Supporting Information for
"Seismicity During the Initial Stages of the Guy-Greenbrier, Arkansas, Earthquake Sequence"

Clara E. Yoon[1], Yihe Huang[1,2], William L. Ellsworth[1], and Gregory C. Beroza[1]

[1]Department of Geophysics, Stanford University, Stanford, California, USA.

[2]Department of Earth and Environmental Sciences, University of Michigan, Ann Arbor, Michigan, USA.

Contents

1. Sections S1 to S5
2. Figures S1 to S12
3. Tables S1 to S8

Additional Supporting Information (Files uploaded separately)

1. Captions for Datasets S1 to S3
2. Captions for Movies S1 to S1

Corresponding author: C. E. Yoon, `ceyoon@stanford.edu`

Section S1: Magnitude estimation

We report local magnitude M_L for all detected events in this study. To calibrate M_L, we first calculate the moment magnitude M_w for a selected group of 54 events with high-quality waveforms, located at different distances from station WHAR, within event clusters 1, 5, 6, and 7 (Figure S3a). To obtain M_w, we calculate seismic moment M_0 in the time domain for P, SH, and SV arrivals on displacement waveforms at WHAR (after removal of instrument response and rotation of the horizontal components to radial and transverse), following *Prejean and Ellsworth* [2001]:

$$M_0 = \frac{4\pi\rho v^3 r\left(\int u\,dt\right)}{2F} \tag{1}$$

where $\rho = 2.5$ kg/m^3 is density, v is velocity (4.6 km/s for P and 2.7 km/s for S waves, estimated using travel times and hypocentral distances for these 54 events), and r is hypocentral distance. $\left(\int u\,dt\right)$ is the integrated area under the displacement pulse for the P, SH, or SV arrival. F is the radiation pattern correction term, which we approximate using the average radiation pattern term $<F> = 0.52$, 0.41, and 0.48 for P, SH, and SV, respectively. Equation 1 has an extra factor of 2 in the denominator compared to *Prejean and Ellsworth* [2001] to approximately account for the amplitude at the free surface. To calculate M_0, we use only arrivals with high-quality displacement pulses: P and SH for Cluster 1 events (taking the average M_0), SV for Cluster 5 and 6 events, and SH for Cluster 7 events. We then calculate moment magnitude M_w from M_0 [*Bormann*, 2012]:

$$M_w = \frac{2}{3}\log_{10} M_0 - 6.07 \tag{2}$$

Next, we solve for the local magnitude distance correction using these 54 calibration events. We assume that the magnitude can be expressed as:

$$\begin{aligned} M_w &= \log_{10}\left(A_{\text{peak}} R^k\right) + C \\ \Rightarrow M_w - \log_{10} A_{\text{peak}} &= k \log_{10} R + C, \end{aligned} \tag{3}$$

with $\log_{10} A_{\text{peak}}$ computed as:

$$\log_{10} A_{\text{peak}} = \frac{1}{2}\left(\log_{10} A_{\text{peak,East}} + \log_{10} A_{\text{peak,North}}\right), \tag{4}$$

where $A_{\text{peak,East}}$ and $A_{\text{peak,North}}$ are peak Wood-Anderson seismogram amplitudes, after applying a 1 Hz high-pass filter to remove low-frequency noise, from the east and north components, respectively [*Bormann*, 2012]. In Equation 3, R is epicentral distance, and the distance correction parameters to estimate are k (representing the effect of geometric spreading and attenuation) and C (constant for the base level, where we solve for a separate C for each station). Since we can obtain A_{peak} from Equation 4 at each station, with the objective of estimating 4 distance correction parameters

k, C_{WHAR}, C_{ARK2}, and C_{ARK1}, we express Equation 3 in matrix form (**d = Gm**), where the design matrix **G** has dimensions 54 events x 3 stations = 162 rows, by 4 columns, and i is the event index:

$$\begin{bmatrix} M_w[i=1] - \log_{10}\left(A_{\mathrm{peak,WHAR}}[i=1]\right) \\ M_w[i=2] - \log_{10}\left(A_{\mathrm{peak,WHAR}}[i=2]\right) \\ \vdots \\ M_w[i=54] - \log_{10}\left(A_{\mathrm{peak,WHAR}}[i=54]\right) \\ \hline M_w[i=1] - \log_{10}\left(A_{\mathrm{peak,ARK2}}[i=1]\right) \\ M_w[i=2] - \log_{10}\left(A_{\mathrm{peak,ARK2}}[i=2]\right) \\ \vdots \\ M_w[i=54] - \log_{10}\left(A_{\mathrm{peak,ARK2}}[i=54]\right) \\ \hline M_w[i=1] - \log_{10}\left(A_{\mathrm{peak,ARK1}}[i=1]\right) \\ M_w[i=2] - \log_{10}\left(A_{\mathrm{peak,ARK1}}[i=2]\right) \\ \vdots \\ M_w[i=54] - \log_{10}\left(A_{\mathrm{peak,ARK1}}[i=54]\right) \end{bmatrix} = \begin{bmatrix} \log_{10}(R_{\mathrm{WHAR}}[i=1]) & 1 & 0 & 0 \\ \log_{10}(R_{\mathrm{WHAR}}[i=2]) & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \log_{10}(R_{\mathrm{WHAR}}[i=54]) & 1 & 0 & 0 \\ \hline \log_{10}(R_{\mathrm{ARK2}}[i=1]) & 0 & 1 & 0 \\ \log_{10}(R_{\mathrm{ARK2}}[i=2]) & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \log_{10}(R_{\mathrm{ARK2}}[i=54]) & 0 & 1 & 0 \\ \hline \log_{10}(R_{\mathrm{ARK1}}[i=1]) & 0 & 0 & 1 \\ \log_{10}(R_{\mathrm{ARK1}}[i=2]) & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ \log_{10}(R_{\mathrm{ARK1}}[i=54]) & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} k \\ C_{\mathrm{WHAR}} \\ C_{\mathrm{ARK2}} \\ C_{\mathrm{ARK1}} \end{bmatrix} \tag{5}$$

Inverting for the best-fit distance correction parameters in a least-squares sense, we get k = 1.5273, C_{WHAR} = 1.7141, C_{ARK2} = 0.4447, and C_{ARK1} = 0.5782. Plugging in these parameters into Equation 3, and assuming $M_w = M_L$, we calculate local magnitude at each station for every detected event, given peak amplitudes from 1 Hz high-pass filtered Wood-Anderson seismograms (Equation 4) and epicentral distances:

$$\begin{aligned} M_{\mathrm{L,WHAR}} &= \log_{10} A_{\mathrm{peak,WHAR}} + k \log_{10} R_{\mathrm{WHAR}} + C_{\mathrm{WHAR}} \\ \Rightarrow M_{\mathrm{L,WHAR}} &= \log_{10} A_{\mathrm{peak,WHAR}} + 1.5273 \log_{10} R_{\mathrm{WHAR}} + 1.7141 \end{aligned} \tag{6}$$

$$\begin{aligned} M_{\mathrm{L,ARK2}} &= \log_{10} A_{\mathrm{peak,ARK2}} + k \log_{10} R_{\mathrm{ARK2}} + C_{\mathrm{ARK2}} \\ \Rightarrow M_{\mathrm{L,ARK2}} &= \log_{10} A_{\mathrm{peak,ARK2}} + 1.5273 \log_{10} R_{\mathrm{ARK2}} + 0.4447 \end{aligned} \tag{7}$$

$$\begin{aligned} M_{\mathrm{L,ARK1}} &= \log_{10} A_{\mathrm{peak,ARK1}} + k \log_{10} R_{\mathrm{ARK1}} + C_{\mathrm{ARK1}} \\ \Rightarrow M_{\mathrm{L,ARK1}} &= \log_{10} A_{\mathrm{peak,ARK1}} + 1.5273 \log_{10} R_{\mathrm{ARK1}} + 0.5782 \end{aligned} \tag{8}$$

For the 1,740 located events (Figure 6, blue), we report local magnitude M_L as the average of M_L estimates at the 3 stations:

$$M_L = \frac{1}{3}\left(M_{\mathrm{L,WHAR}} + M_{\mathrm{L,ARK2}} + M_{\mathrm{L,ARK1}}\right) \tag{9}$$

Figure S3b shows that the M_L values from Equation 9 are reasonably similar to M_w estimates for the 54 calibration events, although they sometimes deviate by as much as 0.7 magnitude units.

For the 6,508 assigned events (Figure 6, black) and 6,356 unassigned events (Figure 6, red), we report local magnitude M_L as equal to $M_{\mathrm{L,WHAR}}$ instead of using the average from Equation 9, because waveforms at stations ARK1 and ARK2 are unreliably noisy for these smaller events. For the assigned events, we calculate the epicentral distance R_{WHAR} using the average latitude and longitude from all located events in that cluster. For the unassigned events, we pick P and S phases on the Wood-Anderson seismograms at WHAR, compute the S-P time, and estimate R_{WHAR} as:

$$R_{\mathrm{WHAR}} = \frac{v_p v_s}{v_p - v_s} \left(t_s - t_p \right) \tag{10}$$

using the same velocities from Equation 2: v_p = 4.6 km/s, v_s = 2.7 km/s.

Section S2: Phase picking procedure

We automatically pick P- and S-wave arrivals on the 3 stations (ARK1, ARK2, WHAR) with the Akaike Information Criteria (AIC) picker [*Maeda*, 1985], manually adjust them as needed, and assign integer weights for pick quality ranging from 0 (best) to 3 (worst). We pick P phases on the vertical component and S phases on the horizontal components. We estimate the origin time of each event using the S-P time on the station with the earliest arrivals, assuming a V_p/V_s ratio of $\sqrt{3}$.

Section S3: VELEST inversion for quarry-constrained velocity model

The layer boundaries in the *Ogwari et al.* [2016] starting velocity model remain the same, but we allow V_p and V_s to change in each layer. We input the notch location (35.2928° N, 92.3973° W, 0 km depth), estimated origin times, and travel times for the 3 quarry blasts as shot data (with fixed coordinate locations), repeated 8 times, for a total of *nshot* = 24 shots, which balances the relative number of shots compared to the number of earthquakes used in the velocity model inversion. We do not fix the origin time of the shots (*nshfix* = 0), do not apply a shot correction (*nshcor* = 0), and do not invert for station corrections (*nsinv* = 0). To ensure a robust result, we run VELEST twice, each time using a different subset of 50 earthquakes with well-distributed locations, and confirm that we get a consistent velocity model for each case. We jointly invert for the velocity model and earthquake locations every *invertratio* = 3 iterations until we reach *ittmax* = 50 iterations, using damping parameters *othet* = *xythet* = *zthet* = 0.03 and *vthet* = 10.

Section S4: Differential travel time calculation for hypoDD

We first compute differential travel times from both catalog P and S picks and cross-correlation-based picks for the 1,229 events where we already have initial absolute locations from VELEST. The P

and *S* picks, originally determined for the VELEST locations, are input into the *ph2dt* program with parameters from Table S2 to get catalog differential times. The VELEST to *ph2dt* pick weight conversion is: 0 to 1 (best), 1 to 0.75, 2 to 0.5, 3 to 0.25 (worst). To get cross-correlation differential times, we cross-correlate short windows around the *P* and *S* arrivals for every event pair separated by less than 2 km within the 1,229 located earthquakes at all 3 stations, keeping only differential times with CC ≥ 0.6. We use a 0.9 s window around the *P* arrival, 0.35 s before and 0.55 s after, on the vertical component. We use a longer 1.5 s window around the *S* arrival, 0.55 s before and 0.95 s after, on both horizontal components, and keep the differential time from the component that results in a higher CC. These time windows are tapered and filtered 2-20 Hz before cross-correlation in the time domain, where we achieve subsample precision of the differential time by fitting a parabola near the peak of the cross-correlation function and interpolating the time at the peak [e.g. *Deichmann et al.*, 1992; *Schaff et al.*, 2004].

We also compute cross-correlation differential times between each of the 1,229 initially located events and the 13,375 remaining unlocated events, with the goal of locating as many of the remaining events as possible. For the unlocated events, we automatically pick *P* phases on the vertical component and *S* phases on the horizontal components with the AIC picker. We cross-correlate short windows around the *P* and *S* arrivals at all 3 stations for each pair containing a located and an unlocated event, keeping only differential times shorter than 0.5 seconds with CC ≥ 0.6. For each unlocated event, we set its starting location as the initial VELEST location of the already-located event with the highest cross-correlation. We use the same window lengths, filter bands, and time-domain cross-correlation with subsample precision mentioned previously.

Section S5: Assignment of unlocated events to existing clusters

We first generate a representative stack waveform at station WHAR for each event cluster 1-16 (Table S5) by stacking all located event waveforms at WHAR belonging to that cluster. Since all events within a cluster are close together relative to their distance to WHAR, their waveforms are reasonably similar. We designate the highest magnitude event in the cluster as the master event. For each located event in the cluster (including the master event), we take a time window of length 8 seconds, starting 1.5 seconds before the *P*-wave, on all 3 components at station WHAR; apply a 1-20 Hz bandpass filter; and normalize each waveform by dividing by its L2 norm. For each non-master event in the cluster, we cross-correlate its normalized waveform with the normalized master event waveform, and align the waveform to the master event using the delay time associated with the peak CC averaged over 3 components. We stack all aligned waveforms, including the master event, and divide by the total number of events to get the representative stack waveform. Figure S6 shows these stack waveforms for the 11 clusters where the number of sub-clusters is 1 (Table S5).

Clusters 1-16 are defined by spatially compact regions, but sometimes waveforms at WHAR within a cluster exhibit significant variability, so combining all aligned waveforms results in a stack waveform that is not similar to most event waveforms in the cluster. In this case, we apply hierarchical clustering with average linkage to group events with similar waveforms together into sub-clusters (e.g. *Harris* [2006], *Leskovec et al.* [2014]). We first compute the total 3-component CC between each pair of located events in the cluster, to use as a similarity metric. We take an agglomerative approach to determine sub-clusters: we start with each event in its own sub-cluster, then we combine sub-clusters by grouping the most similar (highest 3-component CC) events together, then repeat the grouping step using the average CC between sub-clusters, until we reach a minimum CC cutoff. Events within a sub-cluster have more similar waveforms (above the minimum CC cutoff), compared to events from different sub-clusters. Clusters 3, 4, 5, 10, 13 have 5, 3, 3, 9, and 11 sub-clusters, respectively (Table S5). The number of sub-clusters depends on the minimum CC cutoff used in hierarchical clustering (Table S6), which we set empirically to get a relatively small number of sub-clusters. For the 5 clusters with multiple sub-clusters, the stack waveforms for sub-clusters within the same cluster exhibit obvious differences (Figures S7 to S11).

We cross-correlate each unlocated event waveform at WHAR on all 3 components with the representative stack waveform from each cluster (or each sub-cluster), and assign the unlocated event to the cluster where we get the highest CC between the unlocated event waveform and stack waveform, provided that the CC exceeds an empirically determined cluster-specific threshold listed as "Minimum CC threshold to assign" in Table S5. Figure S12 explains how to set this threshold to assign events to a cluster, for the example of Cluster 1 where the threshold is CC $\geq$ 0.5 (Table S5). We first calculate a CC distribution by cross-correlating the cluster stack waveform with each located event waveform belonging to the same cluster (Figure S12a) because we want the CC for these events in the same cluster to exceed the threshold. We then calculate a CC distribution between the cluster stack waveform and each located waveform belonging to a different cluster or sub-cluster. Since we have a total of 42 cluster or sub-cluster stack waveforms (Table S5; Figures S6 to S11), we compute 41 such distributions for every cluster. Figure S12b shows one out of these 41 CC distributions, between the Cluster 1 stack waveform and all located waveforms from Cluster 2; we want the CC for these events in a different cluster to remain below the threshold. We repeat this entire process for the stack waveform from the 41 remaining clusters or sub-clusters, obtaining a cluster-specific threshold that separates events belonging to the cluster from events belonging to other clusters. Once again, we compute the average CC over all 3 components of station WHAR, using a time window of length 8 s, starting 1.5 s before the *P*-wave, filtered 1-20 Hz, and normalized by the L2 norm of the waveform.

Many previous studies have applied a CC threshold to group similar event waveforms into clusters, also called "multiplets" or "families". The choice of CC threshold depends on the time window length, frequency band, number of stations, noise content, and source characteristics of the event waveform. Most studies in induced seismicity [*Schultz et al.*, 2015a, 2016], volcano seismology [*Rowe et al.*, 2004; *Green and Neuberg*, 2006; *Petersen*, 2007], glacial seismicity [*Thelen*, 2013], and tectonic microseismicity [*Cattaneo et al.*, 1999; *Massa et al.*, 2006; *Bisrat et al.*, 2012] set a high CC threshold between 0.7 and 0.8 in order to include only the most similar events within each cluster. However, our goal was to include as many events as possible within a cluster, especially the noisier small events, so we set a lower CC threshold that ranged from 0.3 to 0.7 (Table S5). In addition to using the CC distributions of events within the same cluster and in different clusters as a guide to empirically set the threshold (Figure S12), we visually inspected all located and assigned waveforms in each of the 42 clusters or sub-clusters, as done for Cluster 1 in Figure 7, to check for waveform similarity within the cluster. *Harris and Dodge* [2011] used a low CC threshold of 0.316 to assign similar aftershocks to a cluster, because it was higher than most CC values in the distribution of unrelated event waveforms.

Figures S1 to S12

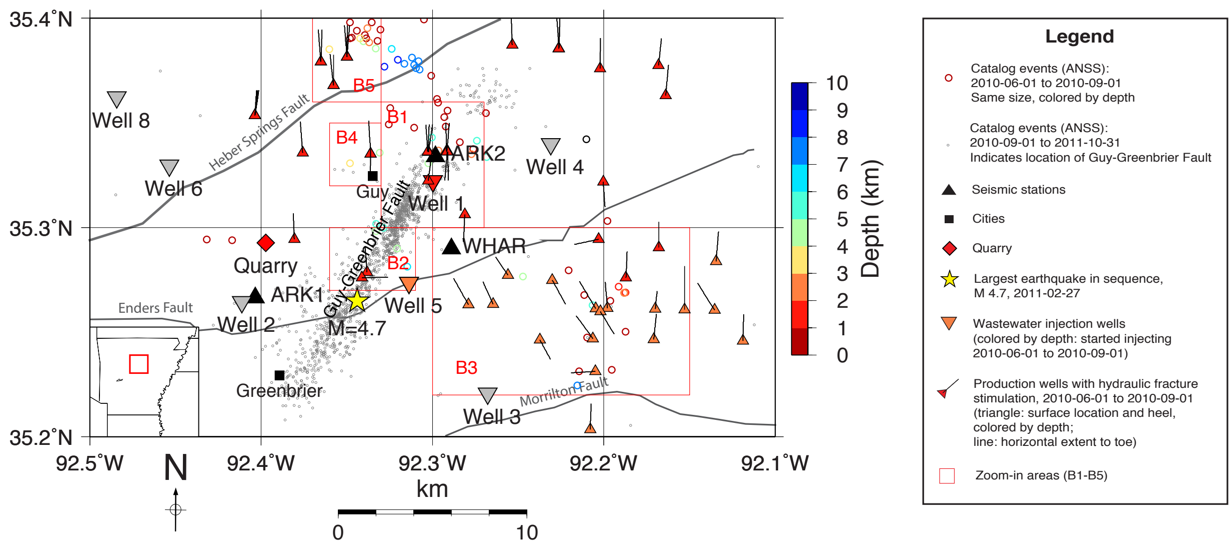

Figure S1. Map of Guy-Greenbrier area in central Arkansas (red box, inset at lower left) with ANSS catalog earthquake locations (75 events in Data Set S1), seismic stations, wastewater injection wells, and production wells with hydraulic fracturing stimulation during the time period 2010-06-01 to 2010-09-01. We plot ANSS catalog events after this time period (small gray dots) to delineate the location of the Guy-Greenbrier Fault. We also include the same red boxes B1-B5 from Figure 1. Fault traces are from *Horton* [2012].

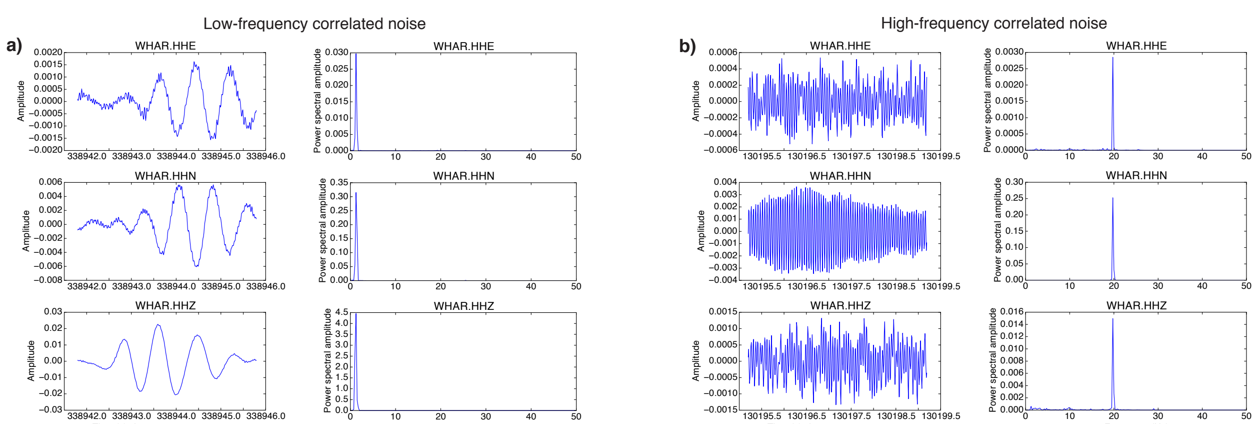

Figure S2. Examples of (a) low-frequency, (b) high-frequency, narrowband noise signals above the FAST similarity event detection threshold of 0.33 (Table S1) that were automatically removed by post-processing. These signals were similar enough to at least another narrowband noise signal at different times in the continuous seismic data to exceed the detection threshold. Time windows are on the left and power spectra are on the right.

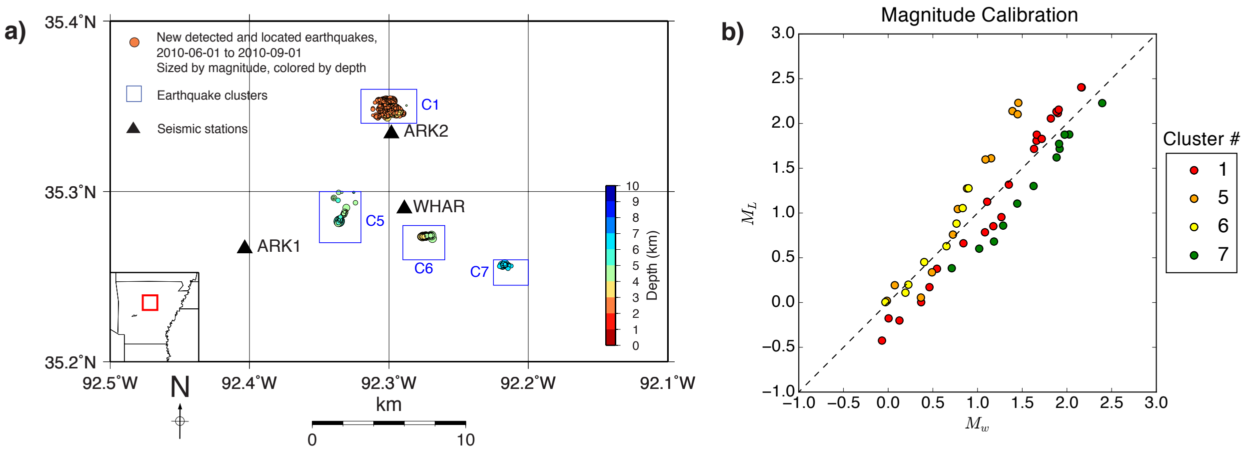

Figure S3. (a) Location of 54 selected high-quality events, belonging to one of the event clusters 1, 5, 6, 7 (blue rectangles) used to calibrate M_L. (b) Magnitude calibration results: comparison of M_L with M_w for the selected 54 events, colored by cluster number.

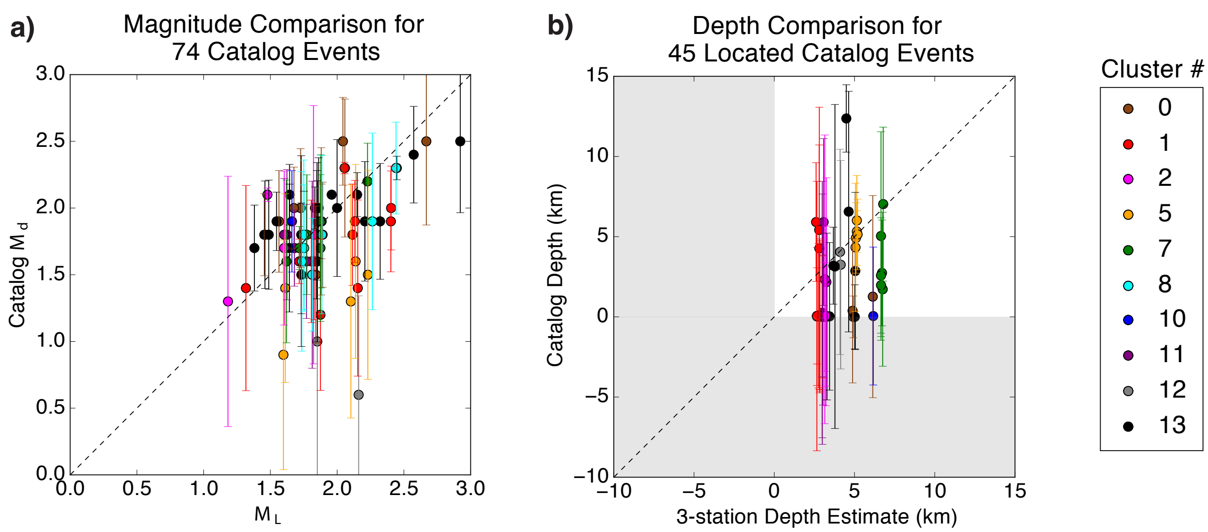

Figure S4. Comparison of our magnitude and depth estimates with those from the ANSS catalog. If an event lies on the dotted line, it matches the catalog value. Events are colored by cluster number (Table S5), with 0 indicating that the event does not belong to any of the 16 clusters. (a) Catalog magnitude M_d, computed from duration of the coda, versus our M_L local magnitude estimates, for 74 events. Although catalog M_d values have large uncertainties, they are reasonably consistent with our M_L estimates (Section S1). (b) Catalog depth versus our 3-station depth estimates for the 45 catalog events that we located (events before 2010-06-11 were not recorded on ARK1 or ARK2). Depths are positive downward, so the gray regions with negative depths are not physical. The depths generally agree, only because of the large depth uncertainties from the catalog.

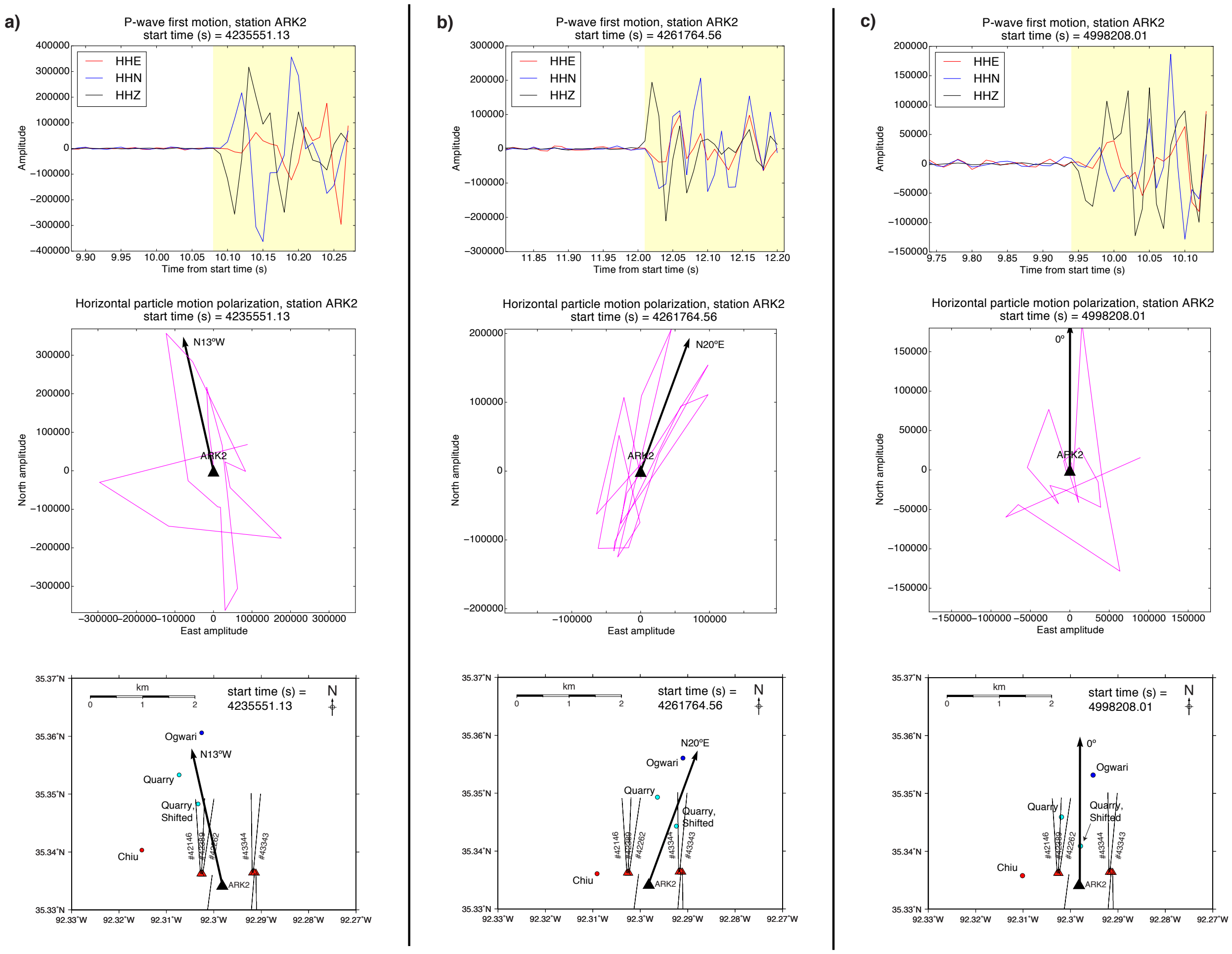

Figure S5. *P*-wave polarization analysis at station ARK2 for three events from Cluster 1: (a) Event at 4235551.13 s, (b) Event at 4261764.56 s, (c) Event at 4998208.01 s, for the purpose of testing different velocity models. Event start times are defined by the number of seconds after UTC 2010-06-01 00:00:00. (Top) 0.4-second window on each component of ARK2, centered on the *P*-wave arrival. We use a 0.2-second time window after the *P* arrival (yellow rectangle) for polarization analysis of the horizontal components (Center), which shows the North (HHN, blue) versus East (HHE, red) particle motion (magenta). To determine the back-azimuth to the event location (black arrow away from ARK2), we estimate the best-fit orientation of the particle motion with singular value decomposition, then resolve the 180° ambiguity with the first motion polarities [*Havskov and Ottemoller*, 2010]. (Bottom) We compare the back-azimuth to the event (black arrow away from ARK2) with the event location determined using three different velocity models: *Chiu et al.* [1984] (red), *Ogwari et al.* [2016] (blue), and our new quarry-constrained model (cyan). For all three events, the locations obtained with the *Ogwari et al.* [2016] and quarry-derived models are reasonably consistent with the back-azimuth direction, but the location from the *Chiu et al.* [1984] model is in the wrong direction. The quarry-model locations shifted ~0.7 km southeast (cyan), as applied in Figure 9, agree with the back-azimuth locations.

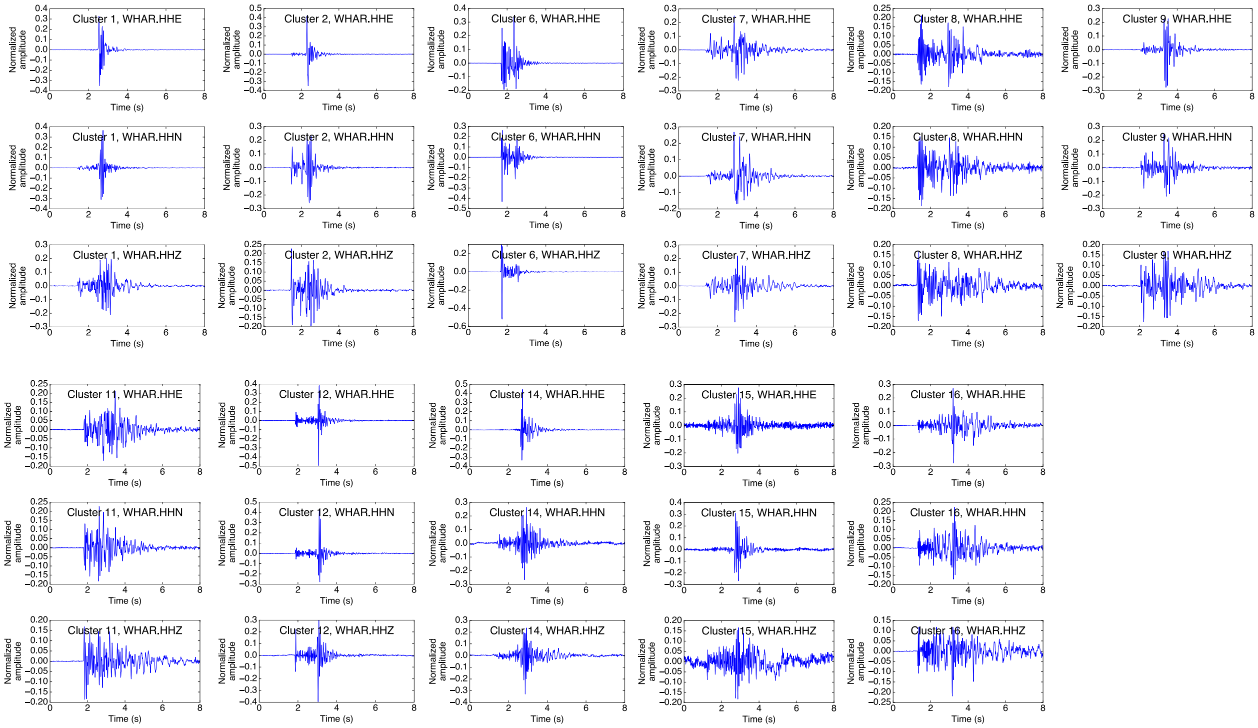

Figure S6. Representative stack waveforms at each component of station WHAR for the 11 clusters from Table S5 with only one sub-cluster. Waveforms at WHAR for all located events in each cluster are normalized and aligned to maximum CC with the master event (with largest magnitude) before stacking. Unlocated events are cross-correlated with these stack waveforms in order to assign them to clusters.

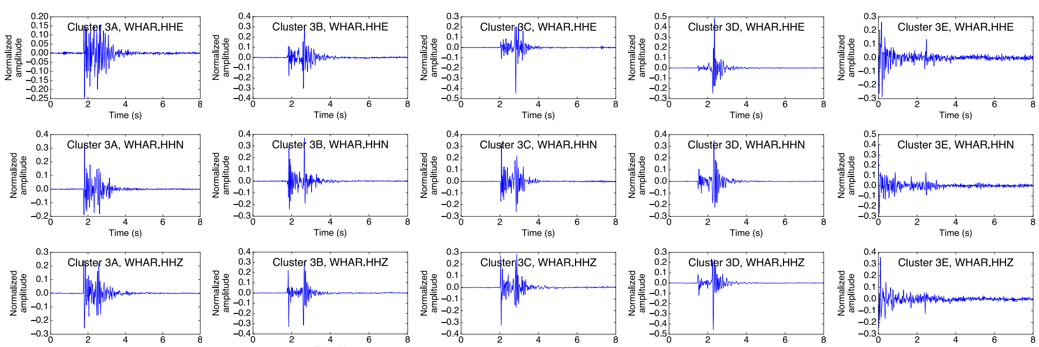

Figure S7. Stack waveforms at each component of station WHAR for the 5 sub-clusters (3A-3E) from Cluster 3 (Table S5). Differences between the stack waveforms for each sub-cluster are evident. Events are assigned to Cluster 3 by cross-correlation with each sub-cluster stack waveform.

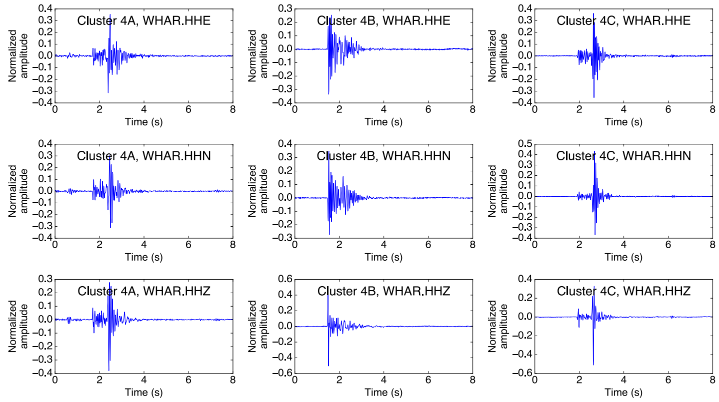

Figure S8. Stack waveforms at each component of station WHAR for the 3 sub-clusters (4A-4C) from Cluster 4 (Table S5). Differences between the stack waveforms for each sub-cluster are evident. Events are assigned to Cluster 4 by cross-correlation with each sub-cluster stack waveform.

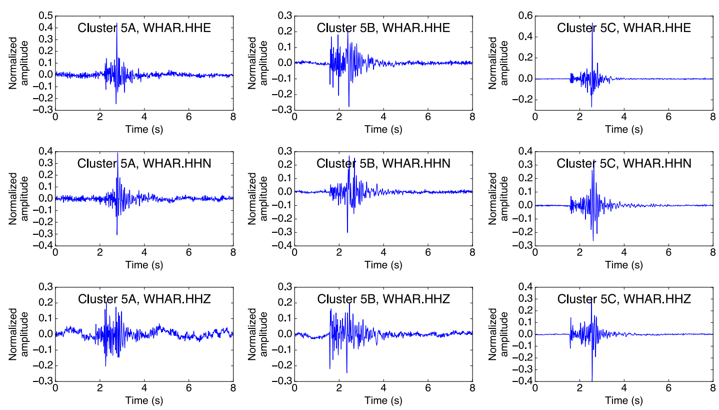

Figure S9. Stack waveforms at each component of station WHAR for the 3 sub-clusters (5A-5C) from Cluster 5 (Table S5). Differences between the stack waveforms for each sub-cluster are evident. Events are assigned to Cluster 5 by cross-correlation with each sub-cluster stack waveform.

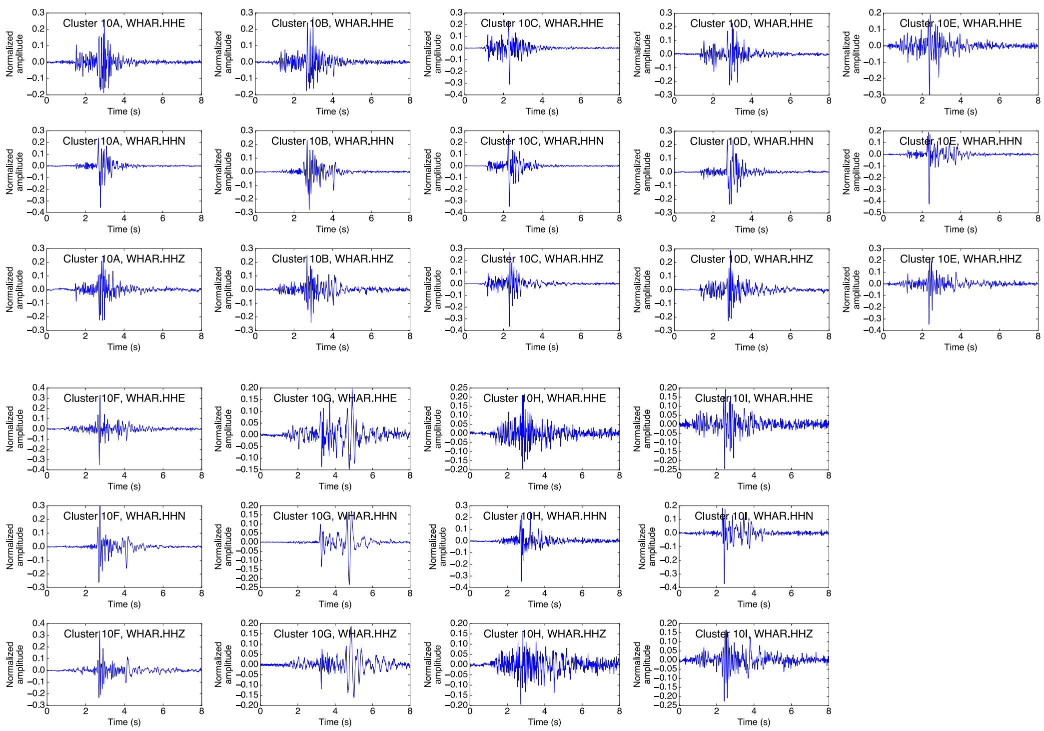

Figure S10. Stack waveforms at each component of station WHAR for the 9 sub-clusters (10A-10I) from Cluster 10 (Table S5). Differences between the stack waveforms for each sub-cluster are evident, though this is not surprising because this cluster includes events from a larger area. Events are assigned to Cluster 10 by cross-correlation with each sub-cluster stack waveform.

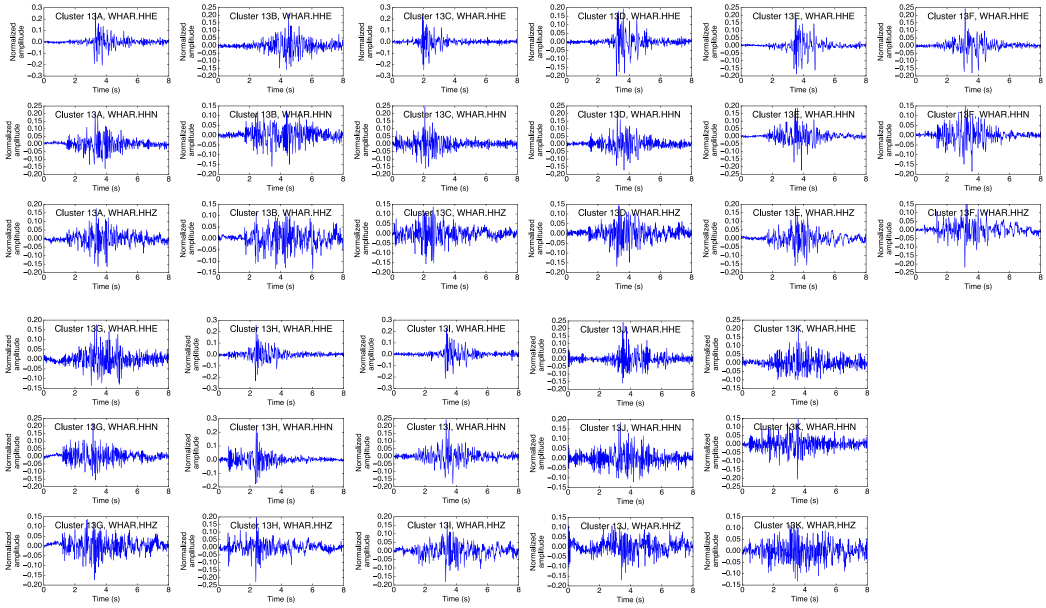

Figure S11. Stack waveforms at each component of station WHAR for the 11 sub-clusters (13A-13K) from Cluster 13 (Table S5). Differences between the stack waveforms for each sub-cluster are evident, especially because these waveforms are noisy. Events are assigned to Cluster 13 by cross-correlation with each sub-cluster stack waveform.

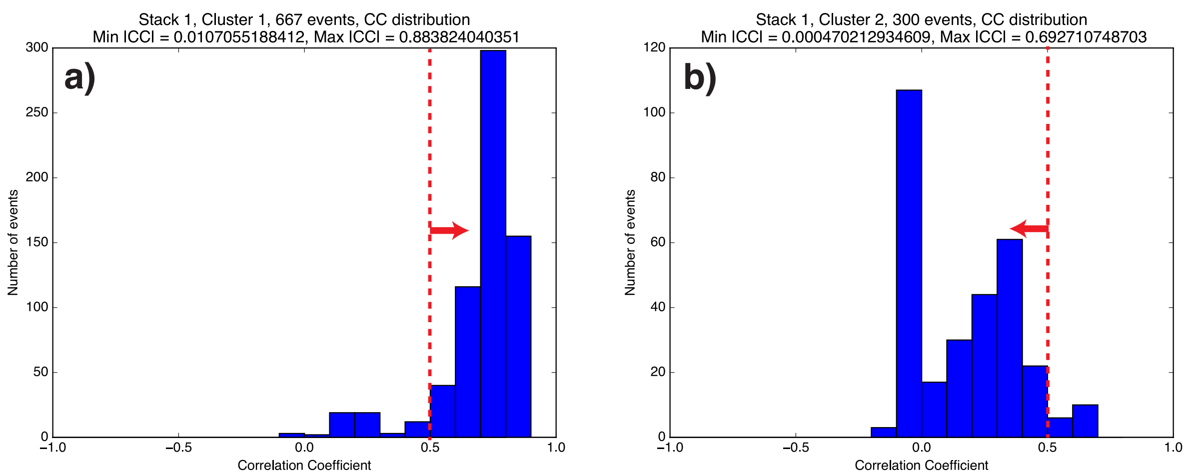

Figure S12. Description of how to set a CC threshold empirically to assign events to a specific cluster. In this example, the threshold is CC ≥ 0.5 (dotted red line) for Cluster 1 (Table S5). (a) CC distribution from cross-correlating the Cluster 1 stack waveform with all located event waveforms in Cluster 1, at 3 components on station WHAR. Almost all CC values exceed the threshold of 0.5 as desired. (b) CC distribution from cross-correlating the Cluster 1 stack waveform with all located event waveforms in Cluster 2, at 3 components on station WHAR. Since these events are in a different cluster from the stack waveform, most CC values are below the threshold of 0.5 as desired.

Tables S1 to S8

Table S1. FAST input parameters [*Yoon et al.*, 2015] used to detect earthquakes in each component of continuous seismic data at station WHAR. The event detection threshold was applied to the total 3-component FAST similarity matrix.

FAST parameter	**Value**
Time series window length for spectrogram generation	300 samples (3 s)
Time series window lag for spectrogram generation	3 samples (0.03 s)
Spectral image window length	64 samples (1.92 s)
Spectral image window lag = fingerprint sampling period	10 samples (0.3 s)
Number of top k amplitude standardized Haar coefficients	800 (out of 2048)
LSH: number of hash functions per hash table r	8
LSH: number of hash tables b	100
Initial pair threshold: number v (fraction) of tables, pair in same bucket	4 (4/100 = 0.04)
Event detection threshold (total 3-component FAST similarity)	0.33
Similarity search: near-repeat exclusion parameter	5 samples (1.5 s)
Near-duplicate pair and event elimination time window	4 s
Template matching comparison time window	6 s

Table S2. Input parameters for *ph2dt* program to compute catalog P and S differential travel times.

Parameter name	**Value**	**Description**
MINWGHT	0	Minimum pick weight allowed
MAXDIST	120	Maximum distance (km) between event pair and stations
MAXSEP	10	Maximum hypocentral separation (km)
MAXNGH	30	Maximum number of neighbors per event
MINLNK	3	Minimum number of links required to define a neighbor
MINOBS	1	Minimum number of links per pair saved
MAXOBS	6	Maximum number of links per pair saved

Table S3. Input parameters for hypoDD program to relocate earthquakes.

Parameter name	Value	Description
IDAT	3	Cross-correlation and catalog
IPHA	3	P and S phases
DIST	400	Maximum distance (km) between cluster centroid and station
OBSCC	2	Minimum number of obs/pair for crosstime data
OBSCT	2	Minimum number of obs/pair for network data
MINDS	0	Minimum distance between individual event pairs and stations
MAXDS	150	Maximum distance between individual event pairs and stations
MAXGAP	-999	Maximum azimuthal gap between individual event pairs and stations (-999: not used)
ISTART	2	From network sources
ISOLV	2	LSQR
IAQ	2	Keep air-quakes, reset depths to those of previous (successful) iteration
NSET	5	Number of sets of iteration

Table S4. hypoDD data weighting parameters, LSQR mode, 24 iterations.

NITER	WTCCP	WTCCS	WRCC	WDCC	WTCTP	WTCTS	WRCT	WDCT	DAMP
4	0.01	0.01	-9	-9	1	1	-9	-9	180
4	0.01	0.01	-9	-9	1	1	5	4	180
4	1	1	-9	2	0.01	0.01	5	4	180
4	1	1	5	2	0.01	0.01	5	4	180
8	1	1	5	0.5	0.01	0.01	5	4	180

Table S5. Information for 16 earthquake clusters with at least 10 events: minimum and maximum latitude and longitude of box defining the cluster boundary; number of sub-clusters; minimum CC threshold to assign events to each cluster; number of located, assigned, and total events in cluster; production well permit numbers (Table S8) within 2 km of and associated with each cluster. For clusters with more than one sub-cluster (Clusters 3, 4, 5, 10, 13), the bottom table contains the minimum CC threshold to assign events to each sub-cluster, and the number of located, assigned, and total events in that sub-cluster.

Cluster number	Minimum latitude (deg)	Maximum latitude (deg)	Minimum longitude (deg)	Maximum longitude (deg)	Number of sub-clusters	Minimum CC threshold to assign	Number of located events	Number of assigned events	Total number of events	Nearby production wells: Permit number
1	35.34	35.36	-92.32	-92.28	1	0.5	667	2525	3192	42069, 42146, 42262, 42389, 43343, 43344, 43375, 43376
2	35.3258	35.34	-92.32	-92.28	1	0.5	300	778	1078	42069, 42146, 42262, 42389, 43343, 43344, 43375, 43376
3	35.315	35.3258	-92.32	-92.3	5	see table below	228	486	714	42069, 42146, 42262, 42389, 43343, 43344, 43375, 43376
4	35.3	35.315	-92.325	-92.31	3	see table below	22	266	288	42069, 43375, 43376
5	35.27	35.3	-92.35	-92.32	3	see table below	45	12	57	43114, 43304
6	35.26	35.28	-92.29	-92.26	1	0.5	80	440	520	43043
7	35.245	35.26	-92.225	-92.22	1	0.5	45	210	255	43153
8	35.225	35.245	-92.24	-92.203	1	0.4	19	365	384	43154, 43258
9	35.235	35.245	-92.27	-92.25	1	0.5	3	9	12	43154
10	35.26	35.3	-92.25	-92.15	9	see table below	32	428	460	43042, 43058, 43203, 43241, 43257, 43272, 43122, 43149
11	35.335	35.35	-92.344	-92.332	1	0.4	55	79	134	43439
12	35.325	35.335	-92.355	-92.345	1	0.7	9	5	14	43433
13	35.365	35.395	-92.368	-92.335	11	see table below	207	848	1055	43252, 43253, 43254, 43255, 43256
14	35.233	35.245	-92.32	-92.3	1	0.5	4	15	19	–
15	35.3	35.325	-92.395	-92.375	1	0.5	2	24	26	43244
16	35.38	35.395	-92.26	-92.235	1	0.4	1	18	19	43219

Sub-cluster number	Minimum CC threshold to assign	Number of located events	Number of assigned events	Total number of events
3A	0.6	11	34	45
3B	0.5	14	74	88
3C	0.5	38	222	260
3D	0.6	164	156	320
3E	0.5	1	0	1
4A	0.6	2	179	181
4B	0.6	12	54	66
4C	0.7	8	33	41
5A	0.5	2	5	7
5B	0.4	2	2	4
5C	0.6	41	5	46
10A	0.5	9	66	75
10B	0.5	4	73	77
10C	0.5	4	83	87
10D	0.8	2	1	3
10E	0.5	3	2	5
10F	0.5	7	169	176
10G	0.5	1	9	10
10H	0.4	1	6	7
10I	0.5	1	19	20
13A	0.4	12	94	106
13B	0.3	2	12	14
13C	0.4	2	28	30
13D	0.4	4	346	350
13E	0.4	63	44	107
13F	0.4	72	171	243
13G	0.3	2	5	7
13H	0.5	20	131	151
13I	0.4	11	12	23
13J	0.35	8	0	8
13K	0.3	11	5	16

Table S6. Minimum CC cutoff for separating event clusters into sub-clusters; also see Table S5.

Cluster number	Minimum CC cutoff	Number of sub-clusters
3	0.2	5
4	0.2	3
5	0.1	3
10	0.4	9
13	0.18	11

Table S7. Class 2 Underground Injection Control (UIC) wastewater injection wells (Figure 1, inverted triangles) active during the study period 2010-06-01 to 2010-09-01. Wells 1 and 5 (Figure 1, inverted triangles colored by depth), located nearest the Guy-Greenbrier Fault, started injecting during the study period. Volume and pressure are peak values observed during the injection period (not the study period). Data are taken from *Horton* [2012], *Ogwari et al.* [2016], and *AOGC* [2017a].

Well Number	Well Name	Permit	Volume (m^3/month)	Pressure (MPa)	Start and Stop Dates	Injection Depth (m)
1	SRE	43266	62,622	11.8	2010-07-07 to 2011-03-03	1821 to 1969
2	Trammel	41079	54,058	15.8	2009-04-15 to 2011-06-20	1982 to 2009
3	Moore	39487	23,435	20.3	2009-06-15 to 2011-07-27	2365 to 3231
4	Underwood	42981	29,573	5.1	2010-01-15 to 2010-10-15	1713 to 1926
5	Edgmon	36380	19,580	19.6	2010-08-16 to 2011-03-03	2379 to 3344
6	Scroggins	42989	18,629	3.2	2010-04-05 to 2011-10-15	678 to 706

Table S8. Data for 53 production wells with hydraulic fracture stimulation during the time period 2010-06-01 to 2010-09-01, located within the map area in Figure 1, retrieved from the public Arkansas Oil and Gas Database [*AOGC*, 2017a,b]. The total depth, top depth, and bottom depth are measured depth values along the well trajectory, while the true vertical depth is measured vertically. Bottom depth is deepest measured depth at the first stage of stimulation, while top depth is shallowest measured depth at the last stage of stimulation.

Permit	Date, first stage of stimulation	Date, last stage of stimulation	Latitude (deg) at surface	Longitude (deg) at surface	Latitude (deg) at well toe	Longitude (deg) at well toe	Elevation (ft) at ground level	True vertical depth (ft)	Total depth (ft)	Top depth (ft)	Bottom depth (ft)	Acid (gallons)	Frac Volume (barrels)	Number of stages
42069	06-14-2010	06-19-2010	35.322345	-92.302538	35.335994	-92.300343	758	5543	10622	5940	10530	15192	138890	10
42146	07-19-2010	07-24-2010	35.336189	-92.302812	35.349317	-92.303790	664	5381	10078	5880	9983	13495	122208	9
42262	07-26-2010	07-29-2010	35.336171	-92.302342	35.349194	-92.299977	664	5362	10144	5935	10052	13495	123949	9
42389	07-19-2010	07-24-2010	35.336178	-92.302577	35.349272	-92.301942	676	5359	9943	5736	9851	13495	123255	9
43010	07-02-2010	07-08-2010	35.263261	-92.279087	35.275764	-92.287730	478	7117	12055	7350	11675	13495	118934	9
43042	07-12-2010	07-20-2010	35.274831	-92.230032	35.263347	-92.222359	594	6980	11668	7222	11483	14994	142486	10
43043	07-06-2010	07-09-2010	35.263451	-92.26466	35.274076	-92.270534	483	7138	11288	7380	11188	13733	121578	9
43058	07-30-2010	08-16-2010	35.294472	-92.202972	35.291944	-92.217222	576	6492	11900	7896	11740	26000	151031	13
43068	06-24-2010	06-29-2010	35.260534	-92.135228	35.273433	-92.145333	716	6702	12183	7145	12082	16386	139414	11
43114	07-30-2010	08-03-2010	35.278709	-92.33822	35.291188	-92.339376	558	5772	10165	6097	10057	16491	124308	11
43122	07-12-2010	07-16-2010	35.261361	-92.204811	35.274272	-92.214154	445	6728	12006	7030	11903	16493	133826	11
43130	06-01-2010	06-05-2010	35.283936	-92.134435	35.298056	-92.133056	719	6575	11747	8118	11672	15000	47266	9
43149	07-23-2010	07-28-2010	35.261529	-92.197928	35.273921	-92.199155	495	6727	11498	7155	11382	28000	164030	14
43150	06-14-2010	06-18-2010	35.246555	-92.17084	35.258701	-92.169322	616	6973	11140	6975	11002	24000	150094	12
43153	08-23-2010	08-28-2010	35.246948	-92.206349	35.257882	-92.216141	539	6954	12647	7720	12505	16493	137875	11
43154	08-02-2010	08-06-2010	35.246407	-92.23741	35.237078	-92.231171	379	6956	10480	7478	10374	11995	85108	8
43157	06-07-2010	06-11-2010	35.387207	-92.2535	35.403056	-92.254167	617	5342	11193	5635	11052	32000	195116	16
43160	07-12-2010	07-14-2010	35.245834	-92.11881	35.258611	-92.118333	622	6922	11140	7360	11042	20000	127363	10
43172	06-07-2010	06-09-2010	35.32187	-92.20038	35.309911	-92.199324	558	5290	9642	5670	8252	9069	79617	6
43203	06-25-2010	07-07-2010	35.276121	-92.18706	35.288611	-92.186389	499	6544	11140	6870	10982	26000	146099	13
43219	06-16-2010	06-18-2010	35.3856	-92.22622	35.401111	-92.227500	698	5036	10851	5277	9535	18170	108268	11
43229	06-11-2010	06-17-2010	35.385589	-92.226237	35.401389	-92.225556	698	5072	10550	5308	10485	22500	143263	14
43230	06-11-2010	06-17-2010	35.385644	-92.226235	35.401389	-92.225556	698	5057	10805	5193	10715	22500	144653	14
43241	06-26-2010	07-02-2010	35.290397	-92.167958	35.302747	-92.167991	607	6433	10616	7000	10086	10555	91252	7
43244	06-09-2010	06-13-2010	35.294495	-92.380725	35.306775	-92.381327	640	5862	10199	6130	10000	16500	107819	10
43252	07-19-2010	07-23-2010	35.367931	-92.357862	35.382900	-92.359842	590	4260	9647	4579	9538	22500	141868	13
43253	07-19-2010	07-23-2010	35.367974	-92.35782	35.382598	-92.357857	591	4258	9222	4506	9110	19500	128390	12
43254	05-28-2010	06-03-2010	35.381441	-92.349975	35.396667	-92.351389	697	4324	9868	4557	9806	20913	127268	13
43255	05-23-2010	05-27-2010	35.381496	-92.349973	35.396667	-92.349722	697	4320	9640	4523	9577	19500	136670	13
43256	05-24-2010	05-27-2010	35.381551	-92.349974	35.396667	-92.348443	699	4306	9694	4508	9630	21059	139043	13
43257	06-14-2010	06-23-2010	35.260774	-92.152792	35.281667	-92.152778	699	6592	13978	7330	13872	40000	236143	20
43258	06-03-2010	06-06-2010	35.231283	-92.204868	35.230556	-92.219167	364	7049	11139	7195	11032	20000	119337	10
43272	07-16-2010	07-21-2010	35.261091	-92.16996	35.269722	-92.168889	646	6716	9715	6900	9474	18000	106571	9
43278	07-09-2010	07-14-2010	35.203558	-92.207971	35.215769	-92.207403	352	8239	11500	8559	11382	16000	89462	8
43304	08-10-2010	08-14-2010	35.276012	-92.341151	35.276268	-92.325925	505	5820	10280	6193	10174	14994	113580	10
43306	07-06-2010	07-08-2010	35.377573	-92.168007	35.390289	-92.166377	771	4700	9140	5020	9040	16650	112500	10
43325	07-23-2010	07-28-2010	35.379226	-92.365244	35.394444	-92.366667	688	4349	9958	4817	9875	21000	130006	13
43326	07-24-2010	07-28-2010	35.379225	-92.365177	35.394444	-92.364444	687	4352	9637	4723	9535	18031	129153	11
43343	07-15-2010	07-21-2010	35.336367	-92.291678	35.350051	-92.290138	724	5357	10200	5680	10092	14987	134400	10
43344	07-16-2010	07-22-2010	35.336375	-92.291912	35.350151	-92.292110	722	5360	10270	5750	10165	15633	132376	10
43375	07-22-2010	07-29-2010	35.336349	-92.291120	35.322514	-92.290833	725	5444	10268	6083	10161	14192	115168	9
43376	07-23-2010	07-28-2010	35.336357	-92.291443	35.322597	-92.292798	724	5468	10265	5811	10160	15280	131466	10
43379	07-26-2010	07-28-2010	35.363129	-92.163863	35.377778	-92.162222	824	4886	10005	5257	9920	19601	120158	12
43433	08-07-2010	08-11-2010	35.335234	-92.336146	35.323000	-92.336433	710	5603	9868	5708	9743	16500	106904	10
43436	08-23-2010	08-30-2010	35.259998	-92.202075	35.249112	-92.194023	468	6812	11131	7028	11071	16833	116720	11
43439	08-07-2010	08-11-2010	35.33523	-92.336079	35.350997	-92.337277	710	5536	11216	5837	11148	21000	138348	13
43465	08-16-2010	08-20-2010	35.306258	-92.281166	35.293181	-92.281918	758	5707	10112	5895	10004	16493	123540	11
43514	08-23-2010	08-25-2010	35.335774	-92.375855	35.350651	-92.377132	544	5667	10706	6305	10500	15186	98459	9
43532	08-10-2010	08-13-2010	35.277262	-92.255864	35.285906	-92.263239	523	7029	10716	7431	10614	13495	100748	9
43554	08-09-2010	08-18-2010	35.353586	-92.403694	35.365625	-92.402707	548	4345	8465	4621	8396	16500	96254	10
43555	08-09-2010	08-19-2010	35.35362	-92.403641	35.365609	-92.402037	548	4356	8550	4715	8486	16000	100711	10
43556	08-09-2010	08-20-2010	35.353655	-92.403591	35.365594	-92.401366	548	4347	8690	4695	8617	16500	98743	10
43627	07-29-2010	07-31-2010	35.376012	-92.202252	35.390556	-92.201944	694	4752	9760	5012	9675	19638	118441	12

Data Set S1.

List of 75 catalog events from the ANSS catalog, with latitude between 35.18°N and 35.42°N and longitude between -92.52°W and -92.08°W, sorted by time. Figure S1 plots these catalog event locations, colored by depth. (To account for location uncertainty, we added 0.02° ≈ 2 km to the map boundaries in Figure S1 to define our catalog search box.) These events were recorded by the Cooperative New Madrid Seismic Network. One event (ID nm607354) occurred during a time gap in the continuous data at WHAR, so we did not detect it. The table below describes each column. Columns 16-19 match columns 1-4 in Data Set S2, and columns 5-8 in Data Set S3.

Column	**Description**
1	Catalog origin time in seconds since UTC 2010-06-01T00:00:00
2	Catalog origin time (UTC) in format YYYY-MM-DDTHH:MM:SS.SSSZ
3	Latitude (deg) from catalog
4	Longitude (deg) from catalog
5	Depth (km) from catalog
6	Magnitude from catalog: M_d calculated from coda duration
7	Number of stations used for catalog location
8	Largest azimuthal gap between azimuthally adjacent stations (deg)
9	Horizontal distance from epicenter to nearest station (deg)
10	RMS travel time residual (s)
11	Catalog event ID
12	Catalog horizontal location error (km)
13	Catalog depth location error (km)
14	Catalog magnitude error
15	Number of stations used for magnitude calculation
16	Event id number = 100 * (Event detection time in seconds since UTC 2010-06-01T00:00:00)
17	Local magnitude M_L
18	Which algorithm detected this event? (0: both FAST and template matching, 1: only template matching, 2: only FAST)
19	Cluster number (from Table S5), including sub-cluster information. 0: does not belong to any cluster.

Data Set S2.

List of 14,604 detected events from 2010-06-01 to 2010-09-01 at single station WHAR, sorted by time. The table below describes each column.

Column	Description
1	Event id number = 100 * (Event detection time in seconds since UTC 2010-06-01T00:00:00)
2	Local magnitude M_L
3	Which algorithm detected this event? (0: both FAST and template matching, 1: only template matching, 2: only FAST)
4	Cluster number (from Table S5), including sub-cluster information. 0: does not belong to any cluster. 17: quarry blast.

Data Set S3.

List of 1,740 located events, sorted by time. The table below describes each column. Columns 5-8 match columns 1-4 in Data Set S2.

Column	Description
1	Latitude (deg)
2	Longitude (deg)
3	Depth (km)
4	Number of days since UTC 2010-06-01T00:00:00
5	Event id number = 100 * (Event detection time in seconds since UTC 2010-06-01T00:00:00)
6	Local magnitude M_L
7	Which algorithm detected this event? (0: both FAST and template matching, 1: only template matching, 2: only FAST)
8	Cluster number (from Table S5), including sub-cluster information. 0: does not belong to any cluster. 17: quarry blast.

Movie S1.

Cumulative time evolution of seismicity in Cluster 1, and hydraulic fracturing stimulation at the 5 nearest production wells (labeled by permit number in Table S8), near north end of the Guy-Greenbrier Fault. We display the shifted event locations from Figure 9b. Earthquakes (circles sized by relative magnitude), as well as stimulated sections of production wells during each stage of hydraulic fracturing, are colored by time with Day 0 defined as 2010-07-16 00:00:00 UTC. Thick colored stages along the well path are currently being stimulated, while thin colored stages are past stimulations. The movie covers the 16-day time period shown in Figure 10e.