

Machine learning based analysis of the Guy-Greenbrier, Arkansas earthquakes: a tale of two sequences

Park Yongsoo¹, Mousavi S. Mostafa¹, Zhu Weiqiang¹, Ellsworth William L.¹, and Beroza Gregory C.¹

¹Stanford University

November 16, 2022

Abstract

We revisited the June, 2010 - October, 2011 Guy-Greenbrier earthquake sequence in central Arkansas using PhaseNet, a deep neural network trained to pick P and S arrival times. We applied PhaseNet to continuous waveform data and used phase association and hypocenter relocation to locate nearly 90,000 events. Our catalog suggests that the sequence consists of two adjacent earthquake sequences on the same fault and that the second sequence may be associated with the wastewater disposal well to the west of the Guy-Greenbrier Fault, rather than the wells to the north and the east that were previously implicated. We find that each sequence is comprised of many small clusters that exhibit diffusion along the fault at shorter time scales. Our study demonstrates that machine learning based earthquake catalog development is now feasible and will yield new insights into earthquake behavior.

1 **Machine learning based analysis of the Guy-Greenbrier,**
2 **Arkansas earthquakes: a tale of two sequences**

3 **Yongsoo Park¹, S. Mostafa Mousavi¹, Weiqiang Zhu¹, William L. Ellsworth¹,**
4 **Gregory C. Beroza¹**

5 ¹Department of Geophysics, Stanford University, Stanford, CA 94305, USA

6 **Key Points:**

- 7 • Deep neural network pickers can increase the efficiency of earthquake cataloging
8 workflow.
- 9 • The sequence appears to be composed of two sub-sequences possibly triggered by
10 different wells.
- 11 • Improved catalog reveals characteristic spatio-temporal seismicity patterns.

Corresponding author: Yongsoo Park, ysp@stanford.edu

Abstract

We revisited the June, 2010 - October, 2011 Guy-Greenbrier earthquake sequence in central Arkansas using PhaseNet, a deep neural network trained to pick P and S arrival times. We applied PhaseNet to continuous waveform data and used phase association and hypocenter relocation to locate nearly 90,000 events. Our catalog suggests that the sequence consists of two adjacent earthquake sequences on the same fault and that the second sequence may be associated with the wastewater disposal well to the west of the Guy-Greenbrier Fault, rather than the wells to the north and the east that were previously implicated. We find that each sequence is comprised of many small clusters that exhibit diffusion along the fault at shorter time scales. Our study demonstrates that machine learning based earthquake catalog development is now feasible and will yield new insights into earthquake behavior.

Plain Language Summary

Finding small earthquake signals from long duration continuous seismic data is a time consuming task, but machine learning algorithms have the potential to accelerate the workflow and improve the results. We reprocessed the seismic data from the area spanning Guy and Greenbrier in central Arkansas in 2010 and 2011 using a machine learning algorithm to re-examine this well-studied earthquake sequence, which is thought to be caused by injection of wastewater from unconventional hydrocarbon production into deep disposal wells. Even using conservative post-processing steps, we were able to locate nearly 90,000 earthquake events. The improved catalog illuminates previously unseen aspects of this earthquake sequence that give new insights into its behavior.

1 Introduction

The Guy-Greenbrier sequence stretched from June, 2010 to October, 2011 in the region spanning Guy and Greenbrier, Arkansas and is thought to have been induced by some combination of hydraulic stimulation of horizontal production wells and injection of wastewater into deep disposal wells. Since the work published by Horton in 2012 (Horton, 2012), this sequence has become an important subject for induced seismicity studies (Horton, 2012; Llenos & Michael, 2013; Huang & Beroza, 2015; Huang et al., 2016; Ogwari et al., 2016; Ogwari & Horton, 2016; Yoon et al., 2017; Mousavi et al., 2017; Yehya et al., 2018).

42 Some of these studies have focused on improving the catalog using different algo-
43 rithms and workflows. Ogwari and co-workers (Ogwari et al., 2016) used the short-time
44 average/long-time average (STA/LTA) trigger for detection, manually picked P and S
45 arrival times, and located 17,395 events from July 7, 2010 to October 20, 2010. From
46 their improved catalog they concluded that seismicity was mostly correlated with the
47 wastewater disposal well located close to the north end of the Guy-Greenbrier Fault. Yoon
48 and co-workers (Yoon et al., 2017) applied the FAST algorithm (Yoon et al., 2015) and
49 found over 10,000 events during the period from June 1, 2010 to August 31, 2010. They
50 were able to locate 1,740 events, which revealed multiple event clusters from hydraulic
51 fracturing operations as well as the initial stages of seismicity at the north end of the
52 Guy-Greenbrier Fault.

53 While these studies provided new insight into the initial stages of the sequence, the
54 activity over the remaining 12 months of the sequence received less attention. Huang and
55 Beroza (Huang & Beroza, 2015) used single station template matching on continuous wave-
56 form from July, 2010 to October, 2011 and detected over 460,000 events. The detections
57 in their study clearly showed two peaks of seismic activity: the first in October-November,
58 2010 and the second in February-March, 2011, however, the hypocenters of the detected
59 events were undetermined.

60 Cataloging small earthquakes from long continuous data involves a trade-off be-
61 tween detecting all the events and avoiding false detections. Avoiding false positive de-
62 tectations is often done through manual inspection in a post-processing stage, and the num-
63 ber of waveforms to inspect can grow quite rapidly. Machine learning algorithms have
64 the potential to play an important role in this application (Bergen et al., 2019). In par-
65 ticular, deep neural network arrival time pickers (Zhu & Beroza, 2018; Ross et al., 2018)
66 have the potential to accelerate the front end of the seismicity analysis workflow by re-
67 ducing the need for human arrival time picking. The robustness of these algorithms can
68 reduce the number of false positives, which can reduce the need for analysts to hand tune
69 parameters and inspect waveforms. In this study, we use the PhaseNet arrival time picker
70 developed by Zhu and Beroza (Zhu & Beroza, 2018), which uses the U-Net architecture
71 (Ronneberger et al., 2015) as trained on the Northern California earthquake catalog. We
72 demonstrate how this tool can accelerate and improve the earthquake cataloging work-
73 flow. The improved catalog that results leads to new insights into this already well stud-
74 ied earthquake sequence.

2 Method

PhaseNet (Zhu & Beroza, 2018) takes a 3-component seismogram or a single channel vertical seismogram as input, and outputs the probability of each time step being a P arrival, an S arrival, or neither. The algorithm is designed such that the probability will peak at the arrival time of the phase. We use data from ten stations located near the Guy-Greenbrier Fault. Station WHAR was operational during the entire period from June 1, 2010 to October 31, 2011, while the other stations covered only a portion of the period (Table S1) resulting in a temporally and spatially variable network. Each component of the data was high pass filtered at 1 *Hz* and windowed into traces with 3,000 samples. We used an overlap of 1,500 samples to prevent earthquake signals being truncated at the edges. We removed duplicate picks in the post-processing step by tracking their timestamps and retained only the picks with probability 0.5 or higher. Traces containing time gaps were discarded. The approximate wall clock time for running the network on a month of continuous data from a single station was around 40 minutes on a single Intel Xeon CPU E5-2683V4 (2.10 *GHz*) hardware.

We associated the picks returned by the neural network by grid searching for a matching synthetic moveout calculated from the travel times, similar to the methods described in (Johnson et al., 1997; Zhang et al., 2019) using the velocity model in (Yoon et al., 2017). For efficient computation, we allowed a maximum separation of 20 seconds per event when forming possible moveouts from the picks, and used a lookup table for the travel times, which was built after discretizing the domain into blocks with latitude and longitude spacing of 0.004 degrees and depth spacing of 0.4 *km*. We allowed a difference between observed and synthetic travel time of up to 0.4 seconds and retained only the cases with at least 4 associated picks from at least 3 stations as candidate events. The number of events after association was 100,092. We used HYPOINVERSE-2000 (Klein, 2002) to locate each event independently and used the grid search solutions to initialize hypocenters. We discarded events with RMS travel time residual greater than 0.2 seconds, horizontal error of greater than 5 *km*, and depth error of greater than 5 *km*. This reduced the number of events to 96,191.

These initial locations placed the majority of the events on the Guy-Greenbrier Fault, with others forming small clusters that appeared to be a combination of natural seismicity, hydraulic fracturing operations, quarry blasts, and scatter due to the event being ei-

107 ther poorly located or a false positive. We separate these events into two groups: the on-
108 fault group, which are spread along the Guy-Greenbrier Fault, and the off-fault group,
109 which are more widely distributed, to facilitate double-difference relocation (Waldhauser
110 & Ellsworth, 2000). The on-fault group was composed of the events within 2 *km* of the
111 Guy-Greenbrier Fault and the remaining events were put into the off-fault group. The
112 on-fault group was further separated into a group of well-constrained events and a group
113 of poorly constrained events, based on the errors computed from HYPOINVERSE-2000.
114 We used a horizontal error of 1 *km* and a depth error of 2 *km* for the separation.

115 We paired all events with the well-constrained events to compute arrival time dif-
116 ferences by waveform cross-correlation. We bandpassed the data from 1 to 25 *Hz* and
117 the windows for cross-correlation were set to 0.2 seconds before and 0.4 seconds after the
118 P arrival, and 0.4 seconds before and 0.8 seconds after the S arrival, with shift length
119 of 0.3 seconds. We set the search radius to 2 *km* when pairing two well-constrained events
120 and to 5 *km* when pairing a poorly constrained event with a well-constrained event af-
121 ter setting the depths of all events to zero. The number of pairs was restricted to 100
122 when pairing each poorly constrained event with well-constrained events to prevent the
123 cluster centroid from moving towards the locations of the poorly constrained events. We
124 paired neighboring events that had at least 4 differential travel time observations from
125 3 stations with cross-correlation coefficient of 0.7 or above. The February 28, 2011 main-
126 shock had enough phase arrival picks to determine the initial hypocenter but it could
127 not be paired with other events using cross-correlation because the data was partially
128 clipped. Therefore, we paired this event with 1,000 other events that were within 2 *km*
129 radius and had highly confident picks (probability of 0.9 or greater). We passed the dif-
130 ferential travel times to the HypoDD double-difference relocation algorithm (Waldhauser,
131 2001) and the group was treated as a single cluster. After relocation, the events that did
132 not belong to the now-sharp Guy-Greenbrier Fault and the scattered events that were
133 appended to the on-fault group that were not paired with other events were placed back
134 to the off-fault group.

135 Most clusters and scattered events in the off-fault group had poor azimuthal sta-
136 tion coverage, i.e., an azimuthal gap greater than 180 degrees. Rather than separating
137 the events into well- and poorly constrained groups, we took a more conservative approach
138 and only retained the events that had at least 5 other events within 1.5 *km* radius and
139 that had at least 4 differential travel time observations from 3 stations with cross-correlation

140 coefficient of 0.7 or higher. The final number of events after the double-difference relo-
141 cation was 86,657 for the on-fault group and 2,499 for the off-fault group, making the
142 total count of 89,156. List of events are provided in Table S3 and the epicenters are plot-
143 ted in Figure 1 along with the seismic stations, wastewater disposal wells, and produc-
144 tion wells. The production wells are colored based on their completion date, which can
145 either be the completion date of drilling or the date after hydraulic stimulation.

146 We estimated local magnitudes with station WHAR alone as 99 percent of the events
147 had phase picks from this station. We used the distance calibration factors from (Yoon
148 et al., 2017) on 1 *Hz* highpass filtered waveforms converted to standard Wood-Anderson
149 response. We measured the average of the zero to peak amplitudes of the two horizon-
150 tal components. We did not calculate magnitudes of the 866 events (fewer than 1% of
151 the total) not picked by station WHAR, and used the previously reported magnitude (M
152 4.7) for the February 28, 2011 mainshock.

153 **3 Evaluation of the Workflow**

154 Monthly counts of P and S arrival time picks returned by the picker, and those even-
155 tually used for locating the events as well as the ratio between the two (yield) are listed
156 in Table S2. The number of picks from station WHAR was the highest (P: 283,248, S:
157 280,331), but the overall yield was only around 0.3. The low yield is due to the limited
158 number and relatively poorer performance of the other stations that were simultaneously
159 in operation. There was a maximum of seven stations and an average of five during the
160 study period. Because we required each event to have at least 4 phase picks from 3 sta-
161 tions during phase association, events that were missed by a few stations were not as-
162 sociated. This is especially apparent when comparing pre- and post- September, 2010
163 in Table S2 where introducing additional stations increased the yield substantially. Note
164 that data quality from stations ARK1 and ARK2 was much lower than station WHAR,
165 which is apparent in the relatively low number of picks over the same period of opera-
166 tion. More stations with high quality data would have greatly increased the yield from
167 station WHAR and the total number of events in the catalog.

168 Figure 2 shows frequency-magnitude distribution of the events along the Guy-Greenbrier
169 fault plotted per batch of 1,000 with magnitude bin size of 0.1. The batch before the Febru-
170 ary sequence, i.e., the distribution on the left to the vertical line at February 15, 2011,

171 contains 337 events and the last batch contains 457 events. To indicate an approximate
 172 level of magnitude of completeness (M_C) for the figure, we used the b-value stability cri-
 173 terion (Cao & Gao, 2002), where the b-values were calculated using the maximum like-
 174 lihood estimation (Aki, 1965), and manually picked the cutoff magnitude that led to the
 175 first stable b-value as M_C . These are shown in solid line in Figure 2. Note that these
 176 are approximated values and are not intended for quantitative analysis, but rather to
 177 support the notion that our catalog is statistically homogeneous enough to make an un-
 178 biased interpretation.

179 **4 The Case for Two Sequences**

180 Figure 3(a) shows the evolution of seismicity along the Guy-Greenbrier Fault, i.e.,
 181 the events colored in grey and the events that coincide with Well X (API: 03045102980000),
 182 as a function of time versus distance from point O along the dashed line in Figure 1. The
 183 sequence along the Guy-Greenbrier Fault is composed of two distinct sub-sequences that
 184 are separated from each other both in space and time. We refer to the two sub-sequences
 185 as the July sequence and the February sequence. Figure 4 shows the evolution of the July
 186 sequence, the initial stage of the February sequence, and the rest of the February sequence
 187 in separate plots based on the distance from point O in Figure 1 versus depth. The events
 188 that occurred in February 15, 2011, which is the start date of the February sequence, are
 189 plotted with filled circles in Figure 4(b). We plot events with magnitude 3 or greater with
 190 unfilled circles scaled by their estimated source radius assuming a constant stress drop
 191 of 9.47 MPa (Huang et al., 2016).

192 The early stages of the July sequence are well explained by previous studies (Ogwari
 193 et al., 2016; Ogwari & Horton, 2016; Mousavi et al., 2017) suggesting it was caused by
 194 injection at Well 1 and 5. However, the explanation for the February sequence and the
 195 time delay between the two sequences is less clear. The separation of the two sequences
 196 was observed in the study by Horton (Horton, 2012) and he hypothesized that the Febru-
 197 ary sequence could have been influenced by the Enders Fault hydraulically connecting
 198 Well 2 and 5 with the Guy-Greenbrier Fault. Note that Well 2 is located to the west of
 199 the Guy-Greenbrier Fault and to the north of the Enders Fault (Figure 1) and was op-
 200 erating since April 2009 with a comparable injection rate to Well 1 (Well 1: $62663 \text{ m}^3/\text{month}$,
 201 Well 2: $54058 \text{ m}^3/\text{month}$) while the injection rate of Well 5 was much lower ($19580 \text{ m}^3/\text{month}$).
 202 The events detected by Huang and Beroza (Huang & Beroza, 2015) also clearly showed

203 the seismically less active period between late December, 2010 and early February, 2011.
204 However, they associated both sequences with Well 1 and 5 instead of suggesting an al-
205 ternative explanation. Later, Yehya and coworkers (Yehya et al., 2018) again associated
206 the February sequence to Well 1 and 5 but suggested that the time delay between the
207 two sequences could be caused by the Enders Fault providing a barrier to pore pressure
208 diffusion along the Guy-Greenbrier Fault.

209 The spatio-temporal evolution observed in our study suggests an alternative ex-
210 planation. The initiation point of the February sequence was more than a kilometer south
211 of the Enders Fault and spread bilaterally - north towards the Enders Fault and the July
212 sequence, as well as to the south. This is displayed in Figure 3(b) and Figure 4(b) in two
213 different views, and more explicitly in the supplementary animation (Movie S1). More-
214 over, the seismicity cluster CL1 in Figure 1 suggests that the medium between the En-
215 ders Fault and the Guy-Greenbrier Fault went through stress perturbation before the
216 February sequence potentially due to the pore pressure diffusion from Well 2. Note that
217 there are no production wells in the vicinity, which makes the possibility of these events
218 being triggered by hydraulic fracturing less likely. We suggest that pressure from Well
219 2 could also have diffused to the Guy-Greenbrier Fault, but not necessarily through the
220 Enders Fault, triggering the February sequence. Such remote triggering of seismicity is
221 also proposed and a well-accepted mechanism in Oklahoma and Kansas (Keranen et al.,
222 2014; Peterie et al., 2018).

223 Our observations also suggest that the July sequence effectively decayed before Wells
224 1 and 5 were shut down. The seismicity cluster near the north end that formed in early
225 March, 2011 is likely to be associated with Well X where its reported date of the first
226 treatment was February 28, 2011 with a completion date of March 7, 2011 (Figure 3).

227 Several clusters exterior to the Guy-Greenbrier Fault are also apparent in Figure
228 1. While the relative locations were improved by the double-difference algorithm, the ab-
229 solute locations of most of these clusters remain uncertain due to the poor azimuthal sta-
230 tion coverage. However, most clusters collocate with production wells both in space and
231 time, based on the well completion date, as discussed by Yoon and coworkers (Yoon et
232 al., 2017). Thus, it is likely that these clusters are associated with hydraulic fracturing
233 operations. The clusters on the west side of Well 3 (annotated with CL2 in Figure 1) and
234 to the northeast of the Guy-Greenbrier Fault (annotated with CL3 in Figure 1), how-

235 ever, do not collocate with any production wells in the vicinity. Similar to CL1, these
 236 could be associated with long-term injection of the nearby wastewater disposal wells, or
 237 to natural seismicity.

238 The depth versus distance plots in Figure 4 also exhibits multiple gaps in seismic-
 239 ity. The linear gap near the vertical line corresponding to the approximate intersection
 240 between the Guy-Greenbrier Fault and the Enders Fault in Figure 4(c) is especially no-
 241 table. While this is spatially correlated with the Enders Fault, precise information of the
 242 fault geometry is required for confirmation.

243 **5 Diffusion of Seismicity at Two Time Scales**

244 The diffusivity of a seismicity front with respect to a point injection source or other
 245 reference point is an important metric for analyzing induced seismicity (Shapiro et al.,
 246 2002; Segall & Lu, 2015). The February sequence showed a smoothly migrating front in
 247 both directions. Approximating the diffusivity with respect to the initiation point of the
 248 February sequence using $r = \sqrt{4\pi Dt}$ yields a diffusivity of $2.4 \text{ m}^2/\text{s}$ for the northern
 249 front and $1.6 \text{ m}^2/\text{s}$ for the southern front (Figure 3(b)).

250 The overall sequence is comprised of a hierarchy of multiple smaller arc-shaped pat-
 251 terns of seismicity in space-time. These smaller seismicity patterns propagate at a some-
 252 what faster rate over the fault than does the large-scale front. Two well-separated arc-
 253 structures that span the south end of the fault are shown in Figure 3(b) and their hypocen-
 254 ters are shown in 3(c) as examples. The implied diffusivity is correspondingly higher -
 255 $8 \text{ m}^2/\text{s}$ in both instances. It is not necessarily clear what quantity is diffusing, with pore
 256 fluid, stress, and aseismic slip being among the possibilities. Regardless of the cause, ob-
 257 serving such detailed features in the seismicity is only possible through the systematic
 258 cataloging of small earthquakes.

259 Yet another aspect of interest, which is most clearly revealed in the animation (Movie
 260 S1), is the propensity of parts of the fault to be illuminated by microearthquake activ-
 261 ity multiple times. Most prominent are parts of the fault from 3-7.5 *km* along strike that
 262 were initially active in February, but were reactivated from March 22-23, and even more
 263 dramatically from April 7-8. Such behavior could occur if small events were driven by
 264 otherwise aseismic slip.

265 6 Conclusion

266 Through the case study of the Guy-Greenbrier seismic sequence, we demonstrate
 267 that machine-learning-based analysis of earthquake sequences is now possible. The PhaseNet
 268 deep neural network arrival time picker can accelerate the front end of earthquake cat-
 269 alog development that in this instance led to precise locations of almost 90,000 events.
 270 Our results suggest that the Guy-Greenbrier seismicity consisted of two sub-sequences,
 271 that the second sequence may have been triggered by a wastewater disposal well previ-
 272 ously not implicated, and revealed diffusive patterns of seismicity at shorter time scales,
 273 and repeated illumination of large parts of the fault by seismicity.

274 Acknowledgments

275 This work is supported by the Stanford Center for Induced and Triggered Seismicity (SC-
 276 ITS) and the Department of Energy. Seismic data for stations ARK1, ARK2, ARK3,
 277 WHAR, X103, X201, X301, and X401 were downloaded from the Incorporated Research
 278 Institutions for Seismology (IRIS) at <http://ds.iris.edu/ds/> and data for stations
 279 Q201 and Q202 were obtained from P. Ogwari through S. M. Mousavi. Well informa-
 280 tion was obtained from the Arkansas Oil and Gas Commission (AOGC) at [http://www](http://www.aogc.state.ar.us/welldata/wells)
 281 [.aogc.state.ar.us/welldata/wells](http://www.aogc.state.ar.us/welldata/wells) and ENVERUS at <https://www.enverus.com>.
 282 Y. Park wants to thank C. E. Yoon for sharing her code scripts for local magnitude and
 283 differential travel time calculations.

284 References

- 285 Aki, K. (1965). Maximum likelihood estimate of b in the formula $\log n = a - bm$ and
 286 its confidence limits. *Bull. Earthq. Res. Inst., Tokyo Univ.*, *43*, 237–239.
- 287 Bergen, K. J., Johnson, P. A., Maarten, V., & Beroza, G. C. (2019). Machine learn-
 288 ing for data-driven discovery in solid earth geoscience. *Science*, *363*(6433),
 289 eaau0323.
- 290 Cao, A., & Gao, S. S. (2002). Temporal variation of seismic b -values beneath north-
 291 eastern japan island arc. *Geophysical research letters*, *29*(9), 48–1.
- 292 Horton, S. (2012). Disposal of hydrofracking waste fluid by injection into subsur-
 293 face aquifers triggers earthquake swarm in central arkansas with potential for
 294 damaging earthquake. *Seismological Research Letters*, *83*(2), 250–260.
- 295 Huang, Y., & Beroza, G. C. (2015). Temporal variation in the magnitude-frequency

- 296 distribution during the guy-greenbrier earthquake sequence. *Geophysical Re-*
297 *search Letters*, *42*(16), 6639–6646.
- 298 Huang, Y., Beroza, G. C., & Ellsworth, W. L. (2016). Stress drop estimates of po-
299 tentially induced earthquakes in the guy-greenbrier sequence. *Journal of Geo-*
300 *physical Research: Solid Earth*, *121*(9), 6597–6607.
- 301 Johnson, C. E., Lindh, A., & Hirshorn, B. (1997). *Robust regional phase association.*
302 US Department of the Interior, US Geological Survey.
- 303 Keranen, K. M., Weingarten, M., Abers, G. A., Bekins, B. A., & Ge, S. (2014).
304 Sharp increase in central oklahoma seismicity since 2008 induced by massive
305 wastewater injection. *Science*, *345*(6195), 448–451.
- 306 Klein, F. W. (2002). *User's guide to hypoinverse-2000, a fortran program to solve*
307 *for earthquake locations and magnitudes* (Tech. Rep.). US Geological Survey.
- 308 Llenos, A. L., & Michael, A. J. (2013). Modeling earthquake rate changes in ok-
309 lahoma and arkansas: Possible signatures of induced seismicity. *Bulletin of the*
310 *Seismological Society of America*, *103*(5), 2850–2861.
- 311 Mousavi, S. M., Ogwari, P. O., Horton, S. P., & Langston, C. A. (2017). Spatio-
312 temporal evolution of frequency-magnitude distribution and seismogenic index
313 during initiation of induced seismicity at guy-greenbrier, arkansas. *Physics of*
314 *the Earth and Planetary Interiors*, *267*, 53–66.
- 315 Ogwari, P. O., & Horton, S. (2016). Numerical model of pore-pressure diffusion
316 associated with the initiation of the 2010–2011 guy–greenbrier, arkansas earth-
317 quakes. *Geofluids*, *16*(5), 954–970.
- 318 Ogwari, P. O., Horton, S. P., & Ausbrooks, S. (2016). Characteristics of in-
319 duced/triggered earthquakes during the startup phase of the guy–greenbrier
320 earthquake sequence in north–central arkansas. *Seismological Research Letters*,
321 *87*(3), 620–630.
- 322 Peterie, S. L., Miller, R. D., Intfen, J. W., & Gonzales, J. B. (2018). Earthquakes in
323 kansas induced by extremely far-field pressure diffusion. *Geophysical Research*
324 *Letters*, *45*(3), 1395–1401.
- 325 Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for
326 biomedical image segmentation. In *International conference on medical image*
327 *computing and computer-assisted intervention* (pp. 234–241).
- 328 Ross, Z. E., Meier, M.-A., & Hauksson, E. (2018). P wave arrival picking and first-

- 329 motion polarity determination with deep learning. *Journal of Geophysical Re-*
330 *search: Solid Earth*, 123(6), 5120–5129.
- 331 Segall, P., & Lu, S. (2015). Injection-induced seismicity: Poroelastic and earthquake
332 nucleation effects. *Journal of Geophysical Research: Solid Earth*, 120(7), 5082–
333 5103.
- 334 Shapiro, S. A., Rothert, E., Rath, V., & Rindschwentner, J. (2002). Characterization
335 of fluid transport properties of reservoirs using induced microseismicity. *Geo-*
336 *physics*, 67(1), 212–220.
- 337 Waldhauser, F. (2001). hypodd—a program to compute double-difference hypocenter
338 locations.
- 339 Waldhauser, F., & Ellsworth, W. L. (2000). A double-difference earthquake location
340 algorithm: Method and application to the northern hayward fault, california.
341 *Bulletin of the Seismological Society of America*, 90(6), 1353–1368.
- 342 Yehya, A., Yang, Z., & Rice, J. R. (2018). Effect of fault architecture and permeabil-
343 ity evolution on response to fluid injection. *Journal of Geophysical Research:*
344 *Solid Earth*, 123(11), 9982–9997.
- 345 Yoon, C. E., Huang, Y., Ellsworth, W. L., & Beroza, G. C. (2017). Seismicity during
346 the initial stages of the guy-greenbrier, arkansas, earthquake sequence. *Journal*
347 *of Geophysical Research: Solid Earth*, 122(11), 9253–9274.
- 348 Yoon, C. E., O'Reilly, O., Bergen, K. J., & Beroza, G. C. (2015). Earthquake de-
349 tection through computationally efficient similarity search. *Science advances*,
350 1(11), e1501057.
- 351 Zhang, M., Ellsworth, W. L., & Beroza, G. C. (2019). Rapid earthquake association
352 and location. *Seismological Research Letters*, 90(6), 2276–2284.
- 353 Zhu, W., & Beroza, G. C. (2018). Phasenet: a deep-neural-network-based seismic
354 arrival-time picking method. *Geophysical Journal International*, 216(1), 261–
355 273.

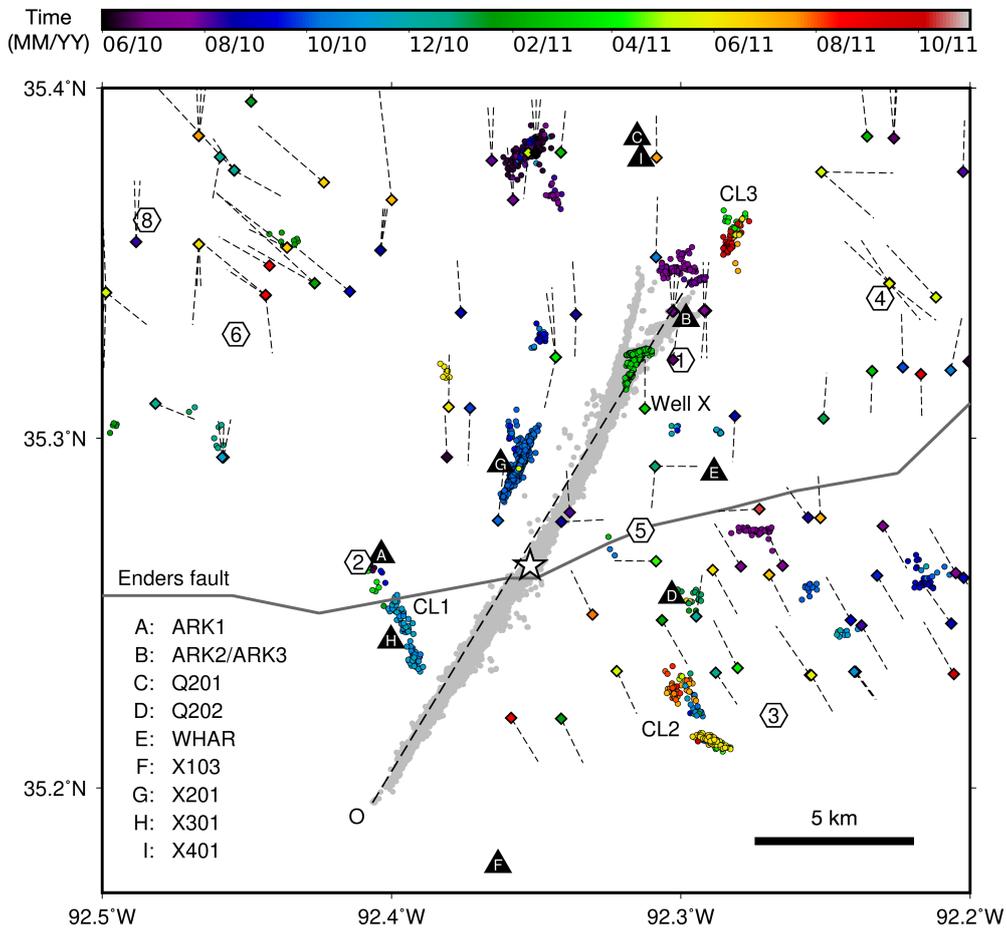


Figure 1. Epicenters of the 89,156 located events. Triangles are seismic stations, star is the February 28, 2011 mainshock, hexagons are wastewater disposal wells with well numbers based on Horton (Horton, 2012), and diamonds are production wells with each dashed line connecting the heel-toe. Events that are exterior to the Guy-Greenbrier Fault and the events that collocate with Well X (API: 03045102980000) in space and time are colored by their origin time. Production wells are colored by their completion date. See text for details.

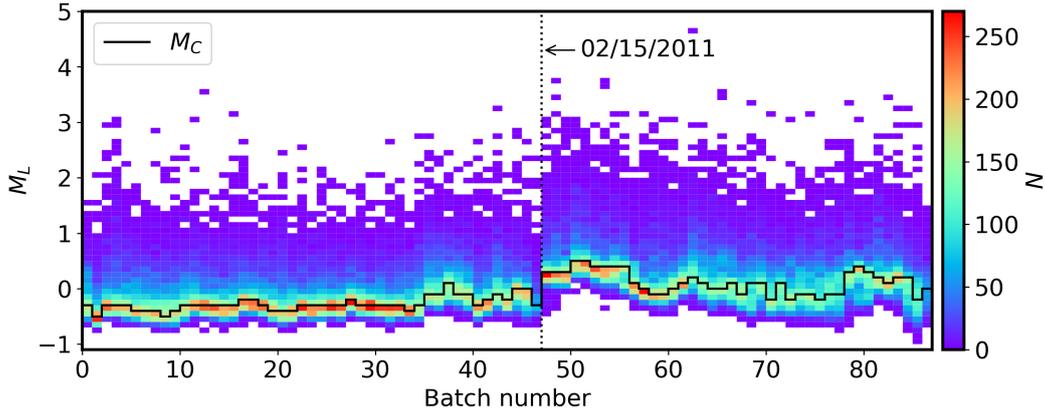


Figure 2. Frequency-magnitude distribution of the events along the Guy-Greenbrier Fault. Each batch contains 1,000 events except the batch before the February 15, 2011 line (337 events) and the last batch (457 events). The magnitude bin size was set to 0.1. Estimated magnitude of completeness (M_C) per batch is shown in solid line.

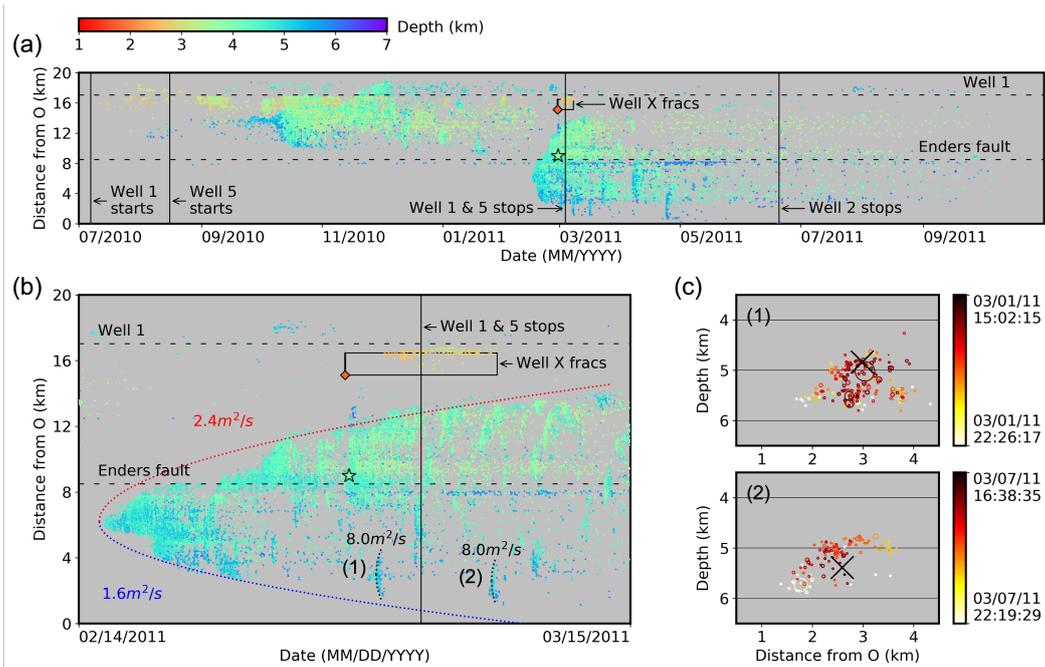


Figure 3. Events along the Guy-Greenbrier Fault plotted by their origin times versus the distance from point O in Figure 1 (a) from July 1, 2010 to October 31, 2011 and (b) from February 14, 2011 to March 14, 2011. Subplot (c) shows the hypocenters of the events forming the two arc-shaped structures in (b) scaled by their source radius. The first event in each plot is marked with \times . Numbers in (b) are approximated diffusivities for each seismicity front. See text for details.

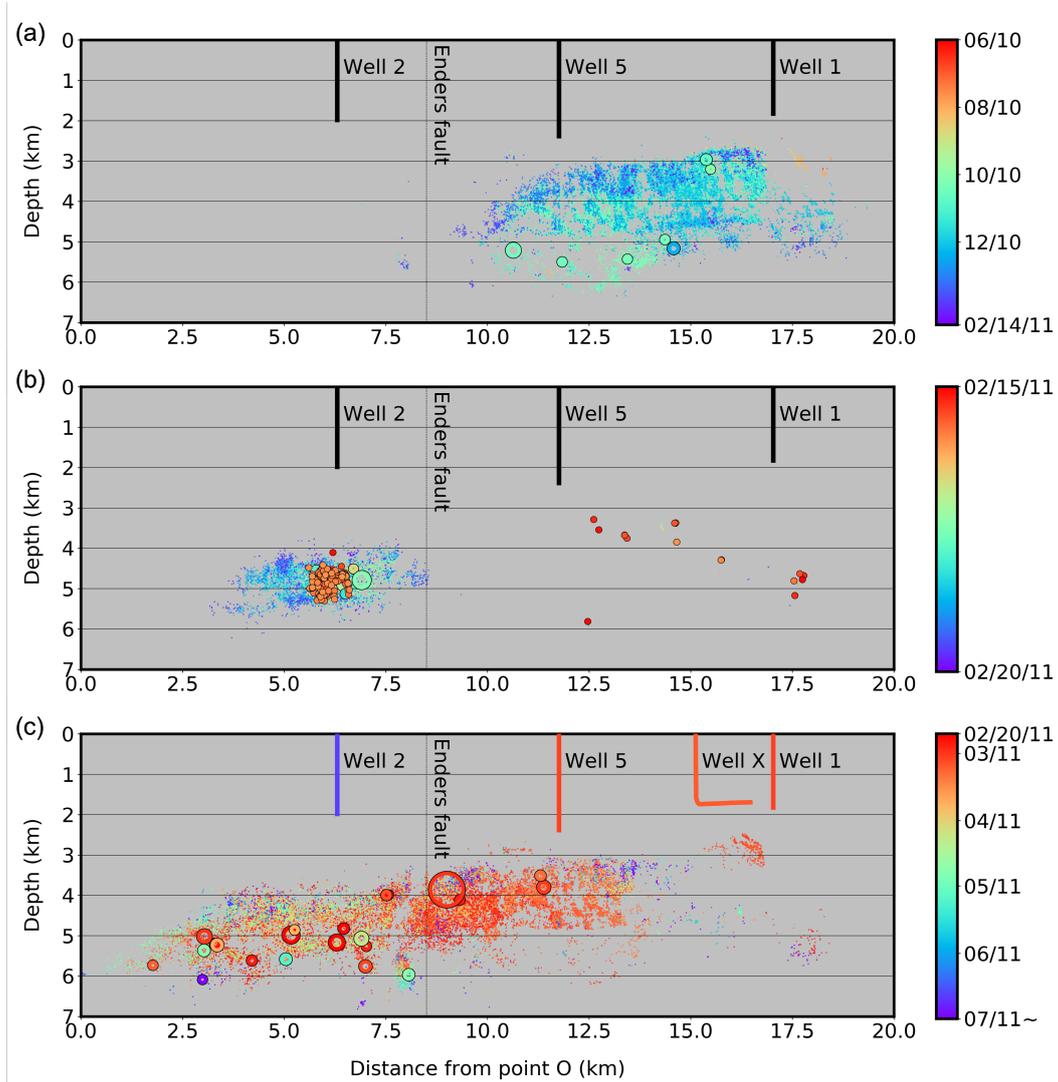


Figure 4. Events along the Guy-Greenbrier Fault plotted by the distance from point O in Figure 1 versus depth (a) from June 1, 2010 to February 14, 2011, (b) from February 15, 2011 to February 20, 2011, and (c) from February 20, 2011 to October 31, 2011. Events from February 15, 2011 (the initiation date of the February sequence) are plotted with filled circles in (b). Well 1, 2, and 5 are colored by their shutdown date and Well X is colored based on its completion date in (c). Events with magnitude 3 or greater are plotted in open circles scaled by their estimated source radius.