AGU100 ADVANCING EARTH AND SPACE SCIENCE

Geophysical Research Letters

RESEARCH LETTER
10.1029/2018GL079712

Key Points:
- Machine learning can model important characteristics of laboratory fault physics by training on finely resolved catalogs of slip events
- Fault physics becomes significantly harder to learn if catalogs are truncated at or above a critical magnitude of completeness

Supporting Information:
- Supporting Information S1

Correspondence to:
P. A. Johnson,
paj@lanl.gov

Citation:
Lubbers, N. E., Bolton, D. C., Mohd-Yusof, J., Marone, C., Barros, K. M., & Johnson, P. A. (2018). Earthquake catalog-based machine learning identification of laboratory fault states and the effects of magnitude of completeness. *Geophysical Research Letters*, *45*. https://doi.org/10.1029/2018GL079712

Received 21 JUL 2018
Accepted 7 NOV 2018
Accepted article online 12 NOV 2018

Earthquake Catalog-Based Machine Learning Identification of Laboratory Fault States and the Effects of Magnitude of Completeness

Nicholas Lubbers[1], David C. Bolton[2], Jamaludin Mohd-Yusof[3], Chris Marone[2], Kipton Barros[1], and Paul A. Johnson[4]

[1]Theoretical Division and CNLS, Los Alamos National Laboratory, Los Alamos, NM, USA, [2]Department of Geosciences, Pennsylvania State University, University Park, PA, USA, [3]Computer, Computational, and Statistical Sciences Division, Los Alamos National Laboratory, Los Alamos, NM, USA, [4]Geophysics Group, Los Alamos National Laboratory, Los Alamos, NM, USA

Abstract Machine learning regression can predict macroscopic fault properties such as shear stress, friction, and time to failure using continuous records of fault zone acoustic emissions. Here we show that a similar approach is successful using event catalogs derived from the continuous data. Our methods are applicable to catalogs of arbitrary scale and magnitude of completeness. We investigate how machine learning regression from an event catalog of laboratory earthquakes performs as a function of the catalog magnitude of completeness. We find that strong model performance requires a sufficiently low magnitude of completeness, and below this magnitude of completeness, model performance saturates.

Plain Language Summary Seismologists analyze faults in the earth by creating earthquake catalogs-records of the times, locations, and sizes of earthquakes. For decades, researchers have attempted to use the these catalogs to predict the timing and size of future earthquakes. Recently, researchers have found that machine learning algorithms can forecast the motion of the fault using subtle "creaking" sounds, both in the laboratory and in the real world. These creaking sounds had previously been thought to be noise and were not commonly cataloged as earthquake activity. We installed a very powerful sensor in a laboratory fault and created a very detailed catalog that captures very small quakes—small enough that they would have looked like noise to a less powerful sensor. We then used machine learning on this catalog to try and forecast the large laboratory earthquakes. We found that machine learning model is successful when small-enough events are part of the catalog. This says that subtle seismic sounds that look like noise may be very small earthquakes that were previously overlooked. These findings suggest that to improve earthquake forecasting, we might broaden our ideas of what signals to label as potential earthquakes and save in catalogs.

1. Introduction

Earthquake catalogs are a key seismological tool for problems involving location, precursors, and earthquake source parameters (Aki & Richards, 2002; Meier et al., 2017; Gutenberg & Richter, 1956; Scholz, 1968). Catalogs are fundamental to determining where, when, and how a fault has ruptured (Aki & Richards, 2002; Scholz, 2002), and they are central to studies of earthquake precursors and seismic hazard based on both field observations (Bouchon & Karabulut, 2008; Bouchon et al., 2013; Huang et al., 2017; Marsan et al., 2014; Meng et al., 2018) and theory (Aki & Richards, 2002; Ferdowsi et al., 2013; Kazemian et al., 2015). Laboratory studies of earthquake precursors also rely on event catalogs (Lei & Ma, 2014; Johnson et al., 2013; Rivière et al., 2018). Such studies include the temporal evolution of the Gutenberg-Richter *b*-value, which is a fundamental parameter relating laboratory and field studies of earthquake physics (Gulia et al., 2015; Kwiatek et al., 2014; Rivière et al., 2018; Scholz, 2015, 2002; Smith, 1980; Wang et al., 2016).

Laboratory-based studies of seismic hazard and earthquake forecasting have traditionally relied on earthquake catalogs. However, recent work shows that lab earthquakes can be predicted based on continuous acoustic data (Hulbert et al., 2018; Rouet-Leduc et al., 2017, 2018a). These works show that statistical characteristics of the continuous seismic signal emanating from lab fault zones can predict the

timing of future failure events as well as the frictional state of the fault zone. Moreover, recent work has extended this approach to field observation by showing that statistical characteristics of continuous seismicity recorded in Cascadia contain a fingerprint of the megathrust displacement rate that can be used to predict the timing of episodic tremor and slip events (Rouet-Leduc et al., 2018b). The methods are based on machine learning (ML) and rely exclusively on continuous measurements of acoustic emission (AE). Thus, a key question involves whether this approach can be extended to catalog-based measurements of seismic activity.

Applications of ML to seismology and geosciences are becoming increasingly common (Asim et al., 2018; Holtzman et al., 2018; Lary et al., 2016; Li et al., 2018; Ueki et al., 2018). Here we apply the ML random forest (RF) method (Breiman, 2001) to study laboratory earthquake catalogs. We find that as the catalog degrades, so does our ability to infer fault physics, as quantified by the ability to predict shear stress and the time to and since failure. We begin by describing the biaxial shearing experiment, our methods to construct a catalog from the continuous waveform data, the statistical descriptors we extract from the catalog, and the RF method used to make inferences from these descriptors. This results in successful models for estimating the fault physics variables using local-in-time catalog data. Then, from the original catalog we construct a series of progressively more degraded catalogs, each of which contains only the events that exceed a variable magnitude of completeness. We examine the performance of this method as a function of the magnitude of completeness and observe that the learned signatures of fault physics degrade dramatically if small enough events are not detected. However, the performance plateaus for low magnitudes of completeness; the smallest 80% events in the catalog can be dropped without sacrificing performance. Our results show that ML methods based on event catalogs can successfully model laboratory faults, with the caveat that predictive power requires a sufficiently complete catalog.

2. Data Collection

2.1. Biaxial Shear Experiment

We used data from laboratory friction experiments conducted with a biaxial shear apparatus (Johnson et al., 2013; Marone, 1998) pictured in Figure 1a. Experiments were conducted in the double direct shear configuration in which two fault zones are sheared between three rigid forcing blocks. Our samples consisted of two 5-mm-thick layers of simulated fault gouge with a nominal contact area of 10×10 cm^2. Gouge material consisted of soda-lime glass beads with initial particle size between 105 and 149 μm. Additional details about the apparatus and sample construction can be found in Anthony and Marone (2005), Rivière et al. (2018), and the supporting information Text S1. Prior to shearing, we impose a constant fault normal stress of 2 MPa using a servo-controlled load-feedback mechanism and allow the sample to compact. Once the sample has reached a constant layer thickness, the central block is driven down at constant rate of 10 μm/s. In tandem, we collect an AE signal continuously at 4 MHz (Figure 1b, yellow curve) from a piezoceramic sensor embedded in a steel forcing block ≈22 mm from the gouge layer.

Figure 1c shows the shear stress measured on the fault interface for a full experiment. Experiments begin with a run-in stage where the shear stress increases and macroscopic shearing of the fault zone transitions from stable sliding to stick-slip failure. The repetitive cycles of loading and failure represent laboratory seismic cycles and transition from periodic to aperiodic as a function of load-point displacement in our experiments (Anthony & Marone, 2005; Johnson et al., 2013). Here we focus on aperiodic slip cycles and measure the slip history for a set of 23 failure events that occur over 268 seconds (Figure 1c). We define large failure events as times for which stress drop exceeds 0.05 MPa within 1 ms.

2.2. Event Catalog

Our event catalog is built from the acoustic signal (example in Figure 1b, yellow) during the time interval noted in Figure 1c, inset. The catalog is composed of event amplitudes $A(t_k)$ and times t_k, where k is an event index, built following methods described in Rivière et al. (2018) and Text S2. We define laboratory magnitudes $m_k = \log_{10} A(t_k)$. The catalog contains $N_{\text{tot}} \approx 3.8 \times 10^5$ slip events with laboratory magnitudes that range from about 1.0 to 4.2. Our lab magnitudes are based only on acoustic signal amplitude, and although they do not connect directly to earthquake magnitude, previous works have demonstrated the connection between seismic b-value and lab frequency magnitude statistics (Scholz, 2002).

To visualize the distribution of magnitudes, we construct the cumulative event distribution

$$N(M) = \sum_{m_k \geq M} 1. \tag{1}$$

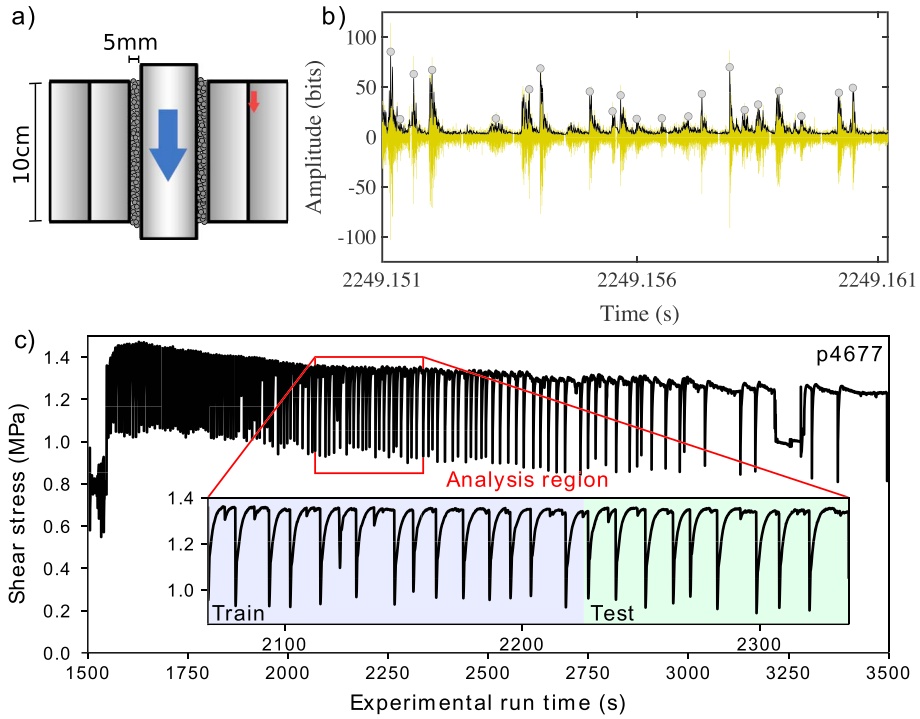

Figure 1. (a) Diagram of biaxial shear apparatus. The blue arrow indicates the direction of shear, and the red arrow locates the piezoceramic sensor that records acoustic emission (AE). (b) Example cataloged events (gray circles) derived from continuously recorded AE (yellow) using the statistics of the smoothed signal envelope (black). (c) Observed shear stress for an experiment at fixed strain rate. Sharp drops in stress correspond to failure events. (inset) The data segment analyzed in this paper. An event catalog is constructed from AE data sampled at 4 MHz. Our models are constructed using event data from the training segment (blue) and evaluated on data from the test segment (green).

That is, $N(M)$ is the count of all events with magnitude at least M. We plot $\log N(M)/N_{\text{tot}}$ in Figure 2a and observe reasonable Gutenberg-Richter scaling with a b-value of 1.66.

Figure 2b shows the rate of events as a function of time to failure (TTF), averaged over all slip cycles. The rate of events increases significantly as failure is approached, similar to what has been observed in prior works (Goebel et al., 2013; Johnson et al., 2013; Ponomarev et al., 1997; Rivière et al., 2018; Weeks et al., 1978). Figure 2c shows that the median event magnitude changes steadily but slowly from approximately $M_{\text{median}} = 1.4$ to $M_{\text{median}} = 1.5$ as a function of time during the lab seismic cycle. The trends of the catalog event count and size reflected in Figures 2b and 2c motivate the creation of ML features that describe the evolution event count and size in more detail.

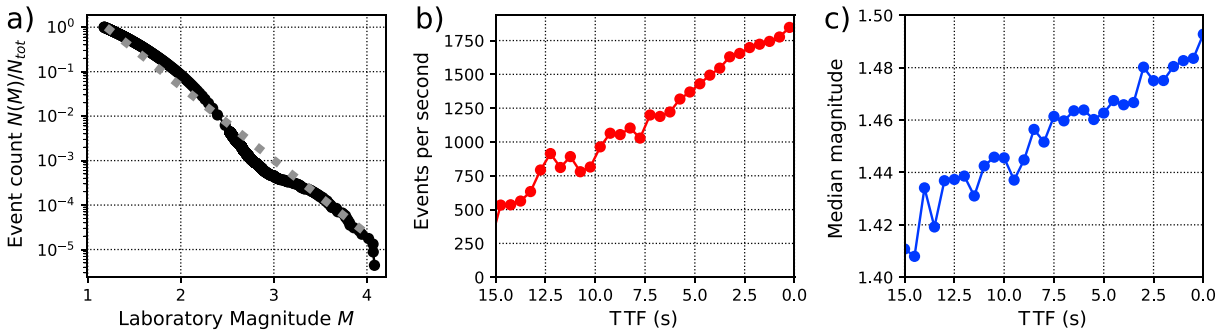

Figure 2. a) Normalized count of training events with magnitude exceeding M. The dashed line represents ideal Gutenberg-Richter scaling with a b-value of 1.66. **b)** Event rates vs. TTF stacked over all 23 failure cycles. **c)** Stacked event statistics for median event magnitude vs. time to failure. TTF = time to the next failure.

3. ML Methods

3.1. Regression With RFs

Abstractly, a regression model maps input points $\mathbf{X}_i$ to continuous output labels $\hat{y}_i$. Input points are represented by a feature vectors $\mathbf{X}_i = (X_{i,1}, X_{i,2}, \ldots X_{i,N_{\text{features}}})$ that are designed to capture the relevant information about $\hat{y}_i$. The training data set $\{(\mathbf{X}_i, y_i)\}$ consists of many points $\mathbf{X}_i$ associated with true labels y_i. From this training set, one aims to learn a model that is successful at predicting labels $\hat{y}_i$ for new data points that were unseen within the training process. In this work, the features will be statistics extracted from the catalog, the labels will describe the macroscopic state of the fault (such as shear stress); both are indexed by time window i.

We use the RF algorithm (Breiman, 2001) as it is implemented in the scikit-learn package (Pedregosa et al., 2011). A single RF model contains many decision trees, each of which is a simple regression model constructed stochastically from the training data. To make a prediction, the decision tree begins at its root node and asks a series of yes/no questions. Each question has the following form: Is some input feature above or below some threshold value? Eventually a unique leaf node is reached, from which the decision tree makes its prediction. The RF makes a prediction by averaging over many decision trees and tends to be more robust than any individual tree. Additional details about our training procedure are available in Text S3. We investigated other regression methods (Friedman, 2001; Hoerl & Kennard, 1970; Tibshirani, 1996; Zou & Hastie, 2005) besides the RF, and although all were successful to some extent, none offered definitive improvement over the RF (see Table S1).

A regression model is successful if its predicted labels $\hat{y}_i$ agree with the true labels y_i for data points in the testing data set, which is separate from the training data set. We use the dimensionless R^2 value to quantify the performance of the model,

$$R^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (\bar{y} - y_i)^2}, \tag{2}$$

where $\bar{y}$ is the mean output label for the testing data set. Let us review the significance of the R^2 metric, which is maximized at $R^2 = 1$, and can take on arbitrarily low values. Using the mean label as a constant predictor for all data points, $\hat{y}_i = \bar{y}$, yields $R^2 = 0$. Thus, $R^2 < 0$ indicates that a model performs worse than a constant learned only from the labels, without utilizing the features; if $R^2 \leq 0$, a model has failed to discover useful correlations between the features and the labels. The condition $R^2 = 1$ would indicate that a model makes perfect predictions for all labels in the test set.

3.2. Feature and Label Creation

Our task is to build a regression model that uses the event distribution within a local time window to understand and make predictions about the stress state and times to and from large failure events. Each time window is labeled, and predictions of the labels will depend only on features from within the window. Thus, each test prediction is independent of prior predictions, and the model cannot exploit the quasi-periodicity of the fault state.

There is significant interest in predicting the TTF for an upcoming earthquake event. Rouet-Leduc et al. (2017) and Rouet-Leduc et al. (2018a) showed that the continuous AE from a fault is remarkably predictive of the TTF and instantaneous friction. In both previous works, $\mathbf{X}_i$ is a collection of statistical features extracted from the continuous AE signal for the time window. Here we build features for each time window using only the event catalog, rather than the continuous acoustic signal.

We divide our full event catalog (cf. section 2.2) into subcatalogs that contain only the events occurring in nonoverlapping time windows of length $\Delta T = 1$ s. The feature vector contains information about events in the subcatalog for this time window. Our analysis region consists of 270 s, which we divide into 1-s windows. We train on the first 60% (159 windows) and test on the remaining 40% (111 windows).

We construct regression models for three labels: (a) the shear stress, (b) the time to the next failure event (TTF), and (c) the time since the last failure event (TSF). The shear stress label is assigned using the mean shear stress over that window. TTF is measured following the end of time window, and TSF is measured preceding the beginning of the time window. For windows including failures, both TTF and TSF are zero.

Abstractly, the features for each time window describe the cumulative statistics of event counts and amplitudes from the subcatalog evaluated at magnitude thresholds extracted from the full catalog. More precisely,

Table 1
R^2 Performance for Regression of Shear Stress, TSF, and TTF for Catalog-RF Models on Complete Catalogs

	Shear stress	TSF	TTF
$\mathbf{X}^{\text{count}}$	0.898	0.842	0.551
$\mathbf{X}^{\text{ampl}}$	0.848	0.882	0.612

Note. The choice of count- or amplitude-based features ($\mathbf{X}^{\text{count}}$ or $\mathbf{X}^{\text{ampl}}$) affects performance slightly. TSF = time since failure; TTF = time to failure.

the feature vector $\mathbf{X}_i$ for time window i will have components $X_{i,1}, X_{i,2}, \ldots X_{i,N_{\text{features}}}$. Each feature component $X_{i,j}$ is associated with a characteristic magnitude M_j. For each j, we define M_j to be the largest observed magnitude that satisfies

$$\frac{N(M_j)}{N_{\text{tot}}} \geq \alpha^j, \tag{3}$$

where the j in the right-hand side is applied as an exponent; α is a parameter of range $0 < \alpha < 1$ that controls the fineness of the magnitude bins and the total number of bins, N_{features}. Recall from equation (1) that $N(M)$ is the cumulative event count of our entire training catalog and that $N_{\text{tot}} \approx 3.8 \times 10^5$. In this work, we select $\alpha = 0.7$, and this leads to $N_{\text{feature}} = 35$ nonempty bins. We emphasize that the characteristic magnitudes M_j are derived from the entire training catalog, independent of window index i.

We now define the feature vector $\mathbf{X}_i$ for each window i from the subcatalog of event magnitudes $m_k^{(i)}$, where k indexes the events within the window. We test two simple schemes for $\mathbf{X}_i$, sensitive to the counts and amplitudes in the catalog. Mathematically, the jth components of these two feature vectors are

$$X_{i,j}^{\text{count}} = \sum_{m_k^{(i)} \geq M_j} 1 \tag{4}$$

$$X_{i,j}^{\text{ampl}} = \sum_{m_k^{(i)} \geq M_j} 10^{m_k^{(i)}} \tag{5}$$

where the sums run over all subcatalog events k with magnitude $m_k^{(i)}$ at least M_j. In words, $X_{i,j}^{\text{count}}$ is the count of events in the ith window with magnitude exceeding M_j (analogous to $N(M_j)$, but restricted to subcatalog i). The features $X_{i,j}^{\text{ampl}}$ measure the total amplitude of all events in window i above the magnitude M_j.

3.3. Catalog Ablation Test

To understand how our catalog-RF model performs with decreasing information, we vary the magnitude of completeness of our catalog by discarding all events with magnitude below a cutoff, M_{cut}. In practice, we implement this ablation by removing features $X_{i,j}$ for indices j that correspond to magnitudes M_j below the cutoff M_{cut}. We can then quantify the performance as a function of the ablation cutoff M_{cut}.

4. Results

4.1. Performance on Full Catalogs

Table 1 shows the R^2 performance of our catalog-RF models for shear stress, TSF, and TTF, trained using the full catalogs. The TTF models ($R^2 = 0.551$, $R^2 = 0.612$) do not capture fine details. The TTF prediction is worse than the TSF and shear stress models ($R^2 > 0.8$)—in other words, it is harder to predict the future than to estimate aspects of the past or present. Table 1 also indicates that the cumulative counts $\mathbf{X}^{\text{count}}$ are better for predicting the shear stress, and the cumulative amplitudes $\mathbf{X}^{\text{ampl}}$ are slightly better for TTF and TSF. The count-based catalog features predict shear stress with an accuracy of $R^2 = 0.898$, nearly equal to the continuous acoustic approach of Rouet-Leduc et al. (2018a), who report $R^2 = 0.922$ using the same data (i.e., same train and test segments on the same experiment). We note that although the same data are analyzed, the methods cannot be compared exactly; for example, the time window in Rouet-Leduc et al. (2018a) is 1.33 s, and each time window has 90% overlap with the previous window. (Recall that in our work the time window length is 1 s, and windows do not overlap.) Figure 3 shows the predictions compared to the true labels over time for the amplitude-based features $\mathbf{X}^{\text{ampl}}$, as well as scatter-plots showing a comparison of the true and predicted labels. Figure S1 shows that models perform nearly as well for shear stress and TSF when analyzing a larger

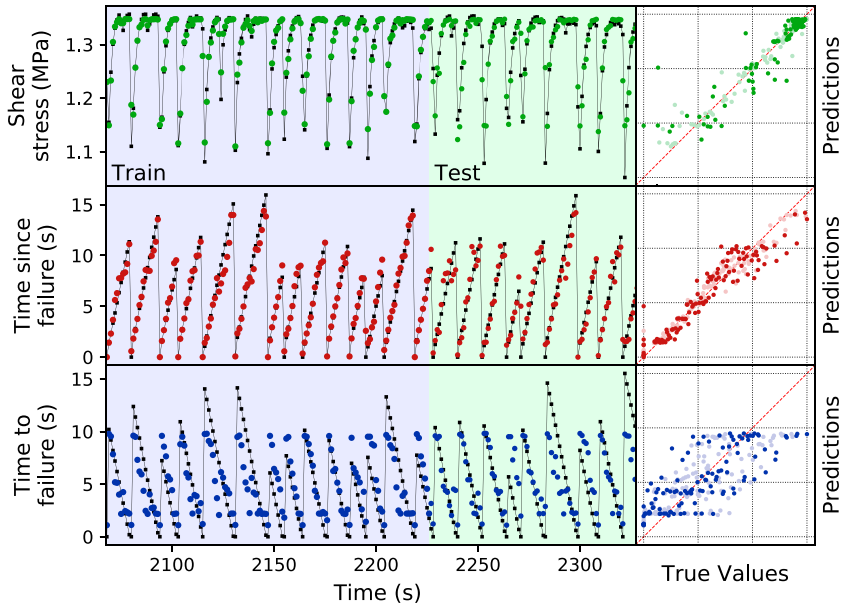

Figure 3. Regression performance for the catalog-ML model using amplitude-based features. The rows show performance for shear stress, TSF, and TTF, respectively. The left panel of each row shows predictions on training and testing data over time. True values are shown as black squares, and predicted values as colored circles. Time windows for features are contiguous and nonoverlapping. Each point is plotted at the end of the time window represented. The right panel of each row plots the prediction versus true values for each label. Testing data are shown with dark markers, and training data are shown with faint markers. ML = machine learning; TSF = time since failure; TTF = time to failure.

region of 573 s of data (44 failure cycles), but with reduced performance for TTF. This fact can be accounted for by drift in cycle length over time; TTF performance is restored by randomly assigning cycles to testing data from throughout the entire data analysis region (Figure S2).

4.2. Performance on Ablated Catalogs

Next we build our catalog-RF models on ablated catalogs that include events only above an imposed magnitude of completeness, M_{cut}. For this study, we use the amplitude features $\mathbf{X}^{ampl}$ of equation 5. Figure 4, left side, shows the R^2 performance as a function of the imposed magnitude cutoff. As one might expect, catalog-RF models generally perform worse when provided less information. However, it is interesting to note that R^2 decays relatively slowly until M_{cut} reaches about 2.0, after which the performance drops precipitously, plateauing near magnitude 3.0. In this regard we say that $M_{cut} = 2.0$ provides a sufficient magnitude of completeness for the catalog-RF model with the given window size. In Figure 4, right side, we plot R^2 as a function

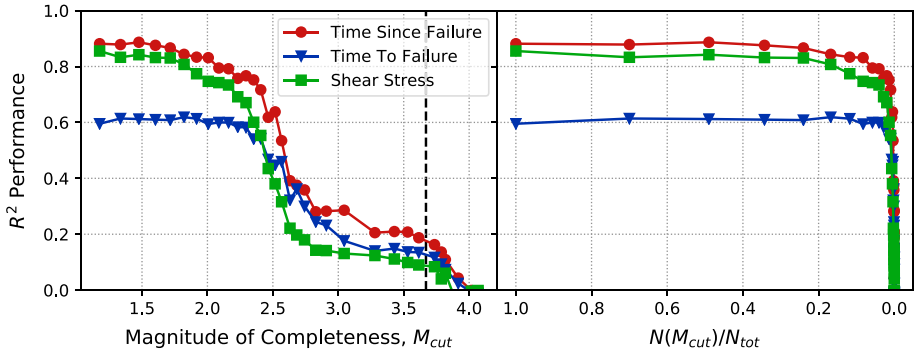

Figure 4. The accuracy of the catalog-RF model decreases as training subcatalogs are artificially limited to varying magnitudes of completeness, M_{cut}. (left) R^2 performance versus M_{cut}. The drop of R^2 for M_{cut} between 2.0 and 3.0 indicates that this range of event magnitudes plays a key role in our RF models. The dashed vertical line at $M_{cut} = 3.67$ denotes the smallest event magnitude associated with a large stress drop. (right) R^2 is plotted against $N(M_{cut})/N_{tot}$, that is, the fraction of events remaining in the full catalog after ablation. One observes that most events contribute very little to the catalog-RF performance. RF = random forest.

of the remaining event fraction $N(M_{cut})/N_{tot}$. This highlights that the performance drop at $M_{cut} \approx 2.0$ corresponds to ablating $\approx 80\%$ of the events from the catalog. Likewise, the sharp decay in performance indicates that events $M \lesssim 2.5$ are required to make an accurate determination of the fault state. The performance of the method for $M_{cut} \gtrsim 3$ plateaus at a nonzero value of R^2 because the algorithm is still able to differentiate between windows that contain a large failure and those that do not, but is otherwise unable to discern useful differences between subcatalogs.

A study of the ablation curve as a function of window size (Figure S3) shows that the sharp decay in performance shifts to smaller cutoff magnitudes as the window size is decreased; while regression can be performed with small window sizes, the sufficient magnitude of completeness is lower. This leads us to hypothesize that the sharp decay in performance may be related to the number of events within a window; to keep the number of events in a window fixed, one must lower the magnitude of completeness along with the window size.

5. Conclusions

We analyze a dense acoustic event catalog obtained from biaxial shearing friction experiment. The device produces dozens of aperiodic stick-slip events during an experiment. Stacking catalog statistics over many cycles shows that as failure approaches, events are more common and typically larger (Figure 2). We then use a ML workflow to show that physical characteristics of the fault (shear stress, TSF, and TTF) can be obtained from catalog statistics over a short time window using a RF regression model. Previous work employed the continuous AE and showed remarkable ability to forecast failure (Rouet-Leduc et al., 2017), as well as to infer the instantaneous shear stress and friction (Rouet-Leduc et al., 2018a). Here we demonstrate that the continuous acoustic waveform is not needed if a catalog with sufficient completeness is available. Our models achieve similar accuracy to the continuous approach.

We note, however, that our catalogs are extraordinarily well resolved. To create these catalogs, we recorded AE at very high frequency (4 MHz) using a sensor very near the fault (distance $\approx$ 22 mm). One would not expect such fidelity in real earthquake catalogs, except perhaps if the sensors happened to be located very close to the slip patch. We find that the smallest events in our catalog contribute little to the prediction accuracy; increasing M_{cut} from 1.0 to 2.0 only slightly diminishes the catalog-RF model performance (visible in Figure 4).

The fact that the smallest events are not required offers hope that a catalog-based ML approach could eventually contribute to the understanding of natural seismicity. We note a few potential differences between laboratory conditions and natural seismicity that may pose challenges. Laboratory data corresponds to a single isolated fault. Tectonic fault zones are typically comprised of many subfaults, and predictions may be vastly more difficult. While precursors are frequently observed in laboratory studies, they are not as reliably observed in natural faults. Moreover, large seismic cycles in the earth are far slower (Ellsworth et al., 1981; Nelson et al., 2006), and thus, fewer cycles are available to train ML models.

We note that Rouet-Leduc et al. (2018b) have applied the continuous approach developed in the laboratory to slow slip in Cascadia with surprisingly good results. Thus we are hopeful that the catalog-based approach described here may prove fruitful to the study of real earthquakes, given sufficient catalog quality. It may have particular importance in regions where continuous data are not yet available, where the continuous data are quite noisy, or for examining archival catalogs.

Acknowledgments
This work is supported by institutional support (LDRD) at the Los Alamos National Laboratory. We gratefully acknowledge the support of the Center for Nonlinear Studies. C. M. was supported by the National Science Foundation (NSF-EAR1520760) and the Department of Energy (DE-EE0006762). We thank Andrew Delorey, Robert A. Guyer, Bertrand Rouet-Leduc, Claudia Hulbert, and James Theiler for productive discussions. The data used here are freely available from the Penn State Rock Mechanics lab at http://www3.geosc.psu.edu/~cjm38/.

References

Aki, K., & Richards, P. G. (2002). *Quantitative seismology* (2nd Ed.). Sausalito, CA: University Science Books.
Anthony, J. L., & Marone, C. (2005). Influence of particle characteristics on granular friction. *Journal of Geophysical Research*, *110*, B08409. https://doi.org/10.1029/2004JB003399
Asim, K. M., Idris, A., Iqbal, T., & Martínez-Álvarez, F. (2018). Seismic indicators based earthquake predictor system using genetic programming and adaboost classification. *Soil Dynamics and Earthquake Engineering*, *111*, 1–7.
Bouchon, M., Durand, V., Marsan, D., Karabulut, H., & Schmittbuhl, J. (2013). The long precursory phase of most large interplate earthquakes. *Nature Geoscience*, *6*(4), 299–302.
Bouchon, M., & Karabulut, H. (2008). The aftershock signature of supershear earthquakes. *Science*, *320*(5881), 1323–1325.
Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.
Ellsworth, W. L., Lindh, A. G., Prescott, W. H., & Herd, D. G. (1981). The 1906 San Francisco earthquake and the seismic cycle. In D. W. Simpson & P. G. Richards (Eds.), *Earthquake Prediction* (pp. 126–140). Washington, DC: American Geophysical Union.
Ferdowsi, B., Griffa, M., Guyer, R. A., Johnson, P. A., Marone, C., & Carmeliet, J. (2013). Microslips as precursors of large slip events in the stick-slip dynamics of sheared granular layers: A discrete element model analysis. *Geophysical Research Letters*, *40*, 4194–4198. https://doi.org/10.1002/grl.50813
Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5), 1189–1232.

Goebel, T. H. W., Schorlemmer, D., Becker, T. W., Dresen, G., & Sammis, C. G. (2013). Acoustic emissions document stress changes over many seismic cycles in stick-slip experiments. *Geophysical Research Letters*, *40*, 2049–2054. https://doi.org/10.1002/grl.50507

Gulia, L., Tormann, T., Wiemer, S., Herrmann, M., & Seif, S. (2015). Short-term probabilistic earthquake risk assessment considering time-dependent b values. *Geophysical Research Letters*, *43*, 1100–1108. https://doi.org/10.1002/2015GL066686

Gutenberg, B., & Richter, C. F. (1956). Magnitude and energy of earthquakes. *Annali di Geofisica*, *9*(1), 1–15.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55–67.

Holtzman, B. K., Paté, A., Paisley, J., Waldhauser, F., & Repetto, D. (2018). Machine learning reveals cyclic changes in seismic source spectra in geysers geothermal field. *Science Advances*, *4*(5).

Huang, F., Li, M., Ma, Y., Han, Y., Tian, L., Yan, W., & Li, X. (2017). Studies on earthquake precursors in China: A review for recent 50 years. *Geodesy and Geodynamics*, *8*(1), 1–12.

Hulbert, C., Rouet-Leduc, B., Ren, C. X., Bolton, D. C., Rivière, J., Marone, C., & Johnson, P. A. (2018). Estimating the physical state of a laboratory slow slipping fault from seismic signals. *arXiv preprint arXiv*, *1801*, 07806.

Johnson, P. A., Ferdowsi, B., Kaproth, B. M., Scuderi, M., Griffa, M., Carmeliet, J., et al. (2013). Acoustic emission and microslip precursors to stick-slip failure in sheared granular material. *Geophysical Research Letters*, *40*, 5627–5631. https://doi.org/10.1002/2013GL057848

Kazemian, J., Tiampo, K. F., Klein, W., & Dominguez, R. (2015). Foreshock and aftershocks in simple earthquake models. *Physical Review Letters*, *114*(8), 088501.

Kwiatek, G., Goebel, T. H. W., & Dresen, G. (2014). Seismic moment tensor and b value variations over successive seismic cycles in laboratory stick-slip experiments. *Geophysical Research Letters*, *41*, 5838–5846. https://doi.org/10.1002/2014GL060159

Lary, D. J., Alavi, A. H., Gandomi, A. H., & Walker, A. L. (2016). Machine learning in geosciences and remote sensing. *Geoscience Frontiers*, *7*(1), 3–10.

Lei, X., & Ma, S. (2014). Laboratory acoustic emission study for earthquake generation process. *Earthquake Science*, *27*(6), 627–646.

Li, Z., Meier, M.-A., Hauksson, E., Zhan, Z., & Andrews, J. (2018). Machine learning seismic wave discrimination: Application to earthquake early warning. *Geophysical Research Letters*, *45*, 4773–4779. https://doi.org/10.1029/2018GL077870

Marone, C. (1998). The effect of loading rate on static friction and the rate of fault healing during the earthquake cycle. *Nature*, *391*, 69–72.

Marsan, D., Helmstetter, A., Bouchon, M., & Dublanchet, P. (2014). Foreshock activity related to enhanced aftershock production. *Geophysical Research Letters*, *41*, 6652–6658. https://doi.org/10.1002/2014GL061219

Meier, M.-A., Ampuero, J. P., & Heaton, T. H. (2017). The hidden simplicity of subduction megathrust earthquakes. *Science*, *357*(6357), 1277–1281.

Meng, X., Yang, H., & Peng, Z. (2018). Foreshocks, b value map, and aftershock triggering for the 2011 M_w 5.7 Virginia earthquake. *Journal of Geophysical Research: Solid Earth*, *123*, 5082–5098. https://doi.org/10.1029/2017JB015136

Nelson, A. R., Kelsey, H. M., & Witter, R. C. (2006). Great earthquakes of variable magnitude at the Cascadia subduction zone. *Quaternary Research*, *65*(3), 354–365.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in python. *Journal of machine learning research*, *12*, 2825–2830.

Ponomarev, A., Zavyalov, A., Smirnov, V., & Lockner, D. (1997). Physical modeling of the formation and evolution of seismically active fault zones. *Tectonophysics*, *277*(1), 57–81.

Rivière, J., Lv, Z., Johnson, P., & Marone, C. (2018). Evolution of b-value during the seismic cycle: Insights from laboratory experiments on simulated faults. *Earth and Planetary Science Letters*, *482*, 407–413.

Rouet-Leduc, B., Hulbert, C., Bolton, D. C., Ren, C. X., Rivière, J., Marone, C., et al. (2018a). Estimating fault friction from seismic signals in the laboratory. *Geophysical Research Letters*, *45*, 1321–1329. https://doi.org/10.1002/2017GL076708

Rouet-Leduc, B., Hulbert, C., & Johnson, P. A. (2018b). Breaking Cascadia's silence: Machine learning reveals the constant chatter of the megathrust, arXiv preprint arXiv:1805.06689v1 [physics.geo-ph].

Rouet-Leduc, B., Hulbert, C., Lubbers, N., Barros, K., Humphreys, C. J., & Johnson, P. A. (2017). Machine learning predicts laboratory earthquakes. *Geophysical Research Letters*, *44*, 9276–9282. https://doi.org/10.1002/2017GL074677

Scholz, C. H. (1968). The frequency-magnitude relation of microfracturing in rock and its relation to earthquakes. *Bulletin of the Seismological Society of America*, *58*(1), 399.

Scholz, C. H. (2002). *The mechanics of earthquakes and faulting*. Cambridge, UK: Cambridge University Press.

Scholz, C. H. (2015). On the stress dependence of the earthquake b value. *Geophysical Research Letters*, *42*, 1399–1402. https://doi.org/10.1002/2014GL062863

Smith, W. D. (1980). The b-value as an earthquake precursor. *Nature*, *289*, 136–139.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*(1), 267–288.

Ueki, K., Hino, H., & Kuwatani, T. (2018). Geochemical discrimination and characteristics of magmatic tectonic settings: A machine-learning-based approach. *Geochemistry, Geophysics, Geosystems*, *19*, 1327–1347. https://doi.org/10.1029/2017GC007401

Wang, J.-H., Chen, K.-C., Leu, P.-L., & Chang, C.-H. (2016). Precursor times of abnormal b-values prior to mainshocks. *Journal of Seismology*, *20*(3), 905–919.

Weeks, J., Lockner, D., & Byerlee, J. (1978). Change in b-values during movement on cut surfaces in granite. *Bulletin of the Seismological Society of America*, *68*(2), 333.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *67*(2), 301–320.