



Explainable machine learning for labquake prediction using catalog-driven features

Sadegh Karimpouli^{a,*}, Danu Caus^{b,c,d}, Harsh Grover^{b,c,d}, Patricia Martínez-Garzón^a, Marco Bohnhoff^{a,e}, Gregory C. Beroza^f, Georg Dresen^{a,g}, Thomas Goebel^h, Tobias Weigel^{b,c,d}, Grzegorz Kwiatek^a

^a Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, Potsdam, Germany

^b DKRZ German Climate Computing Center, Hamburg, Germany

^c Helmholtz Center Hereon, Geesthacht, Germany

^d Helmholtz AI, Germany

^e Department of Earth Sciences, Free University Berlin, Berlin, Germany

^f Department of Geophysics, Stanford University, Stanford, CA, USA

^g Institute of Earth and Environmental Sciences, Universität Potsdam, Potsdam, Germany

^h Center for Earthquake Research and Information, University of Memphis, Memphis, USA

ARTICLE INFO

Article history:

Received 6 June 2023

Received in revised form 21 August 2023

Accepted 5 September 2023

Available online 11 October 2023

Editor: R. Bendick

Keywords:

labquake prediction

explainable ML

catalog-driven features

time to failure

ABSTRACT

Recently, Machine learning (ML) has been widely utilized for laboratory earthquake (labquake) prediction using various types of data. This study pioneers in time to failure (TTF) prediction based on ML using acoustic emission (AE) records from three laboratory stick-slip experiments performed on Westerly granite samples with naturally fractured rough faults, more similar to the heterogeneous fault structures in the nature. 47 catalog-driven seismo-mechanical and statistical features are extracted introducing some new features based on focal mechanism. A regression voting ensemble of Long-Short Term Memory (LSTM) networks predicts TTF with a coefficient of determination (R^2) of 70% on the test dataset. Feature importance analysis revealed that AE rate, correlation integral, event proximity, and focal mechanism-based features are the most important features for TTF prediction. Results reveal that the network uses all information among the features for prediction, including general trends in high correlated features as well as fine details about local variations and fault evolution involved in low correlated features. Therefore, some highly correlated and physically meaningful features may be considered less important for TTF prediction due to their correlation with other important features. Our study provides a ground for applying catalog-driven to constrain TTF of complex heterogeneous rough faults, which is capable to be developed for real application.

© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

1. Introduction

Earthquake forecasting and improvement of probabilistic seismic hazard assessment are persistent and challenging problems in geoscience, primarily due to the complexity of deformation processes, evolving fault structure and varying mechanical behavior of geomaterials. Recent advances in machine learning (ML) algorithms and hardware provided new perspectives and tools to the seismology community (Johnson et al., 2021). Basic signal processing analysis is within the scope of ML applications, including earthquake event detection (Mousavi et al., 2020), phase picking (Zhu et al.,

2019) and association (McBrearty et al., 2019), as well as hypocenter determination (Mousavi and Beroza, 2020), among others (as reviewed by Ren et al. (2020)). At the same time, data-driven ML-based approaches were applied to predict the time-to-failure (TTF) in laboratory experiments (e.g. Rouet-Leduc et al., 2017) using Acoustic Emission (AE) data and associated measurements. In general, approaches to predict TTF in laboratory experiments may be categorized into three groups based on the input features fed into the ML network as: a) AE-driven features, extracted directly from continuous AE signals, b) Geodetic-driven features, extracted from geodetic measurements, and c) Catalog-driven features, extracted from earthquake or seismicity catalogs.

Recently, several studies applied ML-based techniques to forecast TTF using stick-slip deformation experiments in a biaxial apparatus on smooth faults, which is an analog of the earthquake

* Corresponding author.

E-mail address: sadegh.karimpouli@gfz-potsdam.de (S. Karimpouli).

cycle (Byerlee and Brace, 1968). In addition to the mechanical data, continuous AE waveforms serve as input in an effort to predict TTF (Rouet-Leduc et al., 2017). The experimental setup enabled recordings of AEs from hundreds of stick-slip cycles during a single experiment, providing large datasets that are well-suited for analysis using ML algorithms. AE-driven features may comprise either raw AE signals (e.g., Pu et al., 2021; Wang et al., 2022) or statistical features such as mean, variance, skewness and kurtosis of signal distribution, time correlation features via the integral of power spectrum over a narrow frequency band and autocorrelation features (Rouet-Leduc et al., 2017). These features have been studied extensively and were used in recent years to successfully predict TTF, shear stress, fault friction, displacement, and slip velocity (e.g., Bolton et al., 2019; Jaspersen et al., 2021; Johnson et al., 2021; Lubbers et al., 2018; Rouet-Leduc et al., 2018, 2017; Wang et al., 2021). Rouet-Leduc et al. (2018) demonstrated that the AE variance calculated from the waveform amplitudes is the most sensitive feature. Shreedharan et al. (2021) showed that active source acoustic data can be used to predict labquakes and Laurenti et al. (2022) utilized AE variance for both TTF prediction and forecasting fault zone shear stress. Recently, Wang et al. (2022) directly fed AE signals into a transformer model to forecast near-future fault friction and Borate et al. (2023) demonstrated how physics-informed models can improve ML/DL methods.

Geodetic-driven features have also been used for prediction of laboratory earthquakes. For instance, Corbi et al. (2020, 2019) recorded synthetic GPS data in an experimental subduction model and predicted slip events (akin to lab earthquakes) in an analog material (TTF prediction and slip-event imminence classification) using features extracted from displacement data. These features include cumulative displacement of points, velocity of points, correlation between subsequent frames, standard deviation, variance, skewness, and kurtosis of the velocity field, threshold of velocity for target points, and cumulative displacement and velocity vector length. Corbi et al. (2019) demonstrated that the most crucial feature was the trench-parallel component of cumulative displacement of points.

Other researchers have utilized catalog-driven features, containing physically explainable patterns during the preparatory phase potentially preceding earthquakes, which is demonstrated from theoretical concepts and laboratory measurements to reservoir-scale and megathrust observations (Picozzi et al., 2023). McBeck et al. (2020) conducted a study to characterize the evolving fracture network in triaxial compression experiments performed with synchrotron X-ray computed tomography and predicted the proximity of sample failure. The authors suggested that evolution of the strain energy density field improved reliable predictions of the proximity to failure compared to other metrics used to measure rock deformation. Also, in a biaxial shear experiment, (Bolton et al., 2019) performed unsupervised clustering of the first two principal components of 43 AE-based features (similar to Rouet-Leduc et al., 2017), where they identified temporal trends and precursors and showed different clusters that determined the preparatory phase of the fault.

Panakkat and Adeli (2007) extracted features from the Gutenberg-Richter (GR) distribution and characteristic earthquake distribution to predict earthquake magnitude using various artificial neural networks (ANN). Reyes et al. (2013) used the GR-law's b -value, Bath's law, and Omori-Utsu's law as the input data for an ANN to determine the probability of an earthquake in the next five days in Chile. Following this work, Asencio-Cortés et al. (2017) used the same features, as well as mean magnitude, coefficient of variation, maximum magnitude over a temporal window, and other features, to classify earthquake magnitude in Tokyo into larger or smaller than $M \sim 5$ using ANNs. Lubbers et al. (2018) used the same AE signals as Rouet-Leduc et al. (2017), but extracted catalog-

driven features such as cumulative statistics of event count and amplitude computed for a specific time window. Using a Random Forest (RF) model, the authors predicted shear stress and time since the last failure event with high accuracy and TTF with a lower performance. Picozzi and Iaccarino (2021) extracted a variety of temporal features from an induced seismicity catalog of The Geysers geothermal field in California. The extracted features included duration of event groups, inter-event time, individual moment magnitudes and moment rate, b -value and completeness magnitude, fractal dimension, proximity and Shannon's information entropy. They used a recurrent neural network (RNN) to predict the preparatory phase of magnitude $M \sim 4$ earthquakes. Their results showed the potential of using catalog-driven features and RNN to predict large induced seismic events.

This study pioneers in employing physics-informed catalog-driven features derived from AE data recorded during stick-slip experiments on rough faults for TTF prediction. The majority of recent laboratory earthquake studies on similar topics focused so far on the double-shear test. Although the experimental setup to generate laboratory quakes has been discussed extensively (Marone, 1998; Niemeijer et al., 2010), it produces repetitive patterns of quasi-periodic slips, since it is a relatively homogeneous system with smooth fault surface (Johnson et al., 2021). This type of experiment favors ML training and consequently the TTF prediction, as the experimental procedure itself produces a huge amount of repetitive data in a framework where roughness is evolving very slowly. However, natural faults display larger roughness and structural complexity (e.g. fault networks) where the individual fault patches interact when major fault slip occurs (Johnson et al., 2021). Thus, our study for the first time attempts to handle the case of a complex rough fault and fault network as typically observable in nature. Employing a regression voting ensemble (An and Meng, 2010), consisting of 10 Long Short-Term Memory networks (LSTMs), we analyze the potential of the developed catalog-driven features for constraining the TTF. Following to applying an explainable ML method, we evaluate and rank the features based on their importance for predicting TTF, also allowing us to interpret the relations between different features and explain the ML predictions.

2. Experimental set-up and data acquisition

2.1. Sample material and stick-slip experiments

We use data from triaxial stick-slip experiments performed on the three samples cored from the same block of Westerly granite (WgN04, WgN05 and WgN07) (Goebel et al., 2012). The mineralogical composition of the samples is 28% quartz, 33% plagioclase, 33% K-feldspar and 5% mica. Initial porosity was 2% and average grain size is 750 microns. Cylindrical samples were precision-ground to 102–107 mm in length and the same diameter of 40 mm. To produce rough faults, all specimens were prepared with 1.5–2.2 cm deep notches inclined at 30° to the specimen axis, subjected to an isostatic confining pressure of 75 MPa, and then loaded at constant piston displacement rate until shear failure occurred, creating a rough fault surface (Fig. S1, in the supporting text). In the following, in all three samples, the confining pressure was raised to constant value of 150 MPa to lock the faults and progressive axial loading at constant displacement rate of 0.33 $\mu\text{m/s}$ was resumed to induce stick-slip events (Fig. 1) (Goebel et al., 2012).

We measured the axial force using an internal load cell with accuracy of ± 0.05 MPa. To measure seismic activity, 14 piezoelectric AE sensors with a resonant frequency of 1 MHz were placed in brass housings and glued directly to the sample surface using a low viscosity epoxy and appropriate holes in the rubber jacket (Fig. S1). Two additional AE sensors were placed in the top and bottom steel plugs. Full waveforms were recorded with 16-bit resolution

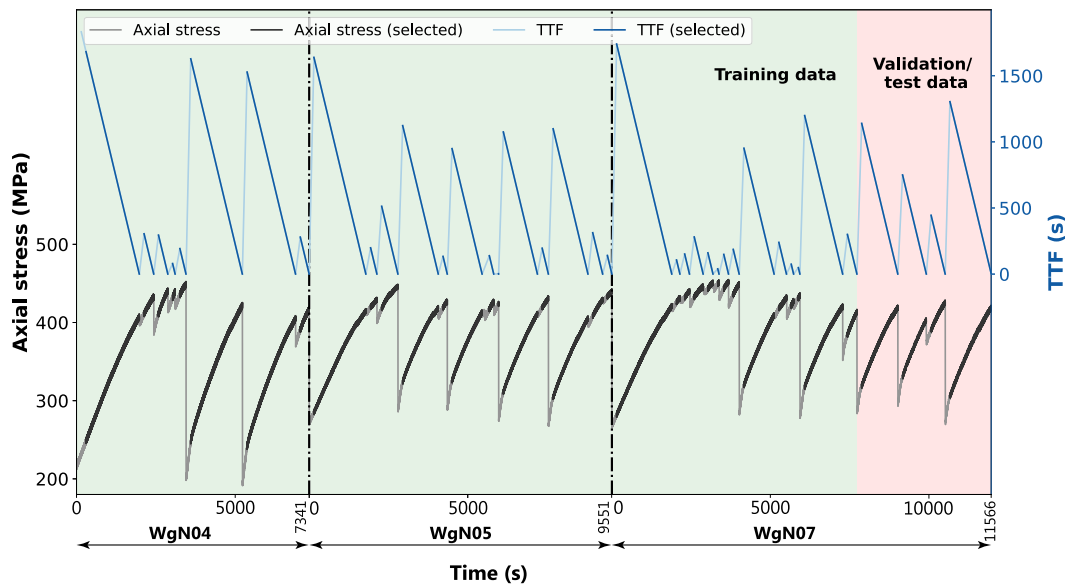


Fig. 1. TTF and axial stress values of three experiments (WgN04, WgN05, WgN07), separated by vertical dashed lines. All data are combined into a long time series and divided into 80% training data (green background) and 20% validation/test data (red background). Each small and large slip event (shown as a stress drop) is considered a failure event which we aim to forecast. As the features are calculated using a moving time window of different widths (see Table 1), data from 150 seconds after each axial stress drop are removed to avoid interference with the previous slip. The selected stress and TTF time series are shown with darker black and blue colors, respectively.

at 10 MHz sampling rate. The recording system was operated in triggered mode, which means the waveform recording is triggered when a specific amplitude threshold is exceeded on at least one AE sensor (see Fig. S2 in the supporting text for some examples of raw acoustic data with different magnitudes). To monitor *P*-wave velocities during damage evolution and to continuously update the velocity model, horizontal and vertical sensors were used periodically as sensor-receiver pairs of ultrasonic pulses (cf. Dresen et al., 2020).

2.2. AE catalog

The central processing steps to extract the time-dependent AE characteristics analyzed here from AE data are summarized in the following. They are described in detail in Kwiątek et al. (2023). First *P*-wave arrivals of AE events were picked automatically using the Akaike Information criterion, followed by pick refinement employing a modified Convolutional Neural Network scheme (Ross et al., 2018). Based on the continuously updated, quasi-anisotropic velocity model, hypocenter locations were determined using a grid search algorithm paired with the Coyote optimization algorithm (Pierezan and Dos Santos Coelho, 2018). We estimate the hypocenter location accuracy to be about ± 2 mm (cf. Dresen et al., 2020). We correct first-arrival *P*-wave amplitudes for hypocentral distance, incidence angle and differences in coupling using an ultrasonic calibration technique (Kwiątek et al., 2014). The AE magnitude was calculated from first *P*-wave amplitudes using:

$$M_{AE} = \log_{10} \sqrt{\frac{1}{N} \sum_{i=1}^n (A_i R_i)^2} \quad (1)$$

where A_i and R_i are corrected first *P*-wave amplitude and source-receiver sensor distance respectively (cf. Dresen et al., 2020; Goebel et al., 2012). The seismicity catalogs for WgN04, WgN05 and WgN07 contained 102540, 240328 and 199255 events, respectively, and contained origin time, location in the local Cartesian coordinate system and AE magnitude. The AE magnitude of completeness have been estimated for all catalogs using a goodness of fit method (Wiemer and Wyss, 2000) leading to $M_C^{AE} = 1.5$.

3. Feature extraction and processing

3.1. AE-catalog-driven features

All features used in this study are summarized in Table 1. We extracted different features over the time, space and magnitude dimensions (or a combination of them) to characterize the preparatory processes preceding the occurrence of slip events known as laboratory earthquakes (or labquakes). The features are described in detail in the supporting text and some of them are illustrated for WgN07 in Fig. 2. For a more general discussion of the physical meaning of selected features, see Kwiątek et al. (2023) for details.

AE event rate (n): The AE event rate refers to the number of events per second using AEs above the magnitude of completeness (Fig. 2) and as such reflects the damage occurring in the sample during loading. Sano et al. (1981) discovered that the AE event rate increased as the loading rate increased in an unconfined compressive strength test on a granite sample. Similar behavior has been shown by Bolton et al. (2021) in double-shear test. Dresen et al. (2020) observed a less pronounced increase in AE rate leading to failure along a rough fault. They also concluded that the AE rate and slip rate are correlated, since intermittent bursts of AE activity prior to and during slip events are followed by an increase in AE rates.

GR law's b -value (b): The Gutenberg-Richter (GR) law expresses a linear relationship between the magnitude and the logarithm of the total number of seismic events below that magnitude. The slope of this relationship is commonly called b -value and indicates the proportion between large and small events. Studies have shown that the b -value depends on stress (Kwiątek et al., 2014; Lei, 2006; Rivière et al., 2018), damage (Main, 1991), as well as strain localization and geometric complexity (Goebel et al., 2017) (Fig. 2).

Correlation integral (c): The correlation integral is a measure of the probability of two points being separated by a spatial distance less than a certain value (Henderson et al., 1999). As such, it is describing the level of point-clustering with 0 and 1 corresponding to point-clustering and random distribution, respectively (Kagan and

Table 1

A summary of all features used in this study.

Parameter		notation	Time windows (s)	Dimension sensitivity	Commentary
AE event rate		<i>n</i>	23, 45 90, 180	time	Calculated directly from AE catalog
GR law's <i>b</i> -value		<i>b</i>	10, 30, 90, 180	time	
Correlation integral		<i>c</i>	90, 180 (r = 5 mm)* 45, 90 (r = 20 mm)	space-time	
Interevent time distribution features	Ratio (deviation from uniform distribution at edges of the empirical distribution)	<i>r</i>	23, 45, 90, 180	time	Calculated directly from AE catalog
	Deviation from uniform distribution (χ^2 test)**	<i>rx2</i>			
Fractal dimension		<i>d2</i>	45, 90, 180	space-time	Calculated from AE catalog, boxcounting method
Magnitude correlation		<i>dm</i>	90	time	Calculated from AE catalog
Clustering and localization features	Product of T and R***	<i>trp</i>	25, 50, 100	space-time-magnitude	Clustering analysis outcomes
	Quotient of T and R	<i>trq</i>			
	Proportion of foreshocks	<i>pfo</i>			
	Proportion of aftershocks	<i>paf</i>			
	Proportion of mainshocks	<i>pma</i>			
Focal mechanisms and deviatoric stress tensor based features	Median fault plane variability	<i>vm</i>	100, 200	space-time	Full moment tensor inversion
	Maximum principal stress plunge	<i>s1d</i>	90, 180		Stress tensor inversion
	Maximum principal stress variance	<i>svar</i>			
	Stress ratio	<i>sr</i>			

* 'r' is a scale limit for spatial distance between coordination of two events. ** ' χ^2 ' is the Chi-squared distribution. *** 'T' and 'R' denote the time and space components of the proximity formulation (for more details see the text and supporting information).

Knopoff, 1980). However, it can also be used as an indicator of the localization of the event hypocenters (Fig. 2).

Interevent time ratio (*r1*, *rx2*): Interevent time ratio introduced by Van Der Elst and Brodsky (2010) is a measure of local-in-time temporal clustering or anti-clustering of seismicity. The probability density function of interevent time ratios is uniform for random (Poissonian) distribution of events in time (cf. Kwiatek et al., 2022). Clustering or anti-clustering is here measured with two features. The first measures the deviation of the observed distribution from the uniform (*r1*) distribution at the edges of the PDF. The second measures the deviation of the observed distribution from uniform distribution using chi-squared (*rx2*) formulation. Both features do not quantify the statistical significance of the deviation from the uniform distribution (cf. testing of significance in Kwiatek et al., 2022).

Fractal dimension (*d2*): It has been suggested that the spatial distribution of AE locations exhibits fractal self-similarity behavior (Hirata et al., 1987). Fractal dimension can be calculated using either the pair correlation integral (Hirata et al., 1987) or the box-counting method (Sadovskiy, 1984). The fractal dimension is highly dependent on the roughness of the fault plane, and its evolution close to failure can either gradually decrease due to fault nucleation, or increase, signifying an overall increase in AE activity over the entire fault surface as a consequence of increased contact area between the two faces of the fault (Kwiatek et al., 2023; Lei and Ma, 2014).

Magnitude correlation (*dm*): For each time window we tested whether the vector of magnitudes ordered in time behaves as if it would be randomly drawn from a Gutenberg-Richter distribution (see e.g. Kwiatek et al. (2022) for details). Magnitude correlations

suggest local-in-time accelerations or decelerations of seismic processes that are not expected in a stationary Poissonian process.

Clustering and localization features (*trp*, *trq*, *pma*, *pfo*, *paf*): Earthquake clustering analysis was developed by Baiesi and Paczuski (2005), who defined a pairwise distance between earthquakes in the space-time-magnitude domain. (Zaliapin et al., 2008) expanded on this by introducing magnitude-normalized time and space components, denoted as T and R. The spatio-temporal proximity between events can be conveniently expressed as the product of their temporal (T) and spatial (R) components, scaled by the magnitude of the earlier event. In this study, we calculated the averaged products and quotients (or ratio) of these two components (*trp* and *trq*). The first parameter (*trp*) describes the general space-time localization process. The *trq* parameter picks up the inverse relation between T and R components (e.g. small T and large R and vice versa, Martínez-Garzón et al. (2019)).

The nearest-neighbor proximities form a bimodal distribution (Zaliapin et al., 2008; Zaliapin and Ben-Zion, 2013). This distribution may be decomposed into background, approximately Poissonian seismicity distributed as the larger event proximities, and clustered seismicity. The separation into background and clustered seismicity allows further decomposition into foreshocks, mainshocks and aftershocks. In this study, we calculate fractions of each event type within a selected space-time window (Fig. 2).

Focal mechanisms and deviatoric stress tensor based features (*vm*, *s1d*, *sr*, *svar*): The inversions of full moment tensors and focal mechanisms allow deciphering the micro-kinematics at the grain-scale (Kwiatek et al., 2014). The stress tensor inversion from AE-derived focal mechanisms allows computing spatial resolution of local directions of principal stress axes (Martínez-Garzón et al., 2014;

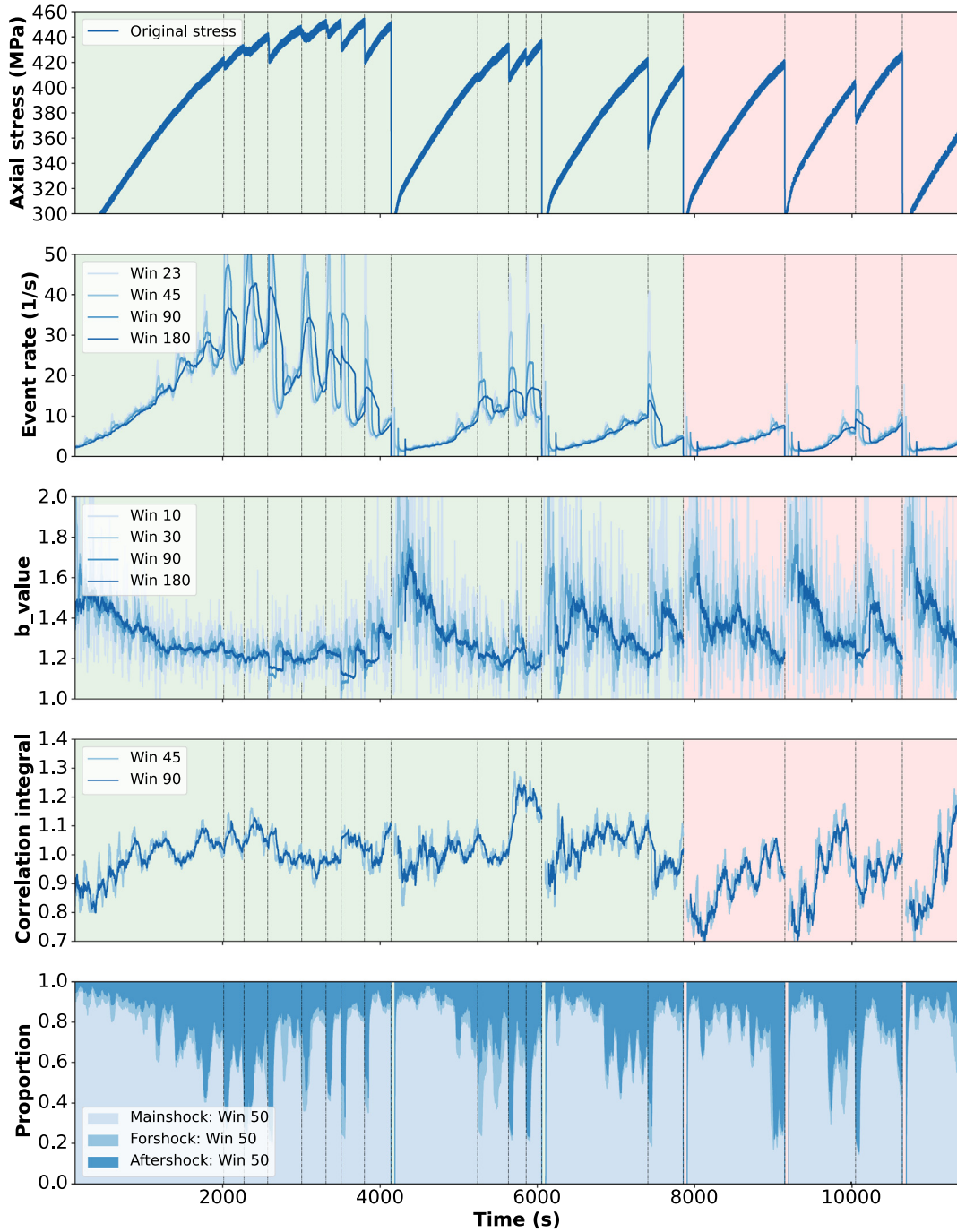


Fig. 2. Representation of several features as an example in WgN07 to show both training and test data. The first figure (axial stress) reveals the failure times. From top to bottom: axial stress, AE event rate (n), b -value (b), correlation integral (c), and proportion of mainshock, foreshock and aftershock. Each feature is computed taking the past data from a time window of selected length, which is indicated by 'Win' in seconds. Green and red background are the same as in Fig. 1.

Vavryčuk, 2014) and obtaining the ratio between magnitudes of principal stress axes (stress shape ratio).

Following Kwiątek et al. (2023), we use focal mechanism data to determine the median fault plane variability (vm). This feature characterizes the heterogeneity in AE-derived focal mechanisms (Goebel et al., 2017). Additionally, using the results of stress tensor inversion of AE focal mechanisms, we compute the plunge of the maximum principal stress ($s1d$), stress shape ratio (sr), and maximum principal stress variance ($svar$). The $s1d$ parameter characterizes the deviation of locally derived stress orientations from the globally imposed stress field by axial loading. It primarily reflects the stress concentration close to the fault surface and the amount

of shear-enhanced compaction (see discussion in Kwiątek et al. (2023)). The sr parameter defines the shape of the principal deviatoric stress ellipsoid. Finally, the $svar$ parameter describes the spatial variability of local stress tensors derived. The larger the value of $svar$ parameter, the more heterogeneous the local stress field is.

3.2. Feature pre-processing

For each experiment, we calculated 47 time series on the basis of 16 unique seismo-mechanical and statistical features using moving time windows of different lengths (see Table 1). We constructed the time series by concatenating all data from the three

samples, as shown in Fig. 1, to provide a dataset spanning a longer period. We assumed small and large axial stress drops to represent stick-slip events, separating different stress cycles including loading and unloading, which formed the basis for all subsequent analyses. For example, for TTF estimates, time was set to zero at each axial stress drop, increasing backward until the previous drop or the beginning of the test (Fig. 1). It should be noted that while calculating the time series, we omitted the first 150 s after each stick-slip event to avoid mixing of data from two different stick-slip cycles and to not pollute the next cycle with the preparatory activity of the previous cycle.

We normalized each feature within each stick-slip cycle between 0 and 1 using min-max normalization. We used a time interval of 150 s in length to construct 19,786 multivariate sequences of time series at every one second. Since we computed all 47 features with a 0.5 s time step, it resulted in an input data size of $19,786 \times 300 \times 47$ for each time interval. All data are divided into 80% training and 20% validation/test data based on the stick-slip cycles. The test data are used for both validation and test phase, since they are not contributing in training and are assumed as unseen data for the network (Saenger et al., 2021). We decided to do this in order to increase the amount of the training data.

Here, we analyze TTF before small and large stick-slip events. Stick-slips are directly detectable in axial load data (Fig. 1) in the form of sudden drops in the axial stress within less than 0.1 s (Kwiatkiewicz et al., 2023). The onset of the stick-slip is always marked with an occurrence of a large AE event(s) whose location can be considered a nucleation point of the following stick-slip. Large stick-slips/axial stress drops (>100 MPa) activate the whole fault surface, as revealed in the distribution of AE aftershocks following the slip. Similarly, most small slip events (<100 MPa) activate the entire fault surface but produce much less slip.

4. ML approach and results

4.1. TTF prediction with LSTM

Selected features characterize the temporal evolution of damage and local stress over and in the vicinity of the fault surface (Dresen et al., 2020). The input data in each time sample can be considered as point data where classical ML methods such as ANNs (Panakkat and Adeli, 2007), Random Forests (Rouet-Leduc et al., 2017), and Gradient Boosted Regression Trees (Ren et al., 2019) could be applied. However, these features are assumed as time series data and one could utilize deep learning methods such as Convolutional Neural Networks (Pu et al., 2021), Recurrent Neural Networks (Picozzi and Iaccarino, 2021), LSTM (Jasperson et al., 2021), Gated Recurrent Units and Transformers (Wang et al., 2022).

We tested a series of classical and deep networks and found the LSTM method (for basics of LSTM see Supporting text) to produce more accurate results in terms of prediction error. Following a regression voting ensemble learning, we use an average of 10 values predicted by 10 LSTMs to produce more robust predictions. We use a many-to-one LSTM training scheme, where we enforce a sequence length of 300 samples and 47 features. The batch size was 1000 and the hidden variables of the LSTM are reset after each batch of sequenced data. All models have two main layers: an LSTM layer with 256 units and a dense layer with 256 neurons. Each of these layers is followed by a batch-normalization and a dropout layer with a rate of 50%. We used the 'Adam' optimizer with a learning rate of 0.001.

Overall, predictions of the training data are more accurate than the validation data during the training phase (Fig. S4, in the supporting text). This means our ML models are highly susceptible to overfitting, which is mostly due to the limited available training data. Short training data and relatively few stick-slip events

are due to limited maximum displacements (up to ~ 4 mm vertical) in the experiments. Larger displacements can damage or break the rubber jacket surrounding the sample. Even if we applied overfitting-mitigation procedures such as application of a high dropout rate (50%) in our models, overfitting is still inevitable.

The results of our TTF prediction are displayed in Fig. 3. We find a statistically significant coefficient of determination (R^2) for the prediction which is more than 70%, with mean square error (MSE) for normalized values of 8.6×10^{-3} (Fig. 3a). Considering the complexity of the preparatory process and limited training data, the results provide a good constraint for TTF of stick-slip events (Fig. 1). To have a simple base model for comparison, we conducted a linear multivariate regression model via the stepwise algorithm (Montgomery and Peck, 1992) with all 47 features. After removing 7 features by the algorithm, the final model with 40 features shows $R^2 = 45.6\%$ and $MSE = 15.9 \times 10^{-3}$. This means that the ML network, which is able to capture complex and nonlinear relations, does produce predictions with about 55% improved accuracy (R^2) and 50% reduced error (MSE). In the following, we evaluate the relative importance of the features and analyze the TTF prediction results using the most important features to explain ML results.

4.2. AE features: importance versus correlation

Multivariate linear regression is mostly based on the linear correlation between dependent and independent variables, where the model coefficients are examined to evaluate the contribution of each variable. However, in a non-linear regression model, like in our case, the most correlated variables do not necessarily play the most important roles for prediction. To assess the importance of individual features and feature combinations, we employed the integrated gradient (IG) method, which computes gradients (i.e., changes) of the output with respect to the input for a deep network (Sundararajan et al., 2017) (see supporting text). More specifically, the IG method computes gradients of the output with respect to interpolated inputs and averaging/integrating the result for all the interpolated input features. To account for the TTF contributions of each feature, we computed the IG magnitudes for 100 test data points for all LSTM models, and calculated averages of the absolute values (Sundararajan et al., 2017). As illustrated in Fig. 3b (right side), the features are ranked based on their role for TTF prediction employing the LSTM networks, for which the AE rate ($n_{90, n_{180}, n_{45}}$), correlation integrals ($c_{rc20_{90}, c_{rc20_{45}}$), quotient of T and R (trq_{100}) and median fault plane variability (vm_{200}) came out as the seven most important features. We performed a TTF analysis using only these most important features (Fig. 3c) and the results suggest that the general TTF trends are captured even by seven features, although the statistical significance R^2 dropped to 56%. It should also be noted that among the most important features affecting the TTF prediction there are similar features, but with different time windows, such as: 45, 90 and 180 s for AE rate (n) and 45 and 90 s for the correlation integral (c). The length of time windows for quotient of T and R (trq) and the median fault plane variability (vm) are 100 and 200 s. This indicates that features calculated over medium to longer time windows are more informative.

We calculated the correlation coefficient between the estimated TTF and all features (Fig. 3d). As expected, the AE rates (n) for different time windows are highly correlated with TTF. However, the next important features derived from the IG method (Fig. 3b), which included the correlation integrals ($c_{rc20_{90}, c_{rc20_{45}}$), quotient of T and R (trq_{100}) and median fault plane variability (vm_{200}) all show a lower correlation with TTF, but still rank higher compared to other highly correlated features, such as proportion of mainshocks and aftershocks (pma, paf), median proxim-

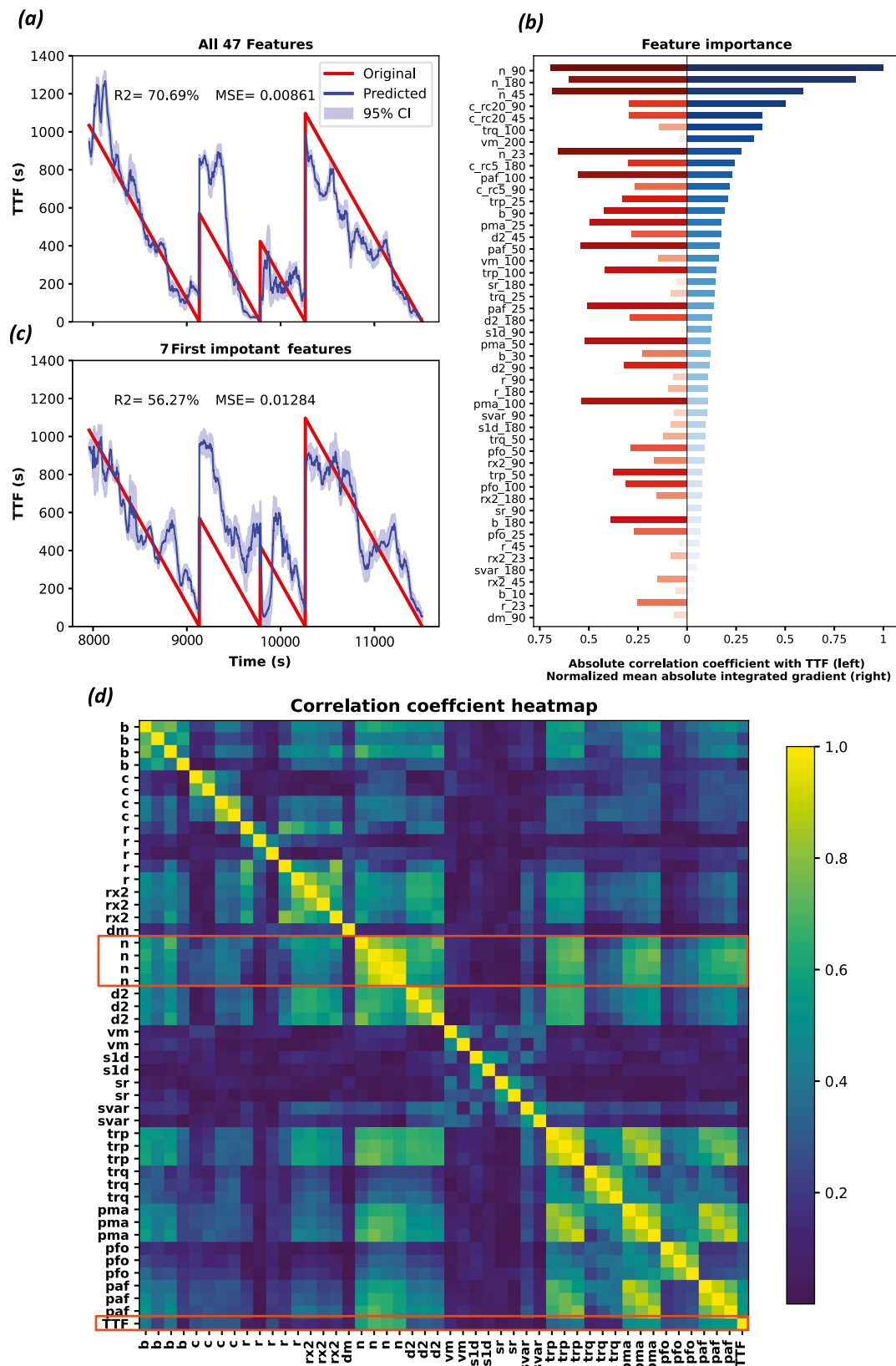


Fig. 3. TTF prediction results using the ensemble LSTM model with (a) all 47 features and (c) only the 7 most important features selected from the ranking list in (b). (b) The importance of each feature was computed using the IG method and sorted (right blue bars), as well as the correlation coefficient (absolute values) of the corresponding feature with TTF (left red bars). (d) Absolute values of correlation coefficients among all features and TTF (details can be found in the accompanying text).

ity (*trp*), and *b*-value. This is due to *n* being highly correlated with *pma*, *paf*, *trp*, fractal dimension (*d2*), and even *b*-value, as shown by a rectangle in the correlation matrix (Fig. 3d). Thus, the latter

features behave similar to *n* and do not seem to provide further independent information to improve accuracy of the model prediction. Consequently, other features with lower correlation with both

TTF and n , more significantly affect the model's performance. This is probably because their small details, which are not contained in n and its correlated features, fill the gaps of the model for a better prediction. See section 6 in the supporting text for more details about the importance of fault plane variability.

5. Discussion

Recently, ML algorithms have been successfully used to enhance seismic catalogs, pioneer prediction of stick-slip events in double-shear friction tests, and to analyze varying seismic characteristics (Mousavi and Beroza, 2023; Ren et al., 2020). In contrast to previous studies, the forecasting approach is applied to stick-slip experiments performed on rough faults, that bear more similarity with heterogeneous fault structures in nature, but make AI based forecasting notoriously difficult. Rough faults are more complicated yet closer to the (generally complex) faults observed in nature, whereas so far relatively simple planar faults were analyzed in TTF studies, based on a multitude of repetitive stick slip cycles. In this study, we predicted TTF using catalog-driven features with the main aims being to: 1) explain physically and statistically each feature individually and its relation with TTF, and 2) evaluate their role in a group of features for TTF prediction. Our results show that the employed ML algorithms allow predicting TTF of laboratory stick-slip events on heterogeneous rough faults with a R^2 of about 70%. The most important features for prediction include AE event rate, correlation integral, median spatio-temporal proximity from clustering analysis, and focal mechanism-based features. Among them, fault plane variability is introduced as an important feature which describe local stress evolution in the sample based on AE data, which to our knowledge have not been yet proposed. Consequently, the features cover a variety of physical processes involving space, time and magnitude dimensions, which control fault damage evolution and the associated labquake processes.

Event rate n is a key parameter which commonly increases when approaching failure in lab tests performed on rough faults (Dresen et al., 2020; Lei and Ma, 2014) and granular gouge (Johnson et al., 2013). It indicates the damage intensity of the granular material forming the fault zone and is commonly associated with slip rate. Without significant b -value variations, event rate n can be considered proportional to the seismic energy release rate, or seismic moment rate. However, this implies that AE event rate is in general expected to vary spanning a broad range, reflecting varying seismic efficiency of the monitored processes as well as monitoring conditions. This may potentially lead to normalization problems when used in ML algorithms due to different seismicity backgrounds and n range in different areas. In such cases, n may be effectively replaced with a (or a combination of) correlated parameter(s) such as trp , pma , paf that do not exhibit comparable normalization problems.

Although the correlation integral is typically used to compute the fractal dimension, it also contains time-space information that may be related to the preparation process (Dresen et al., 2020). For rough faults, clustering of AEs is observed at high load, indicating breakdown of cm-scale asperities forming the fault surfaces ahead of a major slip event (Goebel et al., 2012).

Features obtained from clustering analysis are significant since they capture time-space-magnitude dimensions physically and are highly correlated with TTF. The features trp and trq indicate localization processes in time and space (Ben-Zion and Zaliapin, 2020) when approaching the earthquake. The varying proportions of mainshocks, foreshocks and aftershocks (pma , pfo and paf) become visibly prominent during loading at high stress levels. Temporal changes of these parameters are controlled by the breakdown of small-scale asperities (mm to cm wavelength) in the sample, but grain scale roughness remains unchanged (Kwiatk et al., 2023).

Each asperity breakdown leads to smoothing out the local stress field in the asperity and local (i.e. within the sample) stress redistribution to the surrounding areas. This is evidenced by clustering of AE and presence of prominent AE foreshock-mainshock-aftershock sequences. Intensive occurrence of clustered AEs indicates therefore progressive expansion of failures of asperities, establishment of long-range stress correlations, which in the presence of high ambient stress are a prerequisite for rupture propagation of large events (Kwiatk et al., 2023).

Features based on focal mechanisms such as Vm are also important. For example, for TTF prediction in rock fracture tests, McBeck et al. (2020) characterized damage evolution monitoring orientation of the smallest dimension of the fractures, distance between fractures, fracture apertures, and anisotropy of fractures. Using these features in their model, the authors arrived at TTF predictions with $R^2 = 50 - 60\%$. This is somewhat less than presented in this study, which is possibly due to the large complexity of damage localization in performed friction tests. Nevertheless, Kwiatek et al. (2023) highlighted the importance of features based on AE micro-kinematics in identifying the approach to major slip. They identified signatures of alignment of AE-derived focal mechanisms expressed with lowered vm values. In addition, they found locally resolved stress fields become more homogeneous before slips, as indicated by reduction in local stress field variability ($svar$). Therefore, the focal-mechanisms based proxies indicate preparation of the fault surface for large slip sensing processes such as roughness removal, smoothing of the stress field and creation of the long-length scale stress correlation, as well as alignment of fractures (Ben-Zion et al., 2003; Dresen et al., 2020; see also discussion on Kwiatek et al., 2023).

5.1. Importance of different features

Employing features with varying time windows improves the TTF prediction (e.g., n or c in Fig. 3b). We also investigated the model performance using just one time window from each type of feature and tested respective model performance using 16 different features and two similar LSTM models; model-I consisting of the 16 most important features (following the order presented in Fig. 3b), and model-II with 16 different-unique features (i.e. we ignore the repetitive features with different windows and always choose the most important time window).

Fig. 4 shows the results from these tests. Although the goodness of fit criteria (R^2 and MSE) reveal that model-I with 16 most important features may achieve better results compared to model-II, the main trends are still captured by both models. In fact, the instability of the predictions with 16 features does not allow for a conclusive judgment, and increasing training data may provide a fairer comparison.

The IG method is also employed to investigate the feature importance for a model with 16 features (Fig. 4c) indicating that n , pma , trp , and $d2$ are the most significant features. Compared to the model employing 47 features (Fig. 3c), these 5 features exhibit a high correlation not only with TTF, but also with each other (Fig. 3d). This implies that highly correlated features with TTF are essential for a model to capture the run-up trend to failure. Similar to the models based on 47 features, the model with 16 different-unique features also employ moderate to low correlation features, such as c , trq , sr , vm , and $s1d$ to improve the prediction.

5.2. Potentials, limitations and future directions

ML methods are considered as data-dependent methods, which means their reliability is conditioned on whether the new test data is scattered within the distribution of the training data or not. For example, we trained the model using AE-derived data from

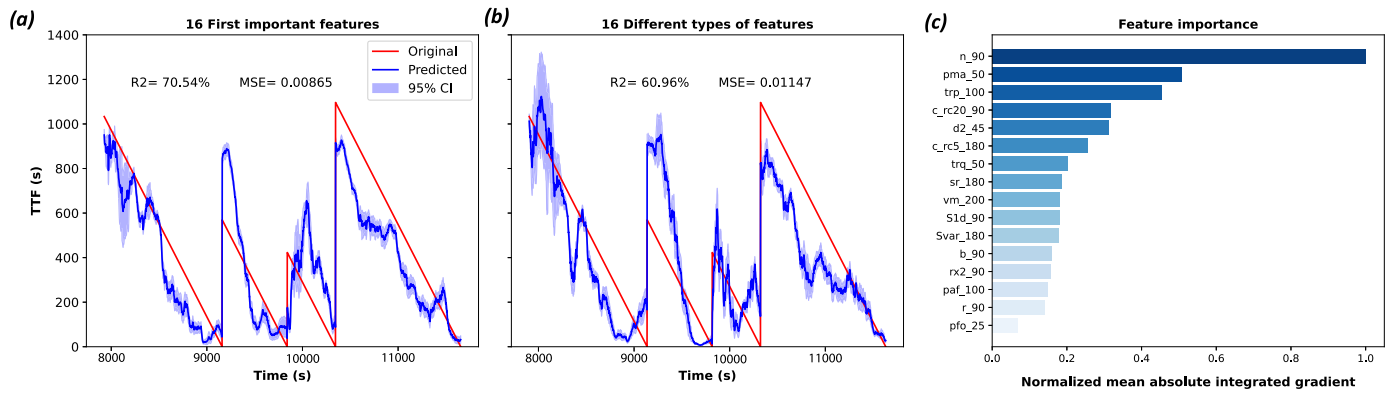


Fig. 4. Prediction results of TTF using ensemble LSTM model with (a) first 16 most important features and (b) 16 different-unique features. (c) Sorted 16 different-unique features based on their importance computed using the IG method.

stick-slip experiments on Westerly granite samples that deform by brittle deformation, i.e. cracking and frictional sliding. It is likely that because of using different sample material, deformation by different mechanisms (i.e. brittle creep, granular flow), with different fault structures and/or roughness, the trained network may not produce accurate predictions. To address this issue, the next step could be to enrich the training dataset with more experimental data that cover varying shear deformation scenarios (e.g. different confinement, axial loading rate), surface roughness (rough, sandpaper-polished, smooth), and/or a variety of rock materials. It is also beneficial to overcome the overfitting issue. However, collecting such a large dataset of stick-slips is a challenging task, regardless of the material involved. Limited training data can be offset by data augmentation methods, including Digital Twins (DT) and numerical simulations. By having access to a larger dataset, more complex networks could be trained, and transfer learning may be utilized to fine-tune models with limited experimental data, as suggested by recent studies (Jaspersen et al., 2021; Wang et al., 2021).

Scaling-up of experimental tests to natural earthquakes poses a key challenge, not least due to differences in boundary conditions and fault heterogeneity. Nevertheless, recent studies of induced and natural seismicity showed that ML and template matching techniques can significantly shrink the scale gap between laboratory and nature (Martínez-Garzón et al., 2023; Park et al., 2020; Picozzi and Iaccarino, 2021; Ross et al., 2018).

Obviously, field seismological data cannot provide similar spatio-temporal resolution down to laboratory scales with detection of events down to $M_W -10$. Likewise, spatial, temporal and characteristic structural scales differ enormously between deca-km-long faults in the field and cm-scale faults in laboratory samples. However, significant advances have been made in recent years implementing near-fault dense monitoring networks (Martínez-Garzón et al., 2021) and high resolution wide-band monitoring techniques (Durand et al., 2022) in nature and in-situ underground experiments. Recent field experiments show that one can monitor fracturing in-situ down to the cm-scale, but this requires dedicated complex networks composed of different sensors covering a limited volume of rocks. Nevertheless, targeted and densified instrumentation enable development of better seismicity catalogs with lower magnitudes of completeness. Lubbers et al. (2018) noted that ML models built on catalog-based features may reach a performance similar to models built on waveform-based features only for a very complete catalog, something currently not possible on Earth observations (Ren et al., 2020). Although this statement is generally correct, it should be noted that recent ML methods showed that they can improve earthquake catalogs using field data up to tens of times as compared to the conventional methods (Mousavi

et al., 2019; Ross et al., 2018). Moreover, ML models based on physics-based features derived from seismic catalogs enable a better understanding of the physical processes leading to earthquakes (Kwiatk et al., 2023). Finally, Mignan (2014) noted that foreshock activity tends to contain useful information on preparatory processes, leading to large earthquakes if the magnitudes of small foreshocks earthquakes are lower by at least 3 units of magnitude with respect to the following mainshock. From that perspective, laboratory earthquakes with moment magnitudes ranging from $M_W -10$ to $M_W -7$ are expected to include precursory information, as the moment magnitude of LSE and SSE observed here is comparable to an $M_W -4$ event (Dresen et al., 2020). Likewise, high-resolution seismic networks, supported with ML/AI based techniques for waveform processing, already provide similar resolution in dedicated field monitoring experiments (Mousavi and Beroza, 2023 and references therein).

6. Conclusions

In this study, we used an ensemble of 47 physics-based features, derived from AE catalogs that describe damage and stress evolution on the fault surface, during laboratory stick-slip experiments, performed on naturally fractured rough faults in Westerly granite samples. Data-driven features extracted from the AE catalogs were used as input for an ensemble of LSTM networks to predict TTF. This study provides the foundations to develop a flowchart for real case applications, which could be universally applied at different scales of the seismic process (from laboratory to the field). In this framework, it is important for us to use the known features and explain their importance for TTF prediction. Thus, we use: 1. physics-based catalog-driven features, which some of them like b-value are known for decades, and 2. explainable ML algorithms to rank the importance of each feature. Our investigations revealed that:

(1) The ML algorithms provide improved TTF predictions for labquakes on natural fractures. The improved predictions are due to the ability of neural networks to map non-linear relationships between disparate data features and the target variable, which is not easily done with, for example, standard multivariate regression.

(2) Feature importance analysis indicated that the most critical seismo-mechanical and statistical features for TTF prediction are AE event rate, correlation integral, median spatio-temporal proximity from clustering analysis, and focal mechanism-based features such as median focal mechanism variability. Among them, there are new features characterizing local stress evolution over the sample surface using moment tensor inversion and stress tensor inversion data from AEs.

(3) The important features for time to failure (TTF) prediction are a combination of high and low correlated features, rather than

solely relying on those highly correlated with TTF. In our case, the AE rate is identified as the most significant feature. Thus, other inter-correlated features with AE rate are deemed less important, even if they exhibit a strong correlation with TTF, since they contain similar data as those included in the AE rate features. The network utilizes small, yet novel details in other features such as correlation integral, proximity, and focal mechanism-based features to fill the prediction gap.

(4) Features calculated with medium to large time windows are generally ranked as more important than those computed with small time windows.

(5) The results emphasize the importance of utilizing features from various physical perspectives for predicting TTF, even if they are not highly correlated with TTF. This allows for the possibility of discovering new important features in the future that may play a prominent role in earthquake prediction.

CRediT authorship contribution statement

Sadeh Karimpouli: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Danu Caus:** Writing – review & editing, Software, Methodology. **Harsh Grover:** Writing – review & editing, Software, Methodology. **Patricia Martínez-Garzón:** Writing – review & editing, Resources, Data curation, Conceptualization. **Marco Bohnhoff:** Writing – review & editing, Validation. **Gregory C. Beroza:** Writing – review & editing, Validation. **Georg Dresen:** Writing – review & editing, Validation, Conceptualization. **Thomas Goebel:** Writing – review & editing, Resources, Data curation. **Tobias Weigel:** Writing – review & editing, Validation. **Grzegorz Kwiatek:** Writing – review & editing, Supervision, Resources, Project administration, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

This work was performed in the frame of EU HORIZON DT-GEO project (HORIZON-INFRA-2021-TECH-01, project number 101058129) and supported by the Helmholtz Association's Initiative and Networking Fund through Helmholtz AI (grant number: ZT-I-PF-5-01). P.M.G. acknowledges funding from the Helmholtz Association in the frame of the Young Investigators Group VH-NG-1232 (SAIDAN) and funding from the European Research Council (ERC) under the European Union Horizon 2020 Research and Innovation Program (grant agreement 101076119 - QUAKEHUNTER).

The authors thank Chris Marone and an anonymous reviewer for evaluating and enhancing the quality of this paper.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.epsl.2023.118383>.

References

An, K., Meng, J., 2010. Voting-averaged combination method for regressor ensemble. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 6215 LNCS, 540–546. https://doi.org/10.1007/978-3-642-14922-1_67/COVER.

- Asencio-Cortés, G., Martínez-Álvarez, F., Troncoso, A., Morales-Esteban, A., 2017. Medium-large earthquake magnitude prediction in Tokyo with artificial neural networks. *Neural Comput. Appl.* 28, 1043–1055. <https://doi.org/10.1007/S00521-015-2121-7/FIGURES/4>.
- Baiesi, M., Paczuski, M., 2005. Complex networks of earthquakes and aftershocks. *Nonlinear Process. Geophys.* 12, 1–11. <https://doi.org/10.5194/NPG-12-1-2005>.
- Ben-Zion, Y., Eneva, M., Liu, Y., 2003. Large earthquake cycles and intermittent criticality on heterogeneous faults due to evolving stress and seismicity. *J. Geophys. Res., Solid Earth* 108, 2307. <https://doi.org/10.1029/2002JB002121>.
- Ben-Zion, Y., Zaliapin, I., 2020. Localization and coalescence of seismicity before large earthquakes. *Geophys. J. Int.* 223, 561–583. <https://doi.org/10.1093/GJI/GGAA315>.
- Bolton, D.C., Marone, C., Shokouhi, P., Rivière, J., Rouet-Leduc, B., Hulbert, C., Johnson, P.A., 2019. Characterizing acoustic signals and searching for precursors during the laboratory seismic cycle using unsupervised machine learning. *Seismol. Res. Lett.* 90, 1088–1098. <https://doi.org/10.1785/0220180367>.
- Bolton, D.C., Shreedharan, S., Rivière, J., Marone, C., 2021. Frequency-magnitude statistics of laboratory foreshocks vary with shear velocity, fault slip rate, and shear stress. *J. Geophys. Res., Solid Earth* 126, e2021JB022175. <https://doi.org/10.1029/2021JB022175>.
- Borate, P., Rivière, J., Marone, C., Mali, A., Kifer, D., Shokouhi, P., 2023. Using a physics-informed neural network and fault zone acoustic monitoring to predict lab earthquakes. *Nat. Commun.* 14, 1–12. <https://doi.org/10.1038/s41467-023-39377-6>.
- Byerlee, J., Brace, W., 1968. Stick slip, stable sliding, and earthquakes—effect of rock type, pressure, strain rate, and stiffness. *J. Geophys. Res.* 73, 6031–6037. <https://doi.org/10.1029/JB073i018P06031>.
- Corbi, F., Bedford, J., Sandri, L., Funicello, F., Gualandi, A., Rosenau, M., 2020. Predicting imminence of analog megathrust earthquakes with machine learning: implications for monitoring subduction zones. *Geophys. Res. Lett.* 47, e2019GL086615. <https://doi.org/10.1029/2019GL086615>.
- Corbi, F., Sandri, L., Bedford, J., Funicello, F., Brizzi, S., Rosenau, M., Lallemand, S., 2019. Machine learning can predict the timing and size of analog earthquakes. *Geophys. Res. Lett.* 46, 1303–1311. <https://doi.org/10.1029/2018GL081251>.
- Dresen, G., Kwiatek, G., Goebel, T., Ben-Zion, Y., 2020. Seismic and aseismic preparatory processes before large stick-slip failure. *Pure Appl. Geophys.* 177, 5741–5760. <https://doi.org/10.1007/S00024-020-02605-X/FIGURES/8>.
- Durand, V., Gualandi, A., Ergintav, S., Kwiatek, G., Haghshenas, M., Motagh, M., Dresen, G., Martínez-Garzón, P., 2022. Deciphering aseismic deformation along submarine fault branches below the eastern Sea of Marmara (Turkey): insights from seismicity, strainmeter, and GNSS data. *Earth Planet. Sci. Lett.* 594, 117702. <https://doi.org/10.1016/j.epsl.2022.117702>.
- Goebel, T.H.W., Becker, T.W., Schorlemmer, D., Stanchits, S., Sammis, C., Rybacki, E., Dresen, G., 2012. Identifying fault heterogeneity through mapping spatial anomalies in acoustic emission statistics. *J. Geophys. Res., Solid Earth* 117, 3310. <https://doi.org/10.1029/2011JB008763>.
- 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, 815–818. <https://doi.org/10.1130/G39147.1>.
- Henderson, J.R., Barton, D.J., Foulger, G.R., 1999. Fractal clustering of induced seismicity in the geysers geothermal area, California. *Geophys. J. Int.* 139, 317–324. <https://doi.org/10.1046/J.1365-246X.1999.00939.X/2/139-2-317-FIG006.JPEG>.
- Hirata, T., Satoh, T., Ito, K., 1987. Fractal structure of spatial distribution of microfracturing in rock. *Geophys. J. Int.* 90, 369–374. <https://doi.org/10.1111/J.1365-246X.1987.TB00732.X>.
- Jasperse, H., Bolton, D.C., Johnson, P., Guyer, R., Marone, C., de Hoop, M.V., 2021. Attention network forecasts time-to-failure in laboratory shear experiments. *J. Geophys. Res., Solid Earth* 126, e2021JB022195. <https://doi.org/10.1029/2021JB022195>.
- Johnson, P.A., Ferdowsi, B., Kaproth, B.M., Scuderi, M., Griffa, M., Carmeliet, J., Guyer, R.A., Le Bas, P.Y., Trugman, D.T., Marone, C., 2013. Acoustic emission and microslip precursors to stick-slip failure in sheared granular material. *Geophys. Res. Lett.* 40, 5627–5631. <https://doi.org/10.1002/2013GL057848>.
- Johnson, P.A., Rouet-Leduc, B., Pyrak-Nolte, L.J., Beroza, G.C., Marone, C.J., Hulbert, C., Howard, A., Singer, P., Gordeev, D., Karaflos, D., Levinson, C.J., Pfeiffer, P., Puk, K.M., Reade, W., 2021. Laboratory earthquake forecasting: a machine learning competition. *Proc. Natl. Acad. Sci.* 118, e2011362118. <https://doi.org/10.1073/PNAS.2011362118>.
- Kagan, Y.Y., Knopoff, L., 1980. Spatial distribution of earthquakes: the two-point correlation function. *Geophys. J. Int.* 62, 303–320. <https://doi.org/10.1111/J.1365-246X.1980.TB04857.X>.
- 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. *Geophys. Res. Lett.* 41, 5838–5846. <https://doi.org/10.1002/2014GL060159>.
- Kwiatek, G., Martínez-Garzón, P., Davidsen, J., Malin, P., Karjalainen, A., Bohnhoff, M., Dresen, G., 2022. Limited earthquake interaction during a geothermal hydraulic stimulation in helsinki, Finland. *J. Geophys. Res., Solid Earth* 127, e2022JB024354. <https://doi.org/10.1029/2022JB024354>.

- Kwiatkiewicz, Grzegorz, Martínez-Garzón, Patricia, Goebel, Thomas, et al., 2023. Complex multi-scale preparatory processes of stick-slip events on rough laboratory faults. *ESS Open Archive*. September 11. <https://doi.org/10.22541/essoar.169447455.58529925/v1>.
- Laurenti, L., Tinti, E., Galasso, F., Franco, L., Marone, C., 2022. Deep learning for laboratory earthquake prediction and autoregressive forecasting of fault zone stress. *Earth Planet. Sci. Lett.* 598, 117825. <https://doi.org/10.1016/j.epsl.2022.117825>.
- Lei, X., 2006. Typical phases of pre-failure damage in granitic rocks under differential compression. *Geol. Soc. Spec. Publ.* 261, 11–29. <https://doi.org/10.1144/GSL.SP.2006.261.01.02>.
- Lei, X., Ma, S., 2014. Laboratory acoustic emission study for earthquake generation process. *Earthq. Sci.* 27, 627–646. <https://doi.org/10.1007/S11589-014-0103-Y/FIGURES/9>.
- 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. *Geophys. Res. Lett.* 45, 13,269–13,276. <https://doi.org/10.1029/2018GL079712>.
- Main, I.G., 1991. A modified Griffith criterion for the evolution of damage with a fractal distribution of crack lengths: application to seismic event rates and b-values. *Geophys. J. Int.* 107, 353–362. <https://doi.org/10.1111/j.1365-246X.1991.tb00830.x>.
- Marone, C., 1998. Laboratory-derived friction laws and their application to seismic faulting. *Annu. Rev. Earth Planet. Sci.* 26, 643–696. <https://doi.org/10.1146/ANNUREV.EARTH.26.1.643>.
- Martínez-Garzón, P., Ben-Zion, Y., Zaliapin, I., Bohnhoff, M., 2019. Seismic clustering in the Sea of Marmara: implications for monitoring earthquake processes. *Tectonophysics* 768, 228176. <https://doi.org/10.1016/j.tecto.2019.228176>.
- Martínez-Garzón, P., Beroza, G.C., Bocchini, G.M., Bohnhoff, M., 2023. Sea level changes affect seismicity rates in a hydrothermal system near Istanbul. *Geophys. Res. Lett.* 50, e2022GL101258. <https://doi.org/10.1029/2022GL101258>.
- Martínez-Garzón, P., Durand, V., Bentz, S., Kwiatkiewicz, G., Dresen, G., Turkmen, T., Nurlu, M., Bohnhoff, M., 2021. Near-fault monitoring reveals combined seismic and slow activation of a fault branch within the Istanbul–Marmara seismic gap in Northwest Turkey. *Seismol. Res. Lett.* 92, 3743–3756. <https://doi.org/10.1785/0220210047>.
- Martínez-Garzón, P., Kwiatkiewicz, G., Ickrath, M., Bohnhoff, M., 2014. MSATSI: a MATLAB package for stress inversion combining solid classic methodology, a new simplified user-handling, and a visualization tool. *Seismol. Res. Lett.* 85, 896–904. <https://doi.org/10.1785/0220130189>.
- McBeck, J.A., Aiken, J.M., Mathiesen, J., Ben-Zion, Y., Renard, F., 2020. Deformation precursors to catastrophic failure in rocks. *Geophys. Res. Lett.* 47, e2020GL090255. <https://doi.org/10.1029/2020GL090255>.
- McBrearty, I.W., Delorey, A.A., Johnson, P.A., 2019. Pairwise association of seismic arrivals with convolutional neural networks. *Seismol. Res. Lett.* 90, 503–509. <https://doi.org/10.1785/0220180326>.
- Mignan, A., 2014. The debate on the prognostic value of earthquake foreshocks: a meta-analysis. *Sci. Rep.* 4, 1–5. <https://doi.org/10.1038/srep04099>.
- Montgomery, D.C., Peck, E.A., 1992. *Introduction to Linear Regression Analysis*. John Wiley & Sons.
- Mousavi, S.M., Beroza, G.C., 2023. Machine learning in earthquake seismology. <https://doi.org/10.1146/annurev-earth-071822-100323>. 51.
- Mousavi, S.M., Beroza, G.C., 2020. Bayesian-deep-learning estimation of earthquake location from single-station observations. *IEEE Trans. Geosci. Remote Sens.* 58, 8211–8224. <https://doi.org/10.1109/TGRS.2020.2988770>.
- Mousavi, S.M., Ellsworth, W.L., Zhu, W., Chuang, L.Y., Beroza, G.C., 2020. Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nat. Commun.* 11, 1–12. <https://doi.org/10.1038/s41467-020-17591-w>.
- Mousavi, S.M., Zhu, W., Sheng, Y., Beroza, G.C., 2019. CRED: a deep residual network of convolutional and recurrent units for earthquake signal detection. *Sci. Rep.* 9, 1–14. <https://doi.org/10.1038/s41598-019-45748-1>.
- Niemeijer, A., Marone, C., Elsworth, D., 2010. Frictional strength and strain weakening in simulated fault gouge: competition between geometrical weakening and chemical strengthening. *J. Geophys. Res.* 115. <https://doi.org/10.1029/2009JB000838>.
- Panakkat, A., Adeli, H., 2007. Neural network models for earthquake magnitude prediction using multiple seismicity indicators. *Int. J. Neural Syst.* 17, 13–33. <https://doi.org/10.1142/S0129065707000890>.
- Park, Y., Mousavi, S.M., Zhu, W., Ellsworth, W.L., Beroza, G.C., 2020. Machine-learning-based analysis of the guy-greenbrier, Arkansas earthquakes: a tale of two sequences. *Geophys. Res. Lett.* 47, e2020GL087032. <https://doi.org/10.1029/2020GL087032>.
- Picozzi, M., Iaccarino, A.G., 2021. Forecasting the preparatory phase of induced earthquakes by recurrent neural network. In: *Forecasting 2021*, vol. 3, pp. 17–36.
- Picozzi, M., Iaccarino, A.G., Spallarossa, D., Bindi, D., 2023. On catching the preparatory phase of damaging earthquakes: an example from central Italy. <https://doi.org/10.21203/RS.3.RS-2957583/V1>.
- Pierezan, J., Dos Santos Coelho, L., 2018. Coyote optimization algorithm: a new metaheuristic for global optimization problems. In: *2018 IEEE Congress on Evolutionary Computation, CEC 2018 - Proceedings*.
- Pu, Y., Chen, J., Apel, D.B., 2021. Deep and confident prediction for a laboratory earthquake. *Neural Comput. Appl.* 33, 11691–11701. <https://doi.org/10.1007/S00521-021-05872-4/FIGURES/9>.
- Ren, C.X., Dorostkar, O., Rouet-Leduc, B., Hulbert, C., Strebel, D., Guyer, R.A., Johnson, P.A., Carmeliet, J., 2019. Machine learning reveals the state of intermittent frictional dynamics in a sheared granular fault. *Geophys. Res. Lett.* 46, 7395–7403. <https://doi.org/10.1029/2019GL082706>.
- Ren, C.X., Hulbert, C., Johnson, P.A., Rouet-Leduc, B., 2020. Machine learning and fault rupture: a review. *Adv. Geophys.* 61, 57–107. <https://doi.org/10.1016/BS.AGPH.2020.08.003>.
- Reyes, J., Morales-Esteban, A., Martínez-Álvarez, F., 2013. Neural networks to predict earthquakes in Chile. *Appl. Soft Comput.* 13, 1314–1328. <https://doi.org/10.1016/J.ASOC.2012.10.014>.
- 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 Planet. Sci. Lett.* 482, 407–413. <https://doi.org/10.1016/j.epsl.2017.11.036>.
- Ross, Z.E., Meier, M.A., Hauksson, E., Heaton, T.H., 2018. Generalized seismic phase detection with deep learning. *Bull. Seismol. Soc. Am.* 108, 2894–2901. <https://doi.org/10.1785/0120180080>.
- Rouet-Leduc, B., Hulbert, C., Bolton, D.C., Ren, C.X., Riviere, J., Marone, C., Guyer, R.A., Johnson, P.A., 2018. Estimating fault friction from seismic signals in the laboratory. *Geophys. Res. Lett.* 45, 1321–1329. <https://doi.org/10.1002/2017GL076708>.
- Rouet-Leduc, B., Hulbert, C., Lubbers, N., Barros, K., Humphreys, C.J., Johnson, P.A., 2017. Machine learning predicts laboratory earthquakes. *Geophys. Res. Lett.* 44, 9276–9282. <https://doi.org/10.1002/2017GL074677>.
- Sadovskiy, M.A., 1984. Characteristic dimensions of rock and hierarchical properties of seismicity. *Izv. Earth Phys.* 20, 87–96. https://doi.org/10.2208/JSCE.1998.596_49.
- Saenger, E.H., Finger, C., Karimpouli, S., Tahmasebi, P., 2021. Single-station coda wave interferometry: a feasibility study using machine learning. In: *Materials 2021*, Vol. 14, p. 3451.
- Sano, O., Ito, I., Terada, M., 1981. Influence of strain rate on dilatancy and strength of Oshima granite under uniaxial compression. *J. Geophys. Res., Solid Earth* 86, 9299–9311. <https://doi.org/10.1029/JB086iB10P09299>.
- Shreedharan, S., Bolton, D.C., Rivière, J., Marone, C., 2021. Machine learning predicts the timing and shear stress evolution of lab earthquakes using active seismic monitoring of fault zone processes. *J. Geophys. Res., Solid Earth* 126, e2020JB021588. <https://doi.org/10.1029/2020JB021588>.
- Sundararajan, M., Taly, A., Yan, Q., 2017. Axiomatic attribution for deep networks. In: *The 34th International Conference on Machine Learning*. PMLR, pp. 3319–3328.
- Van Der Elst, N.J., Brodsky, E.E., 2010. Connecting near-field and far-field earthquake triggering to dynamic strain. *J. Geophys. Res., Solid Earth* 115, 7311. <https://doi.org/10.1029/2009JB006681>.
- Vavryčuk, V., 2014. Iterative joint inversion for stress and fault orientations from focal mechanisms. *Geophys. J. Int.* 199, 69–77. <https://doi.org/10.1093/GJI/GGU224>.
- Wang, K., Johnson, C.W., Bennett, K.C., Johnson, P.A., 2022. Predicting future laboratory fault friction through deep learning transformer models. *Geophys. Res. Lett.* 49, e2022GL098233. <https://doi.org/10.1029/2022GL098233>.
- Wang, K., Johnson, C.W., Bennett, K.C., Johnson, P.A., 2021. Predicting fault slip via transfer learning. *Nat. Commun.* 2021, 1–11. <https://doi.org/10.1038/s41467-021-27553-5>.
- Wiemer, S., Wyss, M., 2000. Minimum magnitude of completeness in earthquake catalogs: examples from Alaska, the Western United States, and Japan. *Bull. Seismol. Soc. Am.* 90, 859–869. <https://doi.org/10.1785/0119990114>.
- Zaliapin, I., Ben-Zion, Y., 2013. Earthquake clusters in southern California I: identification and stability. *J. Geophys. Res., Solid Earth* 118, 2847–2864. <https://doi.org/10.1002/JGRB.50179>.
- Zaliapin, I., Gabriellov, A., Keilis-Borok, V., Wong, H., 2008. Clustering analysis of seismicity and aftershock identification. *Phys. Rev. Lett.* 101, 018501. <https://doi.org/10.1103/PHYSREVLETT.101.018501/FIGURES/5/MEDIUM>.
- Zhu, L., Peng, Z., McClellan, J., Li, C., Yao, D., Li, Z., Fang, L., 2019. Deep learning for seismic phase detection and picking in the aftershock zone of 2008 Mw7.9 Wenchuan Earthquake. *Phys. Earth Planet. Inter.* 293, 106261. <https://doi.org/10.1016/j.pepi.2019.05.004>.