3-1-2013

Clustering Revisited: A Spectral Analysis of Microseismic Events

Deborah Fagan
Boise State University

Kasper van Wijk
Boise State University

James Rutledge
Schlumberger Limited

GEOPHYSICS, VOL. 78, NO. 2 (MARCH-APRIL 2013); P. KS41–KS49, 14 FIGS.
10.1190/GEO2012-0323.1

Clustering revisited: A spectral analysis of microseismic events

Deborah Fagan[1], Kasper van Wijk[2], and James Rutledge[3]

ABSTRACT

Identifying individual subsurface faults in a larger fault system is important to characterize and understand the relationship between microseismicity and subsurface processes. This information can potentially help drive reservoir management and mitigate the risks of natural or induced seismicity. We have evaluated a method of statistically clustering power spectra from microseismic events associated with an enhanced oil recovery operation in southeast Utah. Specifically, we were able to provide a clear distinction within a set of events originally designated in the time domain as a single cluster and to identify evidence of en echelon faulting. Subtle time-domain differences between events were accentuated in the frequency domain. Power spectra based on the Fourier transform of the time-domain autocorrelation function were used, as this formulation results in statistically independent intensities and is supported by a full body of statistical theory upon which decision frameworks can be developed.

INTRODUCTION

Monitoring induced seismicity provides valuable information regarding fluid transport properties in hydrocarbon (Maxwell and Urbancic, 2001) and geothermal (Saar and Manga, 2003) reservoirs, geomechanical effects in carbon sequestration (Streit et al., 2005), and hydraulic properties in aquifer recharge and recovery. In energy extraction operations, identifying, through microseismicity, individual faults, fault systems, and associated structures influence the location and depth of production wells to ensure that the correct fluid conducting faults are tapped. Similarly, the location of microseismicity is important for injection wells to ensure that locations, depths, and injection rates will generate fractures in desired fluid-bearing layers.

Waveform correlations in the time domain are used to detect low-magnitude and microseismic events (Schaff and Beroza, 2004; Schaff and Richards, 2004, 2011), refine phase arrivals and event hypocenters (Rowe et al., 2002; Snieder and Vrijlandt, 2005; Hansen et al., 2006; Song et al., 2010), determine event focal mechanisms (Hansen et al., 2006), and to classify unknown source events (Harris, 1991; Shumway, 2003; Shumway and Stoffer, 2006). In this paper, we explore correlations in the power spectra of events originally clustered based on a high correlation in the time domain. A permanent vertical borehole array recorded over 3800 microseismic events in 2008 and 2009 at the Aneth field, located in southeast Utah. Differences in the power spectra of microseismic events from a cluster are small, but consistent in terms of their spatial distribution. As a result, we suggest that these events should be subclustered before further location analysis with more advanced techniques, such as double-difference tomography (Waldhauser and Ellsworth, 2000; Zhang and Thurber, 2003).

Geologic background and production history

The Aneth unit is one of four oil-producing units located within the Greater Aneth oil field of the Paradox Basin in southeast Utah (see Figure 1). The lower portion of a stratigraphic column for the Greater Aneth field from Hintze and Kowallis (2009) is shown in Figure 2. Of interest here are the Pennsylvanian and Mississippian subperiods, from which oil and gas are extracted, into which CO_2 is injected, and where microseismicity occurs. The formations of interest, from oldest to youngest, are the Mississippian Leadville (two strata: lower dolostone and upper lime mudstone to peloidal to crinoidal lime wackestone), Pennsylvanian Molas (poorly stratified silt and sandstones), Pinkerton Trail (alternating thin beds of mudstone and shales interbedded with limestone), and Paradox (cyclic intervals of dolostone, shale, and salt). Within the Aneth unit, the Paradox Formation has been informally divided into

Manuscript received by the Editor 14 August 2012; revised manuscript received 26 October 2012; published online 1 March 2013.
[1]Boise State University, Department of Geosciences, Physical Acoustics Laboratory, Boise, Idaho, USA. E-mail: mtfishergirl2001@yahoo.com.
[2]Boise State University, Department of Geosciences, Physical Acoustics Laboratory, Boise, Idaho, USA and University of Auckland, Department of Physics, Auckland, New Zealand. E-mail: kaspervanwijk@boisestate.edu.
[3]Presently Schlumberger Limited, Microseismic Services, Houston, Texas, USA; formerly Los Alamos National Laboratory, Los Alamos, New Mexico, USA. E-mail: rutledge@swcp.com.

"production zones," from oldest to youngest, named Alkali Gulch, Barker Creek, Akah, Desert Creek, and Ismay. Oil is produced from the Desert Creek and Ismay zones; the lowest three are referred to as "Paradox salts" (Carney, 2010).

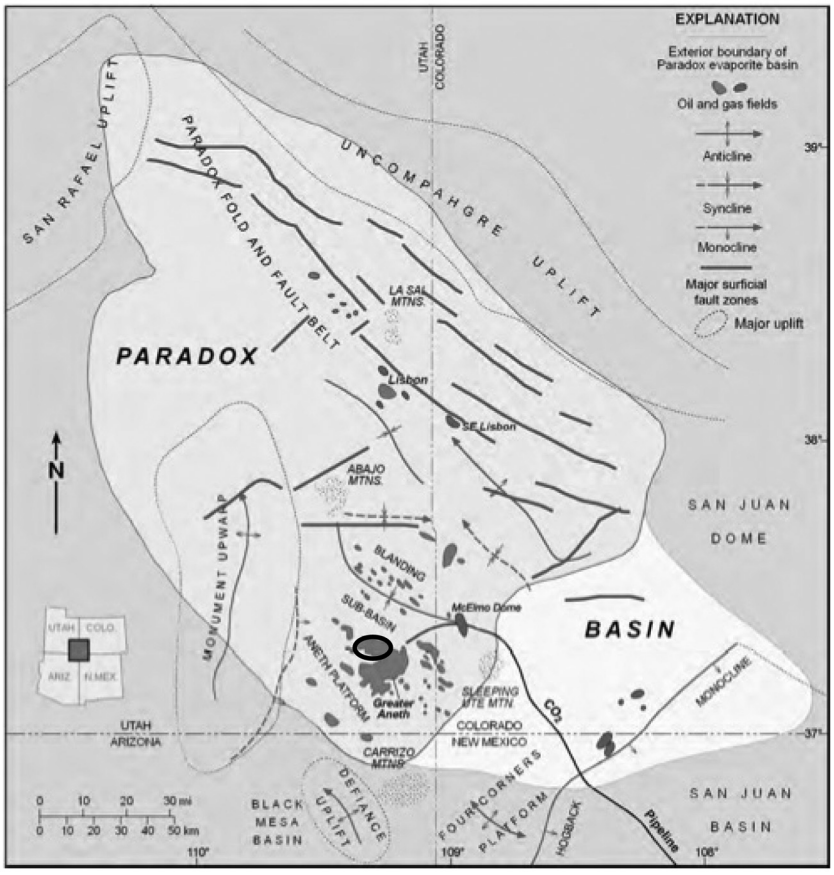

Figure 1. Paradox Basin oil and gas fields (modified from Chidsey, 2009). The study area is in the Greater Aneth and marked by the thick black oval.

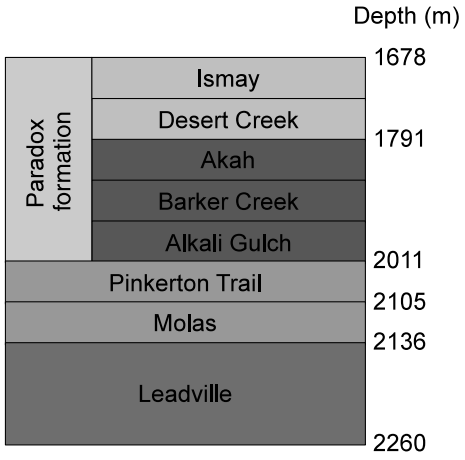

Figure 2. The lower portion of a stratigraphic column for Aneth (based on Hintze and Kowallis, 2009). Oil is produced from the Isamy and Desert Creek zones within the Paradox Formation. The lower zones of the Paradox Formation (Akah, Barker Creek, and Alkali Gulch) are referred to as the Paradox salts, as seen in Figure 4.

Chidsey (2009) provides a detailed discussion of the surface and subsurface geology and production history of Aneth; the relevant details are summarized here. Structural orientation within the Aneth unit is consistent with that of the Paradox Basin and Greater Aneth (Chidsey, 2009; Carney, 2010). Figure 1 shows the northwest–southeast-oriented fold and fault belt along the northeast portion of the Paradox Basin. Surface deformation bands within Greater Aneth indicate a northwest–southeast trend. Similarly, structure contour maps of Paradox Formation strata identify the Aneth area as a northwest–southeast contour orientation. The 3D seismic studies of Greater Aneth show basement faults of Mississippian and Pennsylvanian age that strike northwest–southeast (Rutledge and Soma, 2010).

This field was discovered in 1956, with water-enhanced oil recovery commencing in 1961. The process of converting water injection wells to gas began in 2008 and coincided with the installation of a permanent, 60-level, downhole geophone array in a former injection well on the western side of the field (Rutledge and Soma, 2010). Geophones were set between 800 and 1700 m at 15-m spacing. With a sampling rate of 1500 Hz, the deepest eighteen geophones are 3C; the other sensors record the vertical component only. The 3C geophones have a right-hand, 3D coordinate system so that a Z signal is positive for a compression arrival from above. The 60-sonde array was built by VCable LLC and consists of OYO Geospace Corporation GS-14-L3 geophones with 24–3000-Hz bandwidth and 290-mV/ips sensitivity.

With the conversion to gas injection, a deep brine injection well was installed to dispose of excess water. The well was drilled approximately 1 km northwest of the geophone array, into Leadville Formation limestone to a depth of approximately 2240 m. The well has four lateral arms, and injection rates varied between 12,000 and 20,000 m^3 per week between March 2008 and March 2009 (Chidsey, 2009).

MICROSEISMIC EVENTS

Passive seismic data were collected at the geophone array between March 2008 and March 2009. Of the 3800 events recorded, P- and S-wave arrival times were obtained for 1212 events at 24 geophones. Events were organized into two clusters on opposite sides of the reservoir; the northeast side includes 44 events, the southwest side, 1168 events, which is the focus of discussion and analysis here. Events in the southwest cluster trend along a northwest–southeast fracture zone (consistent with the geologic structure of the area) at least 1500 m long within a depth range of 1830 to 2037 m, which places them in either the Paradox salt layer or the Pinkerton Trail Formation (Figure 2). Figure 3 shows the southwest cluster of events along with the location of the geophone array. The asterisks in Figure 3 identify the location of the 67 events that belong to multiplet 18, discussed in the next section.

On the 16 shallower receivers, the P- and S-wave phases tend to be impulsive and easy to identify. Events were initially screened by manually picking P-wave arrivals on waveforms from just a few receivers. Using a correlation threshold of 0.8 in the time domain, these waveforms were then clustered and repicked using the methods of Rowe et al. (2002), which resulted in 47 sets of events with similar wavefields called *multiplets*, using a threshold correlation coefficient of 0.8.

The velocity model used for location was estimated from a 100-ft (30-m) median smooth of a sonic log obtained from the deep borehole prior to commencing saltwater injection. Figure 4 shows the sonic log, velocity model, and geophone depths for the receivers used to estimate event locations.

Rutledge and Soma (2010) apply a master-slave location scheme to tie the weaker events to the first arrivals of the stronger events. High signal-to-noise ratio master events were formed by stacking multiplets or were chosen from a single large-magnitude event from a multiplet group so that the true first arrivals could be reliably picked. For each master event, the source locations (radial and depth positions with respect to the vertical receiver array) were determined using an iterative least-squares method to best fit the P- and S-wave arrival times (e.g., Rutledge and Phillips, 2003) plus reflected SV phases (similar to Phillips et al., 1998). The traveltime residuals for the master event locations were then applied as station corrections to the remaining events, while omitting the use of any arrivals from the lower eight stations of the array.

Azimuths to the sources are determined from the horizontal-component P-wave particle-motion trajectories (Rutledge and Soma, 2010). The particle-motion data from the array were highly linear and consistent between receiver levels. Azimuths were obtained by averaging over multiple receiver levels. Relative azimuth error was about 1°, measured as the standard error of the mean azimuth (Rutledge and Soma, 2010). Errors in estimating azimuths obtained from the particle-motion data, though on average small, result in the largest component of the relative location error when translated over the large receiver-source distances. Relative depth errors average 3.35 m, and relative radial errors average 7.62 m (Rutledge and Soma, 2010).

An Mw 3.7 earthquake (the Bluff earthquake) struck the region on 6 June 2008. It was located approximately 13 km northwest of the geophone array at a depth of 9.5 km. This earthquake was used to verify the azimuths to the events shown in Figure 3. The moveout of first arrivals for the earthquake clearly indicates an upward-propagating arrival across the array, and given the polarity of the vertical components on the array, indicates dilatational first motion. Similarly, for the microseismic events, which all occur below the bottom of the array, the vertical component at a subset of geophones show dilatational first motion. Horizontal-component first motions are all opposite of the Bluff earthquake, putting the microseismic events in the southeast quadrant with respect to the array.

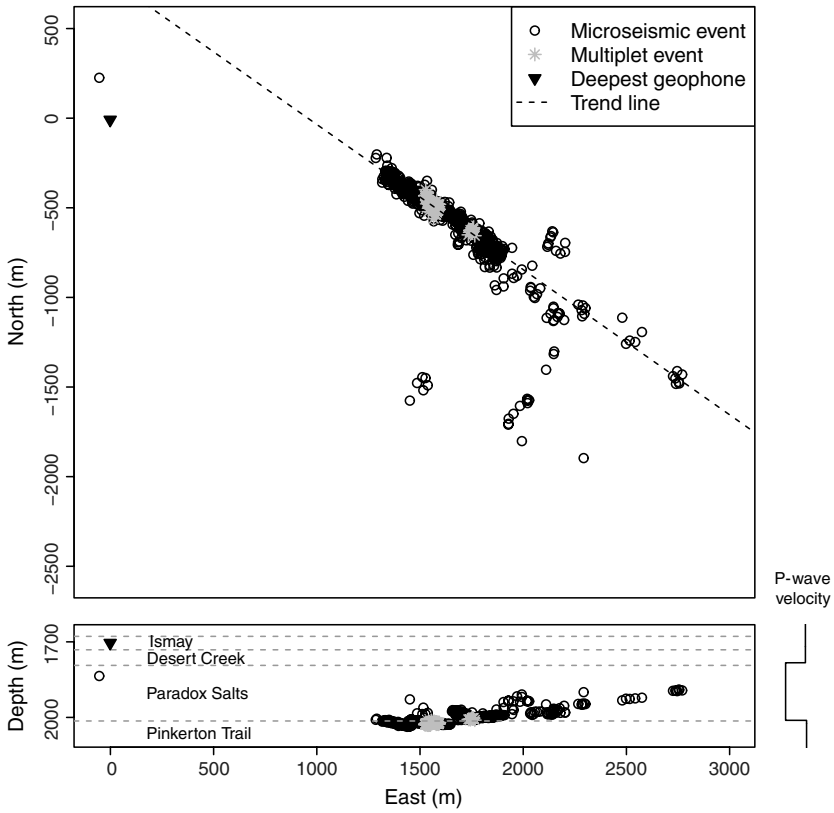

Figure 3. The 1168 events located on the southwest boundary of the reservoir follow a northwest–southeast orientation. The 67 events shown as asterisks belong to multiplet 18 and separate into two spatial groupings.

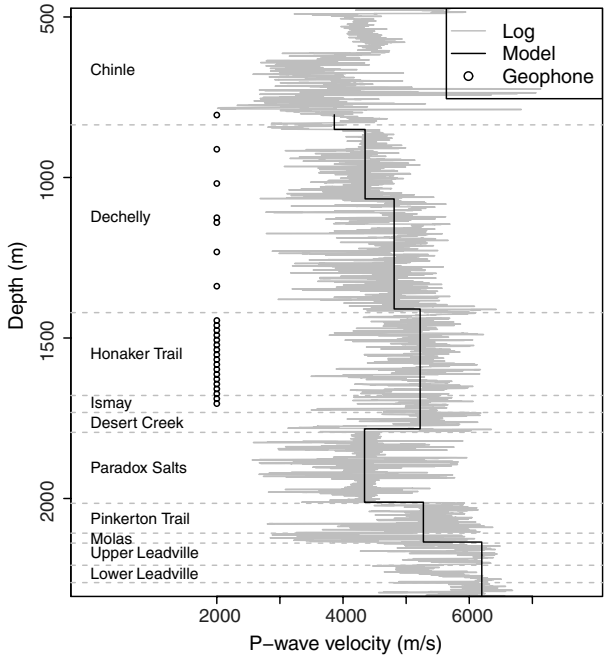

Figure 4. Sonic log, P-wave velocity model, and geophone depths used to estimate hypocenters. The S-wave velocity is assumed to be $V_P/1.625$ m/s. Paradox salts correspond to the Akah, Barker Creek, and Alkali Gulch zones of the Paradox Formation in Figure 2.

Rutledge and Soma (2010) study the space-time development of microearthquake occurrence and the correlation between seismicity and injection/production activities. Using the Gutenburg-Richter law $\log(N) = a - bM$, where N is the number of events having magnitude that exceeds M, Rutledge and Soma (2010) fit a value $b = 2.0$. Tectonic seismicity typically results in $b \approx 1$. The high b value indicates a higher proportion of small events to large ones, characteristic of fluid-injection induced seismicity or natural swarm seismicity. Natural swarm seismicity typically precedes volcanic eruption or is associated with migration of crustal fluids through fracture networks (Lay and Wallace, 1995). The population of earthquakes dominated by small events is generally considered an indication of many small discontinuous faults accommodating strain accumulation.

In a temporal analysis of events, Rutledge and Soma (2010) observe that higher magnitude events occur near gaps in seismicity, indicating that the structure is composed of discrete segments, such as en echelon jogs where stress would tend to concentrate near discontinuities. Up to this point, further inferences regarding subsurface structure are limited, however, because of the single-well-array geometry and large source-array differences.

Based on the distance between the saltwater injection well and microseismicity (1200 m), no discernible correlation between injection rates and seismicity, and the long history of injection at Aneth, Rutledge and Soma (2010) conclude that events occurring within the study period were not a direct result of deep saltwater injection into the Leadville or CO_2 injection into the Desert Creek zone, but rather a stress release related to the overall reservoir reduction over the production life of the area.

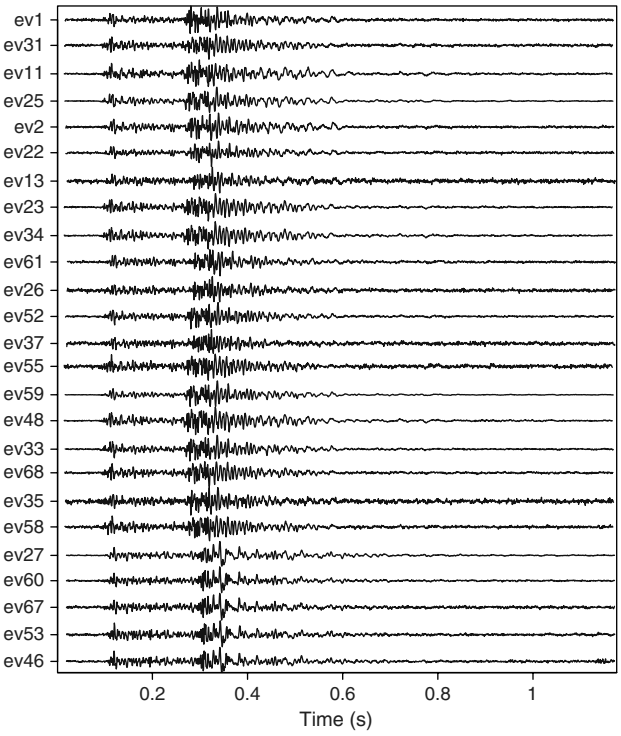

Figure 5. Waveforms from multiplet 18 from geophone 1, the deepest receiver, aligned by the lag of the maximum correlation coefficient around the maximum amplitude. Each of the 25 waveforms starts 0.1 s prior to the P-wave arrival and is normalized by its maximum amplitude.

MULTIPLET ANALYSIS

Of the 47 multiplets located on the southwest boundary of the field, multiplet 18 was chosen to research because of its highly correlated waveforms that cover the 12-month study period. This multiplet also has numerous events with high-confidence picks used for multiplet identification and location. Multiplet 18 includes 73 events; however, the number of P-wave picks varies from receiver to receiver along the array (as well as from component to component). For instance, 70 picks are available from the east-component data for the geophone at 1552 m (the 11th geophone from the bottom), whereas 67 picks are available from the deepest receiver (geophone 1).

We investigated statistical clustering of multiplet 18 events from geophone 1. Data from this receiver were excluded in the location estimation by Rutledge and Soma (2010) because of complications likely stemming from guided-wave arrivals. Although this was an issue with the location algorithm used (guided-wave), data from this geophone may contain information on the relatively uncharacterized volume underlying the production reservoirs.

Rotated east-component waveforms from the deepest geophone for 25 events from multiplet 18 are shown in Figure 5. Peak-normalized waveforms starting 0.1 s prior to the P-wave arrival are aligned according to the maximum correlation in the first 0.25 s, similar to the process Rowe et al. (2002) use to create the mulitplets. At geophone 11 (1552-m depth), waveform correlation coefficients for multiplet 18 events exceed 0.90. Crosscorrelation coefficients at geophone 1 (1704 m) range between 0.68 and 0.98. Rutledge and Soma (2010) suggest that raypaths for these events to the deepest eight geophones — deeper than 1590 m — include critically refracted paths in the Leadville Formation. Defined by their strong wavefield correlation in the time domain, these events belong to the same multiplet 18 and are located using a single master event.

Spectral correlations

The spectra of the events in multiplet 18 were clustered via hierarchical agglomerative clustering methods using spectral distance as the dissimilarity metric; see Appendix A for details, but in essence the spectral distance is defined as the Euclidean distance using the normalized spectral power by frequency. Each event starts out as its own cluster, with similar clusters being iteratively combined until only a single cluster remains. The final number of clusters chosen is based on a dendrogram. Because it works on a distance matrix, the algorithm is robust to the order of the events. It is a widely used algorithm, successfully used in other seismic applications (Rowe et al., 2002; Bardainne et al., 2006).

Using Fourier frequencies in the range of 0 to 400 Hz, event spectra organize spatially, consistent with location. The cluster dendrogram is shown in Figure 6, with rectangles corresponding to cluster membership. The height on the y-axis represents dissimilarity: A higher horizontal agglomeration line indicates greater spectral dissimilarity than a lower one. To study the spectral characteristics that affect clustering, it is useful to examine events corresponding to individual agglomerations in the dendrogram of Figure 6. Events in group 3 in Figure 6 are agglomerated early in the algorithm, but they only join the rest of the events at the last agglomeration step. Figure 7 shows that group 3 events are geographically separated from the rest of the events in the multiplet. Examination of the mean spectra in Figure 8 identifies the same range of contributing

frequencies, but intensities differ markedly, lending insight to how the clustering algorithm uses spectral distance to differentiate groups of events.

The next branch in the dendrogram in Figure 6 reveals additional geographic organization of events. All events disjointed from group 3 subdivide into a western and eastern cluster, as seen in Figure 9. Figure 10 illustrates that the shapes of the western and eastern clusters mean that spectra are highly similar, but a frequency shift appears to be the factor in their cluster membership. Figure 11 shows that the maximum crosscorrelation between these two spectra occurs at an ~2-Hz lag. Combining this information with the clear spatial correlation within a subcluster leads us to conclude that the lag at maximum correlation is small but significant. Figure 12 shows individual events crosscorrelated with the western cluster mean. Every event in the western cluster maximizes the crosscorrelation with its mean at zero lag, and all eastern events show a positive lag. The eastern cluster events show positive lags at up to 3 Hz, consistent with the ~2-Hz lag indicated in the crosscorrelation of the means.

Finally, the western cluster appears to subdivide in groups 1, 2, and 5 (Figure 6), although the eastern cluster is made up of groups 4 and 6. As with many exploratory applications of statistical clustering, determining the final number of clusters is somewhat subjective, even with the calculation of additional statistical parameters, as in Rowe et al. (2002). Hard rules can be established in discriminant analysis applications when phenomenology provides a definitive basis. However, in this application, a subdivision into six clusters resulted in mean spectral differences that were spatially consistent within each subcluster, as shown in Figure 13. Maximum crosscorrelations between individual events and the westernmost cluster mean (group 5) are shown in Figure 14. Frequency shifts are more complicated, but they generally follow the trend observed in Figure 10. Although lags show more variability within clusters, the geographical organization of the clusters is a compelling reason

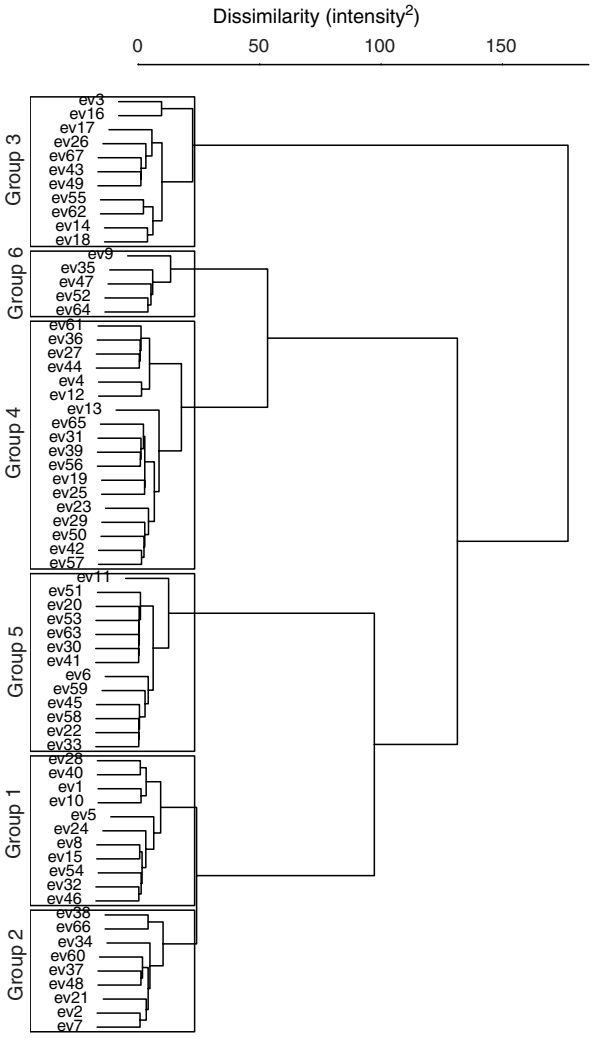

Figure 6. Dendrogram for the spectra of multiplet 18 events from geophone 1, with rectangles identifying cluster membership. The longest branch of the dendrogram indicates that the spectra of group 3 members are most disparate; the shortest branch shows that groups 1 and 2 event spectra are the most similar.

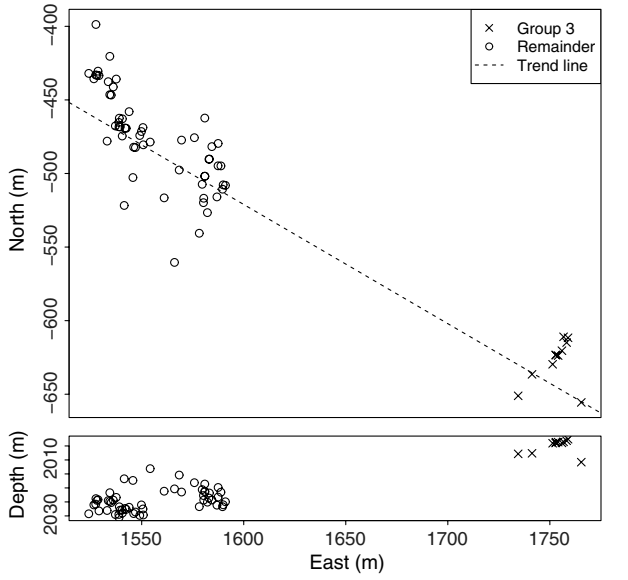

Figure 7. Spectral clustering results when two clusters are identified. Group 3 corresponds to the events in the leftmost rectangle on Figure 6 and form a geographically separate region.

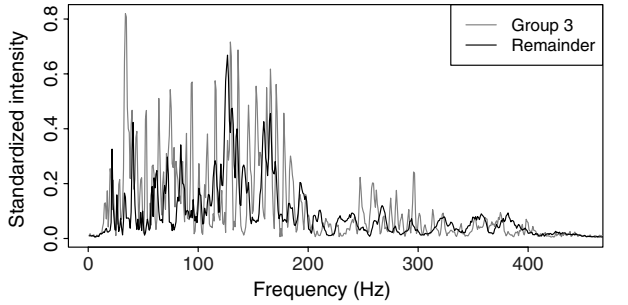

Figure 8. Mean peak normalized power spectra for the two clusters of events identified in Figure 7. The shapes are markedly different, providing insight to the spectral clustering process.

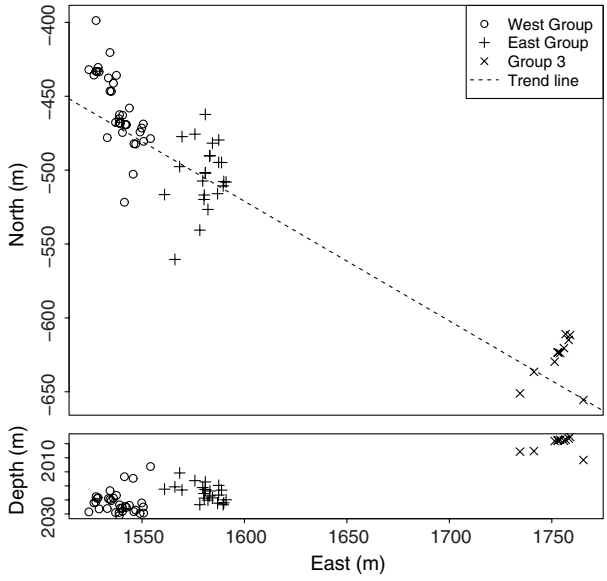

Figure 9. Results for spectral clustering with three clusters; the group of circles in Figure 7 has been subdivided into western and eastern clusters based on spectral clustering.

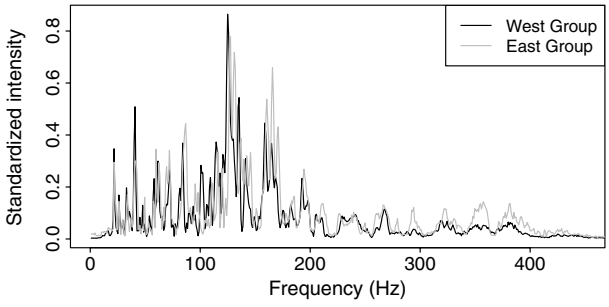

Figure 10. Mean normalized power spectra for the clusters identified in Figure 9. Peak locations hint at a subtle spectral shift between western and eastern groups.

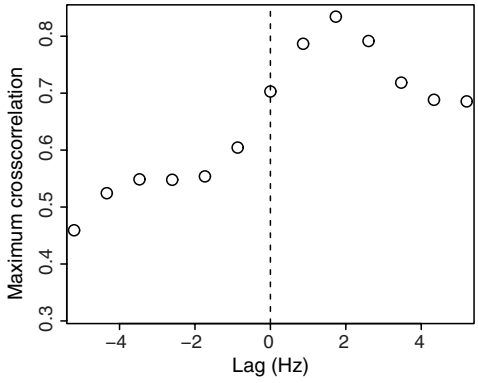

Figure 11. Crosscorrelation for the mean spectra of Figure 10 shows a maximum at 2 Hz.

to consider these as legitimate subclusters. In the next section, we present geologic arguments for the significance of these subclusters.

DISCUSSION

Although we observe spectral differences within a cluster, there are no other particularly distinguishing features that differentiate these clustered events. A time-domain analysis of the same data in Fagan (2012) does not reveal consistent agglomeration of events with the resolution observed in the frequency domain. This result may seem surprising, given that the time- and frequency-domain analyses are based on the same data. However, it may be that a focus on the power spectra (ignoring phase information) makes small temporal differences easier to identify.

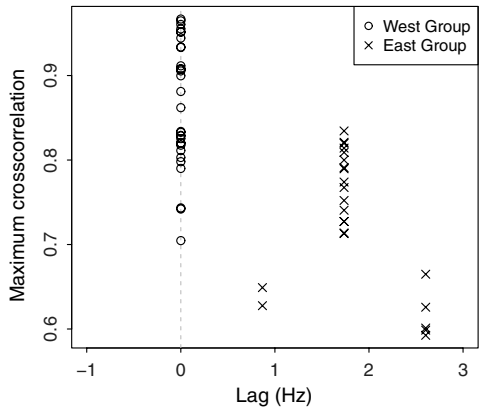

Figure 12. Maximum crosscorrelations between the western cluster mean and individual events, coded by cluster membership. Positive lags to approximately 3 Hz support the mean frequency shift indicated in Figure 11.

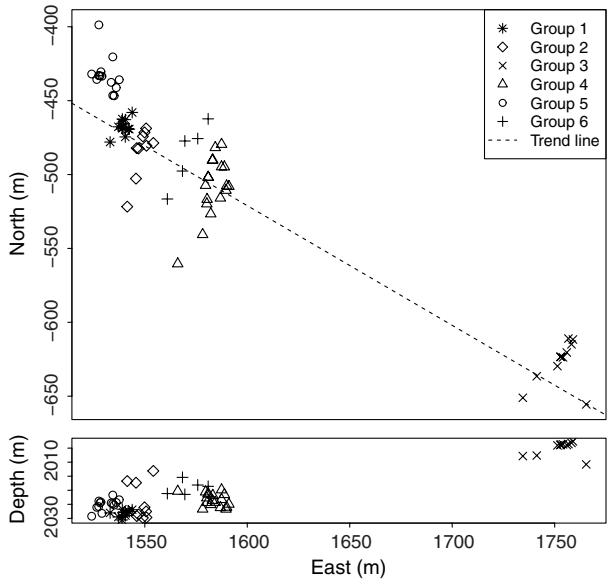

Figure 13. Further subclustering leads to event locations coded by spectral cluster membership of Figure 6. The identified subclusters align into what appear to be individual en echelon fractures.

Differentiation within the cluster with spectral intensities in Figure 8 can be attributed to source and/or path effects. These events are nearly colocated relative to the sensor array (with the exception of group 3), thereby minimizing path effects to the extent possible. However, there is sufficient ambiguity in subsurface structure at the events to preclude ruling out a path effect yet. Sensor locations are not sufficient to determine the orientation of events or focal mechanism, but the analysis of Rutledge and Soma (2010) indicates nontectonic seismicity. If attenuation was a factor in the spectral separation of these events, one would expect the opposite of what is observed: Events farther from the receiver would have less spectral energy in the higher frequencies than events farther afield.

After a more detailed subclustering in Figure 13, we can distinguish the individual en echelon faults that are initially hypothesized by Rutledge and Soma (2010), based upon "gaps" in the seismicity. Uncertainty ellipses are oriented southwest–northeast, generally perpendicular to the linear orientation of events (see Fagan, 2012), but their size blurs the distinction between subgroups. That, along with high time-domain waveform correlation, made distinguishing detailed subsurface structure difficult. Spectral correlation, however, reveals event subclusters suggesting separate parallel fractures striking north-northeast–south-southwest, oblique to the main fault lineament and characteristic of en echelon fracturing.

En echelon fractures are planar structures. Although the formation depths below the oil-producing zones are not well understood in the region of microseismicity, assuming the layer thickness contour plot in Chidsey (2009) represents thicknesses patterns of the underlying formations, the consistent depths of multiplet 18 events indicate that they occur within the same structure and are therefore planar as well. This collection of evidence leads to the characterization of en echelon faulting as a result of stress changes associated with Aneth's 50-plus year production history. In future work, we can test this model. Popular techniques to refine hypocenter distributions include *hypodd* (Waldhauser and Ellsworth, 2000) and *tomodd* (Zhang and Thurber, 2003), based on the double differences between P- and S-wave arrivals of closely spaced events. The methods rely on interevent distances not to exceed the scale length of the velocity model, or bias may be introduced. An assumption is that events are from the same fracture or medium. We propose that our spectral method helps predefine subsets of hypocenters, which in turn can then be relocated via double-difference tomography.

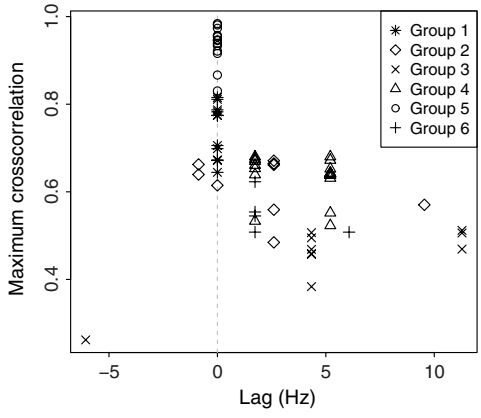

Figure 14. Maximum crosscorrelations for each multiplet 18 event crosscorrelated with the westernmost spectral cluster mean (group 5). Events are generally less correlated away from the group 5 centroid (the westernmost group), but lags follow a generally positive trend, an indication of a frequency shift with distance from the receiver.

CONCLUSIONS

Microseismicity in areas of oil and gas production, as well as CO_2 injection, provides valuable information about time-lapse reservoir conditions. Based on an analysis of events clustered in the time domain, a novel spectral clustering technique reveals more subtle variations in the distribution of hypocenters than is available through time-domain analysis. The dissimilarity metric upon which this spectral clustering is based illuminates subtle, potentially undetectable differences in the time domain and is supported by a fully developed body of statistical theory upon which future applications of statistical decision making using event spectra may be based.

We hypothesize that the new fracture patterns revealed by the spectral clustering technique are characteristic of stress release as en echelon faulting. These details may be valuable as a preprocessing step for high-resolution time-domain location algorithms such as double-difference tomography. Further study is needed to determine whether these results can be observed for other clusters within this data set and whether other field sites yield similar results.

ACKNOWLEDGMENTS

We would like to acknowledge Schlumberger and, in particular, I. Bradford and D. Raymer for their support in this research. In addition, we thank L. Adam, J. Pelton, and M. Smith for their valuable feedback on the contents of this manuscript.

APPENDIX A

DISCUSSION OF HIERARCHICAL AGGLOMERATIVE CLUSTERING

Here, we summarize hierarchical agglomerative clustering using the discrete Fourier transform of the autocovariance function as the power spectrum. See Diggle (1991), R Development Core Team (2008), and Venables and Ripley (1999), for details of calculating the spectrum and Kaufman and Rosseeuw (1990), and R Development Core Team (2008), for additional details of hierarchical agglomerative clustering.

In agglomerative clustering, each event starts out as its own cluster; clusters are agglomerated (combined) into larger ones, culminating in a single cluster that contains all events. Agglomeration decisions are based on a dissimilarity metric, among which there are many to choose, including the Euclidean distance, Mahalanobis distance, and Kullback-Liebler divergence, to mention a few. This application uses squared spectral distance, defined below, as the dissimilarity metric.

Among clustering algorithms, hierarchical agglomerative clustering is attractive for several reasons: (1) unlike k-means, the number of clusters is not required in advance, (2) because it works on a distance matrix, it is robust to the order of events, and (3) interactive determination of the number of clusters provides flexibility, with the added advantage of being able to quickly identify outliers and singletons. Hierarchical agglomerative clustering is a widely used algorithm, successfully applied in other seismic applications (Rowe et al., 2002; Bardainne et al., 2006).

For an observed waveform (event) $\{y_t\colon t = 1, 2, \ldots, n\}$ with zero mean, the sample autocovariance for lag k is

$$f(k) = \frac{1}{n}\sum_{t=k+1}^{n} y_t y_{t-k}, \qquad k = 0, 1, \ldots, n-1. \qquad \text{(A-1)}$$

The spectrum is the discrete Fourier transform of f and has the form

$$F(\omega) = f(0) + 2\sum_{k=1}^{n-1} f(k)\cos(k\omega), \qquad \text{(A-2)}$$

where $f(0)$ is the variance of y and $\omega = 2\pi j/n$ for positive integer $j < n/2$.

Starting with the P-wave arrival, each waveform is trimmed to the length of the shortest duration event so that the Fourier frequencies are constant between events and interpolation is not necessary. The first 400 Fourier frequencies are then used for clustering. The shortest length event is 1598 samples, which is then padded with zeros to $n = 1600$ to optimize the discrete Fourier transform. The result is 800 Fourier frequencies with a delta frequency of 0.937 Hz.

The squared spectral distance is defined as the squared Euclidean distance using the first 400 spectrum amplitudes. The squared spectral distance for events a and b is

$$d^2(a, b) = [F_a(\omega) - F_b(\omega)]^T[F_a(\omega) - F_b(\omega)], \qquad \text{(A-3)}$$

with F from equation A-2.

At each agglomeration step, the dissimilarity measure must be updated. We use Ward's minimum variance criterion (Ward, 1963), which minimizes $\mathrm{Var}[d^2]$ over all candidate agglomerations. Ward's minimum variance only works with true distance measures, which is our motivation for using spectral distance. The Lance-Williams algorithm provides an efficient method of performing these calculations, and it is discussed in detail in Kaufman and Rosseeuw (1990). If two clusters C_1 and C_2 have been agglomerated, the dissimilarity between their union and another cluster M is given by

$$D(C_1 \cup C_2, M) = D(C_1, M) + D(C_2, M) - 0.5 * D(C_1, C_2), \qquad \text{(A-4)}$$

where D is $\mathrm{Var}[d^2]$, with d^2 as defined in equation A-3.

As noted, we use the Fourier transform of the autocorrelation function (ACF) as our spectral estimate. The statistical properties of this estimator lend themselves to future research and statistical decision-making opportunities. The statistical instability of intensity estimates is offset by fine tuning the taper and smoothing parameters, which, if varying between events, could result in different peak locations. We mitigate variability between event spectra by using the same parameters among events as well as by making all events the same length. Each spectrum has identical Fourier frequencies and an equal number of observations for each intensity estimate across events. Slight differences between spectra may be present at the extreme frequencies, which may be impacted by edge effects, but this is also mitigated by clustering over a range of frequencies inside the extremes.

The ordinates (autocorrelation coefficients) of the ACF are not statistically independent, whereas the Fourier transform of the ACF is an orthogonal partitioning of the sum of squares that results in statistically independent ordinates (intensities). The absolute value of the ACF can give an indication of whether certain frequencies are prevalent in the waveform, but because spectrum coefficients (intensities) are independent, we can use fully established theory to statistically compare spectral intensities. It is clear that the event spectra do not cluster randomly, as evidenced by their spatial organization. Statistical analysis of cluster membership is a topic of future research, as requirements of statistical decision making in this application are developed.

REFERENCES

Bardainne, T., P. Gaillot, N. Dubos-Sallée, J. Blanco, and G. Sénéchal, 2006, Characterization of seismic waveforms and classification of seismic events using chirplet atomic decomposition. Example from the Lacq gas field (Western Pyrenees, France): Geophysical Journal International, **166**, 699–718, doi: 10.1111/j.1365-246X.2006.03023.x.

Carney, S., 2010, Subsurface structural and thickness mapping: Technical report, Chapter 3, Utah Geological Survey.

Chidsey, T. C. Jr., 2009, Surface and subsurface geological characterization of the Aneth Unit, Greater Aneth field, Paradox Basin, Utah — Final report: Technical report, Utah Geological Survey.

Diggle, P. J., 1991, Time series: A biostatistical introduction: Clarendon Press.

Fagan, D. K., 2012, Statistical clustering of microseismic event spectra to identify subsurface structure: M.S. thesis, Boise State University.

Hansen, S. E., S. Y. Schwartz, H. R. DeShon, and V. González, 2006, Earthquake relocation and focal mechanism determination using waveform cross correlation, Nicoya Peninsula, Costa Rica: Bulletin of the Seismological Society of America, **96**, 1003–1011, doi: 10.1785/0120050129.

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

Hintze, L. F., and B. J. Kowallis, 2009, Geologic history of Utah: Brigham Young University Geology Studies Special Publication 9.

Kaufman, L., and P. J. Rosseeuw, 1990, Finding groups in data: An introduction to cluster analysis: Wiley.

Lay, T., and T. C. Wallace, 1995, Modern global seismology: Academic Press.

Maxwell, S. C., and T. I. Urbancic, 2001, The role of microseismic monitoring in the instrumented oil field: The Leading Edge, **20**, 636–639, doi: 10.1190/1.1439012.

Phillips, W. S., T. D. Fairbanks, J. T. Rutledge, and D. W. Anderson, 1998, Induced microearthquake patterns and oil-producing fracture systems in the Austin Chalk: Tectonophysics, **289**, 153–169, doi: 10.1016/S0040-1951(97)00313-2.

R Development Core Team, 2008, R: A language and environment for statistical computing, R Foundation for Statistical Computing.

Rowe, C. A., R. C. Aster, B. Borchers, and C. J. Young, 2002, An automatic, adaptive algorithm for refining phase picks in large seismic data sets: Bulletin of the Seismological Society of America, **92**, 1660–1674, doi: 10.1785/0120010224.

Rutledge, J., and N. Soma, 2010, Microseismic monitoring of CO_2 enhanced oil recovery in the Aneth Field: Technical report LA-UR-10-08213, Chapter 12, Los Alamos National Laboratory.

Rutledge, J. T., and W. S. Phillips, 2003, Hydraulic stimulations of natural fracture as revealed by induced microearthquakes, Carthage Cotton Valley gas field, east Texas: Geophysics, **68**, 441–452, doi: 10.1190/1.1567214.

Saar, M. O., and M. Manga, 2003, Seismicity induced by seasonal groundwater recharge at Mt. Hood, Oregon: Earth and Planetary Science Letters, **214**, 605–618, doi: 10.1016/S0012-821X(03)00418-7.

Schaff, D. P., and G. C. Beroza, 2004, Coseismic and postseismic velocity changes measured by repeating earthquakes: Journal of Geophysical Research, **109**, B10302, doi: 10.1029/2004JB003011.

Schaff, D. P., and P. G. Richards, 2004, Repeating seismic events in China: Science, **303**, 1176–1178, doi: 10.1126/science.1093422.

Schaff, D. P., and P. G. Richards, 2011, On finding and using repeating seismic events in and near China: Journal of Geophysical Research, **116**, B03309, doi: 10.1029/2010JB007895.

Shumway, R. H., 2003, Time-frequency clustering and discriminant analysis: Statistics & Probability Letters, **63**, 307–314, doi: 10.1016/S0167-7152(03)00095-6.

Shumway, R. H., and D. S. Stoffer, 2006, Time series analysis and its applications: With R examples, 2nd ed.: Springer-Verlag.

Snieder, R., and M. Vrijlandt, 2005, Constraining relative source locations with coda wave interferometry: Theory and application to earthquake

doublets in the Hayward Fault, California: Journal of Geophysical Research, **110**, B04301, doi: 10.1029/2004JB003317.

Song, F., H. S. Kuleli, M. N. Toksöz, E. Ay, and H. Zhang, 2010, An improved method for hydrofracture-induced microseismic event detection and phase picking: Geophysics, **75**, no. 6, A47–A52, doi: http://link.aip.org/link/doi/10.1190/1.3484716.

Streit, J. E., A. F. Siggins, and B. J. Evans, 2005, Predicting and monitoring geomechanical effects of CO_2 injection, *in* D. C. Thomas, Carbon dioxide capture for storage in deep geologic formations, vol. 2: Elsevier Ltd., 751–766.

Venables, W. N., and B. D. Ripley, 1999, Modern and applied statistics with S-PLUS, 3rd ed.: Springer-Verlag.

Waldhauser, F., and W. L. Ellsworth, 2000, A double-difference earthquake location algorithm: Method and application to the northern Hayward Fault, California: Bulletin of the Seismological Society of America, **90**, 1353–1368, doi: 10.1785/0120000006.

Ward, J. H. Jr., 1963, Hierarchical grouping to optimize an objective function: Journal of the American Statistical Association, **58**, 236–234, doi: 10.1080/01621459.1963.10500845.

Zhang, H., and C. H. Thurber, 2003, Double-difference tomography: The method and its application to the Hayward Fault, California: Bulletin of the Seismological Society of America, **93**, 1875–1889, doi: 10.1785/0120020190.