Probabilistic micro-earthquake location for reservoir monitoring

Ran Xuan and Paul Sava, Center for Wave Phenomena, Colorado School of Mines*

SUMMARY

Micro-earthquake sources and onset times are identified within the general framework of Bayesian inversion using time reversal. Given an assumption about the possible locations of micro-earthquakes, we use recorded data to evaluate the feasibility 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. We simulate wavefields corresponding to possible sources distributed in the model and evaluate their match with the wavefield reconstructed from field data. The method operates like a pattern-recognition procedure and can exploit a wide variety of techniques designed for this purpose. We use simple cross-correlation to take advantage 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 our method is able to identify not only the position 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.

INTRODUCTION

High-pressure fluid injected into oil and gas reservoirs causes time-invariant stress and strain changes. When the stress exceeds a threshold characterizing the resistance of rock materials to stress, micro-seismicity is triggered by the release of pressure along pre-existing fractures or through the creation of new fractures (Maxwell and Urbancic, 2001; Duncan, 2005). Precisely locating micro-seismic events can be used to monitor the hydraulic fracturing and for reservoir characterization (Rentsch, 2004).

Most location methods currently used require the identification of seismic arrivals which involves accurate picking of P- and S-wave arrival times. The onset time 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. The source is located by optimizing a misfit function between measured and calculated quantities (Pujol, 2004; Lay and Wallace, 1995; Thurber and Rabinowitz, 2000). Such location methods strongly depend on the picking accuracy of P- and S-wave arrival times and have a low degree of automation (Rentsch, 2007).

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

Within the context of migration-based methods, we advocate in this paper a new methodology which can be used to automatically identify and locate sources of micro-seismicity by using time-reversal based on numeric solutions to the acoustic wave-equation to reconstruct seismic wavefields.

MICRO-EARTHQUAKE LOCATION

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. The posterior combines our prior understanding of the model parameters, independent of any subsequent measurement we make, with additional information derived from measurements and assumptions about the physical laws governing the process under investigation. As discussed by Tarantola (2005), the posterior state of information is obtained using the Bayes theorem

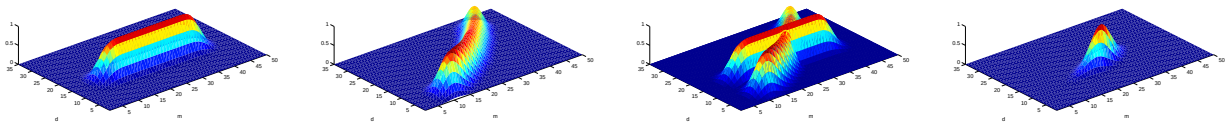
$$\sigma(\mathbf{m},\mathbf{d}) = k\frac{\rho(\mathbf{m},\mathbf{d})\Theta(\mathbf{m},\mathbf{d})}{\mu(\mathbf{m},\mathbf{d})}\ , \tag{1}$$

where $\mathbf{m}$ is the model, $\mathbf{d}$ is the data, $\rho(\mathbf{m},\mathbf{d})$ is the a priori state of information, $\Theta(\mathbf{m},\mathbf{d})$ is the theoretical state of information, $\mu(\mathbf{m},\mathbf{d})$ is the homogeneous state of information, k is a normalization constant, and $\sigma(\mathbf{m},\mathbf{d})$ is the a posteriori state of information. According to equation 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 $\rho(\mathbf{m},\mathbf{d})$, Figures 1(a)-1(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:

$$\sigma_{\mathfrak{M}}(\mathbf{m}) = \int_{\mathfrak{D}} \sigma(\mathbf{m},\mathbf{d})d\mathbf{d}\ , \tag{2}$$

where $\mathfrak{D}$ is the data space. When the uncertainty of the kernel function is negligible, $\sigma_{\mathfrak{M}}(\mathbf{m})$ is

$$\sigma_{\mathfrak{M}}(\mathbf{m}) = \nu\rho_{\mathfrak{M}}(\mathbf{m})\rho_{\mathfrak{D}}\left(G(\mathbf{m}),\mathbf{d}^{obs}\right)\ , \tag{3}$$



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

Figure 1: (a) A priori probability density function $\rho(\mathbf{m},\mathbf{d})$. (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.

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.

We apply the Bayesian inversion theory to solve the problem of micro-earthquake identification. The locations corresponding to micro-seismic events are provided by incorporating information from recorded seismograms with prior information on the events. 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 $\mathbf{d}$ represent the wavefields reconstructed from recorded seismograms in the space around the most probable micro-earthquake source locations and onset times. The data parameters $D(\mathbf{x}, t)$ are voxels of the 4-D wavefield. The kernel function $G(\mathbf{m})$ describes the theoretical relationship between the data and model. In our case, the kernel function is based on the two-way acoustic wave-equation. For a given model, we refer to the data calculated using this procedure as the predicted data $D^{pre} = G(\mathbf{m})$. In contrast, we refer to the wavefields reconstructed from field seismograms as the observed data D^{obs}.

Prior information on the model $\rho_{\mathfrak{M}}(\mathbf{m})$
We assume that the hypocenter and the onset time of micro-seismic events may be located with equal probability within a certain space domain and time interval. Thus, the prior probability $\rho_{\mathfrak{M}}(\mathbf{m}) = 1$ within this region of space and time and 0 elsewhere. We uniformly sample the model space within this region. The assumption we make here is that the observed data characterize with equal probability all models within our selected model space, as illustrated in Figure 1(a).

We demonstrate our methodology on an example with multiple overlapping micro-earthquake events. In this numeric experiment, we assume that the micro-seismic events are distributed along a fracture located at $z = 0.326$ km, as indicated in Figure 2. Figure 3 shows the simulated seismogram characterizing all the events. The signal-to-noise ratio (SNR) for this example is as low as 0.1. We assume that measurement uncertainties are 0.64 because of the low SNR. We simulate 20 micro-earthquakes and attempt to locate all of them with our methodology, as seen in Figures 6(a)-6(h).

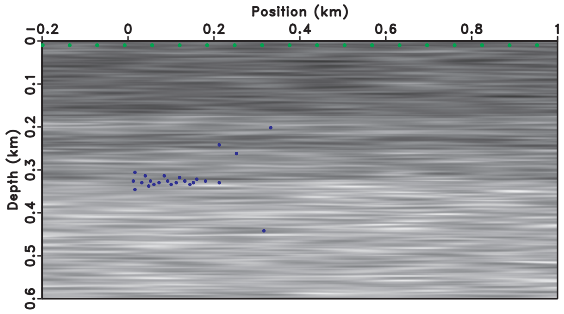


Figure 2: Synthetic model. The dots located at about $z = 0$ km represent receivers, the other dots represent the micro-seismic events.

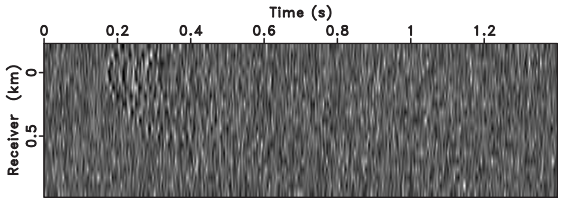


Figure 3: Synthetic seismogram. A total of 7000 time steps with an increment of 0.2 ms are calculated.

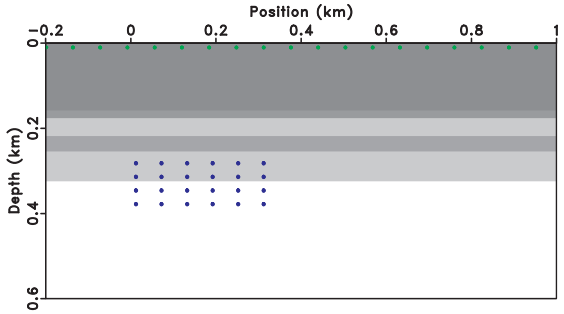


Figure 4: A schematic representation of the a priori information on the model space. We assume that the micro-earthquakes occur in the region indicated by the dense dots between $x = 0 - 0.35$ km and $z = 0.25 - 0.35$ km. A similar discrete set of possible onset times are defined along the time axis. The acquisition array consists of 19 receivers at the surface.

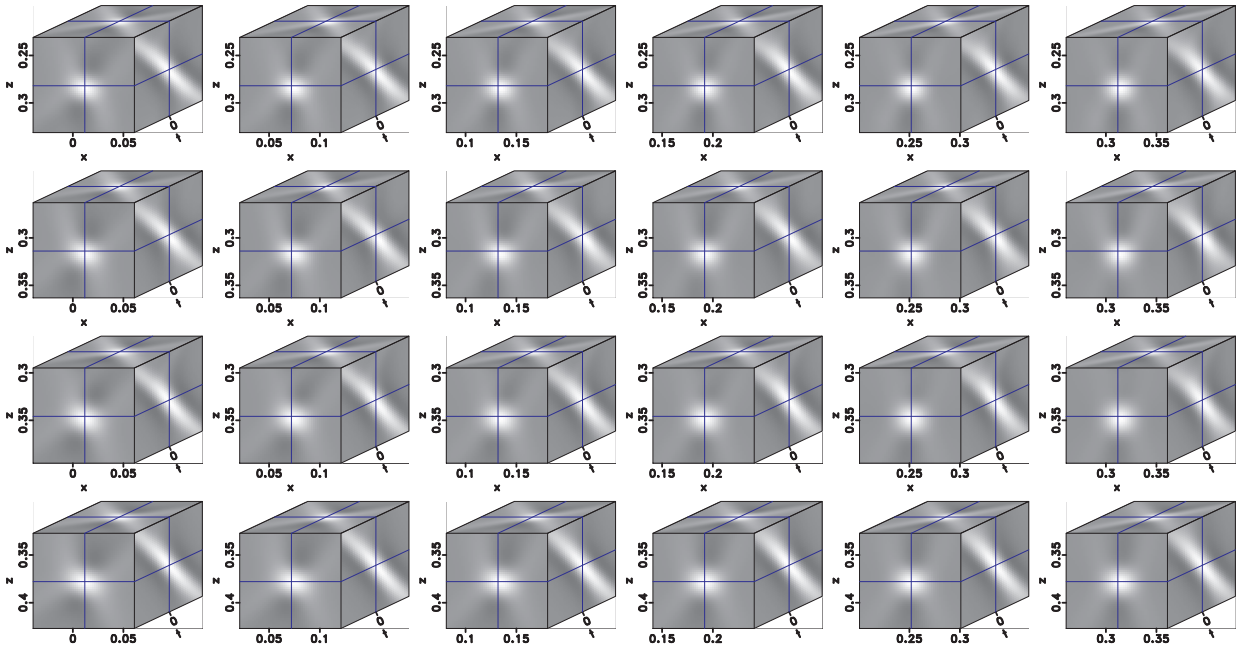


Figure 5: Seismic wavefield database. The center of each cube is the source location of the predicted data. The velocity model and acquisition used to build this database is shown in Figure 4.

Theoretical information on model and data $G(\mathbf{m})$
For all models within our selected region of space, we compute the corresponding wavefield by using the kernel function $G(\mathbf{m})$ which links theoretically individual members of the model and data spaces. This construction corresponds to the illustration in Figure 1(b). We assemble all those calculations in a database of wavefields which is used to compare with the observed data.

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 data for one onset time and then compare this simulated wavefield with the wavefields reconstructed from field seismographs at various times. Thus we can not only identify the location of the source, but we can also identify the onset time. Figure 5 illustrates the wavefield database at some locations in the subsurface indicated by dots in Figure 4, for fixed subsurface velocity and acquisition geometry. The size of the cubes in Figure 5 is two wavelengths in space and two periods in time.

Prior information on the data $\rho_{\mathfrak{D}}(\mathbf{d})$

We loop over the models provided by $\rho_{\mathfrak{M}}(\mathbf{m})$ and look for the times and locations in space where the observed data match the simulated data at locations where we assume that micro-seismic events are located. In this work, we use normalized cross-correlation (N_{cc}) to measure the similarity between the predicted data $G(\mathbf{m})$ and the observed data D^{obs}. By definition, we compute normalized cross-correlation by the expression

$$N_{cc}(\mathbf{m}) = \frac{\sum_{\mathbf{x}-M_{\mathbf{x}}}^{\mathbf{x}+M_{\mathbf{x}}}\sum_{t-M_t}^{t+M_t}(D^{obs}-\overline{D^{obs}})(G(\mathbf{m})-\overline{G(\mathbf{m})})}{\sqrt{\sum_{\mathbf{x}-M_{\mathbf{x}}}^{\mathbf{x}+M_{\mathbf{x}}}\sum_{t-M_t}^{t+M_t}(D^{obs}-\overline{D^{obs}})^2\sum_{\mathbf{x}-M_{\mathbf{x}}}^{\mathbf{x}+M_{\mathbf{x}}}\sum_{t-M_t}^{t+M_t}(G(\mathbf{m})-\overline{G(\mathbf{m})})^2}}, \tag{4}$$

where the overline denotes average values. The size of the averaging window is $(2M_{\mathbf{x}}+1)\times(2M_t+1)$, where $\mathbf{x}=\{x,y,z\}$. Values of N_{cc} are between -1 and $+1$. In terms of Bayesian inversion, this process of identification of the most probable model corresponds to the conjunction of the prior and theoretical probabilities, as indicated in Figures 1(c) and 1(d). The size of the window affects the value of $N_{cc}(\mathbf{m})$ and its identification power.

Instead of selecting a single model, which generates the highest N_{cc} value, as the answer, we construct a probability distribution characterizing various models and their likelihood to represent sources of micro-earthquakes. Based on equation 3, we assume that errors are introduced by the measurement uncertainty of the observed data and the error is distributed in a Gaussian distribution. The probability density function based on the normalized cross-correlation given by equation 4 is

$$\sigma_{\mathfrak{M}}(\mathbf{m}) = \rho_{\mathfrak{M}}(\mathbf{m})e^{-\frac{(1-N_{cc}(\mathbf{m}))^2}{2C_d}}, \tag{5}$$

where C_d is the measurement uncertainty. 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.

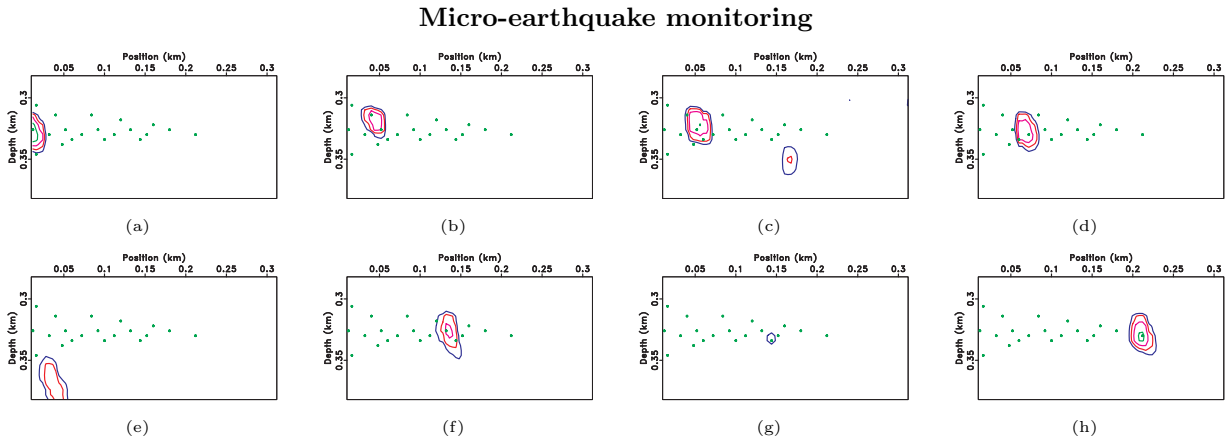


Figure 6: Snapshots of the probability distribution of micro-seismic events. The value of the outer contour is 0.8, the contour interval is 0.05. The dots represent the locations of micro-earthquakes.

Figure 6(a) shows the results 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 included in the contour corresponding to 0.95 confidence, indicating the success of our procedure. Figure 6(b) 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 our procedure may not have the resolution to distinguish among events that are located too close to one-another. The true location of the second events is indicated by the contours shown in Figure 6(c). However, this apparent mismatch is not a big problem because both regions indicate the presence of micro-earthquakes and they overlap in space and time. Other choice of parameters, e.g. in the definition of the wavefield database, might increase the spatial resolution of the identification. This example also seems to indicate 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 our simulation. Figure 6(d) 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.

Figure 6(e) shows that the event at $\mathbf{m}_{12} = \{0.112$ km, 0.33 km, 0.52 s$\}$ is not located within the contours obtained by our inversion. In this case, our 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. Figure 6(f) 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 6(g) 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 we successfully identify this event and the experiment seems to place it at the right position, we have lower confidence in its position. The lower confidence is likely due to the low SNR in the seismogram, which is approximately 0.1. Finally, Figure 6(h) 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. This is another example of successful identification of a seismic source.

Although in this example we do not locate all existing micro-earthquakes, we 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.

CONCLUSIONS

We present a method for automatic micro-earthquake location using Bayesian inversion theory. Our 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. Our method not only identifies the spatial location of the source, but also specifies the onset time of the source. Furthermore, since we are using a probabilistic approach to micro-earthquake location, our 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 when micro-earthquakes occur in close succession in a small region in space and are recorded in a noisy environment with a SNR as low as 0.1.

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 access to data, technical discussions and financial support of this project.