Machine Learning Predicts Aperiodic Laboratory Earthquakes

Olha Tanyuk, Daniel Davieau, Charles South and Daniel W. Engels

Southern Methodist University, Dallas TX 75205, USA otanyuk@mail.smu.edu, danieldavieau@mail.smu.edu, csouth@mail.smu.edu, dwe@lyle.smu.edu

Abstract. Our goal is to find the pattern of aperiodic seismic signals that precede earthquakes at any time in the earthquake's cycle using a small window of time. We use data collected from a laboratory experiment which exhibits similar behavior to natural earthquakes. We apply machine learning to data set, that comes from a classic laboratory experiment involving repeated stick-slip displacement ("earthquake") on a sliding interface, a type of experiment which has been studied in depth as a tabletop analog of seismogenic faults for decades. Here we show that by listening to the acoustic signal emitted by a laboratory fault, an algorithm tuned through machine learning can predict the time remaining before it fails. 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 showed that the prediction of laboratory earthquakes from continuous seismic data is possible in the case of quasi-periodic laboratory seismic cycles.

1 Introduction

Earthquakes cause mass destruction and loss of life. A traditional method to predict earthquakes is to look to past recurrence intervals. Because the recurrences are not constant, predictions can only be made within broad time windows. One such model predicted that a strong earthquake would occur between 1985 and 1993 in the Parkfield California area but no significant event occurred until 2004 [1].

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 such as slow slips. Slow Slip Earthquakes (SSE) 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 [2]. There is evidence to suggest that there is a relationship between slow slip earthquakes and more noticeable regular earthquakes [3].

Researchers imitate natural slow slip earthquakes in the laboratory by placing rocky material between steel blocks and applying shear stress to induce slipping.

Recent improvements in the instruments [5] used to measure signals have enabled the collection of larger volume data from more realistic and unpredictable laboratory earthquakes. However, processing the data and detecting patterns in it has become more difficult to work with. In this paper we demonstrate that machine learning can be used to detect patterns in the more realistic data and predict laboratory earthquakes.

In this paper, given seismic signal data with considerably more aperiodic slow-slip laboratory earthquake failures, we find the pattern of acoustic signals to predict the time at which laboratory earthquakes will occur at any time in the slip cycle.

We use acoustic data, which was provided by LANL in January 2019, as part of Kaggle project [5], which also represent laboratory slow-slip earthquakes, but very aperiodic and it has more natural behavior compared to the data LANL studied earlier, with earthquakes occurring very irregularly in time [5]. The results of this experiment are potentially applicable to the field of real world earthquakes [6]. Other potential applications include avalanche prediction or failure of machine parts [6].

2 Background

2.1 Statistical Values to Evaluate Predictions

R2	Proportion of the variance in the dependent variable that is predictable from the independent variable	$R^2 \equiv 1 - \frac{SS_{\rm res}}{SS_{\rm tot}}$ where SSres is the sum of the squares of the residuals and SStot is the sum of the squared differences of each observation from the overall mean
MAE	It measures the average magnitude of the errors in a set of predictions, without considering their direction.	$\mathrm{MAE} = \frac{1}{n}\sum_{j=1}^{n} \lvert y_j - \hat{y}_j \rvert$ where yj is an actual value, ŷj is a predicted value, n is the total number of observations

Fig. 1. The values used to evaluate predictions in our work are the coefficient of determination r^2 and mean absolute error (MAE).

2.2 Los Alamos National Laboratory's Findings

In 2017 Los Alamos National Laboratory (LANL) researchers discovered a way to successfully predict Slow Slip Earthquakes (SSE) in a laboratory experiment that simulates natural conditions. The team trained a computer to pinpoint and analyze quasi-periodic seismic and acoustic signals emitted during the movements along the fault. 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, using time window of 1.8 sec of the data to make each prediction, with 89% coefficient of determination [4]. This result they achieved using machine learning technique Random Forest Regression. Their results were achieved using quasi-periodic data and can not be generalized over aperiodic data.

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 [4]. 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 [4]. The team's lab results reveal that the fault does not fail randomly but in a 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 [4]. So far seismologists and Earth scientists have mostly relied on catalogues of historical data to try to characterize the state of faults. These catalogs contain a minute fraction of seismic data, and remaining seismic data is discarded during analysis as useless noise. The authors discovered that—in the case of their laboratory faults–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 [4].

2.3 Experimental Setup

The laboratory system is a two-fault configuration that contains fault gouge material submitted to double direct shear [5].

Two fault gouge layers are sheared simultaneously while subjected to a constant normal load and a prescribed shear velocity [5]. The laboratory faults fail in repetitive cycles of stick and slip that is meant to mimic the cycle of loading and failure on tectonic faults [5]. While the experiment is considerably simpler than a fault in Earth, it shares many physical characteristics [5].

The driving piston displaces at a very constant velocity during the inter-event time and accelerates briefly during slip [6]. An accelerometer records the acoustic emission emanating from the shearing layers [6]. The steel blocks are extremely stiff, so the deformation takes place largely in the gouge [6]. Under a broad range of load and shear velocity conditions, the apparatus stick-slips quasi-periodically for hundreds of stress cycles during a single experiment and in general follows predictions from rate and state friction [6]. The rate of im-

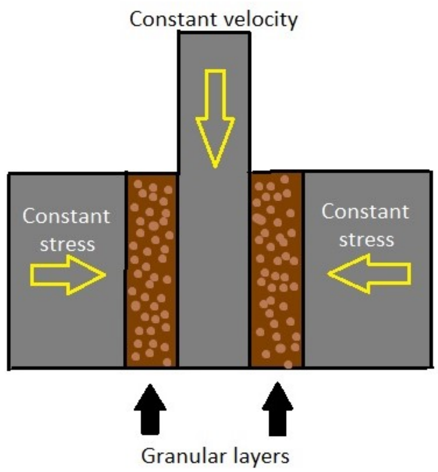

Fig. 2.

pulsive precursors accelerates as failure approaches, suggesting that upcoming laboratory earthquake timing could be predicted [6].

Experimental 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.

2.4 Extra Trees Regressor Overview

The Extra-Trees algorithm builds an ensemble of unpruned regression trees according to the classical top-down procedure. Its two main differences with other treebased ensemble methods are that it splits nodes by choosing cut-points fully at random and that it uses the whole learning sample (rather than a bootstrap replica) to grow the trees [7].

The Extra-Trees splitting procedure uses two parameters: K (the number of attributes randomly selected at each node) and nmin (the minimum sample size for splitting a node). It is used several times with the (full) original learning sample to generate an ensemble model (we denote by M the number of trees of this ensemble) [7].

The predictions of the trees are aggregated to yield the final prediction, by majority vote in classification problems and arithmetic average in regression problems [7]. From the bias-variance point of view, the rationale behind the Extra-Trees method is that the explicit randomization of the cut-point and attribute combined with ensemble averaging should be able to reduce variance more strongly than the weaker randomization schemes used by other methods [7]. The usage of the full original learning sample rather than bootstrap replicas is motivated in order to minimize bias [7]. From a computational point of view, given the simplicity of the node splitting procedure we expect the constant factor to be much smaller than in other ensemble based methods which locally optimize cut-points [7]. The parameters K, nmin and M have different effects: K determines the strength of the attribute selection process, nmin the strength of averaging output noise, and M the strength of the variance reduction of the ensemble model aggregation. These parameters could be adapted to the problem specifics in a manual or an automatic way (e.g. by cross-validation) [7].

The sklearn.ensemble module in Python programing language includes averaging algorithm based on randomized decision trees: Extra-Trees method.

3 Data

The data used in this work is a 157.275 second chunk of seismic data (ordered in time), which is recorded at 4MHz, hence 629,145,480 data points accompanied by the 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 (henceforth we will use the term acoustic signal). The seismic data, which serves as the input to our analysis, is this recorded voltage, in integers.

Acoustic signal is voltage upon deformation by incoming seismic waves.

Time to failure is the remaining time in seconds until an actual stick-slip failure occurred.

Table 1.

Acoustic Signal	Time to Failure
12	1.469099998474121
6	1.469099998474121
8	1.469099998474121
5	1.469099998474121
8	1.469099998474121

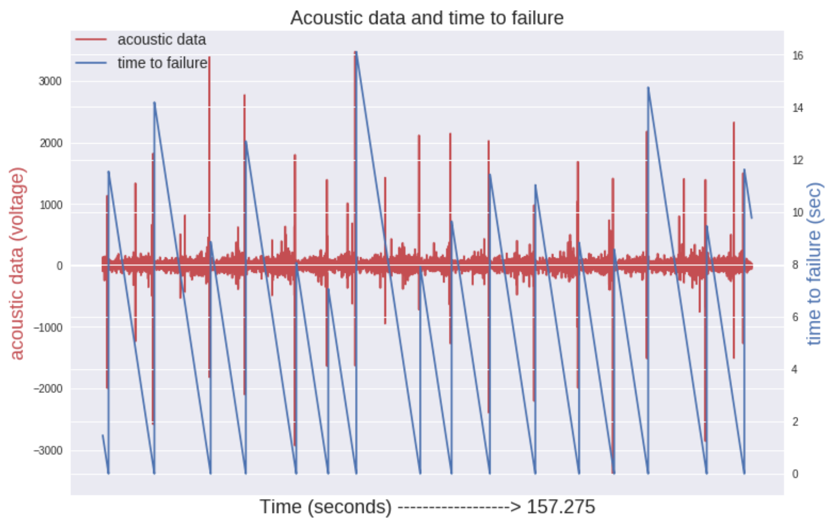

Fig. 3. We can see that usually acoustic data shows huge fluctuations just before the failure. Another important point: visually failures can be predicted as cases when huge fluctuations in signal are followed by small signal values.

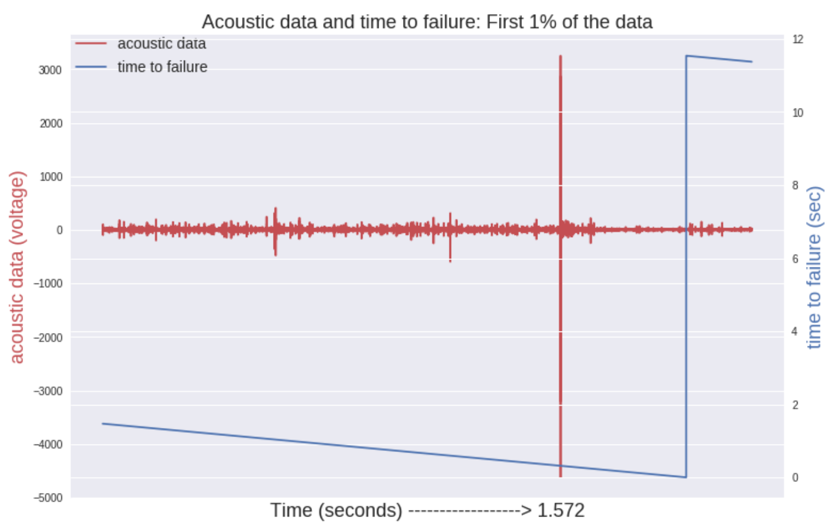

Fig. 4. On zoomed-in-time plot we can see that the large oscillation before the failure is not in the last moment. There are trains of intense oscillations preceding the large one and also some oscillations with smaller peaks after it.

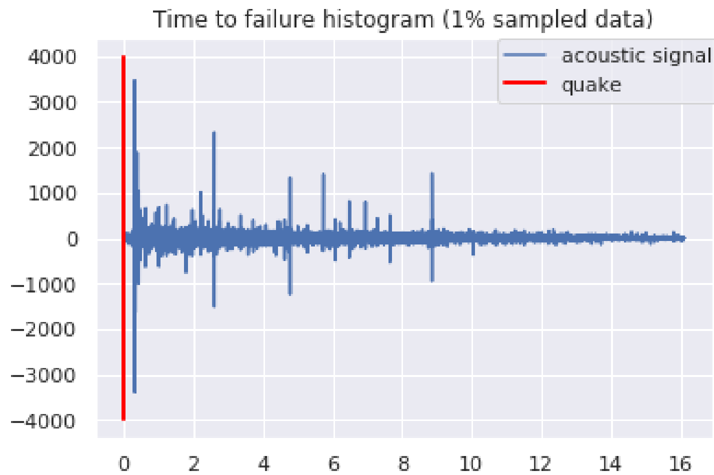

Fig. 5. The voltage amplitude of acoustic precursors accelerates as failure approaches, suggesting that upcoming laboratory earthquake timing could be predicted. We used 1% sample of the data. Red line indicates, that quake occurs, when time to failure approaches to 0. Minimum time remaining until the quake in the data is -5.5150e+03 sec.

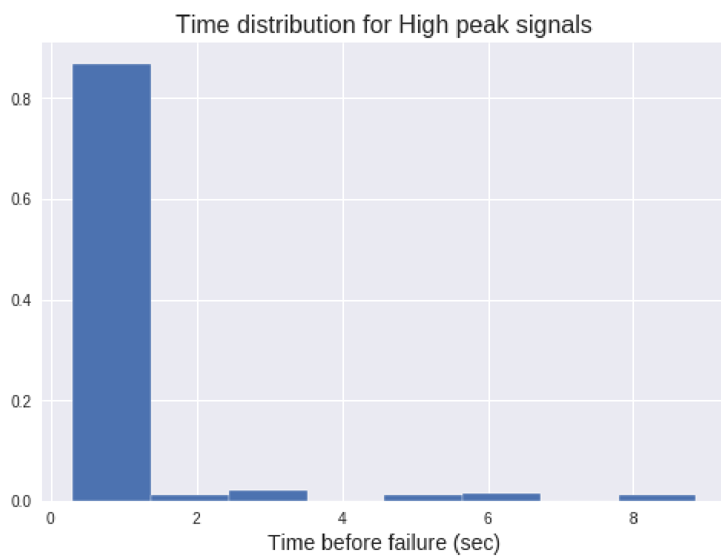

Fig. 6. We found that more than 90% of high acoustic values (absolute value greater than 1000) are around 0.31 seconds before an earthquake!

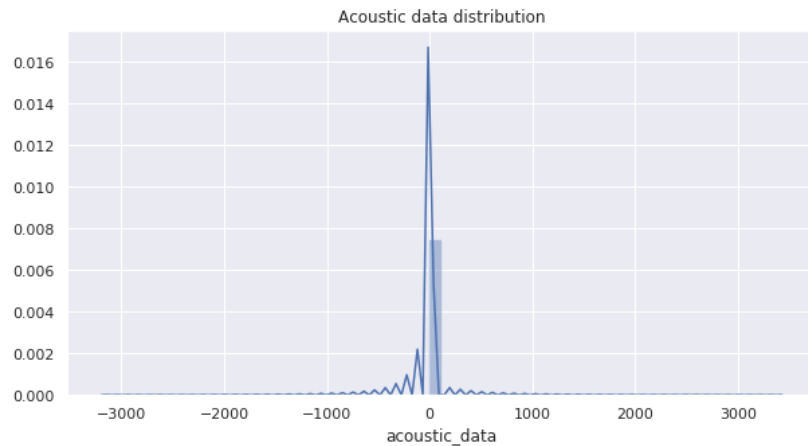

Fig. 7. Acoustic data distribution has a very high peak and we see outliers in both directions

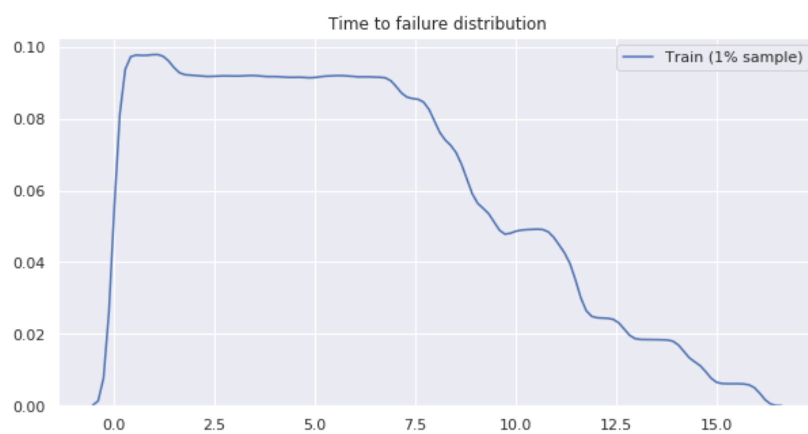

Fig. 8. Distribution of time to failure seems to be right skewed. It should be normalized.

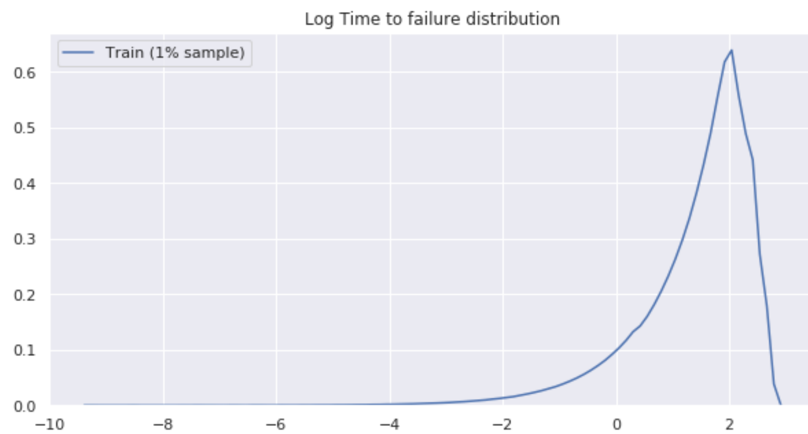

Fig. 9. Log transformed time distribution looks more left skewed than normal.

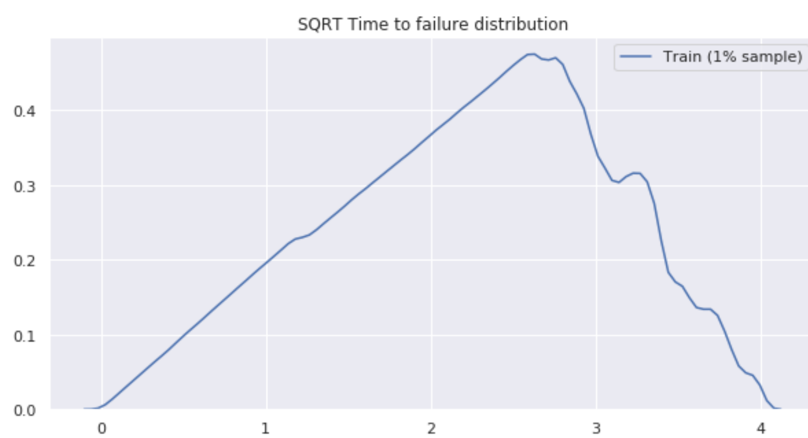

Fig. 10. The distribution of the time after square root transformation is still not ideally normal, but looks much better than the distribution of the time after log transformation.

4 Feature Engineering

LANL, working with quasi periodic seismic signals, achieved 0.89 coefficient of determination, using Random Forest technique and dividing data by 1.8 seconds time windows [6]. The most important features in LANL model were variance, kurtosis and threshold. We used a similar approach. 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 into time windows that each contain 0.3 seconds of data (1,500,000 observations), which is small enough compared with lab quake cycle that we have (8 to 16 sec). As we discussed above, more than 90% of high acoustic values (absolute value greater than 1000) are around 0.31 seconds before an earthquake. It makes sense to divide our data by 0.3 sec windows to reduce error at the end of the quake cycle.

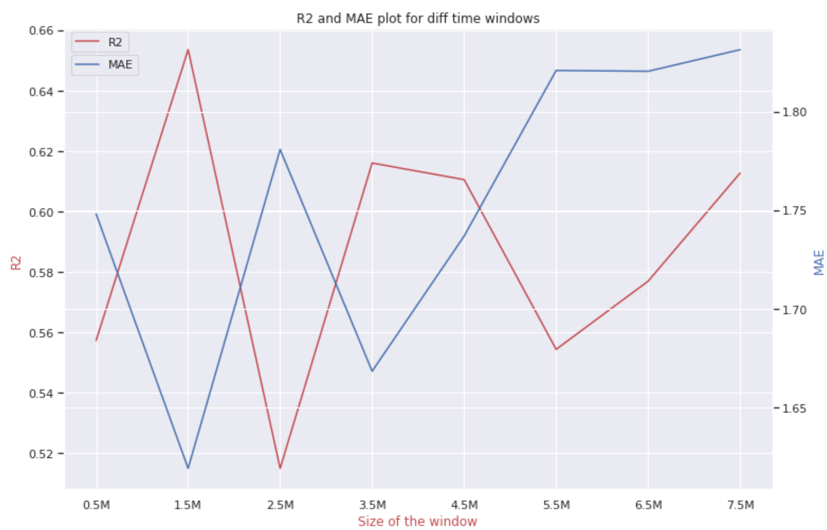

Fig. 11. We checked how sensitive our results are to the size of the window, and found that the highest R2 and smallest MAE we were able to get with 1.5M observations in each time window.

So our new data set is 419 time windows (0.3 sec long each). From each time window, we compute a set of 95 potentially relevant statistical features (e.g., mean, variance, kurtosis, min/max, threshold and so on), but using feature importance we found that only the following (in Table 3) were important:

Table 2. List of Engineered Features

Standard Deviation
90% Quantile
95% Quantile
99% Quantile
Absolute Standard Deviation
Average Rolling Standard Deviation for 100, 1000 and 10000 observations
Variance of Rolling Standard Deviation for 100, 1000 and 10000 observations
Minimum Rolling Standard Deviation for 100, 1000 and 10000 observations
1% Quantile of rolling standard deviation for 100, 1000, 10000 observations
5% Quantile of rolling standard deviation for 100, 1000, 10000 observations
10% Quantile of rolling standard deviation for 100, 1000, 10000 observations
90% Quantile of rolling standard deviation for 100, 1000, 10000 observations
95% Quantile of rolling standard deviation for 100, 1000, 10000 observations
99% Quantile of rolling standard deviation for 100, 1000, 10000 observations
Variance of Rolling Absolute Mean for 100, 1000 and 10000 observations

We apply different 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.

In order to avoid correlation between new features we applied principal component analysis. Instead of using 35 features, we created just 5 that represented 99.9% of the full data variation.

We use a 50/50 continuous split of the full time series for use as training and testing data sets respectively. Contiguity of train and test data sets is important, since we want to minimize contamination of the training data with information about the test data.

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 different techniques on a training data set (50% of the full data) before principal component analysis and after. Principal component analysis did not improve our findings significantly. We quantify the accuracy of our model using r^2 (the coefficient of determination) and MAE (mean absolute error), applying predicting model on test data set. The best results we achieved using Extra Trees Regressor technique: MAE: 1.61, $r^2 = 0.65$.

Hyper-parameters for this technique are 'maxdepth': 25, 'maxfeatures': 'auto', 'minsamplesleaf': 16, 'Minimum Sample Split': 30, 'minweightfractionleaf': 0.06225661421078982, 'nestimators': 400. The Extra Trees Regressor overview is presented in the section 2.2.

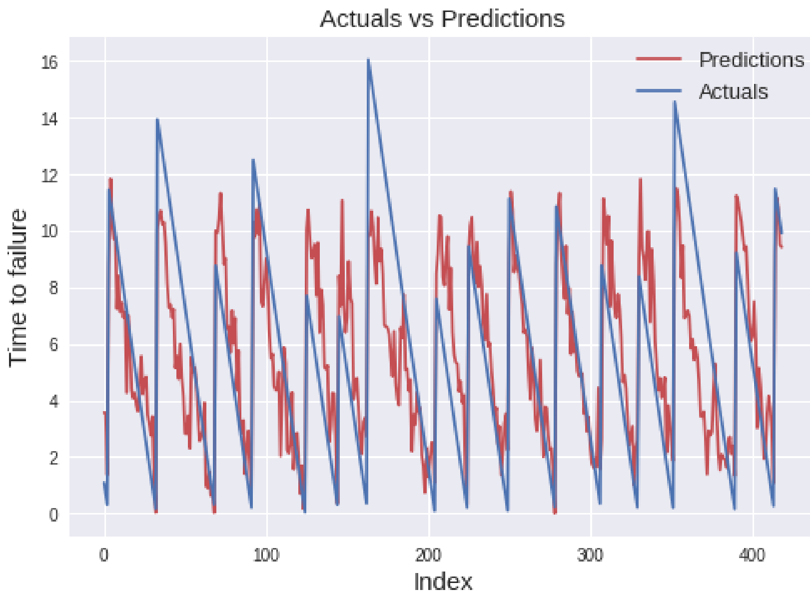

Fig. 12. When making a prediction (red curve), we emphasize that there is no past or future information considered: each prediction uses only the information within one single time window of the acoustic signal.

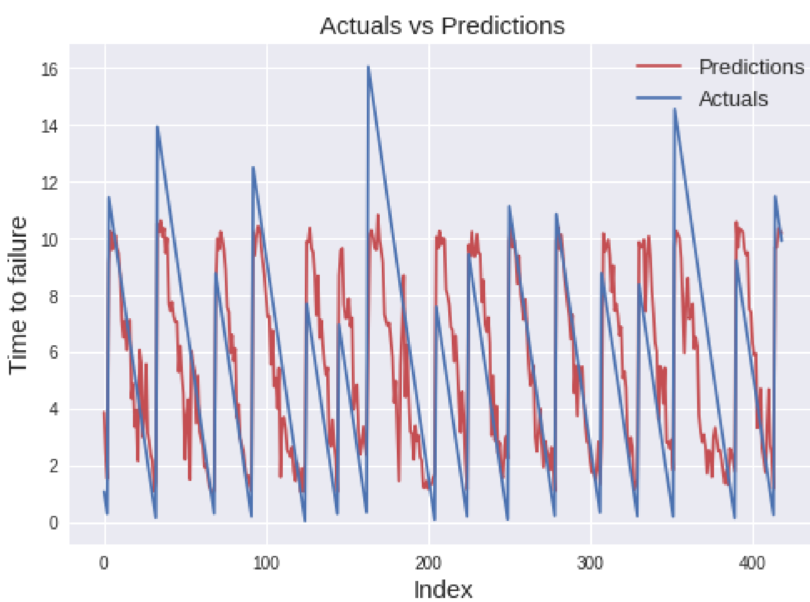

Fig. 13. Results we achieved using Ada Boost Regressor: MAE: 1.67, r2score: 0.62. Hyper-parameters for this technique are:'learningrate': 0.018261985736355728, 'loss': 'square', nestimators = 500, baseestimator=Ridge(alpha=1).

6 Analysis

The most accurate results with coefficient of determination 0.65 and mean absolute error 1.61 seconds we got using Extra Trees Regressor. The most important features are shown on Fig. 12. Top 5 are 90% and 95% Quartiles rolling standard deviations, standard deviation of rolling absolute mean, average rolling standard deviation.

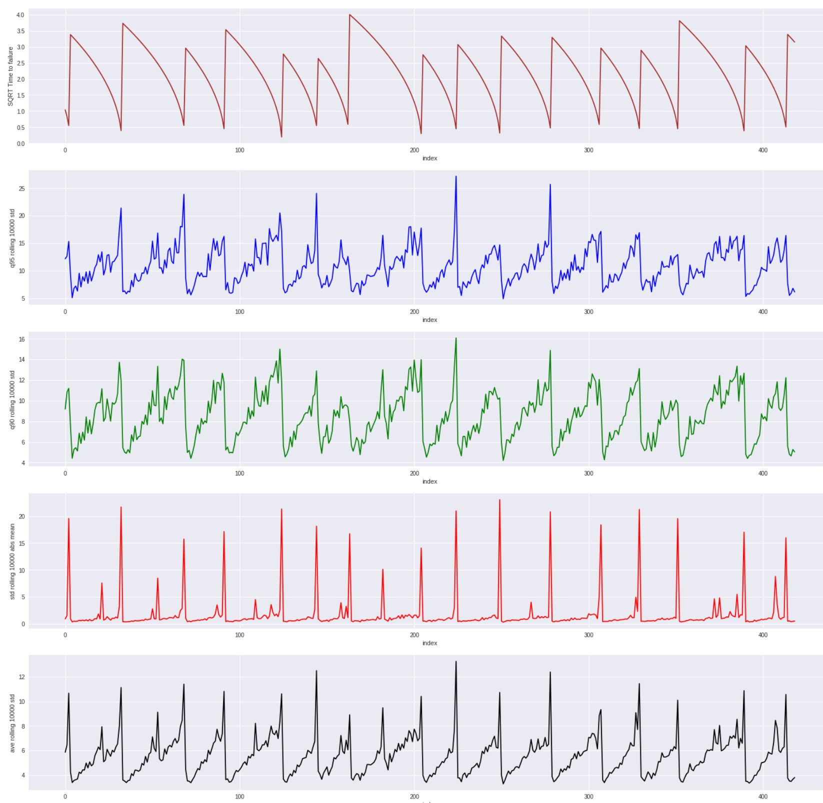

Fig. 14.

7 Ethics

Our responsibility to report our results has to be weighed with our responsibility not to cause social disturbance. If the method demonstrated in this paper is scaled and applied to predict natural earthquakes this balance must be considered. Unwarranted predictions could have affects on personal property value while failure to report warranted predictions could result in loss of life or avoidable property damage [8].

8 Conclusions

Given the more realistic data, Machine Learning can provide failure forecasts based on small time windows. The acoustic signal measurement is an indicator of imminent failure. The recurrence interval is not needed to make a prediction. The Extra Trees Regressor machine learning technique provides the most accurate results.

In this study we are using only 157 seconds of data. Future work should introduce higher volumes of data to determine if accuracy can be improved. Also the combination of the acoustic signal and the recurrence interval should be tested to determine it's affect on accuracy.

9 Acknowledgment

This research was supported by Dr. Michael L. Blanpied, U.S. Geological Survey. We thank Dr. Michael L. Blanpied for comments that greatly improved our work.

References

1. Jackson, D.D.: The 2004 parkfield earthquake, the 1985 prediction, and characteristic earthquakes: Lessons for the future. Bulletin of the Seismological Society of America **96**(4B) (2006) 397–409 https://doi.org/10.1785/0120050821.
2. Ikari, M.J.: Slip weakening as a mechanism for slow earthquakes. Nature Geoscience **6** (May 2013) 468 https://doi.org/10.1038/ngeo1818.
3. Rousset, B., Bürgmann, R., Campillo, M.: Slow slip events in the roots of the san andreas fault. Science Advances **5**(2) (2019)
4. Rouet-Leduc, B.: Los Alamos machine learning discovers patterns that reveal earthquake fault behavior. Los Alamos National Laboratory, P.O. Box 1663 Los Alamos, NM 87545. Science briefs 2018 edn. (March 2018)
5. Kaggle: Lanl earthquake prediction. https://www.kaggle.com/lanl (March 2019)
6. Rouet-Leduc, B.: Machine learning predicts laboratory earthquakes. Geophysical Research Letters **44**(18) (August 2017)
7. Geurts, P.: Extremely randomized trees. Springer Science + Business Media, Inc, (March 2006)
8. Ayhan Sol, H.T.: The ethics of earthquake prediction. Science and Engineering Ethics **10**(4) (December 2004) 655–666