Interactive data visualization for microseismic characterization in hydraulic fracturing domain

Michael J. Nava, Civil Systems Department of Civil and Environmental Engineering University of California, Berkeley

Abstract

Hydraulic Fracturing processes are very large and increasingly complex endeavors. The accompanying dataset generated by surveying techniques is equally expansive. As such, there is a growing difficulty in processing the data and interpreting the information in a concise and organized manner. Moreover, there is a disparity between academic software and proprietary computing techniques that curb academic advancement in this research area. The purpose of this project is to bridge the gap between academia and high-cost, proprietary software through interactive data analysis. Specifically, a suite of linked graphical user interfaces is created with the domain of hydraulic fracturing in mind. The graphical user interfaces allow the user to interactively analyze the data through visualizations in three main views. First, traditional analysis of seismograms is improved by allowing the user to quickly analyze raw data, normalized data, and filtered data. Second, locations of microseismic data are visualized in three-dimensions that allow the user to show events sorted by multiple process parameters like magnitude, distance, and stage. Finally, the frequency domain can be analyzed by displaying the average event spectra in order to gain a richer understanding of microseismic source mechanisms.

Introduction

Containing over one thousand trillion cubic feet of natural gas, the Marcellus shale is said to be one of the largest shale plays in the world (Duncan, 2010). As such, there has been a recent surge in hydrocarbon exploration in the area, most notably through hydraulic fracturing processes. An important requirement of this exploration is the identification and characterization of microseismic events. Until recently, the emphasis of microseismic imaging has been on classifying the location of events (Maxwell et al., 2010). However, the location alone, while useful, has limited applicability when attempting to understand the subsurface fracture mechanisms, the formation of fracture networks, the reactivation of existing faults, and the overall quality of production.

In order to go a step further, it is necessary to consider more than simply event location and investigate traditional views of seismic data as well as spectral content. The difficulty in doing so is that it is challenging to analytically characterize these data. As a result, it is necessary to turn to visualization techniques for data analysis.

Hydraulic Fracturing: A Brief Overview

Hydraulic fracturing is a technique employed in order to retrieve hydrocarbon-rich material like natural gas and oil from rock formations – typically shale – by drilling and pressurizing horizontal boreholes (Hubbert, 1976). There are generally at least two boreholes drilled: a treatment well and an observation well. The treatment well is pressurized in order to cause fractures in the shale formation, while the observation well is not pressurized. Instead, the observation well is used to house geophones, usually arranged in a linear array. Geophones are sensitive monitoring devices that measure acoustic emissions from fracturing events. Specifically, geophones measure ground displacement as a function of time and the output of these sensors can be read on a seismogram (Hoshiba, 1993).

Both the treatment well and the observation well are sealed with thick layers of cement casing in order to prevent negative environmental effects. Fracturing fluid, which is typically water that contains sand, or proppant, flows into the treatment well and the pressure is increased until fractures occur in the target formation. The purpose of proppant is to remain embedded in the fractures in order to keep them open so that oil and natural gas can flow through.

The main advantage to hydraulic fracturing over conventional methods of retrieving hydrocarbon-rich materials is that hydraulic fracturing processes can occur in formations with very little porosity and permeability. As such, there are significantly more possible sites for

hydrocarbon retrieval. However, there are also disadvantages to this approach. For example, since the fractures are directly related to hydrocarbon retrieval and those fractures occur non-deterministically, there is a relatively high amount of uncertainty regarding overall success. Moreover, analysis of the data is difficult due to a number of factors including the sheer volume of data. This project attempts to aid in that regard.

Motivation

While sophisticated processing and analysis techniques exist, one of the largest difficulties in microseismic analysis of hydraulic fracturing processes is data handling. For example, a single hydraulic fracturing process can last a number of days. The entire process is continually monitored by sensors at the surface, typically measuring surface pressure, slurry flow rate, and proppant concentration. Additionally, downhole geophone arrays monitor acoustic emissions from nearby fracture events. All of these sensors are broadband monitoring tools and have high sample rates. As such, it is not uncommon to generate more than 100 gigabytes of data from a hydraulic fracturing process.

Despite the large volume of information generated, the difficulty in interpreting and understanding the data stems from the disparate nature of the datasets. For example, since there are a number of different sensors, or sensor arrays, tasked with monitoring a specific aspect of the hydraulic fracturing process, establishing relationships between the datasets is quite challenging. Moreover, the available software packages for the academic researcher are fairly limited. Conversely, there are a number of very powerful software packages in industry; however, processing techniques employed in this type of software are typically proprietary and highly guarded. As such, there is a growing disparity between academic visualization packages and proprietary software.

Although there is a deficiency of available software for the academic researcher, there is certainly not a total absence of functional packages. For example, many in academia focusing on seismic research tend to use Matlab, which is a powerful computing platform specializing in matrix manipulation. Since data from geophone arrays, typical in seismic imaging techniques, are generally stored as matrices, Matlab is the logical choice for hydraulic fracturing analysis as it allows for fast computing of large data in this form (Lovett et al.,2001). There are a number of available toolkits for Matlab that enable the reading of seismic file types.

Current Software Packages and Limitations

CREWES (Consortium for Research in Elastic Wave Exploration Seismology) has a downloadable toolkit that enables the user to read in multiple types of seismic data – SEG2, SEGY, etc. – and perform complex seismic analysis techniques by calling pre-built functions (Crewes.org, 2014). SegyMAT and SeisMat are similar toolkits that provide the ability to read in seismic data and perform more basic processing and plotting through the use of bundled functions (SegyMAT.sourseforge.net, 2014).

While all of these toolkits are useful in converting specialized seismic file types to readable data types, there is a limitation regarding hydraulic fracturing process interpretation. This is due to the fact that all of these toolkits were created for more conventional seismic methods – earthquake analysis, near-surface exploration, as well as other geophysical approaches.

Moreover, there is a fundamental difference between traditional seismic imaging techniques and those used in the hydraulic fracturing arena. Consider for example, a typical earthquake that occurs and is monitored by a sparse network of stations containing tri-axial sensors (much like those used in downhole monitoring of hydraulic fracturing processes). A single seismic event, or earthquake, occurs and the acoustic emissions from that event are carried large distances. These emissions are captured by seismic monitoring stations that are statically deployed throughout the world (Figure 1).



Figure 1: Map view of monitoring stations used to record a seismic event (earthquake) off the coast of Northern California. The event is shown as the red and white circle and the monitoring stations are represented by the inverted triangles. Note the large distances and wide azimuthal distribution of the monitoring stations.

From here, a single "trace" is generated that describes the ground motion at each monitoring site. Multiple traces from monitoring stations (Figure 2) are used in order to define the fracture mechanism of the event – what type of fault caused the earthquake, what the magnitude of the event was, and what the directionality of the movement was. All of these things help classify the seismic event through a well-formed analytical technique called moment tensor inversion.



Figure 2: Multiple traces capturing a large earthquake that occurred off the coast of Northern California in 2014. This is an example of a complete dataset used to interpret a single seismic event.

Monitoring hydraulic fracturing processes is abundantly different. For example, where there was a single event in the conventional example, there are thousands of events in the hydraulic fracturing case. Where the magnitude of a typical earthquake can range from 3 M_W to 7 M_W or higher, the typical events in a hydraulic fracturing process can be as low as -1 M_W to -5 M_W (magnitude is measured on a log scale so this an exceedingly large difference). This lower magnitude becomes an issue due to the increasing impact of noise on data quality. Where a typical earthquake can be monitored with stations at varying distances, and azimuth, hydraulic fracturing events are monitored by a linear array usually positioned in a monitoring well that parallels the fracturing borehole (Figure 3). This geometrical constraint is a significant limitation since it prevents a rich understanding of source mechanism. This is a consequence of survey geometry, as it does not allow for an accepted analytical technique known as moment tensor inversion.



Figure 3: Map view of a hydraulic fracturing process. Treatment well is shown in blue. Observation well is shown in red. Events are colored to distinguish events by stage. The geophone array locations are shown by multi-colored stars on the red observation well. Note the relatively short distance and also note that the geophones are directly across from the treatment well. This leads to a narrow aperture and subsequent analysis limitations.

The result of all these limitations and deficiencies leads to only one thing – the need for a data analysis package specifically designed to allow the user to interact with large amounts of seismic data from a hydraulic fracturing process.

Approach

The purpose of this project is to bridge the gap between academia and high-cost, proprietary software through interactive data analysis. Specifically, by incorporating processing steps that are designed for the unique domain of hydraulic fracturing seismic imaging, the user is now able to analyze a large portion of the data processing by a button click. Rather than relying on traditional Matlab plotting functions, which save static images for later reference, the interactive graphical user interface enables real-time comparison of seismic data in a number of interrelated views.

The main control panel allows the user to perform highlevel processing and visualization steps. Figure 4 shows the main interface, which contains six buttons that allow the user to access all aspects of the software package.



Figure 4: Home screen of graphical user interface. All aspects of the software package are accessible from this panel. Each of the highlighted buttons opens a new, resizable graphical user interface so that the user can position the windows in order to best suit his/her specific needs.

Read Data

This uses the pre-configured Matlab toolkit, SegyMAT, in order to read in the data. The test dataset used Society of Exploratory Geophysicist's SEG-Y data format. However, CREWES can also be implemented to read other data types like SEG-D or SEG2. These pre-configured Matlab toolkits read the specialized files and store the values in a cell array. Trace and file headers are also stored as structures for follow-on computation. Once the data are in a format that is compatible with Matlab, this aspect of the system is complete. Next, the user can process the data.

Process Data

This portion of the software package performs a number of processing techniques and stores the data for interpretation through interactive visualizations. Specifically, the raw data is read in and stored for examination and interpretation. Raw data is important since it gives the most complete view of the events. While this is an important aspect of interpretation, the raw data also has the potential to occlude a number of important attributes of the waveform. As such, the data are processed with a number of objectives in mind.

First, the data are filtered by a standard Low Pass Butterworth filter with a cutoff frequency of 150Hz. Next, there is a normalization schema employed that aims to remove noise that is specific to each location of the geophone array. Finally, the data are stored in order to be displayed for interpretation. Although this approach is included in an attempt to minimize external effects, it may also be valuable to simply examine the data with various filters applied. As such, the data are also filtered by standard Low Pass and High Pass Butterworth filters with cutoff frequencies at 10Hz, 50Hz, 100Hz, 150Hz, and 200Hz. This is done in attempt to allow the user to customize his/her own approach for data analysis.

After the normalization of data, a Fast Fourier Transform (FFT) is performed on each trace in the dataset. The FFT is a digital adaptation of the analytical Fourier Transform and is defined as:

$$X(k) = \sum_{i=1}^{N} x(i) \omega_N^{(j-1)(k-1)}$$
(1)

where,

$$\omega_N = e^{(-\pi i)/N}$$

There are eleven geophones that monitor the hydraulic fracturing process simultaneously; as such, there are eleven spectral responses per event. Attempting to visualize each of the eleven spectral responses for more than 1,200 events creates a very cluttered image. As such, a two-dimensional average of the spectral responses was used as a representative measure of the spectrum of each event.

Another useful aspect of the data processing is sorting by independent parameters like stage, event magnitude, distance from event to treatment well, and event time. This sorting allows the user to visualize event location and spectral response based on each of these parameters.

With the data processed, the visualizations can now be generated quickly in order to produce a frame rate high enough for interactivity to occur smoothly (Hochheiser, 2004). From here, the user can choose his/her own workflow. The graphical user interface provides three different views of various aspects of the data – seismograms show the seismic traces and can be viewed, the locations of microseismic events are shown and can be sorted, and finally, average spectral responses of events are shown and can also be sorted according to independent parameters.

Seismograms

At the most basic level of data analysis, seismograms, which are records of ground motion due to acoustic emissions from rock fractures as a function of time, allow the user to investigate a number of phenomena. For example, the event in Figure 5 shows the seismogram from raw data. Here, a first arrival at approximately 0.25 seconds can be seen with a normal moveout, which would enable a basic estimate of the event location. Toward the bottom of the figure, large amplitude, high frequency

ringing can be seen. The user can select a "tag" listed at the bottom of the graphical user interface in order categorize pre-defined aspects of the event in order to later sort the data by these parameters. In this case, the user should select "Ringing" to characterize this event in its current view.



Figure 5: Seismogram showing raw data from an event that contains a number of interesting phenomena. From this view, high amplitude and frequency wavelets can be seen throughout the record.

Going further, by clicking the "Normalize Data" radio button, the user can view the same event, but postprocessed. The difference, shown in Figure 6, is quite dramatic.



Figure 6: Seismogram of location-specific, normalized data. It can be seen that the high frequency ringing is no longer present, indicative of the presence of tube waves and poor geophone coupling. Also visible is a lower amplitude secondary event.

For example, the large amplitude, high frequency ringing both at the bottom of the record and below the first arrival is no longer present. This is an indicator of the presence of tube waves. A tube wave is a compressional wave that travels radially through the observation well. These waves are evident on seismograms when there is poor coupling between the geophone and the observation well. This realization is quite important since coupling issues can lead to spurious results, incorrect estimates of event location, and a mischaracterization of microseismic source mechanisms. Additionally, there seems to be a secondary event (circled in red) that occurs at nearly the same time, but from a very different location.

This secondary event is even more prominent at lower frequencies, shown in Figure 7, indicating that the event may be at a greater distance than the primary event (Romanowicz, 1984). The user may now select the "Revisit" tag at the bottom of the window for later sorting since there are multiple phenomena in this record.



Figure 7: This seismogram is of the same event, but the data are only low pass filtered without normalization. It can be seen that the secondary event is more pronounced, indicating that there is greater content from lower frequencies.

The advantage of incorporating these tags for microseismic attribute identification is that as the user examines the data, which can be done over the course of days or months, he/she can keep record of interesting phenomena in the data. The data can later be sorted in order to determine if there is any relationship between those attributes and other parameters. For example, it may be possible to determine if ringing is associated with a specific azimuthal distribution or perhaps there is some correlation between magnitude and ringing. These are questions still to be resolved; however, with this capability in place, those answers may come sooner than previously possible.

At the bottom right of the window (Figure 7), two rectangular toggle buttons can be seen. The top button allows the user to view the current event in the *Event Locations* interface. The bottom button allows the user to see the current event's spectral response. Through the use of linking, the data is shared between graphical user interfaces. This is valuable because it allows the user to toggle between the different views. Moreover, it allows the user to get real-time feedback regarding possible relationships between process parameters. For instance, if the user notices a number of seismograms have a similar quality and would like to see if they share a similar distance from the treatment well or have a similar spectral response, he/she can confirm whether or not that is the case with a button click.

Event Locations

Another critical component to understanding hydraulic fracturing processes is the identification and classification of microseismic event locations. The distribution of events is one of the most important aspects of hydraulic fracturing because the purpose of the overall process is to increase porosity and permeability in the targeted shale formation.

Porosity and permeability are important because both parameters give an indication of production potential. The overall goal of a hydraulic fracturing process is to create new areas of pore space (porosity) through which hydrocarbons can flow (permeability). This is necessary in order to successfully retrieve oil and natural gas from the shale formation.

While there is still a significant limitation in fully understanding specific values of porosity and permeability within hydraulic fracturing processes, information regarding these parameters can be gleaned from the knowledge of event locations. Although it is difficult to say with confidence whether event locations that are closely grouped are, in fact, connected, it is acceptable to infer at least a first order estimate of the fracture network.

Figure 8 shows event locations for the entire dataset. It is clear that although all the data is shown in this view, it is quite difficult to understand specific details. Despite this limitation, it is evident that the distribution of event locations toward the right side of the window is not as closely grouped.



Figure 8: Event locations for all events in the dataset. Although this image is difficult to interpret when looking for specific events, the advantage is that it shows the events at the "toe" of the treatment well are more widely distributed. Lower magnitude events are shown in red, higher magnitudes are shown in orange.

An advantage of the interactive graphical user interface is that it gives the user the ability to sort the events by stage, event magnitude, distance from the treatment well and also by time. Additionally, the full capabilities of Matlab's three-dimensional graphics renderer are used in this interface. As such, the user can rotate the data (Figure 9), zoom in, and even manipulate the color mapping used in order to identify various aspects of the data.



Figure 9: Rotated and zoomed view of the event locations, sorted by magnitude. It can be seen in this view that the events are clustered above the treatment well. Lower magnitude events are shown in red, higher magnitudes are shown in orange. This uses a standard color encoding within Matlab; however, the user is able to change and customize the encoding.

Furthermore, since all of the interfaces are resizable, the user can control the size of the image, which allows extreme zooming. This is important in order to understand the distribution of events based on external parameters. For example, sorting the events by only showing the highest magnitudes, then rotating and zooming in reveals that there may be some correlation between high magnitudes and perhaps a pre-existing fault located close to the treatment well.

As before, the user can select the toggle button at the bottom right of the screen to view the selected events in the Spectral Analysis interface. While this shares the same basic functionality as the Seismogram interface, now the use can view the spectral response for a range of sorted data. That is, if the user were investigating the locations of events with magnitudes ranging from -2.86 Mw to -1.14 Mw, he/she can also investigate the spectral response for that same range of magnitudes with a button click. This capability may reveal the case where events sharing a similar spatial grouping may also have similar bandwidth in the spectral domain. This is important because that particular combination may reveal an area that has the potential for fault reactivation - a negative effect of hydraulic fracturing, which may lead to contamination of the water table.

Spectral Analysis

Finally, another important aspect of analysis is consideration of spectral content in event data. One of the main difficulties of monitoring hydraulic fracturing processes is the limited aperture that comes as a direct result of the survey geometry. Since, unlike conventional seismic imaging endeavors, the geophones are located in a relatively condensed area and in close proximity to the microseismic events, it is nearly impossible to use moment tensor inversion techniques to understand the source mechanism of fracture events. As such, an attempt is being made to use the information contained in the frequency domain in order to understand source mechanisms of fracture events. The spectral content can be seen in Figure 10 for the entire dataset.





In order to better understand the spectral content, the graphical user interface enables the user to sort by stage, magnitude (Figure 11), distance from treatment well (Figure 12), and by gas production. This capability enables the user to focus on a specific range of parameter values in order to correlate physical attributes of the process to spectrum. In other work performed on the test data, it was postulated that there is a correlation between change in surface pressure and a drop in frequency. Through the use of the interface, it may be possible to confirm this hypothesis by considering other parameters along with event spectrum.

Considering Figures 11 and 12, it can be seen that when sorted by magnitude, there is a clustering of broadband events at lower magnitudes. This may give some insight into the fracture mechanism of lower magnitude events.



Figure 11: Spectral content of all events sorted by magnitude. There is a subtle shift in broadband events at lower magnitudes, which may give insight into source mechanism.



Figure 12: Spectral content of all events sorted by distance from treatment well. It can be seen that there is slightly more high-frequency content at greater distances from the treatment well.

While there is certainly more analysis that must be done in order to fully understand the relationship between frequency content and other parameters like magnitude and distance, there can already be inferences made from basic sorting. Zooming in and changing the color mapping to focus on peak frequencies or broadband signals may also aid in interpretation.

Export Workflow

In any exploration process, it is important to be able to reproduce results. One valuable aspect of the graphical user interface that is still in the planning phase is the ability to keep record of analysis steps for later recall. Given that the dataset is so large, this capability will enable a user to keep a history of the event he/she investigates, which will enable the user to revisit events with a new perspective hours or days later. Moreover, keeping a record of the analysis steps will enable the user to export a workflow diagram showing the steps taken. This will be valuable for reporting and transparency.

Results

A prototype version of this graphical user interface was used to analyze a set of test data. Within a very short time, it was evident that the ability to quickly switch between visualizations and to progress from investigating seismograms to location information and finally spectral responses was an invaluable asset. In the past, the method in which visualizations were shared between researchers was in the form of multiple folders containing static images. For the test data, in order to capture seismograms showing raw data, normalized data, and filtered data, the result would be three folders containing more than 1,200 images each. Comparing multiple views at the same time was incredibly cumbersome and dramatically inhibited the visual investigation of seismic data.

The addition of the graphical user interface enabled what would have taken days in the past to occur nearly instantaneously. In a single window, seismograms were compared, subtle differences were noted, and large discoveries occurred. The use of the *Event Locations* interface enabled sorting of event locations by magnitude and other parameters. Although the test data had been visualized in the past, this had not been done. As a result of redundant encoding of visual variables, correlations between magnitude and location have been noted. Additionally, cross-referencing sorted datasets will allow for further intuitive leaps that may occur where analytical solutions do not exist.

The advantage of interactivity in data analysis through visualizations is significant. Due to the amount of data generated as a result of hydraulic fracturing processes, it is difficult to create static images that capture the macro view as well as showcase specific phenomena (Zockler et al., 1996). However, through the use of user-defined views, configurable visualizations enable the user to begin by examining the high-level data and then incrementally progress through the data. The result is a more rich understanding of the information presented through the visualizations.

Moreover, although there has been a large amount of work in defining analytical solutions to seismic phenomena, the fact remains that some problems are not well formed and cannot be adequately understood through these traditional approaches (Aki, 2009). This is especially true in the hydraulic fracturing domain. Due to a number of factors like heterogeneity of the target formation and surrounding rock structures as well as vertical anisotropy, traditional solutions simply fail to adequately characterize the subsurface mechanisms. As such, a new methodology is needed to understand hydraulic fracturing.

Future Work

Even in the testing phase of a prototype version of the graphical user interface, significant advances in the analysis and interpretation of the test data were made very quickly. As such, this implementation will continue to be improved.

An important aspect of data analysis that has not been mentioned is the picking of first arrivals in seismograms. The first arrivals are simply the time at which the fastest wave reaches the geophones. This is a fundamental aspect of seismic analysis and is used to understand the type of wave that is being monitored, the speed and direction at which it is propagating through the medium, and also aids

in event location determination. There are automatic picking algorithms that are widely accepted; however, there is an accompanying loss of confidence due to inherent error (Sleeman, 1999). As such, another functional aspect that will be added is the ability to manually pick these first arrivals and to incorporate those changes automatically for subsequent analysis.

Moreover, continuing with the idea of using microseismic attribute tags for characterization, another addition will be an area for the user to input comments for each event. This would store a string of data in event headers for later recall. While this may seem like a basic function, it will be very valuable for the user to be able to recall specific thoughts later in the analysis steps. Given the amount of data involved in hydraulic fracturing processes, it is not at all unlikely that the analysis steps will be performed in multiple sessions with some time in between. Being able to review one's comments before continuing the analysis will enable fluency in the analysis approach.

References

Aki, K., & Richards, P. (2009). *Quantitative seismology* (2nd ed.). Sausalito, CA: University Science Books.

CREWES Matlab Toolbox. (n.d.). Retrieved October 2, 2014 from http://www.crewes.org

Duncan, P. and Eisner, L. (2010). "Reservoir characterization using surface microseismic monitoring." GEOPHYSICS, 75(5), 75A139–75A146.

Hochheiser, H., & Shneiderman, B. (2004). Dynamic query tools for time series data sets: timebox widgets for interactive exploration. *Information Visualization*, *3*(1).

Hoshiba, M. (1993). Separation of scattering attenuation and intrinsic absorption in Japan using the multiple lapse time window analysis of full seismogram envelope. *Journal* of Geophysical Research: Solid Earth (1978–2012).

Hubbert, M. K., & Willis, D. G. (1972). Mechanics of hydraulic fracturing.

Lovett, T., Monti, A., Santi, E., & Dougal, R. A. (2001). A multilanguage environment for interactive simulation and development controls for power electronics. In Power Electronics Specialists Conference, 2001. PESC. 2001 IEEE 32nd Annual (Vol. 3, pp. 1725-1729). IEEE.

Maxwell, S., Rutledge, J., Jones, R., and Fehler, M. (2010). "Petroleum reservoir characterization using downhole microseismic monitoring." GEOPHYSICS, 75(5), 75A129–75A137.

Romanowicz, B., & Guillemant, P. (1984). An experiment in the retrieval of depth and source mechanism of large earthquakes using very long-period Rayleigh wave data. Bulletin of the Seismological Society of America, 74(2), 417-437.

SegyMAT : Read and Write SEG-Y files using Matlab and Octave. (n.d.). Retrieved August 1, 2012, from http://segymat.sourceforge.net

Sleeman, R., & van Eck, T. (1999). Robust automatic Pphase picking: an on-line implementation in the analysis of broadband seismogram recordings. *Physics of the earth and planetary interiors*, *113*(1), 265-275.

Zockler, M., Stalling, D., & Hege, H. C. (1996, October). Interactive visualization of 3D-vector fields using illuminated stream lines. In Visualization '96. Proceedings. (pp. 107-113). IEEE.