LABORATORY EARTHQUAKE ANALYSIS

Olha Tanyuk[1], Daniel Davieau[1], Dr. Michael L. Blanpied[2], Dr. Charles South[1] and Dr. Daniel W. Engels[1]

[1] Southern Methodist University, Dallas, TX 75205, USA
[2] U.S. Geological Survey Headquarters, Reston, VA 20192, USA

Abstract. We apply machine learning to data set, that comes from a classic laboratory earthquake experiment, that has been studied in depth as a tabletop analog of seismogenic faults for decades. Our goal is to find the pattern of seismic signals that precede earthquakes. Here we show that by listening to the acoustic signal emitted by a laboratory fault, machine learning can predict the time remaining before it fails with a good accuracy. These predictions are based solely on the instantaneous physical characteristics of the acoustical signal and do not make use of its history. Los Alamos' initial work [1] showed that the prediction of laboratory earthquakes from continuous seismic data is possible in the case of quasi-periodic laboratory seismic cycles. In this work we use a much more challenging dataset with considerably more aperiodic earthquake failures with more realistic behavior.

1 INTRODUCTION

A traditional way to determine that an earthquake may take place is based on the recurrence interval for characteristic earthquakes, that repeat periodically. But the earthquake recurrence is not constant for a given fault, event occurrence can only be inferred within large error bounds. Over the last 15 years, there has been renewed hope that progress can be made regarding forecasting owing to tremendous advances in instrumentation quality and density. These advances have led to exciting discoveries of previously unidentified slip processes that include slow slip, low frequency earthquakes and Earth tremor, that occur deep in faults. These discoveries form a new understanding of fault slip and may lead to advances in predicting.[1]

In August 2017 Los Alamos National Laboratory (LANL) conducted an experiment [1], that illuminate the mechanics of slow-slip phenomena. They predicted the remaining time until laboratory earthquakes occur with 89% accuracy. In this paper we use acoustic data, that was provided by LANL in January 2019, as part of Kaggle project, which also represent laboratory slow-slips. Data from this experiment exhibits a very aperiodic and more realistic behavior compared to the data they studied earlier, with earthquakes occurring very irregularly.[2] The results of this experiment are potentially applicable to the field of real world earthquakes. Other potential applications include avalanche prediction or failure of machine parts. "If this challenge is solved and the physics are ultimately shown to scale from the laboratory to the field, researchers will have the potential to improve earthquake hazard assessments that could save lives and billions of dollars in infrastructure.[2]" Given seismic signal data with considerably more a-periodic laboratory earthquake failures and modern computing hardware; we find the pattern of acoustic signals to determine when laboratory earthquakes will occur.

2 TUTORIAL MATERIAL

We hear about earthquakes mostly via news media when there is a large seismic event which is noticeable, causes death and destruction. These are stick–slip events that radiate seismic energy along the seams (fault lines) between tectonic plates. In this study we refer to these as Regular Earthquakes. Regular earthquakes are caused by a sudden slip on a fault. Tectonic plates are always slowly moving, but they get stuck at their edges due to friction. When the stress on the edge overcomes the friction, there is an earthquake that releases energy in waves that travel through the earth's crust and cause the shaking that we feel.[5]

Another type of earthquake we refer to in this study is a Slow Slip Earthquake (SSE). SSE's are fault behaviors that occur slowly enough to make them undetectable without instrumentation. They do not shake the ground and cause widespread destruction like regular earthquakes do. They occur near the boundaries of large earthquake rupture zones[3]. There is evidence to suggest that there is a relationship between slow slip earthquakes and more noticeable regular earthquakes[4].

LANL researchers discovered a way to successfully predict SSE in a laboratory experiment that simulates natural conditions. In 2017, this team discovered a way to train a computer to pinpoint and analyze seismic and acoustic signals emitted during the movements along the fault to predict an earthquake. They processed massive amounts of data and identified a particular sound pattern previously thought to be noise that precedes an earthquake. The team was able to characterize the time remaining before a laboratory earthquake at all times.[6]

In the lab, the team imitated a real earthquake using steel blocks interacting with rocky material (fault gouge) to induce slipping that emitted seismic sounds. An accelerometer recorded the acoustic emission emanating from the sheared layers.[6] For the first time, researchers discovered a pattern that accurately predicted when a quake would occur. The team acknowledges that the physical traits of the lab experiment (such as shear stresses and thermal properties) differ from the real world but the application of the analysis to real earthquakes to validate their results is ongoing. This method can also be applied outside of seismology to support materials' failure research in many fields such as aerospace and energy.[6] The team's lab results reveal that the fault does not fail randomly but in a highly predictable manner. The observations also demonstrate that the fault's critical stress state, which indicates when it might slip, can be determined using exclusively an equation of state.[6] So far seismologists and Earth scientists have relied exclusively on catalogues of historical data to try to characterize the state of faults. These catalogues contain a minute fraction of seismic data, and remaining seismic data is discarded during analysis as useless noise. The authors discovered that hidden in this noiselike data there are signals emitted by the fault that inform them of the state of the fault much more precisely than catalogues.[6] "Our work shows that machine learning can be used to extract new meaningful physics from a very well studied system," said Bertrand Rouet-Leduc, Los Alamos Earth and Environmental Sciences Division scientist and the paper's lead author. "It also shows that seismogenic faults are continuously broadcasting a signal that precisely informs us of their physical state and how close they are to rupture, at least in the laboratory."

2.1. Experimental setup [1]

A three-block assembly with two gouge layers is placed in a bi-axial stress configuration. Two 5 mm thick fault gouge layers are placed between the three blocks, which are held in place by a fixed normal load. The gouge material is comprised of Class IV beads with diameter 105-149 µm. The central block is sheared at constant displacement rate. The two data streams recorded for our purposes here are the shear stress and the acoustic signal. At some time while the gouge material is in a critical shear stress regime, the shear stress abruptly drops, indicating gouge failure. These large drops in shear stress are laboratory quakes. As applied load progressively increases, the inter-event time (recurrence) of laboratory earthquakes progressively decreases. At smaller applied loads the slips become aperiodic. In all cases, the rate of impulsive precursors accelerates as failure is approached. The acoustic particle acceleration Ü is measured on the central block and can be readily converted to dynamic strain used in the ML analysis:

$$\epsilon = \frac{\dot{u}}{c}, \text{ where } \dot{u} = \frac{\ddot{u}}{\omega}, \text{ and } \omega = 2\pi f$$

with c ≈ 700m/s the average measured wave speed in the granular material, and f ≈ 40.3kHz. The sampling rate of the acoustic data is 330kHz. We are band-limited by the 20 accelerometer (the frequency response is poor above about 50kHz). Therefore we select one of the system mechanical resonances within this band occurring at 44 kHz. Using this peak we improve our signal to noise ratio. The 'noise' is of very different character when the piston is stopped, and reflects primarily the mechanical resonances of the system. In short, we are certain the signal we analyze is the acoustic signal emanating from the fault, and the electromagnetic and system noise play no role in the predictions. [1]

3 DATA

The data used in this work is a chunk of 157.275 seconds of seismic data (ordered in time), which is recorded at 4MHz, hence 629,145,480 data points, and the output is time remaining until the following lab earthquake, in seconds.
The seismic data is recorded using a piezoceramic sensor, which outputs a voltage upon deformation by incoming seismic waves. The seismic data of the input is this recorded voltage, in integers. (Table 1)

index	acoustic_data	time_to_failure
0	12	1.469099998474121
1	6	1.469099998474121
2	8	1.469099998474121

3	5	1.469099998474121
4	8	1.469099998474121

The voltage rate of acoustic precursors accelerates as failure approaches, suggesting that upcoming laboratory earthquake timing could be predicted (Fig.1).

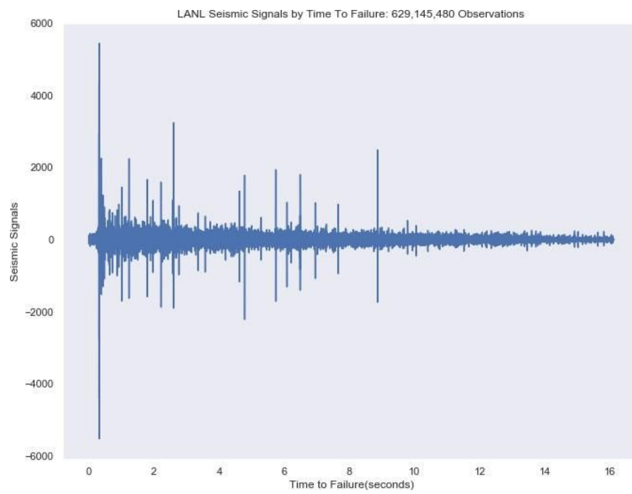


Fig. 1.

The data has 16 earthquakes. The shortest time to failure is 1.5 seconds for the first earthquake and 7 seconds for the 7th, while the longest is around 16 seconds. We can see that usually acoustic data shows huge fluctuations just before the failure and the nature of data is cyclical. Another important point: visually failures can be predicted as cases when huge fluctuations in signal are followed by small signal values (Fig.2)

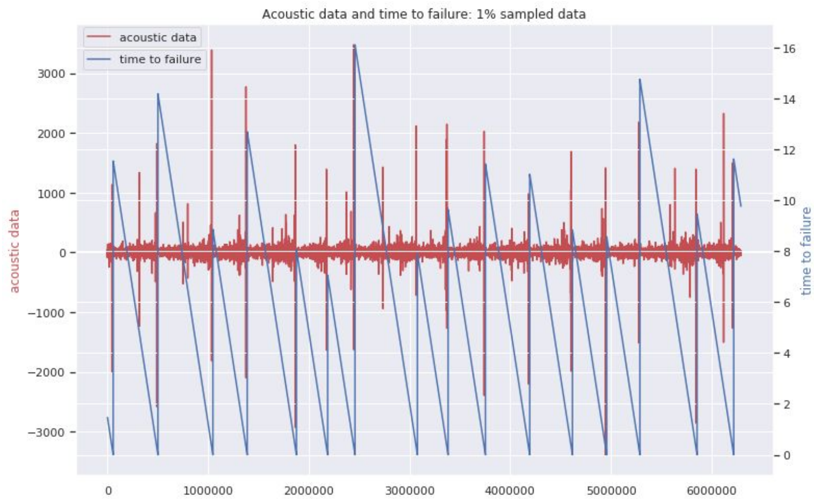


Fig. 2

On zoomed-in-time plot (first 1% of the data) we can see that the large oscillation before the failure is not in the last moment. There are trains of intense oscillations preceding the large one and also some oscillations with smaller peaks after it. (Fig. 3)

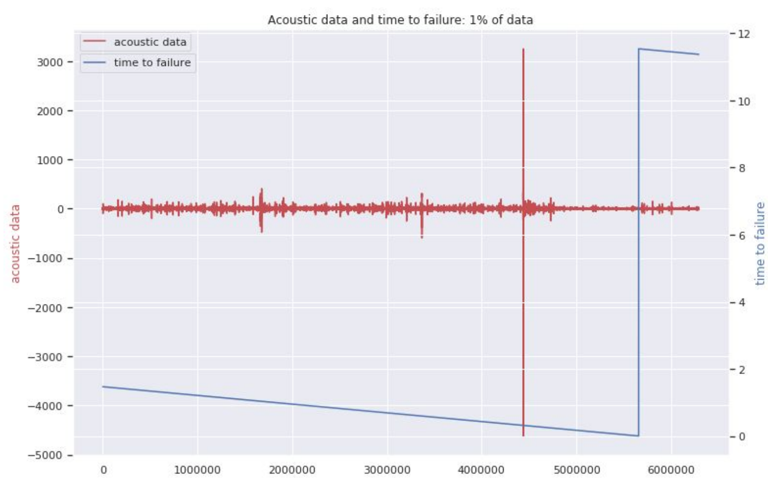


Fig. 3

Acoustic data distribution has a very high peak and we see outliers in both directions (Fig. 4).

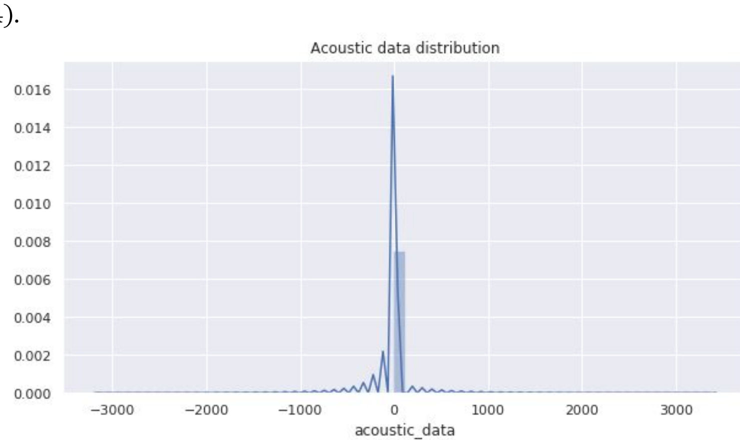


Fig. 4. Distribution of Seismic Signal Measurements by LANL

Distribution of time to failure seems to be right skewed (Fig.5). We decided to use square root of time feature for our analysis in order to normalize it and improve prediction models. It is still not ideally normal, but looks much more better after square root transformation. (Fig 6)

Fig. 5

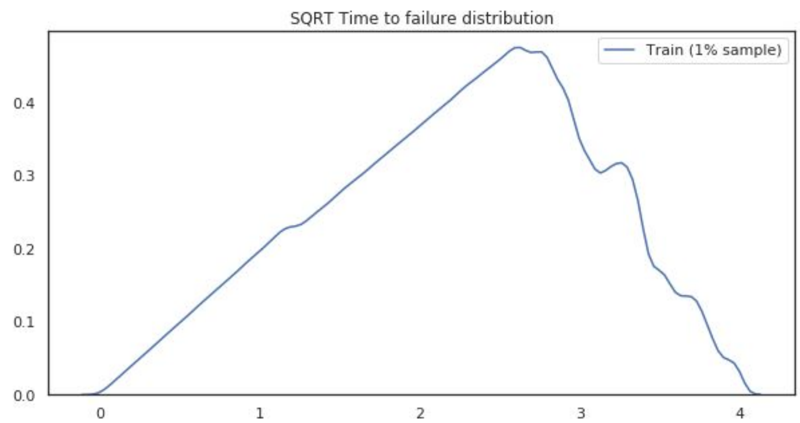


Fig. 6

4 METHODS AND EXPERIMENTS

Our goal is to predict the time remaining before the next failure using only moving time windows of the acoustic data. We divided our data by windows that contains 150,000 observations each (0.0375 seconds of seismic data), so our new data set is 4194 windows. From each time window, we compute a set of 98 potentially relevant statistical features (e.g., mean, variance, kurtosis).
We apply a machine learning techniques such as the Random Forest Regressor, XGB Regressor, Decision Tree Regressor, LGBM Regressor, Extra Trees Regressor to the new continuous values that we got, analysing acoustic time series data.
New features can be separated into two main classes:

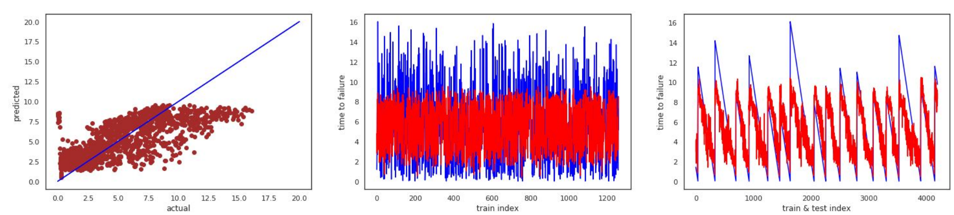
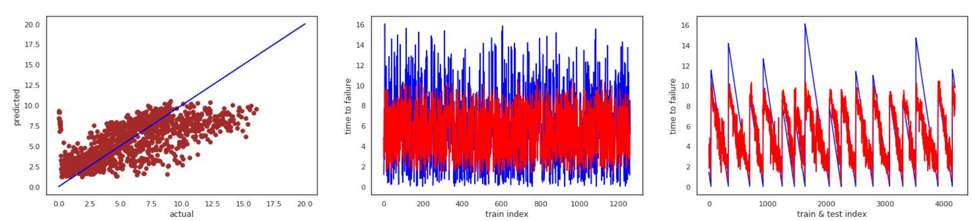
- Distribution of signal's energy: we use couple of higher order moments of the acoustic data to capture the evolution of the signal's energy. Within each time window we compute the signal's normalized mean, minimum, maximum and higher moments (variance, skewness, kurtosis).
- Precursors: the system enters a critical state when close to failure. We rely on different percentiles and thresholds to monitor this precursory activity. We use the 1st - 9th and 91th - 99th percentiles. Our thresholds measure the count of observations that the acoustic signal spends over a threshold value f0 and under a threshold value f1.

In order to avoid correlation between new features we applied principal component analysis. Instead of using 98 features, we created just 10 that

represented 99.9% of the full data variation.
We use a 70/30 random split of the full time series for use as training and testing data sets respectively. We computed regularization hyper-parameters for each machine learning predicting techniques by random grid search based on a 3-fold cross-validation.

5 RESULTS

We run 5 different techniques on a train data set (70% of the full data) before principal component analysis and after. Principal component analysis did not improve our findings significantly, that is why we are not represent those results here. For each model we provide hyper-parameters details for future reproducibility (Table 2).
When making a prediction (red curve), we emphasize that there is no past or future information considered: each prediction uses only the information within one single time window of the acoustic signal. We quantify the accuracy of our model using R2 (the coefficient of determination) and MAE (mean absolute error), applying predicting model on a 30% of the full data (test data).

Method	Results
RandomForestRegressor 'min_samples_split': 12, 'min_samples_leaf': 4, 'max_features': 'sqrt', 'max_depth': 8 n_estimators=1000	mean_absolute_error: 2.0322193397002044 r2_score: 0.4953230388182709 
XGBRegressor 'colsample_bytree': 0.9059046706163597, 'gamma': 4.395361343977617, 'learning_rate': 0.22367318773362826, 'max_depth': 21, 'min_child_weight': 69.30113404571125, 'reg_alpha': 7.681525516492065, 'subsample': 0.8919123509453631 nthreads=-1, n_estimators=1000	mean_absolute_error: 2.0634676560951557 r2_score: 0.48347777795889124 
DecisionTreeRegressor 'max_depth': 4, 'max_features': 'sqrt', 'min_samples_leaf': 12, 'min_samples_split': 70	mean_absolute_error: 2.066672622791363 r2_score: 0.4703327918288054

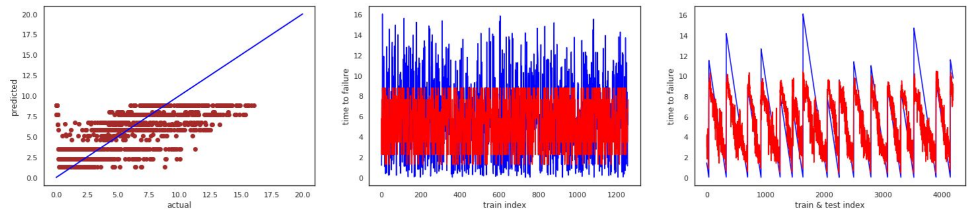
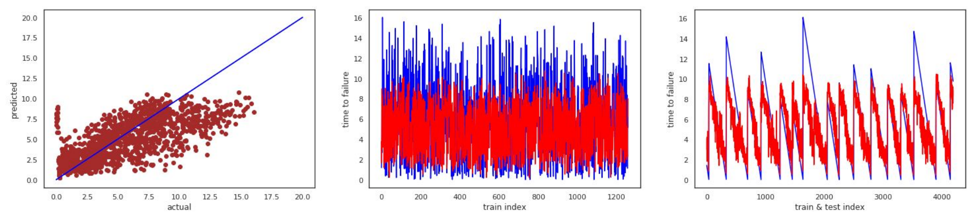
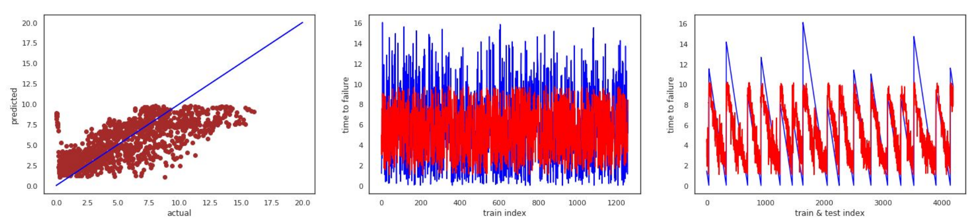
	
lgb.LGBMRegressor 'feature_fraction': 1, 'lambda_l1': 0.6, 'lambda_l2': 0.4, 'learning_rate': 0.05, 'max_depth': 8, 'min_data_in_leaf': 10, 'min_gain_to_split': 0.01, 'num_leaves': 44, 'random_seed': 13, 'subsample': 0.9 n_estimators=10000	mean_absolute_error: 2.089058482904736 r2_score: 0.43881088754719744 
ExtraTreesRegressor 'max_depth': 11, 'max_features': 'auto', 'min_samples_leaf': 40, 'min_samples_split': 200, 'min_weight_fraction_leaf': 0.010872755116371424, 'n_estimators': 200	mean_absolute_error: 2.0338440361453003 r2_score: 0.4950754543622904 

Table 2

6 ANALYSIS

The most accurate results with coefficient of determination 0.5 and mean absolute error 2.03 we got using Random Forest Regressor. The most important features, shown on Fig. 7, suggest to check our data for correlated features, since rolling standard deviations show the same patterns. Plus in order to improve our predicting model, we need to get rid of features that have importance equal to 0 or close to 0, such as hypermoment, minimum, maximum and some of the rolling means. Currently we are working on improving of the model to make prediction for the beginning and the end of the one quake cycle. As you can see on a Fig. 1, seismic waves with small voltage are in our interest. Random forest model accurately predicts failure across load level, but hardly can predict outliers. It means that we still need to think about new features that we missing. Precursors that we choose during feature engineering were not good enough when the system enters a critical state close to failure.

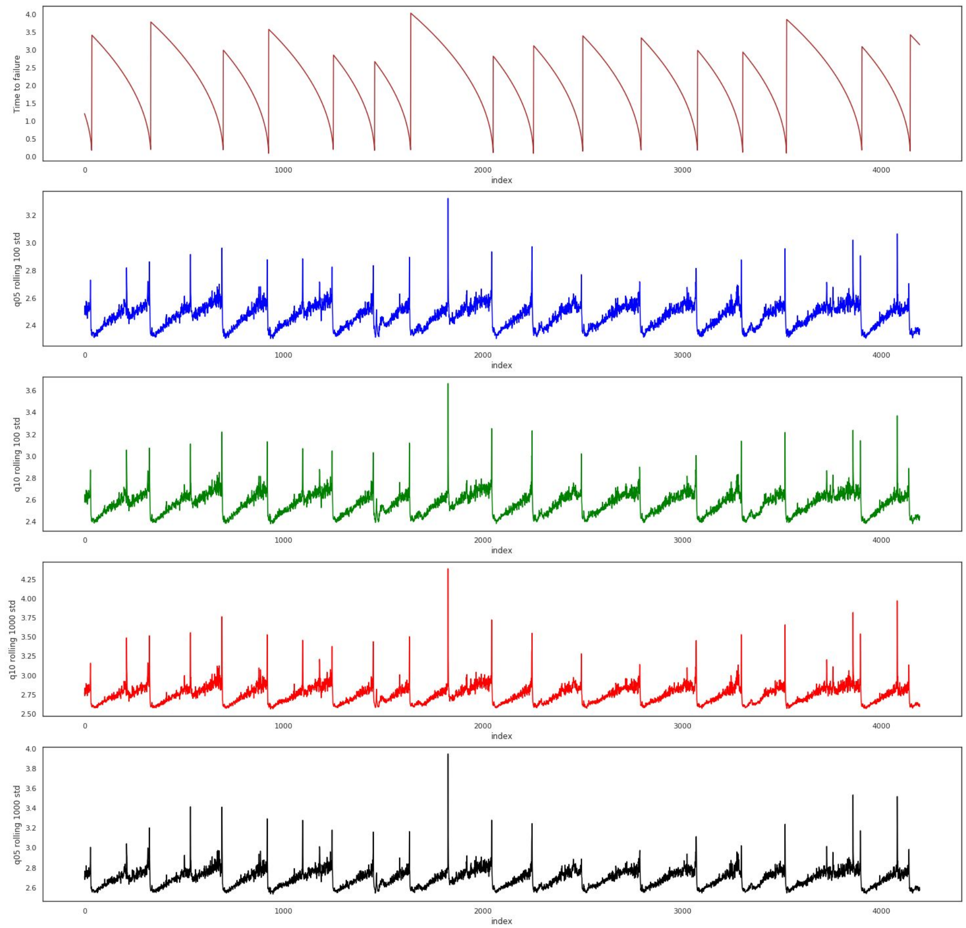


Fig. 7

7 ETHICS

Scientists' responsibility to inform the public about their results may conflict with their responsibility not to cause social disturbance by the communication of these results. A study of the well-known Brady-Spence and Iben Browning earthquake predictions illustrates this conflict in the publication of scientifically unwarranted predictions. Furthermore, a public policy that considers public sensitivity caused by such publications as an opportunity to promote public awareness is ethically problematic from (i) a refined consequentialist point of view that any means cannot be justified by any ends, and (ii) a rights view according to which individuals should never be treated as a mere means to ends. [7]

8 CONCLUSION (I used LANL conclusion and put our numbers. Can you please paraphrase it? I changed it not enough.Original conclusion you can see here https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2017GL074677)

In our work we show that ML applied to this experiment provides failure forecasts with 2.03 sec mean absolute error and 0.5 coefficient of determination based on the instantaneous analysis of the acoustic signal at any time in the slip cycle, except beginning and end of the cycle. These results should suffice to encourage ML analysis of seismic signals in Earth. This is one of the first application of ML to continuous acoustic/seismic data with the goal of inferring failure times. These results suggest that previous analyses based exclusively on earthquake catalogs may be incomplete. In particular, ML-based approaches mitigate human bias by automatically searching for patterns in a large space of potentially relevant variables. Our current approach is to progressively scale from the laboratory to the Earth by applying this approach to Earth problems that most resemble the laboratory system. An interesting analogy to the laboratory may be faults that exhibit small repeating earthquakes. For instance, fault patches located on the San Andreas Fault near Parkfield (Nadeau & McEvilly, 1999; Zechar & Nadeau, 2012) exhibit such behavior. Repeaters at these fault patches may be emitting chattering in analogy to the laboratory. If so, can this signal be recorded by borehole and surface instruments? Whether ML approaches applied to continuous seismic or other geophysical data succeed in providing information on timing of earthquakes (not to mention the challenge of predicting earthquake magnitude), this approach may reveal unidentified signals associated with undiscovered fault physics. Furthermore, this method may be useful for failure prediction in a broad spectrum of industrial and natural materials. Technology is at a confluence of dramatic advances in instrumentation, machine learning, the ability to handle massive data sets and faster computers. Thus, the stage has been set for potentially marked advances in earthquake science.

References

1. Bertrand Rouet-Leduc, Claudia Hulbert, N.L.K.B.C.J.H.P.A.J.: Machine learning predicts laboratory earthquakes
2. Kaggle, R.: Lanl earthquake prediction
3. Ikari Matt J, Marone Chris, S.D.M.K.A.J.: Slip weakening as a mechanism for slow earthquakes
4. Baptiste Rousset, Roland Burgmann, M.C.: Slow slip events in the roots of the san andreas fault
5. USGS: What is an earthquake and what causes them to happen? (March 2019)
6. Rouet-Leduc, B.: Los Alamos machine learning discovers patterns that reveal earthquake fault behavior.
7. Ayhan Sol, Halil Turan: The ethics of earthquake prediction. Volume 10, Issue 4, pp 655–666 (December 2004)