CWP-640
September 2009

Probabilistic micro-earthquake location for reservoir monitoring

Ran Xuan

— Master's Thesis —

Geophysics

Defended on September 18, 2009

Advisor:	Prof. Paul Sava
Co-advisor:	Prof. Roel Snieder

Center for Wave Phenomena
Colorado School of Mines
Golden, Colorado 80401
(1) 303 384-2178

Abstract

Micro-seismicity is used to monitor fluid migration during reservoir production and hydro-fracturing operations. This is usually done with sparse networks of seismic sensors located in boreholes. The data used for micro-earthquake monitoring are usually corrupted by noise which reduces the signal-to-noise ratio to values as low as 0.1. Monitoring methods based on travel-time picking of various wave modes (P or S) cannot deal with this level of noise and require extensive user interaction. An alternative class of methods uses time reversal to focus micro-earthquake information at the source position. These methods can handle noisier signal but are also costlier to run. The technique advocated in this paper exploits time-reversal within the general framework of Bayesian inversion. Given an assumption about the possible locations of micro-earthquakes, I use recorded data to evaluate the probability of micro-earthquakes occurring at various locations in the Earth. The method takes into account imaging imperfections due to unknown components of the model or acquisition array aperture. I simulate wavefields corresponding to possible sources distributed in the model and evaluate their match with the wavefield reconstructed from data recorded in the field. The method operates like a pattern-recognition procedure and can exploit a wide variety of techniques designed for this purpose. I use simple cross-correlation because of the speed and robustness of this technique. The wavefields reconstructed at various locations are used to scan over time the wavefield constructed from field data; thus, my method is able to identify not only the positions of the micro-earthquakes but also their onset times. The final outcome of this automated process is a map of probabilities indicating the confidence of micro-earthquake occurrence at various positions and times.

Table of Contents

Acknowledgments

This work was supported by ExxonMobil URC. We are grateful to Mike Payne, Jie Zhang, Rongrong Lu, Alex Martinez, and Anupama Venkataraman at ExxonMobil URC for allowing us access to data and for their technical discussions and financial support of this project. I would like to thank my adviser Paul Sava for patiently guiding this research, my co-adviser Roel Snieder for the technical discussions, and all of my colleagues.

Chapter 1

Introduction

High-pressure fluid injected into oil and gas reservoirs causes temporal stress and strain changes. When the stress exceeds a threshold that characterizes the resistance of rock materials, micro-seismicity is triggered by the release of pressure along pre-existing fractures or through the creation of new fractures (King & Willis, 1957; Tabeli & Cornet, 1987; Talebi *et al.*, 1998a,b; Maxwell & Urbancic, 2001; Duncan, 2005). Two kinds of physical processes within the reservoir may results depending on the injection pressure (Fischer *et al.*, 2008). The first possibility is that when the introduced pressure is less than the minimum compressive stress, the fluid penetrates the reservoir along the pre-existing fracture (House, 1987). The second possibility is that when the introduced pressure is larger than the minimum compressive stress, the fluid propagates along a tensile fracture which is caused by imbalanced pressure (Evans *et al.*, 2005).

By accurately locating the source of micro-seismic events, we can map micro-seismicity and determine the width, length, and orientation of the fracture which is either pre-existing or created by the introduced pressure (Phillips, 1998; Duncan, 2005). Thus, precise locations of micro-seismic events can be used to monitor hydraulic fracturing (Rentsch, 2004).

There are two categories of methods for locating micro-seismic events. One is the arrival-time-based method; the other is the migration-based method. The arrival-time-based methods are the traditional methods used in locating earthquakes. This category of method requires the identification of seismic arrivals which involves accurate picking of P- and S-wave arrival times. The onset times and the coordinates of the hypocenter of micro-seismic events are given by calculations which require accurate knowledge of the velocity model and of the physical relationships describing wave propagation in the subsurface.

By assuming that there is a linear relationship between the travel time t and the distance

between the location of an earthquake and the station point, we calculate the predicted travel time by using the equation $t = ls$, where s is the slowness which characterize the Earth model. The source is located by optimizing a misfit function between measured and predicted travel times (Shearer, 1997; Aki & Richards, 2002). However, this category of methods is non-linear and uses least-square hypocentral estimates. Therefore the method strongly depends on the network of stations, the accuracy of the arrival time picking, the velocity model, and the theory accuracy (Pavlis, 1986). In order to eliminate or minimize the imprint of the velocity model, relative earthquake location methods were developed based on the assumption that their waveforms are similar or identical to each other for a cluster of micro-earthquakes. By calculating the arrival time difference of a pair of micro-earthquakes from the same station, we locate the micro-earthquake locations relative to the reference event (Poupinet *et al.*, 1984; Fremont & Malone, 1987; Waldhauser, 2000). The relative earthquake location methods also depend on the picking accuracy of arrival times.

In order to achieve high picking accuracy, manual picking of the arrival time is performed. However, this method is time consuming and has low automation. Thus, many automatic picking algorithms have been developed in recent years to replace the manual picking method. One automatic picking method used for relative earthquake methods is waveform cross-correlation. Assuming that the waveform of the events in the same cluster is similar, we cross-correlate the seismogram with the reference waveform and pick the arrival times (Schaff *et al.*, 2004; Schaff & Waldhauser, 2005; Schaff, 2008). Short-term-average and long-term-average (STA/LTA) is another category of methods commonly used. STA/LTA is a measurement of local signal-to-noise ratio. Four parameters are required before the calculation: the STA and LTA window lengths, the length of the Hanning filter which is used to smooth the STA/LTA, and a threshold value which is used to pick the arrival time. When the smoothed STA/LTA exceeds the threshold value, a corresponding arrival time is calculated (Earle & Shearer, 1994). In addition, there are other automatic picking algorithms which are based on energy analysis (Earle & Shearer, 1994), polarization analysis (Vidale, 1986), statistical analysis, wavelet analysis, frequency analysis, and autoregressive techniques (Leonard & Kennett, 1999; Leonard, 2000).

In contrast to the methods based on picked arrival times, migration-based methods use the full seismic wavefield to locate micro-seismic events. The main advantage provided by the methods in this category is that they are independent of the picking of specific arrivals and can locate weak events by focusing energy at the source using time-reversal (Baysal, 1983; Levin, 1984; Gajewski,

2005). The drawback of methods in this category is high computational cost which limits their practical use. However, this drawback becomes less of a problem due to the significant advances in computer hardware, which allows for fast processing of large volumes of data.

Within the context of migration-based methods, I advocate in this thesis a new methodology which can be used to automatically identify and locate micro-seismic events. I use time-reversal based on numerical solutions to the acoustic wave-equation to reconstruct seismic wavefields in the space around the possible source position and at times around the possible onset time of seismic events. Ideally, the energy corresponding to a micro-seismic event focuses at a time and a position in space. However, due to the incomplete acquisition, inaccuracies of the velocity model, and noise present in data, this focus is hard to identify or may not even form at all.

I address this problem by constructing, at various locations in the model and at various times, "ideal" wavefields which we would record given the field acquisition geometry and the velocity model in the region under investigation if a micro-earthquake source would be located at that position and time. Then, I employ a pattern matching procedure to compare the reconstructed wavefields based on field data with the simulated wavefields and identify the most probable location and onset time of the micro-seismic event. This process exploits the kinematic properties of the focused wavefields and takes advantage of the distortions caused by the incomplete acquisition in the area under investigation.

Given the uncertainty characterizing the process of micro-earthquake identification, I do not construct our results as precise positions and onset times. Instead, I construct maps of probability indicating the likelihood that a micro-seismic source exists at a given position in space and a given time. I construct the solutions using Bayesian inversion theory which states that inversion is a process which refines our state of knowledge. Bayesian theory discusses a priori state of knowledge, which incorporates our knowledge about our system in absence of any measurements, and a posterior state of knowledge, which combines the prior with additional knowledge acquired from new measurements.

In Chapter 2, I first provide a method for automatic micro-seismic location in the migration-based category. The method discussed in this chapter is illustrated with a 2D velocity model with only one micro-earthquake. Although this method automatically provides a source for a synthetic seismogram, it does not take the uncertainties of measurement, theory, and S/N ratio into consideration. Thus, in Chapter 3, I incorporate Bayesian theory into the method and calculate maps which illustrate the probability of a micro-earthquake triggered at a certain location and time. In this

chapter, I also apply the method to a situation with many micro-earthquakes and calculate their probability maps. Finally, in Chapter 4, I apply the method to field data with controlled explosive sources provided by ExxonMobil. In this chapter, I illustrate the method with a 3D velocity model and prove the robustness of the method for field data.

Chapter 2

A methodology of automatically micro-seismic location

In this chapter, I introduce a new methodology which automatically provides the location and onset time of a micro-earthquake in a 2D velocity model. A similar methodology is applied to 3D as discussed in chapter 4. By solving the acoustic two-way wave-equation using reverse time migration and using a finite difference algorithm, I reconstruct the wavefield in the time domain from a given seismogram. The hypocentre and excitation time of a micro-seismic event is detected from the reconstructed wavefield by comparing the wavefield with the simulated wavefields at possible source locations and origin times.

I exploit four methods, which are normalized cross-correlation, sum of squared difference, ratio image uniformity, and mutual information, to compare the similarity between the reconstructed wavefield and simulated wavefields. After discussing the advantages and disadvantages of the four methods, I select normalized cross-correlation as the method used in the following thesis. At the end of this chapter, I discuss the robustness of the selected method in the noisy environments. The advantages of using migration-based methods are that they are independent of the picking of arrival times and that they can enhance the S/N ratio of the reconstructed wavefield.

2.1 Introduction

Due to the low magnitude of a reservoir micro-seismic event, it is frequently difficult to pick any event from the seismogram. However, by applying a two-way acoustic wave equation to the recorded seismograms, we avoid the picking of arrival times. Furthermore, using reverse time migration (RTM) to solve the wave-equation exploits the redundant and noisy information in seismograms as

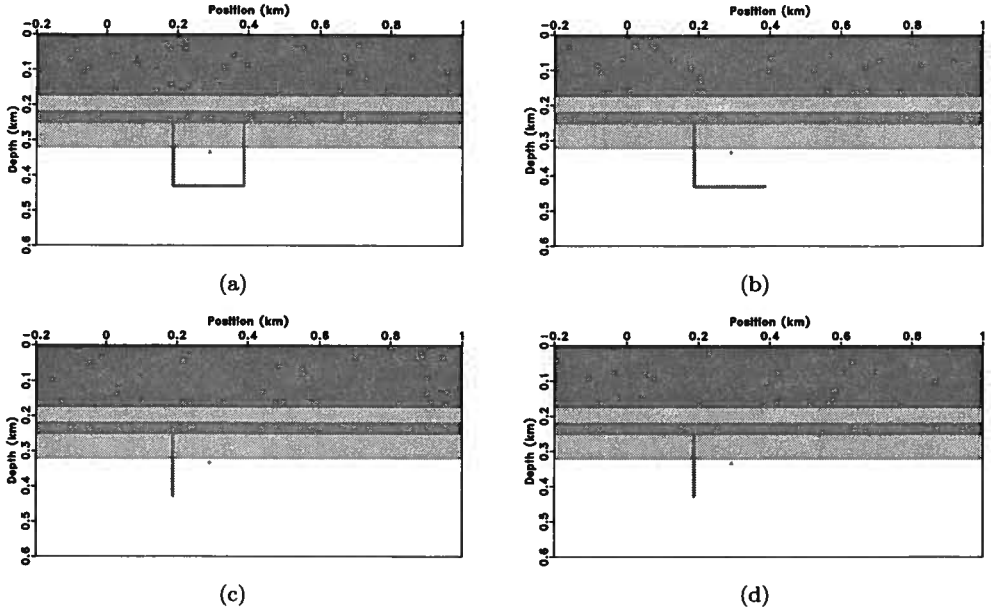


Figure 2.1. Four different layouts of receivers.

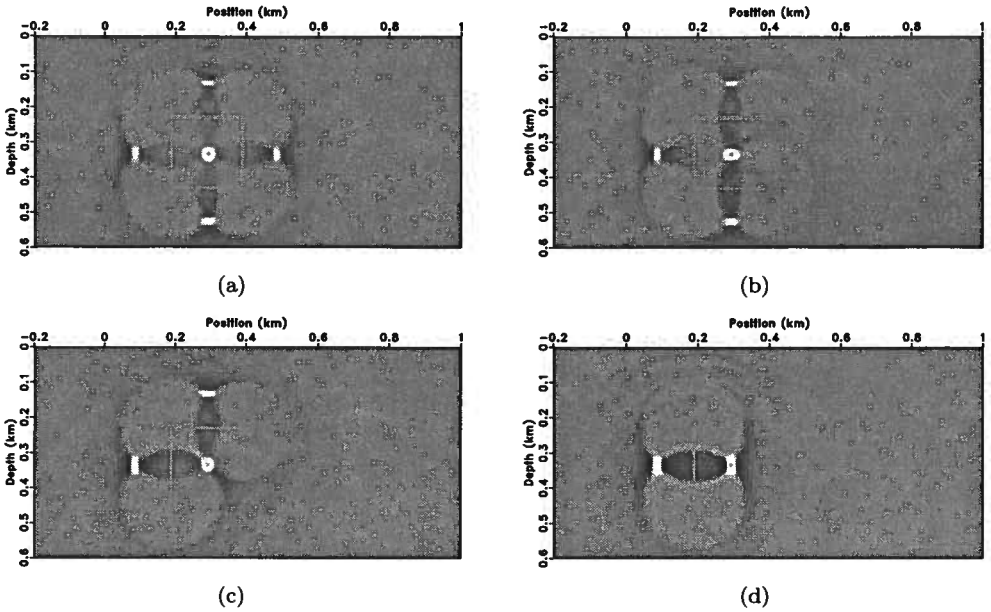
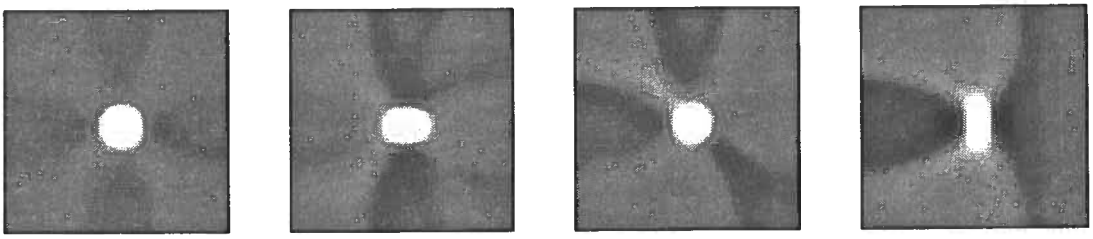


Figure 2.2. Snapshots of back propagated wavefields at the source onset time.



(a) (b) (c) (d)

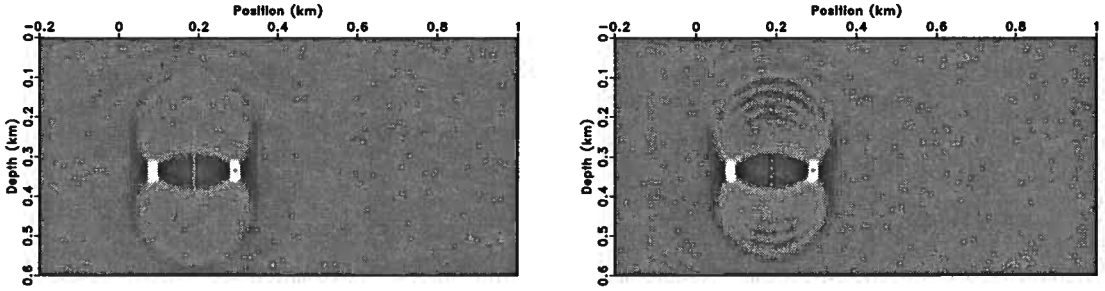
Figure 2.3. Zoomed-in snapshots of reconstructed wavefields at the source onset time. The panels correspond to the box shown in Figure 2.2(a).

an input to calculate the wavefield (Gajewski, 2005). Similar to stacking in seismic exploration, RTM sums all the recorded information, thereby enhancing the correlated micro-seismic signals and canceling the uncorrelated signal of noise. Therefore, even if we cannot identify any event in individual seismograms, RTM has the potential to focus all the energy at the correct location of the source.

For a wavefield reconstructed from a seismogram by using RTM, we commonly expect that the energy of the wavefield is focused on the hypocentre of the seismic event. However, this focus strongly depends on the density, coverage, and number of receivers. The larger the coverage of receivers, the better the focus of energy. Figures 2.1-2.3 show the relationship between the coverage of receivers and the focus of the wavefield.

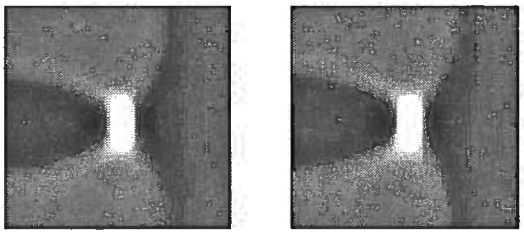
Figure 2.1 describes four different layouts of receivers. The gray background is the velocity model. There are six homogeneous, isotropic, and horizontal layers. Lines represent receivers. The receiver sampling distance is 2 meters, which is very dense. There are 100 receivers on each side. In Figures 2.1(a) - 2.1(d), there are 400, 300, 200, and 100 receivers, respectively. The dot shows the position of the seismic event.

Figure 2.2 shows snapshot of the reconstructed wavefields at the origin time of the event for the four different receiver layouts. Because the wave-equation does not have a direction, it sends signals symmetric to its receivers. Thus, there is a symmetric wavefield along the receiver line. In this research, I only take into consideration the right hand side of the receiver line. In Figure 2.2(a), we see a perfect circle at the source position which is represented by the dot. Figure 2.3(a) shows a zoomed-in plot of Figure 2.2(a). This perfect point is distorted when the receiver coverage is decreased. When there is a line of receivers, the shape of focus (shown in white) at the source point



(a) (b)

Figure 2.4. Snapshot of reconstructed wavefields at the source onset time for dense and sparse layouts of receivers, respectively.

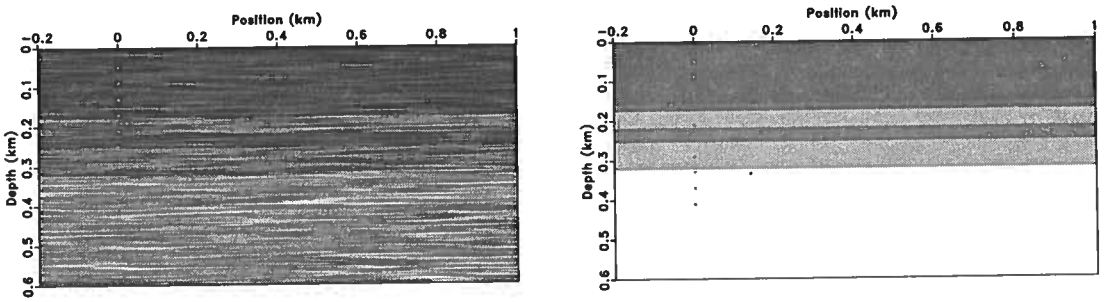


(a) (b)

Figure 2.5. Zoomed-in snapshot of reconstructed wavefields at the source onset time for dense and sparse layouts of receivers, respectively.

is similar to a rectangle rather than a circle. Therefore, if we assume that the shape of the focus is a circle when there is limited receiver coverage and use this assumption to determine the hypocentre of the event, we may not find the source position of the event.

In practice, the density of receivers is much lower than that shown in Figure 2.1. In reality, several receivers would be sparsely placed along a certain section of the borehole. In Figure 2.4(b), the density of receivers is one tenth of that shown in Figure 2.4(a). There are only 10 receivers in Figure 2.4(b), and the distance between two nearby receivers is 20 meters. Figure 2.5 shows the shapes of the wavefield focus for the two different layouts of receivers at the source excitation time and indicates that these two shapes are similar. However, due to the sparsely placed receivers, artifacts are introduced into the reconstructed wavefield as shown in Figure 2.4(b). Therefore, in practice, the focus of the wavefield cannot be applied to determine the hypocentre and the excitation time of a seismic event. In this research, I introduce a new method which uses the shape of the



(a) (b)

Figure 2.6. (a) The velocity model with random variability used to simulate data and (b) the velocity model used for the wavefield reconstruction. The random velocity in panel (a) is the sum of the velocity in model (b) with correlated Gaussian noise with a magnitude of 40% of the background.

reconstructed wavefield as a criterion to identify the source of a micro-earthquake.

2.2 The reconstructed wavefield database

In this research, I fix the velocity models and receiver geometry shown in Figure 2.6. The velocity shown in Figure 2.6(a) is used to simulate the seismogram recorded in reality. I assume that the surface is absorbing because normally micro-seismic signals decay before arriving at the surface. The velocity (shown in Figure 2.6(b)), from the top layer to the bottom is: 1900 m/s, 1950 m/s, 2150 m/s, 2000 m/s, 2150 m/s, and 2350 m/s, respectively. The medium is isotropic and heterogeneous. This velocity model is sampled every 2 m and the calculation is done on a numerical grid of 600 $\times$ 300 nodes. The acquisition array consists of 10 receivers placed at $x = 0$ km with a spacing of 40 m, which are represented by dots. I simulate an explosive seismic source (represented by a dot) at $x = 0.140$ km, $z = 0.334$ km, and $t = 0.04$ s, using a Ricker wavelet with a dominant frequency of 50 Hz. A total of 2000 time steps with a time increment of 0.2 ms are calculated.

The wavefield simulated for this event is calculated using,

$$\frac{1}{v^2}\ddot{U} = \nabla^2 U + f(\mathbf{m}_t, t)\ , \tag{2.1}$$

where $f(\mathbf{m}_t, t)$ is the designed source wavelet at a particular micro-seismic source location at $\mathbf{m}_t = \{x_t, y_t, z_t, t_t\}$, where the subscript t indicates the "true" position of the source, $U(\mathbf{x}, t)$ represents the seismic wavefield simulated forward in time where $\mathbf{x} = \{x, y, z\}$, and $v(\mathbf{x})$ is the velocity shown

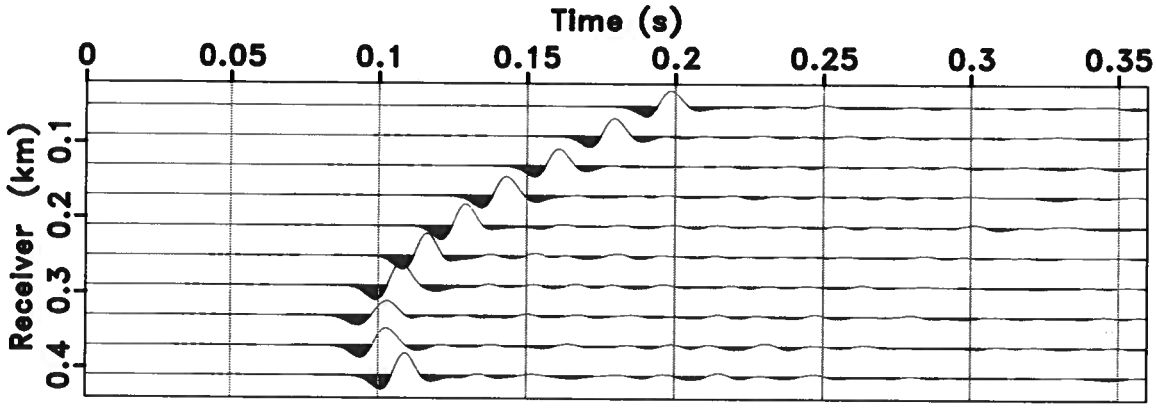


Figure 2.7. The synthetic seismogram simulated using the velocity shown in Figure 2.6(a).

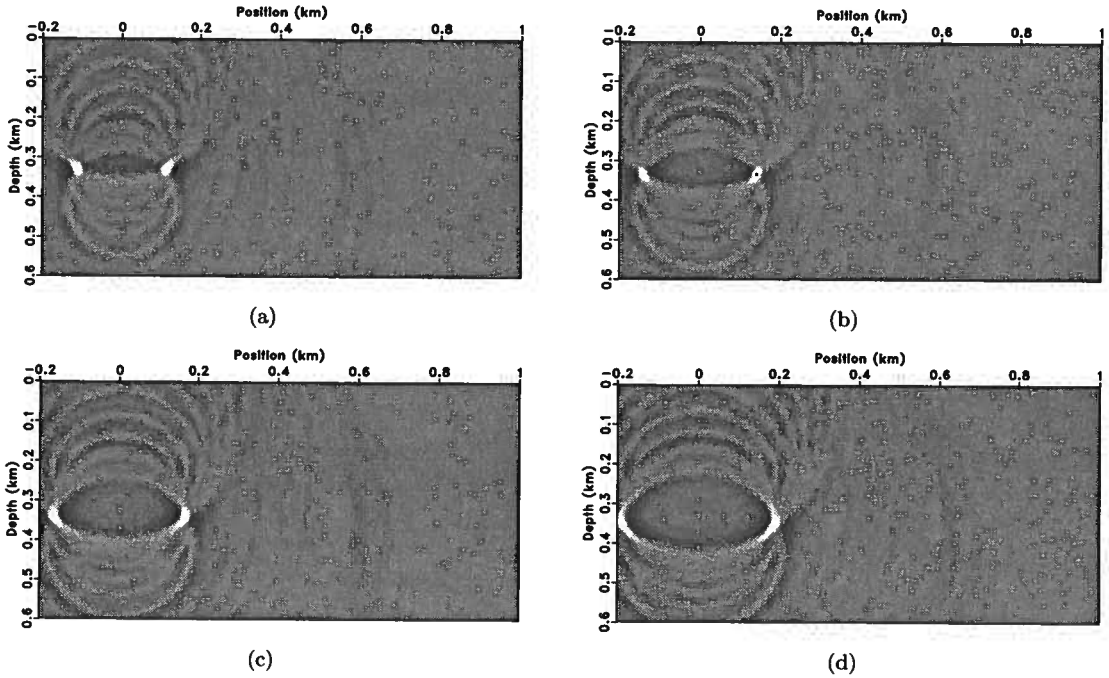


Figure 2.8. Four snapshots of the reconstructed wavefield. (b) shows the focus of reconstructed wavefield.

in Figure 2.6(a). I use this velocity to simulate the seismogram recorded in reality but I assume that this velocity is unknown during the reconstruction process. The simulated synthetic seismogram is shown in Figure 2.7.

This simulated seismogram is used to calculate the reconstructed wavefield using the velocity shown in Figure 2.6(b). Four snapshots of the reconstructed wavefield are shown in Figure 2.8. I only take the right hand side of the reconstructed wavefield into consideration. The wavefield defocuses when it moves away from the source location. In Figure 2.8(b), the hypocentre of the micro-seismic event is located at the focus point. If we use the best focus of the reconstructed wavefield as a criterion to locate the source of a micro-seismic event, it is difficult to choose between Figures 2.8(a) and 2.8(b). Thus, it is necessary to define an efficient method to locate the hypocentre and origin time of a micro-seismic event from the reconstructed wavefield.

There are approximately three categories of methods that can be used to identify the source of a micro-earthquake from a reconstructed wavefield. The first category involves correctly removing the geometric spreading during the process of calculating back propagation; the second involves focusing of wavefield energy during back propagation as a rule to find the source information; the third involves mathematically calculating the properties of the reconstructed wavefields at the hypocentre of earthquakes.

For the first category, there are mainly two strategies for addressing the amplitude preservation problem. One strategy which is commonly used is crudely implemented by a simple time scaling of the images at every time step (Tygel *et al.*, 1992). Another strategy is the removal of geometrical spreading using a wavefield continuation method for back propagation, which is based on the high frequency asymptotic assumption (Gajewski, 2005). One disadvantage of the second strategy is that this method straightforwardly uses parameters derived from the ray-based true amplitude even though most of the parameters involved in this migration cannot be derived from a full wavefield (Plessix & Mulder, 2002; Joncour *et al.*, 2005b,a).

For the second category, as discussed previously, the focus of a back propagated wavefield depends on the angle coverage of receivers; the higher the aperture, the better the wavefield focuses. Due to limited the coverage of receivers (for example as shown in Figure 2.6), the shape of the focus varies with source location. For example, in Figure 2.9, one source is placed close to the receivers and another source is placed far from the receivers. These two sources are indicated by two dots. Parts of the snapshots of the two reconstructed wavefields centered at the source positions are shown

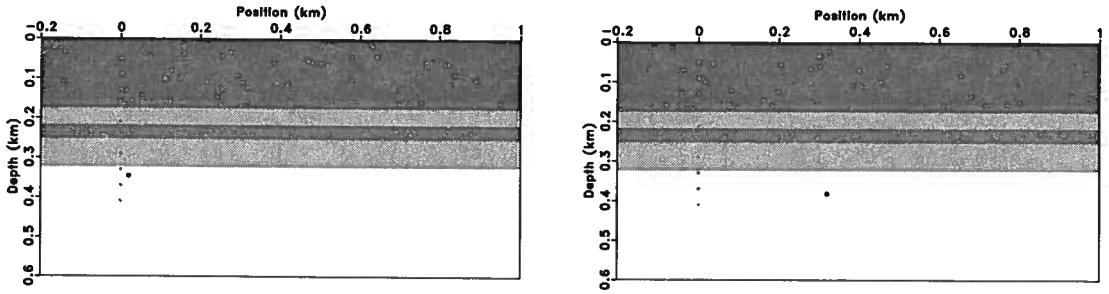


Figure 2.9. Sources at different positions. In Figure (a), source is close to the receivers and in Figure (b), source is far from the receivers.

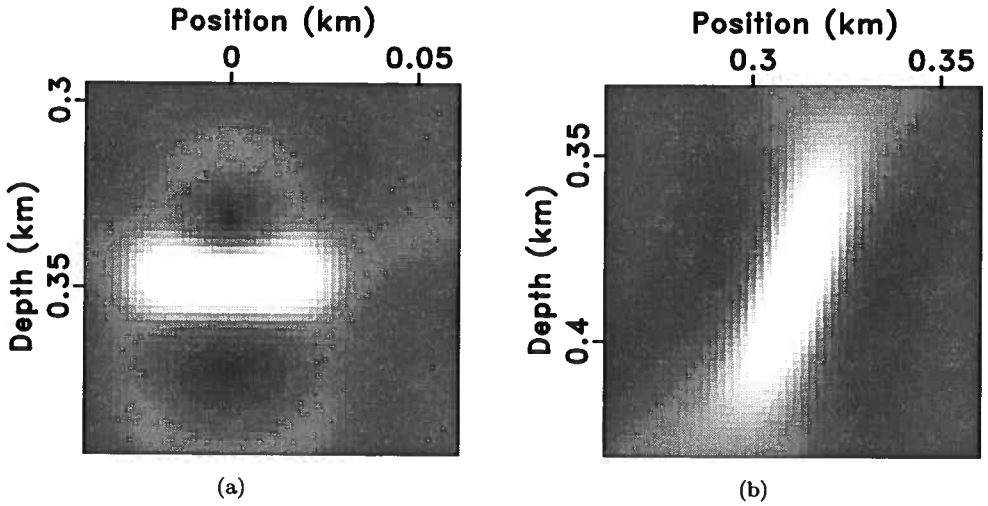


Figure 2.10. Snapshots of the reconstructed wavefields at the source onset time. The source location (a) is close to the receivers and (b) is far from the receivers.

in Figures 2.10(a) and 2.10(b). A comparison of these two plots shows that the shape of the focus varies with location and that neither of these two foci is a circular.

The third category is based on the mathematical analysis of the two-way acoustic wave equation. By solving the wave equation in the frequency domain based on the high frequency asymptotic, we detect some properties at the focused source point. The properties are used to detect the source location. One approach for using these properties is to calculate a predicted focal spot size based on the effective aperture which is associated with the statistical properties of the random medium. This spot size is then compared to the refocused spot size calculated from the field seismogram (Papanicolaou *et al.*, 2004; Borcea *et al.*, 2003). Another approach is to analyze the pulse refocused on a ring at the depth of the source and at a critical time. Thereby, we identify a relationship between this ring and the source location (Fouque *et al.*, 2004). Since these methods assume that the media is randomly layered, these methods can be used in ultrasound and underwater acoustic environments where the velocity structure is much simpler than that in the Earth.

In this research, I expect that the method I provide is able to handle the narrow aperture of receivers and be easy to implement. I also expect that I can use this method to determine the source locations of multiple micro-seismic events in 3D velocity models. However, none of the above three methods satisfied our expectations. The method developed in this chapter exploits the uniqueness of the shape of the refocused wavefield to identify the hypocentre and the critical time of a micro-seismic event. This method is composed of two steps: 1) constructing a reconstructed wavefield database and 2) similarity comparison. Each of these steps is discussed separately in the following section. The first step can be further divided into two parts: 1) selecting possible micro-seismic event locations and excitation times and 2) reconstructing wavefields for all selected micro-seismic event locations for a known origin time.

2.2.1 Selecting possible micro-seismic event locations

We can select a probable area for micro-earthquakes by collecting different information. Geologic information about a given reservoir or preliminary results of micro-earthquakes provided by other simple methods are two possible sources of information.

According to geologic information, there are three geologic layers in the reservoir under investigation: Colorado shales from 0 to 250 m; Grand Rapid from 250 to 320 m; and the Clearwater reservoir which is located at the bottom (Miyazawa, 2007). As I consider this geologic information

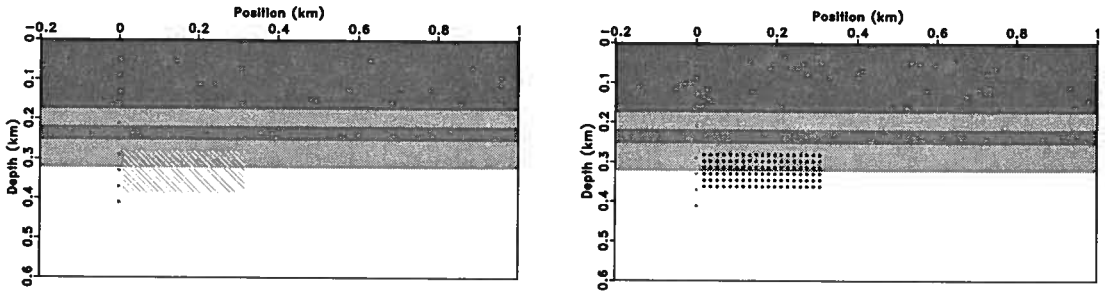


Figure 2.11. (a) A schematic representation of possible source area; (b) a schematic representation of possible source locations. I assume that the micro-earthquakes occur in the region indicated by the dense dots between $x = 0.012 - 0.312$ km and $z = 0.282 - 0.382$ km. A similar discrete set of possible onset times are defined along the time axis.

along with the fact that a micro-seismic event can be triggered along a pre-existing fault (Rothert & Shapiro, 2003), I assume that micro-earthquakes can be triggered along the interfaces of any two geologic layers since the resistance to stress and strain around the interfaces of two geologic layers is much weaker than that in other areas.

Using information provided by other methods (discussed below), I identify an area which is shown by the yellow box in Figure 2.11(a). These methods include grid-searching of travel-times which can generate a general idea of the distribution of sources although the uncertainty of this method is great; Gaussian beam migration which can provide information from a 3D velocity model (Rentsch, 2004, 2006, 2007); and the double-difference (hypoDD) method (Waldhauser, 2000) that has been applied to the same reservoir (Miyazawa, 2007). I assume that all micro-seismic events analyzed in this research are triggered within the identified area in Figure 2.11(a). The a priori probability of a micro-seismic event being triggered within the yellow box is 1; outside the box the probability is 0.

After selecting an area and a range of time within which a micro-earthquake can be triggered, I design a walk within this area. For simplicity, I uniformly sample the possible source locations and onset times within this region. The number of sources located in this space is $n_x \times n_z$, where n_x and n_z represent the numbers of sources along the x and z axes, respectively. In these examples, the possible source sampling space is 4 m, there are 76×26 sources, and a selection of possible sources in space are shown in Figure 2.11(b). By selecting a finite number of possible source positions, I not

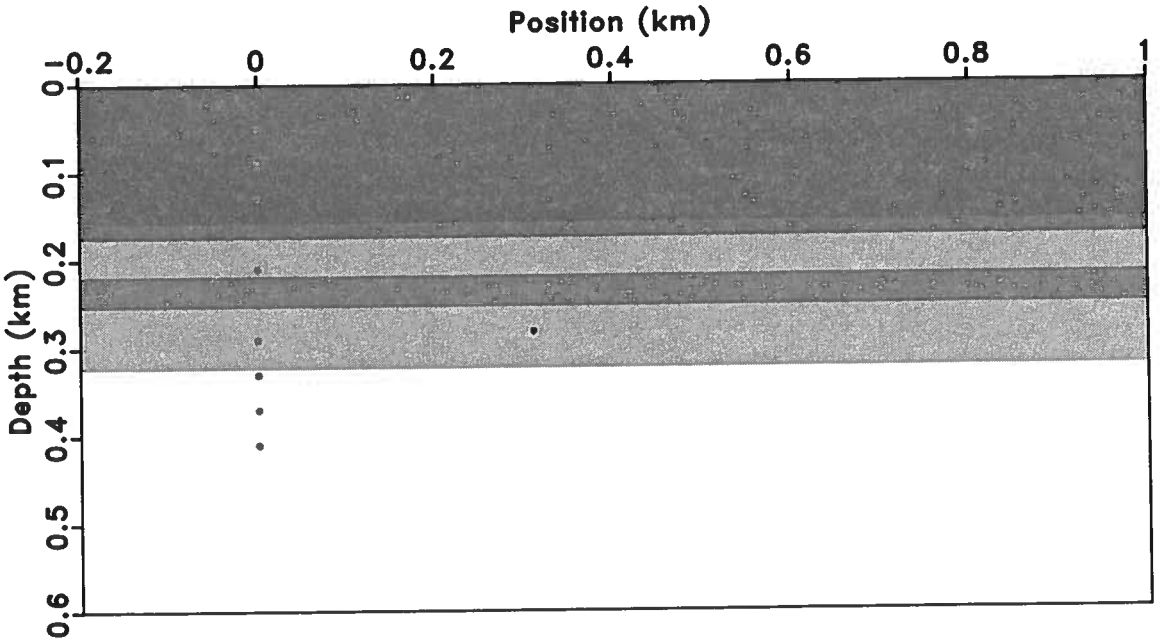


Figure 2.12. One selected possible source location which is represented by the dot located at $\{x, z\}$ = {308 m, 303 m}.

only use the knowledge about the true source location but also reduce the computational time. The assumption I make is that the reconstructed wavefield corresponding to the recorded seismograms is characterized by equal probability for all sources within the region where micro-seismic events are assumed to be located.

2.2.2 Calculating a reconstructed wavefield database

For all sources within the selected region in space, I compute the corresponding wavefield using a forward/backward procedure. In this case, this procedure is based on the two-way acoustic wave-equation outlined in the following steps:

- First, for a given source of a micro-earthquake ($\mathbf{m}_d$) characterized by the onset time t_d and location $\mathbf{x}_d$, I propagate waves forward from the source using the velocity $c(\mathbf{x})$ shown in Figure 2.6(b) according to the two-way acoustic wave-equation

$$\frac{1}{c^2}\ddot{U} = \nabla^2 U + f(\mathbf{m}_d, t)\,, \tag{2.2}$$

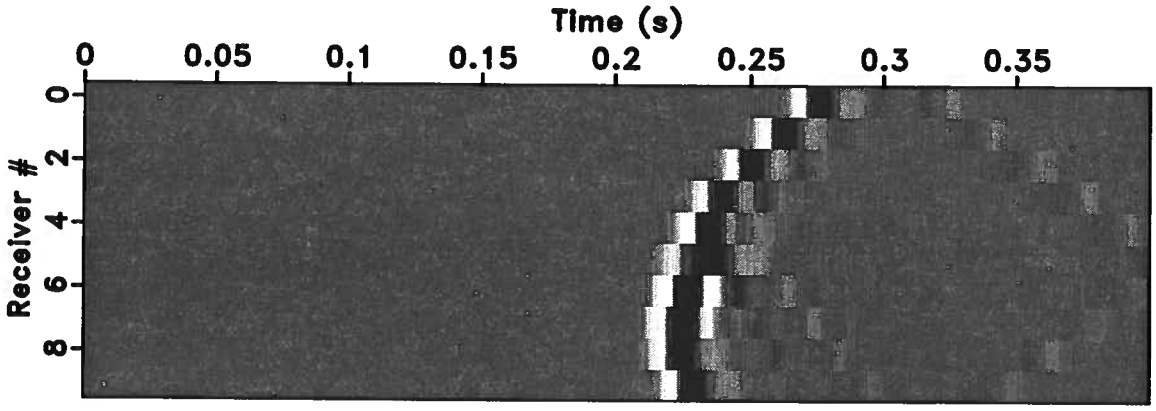


Figure 2.13. Seismogram calculated by using equation 3.7 and equation 3.8

where $f(\mathbf{m}_d, t)$ is the considered source wavelet and $U(\mathbf{x}, t)$ represents the seismic wavefield simulated forward in time. For example, one source within the selected possible source locations is $\mathbf{m}_d = \{\mathbf{x}_d, t_d\} = \{308 \text{ m}, 303 \text{ m}, 0 \text{ s}\}$ represented by a dot in Figure 2.12.

- Second, the forward wavefield is sampled at receiver coordinates $\mathbf{x}_0$ arbitrarily located in space:

$$R(\mathbf{x}_0, t) = U(\mathbf{x} = \mathbf{x}_0, t) \,, \tag{2.3}$$

where $R(\mathbf{x}_0, t)$ represents the recorded seismogram. In this example, the seismogram (shown in Figure 2.13) is recorded by receivers located at $x = 0$ m represented by dots in Figure 2.12. In this seismogram, we can identify the direct P wave. There are some reflected waves from the boundary of layers but no reflection from the top of the surface.

- Third, the recorded seismogram is time reversed due to the space-time reciprocity and propagated backward in time from the acquisition array to reconstruct the wavefield at all locations in space and all times. This step is also based on the two-way acoustic wave-equation

$$\frac{1}{c^2}\ddot{D} = \nabla^2 D + R(\mathbf{x}_0, -t) \,, \tag{2.4}$$

where $D(\mathbf{x}, t)$ are the reconstructed wavefield. Because I am only concerned with the focus part of the reconstructed wavefield around the source location and onset time, I store a cube from the reconstructed wavefield (shown in Figure 2.14). The size of this cube is two wavelengths in

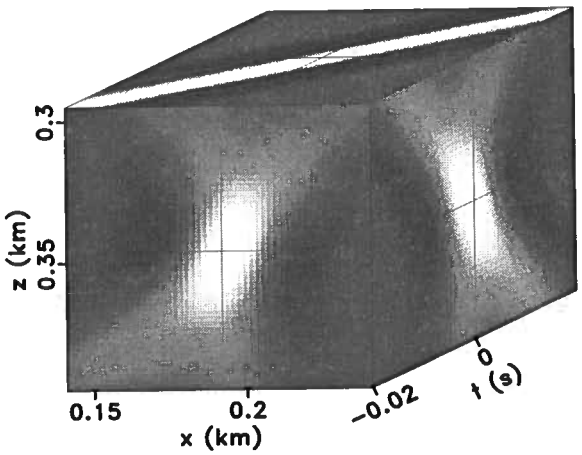


Figure 2.14. Part of the wavefield reconstructed from the possible source location. The intersections between lines indicate the source location.

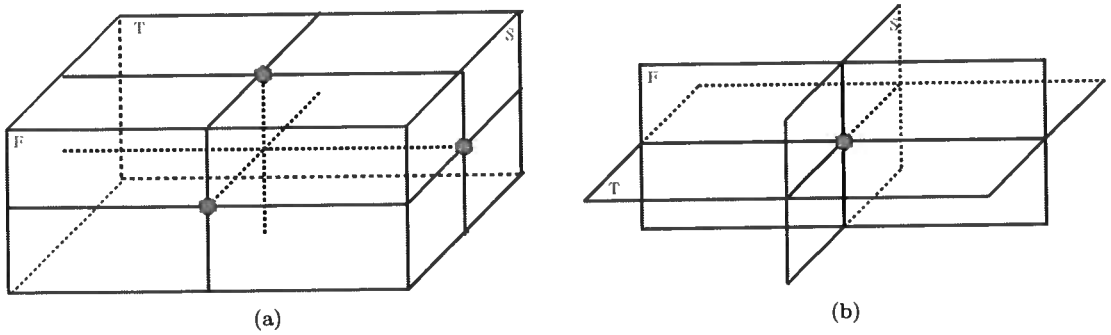


Figure 2.15. Illustration of coordinates of Figure 2.14. (a) shows a schematic plot of Figure 2.14 and the true coordinates of the three faces of (a) are shown in (b).

space and two periods in time. The coordinates of Figure 2.14 are illustrated by Figure 2.15. I use "F", "T", and "S" to mark the front, top, and side plots of Figure 2.15(a), respectively. The true locations of these three faces are indicated by "F", "T", and "S" in Figure 2.15(b). The polygons represent where these three faces intersect. The center of Figure 2.14 is the source location represented by the intersections of lines.

For a given source location, I refer to the wavefield reconstructed from a field seismogram or a simulated synthetic seismogram as the observed reconstructed wavefield. In contrast, I refer to the wavefield calculated using the three previous described steps as the predicted reconstructed wavefield.

The above three steps are applied to all the selected sources. Then I assemble resulting calculations in a database of wavefields which is used to compare with the wavefield reconstructed from the "field" seismograms. Figure 2.16 shows part of the reconstructed wavefield database. This database only needs to be calculated once as long as we keep the subsurface velocity and acquisition geometry fixed.

If two sources are located at the same position but have different onset times, the refocused fields of these two events are the same. It is sufficient to compute reconstructed wavefields for one onset time and then compare these simulated wavefields with the wavefields reconstructed from field seismographs at various times. This procedure can be conceptually illustrated by Figure 2.17, where the big box represents the wavefields reconstructed from the "field" seismograms, the small cubes represent predicted reconstructed wavefields, and the dot within each small cube represents the source location of the predicted reconstructed wavefield. By design, the source onset times of the predicted reconstructed wavefields are all zero. By comparing the wavefield reconstructed from the "field" seismogram with predicted reconstructed wavefields at the corresponding locations along the time axis, we can not only identify the location of the source, but we can also identify the onset time.

In an ideal case, if the sources are completely surrounded by receivers, the wavefronts reconstructed by time reversal focus into a point at the source location and at the onset time. However, typical acquisition arrays are limited to a relatively small region of space. In the example shown in Figure 2.11(b), 10 receivers are placed along a borehole located at $x = 0$ km. For limited aperture of receivers, the wavefronts do not focus at the source location.

As illustrated in Figure 2.18, I define the range L as the shortest distance between the source and the receiver line and the cross-range C as the shortest distance between the source and the line which is perpendicular to the line of receivers in the middle. According to Figure 2.16, the cross-range resolution C_r of the refocused wavefield at $t = 0$ s increases when the range decreases, but the range resolution does not change much. The relationship between the cross-range resolution C_r and the range L in homogeneous media can be approximated by $C_r = \lambda_0 L/a$, where λ_0 is the wavelength and a is the array aperture (Born & Wolf, 2003). The wavefront shape at $t = 0$ s is controlled by the range and the cross-range for a fixed array aperture and wavelength. It is apparent that not only the focus is incomplete but also that the shape of the reconstructed wavefields changes in space corresponding to the relative positions to the acquisition array as well as the velocity distribution in space. This proves that the shape of focus varies with the location of source and that the shape is unique.

2.3 Methods of Similarity Measurement

I have shown in the previous section that the shape of the refocused wavefield varies with the source location. Thus, I expect that the predicted reconstructed wavefield with the center at the true source location and onset time should be the same as the observed reconstructed wavefield at the common areas in space and time.

Figure 2.19 shows three pairs of predicted reconstructed wavefields and parts of the observed reconstructed wavefield. In this figure, the size of the predicted reconstructed wavefields is two wavelengths in space and two periods in time; the size of the parts of the observed reconstructed wavefield is the same. For simplicity, I refer to any part of the observed wavefield which has the same size as the predicted wavefield as the windowed observed wavefield.

In Figure 2.19, the left column shows three predicted reconstructed wavefields with the source located at $\{x,\ z\} = \{132$ m, 334 m$\}$, $\{108$ m, 334 m$\}$, and $\{132$ m, 334 m$\}$; the right column shows three windowed observed wavefields with the center at $\{x,\ z,\ t\} = \{132$ m, 334 m, 0.04 s$\}$, $\{108$ m, 334 m, 0.04 s$\}$, and $\{132$ m, 334 m, 0.045 s$\}$; the explosive source location and excitation time is $\{140$ m, 334 m, 0.04 s$\}$. From Figure 2.19, we can conclude that the pair of the predicted reconstructed wavefield and windowed observed wavefield are most similar around the true source location and onset time.

As discussed earlier, I refer to the predicted reconstructed wavefields as the wavefields reconstructed at all possible source locations and within a given time window. These wavefields are calculated using equations 3.7 - 3.9 and they are stored around their hypocentres and onset times. I loop over the possible source locations and excitation times and search for the times and locations in space where the wavefield reconstructed from field seismograms or from synthetic field seismograms match the predicted reconstructed wavefield at assumed locations for micro-seismic events. I conceptually describe the process as measuring the similarity between multi-dimensional images.

Automatic similarity comparison methods are divided into two main categories: area-based methods and feature-based methods. Feature-based methods require salient structures which can be extracted from images. This salient structure should be distinct, spread across the image, and efficiently detectable in both predicted and observed images. However, in this experiment, there is no such feature that is distinct and spread across the reconstructed wavefield database. On the other hand, area-based methods, which put emphasis on feature matching rather than on feature detection, are suitable for my research since I expect to use all predicted reconstructed wavefields as criteria to compare with the observed reconstructed wavefield.

Area-based methods are widely used in medical image registration. Normalized cross-correlation, sum of squared difference, and ratio image uniformity are the three most common feature-based methods used in comparing images of the same modality. Mutual information and its variant are the most popular image similarity measures for registration of multi-modality images (Zitova & Flusser, 2003). In this work, I apply four methods (normalized cross-correlation, sum of squared difference, ratio image uniformity, and mutual information) to calculate the similarity between the predicted wavefields and the observed reconstructed wavefield.

2.3.1 Normalized cross-correlation

The normalized cross-correlation (N_{cc}) is defined as

$$N_{cc}(\mathbf{x},t) = \frac{\sum_{\mathbf{x}-M_{\mathbf{x}}}^{\mathbf{x}+M_{\mathbf{x}}}\sum_{t-M_t}^{t+M_t}(\langle D^{obs}\rangle - \overline{\langle D^{obs}\rangle})(D^{pre} - \overline{D^{pre}})}{\sqrt{\sum_{\mathbf{x}-M_{\mathbf{x}}}^{\mathbf{x}+M_{\mathbf{x}}}\sum_{t-M_t}^{t+M_t}(\langle D^{obs}\rangle - \overline{\langle D^{obs}\rangle})^2 \sum_{\mathbf{x}-M_{\mathbf{x}}}^{\mathbf{x}+M_{\mathbf{x}}}\sum_{k=t-M_t}^{t+M_t}(D^{pre} - \overline{D^{pre}})^2}}, \tag{2.5}$$

where the overline denotes averaging values below, $\langle D^{obs} \rangle$ is the windowed observed wavefield, and D^{pre} is the predicted wavefield. The size of the averaging window is $(2M_{\mathbf{x}} + 1) \times (2M_t + 1)$, where $\mathbf{x} = \{x, y, z\}$. The normalized cross-correlation is evaluated at the locations of every possible source location and onset time. In this example, the size of $M_{\mathbf{x}}$ is one wavelength and the length of M_t is one period. The values of N_{cc} are between -1 and $+1$. When $N_{cc} = 1$, the shapes of the predicted wavefield and the windowed predicted wavefield are identical. Thus, using this equation, I pinpoint the location and time which generate the highest N_{cc}. However, the size of the predicted reconstructed wavefield affects the value of $N_{cc}(\mathbf{x}, t)$.

After calculating the similarity between the observed wavefield (four snapshots are shown in Figure 2.8) and the reconstructed wavefield database (part of this is shown in Figure 2.16) by using equation 2.5, I get the normalized cross-correlation values shown in Figure 2.20. Each cube in Figure 2.20 represents one possible source location and onset time which I previously selected. The source location pinpointed from N_{cc} and the true source location are shown in Figure 2.21.

In this example, the detected and true source locations are close to each other; the distance between them is 8 m which is 2 sampling spaces of possible sources. This error is much less than the wavelength of the wavelet which is about 40 m. Thus, the accuracy is less than 1/5 of a wavelength by using normalized cross-correlation.

In this section, normalized cross-correlation rather than cross-correlation is used to calculate the similarity between two images because cross-correlation (CC) fails in feature recognition. Figure 2.22 shows a Ricker wavelet called the sensed wavelet with a dominant frequency of 40 Hz and a magnitude of 1. I expect to find a Ricker wavelet also with a dominant frequency of 40 Hz from the two reference wavelets A and B shown in Figure 2.22. For these two wavelets A and B, the dominant frequencies are 40 Hz and 20 Hz and the magnitudes are 0.6 and 0.8, respectively. I apply normalized cross-correlation and cross-correlation to calculate the similarities between the sensed wavelet and the two reference wavelets. The N_{cc} value of the sensed wavelet and reference wavelet A is 1; and the value of the sensed wavelet and the reference wavelet B is 0.5724. The CC value of the sensed wavelet and reference wavelet A is 11.2203; and the value of the sensed wavelet and the reference wavelet B is 12.111. I detect the reference wavelet A with a dominant frequency of 40 Hz based on the information provided by N_{cc}. However, according to the information provided by CC, I detect that the reference wavelet B is more similar to the sensed wavelet than the reference wavelet A. The reason for this failure is that CC takes the amplitude of wavelets into consideration. However, in this

research, I expect to find the pair of the predicted wavefield and the windowed observed wavefield that produces the highest shape similarity without considering the amplitude of the wavefields. Therefore, normalized cross-correlation is a more appropriate way to calculate the shape similarity.

2.3.2 Minimizing intensity difference

The minimizing intensity difference (MID) method is one of the simplest methods that calculates the voxel similarity. It sums the squared intensity difference between the observed and predicted images, and we can apply a threshold criterion to the MID value (Viola, 1995). The modified minimizing intensity difference equation is expressed as

$$MID(\mathbf{x},t) = \frac{1}{N}\sum_{\mathbf{x}-M_{\mathbf{x}}}^{\mathbf{x}+M_{\mathbf{x}}}\sum_{t-M_t}^{t+M_t}(D_n^{pre} - \langle D^{obs}\rangle_n)^2\,, \tag{2.6}$$

where N is the number of voxels of one predicted reconstructed wavefield and the subscript n of D^{pre} and $\langle D^{obs}\rangle$ denotes their normalized values. The value of MID is greater than 0, where 0 means that the normalized predicted wavefield is the same as the windowed observed wavefield after normalizing.

Because MID is sensitive to a small number of voxels that have large intensity differences between two images (Hill *et al.*, 2001), the plot of MID without a threshold (shown in Figure 2.23(a)) does not provide us with visual information about the source location. However, after applying a threshold criterion with a value of 0.01 (this value will vary with noise level), we can visually detect the source location. The source location I picked and the true source location are shown in Figure 2.24. The source location provided by MID is the same as that of N_{cc}. Thus, MID generates an accurate result.

2.3.3 Ratio image uniformity

The ratio image uniformity (RIU) algorithm was originally introduced by Woods *et al.* (1992) for the registration of serial positron emission tomography (PET) studies, but it has more recently been widely used for serial magnetic resonance (MR) registration (Woods *et al.*, 1998a,b) and is available in an automated image registration (AIR) package (Hajnal *et al.*, 2001). Ratio image uniformity can be interpreted as the normalized standard derivation (Beutel *et al.*, 2000). In this algorithm, the higher the similarity between two images, the smaller the normalized standard derivation. RIU

is defined as

$$RIU(\mathbf{x},t) = \frac{\sqrt{\frac{1}{N}\sum_{\mathbf{x}-M_{\mathbf{x}}}^{\mathbf{x}+M_{\mathbf{x}}}\sum_{t-M_t}^{t+M_t}(R-\overline{R})^2}}{\overline{R}}, \tag{2.7}$$

where

$$R = \frac{\langle D^{obs}\rangle}{D^{pre}} \tag{2.8}$$

is the ratio of the windowed observed wavefield and the predicted wavefield at every voxel of the wavefields and

$$\overline{R} = \frac{1}{N}\sum_{\mathbf{x}-M_{\mathbf{x}}}^{\mathbf{x}+M_{\mathbf{x}}}\sum_{t-M_t}^{t+M_t} R \tag{2.9}$$

is the average value of R, where N is the number of voxels of one predicted reconstructed wavefield. When $\langle D^{obs}\rangle$ and D^{pre} are identical, the standard derivation equals zero, which is the minimum value of RIU.

The result of RIU shown in Figure 2.25 provides us with no visual information. By searching within this result, I find that the source detected from RIU shown in Figure 2.26 is several wavelengths away from the true source location. Thus, RIU is not accurate for locating the source and onset time. One reason for this failure is the division between the windowed observed wavefield and the predicted wavefield, which causes instability within this algorithm.

2.3.4 Mutual information

Measuring the similarity between two images can also be interpreted as measuring the amount of information shared by two images or measuring how well one image explains the other (Shannon, 1948). In communication theory, mutual information, which describes the "rate of transmission information", is developed in order to quantitatively analyze the common information shared by two images. Collignon *et al.* (1995), Viola (1995), Wells *et al.* (1997), and Maes *et al.* (1997) independently and simultaneously proposed the mutual information for intermodality medical image registration.

The equation is used to calculate the mutual information in this research is

$$MI(\mathbf{x},t) = \sum_a\sum_b P(a,b)log\frac{P(a,b)}{P(a)P(b)}, \tag{2.10}$$

where $P(a,b)$ is the joint probability density function (PDF) of the windowed observed wavefield and the predicted reconstructed wavefield, $P(a)$ is the intensity PDF of the predicted reconstructed wavefields, and $P(b)$ is the intensity PDF of the windowed observed wavefields.

The joint PDF $P(a,b)$ denotes the probability of occurrence of the intensity pair (a, b) with intensity a and b located at the predicted wavefield and windowed observed wavefield, respectively. The positions of intensities are the same. This joint PDF $P(a,b)$ can be approximated by

$$P(a,b) = \frac{N(a,b)}{N}. \tag{2.11}$$

Here, $N(a,b)$ is the number of voxels that have the intensity a (in the predicted wavefield) and b (in the windowed observed wavefield) at the same locations for both wavefields; N is the number of voxels in the wavefield. The intensity of the PDF of the predicted wavefield and the windowed observed wavefield can be expressed as

$$P(a) = \sum_b P(a,b) = \frac{1}{N}\sum_b N(a,b) = \frac{N(a)}{N} \tag{2.12}$$

and

$$P(b) = \sum_a P(a,b) = \frac{1}{N}\sum_a N(a,b) = \frac{N(b)}{N}\,, \tag{2.13}$$

respectively, which can be calculated directly by projecting the joint PDF on two orthogonal axes. The intensity of PDF can also be interpreted as the number of voxels with an intensity value of a or b divided by the number of voxels in the wavefield.

Figure 2.27 shows four pairs of predicted wavefields (left column) and windowed observed wavefields (right column); their corresponding histograms are shown in Figure 2.28. The first row shows a duplicated predicted wavefield. Because these two figures are identical, their joint PDF shown in Figure 2.28(a) is distributed along the diagonal. In the second row, the source location of the predicted wavefield is {132 m, 334 m} and the center of the windowed observed wavefield is {132 m, 334 m, 0.04 s}. The joint PDF of these two wavefields (shown in Figure 2.28(b)) is located close to the diagonal line but is not precisely concentrated along that line. Due to various errors (for example, noisy seismograms and incorrect velocity models), it is impossible to generate the same histogram as shown in Figure 2.28(a). In the third panel, both wavefields are centered at the source onset time but not at the true source location. In the fourth panel, two wavefields are located at

the source location but not at the source onset time. Based on Figure 2.28, I find that the joint histogram dispels or "blurs" when the wavefields are not identical; for example, the brightest regions of the histogram become less bright and the number of dark regions are reduced. The ideal case is shown in Figure 2.28(a) where only one line is located along the diagonal direction.

I apply equation 2.10 to the calculated joint PDFs to compute the mutual information, shown in Figure 2.29. Based on the information shown in Figure 2.29, we can tell that mutual information produces a higher resolution than other methods for identifying the source location. Because mutual information can be thought of as a measurement of how well one image explains another, or how much one image shares information with another, this method achieves its maximum value when two images are in optimal alignment. Therefore, I search for the maximum value within mutual information. The detected source location is shown in Figure 2.30. The source location provided by this method is identical with locations provided by N_{cc} and MID. Therefore, the location accuracy of this method is also less than 1/5 of a wavelength.

2.3.5 Comparison of similarity measurements

In this section, four methods for similarity comparison are discussed: normalized cross-correlation, minimizing intensity difference, ratio image uniformity, and mutual information. Ratio image uniformity is the only method that failed in providing a correct source location. The location accuracy of the other three methods is similar. However, the computation cost of mutual information is high because this method requires the extra step of calculating the joint PDF. Although the resolution of mutual information is the highest among these three methods, I prefer normalized cross-correlation and minimizing intensity difference for calculating the similarity. When the observed and predicted images are identical, their minimizing intensity difference value is 0 and normalized cross-correlation is 1. When two images are different, however, their normalized cross-correlation value ranges between -1 and $+1$; there is no upper limit for minimizing intensity difference value. Thus, by applying normalized cross-correlation, I can quantitatively understand the similarity between the two images; in contrast, minimizing intensity difference fails to provide quantitative information.

After analyzing the four methods, I prefer normalized cross-correlation for calculating the similarity between the observed reconstructed wavefield and the predicted wavefields.

2.4 Signal-to-noise ratio

In this section, I discuss the robustness of the method in noisy data because in reality, the signal to noise (S/N) ratio is much lower than that shown in Figure 2.7. Thus, in order to simulate seismograms recorded in reality, random noise with a frequency band from 0 Hz to 140 Hz is added to the seismogram shown in Figure 2.7. Figure 2.31 shows the three seismograms with different S/N ratios. The SNR is calculated by

$$SNR = \left(\frac{A_{signal}}{A_{noise}} \right)^2 , \tag{2.14}$$

where A represents the root-mean-square. The SNR of these three seismograms are 0.185, 0.05, and 0.013. After I apply the method, three calculated normalized cross-correlations are shown in Figure 2.32. The intersections of lines indicate the true micro-earthquake locations. The distributions of these three plots are similar. The identified source locations are shown in the right column of Figure 2.33, and their snapshots of the reconstructed wavefields at the source onset time are shown in the left column. The focus of the wavefield is less obvious when the S/N ratio decreases. As shown in Figure 2.33(f), the source location I detected is incorrect. Nevertheless, I regard my method as robust because I can provide the source location with an accuracy of less than 1/5 of a wavelength even when the S/N ratio is as low as 0.05.

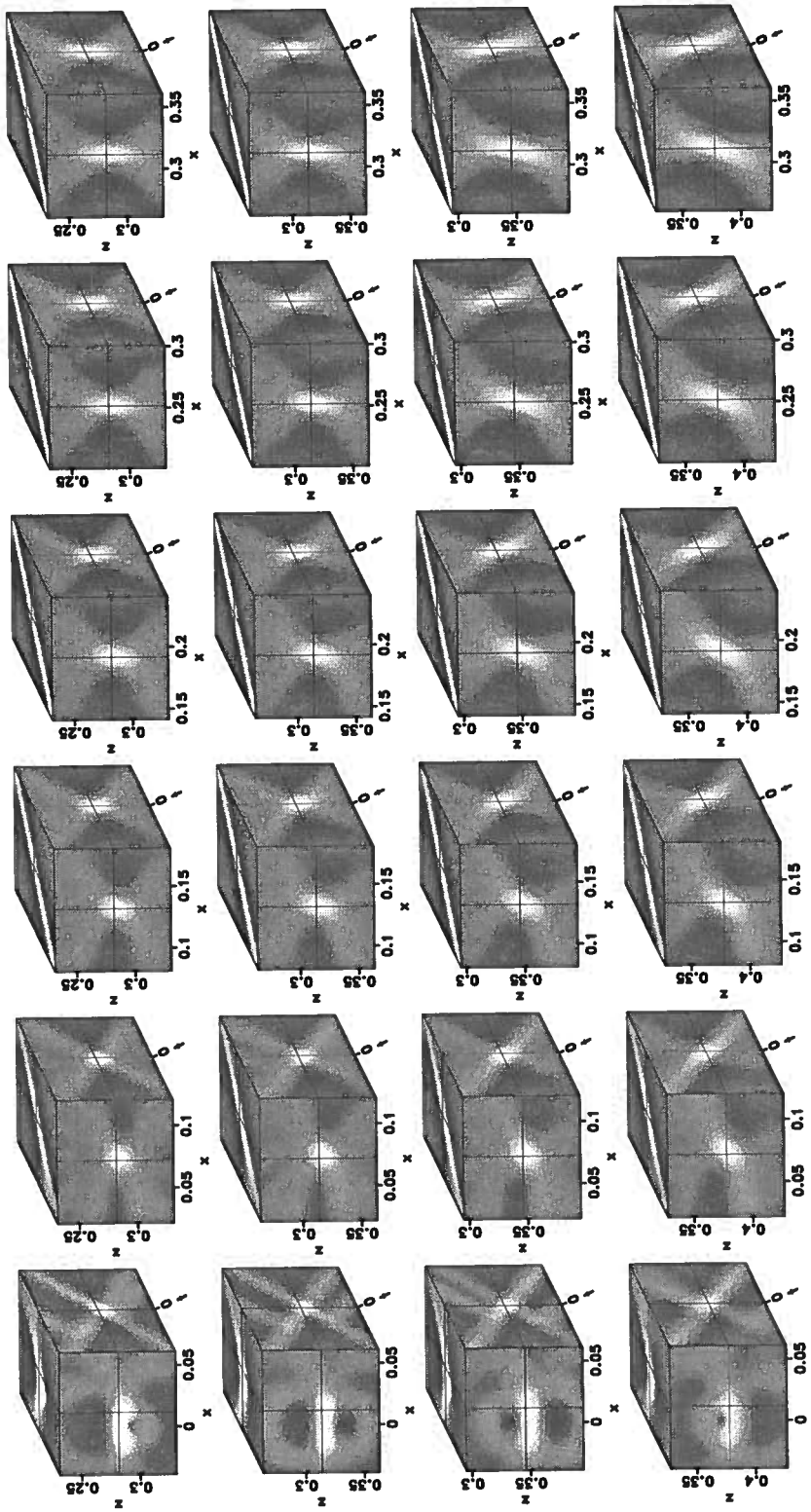


Figure 2.16. Reconstructed wavefield database. For each predicted reconstructed wavefield, the intersections of lines indicate the source location.

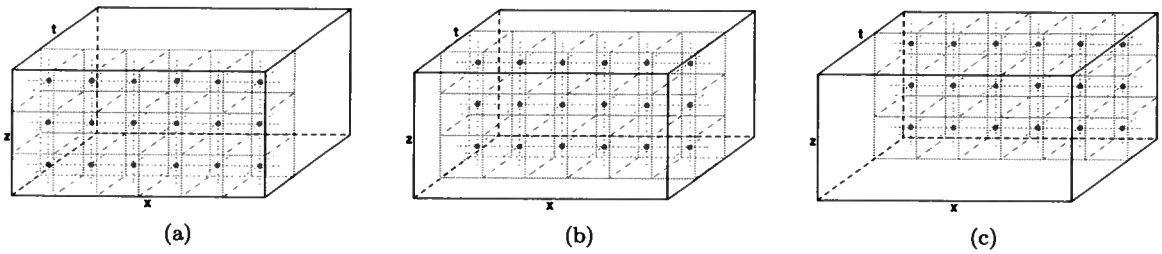


Figure 2.17. Similarity comparison between the observed reconstructed wavefield and the predicted reconstructed wavefields at different times. Time increases from (a) to (c).

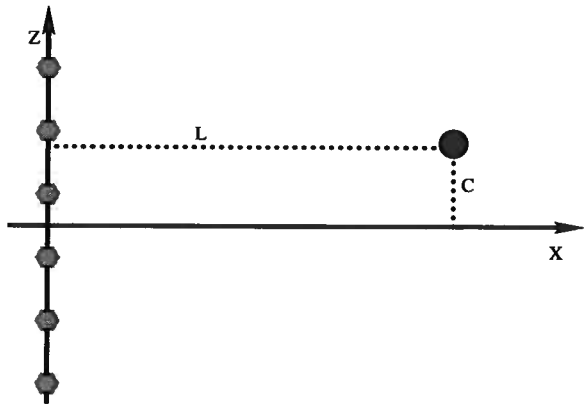


Figure 2.18. Range L and cross-range C. The dot represents the source and the polygons represent the receivers.

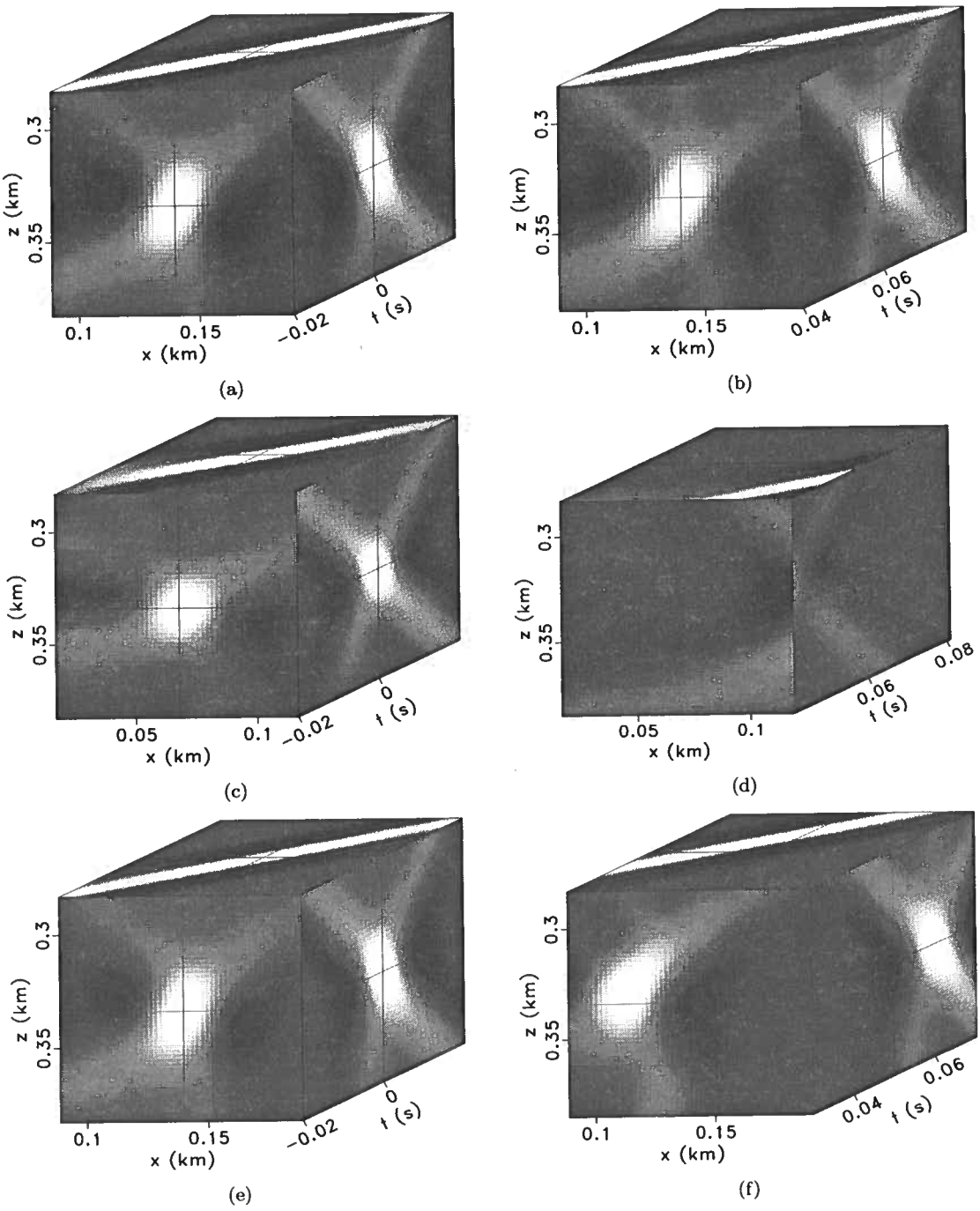


Figure 2.19. Similarity comparison between three pairs of predicted reconstructed wavefields and windowed observed wavefields. (a) and (b) are located close to the source location and onset time. (c) and (d) are located at the source onset time but not at the source explosive location. (e) and (f) are located at the source location but not at the source onset time.

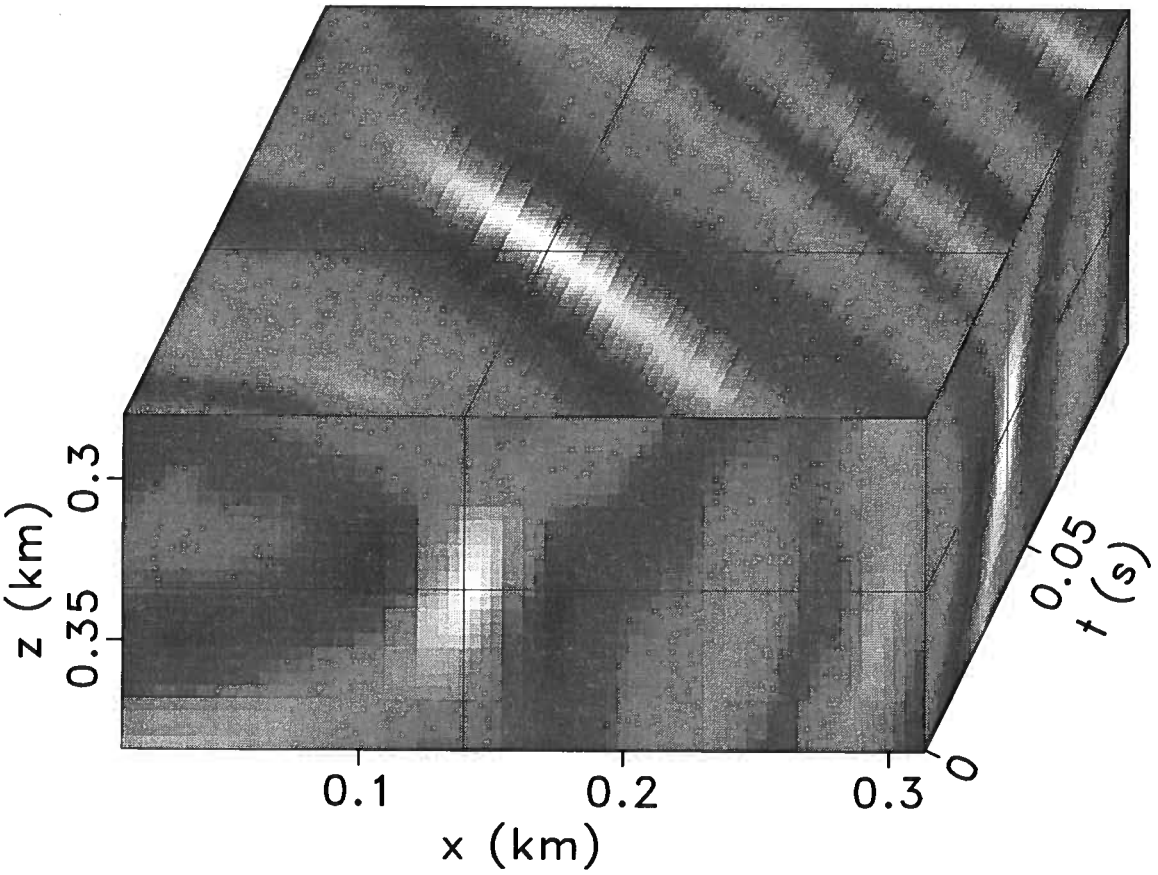


Figure 2.20. Normalized cross-correlation. The intersections of the lines indicate the location with the maximum value of N_{cc}.

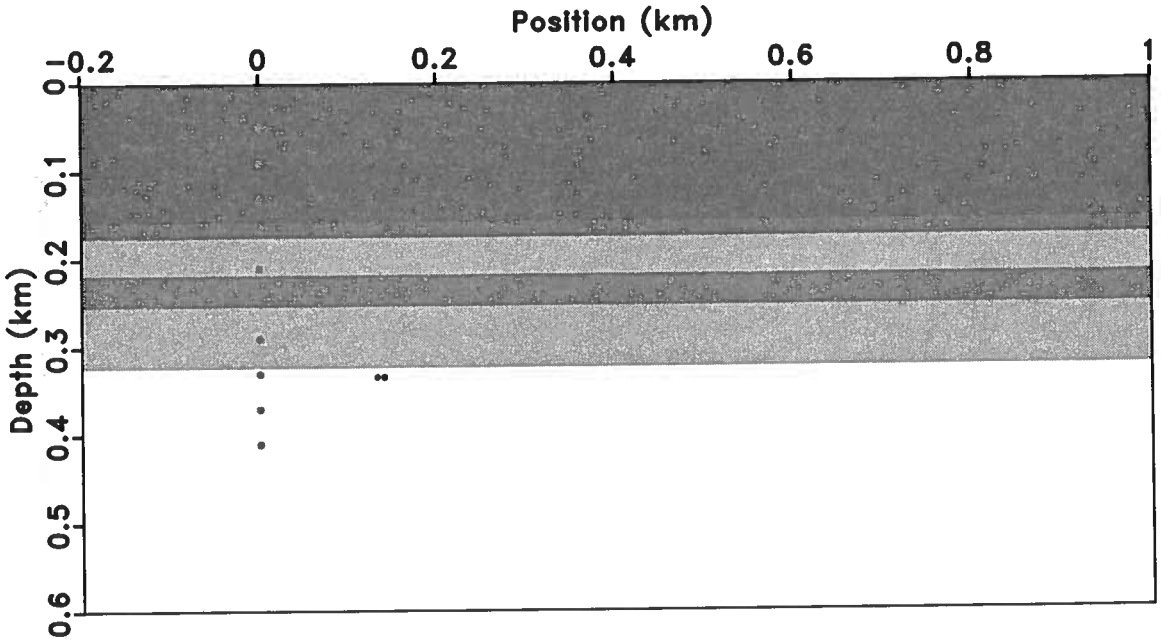


Figure 2.21. A comparison between the picked source location, represented by the left dot of the two dots located at $x \approx 130$ m, and the true source location, represented by the right dot of the two dots.

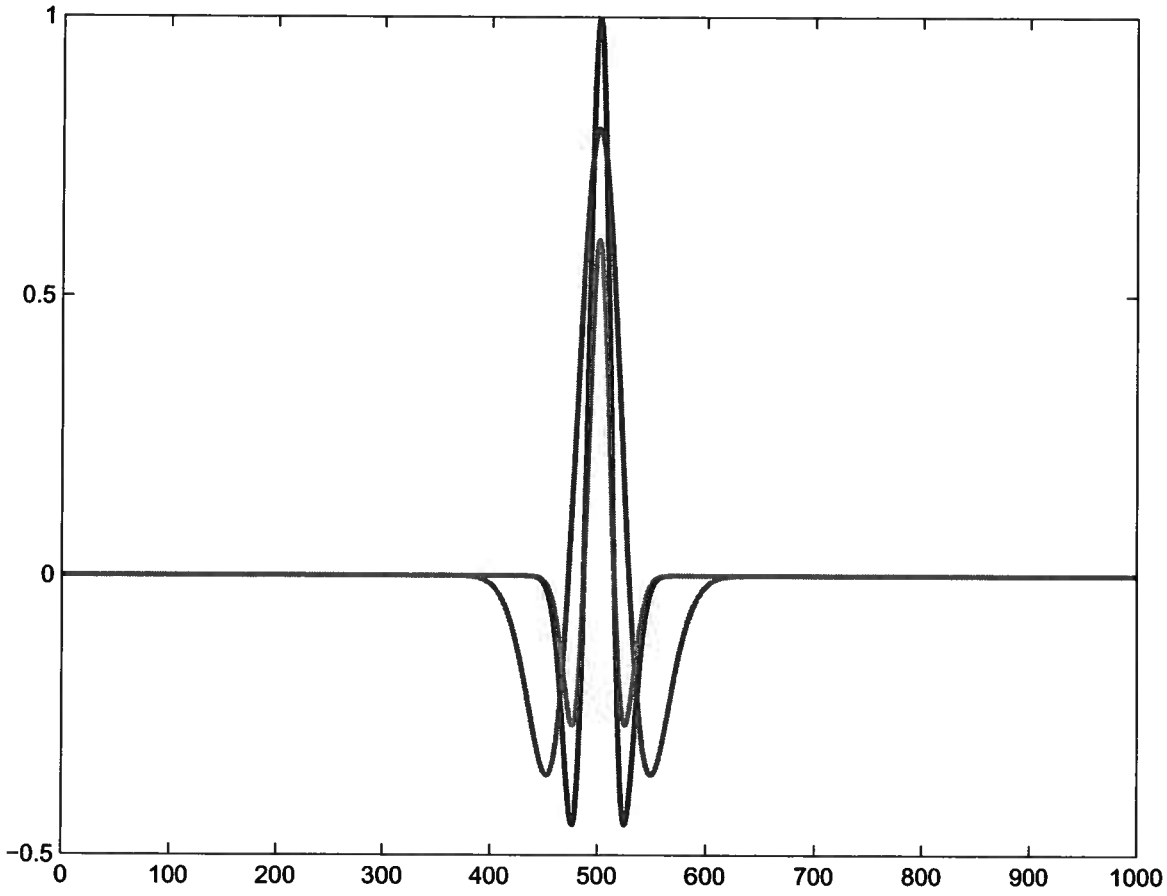


Figure 2.22. The sensed wavelet is represented by the tallest line, the reference wavelet A is represented by the lowest one, and the reference wavelet B is represented by the middle one.

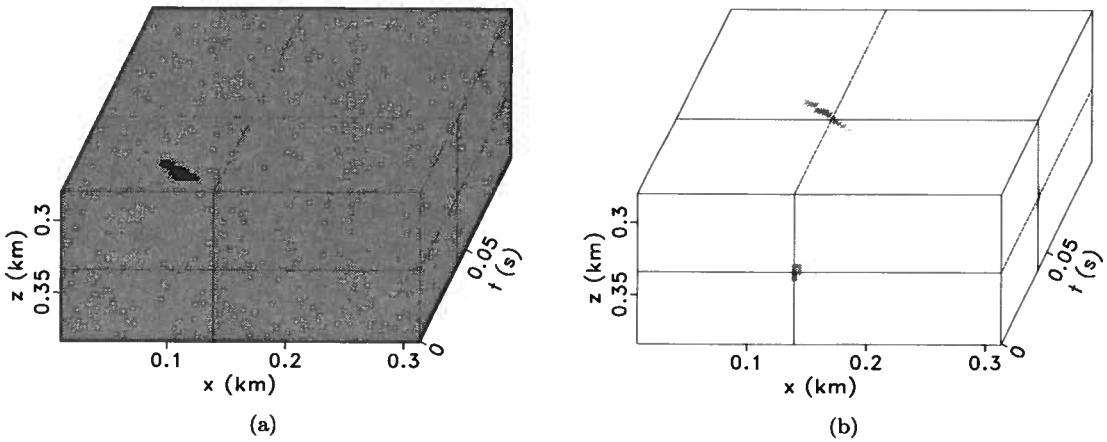


Figure 2.23. Minimizing intensity difference. (a) MID without a threshold, (b) MID with a threshold with the value of 0.01

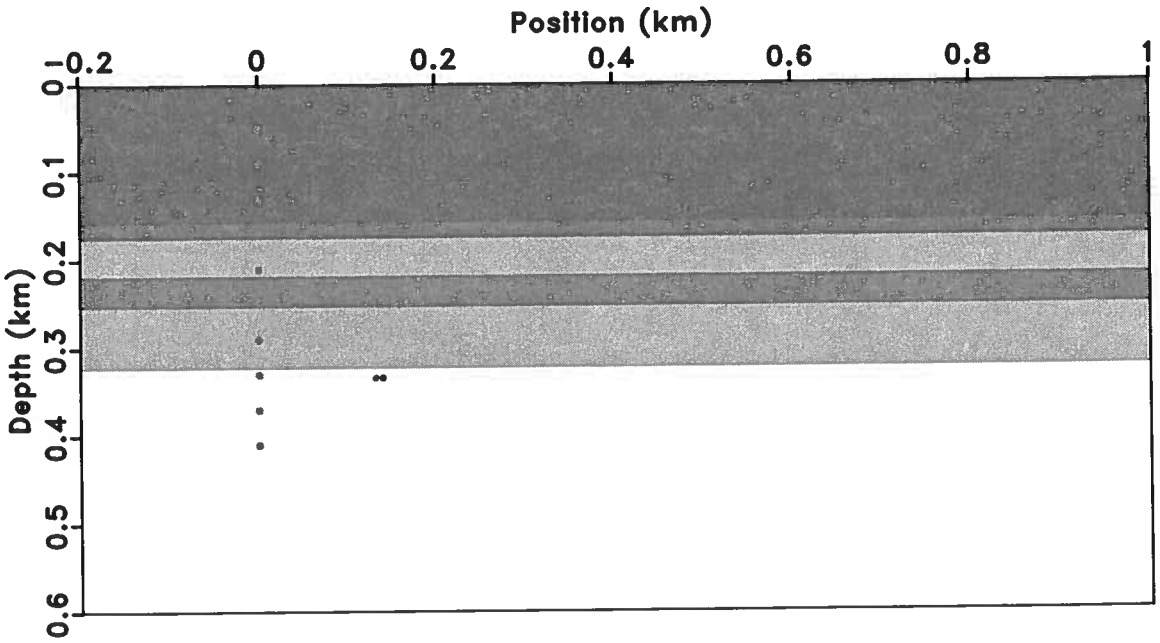


Figure 2.24. A comparison between the true source location, represented by the right dot of the two dots located at $x \approx 130$ m, and the detected source location, represented by the left dot of the two dots.

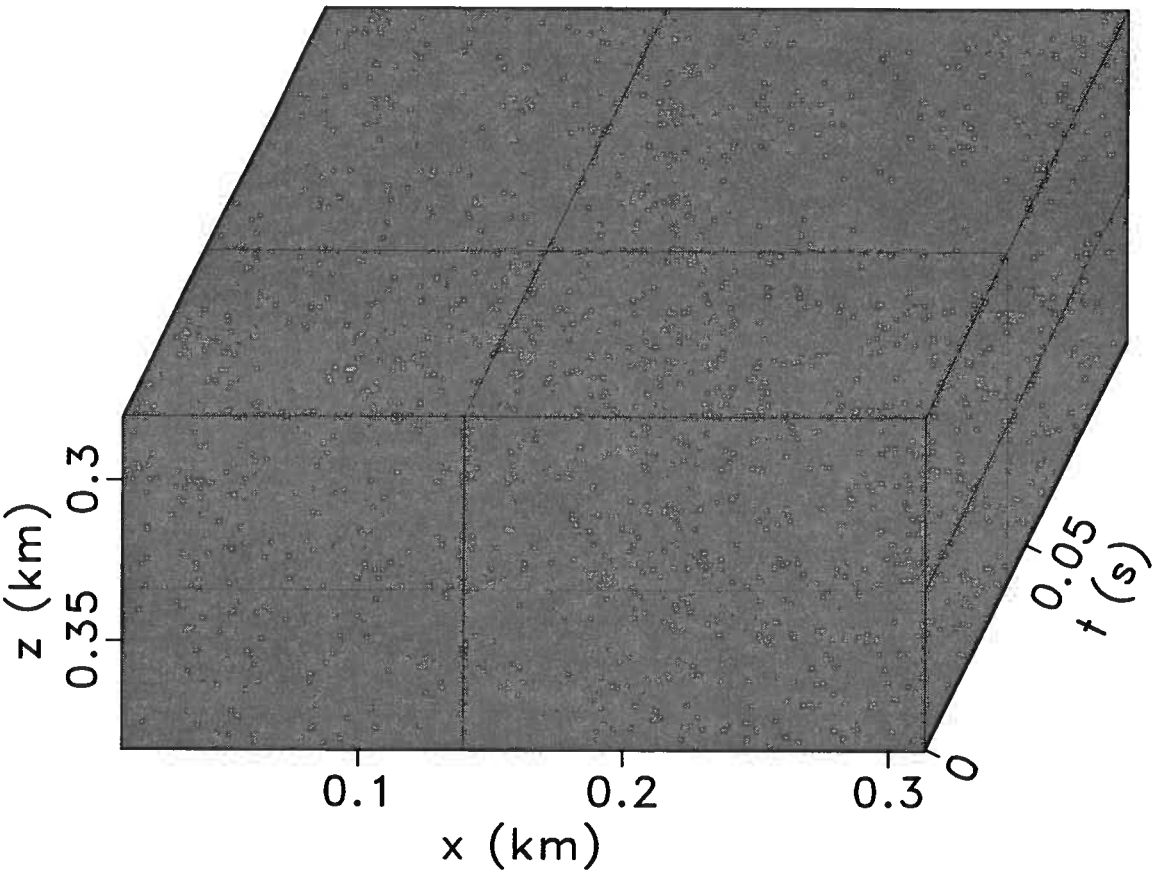


Figure 2.25. Ratio image uniformity where the intersections of lines indicate the true source location. From this plot, it is difficult to get information about the estimated source position.

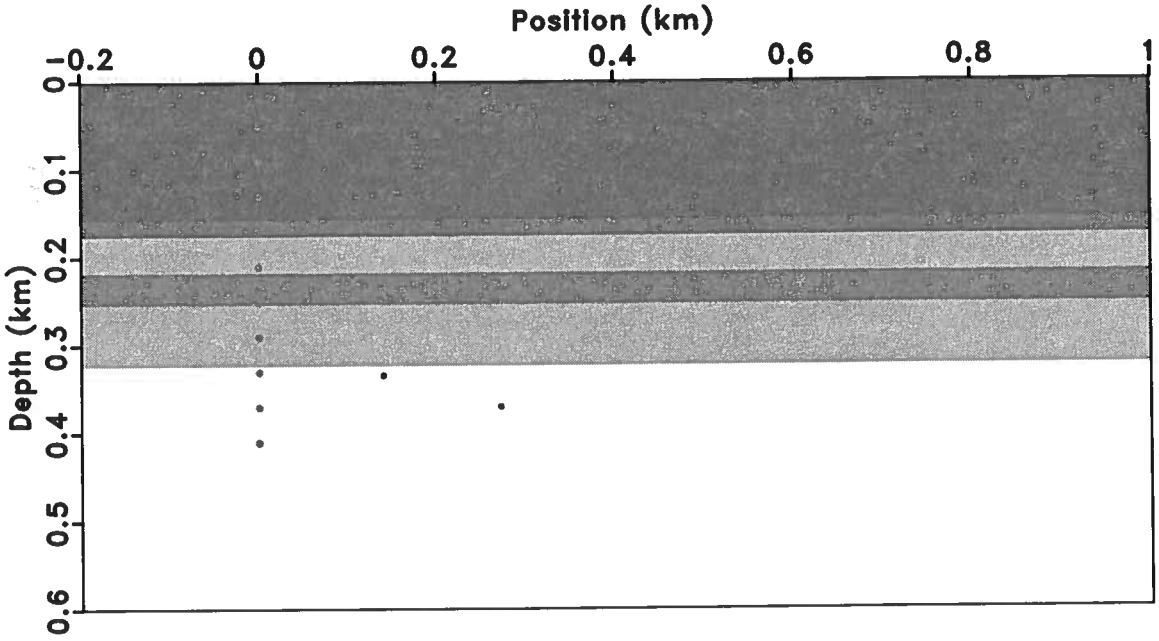


Figure 2.26. A comparison between the true source location, represented by the dot located at $x \approx 130$ m, and the detected source location, represented by the dot located at $x \approx 300$ m .

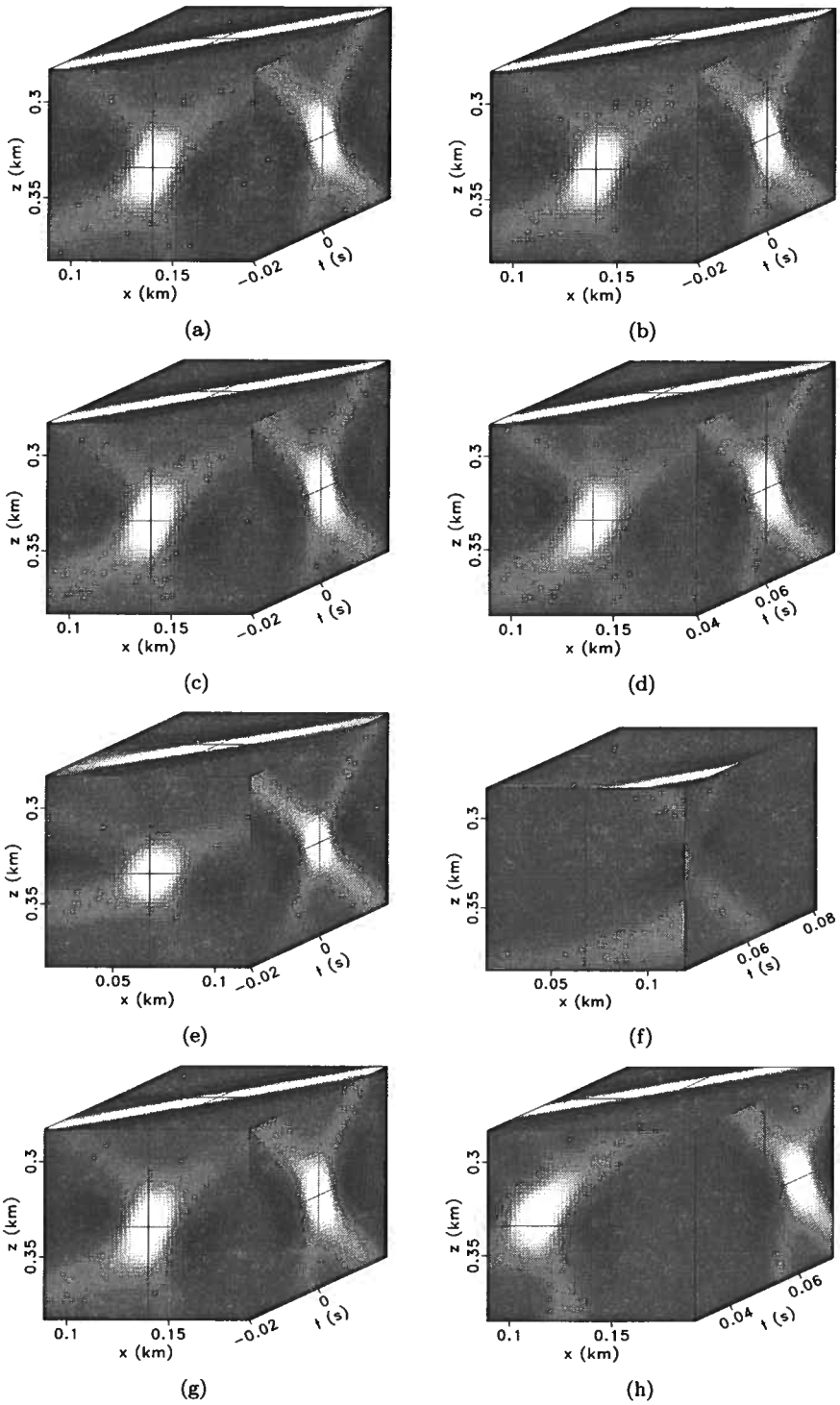


Figure 2.27. Four pairs of the predicted wavefields and windowed observed wavefields where (a) and (b) are the same, (c) and (d) are centered around the true source location and onset time, (e) and (f) are centered at the source onset time, and (g) and (h) are centered at the source location.

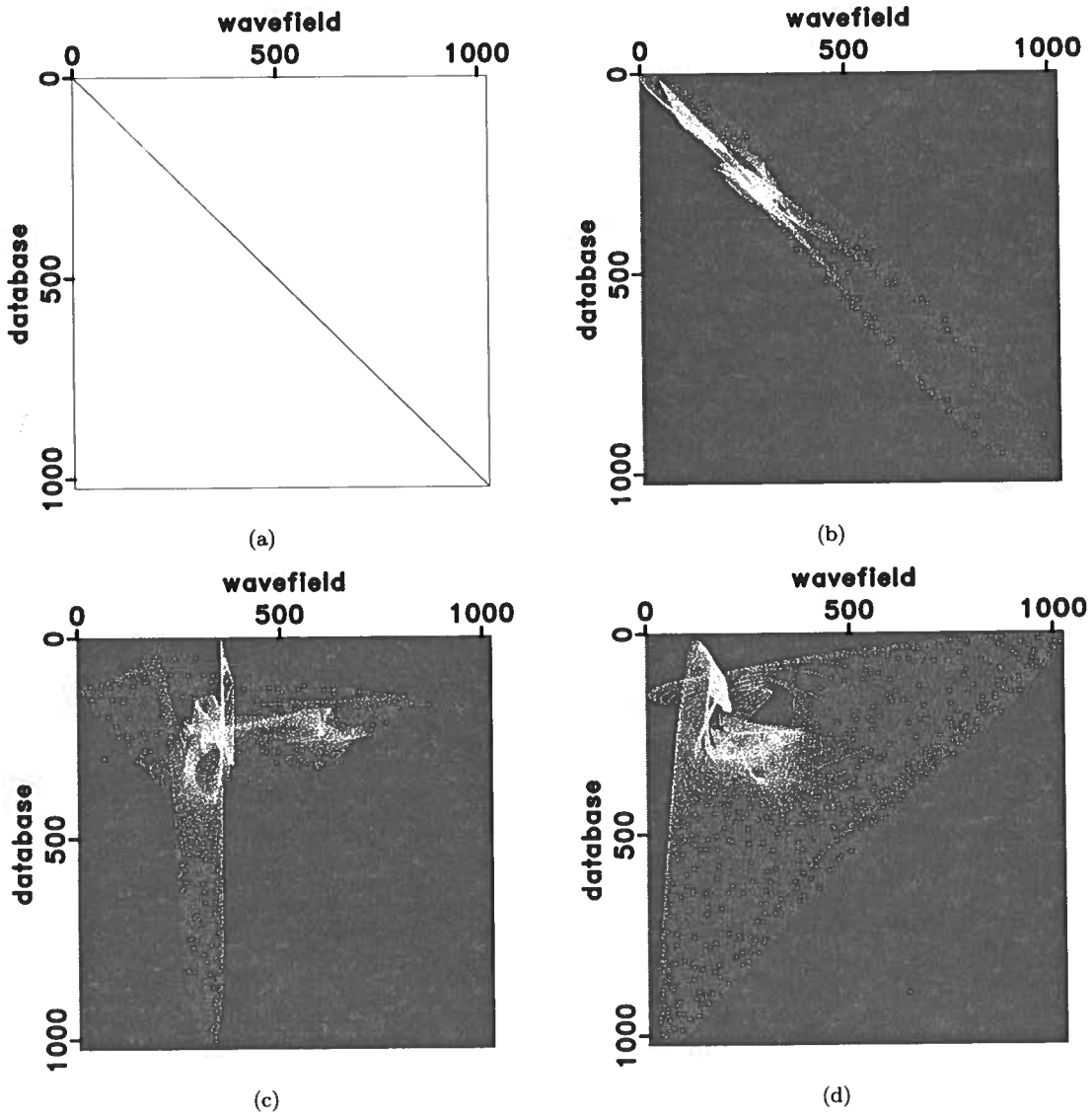


Figure 2.28. Joint histograms. (a) is the histogram of plots (a) and (b) of Figure 2.27, (b) shows the histogram of (c) and (d), (c) illustrates the histogram of (e) and (f), and (d) shows the histogram of (f) and (h).

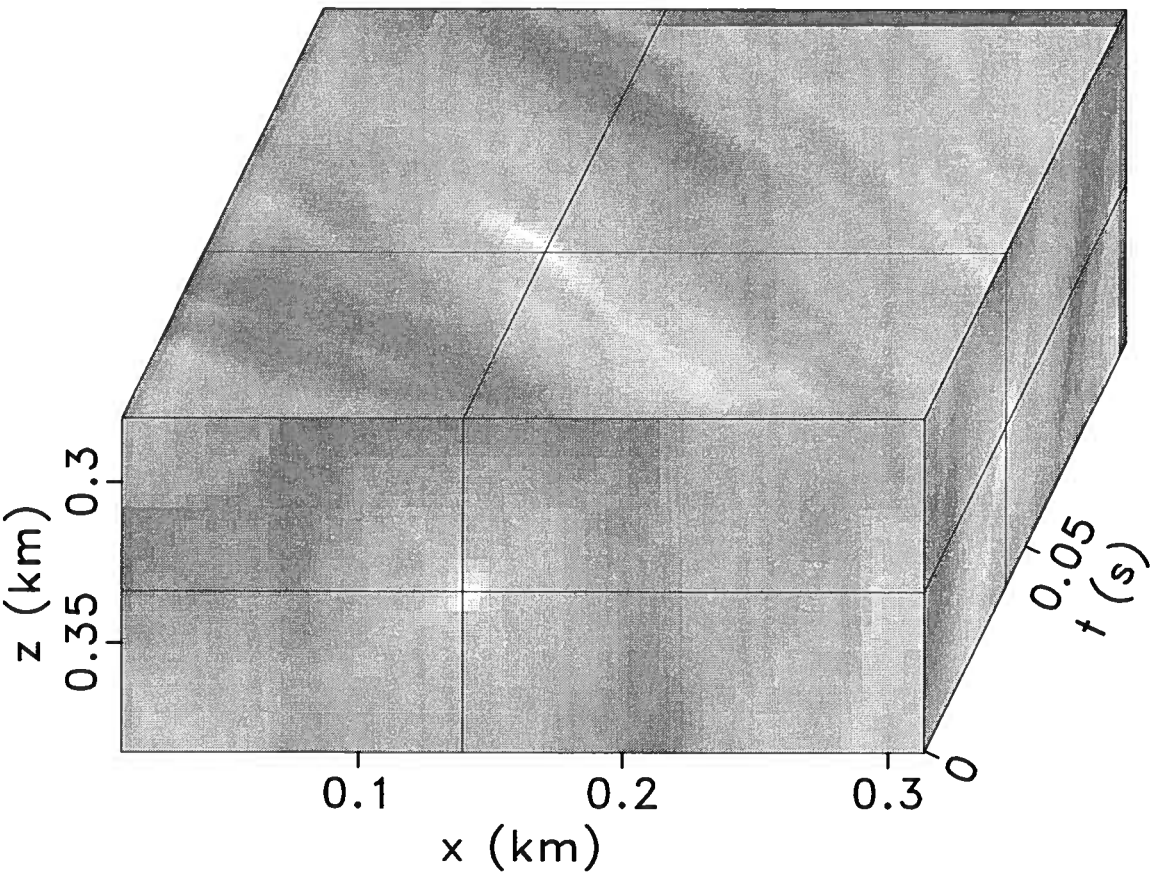


Figure 2.29. Mutual information, where the intersections of lines indicate the true source location.

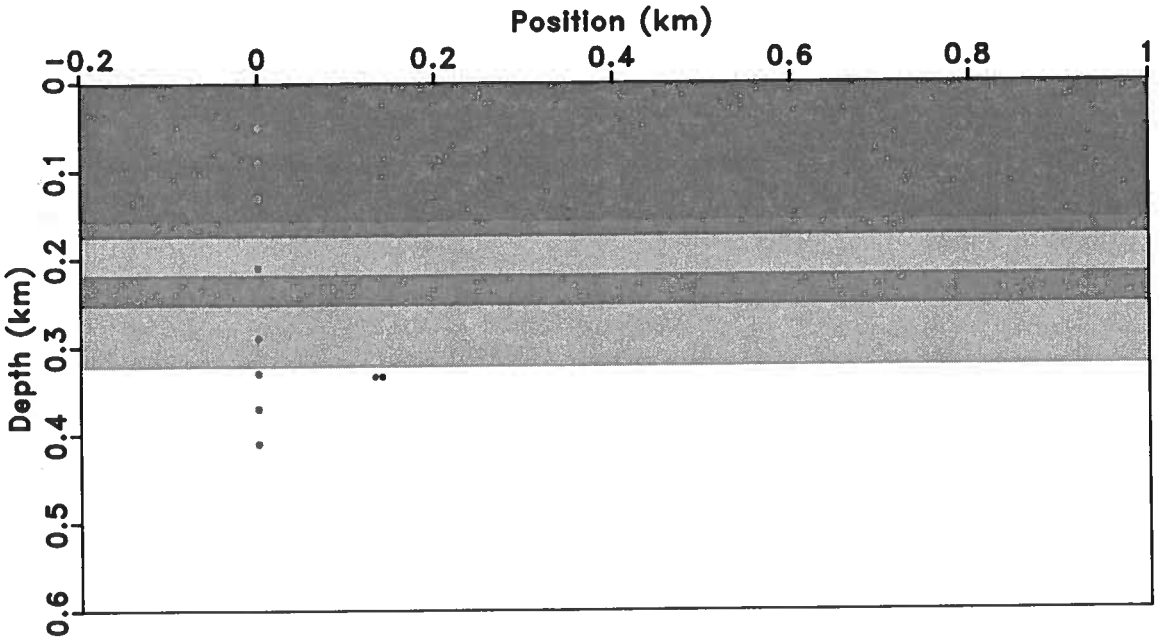


Figure 2.30. A comparison between the true source location, represented by the right dot of the two dots located at $x \approx 130$ m, and the detected source location, represented by the left dot of the two dots.

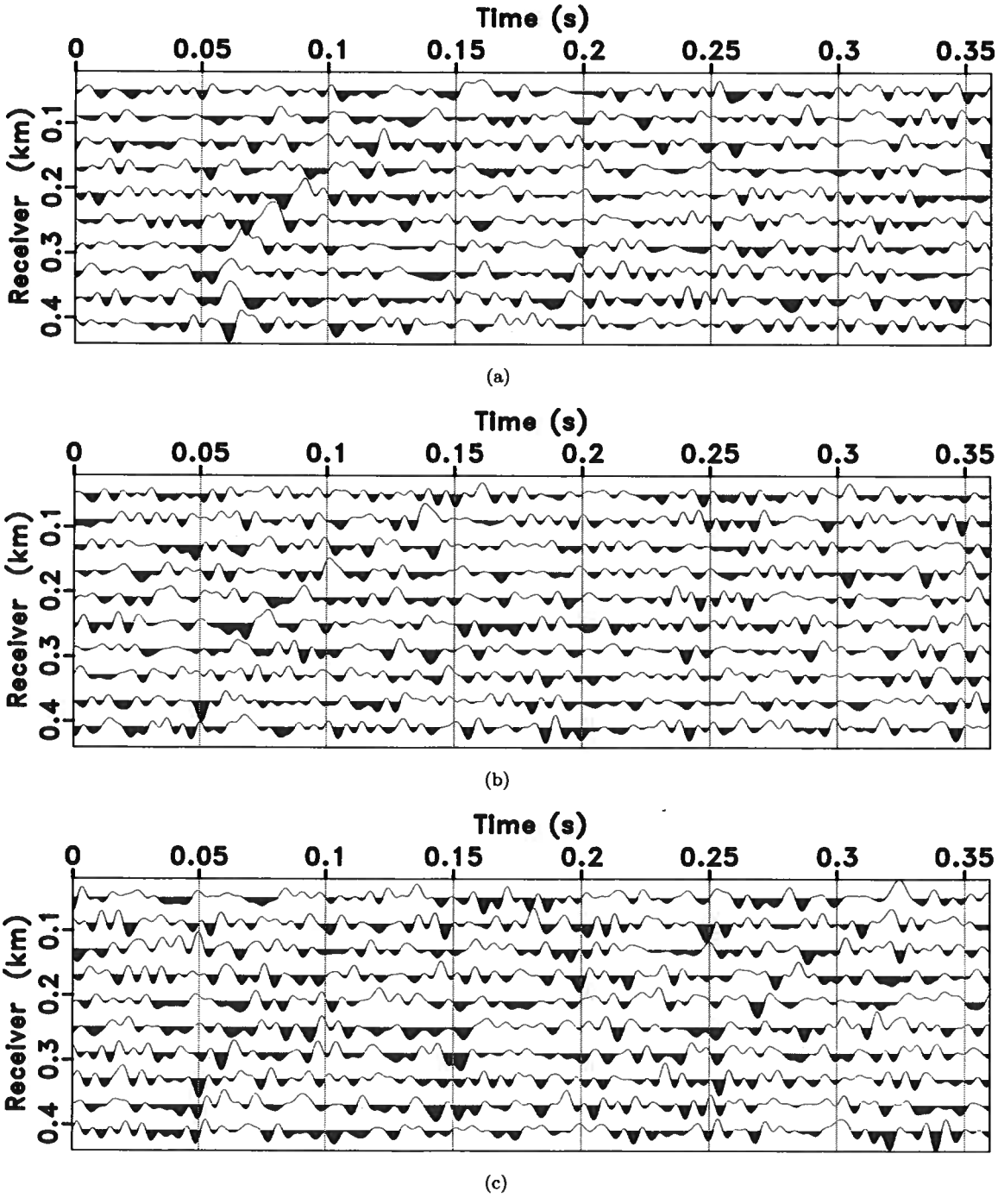


Figure 2.31. Noisy seismograms where S/N rations are (a) 0.185, (b) 0.05, and (c) 0.013.

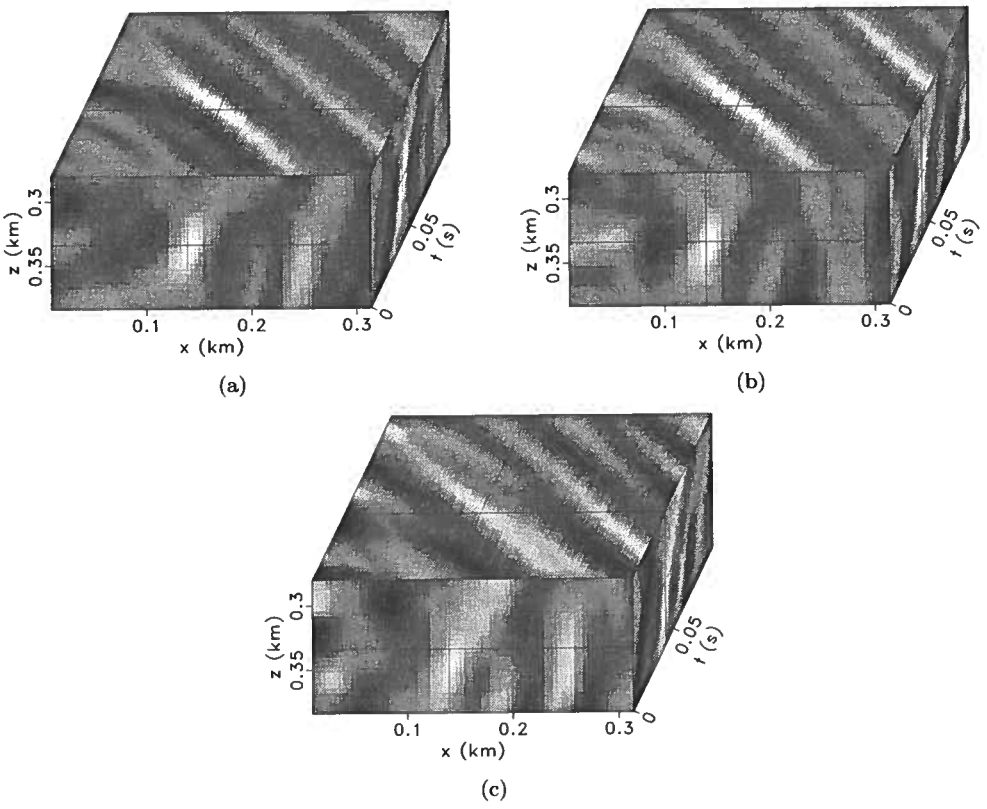


Figure 2.32. Normalized cross-correlations corresponding to the seismograms shown in Figure 2.31.

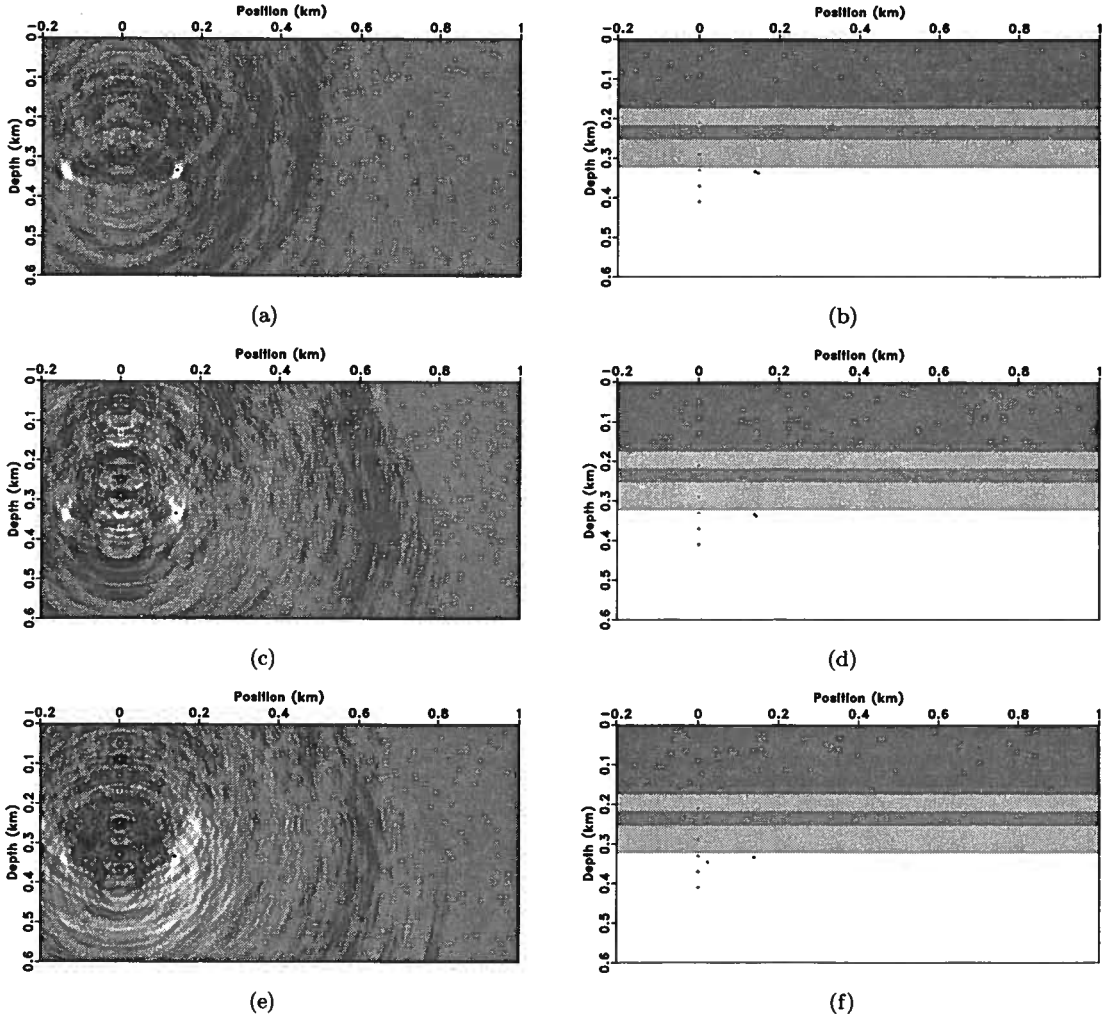


Figure 2.33. The left column shows snapshots of the reconstructed wavefields; the right column shows the true source locations at $\{x, z\} = \{140 \text{ m}, 334 \text{ m}\}$ and the detected source locations. (a) and (b) indicate the result of the seismogram shown in Figure 2.31(a) with a SNR which is 0.185; (c) and (d) indicate the result of the seismogram shown in Figure 2.31(b) with a SNR which is 0.05; (e) and (f) indicate the result of the seismogram shown in Figure 2.31(c) with a SNR which is 0.013. The focus of reconstructed wavefield becomes less obvious when S/N ratio decreases.

Chapter 3

Probability of a micro-seismic event location

In the preceding chapter, I assumed that the location and onset time that produce the maximum similarity are the detected hypocentre and excitation time of a micro-earthquake. However, due to various errors, such as noisy seismograms and incorrect velocity, we are not able to provide the exact source location and onset time. Therefore, in order to quantitatively describe the uncertainty of the detected the source location, I exploit Bayesian inversion theory. This allows me to illustrate the probability of the source and onset time I located for a micro-seismic event. I also apply this method to locate multiple micro-earthquakes which are triggered within a certain area and a fixed range of time. By studying the distribution of micro-earthquakes, we can get a sense of the growth of fractures.

3.1 Bayesian inversion theory

Bayesian inversion theory uses probability density functions as a measurement of a given state of knowledge (Jaynes, 2003). A crucial feature of Bayesian theory is that a probability can be assigned to a hypothesis. According to Bayesian theory, we begin investigation with a certain state of knowledge which is independent of any subsequent measurement we make. We assign to any possible model acceptable to the prior state of knowledge a likelihood which quantifies how probable it is for a particular model to represent reality (Tarantola, 2005). The posterior combines the prior understanding of the model parameters with additional information derived from measurements and assumptions about the physical laws governing the process under investigation. This approach allows us to incorporate any a priori information into the data, parameters, and theory. As discussed by

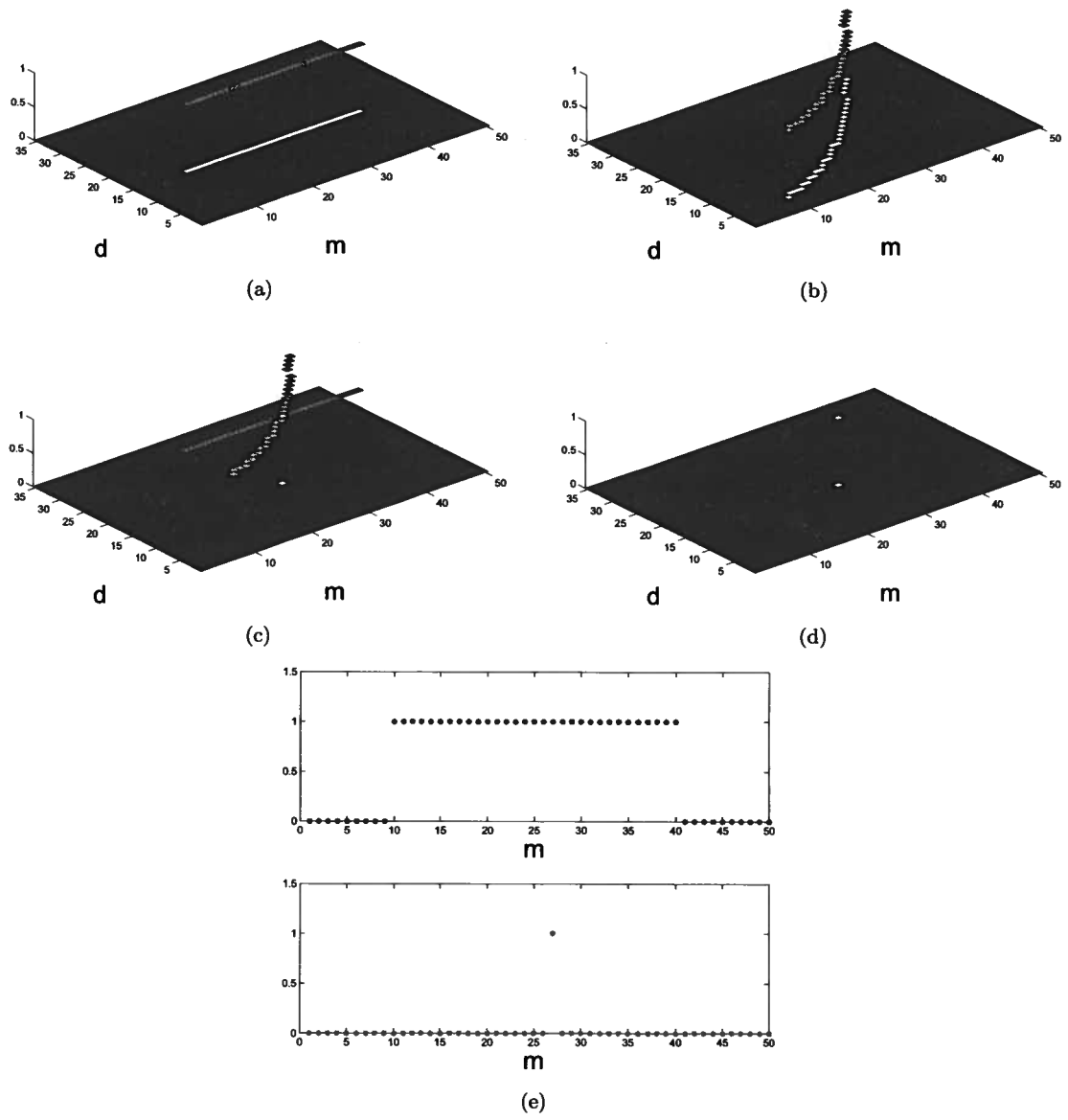


Figure 3.1. (a) A priori probability density function $\rho(\mathbf{m}, \mathbf{d})$ represented by a delta function corresponding to $\rho_{\mathfrak{M}}(\mathbf{m})$ which is also represented by a step function. (b) Theoretical probability density function $\Theta(\mathbf{m}, \mathbf{d})$ representing the physical relationship given by $G(\mathbf{m})$ and assumed to be exact. (c) Conjunction between $\rho(\mathbf{m}, \mathbf{d})$ and $\Theta(\mathbf{m}, \mathbf{d})$. (d) A posteriori probability density function $\sigma(\mathbf{m}, \mathbf{d})$ indicating that the solution to the Bayesian inversion problem is given by a single model. (e) A comparison between the marginal probabilities showing $\rho_{\mathfrak{M}}(\mathbf{m})$ in the upper panel and $\sigma_{\mathfrak{M}}(\mathbf{m})$ in the lower panel.

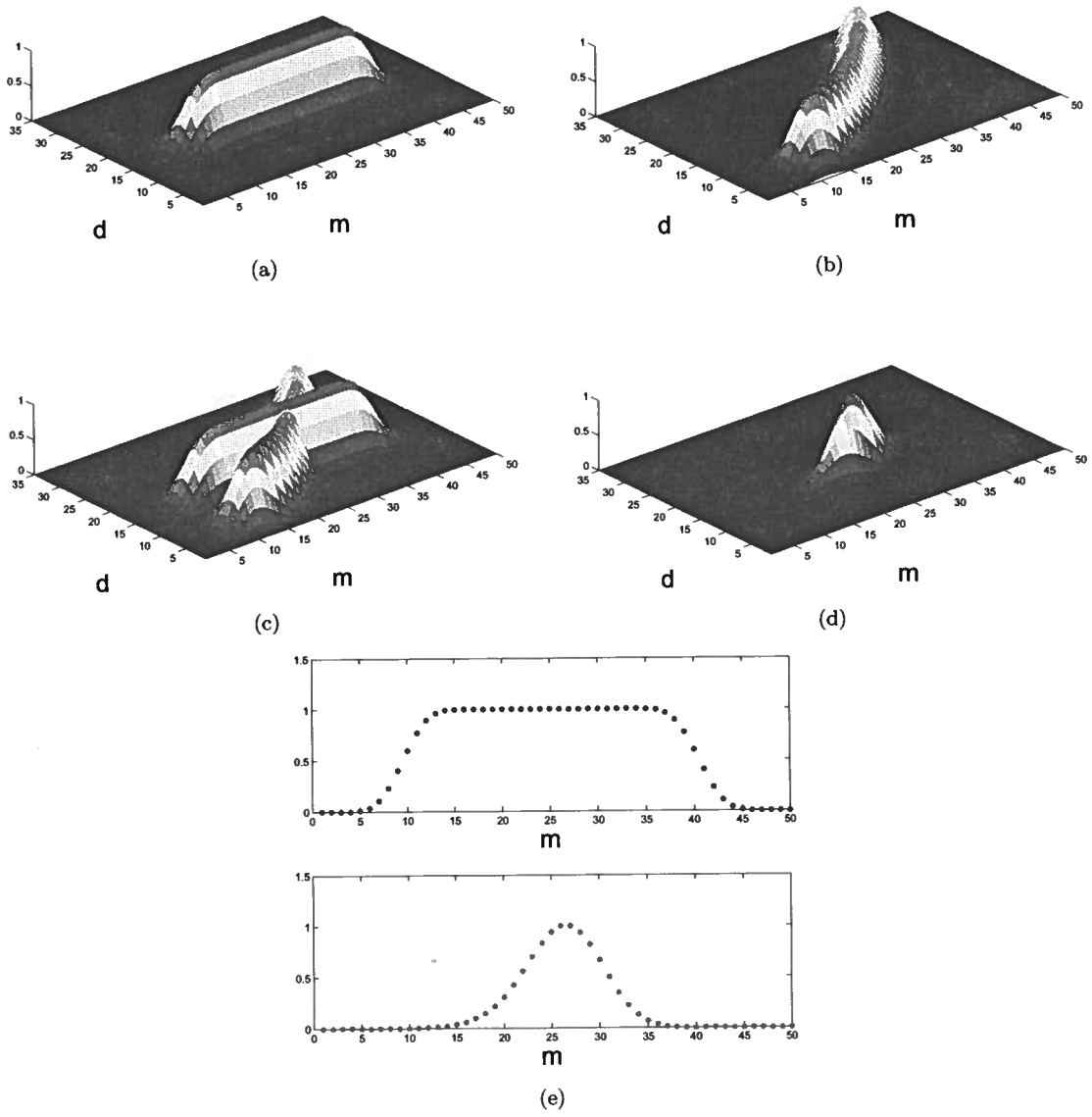


Figure 3.2. (a) A priori probability density function $\rho(\mathbf{m}, \mathbf{d})$ represented by a Gaussian function corresponding to $\rho_{\mathfrak{M}}(\mathbf{m})$ represented by a tapered step function. (b) Theoretical probability density function $\Theta(\mathbf{m}, \mathbf{d})$ representing the physical relationship given by $G(\mathbf{m})$ and assumed to be uncertain. (c) Conjunction between $\rho(\mathbf{m}, \mathbf{d})$ and $\Theta(\mathbf{m}, \mathbf{d})$. (d) A posteriori probability density function $\sigma(\mathbf{m}, \mathbf{d})$ indicating that the solution to the Bayesian inversion problem is given by several models corresponding to the uncertainties of the prior and theoretical probability density functions. (e) A comparison between the marginal probabilities showing $\rho_{\mathfrak{M}}(\mathbf{m})$ in the upper panel and $\sigma_{\mathfrak{M}}(\mathbf{m})$ in the lower panel.

Tarantola (2005), the posterior state of information is obtained using the Bayes theorem

$$\sigma(\mathbf{m},\mathbf{d}) = k\frac{\rho(\mathbf{m},\mathbf{d})\Theta(\mathbf{m},\mathbf{d})}{\mu(\mathbf{m},\mathbf{d})}\,, \tag{3.1}$$

where $\mathbf{m}$ represents the model, $\mathbf{d}$ represents the data, $\rho(\mathbf{m},\mathbf{d})$ represents the a priori state of information, $\Theta(\mathbf{m},\mathbf{d})$ represents the theoretical state of information, $\mu(\mathbf{m},\mathbf{d})$ represents the homogeneous state of information, k represents a normalization constant, and $\sigma(\mathbf{m},\mathbf{d})$ represents the a posteriori state of information.

We can define two states of knowledge based on equation 3.1. The *prior* characterizes our knowledge about the model and data parameters before any measurements are taken. This information is based on the general knowledge of the phenomenon under investigation, of the distribution of model parameters, and on the assumptions about the reliability of the measurements. The *posterior* characterizes our knowledge about the model and data parameters after measurements are made and data are processed and interpreted. Both states of knowledge are characterized by probability density functions (PDFs) linking model and data parameters. Ideally, the measurements help refine the prior into the posterior state of knowledge which provides a tighter connection between model and data parameters.

According to equation 3.1, the a posteriori state of information $\sigma(\mathbf{m},\mathbf{d})$ is computed by combining the theoretical function $\Theta(\mathbf{m},\mathbf{d})$ with the prior information $\rho(\mathbf{m},\mathbf{d})$. The solution to the inversion problem is given by $\sigma_{\mathfrak{M}}(\mathbf{m})$ which represents the posterior marginal probability on the model space and is obtained by integrating the a posteriori probability density function over data:

$$\sigma_{\mathfrak{M}}(\mathbf{m}) = \int_{\mathfrak{D}} \sigma(\mathbf{m},\mathbf{d})d\mathbf{d}\,, \tag{3.2}$$

where $\mathfrak{D}$ is the data space. When the uncertainty associated with the theoretical state of information is negligible, $\sigma_{\mathfrak{M}}(\mathbf{m})$ is

$$\sigma_{\mathfrak{M}}(\mathbf{m}) = \nu\rho_{\mathfrak{M}}(\mathbf{m})\rho_{\mathfrak{D}}(G(\mathbf{m}),\mathbf{d}^{obs})\,, \tag{3.3}$$

where $\rho_{\mathfrak{M}}(\mathbf{m})$ is the prior state of information on the model space, $\rho_{\mathfrak{D}}(\mathbf{d})$ is the prior information on the data space, and ν is a normalization constant. In order to describe Bayesian inversion theory, I need to discuss the following elements: model ($\mathbf{m}$) and model space, data ($\mathbf{d}$) and data space, the prior state of knowledge (a priori PDF $\rho(\mathbf{m},\mathbf{d})$), the theoretical relations between model and data

(theoretical PDF $\Theta(\mathbf{m},\mathbf{d})$), the posterior state of knowledge (a posteriori PDF $\sigma(\mathbf{m},\mathbf{d})$), and the posterior marginal probability on the model space(a marginal PDF $\sigma_{\mathfrak{M}}(\mathbf{m})$).

3.1.1 Model and data spaces

The model space ($\mathfrak{M}$) is composed of a set of individual models which can be denoted by $\mathbf{m}_i$ ($i \in \{1,2,...,N\}$). A particular parameterization has to be chosen in order to quantitatively describe the model, and these parameters correspond to the coordinates of a system. The model space represents all the models that can possibly characterize the physical phenomenon under investigation.

In this research, the model parameters used for the micro-earthquake location problem are $\mathbf{m} = \{\mathbf{x}, t\}$, where $\mathbf{x} = \{x, y, z\}$ describes the position of a micro-seismic in space and t describes the onset time.

The data space $\mathfrak{D}$ is composed of all instrumental responses to the investigated models which can be denoted by $\mathbf{d}_i$ ($i \in \{1,2,...,M\}$). The data space represents all data that can possibly be recorded and characterizes the physical phenomenon under investigation.

In this research, the data $\mathbf{d}$ represent the wavefields reconstructed from recorded seismograms in the space around the most probable micro-earthquake source locations and onset times. In this context, it is irrelevant whether the data are acquired in a borehole, on the surface, or using a combination of borehole and surface arrays. The data parameters $D(\mathbf{x}, t)$ are samples of the 4-D wavefield. Each model has only one datum because there is one wavefield characterizing one micro-seismic event in space and time.

3.1.2 A priori probability density $\rho(\mathbf{m},\mathbf{d})$

We describe $\rho(\mathbf{m},\mathbf{d})$ by a joint PDF over the model and data spaces. This prior information is independent of any measurement. Furthermore, in the special case when the prior information on the model space $\rho_{\mathfrak{M}}(\mathbf{m})$ is independent of the prior information on the data space $\rho_{\mathfrak{D}}(\mathbf{d})$, the a priori probability density function $\rho(\mathbf{m},\mathbf{d})$ is (Tarantola, 2005)

$$\rho(\mathbf{m},\mathbf{d}) = \rho_{\mathfrak{D}}(\mathbf{d})\rho_{\mathfrak{M}}(\mathbf{m}) \,, \tag{3.4}$$

where $\rho_{\mathfrak{D}}(\mathbf{d})$ describes the measurement uncertainty of the observed data, and $\rho_{\mathfrak{M}}(\mathbf{m})$ illustrates our confidence in the chosen model parameters.

Two examples of a priori probability density functions are shown in Figures 3.1(a) and 3.2(a). The vertical axis represents the value of the probability that a particular model characterizes for the physical system under investigation. In Figure 3.1(a), $\rho(\mathbf{m},\mathbf{d}) = \rho_{\mathfrak{M}}(\mathbf{m})\delta(\mathbf{d} - D^{obs})$, where D^{obs} denote the observed data, and the a priori probability in the model space is a step function. In this case, no error is associated with the observed data which means we are extremely confident in the data we observe. In Figure 3.2(a), $\rho(\mathbf{m},\mathbf{d}) = \rho_{\mathfrak{M}}(\mathbf{m})e^{-\frac{(\mathbf{d}-D^{obs})^2}{2C_d}}$, where $\rho_{\mathfrak{M}}(\mathbf{m})$ is a tapered step function, and C_d is the measurement uncertainty of the data. This equation demonstrates that the higher the value of C_d, the lower the confidence in the observed data. We assume that the data are distributed following a Gaussian function and provide the highest probability for D^{obs}.

For the prior probability $\rho_{\mathfrak{M}}(\mathbf{m})$ in this research, I assume that the hypocentre and the onset time of micro-seismic events may be located with equal probability within a certain space domain and time interval. Thus, I represent this prior knowledge on the model space as a boxcar function in space and time. The prior probability $\rho_{\mathfrak{M}}(\mathbf{m}) = 1$ within this region of space and time and $\rho_{\mathfrak{M}}(\mathbf{m}) = 0$ elsewhere. I uniformly sample the model space within this region. Figure 2.11(b) shows the $\rho_{\mathfrak{M}}(\mathbf{m})$ designed for this example. Here, I assume that the data (the reconstructed wavefield corresponding to the recorded seismograms) characterize all models within the selected model space with equal probability, as illustrated in Figure 3.1(a). I will discuss the design of the prior information on the data space later in this paper.

3.1.3 Homogeneous probability density $\mu(\mathbf{m},\mathbf{d})$

In order to remove the effect of the discretization on the model and data parameters, we define a homogeneous PDF $\mu(\mathbf{m},\mathbf{d})$ whose role is to balance the discrete model and data spaces. By definition, the homogeneous probability distribution states that a probability assigned to each region of the space is proportional to the volume of the region (Mosegaard & Tarantola, 2002). In the special case when the homogeneous probability densities on the model and data spaces are independent, $\mu(\mathbf{m},\mathbf{d})$ is expressed as

$$\mu(\mathbf{m},\mathbf{d}) = \mu_{\mathfrak{D}}(\mathbf{d})\mu_{\mathfrak{M}}(\mathbf{m}) , \tag{3.5}$$

where $\mu_{\mathfrak{M}}(\mathbf{m})$ and $\mu_{\mathfrak{D}}(\mathbf{d})$ are the homogeneous probability density on the model and data space, respectively.

For example, for 3-D Cartesian coordinates, $dV(x,y,z) = dxdydz$ and μ is a constant. According to the experiment, the system is built on the Cartesian coordinates. Thus, for simplicity, I assume $\mu = 1$ in this section.

3.1.4 Theoretical probability density $\Theta(\mathbf{m},\mathbf{d})$

The theoretical probability density $\Theta(\mathbf{m},\mathbf{d})$ describes the relation between the model and data parameters. The defined relationship corresponds to a physical law and may incorporate uncertainties associated with theory, for example due to assumptions or simplifications of the physical law or due to imperfect knowledge of the physical parameters underlying it. The theoretical probability density $\Theta(\mathbf{m},\mathbf{d})$ can be expressed as

$$\Theta(\mathbf{m},\mathbf{d}) = \theta(\mathbf{d}|\mathbf{m})\mu_{\mathfrak{M}}(\mathbf{m}) \,, \tag{3.6}$$

where $\theta(\mathbf{d}|\mathbf{m})$ describes the probability distribution of data for the given kernel function and model and $\mu_{\mathfrak{M}}(\mathbf{m})$ is the marginal probability defined in the preceding section. In Figure 3.1(b), $\theta(\mathbf{d}|\mathbf{m}) = \delta(\mathbf{d} - G(\mathbf{m}))$, where $G(\mathbf{m})$ is the kernel function which describes the physical relationship between $\mathbf{d}$ and $\mathbf{m}$. No error is associated with this kernel function. In Figure 3.2(b), $\theta(\mathbf{d}|\mathbf{m}) = e^{-\frac{(\mathbf{d}-G(\mathbf{m}))^2}{2C_t}}$, where C_t describes the theory uncertainty. The error associated with this kernel function follows a Gaussian distribution.

In this research, the kernel function is based on the two-way acoustic wave-equation and is constructed using the following steps:

- First, for a given model ($\mathbf{m}_d$) characterized by the onset time t_d and location $\mathbf{x}_d$, I propagate waves forward from the source using the velocity $c(\mathbf{x})$ according to the two-way acoustic wave-equation

$$\frac{1}{c^2}\ddot{U} = \nabla^2 U + f(\mathbf{m}_d, t) \,, \tag{3.7}$$

  where $f(\mathbf{m}_d, t)$ is the considered source wavelet and $U(\mathbf{x}, t)$ represents the seismic wavefield simulated forward in time.

- Second, the forward wavefield is recorded at receiver coordinates $\mathbf{x}_0$ arbitrarily located in space:

$$R(\mathbf{x}_0, t) = U(\mathbf{x} = \mathbf{x}_0, t) \,, \tag{3.8}$$

where $R(\mathbf{x_0}, t)$ represents the recorded seismogram.

- Third, the recorded seismogram is time-reversed and propagated backward in time from the acquisition array to reconstruct the wavefield at all locations in space and all times within the window considered for the prior model. This step is also based on the two-way acoustic wave-equation

$$\frac{1}{c^2}\ddot{D} = \nabla^2 D + R(\mathbf{x_0}, -t)\,, \tag{3.9}$$

where $D(\mathbf{x}, t)$ are the data parameters, as explained earlier.

The kernel functions constructed with this procedure are functions of the velocity, the acquisition geometry, and the source type. For a given model, I refer to the data calculated using this procedure centered in the model with the fixed size as the predicted data D^{pre}. In contrast, I refer to the wavefields reconstructed from field seismograms centered in the model with the fixed size as the observed data D^{obs}.

For all models within the selected region of space, I compute the corresponding wavefield using the kernel function. In this case, I assume that each model corresponds to one dataset and that all simulated data have equal probability for explaining the recorded data. This construction, which corresponds to the illustration in Figure 3.1(b), applies the kernel function G linking theoretically individual members of the model and data spaces. As discussed in the previous chapter, it is sufficient to compute data for one onset time and to calculate the database only once as long as we keep the subsurface velocity and acquisition geometry fixed. I assemble all these calculations in a database of wavefields which is used to compare with data reconstructed from the "field" seismograms. Figure 2.16 illustrates the wavefield database at all locations indicated by dots in Figure 2.11(b), and the size of the predicted data in Figure 2.16 is two wavelengths in space and two periods in time.

3.1.5 A posteriori probability density $\sigma(\mathbf{m}, \mathbf{d})$

The conjunction between the a priori state of information $\rho(\mathbf{m}, \mathbf{d})$ and the theoretical probability density function $\Theta(\mathbf{m}, \mathbf{d})$ provides the a posteriori state of information $\sigma(\mathbf{m}, \mathbf{d})$:

$$\sigma(\mathbf{m}, \mathbf{d}) = k\frac{\rho(\mathbf{m}, \mathbf{d})\Theta(\mathbf{m}, \mathbf{d})}{\mu(\mathbf{m}, \mathbf{d})}, \tag{3.10}$$

where k is a normalization constant that serves the purpose of keeping the area under the graph of $\sigma(\mathbf{m},\mathbf{d})$ constant. The expression for k is

$$k = \frac{1}{\int_{\mathfrak{X}} d\mathfrak{X}\rho(\mathbf{m},\mathbf{d})\Theta(\mathbf{m},\mathbf{d})/\mu(\mathbf{m},\mathbf{d})}, \tag{3.11}$$

where $\mathfrak{X} = (\mathfrak{D},\mathfrak{M})$ is the joint space of the data and model.

The posterior information $\sigma(\mathbf{m},\mathbf{d})$ is computed based on observations. In the special case when both the model space and the data space are described in Cartesian coordinates, the posterior information is proportional to the conjunction between the prior information and the theoretical information:

$$\sigma(\mathbf{m},\mathbf{d}) \propto \rho(\mathbf{m},\mathbf{d})\Theta(\mathbf{m},\mathbf{d}) \,. \tag{3.12}$$

Figures 3.1(c) and 3.2(c) schematically show the conjunction between the $\rho(\mathbf{m},\mathbf{d})$ and $\Theta(\mathbf{m},\mathbf{d})$, and the results of the conjunction are shown in Figures 3.1(d) and 3.2(d). In Figure 3.1(d), $\sigma(\mathbf{m},\mathbf{d}) \propto \delta(D^{obs} - G(\mathbf{m}))$ indicates that $\sigma(\mathbf{m},\mathbf{d}) = 1$ only when the observed data are exactly the same as the data calculated from the kernel function. In Figure 3.2(d), $\sigma(\mathbf{m},\mathbf{d}) \propto e^{-\frac{(D^{obs}-G(\mathbf{m}))^2}{2C}}$, where $C = C_d + C_t$. The probability distribution in Figure 3.2(d) shows how well the data calculated from the kernel function explain the observed data.

3.1.6 The marginal probability density $\sigma_{\mathfrak{M}}(\mathbf{m})$

In order to provide a solution to the inversion problem, we have to transfer the information provided by $\sigma(\mathbf{m},\mathbf{d})$ to the model space. We obtain the marginal probability density in the model space $\sigma_{\mathfrak{M}}(\mathbf{m})$ by projecting $\sigma(\mathbf{m},\mathbf{d})$ onto the model space:

$$\sigma_{\mathfrak{M}}(\mathbf{m}) = \int_{\mathfrak{D}} \sigma(\mathbf{m},\mathbf{d}) d\mathbf{d} \,. \tag{3.13}$$

The probability density function provided by $\sigma_{\mathfrak{M}}(\mathbf{m})$ indicates which models simultaneously satisfy our prior knowledge on the distribution of model parameters, the theoretical relationships between model and data, and the measurement uncertainties. A comparison between $\rho_{\mathfrak{M}}(\mathbf{m})$ and $\sigma_{\mathfrak{M}}(\mathbf{m})$ is shown in Figures 3.1(e) and 3.2(e). The distribution of $\sigma_{\mathfrak{M}}(\mathbf{m})$ is narrower on the posterior relative to the prior indicating the tighter connection introduced by the recorded data. In theory, I expect that I will get the same distribution of $\sigma_{\mathfrak{M}}(\mathbf{m})$ as shown in Figure 3.1(e). However, in practice, this

does not hold true due to various uncertainties in reality.

3.1.7 Negligible theoretical uncertainties

When I assume that no uncertainty is associated with the kernel function $G(\mathbf{m})$ which describes the physical relationship between the data and the model, the theoretical probability density $\Theta(\mathbf{m}, \mathbf{d})$ can be expressed as

$$\Theta(\mathbf{m}, \mathbf{d}) = \delta(\mathbf{d} - G(\mathbf{m}))\mu_{\mathfrak{M}}(\mathbf{m}) \,, \tag{3.14}$$

where $\mu_{\mathfrak{M}}(\mathbf{m})$ is a constant for models parameterized in Cartesian coordinates. The theoretical probability density $\Theta(\mathbf{m}, \mathbf{d}) = 1$ when the data equal the value calculated from the kernel function and $\Theta(\mathbf{m}, \mathbf{d}) = 0$ otherwise.

In this case, the marginal probability in the model space in equation 3.13 can be expressed as

$$\sigma_{\mathfrak{M}}(\mathbf{m}) = \int_{\mathfrak{D}} \sigma(\mathbf{m}, \mathbf{d}) d\mathbf{d} = k \int_{\mathfrak{D}} \frac{\rho(\mathbf{m}, \mathbf{d})\Theta(\mathbf{m}, \mathbf{d})}{\mu(\mathbf{m}, \mathbf{d})} d\mathbf{d} \,; \tag{3.15}$$

when $\Theta(\mathbf{m}, \mathbf{d})$ is given in equation 3.14, with $\rho(\mathbf{m}, \mathbf{d}) = \rho_{\mathfrak{D}}(\mathbf{d})\rho_{\mathfrak{M}}(\mathbf{m})$ and $\mu(\mathbf{m}, \mathbf{d}) = \mu_{\mathfrak{D}}(\mathbf{d})\mu_{\mathfrak{M}}(\mathbf{m})$, we can write

$$\sigma_{\mathfrak{M}}(\mathbf{m}) = k\rho_{\mathfrak{M}}(\mathbf{m}) \int_{\mathfrak{D}} \frac{\rho_{\mathfrak{D}}(\mathbf{d})\delta(\mathbf{d} - G(\mathbf{m}))}{\mu_{\mathfrak{D}}(\mathbf{d})} d\mathbf{d} \,, \tag{3.16}$$

where $\rho_{\mathfrak{D}}(\mathbf{d})$ is a function of $\mathbf{d}$ and the observed data, and $\mu_{\mathfrak{D}}(\mathbf{d})$ is a constant for data discretized in Cartesian coordinates. Thus, equation 3.13 for an ideal kernel function can be written as

$$\sigma_{\mathfrak{M}}(\mathbf{m}) = \nu\rho_{\mathfrak{M}}(\mathbf{m})\rho_{\mathcal{D}}(G(\mathbf{m}), D^{obs}) \,, \tag{3.17}$$

where

$$\nu = \int_{\mathfrak{M}} \rho_{\mathfrak{M}}(\mathbf{m})\rho_{\mathcal{D}}(\mathbf{d}, D^{obs}) d\mathbf{m} \tag{3.18}$$

is a constant. In equation 3.17, $\sigma_{\mathfrak{M}}(\mathbf{m})$ is proportional to the product of the prior probability on the model and data spaces.

In this research, I assume that there is no error associated with the theoretical information. Therefore, errors of $\sigma_{\mathfrak{M}}(\mathbf{m})$ are introduced by the measurement uncertainty of the observed data.

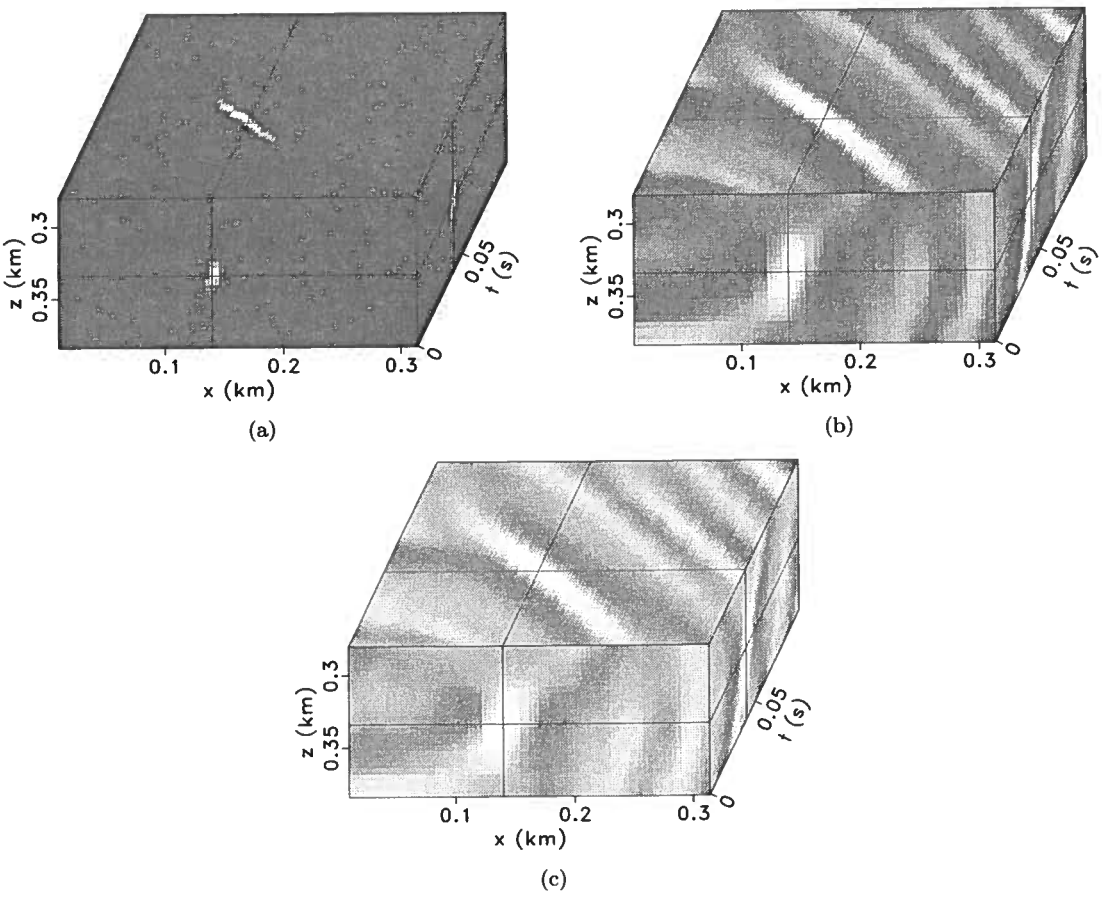


Figure 3.3. Probabilistic distributions of normalized cross-correlation. The measurement uncertainties are (a) $C_d = 0.1$, (b) $C_d = 0.49$, and (c)$C_d = 1$. The distribution of $\sigma_{\mathfrak{M}}(\mathbf{m})$ becomes flatter when measurement uncertainty increases.

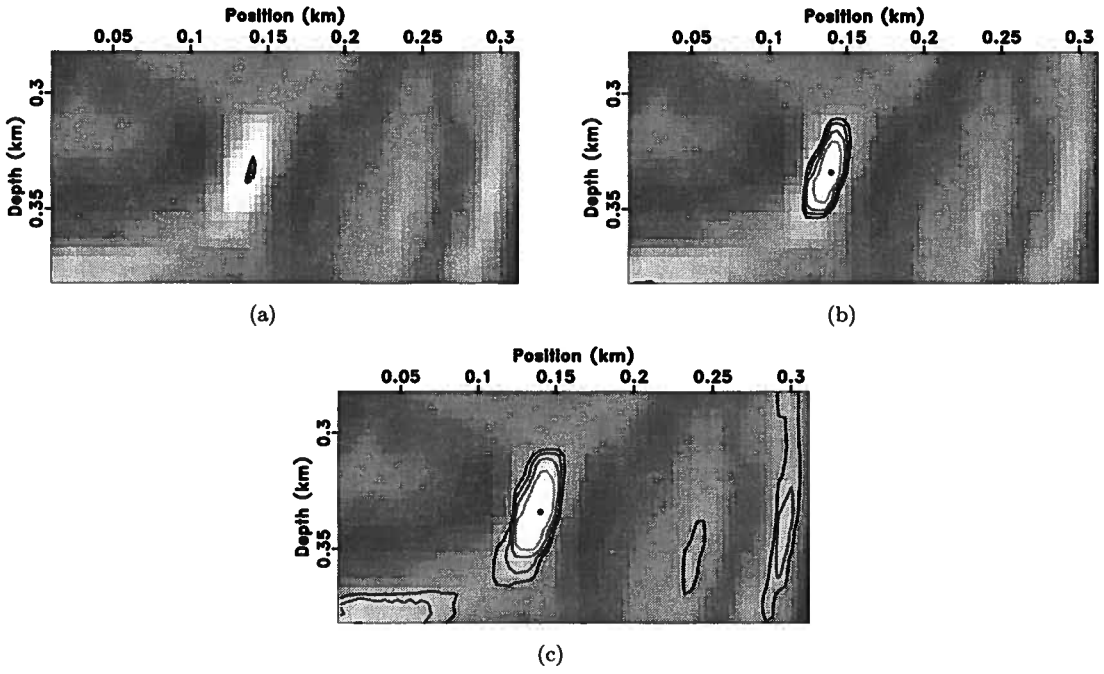


Figure 3.4. Snapshots of the probability density function where dots are the hypocentre of the micro-earthquake. The outer contour has a probability of 0.8 and the contour interval is 0.05. The measurement uncertainties of these three plots are (a) $C_d = 0.1$, (b) $C_d = 0.49$, and (c) $C_d = 1$.

3.1.8 The prior information on the data

As discussed earlier, I refer to data as windowed reconstructed wavefields at all locations in space and within a given time window. I loop over the models provided by $\rho_{\mathfrak{M}}(\mathbf{m})$ and look for the times and locations in space where data reconstructed from recorded seismograms match the simulated data at locations where I assume that micro-seismic events are located. We can conceptually describe the process as measuring the similarity between multi-dimensional images. In the preceding chapter, I discussed various methods that can be used to compare the similarity between the predicted data and observed data. The preferred method of normalized cross-correlation (N_{cc}) measures the similarity between theoretical wavefields and wavefields reconstructed from field seismograms. The values of N_{cc} are between -1 and $+1$.

For simplicity, I assume that the error of the observed data has a Gaussian distribution, and I give the highest probability to the model which generates the smallest data misfit. The probability density function based on the normalized cross-correlation in equation 2.5 is

$$\sigma_{\mathfrak{M}}(\mathbf{m}) = \rho_{\mathfrak{M}}(\mathbf{m}) e^{-\frac{(1-N_{cc}(\mathbf{m}))^2}{2C_d}}, \tag{3.19}$$

where C_d is the measurement uncertainty and $N_{cc}(\mathbf{m})$ is the normalized cross-correlation function. $N_{cc}(\mathbf{m})$ can be expressed as

$$N_{cc}(\mathbf{m}) = \frac{\sum_{\mathbf{x}-M_{\mathbf{x}}}^{\mathbf{x}+M_{\mathbf{x}}} \sum_{t-M_t}^{t+M_t} (D^{obs}(\mathbf{m}) - \overline{D^{obs}(\mathbf{m})})(G(\mathbf{m}) - \overline{G(\mathbf{m})})}{\sqrt{\sum_{\mathbf{x}-M_{\mathbf{x}}}^{\mathbf{x}+M_{\mathbf{x}}} \sum_{t-M_t}^{t+M_t} (D^{obs}(\mathbf{m}) - \overline{D^{obs}(\mathbf{m})})^2 \sum_{\mathbf{x}-M_{\mathbf{x}}}^{\mathbf{x}+M_{\mathbf{x}}} \sum_{k=t-M_t}^{t+M_t} (G(\mathbf{m}) - \overline{G(\mathbf{m})})^2}}, \tag{3.20}$$

which is a function of the predicted data and observed data, where the overline denotes average values.

In equation 3.19, the larger the value of C_d, the flatter the distribution of $\sigma_{\mathfrak{M}}(\mathbf{m})$. When C_d is small, $\sigma_{\mathfrak{M}}(\mathbf{m})$ tends toward a delta function. Three probability distributions of normalized cross-correlation are shown in Figure 3.3 with different measurement uncertainties. After examining these three figures, I find that the distribution of probability follows the role I discussed previously. Figure 3.4 explicitly shows this property. Snapshots of Figure 3.3 are picked from Figure 3.4 at the time of the maximum probability value.

I expect that for the distribution of $\sigma_{\mathfrak{M}}(\mathbf{m})$, the radius of the contour with a value of 0.9 will be half the wavelength because the location error of the N_{cc} method is less than half of the wavelength. Thus, I choose $C_d = 0.49$ because the radius of the contour (shown in Figure 3.4(b)) with a value of 0.9 is about 20 m. I compute normalized cross-correlation by comparing the pre-calculated wavefield database with the observed reconstructed wavefield, as shown in Figure 2.20. The maximum N_{cc} value is 0.991 at $x = 132$ m, $z = 334$ m, $t = 0.044$ s. However, the N_{cc} value is 0.985 at $x = 136$ m, $z = 334$ m, $t = 0.042$ s. These two values are very close and it would be unwise to pick one without taking the other into consideration. Therefore, instead of selecting a single model as the answer, I construct a probability distribution characterizing various models and their likelihood to represent sources of micro-earthquakes.

3.2 Synthetic example with multiple micro-earthquakes

In the preceding section, I present micro-earthquake location in the case when only one seismic event occurs in a region during a time interval. However, a more common situation is that multiple micro-seismic events occur more or less at the same time and overlap in space during a given time interval. In this section, I demonstrate the methodology on another example with multiple overlapping events.

3.2.1 Distribution of micro-earthquakes

The motivation for such an experiment derives from field observations about how fractures propagate during fluid injection. By spatially mapping the signals emitted by micro-earthquakes, we can monitor the extensions of fractures (Keppler *et al.*, 1982; Shapiro *et al.*, 2006). Numerous field observations indicate that the seismic zone surrounds the fractures and forms an elongated pattern (House, 1987; Phillips, 1998; Rutledge & Phillips, 2001). Thus, the fluid-induced microseismicities are concentrated in a spatial domain close to the hydraulic fracture (Shapiro *et al.*, 2006).

In this example, I collected a variety of information about micro-seismic events from the same injection experiment. Based on this information, I assume that the micro-seismic events are distributed along a fracture which is located at $z = 0.326$ km. The opening of this fracture, which is along the x direction where x increases, is caused by high pressure steam injected into the borehole located at $x = 0$ km. In the case of a straight, planar, height-fixed fracture, the half-length L of the

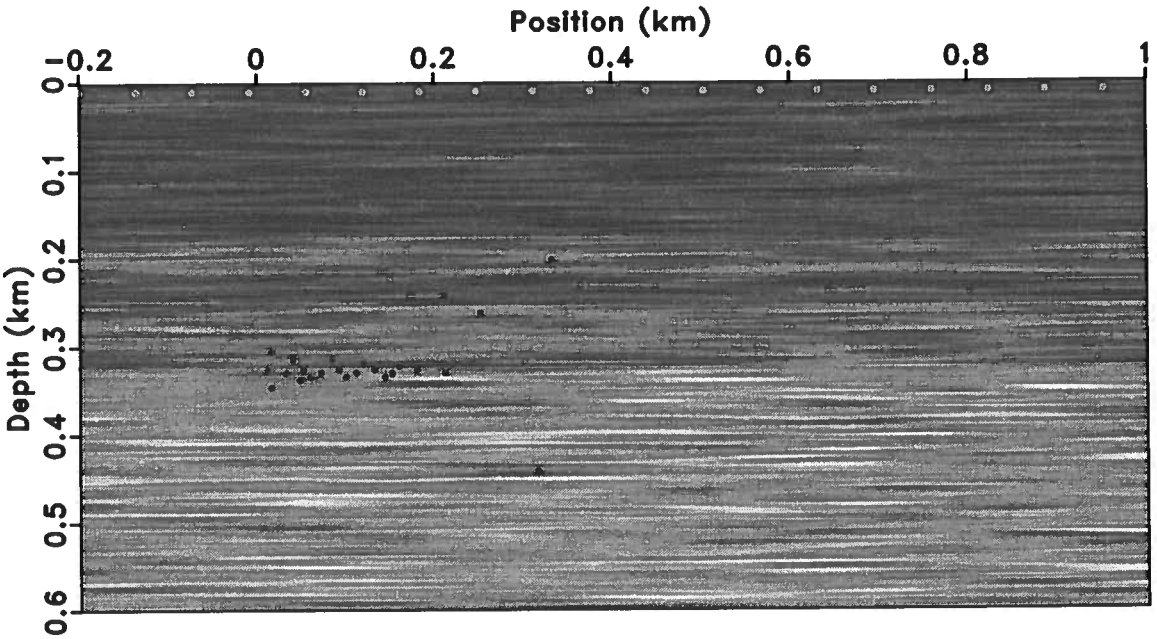


Figure 3.5. Synthetic model. The dots located at about $z = 0$ km represent receivers; the other dots represent the micro-seismic events.

fracture is given as a function of the injection time t by

$$L(t) = \frac{t}{c_3 + c_4\sqrt{t}} \tag{3.21}$$

with constants $c_3 = 2hw/q$ and $c_4 = 4\sqrt{2}hC/q$, where h is the average height of the fracture, q is the average injection rate of the treatment fluid, C is the leak-off coefficient, w is the fracture width which is a constant, c_3 describes the fluid loss from the fracture into the rock, and c_4 describes the contribution of the effective fracture volume and depends mainly on the geometry of the fracture vertical cross-section (Economides, 1992; Fischer *et al.*, 2008). By design, I assume that the growth of the fracture has a linear relationship with the injection time and that the microseismicities are triggered around the tips of the opening fracture. Therefore, the triggering time of one micro-seismic event has a linear relationship with the distance between the microseismic location and the injection point. The difference between the onset times of two nearby micro-seismic events is greater than one period.

I also assume that the magnitudes of all micro-earthquakes are different and the sources are

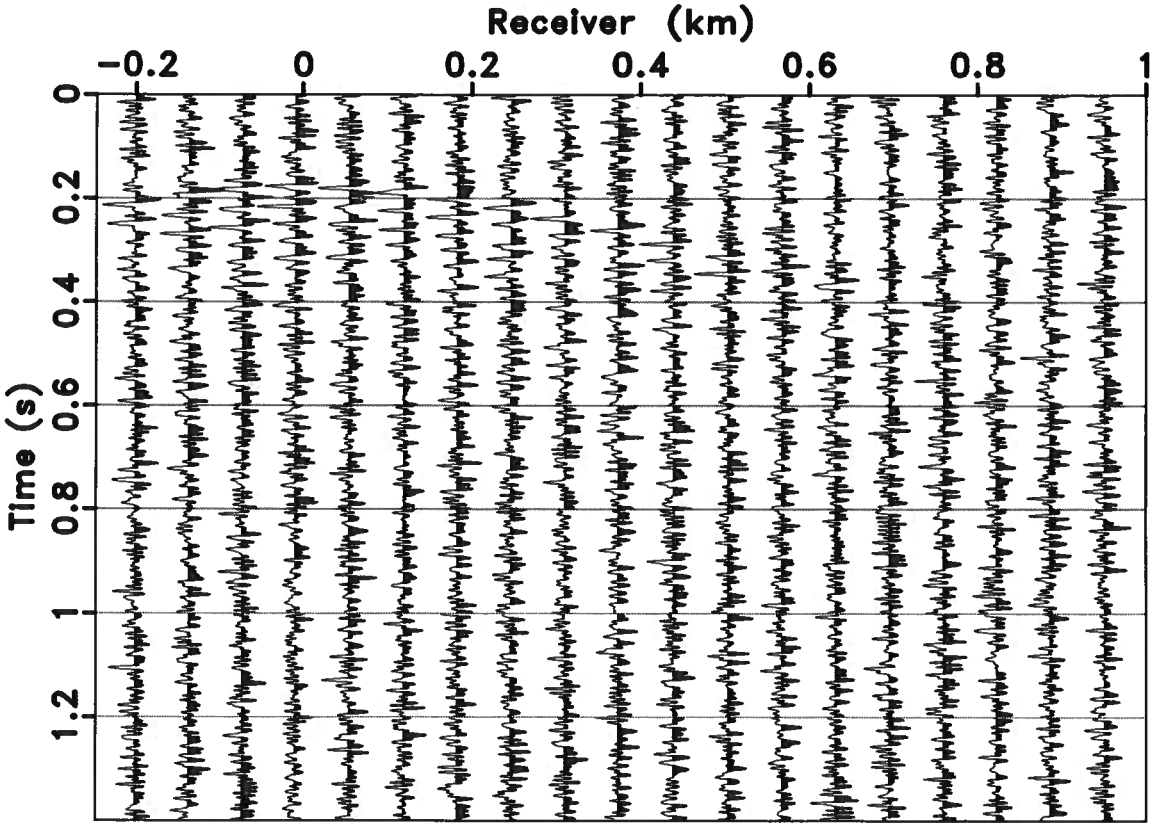


Figure 3.6. Synthetic seismogram. A total of 7000 time steps with an increment of 0.2 ms are calculated.

characterized by Ricker wavelets with a dominant frequency of 50 Hz. Figure 3.6 shows the simulated seismogram characterizing all the events. The seismogram is contaminated with two types of noise. The first type of noise is due to model heterogeneities which I assume that I cannot identify in the imaging process. The second type of noise corresponds to the fact that the experiment is run in a noisy environment. The frequency band of the random noise is $0 - 140$ Hz overlapping with the micro-earthquake band. The SNR for some events in this example is as low as 0.1. I assume that measurement uncertainties are 0.64 because of the low SNR. Since the larger the size of the predicted data, the higher the identification of a single micro-earthquake and the lower the resolution of two nearby micro-earthquakes; thus, in order to achieve high resolution and maintain high identification, I select the predicted data with the size of two wavelengths in space and two periods in time. I simulate 20 micro-earthquakes and attempt to locate all of them with my methodology.

3.2.2 Analyzing probability density function of micro-earthquakes

After applying the method to the seismogram shown in Figure 3.6, I pick 20 snapshots (shown in Figures 3.7, 3.8, and 3.9) from the probability distribution. The contours shown indicate the probability of earthquake occurrence at various times and range from 0.8 to 0.95 with a contour interval of 0.05. One snapshot corresponds to a single micro-earthquake I located.

Figure 3.7(a) shows the source location obtained for a micro-earthquake occurring at $\mathbf{m}_0 = \{x, z, t\} = \{0.012$ km, 0.326 km, 0.02 s$\}$. This point is located within a contour with a confidence level of 0.95. Figure 3.7(b) shows that the event $\mathbf{m}_1 = \{0.016$ m, 0.306 m, 0.10 s$\}$ is located within the contour with a value of 0.85. I assume that micro-earthquakes can be triggered within an area with a probability value of 0.85 or greater due to the low SNR of this example. Figure 3.7(c) shows the event $\mathbf{m}_2 = \{0.016$ m, 0.346 m, 0.05 s$\}$ located within the contour with a value of 0.9. I locate this event accurately. However, this plot also shows another area with a value of 0.85. This may be caused by artifacts of the reconstructed wavefield. Figure 3.7(d) shows an area characterized by the probability value of 0.85. This area contains the event $\mathbf{m}_3 = \{0.032$ m, 0.33 m, 0.12 s$\}$.

Figure 3.7(e) shows two events, $\mathbf{m}_4 = \{0.04$ km, 0.314 km, 0.16 s$\}$ and $\mathbf{m}_5 = \{0.052$ km, 0.326 km, 0.24 s$\}$, located within the contour with a confidence level of 0.9. This example shows that the procedure may not have the resolution to distinguish among events that are located too close to one another. Event $\mathbf{m}_5$ is also located by contours shown in Figure 3.7(f). Other choices of parameters, e.g. in the definition of the wavefield database, might increase the spatial resolution of

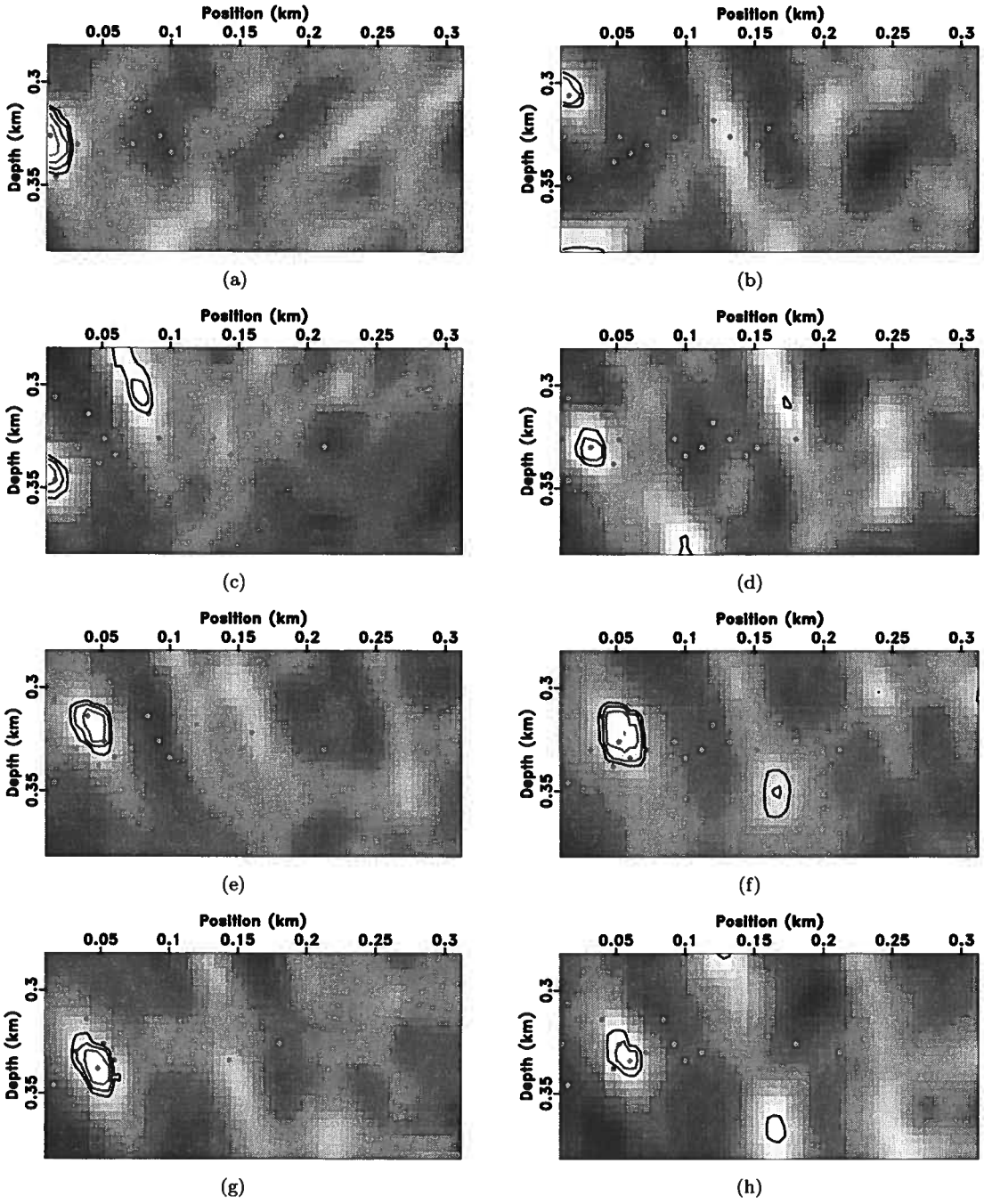


Figure 3.7. Snapshots of the probability distribution of micro-seismic events. The value of the outer contour is 0.8 and the contour interval is 0.05. The dots represent the locations of micro-earthquakes.

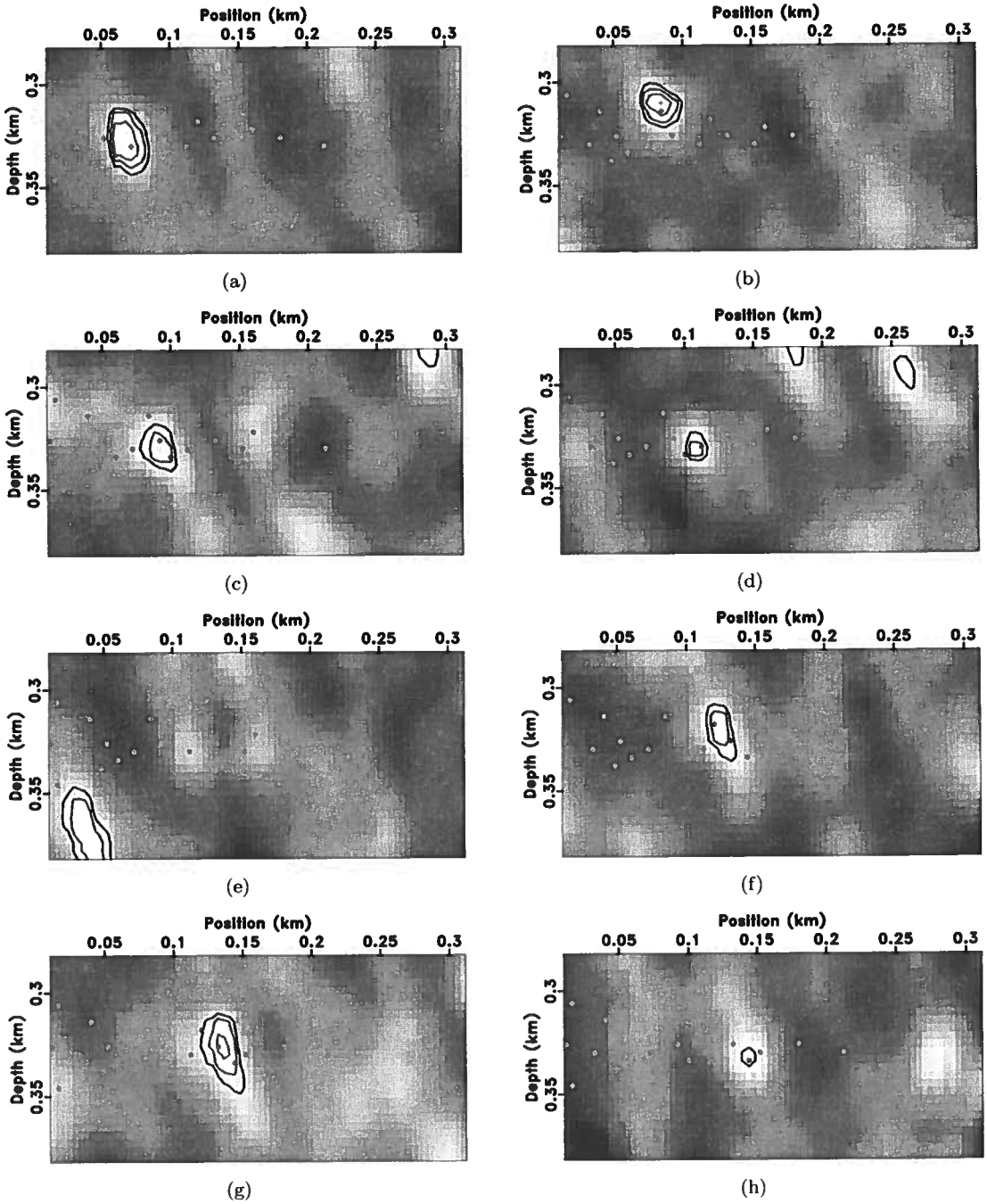


Figure 3.8. Snapshots of the probability distribution of micro-seismic events. The value of the outer contour is 0.8 and the contour interval is 0.05. The dots represent the locations of micro-earthquakes.

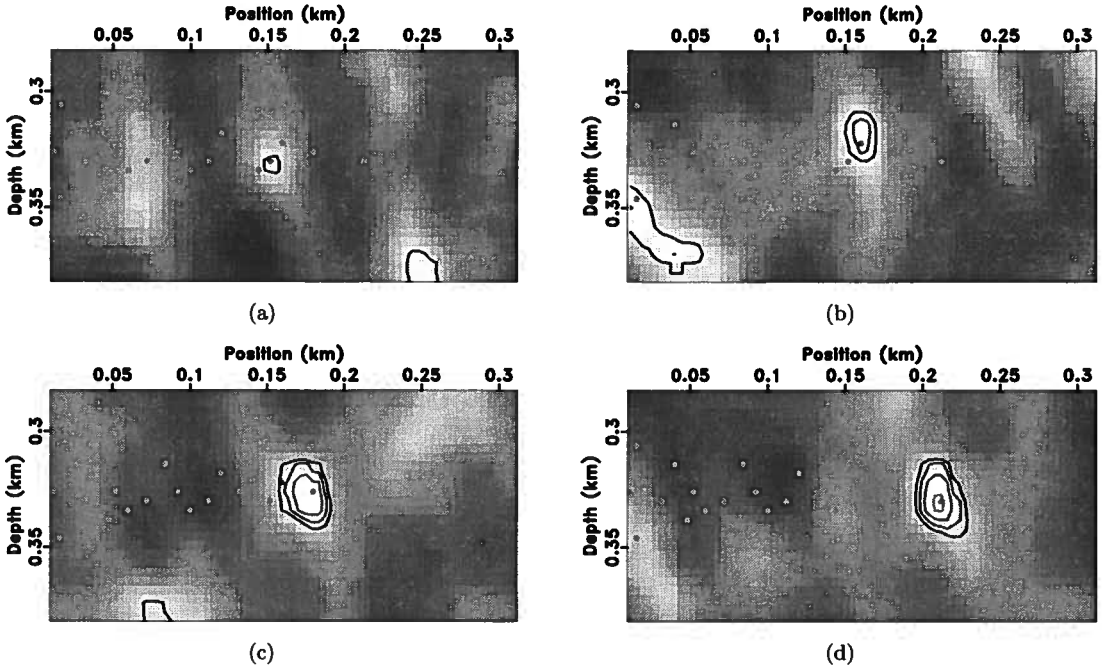


Figure 3.9. Snapshots of the probability distribution of micro-seismic events. The value of the outer contour is 0.8 and the contour interval is 0.05. The dots represent the locations of micro-earthquakes.

the identification. This example indicates the presence of another event located inside the contour centered at about $\{x = 0.17$ km, $z = 0.35$ km$\}$. This is simply an artifact since no event occurs at this position and time in the simulation. Figure 3.7(g) indicates the event $\mathbf{m}_6 = \{0.048$ km, 0.338 km, 0.20 s$\}$ located within the contour with a value of 0.9. The spatial position of the event $\mathbf{m}_6$ is also located within the contour with a value of 0.85. In Figure 3.7(h), two events $\mathbf{m}_5$ and $\mathbf{m}_7$ are included within the contour with a value of 0.85. Figures 3.7(g) and 3.7(h) further demonstrate the low spatial resolution of identification.

Figure 3.8(a) shows the event $\mathbf{m}_8 = \{0.072$ km, 0.33 km, 0.32 s$\}$ located within the contour with a confidence value of 0.9. In Figure 3.8(b), event $\mathbf{m}_9 = \{0.084$ km, 0.314 km, 0.38 s$\}$ is located within the contour with a confidence value of 0.9. Figure 3.8(c) shows event $\mathbf{m}_{10} = \{0.092$ km, 0.326 km, 0.42 s$\}$ located within the contour with a probability value of 0.85. The spatial location of event $\mathbf{m}_{11} = \{0.1$ km, 0.334 km, 0.46 s$\}$ is also located within this contour. However, in Figure 3.8(d), the event $\mathbf{m}_{11}$ that I try to identify is not located within the contour with the value of 0.8. Instead, the spatial location of event $\mathbf{m}_{12} = \{0.112$ km, 0.33 km, 0.52 s$\}$ is located on the contour line with

the value of 0.85.

Figure 3.8(e) shows that the event at $\mathbf{m}_{12}$ is not located within the contours obtained by the inversion. In this case, the method fails. This phenomenon is likely due to the fact that multiple micro-earthquakes overlap in this region creating false foci that masquerade as real events. The event at $\mathbf{m}_{13} = \{0.12$ km, 0.318 km, 0.56 s$\}$, shown in Figure 3.8(f), is located within the contour with a value of 0.85. Figure 3.8(g) shows the event $\mathbf{m}_{14} = \{0.132$ km, 0.326 km, 0.62 s$\}$ located within the contour with a confidence value of 0.9. Figure 3.8(h) shows the event at $\mathbf{m}_{15} = \{0.144$ km, 0.334 km, 0.68 s$\}$ located within the contour with a confidence value of 0.8. Although I successfully identify this event and the experiment seems to place it at the right position, I have lower confidence in its position due to the low SNR in the seismogram, which is approximately 0.1.

Similar to the situation shown in Figure 3.8(h), Figure 3.9(a) shows the event $\mathbf{m}_{16} = \{0.152$ km, 0.33 km, 0.72 s$\}$ located within the contour with a confidence value of 0.8. Figure 3.9(b) shows the event $\mathbf{m}_{17} = \{0.16$ km, 0.322 km, 0.76 s$\}$ located within the contour with a confidence value of 0.85. Figure 3.9(c) shows the event $\mathbf{m}_{18} = \{0.18$ km, 0.326 km, 0.86 s$\}$ located within the contour with a confidence value of 0.9. Finally, Figure 3.9(d) shows the event at $\mathbf{m}_{19} = \{0.212$ km, 0.33 km, 1.02 s$\}$ located well within the contour with a confidence value of 0.95. These last three figures show examples of successful identification of a seismic source.

Although in this example I do not locate all existing micro-earthquakes accurately, I regard this experiment as a success because we are getting a good representation of the direction and speed of fracture propagation in a difficult setting characterized by low SNR and sparse data acquisition.

Chapter 4

Field data analysis

In this chapter, I demonstrate the technique with field datasets corresponding to string shots and perf shots lowered in a borehole. In reality, the magnitude of a micro-earthquake is too small to be identified by using travel-time based methods. However, my method reveals its robustness in low SNR situations and demonstrates its feasibility with the field data.

This chapter is composed of three parts: first, I locate the source of perf shots based on the 2-D velocity model; second, I locate the source of string shots based on the same model; third, I relocate the source of string shots based on the 3-D velocity model.

The perf and string shots shown here are two explosive apparatuses which are used to open a borehole. The strength of these two shots is under control. The controlled sources allow me to quantitatively analyze the method. When one shot is triggered, its signal is recorded by the receivers which are spread on the surface. The layout of receivers (shown in Figure 4.1) for both perf and string shots are the same. The dark blue dots in this figure represent the receivers; the green dot represents the wellhead. The origin of the receiver coordinate is the location of the wellhead which is {0 m,0 m}.

For the first two sections of this chapter, I pick a line from these receivers and calculate the source locations. The line I pick is the one closest to the location of the explosive source projected on the surface. The geometry of these experiments is shown in Figure 4.2, where green dots represent receivers.

4.1 Perf shots

I test my method using the data acquired along a line which is close to the borehole. Part of the field data of the two perf shots is shown in Figure 4.3. There are 1000 time samples, the

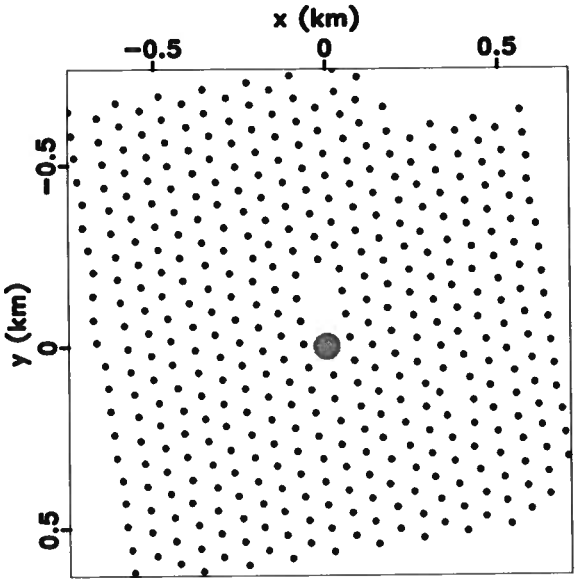


Figure 4.1. Receiver layout where the red dot represents the wellhead and the green dots represent the receivers. The distribution of receivers is close to uniform sampling.

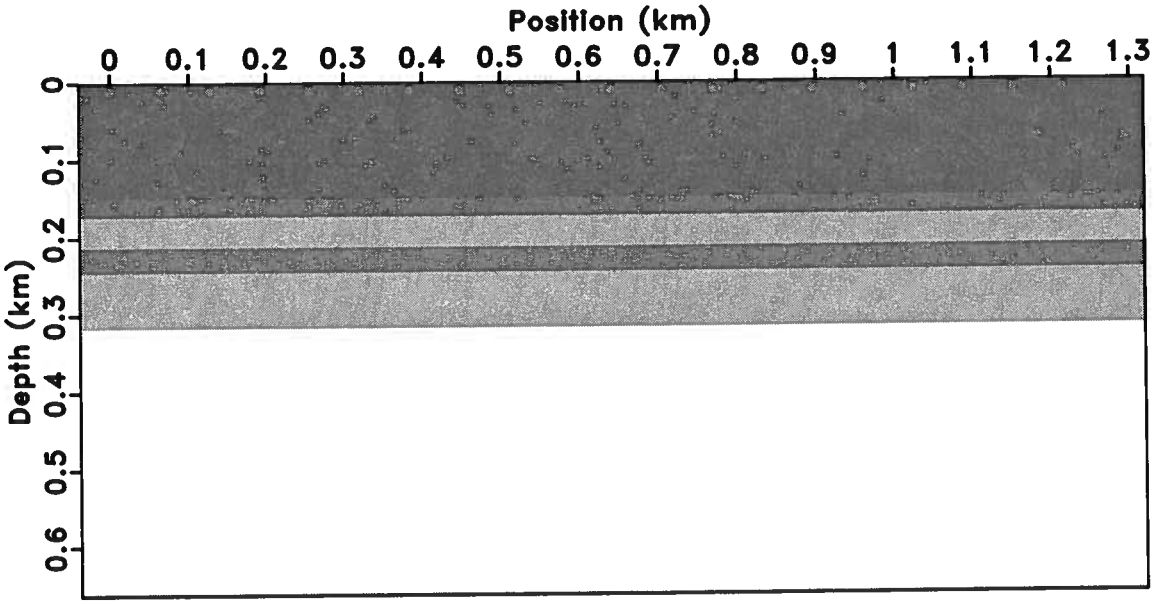


Figure 4.2. Velocity model and receiver layout.

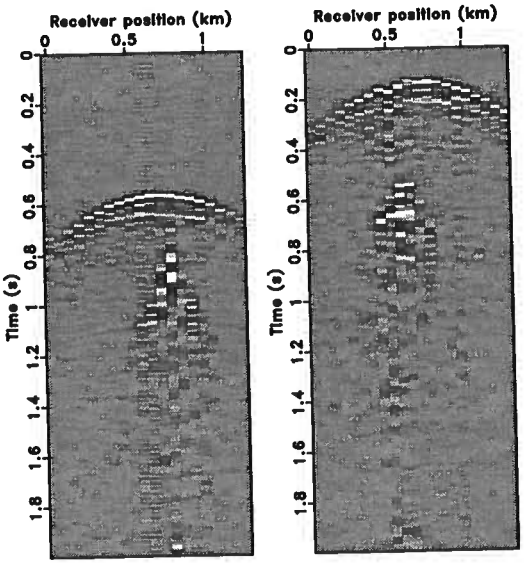


Figure 4.3. (a) and (b) are part of the field data of two perf shots where (a) records perf I shot and (b) records perf II shot.

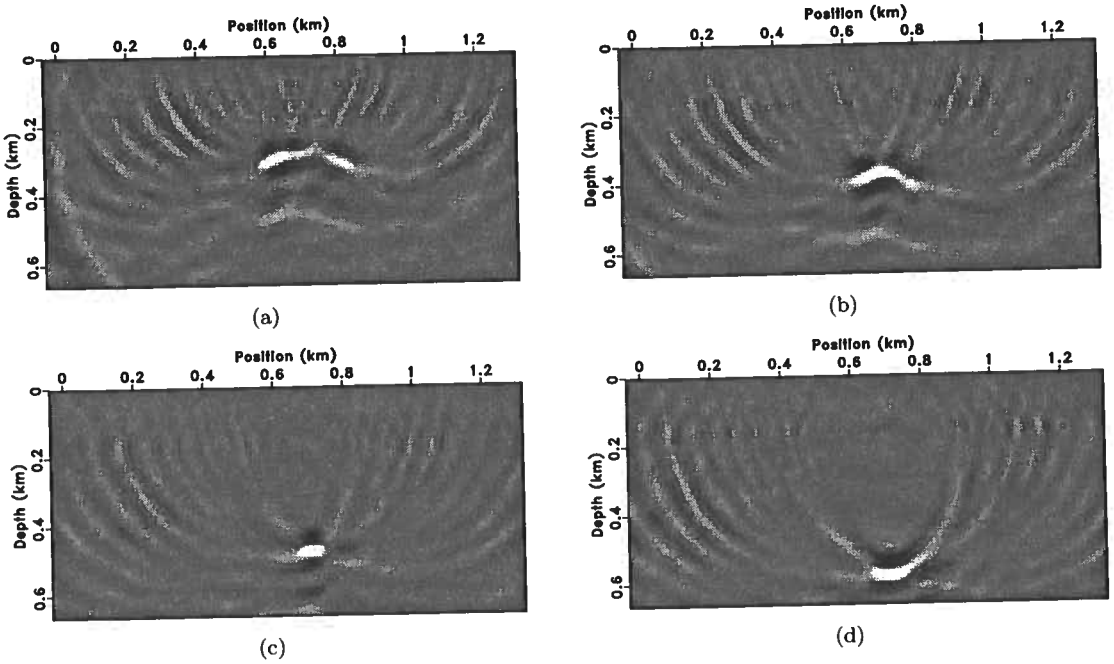


Figure 4.4. Four snapshots of the wavefield reconstructed from perf I. (c) shows the focus of the reconstructed wavefield.

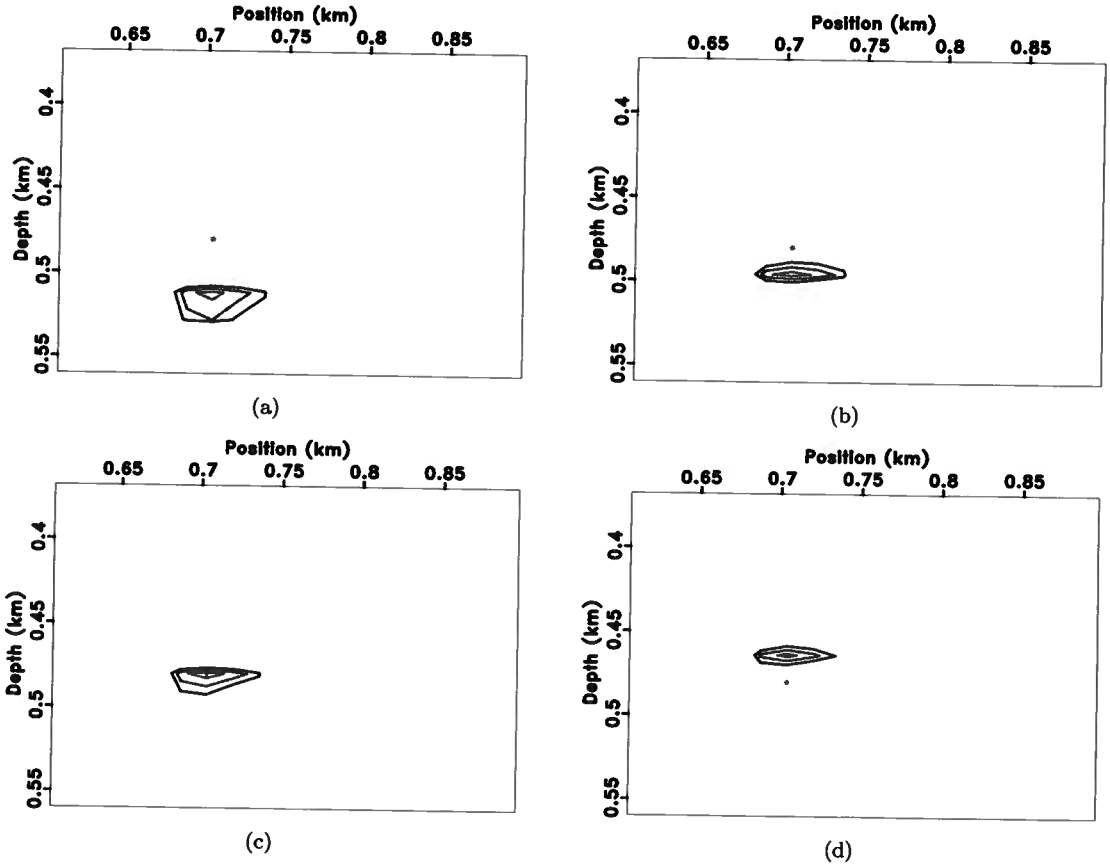


Figure 4.5. Four snapshots of the probability map of perf I where the green dots represent the detected source location, and the probability value of the inner contour is 0.95.

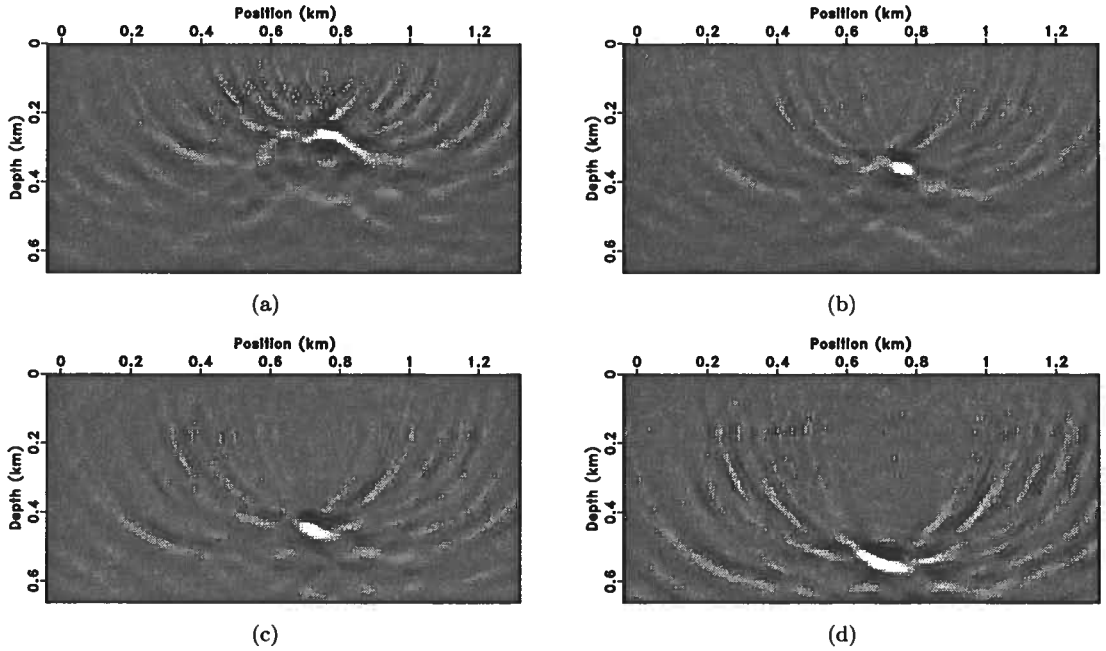


Figure 4.6. Four snapshots of the wavefield reconstructed from perf II. (b) shows the focus of the reconstructed wavefield.

time sampling is 2 ms, and the dominant frequency of the signals is 30 Hz. The magnitude of the explosive shots is strong, and we can identify the P-wave first arrival times. The calculation is done on a numerical grid of 171 × 84 nodes and the grid spacing is 8 m.

I begin my method by reconstructing the wavefields of the two seismograms. Four snapshots corresponding to the perf I shot are shown in Figure 4.4. Because of the high SNR, we can easily identify the position and time where the reconstructed wavefield best refocuses (Figure 4.4(c)). The a priori state of information on the model space in this experiment is in the vicinity where the reconstructed wavefield focuses. The space sampling of the possible source locations is 16 m. Applying the method, I calculate the probability. Figure 4.5 shows four snapshots of the probability, where green dots indicate the location with the highest probability value. The probability value of the outer contour is 0.85 and the contour increment is 0.05. The source location I detected by using the method is {x,z} = {0.7 km, 0.48 km}, which is characterized by a high probability value.

The same processes are applied to the perf II seismogram which is shown in Figure 4.3. After observing the reconstructed wavefield shown in Figure 4.6, we see that the best focus location is

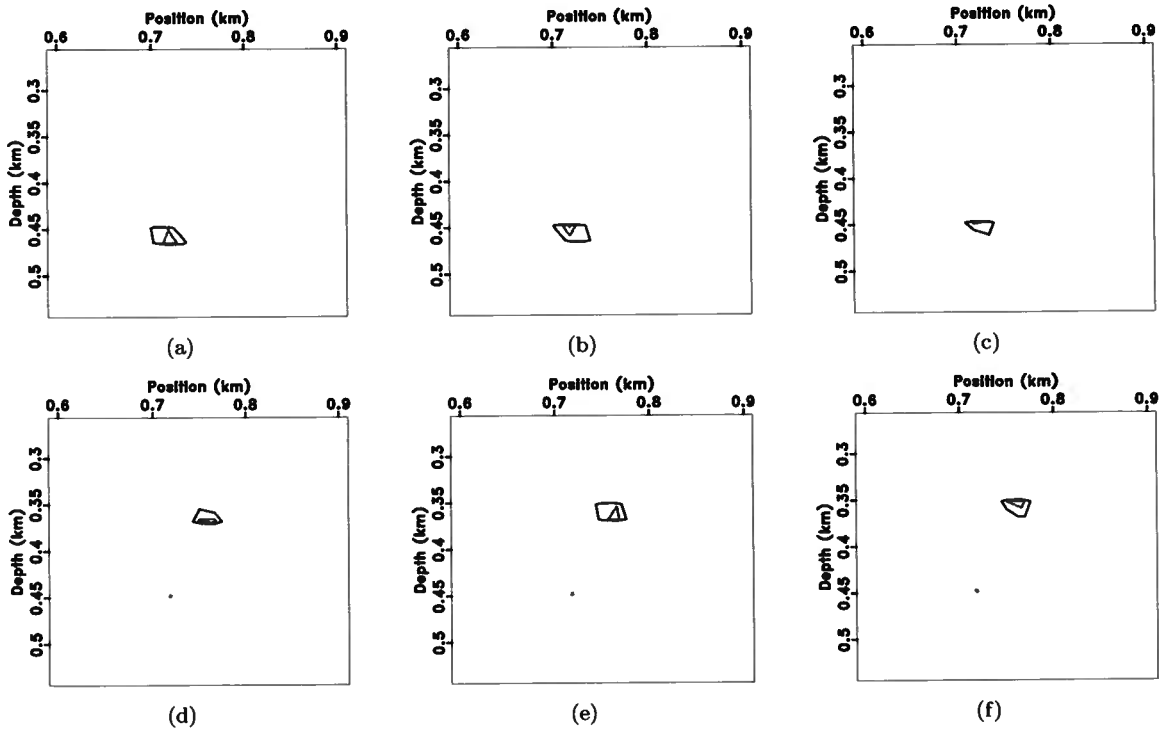


Figure 4.7. Six snapshots of the probability map of perf II where the green dots represent the detected source location, and the probability value of the inner contour is 0.9.

Table 4.1. A comparison between the true and detected source locations. The top lines of perf I and II show the true source locations. The bottom lines show the detected source locations.

	East(m)	North(m)	Depth(m)
perf I	50.94	-31.28	455.00
	46.1	-151.65	480.00
perf II	-155.3	-174.11	455.43
	-153.3	-218.07	368.00

around {0.8 km, 0.3 km}. However, after calculating the probability, I discovered two areas where probabilities are greater than 0.9. The first area is shown by snapshots in Figures 4.7(a) - 4.7(c); the second area is shown by snapshots in Figures 4.7(d) - 4.7(f). The green dots represent the location with the maximum probability value. In this experiment, lacking other sources of information, I am unsure of the best locations to pick for the perf shot. However, by combining the probability with the reconstructed wavefield, I pick the source location which corresponds to the location of the best focus of the reconstructed wavefield. The explosive source location picked is {0.75 km, 0.368 km}.

When I convert the source locations into a common 3-D coordinate framework as shown in Figure 4.1, the source location provided by the perf I seismogram is {East, North, Depth} = {46.10 m, −151.65 m, 480 m} and the source location provided by the perf II seismogram is {−153.32 m, −218.07 m, 368 m}. A comparison of the detected sources with the true source locations, which is shown in Table 4.1, shows that there is a significant positioning difference in the north direction. This is due to the 2D nature of the inversion which does not take into account the fact that the sources might be located off-line.

4.2 String shots

In this section, I demonstrate the technique with several field datasets corresponding to string shots lowered in a borehole. This experiment is meant to demonstrate the robustness of the method with low SNR.

The greatest challenge in this experiment is the fact that these string seismograms may record more than just the string shot. The seismogram may record many micro-earthquakes which are triggered by the string shot. One possible explanation for the mechanism of micro-earthquakes recorded by this seismogram is that when the modified stress and strain fields, caused by the string

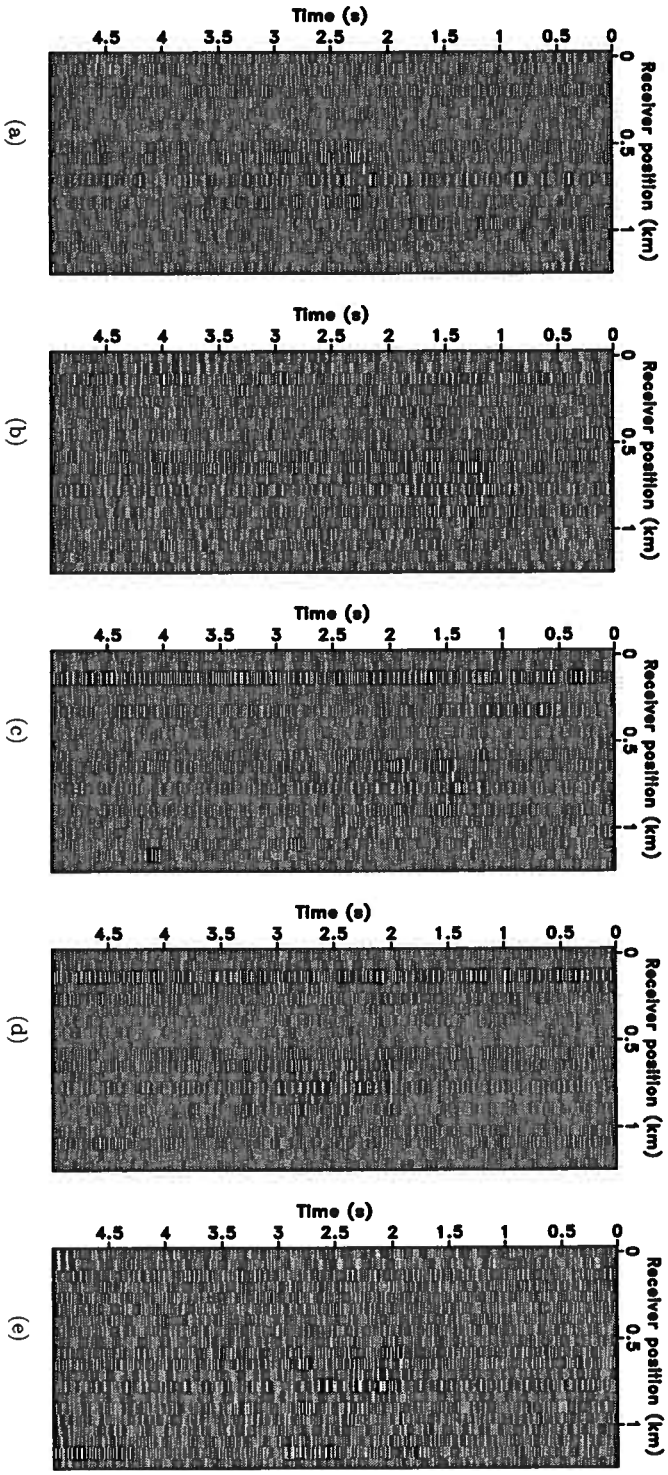


Figure 4.8. (a) - (e) are part of the seismograms of the five string shots where (a) - (e) records string shots I - V, respectively. The S/N ratio decreases from (a) - (e). It is easy to identify the P-wave from (a).

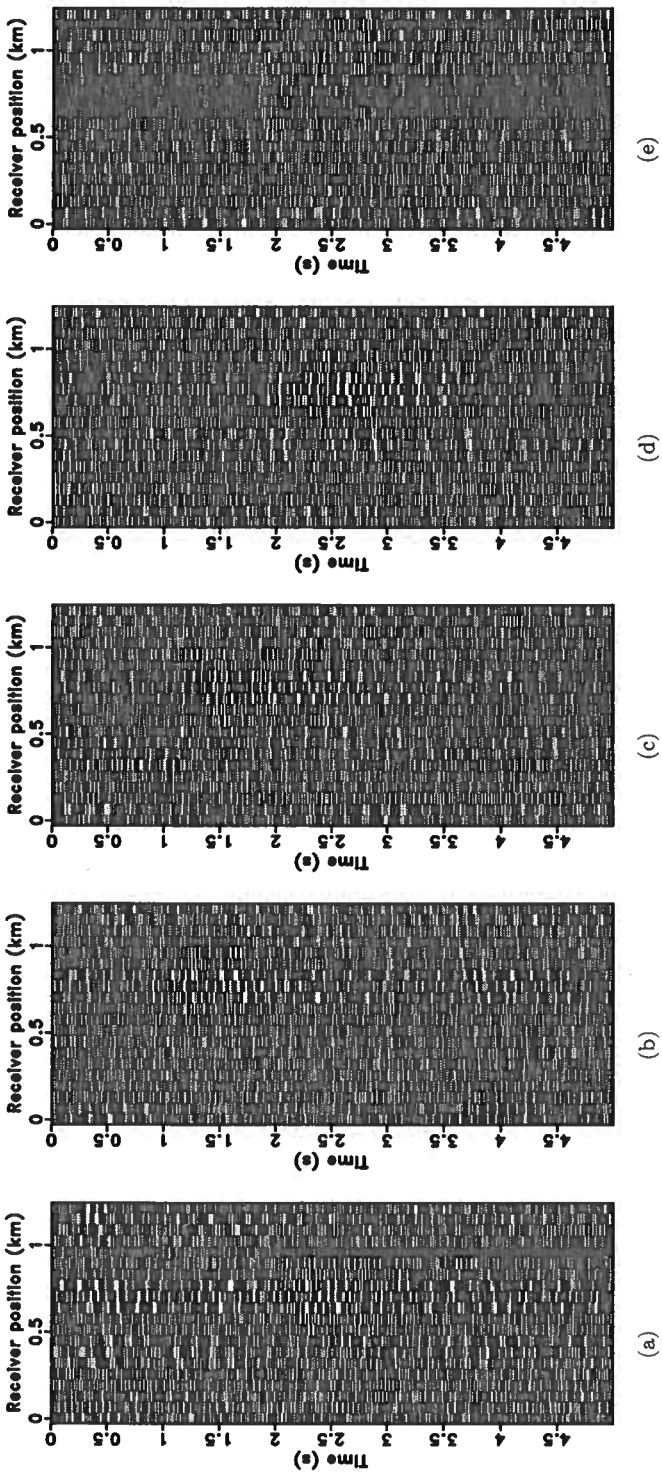


Figure 4.9. Seismograms from Figure 4.8 after amplitude balancing has been applied. (a) - (e) show the results of seismograms of string shots I - V, respectively. The S/N decreases from (a) - (e).

shot, exceed their criteria, micro-earthquakes can be triggered. Therefore, my method may provide both the source location of the string shot and source locations of natural micro-earthquakes.

My task is to identify the source of the string shot among all the sources I have located since I can only verify my method by comparing the detected with the known source location. In order to separate the string shot from micro-earthquakes, other kinds of information are required. For example, I can employ the property of the magnitude of string shots because this magnitude is greater than that of micro-earthquakes. One deficiency of my method, however, is that it cannot detect the magnitude of a micro-seismicity since the method measures the similarity between two reconstructed wavefields. I can solve this problem by combining the probability density function with the reconstructed wavefield as discussed in the previous section. By checking the locations provided by the probability map in the reconstructed wavefield to see whether there is a refocused field, I can select the proper location for the string shot.

I test the method using data recorded along another line which is also close to the head of the borehole for all string shot seismograms. The respective seismograms are shown in Figure 4.8. For all seismograms, there are 5000 time samples, the time sampling interval is 1 ms, and the dominant frequency of the signal is about 30 Hz. In Figure 4.8, the hyperbolas corresponding to the string shots are not obvious. However, for these five seismograms, some strong signals are located in time between $1 - 1.5$ s from trace 9 to 14. This kind of signal may be caused by the Stonely wave that propagates along the borehole and transforms into the surface wave near the top of the surface. Although the magnitude of this kind of wave is strong, it does not affect the location accuracy because there is no focus produced by this kind of signal. After I apply amplitude balancing, which normalizes each trace of the seismogram by its root-mean-value, to the recorded seismogram, I use the seismograms (shown in Figure 4.9) to calculate the reconstructed wavefields.

The a priori information on the model space is shown in Figure 4.10. Although for this section I could have restricted our attention to the immediate vicinity of the borehole, I assume that little prior knowledge about the location of the source is known to us. The purpose of this assumption is to simulate the situation in real cases where we know nothing about the location of a micro-earthquake. The sampling space of the possible source location is 16 m. Blue dots in Figure 4.10 represent the possible source locations. The size of each reference reconstructed wavefield is four wavelengths in space and four periods in time. The measurement uncertainty picked in this section is 0.95 due to the low S/N ratio, the large size of reference reconstructed wavefield, and the expected high source

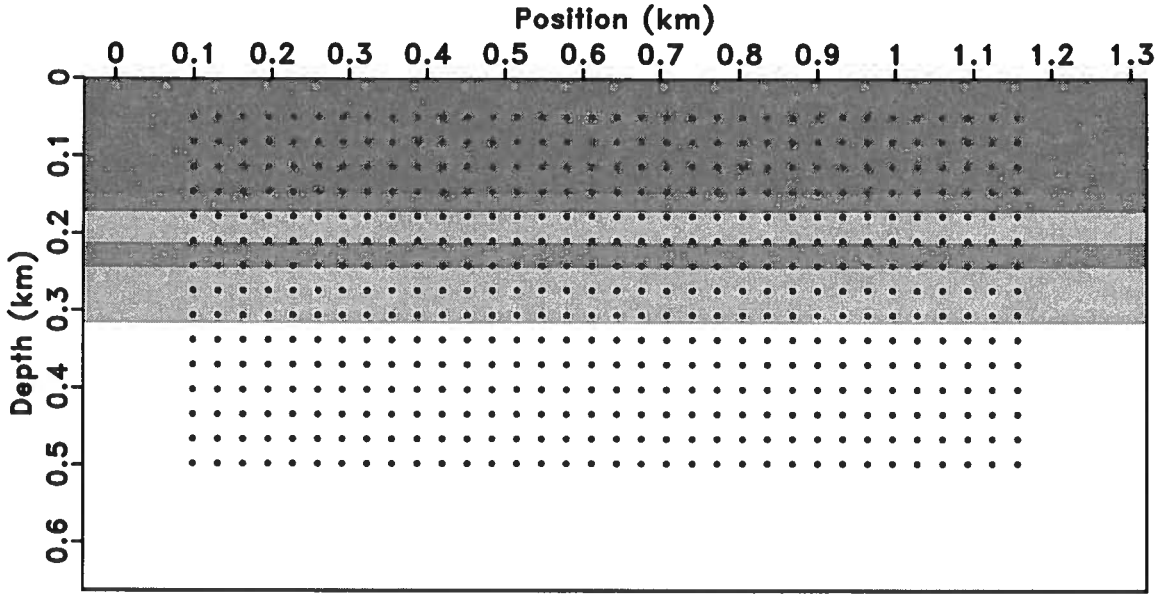


Figure 4.10. The prior information on the model space.

identification. I apply the method to string shots I, II, III, IV, and V and explain the location accuracy of the method in the following section.

For string shot I, I failed to detect the source location because the maximum probability value is less than 0.5 and no refocused field exists in the reconstructed wavefield at the corresponding location that generates the highest probability value. The failure is may caused by the low S/N ratio.

For string shot II, several snapshots of the reconstructed wavefield corresponding to the seismogram shown in Figure 4.8(b) are displayed in Figures 4.11(a)-4.11(d). Due to the low SNR, the noise in the reconstructed wavefield is strong. However, due to the wide coverage of the receivers, the refocused wavefield is easy to identify. Several snapshots of the probabilistic inversion result are shown in Figures 4.12(a)-4.12(d). In this sequence of snapshots, there are two possible source locations: one is indicated by Figure 4.12(b); the other is indicated by Figure 4.12(d). Both locations are characterized by a high degree of confidence. However, after considering these two source choices and the reconstructed wavefield, I select {756 m, 290 m} as the explosive source location for this seismogram. When I convert the detected source locations into the common 3-D coordinate framework shown in Figure 4.8(b), the source location is {East, North, Depth} = {40 m, −201.5 m,

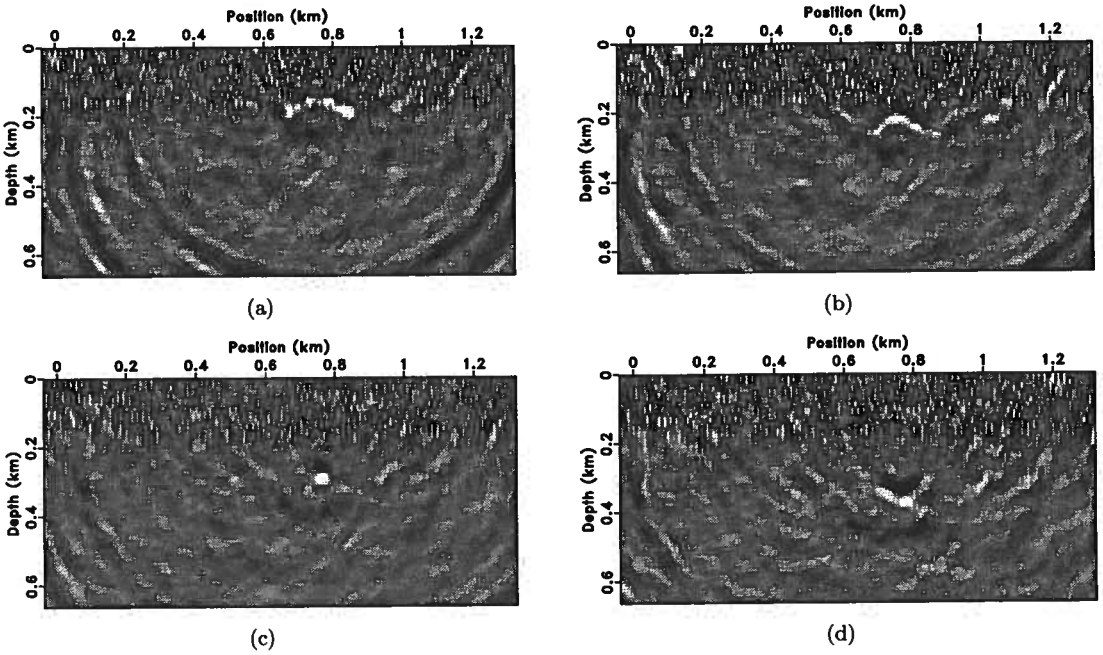


Figure 4.11. Four snapshots of the wavefield reconstructed from string shot II. (c) shows the focus of the wavefield.

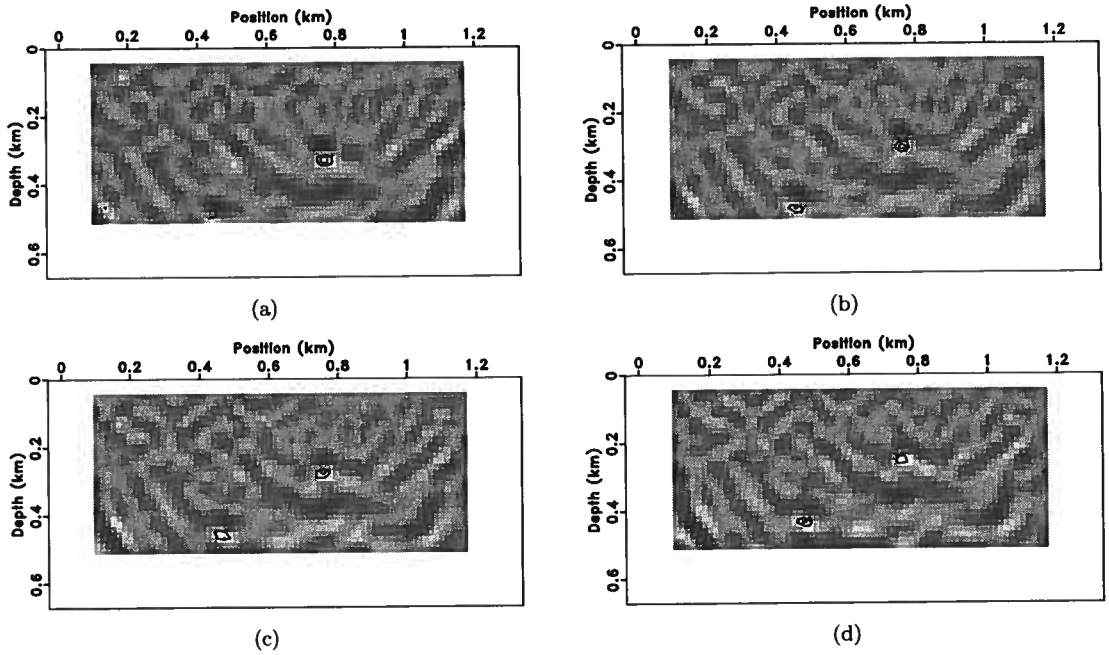


Figure 4.12. Four snapshots of the probability calculated from string shot II. The probability value of the outer contour is 0.75, and the contour increment is 0.05.

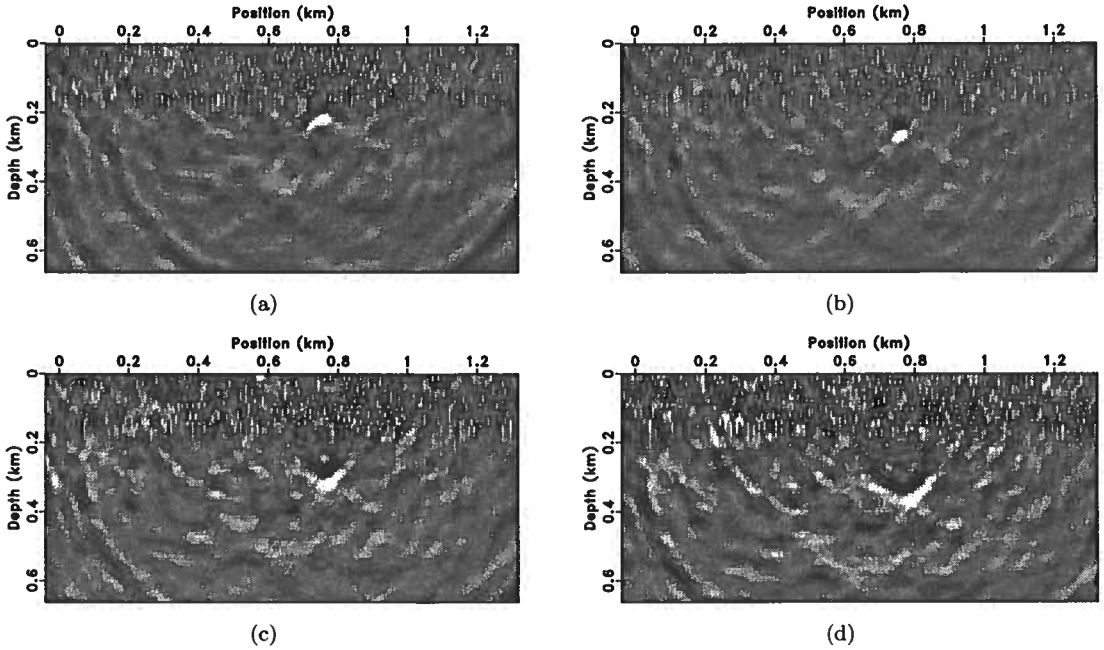


Figure 4.13. Four snapshots of the wavefield reconstructed from string shot III. (b) shows the focus of wavefield.

290 m}. Because of the 2D nature of the inversion which does not take into account the fact that the sources might be located off-line, a geometric correction is applied which moves the source location in space. The corrected source location is {31.70 m, −134.5 m, 282 m}. After this correction, I compare the true string shot location and the corrected source location which are shown in column II of Table 4.2. Although the location error in the depth direction is about 30 m, this error is less than 2 grid samplings of $\rho_{\mathfrak{M}}(\mathbf{m})$.

For string shot III, the same procedure applied to the seismogram shown in Figure 4.8(c) yields four snapshots of this reconstructed wavefield shown in Figure 4.13. Four snapshots of the probabilistic inversion result are shown in Figures 4.14(a)-4.14(d). In this experiment, the location of the highest probability value (shown in Figure 4.14(b)) is in the vicinity of the borehole. When I convert the source locations into the common 3-D coordinate framework, the source location is {East, North, Depth} = {40 m, −201.5 m, 258 m}. A geometric correction moves the source location in space to {31.70 m, −134.5 m, 249 m}. A comparison between the true string shot location and the corrected source location is shown in column III of Table 4.2. The location error is about 1.5

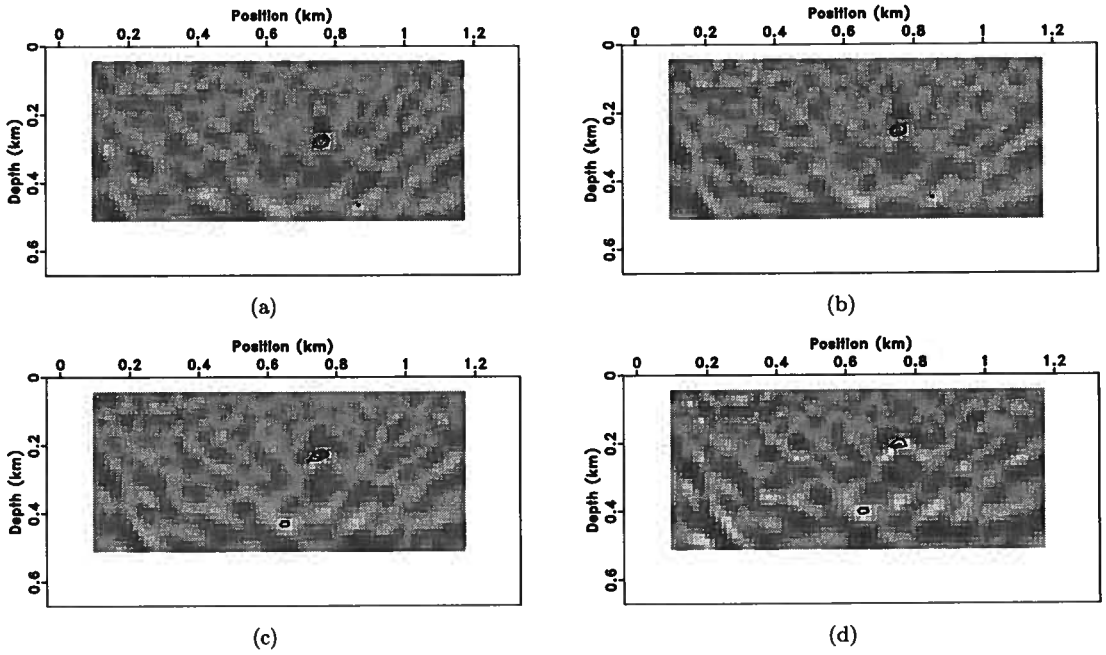


Figure 4.14. Four snapshots of the probability calculated from string shot III. The probability value of the outer contour is 0.75, and the contour increment is 0.05.

grid samplings.

For string shot IV, the same procedure used in the two previous seismograms is applied to this seismogram shown in Figure 4.8(d). Four snapshots of this reconstructed wavefield and four snapshots of the probabilistic inversion result are shown in Figures 4.15 and 4.16(a)-4.16(d), respectively. The location of the highest probability value is indicated by Figure 4.16(b). After converting the source locations into the common 3-D coordinate framework, the source location is {East, North, Depth} = {40 m, −201.5 m, 210 m}. A geometric correction moves the source location in space and the corrected source location becomes {31.70 m, −134.5 m, 189 m}. A comparison between the true string shot location and the corrected source location of this seismogram is shown in column IV of Table 4.2. The depth positions of the source provided by this seismogram is accurate.

For string shot V, after I apply the process to the seismogram shown in Figure 4.8(e), I get four snapshots of this reconstructed wavefield shown in Figure 4.17. Four snapshots of the probabilistic inversion result are shown in Figures 4.18(a)-4.18(d), where Figure 4.18(c) yields the location of the highest probability. When I convert the source locations into the common 3-D coordinate framework,

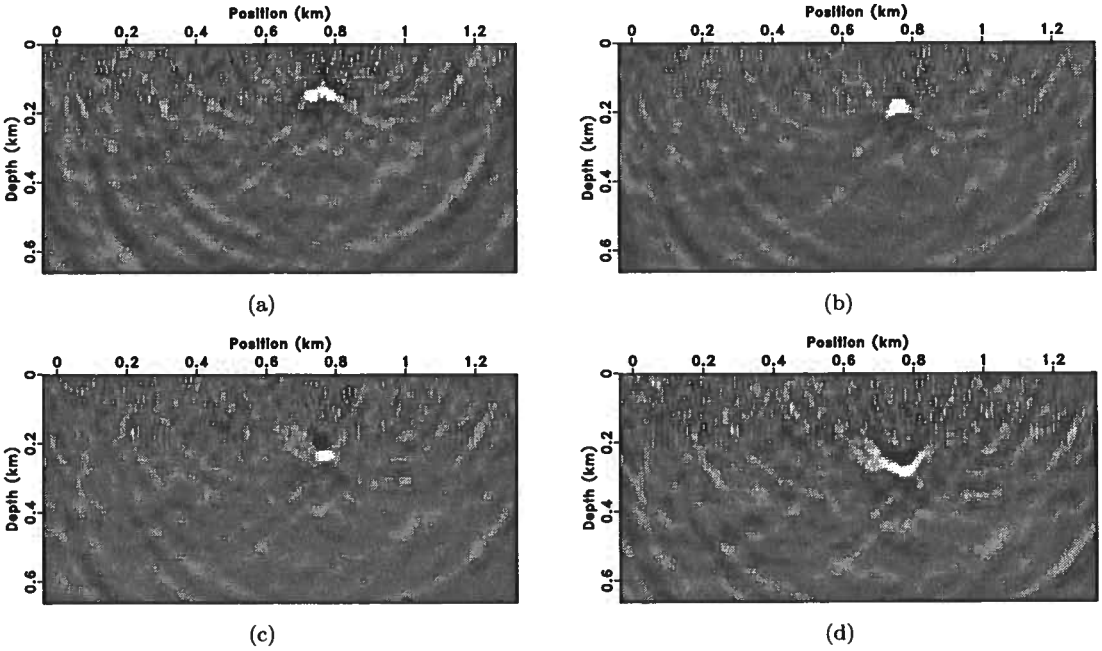


Figure 4.15. Four snapshots of the wavefield reconstructed from string shot IV. (c) shows the focus of the wavefield.

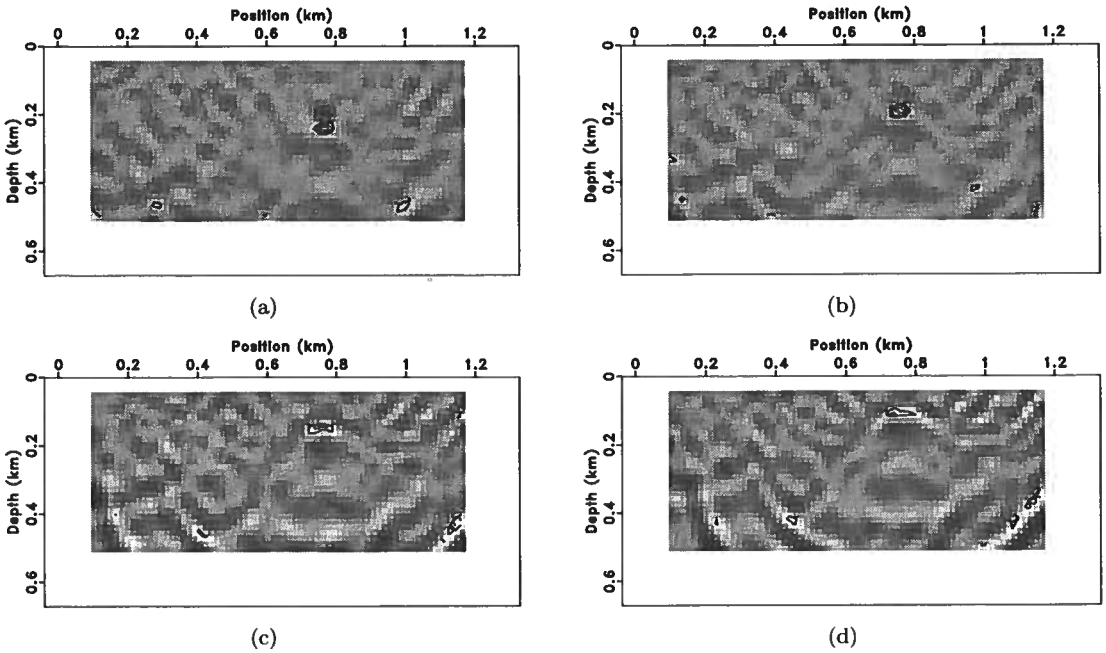


Figure 4.16. Four snapshots of the probability calculated from string shot IV. The probability value of the outer contour is 0.75, and the contour increment is 0.05.

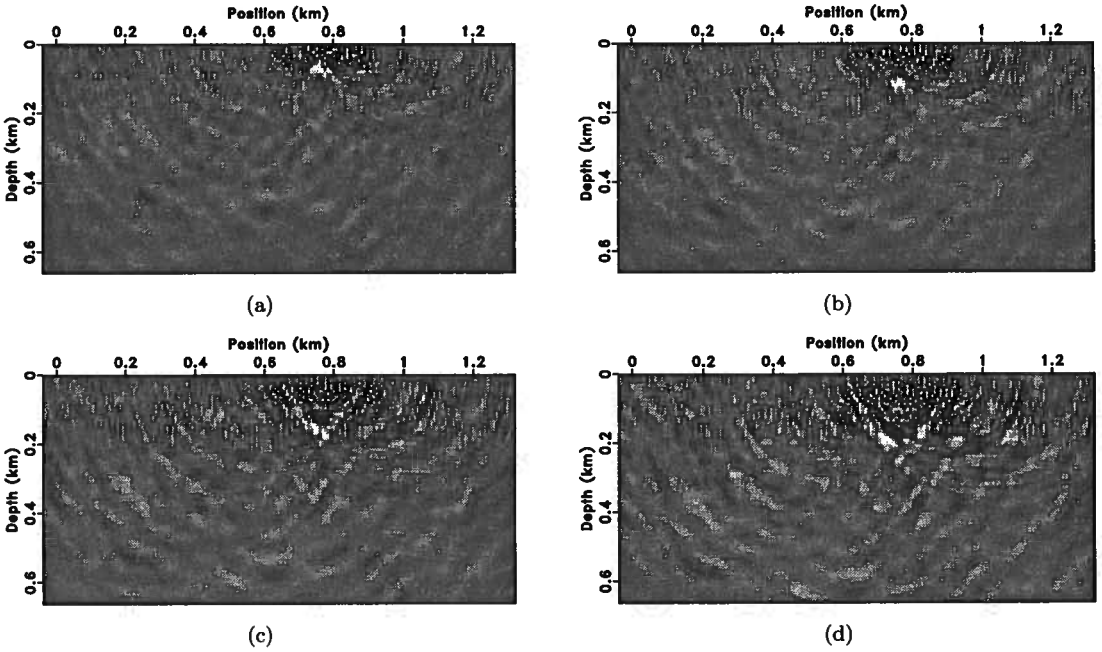


Figure 4.17. Four snapshots of the wavefield reconstructed from string shot V. (b) shows the focus of the wavefield.

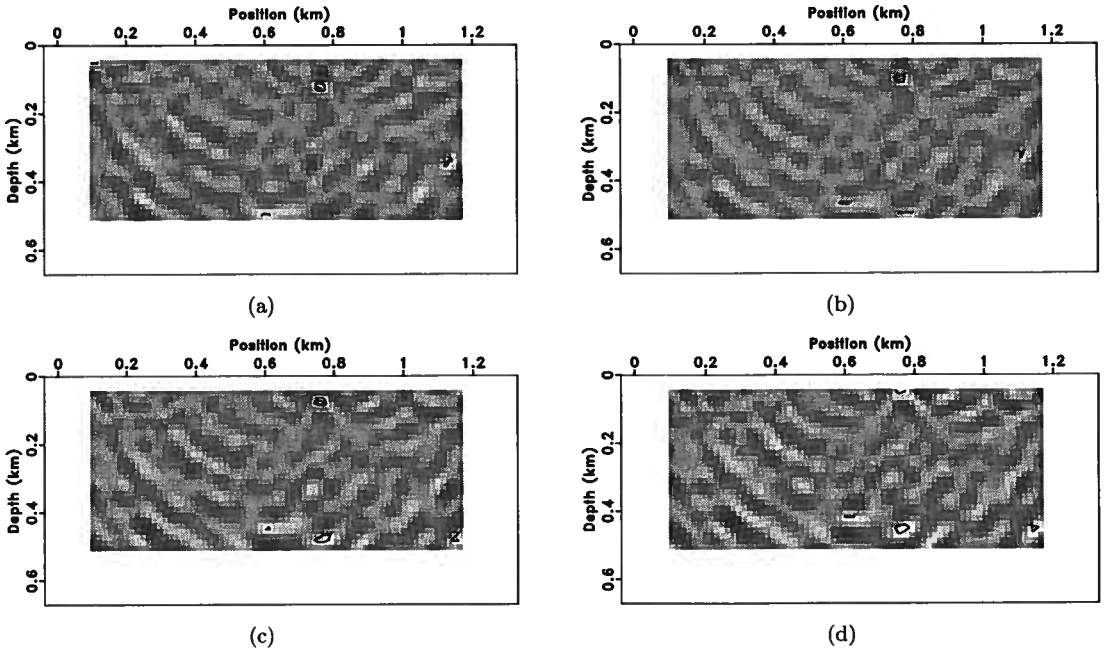


Figure 4.18. Four snapshots of the probability calculated from string shot V. The probability value of the outer contour is 0.75, and the contour increment is 0.05.

Table 4.2. A comparison between the true and corrected source locations. The top lines of East, North, and Depth are the true source locations. The bottom lines show the corrected source locations.

File ID	I	II	III	IV	V
East(m)	18.71	14.00	11.00	11.02	10.35
		31.70	31.70	31.70	31.70
North(m)	-66.85	-73.12	-76.11	-76.12	-76.78
		-134.5	-134.5	-134.5	-134.5
Depth(m)	291.87	256.78	222.00	186.00	97.00
		282	249	189	71

the source location becomes {East, North, Depth} = {40 m, −201.5 m, 98 m}. The source location with the applied geometric correction is {31.70 m, −134.5 m, 71 m}. A comparison between the true string shot location and the corrected source location is shown in column V of Table 4.2.

After correction, the source locations detected from the seismograms of string shots I, II, III, IV, and V are located along a vertical line which is in the vicinity of the borehole. The position difference in the north direction is still significant even after applying the geometric correction. However, errors of the depth positions are less than 3 grid samplings of the possible source location. Thus, we consider that the depth positions provided by the method are correct. An alternative to this geometric correction would be to run the inversion in 3D, thus reducing the spatial position ambiguity.

4.3 String shots (3D)

In this section, I apply the technique to the string shots using the 3D velocity model. This experiment is meant to demonstrate the feasibility of the technique in reality since the real world is a 3D space.

The layout of receivers is shown in Figure 4.1, and the velocity model shown in Figure 4.19 is the horizontal extension of the velocity used in the 2D model in the y direction. The calculation is done on a grid of x × y × z = 183 × 175 × 91 nodes with a grid spacing of 8 m. Because the calculation in the 3D space is much higher than that in the 2D space, the a priori state of information on the model space in this experiment is designed in the vicinity of the borehole and centered on the borehole. The number of possible source locations is x × y × z = 9 × 9 × 14 with a grid

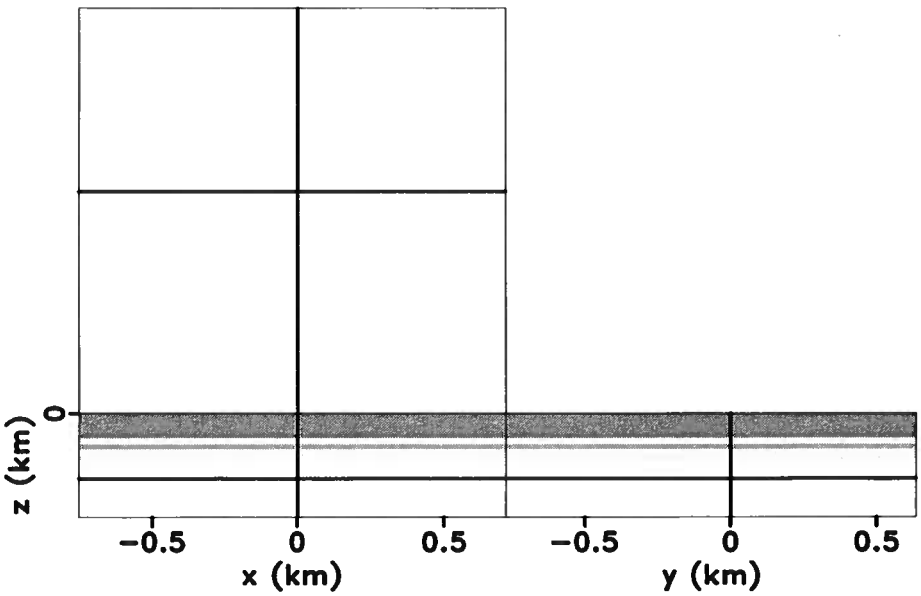


Figure 4.19. 3D velocity model.

spacing of 16 m. In order to maintain consistency with the previous section, I choose a measurement uncertainty of 0.95.

Due to the wide coverage of receivers, I expect the predicted reconstructed wavefield to be similar for different source locations around the borehole. Thus, instead of calculating a reconstructed wavefield database for all the selected sources in $\rho_{\mathfrak{M}}(\mathbf{m})$, I calculate only one reconstructed wavefield for the database and compare the wavefield reconstructed from string shots with this database at locations provided by $\rho_{\mathfrak{M}}(\mathbf{m})$. This strategy not only reduces the time used in calculating the reconstructed wavefield database but also saves the space used to store the database. However, one deficiency of this strategy may be low in location accuracy because no matter how wide the coverage of receivers is, the focus of wavefield various with the source location. Therefore, we should take the coverage of receivers and location accuracy into consideration before applying this strategy.

As discussed in the preceding section, the procedure used to select the most reliable string source from the source possibilities involves combining the probability map with the reconstructed wavefield. However, for the 3D velocity model, it is impractical for us to manually check the best focus of the reconstructed wavefield. Therefore, I first apply the method to the string shots to locate source possibilities. Then I check whether reconstructed wavefields focus at the locations provided

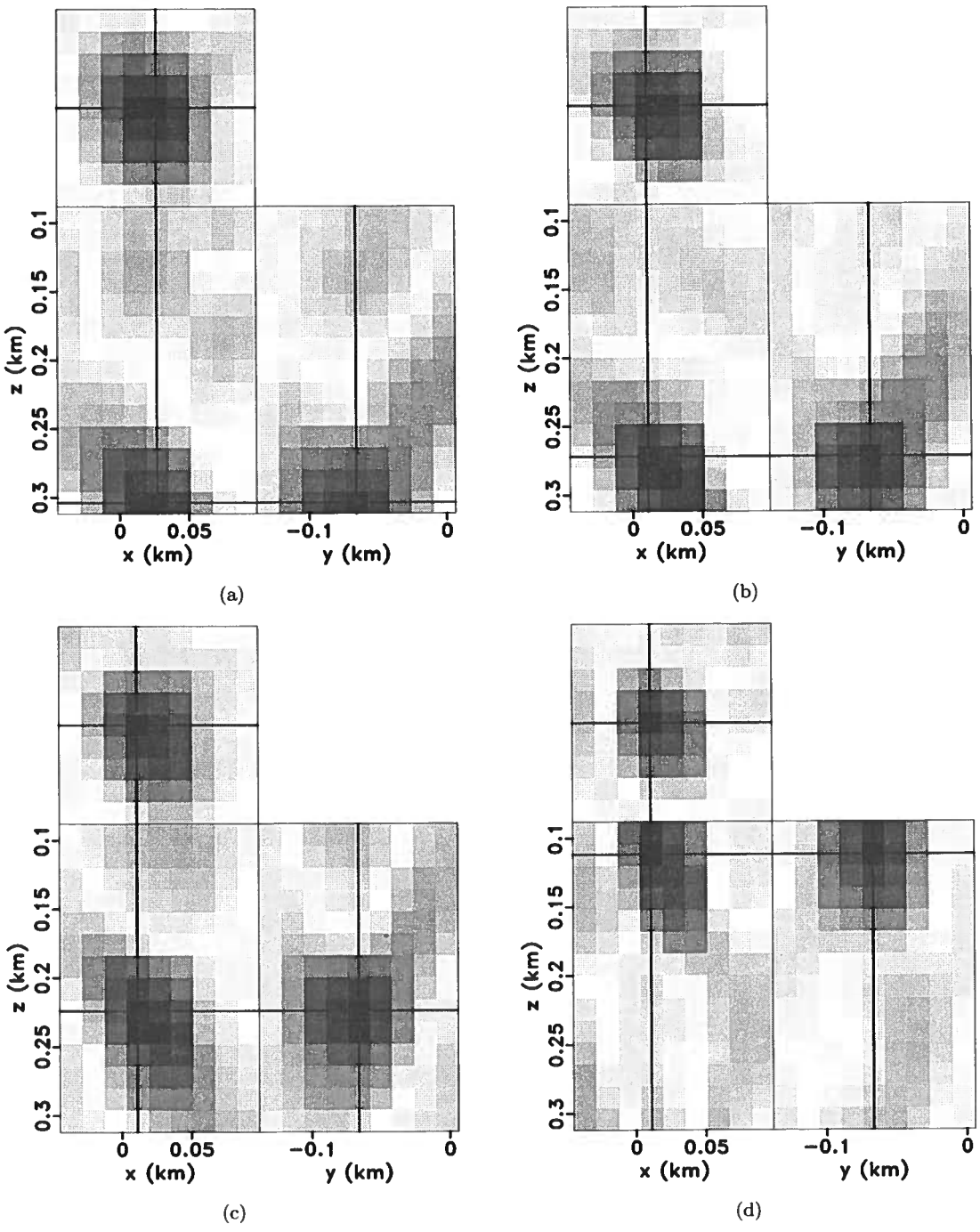


Figure 4.20. Probability values for (a) string shot I, (b) string shot II, (c) string shot III, (d) string shot IV, and (e) string shot V. The value of the white color is 0.6; the blue color shows values that are greater than 0.6; the red color shows values that are lower than 0.6.

by these choices. The adjusted procedure is applied to string shots I, II, III, IV, and V. For string shot I, I failed to find the source location because of the low maximum probability value.

For string shot II, one snapshot of the probability is shown in Figure 4.20(a). The intersections of the blue lines indicate that the location of the maximum probability value is about 0.87. This location is in the neighborhood of the borehole and is {East, North, Depth} = {27 m, −66 m, 304 m}. Six snapshots of the reconstructed wavefield are shown in Figure 4.21, where the intersections of the blue lines indicate the location with the highest probability value. Based on observation of these snapshots, I find that the wavefront focuses around the location with the highest probability value and then defocuses. However, although the coverage of receivers is wide, the wavefield does not focus into a point which may indicate that the explosive energy does not uniformly propagate from the string shot.

For string shot III, one snapshot of the probability is shown in Figure 4.20(b). The value of the maximum probability is about 0.88. The detected location is in the neighborhood of the borehole and is {East, North, Depth} = {11 m, −66 m, 272 m}. Six snapshots of the reconstructed wavefield are shown in Figure 4.22, where the intersections indicate the location with the highest probability value.

For string shot IV, one snapshot of the probability is shown in Figure 4.20(c). The value of the maximum probability is about 0.89. The detected location is in the neighborhood of the borehole and is {East, North, Depth} = {11 m, −66 m, 224 m}. Six snapshots of the reconstructed wavefield are shown in Figure 4.23, where the intersections indicate the location with the highest probability value.

For string shot V, one snapshot of the probability is shown in Figure 4.20(d). The value of the maximum probability is about 0.85. The detected location is in the neighborhood of the borehole and is {East, North, Depth} = {11 m, −66 m, 112 m}. Six snapshots of the reconstructed wavefield are shown in Figure 4.24, where the intersections indicate the location with the highest probability value. Due to the incorrect velocity information near the top of the surface, the variation of the reconstructed wavefield in this area is great. This variation affects the location accuracy especially when the explosive source is located close to the surface. This explains why the probability of the detected source of string shot V is lower than the other three shots.

For string shots II, III, IV, and V, the selected source locations all correspond to the location with the highest probability. A comparison between the true source locations and the detected

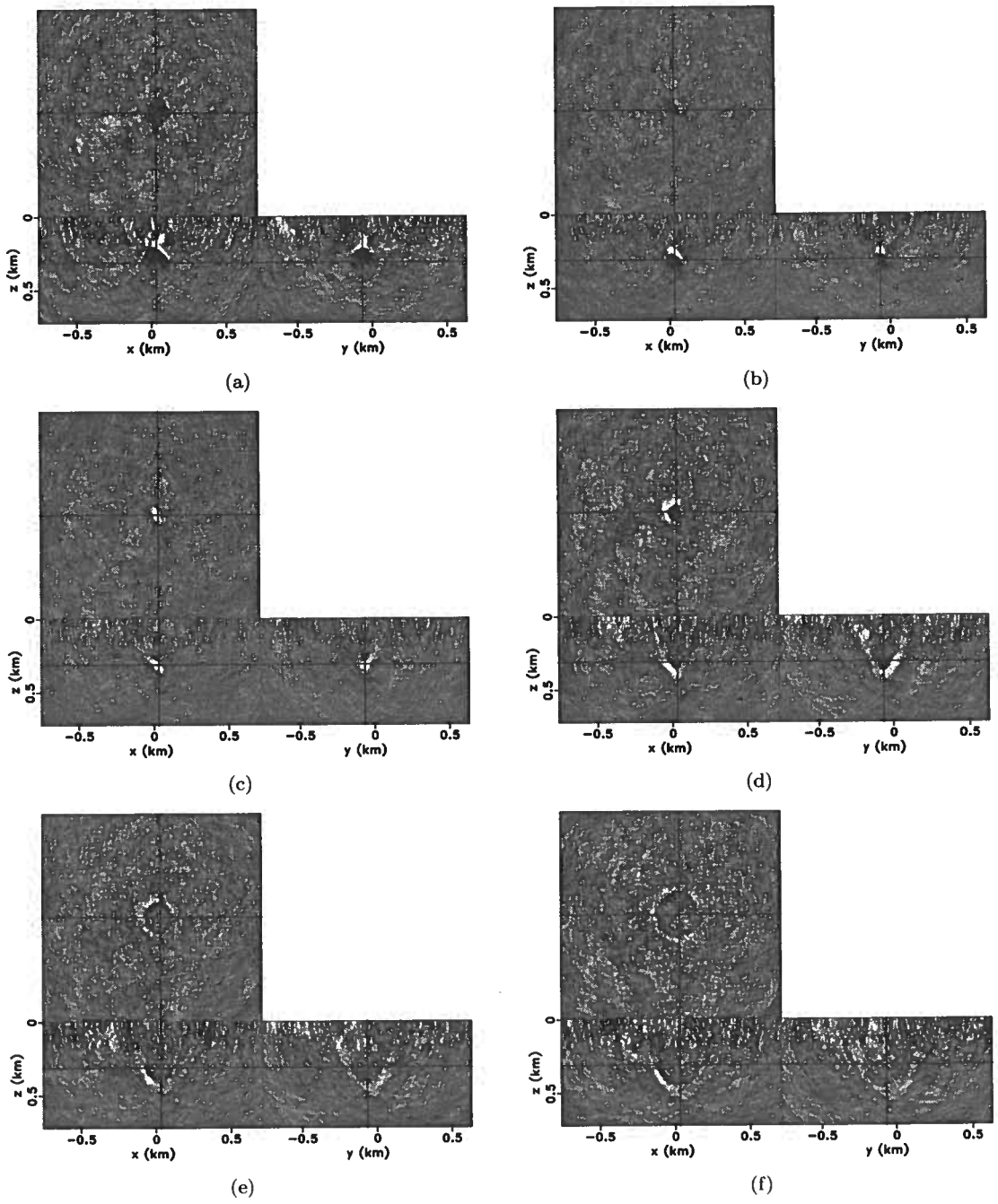


Figure 4.21. Six snapshots of the wavefield reconstructed from string shot II. (c) shows the focus of the wavefield.

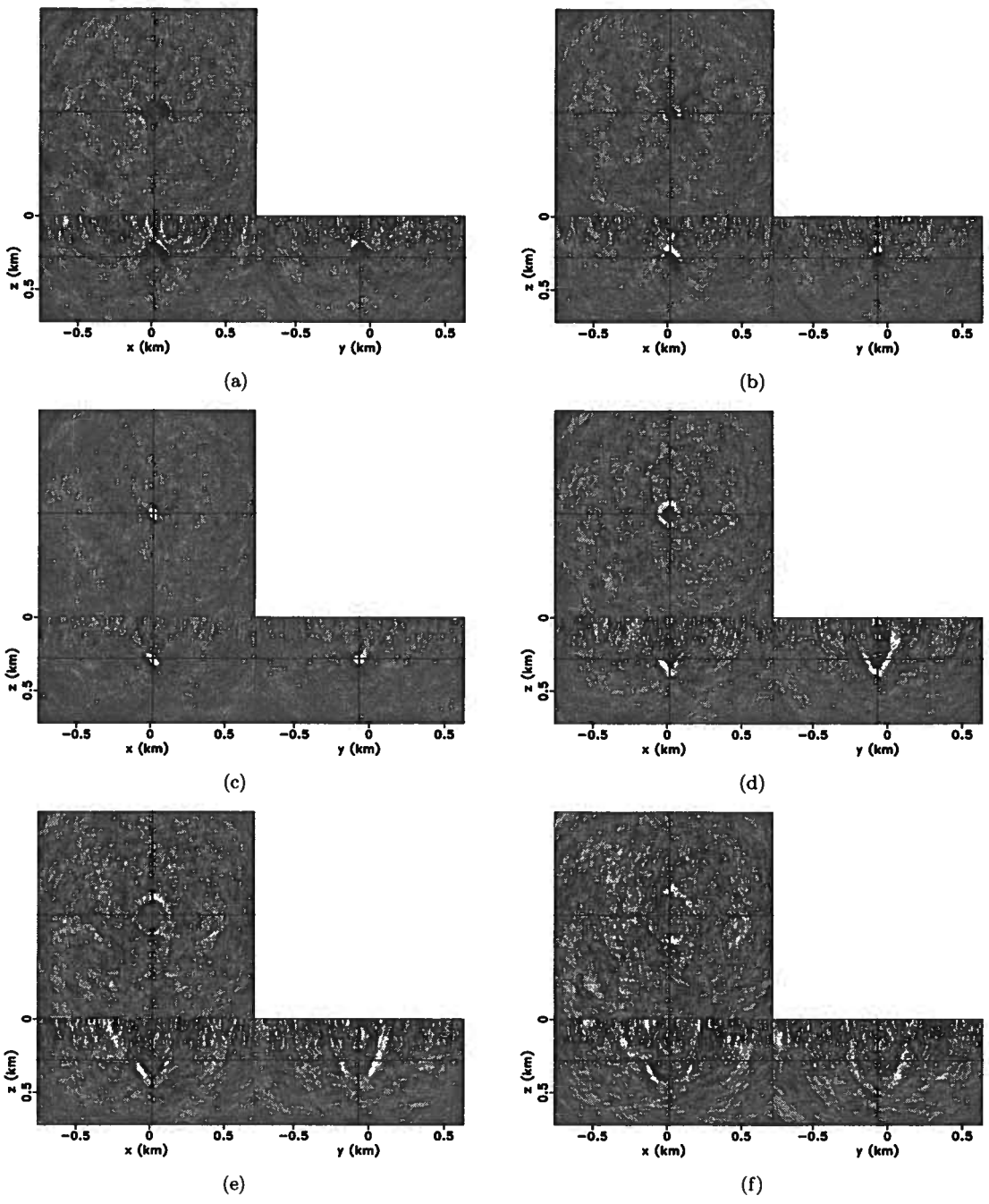


Figure 4.22. Six snapshots of the wavefield reconstructed from string shot III. (c) shows the focus of the wavefield.

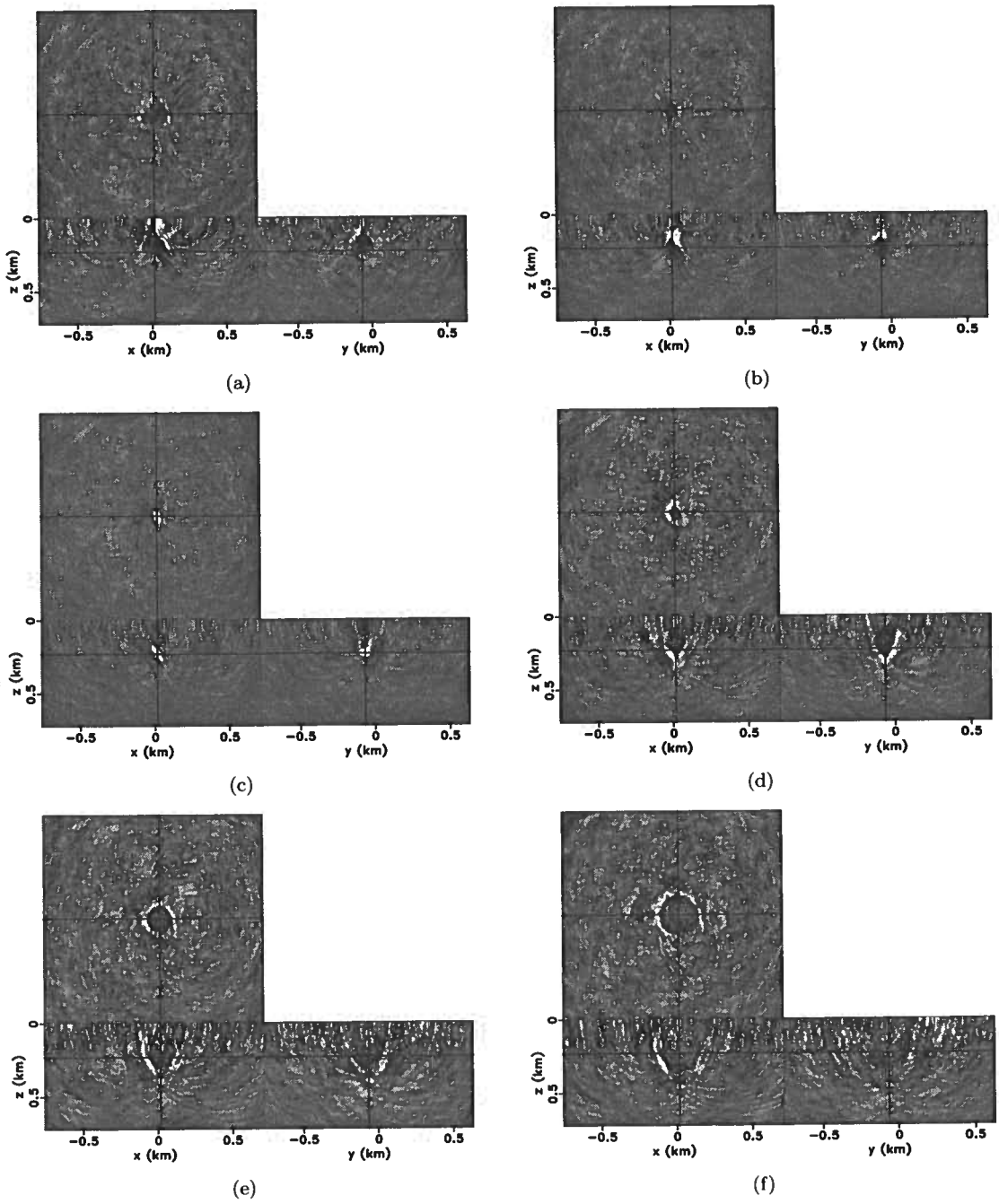


Figure 4.23. Six snapshots of the wavefield reconstructed from string shot IV. (c) shows the focus of the wavefield.

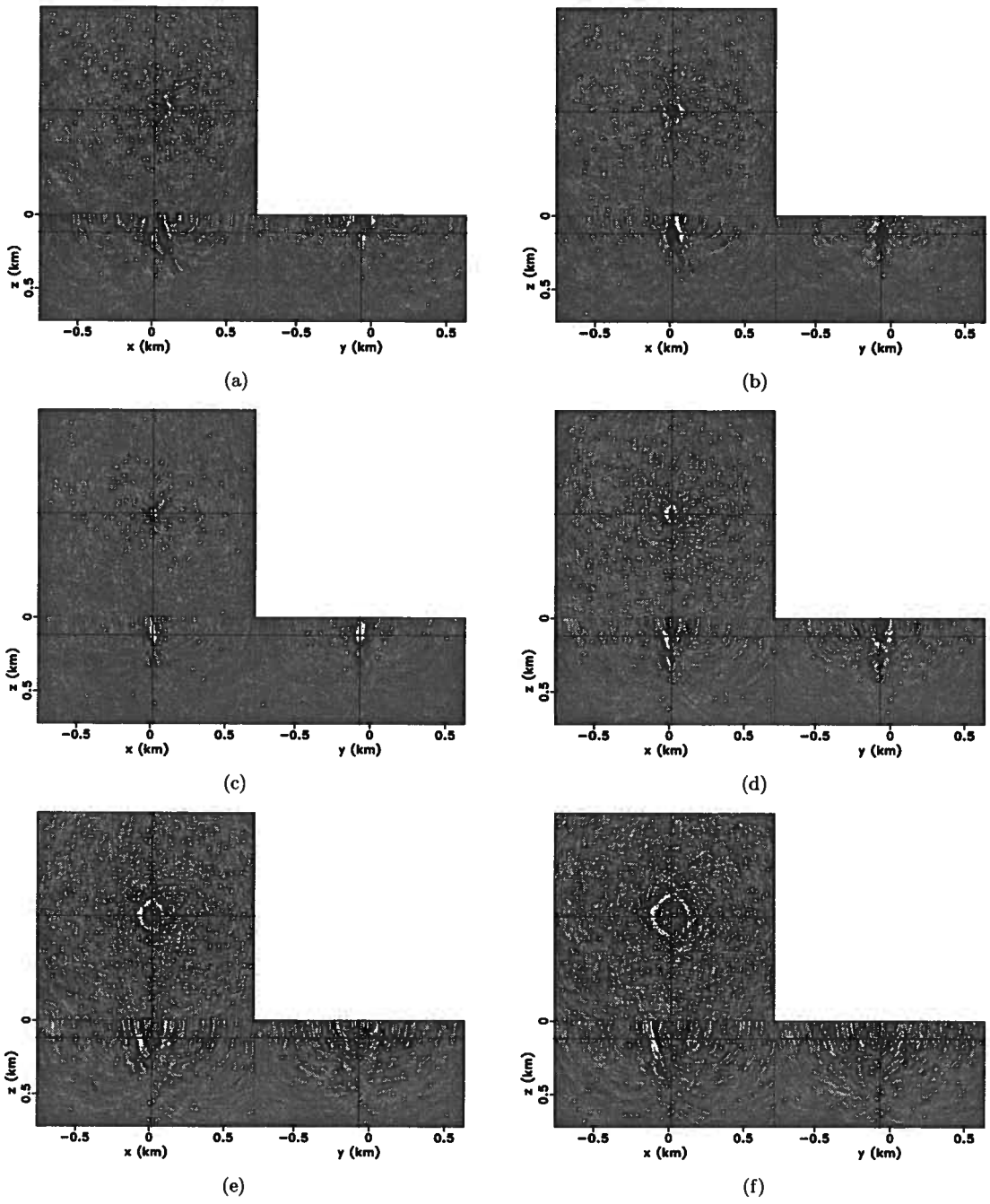


Figure 4.24. Six snapshots of the wavefield reconstructed from string shot V. (c) shows the focus of wavefield.

Table 4.3. A comparison between the true and corrected source locations. The top lines of East, North, and Depth are the true source locations. The bottom lines show the detected source locations.

File ID	I	II	III	IV	V
East(m)	18.71	14.00	11.00	11.02	10.35
		27.00	11.00	11.00	11.00
North(m)	-66.85	-73.12	-76.11	-76.12	-76.78
		-66.00	-66.00	-66.00	-66.00
Depth(m)	291.87	256.78	222.00	186.00	97.00
		304	272	224	112

source locations, using the 3D velocity model, is shown in Table 4.3. The location error in the north and east directions is small and less than one grid sampling. However, the location error in the depth direction is greater than that in the other two directions. Nevertheless, these location errors are less than 4 grid samplings and less than one wavelength which is about 67 m. Thus, I consider that I locate these sources accurately since the sources are located within the expected region of error which is one wavelength.

For the 3D velocity model, I expect a much higher location accuracy than I find here, especially in the depth direction. One reason for the location error in this experiment may be that there is only one reconstructed wavefield in the database. However, the location accuracy does not improve significantly even when the database is composed of the wavefields calculated from sources provided by $\rho_{\mathfrak{M}}(\mathbf{m})$. Two other possible causes for the error are: the low SNR of the reconstructed wavefield (because noise can distort the wavefront) or the source type (because the wavelet I used to calculate the database is the Ricker wavelet).

In this section, I show that the method can be easily extended to 3D space and it improves the location accuracy in the north and east directions without the requirement of geometric correction.

Chapter 5

Conclusions

I present a method for automatic micro-earthquake location using Bayesian inversion theory. The method exploits the unique shape of wavefields partially refocused at the source position. This method does not require picking of arrival times but relies on wavefield focusing obtained by time reversal. The method not only identifies the spatial location of the source but also specifies the onset time of the source. Furthermore, since I am using a probabilistic approach to micro-earthquake location, the method provides confidence maps of micro-earthquake locations which can be used for risk assessment.

Synthetic data examples demonstrate the robustness of the method and its applicability to situations where micro-earthquakes occur in close succession in a small region in space and are recorded in a noisy environment with SNR as low as 0.1. I also illustrate the technique with field data in a noisy environment using both 2D and 3D velocity models. The results demonstrate the robustness of the method for both perf and string shots.

Due to limited time, I have focused only on the explosive sources triggered in acoustic media. However, there are further possibilities for future study: 1) we should be able to apply this method to elastic media with double-couple sources; 2) By reducing the computation cost of this method, we can apply it to real-time monitoring reservoirs; 3) By incorporating this method with others in the travel-time based category, we can not only identify the magnitude of micro-earthquakes but also detect the stress and strain orientations of faults.

References

Aki, K, & Richards, P. 2002. *Quantitative Seismology.* University Science Books.

Baysal, E. 1983. Reverse time migration. *Geophysics*, **48**(11), 1514–1524.

Beutel, J, Sonka, M, Kundel, H, Fitzpatrick, J, & Metter, R. 2000. *Handbook of Medical Imaging: Medical Image Processing and Analysis.* SPIE Press.

Borcea, L, Papanicolaou, G, & Tsogka, C. 2003. Theory and applications of time reversal and interferometric. *PACS*, 1–24.

Born, M, & Wolf, E. 2003. *Principles of optics: Electromagnetic theory of propagation, interference and diffraction of light.* Cambridge University Press.

Collignon, A, Maes, F, Delaere, D, Vandermeulen, D, Suetens, P, & Marchal, G. 1995. Automated multi-modality image registration based on information theory. *Medical Imaging*, 263–274.

Duncan, M. 2005. Is there a future for passive seismic? *First Break*, **23**, 111–115.

Earle, P, & Shearer, P. 1994. Characterization of Global Seismograms Using an Automatic-Picking Algorithm. *Bulletin of the Seismological Society of America*, **84**(2), 366–376.

Economides, M. 1992. *A Practical Companion to Reservoir Stimulation.* Elsevier.

Evans, K, Moriya, H, Niitsum, H, Jones, R, Phillips, W, Genter, A, Sausse, J, Jung, R, & Baria, R. 2005. Microseismicity and permeability enhancement of hydrogeologic structures during massive fluid injections into granite at 3 km depth at the Soultz HDR site. *Geophysical Journal International*, **160**(1), 389–412.

Fischer, T, Hainzl, S, Eisner, L, Shapiro, S, & Calvez, J. 2008. Microseismic signatures of hydraulic fracture growth in sediment formations: Observations and modeling. *Journal of Geophysical Research Solid Earth*, **113**, 1–12.

Fouque, J, Garnier, J, Nachbin, A, & Solna, K. 2004. Time Reversal Refocusing for Point Source in Randomly Layered Media. 1–32.

Fremont, M, & Malone, S. 1987. Hige precision Relative Locations of Earthquakes at Mount St. Helens, Washington. *Journal of Geophysical Research*, **92**(B10), 10223–10236.

Gajewski, D. 2005. Reverse modelling for seismic event characterization. *Geophysical Journal International*, **163**(1), 276–284.

Hajnal, J, Hill, D, & Hawkes, D. 2001. *Medical Image Registration.* CRC Press.

Hill, D, Batchelor, P, Holden, M, & Hawkes, D. 2001. Medical image registration. *Physics in Medicine and Biology*, **46**(2001), R1–R45.

House, L. 1987. Locating Microearthquakes Induced by Hydraulic Fracturing in Crystalline Rock. *Geophysical Research Letters*, **14**(9), 919–921.

Jaynes, E. 2003. *Probability Theory: The Logic of Science.* Cambridge University Press.

Joncour, F, Lambare, G, Svay-Lucas, J, & Duquet, B. 2005a. True amplitude migration by wavefield continuation. *SEG Expanded Abstracts*, **24**, 1942–1945.

Joncour, F, Svay-lucas, J, Lambare, G, & Duquet, B. 2005b. True amplitude wave equation. *EAGE 67th conference and technical exhibition*, 1–4.

Keppler, H, Pearson, C, Potter, R, & Albright, J. 1982. Microearthquakes induced during hydraulic fracturing at the Fenton Hill HDR site: the 1982 experiments. *1983 Geothermal Resources Council's annual meeting*, 1–9.

King, M, & Willis, G. 1957. Mechanics of hydraulic fracturing. *Trans. Am. Inst. of Mining, Metallurgical, and Petroleum Eng.*, **210**, 175–190.

Leonard, M. 2000. Comparison of manual and automatic onset time picking. *Bulletin of the Seismological Society of America*, **90**(6), 1384–1390.

Leonard, M, & Kennett, B. 1999. Multi-component autoregressive techniques for the analysis of seismograms. *Physics of the Earth and Planetary Interiors*, **113**, 247–263.

Levin, S. 1984. Principle of reverse-time migration. *Geophysics*, **49**, 581–583.

Maes, F, Collignon, A, Vandermeulen, D, Marchal, G, & Suetens, P. 1997. Multimodality image registration by maximization of mutual information. *Medical Imaging*, **16**(2), 187–198.

Maxwell, S, & Urbancic, T. 2001. The role of passive microseismic monitoring in the instrumented oil field. *The Leading Edge*, 636–639.

Miyazawa, M. 2007. Analysis of Micro-earthquake Data at Cold Lake and its Applications to Reservoir Monitoring. *SEG/San Antonio 2007 Annual Meeting.*

Mosegaard, K, & Tarantola, A. 2002. *Probabilistic Approach to Inverse Problems.* Academic Press.

Papanicolaou, G, Ryzhik, L, & Solna, K. 2004. Statistical Stability in Time Reversal. *Society for Industrial and Applied Mathematics*, **64**(4), 1133–1155.

Pavlis, G. 1986. Appraising Earthquake Hypocenter Location Errors: A Complete, Practical Approach for Single-event Locations. *Bulletin of the Seismological Society of America*, **76**(6), 1699–1717.

Phillips, W. 1998. Induced microearthquake patterns and oil-producing fracture systems in the Austin chalk. *Tectonophysics*, **289**, 153–169.

Plessix, R, & Mulder, W. 2002. Amplitude-preserving finite-difference migration based on a least-squares formulation in the frequency domain. *SEG Int'l Exposition and 72nd Annual Meeting*, 1–4.

Poupinet, G, Ellsworth, W, & Frechet, J. 1984. Monitoring Velocity Variations in the Crust Using Earthquake Doublets: An application to the Calaveras Fault, California. *Journal of Geophysical Research*, **89**(B7), 5719– 5731.

Rentsch, S. 2004. Location of seismicity using Gaussian beam type migration. *SEG Expanded Abstracts*, **23**, 354–357.

Rentsch, S. 2006. Passive seismic monitoring using Gaussian Beams with application to borehole data from the San Andreas Fault at Parkfield ,California. *SEG/New Orleans 2006 Annual Meeting*, 1580–1583.

Rentsch, S. 2007. Fast location of seismicity:A migration-type approach with application to hydraulic-fracturing data. *Geophysics*, **72**(1), S33–S40.

Rothert, E, & Shapiro, S. 2003. Microseismic monitoring of borehole fluid injections: Data modeling and inversion for hydraulic properties of rocks. *Geophysics*, **68**(2), 685–689.

Rutledge, J, & Phillips, W. 2001. Hydraulic Stimulation of Natural Fractures as Revealed by Induced Microearthquakes, Carthage Cotton Valley Gas Field, East Texas. *Geophysics*, 1–37.

Schaff, D. 2008. Semiempirical Statistics of Correlation-Deteror Performance. *Bulletin of the Seismological Society of America*, **98**(3), 1495–1507.

Schaff, D, & Waldhauser, F. 2005. Waveform Cross-Correlation-Based Differential Travel-Time Measurements at the North California Seismic Network. *Bulletin of the Seismological Society of America*, **95**(6), 2446–2461.

Schaff, D, Bokelmann, G, Ellsworth, W, Zanzerkia, E, Waldhauser, F, & Beroza, G. 2004. Optimizing Correlation Techniques for Improved Earthquake Location. *Bulletin of the Seismological Society of America*, **94**(2), 705–721.

Shannon, C. 1948. A mathematical theory of communication. *Bell system technical journal*, **27**, 379–423, 623–656.

Shapiro, S, Dinske, C, & Rothert, E. 2006. Hydraulic-fracturing controlled dynamics of microseismic clouds. *Geophysical Research Letter*, **33**(14), 1–5.

Shearer, P. 1997. Improving Local Earthquake Locations using the L1 Norm and Waveform Cross Correlations: Application to the Whittier Noarrows, California, Aftershock Sequence. *Journal of Geophysical Research*, **102**(B4), 8269–8283.

Tabeli, S, & Cornet, H. 1987. Analysis of the microseismicity induced by a fluid injection in a granite rock mass. *J. Geophys. Res.*, **14**, 227–230.

Talebi, S, Boone, T, & Nechtschein, S. 1998a. A seismic model of casing failure in oil fields. *Pure Appl. Geoph.*, **153**, 197–217.

Talebi, S, Nechtschein, S, & Boone, J. 1998b. Seismicity and casing failures due to steam injection in oil sands. *Pure Appl. Geoph.*, **153**, 219–233.

Tarantola, A. 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation.* Society for Industrial and Applied Mathematics.

Tygel, M, Schleicher, J, & Hubral, P. 1992. Geometrical spreading corrections of offset reflections in laterally inhomogeneous earth. *Geophysics*, **57**(8), 1054–1063.

Vidale, J. 1986. Complex polarization analysis of particle motion. *Bulletin of the Seismological Society of America*, **76**(5), 1393–1405.

Viola, P. 1995. Alignment by Maximization of Mutual Information. *A.I Technical Report*, **1548**, 1–156.

Waldhauser, F. 2000. A Double-Difference Earthquake Location Algorithm: Method and Application to the Northern Hayward Fault, California. *Bulletin of the Seismological Society of America*, **90**(6), 1353–1368.

Wells, W, Viola, P, Atsumi, H, Nakajima, S, & Kikinis, R. 1997. Multi-modal image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging.*

Woods, R, Cherry, S, & Mazziotta, J. 1992. Rapid automated algorithm for aligning and reslicing PET images. *Journal of computer assisted tomography*, **16**(4), 620–33.

Woods, R, Grafton, S, Holmes, C, Cherry, S, & Mazziotta, J. 1998a. Automated image registration: I. general methods and intrasubject, intramodality validation. *Journal of computer assisted tomography*, **22**(1), 139–152.

Woods, R, Grafton, S, Watson, J, Sicotte, G, & Mazziotta, J. 1998b. Automated image registration: II. intersubject validation of linear and nonlinear models. *Journal of computer assisted tomography*, **22**(1), 153–165.

Zitova, B, & Flusser, J. 2003. Image registration methods: a survey. *Image and Vision Computing*, **21**, 977–1000.