Deep Learning Can Predict Laboratory Quakes From Active Source Seismic Data

Parisa Shokouhi[1], Vrushali Girkar[2], Jacques Rivière[1], Srisharan Shreedharan[3], Chris Marone[3], C. Lee Giles[2,4], and Daniel Kifer[2]

[1]Engineering Science and Mechanics, The Pennsylvania State University, University Park, PA, USA, [2]Computer Science and Engineering, The Pennsylvania State University, University Park, PA, USA, [3]Geosciences, The Pennsylvania State University, University Park, PA, USA, [4]Information Sciences and Technology, The Pennsylvania State University, University Park, PA, USA

Abstract Small changes in seismic wave properties foretell frictional failure in laboratory experiments and in some cases on seismic faults. Such precursors include systematic changes in wave velocity and amplitude throughout the seismic cycle. However, the relationships between wave features and shear stress are complex. Here, we use data from lab friction experiments that include continuous measurement of elastic waves traversing the fault and build data-driven models to learn these complex relations. We demonstrate that deep learning models accurately predict the timing and size of laboratory earthquakes based on wave features. Additionally, the transportability of models is explored by using data from different experiments. Our deep learning models transfer well to unseen datasets providing high-fidelity models with much less training. These prediction methods can be potentially applied in the field for earthquake early warning in conjunction with long-term time-lapse seismic monitoring of crustal faults, CO_2 storage sites and unconventional energy reservoirs.

Plain Language Summary Laboratory experiments and field observations show that wave velocity, amplitude and frequency vary systematically over time during seismic cycles. These wave characteristics drop before failure (shear stress drop) albeit at different times and thus are believed to contain precursory information about the upcoming failure event. Here, we continuously record ultrasonic data during a series of experiments designed to simulate earthquakes in the laboratory or laboratory quakes. We investigate whether machine learning can predict the occurrence of laboratory quakes from ultrasonic data. We apply XGBoost and a suite of deep learning methods to this data and present models that can accurately predict the laboratory quake timing, size or both. We compare the performance of different models in terms of accuracy and training time. Also, we interpret the developed models. Finally, we show that these models successfully transfer from one data set to another obtained using different experimental constraints. Consequently, the training time and amount of data necessary to develop models for new datasets are significantly reduced. The developed prediction models can be used for seismic hazard assessment and warning, safe management of CO_2 storage sites and unconventional energy reservoirs in conjunction with continuous and long-term seismic monitoring.

1. Introduction

Laboratory studies in rock mechanics, including friction experiments, have been transformational in our understanding of earthquake physics. Earthquake nucleation arises from frictional instability on fault planes that can be modeled at the macro-scale using small (second-order) changes in friction coefficient (Ampuero & Rubin, 2008; Dieterich & Kilgore, 1994; Marone, 1998; Ruina, 1983; Rubin & Ampuero, 2005). To help relate the overall frictional response to the numerous physical mechanisms taking place at the micro-scale (healing, wear development, grain sliding/crushing, among others), laboratory settings are often instrumented with ultrasonic transducers for recording acoustic emissions (AE) (equivalent to microseismicity in the field) (Goebel et al., 2017; Kwiatek et al., 2014; Lockner et al., 1991; McLaskey et al., 2014; Passelegue et al., 2013; Rivière et al., 2018; Marty et al., 2019; Scholz, 1968) and time-lapse active source monitoring to probe fault zone properties (Kaproth & Marone, 2013; Kilgore et al., 2012; Nagata et al., 2014, 2008; Scuderi et al., 2016; Shreedharan et al., 2020). Such instrumentation can provide temporal (Bolton et al., 2020; Lubbers et al., 2018; Passelegue et al., 2013) or spatio-temporal evolution (Dresen et al., 2020;

Goebel et al., 2017; McLaskey & Lockner, 2016; McLaskey et al., 2014; Trugman et al., 2020) of AE events (equivalent to seismic catalogs) as well as local changes in elastic wave velocity and/or amplitude (Brantut, 2018; Shreedharan et al., 2021).

We focus here on unstable laboratory faults and document a series of slow and fast stick-slips, analogous to seismic cycles in nature. Past studies have reported that wave velocities and amplitudes vary systematically along the seismic cycle, with an increase during the initial phase of elastic loading, followed by a distinct reduction prior to fault failure (Hedayat et al., 2014a, 2014b, 2018; Kaproth & Marone, 2013; Shreedharan et al., 2019, 2020), as the fault slowly unlocks. The precursor time at which wave velocity and amplitude start to decrease often differ (Shreedharan et al., 2020). This delay seems to be associated with the competing effect of stress increase as failure approaches. While wave amplitude seems to be primarily sensitive to slip rate on the fault, wave velocity is sensitive to both slip rate and surrounding stress (Shreedharan et al., 2020). During most of the pre-seismic phase, when slip rate is increasing but still relatively small, wave velocity continues to increase due to large surrounding stress, while wave amplitude has already started to decrease (see Shreedharan et al., 2021 for further details).

In recent years, machine learning (ML) has been used on AE data to predict shear failure during laboratory friction experiments. Unsupervised approaches were successful at clustering the initial loading phase from the critical phase that precedes failure (Bolton et al., 2019). Supervised approaches have shown that the timing (Hulbert et al., 2019; Rouet-Leduc et al., 2017), shear stress (Hulbert et al., 2019; Rouet-Leduc et al., 2018) as well as magnitude and duration (Hulbert et al., 2019) of laboratory earthquakes could be well predicted, and that the variance of AE time-series was by far the most predictive feature. These ML approaches have also been tested on field data (slow slip sequences along the Cascadia Subduction zone (Hulbert et al., 2020; Rouet-Leduc et al., 2019, 2020). Beyond the direct applicability to field prediction of seismic events (Niu et al., 2008), these approaches are of great interest at the laboratory scale, to help unravel the mechanisms that dictate earthquake nucleation. The large quantity of constrained laboratory data also enables one to carefully examine the capabilities/limits of these ML approaches—to test, for instance, how predictive models perform under differing experimental conditions.

In this study, we use ML on active source ultrasonic data to predict frictional failure based on the evolution of the ultrasonic wave features. We compare the performance of several models that predict shear stress, time to failure or both given wave velocity, amplitude and frequency. We demonstrate that deep learning models can learn the complex relationships between wave features and shear stress and accurately predict both the timing and size of an irregular sequence of laboratory quakes. Among the methods applied, long short-term memory (LSTM) network (Hochreiter & Schmidhuber, 1997) that takes advantage of the time evolution of the features produces the most accurate predictions followed by multilayer perceptron (MLP) and XGBoost. Prediction models that use both velocity and amplitude features perform better than those that only use either velocity or amplitude features. In addition, we demonstrate that a model trained on a data set collected from one experiment can be adapted to a different experiment by further training the model on a relatively small amount of extra data (without this pre-training from the first experiment, a much larger training data set is needed from the second experiment). This indicates that the models are generalized and transferable across different datasets.

2. Data and Methods

We use several machine learning methods to predict failure in a set of friction experiments with continuous recording of elastic wave characteristics. Details of the experiments were reported previously (Shreedharan et al., 2020, 2019). The friction experiments use a double direct shear (DDS) configuration and rough, Westerly granite surfaces coated with a thin layer of quartz powder. The fault normal stress was held constant at 10 MPa and a constant shear rate of 11 μm/s was imposed. The nominal frictional and real contact areas remain constant throughout the experiment (5×5 cm^2 for each fault). In these experiments the loading stiffness was adjusted to produce a spectrum of lab earthquakes from fast to slow (Shreedharan et al., 2019) and a complex slip history as expected near the frictional stability boundary (e.g., Leeman et al., 2016). We focus on measurements of shear stress and fault shear displacement measured with a displacement transducer mounted on the central block. All stresses and displacements are continuously recorded at 10 kHz

throughout the experiment and averaged to rates of 100 Hz. The experiments included continuous recording of elastic waves that traversed the lab fault zones and serve as probes of the evolution of fault elastodynamic properties. The elastic waves were generated and recorded with compressional wave piezoelectric transducers embedded within the loading platens of the DDS assembly (inset Figure 1a). Half-sine pulses of a nominal center frequency 500 kHz were transmitted every 1 ms across the two faults on one side, received by the other transducer within the opposite side block and digitized at 25 MHz throughout the experiment (Figures 1b and 1c). Three distinct features are extracted from each recorded ultrasonic waveform: Wave velocity, spectral amplitude (at 400 kHz, close to the center frequency of the waveforms) as well as the dominant frequency of the first-arrived wave packet, which is around 450 kHz. The systematic variation of these wave characteristics over the seismic cycle (as shown in Figure 1b for the amplitude) suggest that they carry precursory information about the stick-slip frictional failure of the laboratory faults. In Figure 1a, shear stress increases during initial load-up and reaches a plateau (stable-sliding regime) that is only interrupted by two shear unload/reload cycles at a load-point displacement u_l of 1.5 and 3 mm. Around $u_l = 7$ mm, the faults slowly transition to a stick-slip regime (Leeman et al., 2016). Complex behavior such as irregular stick-slips and period doubling can be observed, as shown in the inset of Figure 1a. Such complex sequences are likely more representative of observations made in nature and pose a greater challenge for predictive ML models, compared to periodic cycles.

We use ML techniques to predict the timing and size of laboratory quakes using the distinct features extracted from ultrasonic data recorded continuously throughout the experiment. The features are: Wave velocity, amplitude, and center frequency (Figure 2). The mathematical expressions describing each feature are given in Equations 1–4 below.

Wave velocity v_i is calculated by dividing the specimen thickness d_i by the sum of reference signal's time of flight TOF_0 and time delay Δt_i as defined in Equation 2

$$v_i = d_i/(TOF_0 + \Delta t_i) \tag{1}$$

$$\Delta t_i = \arg\max_{\tau} \int_{t_i+w_1}^{t_i+w_2} S_0(t)S_i(t+\tau)dt \tag{2}$$

where Δt_i is the time delay calculated using a cross correlation between S_0, the reference ultrasonic signal, and S_i, the ith recorded ultrasonic record at t_i within a finite window of size T extending from time w_1 to $w_2 = w_1 + T$. Here, we use a window of size $T = 4.4$ μs extending from $w_1 = 20.76$ μs to $w_2 = 25.16$ μs. This time window includes the first arrival wave packet arriving at $TOF_0 = 20.74$ μs (highlighted in red in the inset of Figure 1c). The reference signal S_0 is a waveform recorded at the beginning of the experiment, once shear stress reaches a plateau ($t = 2,065$ s and $u_l \approx 7.4$ mm in Figure 1). Time delays Δt_i can be calculated using cross-correlation because waveform shapes change little throughout the experiment (the cross-correlation coefficient remains always greater than 0.97).

The second feature is spectral amplitude A_i calculated as the amplitude (absolute value) of the Fourier transform ($\mathcal{F}$) of the windowed signal S_i at frequency $f = 400$ kHz.

$$A_i = |\mathcal{F},(S_i(t_i + w_1 : t_i + w_2))|_{f=400\ kHz} \tag{3}$$

The third feature corresponds to the central frequency f_i of the received wave packet that is, frequency at which the spectrum is maximum:

$$f_i = \arg\max_{f} |\mathcal{F},(S_i(t_i + w_1 : t_i + w_2))| \tag{4}$$

The pre-processing steps include smoothing the data by low-pass filtering of the raw feature time histories using a 10-point backward looking moving average window. The temporal evolution of these filtered features is shown in Figure 2 along a few seismic cycles. In addition, in order to remove the long term trends in data caused by the gradual development of wear product on the fault planes, a second set of features are
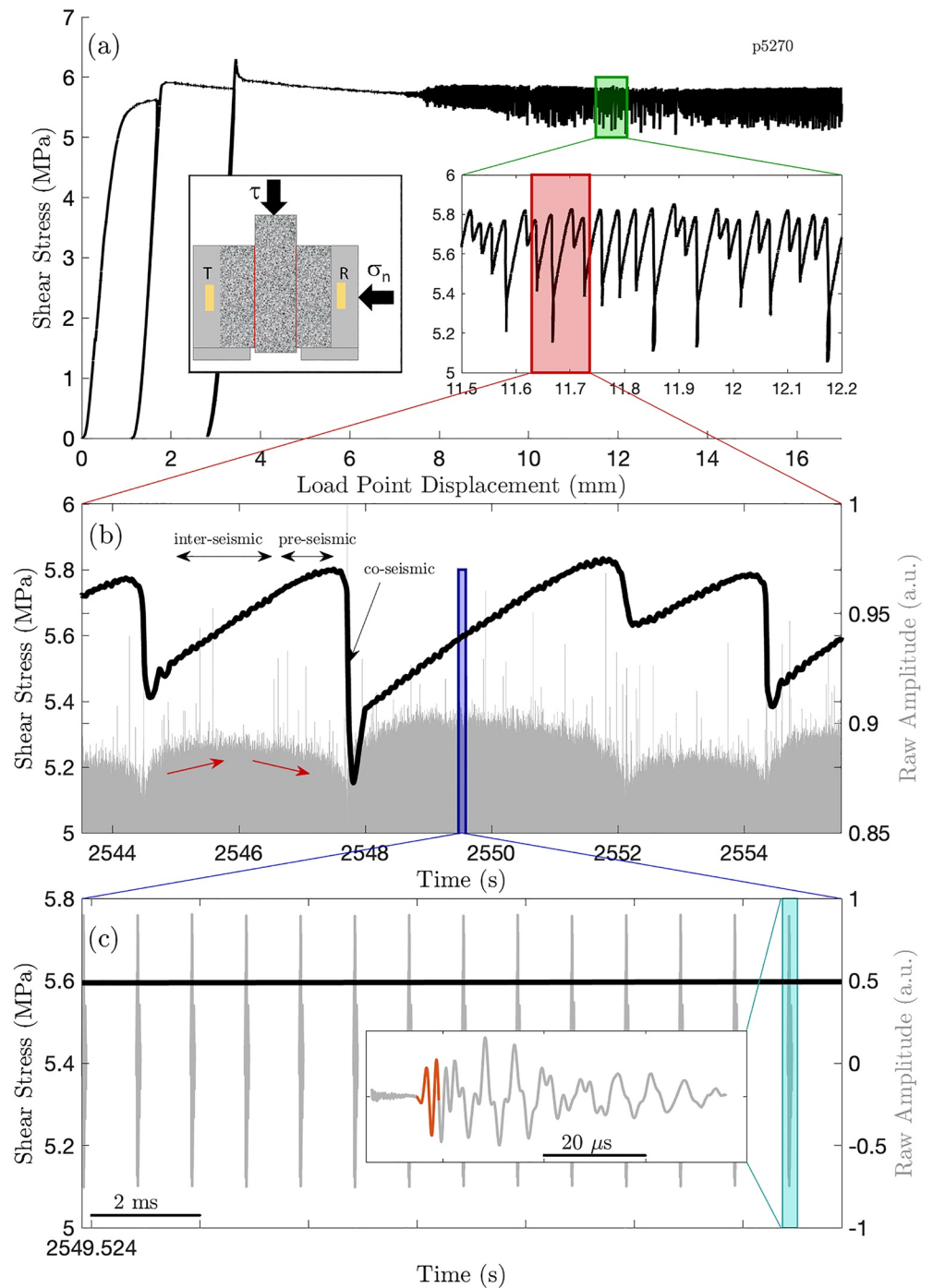
Figure 1. Experimental data and setup. (a) Shear stress (τ) as a function of load-point displacement u_l. Normal stress (σ_n) is held constant throughout the experiment. Following initial loading and a stable-sliding phase, the laboratory faults transition to a stick-slip regime (starting around $u_l = 8$ mm). One inset on the left shows a sketch of the double-direct shear configuration with two ultrasonic transducers (one transmitter (T) and one receiver (R) depicted as yellow rectangles) used to probe the laboratory faults (red vertical bars). The second inset shows a close-up view of a few stick-slips. Notice the irregular stick-slip cycles. (b) A close-view of shear stress for three cycles superimposed on the raw ultrasonic waveforms. Only the peak of these waveforms is shown (amplitude ranging from 0.85 to 1) to highlight the amplitude changes as shear stress evolves (red arrows). (c) A close-up view on a few waveforms (looking like gray vertical bars at this scale) that are recorded every millisecond. Shear stress is essentially constant at this time scale. The inset shows a close-up view of a single waveform. Features used as inputs for the machine learning (ML) models are extracted from the first-arrived wave packet (highlighted in red).
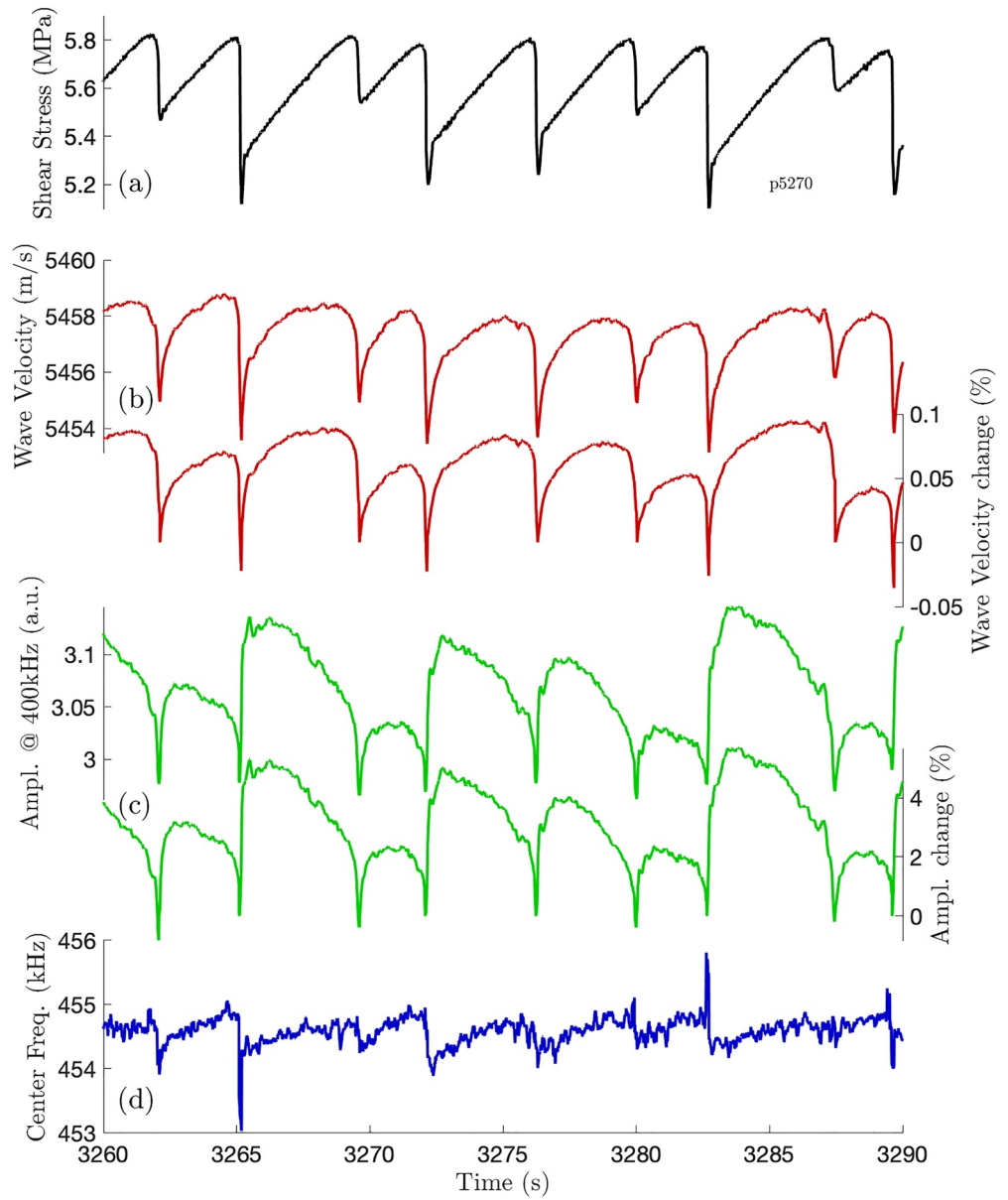
Figure 2. Temporal evolution of wave velocity, amplitude and frequency, all used as features in the machine learning (ML) models. (a) Shear stress. (b) Wave velocity (m/s) (left) and percent change in wave velocity referenced to the beginning of each seismic cycle (right). (c) Wave amplitude at 400 kHz (left) and percent change in wave amplitude referenced to the beginning of each seismic cycle (right). (d) Central frequency of the received ultrasonic pulse.

defined by normalizing two of the three features above (i.e., V_i and A_i) with respect to the corresponding value at the onset of each seismic cycle as described in Equation 5 for feature x_i:

$$\hat{x}_i = (x_i - x_0)/x_0 \tag{5}$$

where x_0 represents the feature value at the beginning of each seismic cycle. These normalized versions are shown for wave velocity and amplitude in Figures 2b and 2c (right-hand side y-axes). Finally, since the ultrasonic signals (and thus corresponding features) are recorded at a different rate (1 kHz, i.e., one ultrasonic waveform every millisecond) than shear stress and other mechanical data (sampled at 100 Hz), the two time series are matched using linear interpolation.

Our goal is to predict shear stress σ and time to failure (target variables) given the recorded ultrasonic features (input variables: V_i, A_i, f_i, $\hat{V}_i$, $\hat{A}_i$). It is worth noting that the relationship between these wave features and shear stress is highly nonlinear and complex (Figure 2), that is, no one-on-one mapping between either wave feature and shear stress exists (as highlighted in Figure S1 where plots of wave velocity and amplitude are shown as a function of shear stress). This makes the laboratory quake prediction challenging, requiring advanced statistical approaches (i.e., ML).

In order to implement ML algorithms, the data are divided into training, validation and testing sections. The training data are used to fit the machine learning model, the validation data are used for hyperparameter tuning (e.g., to help select values for non-differential parameters), while the testing data are used for evaluation of the model. Since the testing data are never used during the learning process, they provide a good estimate of the performance of the model. We note that the validation set is a crucial part of machine learning methodology. Hyperparameters should never be chosen using the test set as that would result in overfitting the test set (and non-reproducible results). Furthermore tuning hyperparameters using the validation set rather than the training set helps avoid overfitting to the training set. For example, the number of training epochs (for deep learning) or trees (in XGBoost) are parameters that can overfit the training data if set too large. By monitoring the error on the validation set, one can decide when additional training epochs (or trees) become harmful to model accuracy.

For an ML model to work efficiently, it is necessary to provide a large training data set. Hence, 72% of the data are reserved for the training purpose, another 8% are used as validation data and the remaining 20% are used for testing the model. Moreover, we treat this problem as a time-series prediction task. Thus, in order to take advantage of the sequential (time) dependency, we do not shuffle the data set while splitting it.

Three different supervised machine learning models are employed namely, XGBoost (Chen & Guestrin, 2016), an LSTM network (Hochreiter & Schmidhuber, 1997) and an MLP network. While implementing these models, we exploit the corresponding seismic history of 300 observations (equivalent to 3 s of data) to make predictions at the current time. Although the shear stress cycles are aperiodic, the average duration of one cycle in this data set is about 3–4 s. Hence, we provide the data history over a period of 3 s prior to the current time to predict the current shear stress and/or time to failure. That is, to predict shear stress (or time to failure) at observation number *n*, we only use features from observations *n-300, ..., n-1*. We implement three different models for each ML algorithm: single output model to predict shear stress, single output model to predict time to failure and a multi-output model to simultaneously predict shear stress and time to failure. Brief descriptions of each method together with the model architecture and hyperparameters are provided in the supplementary materials.

3. Results and Discussion

The first analysis step consists of establishing the baseline performance; we use basic methods to predict shear stress in order to provide a meaningful basis for comparison. We employ a number of baseline methods including simple linear regression as well as several autoregressive methods such as average method, rolling average, naive forecast, and seasonal persistence methods. Among the autoregressive methods, rolling average method gives the best test data R^2 score of -0.0003 which clearly indicates that the prediction task is not an autoregression problem but a feature-dependent regression problem. A linear regression analysis results in an R^2 score of 0.85054 for shear stress on the test data. This serves as the baseline performance in this study.

Next, we apply XGBoost, MLP, and LSTM to predict shear stress or/and time to failure from the five ultrasonic features defined in Equations 1–5 and shown in Figure 2 namely, wave velocity, normalized wave velocity, wave amplitude and normalized wave amplitude as well as the center frequency. All three models are trained and validated using the very same features. Figure 3 compares the performance of these three methods in predicting shear stress (see Figures 3a, 3b and 3c) or time to failure (see Figures 3d and 3e). For shear stress predictions (using single output models), LSTM and MLP models with test R^2 scores of 0.94825 and 0.92782 clearly outperform the XGBoost model that gives a test R^2 score of 0.90497. A close examination of the three predictions and the residuals in Figures 3b and 3c indicate that the XGBoost model especially
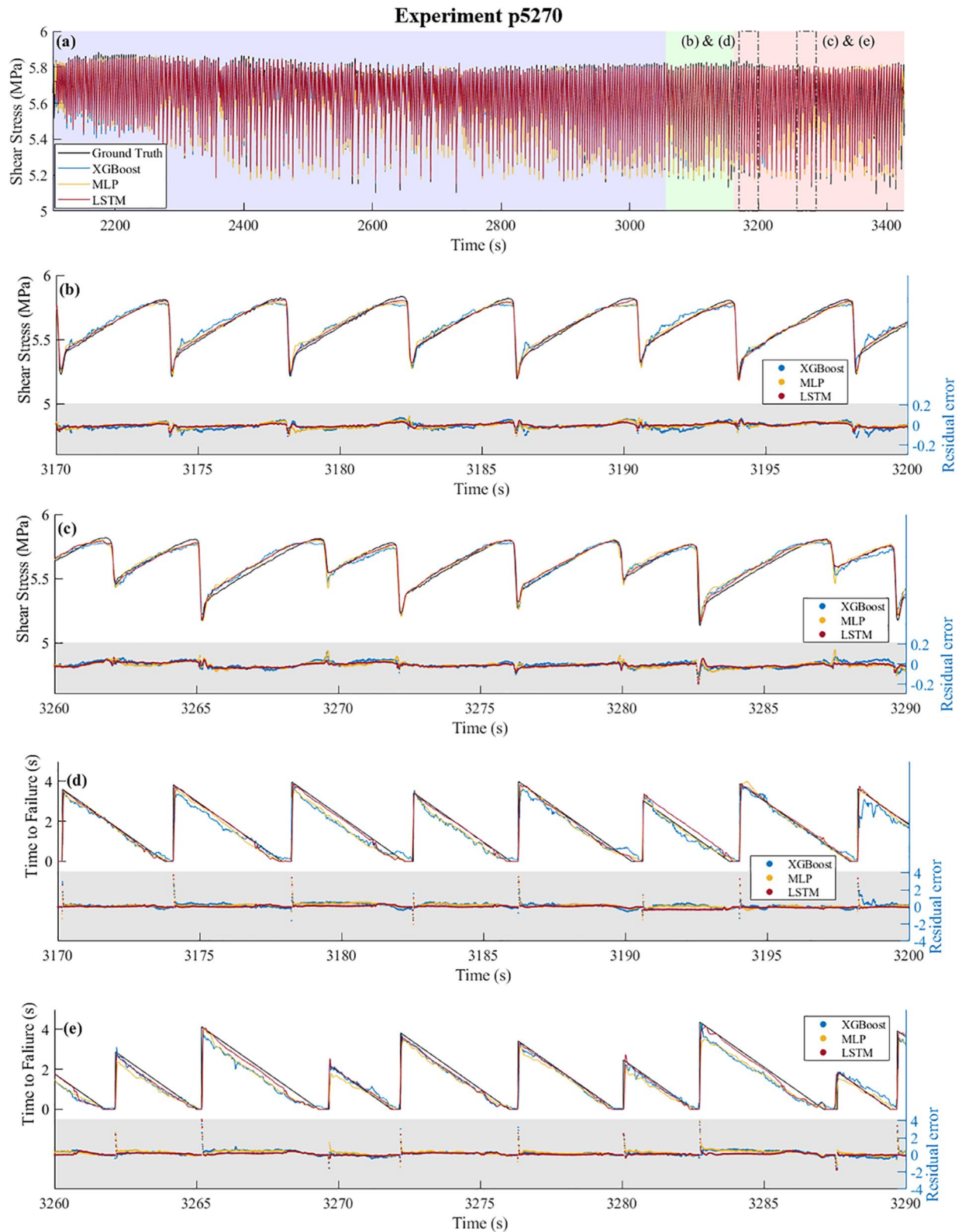
Figure 3. Performance of single output models. (a) Shear stress prediction using the XGBoost, MLP and LSTM models. The blue region corresponds to the training set (72%), the green region corresponds to the validation set (8%) and the pink region corresponds to the test set (20%). (b) Detailed view for regular cycles of shear stress with the corresponding residual error. (c) Detailed view for irregular cycles of shear stress with the corresponding residual error. (d) Detailed view for regular cycles of time to failure with the corresponding residual error. (e) Detailed view for irregular cycles of time to failure with the corresponding residual error. LSTM, long short-term memory; MLP, multilayer perceptron.

underpredicts the shear stress toward the end of seismic cycles, shortly before failure. When predicting time to the next failure event, the LSTM model with a test R^2 score of 0.90601 again outperforms both MLP and XGBoost models that show test R^2 scores of 0.87780 and 0.83184, respectively. The model performances are compared in Table S1.

LSTM models outperforming the MLP and XGBoost models is not surprising considering the time sequence nature of the features and prediction tasks. In other words, due to their design, LSTM models are expected to better capture the precursory information in the systematic time-evolution of the wave velocity, amplitude and frequency features. The XGBoost models appear most susceptible to overfitting the training data—showing good performance in training and validation phases, but poor performance on testing data. In contrast, MLP and LSTM perform very well without overfitting. We note that the training and prediction time required for these models vary widely. The XGBoost model can be trained in ∼20 min and can predict the test shear stress data in 0.4 s. The MLP model takes ∼2.5 min to train and 1 s to predict the test data. The LSTM model requires 10 min for training and 8 s for prediction. It is evident that including the data history significantly improves the predictions at the cost of extra training/prediction time. All the models reported here are run using *Google Colab* which provides GPU acceleration with 2.20 GHz Intel(R) Xeon(R) processor and 16 GB memory.

Comparing the performance of the LSTM models described above with models that use only amplitude or velocity features (Figures S2 and S3), we find that exploiting all the extracted ultrasonic features provides the most accurate predictions for shear stress (test R^2 score of 0.94825). LSTM models that only rely on velocity features provide better predictions (test R^2 score of 0.93878) than those that only use amplitude features (test R^2 score of 0.92491). We argue that this is due to the higher sensitivity of wave velocity to shear stress than wave amplitude. Shreedharan et al. (2021) recently showed that during the seismic cycle, the changes in transmitted wave amplitude are primarily correlated with slip rate while velocity changes are dictated by both slip rate and stress changes. Figure S4 compares the output of models that simultaneously predict both shear stress and time to failure, which we refer to as multi-output models. The multi-output models perform as well as the single-output models. Similar to single-output models, LSTM models show the most accurate predictions while MLP predictions outperform XGBoost models. Again, XGBoost model is slowest to train and test followed by LSTM and MLP models. The observed similarity between the shear stress and time to failure prediction models is expected given that time to failure is deduced from shear stress history. However from practical standpoint, it is useful to have models that predict both timing and size of laboratory quakes (equivalent to the shear stress drops) simultaneously. Finally, we note that although including normalized features (Equation 5) slightly improves the model performances, it is not necessary to obtain models with good performance (Figure S5).

Next, we explore the transportability of the models to a distinct data set. To that end, we use the XGBoost and LSTM model trained for the original data set (experiment p5270) to make predictions for a second data set pertaining to another experiment (p5271) with a different loading stiffness. This process is known as transfer learning, where a pre-trained model is used for a different prediction task having similar features and target variable(s). For neural network models, this is most easily accomplished using a practice called *fine-tuning*, where the parameters of the pre-trained model are updated by further training on the new experiment using a small number of iterations. This is not possible for XGBoost as its primary goal in training is to add new trees that perform an additive correction to the predictions of the previous trees (rather than modifying the structure and parameters of the previous trees). When an XGBoost model is further trained on new data, additional trees are added that essentially model the difference between the old and new prediction problems.

Experiment p5271 is different from experiment p5270 in that a different machine loading stiffness is applied. This is realized by using acrylic springs of different cross-sectional areas in series with the shear loading piston. Transfer learning is achieved by initializing the model using the parameters associated with the model trained on p5270 and then further training the model on p5271. Generally, when we train a deep learning model with random initialization, it takes a considerable amount of input data and training time until the model converges. However, if we "reuse" a pre-trained model on another data set with similar features (experiment p5271), the initial weights associated from the models trained on a source data set (experiment p5270) serve as good starting points for further training on the new experiment. Hence,
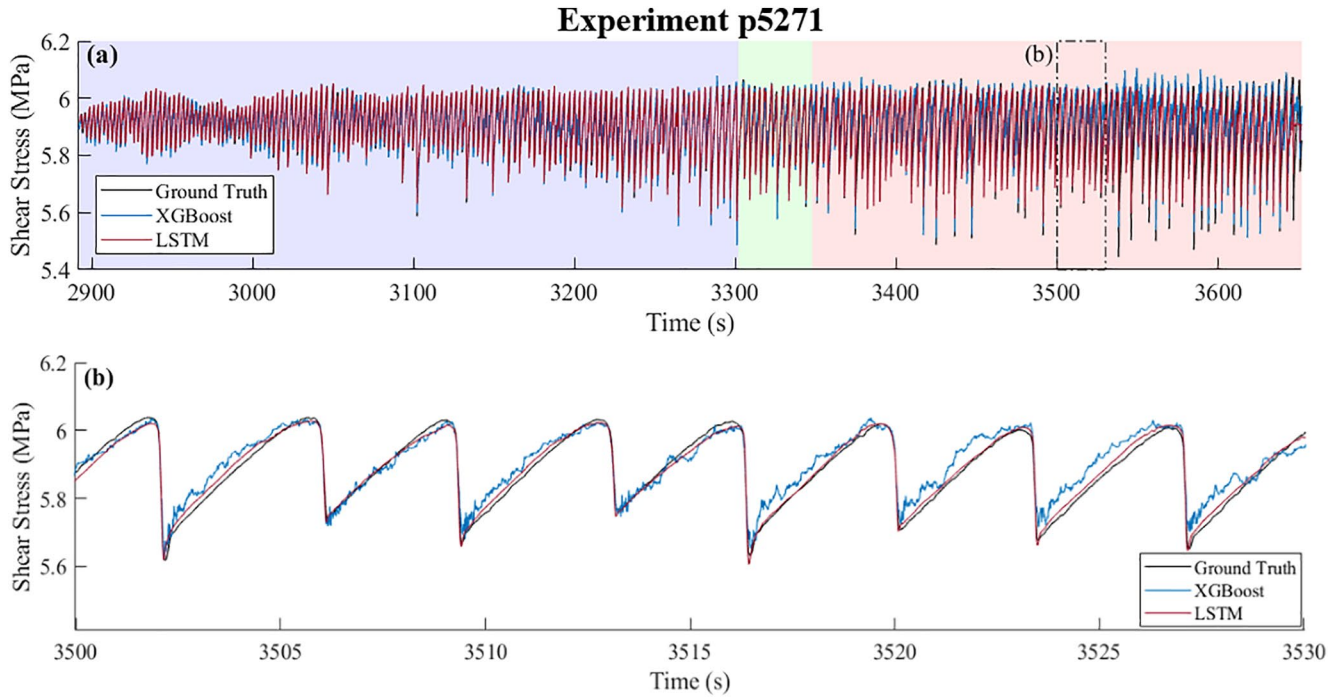
Figure 4. Transfer Learning on Experiment p5271 for shear stress prediction. (a) The predictions of XGBoost and LSTM models, which are pre-trained on the 80% (train and validation sets) of the p5270 experiment data set, are reused as a starting point for predicting shear stress in the p5271 experiment. Even with 54% data set reserved for training (blue region) and 6% for validation (green region), we can acquire an R^2 score of 0.92832 for the test set (pink region). The XGBoost model gives R^2 score of 0.86723 on the test data. (b) Detailed view of the predictions. LSTM, long short-term memory.

even with less training data and fewer number of epochs i.e., with fewer iterations, we can acquire good predictions faster. Here we only use the validation set to determine when to stop training. That is, we use validation error after every epoch to decide when to stop training the deep learning models and similarly we use validation error to decide how many additional trees to use with XGBoost. We compare the transportability of the deep learning models with the transportability of XGBoost. For XGBoost, the re-training process for p5271 data set takes approximately the same amount of time as that for the training on p5270 data. As shown in Figure 4, the use of transfer learning via fine-tuning speeds up the LSTM training process and even with 54% (down from 72%) of the new data reserved for training, a good prediction model with an R^2 score of 0.92832 is attained. In comparison, the XGBoost model does not transfer as well giving a lower R^2 score of 0.86723. LSTM model's superior transferability to a different data set again reveals the power of deep learning in capturing the essential features of the data.

Our findings suggest that the features from active source seismic data can accurately predict both the timing and size of shear failures. The good performance of the deep learning models is not surprising considering the friction evolution laws and the conceptual connection between the state variable and the seismic features (Marone, 1998). The rate and state friction law is written in the form:

$$\mu = \mu_0 + a\, ln(V / V_0) + b\, ln(V_0 \theta / D_c) \tag{6}$$

where μ_0 is a constant representing the stick-slip at velocity V_0, V is the frictional slip rate, θ is a state variable and a and b are empirical constants. The state variable θ is commonly interpreted as a characteristic contact time, which can be indirectly probed using seismic/ultrasonic data. That is, a longer contact time would increase the interfacial stiffness k_I, which in turn would influence both wave velocity v and amplitude A through the following relationships (Tattersall, 1973):

$$|T| = \frac{A}{A_0} = \frac{2k_I}{\sqrt{((\omega \rho v)^2 + 4k_I^2)}} \qquad TOF = tan^{-1}\frac{(\omega \rho v\, /\, 2k_I)}{(\omega)} \tag{7}$$

where T is the transmission coefficient, A_0 is a reference amplitude, ρ is the density of the material surrounding the fault and $\omega = 2\pi f$ is the angular frequency. Additionally, to describe the variations in frictional strength, the constitutive law in Equation 6 must be coupled with another relation to include elastic interactions between the frictional surfaces of the fault and their surroundings:

$$d\mu/dt = k(V_l - V) \qquad (8)$$

where V_l is the velocity of the shear load point and k is the stiffness of the system. The system stiffness k is a summation of the loading apparatus stiffness K and the interfacial stiffness k_l, which is related to both wave amplitude and velocity following Equation 7. This relation suggests an additional physical underpinning for our data-driven models.

In sum, our results demonstrate that deep learning on time-lapse active ultrasonic monitoring data can accurately predict the timing and size of laboratory fault failure. This suggests that active seismic monitoring of tectonic faults could in theory yield similar predictions in nature. However, it must be acknowledged that such an approach is not readily applicable in nature. Compared to laboratory experiments, instrumentation in the field is typically far from the targeted fault. Further, field surveys often probe multiple fault strands and fracture zones with complex stress patterns, at far lower frequencies (orders of magnitude, 1–1,000 Hz) than the lab-based techniques, all of which could affect the models. Additionally, while multiple seismic cycles can be used in the lab for training, it is generally not the case in nature, with one notable exception being the small earthquake repeaters that have been detected in great numbers over the last two decades. Beyond earthquake repeaters, one possible pathway would be to create ML models based on field data from multiple tectonic sites worldwide (Pritchard et al., 2020), as well as laboratory data. Our transfer learning approach is a step in this direction.

Finally, the excellent performance of the presented models indicate a strong physical link between the extracted seismic features and friction constitutive laws. The developed data-driven models may help better understand the relation between frictional state variable and seismic data.

4. Conclusions

We demonstrate that machine learning models can predict the timing and/or size of laboratory quakes from continuously recorded active source ultrasonic data where deep learning models yield accurate predictions despite irregular seismic cycles. These models rely on ultrasonic wave velocity, amplitude and frequency features. Prediction models that use all the features produce more accurate results than those that use velocity or amplitude features alone. The LSTM model, which takes into account the data history and captures the time-evolution of the features gives the most accurate predictions. Our transfer learning study shows that unlike the XGBoost model, the LSTM model trained on one data set can easily transfer to a distinct data set, requiring only a small amount of fine-tuning that significantly reduces the training data size and time (compared to training from scratch) while still providing accurate predictions. We conclude that deep learning in conjunction with time-lapse active source seismic monitoring data can accurately predict the time of occurrence and size of laboratory quakes. This finding could have important implications for active seismic monitoring of carbon storage sites as well as geothermal and unconventional reservoirs. Additionally, these models can provide insight into the connection between seismic data and friction constitutive laws.

Data Availability Statement

References

Ampuero, J.-P., & Rubin, A. M. (2008). Earthquake nucleation on rate and state faults—Aging and slip laws. *Journal of Geophysical Research*, *113*(B1), B01302. https://doi.org/10.1029/2007JB005082

Bolton, D. C., Shokouhi, P., Rouet-Leduc, B., Hulbert, C., Rivière, J., Marone, C., & Johnson, P. A. (2019). Characterizing acoustic signals and searching for precursors during the laboratory seismic cycle using unsupervised machine learning. *Seismological Research Letters*, *90*(3), 1088–1098.

Bolton, D. C., Shreedharan, S., Rivière, J., & Marone, C. (2020). Acoustic energy release during the laboratory seismic cycle: Insights on laboratory earthquake precursors and prediction. *Journal of Geophysical Research: Solid Earth*, *125*(8), e2019JB018975. https://doi.org/10.1029/2019JB018975

Brantut, N. (2018). Time-resolved tomography using acoustic emissions in the laboratory, and application to sandstone compaction. *Geophysical Journal International*, *213*(3), 2177–2192. https://doi.org/10.1093/gji/ggy068

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*.

Dieterich, J. H., & Kilgore, B. D. (1994). Direct observation of frictional contacts: New insights for state-dependent properties. *Pure and Applied Geophysics*, *143*(1), 283–302.

Dresen, G., Kwiatek, G., Goebel, T., & Ben-Zion, Y. (2020). Seismic and aseismic preparatory processes before large stick-slip failure. *Pure and Applied Geophysics*, *177*(12), 5741–5760.

Goebel, T. H. W., Kwiatek, G., Becker, T. W., Brodsky, E. E., & Dresen, G. (2017). What allows seismic events to grow big? Insights from b-value and fault roughness analysis in laboratory stick-slip experiments. *Geology*, *45*(9), 815–818. https://doi.org/10.1130/G39147.1

Hedayat, A., Haeri, H., Hinton, J., Masoumi, H., & Spagnoli, G. (2018). Geophysical signatures of shear-induced damage and frictional processes on rock joints. *Journal of Geophysical Research: Solid Earth*, *123*(2), 1143–1160. https://doi.org/10.1002/2017JB014773

Hedayat, A., Pyrak-Nolte, L. J., & Bobet, A. (2014a). Multi-modal monitoring of slip along frictional discontinuities. *Rock Mechanics and Rock Engineering*, *47*(5), 1575–1587. https://doi.org/10.1007/s00603-014-0588-7

Hedayat, A., Pyrak-Nolte, L. J., & Bobet, A. (2014b). Precursors to the shear failure of rock discontinuities: Precursors to Rock Frictional Sliding. *Geophysical Research Letters*, *41*(15), 5467–5475. https://doi.org/10.1002/2014GL060848

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.

Hulbert, C., Rouet-Leduc, B., Johnson, P. A., Ren, C. X., Rivière, J., Bolton, D. C., & Marone, C. (2019). Similarity of fast and slow earthquakes illuminated by machine learning. *Nature Geoscience*, *12*(1), 69. https://doi.org/10.1038/s41561-018-0272-8

Hulbert, C., Rouet-Leduc, B., Jolivet, R., & Johnson, P. A. (2020). An exponential build-up in seismic energy suggests a months-long nucleation of slow slip in Cascadia. *Nature Communications*, *11*(1), 4139. https://doi.org/10.1038/s41467-020-17754-9

Kaproth, B. M., & Marone, C. (2013). Slow earthquakes, preseismic velocity changes, and the origin of slow frictional stick-slip. *Science*, *341*(6151), 1229–1232. https://doi.org/10.1126/science.1239577

Kilgore, B., Lozos, J., Beeler, N., & Oglesby, D. (2012). Laboratory observations of fault strength in response to changes in normal stress. *Journal of Applied Mechanics*, *79*(3). https://doi.org/10.1115/1.4005883

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: Moment tensor and b-value in stick-slips. *Geophysical Research Letters*, *41*(16), 5838–5846. https://doi.org/10.1002/2014GL060159

Leeman, J., Saffer, D. M., Scuderi, M. M., & Marone, C. (2016). Laboratory observations of slow earthquakes and the spectrum of tectonic fault slip modes. *Nature Communications*, *7*, 11104. https://doi.org/10.1038/ncomms11104

Lockner, D. A., Byerlee, J. D., Kuksenko, V., Ponomarev, A., & Sidorin, A. (1991). Quasi-static fault growth and shear fracture energy in granite. *Nature*, *350*(6313), 39.

Lubbers, N., Bolton, D. C., Mohd-Yusof, J., Marone, C., Barros, K., & 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*(24), 13269–13276. https://doi.org/10.1029/2018GL079712

Marone, C. (1998). Laboratory-derived friction laws and their application to seismic faulting. *Annual Review of Earth and Planetary Sciences*, *26*(1), 643–696.

Marty, S., Passelègue, F. X., Aubry, J., Bhat, H. S., Schubnel, A., & Madariaga, R. (2019). Origin of high-frequency radiation during laboratory earthquakes. *Geophysical Research Letters*, *46*(7), 3755–3763. https://doi.org/10.1029/2018GL080519

McLaskey, G. C., Kilgore, B. D., Lockner, D. A., & Beeler, N. M. (2014). Laboratory Generated M -6 Earthquakes. *Pure and Applied Geophysics*, *171*(10), 2601–2615. https://doi.org/10.1007/s00024-013-0772-9

McLaskey, G. C., & Lockner, D. A. (2016). Calibrated acoustic emission system records m 3.5 to m 8 events generated on a saw-cut granite sample. *Rock Mechanics and Rock Engineering*, *49*(11), 4527–4536. https://doi.org/10.1007/s00603-016-1082-1

Nagata, K., Kilgore, B., Beeler, N., & Nakatani, M. (2014). High-frequency imaging of elastic contrast and contact area with implications for naturally observed changes in fault properties: Fault elastic contrast and contact area. *Journal of Geophysical Research: Solid Earth*, *119*(7), 5855–5875. https://doi.org/10.1002/2014JB011014

Nagata, K., Nakatani, M., & Yoshida, S. (2008). Monitoring frictional strength with acoustic wave transmission. *Geophysical Research Letters*, *35*(6). https://doi.org/10.1029/2007GL033146

Niu, F., Silver, P. G., Daley, T. M., Cheng, X., & Majer, E. L. (2008). Preseismic velocity changes observed from active source monitoring at the Parkfield SAFOD drill site. *Nature*, *454*(7201), 204.

Passelegue, F. X., Schubnel, A., Nielsen, S., Bhat, H. S., & Madariaga, R. (2013). From sub-Rayleigh to supershear ruptures during stick-slip experiments on crustal rocks. *Science*, *340*(6137), 1208–1211. https://doi.org/10.1126/science.1235637

Pritchard, M. E., Allen, R. M., Becker, T. W., Behn, M. D., Brodsky, E. E., Bürgmann, R., et al. (2020). New opportunities to study earthquake precursors. *Seismological Research Letters*, *91*(5), 2444–2447. https://doi.org/10.1785/0220200089

Rivière, J., Lv, Z., Johnson, P. A., & 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., Riviere, J., Marone, C., et al. (2018). Estimating fault friction from seismic signals in the laboratory. *Geophysical Research Letters*, *45*(3), 1321–1329. https://doi.org/10.1002/2017GL076708

Rouet-Leduc, B., Hulbert, C., & Johnson, P. A. (2019). Continuous chatter of the Cascadia subduction zone revealed by machine learning. *Nature Geoscience*, *12*(1), 75.

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*(18), 9276–9282. https://doi.org/10.1002/2017GL074677

Rouet-Leduc, B., Hulbert, C., McBrearty, I. W., & Johnson, P. A. (2020). Probing slow earthquakes with deep learning. *Geophysical Research Letters*, *47*(4), e2019GL085870. https://doi.org/10.1029/2019GL085870

Rubin, A. M., & Ampuero, J.-P. (2005). Earthquake nucleation on (aging) rate and state faults. *Journal of Geophysical Research*, *110*(B11), B11312. https://doi.org/10.1029/2005JB003686

Ruina, A. (1983). Slip instability and state variable friction laws. *Journal of Geophysical Research*, *88*(B12), 10359–10370. https://doi.org/10.1029/JB088iB12p10359

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), 399–415.

Scuderi, M. M., Marone, C., Tinti, E., Di Stefano, G., & Collettini, C. (2016). Precursory changes in seismic velocity for the spectrum of earthquake failure modes. *Nature Geoscience*, *9*(9), 695–700.

Shreedharan, S., Bolton, D. C., Rivière, J., & Marone, C. (2020). Preseismic fault creep and elastic wave amplitude precursors scale with lab earthquake magnitude for the continuum of tectonic failure modes. *Geophysical Research Letters*, *47*(8), e2020GL086986. https://doi.org/10.1029/2020GL086986

Shreedharan, S., Bolton, D. C., Rivière, J., & Marone, C. (2021). Competition between preslip and deviatoric stress modulates precursors for laboratory earthquakes. *Earth and Planetary Science Letters*, *553*, 116623.

Shreedharan, S., Rivière, J., Bhattacharya, P., & Marone, C. (2019). Frictional state evolution during normal stress perturbations probed with ultrasonic waves. *Journal of Geophysical Research: Solid Earth*, *124*(6), 5469–5491. https://doi.org/10.1029/2018JB016885

Tattersall, H. (1973). The ultrasonic pulse-echo technique as applied to adhesion testing. *Journal of Physics D: Applied Physics*, *6*, 819.

Trugman, D. T., McBrearty, I. W., Bolton, D. C., Guyer, R. A., Marone, C., & Johnson, P. A. (2020). The spatiotemporal evolution of granular microslip precursors to laboratory earthquakes. *Geophysical Research Letters*, *47*(16), e2020GL088404. https://doi.org/10.1029/2020GL088404