# Classification of microseismic events from bitumen production at Cold Lake, Alberta

Jeffrey F. Tan, Henry C. Bland and Robert R. Stewart


## ABSTRACT

This work develops and tests microseism event-classification techniques. Research was performed in collaboration with CREWES using microseismic data from Cold Lake, Alberta that was provided by Imperial Oil Ltd. The objective was to develop passive-seismic signal classification algorithms capable of precisely and automatically distinguishing between microseismic events warranting further investigation from noise events that are generally not of interest. Novel methods involving frequency-filtering, event-length detection, and statistical analysis were developed.

After extensive testing, it was found that developed statistical analysis algorithms performed best. Principal components analysis was applied to statistical analysis algorithm outputs to optimize classification.

A MATLAB® implementation scheme was created that yielded classification accuracies between 90% and 99.5% when tested on a wide range of datasets. Given that up to tens of thousands of microseismic events are detected daily at Cold Lake, this work could have significant future impact.


## INTRODUCTION

### Background

Passive-seismic monitoring listens for small earthquakes (microseisms) that can occur when there are stress changes in a reservoir (Maxwell and Urbanic, 2001). As opposed to conventional surface-seismic acquisition techniques, passive-seismic monitoring does not employ a source to create elastic waves. Instead, spontaneous subsurface microseismic events are detected with sensors when they occur.

Imperial Oil Limited is involved in oil sands production at Cold Lake, Alberta, Canada. At Cold Lake, hydrocarbon production comes from the Clearwater formation, which has a primarily sandstone lithology. This producing formation is buried over 400 m deep at Cold Lake, and the bitumen contained within it has an American Petroleum Institute (API) gravity index of approximately 8° to 9°. Cyclic steam stimulation (CSS), which creates pressures and temperatures of approximately 320°C and 11 MPa, respectively in the Clearwater formation (Campbell, 2005), is required to extract the viscous bitumen. Mechanical issues in producing wells such as cement cracks or casing failures can result from the high pressures and temperatures associated with CSS. If undetected, these production issues could result in large cleanup costs, in addition to potential legal implications. A microseismic earthquake with its focus near the damaged area is created when these mechanical issues occur. Imperial Oil Limited operates a passive-seismic monitoring system at Cold Lake to proactively detect these microseisms. The Consortium for Research in Elastic Wave Exploration Seismology (CREWES)

Project at the University of Calgary was involved in this passive-seismic research with Imperial Oil.

The passive-seismic monitoring system implemented at Cold Lake is present on approximately 75 production pads. Each pad has a centrally located monitoring well that records ground vibrations (including microseisms). The monitoring well is instrumented by a down-hole array of five or eight 3-component (3-C) geophone sondes connected to seismic recorders at the surface (Tan et al., 2006). Seismic recorders listen for discrete seismic events and store them as microseismic event files to disk for later review. For an array of five (eight) geophones, these digital event files contain fifteen (twenty-four) traces that display 1.365 seconds (1.5 seconds) of microseismic activity recorded by the 3-component geophone sondes. Three traces are outputted from each 3-C geophone.

Vendor-supplied event-classification software analyzes each created microseismic file and assigns a classification. If a file is classified as "good", this indicates that the software has decided that the event file warrants further investigation; conversely, if a file is classified as "noise", it is supposedly an event that is not of interest (Tan et al., 2006). Approximately 99% of all detected events are noise. Examples of "good" events worth further investigation include cement cracks around the casing in the wells, and casing failures. Examples of noise events include noise created by pump rods and passing vehicles (Campbell, 2005). When a “good” event is automatically detected followed by manual confirmation, an attempt is made to locate its hypocenter, its point of origin.

Many microseismic “good” events tend to have impulsive arrivals, high P-wave arrival frequencies, and decreasing signal frequency with increasing time (Lee and Stewart, 1981). Figure 1 shows an example of a “good” event from the Cold Lake dataset, with P- and S-wave arrivals indicated. Empirically, it can be determined that this event is “good” due to the distinct and impulsive P- and S-wave arrivals. The decrease in frequency of the S-wave arrival compared to the P-wave can be seen. Figure 2 shows an example noise event, which is generally random and does not follow deterministic properties. The microseismic traces shown in Figures 1 and 2 have been normalized to the largest data value (in magnitude), and have any direct-current (DC) offset removed.

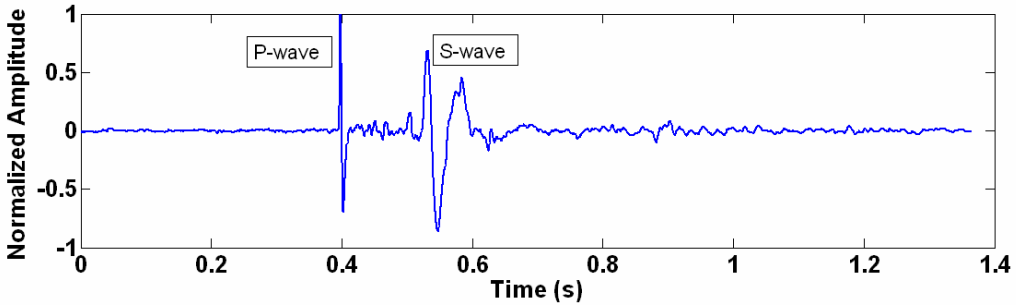


FIG. 1. Example of a “good” event, with P- and S-wave arrivals indicated. Empirically, this event is “good” due to the distinct and impulsive P- and S-wave arrivals. The decrease in frequency of the S-wave arrival compared to the P-wave can be seen.

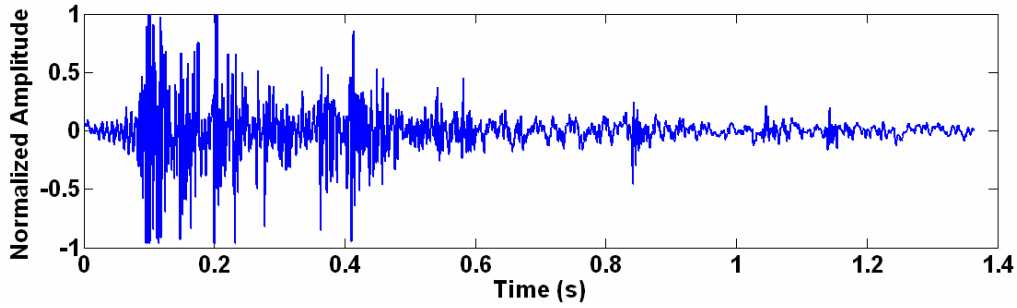


FIG. 2. Example of a noise event. Noise events are generally random and do not follow deterministic properties of "good" events.

## Objective

The current event-file classification software present on the Cold Lake passive-seismic monitoring system has been known to misclassify a large portion of the received event files. This has resulted in many noise events being incorrectly identified as "good" events, often referred to as "false positives". Manual investigation of misclassified files can become very costly and time-consuming.

The purpose was to develop automatic and robust microseismic signal analysis algorithms capable of precisely classifying the microseismic event files generated by the passive-seismic monitoring system at Cold Lake. Compared to noise events, I observed that many "good" events generally have lower dominant frequency content, shorter P-wave event-lengths, and flatter time-domain characteristics. Based on these observations, I developed classification algorithms involving frequency-filtering, event-length detection, and statistical analysis.

After extensive testing on microseismic data from Cold Lake, it was determined that the developed statistical analysis algorithms demonstrated the strongest classification performance. Principal components analysis (PCA), a multivariate data reduction technique, was applied to statistical algorithm outputs for optimized microseismic file classification. A MATLAB® application was developed and tested on three microseismic datasets from Cold Lake, yielding classification accuracies between 90% and 99.5%. Given that up to tens of thousands of microseismic events are detected daily at Cold Lake, this application could have significant future impact.

## ALGORITHMS

Compared to noise events, I observed that many "good" events generally have lower dominant frequency content, shorter P-wave event-lengths, and flatter time-domain characteristics. Based on these observations, classification algorithms involving frequency-filtering, event-length detection, and statistical analysis were developed.

## Frequency filtering

In practical applications, there are four main filter responses that rely on approximations (e.g. Maundy, 2005): the Butterworth, Chebyshev, Inverse-Chebyshev, and Cauer. Three of them -- the Butterworth, Chebyshev, and Inverse-Chebyshev responses -- were used to distinguish "good" events from noise.

The Butterworth response has a maximally flat attenuation characteristic in the passband as frequency increases (an advantage) and a monotonically increasing attenuation characteristic in the stopband (e.g. Haykin and Van Veen, 2003). The transition region between the passband and stopband limits of the Butterworth filter, however, is less sharp, a disadvantage, than that of a different response, such as the Chebyshev.

The Chebyshev response has an "equiripple" attenuation characteristic in the passband (e.g. Zhou and McMechan, 1999) and a monotonically increasing attenuation characteristic in the stopband. The passband equiripple characteristic pertains to oscillations in the frequency response, which is a disadvantage of the Chebyshev filter; however, this response provides greater stopband attenuation than the Butterworth response for a given filter order, which is advantageous (e.g. Maundy, 2005).

The Inverse-Chebyshev response contains a maximally flat attenuation characteristic in the passband and an equiripple characteristic in the stopband. For a given minimum stopband attenuation, an Inverse-Chebyshev filter will require a lower order than a Butterworth filter (an advantage). When designing band-elimination topologies, however, the Inverse-Chebyshev response requires circuit realizations with high complexity (e.g. Maundy, 2005).

To provide a diverse set of frequency responses and optimize the aggregate performance of these classification algorithms, each of the Butterworth, Chebyshev, and Inverse-Chebyshev frequency response approximations were modeled as filters in MATLAB®. Results obtained from the Inverse-Chebyshev and Butterworth filters follow.

Detailed mathematical theory pertaining to filter frequency responses is provided in literature by Sheriff and Geldart (1995); Haykin and Van Veen (2003); as well as Maundy (2005), among others. For conciseness and brevity, this theory is not repeated here.

### *Low-pass Inverse-Chebyshev filter*

A fourth-order low-pass Inverse-Chebyshev filter with a 100 Hz lower stopband limit and 36 Hz upper passband limit was created. This filter provides at least 40 dB attenuation in the stopband and does not attenuate passband frequencies by more than 0.25 dB. Figure 3 shows the magnitude response for this filter.

Figures 4 and 5 show the "good" and noise traces from Figures 1 and 2 after low-pass filtering. The magnitude of the peak data value in Figure 4 (approximately 0.5) is significantly higher than that of Figure 5, which is approximately 0.13.

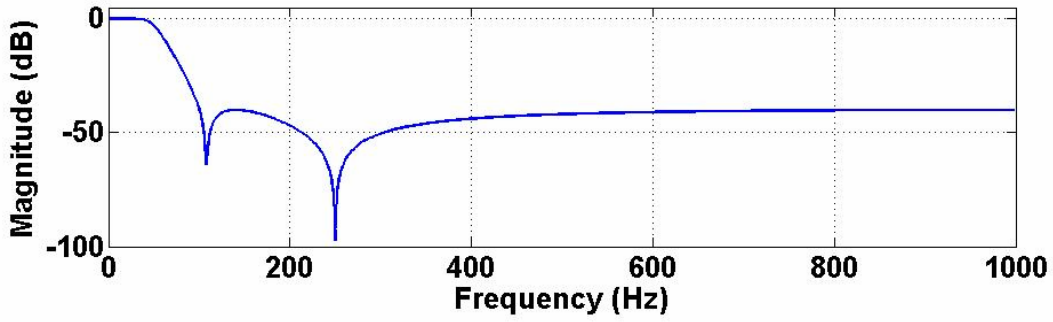


FIG. 3. Amplitude response of low-pass Inverse-Chebyshev filter.

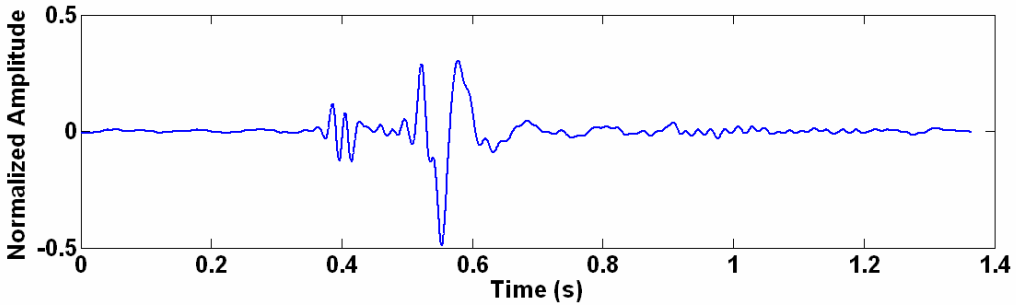


FIG. 4. Result of low-pass filtering the "good" trace in Figure 1.

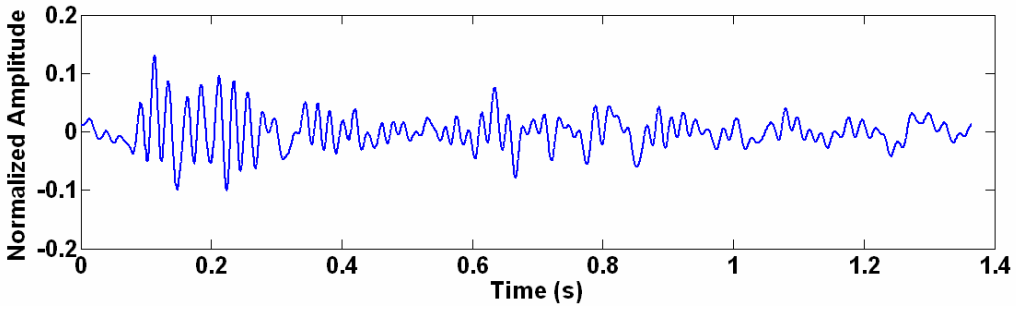


FIG. 5. Result of low-pass filtering the noise trace in Figure 2.

*High-pass Butterworth filter*

A fourth-order high-pass Butterworth filter with a 215 Hz upper stopband limit and 398 Hz lower passband limit was created. This filter provides at least 25 dB attenuation in the stopband and does not attenuate passband frequencies by more than 3 dB. Figure 6 shows the magnitude response for this filter.

Figures 7 and 8 show the "good" and noise traces from Figures 1 and 2 after high-pass filtering. The magnitude of the peak data value in Figure 7 (approximately 0.05) is significantly lower than that of Figure 8, which is approximately 0.7.

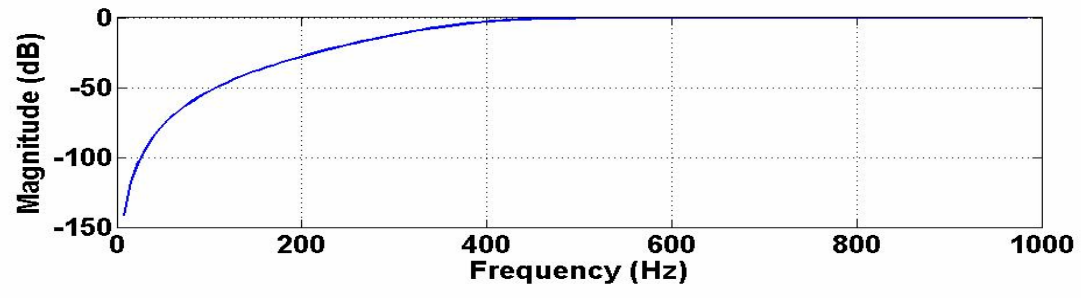


FIG. 6. Amplitude response of high-pass Butterworth filter.

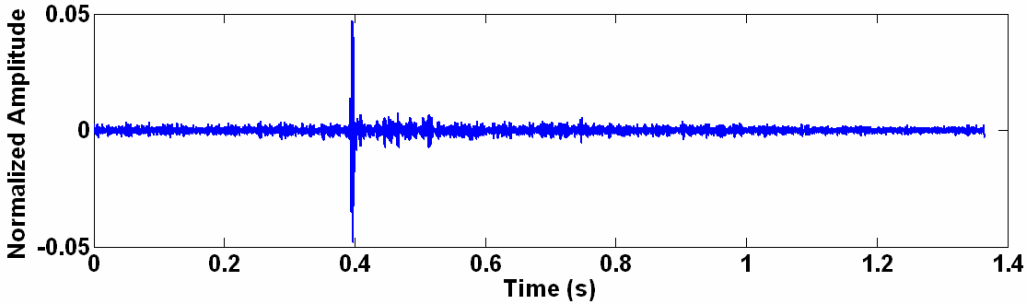


FIG. 7. Result of high-pass filtering the “good” trace in Figure 1.

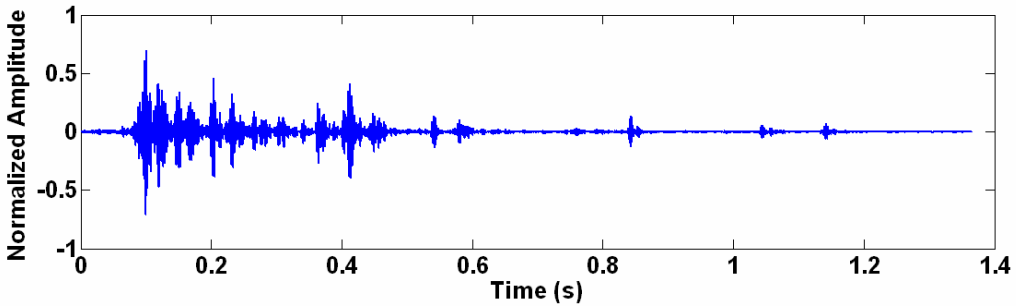


FIG. 8. Result of high-pass filtering the noise trace in Figure 2.

## Event-length detection

The impulsive P-wave event-lengths in many “good” traces are usually significantly shorter than noise event-lengths. This set of classification algorithms determines the first-arrival event-lengths.

*STA / LTA algorithm*

This time-domain algorithm continuously calculates ratios of short-term averages (STA) to long-term averages (LTA) of microseismic energy. This is the STA/LTA technique (Ambuter and Solomon, 1974) demonstrated by Munro (2005). The STA/LTA ratio will significantly increase at the onset of a microseismic event; conversely, this ratio will significantly decrease at the event's termination. Calculating the time interval between the onset and termination of a microseismic event yields its approximate length.

To demonstrate an application of this algorithm to microseismic event detection, assume there are $N$ data samples in the microseismic trace under examination. Take the sampling frequency to be $f_s$. Define $S$ as the length of the STA window in seconds containing $p$ data points. Similarly, define $L$ as the length of the LTA window in seconds containing $q$ data points. Then,

$$S = \frac{p}{f_s}; \tag{1}$$

and

$$L = \frac{q}{f_s}.$$

Let $a_i$ represent the amplitude of an arbitrary data point in the microseismic trace with $i = 0, 1, 2, \ldots, N\text{-}1$. The relationship between a point in time $t_i$ and the index "$i$" is

$$t_i = \frac{i}{f_s} \tag{2}$$

The following application of this algorithm was found to work best for microseismic event-length determination. To perform the first STA / LTA calculation, the LTA window was positioned such that it calculated a long-term energy average from the beginning of the trace at $t = 0$ to the time $t = L$. Using equation 2, this corresponds to data point indices from $i = 0$ to $i = Lf_s$. The STA window was positioned such that it calculated a short-term energy average of the trace from the time $t = L - S$ to the time $t = L$. This corresponds to data point indices from $i = (L\text{-}S)f_s$ to $i = Lf_s$. Let $STA_x$ and $LTA_x$ represent the $x^{th}$ short-term and long-term energy calculations, respectively. Let $R_x$ signify the $x^{th}$ calculated STA / LTA ratio. The STA / LTA algorithm applied here starts with $x = 1$ followed by subsequent single data point energy window advancements until a significant increase in $R_x$ is seen, indicating the detection of the P-wave arrival. The termination of the event is detected through a significant decrease in $R_x$. If no event is detected, then calculations of $R_x$ terminate when the energy windows reach the end of the microseismic trace. The corresponding equations for $STA_x$, $LTA_x$, and $R_x$ are

$$STA_x = \frac{\sum_{i=(L-S+x-1)f_s}^{Lf_s+x-1} a_i^2}{p}; \qquad (3)$$

$$LTA_x = \frac{\sum_{i=x-1}^{Lf_s+x-1} a_i^2}{q};$$

and

$$R_x = \frac{STA_x}{LTA_x}.$$

*Continuous-time frequency analysis*

I observed that the high-frequency content of many microseismic traces usually increased significantly at the onset of an event and decreased significantly at the event's termination.

A developed method to pick the onset and termination times of microseismic events was to continually analyze the frequency characteristics of a select number of points in the microseismic trace. A continuous-time frequency analysis was performed by supplying a moving frequency transform window. This developed technique is similar to the Gabor transform (e.g. Feichtinger and Strohmer, 1998) and S-transform (Stockwell et al., 1996) techniques which represent frequency transform methods that can be localized in time. These transform methods are useful when dealing with signals that are not stationary, such as earthquake seismograms (e.g. Stockwell et al., 1996).

For discrete signals, the discrete Fourier transform (DFT) is used to determine frequency characteristics. The generalized definition of the DFT is

$$f_m = \sum_{n=0}^{N-1} f_n e^{\frac{-j2\pi mn}{N}};$$

$$j = \sqrt{-1} \qquad (4)$$

(e.g. Margrave 2007). In equation 4, $n$ corresponds to the indices of the time-series elements $f_n$ (with $n$ = 0, 1, 2, 3, …, $N$-1), $m$ corresponds to the indices of the frequency elements $f_m$ (with $m$ = 0, 1, 2, 3, …, $N$-1), and $N$ is the number of time series data points in the examined signal.

For the developed algorithm, equation 4 was localized to select groups of time series data points in a microseismic trace. Let an examined microseismic trace sampled at a frequency $f_s$ contain $N$ data points. Let $d_i$ represent an arbitrary data point in this trace with $i$ = 0, 1, 2,…, $N$-1. Let $W$ represent a fast Fourier transform (FFT) window containing $k$ data points where $k < N$. Define $W_p$ as the $p^{th}$ data point contained in $W$, with $p$ = 0, 1, 2, … , $k$-1. The fast Fourier transform (FFT) is a built-in MATLAB® function

that calculates the discrete Fourier transform using an optimized method to reduce computation time.

To start, $W$ is positioned at the beginning of an examined microseismic trace, and an FFT of data points from $d_0$ to $d_{k-1}$ is applied. The $k$ time-series elements are transformed to $k$ frequency elements. Let $f_m$ represent the $m^{th}$ frequency element of the transformed data (with $m$ = 0, 1, 2, 3, …, $k$-1) contained in the window $W$. For this algorithm, equation 4 can be rewritten as

$$f_m = \sum_{p=0}^{k-1} W_p e^{\frac{-j2\pi mp}{k}} \quad . \tag{5}$$

The Nyquist frequency is defined as

$$f_{NYQ} = \frac{f_s}{2}. \tag{6}$$

Let $F_m$ represent the frequency (in Hz) that each corresponding transformed frequency element $f_m$ pertains to. The expression for $F_m$ will depend on whether $k$ is odd or even. If $k$ is even, the equation is

$$F_m = \frac{mf_s}{k}, \qquad m \le \frac{k}{2}; \tag{7a}$$

$$F_m = \frac{mf_s}{k} - 2f_{NYQ}, \qquad m > \frac{k}{2}, \tag{7b}$$

whereas if k is odd, the expression is

$$F_m = \frac{mf_s}{k}, \qquad m \le \frac{k-1}{2}; \tag{8a}$$

$$F_m = \frac{mf_s}{k} - 2f_{NYQ}, \qquad m > \frac{k-1}{2}. \tag{8b}$$

Equations 7a and 8a correspond to positive frequencies, while equations 7b and 8b correspond to negative frequencies. Comparing these equations with equation 6, note that $|F_m| \le f_{NYQ}$ for all $m$.

After the FFT step is complete, the power spectral density (PSD) of the frequency elements $f_m$ is calculated. Let $P_m$ represent the PSD magnitude corresponding to a transformed frequency element $f_m$ given by

$$P_m = f_m f_m^* = |f_m|^2 \tag{9}$$

where "*" represents the complex conjugate operator. If the time series contained in $W$ is real, as expected for a microseismic trace, then it can be shown that the amplitude spectrum of the transformed elements, $|f_m|$, is an even function in frequency (e.g. Haykin and Van Veen, 2003), implying that $P_m$ is an even function in frequency as well. Thus, transformed elements corresponding to negative frequencies can effectively be ignored,

as they are redundant, and this analysis can be restricted to positive frequencies to reduce computation time. The elements in $P_m$ are then placed in a single column of a 2-D matrix whose rows and columns correspond to frequency and time, respectively.

The above procedure is repeated as the window $W$ moves forward by single data point increments until the end point of this window reaches $d_{N-1}$. High frequency content in the calculated PSD can be examined to determine the onset and termination times of an event.

Figure 9 displays the "good" microseismic trace shown in Figure 1 with this continuous-time Fourier transform algorithm applied to it. This trace has $N$ = 4096 and $f_s$ = 3000 Hz. The size of the fast Fourier transform window applied is $k$ = 100. The vertical lines near a time of 0.4s depict the onset and termination temporal locations of the P-wave arrival. The lowest (highest) PSD amplitudes are depicted in blue (red). There is a sharp increase in high frequency content at the onset of the event, and a sharp decrease in high frequency content at the event's termination. The calculated length of the P-wave event was 35 ms. Similarly, Figure 10 shows the noise trace in Figure 2 with this transform applied. The noise event-length was calculated to be 547 ms, significantly longer than the calculated P-wave event-length of 35 ms.

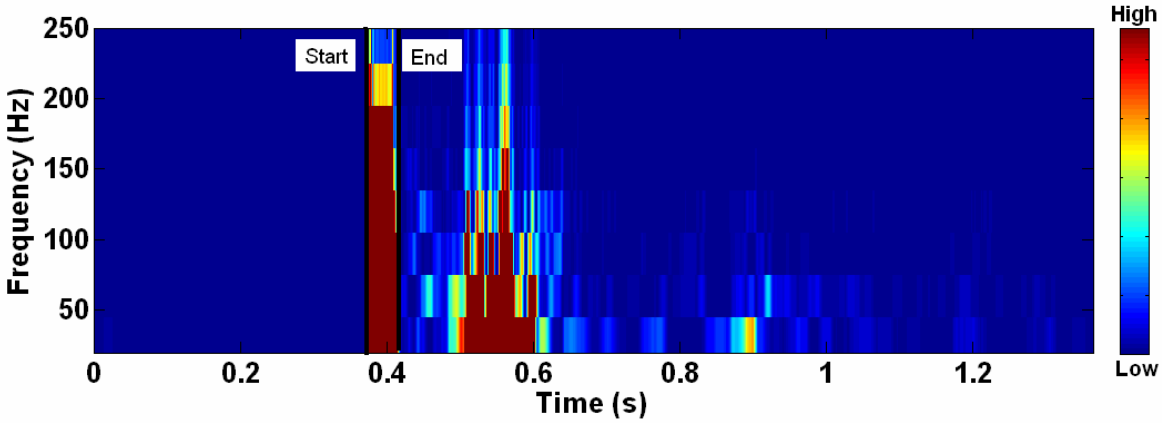


FIG. 9. Continuous-time Fourier transform PSD plot of "good" trace shown in Figure 1 using an FFT window size of $k$ = 100. The color scale ranges from blue (low amplitude) to red (high amplitude). A sharp increase in high frequency content can be seen at the onset of the P-wave arrival, while a sharp decrease in high frequency content can be seen at the event's termination. The calculated event-length was 35 ms.

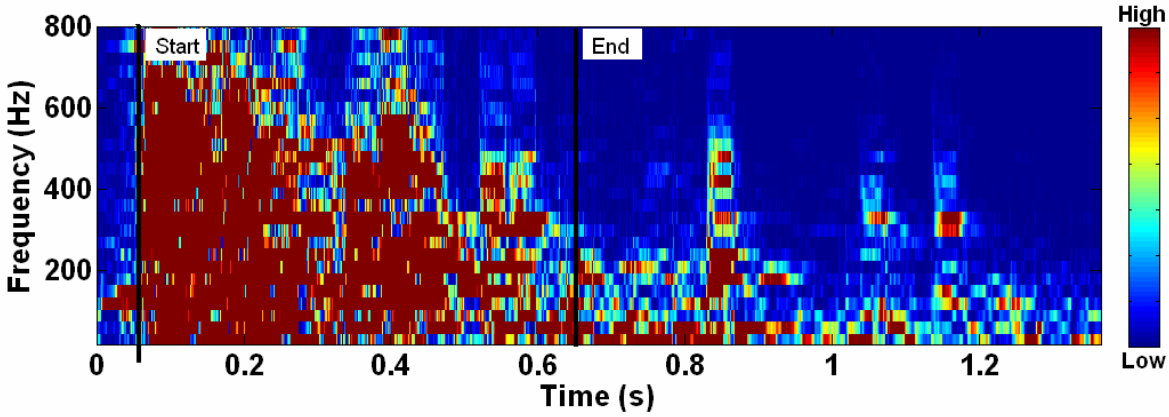


FIG. 10. Continuous-time Fourier transform PSD plot of noise trace shown in Figure 2 using an FFT window size of $k$ = 100. The color scale ranges from blue (low amplitude) to red (high amplitude). A sharp increase in high frequency content can be seen at the onset of the noise event, while a sharp decrease in high frequency content can be seen at the event's termination. The calculated event-length was 547 ms.

## Statistical analysis

Compared to noise traces, I observed that "good" microseismic traces generally contain lower signal variance, higher central data distribution, more frequent oscillations, and greater signed amplitude differences between adjacent time-series data points. Statistical analysis classification algorithms were developed based on these observations.

### *Threshold algorithm*

This algorithm was developed based on "good" microseismic traces generally containing lower signal variance than noise traces. As an example, define $g_i$ ($n_i$) as the signed amplitude of the $i^{th}$ time series data point corresponding to the "good" (noise) trace shown in Figure 1 (2), for $i$ = 0,1, 2,…, $N$-1, where $N$ is the total number of data points in the time series. For the microseismic traces depicted in Figures 1 and 2, $N$ = 4096.

Define variance of the time series corresponding to the "good" (noise) trace as $\sigma^2_g$ ($\sigma^2_n$). Then, for the zero-mean signal (DC offset has been removed) depicted in Figure 1, for example,

$$\sigma^2_g = \frac{\sum_{i=0}^{N-1} g_i^2}{N-1}. \qquad (10)$$

The variance of the noise trace $\sigma^2_n$ simply corresponds to replacing $g$ with $n$ in equation 10.

The "good" microseismic trace in Figure 1 has a variance of 0.0128, while the noise trace in Figure 2 has a variance of 0.0293. By empirically examining these two figures, a higher noise trace variance would be expected. Over the whole trace, the noise trace contains visibly larger deviations from the mean compared to the "good" trace. The likelihood of dataset points existing a given distance away from its mean is proportional

to variance and can be quantified by applying "Chebyshev's inequality". For detailed theory pertaining to Chebyshev's inequality refer to literature authored by Lange (2003); Therrien and Tummala (2004); Miller and Childers (2004); Mitzenmacher and Upfal (2005); and Suhov and Kelbert (2005). For brevity and conciseness, this theory is not repeated here.

Chebyshev's inequality is given as

$$\Pr(|\,\mathrm{X} - \mathbf{E}[\mathrm{X}]\,| \;\geq a) \leq \frac{\mathbf{VAR}[\mathrm{X}]}{a^2} \tag{11}$$

(e.g. Suhov and Kelbert, 2005), where X is a random variable with expected value E[X] and variance VAR[X], $a$ is a strictly positive constant ($a > 0$), and "Pr()" represents event probability. Within the context of this discussion corresponding to microseismic signal analysis, X pertains to arbitrary time series data points, E[X] = 0, and VAR[X] pertains to microseismic signal variance.

By examining equation 11, it can be seen that Chebyshev's inequality states that a larger dataset variance corresponds to an increase in the expected maximum number of data points lying outside a mean-centered window of width $2a$ (or single-sided width $a$). The developed "Threshold" algorithm applies equation 11 in a simple manner.

Define a window $w$ existing from $-a \leq w \leq a$. This window is centered at the mean of examined microseismic traces, which is zero. The "Threshold" algorithm simply determines the fraction of time series data points that lie outside the threshold limits $\pm a$ of the window $w$. This fraction can then be used for microseismic signal classification. The "good" trace in Figure 1 has 20.8% of its time series data points outside the window $w$, while the noise trace in Figure 2 has 68.2% outlying data points.

*Histogram algorithm*

I observed that the time series signed amplitude distribution of "good" traces tended to be more heavily concentrated near the time axis than noise traces, as can be seen in the "good" and noise traces in Figures 1 and 2, respectively. One tool that can be used to examine and quantify data distribution is a histogram.

Assume that the signed amplitude of the $i^{th}$ data point in an $N$-point discrete time series is $d_i$. For example, this time series could be one of the microseismic signals shown in Figures 1 or 2. Assume that both positive and negative values are present in this time series. Let $d_{max}$ denote the largest positive value in the time series, and $d_{min}$ denote the largest negative value. Assume there are $n$ histogram bins each with a bin width of $b$. In this simple case, the total number of histogram bins, $n$, can then be calculated as

$$n = \left\lceil \frac{d_{max} - d_{min}}{b} \right\rceil, \tag{12}$$

where "$\lceil\ \rceil$" signifies rounding upwards to the closest integer. Let $m_k$ represent histogram bin $k$ for $k = 0,1,2,...n-1$. One possible method is to position the lower boundary of the

first bin $m_0$ on $d_{min}$, the second bin $m_1$ on $d_{min} + b$, and so on. Thus, $m_k$ would record the total number of points that fall in the range $d_{min} + kb \le d_i < d_{min} + (k+1)b$. The notation

$$m_k \rightarrow [d_{min} + kb \le d_i < d_{min} + (k+1)b]; \quad k = 0,1,2,...,n-1 \tag{13}$$

will be used to signify this.

An issue that arises when creating a histogram includes determining the optimal bin width $b$ (or equivalently, through equation 12, the total number of bins $n$). It is desired to choose $b$ (or $n$) such that the data distribution is well represented. If $b$ is too large, critical information may be smeared across a single bin. That is, the frequency structure of the data may not be appropriately emphasized. Conversely, if $b$ is too small, fluctuations in the histogram that may appear important could be the simple result of random data variation over a small interval.

Statistics literature has attempted to determine how to optimally choose $b$. Three well-known equations to estimate $b$ have been developed by Sturges (1926), Scott (1979), as well as Freedman and Diaconis (1981). These equations are not optimally applicable for all datasets, as they are typically applied for estimating $b$ (or equivalently $n$) when time constraints do not allow for empirical determination of this bin width. In general, if time permits, it is best to empirically choose $b$ to suit the dataset of interest.

A "Histogram" microseismic classification technique was developed. The histogram parameters were empirically chosen to best fit the given microseismic dataset. A total of $n = 99$ histogram bins, each of width $b = \frac{2}{99}$ were applied. The data limits were taken to be $d_{min} = -1$ and $d_{max} = +1$ to obtain a symmetrical histogram about the zero-mean signals. Referring to equation 13, histogram bin $k$ ($k = 0,1,2,\ldots,98$), represented by $m_k$, would thus record the total number points that fall in the range $-1 + \frac{2k}{99} \le d_i < -1 + \frac{2(k+1)}{99}$. Using the notation in equation 13 for this example gives

$$m_k \rightarrow \left[ -1 + \frac{2k}{99} \le d_i < -1 + \frac{2(k+1)}{99} \right]; \quad k = 0,1,2,...,98. \tag{14}$$

After histogram generation, the data occurrence recorded in a central bin (corresponding to $k \approx \frac{n}{2}$, roughly speaking) can be examined for microseismic signal classification.

If the 50$^{th}$ histogram bin ($k = 49$) is examined, the "good" trace depicted in Figure 1 has 34.6% of its time series data points lying in the 50$^{th}$ bin range, while the noise trace in Figure 2 only has 10.7% of its points in this range. Figures 11 and 12 show histogram plots corresponding to the "good" and noise traces depicted in Figures 1 and 2, respectively, plotted on consistent vertical axes for comparison. Note the higher central concentration of data points pertaining to the data point distribution of the "good" trace,

as compared to the noise trace, which shows data distribution spread out to a larger degree.

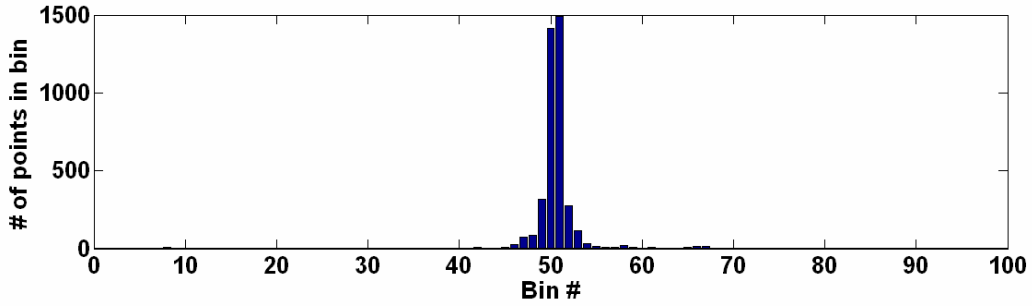


FIG. 11. Histogram plot of "good" trace depicted in Figure 1 with 99 histogram bins corresponding to amplitude ranges from -1 to +1.

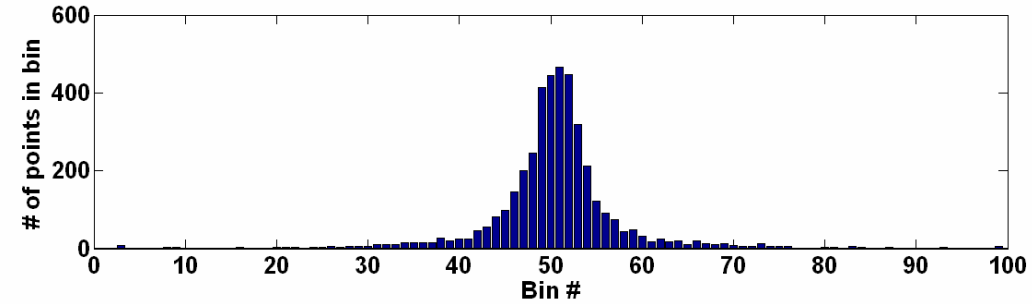


FIG. 12. Histogram plot of noise trace depicted in Figure 2 with 99 histogram bins corresponding to amplitude ranges from -1 to +1.

*Specialized Zero-Crossing Count algorithm*

I observed that microseismic noise signals tend to oscillate more frequently about the time axis. Additionally, I observed that the magnitudes of signed amplitude differences between adjacent time-series data points were generally greater for noise microseismic traces compared to "good" traces. Essentially, this observation implies that "good" traces generally contain minimal signal noise. Noise that was observed in microseismic "good" traces had relatively low amplitude compared to microseismic noise traces. Alternatively stated, it was observed that, compared to noise traces, "good" traces often have less sporadic sequential time-series behaviour about its mean.

To illustrate the above observations, the "good" and noise traces in Figures 1 and 2 are plotted as discrete time-series in Figures 13 and 14 from an example time interval of 1.14s to 1.15s. A very fine time interval is used in these figures to illustrate the signed amplitude difference between adjacent time-series data points corresponding to the example "good" and noise traces. Note that in Figure 14, corresponding to the noise trace, there are more instances where adjacent data points have opposite signs compared to the "good" trace in Figure 13. This is related to the observation that "good" traces oscillate less than noise traces. An additional important observation is that the magnitudes of signed amplitude differences between adjacent time-series data points are larger for the noise trace than the "good" trace, including areas where adjacent data

points change polarity. Based on these observations the "Specialized Zero-Crossing Count" algorithm was developed.

Define a window $v$ that exists in the signed-amplitude range $-z \le v \le z$. First, all the time-series data points that fall within $v$ are set to zero. Essentially, this zeroing step sets all data points close to the time axis to zero and removes low-amplitude noise. As the signed amplitudes of adjacent time-series data points in microseismic noise signals generally vary to greater degrees compared to "good" traces, this zeroing step will tend to preserve polarity reversals (sign changes) in adjacent data points on noise traces, but eliminate many of these polarity reversals on "good" traces. Thus, this step helps to further improve the discrepancy between "good" and noise traces to improve classification accuracy.

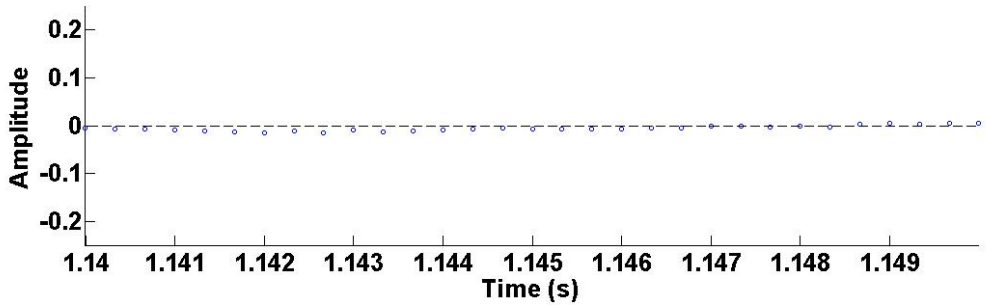


FIG. 13. Discrete time-series plot of example "good" trace for $t$ = 1.14s to $t$ = 1.15s. The horizontal line represents zero amplitude to better view polarity reversals. Fewer polarity reversals between adjacent data points are seen compared to the noise trace in Figure 14. Additionally, the magnitudes of signed amplitude differences between adjacent time-series data points are smaller compared to the example noise trace.

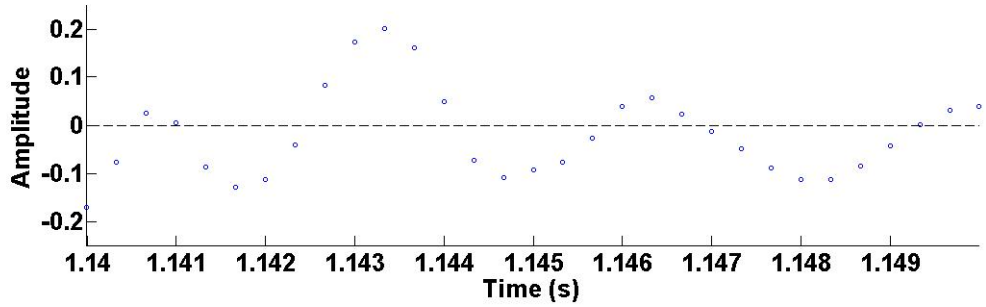


FIG. 14. Discrete time-series plot of example noise trace for $t$ = 1.14s to $t$ = 1.15s. The horizontal line represents zero amplitude to better view polarity reversals. More polarity reversals between adjacent data points are seen compared to the "good" trace in Figure 13. Additionally, the magnitudes of signed amplitude differences between adjacent time-series data points are larger compared to the example "good" trace.

After the zeroing step is applied, the total number of valid polarity reversals between adjacent data points is totaled, and that total is divided by the total number of trace data points to obtain a fractional measurement. A valid polarity reversal corresponds to adjacent data points changing from a strictly positive to a strictly negative value, or vice-versa.

With the zeroing step applied for z = 0.01, the "good" trace in Figure 1 had only one valid polarity reversal out of 4096 total data points, corresponding to a percentage of 0.0244%. The noise trace in Figure 2 had 298 valid polarity reversals corresponding to a percentage of 7.28%.

To illustrate the value of applying the zeroing step to eliminate low-amplitude noise present in the time series, consider the case where it is omitted. If this zeroing step is not applied, the "good" trace in Figure 1 would have 302 polarity reversals out of 4096 points (7.37%), while the noise trace in Figure 2 would have 550 polarity reversals (13.43%). Thus, omitting the zeroing step, only 1.82 times more polarity reversals would be seen in the noise trace compared to the "good" trace. With the zeroing step applied, 298 times more polarity reversals are seen in the noise trace compared to the "good" trace.

## MULTIVARIATE DATA REDUCTION

After extensive testing, it was found that the developed statistical analysis algorithms have the highest potential for maximally accurate microseismic file classification. These algorithms were the Threshold, Histogram, and Specialized Zero-Crossing Count methods.

It is desired to examine the outputs from these three statistical algorithms for accurate microseismic file classification. As these three algorithm outputs result in a three-dimensional data space, multivariate data reduction (i.e. reducing the effective dimensionality of the data from three to one or two) would simplify file classification. The multivariate data reduction technique employed is referred to as principal components analysis.

Principal components analysis (PCA) is a linear technique that transforms a dataset with many variables to a new set of variables that are orthogonal and uncorrelated, called the "principal components" of the dataset (e.g. Jackson, 1991). PCA can also be thought of as representing an $N$ dimensional dataset with $N$ orthogonal basis vectors (principal components) such that data projected onto the first principal component have the highest variance and thus best characterize the dataset (e.g. Shlens, 2003). The Threshold, Histogram, and Specialized Zero-Crossing Count statistical algorithms yielded the most accurate classification results. Thus, PCA was applied to the outputs of these three algorithms for improved classification accuracy. Restricting PCA to the three statistical algorithm outputs was seen to yield improved classification performance over applying PCA to all developed algorithm outputs.

Refer to literature authored by Dunteman (1989), Jackson (1991), Lupton (1993), Jolliffe (2002), Smith (2002), and Shlens (2003), for mathematical theory pertaining to

PCA. Application of PCA theory from this literature to microseismic file classification will be shown here.

To illustrate the application of PCA to microseismic file classification, a diverse example test dataset of 540 microseismic files from 28 different production pads will be used. The correct file classification is known, as these files were previously manually classified as "good" or noise.

Define a matrix A containing normalized data measurements pertaining to the statistical analysis algorithm outputs for this example test dataset. Normalization is a necessary step when applying PCA. In this case, normalization pertains to dividing each row of A by the largest value found in that row. The matrix A will be a $3 \times 540$ matrix whose rows correspond to different algorithm outputs, and whose columns correspond to different microseismic files. Assume that the normalized results from the Threshold, Histogram, and Specialized Zero-Crossing Count algorithms are placed in the first, second, and third rows of A, respectively. Each row of A contains the normalized average algorithm outputs corresponding to the three traces in the microseismic file that have the strongest "good" characteristics. For example, the element $a_{15}$ is the normalized average fraction of data points lying outside a predefined window (Threshold algorithm) corresponding to the three traces in the fifth microseismic file that contain the lowest fraction of outlying points. Figures 15, 16, and 17 show plots of normalized algorithm measurements contained in the first, second, and third rows of A, respectively.

Figure 15 shows the normalized Threshold algorithm output, which corresponds to the normalized fraction of data points that lie outside a predefined window. This window is defined from -0.006 to +0.006 in this example. Each gray square (black diamond) corresponds to the calculated and normalized Threshold algorithm output pertaining to a manually confirmed "good" (noise) file. The calculated and normalized results from the known "good" and noise files are plotted in overlapping horizontal regions to best view vertical overlap between the normalized "good" and noise results. Ideally, no overlap between calculated and normalized algorithm outputs of known "good" and noise files is desired to perform perfect classification. In Figure 15, reasonable vertical separation is seen between the "good" and noise points; however, vertical overlap can be seen. For example, in this figure there are 6 gray squares ("good" file outputs) that are above the lowest black diamond (a noise file output). From a different perspective, there are 32 black diamonds (noise file outputs) that are below the highest gray square (a "good" file output).

Figure 16 shows the normalized Histogram algorithm output, which corresponds to the normalized fraction of points that are contained in a "middle" data bin. In this example, 99 evenly spaced bins corresponding to amplitude ranges from -1 to 1 were created, and the normalized fraction of data points that fell within the range of the 51$^{st}$ bin is shown. Reasonable vertical separation is again seen between "good" and noise outputs; however, some vertical overlap still exists. There are 7 gray squares ("good" file outputs) that are below the highest black diamond (a noise file output). From a different viewpoint, there are 50 black diamonds that are above the lowest gray square.

Figure 17 shows the normalized Zero-Crossing Count algorithm output, which corresponds to the normalized fraction of signal polarity reversals after low-amplitude noise is removed. In this example, data points with a magnitude less than 0.006 were set to zero. Vertical separation is seen between "good" and noise outputs. Vertical overlap is present, however, as is seen with 68 gray squares ("good" file outputs) above the lowest black diamond (a noise file output). Additionally, 19 black diamonds fall below the highest gray square.

PCA aims to reduce the vertical overlap between normalized algorithm outputs pertaining to known "good" and noise files to obtain optimal microseismic file classification. This was achieved by projecting the measured data shown in Figures 15 to 17 onto the dataset's principal components. It has been shown in the cited literature that the principal components of a dataset are equivalent to the unit-length eigenvectors of the dataset's covariance matrix.

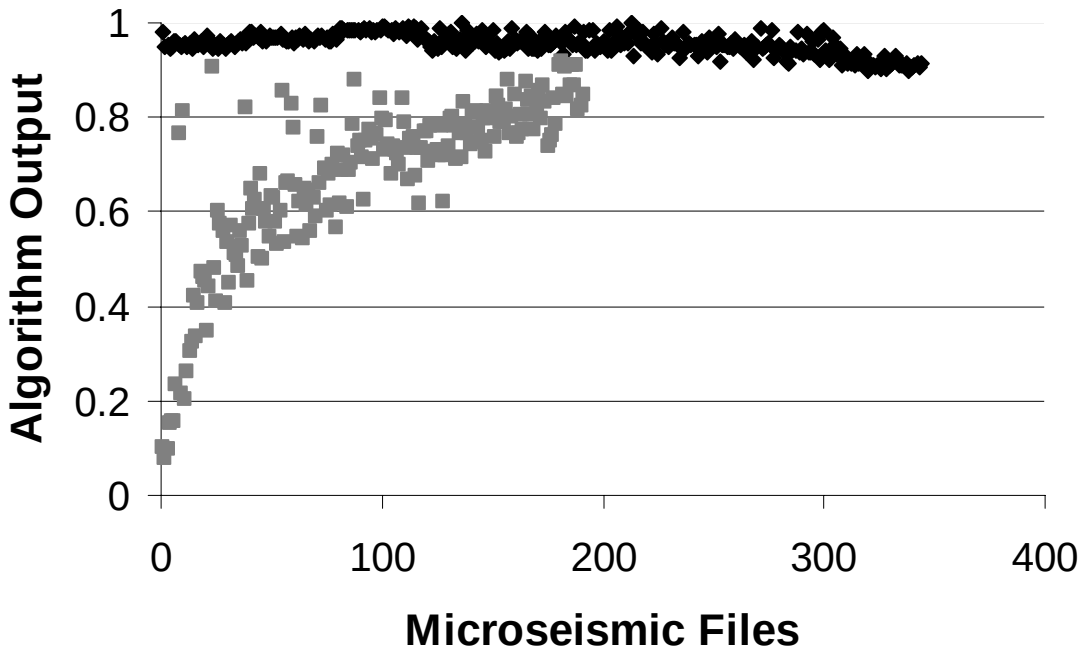


FIG. 15. Normalized Threshold algorithm output. There are 6 gray squares ("good" file outputs) above the lowest black diamond (a noise file output) and 32 black diamonds below the highest gray square.

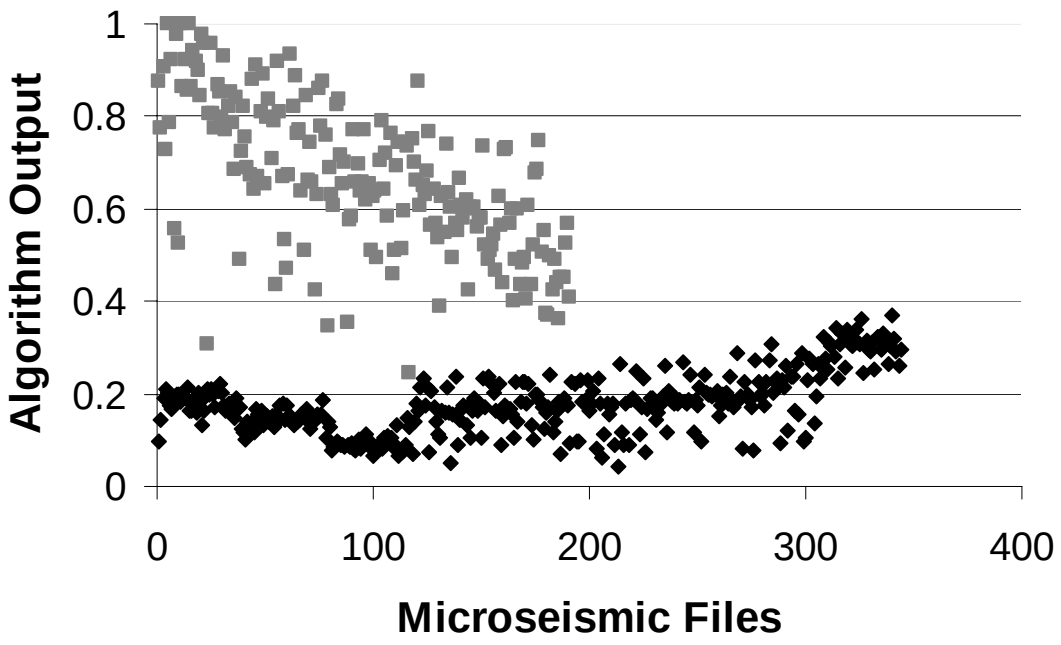


FIG. 16. Normalized Histogram algorithm output. There are 7 gray squares ("good" file outputs) below the highest black diamond (a noise file output) and 50 black diamonds above the lowest gray square.

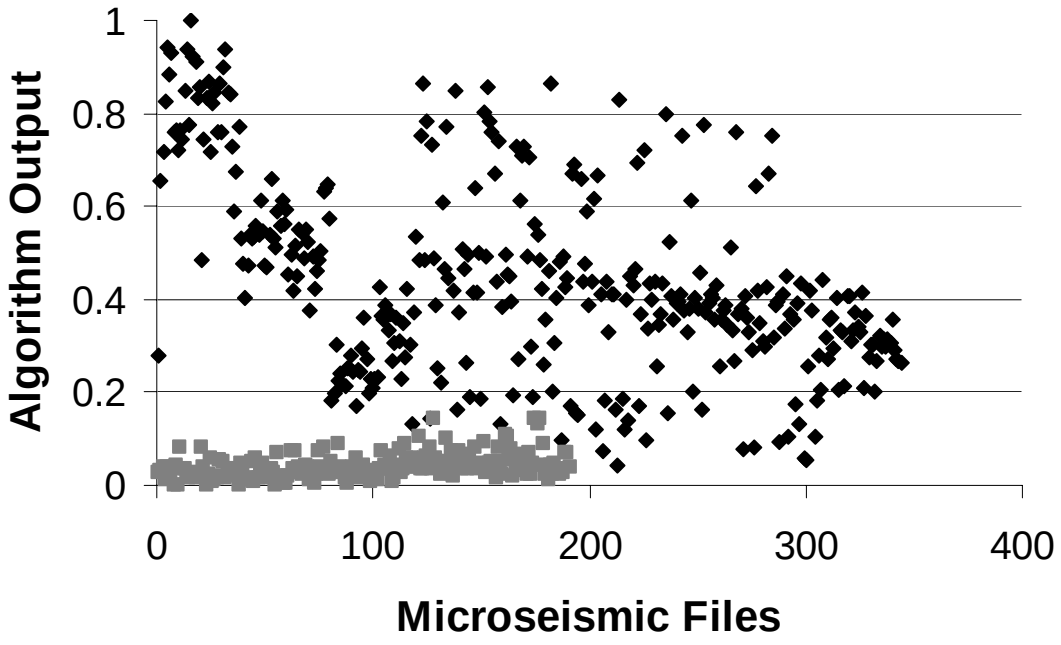


Figure 17. Normalized Zero-Crossing Count algorithm output. There are 68 gray squares ("good" file outputs) above the lowest black diamond (a noise file output) and 19 black diamonds below the highest gray square.

For example, let $c_{TT}$, $c_{HH}$, and $c_{ZZ}$ represent the calculated variance of the normalized Threshold, Histogram, and Specialized Zero-Crossing Count algorithm outputs, respectively, over all 540 files in the example dataset. Also let $c_{TH}$, $c_{TZ}$, and $c_{HZ}$ represent the calculated covariance between the normalized Threshold-Histogram, Threshold-Zero Crossing Count, and Histogram-Zero Crossing Count algorithm outputs, respectively, over all 540 files in the example dataset. It is also necessary to mean-correct the rows of the measurement matrix A to zero prior to formulating the covariance matrix. The symmetric $3\times3$ covariance matrix C of the dataset can then be defined as

$$\mathbf{C} = \begin{bmatrix} c_{TT} & c_{TH} & c_{TZ} \\ c_{TH} & c_{HH} & c_{HZ} \\ c_{TZ} & c_{HZ} & c_{ZZ} \end{bmatrix}. \tag{15}$$

The first row of C contains covariance values that involve the Threshold algorithm outputs; the second row contains values that involve the Histogram algorithm outputs; and the third row contains values that involve the Specialized Zero-Crossing Count algorithm outputs.

Thus, for this symmetrical $3\times3$ covariance matrix, there will be three corresponding unit-length orthogonal eigenvectors each representing a single principal component of the dataset. The eigenvalue of each eigenvector is equivalent to the data variance when projected onto the eigenvector. The eigenvector with the largest corresponding eigenvalue represents the first principal component of the dataset and is orientated in the direction of maximum data variance. The eigenvector with the second largest corresponding eigenvalue represents the second principal component of the dataset, and so on. For detailed mathematical theory pertaining to PCA, refer to the cited literature.

After mean-correction, the normalized measurement data contained in the matrix A was projected onto the principal components of this example dataset containing 540 microseismic files. Figures 18, 19, and 20 show this data after projection onto the first, second, and third principal components, respectively.

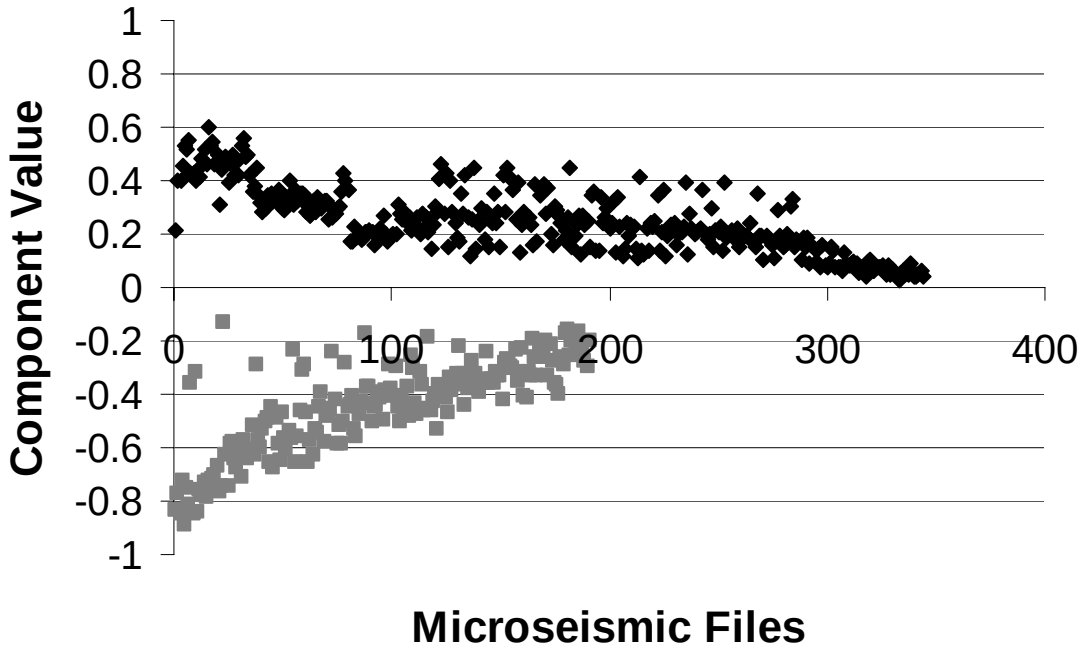


FIG. 18. Dataset projected onto first principal component. No vertical overlap is seen between "good" (gray squares) and noise (black diamonds) points. For this specific test dataset, perfect classification accuracy could be achieved using PCA and examining this component.

In Figure 18, no vertical overlap exists between the "good" and noise data points, which clearly suggests that file classification using Figure 18 corresponding to PCA application would be improved over attempting to empirically classify files with the normalized raw measurements shown in Figures 15 to 17. Figures 19 and 20 correspond to noise components in the data and are not useful for microseismic file classification. Thus, for this dataset, PCA has reduced the effective dimensionality from three to one

.

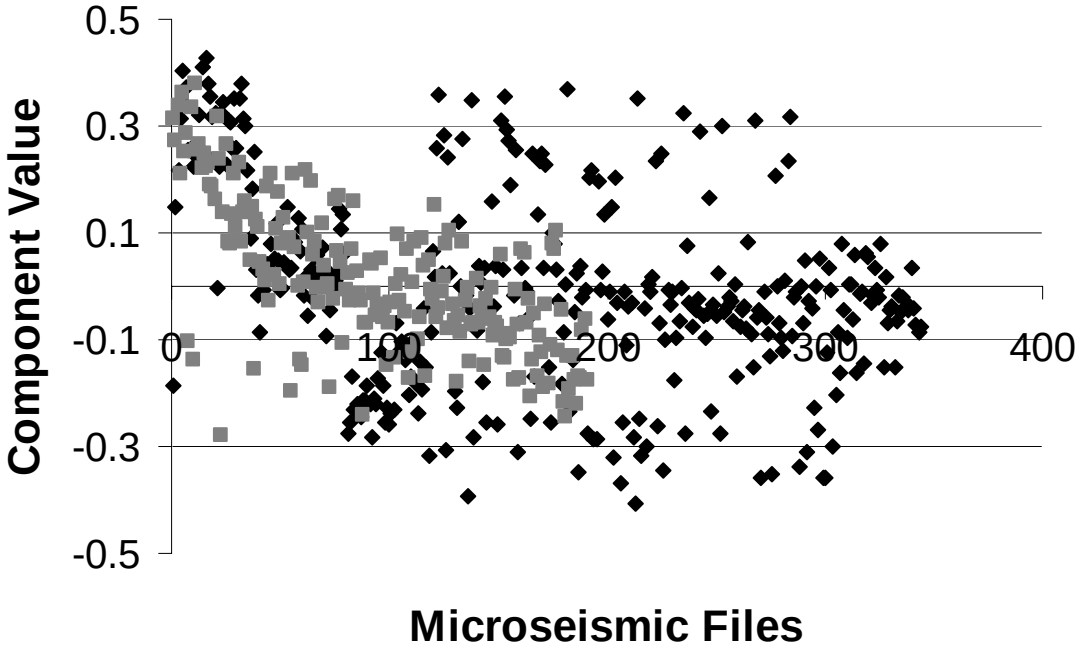


FIG. 19. Dataset projected onto second principal component. No useful information can be extracted in this example.

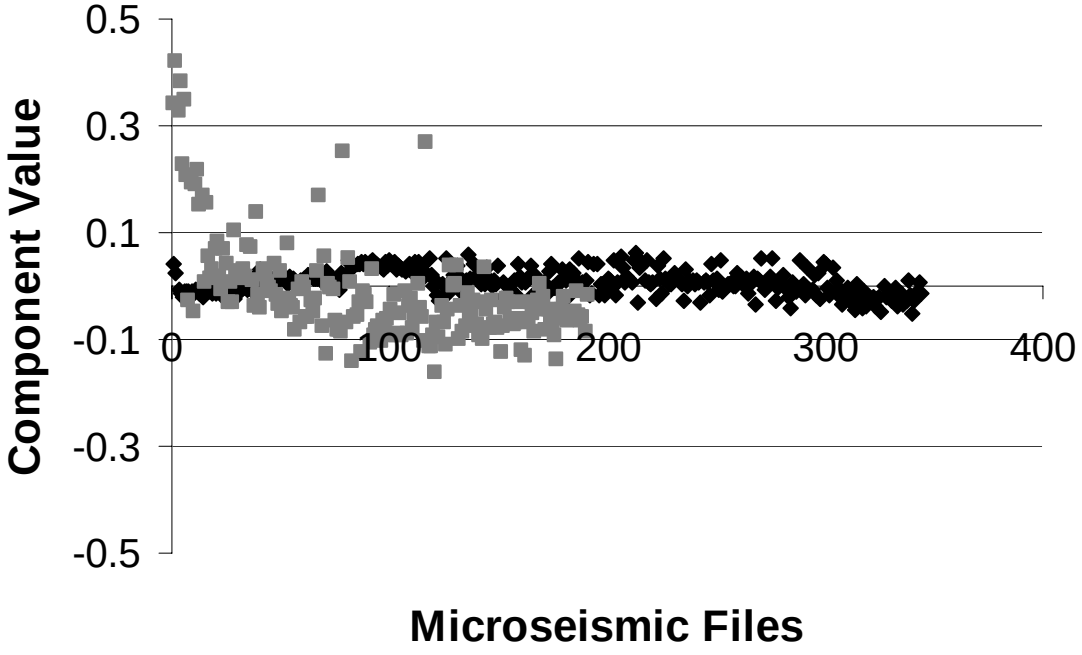


FIG. 20. Dataset projected onto third principal component. No useful information can be extracted in this example.

Figure 18 suggests that if PCA is used for multivariate data reduction, all 540 files from 28 different pads in this specific example dataset could be classified to perfect accuracy. This obviously will not always be the case for all datasets, but applying PCA does result in improved classification accuracy over simply observing individual algorithm outputs for classification.

## IMPLEMENTATION RESULTS

In addition to strong classification performance, the statistical analysis algorithms swere also found to be the most computationally efficient, a required characteristic when potentially classifying tens of thousands of microseismic files daily. I created a MATLAB® program implementation that applied principal components analysis to statistical algorithm outputs for file classification.

First, it was required to obtain principal components from a reference dataset. To ensure classification robustness, this dataset should be as diverse as possible. The statistical algorithm outputs pertaining to an incoming microseismic file to be classified were then projected onto the principal components obtained from the reference dataset. Depending on the principal components found from the reference dataset, data projected onto one or more of these principal components can be analyzed for microseismic file classification.

When the developed program implementation was applied to a specific dataset where most files originated from five production pads, a microseismic file classification accuracy of 99.5% was obtained. Testing on a more diverse dataset with files from 28 different pads yielded a classification accuracy of 98.8%. An exhaustive test on microseismic files from 72 different production pads resulted in a 90% accuracy. Given that up to tens of thousands of microseismic events are detected daily at Cold Lake, this application could have significant future impact.

## CONCLUSIONS

Microseismic signal classification algorithms involving frequency filtering, event-length detection, and statistical analysis were developed and tested on microseismic data from Cold Lake, Alberta, Canada. These algorithms were based on the observations that, compared to microseismic noise, “good” microseismic signals pertaining to events of interest generally contain lower dominant frequencies, shorter first-arrival event-lengths, lower signal variance, higher central data distribution, and less sporadic sequential time-series behaviour about its mean.

After extensive testing, we have seen that the developed statistical analysis algorithms perform best. A MATLAB® program implementation was created that applied principal components analysis, a multivariate data reduction technique to statistical analysis algorithm outputs. This application demonstrated classification accuracies between 90% and 99.5% when tested on various microseismic datasets from Cold Lake. Given that up to tens of thousands of microseismic events are detected daily at Cold Lake, this work could have significant future impact.

## ACKNOWLEDGEMENTS

We would like to thank all CREWES staff, students, and sponsors for their support throughout this project. We are also appreciative to Colum Keith, Richard Smith, and Sophia Follick from Imperial Oil, who provided us with microseismic data, performed a wide variety of testing on the developed program, and consistently provided valuable feedback.

## REFERENCES

Ambuter, B.P., and S.C. Solomon, 1974, An event-recording system for monitoring small earthquakes: Bulletin of the Seismological Society of America, **64**, 1181-1188.

Campbell, G., 2005, Velocity model improvements for use in Imperial Oil's microseismic analysis program: Imperial Oil Resources & University of Alberta.

Dunteman, G.H., 1989, Principal components analysis: Sage Publications.

Feichtinger, H.G., and T. Strohmer, 1998, Gabor analysis and algorithms: theory and applications: Birkhäuser.

Freedman, D., and P. Diaconis, 1981, On the histogram as a density estimator: $L_2$ theory: Probability Theory and Related Fields, **57**, 453-476.

Haykin, S., and B. Van Veen, 2003, Signals and systems, 2nd edition: John Wiley & Sons.

Jackson, J.E., 1991, A user's guide to principal components: John Wiley & Sons.

Jolliffe, I.T., 2002, Principal component analysis, 2nd edition: Springer.

Lange, K., 2003, Applied probability: Springer.

Lee, W.H.K., and S.W. Stewart, 1981, Principles and applications of microearthquake networks: Academic Press.

Lupton, R., 1993, Statistics in theory and practice: Princeton University Press.

Margrave, G.F., 2007, Methods of seismic data processing. Geophysics 557 course lecture notes, Winter 2007: The CREWES Project, University of Calgary.

Maundy, B., 2005, ENEL 559 course notes, University of Calgary.

Maxwell, S.C., and T.I. Urbancic, 2001, The role of passive microseismic monitoring in the instrumented oil field: The Leading Edge, **20**, 636-639.

Miller, S.L., and D.G. Childers, 2004, Probability and random processes with applications to signal processing and communications: Elsevier.

Mitzenmacher, M., and E. Upfal, 2005, Probability and computing: randomized algorithms and probabilistic analysis: Cambridge University Press.

Munro, K. A., 2005, Analysis of microseismic event picking with applications to landslide and oil-field monitoring settings: MSc. thesis, University of Calgary.

Scott, D.W., 1979, On optimal and data-based histograms: Biometrika, **66**, 605-610.

Sheriff, R.E., and L.P. Geldart, 1995, Exploration seismology, 2nd edition: Cambridge University Press.

Shlens, J., 2003, A tutorial on principal component analysis: derivation, discussion, and singular value decomposition: University of California, San Diego, electronic PDF document.

Smith, L.I., 2002, A tutorial on principal components analysis: University of Otago, New Zealand, electronic PDF document.

Stockwell, R.G., L. Mansinha, and R.P. Lowe, 1996, Localization of the complex spectrum: The *S* Transform: IEEE Transactions On Signal Processing, **44**, 998-1001.

Sturges, H., 1926, The choice of a class-interval: Journal of The American Statistical Association, **21**, 65-66.

Suhov, Y., and M. Kelbert, 2005, Probability and statistics by example: Cambridge University Press.

Tan, J.F., H.C. Bland, and R.R. Stewart, 2006, Passive seismic reservoir monitoring techniques applied to heavy oil production: CREWES Research Report, **18**, University of Calgary.

Therrien, C.W., and M. Tummala, 2004, Probability for electrical and computer engineers: CRC Press.

Zhou, H., and G.A. McMechan, 1999, Parallel Butterworth and Chebyshev dip filters with applications to 3-D seismic migration: Geophysics, **64**, 1573-1578.