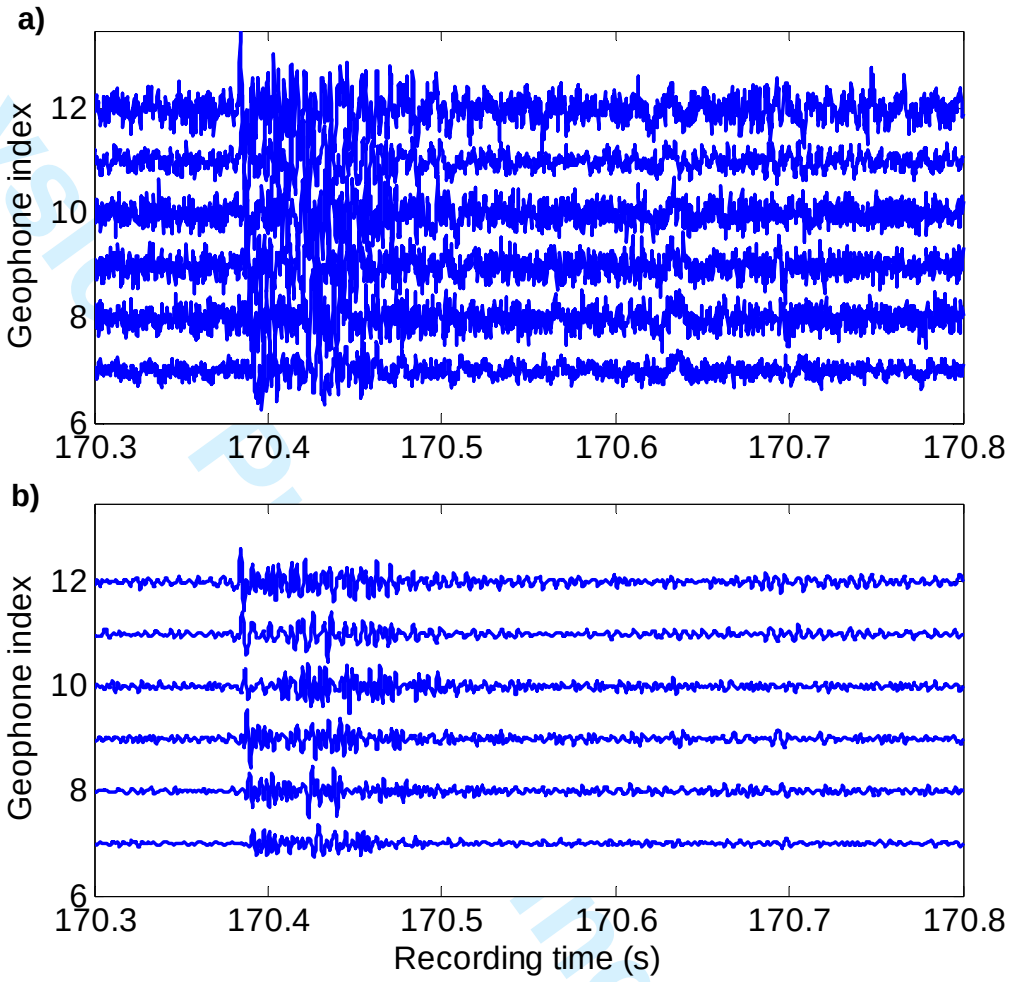

Figure 4. (a) The raw x component data of a 0.5s continuous record from geophones 7-12 in well 13B. (b) The [100, 400] Hz band-pass filtered result of the panel (a).

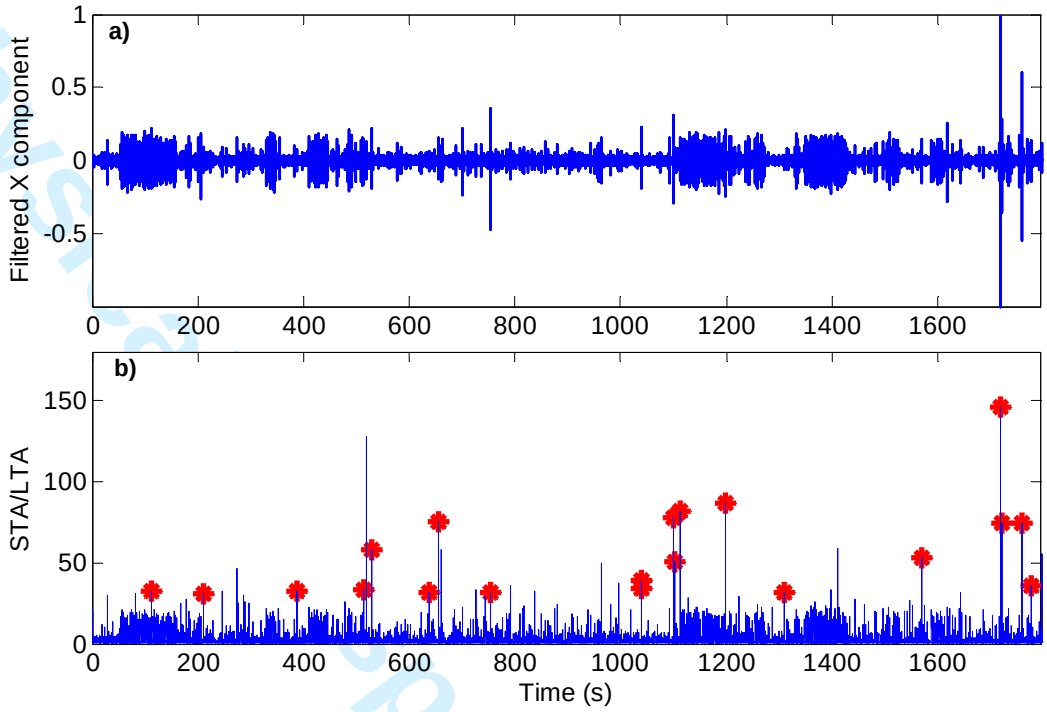

Figure 5. Array STA/LTA detection on a 30-min continuous record from far well 13B. a) The x component [100, 400] Hz band-pass filtered continuous data from one geophone in well 13B. b) The STA/LTA detection results on the channel-multiplexed data. The x, y, z component data from geophones 7-12 are used in the STA/LTA detection. The template event library for the subspace detector, comprising the M = 20 identified events using a conservative STA/LTA threshold of 30, is plotted in red stars.

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

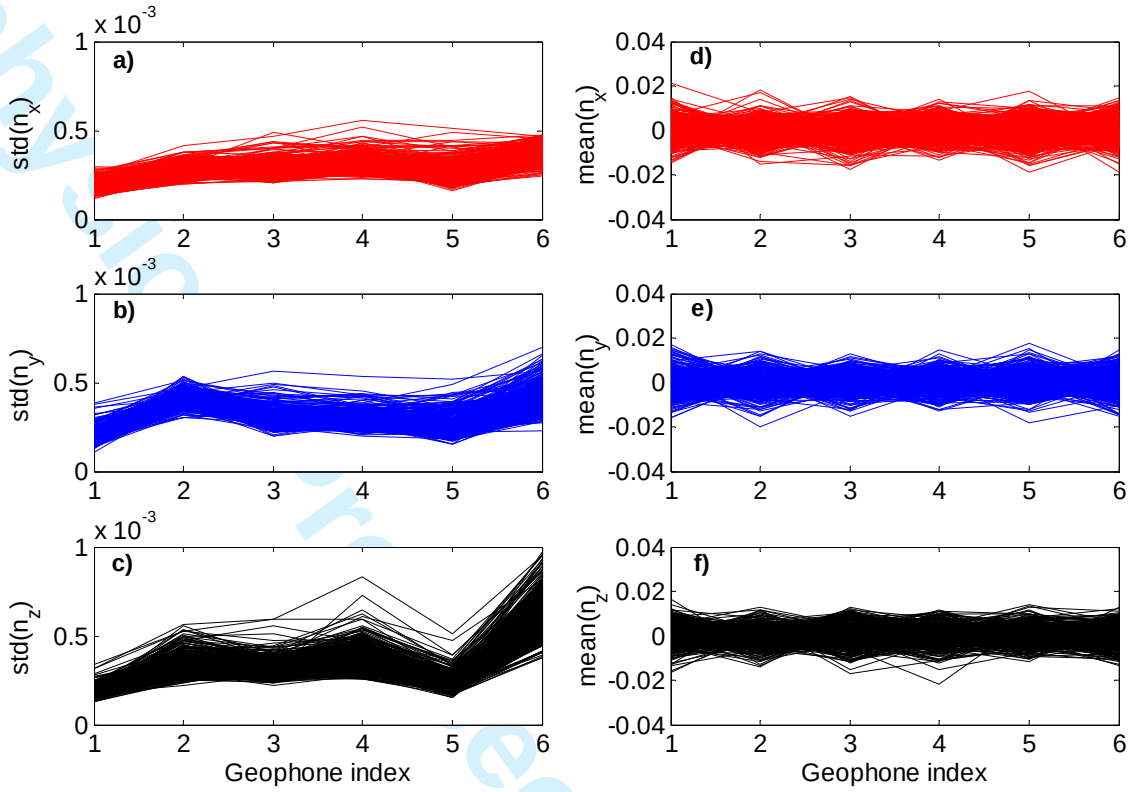

Figure 6. The standard deviation and mean of identified 454 noise data files across the six geophones (geophones 7-12 from well 13B). Left columns: noise standard deviation. a) x component. b) y component. c) z component. Right columns: noise mean as a multiple of its corresponding absolute maximum value. d) x component. e) y component. f) z component.

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

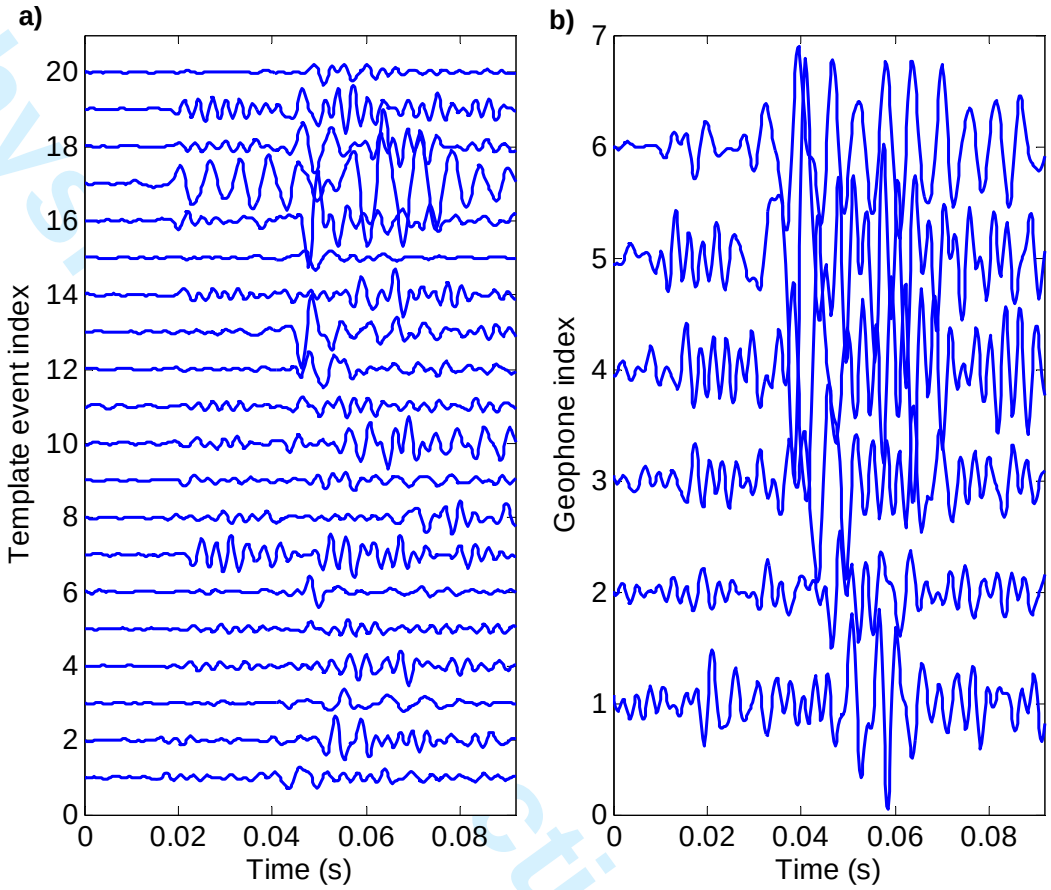

Figure 7. Waveform plot of the detected 20 template events (as described in Figure 5) after noise standard deviation normalization. a): Band-pass filtered unaligned waveforms of all 20 events from one geophone in well 13B. b): Band-pass filtered unaligned waveforms of one template event from all six geophones in well 13B (geophones 7-12).

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

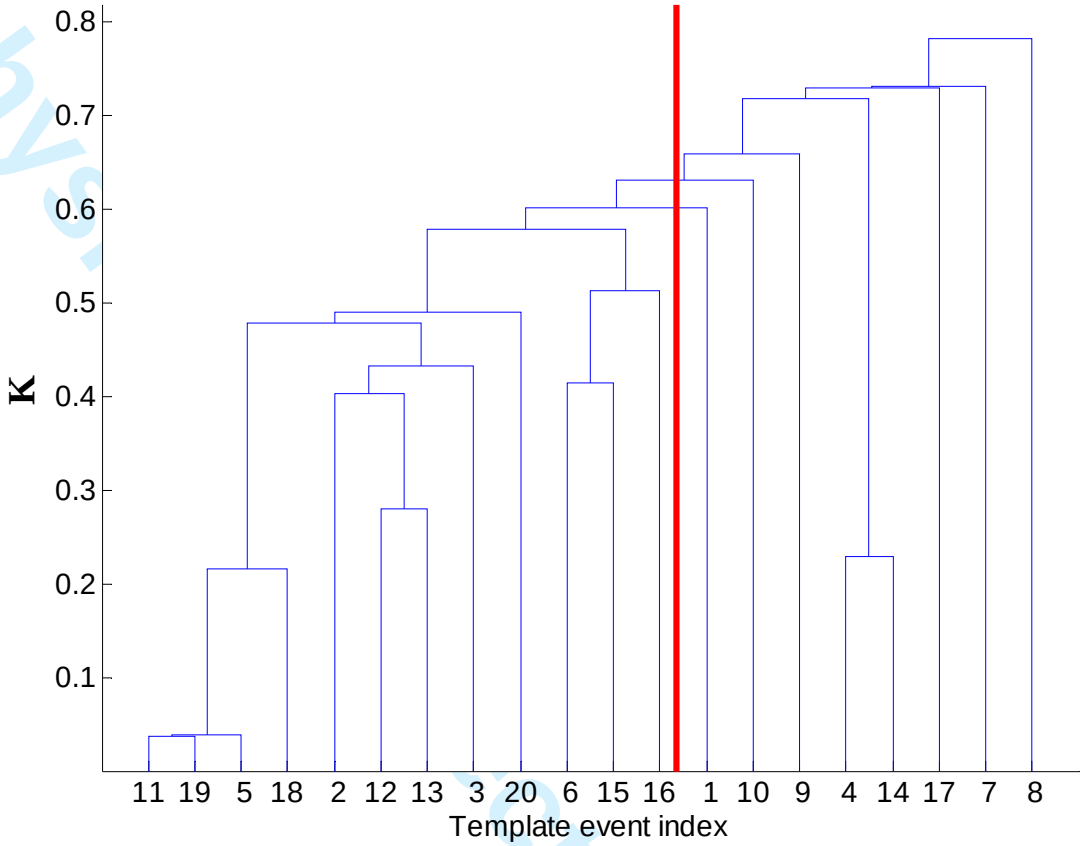

Figure 8. Template event clustering and design set event selection through the dendrogram using the single-link algorithm. The red line shows the termination of clustering with a maximum event dissimilarity distance of 0.6, which gives a design set comprising D = 12 events (events 11 to 16).

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

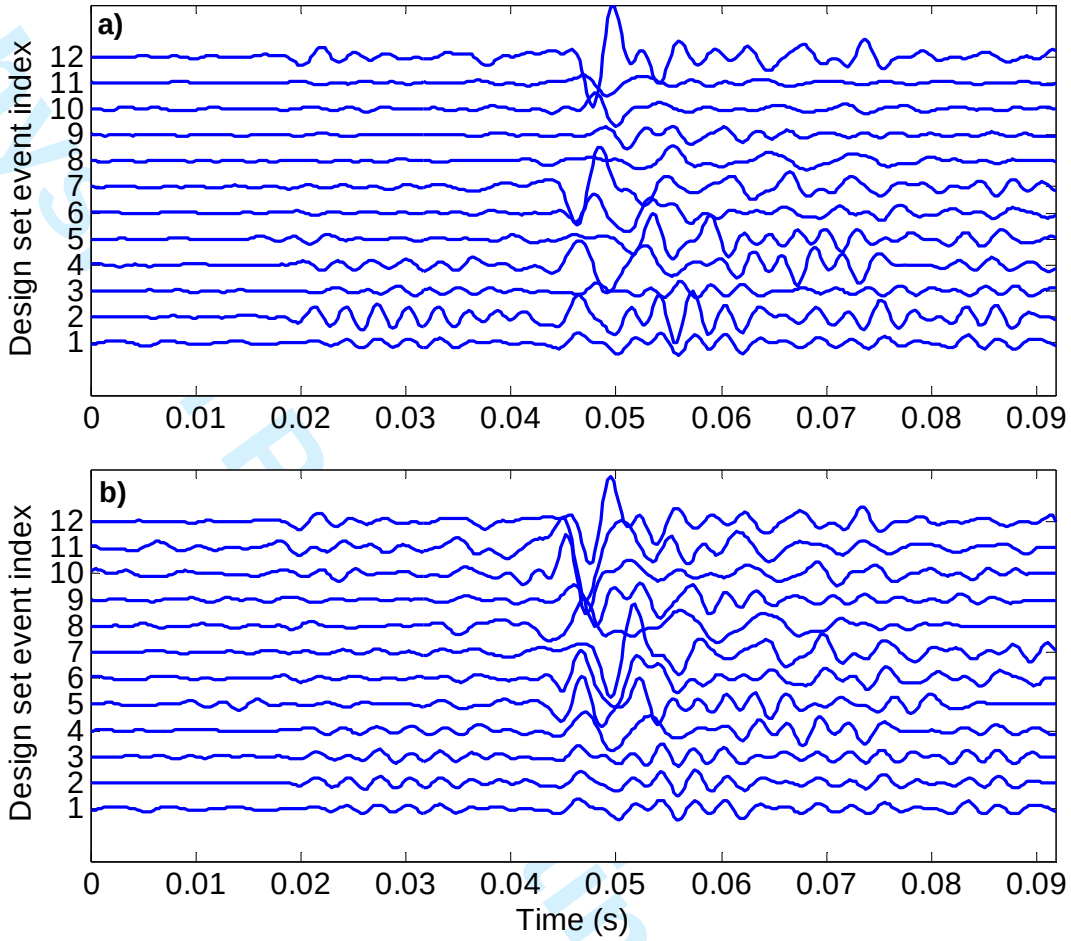

Figure 9. The waveform alignment of design set events using the single-link algorithm. a) The unaligned z component waveform plot from one geophone. b) The waveform plot of panel a) after alignment.

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

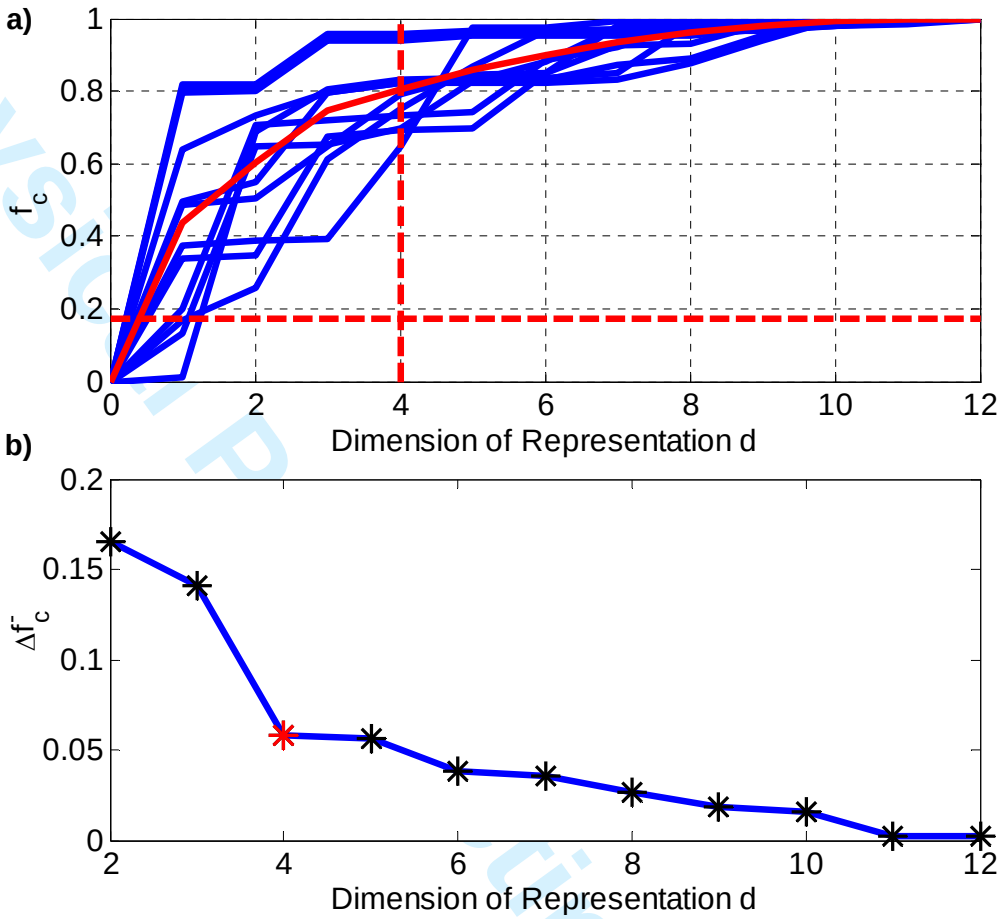

Figure 10. a) Fractional energy captured f_c as a function of dimension of representation d (also known as the signal subspace dimension) for each design set event is plotted in blue, while the average fractional energy capture $\bar{f}_c$ for all D=12 design set events as a function of d is shown in the red curve. A threshold of at least 80% average fractional energy capture plotted as the vertical red line gives an optimal subspace dimension d = 4. The horizontal red line shows the theoretical detection threshold for the subspace detector with d = 4, and false alarm rate of $P_F = 10^{-15}$. b) The increase in the average fractional energy capture $\Delta\bar{f}_c$ as a function of an increased subspace dimension d.

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

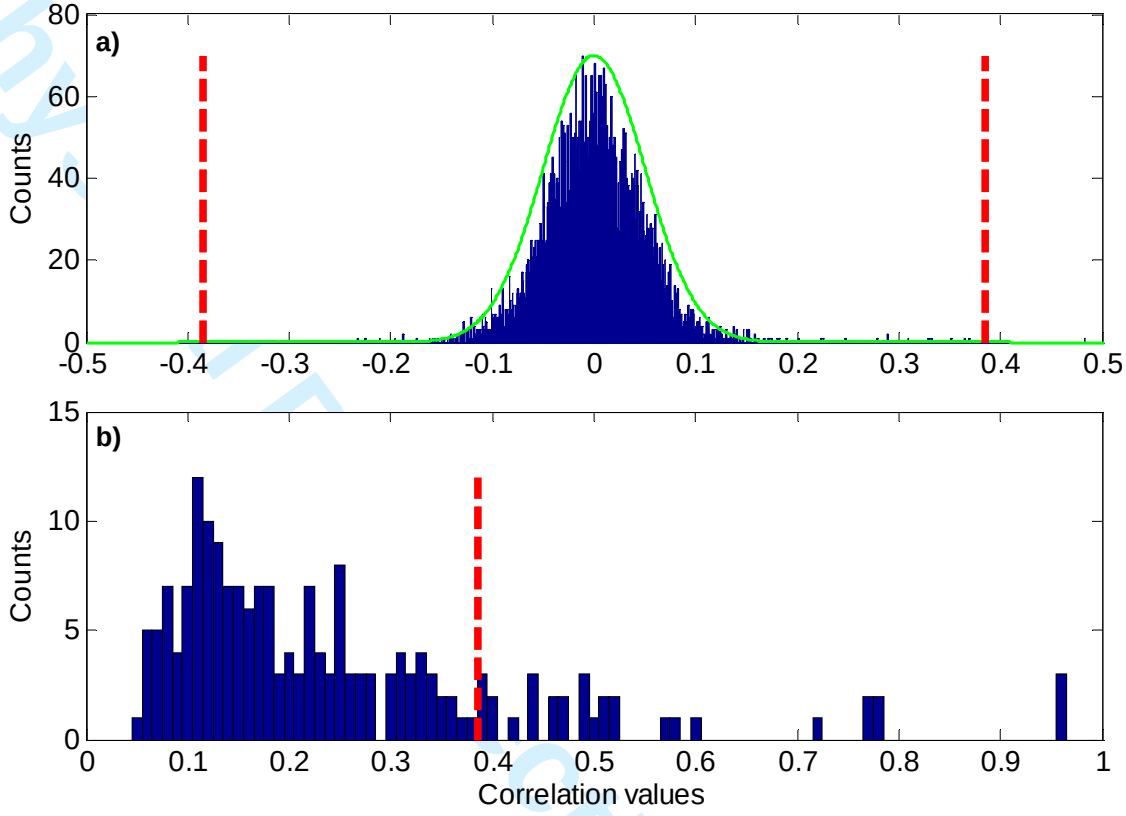

Figure 11. a) The histogram of correlation values between template event and noise. b) The histogram of correlation values between template events.

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

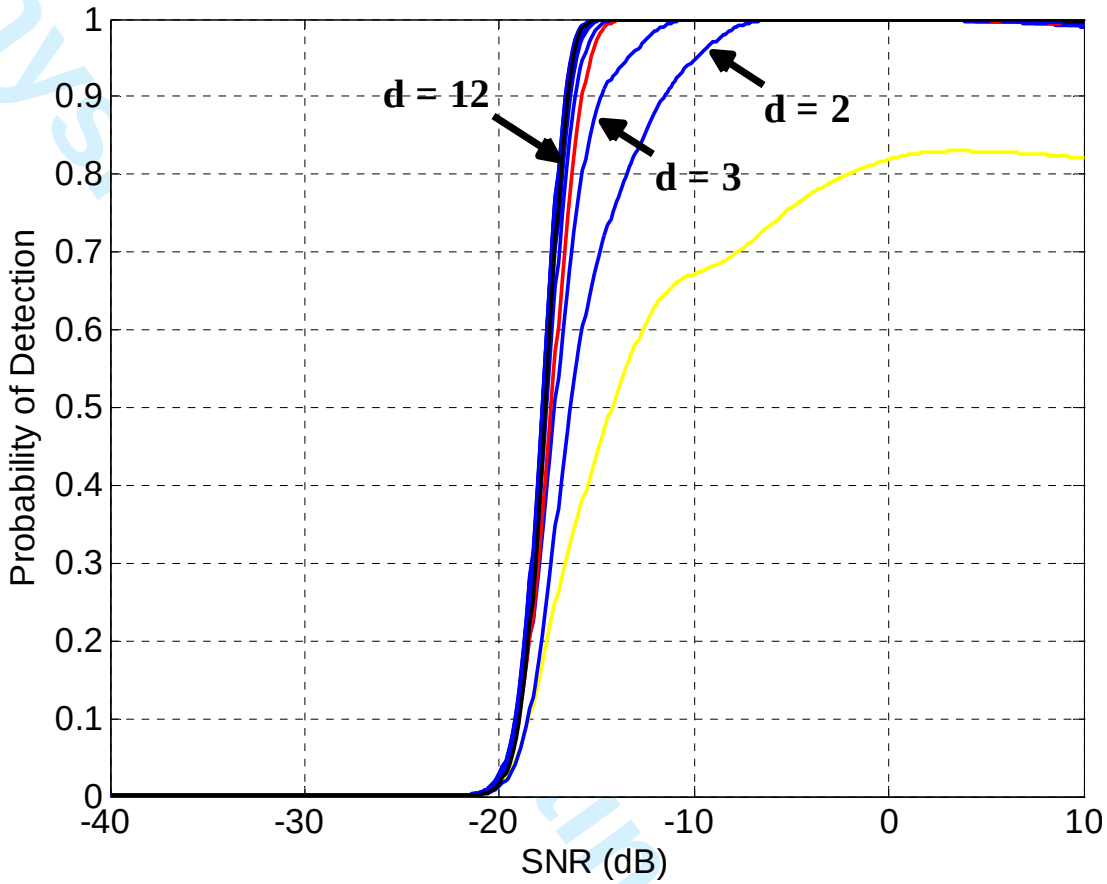

Figure 12. The probability of detection as a function of the SNR at a fixed false alarm rate $P_F = 10^{-15}$. In this case, the detection probabilities are calculated as a function of SNR for subspace dimensions ranging from 1 to 12. The detection probability curve for the selected subspace detector with d = 4 is plotted in red, while the yellow and black curves demonstrate the detection probability curves for the subspace detector with d = 1 and d = 12, respectively.

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

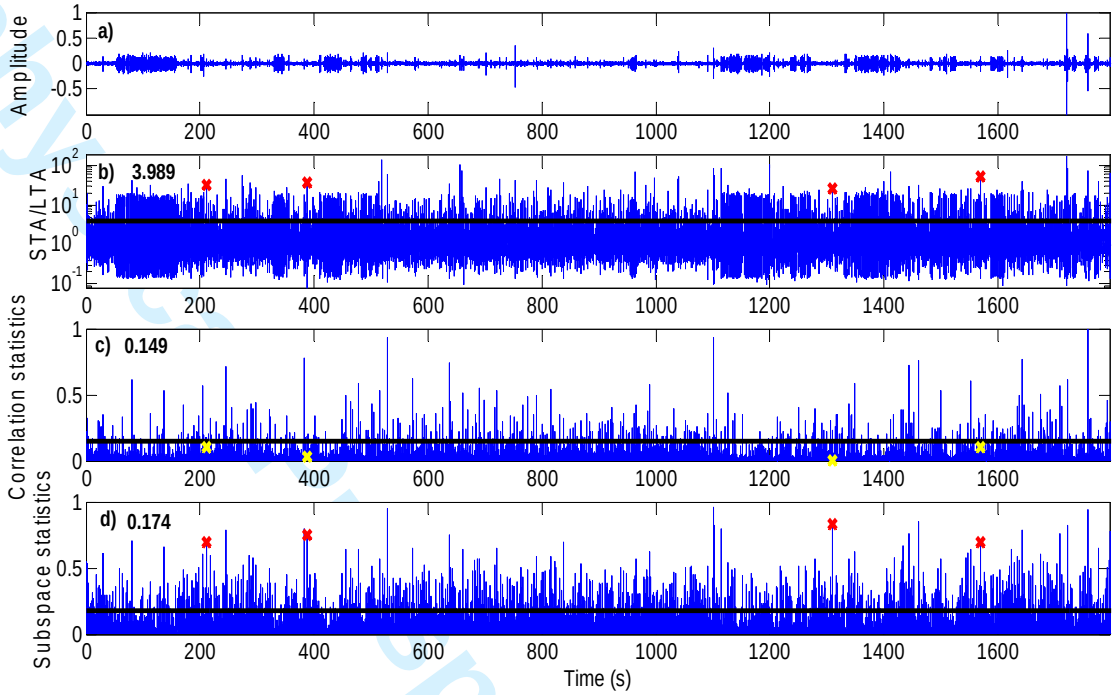

Figure 13. The comparison of detection results on a 30-min continuous record in far well 13B at a fixed false alarm rate $P_F = 10^{-15}$. The new channel-multiplexed data, formed by the x, y, z component data from geophones 7-12 after noise standard deviation normalization, are used in the detection. a) The [100, 400] Hz band-pass filtered x component data from one geophone in well 13B. b) The STA/LTA detection, c) the correlation detection, and d) the subspace detection (d=4) results on the new channel-multiplexed data. The threshold values at $P_F = 10^{-15}$, plotted as the black horizontal lines, are 3.989, 0.149, and 0.174 for the STA/LTA, correlation, and subspace detector, respectively. The four design set events missed by the correlation detector, but captured by STA/LTA and subspace detectors, are plotted as yellow and red crosses.

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

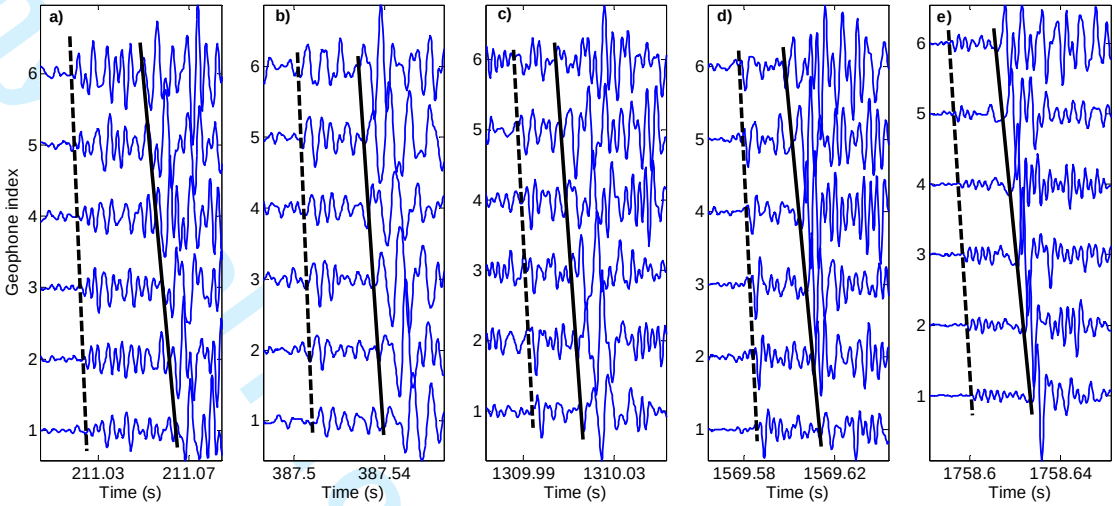

Figure 14. The band-pass filtered x component waveform plot. The dashed and solid black lines represent the P and S arrival picks on geophones 7-12 (geophone index: 1-6) in well 13B. a-d) The four design set events missed by the correlation detector, but captured by STA/LTA and subspace detectors at $P_F = 10^{-15}$. e) The correlation template event.

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

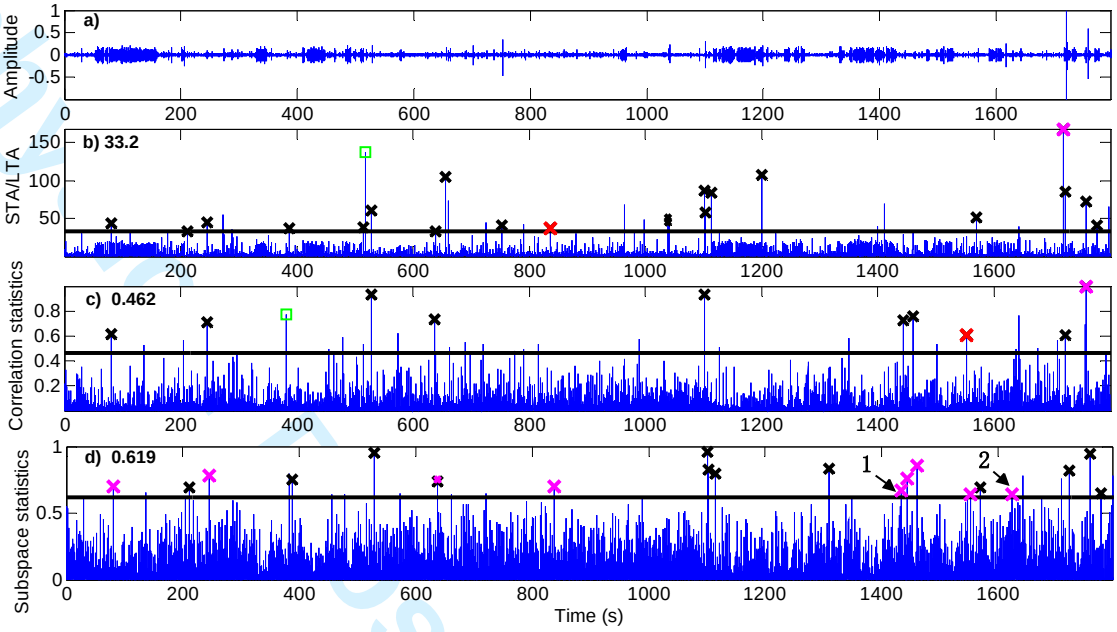

Figure 15. The comparison of the largest 35 triggers on a 30-min continuous record in far well 13B. The new channel-multiplexed data, formed by the x, y, z component data from geophones 7-12 after noise standard deviation normalization, are used in the detection. a) The [100, 400] Hz band-pass filtered x component data from one geophone in well 13B. b) The STA/LTA detector gives 21 events plotted as crosses, with the minimum detected event denoted as the red cross. The false alarm with the largest STA/LTA statistics is shown in the green square. One STA/LTA event missed by the subspace detector is plotted as the magenta cross. c) The correlation detector gives 10 events plotted as crosses, with the minimum detected event and correlation template event denoted as the red and magenta crosses, respectively. The false alarm with the largest correlation statistics is shown in the green square. d) The subspace detector with d=4 generates 21 events plotted as crosses, with 12 out of them being the design set events shown in black, and 9 additional detected events are plotted in magenta. Two events, detected by the subspace detector but missed by both STA/LTA and correlation detectors, are marked as 1, 2 on panel d).

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

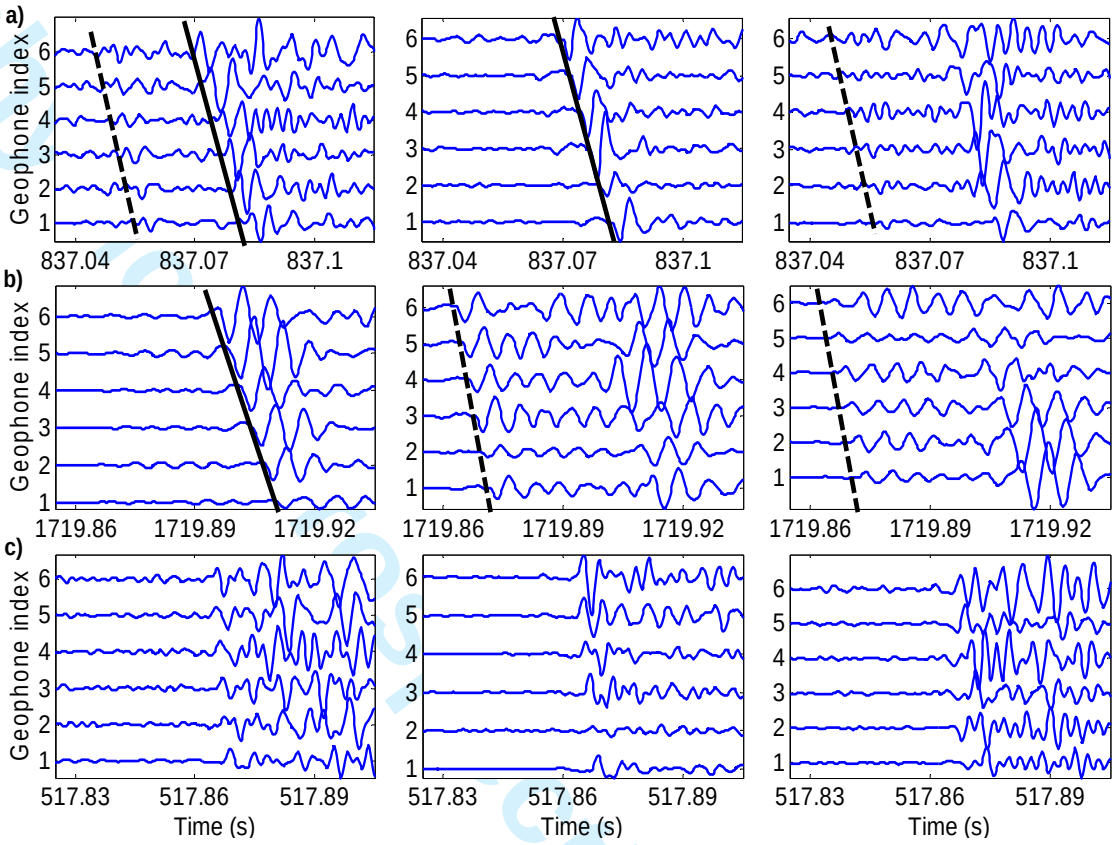

Figure 16. The waveform plot of the band-pass filtered data (columns from left to right: x, y, z components). The dashed and solid black lines represent the P and S arrival picks on geophones 7-12 (geophone index: 1-6) in well 13B. a) The minimum detected event from the array STA/LTA detector (see the red cross on Figure 15b). b) The STA/LTA event missed by the subspace detector (see the magenta cross on Figure 15b). c) The false alarm with the largest STA/LTA statistics (see the green square on Figure 15b).

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

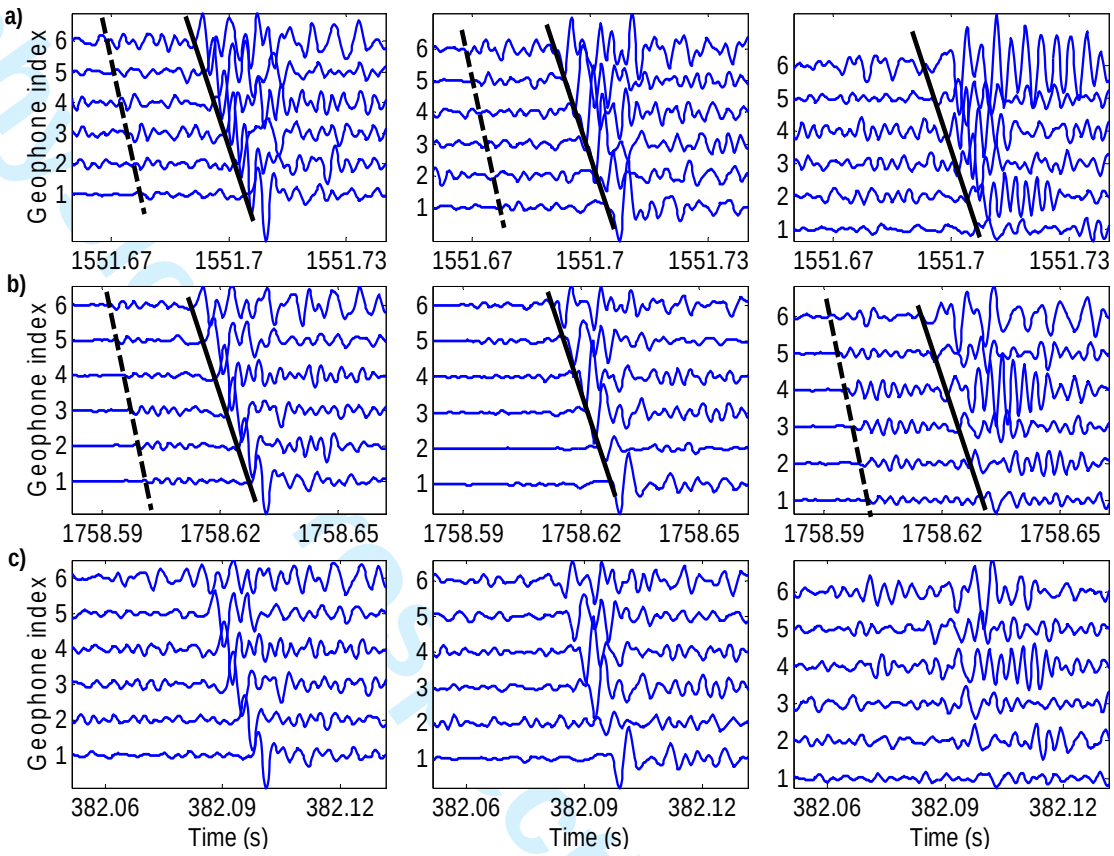

Figure 17. The waveform plot of the band-pass filtered data (columns from left to right: x, y, z components). The dashed and solid black lines represent the P and S arrival picks on geophones 7-12 (geophone index: 1-6) in well 13B. a) The minimum detected event from the array correlation detector (see the red cross on Figure 15c). b) The correlation template event of the array correlation detector (see the magenta cross on Figure 15c). c) The false alarm with the largest correlation statistics (see the green square on Figure 15c).

.

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

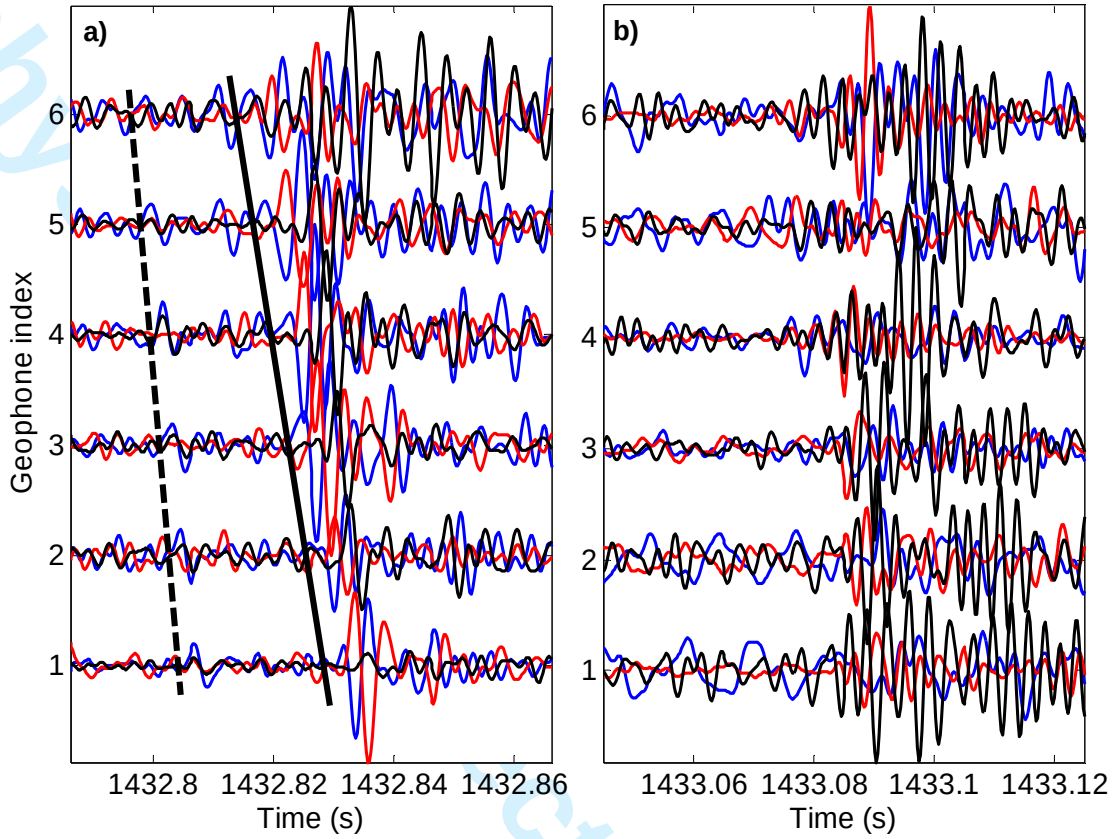

Figure 18. The three-component waveform plot of event 1 on Figure 15d, detected by the subspace detector, but missed by both array STA/LTA detector and array correlation detector (x in blue, y in red, z in black). The dashed and solid black lines represent the P and S arrival picks. a) The band-pass filtered data from geophones 7-12 (geophone index: 1-6) in the far well 13B. b) The corresponding detected waveforms from geophones 7-12 (geophone index: 1-6) in the nearby well 24C. The time difference between a) and b) is to account for the possible arrival time difference between the far well 13B and nearby well 24C.

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

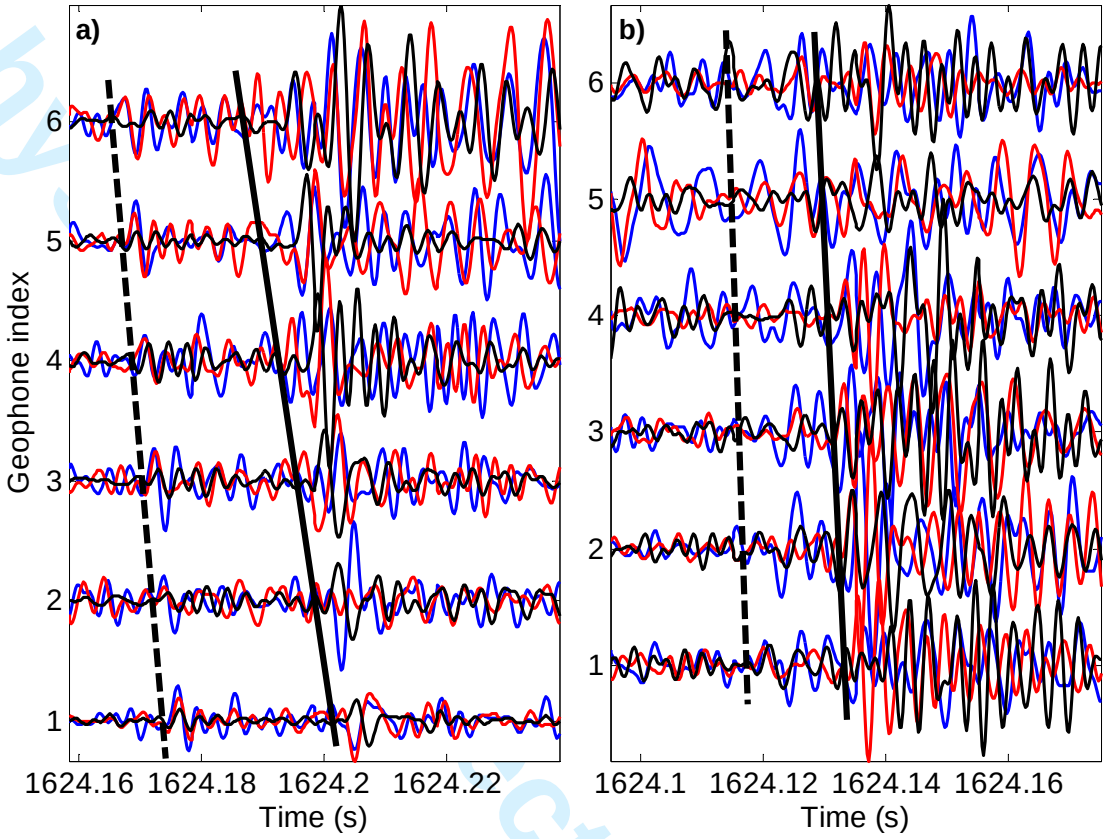

Figure 19. The three-component waveform plot of event 2 on Figure 15d, detected by the subspace detector, but missed by both array STA/LTA detector and array correlation detector (x in blue, y in red, z in black). The dashed and solid black lines represent the P and S arrival picks. a) The band-pass filtered data from geophones 7-12 (geophone index: 1-6) in the far well 13B. b) The corresponding detected waveforms from geophones 7-12 (geophone index: 1-6) in the nearby well 24C. The time difference between a) and b) is to account for the possible arrival time difference between the far well 13B and nearby well 24C.

EAGE Publications B.V., PO Box 59, 3990 DB, Houten, The Netherlands

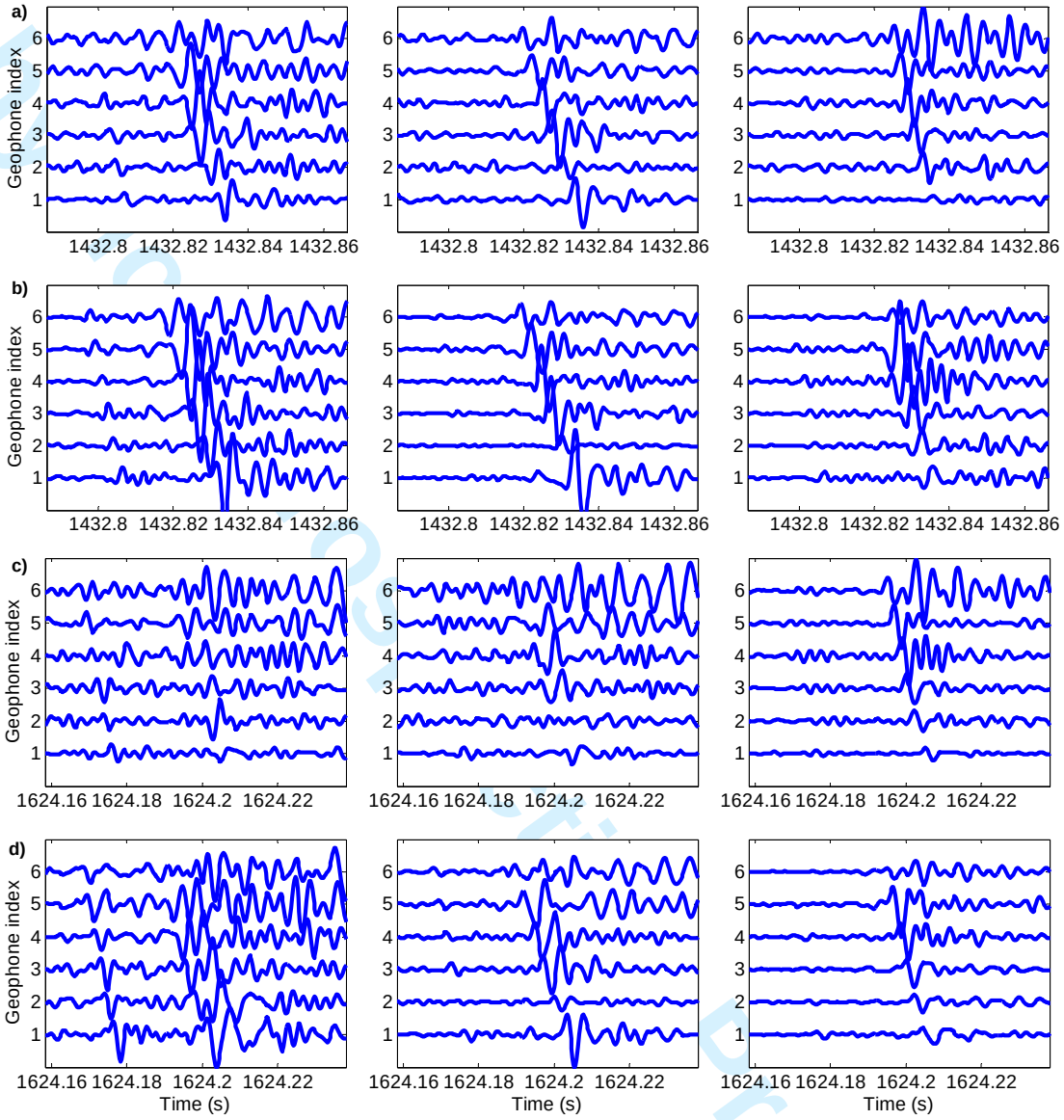

Figure 20. The subspace projection approach for microseismic signal enhancement. The waveform plot of the band-pass filtered (x, y, z) component data from geophones 7-12 (geophone index: 1-6) in the far well 13B (columns from left to right: x, y, z components). a) Data from the detected event 1 as shown in Figure 18, before signal enhancement. b) Data from the detected event 1 as shown in Figure 18, after signal enhancement. c) Data from the detected event 2 as shown in Figure 19, before signal enhancement. d) Data from the detected event 2 as shown in Figure 19, after signal enhancement.