Focal Mechanism Determination Using High Frequency Waveform Matching and Its Application to Small Magnitude Induced Earthquakes

Junlun Li, Haijiang Zhang, H. Sadi Kuleli, and M. Nafi Toksoz

junlunli@mit.edu

Earth Resources Laboratory
Department of Earth, Atmospheric and Planetary Sciences
Massachusetts Institute of Technology
77 Massachusetts Avenue, Cambridge, MA 02139

Summary

We present a new method using high frequency full waveform information to determine the focal mechanisms of small, local earthquakes monitored by a sparse surface network. During the waveform inversion, we maximize both the phase and amplitude matching between the observed and modeled waveforms. In addition, we use the polarities of the first P-wave arrivals and the average S/P amplitude ratios to better constrain the matching. An objective function is constructed to include all four criteria. An optimized grid search method is used to search over all possible ranges of source parameters (strike, dip and rake). 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. Optimizations in filtering and cross-correlation are performed to further speed the grid search algorithm. The new method is tested on a 5-station surface network used for monitoring induced seismicity at a petroleum field. The synthetic test showed that our method is robust and efficient to determine the focal mechanism when using only the vertical component of seismograms in the frequency range of 3 to 9 Hz. The application to dozens of induced seismic events showed satisfactory waveform matching between modeled and observed seismograms. The majority of the events have a strike direction parallel with the major NE-SW faults in the region. The normal faulting mechanism is dominant, which suggests the vertical stress is larger than the horizontal stress.

Introduction

In this study we use high frequency seismograms for determining the focal mechanisms of small earthquakes recorded by a sparse network of seismic stations. The method involves calculating synthetic seismograms for a series of moment tensors and finding the source mechanism where the observed and synthetic seismograms match the best. The method is especially useful for determining the source mechanisms of induced seismic events.

Induced seismicity is a common phenomenon in oil/gas reservoirs and in reservoirs where activities are in progress and internal stress distribution is changed due to water injection, fluid extraction etc. (Rutledge & Phillips, 2003; Rutledge et al., 2004; Chan & Zoback, 2007; Deichmann & Giardini, 2009; Bischoff et al., 2010; Segall, 2010; Suckale, 2010). For example, the gas/oil extraction can cause reservoir compaction and reactivate preexisting faults and induce earthquakes (e.g., Chan & Zoback, 2007; Miyazawa et al., 2008; Sarkar et al., 2008). By studying the patterns of the induced seismicity over an extended time period (e.g. location and focal mechanism), a time-lapse history of the stress changes in the fields may be reconstructed.

Induced earthquakes usually have small magnitudes and are recorded at local stations. Since the monitoring networks at these production sites are usually sparse, it is challenging, if not impossible, to use the conventional P-wave polarity information and/or S/P amplitude ratios to constrain the focal mechanism of the induced earthquakes (Hardebeck & Shearer, 2002, 2003). Waveform matching has been used to determine earthquake focal mechanisms on a regional and global scale using low frequency waveform information (Zhao & Helmberger, 1994; Zhu & Helmberger, 1996; Pasyanos et al., 1996; Dreger et al., 1998; Tan & Helmberger, 2007). However, in induced seismicity cases, waveforms usually have higher frequencies.

High frequency waveform matching, in addition to polarity information, has been used to determine the focal mechanism of induced earthquakes in a mine with a dense network of 20 stations (Julià et al., 2009). Julià et al. used a constant velocity model to calculate the Green's functions, and performed the focal mechanism inversion in frequency domain without phase information in a least square sense between the synthetic and filtered observed data generally below 10 Hz.

To retrieve reliable solutions, our study uses high frequency, full waveform information (both P and S) to determine the focal mechanism of small earthquakes recorded by a sparse 5-station network. Using the known velocity structure, we calculate the Green's functions for all moment tensor components of the source and for each location (hypocenter) and then the synthetic seismograms. To find the best match between the observed and synthetic seismograms, we formulate an objective function that incorporates information from different attributes in the waveforms: the cross correlation values between the modeled waveforms and the data, the L_2 norms of the waveform differences, the polarities of the first P arrivals and the S/P average amplitude ratios.

For real application we use data from a 5-station network monitoring induced seismicity at a petroleum field (Sarkar, 2008; Zhang et al., 2009). The method is tested with data from a synthetic seismic event and then applied to real events selected from an oil and gas field.

Method

The focal mechanism can be represented by a 3 by 3 second order moment tensor (Stein & Wysession, 2003). Usually, as there is no rotation of mass involved in the rupturing process, the tensor is symmetric and has only six independent components. Here we assume the focal

mechanism of the small induced events can be represented by pure double couples (Rutledge & 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 couple (DC) has additional advantages. For example, anisotropy, which is not included by the isotropic Green's functions, often raises spurious non-DC components in focal mechanism determination. By assuming focal mechanism to be DC, this spurious non-DC component can be ruled out while the DCs can be well recovered (_ílen_ & Vavry_uk, 2002). In our analysis we describe the source in terms of its strike, dip and rake, and determine double couple components from these three parameters. For each component of a moment tensor, we can use the Discrete Wavenumber Method (DWN) (Bouchon, 1981, 2003) to calculate its Green's functions $G^n_{ij,k}(t)$. The structure between the earthquake and the station is represented as a 1-D layered medium. The modeled waveform from a certain combination of strike, dip, and rake is expressed as a linear combination of weighted Green's functions:

$$V_i^n = \sum_{j=1}^{3}\sum_{k=1}^{3} m_{jk} G^n_{ij,k}(t) * s(t) \qquad (1)$$

where V_i^n is the modeled i^{th} (north, east or vertical) component at station n; m_{jk} is the moment tensor component and is determined by the data from all stations; $G^n_{ij,k}(t)$ is the i^{th} component of the Green's functions for the (*j, k*) entry at station *n*, and *s(t)* is the source time function. In this study, a smooth ramp is used for *s(t)* (Bouchon, 1981).

Earthquake locations are usually provided by the traveltime location method. However, due to uncertainties in velocity model and arrival times, the seismic event locations may have errors, especially in focal depth. While matching the modeled and observed waveforms, we also search for an improved location around the catalog location.

Before the grid search is performed, we build a Green's functions library. We pre-calculate all $G_{ij,k}^{n}(t)$'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 $G_{ij,k}^{n}(t) * s(t)$, each of which is weighted by m_{jk}.

To determine the best solution, we construct an objective function that characterizes the similarity between the modeled and observed waveforms. We use the following objective function, which evaluates four different aspects of the waveform information:

$$\begin{aligned} &\text{maximize}(J(x,y,z,str,dip,rake,ts)) = \\ &\sum_{n=1}^{N}\sum_{j=1}^{3}\{\alpha_1 \max(\tilde{d}_j^n \otimes \tilde{v}_j^n) - \alpha_2 \left\| \tilde{d}_j^n - \tilde{v}_j^n \right\|_2 \\ &+ \alpha_3 f(pol(\tilde{d}_j^n), pol(\tilde{v}_j^n)) + \alpha_4 h(rat(\frac{S(d_j^n)}{P(d_j^n)}), rat(\frac{S(v_j^n)}{P(v_j^n)}))\} \end{aligned} \tag{2}$$

Here $\tilde{d}_j^n$ is the normalized data and $\tilde{v}_j^n$ is the normalized modeled waveform. Since it is difficult to obtain accurate absolute amplitudes due to site effects in many situations, we normalize the filtered observed and modeled waveforms before comparison. 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:

$$\tilde{d}_j^n = \frac{d_j^n}{\sqrt{\int_{t_1}^{t_2} (d_j^n)^2 \, dt}} \tag{3}$$

where t_1 and t_2 are the boundaries of the time window.

The objective function J in equation (2) consists of 4 terms. α_1 through α_4 are the weights for each term. Each weight is a positive scalar number and is optimally chosen in a way such that no

single term will over-dominate the objective function. The first term in Equation (2) evaluates the maximum cross correlation between the normalized data ($\widetilde{d}_j^n$) and the normalized modeled waveforms ($\widetilde{v}_j^n$). From the cross-correlation, we find the time-shift to align the modeled waveform 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 misalign with wiggles of the next cycle in the modeled waveforms. Therefore, allowed maximum time-shift should be predetermined by the central frequency of the waveforms. The second term evaluates the L_2 norm of the direct differences between the aligned modeled and observed waveforms (note the minus sign of the 2nd term). The reason for maximizing the cross correlation value and minimizing the direct difference between the observed and modeled waveforms is to match both the similarity of the waveforms and the actual amplitudes. The third term evaluates whether the polarities of the first P-wave arrivals as observed in the data are consistent with those in the modeled waveforms. *pol* is a weighted sign function which can be {_, - _, 0}, where _ is a weight reflecting our confidence in picking the polarities of the first P-wave arrivals in the observed data. Zero (0) means undetermined polarity. *f* is a function that penalizes the polarity sign inconsistency in such a way that the polarity consistency gives a positive value while polarity inconsistency gives a negative value. The matching of the first P-wave polarities between modeled and observed waveforms is an important condition for determining the focal mechanism, when the polarities can be clearly identified. Polarity consistence at some stations can be violated if the polarity is not confidently identified (small _) and the other three terms favor a certain focal mechanism. Therefore, the polarity information is incorporated into our objective function with some flexibility. By summing over the waveforms in a narrow window around that arrival time and checking the sign of the

summation, we determine the polarities robustly for the modeled data. For the observed data, we determine the P-wave polarities manually.

The S/P amplitude ratio is also very important in determining the focal mechanism. The fourth term in the objective function is to evaluate the consistency of the S/P amplitude ratios in the observed and modeled waveforms (Hardebeck & Shearer, 2003). The "*rat*" is the ratio evaluation function and it can be written as:

$$rat = \frac{\int_{T_2}^{T_3} \left| r_j^n(t) \right| dt}{\int_{T_1}^{T_2} \left| r_j^n(t) \right| dt} \tag{4}$$

where $[T_1\ T_2]$ and $[T_2\ T_3]$ define the time window of P- and S-waves, respectively , and r_j^n denotes either d_j^n or v_j^n. The term h is a function which penalizes the ratio dif ferences so that the better matching gives a higher value. Note that here we use the un-normalized waveforms d_j^n and v_j^n.

In general, the amplitudes of P-waves are much smaller than those of S-waves. To balance the contribution between P- and S-waves, we need to fit P- and S-waves separately using the first two terms in Equation (2). Also, by separating S- from P-waves and allowing an independent time-shift in comparing observed data with modeled waveforms, it is helpful to deal with incorrect phase arrival time due to incorrect Vp/Vs ratios (Zhu & Helmberger, 1996). We calculate both the first P and S arrival times by the finite difference Eikonal solver (Podvin & Lecomte, 1991). The wave train is then separated into two parts at the beginning of the S wave. To reduce the effect of uncertainty in the origin time, we first align the modeled and observed data using first arrivals and then define the alignment by cross-correlation.

The processing steps can be summarized as follows:

1. Use the known velocity structure to generate a Green's function library;
2. Calculate the first P- and S-wave arrival times using the finite-difference travel time solver (Podvin & Lecomte, 1991), and separate the S-wave segment from the P-wave segment according to the travel time information;
3. For separated P- and S-wave segments
    a. Determine the time-shift by cross-correlation between modeled and observed data
    b. Evaluate the maximum cross correlation value and L_2 norm between the aligned modeled and observed data
    c. Identify the first arrival polarities
    d. Calculate the average amplitudes of the P-wave and S-wave segments;
4. Determine the best fit mechanism by maximizing the objective function.

To find the similarity between modeled and observed waveforms, we do two kinds of basic computations: filtering and cross-correlation. These two computations are very time-consuming when millions of modeled traces are processed. To expedite the computation we use the following manipulations:

$$\begin{aligned} v_i^n &= F * V_i^n = F * \sum_{j=1}^{3}\sum_{k=1}^{3} m_{jk} G_{ij,k}^n(t) * s(t) \\ &= \sum_{j=1}^{3}\sum_{k=1}^{3} m_{jk}^n [F * (G_{ij,k}^n(t) * s(t))] \end{aligned} \tag{5}$$

$$\begin{aligned} d_i \otimes v_i &= d_i \otimes \sum_{j=1}^{3}\sum_{k=1}^{3} m_{jk} F * G_{ij,k} * s(t) \\ &= \sum_{j=1}^{3}\sum_{k=1}^{3} m_{jk} [d_i \otimes (F * G_{ij,k} * s(t))] \end{aligned} \tag{6}$$

where F denotes the impulse response of a filter; "$*$" denotes time domain convolution; d_i^n and v_i^n denote the i^{th} component of the filtered observed and modeled data at station n, respectively; "$\otimes$" denotes the cross-correlation. These two equations indicate that we can apply the filtering and cross correlation into the summation to avoid filtering and cross correlation repetitively during the search over all strikes, dips and rakes. A large amount of time is thereby saved, and the searching speed is boosted by an order of magnitude.

By pre-calculating the library of Green's functions and manipulating the filtering and cross correlation, we greatly 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 10 minutes. The computation of the Green's function library using DWN takes more time, but it only needs to be computed once.

Synthetic data test

We first test the accuracy and robustness of our method using a synthetic data set. We use the same station distribution and velocity model from a field shown in Fig. 1 (Sarkar, 2008). The layered velocity model is obtained from nearby well logs.

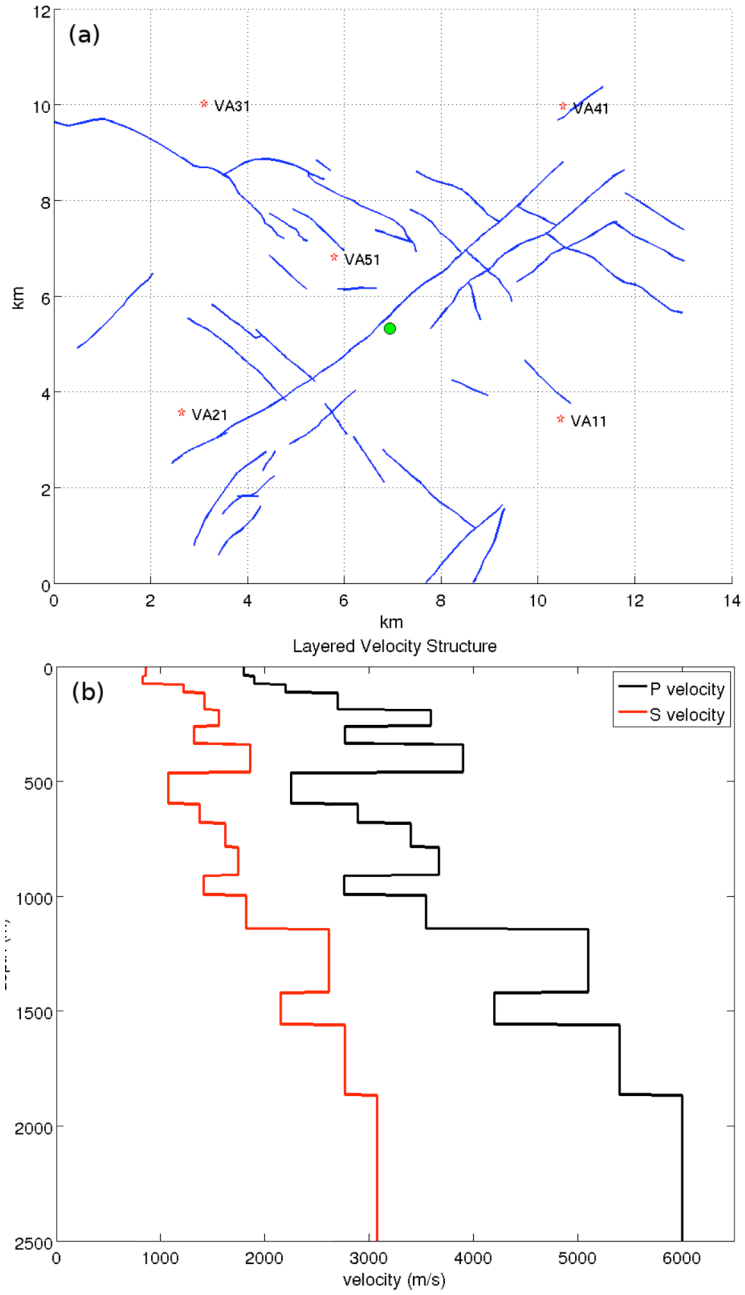

Fig. 1. (a) Locations of stations (red stars) and the synthetic event (green dot). The identified faults, stations and the synthetic event are plotted in a local reference system; (b) Layered velocity structure for P- and S-waves for this region. Velocity data are derived from nearby well logs.

We use the DWN method to generate clean synthetic seismograms and manually pick the first P arrival times, as we will do on the real field data. A source located at 1227 m beneath the surface, with a focal mechanism of strike of 210°, dip of 50° and rake of -40°, is used to generate the synthetic data. The synthetic seismograms are shown in Fig. 2. We use a horizontal grid spacing of 150 m and vertical grid spacing of 50 m with the search range of -900m≤ X,Y ≤900m and -400m≤ Z ≤400m. The reason for choosing smaller vertical grid spacing is that the seismograms are very sensitive to focal depth. Any vertical shift in hypocenter changes the multiple reflection and refraction patterns. The frequency band used in our study is 3-9 Hz, which contains the dominant energy in waveforms recorded from typical induced earthquakes in that field. The searching interval in strike, dip and rake is 10° in this test and, hereafter, in the real data test. This spacing choice indicates that our resolution is 5° at best. Because the auxiliary plane solution and the fault plane solution give the identical waveform, this means that half of the model space [0°≤ strike ≤360°; 0°≤ dip ≤90°; -180°≤ rake ≤180°] is redundant. Therefore, by constraining the model space in [0°≤ strike ≤360°; 0°≤ dip ≤90°; -90°≤ rake ≤90°] (Zhao & Helmberger, 1994), we can eliminate the redundancy and further shorten the search time by half. The weights $_1$ through $_4$ in the objective function (Equation 2) were tried with different values, and we selected ones that balance different terms. We used $_1$=3, $_2$=3, $_3$=1 and $_4$=0.5 for the synthetic tests and real events later. We also found that the final solutions are not very sensitive to small changes in the weights. The results for the synthetic test are summarized in Table 1. The first 200 best solutions are used for statistical analysis of strikes, dips, rakes and locations. It is shown that even for the perfect data, the ambiguities (one standard deviation) in strike, dip and rake are about 10º because only vertical components are involved, and the station coverage is sparse. Among strike, dip and rake, dip has the least standard deviation while strike has the largest. As a small variation in depth changes the

reflection and refraction patterns considerably, our algorithm determines the depth (Z) without much ambiguity.

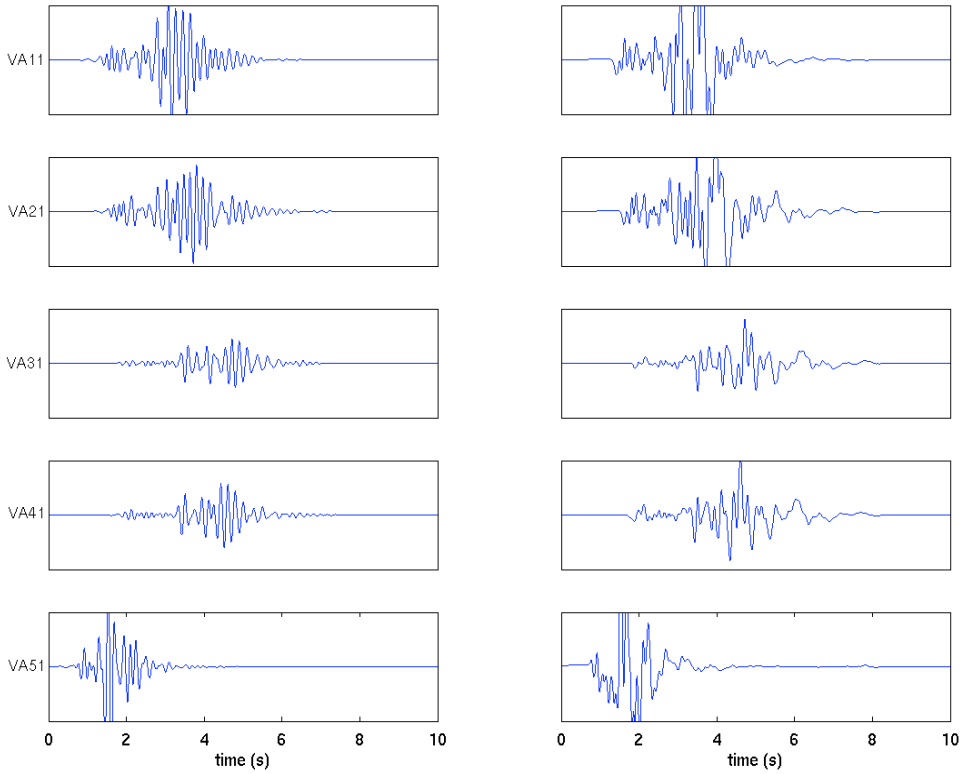

Fig. 2. Synthetic seismograms (vertical component) at the 5 stations. The left column shows the filtered seismograms (3~9 Hz), and the right column shows the unfiltered ones. The source is at 1227 m in depth, and has a strike of 210°, dip of 50° and rake of -40°.

	Strike (°)	Dip (°)	Rake (°)	X (m)	Y (m)	Z (m)
Mean	203	47	-45	-30	10	0
Std.	13	9	11	110	130	0

Table 1. Statistics of focal mechanism parameters in the "clean" synthetic test. The hypocenter is re-coordinated as (0, 0, 0) m for discussing the mean and standard deviation, and the source has a

strike of 210°, dip of 50° and rake of -40°. The best solution has the identical focal parameters with the source.

We further test the robustness of the method by adding noise to the synthetic data. We add white spectrum Gaussian noise to each trace, with zero mean and a standard deviation of 5% of the maximum absolute amplitude of that trace. This level corresponds to the typical noise level we encounter for real data.

Fig. 3 shows the focal mechanism determined using waveform information and 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) matches the correct solution. Fig. 4 shows the comparison between the modeled and synthesized waveforms with noise contamination. The "shift" in the title of each subplot indicates the time shifted in the data to align with the synthetic waveforms. The reasons for having some time shift are as follows: 1) we introduced some artificial error in arrival time by manually picking the first P arrival in the synthetic data; 2) scattering noise can change the maximum cross correlation position (Nolet et. al., 2005). In the left column, the "+" or "-" signs indicate the first arrival polarities of P-waves in the data and those in the synthetics; the upper ones are signs for the synthetic data while the lower ones are signs for the modeled data. The modeled traces all have the identical polarities as their counterparts in the synthetic data. Note that for the evaluation of the polarities, we use the unfiltered waveforms, as filtering usually blurs or distorts the polarities. In the right column, the number to the left of the slash denotes the S/P ratio for the data, and the number to the right of the slash denotes the ratio for the modeled waveform. They are quite close in most cases.

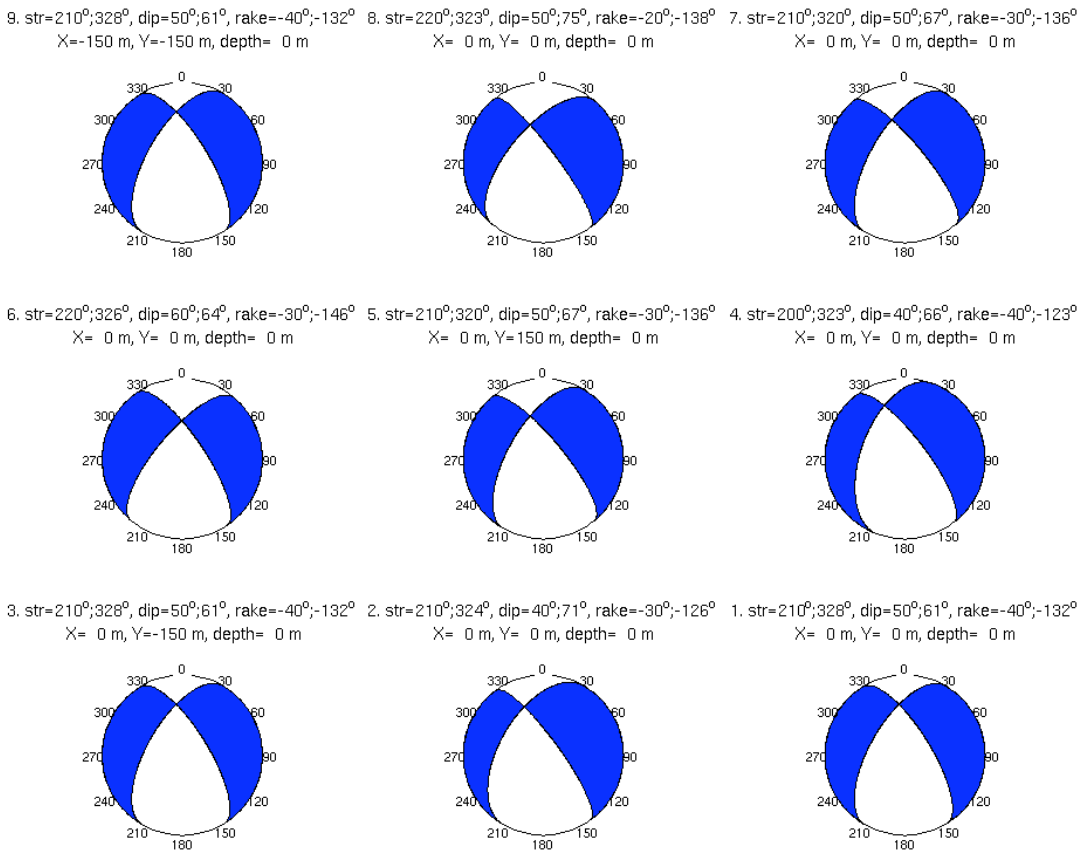

Fig. 3. Nine best solutions from contaminated synthetic data; the number before "str" is the order; "1" means the best solution and "9" means the worst solution among these nine.

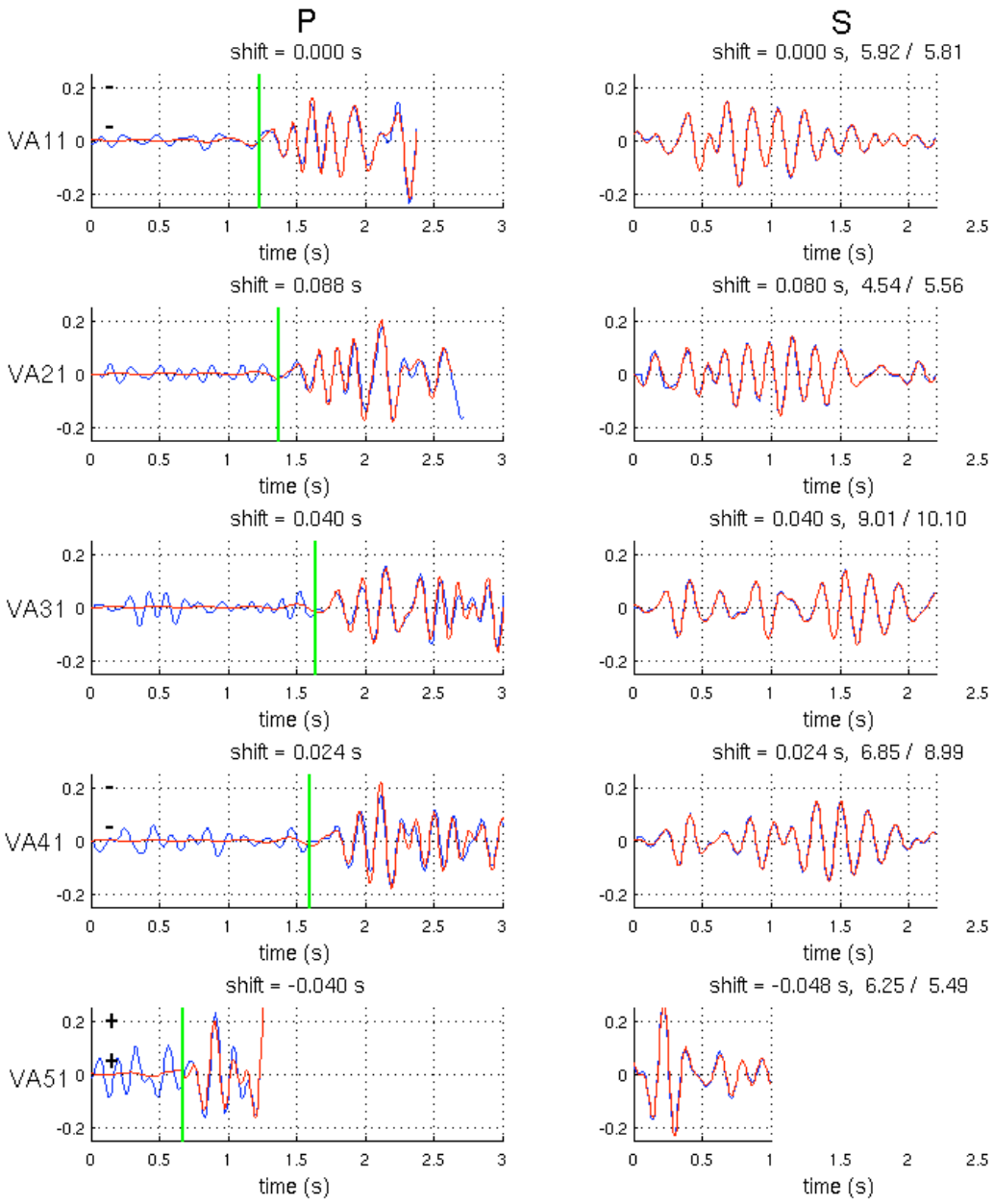

Fig. 4. Comparison between modeled waveforms (red) and noisy synthetic data (blue) at 5 stations. From top to bottom waveforms from the vertical components at stations 1 through 5, respectively, are shown. The left column shows P-waves and right column shows S-waves. The green lines indicate the first P arrival times. For P-waves, zero time means the origin time, and for S-waves, zero time means the S-wave arrival time predicted by the calculated travel time.

We further analyzed the distribution of strikes, dips, rakes, and locations from the first 200 best determined solutions (Table 2). The strike, dip, and rake all have a mean quite close to the correct solution (210°, 50° and -40°, respectively). Among the strike, dip, and rake, dip has the least standard deviation. We find that in the X and Y directions the variation is much larger than that in the Z direction, similar to the case of clean data. The different standard deviations might be an indication of the sensitivity of these model parameters. Therefore, using waveform information makes it possible to obtain accurate focal depth.

	Strike (°)	Dip (°)	Rake (°)	X (m)	Y (m)	Z (m)
Mean	203	45	-41	-20	-10	0
Std.	13	10	11	120	130	0

Table 2. Statistics of focal mechanism parameters in the "noisy" synthetic test. The source location and mechanism are the same as the test in Table I. The best solution has the identical focal parameters with the source.

Application to induced seismicity at an oil field

We applied this method to study earthquakes at a petroleum field which are induced by stress change due to water injection and gas/oil extraction. The earthquakes are small ($-0.50 < M_w < 1.0$), and the dominant energy in the recorded seismograms is between 3 and 15 Hz. Fig. 5 shows a typical event recorded at these stations and its spectrograms. During the period of 1999 to 2007, over 1500 induced earthquakes were recorded by a 5-station near-surface network, and their occurrence frequency was found to be correlated with the amount of gas production (Sarkar, 2008).

Two major fault systems have been identified in this area, with one oriented in the NE-SW direction, and the other oriented in the conjugate direction (NW-SE) (Fig. 6).

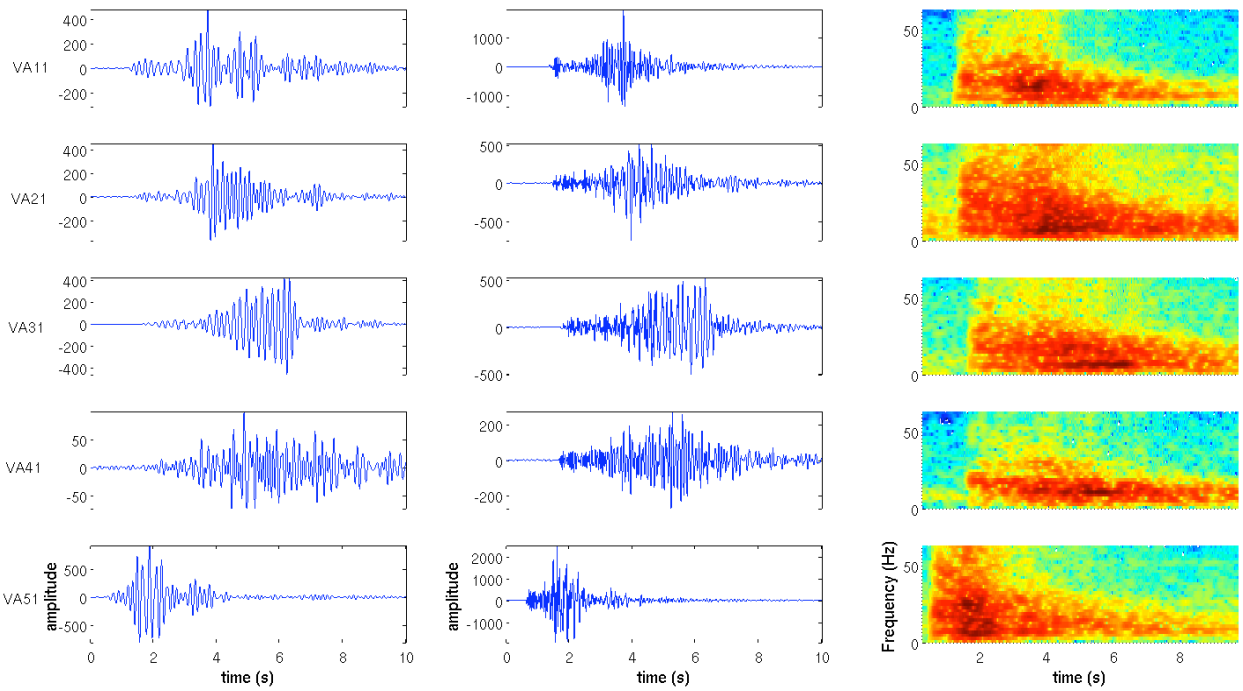

Fig. 5. A typical event used in the focal mechanism determination and its spectrograms. The seismograms are from the vertical components of these five stations. The filtered seismograms (3~9 Hz) are at the left column; the original seismograms are in the middle; the spectrograms of the original seismograms are at the right.

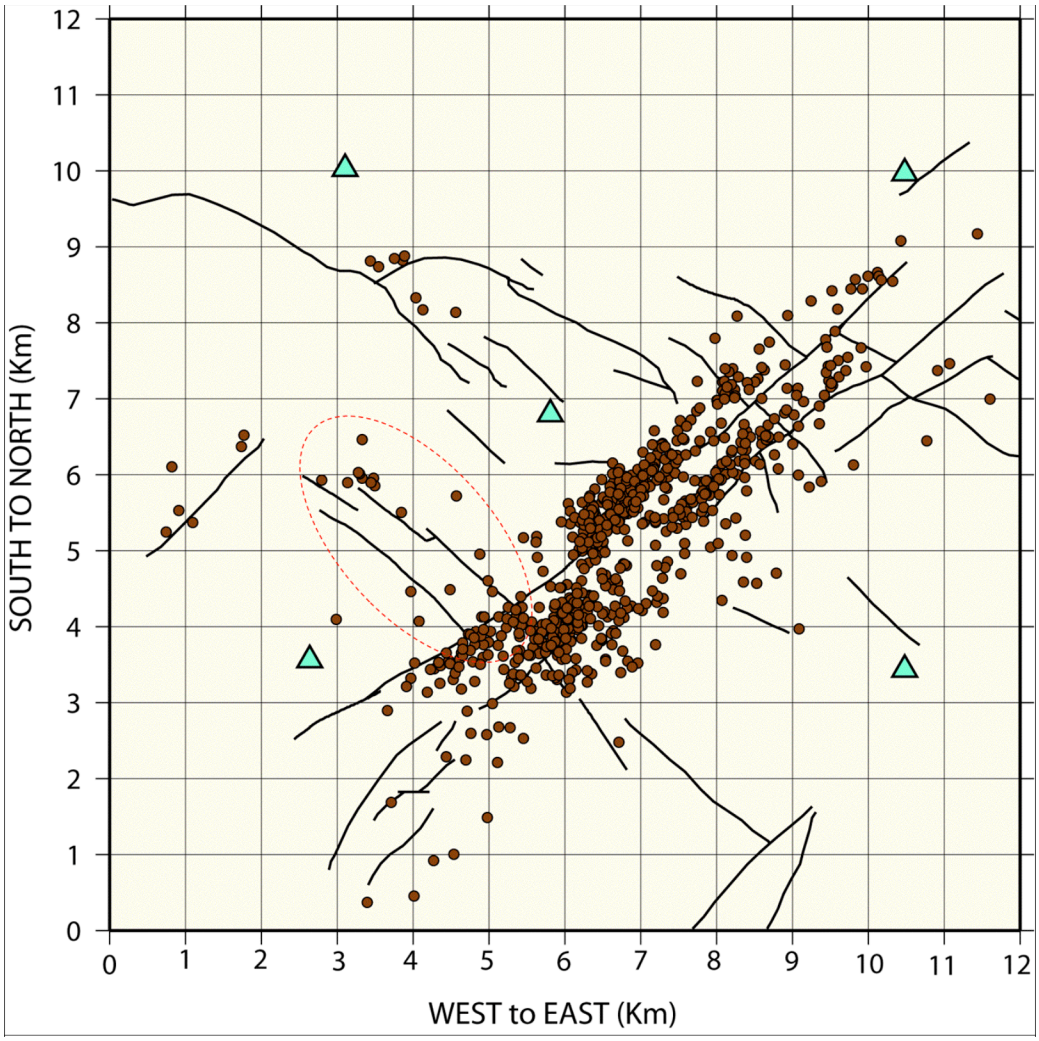

Fig. 6. Distribution of over 1000 located induced earthquakes. Note that the majority of these earthquakes are in the proximity of the NE-SW fault. A group of induced earthquakes within the dashed circle may indicate the activation of a conjugate fault (Sarkar, 2008). The five stations are indicated with green triangles, and station names are shown in Fig. 1(a).

The distribution of induced events in the field is shown in Fig. 6 (Sarkar, 2008; Sarkar et al., 2008; Zhang et al., 2009). All the events have a residual travel time less than 30 ms, indicating they are well located. First we describe the analysis procedure for one event. Fig. 7 shows the change of objective function value with the best solution order. For this event, the value is about 13 for the

best solution and decreases to about 6 for the 200th best solution. The objective function value decreases quickly from the 1st to the 200th best solution but relatively slowly beyond this range. Therefore, we choose 200 as the pool size for evaluating the statistics of focal mechanism parameters. The synthetic tests shown in the previous section have a similar objective function value distribution to the real data. Fig. 8 shows the beachballs of the nine best solutions out of millions of trials. Our best solution (the one at the bottom right, reverse strike-slip) has a strike of 199°, which is quite close to the best known orientation 219° of the NE-SW fault (Fig. 1). However, the faulting (auxiliary direction 325°) could occur in the conjugate NW-SE fault instead, as the auxiliary plane has a strike almost parallel with the conjugate fault. Using our new algorithm, the epicenter is shifted northward by about 750 m, eastward by about 300 m and the depth is shifted 50 m deeper. The shift in epicenter may be biased by errors in first P arrival picking or biased by inaccuracy in the velocity model. As has been discussed before, the shift in epicenter may compensate the phase shift in the modeled seismograms due to inaccuracy in the velocity model.

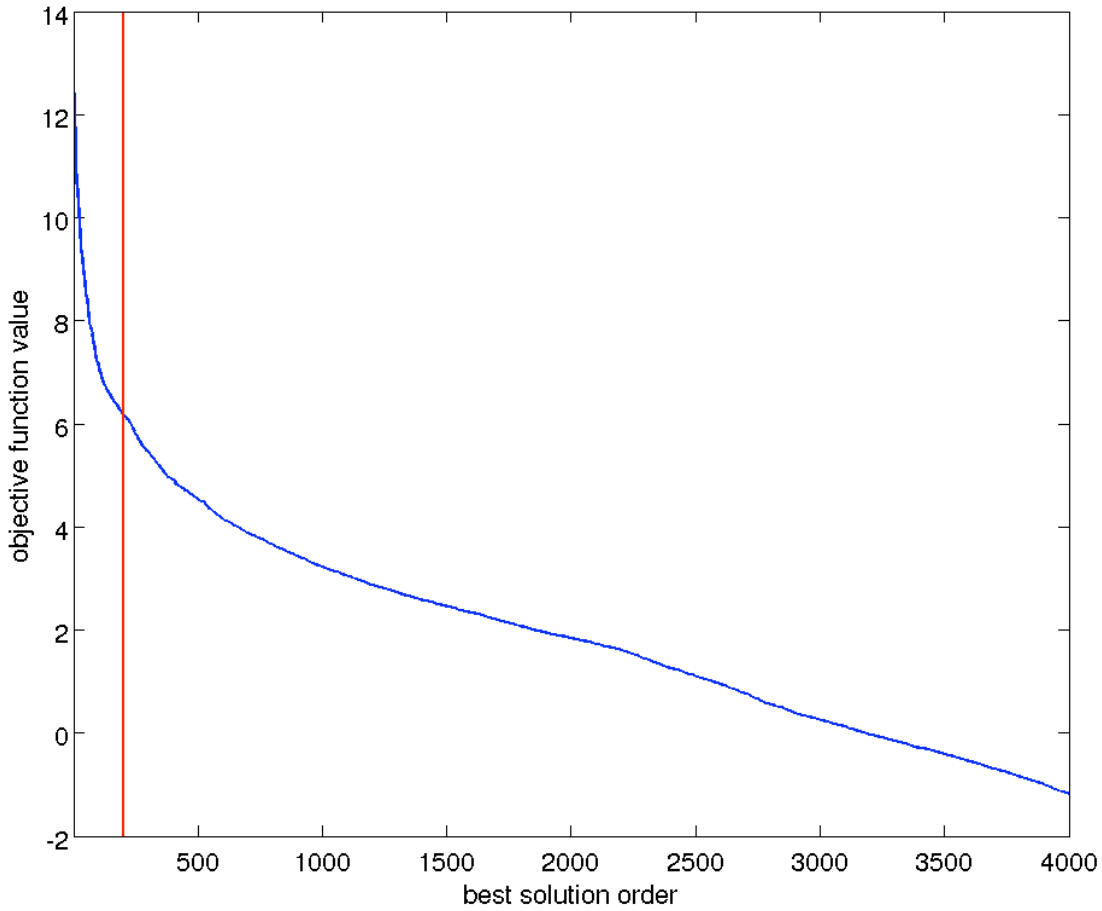

Fig. 7. Objective function value vs. best solution order. To the left of the red line are those solutions used to evaluate the statistics of focal mechanism and location parameters.

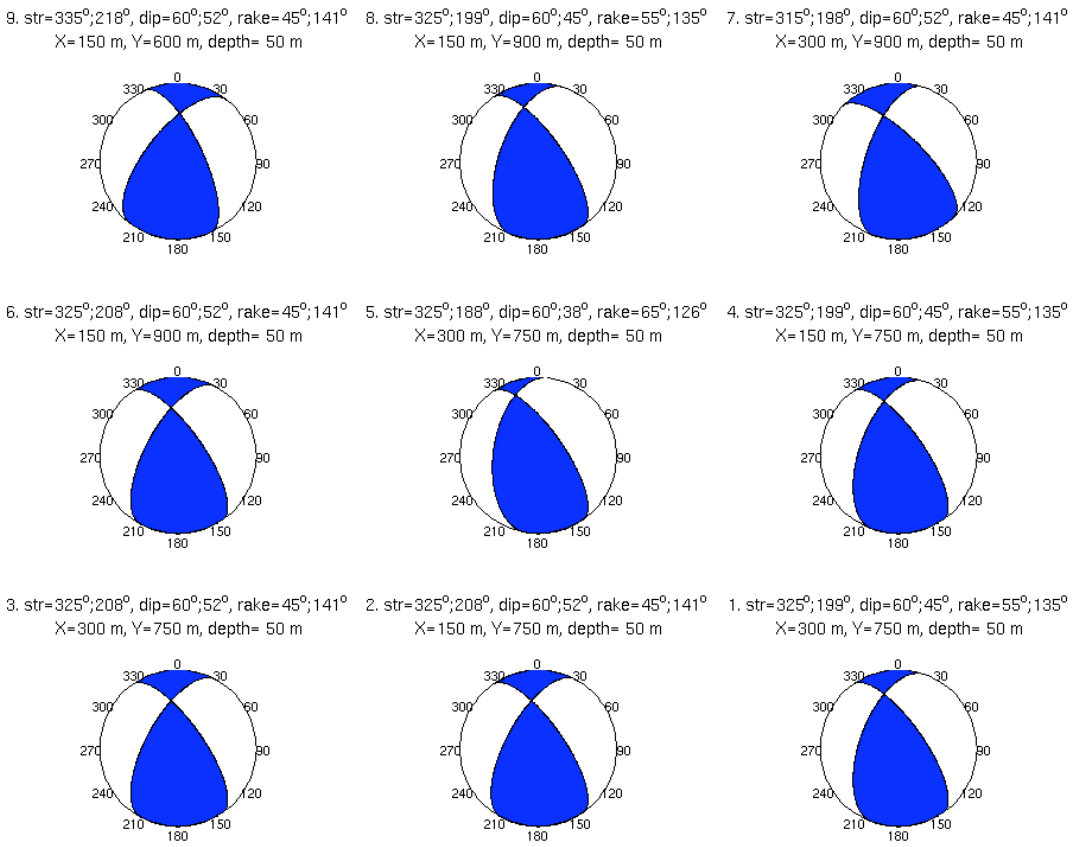

Fig. 8. Focal mechanism solutions for the event 20010047. The one at the bottom right (#1) is the best solution with maximum objective function value.

Fig. 9 shows the comparison between the modeled and the observed data for event 20010047. The waveform similarities between the modeled and observed data are good. Additionally, the S/P waveform amplitude ratios in the modeled and observed data are quite close, and the first P arrival polarities are identical in the modeled and observed data for each station. In this example, all four criteria in Equation (2) are evaluated, and they are consistent between the modeled and observed data. In some situations, the observed and modeled waveforms may look similar, even when the S/P amplitude ratio and first P-wave polarity do not match. When this

happens, our method would not accept this solution. This indicates that it is sometimes misleading to use only waveform matching to determine the focal mechanism, as a wrong solution might still give satisfactory waveform matching, especially when data from a sparse network are used. Table 3 shows the distribution of strike, dip, rake and location. Again, we find dip has the minimum standard deviation. Similar to the synthetic test, depth has the least variation in this case. As discussed before, variation in epicenter (X and Y) can be compensated by shifting the observed waveforms to find a better alignment with the modeled data. Therefore, the constraint on lateral shift is weaker compared to that in the vertical direction.

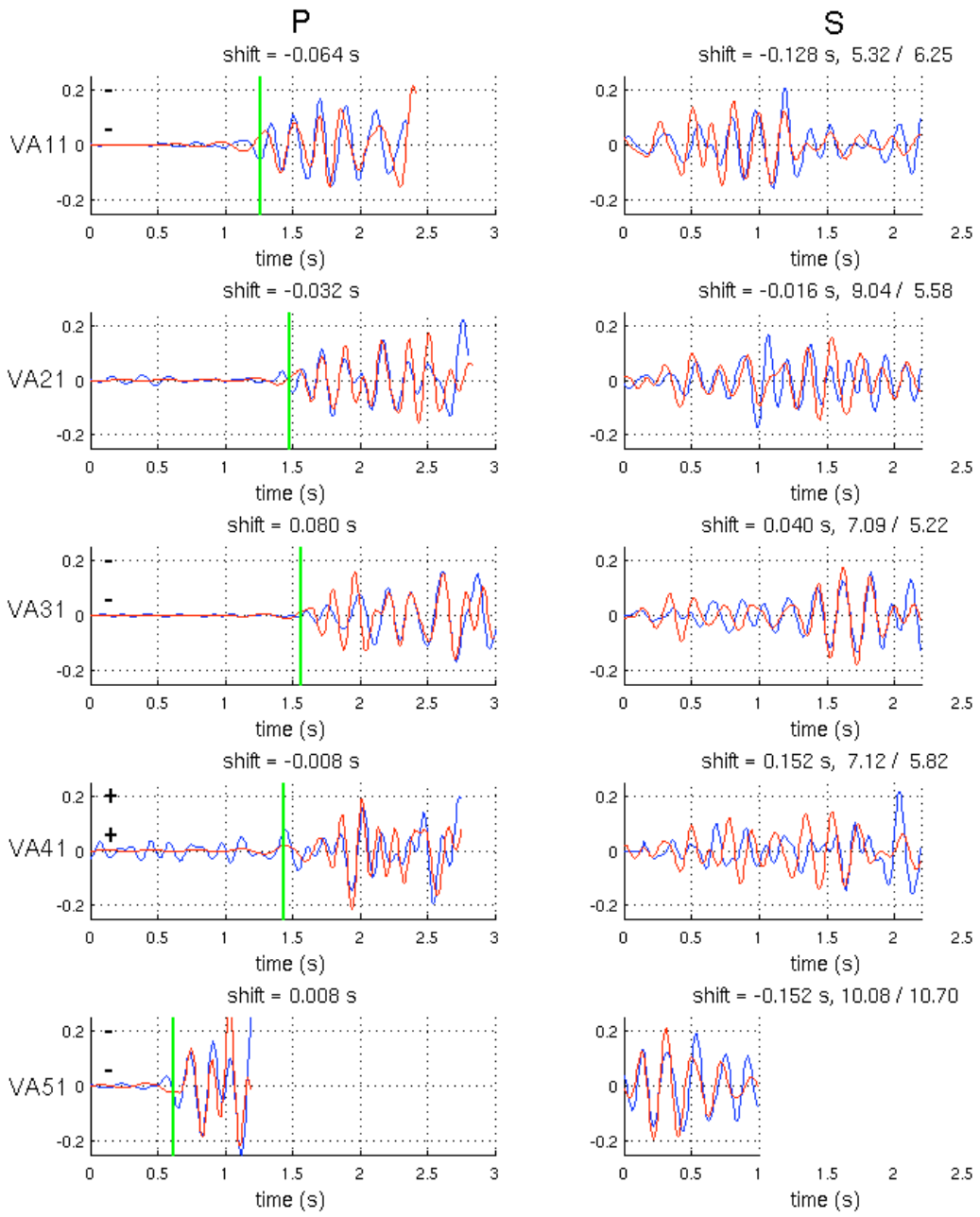

Fig. 9. Comparison between the modeled waveforms (red) and the real data (blue) at 5 stations for the event 20010047. For P-waves, zero time means the origin time, and for S-waves, zero time means the S-wave arrival time predicted by the calculated travel time.

	Strike (°)	Dip (°)	Rake (°)	X (m)	Y (m)	Z (m)
Mean	331	56	59	210	700	50
Std.	12	8	13	100	160	0

Table 3. Statistics of focal mechanism parameters for the event 20010047. The best solution has a strike of 325°, dip of 60°, and rake of 55°. In discussing the mean and standard deviation, the hypocenter is re-coordinated as (0, 0, 0) m. The epicenter of event 20010047 is shown in Fig. 10 in a local coordinate.

Using this method, we have studied 22 earthquakes distributed along the NE-SW fault at this petroleum field. These 22 events are from the data collected between 2000 and 2002. The distribution of located induced earthquakes is shown in Fig. 6. The determined focal mechanisms are shown in Fig. 10. In this petroleum field, we expect some lateral velocity variations, especially across the faults (Zhang et al., 2009). We tested the robustness of our method by perturbing the P and S velocity model for each station independently to simulate lateral velocity heterogeneity (Appendix A). The test showed that we are able to obtain a focal mechanism that is quite close to the correct solution when lateral velocity variation exists. This is because the new method incorporates different aspects of waveform information and compensates for the phase shift due to velocity heterogeneity using the time-shift.

Fig. 10 shows that the majority of the events primarily have the normal faulting mechanism, although there are also a few reverse faulting events. The dominance of the normal faulting mechanism suggests that the vertical stress is greater than the horizontal stress oriented in the NE-SW direction, parallel with strike of the main fault. Among these events, over half have a strike oriented in the NE-SW direction. However, the strike of some faults is in the conjugate fault

direction (NW-SE). The dual orientations suggest that both fault systems are probably still active, and the seismicity pattern shown in Fig. 6 also supports this indication (e.g., the induced earthquakes in the red circle).

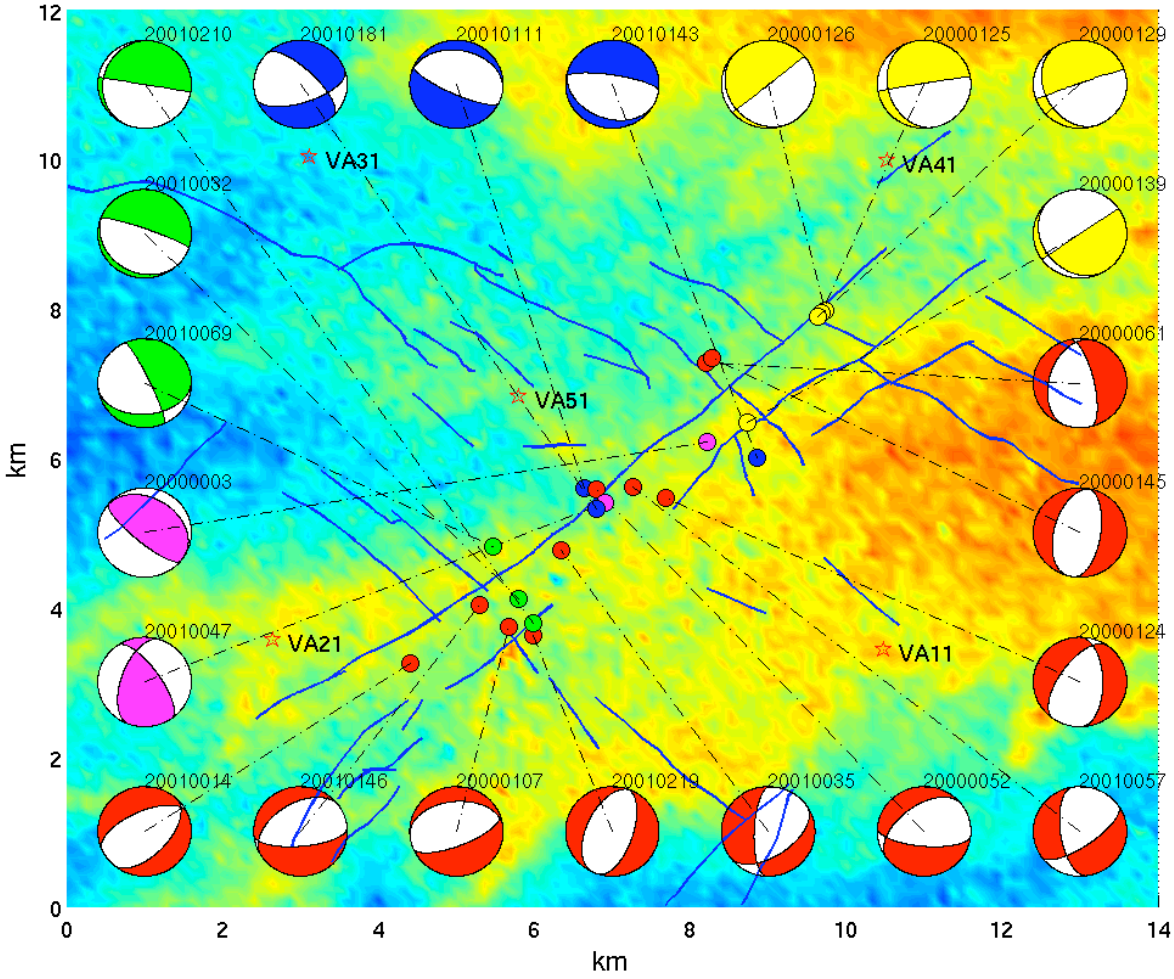

Fig. 10. Focal mechanisms of the 22 events inverted in this study for an oil field. The color in the map indicates the local change in surface elevation with a maximum difference of about 10 m. Different focal mechanisms are grouped in several colors.

Conclusions

In this study, we showed that combining the high frequency seismograms with a fast optimized grid-search algorithm leads to determination of focal mechanisms and locations of small earthquakes, where subsurface velocity information is available. This 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, or only the vertical components are usable. Additionally, because of the normalization in our algorithm, the method does not require the absolute amplitude to be available and, therefore, mitigates the influence of the site effects. The objective function, formulated to include matching phase and amplitude information, first arrival P polarities and S/P amplitude ratios between the modeled and observed waveforms, yields stable solutions. The synthetic tests prove that the method is robust: giving correct solutions in the case of noise contamination or lateral velocity variations.

For real events, we find that the focal mechanisms are consistent with local geological structure and are indicative of local stress distribution. Focal mechanisms for 22 induced earthquakes are mostly normal faulting. A majority of the events are on the NE-SW striking faults, and their mechanisms are consistent with these faults. A few events have strikes in the NW-SE direction. These are events on the conjugate faults. In the region where both the NE trending faults and the conjugate (NW trending) faults exist, the focal mechanisms make it possible to determine with which faults the seismic events were associated.

Acknowledgement

The authors want to thank the support of the ERL consortium members.

References

Bischoff M., Cete A., Fritschen R. & Meier T., 2010. Coal Mining Induced Seismicity in the Ruhr Area, Germany, *Pure and Applied Geophysics*, **167**, 63-75.

Bouchon, M., 1981. A simple method to calculate Green's functions for elastic layered media, *Bull. Seism. Soc. Am.*, **71**, 959-971.

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

Chan, A.W. & Zoback, M.D., 2007. The role of Hydrocarbon production on land subsidence and fault reactivation in the Louisiana coastal zone, *Journal of Coastal Research*, **23**, 771-786.

Deichmann, N. & Giardini, D., 2009. Earthquakes induced by the stimulation of an enhanced geothermal system below basel (Switzerland), *Seismological Research Letters*, **80**, 784-798.

Dreger, D., Uhrhammer, R., Pasyanos, M., Franck, J. & Romanowicz, B., 1998. Regional and far-regional earthquake locations and source parameters using sparse broadband networks: a test on the Ridgecrest sequence, *Bull. Seism. Soc. Am.*, **88**, 1353-1362.

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

Hardebeck, J.L. & Shearer, P.M., 2003. Using S/P amplitude ratios to constrain the focalmechanisms of small earthquakes, *Bull. Seis. Soc. Am.*, **93**, 2434-2444.

Julià, J. & Nyblade, A.A., 2009. Source mechanisms of mine-related seismicity, Savuka mine, South Africa, *Bull. Seism. Soc. Am.*, **99**, 2801-2814.

Miyazawa, M., Venkataraman, A., Snieder, R. & Payne, M.A., 2008. Analysis of microearthquake data at Cold Lake and its applications to reservoir monitoring, *Geophysics*, **73**, O15-O21.

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

Pasyanos, M.E., Dreger, D.S. & Romanowicz, B, 1996. Toward real-time estimation of regional moment tensors, *Bull. Seism. Soc. Am.*, **86**, 1255-1269.

Podvin, P. & Lecomte, I., 1991. Finite difference computation of traveltimes in very contrasted velocity models: a massively parallel approach and its associated tools, *Geophys. J. Int.*, **105**, 271-284.

Rutledge, J.T. & Phillips, W.S., 2002. A comparison of microseismicity induced by gel-proppant and water-injected hydraulic fractures, Carthage Cotton Valley gas field, East Texas, *72nd Annual International Meeting, SEG, Expended Abstracts*, 2393-2396.

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

Rutledge, J.T., Phillips, W.S. & Mayerhofer, M.J., 2004. Faulting induced by forced fluid injection and fluid flow forced by faulting: and interpretation of hydraulic-fracture microseismicity, Carthage Cotton Valley gas field, Texas, *Bull. Seism. Soc. Am.*, **94**, 1817-1830.

Sarkar, S., 2008. Reservoir monitoring using induced seismicity at a petroleum field in Oman, PhD thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, US.

Sarkar, S., Kuleli, H.S., Toksoz, M.N., Zhang, H.J., Ibi, O., Al-Kindy, F. & Al Touqi, N, 2008. Eight years of passive seismic monitoring at a petroleum field in Oman: a case study, *78th Annual International Meeting, SEG, Expended Abstracts*, 1397-1401.

Segall, P., 2010. Earthquake and volcano deformation, Princeton University Press, Princeton, New Jersey, US.

ílen, J., & Vavry_uk, V., 2002. Can unbiased source be retrieved from anisotropic waveforms by using an isotropic model of the medium? *Tectonphysics*, **356**, 125-138.

Stein, S. & Wysession, M., 2003. An introduction to seismology, earthquakes, and earth structure, Blackwell publishing, Malden, Massachusetts, US.

Suckale, J., 2010. Induced seismicity in hydrocarbon fields, Chapter 2, *advances in Geophysics*, **51**.

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

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

Zhang, H.J., Sarkar S., Toksoz, M.N., Kuleli H.S. & Al-Kindy F., 2009. Passive seismic tomography using induced seismicity at a petroleum field in Oman, *Geophysics*, **74**, WCB57-WCB69.

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

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

Zoback M.D., 2007. Reservoir Geomechanics, Cambridge University Press, Cambridge, UK.

Appendix A

The effect of the velocity model on the focal mechanism determination

We test our new algorithm in the case when the 1-D velocity model is not a satisfactory approximation of the realistic subsurface structures, as lateral velocity variation exists locally. In general, we have a reliable velocity model for both P- and S- waves from well logs (Sarkar, 2008). However, because lateral velocity heterogeneity is inevitable, the influence of inaccuracy in the velocity model on the focal mechanism determination also needs to be examined. We still use a 1-D layered model to generate the synthetic data from the event to a station. However, for each layer, a perturbation on the velocity equal to 5% of the layer's reference velocity is added, and the perturbation is independent for all five stations. The density here is not perturbed in this test, as the velocity perturbation is dominant in determining the characteristics of the waveforms. Also, the layer thickness is not perturbed, as perturbation in either layer velocity or thickness generates equivalent phase distortions from each layer. The perturbations in both P and S velocity models are shown in Fig. A1. Note the perturbations in P and S velocity are independent of each other, and are also independent for each layer and station.

Fig. A1 shows that the variation in velocity for a certain layer from one station to another can be rather large. Considering the region in our study is less than 10 km by 10 km, and it is mainly composed of flat sedimentary layers, this variation should be a reasonable approximation of the real maximum lateral heterogeneity. Here we still use only three first P arrival polarities. Fig. A2 shows the best nine beachball solutions. Encouragingly, the solutions do not differ significantly from the correct one (strike = 210°, dip=50°, rake=-40°). The strikes, dips and rakes in these solutions differ from the correct one by about 10° and one grid spacing (150 m) in the horizontal direction. All of these solutions find the correct depth.

Fig. A3 shows the comparison between the synthetic waveforms generated from the reference model and the synthetic 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 larger than in the previous case, especially for the S-wave comparison). 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 that is quite close to the correct solution.

Table A1 shows the distribution of strike, dip, and rake of the focal plane solution and X, Y, and depth of the location from the first 200 best solutions. The mean values are rather close to the correct solutions, and the standard deviations are small, considering that only the vertical components of 5 stations and 3 polarities are used in determining the focal mechanism. Again, the depth variation is strongly constrained, and this indicates that using waveform information can greatly help locate the depth of an event, which is most difficult to constrain in the traditional traveltime based location method.

	Strike (°)	Dip (°)	Rake (°)	X (m)	Y (m)	Z (m)
Mean	194	42	-46	80	10	0
Std.	17	10	14	150	170	0

Table A1. Statistics of focal mechanism parameters in the velocity perturbed synthetic test. The source location and focal mechanism are the same as the clean synthetic test.

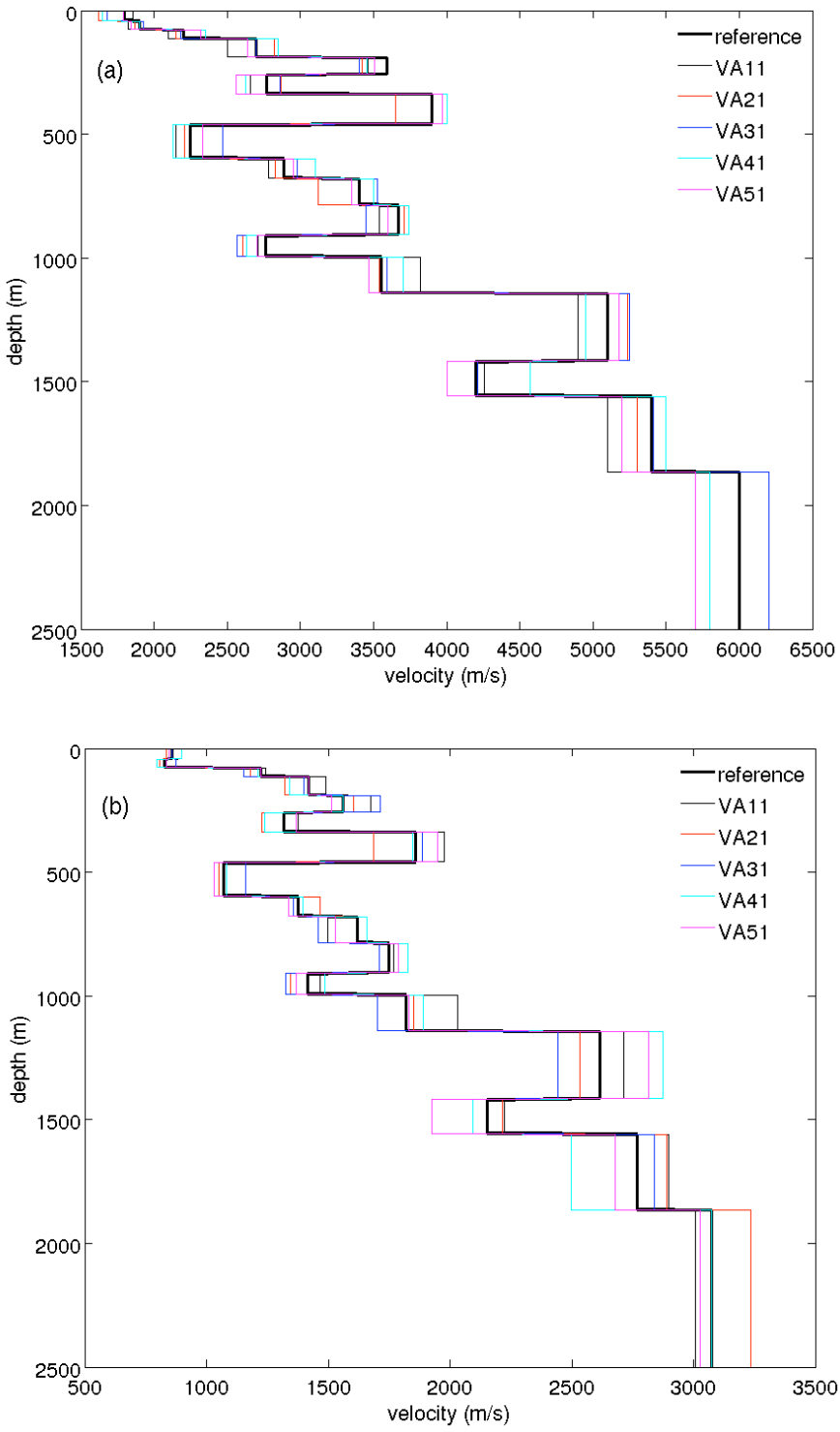

Fig. A1. Perturbations in the P and S velocity model. a) P-wave velocity perturbation. b) S-wave velocity perturbation. The bold black line is the reference model, while the colored lines are perturbed velocity structure for modeling seismograms in each station.

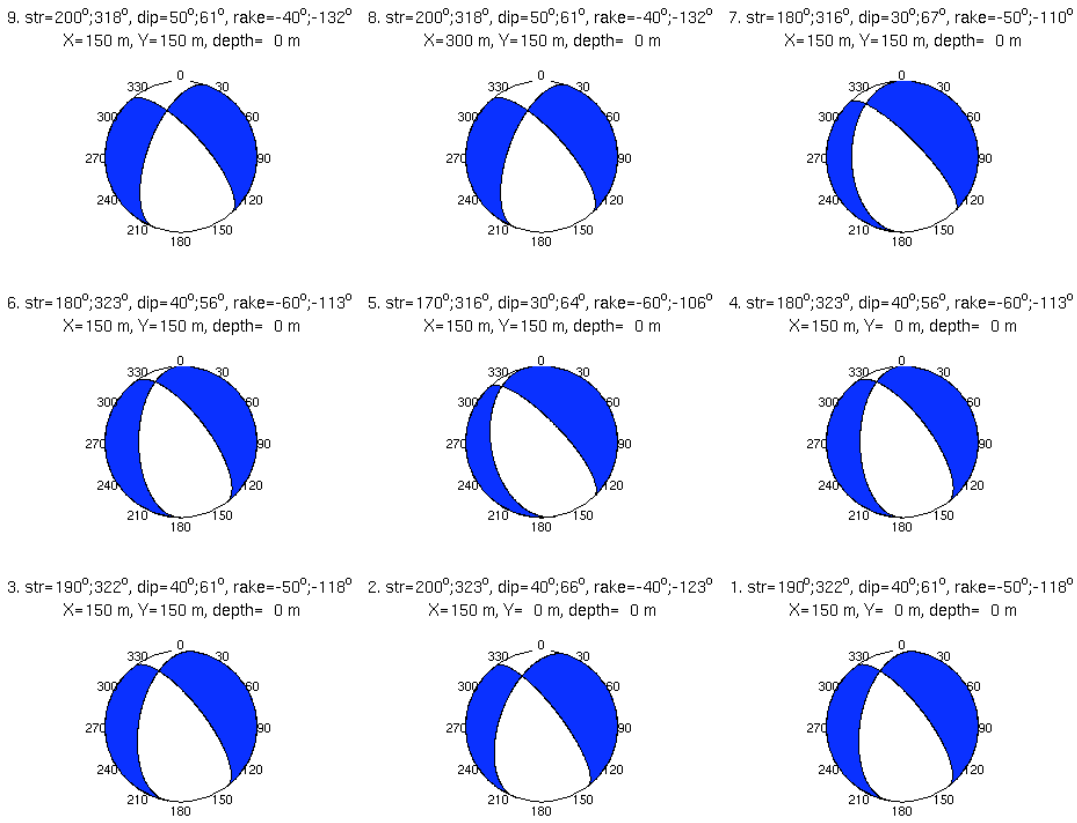

Fig. A2. Focal mechanism solutions when lateral velocity variation exists. The best solution here (#1) does not perfectly match the correct solution but is quite close.

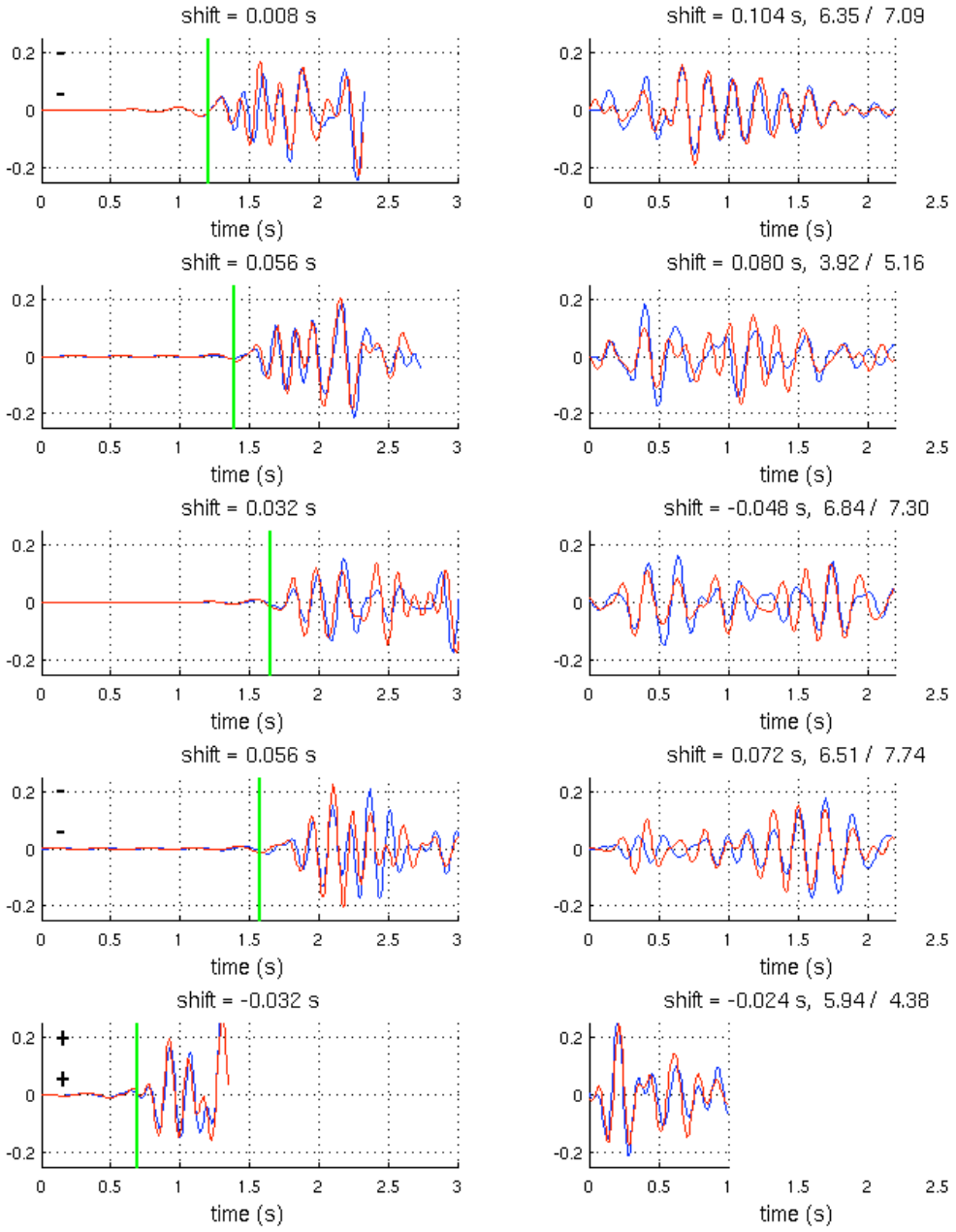

Fig. A3. Comparison between the modeled and synthetic data when lateral velocity variation exists. Phase mismatch between the modeled and synthetic data is similar to the difference in the real event 20010047.