Focal Mechanism Determination using High Frequency, Full Waveform Information

Junlun Li*, Haijiang Zhang, H. Sadi Kuleli and M. Nafi Toksöz
Earth Resources Laboratory
Department of Earth Atmospheric and Planetary Sciences
Massachusetts Institute of Technology
Cambridge, MA 02139

Abstract

In this research, we use high frequency waveform information to determine the focal mechanisms of small local earthquakes at an oil reservoir. During the waveform inversion, we maximize both the phase and amplitude matching between the observed and synthetic waveforms. In addition, we use the polarities of the first P-wave arrivals and the S/P amplitude ratios to better constrain the matching between the synthetic and observed waveforms. The objective function is constructed to include all four criteria. Due to the complexity in the objective function, it is almost impossible to directly perform an inversion with derivative techniques. Instead, an optimized grid search method is used to search over all possible ranges of fault strike, dip and rake, as well as a predetermined range of earthquake locations. To speed up the algorithm, a library of Green's functions is pre-calculated for each of the moment tensor components and possible earthquake locations. Careful optimizations in filtering and cross-correlation are performed to further improve the grid search algorithm, such that no filtering and cross correlations are performed in searching through the parameter space of strike, dip, and rake. Consequently, speed is boosted tenfold by these optimizations in filtering and cross correlation.

We apply the new method to induced seismic events in an oil reservoir. Satisfactory matching between synthetic and observed seismograms is obtained, as well as reasonable focal mechanisms, considering the local geological structure and possible causes for induced seismicity.

Introduction

Induced Seismicity is a common phenomenon in oil reservoirs when production activities are in progress. It is mainly caused by water injection and gas and oil extraction, which change the stress distribution (Sarkar, 2008). By studying the patterns of the induced seismicity (e.g. location and focal mechanism), a time-lapse history of the stress changes in the oil reservoirs may be reconstructed and the influence of oil production on reservoirs may be better understood.

Because the monitoring networks at the oil reservoirs are usually sparse, it is very challenging or impossible to use only polarity information to constrain the focal mechanism of the induced seismicity (Hardebeck and Shearer, 2002). Some previous research used waveform information (Zhao and Helmberger, 1994; Zhu and Helmberger, 1996; Tan and Helmberger, 2007); however, they all used dense station network with low frequency range, and only part of the waveform information was utilized in the inversion. Sze (2005) used only P wave information and a wavelet-based approach to invert for focal mechanisms from a sparse network. Our study uses high frequency, full waveform information to constrain the focal mechanism for induced seismicity. We use the known velocity structure to calculate the Green's functions for all moment tensor components of the source and for each location (hypocenter). To compare these with observed seismograms we design an objective function, which incorporates various information: the cross correlation values between the synthetic waveforms and the data, the L_2 norms of the

waveform differences, the polarities of the first P arrivals, and the S/P ratios. In this way, we can find robust solutions to the focal mechanisms of induced seismicity by a sparse shallow surface network in an oil field, even when only vertical component data are available.

Method

Earthquake locations are usually provided by the traveltime method. However, in most cases, due to an inaccurate velocity model and picking erroneous arrival times, the estimated seismic event locations from traveltime may have errors, which are greater especially at depth. Therefore, to find a better agreement between synthetic and observed waveforms, we also search for better locations around the catalog location. The search domain for a better location is shown in Figure 1. The red dot denotes the location given by the traveltime locating method, e.g., the double difference method (Zhang and Thurber, 2003), and the surrounding green dots (including the red dot) denote the possible locations in our search.

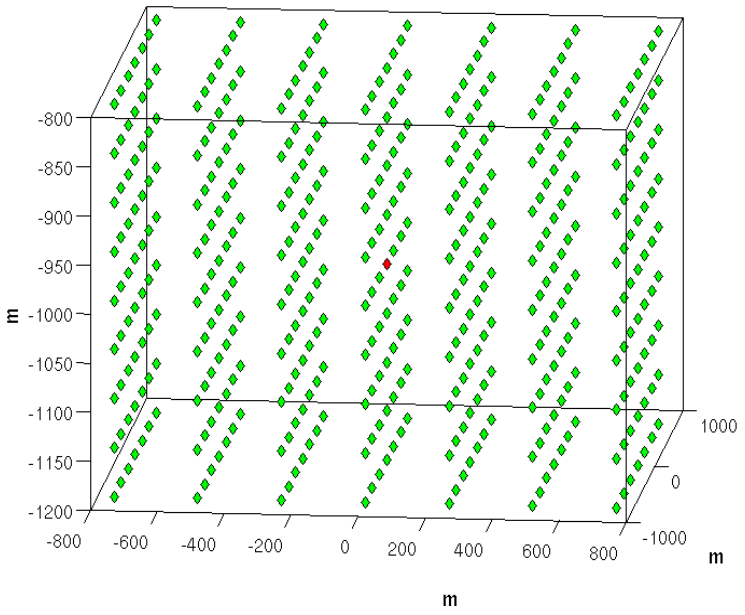

Figure 1. Search domain for a better event location.

The focal mechanism can be simplified and represented by a 3 by 3 moment tensor (Stein and Wysession, 2003). Usually, as there is no rotation of mass involved in the rupturing process, the tensor is symmetric, and only has six independent components. Here we assume the focal mechanism of the small induced seismicity can be represented by pure double couples (Rutledge and Phillips, 2002), though it is possible that a volume change or Compensated Linear Vector Dipoles (CLVD) part may also exist. The constraining of focal mechanism as double couples (DCs) has additional advantages. Specifically, the anisotropy, which is not described by the isotropic Green's functions, often raises spurious non-DC components in focal mechanism determination (Eisner et. al., 2007); thus, by assuming focal mechanism to be DCs, this spurious non-DC components can be ruled out. Instead of searching through all possible double couple combinations, we search through all possible strikes, dips, and rakes for an event, and decompose these three parameters as double couples. For each of these six components, we can use a Discrete Wavenumber Method (DWN; Bouchon, 2003) to calculate its Green Functions. Then the synthetic waveform from a certain combination of strike, dip, and rake is expressed as a linear combination of weighted Green's functions:

$$v_i^n = \sum_{j=1}^{3}\sum_{k=1}^{3} m_{jk}^n G_{jk,i}^n(t) * s(t) \quad , \tag{1}$$

where v_i^n is the synthetic ith (north, east or vertical) component at station n; m_{jk}^n is the moment tensor component, $G_{jk,i}^n(t)$ is the ith component of the Green's function for the (j, k) entry at station n (note for different locations shown in Figure 1, $G_{jk,i}^n(t)$ is different), and $s(t)$ is the source time function. In this study, a smooth ramp is used as $s(t)$.

Before the grid search is performed, we build a Green's function library. We pre-calculate all $G^n_{jk,i}(t)$'s for all possible event locations at all stations and store them on the disk. When we perform the grid search, we simply need to do a linear combination of $G^n_{jk,i}(t) * s(t)$, each of which is weighted by m^n_{jk}.

To find the similarity between synthetic and observed waveforms, we need to do two kinds of basic computations: filtering and cross correlation. Even on modern computers, these two computations are very time-consuming, especially when they are executed repetitively. Therefore, the following manipulations are performed to solve this issue:

$$\begin{aligned} F * v^n_i &= F * \sum_{j=1}^{3}\sum_{k=1}^{3} m^n_{jk} G^n_{jk,i}(t) * s(t) \\ &= \sum_{j=1}^{3}\sum_{k=1}^{3} m^n_{jk} [F * (G^n_{jk,i}(t) * s(t))] \end{aligned} \tag{2}$$

$$\begin{aligned} D^n_i \otimes v^n_i &= D^n_i \otimes \sum_{j=1}^{3}\sum_{k=1}^{3} m^n_{jk} G^n_{jk,i}(t) * s(t) \\ &= \sum_{j=1}^{3}\sum_{k=1}^{3} m^n_{jk} [D^n_i \otimes (G^n_{jk,i}(t) * s(t))] \end{aligned} \tag{3}$$

where F denotes the impulse response of a filter; "$*$" denotes time domain convolution; D^n_i denotes the ith component of the data at station n; "$\otimes$" denotes the cross correlation. These two equations indicate that after convolving the source time function with the Green's functions, we can apply the filtering or cross correlate the observed data to each of the convolved Green's functions. By inserting the filtering and cross correlation into the summation, we do not need to

do filtering and cross correlation repetitively in the search over all strikes, dips, and rakes. A large amount of time is thereby saved, and the searching speed can be boosted tenfold.

To determine if a solution is correct, we need to select an objective function that characterizes the similarity between the synthetic and observed waveforms. We design the following objective function, which utilizes four different aspects of the waveform information:

$$\begin{aligned}&\text{maximize}(J(x,y,z,str,dip,rake,ts)) = \\ &\sum_{n=1}^{N}\sum_{j=1}^{3}\{\alpha_1 \max(\tilde{d}_j^n \otimes \tilde{v}_j^n) - \alpha_2 \left\| \tilde{d}_j^n - \tilde{v}_j^n \right\|_2 \\ &+ \alpha_3 f(pol(\tilde{d}_j^n), pol(\tilde{v}_j^n)) + \alpha_4 h(rat(\frac{S(d_j^n)}{P(d_j^n)}), rat(\frac{S(v_j^n)}{P(v_j^n)}))\}\end{aligned} \tag{4}$$

Usually, the amplitudes of P waves are much smaller than those of S waves. To balance the contribution from P and S waves, we need to fit P and S waves separately using equation (3). Also, separating S from P waves and allowing an independent time shift in comparing observed data with synthetic waveforms can help deal with incorrect phase arrival time due to incorrect P/S velocity ratios. We compute the first P and S arrival traveltime using the level set method and then the wave train is separated into two parts at about the beginning of the S wave. The level set method is able to solve the Eikonal equation in heterogeneous medium accurately and stably. As there usually is a tradeoff between the origin time and the event depth, we eliminate the influence of inaccurate origin time information by the following method: we align the first P arrival moment in the synthetic data with the manual pick of first P arrival in the observed data automatically, and any later time shift by cross correlation is performed according to this alignment. The objective function J consists of 4 terms; α_1 through α_4 are the weights for each

term. The first term evaluates the maximum cross correlation between the normalized data ($\tilde{d}_j^n$) and the normalized synthetic waveforms ($\tilde{v}_j^n$). The normalization used here is the energy normalization, such that the energy of the normalized wave train within a time window adds to unity. In a concise form, this normalization can be written as:

$$\tilde{d}_j^n = \frac{d_j^n}{\int_{t_1}^{t_2} (d_j^n)^2 \, dt} \quad , \tag{5}$$

where t_1 and t_2 are the boundaries of the time window. From the cross correlation, we find the time-shift to align the synthetic with the observed waveform. In high frequency waveform comparisons, cycle-skip is a special issue requiring extra attention: over-shifting the waveform makes the wiggles in the data misaligned with wiggles of next cycle in the synthetic waveforms. Therefore, allowed maximum time shift should be predetermined by the central frequency in the waveforms. The second term evaluates the L_2 norm of the direct differences between the aligned synthetic and observed waveforms. The reason for both maximizing the cross correlation value and minimizing the direct difference between the observed and synthetic waveforms is to reduce the effect of attenuation and inaccurate density information, which affect the amplitudes of the synthetic data. The third term evaluates whether the polarities of the first P-wave arrivals as observed in the data are consistent with those of the synthetic waveforms. In numerical implementation, we first calculate the traveltimes for the first P arrivals, and then by summing over the waveform in a narrow window around that arrival time and checking the sign of the summation we determine the polarities robustly. For the observed data, we manually determine the polarities. Because it is difficult to obtain accurate absolute amplitudes due to site effects, we normalize the data and synthetic waveforms before comparison but lose the ratio and amplitude

information by doing so. We retain the radiation pattern information by incorporating the fourth term in the objective function, which evaluates the consistency of the S/P ratios in observed and synthetic waveforms. The "rat" is the ratio evaluation function and it can be written as:

$$rat = \frac{\int_{T_2}^{T_3} \left| r_j^n(t) \right| dt}{\int_{T_1}^{T_2} \left| r_j^n(t) \right| dt} \quad , \tag{6}$$

where $[T_1 \ \ T_2]$ and $[T_2 \ \ T_3]$ define the time window of P and S wave, respectively, and r_j^n denotes either d_j^n or v_j^n. The term h evaluates the ratio differences. Note here we use the un-normalized waveforms d_j^n and v_j^n.

By pre-calculating the Green's function library and manipulating the filtering and cross correlation, we speed up the grid search process. Searching through all possible X, Y, and Z for location and strikes, dips, and rakes for focal mechanisms often results in over 10 million different waveforms to be compared with the data. Since the grid search can be easily parallelized, it can be done on a multicore desktop machine within 8 minutes. The computation of the Green's function library takes more time, but it only needs to be computed once.

Result

I. Numerical Test

1.1 Noise Contamination

We first test the reliability and robustness of our algorithm. The station distribution and velocity structure in an oil reservoir is used for this discussion (Figure 2). The stations are equipped with

3-component short-period seismometers with a natural resonance frequency of 4.5 Hz. However, since these stations are located 150 m beneath the surface in shallow boreholes, the horizontal orientations of these stations are generally unknown. Therefore, only records from the vertical components are used. This constraint, which introduces many more difficulties into our focal mechanism determination, however, is a realistic constraint in most field oil reservoir monitoring. Therefore, our method can also be applied when only geophones are available. Two scenarios are tested: 1). the synthetic data are contaminated by Gaussian distribution random noise to simulate scattering; 2). we perturb the P and S velocity model for each station independently to simulate lateral velocity heterogeneity. For the first scenario, we use DWN to generate synthetic seismograms and then add Gaussian distribution noises to them. The noise on each trace has a standard deviation of 5% the maximum absolute amplitude of that trace, and this level corresponds to the largest noise level we encounter for valid data in the testing oil field. A source located 1227 m beneath the surface, with a strike of 210^o, dip of 50^o, and rake of -40^o is used to generate the synthetic seismograms. In the later analysis, the hypercenter is re-coordinated as (0, 0, 0) for X, Y and depth for studying the locating precision of this algorithm. Hereafter, we use a horizontal grid spacing of 150 m, and vertical grid spacing of 50 m. The reason for choosing smaller vertical grid spacing than in the horizontal direction is that the seismogram is not very sensitive to location variation in the horizontal direction, but very sensitive to the variation in the vertical direction due to multiple reflections and refractions. The frequency band used in our research is 3 ~ 9 Hz, which contains the majority of the energy in the typical induced seismicity (the spectrograms for a typical event will be shown in the real data study figures). The searching interval in strike, dip, and rake is 10^o in this test and hereafter in the real data test. This spacing choice indicates that our resolution is 5^o at the best situation. Because

the auxiliary plane solution and the fault plane solution give the identical waveform, it means that half of the model space $[0^\circ \le strike \le 360^\circ, 0^\circ \le dip \le 90^\circ, -180^\circ \le rake \le 180^\circ]$ is redundant. Therefore, by constraining the model space in $[0^\circ \le strike \le 360^\circ, 0^\circ \le dip \le 90^\circ, -90^\circ \le rake \le 90^\circ]$ (Zhao and Helmberger, 1994) we can rule out the redundancy and further shorten the search time by half.

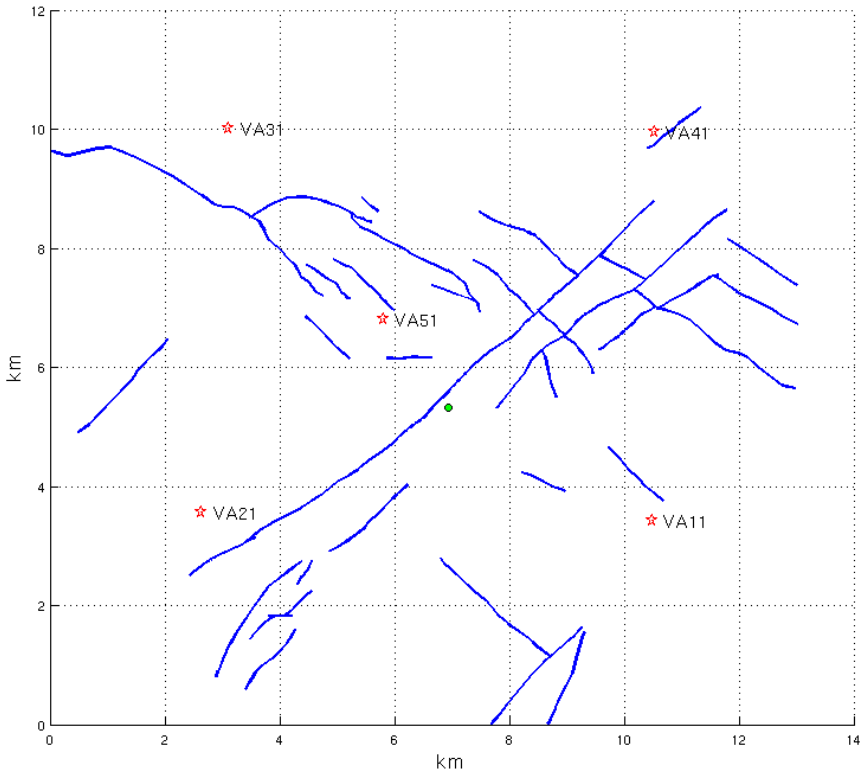

Figure 2. Station distribution (red star) and the synthetic event (green dot); the faults, stations and event are plotted in a local reference system

Figure 3 shows the focal mechanism determination results using only three first P arrival polarities (we assume two polarities out of five are not identifiable due to noise contamination). The best solution here (#1) almost matches the correct solution. The slight difference in strike is caused by our strike spacing choice: having an interval of 10^o, the closest solution the grid search

gives is 205^o or 215^o. Figure 4 shows the comparison between the synthetic and observed contaminated waveforms. The "shift" in the title of each subplot indicates the time shifted in the data to align with the synthetic waveforms. We still have some time shift due to two reasons: 1) we introduce some artificial error in arrival time by manually picking the first P arrival in the observed data; 2) scattering noise can change the maximum cross correlation position according to the Banada-doughnut theory (Nolet et. al., 2005). The number to the left of the slash in the right column denotes the S/P ratio for the data, and the number to the right of the slash denotes the ratio for the synthetic waveform. They are quite close in most cases. The "+" or "-" signs in the left column indicate the first arrival polarities of P waves of the data and those of the synthetic waveforms; the upper signs are those for the data while the lower are those for the synthetic waveforms. We can see they are identical. Note that for evaluation of the polarities, we use the unfiltered waveforms, as filtering usually blurs or destroys the polarities.

We further show the distribution of strikes, dips, and rakes from the determination. The fault plane solutions among the first 200 best solutions are evaluated and shown in Figure 5. The strike, dip, and rake all have a mean quite close to the correct solution (210^o, 50^o and -40^o, respectively). Among the strike, dip, and rake, dip has the least standard deviation (9^o). This might be an indication of the sensitivity of these model parameters. Figure 6 shows the distribution of the relocated positions. We find in the X and Y direction that the error deviation is much larger than that in the Z direction. This is because a small change in the vertical direction changes the reflection and transmission pattern considerably in this layered medium, while horizontal dislocation only makes small variations in the waveform. Thus, this algorithm works quite well in determining depth. The capability to precisely relocate the depth indicates that using

waveform information puts strong constraints on depth variation; therefore, this method could potentially be a new approach to constrain earthquake depth.

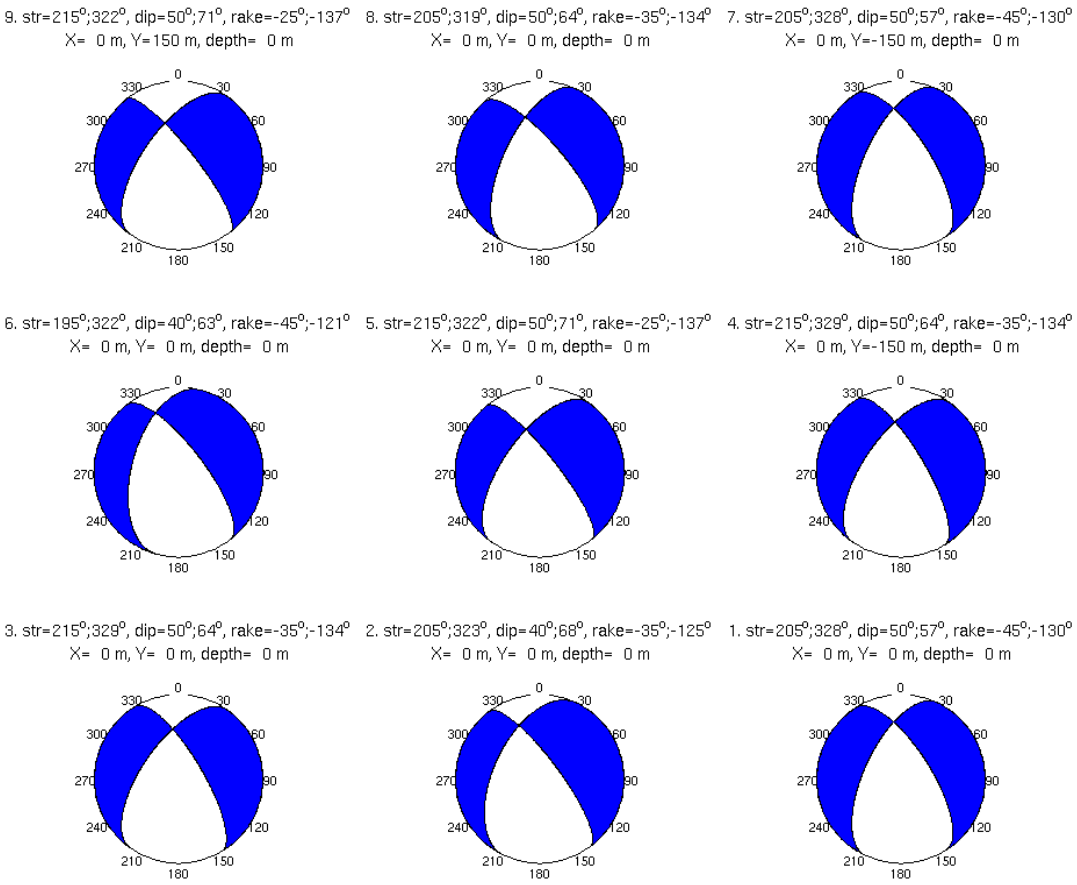

Figure 3. Nine best solutions from contaminated data; the number before "str" is the order: "1" means the best solution and "9" means the relatively worst among these nine.

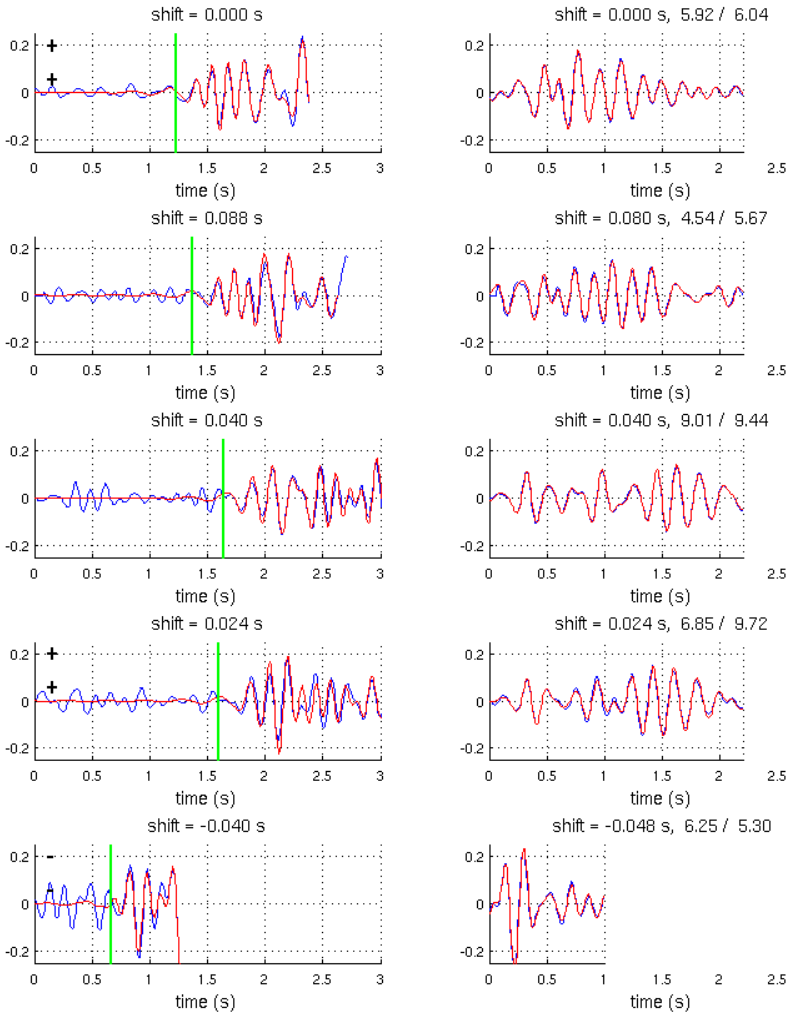

Figure 4. Comparison between synthetic waveforms (red) and data (blue) at 5 stations. From top to bottom show waveforms at station 1 through 5, respectively. Left column shows P waves and right column shows S waves.

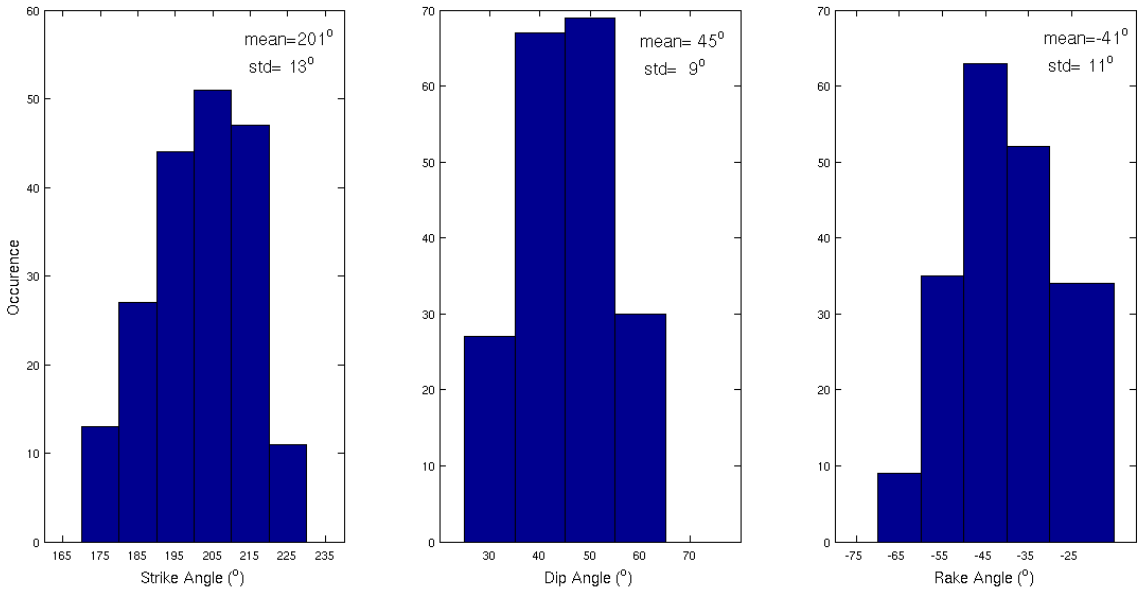

Figure 5. Distribution of the strikes, dips, and rakes among the first 200 best solutions.

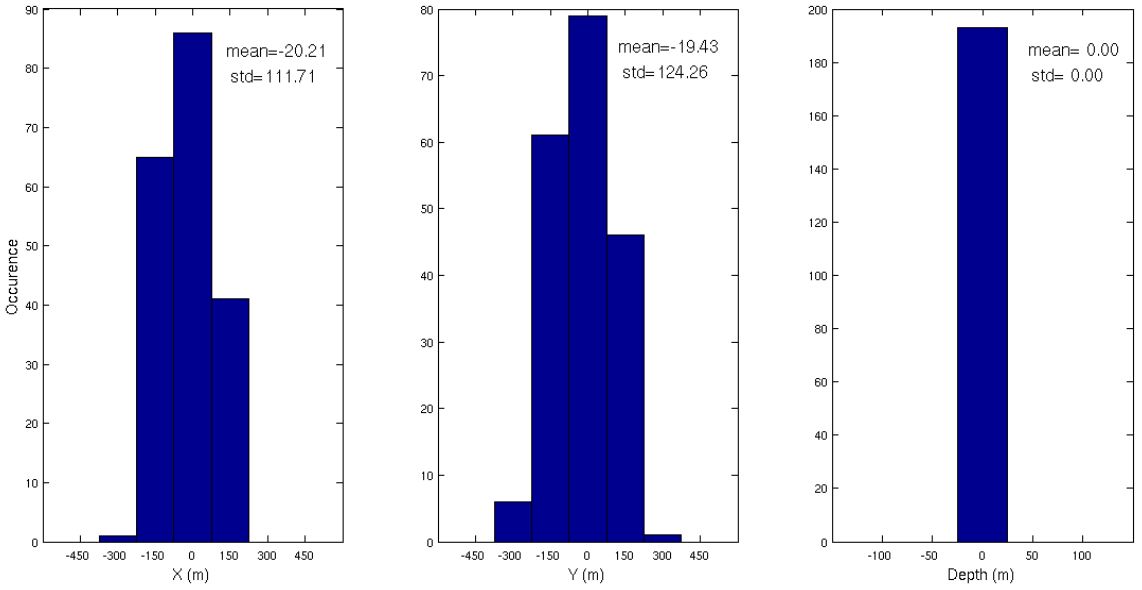

Figure 6. Distribution of the best X, Y, and depth among the first 200 best solutions.

1.2 Lateral Velocity Variation

We further test the reliability and robustness of this new algorithm in the case where the velocity model has errors due to lateral variation. In general, we have a reliable velocity model for both P and S waves from well logs (Sarkar, 2008). However, since lateral velocity heterogeneity is inevitable the influence of perturbation in the velocity model on the focal mechanism determination has also been examined. We still use a 1-D layered structure to generate the synthetic observed seismogram from the event to a station. However, for a certain layer, a random Gaussian perturbation with a standard deviation of 5% of the velocity of that layer is added. The perturbation is independent for all five stations. The density here is not perturbed, as the variation in density only has a slight influence on the waveform's amplitudes but not on the phases, which have a dominant influence on the waveform. Also, the layer thickness is not perturbed, as perturbation in either layer velocity or thickness generates equivalent phase distortions in each layer. The perturbations in P and S velocity models are shown in Figure 7. Note the perturbations in P and S velocity are also independent for each layer and station.

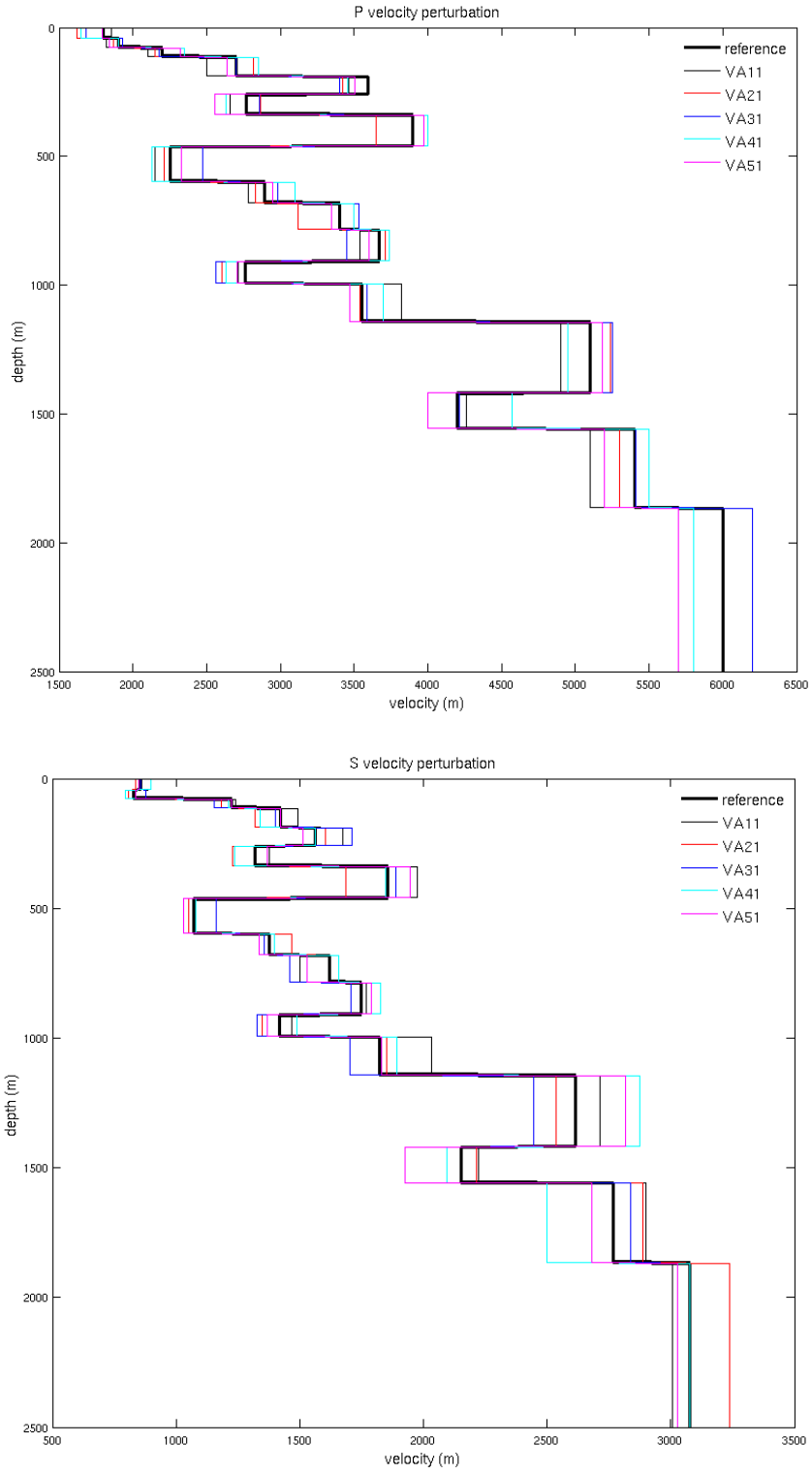

Figure 7. Perturbations in P and S velocity models.

From Figure 7 we can find the variation in velocity for a certain layer is quite large from one station to another. Considering our study region is less than 10 km by 10 km, and this region is mainly composed of flat sedimentary rocks, this variation should be a reasonable approximation to the real maximum lateral heterogeneity. Here we still use only three first P arrival polarities. Figure 8 shows the best nine beachball solutions. Surprisingly, the solutions do not differ significantly from the correct one (strike = 210^o, dip=50^o, rake=-40^o), and they are all very close to the correct solution by an error in strike, dip, or rake by about 10^o and one grid spacing (150 m) in the horizontal direction. All of these solutions find the correct depth. Figure 9 shows the comparison between the synthetic waveforms generated from the reference model and the 'observed' waveforms generated from the perturbed velocity model. Due to the different phase shift by velocity variation, many phases in the waveform have been distorted. However, allowing time shift in the observed data compensates for much of the phase shift and distortion caused by velocity variation (note the "shift" here is much larger than in the previous case). Therefore, by incorporating information from different aspects in the waveform, and compensating for the phase shift we are still able to obtain a focal mechanism quite close to the correct one. Figure 10 shows the distribution of strike, dip, and rake of the focal plane solution from the first 200 best solutions. The mean values are quite close to the correct ones, and the standard deviations are small, considering that only vertical components in 5 stations and 3 polarities are used to determine the focal mechanism. Figure 11 shows the distribution of X, Y, and depth. Similar to the previous case, the deviation in depth is much smaller than in the X and Y direction; in fact, all results have the correct depth. Again, the depth variation is strongly constrained, and this indicates that using waveform information can greatly help to locate the depths of events, which are most difficult to constrain in the traditional traveltime method.

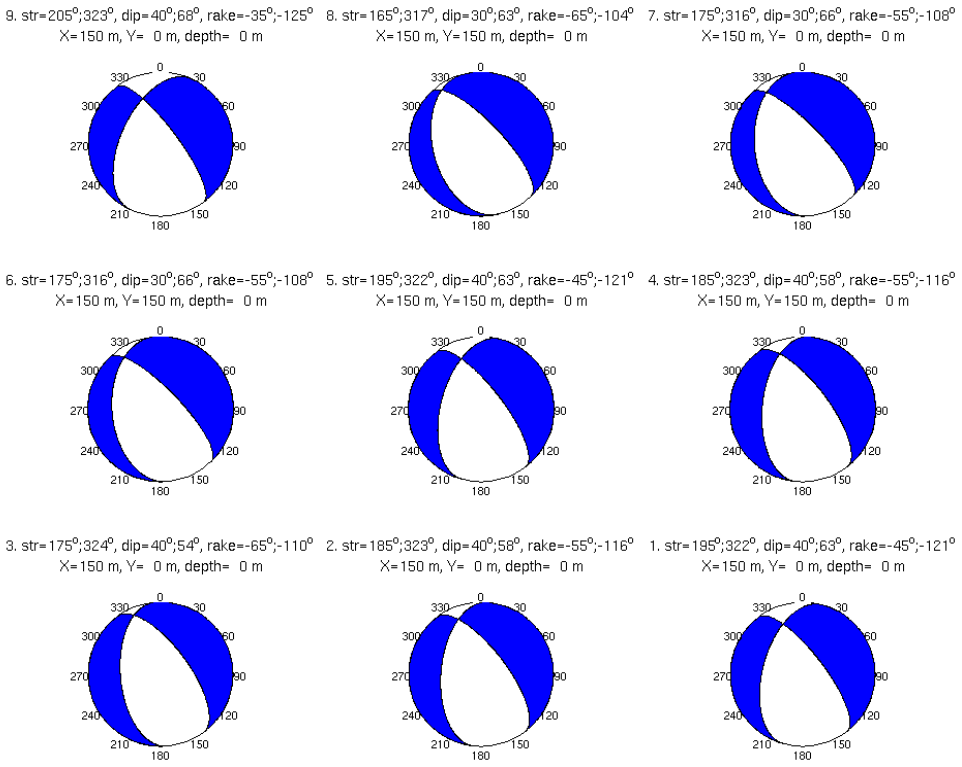

Figure 8. Focal mechanism solutions with lateral velocity perturbation.

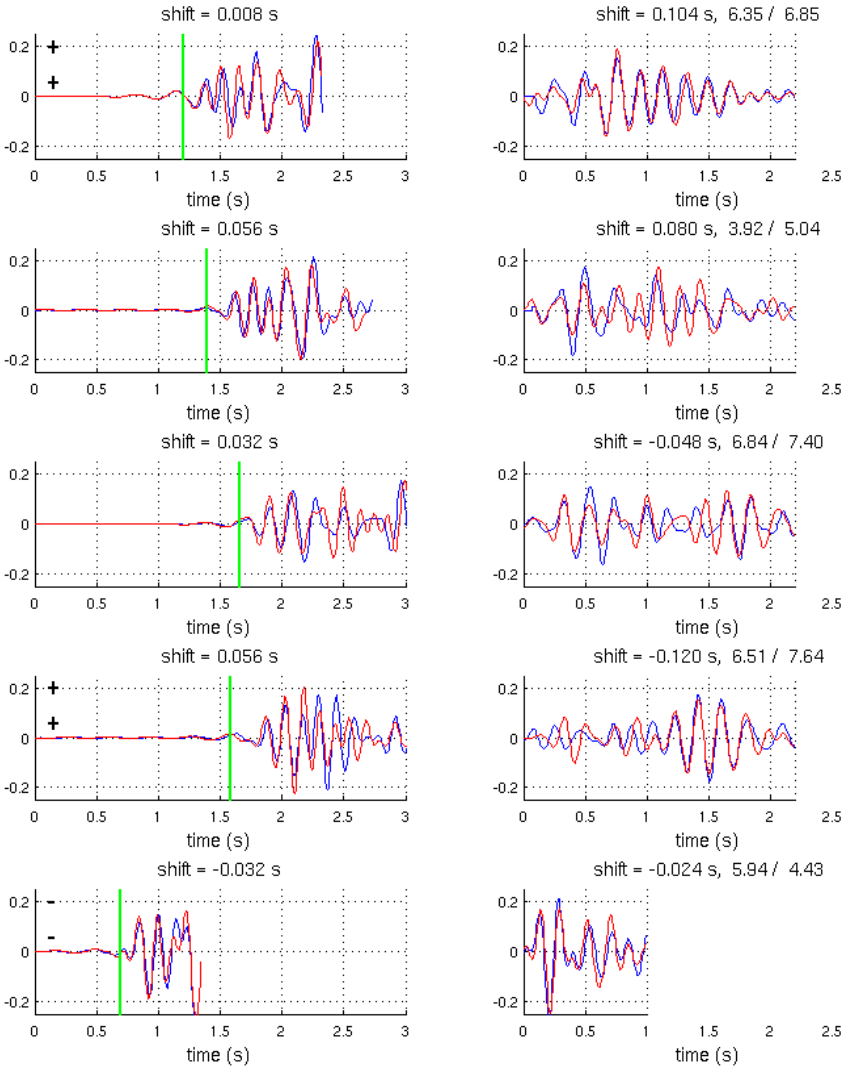

Figure 9. Comparison between synthetic and the observed data with lateral velocity variation.

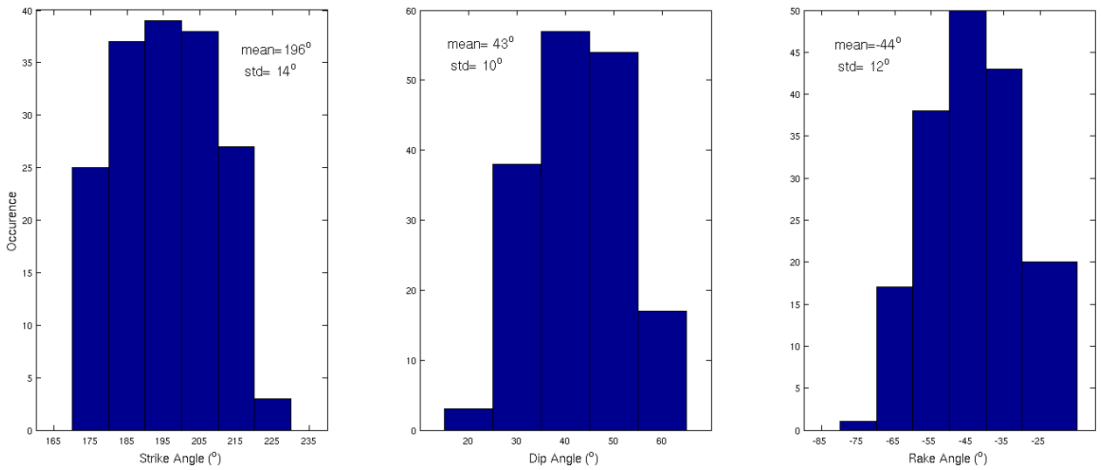

Figure 10. Distribution of strike, dip, and rake with lateral velocity variation.

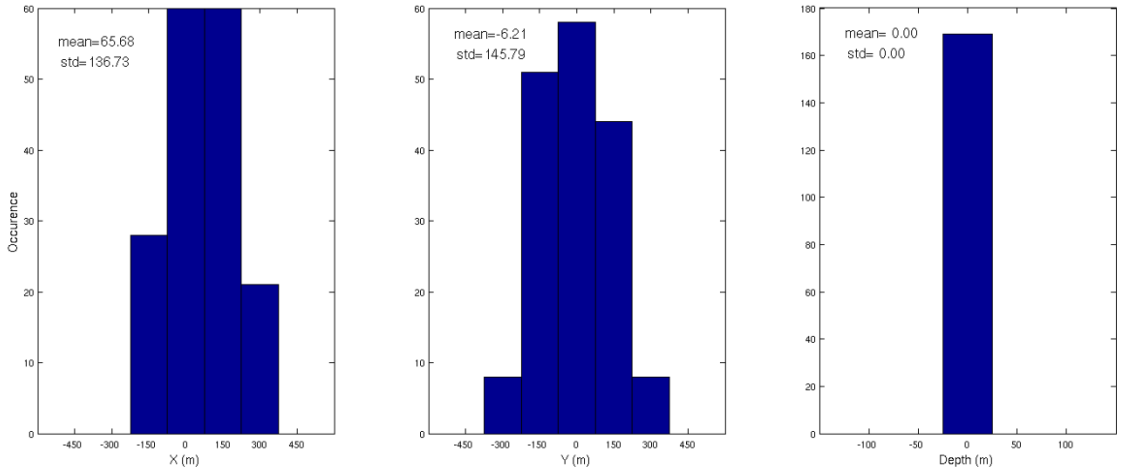

Figure 11. Distribution of X, Y, and depth with lateral velocity variation.

II. Real Data Study using an Induced Event

We also study an induced microseismic event recorded by a five station shallow borehole network in an oil reservoir using our method. Figure 12 shows a typical event recorded at one of these stations and its spectrograms. We can see a majority of the waveform energy is between 3 and 15 Hz.

The layered velocity model is obtained from well logs, and is shown as the reference graph in Figure 7. The event studied in the reservoir occurred on March 4, 2001 (tagged 20010047). First, a double-difference traveltime locating routine is used to locate the hypercenter of the event (Sarkar, 2008). Then, within the neighborhood of this traveltime located hypercenter, we calculate the Green's function with DWN as shown in Figure 1.

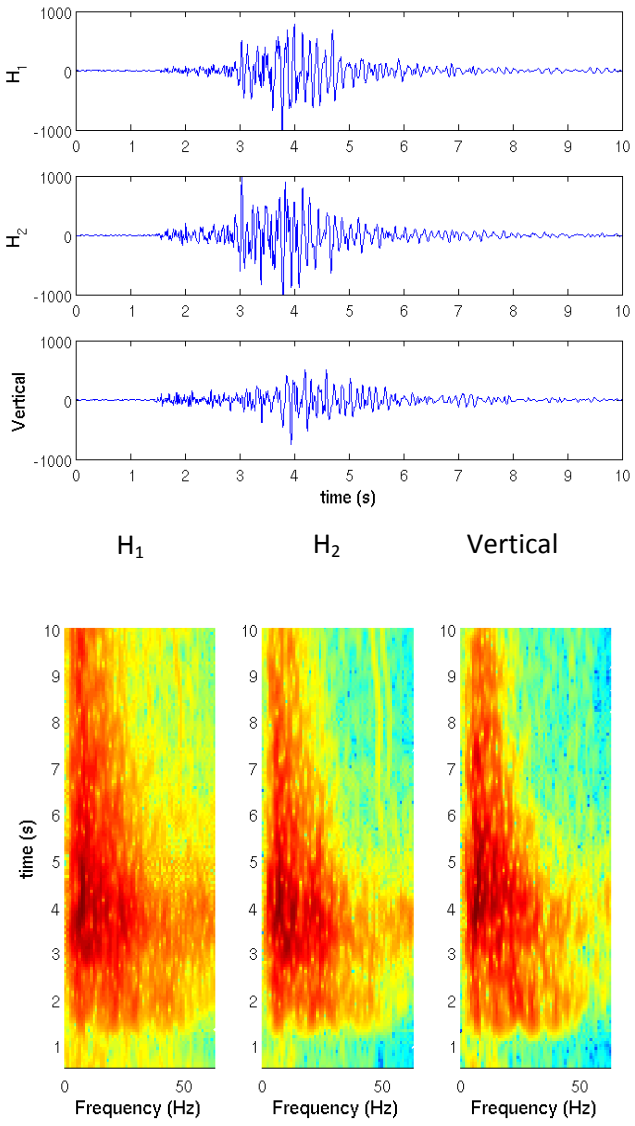

Figure 12. A typical event used in the focal mechanism determination and its spectrograms.

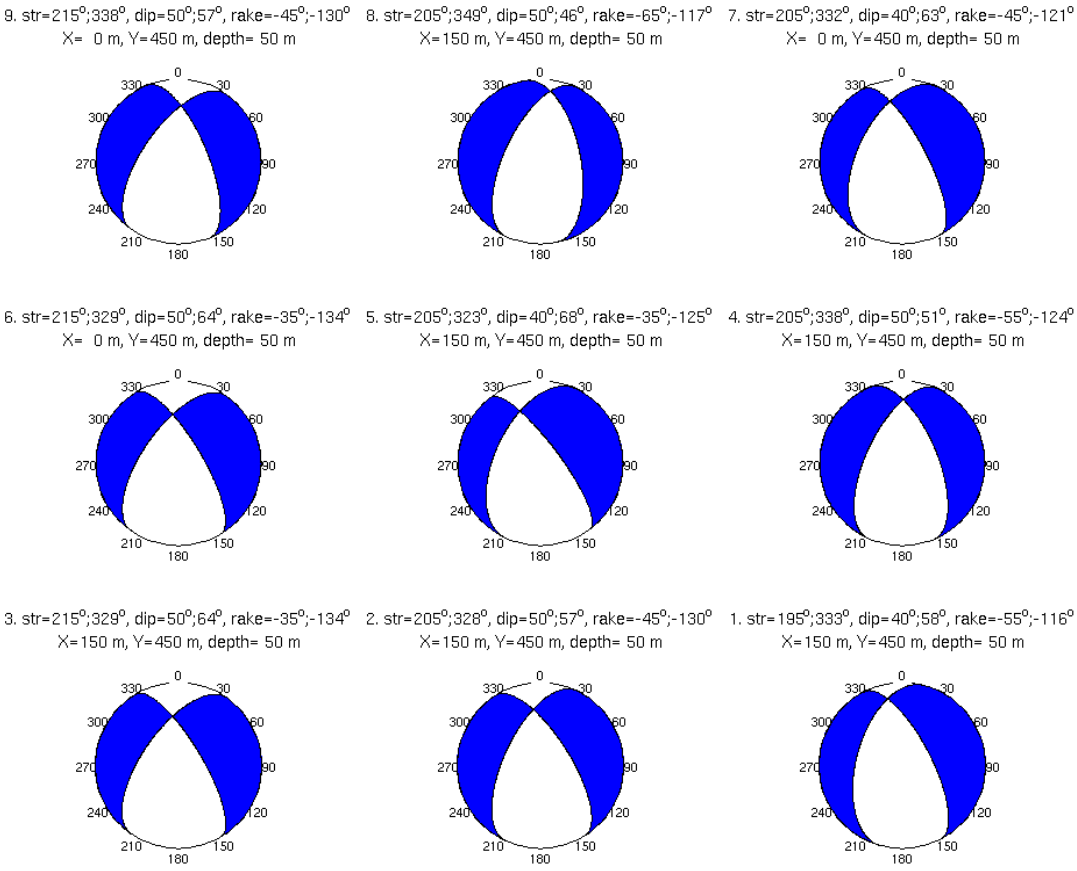

Figure 13. Focal mechanism solution for event 20010047.

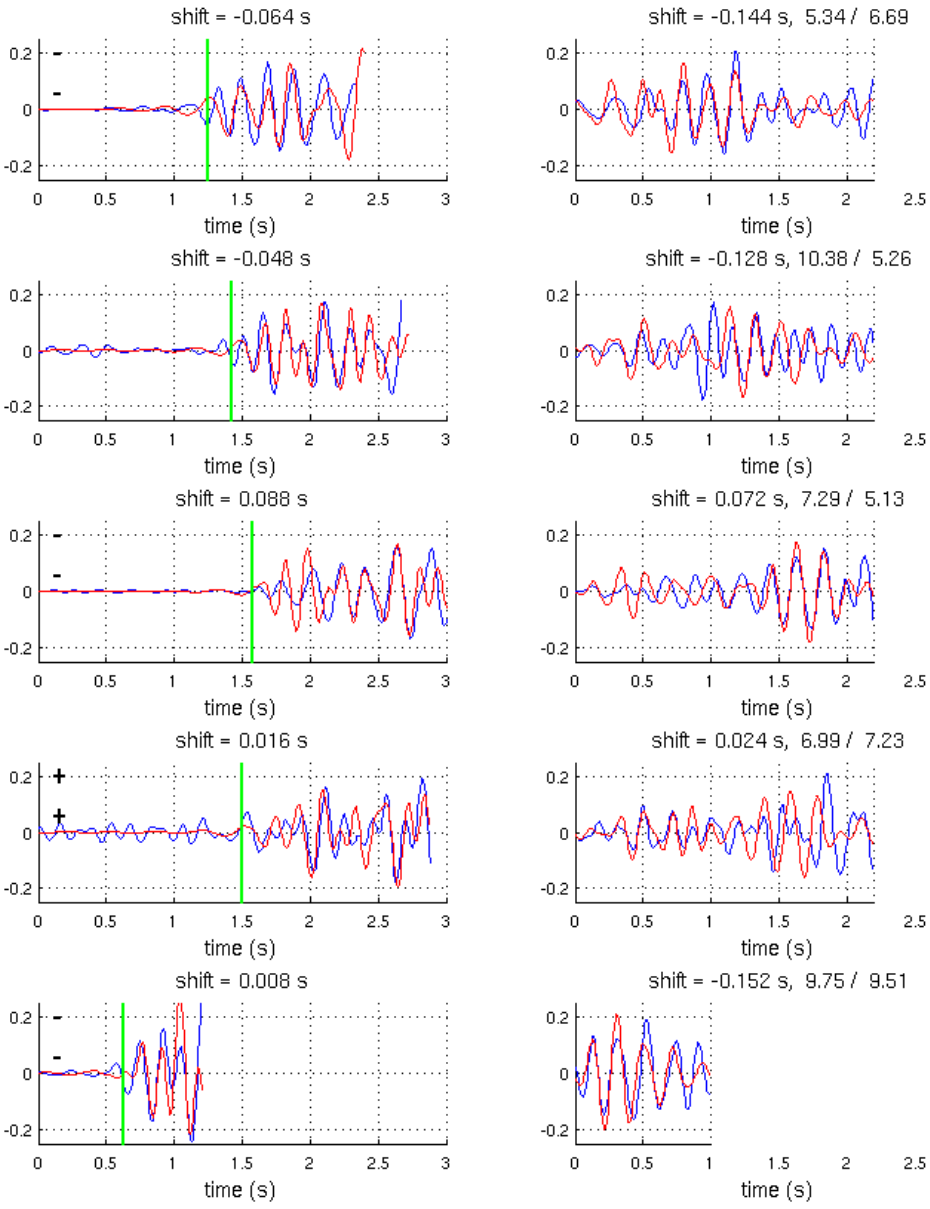

Figure 14. Comparison between synthetic waveforms (red) and real data (blue) at 5 stations.

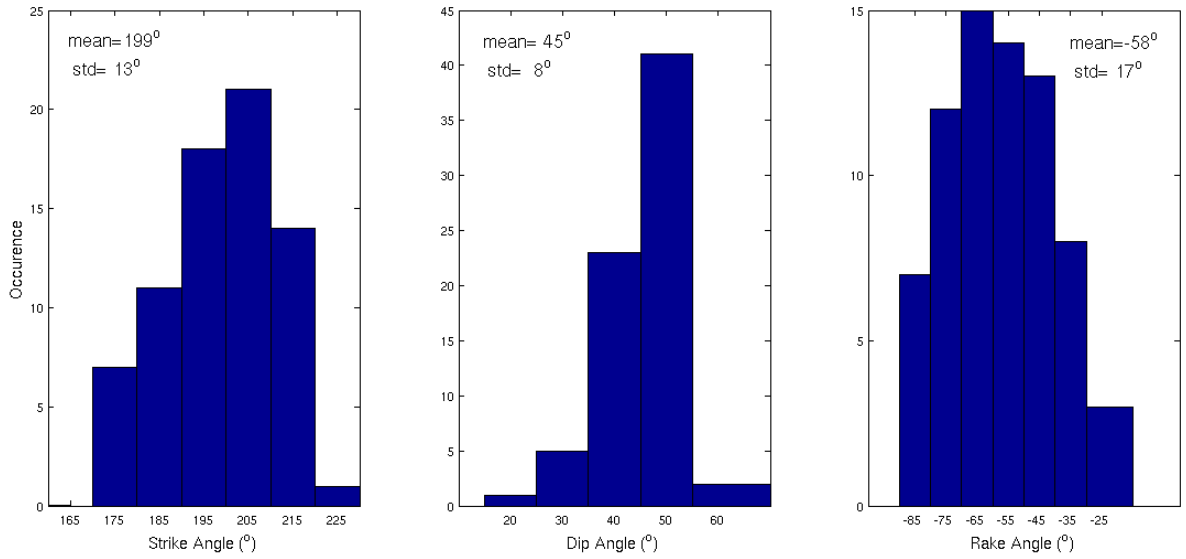

Figure 15. Distribution of strike, dip, and rake among the first 200 best solutions for event 20010047.

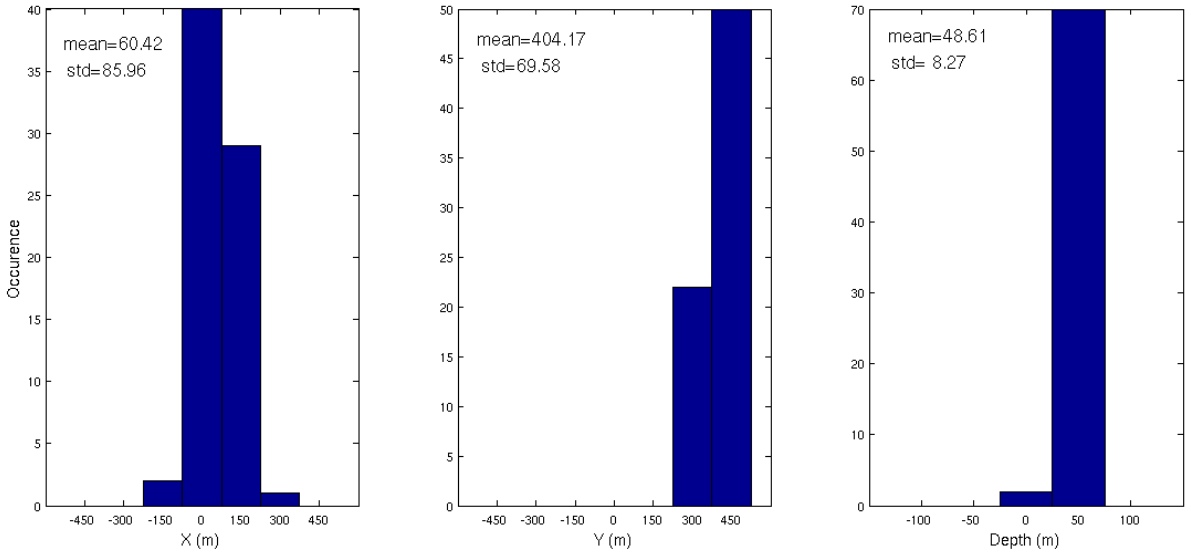

Figure 16. Distribution of X, Y, and depth among the first 200 best solutions for event 20010047.

Figure 13 shows the beachballs of the nine best solutions out of millions of trials. Our best solution (the one at the bottom right) has a strike of 195^o, which is quite close to the best known orientation of the fault (219^o). Also, this event is determined as a dip-strike event, which is a reasonable solution considering local oil production and water injection. Using our new algorithm, the epicenter is shifted northward by about 450 m, eastward by about 150 m, and the depth is shifted 50 m deeper. Figure 14 shows the comparison between the synthetic and the observed data. Comparing Figure 14 with Figure 9, the degree of matching between the synthetic and observed waveforms is quite close in both the synthetic test and the real data test. Figure 15 shows the distribution of strike, dip, and rake. Again, we find dip has the minimum standard deviation. Figure 16 shows the distribution of X, Y, and depth. Our algorithm shows excellent ability in locating the depth of event. However, as discussed before, variation in hypocenter (X and Y direction) can be compensated by shifting the waveforms to find a better alignment with the synthetic data, and the phases in the waveforms are not very sensitive to the horizontal shift. Therefore, the constraint on lateral shift in weaker compared to that on the vertical direction.

Conclusions

We show that combining the vertical component of high frequency seismograms with a fast grid-search algorithm leads to close determination of focal mechanism and locations of small earthquakes where subsurface velocity information is available. These conditions generally apply to petroleum fields. Therefore, the method is especially applicable to the study of induced earthquakes recorded by a small number of stations, even when some first P arrival polarities are not identifiable due to noise contamination. The new objective function, which includes matching phase and amplitude information, first arrival P polarities, and S/P amplitude ratios

between synthetic and observed waveforms, is designed to utilize more information from the data, yielding very stable and reasonable solutions. The synthetic tests prove that our algorithm has strong reliability and robustness. Correct solutions can still be found even when noise contamination or lateral velocity variation exists and the 1-D layered velocity model becomes only an approximation to the real structure. For real events, focal mechanisms are found to be reasonable, considering the local geological structure and possible causes for those induced seismic events.

Acknowledgement

This work was partially supported by the MIT Earth Resources Laboratory Founding Member Consortium.

References

Bouchon, M. (2003). A review of the discrete wavenumber method, Pure and applied Geophysics, **160**, 445-465.

Hardebeck, J.L. and P. M. Shearer (2002). A new method for determining first-motion focal mechanisms, Bull. Seism. Soc. Am. **93**, 1875-1889.

Nolet G., F.A. Dahlen and R. Montelli (2005). Traveltimes and amplitudes of seismic waves: a re-assessment, Array analysis of broadband seismograms, AGU monograph series.

Osher, S. and J. A. Sethian (1988). Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations, J. Comput. Phys. **79**: 12–49.

Rawlinson, N. and M. Sambridge (2004). Multiple reflection and transmission phases in complex layered media using a multistage fast marching method, Geophysics, **69**.

Rutledge, J.T. and W.S. Phillips (2002). A comparison of microseismicity induced by gel-proppant- and water-injected hydraulic fractures, Carthage Cotton Valley gas field, East Texas, 72nd Ann. Internat. Mtg: Soc. Of Expl. Geophys., 2393-2396.

Sarkar, S. (2008). Reservoir monitoring using induced seismicity at a petroleum field in Oman: Ph.D. thesis, Massachusetts Institute of Technology.

Sze, E, M.N. Toksöz, and D.R. Burns (2004). Characterization of induced seismicity in a petroleum reservoir: a case study, 2004 ERL consortium meeting report.

Tan, Y. and D.V. Helmberger (2007). A new method for determining small earthquake source parameters using short-period P waves, Bull. Seism. Soc. Am. **97**, 1176-1195.

Tromp J., C. Tape, and Q. Liu (2005). Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, Geophys. J. Int. **160**, 195-216.

Zhang, H.J. and C.H. Thurber (2003). Double-difference tomography: The method and its application to the Hayward fault, California, Bull. Seism. Soc. Am. **93**, 1875-1889.

Zhao, L.S. and D.V. Helmberger (1994). Source estimation from broadband regional seismograms, Bull. Seism. Soc. Am. **84**, 91-104.

Zhu, L. and D.V. Helmberger (1996). Advancement in source estimation techniques using broadband regional seismograms, Bull. Seism. Soc. Am. **86**, 1634-1641.