A technique for identifying microseismic multiplets and application to the Valhall field, North Sea

Stephen J. Arrowsmith[1] and Leo Eisner[2]

ABSTRACT

A fast, fully automatic technique to identify microseismic multiplets in borehole seismic data is developed. The technique may be applied in real time to either continuous data or detected-event data for a number of three-component receivers and does not require prior information such as P- or S-wave time picks. Peak crosscorrelation coefficients, evaluated in the frequency domain, are used as the basis for identifying microseismic doublets. The peak crosscorrelation coefficient at each receiver is evaluated with a weighted arithmetic average of the normalized correlation coefficients of each component. Each component is weighted by the maximum amplitude of the signal for that component to reduce the effect of noise on the calculations. The weighted average correlations are averaged over all receivers in a time window centered on a fixed lag time. The size of the time window is determined from the dominant period in the signal, and the lag time is the time that maximizes the average correlation coefficient. The technique is applied to a three-component passive seismic data set recorded at the Valhall field, North Sea. A large number of microseismic doublets are identified that can be grouped into multiplets, reducing the total number of absolute event locations by a factor of two. Seven large multiplets reflect the repeated multiple rerupturing (up to 30 times on a single fault) and significant stress release. Two major faults dominate the seismic activity, causing at least one-fourth of the observed events.

INTRODUCTION

Microseismic events (small earthquakes and acoustic emissions) can be induced in hydrocarbon reservoirs as a result of the changing stress distribution resulting from oil extraction or hydraulic fracturing. They can occur anywhere in the reservoir or surrounding rock, but often they rerupture the same fault. In such a case, the events may have similar hypocenter locations and focal mechanisms, producing similar recordings at receivers. Two such similar events are called doublets, and a group of more than two similar events is called a multiplet. We define a multiplet as a cluster of n events ($n \geq 2$) where each event is a doublet with at least one other event in the multiplet. We feel this is a more natural definition for multiplets than the requirement that all events in a multiplet be mutually similar. Multiplets were first observed in natural earthquakes (Geller and Mueller, 1980), and the recorded events are correlated only for limited durations and for low-pass filtered data. For highly heterogeneous media, high waveform correlation is conditional upon events having similar source mechanisms and point-source-location separations less than one-fourth of the dominant wavelength. The doublet definition is therefore frequency dependent. Furthermore, two events comprising a doublet are similar only over a limited-time interval because multiple scattered late-arriving waves reflect small discrepancies in source radiation or position.

Multiplets have many practical uses. First, the similarity of waveforms in multiplets may be exploited to locate events relative to each other (Poupinet et al., 1984; Moriya et al., 1994; Waldhauser and Ellsworth, 2000). Such locations can provide a much higher relative location precision than is achieved from absolute locations and can be used to identify fault zones that are not apparent with absolute locations (Moriya et al., 1994) or to estimate fracture density (Lees, 1998). Multiplets can be also used to monitor the reactivation of faults, fractures, or flow obstacles during hydrofracturing (e.g., Rutledge et al., 2004). They can provide information on the noise level in the data by separating signal and noise. Perhaps the most important advantage is that multiplet identification lets us group events of the same origin based on the physics of the source.

Figure 1 is a proposed scheme for the real-time location of microseismic events. Crosscorrelating every event with every other event in a data set would scale with the square of the number of events and would quickly cause the algorithm to become very expensive (J. Rickett, 2005, personal communication). The scheme in Figure 1 avoids this problem. When a microseismic event is detected, the waveform is compared with a library of master events to determine if the event is a doublet of a master trace. If the event is not the doublet of a master trace, it is located using a conventional location algorithm and added to the library of master events. If the event is the doublet of a master trace, it is located relative to a master event for that group. Next, if the event has a higher S/N ratio than the master event, it replaces the old master.

For tectonic earthquakes, multiplets are especially common in the creeping zones of faults (e.g., Nadeau et al., 1995; Waldhauser and Ellsworth, 2000; Stich et al., 2001). Multiplets also are observed commonly during hydraulic fracture monitoring in hot, dry rocks (e.g., Lees, 1998; Li et al., 1998; Fréchet et al., 1989; Moriya et al., 2003) or oil and gas reservoirs (e.g., Moriya et al., 1994; Rutledge and Philips, 2003). Multiplet analyses from large industrial data sets require the identification of multiplets in a data set. The practical real-time application of multiplet identification in large data sets, such as those recorded in hydraulic fracturing, requires a fully automatic, fast technique. Previous techniques for the automatic detection of multiplets (Maurer and Deichmann, 1995; Cattaneo et al., 1999; Stich et al., 2001; Schaff and Richards, 2004) focus on earthquake seismology and use receivers distributed on the free surface of the earth. These studies use data recorded at networks of seismic receivers. Only the vertical components of the recorded wavefields are used on account of generally worse S/N ratios on horizontal components. In every case, they use preprocessed data with prepicked P- and S-wave phases.

We present a novel identification technique for identifying multiplets which is suitable for borehole data sets such as those recorded during hydraulic fracturing or passive seismic monitoring, and it uses all three components. Using all three components helps to compensate for the lack of receiver distribution (compared to surface seismic data sets) and is a natural generalization to borehole data where there is no preferred component. The technique does not require previously picked P- or S-wave arrivals. We then apply the technique to a passive seismic data set recorded in the Valhall field, North Sea.

TECHNIQUE FOR IDENTIFYING DOUBLETS

We define a doublet as two events that have highly correlated waveforms. Thus, the time and polarization of the P- and S-waves are very similar, and we use the crosscorrelation coefficient to help identify doublets. The normalized crosscorrelation coefficient is evaluated in the frequency domain by analogy with the convolution theorem (Ifeachor and Jervis, 1993). For two traces, $x_1(t)$ and $x_2(t)$, the normalized crosscorrelation function may be evaluated as

$$C_x(\tau_i) = \frac{F_D^{-1}(X_1^*(f)X_2(f))}{\sqrt{\sum x_1^2(t) \sum x_2^2(t)}}, \tag{1}$$

where F_D^{-1} denotes the inverse discrete Fourier transform, $X_1^*(f)$ is the complex conjugate of the Fourier transform of $x_1(t)$, and $X_2(f)$ is the Fourier transform of $x_2(t)$.

The information provided by three-component receivers is exploited in such a way that the effect of noise on the calculations is down weighted. This is particularly important when we apply this algorithm in real-time processing to randomly oriented borehole receivers. A simple crosscorrelation of components with low S/N ratios may distort the averaged coefficient. Assuming the maximum amplitude of each component reflects the S/N ratio, we average the normalized crosscorrelation functions from all components with a weight of the maximum amplitude on each component. Subsequently, for receiver i, the signal weighted average of the crosscorrelation functions from the different components is given by

$$C_{Ri}(\tau_i) = \frac{A_x C_x(\tau_i) + A_y C_y(\tau_i) + A_z C_z(\tau_i)}{A_x + A_y + A_z}, \tag{2}$$

where C_x, C_y, C_z are the normalized crosscorrelation functions for each component; A_x, A_y, A_z are the maximum amplitudes for each component; and τ_i is the lag time of the crosscorrelation for the ith receiver. Here, all times are defined as when the events are actually recorded. Amplitude weighting significantly reduces the effect of noise in our computations by down-weighting components where the S/N ratio is poor.

When multiple receivers exist, the peak crosscorrelation coefficient for two events in a doublet occurs at the same time lag at each receiver within a time window of length Δt. The time window Δt is less than one-fourth of the dominant period in the signal, because we define a doublet as two similar events with a maximum event separation of one-fourth of the dominant wavelength. We find the peak crosscorrelation coefficient C for all receivers by averaging the peak crosscorrelation functions for each separate receiver for τ_i within one-fourth
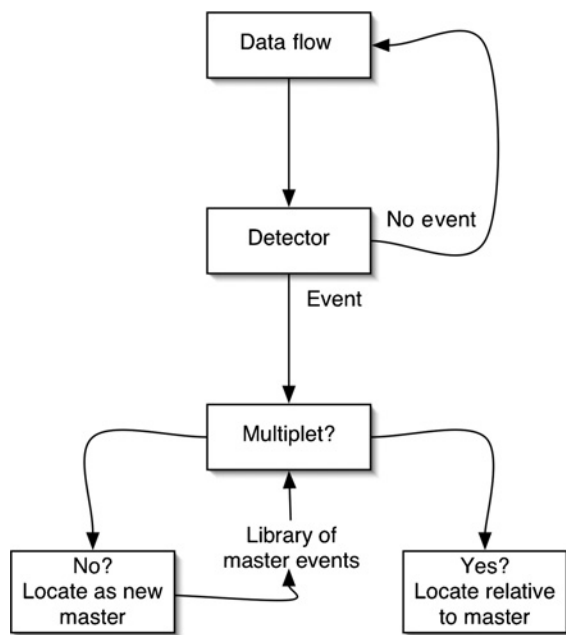

Figure 1. A decision tree for the real-time application of the multiplet-identification algorithm in analyses of microseismic events.

period. Therefore, we seek to find the coefficient given by

$$C = \max_{\tau_i} \left\{ \frac{\sum_{i=1}^{m} C_{Ri}(\tau_i)}{m} \right\}, \qquad (3)$$

where C_{Ri} is the peak crosscorrelation function for the ith receiver; $\tau_i = \tau \pm \Delta t_i$ for all i, the lag times of the crosscorrelation (where $i = 1, \ldots, m$); and m is the number of receivers.

This technique identifies doublets at different correlation threshold levels dependent on the application; e.g., automated processing may use a higher threshold to ensure waveforms are consistent, while a lower threshold may be used for manual preprocessing and a detailed study of the connectivity and relative relations between different multiplet groups. The technique should use a time window that encompasses the P- and S-wave arrivals to correlate both direct waves from the sources. Time-windowing of the data eliminates multiple scattered incoherent waves and parts of the records with noise (such as signal before the P-wave arrival). In addition, low-pass filtering can eliminate high-frequency noise that would decrease the crosscorrelation coefficients between doublets. The peak crosscorrelation coefficients are dominated by large amplitudes (effectively squared traces); therefore, the correlation will be small if there is a small phase shift between two large arrivals in the recordings. This can be true for recordings of a doublet with a different P- to S-wave traveltime delay and comparable amplitudes of P- and S-waves.

To identify doublets with larger shifts and comparable amplitudes of P- and S-waves, we need to use a lower threshold or modify the technique and evaluate the crosscorrelations on prepicked P- and S-waveforms independently. This implementation could improve the algorithm and would increase the number of identified multiplets. However, it would better suit a more complex study of multiplets (not real-time analysis).

The definition of doublet is largely subjective because the requirement of similarity may depend upon the specific application. Furthermore, doublet identification is dependent upon the data bandwidth and upon the complexity of the medium between the sources and receivers. However, to identify doublets automatically, we require some threshold measure of similarity. One way to determine such a threshold is to conduct a synthetic test using a realistic bandwidth and noise distribution.

DETERMINING A DOUBLET DETECTION THRESHOLD: APPLICATION TO A SYNTHETIC DATA SET

We have applied the technique described above to a synthetic data set with a range of known event locations and source mechanisms. The 3D synthetic model used in this test contains a layered structure typical of sedimentary basins with an additional 3D body. Three receivers are located at 50-m intervals along a vertical line to simulate a borehole acquisition. Twenty source locations are used. For each location we simulate seven different source mechanisms. Waveforms are computed using a finite-difference algorithm with five spatial gridpoints-per-minimum wavelength at 100 Hz and a delta source-time function (Figure 2). The data are filtered using

a band-pass Butterworth filter between 5 and 100 Hz. We evaluate the peak crosscorrelation coefficient of equation 3 for every event pair in the data set using the technique outlined above. Only events with the same source mechanisms are crosscorrelated. The length of the time windows used is 300 ms.

Figure 3 shows the resultant plot of peak crosscorrelation coefficient $C(\tau)$ against the spatial separation between every event pair. In this data set the dominant frequency in the seismograms is approximately 100 Hz, as shown in Figure 4; therefore, the one-fourth wavelength maximum separation of doublets is 10 m for an average S-wave velocity of 4000 m/s at the source region. Figure 3 shows that for most all events separated by less than 10 m, the peak crosscorrelation coefficients of equation 3 are greater than 0.9. At greater event
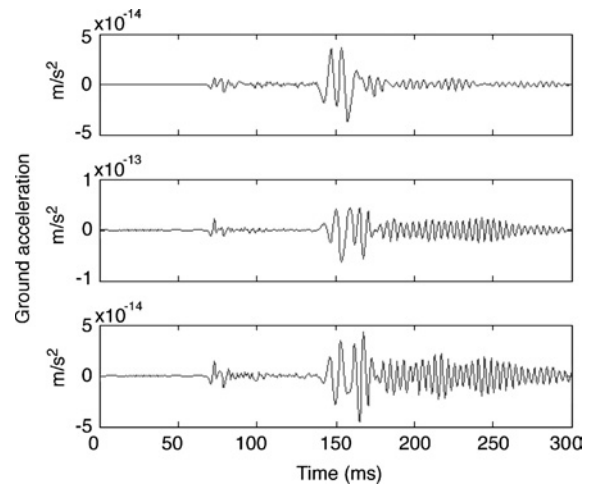

Figure 2. Particle acceleration synthetic seismograms for two events identified as doublets with no added noise.
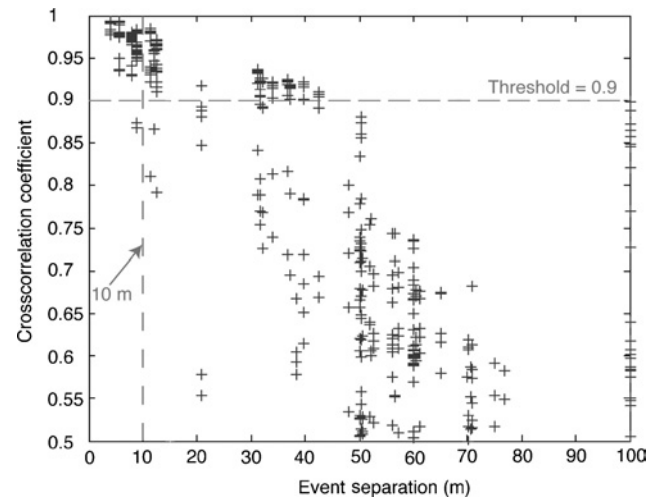

Figure 3. The peak crosscorrelation coefficients of equation 3 obtained for each pair of events, plotted against the event separation. A threshold of 0.9 separates event pairs closer than 10 m apart (doublets) from event pairs separated by greater distances. Only the positive coefficients are shown. There are no negative coefficients at event separations less than 10 m in any of the cases.

separations, where the event pairs are not doublets, and the peak crosscorrelation coefficients are generally lower than
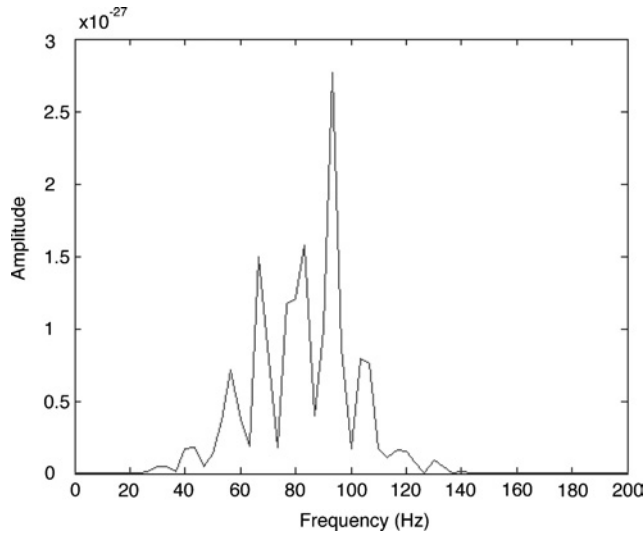

Figure 4. Spectrogram illustrating the frequency content of a typical synthetic seismogram used in this experiment.
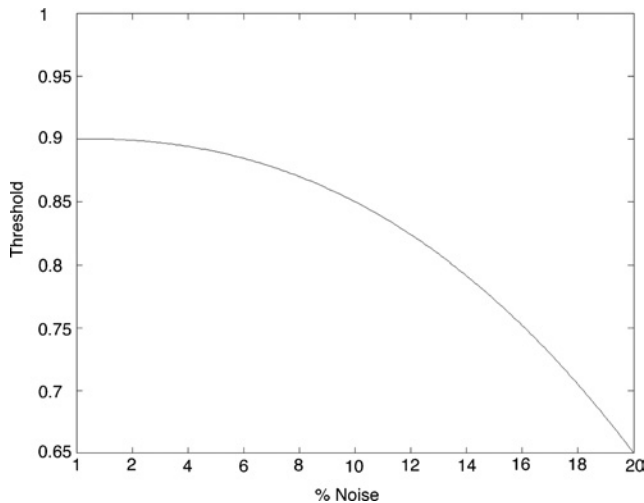

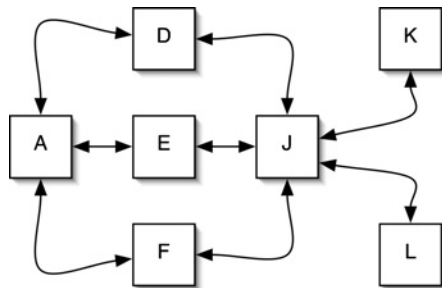Figure 5. The peak crosscorrelation threshold for identifying doublets versus noise level in the data.

Figure 6. Schematic showing the concept of the clustering algorithm used to identify families of doublets (multiplets). The lettered boxes represent seismic events, and the lines connecting events represent identified doublets.

0.9. Therefore, a peak crosscorrelation threshold of 0.9 would identify 97% of the doublets in this synthetic data set.

The threshold of 0.9 also identifies a few event pairs with separations between 10 and 50 m as multiplets. Since these events lie at greater separations than one-fourth of the dominant wavelength, they are not technically doublets. However, the relative location algorithm may still be able to locate them because the waveforms are similar and relative P-to-S-wave times can be determined. The waveforms in this synthetic data set are probably simpler than real data because the synthetic model does not represent the complexity of the real earth.

To apply this algorithm to real data, we must study the effect of noise on the peak crosscorrelation coefficients. Random noise has been added to the data with a standard deviation scaled by the maximum amplitude of the seismograms. Different S/N ratios have been studied to examine the effect of varying noise on the threshold for identifying doublets. With increasing noise level, the magnitude of the peak crosscorrelation coefficients decreases, and subsequently we need to lower the threshold for identification of doublets. This is shown in Figure 5, which is a result of multiple numerical tests with increasing synthetic noise level. Beyond a 20% noise level, this technique becomes unsuitable for identifying doublets.

CLUSTERING DOUBLETS INTO MULTIPLETS

Similar groups of doublets are clustered into larger families called multiplets using graph theory. Previous publications (e.g., Maurer and Deichmann, 1995) use an $N \times N$ matrix (where N is the number of events) to visualize the doublets among the events in the data set. The multiplets are then selected manually. We present a new method for clustering the doublets into multiplets. Multiplet identification is the same as a search through a graph, with the events represented by the nodes of a graph and the identified doublets represented by the edges connecting the nodes. The connected part of this graph is the multiplet. Our definition of multiplets does not require mutual similarity among all events within a multiplet but allows for chainlike similarity. This is particularly suitable for hydraulic fracture monitoring where similar faults are activated as the fracturing progresses (Rutledge et al., 2004). For example, in Figure 6 node (event) A is directly connected with (the doublet of) nodes D, E, and F but not nodes J, K, or L. However, all of the nodes are part of the same multiplet. The connected part of a graph can be sought using the breadth-first search algorithm (Weiss, 1994), where the cost function is zero if the vertices are connected and a finite number if vertices are disconnected.

APPLICATION TO VALHALL FIELD, NORTH SEA

The Valhall field, located in the Norwegian part of the North Sea, has been in production since 1982. The field is an overpressured, undersaturated Upper Cretaceous chalk reservoir that forms a double-plunging north-northwest–southsoutheast anticline. The reservoir suffers from significant compaction and subsidence of the overburden from extraction of oil. For a summary of the Valhall field, including maps and details of the geology and field operations, see, e.g., Barkved et al. (2003).

Data set

To understand stress regimes in the field, a microseismic experiment was undertaken in 1998. Kristiansen et al. (2000) provide a detailed discussion of the experiment. Six CGG-SST500 three-component instruments were placed in a borehole at 2200–2300 m depth. Because of a restriction in the borehole, all instruments were located above the recorded events, which were at depths of 2370–2390 m. The instruments were operated on a triggering algorithm designed to store only events with a good S/N ratio. During the 57-day monitoring period, 572 events were detected, of which 324 events could be reliably located. There appears to be a high attenuation layer of the low-velocity sequence between the geophones and the reservoir, which means that only nearby or very high-energy events were detected. Dyer et al. (1999) and Zoback and Zinke (2002) found that all the events unrelated to the drilling process were located within a 50-m-thick zone. Their analysis of event locations indicated two dominant structures. Zoback and Zinke (2002) inverted focal mechanisms and found a significant normal-faulting component of the induced events consistent with production-induced stress changes. We do not discuss the absolute or relative locations of microseismic events in this article. For maps showing the locations of events and receivers in the data set, refer to Dyer et al. (1999), Kristiansen et al. (2000), and Zoback and Zinke (2002).

Figure 7 shows a doublet identified in this data set. The seismic traces were low-pass filtered with a Butterworth filter using a cutoff frequency of 30 Hz. The two traces have matching waveforms of the P- and S-waves. To obtain such a close match of P- and S-waveforms, both in P-to-S-wave time and phase, the sources need to be at an almost identical location and have very similar source mechanisms. The near-source heterogeneity and the scattered (indirect) arrivals may vary with small changes in the source position (or mechanism). This is seen in Figure 7 for the phases arriving both between the P- and S-waves and after the S-waves.

Preprocessing the data

The data are preprocessed first by taking a time window to eliminate multiply scattered incoherent waves and parts of the records with noise (such as signal before the P-wave arrival). Then we low-pass filter with various frequency thresholds. The time interval for crosscorrelations is set to contain both P- and S-wave arrivals since their similarity is needed for calculating relative locations. To do this in an automated fashion, we manually estimate the maximum P-to-S-wave traveltime t_{PS} and automatically find the time of arrival of the maximum amplitude (using data from all three components), t_a. Then we set the time window in which we calculate crosscorrelations to be $(t_a - t_{PS}, t_a + t_{PS})$. Even if the maximum amplitude is the P-wave arrival, the correlated signal will contain the S-wave arrival. In any case, such a time interval contains the best S/N ratio of the recording, as the time window is centered on the maximum amplitude.

This study has used the 318 locatable events from the 572 events detected during the Valhall experiment. However, the waveforms of 59 of the 318 events were corrupted, so we removed them from our analysis. Therefore, we only used 259 events for doublet detection. Receiver six was very noisy (probably from poor coupling) and has been omitted from the study. For each event, 10 000 data points were recorded at a sampling frequency of 1000 Hz. The time lag between P- and S-wave arrivals at each receiver is typically around 0.4 s. Figure 8 shows a typical amplitude spectrum of the recorded data. The amplitude spectra of the signal usually peak below 30 Hz. Thus, there is a large amount of redundant data, both in time (long records with no signal) and in frequency (data are oversampled). The time intervals with no signal are comprised of uncorrelated noise, and these would downward bias our peak
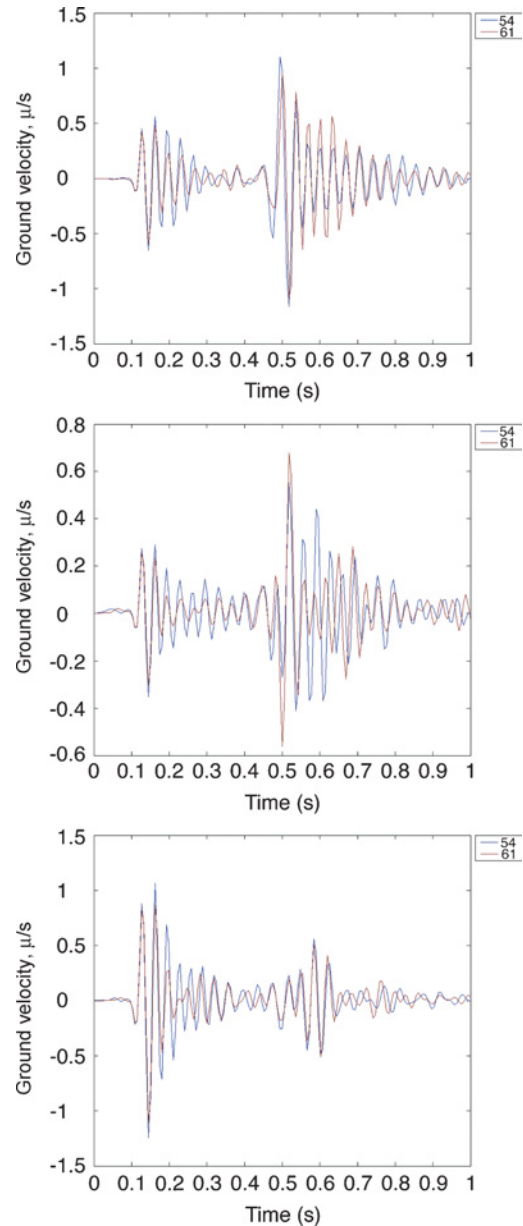

Figure 7. Components of particle velocity for an identified doublet (events 54 and 61). The P-wave arrival is at approximately 0.12 s and the S-wave arrival is at 0.46 s. The x-, y-, z-components have no specific meaning, except that they are mutually orthogonal (coordinates were not rotated). The seismograms were low-pass filtered at 30 Hz.

crosscorrelation coefficients. Analogously, the high-frequency noise above 60 Hz would also decrease the crosscorrelation coefficients between doublets.

To speed up the algorithm and avoid correlating noise, we time window, low-pass filter, and resample the data. We
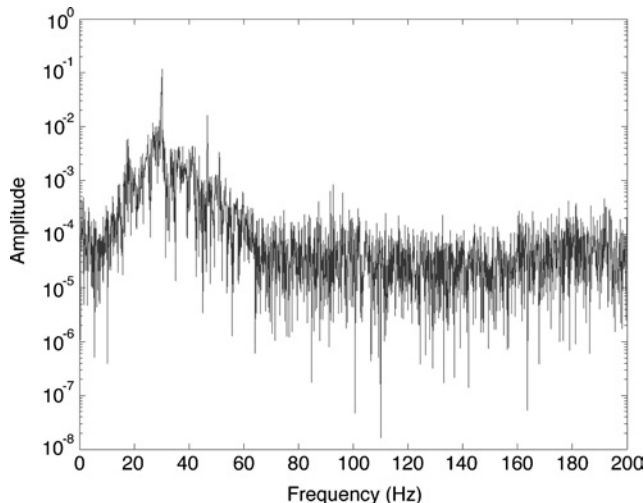

Figure 8. Amplitude spectrum of a typical event recorded at the Valhall data set.
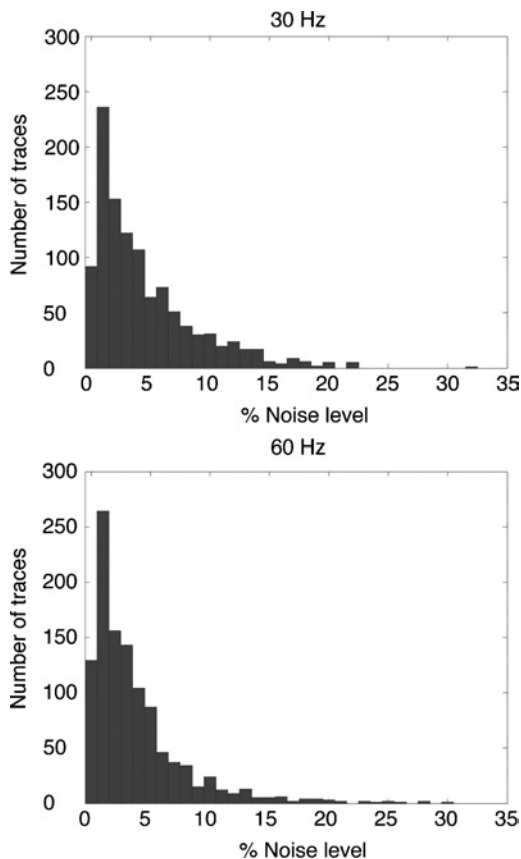

Figure 9. Histograms showing the quantities of traces with different noise levels (measured as a percentage of the maximum amplitudes) for low-pass Butterworth filtering at 30 and 60 Hz.

low-pass filter the data using a Butterworth filter and then search for the maximum waveform amplitude on each receiver. This amplitude is contained typically within the S-wave waveform. The technique then takes a time window of data around the maximum amplitude. A time window of 1 s (0.5 s either side of the maximum) always contains both the P- and S-wave arrivals. Finally, we resample the data to reduce the frequency redundancy after low-pass filtering. The resampling rate is set to have five samples for the shortest period set by the low-pass filter threshold. Preprocessing reduces the number of data points in each trace to 334 when low-pass filtering at 60 Hz and to 167 when low-pass filtering is at 30 Hz.

Results

The synthetic test shows that the doublet detection threshold is dependent upon the level of noise in the data. To evaluate the noise level in the Valhall data set, we filter the traces and compute the ratio of the standard deviation of the noise to the maximum amplitude of each trace to mimic the noise definition of the synthetic data set. The standard deviation of the noise is computed in a time interval before the onset of the P-wave. The noise level depends on the frequency band contained in the data. In general, the narrower the frequency band, the more similar the waveforms and the lower the S/N ratio. To compensate for this when using a lower maximum frequency in the data, we need to increase the detection threshold.

However, for a constant S/N level and threshold, the lower the maximum frequency contained in the data, the larger the number of sources classified as doublets (because we effectively increase the one-fourth wavelength criterion for the distance between two events within a doublet). Therefore, we seek the lowest limit of low-pass filtering that gives us a good S/N ratio in the recorded data. Figure 8 illustrates that the S/N ratio is constant for low-pass filtering above 60 Hz. Furthermore, Figure 9 shows that low-pass filtering at 60 and 30 Hz has a similar effect on the S/N ratios (because most of the signal is below 30 Hz). To show the effect of the low-pass filtering, we identifiy doublets using low-pass filter limits at 60 and 30 Hz. Therefore, for the same threshold of similarity, the algorithm would identify more doublets for data filtered to a maximum frequency of 30 Hz. In this study the analysis of doublets is focused on data that have been low-pass filtered at 30 Hz. Figure 9 shows that there is variable noise, with most traces containing less than 5% noise but some traces as much as 32%. The average noise level is 4%. However, the peak crosscorrelation coefficient between two traces containing different noise levels is dominated by the trace with higher noise level. The choice of 0.8 gives good identification of multiplets in this study based on similarity of the P- and S-waves in identified doublets.

The algorithm, run on the full data set (for 777 traces of the 259 events), took 2–3 hours to run on a single Solaris 8 processor (clock speed of 300 MHz). The number of earthquake doublets identified using low-pass filtering at 60 Hz was 170 with a threshold of 0.8. A threshold of 0.9 would have found 61 earthquake doublets, and a threshold of 0.7 would have identified 439 doublets. Using low-pass filtering at 30 Hz, the algorithm identified 271 doublets with a threshold of 0.8, 75 with a threshold of 0.9, and 842 with a threshold of 0.7. As discussed

earlier, the thresholds of 0.9, and 0.8 consistently identify doublets with similar waveforms of P- and S-waves as needed for relative location. For the rest of this study, we use a threshold of 0.8 to identify the doublets.

Using this threshold, the algorithm identifies 115 out of the 259 events as members of multiplets or doublets when low-pass filtered at 60 Hz, and 130 events when low-pass filtered at 30 Hz. The largest multiplet identified with low-pass filtering at 60 Hz contained 22 events, and with low-pass filtering at 30 Hz, the largest multiplet contained 31 events. The number of absolute locations it is necessary to compute is equals the number of events that are not members of multiplets or doublets plus the total number of multiplets and doublets. The remaining locations can be computed relative to those absolute locations. With low-pass filtering at 60 Hz, there are 28 multiplets and 144 single events ($259 - 115$). Therefore, 172 absolute locations ($144 + 28$) and 87 relative locations ($115 - 28$) would be required. About 33% of the locations could be calculated using a relative location technique. With low-pass filtering at 30 Hz, there are 129 single events and 22 multiplets. Thus, we need 151 absolute locations, i.e., about 60% of the 259 absolute locations that would be required without identification of multiplets.

Figure 10 shows the distribution of the number of multiplets with respect to the number of events in each multiplet with low-pass filtering at 30 Hz. There is a large number of individual doublets (14), and there are five multiplets with 6 to 8 events and two large multiplets with 30 and 31 events each. Low-pass filtered seismograms of one of the medium-sized (seven events) multiplets are shown in Figure 11. The waveforms show a high degree of similarity from trace to trace on all components. The seismograms were aligned on the maximum crosscorrelation of the x-component, but all of the components show excellent alignment. The S-wave has the largest amplitude and dominates the crosscorrelation. The weaker P-wave signals are also similar and show the same polarity for all events of this multiplet. The variation in the P- to S-wave traveltime among the events within this multi-

plet determines the relative distance between events. For this multiplet, we see no variation in the relative arrivals of P-waves within 0.01 s. With this upper limit we can estimate the relative distance of events within the multiplet. In a homogeneous isotropic medium the absolute distance of two events from a receiver is

$$x_a = \frac{(t_s^a - t_p^a)}{\dfrac{1}{\beta} - \dfrac{1}{\alpha}} \text{ and } x_b = \frac{t_s^b - t_p^b}{\dfrac{1}{\beta} - \dfrac{1}{\alpha}}. \tag{4}$$

Here, x_a is the distance to the first event (a), and t_p^a and t_s^a are the arrival times of the P- and S-waves, respectively, for the first event. Analogous notation is used for the second event (b); α and β are the P- and S-wave velocities, respectively. For two events in a multiplet, we assume the variation in the P- to S-wave time is from the relative distance in a homogeneous source region with velocities α and β. Thus, the relative

Figure 11. Three components of the particle velocity for the medium-sized multiplet four recorded on the second receiver (seven events). The numbers on the right side of each trace correspond to the size of the true maximum amplitude (in microns/second) before rescaling to unit size for plotting. The black vertical lines with letters P and S represent approximate P- and S-wave arrival. Seismograms were low-pass filtered at 30 Hz. (a) x-component, (b) y-component, (c) z-component.
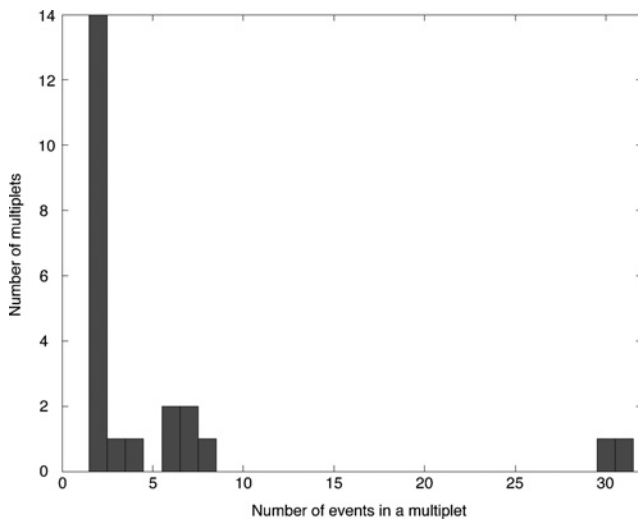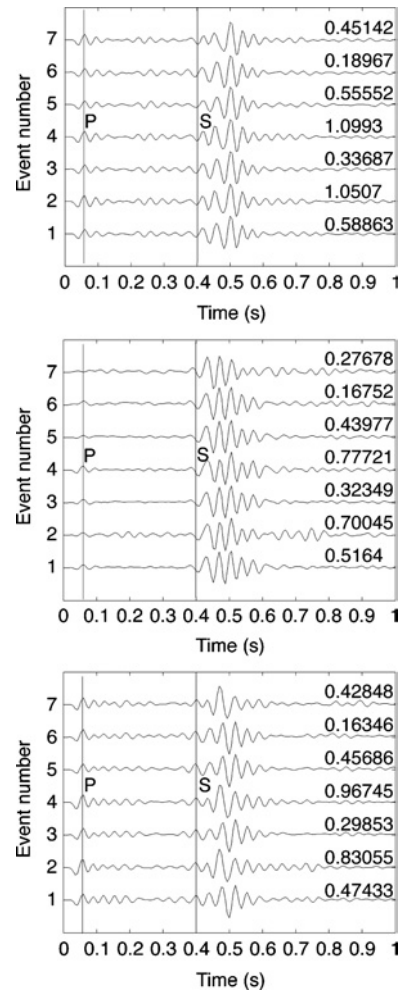

Figure 10. Histograms showing the number of multiplets as a function of the number of events in each multiplet. Low-pass filtering at 30 Hz and a threshold of 0.8 were used for doublet detection.

distance between events a and b is

$$dx = \frac{(t_S^b - t_P^b) - (t_S^a - t_P^a)}{\frac{1}{\beta} - \frac{1}{\alpha}}. \qquad (5)$$

This is a minimum distance. If the sources are perpendicular to raypaths from the receiver array (typical borehole acquisition), the relative distance can be larger because it does not project to the relative P-to-S-wave times but to polarization of the arriving waves. The velocities in the source region of the Valhall data set can be estimated as the velocities of the Balder geologic interval (Dyer et al., 1999): $\alpha = 2100$ m/s and $\beta = 750$ m/s. Therefore, the distances between events within this multiplet (four) are less than 12 m.

If two events are perfect doublets, then their particle motions will be scaled relative to each other; i.e., if the maximum amplitude of the x-component of event a is twice as big as the maximum amplitude of the x-component of event b, then the same should be valid for the y- and z-components. This is approximately true for the multiplet shown in Figure 11 (refer to the numbers on the right of the traces). We can also compare the relative sizes of ruptured areas within the cluster by taking the ratios of observed seismograms. The scaling of seismograms varies up to a factor of five between events four and six. The seismic moment is linearly proportional to recorded amplitudes. Therefore, the seismic moment varies by a factor of five between events four and six. For a constant-stress drop, the source radius is proportional to the third root of the moment (assuming a circular source). Therefore, the source radius varies by a factor of 1.7, and the source area varies by a factor of three.

Figure 12 illustrates the connectivity of the multiplet of Figure 11. Events in this multiplet are mutually similar. The least connected event is event seven, which does not show the same scattering after the S-wave arrival as the other events (see y- and z-components of Figure 11). This event is also the least correlated on several other receivers (Figure 11 represents only the second receiver). For example, for receiver one, the P-to-S traveltime is slightly different from the rest of the multiplet, indicating the event has a greater relative distance from

events one through six. However, the P- and S-waveforms are similar, suggesting it has a source mechanism similar to the other events in the multiplet.

The medium-sized multiplets (three to seven events per multiplet) have similar traces and comply with tests shown for the multiplet of Figure 11. The amplitude variation of events in the multiplets varies from a factor of three to 50. To the best of our knowledge, this is a new observation, as previously published observations in hydrocarbon reservoirs observed multiplets with events of similar size (e.g., Rutledge et al.,

Figure 13. Three-component particle velocity seismograms for the medium-sized multiplet 10 recorded on the first receiver (four events). The numbers on the right side of each trace correspond to the size of the true maximum amtitude (in microns/second) before rescaling to unit size for plotting. Seismograms were low-pass filtered at 30 Hz. (a) x-component, (b) y-component, (c) z-component.
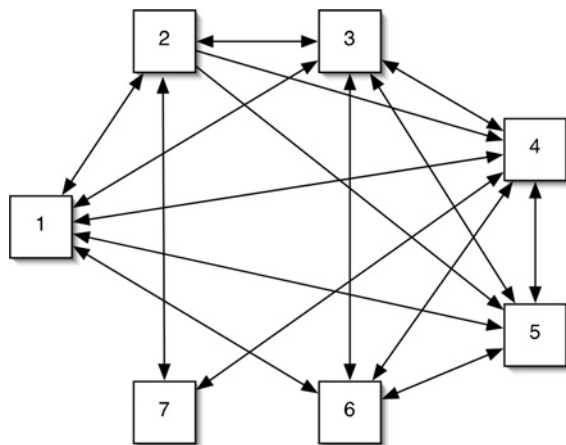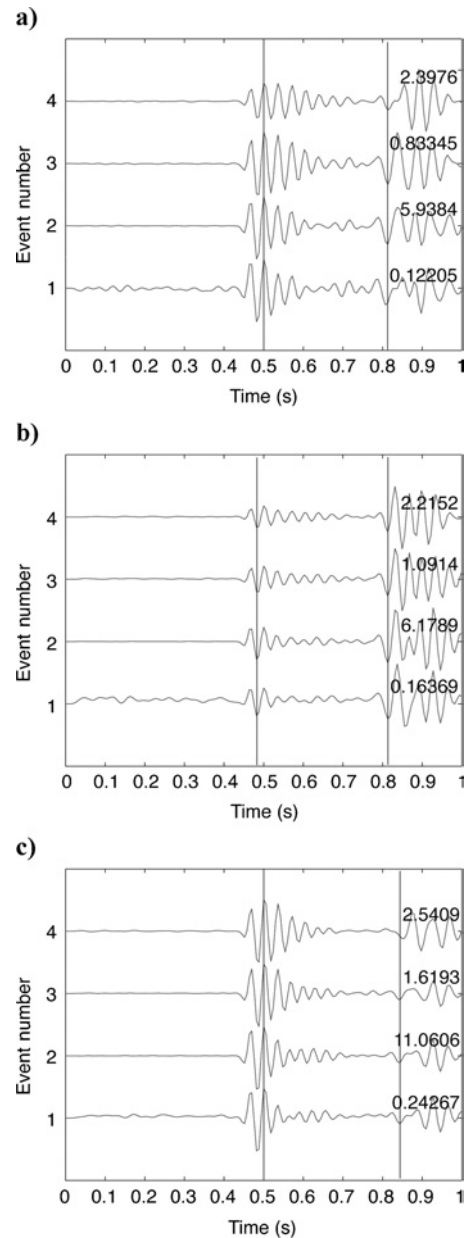
Figure 12. Connectivity of the multiplet shown in Figure 5. Each event is represented by a numbered square, solid lines represent identified doublets.

2004). Figure 13 shows the events of a multiplet where the ratio of the maximum amplitude of event 2 to event 1 is 56, 30, and 43 on the *x*-, *y*-, and *z*-components, respectively. The time window for this multiplet is centered on the P-wave arrival because it has the largest amplitude for this receiver geometry and source mechanism.

Figure 14 shows the *y*-component recordings of a medium-sized multiplet (six events). The enlarged P-wave arrivals in Figure 14b show the variation in P-wave arrival times (e.g., events 4 and 6). The inconsistent P-waveforms for events 1, 2, and 3 are caused by noise (compare the maximum amplitudes of events 1, 2, and 3 with events 4, 5, and 6). The plots on Figure 14a show all events aligned on the S-wave arrival; therefore, the variation in the P-wave arrival leads to the variation in the P-to-S-wave traveltime. The time delay is 0.05 s shorter for events 3 and 4 than for event 6. This variation implies that event 6 is about 100 m away from events 3 and 4, yet they still can be located relative to each other because of the similarity of the P- and S-waveforms. This distance is an order of magnitude larger distance than one-fourth of a wavelength separation at the source region (estimated by Geller and Mueller, 1980), indicating this criterion is rather conservative.

The two largest multiplets, containing 31 and 30 events, respectively, were the most complicated multiplets among those identified. The structure of the 30-event multiplet is illustrated in Figure 15. Twenty-one highly similar events are within 20 m of each other (their P- and S-waves are aligned within 0.01 s). The remaining nine events are similar to only some of the events in this cluster. The variation of maximum amplitude ratios within this multiplet is up to a factor of 40. A possible interpretation of this multiplet is a repeated (21 times) rupture of a main fault with nine ruptures branching off this main fault (but still along the same fault plane, as the source mechanisms are similar). The P-to-S traveltime varies for events 22–30 up to 0.07 s with respect to events one through 21, indicating that the branching ruptures reached up to 140 m away from the main fault. The 31-event multiplet is more complex and consists of two large similar groups containing 10 and 11 events each; these two submultiplets are connected by only one event. There are also side branches of the two main groups, similar to the branches in the 30-event multiplet. The two main groups in the 31-event multiplet can be interpreted as two nearby, parallel, active faults.

DISCUSSION

Why do we observe such a large number of multiplets in the microseismic events recorded during a passive seismic experiment? According to Dyer et al. (1999), these events reflect subsidence of the overburden as a result of oil extraction (and decreased reservoir pressure). We can expect a large number of events to be located on a few faults because the induced events reflect quasi-static changes in the reservoir pressure, not fluid or stress migration. Thus, our observation of a large number of multiplets is consistent with blocks subsiding
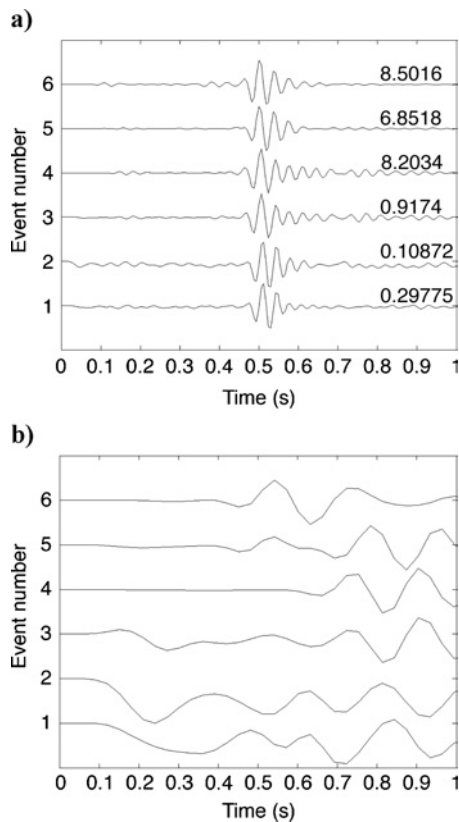

Figure 14. The component recordings of the particle velocity for the medium-sized multiplet 9 recorded at the third receiver. The numbers on the right side of each trace (a) correspond to the size of the true maximum amplitude (in microns/second) before rescaling to unit size for plotting. Seismograms were low-pass filtered at 30 Hz. (b) Enlarged view, showing the arrival of the P-waves.
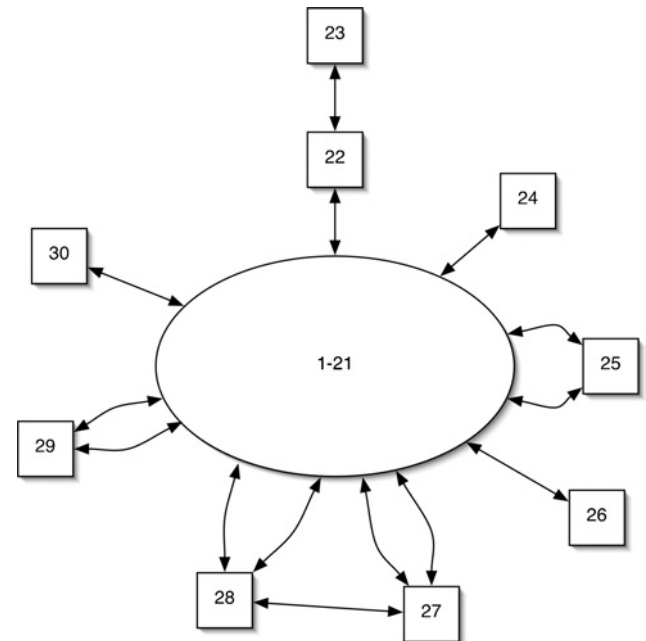

Figure 15. Illustration of the connectivity of multiplet 6. Events 1–21 are very well connected; each event has a minimum of three connections. Events 22–30 are doublets with only 1 or 2 events from the cluster of events 1–21.

along only a few seismically active faults, analogous to plate tectonics on a different scale. The two large multiplets probably reflect the core of the two structures found by Dyer et al. (1999). These two structures account for 61 out of the 259 events tested, or about one-fourth of the seismic activity.

CONCLUSIONS

The algorithm described in this study provides a fast, automatic procedure to identify multiplets in microseismic data sets. Waveform similarity, which is measured using peak cross-correlation coefficients, provides a means of identifying pairs of events with similar source mechanisms and hypocenter locations (doublets). These doublets are grouped into multiplets using graph theory. Using a synthetic data set evaluated for a 3D model, doublets are identified by the peak crosscorrelation coefficient threshold. Noise in the data lowers the threshold of doublet identification. This technique is applicable to events with up to a 20% noise level.

We applied the multiplet detection algorithm to the Valhall passive seismic data set. The algorithm detected a large number of similar events, implying that the number of absolute locations required can be reduced by at least 40% using a relative location algorithm. Most of the similar events are located in seven clusters with spatial dimensions of less than 20 m. Two major active faults are rerupturing more than 20 times and account for one-fourth of the detected seismicity.

ACKNOWLEDGMENTS

REFERENCES

Barkved, O., P. Heavey, R. Kjelstadli, T. Kleppan, and T. G. Kristiansen, 2003, Valhall field — Still on plateau after 20 years of production: Offshore Europe Conference, Society of Petroleum Engineers, Proceedings, 83957.

Cattaneo, M., P. Augliera, D. Spallarossa, and V. Lanza, 1999, A waveform similarity approach to investigate seismicity patterns: Natural Hazards, **19**, 123–138.

Dyer, B. C., R. H. Jones, J. F. Cowles, O. Barkved, and P. G. Folstad, 1999, Microseismic survey of a North Sea reservoir: World Oil, **220**, no. 3, 74–78.

Fréchet, J., L. Martel, L. Nikolla, and G. Poupinet, 1989, Application of the cross-spectral moving-window technique (CSMWT) to the seismic monitoring of forced fluid migration in a rock mass: International Journal of Rock Mechanics Mineral Science Geomechanics Abstracts, **26**, 221–233.

Geller, R. J., and C. S. Mueller, 1980, Four similar earthquakes in central California: Geophysical Research Letters, **7**, 821–824.

Ifeachor, E. C., and B. W. Jervis, 1993, Digital signal processing — A practical approach: Addison-Wesley Publ. Co.

Kristiansen, T. G., O. Barkved, and P. D. Pattillo, 2000, Use of passive seismic monitoring in well and casing design in the compacting and subsiding Valhall field, North Sea: European Petroleum Conference, Society of Petroleum Engineers, Proceedings, 65134.

Lees, J. M., 1998, Multiplet analyses at Coso geothermal: Bulletin of the Seismological Society of America, **88**, 1127–1143.

Li, Y., C. Cheng, and M. Toksoz, 1998, Seismic monitoring of the growth of a hydraulic fracture zone at Fenton Hill, New Mexico: Geophysics, **63**, 120–131.

Maurer, H., and N. Deichmann, 1995, Microseismic cluster detection based on waveform similarities, with an application to western Alps: Geophysical Journal International, **123**, 588–600.

Moriya, H., K. Nagano, and H. Niitsuma, 1994, Precise source location of AE doublets by spectral matrix analysis of the triaxial hodogram: Geophysics, **59**, 36–45.

Moriya, H., H. Niitsuma, and R. Baria, 2003, Multiplet-clustering analyses reveals structural details within the seismic cloud at the Soultz geothermal field, France: Bulletin of the Seismological Society of America, **93**, 1606–1620.

Nadeau, R. M., W. Foxall, and T. V. McEvilly, 1995, Clustering and periodic recurrence of microseismicities on the San Andreas fault at Parkfield, California: Science, **267**, 503–507.

Poupinet, G., W. L. Ellsworth, and J. Frechet, 1984, Monitoring velocity variations in the crust using earthquake doublets: An application to the Calaveras fault, California: Journal of Geophysical Research, **89**, 5719–5731.

Rutledge, J. T., and W. S. Phillips, 2003, Hydraulic stimulation of natural fractures as revealed by induced microearthquakes, Carthage Cotton Valley gas field, East Texas: Geophysics, **68**, 441–452.

Rutledge, J. T., W. S. Phillips, and M. J. Mayerhofer, 2004, Faulting induced by forced injection and fluid flow forced by faulting: An interpretation of hydraulic-fracture microseismicity, Carthage Cotton Valley gas field: Bulletin of the Seismological Society of America, **94**, 1817–1830.

Schaff, D. P., and P. G. Richards, 2004, Repeating seismic events in China: Science, **303**, 1176–1178.

Stich, D., G. Alguacil, and J. Morales, 2001, The relative locations of multiplets in the vicinity of the western Almeria (southern Spain) earthquake series of 1993–1994: Geophysical Journal International, **146**, 801–812.

Waldhauser, F. M., and W. L. Ellsworth, 2000, A double-difference earthquake location algorithm: Method and application to the northern Hayward fault, California: Bulletin of the Seismic Society of America, **90**, 1353–1368.

Weiss, M. A. 1994, Data structures and algorithm analyses: Cumblings Publishing Company.

Zoback, M. D., and J. C. Zinke, 2002, Production-induced normal faulting in the Valhall and Ekofisk oil fields: Pageoph, **159**, 403–420.