Estimation of passive microseismic event location using random sampling based curve fitting

Lijun Zhu, Yang Zhao†, Weichang Li†, Entao Liu*, James H. McClellan*, Zefeng Li‡, and Zhigang Peng‡*
* *Center for Energy and Geo Processing at Georgia Tech and King Fahd University of Petroleum and Minerals*
† *Aramco Services Company: Aramco Research Center—Houston*
‡ *School of Earth and Atmospheric Sciences, Georgia Institute of Technology*

SUMMARY

Characterization of microseismic activities has become a useful tool for both hydraulic fracturing and seismic hazard monitoring. The microseismic data received by a sensor array typically have relative weak signals in the presence of noise and interference. Estimation of microseismic event locations using such noisy data is always a major challenge of passive monitoring, especially for a surface array.

To overcome a large number of false picks due to background noise, a random sampling method is applied to classify picked arrival times into event and nonevent clusters. By isolating false picks into nonevent clusters, picks corresponding to seismic events are highlighted which can then be used to estimate the event location using a background velocity model. A synthetic study shows this method is robust even in the presence of significant lateral and vertical velocity variations using the Marmousi model. Field data of a 5200-element dense array in Long Beach, CA was utilized to evaluate the estimation of microseismic event locations.

INTRODUCTION

Locations of microseismic events provide valuable information for reservoir monitoring during hydraulic fracturing (Duncan, 2005). Identifying microseismic events where severe background noise is present is always a challenging task. Vesnaver et al. (2011) combined both surface and borehole arrays to mitigate the noise challenge. Even though microseismic events were successfully detected by this array setup, event localization was not feasible due to the overwhelming noise observed on both arrays (Vesnaver et al., 2012; Menanno et al., 2013).

Since the velocity model is generally not perfectly known during microseismic monitoring, travel-time based location methods are applied under an initial effective homogeneous medium assumption (Grechka and Zhao, 2012). Blias and Grechka (2013) gives an analytic solution to jointly estimate the event locations and an effective velocity model for microseismic applications by fitting a parabola curve to travel time differences between P-wave and S-wave arrivals. When only a P-wave phase is available, a hyperbola can be used instead to model the moveout curve on the monitoring array (Dix, 1955). However, due to background noise, a large number of false picks will contaminate picked arrival times and make such curve fitting results unreliable.

To overcome this challenge and eliminate false picks, Zhu et al. (2016) proposed Random-sampling-based Arrival Time Event Clustering (RATEC) — a method to find the most reliable moveout curve. The moveout curve is isolated from false picks by repeatedly trying different hypothesized curves and finding the best curve that agrees with the largest number of picked arrival times. In this study, we show successful application of RATEC on a dense surface array setup for earthquake monitoring (Zhu et al., 2017), as well as synthetic examples with P- and S-wave phases and complex velocity models.

METHOD

RANdom SAmpling Consensus (RANSAC)

Random sampling consensus was first proposed by Fischler and Bolles (1981) and then improved by many others (Stewart, 1995; Torr and Zisserman, 2000; Chum and Matas, 2002; Tordoff and Murray, 2005; Chum and Matas, 2005). Despite many variations and adaptations of random sampling (Choi et al., 2009), there are essentially two steps per iteration which are repeated many times to determine the best fit to the data:

- *Hypothesize*: A **minimal** sample subset (MinSet, denoted as Ω_{M}^k) is randomly selected from the dataset and the **unique** model parameters ($\mathbf{p}^k$) are computed for Ω_{M}^k.
- *Test*: Elements in the dataset (Ω_{D}) are evaluated to determine which ones can be labeled as *inliers*, i.e., consistent with the hypothesized model in the sense that the distance from the model's moveout curve is less than some prescribed value (δ). The set of all such inliers is called a consensus set (ConSet, denoted as Ω_{C}^k).

Note that $\Omega_{\mathrm{M}}^k \subset \Omega_{\mathrm{C}}^k \subset \Omega_{\mathrm{D}}$. A set Ω_{M}^k consists of only the minimal number of samples required to uniquely determine a model, e.g., two samples for a line and three for a circle. The more elements in Ω_{C}^k, the better the model we have obtained for the kth hypothesis. Based on the fitted hyperbolic model, the picked arrival times are, in effect, clustered into event groups and nonevent groups. Such clustering not only separates picked arrival times into different phases (e.g., P-wave and S-wave phases), but also improves the accuracy of localization results by eliminating false picks due to noise.

Each RANSAC iteration requires very little computation and there exists a unique solution for each chosen MinSet. In this way, we can afford to use a large number of iterations to discover a consistent model. The number of iterations $\hat{N}$ to guarantee that at least one Ω_{M} will only contain true picks with probability p is

$$\hat{N} = \frac{\log(1-p)}{\log(1-u^m)}, \tag{1}$$

where u is the fraction of all picks that are close to a hyperbola, and m is the size of a MinSet, which is 5 for a hyperbola and

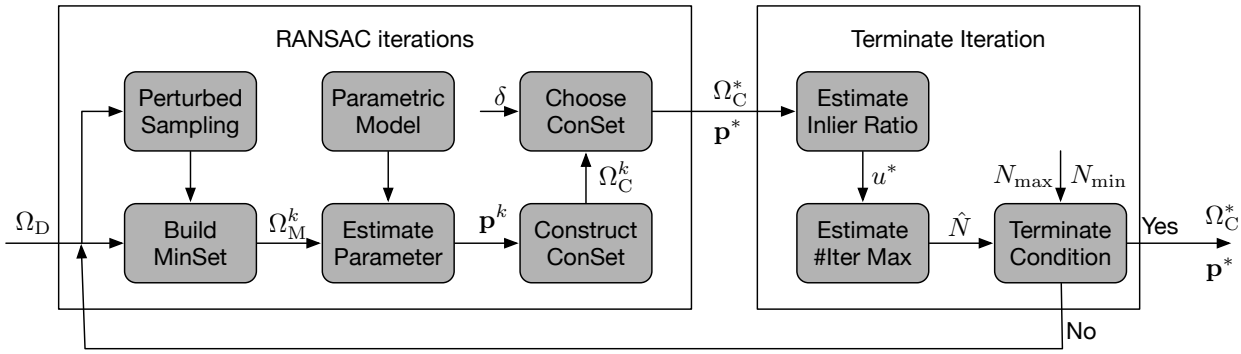

Figure 1: Complete workflow of proposed random sampling scheme.

9 for a hyperboloid. For example, when $u = 0.5$ and $m = 5$, we can guarantee a 99% chance of finding an outlier-free Ω_{M} ($p = 0.99$), if we run $\hat{N} = 145$ iterations.

The overall process of RATEC is summarized in a flow chart in Figure 1. After the kth iteration, the current best ConSet Ω_{C}^* is updated with the kth ConSet if Ω_{C}^k has more inliers. Then Ω_{C}^* is used to estimate the current best inlier ratio u^*. Based on Equation (1), the number of iterations required, $\hat{N}$, can be updated (Tordoff and Murray, 2005). The current $\hat{N}$ is also compared against preset minimum and maximum values $N_{\min}$ and $N_{\max}$. Once the termination condition is satisfied, the best ConSet Ω_{C}^* and model parameter $\mathbf{p}^*$ are returned; otherwise, the iteration loop will continue.

Parameter estimation for moveout curve

The proposed method uses a quadratic model to estimate the parameters of a hyperbolic curve, which takes the following form:

$$\mathscr{P}(x,y;\mathbf{p}) = ax^2 + bxy + cy^2 + dx + ey + f = 0, \qquad (2)$$

where $\mathbf{p}$ has six real elements $(a,\cdots,f)$. There are actually only five free parameters, since one of the nonzero elements can be always normalized to 1. When the determinant

$$\Delta_1 = 4\begin{vmatrix} a & b/2 & d/2 \\ b/2 & c & e/2 \\ d/2 & e/2 & f \end{vmatrix} \qquad (3)$$

is nonzero ($\Delta_1 \neq 0$), Equation (2) defines a non-degenerate conic section. To verify that it is a hyperbola, we must also check a second determinant

$$\Delta_2 = 4\begin{vmatrix} a & b/2 \\ b/2 & c \end{vmatrix} = b^2 - 4ac. \qquad (4)$$

When $\Delta_2 > 0$, Equation (2) defines a hyperbola.

Given n arrival time picks, (x_i, y_i) for $i = 1,\ldots,n$, we form a $n\times 6$ data matrix $\mathbf{D}_n$ and a 6×1 coefficient vector $\mathbf{p}$, such that $\mathbf{D}_n\mathbf{p}$ is the model $\mathscr{P}(x,y;\mathbf{p})$ evaluated at the time picks. With measurement error, there is a nonzero residual $\mathbf{r}$, i.e.,

$$\begin{bmatrix} x_1^2 & x_1y_1 & y_1^2 & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n^2 & x_ny_n & y_n^2 & x_n & y_n & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \\ e \\ f \end{bmatrix} = \begin{bmatrix} r_1 \\ \vdots \\ r_n \end{bmatrix} \qquad (5)$$
$$\Leftrightarrow \quad \mathbf{D}_n\mathbf{p} = \mathbf{r}.$$

If all picks in Ω_{M} are true picks, the residual term $\mathbf{r}$ is negligible. Then we solve the linear system $\mathbf{D}_5\mathbf{p} = \mathbf{0}$, which finds the null space of $\mathbf{D}_5$. From the singular value decomposition (SVD) of $\mathbf{D}_5$, it is easy to see that the last right singular vector $\mathbf{v}_6 \in \mathrm{null}(\mathbf{D}_5)$. To further stabilize the algorithm, a random perturbation is added to y_i when MinSets are formed.

Extension to 2-D surface arrays

This fitting method can be easily extended to surface arrays by changing the underlying hyperbolic curve model to a hyperboloid surface model. Similar to Equation (2), a hyperboloid surface can be defined using a quadratic equation in (x,y,z) that takes the following general form:

$$\mathscr{P}(x,y,z;\mathbf{p}) = [x \quad y \quad z]\begin{bmatrix} a & b/2 & d/2 \\ b/2 & c & e/2 \\ d/2 & e/2 & g \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} + [g \quad h \quad i]\begin{bmatrix} x \\ y \\ z \end{bmatrix} + j = 0 \qquad (6)$$

Using (6), the RANSAC framework can be adapted to hyperboloid surface fitting by finding the parameter vector $\mathbf{p} = [a,b,c,d,e,f,g,h,i,j]$ in a 10-dimensional space. Although this is a somewhat larger parameter space for the MinSets, it adds very little burden to the search process because RATEC still searches the n time picks using models formed from the randomly chosen MinSets which now have 9 elements.

Event location through moveout curve

Various location methods based on travel time can be used on the time picks in the event cluster after RATEC is applied. Below, we offer one simple location estimation by refitting the moveout curves based on an effective homogeneous media assumption. We choose the trust region method (Berghen, 2004) over the least squares method for refitting due to its robustness in the presence of general background noise. Unpacking Equation (2), the event location can be recovered from the hyperbola coefficients as follows:

$$x_0 = -\frac{d}{2a} \quad \text{and} \quad z_0 = \sqrt{\frac{f}{a} - \frac{d^2}{4a^2} - \frac{ce^2}{4ac^2}}\,. \qquad (7)$$

where x_0 is the event horizontal offset and z_0 is the event depth.

SYNTHETIC DATA EXAMPLE

Recorded seismic trace in layered model

Recorded seismic traces from borehole sensors are used in this example. The P- and S-wave phases are manually picked and then delayed according to the travel time calculated from a layered velocity model. The layered velocity model used in this example, shown in Figure 2a, is taken from the Marmousi2 elastic velocity model (Martin et al. 2006). The top water layer in Marmousi2 is removed and the event source is located at a depth of 2.5 km. A 25-receiver nonuniform surface array is used on top of the medium for monitoring underground seismic events occurring below the array center (at 2500 m).

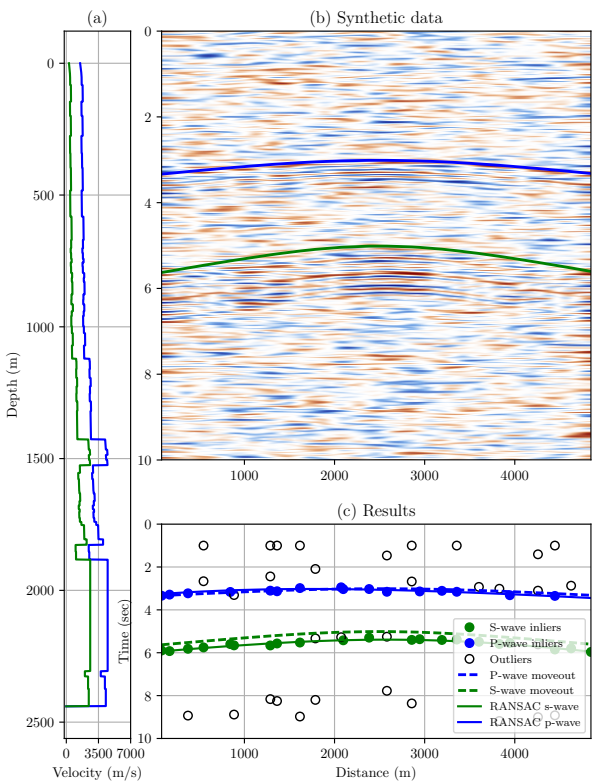

Figure 2: Layered velocity model example using recorded seismic trace with P-wave and S-wave phases: (a) 1-D velocity model from Marmousi2; (b) noisy raw data with PSNR = 10 dB with respect to the S-wave peak after a 20 Hz low-pass filter; (c) fitted moveout curve and classification results comparing to true P-wave and S-wave moveout curves.

Since the dominant frequency of the arrival event is estimated to be 10 Hz from the spectrogram, a low-pass filter with cut-off frequency at 20 Hz is used as pre-processing. Both P-wave and S-wave arrivals are observed in Figure 2b with low-pass filtering. Figure 2c shows the result of applying the RATEC method, where moveout curves were generated by fitting the classified and corrected arrival time picks. RATEC is used repeatedly in this example to extract all possible event phases: after one moveout curve is detected and identified, its outliers are used as the input for the next iteration to search for more curves until there are not enough time picks to successfully define a moveout curve. Here, two iterations are sufficient to identify both P-wave (blue) and S-wave (green) phases with most of the true arrival times labeled correctly.

Ricker wavelet in Marmousi model

Although the proposed method is based on a homogeneous initial velocity model assumption, it is robust enough to handle more complex models. In Figure 3a, the acoustic Marmousi model is used to study the proposed method when there are horizontal variations in the velocity model. A finite-difference time-domain numerical simulation is used to generate the noisy receiver data shown in Figure 3b with 10 dB PSNR. Since each trace has a different peak value, which is common in real seismic data, the PSNR defined here uses the global peak of all the traces. A 25-receiver surface line array with 20 m spacing is used to monitor an event at $x = 400$ and $z = 1000$ m indicated by the red arrow in Figure 3a.

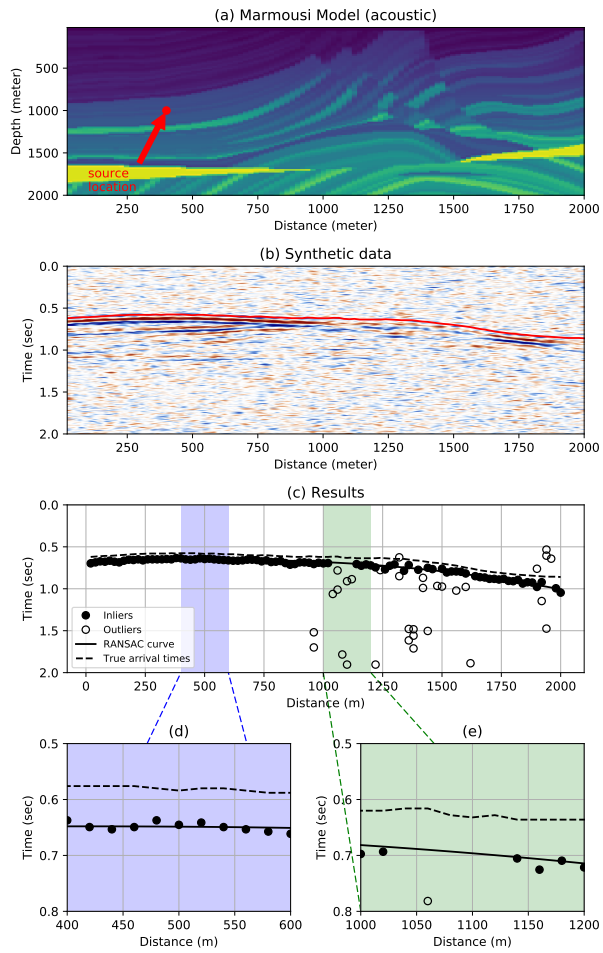

Figure 3: Marmousi model example under 10 dB PSNR: (a) velocity model (red dot indicates source location), (b) synthetic data (red line indicates true arrival times), (c) RATEC results, (d) zoomed-in results between 400 m and 600 m and (e) zoomed-in results between 1000 m and 1200 m.

After applying RATEC, Figure 3c shows the results of moveout curve prediction and arrival time labels. Even though the

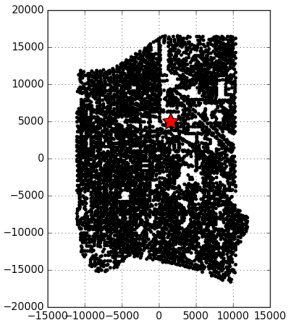

Figure 4: Top view of the sensor array with located event indicated by red star.

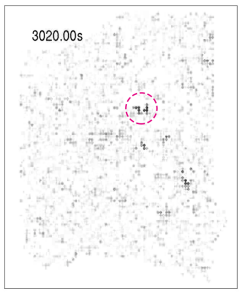

Figure 5: Snapshot of the seismic dataset at time $t = 3020.00\,\mathrm{s}$; the visible event lies inside the red circle.

true moveout is not exactly a hyperbola, RATEC is able to label all the true arrival times within a small distance of the moveout curve. Zooming in around the horizontally layered region, good prediction and perfect labeling are observed in Figure 3d. Notice that there is a consistent offset between picked and true arrival times due to the delay of STA/LTA pickings (Zhao et al., 2008) used here. Figure 3e shows the results in a more complex region where the SNR is worse. Even though many picks in that region are false picks, RATEC successfully eliminates most of the picks far away from the true moveout curve and labels the true time picks correctly.

FIELD DATA EXAMPLE

The proposed method was also tested on a data set of 50 sec collected by the Long Beach nodal dense array (Inbal et al., 2016) in southern California which contains 5200 sensors. The top view of the sensor array is shown in Figure 4. Prior to applying the RATEC scheme, no reliable location estimate can be obtained from the picked arrival times due to a large number of false picks which are visible in Figure 6. Instead of STA/LTA, local similarity (Li and Peng, 2016) is used to improve the time picking accuracy based on the stacked waveform similarity between adjacent traces in a small area.

Using the clustered picks found by RATEC, this seismic event is recognized as a surface event whose location is shown by marking its epicenter with the red star in Figure 4. In order to verify our result, we schematically show the corresponding clipped data amplitude snapshot on the sensor array in Figure 5. The gray-scale of the dots indicates the clipped signal amplitude on the corresponding sensor. The red circle in Figure 5 confirms that in the inverted time and location using the classified true picks, there is indeed a weak event that is barely visible in the raw array data. Moreover, the work log shows that there is a surface source in the estimated area but the local earthquake catalog has no record of earthquakes during the event time. In Figure 6 we show the time picking results that contain a large number of false picks. The best-fitted hyperboloid surface from 3-D RATEC is shown as the red surface. On a laptop, RATEC takes just 31 sec to process 50 sec of data

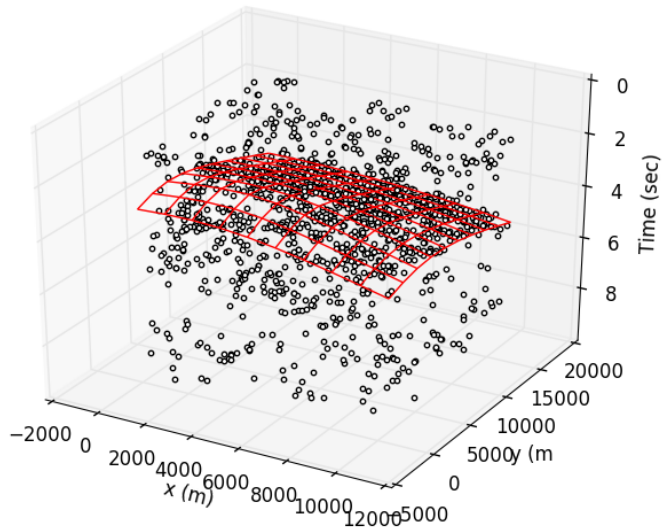

Figure 6: 3-D view of the noisy picks (○) from the 2-D sensor array. Fitted surface obtained via RATEC highlighted in red.

which is sufficient for real-time processing.

CONCLUSION

In this paper, we reformulated the arrival time picking problem into a model fitting framework and extended the RANSAC-based curve fitting and surface fitting methods to classify picked arrival times. The effectiveness and robustness of the proposed method are validated by tests on both synthetic and real data sets. The localization result is significantly improved when RATEC eliminates false picks from the data set. In the 2D surface array example on real microseismic data, an accurate hypocenter is successfully inverted.

ACKNOWLEDGEMENTS

The seismic data analyzed in this study are owned by Signal Hill Petroleum, Inc. and acquired by NodalSeismic LLC. We thank NodalSeismic LLC for making the one-week data available in this study. This work is supported by the Center for Energy and Geo Processing (CeGP) at Georgia Tech and by King Fahd University of Petroleum and Minerals. Zefeng Li and Zhigang Peng are supported by National Science Foundation grant EAR-1551022.

EDITED REFERENCES

Note: This reference list is a copyedited version of the reference list submitted by the author. Reference lists for the 2017 SEG Technical Program Expanded Abstracts have been copyedited so that references provided with the online metadata for each paper will achieve a high degree of linking to cited sources that appear on the Web.

REFERENCES

Berghen, F. V., 2004, Condor: A constrained, non-linear, derivative-free parallel optimizer for continuous, high computing load, noisy objective functions: These de doctorat, Université Libre de Bruxelles.

Blias, E., and V. Grechka, 2013, Analytic solutions to the joint estimation of microseismic event locations and effective velocity model: Geophysics, **78**, no. 3, KS51–KS61, https://doi.org/10.1190/geo2012-0517.1.

Choi, S., T. Kim, and W. Yu, 2009, Performance evaluation of RANSAC family: Journal of Computer Vision, **24**, 271–300, https://doi.org/10.5244/C.23.81.

Chum, O., and J. Matas, 2002, Randomized RANSAC with T(d, d) test: Proceedings of the British Machine Vision Conference, 448–457, https://doi.org/10.5244/C.16.43.

Chum, O., and J. Matas, 2005, Matching with PROSAC-progressive sample consensus: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), IEEE, 220–226, https://doi.org/10.1109/CVPR.2005.221.

Dix, C. H., 1955, Seismic velocities from surface measurements: Geophysics, 20, 68–86, https://doi.org/10.1190/1.1438126.

Duncan, P. M., 2005, Is there a future for passive seismic?: First Break, **23**.

Fischler, M. A., and R. C. Bolles, 1981, Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography: Communications of the ACM, **24**, 381–395, https://doi.org/10.1145/358669.358692.

Grechka, V., and Y. Zhao, 2012, Microseismic interferometry: The Leading Edge, **31**, 1478–1483, https://doi.org/10.1190/tle31121478.1.

Inbal, A., J. P. Ampuero, and R. W. Clayton, 2016, Localized seismic deformation in the upper mantle revealed by dense seismic arrays: Science, **354**, 88–92, https://doi.org/10.1126/science.aaf1370.

Li, Z., and Z. Peng, 2016, Automatic detection and classification of seismic events: Seismological Research Letters, **87**, 501.

Martin, G. S., R. Wiley, and K. J. Marfurt, 2006, Marmousi2: An elastic upgrade for Marmousi: The Leading Edge, **25**, 156–166, https://doi.org/10.1190/1.2172306.

Menanno, G., A. Vesnaver, and M. Jervis, 2013, Borehole receiver orientation using a 3D velocity model: Geophysical Prospecting, **61**, 215–230, https://doi.org/10.1111/j.1365-2478.2012.01106.x.

Stewart, C. V., 1995, MINPRAN: A new robust estimator for computer vision: IEEE Transactions on Pattern Analysis and Machine Intelligence, **17**, 925–938, https://doi.org/10.1109/34.464558.

Tordoff, B. J., and D. W. Murray, 2005, Guided-MLESAC: Faster image transform estimation by using matching priors: IEEE Transactions on Pattern Analysis and Machine Intelligence, **27**, 1523–1535, https://doi.org/10.1109/TPAMI.2005.199.

Torr, P. H., and A. Zisserman, 2000, MLESAC: A new robust estimator with application to estimating image geometry: Computer Vision and Image Understanding, 78, 138–156, https://doi.org/10.1006/cviu.1999.0832.

Vesnaver, A., G. Menanno, S. I. Kaka, and M. Jervis, 2011, 3D polarization analysis of surface and borehole microseismic data: 81th Annual International Meeting, SEG, Expanded Abstracts, 1663–1668, https://doi.org/10.1190/1.3627523.

Vesnaver, A., G. M. Menanno, and M. Jervis, 2012, Accuracy analysis for the hypocenter estimate in a microseismic survey: Presented at the GEO 2012, EAGE.

Zhao, Y., C. Tang, and J. A. Rial, 2008, A new approach for the automatic detection of shear-wave splitting: M.S. thesis, The University Of North Carolina At Chapel Hill.

Zhu, L., Z. Li, Z. Peng, E. Liu, and J. H. McClellan, 2017, Weighted random sampling in seismic event detection/location (WRASED): Applications to local, regional, and global seismic networks: Annual Meeting Seismological Society of America, Seismological Research Letters, 88.

Zhu, L., E. Liu, and J. H. McClellan, 2016, An automatic arrival time picking method based on RANSAC curve fitting: 78th Annual International Conference and Exhibition, EAGE, Extended Abstracts, https://doi.org/10.3997/2214-4609.201601481.