JGR | Machine Learning and Computation

RESEARCH ARTICLE
10.1029/2024JH000398

Machine Learning Predicts the Slip Duration and Friction Drop of Laboratory Earthquakes in Sheared Granular Fault

Mengfan Wei[1] and **Ke Gao**[1,2]

[1]Department of Earth and Space Sciences, Southern University of Science and Technology, Shenzhen, China, [2]Guangdong Provincial Key Laboratory of Geophysical High-Resolution Imaging Technology, Southern University of Science and Technology, Shenzhen, China

Key Points:

- The prediction of slip duration and friction drop of laboratory earthquakes are explored using ensemble learning
- The fourth-order moment of the plate motion signals during the initial slip stage is among the best estimators of slip duration
- The second moment of the plate motion signals during the entire slip stage is the best predictive feature for friction drop

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

K. Gao,
gaok@sustech.edu.cn

Citation:

Wei, M., & Gao, K. (2024). Machine learning predicts the slip duration and friction drop of laboratory earthquakes in sheared granular fault. *Journal of Geophysical Research: Machine Learning and Computation*, *1*, e2024JH000398. https://doi.org/10.1029/2024JH000398

Received 14 AUG 2024
Accepted 27 SEP 2024

Abstract Predicting laboratory earthquakes using machine learning has progressed markedly recently. Previous related studies mainly focus on predicting the occurrence time and shear stress of laboratory earthquakes using acoustic emission signals. Here, based on numerical simulations, we use machine learning to show that statistical features of plate motion signals contain information about the slip duration and friction drop of laboratory earthquakes. We find that the plate motion signals during the initial slip stage contain the precursor information about the slip duration of laboratory earthquakes. However, to accurately predict the friction drop, we need to incorporate the plate motion signals during the entire slip stage. The results demonstrate that the high-order moment and variance of plate motion signals are respectively among the best predictors for the slip duration and friction drop of laboratory earthquakes. Our work provides new insights for future investigations into natural earthquake prediction through machine learning.

Plain Language Summary Earthquake prediction is an important but challenging task. The booming of machine learning brings new hope for earthquake prediction. However, the direct utilization of machine learning for earthquake prediction is still tricky, simply because the available data our human beings currently have can only cover a limited number of earthquake cycles. In this study, we use numerical simulations to represent earthquakes as frictional slips and record hundreds of simulated laboratory earthquake cycles alongside their corresponding fault motion data. We then train the motion data using machine learning to predict the slip duration and friction drop of upcoming laboratory earthquakes. The results show that the slow pre-seismic slip stage, similar to that observed in natural earthquakes, has the potential to indicate seismic nucleation. Particularly, we find that the fault motion signals during the initial slip stage contain the precursor information about the slip duration of laboratory earthquakes. Our findings may contribute to predicting the slip duration and friction drop of natural earthquakes.

1. Introduction

A slow and accelerated slip phase of weeks (Ruiz et al., 2017) to months (Socquet et al., 2017) has been observed before large earthquakes, and seismic swarms triggered at the slip process may be indicative of nucleation process (Kato et al., 2012; Ruiz et al., 2014). Similar slip phases prior to laboratory earthquakes (fault stick-slip failure) have also been observed in laboratory experiments (Guérin-Marthe et al., 2019). Due to its ease of observation, short nucleation periods and reproducibility, laboratory earthquake research can compensate for the restrictions in studying natural earthquakes. Exploring the correlation between slow and accelerated slip of faults and subsequent laboratory earthquake events can improve our understanding of natural fault nucleation processes and provide new insights into earthquake prediction and early warnings.

Compared with most laboratory experiments, sheared gauge fault tests based on numerical simulation methods can provide similar b values (Dahmen et al., 2011; Rivière et al., 2018), more microscale information (Dorostkar et al., 2017), and it is easier to obtain detailed stress evolution and motion inside/near the fault system (Brzinski Iii & Daniels, 2018). In addition, the statistical distribution of velocity fluctuations in sheared granular gouges has been shown to be consistent with observations of interseismic plate motion fluctuations (Meroz & Meade, 2017). This supports that the sheared granular system, based on numerical simulation, can be used to study earthquakes, and the statistical distribution of particle motion fluctuations may be related to the characteristic duration of earthquakes and the frictional state of plate boundaries (Ren et al., 2019).

Due to the heterogeneity and anisotropy of faults, traditional seismic data analysis methods are becoming increasingly difficult to meet research needs. The recent applications of machine learning methods in seismology-related fields have demonstrated machine learning's superior capability over traditional methods in excavating deep correlations between variables from massive data and fitting strong nonlinear mapping relationships between variables in high dimensions. Mousavi and Beroza (2020) employed a deep learning regression algorithm to predict earthquake magnitudes based on raw waveforms from a single station and found that the model can directly learn distance-dependent and location-dependent functions from the training data. Rouet-Leduc et al. (2017) used a random forest algorithm to predict the timing of shear instability based on acoustic emission signals collected from biaxial shear experiments, demonstrating that machine learning algorithms can identify previously unrecognized signals. Corbi et al. (2019) applied the gradient boosting tree algorithm to study data from multiple earthquake cycles in a laboratory-scale subduction zone, showing that by reconstructing and correctly interpreting the system's spatiotemporal complex loading history, it is possible to accurately predict the timing and magnitude of laboratory earthquakes. Wang et al. (2021) were the first to apply transfer learning to transfer the potential fault stick-slip patterns learned from simulated data to the prediction of fault frictional states in physical experiments, improving the generalization ability of machine learning models for laboratory earthquake prediction to a certain extent.

Here, we utilize the Light Gradient Boosting Machine (LightGBM) to investigate whether fault plate motion signals contain information about the slip duration and friction drop of upcoming laboratory earthquakes in a shear granular fault simulated using the combined finite-discrete element method (FDEM). The study shows that the plate motion signals at the initial slip stage contain the precursor information about the slip duration of laboratory earthquakes, while predicting friction drop needs to incorporate the whole plate motion signals during the entire slip stage. Additionally, the high-order moment and variance of the fault motion signals are among the best predictors of the slip duration and friction drop of laboratory earthquakes, respectively. This work deepens our understanding of the seismogenic mechanism of laboratory earthquakes and provides new insights for the future study of machine learning in earthquake prediction.

2. Materials and Methods

2.1. FDEM and Model Setup

The combined finite-discrete element method (FDEM) was initially developed by Munjiza (1992), which combines the algorithmic strengths of both the finite element method (FEM) and discrete element method (DEM) and allows the simulation of material behaviors transitioning from a continuous to a discontinuous state. The core principles of FDEM involve governing equations, deformation description, contact detection, and contact interaction (Lei et al., 2016; Munjiza et al., 2006). The governing equation of FDEM is used to solve the dynamic response of solid materials under external forces, which can be expressed as

$$\mathbf{M}\ddot{\mathbf{x}} + \mathbf{C}\dot{\mathbf{x}} = \mathbf{f}, \tag{1}$$

where $\mathbf{M}$ represents the lumped mass matrix, $\mathbf{x}$ denotes the displacement vector, $\mathbf{C}$ is the damping matrix, and $\mathbf{f}$ is the equivalent force vector acting on each FEM node (Munjiza, 2004). To calculate the temporal evolution of Equation 1, an explicit time integration scheme based on a central difference method is employed to capture the transient behavior of the system. The deformation of finite elements is described using a formula based on multiplicative decomposition (Munjiza et al., 2014). The NBS (Non-Binary Search) algorithm is used to detect contact between discrete elements (Munjiza & Andrews, 1998), and all possible contact pairs are recorded. Finally, the contact force for each contact pair is calculated using a penalty-based method (Munjiza, 2004; Munjiza et al., 2011).

FDEM can simulate the linear elasticity within the shear plate and the particles composing the fault gouge through the FEM module. The interactions between the fault gouge particles and between the fault gouge and the shear plate are handled using the DEM method. Compared to numerical simulation methods based purely on continuous media, such as FEM, the boundary element method, and the finite difference method, FDEM provides a more detailed depiction of the deformation and motion of fault gouge particles and the shear plate during the shearing process. It can simulate the dynamic evolution of deformation and rupture in the shear plate and fault gouge and obtain information on the stress and strain fields within these materials. Additionally, FDEM incorporates

10.1029/2024JH000398

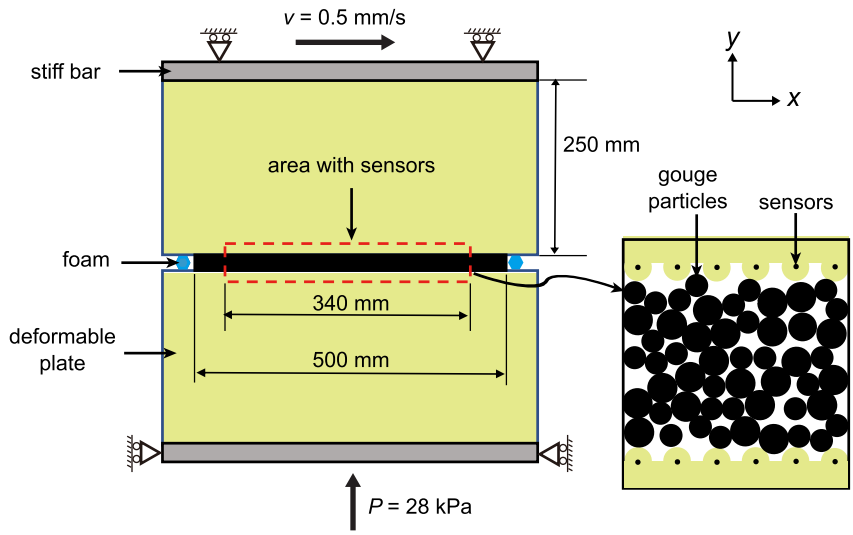

Figure 1. Model setup. The gouge is sandwiched by two identical deformable plates, with foam-like particles (blue color) on its left and right sides to restrict particles from escaping. Each particle is further meshed into finite elements to capture its deformation, and there are 2,817 particles in the gouge. A total of 143 "sensors" are placed on the top and bottom plates, respectively, in places immediately adjacent to the gouge to track the plate motion during stick-slip cycles.

multiple contact models, effectively capturing the complex contact behaviors between fault gouge particles and between the fault gouge and the shear plate during fault shearing (Gao et al., 2018).

Here, we use FDEM to simulate a 2D sheared fault system with granular gouges (Gao et al., 2018) built based on a photoelastic sheared laboratory experiment (Geller et al., 2015). The geometry of the model is shown in Figure 1, and the parameters of the material and numerical simulation are shown in Table 1. The finite element method (FEM) module in FDEM is used to simulate deformations in the shear plates and gouge particles, and the interactions among gouge particles and between the shear plate and particles are processed by the DEM module. Compared to the laboratory earthquakes traditionally simulated using DEM, FDEM can capture more detailed deformation and motion in the particles and plates during the stick-slip cycles (Dratt & Katterfeld, 2017; Ma et al., 2016) (see Text S1 in Supporting Information S1 for more details).

The model consists of 2,817 circular gouge particles randomly placed between two identical deformable plates, with foam-like particles (blue color) on their left and right sides to restrict particles from escaping. In the FDEM model, three-node constant-strain triangular elements are used for the particles and the plates. A total of ~90,000 triangle elements are generated in the simulation domain. Each particle is composed of 24 elements of approximately equal size. This number of elements allows each particle to maintain a nearly circular shape after meshing and is sufficient to precisely capture the particle deformation while assuring the model is not too computationally expensive. Bi-disperse diameters (1.2 or 1.6 mm) for these particles are adopted to avoid crystallization (Tsai et al., 2003). The normal stress P and shear velocity v are applied through two stiff bars under the lower plate and above the upper plate, respectively. At the interface between the plate and particles is a semicircular row of "teeth" with a total of 286 sensors (143 on each plate) in the center to track the motion of the two plates. Specifically, the displacements and velocities at these sensor points are recorded with a frequency of every 1 ms.

First, the top and bottom stiff bars move vertically toward each other to consolidate the gouge particles and ensure they are well contacted. Then, a constant horizontal shearing velocity $v = 0.5$ mm/s, corresponding to a strain rate of $\dot{\varepsilon} = 0.001\ \mathrm{s}^{-1}$, is applied on the top stiff bar to realize the shear, while a normal stress $P = 28$ kPa is maintained on the bottom stiff bar throughout the simulation. During the shearing, the top plate can only move in the x direction, while the bottom plate can only move in the y direction. The inertial number is $I = \frac{\dot{\varepsilon} d}{\sqrt{P/\rho}} \approx 9.0 \times 10^{-6} < 10^{-3}$ (d is the average diameter of particles, and $\rho = 1{,}150\ \mathrm{kg/m}^3$ is the particle density), thus ensuring a quasi-static shearing state. The model is run for roughly 3.0×10^8 time steps, and the total shearing time is approximately 30 s, giving an overall shear displacement of roughly 20 mm (including the pre–steady state). The tangential contact forces between the plate and each gouge particle are summed, and this resultant

Table 1
Material and Numerical Simulation Parameters

Parameter	Value
Particle diameter	1.2 or 1.6 mm
Particle density	1,150 kg/m^3
Particle Young's modulus	0.4 GPa
Particle Poisson's ratio	0.4
Particle-particle friction coefficient	0.15
Number of Particles	2,817
Main plate density	1,150 kg/m^3
Main plate Young's modulus	2.5 MPa
Main plate Poisson's ratio	0.49
Particle-plate friction coefficient	0.15
Stiff bar density	2,800 kg/m^3
Stiff bar Young's modulus	30 GPa
Stiff bar Poisson's ratio	0.33
Foam density	1,150 kg/m^3
Foam Young's modulus	1 MPa
Foam Poisson's ratio	0.4
Contact penalty	4 GPa
Timestep	10^{-7} s
Normal stress P	28 kPa
Shear velocity v	0.5 mm/s

force is then divided by the length of the gouge-plate boundary to obtain the shear stress between the deformable plates and the gouge. The shear stress values from the top and bottom plates are averaged, yielding the shear stress used in our analysis. The normal stress is obtained in a similar manner. The displacement and velocity in the x and y directions of each sensor are monitored during the shearing process. We select the data that the model reached a stable state after the first 5 s for further analyses (see Text S2 in Supporting Information S1 for more details).

The evolution of normalized shear stress, defined as the ratio of shear stress to the applied normal stress, is shown in Figure 2a. It can be observed that the normalized shear stress exhibits more irregular stick-slip behavior compared to the previous double direct shear experiments (Rouet-Leduc et al., 2017). We suspect this is because most fault shear stick-slip experiments are conducted at relatively small scales (~100 mm). Once a slip occurs, it spans the entire fault, leading to relatively constant slip event durations and drops in stress. Our FDEM model is based on the experimental setup of Geller et al. (2015), which simulates relatively large-scale stick-slip experiments. In this setup, the plates use photoelastic materials with stiffness in the MPa range, which is significantly lower than the stiffness of rocks (in the GPa range) typically used in small-scale experiments. Consequently, numerous local slips occur during the shear process, leading to a broader range of variation in the duration and stress drops of slip events. This results in more irregular stick-slip behavior. These irregular stick-slips are likely more representative of natural fault observations and pose a significant challenge to accurately building predictive machine learning models. To eliminate minor rearrangement of particles presented as background fluctuations in the normalized shear stress signal, we use the Wiener filter, an optimal linear filter, to denoise the signal. The Wiener filter can be implemented using the Wiener function from the Scipy library in Python. The disparity between the raw data and the denoised data is illustrated in Figure 2b, where the filter has removed numerous small and frequent noises but still maintains the detailed evolution of stick-slips.

Slip events are extracted by setting a threshold for both the slip duration and friction drop (i.e., the normalized shear stress drop). The definition of slip duration and friction drop is sketched in Figure 2c on the basis of a typical slip event. Only slip events with durations greater than 4 ms and friction drops greater than 5×10^{-5} are considered. In general, in the initial stage of a slip event, the normalized shear stress gradually declines while the plate accelerates until reaching a critical slip distance. The seismic nucleation model proposed based on rock friction experiments (Ohnaka, 1996) and numerical simulation (Dieterich, 1992) posits that localized pre-slip occurs along the fault prior to an earthquake, resulting in a decrease in stress. Nucleation initiates and propagates along the fault plane until it reaches the critical nucleation size. Subsequently, dynamic fault slips (i.e., earthquake rupture) occur. The gradual reduction in normalized shear stress observed in Figure 2c indicates quasi-static nucleation. The proportion of the gradual decline stage in a whole slip varies (see Figure S2 in Supporting Information S1). A total of 296 slip events are extracted during the shearing of the granular gouge.

To quantify the overall plate motion during the stick-slip cycle, the velocities and displacements at the 143 sensor points in the x and y directions on each plate are averaged (see Figures 2d and 2e), and we denote them as the macroscopic plate velocity and displacement herein. When the normalized shear stress suddenly decreases (indicating a slip event), there is a sudden increase in the magnitude of displacement and velocity in the shear direction (Figure S3 in Supporting Information S1). Note that we calculate the x displacement of the top plate (D_{xt}) by subtracting its background shearing displacement. Since the velocities and displacements of the top and bottom plates are nearly symmetric (Figure S4 in Supporting Information S1), we only consider data obtained from the top plate in the following analyses.

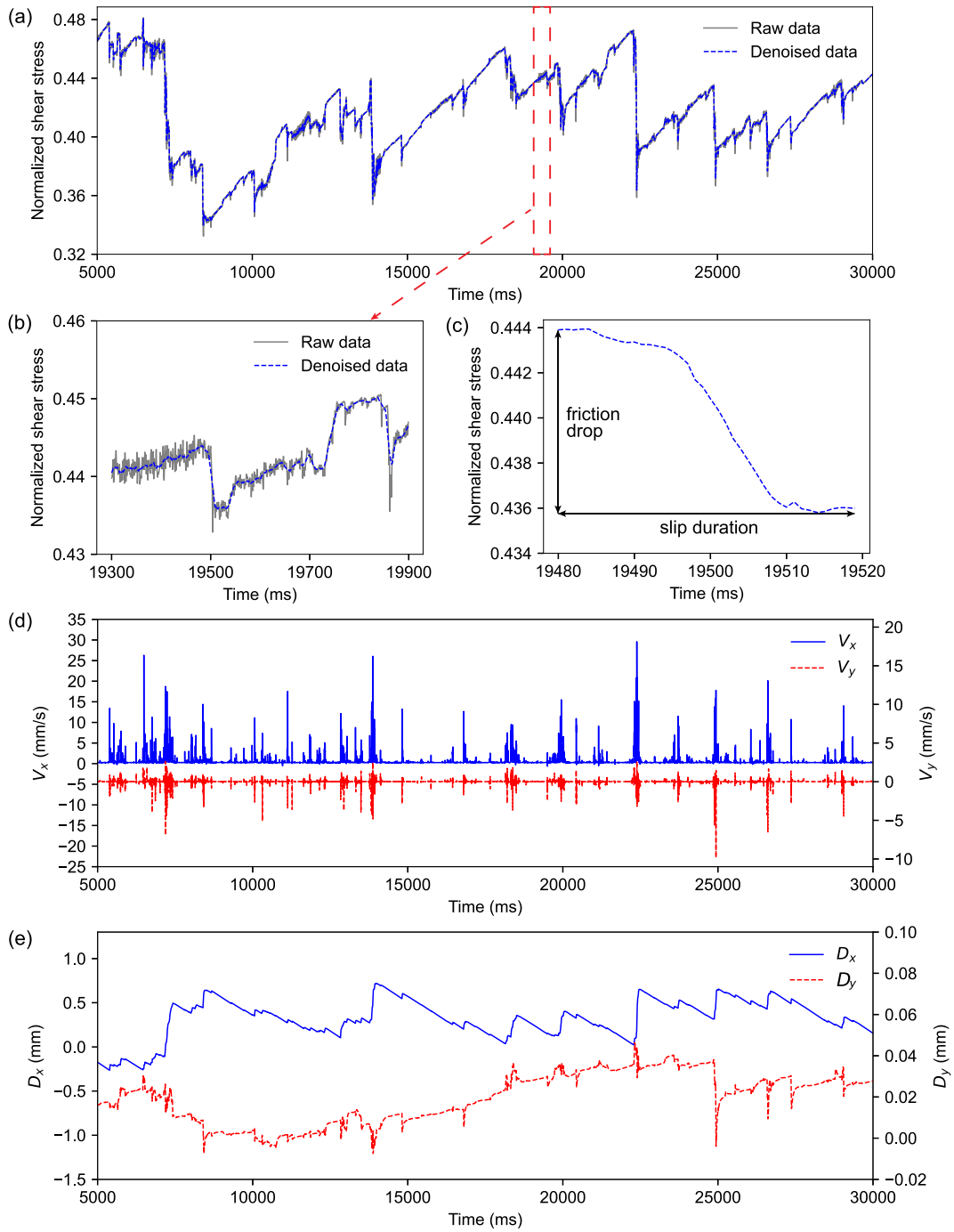

Figure 2. The simulation results. (a) The normalized shear stress as a function of shearing time. The raw data (gray curve) and the denoised data (blue curve). (b) A close-up view of the evolution of normalized shear stress and the difference between the raw and the denoised data. (c) A close-up view of a typical slip event, and the definition of friction drop and slip duration. (d) The *x*- and *y*-velocities of the top plate. (e) The *x*- and *y*-displacements of the top plate.

2.2. Light Gradient Boosting Machine (LightGBM)

We use the LightGBM (Ke et al., 2017), an implementation of the Gradient Boosting Decision Tree (GBDT) algorithm (Friedman, 2002), to establish the quantitative relation between plate motion and the friction drop and slip duration of laboratory earthquakes. GBDT algorithms make predictions by "integrating" a set of simple decision trees. Although the prediction of an individual decision tree may be inaccurate, the errors will be

canceled between trees, and the integrated prediction can reach remarkable accuracy. Because of its strong generalization and good interpretability, GBDT algorithm is an extensively used ensemble learning method.

Compared to other common GBDT algorithms, LightGBM is significantly faster and delivers almost equivalent performance. LightGBM splits the tree leaf-wise with the best fit, unlike other boosting algorithms in which depth-wise or level-wise approaches are used. The leaf-wise algorithm can produce much more complex trees, reduce more loss than the level-wise algorithm, and hence result in much better accuracy that is not easily achievable by other boosting algorithms, such as the XGBoost (Chen et al., 2015). Besides, LightGBM uses a histogram-based algorithm that buckets continuous feature values into discrete bins to speed up the training procedure. The possibility of overfitting can be reduced by setting the max depth parameter. Additionally, LightGBM uses a novel Gradient-based One-Side Sampling (GOSS) technique to filter out the data instances to find appropriate split values. In contrast, XGBoost uses a pre-sorted algorithm to compute the best split. In general, the histogram-based split algorithm, where the bins of all data points are used to find the split values of the histogram, is more efficient than the pre-sorted algorithms where all possible split points on the pre-sorted feature values are enumerated.

In each iteration of LightGBM, the negative gradients of the loss function concerning the output of the model are denoted as g_i, which can be illustrated as

$$\begin{aligned} g_i &= \partial_{\hat{y}^{(t-1)}} \text{loss}\left(y_i, \hat{y}^{(t-1)}\right) \\ h_i &= \partial^2_{\hat{y}^{(t-1)}} \text{loss}\left(y_i, \hat{y}^{(t-1)}\right) \end{aligned}. \tag{2}$$

The decision tree model splits each node with the largest information gain for each feature. With the GOSS method, the information gain measured by the variance of splitting features j at point d for node n can be defined as

$$V_{j|O}(d) = \frac{1}{n}\left(\frac{\left(\sum_{\{x_i \in A_l : x_{ij} \le d\}} g_i + \frac{1-a}{b}\sum_{xi \in B_l} g_i\right)^2}{n_l^j(d)} + \frac{\left(\sum_{\{x_i \in A_r : x_{ij} \le d\}} g_i + \frac{1-a}{b}\sum_{xi \in B_r} g_i\right)^2}{n_r^j(d)}\right). \tag{3}$$

The GOSS method keeps the top $a \times 100\%$ instances with larger gradients and yields an instance subset A. After that, the remaining set A^c consisting of $(1-a) \times 100\%$ instances have smaller gradients. GOSS method randomly samples a subset B with size $b \times A^c$. Thus, the method estimates information gain over a much smaller sample subset, in contrast to other methods such as pre-sort sampling used in XGBoost (Chen et al., 2015) that calculate accurate information gain through all samples. As a result, the computation cost of LightGBM can be significantly reduced. Most importantly, in contrast to deep learning approaches, LightGBM is explicit and can track the contribution of each input feature to the prediction, thus helping us effectively extract the physical information about the gouge shearing system.

We divide the 296 recorded slip events into the training set (the first 60%, comprising 177 slip events) and testing set (the remaining 40%, a total of 119 slip events) according to the shearing time. We use three-fold cross-validation to train the machine learning model (i.e., tuning hyperparameters) based on the training data set. The Hyperopt python package (Bergstra et al., 2013) is employed to automatically tune the hyperparameters of LightGBM. Hyperopt uses an algorithm called Tree-based Parzen Esimtors, which efficiently searches through complex hyperspace to reach optimums by simply defining the functional form and bounds of each hyper-parameter. Once the optimal set of hyperparameters is determined, the test set is used to evaluate the prediction performance of the tuned LightGBM model. We use the coefficient of determination to measure the performance of the model.

2.3. Shapely Additive Explanation (SHAP)

After reaching a decent LightGBM prediction model, we utilize the Shapely Additive Explanation (SHAP), a post-hoc interpretation method based on game theory (Lundberg & Lee, 2017), to evaluate the relative importance of each input feature. The SHAP value for each feature is the average marginal contribution of the feature across

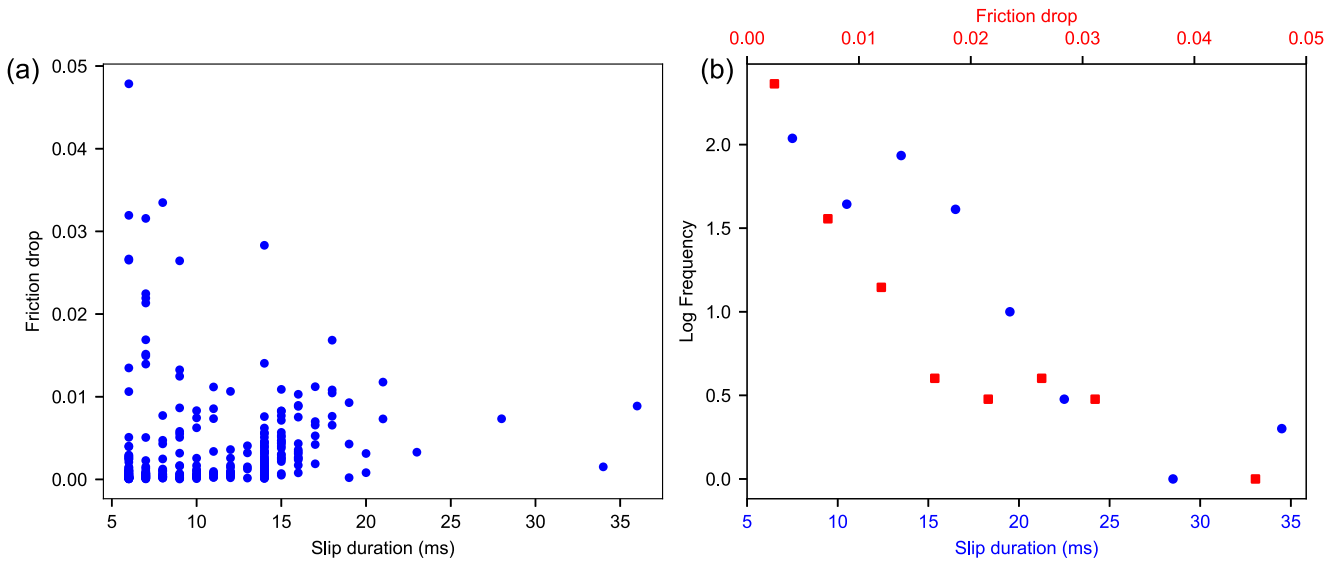

Figure 3. (a) Friction drop versus slip duration for all slip events; (b) Logarithmic frequency of friction drop and slip duration of all slip events.

all possible coalitions of input features, representing their contribution toward a higher or lower final prediction. Compared to other post-hoc interpretability methods, such as LIME (Ribeiro et al., 2016), which primarily focuses on local explanations, the SHAP method provides both global and local explanations. Additionally, unlike feature importance methods that consider only the importance of individual features, SHAP accounts for feature interactions, offering a more detailed and accurate assessment of feature importance.

The calculation of SHAP values is based on Shapley values, also known as expected marginal contributions. The SHAP method represents the influence of each feature on the model's output through an additive approach. For a single sample ($\mathbf{x}$, y), where $\mathbf{x}$ represents the feature vector and y is the label, the post-hoc interpretation model g can be expressed as

$$g(\mathbf{x}) = \phi_0 + \sum_{i=1}^{n} \phi_i. \tag{4}$$

Here, ϕ_0 is the average value of the machine learning model f over all samples; n is the number of features in f; ϕ_i is the Shapley value of the ith feature x_i, representing the expected marginal contribution.

The post-hoc interpretability model and the machine learning model yield the same prediction value for a single sample. Therefore, we have

$$f(\mathbf{x}) = \phi_0 + \sum_{i=1}^{M} \phi_i. \tag{5}$$

It can be seen that the prediction result $f(\mathbf{x})$ of the machine learning model can be expressed as the sum of the Shapley values of all features. The SHAP value for each feature can be calculated for each sample. Features with positive SHAP values contribute to the prediction in a positive manner, and vice versa, and the magnitude of a SHAP value indicates the strength of the contribution of a feature. By averaging the SHAP values of a feature across all samples, the importance of that feature can be measured.

3. Results

The friction drop of these slip events against slip duration is shown in Figure 3a. It is found that, in general, the longer the duration of a slip event, the lower the maximum friction drop. Events with short slip durations and large friction drops are considered as fast slips, while those with long slip durations and small friction drops can be treated as slow slips. Events with short slip duration and small friction drop are presumed to resemble the small

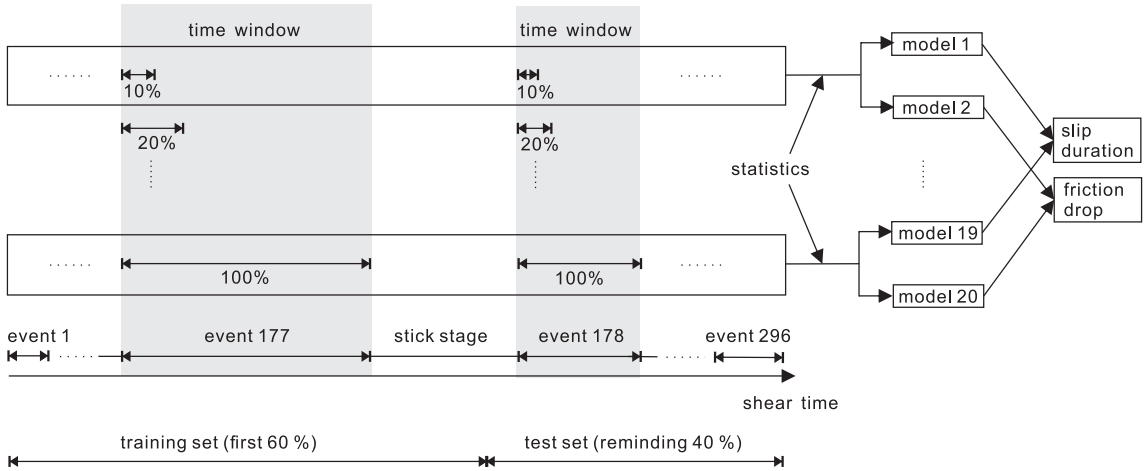

Figure 4. Sketch of the procedure through which the machine learning algorithm derives the slip duration and friction drop of slip events from the plate's motion signal.

transient slip (Frank, 2016), which is expected to occur if the magnitude-frequency distribution of slow slips follows a power law distribution (Wech et al., 2010). The evolution of the displacement in the x direction of the deformable plate for different types of events is shown in Figure S5 in Supporting Information S1. An earthquake-like magnitude-frequency distribution was observed in slow-slip events recorded in the north of Cascadia (Wech et al., 2010), and there is roughly a similar trend in our data (Figure 3b), that is, the greater the friction drop or longer the slip duration of an event, the lower its frequency.

For each slip event, we define 10 time windows, spanning from the first 10%–100% of an event's slip duration time with a 10% increment, as shown in Figure 4. Within each time window, eight statistical values for the velocity and displacement of the plate in the x and y directions are calculated. These statistics include mean, variance, median, higher-order moments and different percentiles. The statistics calculated using the data from the same time proportional window are used as input for a LightGBM model. The output of the model is the slip duration or friction drop of the corresponding slip event. Previous work primarily used continuous sliding time windows to calculate statistical values as input features (Rouet-Leduc et al., 2017), mainly because these studies aimed to predict continuous time series data, which requires a one-to-one correspondence between stress and acoustic data over time in a continuous manner. In contrast, our approach predicts the duration and friction drop of discrete slip events based on plate motion data from different time periods preceding each slip event, which is fundamentally different from the previous methods.

A total of 20 machine learning models are trained—10 models correspond to each of the 10 time windows data used for slip duration prediction, and the other 10 models for friction drop prediction. The performance of these models on the test set is shown in Figure 5a. Specifically, for the prediction of slip duration, when the time window is the first 10%–40%, the performance of the models is all acceptable, and R^2 is above 0.8. As the time window expands from 50% to 100%, the accuracy of the slip duration prediction progressively deteriorates. This suggests that the plate motion signals during the initial slip stage contain important information about the slip event's duration time, whereas the motion signals followed may contain noises for such prediction. For the friction drop prediction, R^2 shows an overall increasing trend with the expansion of the time window, which indicates that the accurate prediction of friction drop of slip events requires the motion signals in the entire slip stage. This also demonstrates the difficulty in friction drop prediction in laboratory earthquakes.

We select the two models with the best predictions for slip duration and friction drop, which are the model trained with the initial 30% time window data for predicting slip duration and the model trained with the 100% time window data for predicting friction drop, and plot the two models' predictive results respectively in Figures 5b and 5c. The performance in predicting the slip duration appears relatively better, with an R^2 above 0.9. In the case of friction drop prediction, the R^2 reaches a maximum of 0.8. It can be seen that the prediction accuracy is higher for slip events with larger friction drops, whereas the predicted values for events with smaller friction drops tend to be higher.

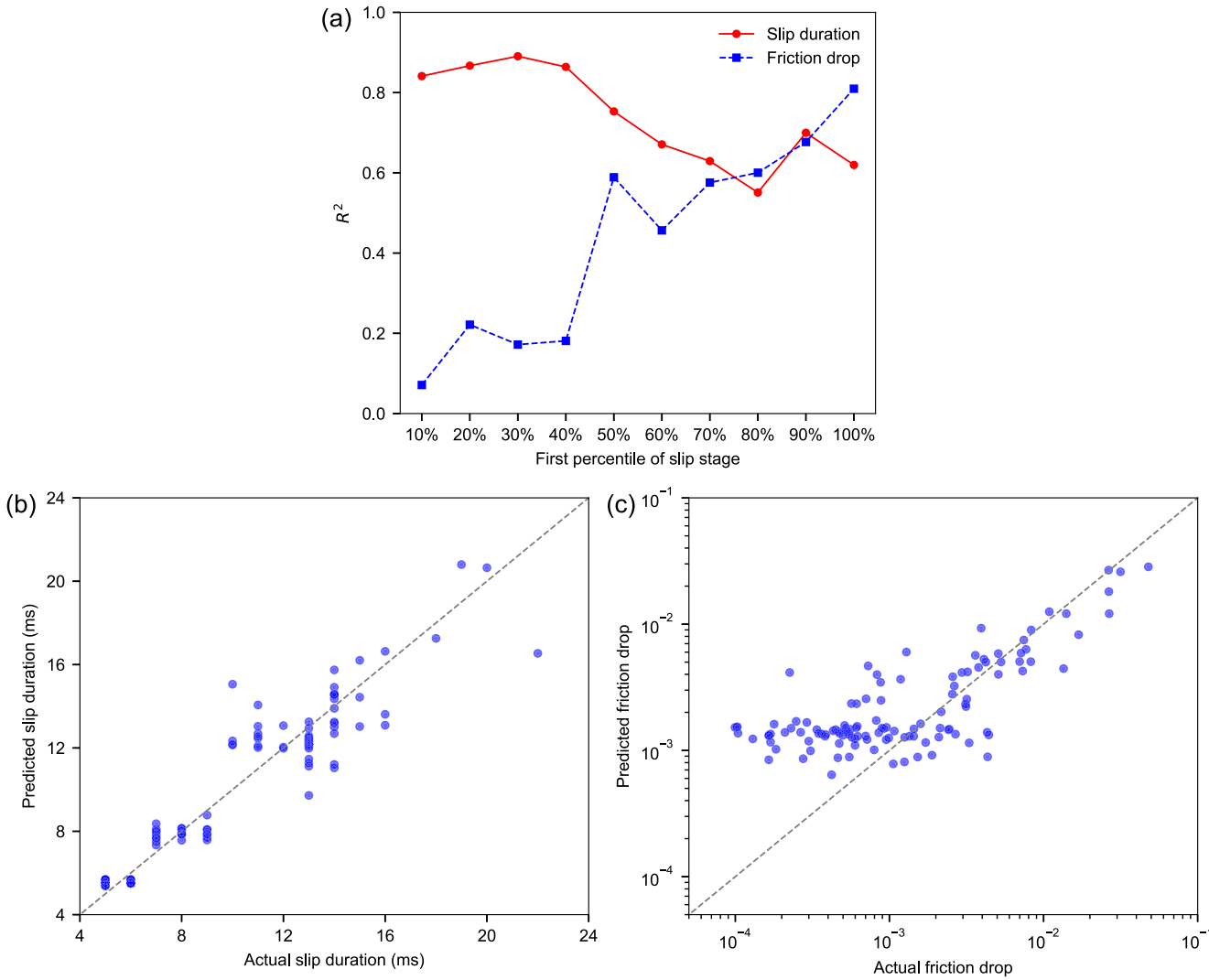

Figure 5. (a) The prediction performance of the models trained using the data spanning different time windows of the slip stage. Detailed prediction performance of best models for predicting, (b) the slip duration (time window is first 30% of the slip duration), and (c) the friction drop (time window is 100% of the slip duration).

4. Discussion

Upon reaching the best prediction, we can further analyze the importance of each feature in the two best models according to the SHAP values. To measure the feature importance, the absolute values of each feature's SHAP value across all slip events in the training set are averaged. Figure 6a and b show the mean absolute SHAP values of the top 10 important features in predicting slip duration and friction drop, respectively. Figure 6c, and d present how these 10 features affect each model's output by showing the SHAP value of the feature versus the corresponding feature value (low to high relative value, represented by the color bar) at each time in the form of violin graphs. It can be seen that the fourth-order moment (i.e., kurtosis) of D_x at the early stage of slip (30% time window of the slip stage) has the highest contribution to the prediction of slip duration, and the higher the fourth-order moment, the larger the SHAP value, that is, the higher positive contribution to the slip duration prediction. Higher kurtosis corresponds to a greater extremity of deviations (or outliers), which indicates that the slip events with longer duration have more extreme motions in the initial slip stage.

The second moment of V_x in the whole slip stage exhibits better predictive performance for friction drop, and a larger second moment of V_x contributes positively to a higher friction drop. The second moment is a statistical measure of the degree of data dispersion. A greater degree of dispersion in the shear direction of fault movement indicates more frequent particle rearrangements, which are also associated with larger seismic magnitudes.

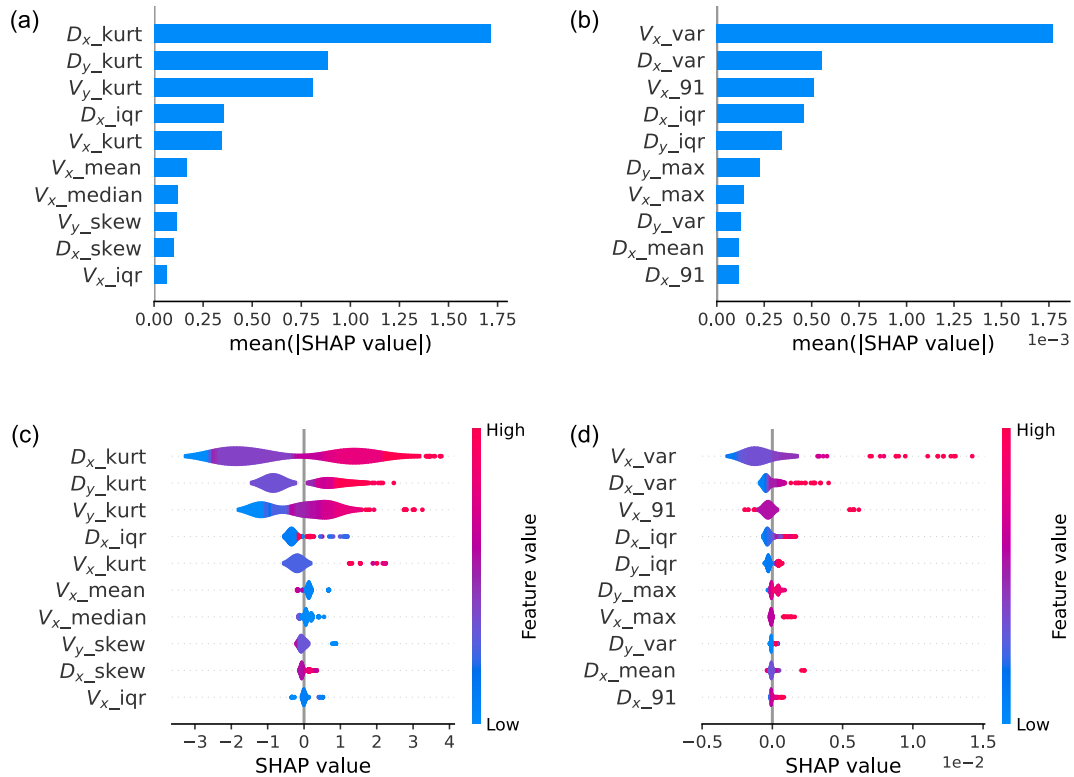

Figure 6. Feature importance analysis. SHAP values for the top 10 most important features in the two best models for predicting (a) slip duration and (b) friction drop, respectively. SHAP summary plots for the top 10 most important features in the two best models for predicting (c) slip duration and (d) friction drop, where each point represents the prediction from the model for a slip event.

Additionally, in both machine learning models, the statistical features of velocity and displacement along the normal direction (*y* direction) also contribute significantly to the model outputs, suggesting that particles perceive laboratory seismic events not only in the shear but also in the normal direction.

In previous research on machine learning predictions of laboratory earthquakes using physical experimental data (Bolton et al., 2019), the second and fourth moments of acoustic emission signals have been proven to have the highest contribution to predicting laboratory earthquakes' timing and shear stress evolution. Since plate boundaries and slowly sheared granular media exhibit a similar statistical distribution of velocity fluctuations (Meroz & Meade, 2017) and sheared granular fault systems can demonstrate comparable stick-slip behavior as natural fault systems (Anderson, 2007), we infer that the fourth moment of movement in natural faults during the nucleation stage may contain precursor information about the slip duration of an upcoming earthquake. The more extreme movements the fault experiences during the nucleation stage, the longer the slip duration of the subsequent earthquake. Estimating the slip duration can aid in distinguishing fast and slow earthquakes in advance (Hulbert et al., 2019), understanding the physical properties of earthquake phenomena, and enhancing earthquake prevention and disaster reduction capabilities (Obara & Kato, 2016). Although in our numerical model, the second moment of fault movement in the shear direction contributes the most to the prediction of friction drop, achieving accurate prediction results requires more movement signals during the late slip stage, which seems to imply the difficulty of predicting earthquake magnitude.

5. Conclusions

We have simulated a sheared granular fault system using FDEM, and use machine learning to predict the friction drop and slip duration of laboratory earthquakes. We show that the plate motion signals in the initial slip stage contain precursor information about the slip duration of laboratory earthquakes, while predicting the friction drop requires the plate motion signals in the entire slip stage. Moreover, we demonstrate that the fourth-order moment

of plate motion in the shear and normal directions during the initial slip stage are among the best estimators of the slip duration of laboratory earthquakes. For friction drop, the second moment of the plate motion signals in the shear direction during the entire slip stage is the best predictive feature. Our work demonstrates the ability of the machine learning approach to characterize highly aperiodic laboratory earthquakes in granular fault systems and identify crucial characteristic parameters for modeling such systems.

Data Availability Statement

The data supporting this paper are available at (Wei & Gao, 2024). The Python code for building the machine learning model is publicly available (Wei, 2024).

Acknowledgments
This work is supported by the National Natural Science Foundation of China (42374070) and the Guangdong Provincial Key Laboratory of Geophysical High-resolution Imaging Technology (2022B1212010002).

References

Anderson, D. L. (2007). *New theory of the Earth* (2 ed.). Cambridge University Press.
Bergstra, J., Yamins, D., & Cox, D. (2013). Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. Paper presented at the International conference on machine learning.
Bolton, D. C., Shokouhi, P., Rouet-Leduc, B., Hulbert, C., Rivière, J., Marone, C., & Johnson, P. A. (2019). Characterizing acoustic signals and searching for precursors during the laboratory seismic cycle using unsupervised machine learning. *Seismological Research Letters*, *90*(3), 1088–1098. https://doi.org/10.1785/0220180367
Brzinski Iii, T. A., & Daniels, K. E. (2018). Sounds of failure: Passive acoustic measurements of excited vibrational modes. *Physical Review Letters*, *120*(21), 218003. https://doi.org/10.1103/physrevlett.120.218003
Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., et al. (2015). Xgboost: Extreme gradient boosting. *R package version 0.4-2*, *1*(4), 1–4.
Corbi, F., Sandri, L., Bedford, J., Funiciello, F., Brizzi, S., Rosenau, M., & Lallemand, S. (2019). Machine learning can predict the timing and size of analog earthquakes. *Geophysical Research Letters*, *46*(3), 1303–1311. https://doi.org/10.1029/2018gl081251
Dahmen, K. A., Ben-Zion, Y., & Uhl, J. T. (2011). A simple analytic theory for the statistics of avalanches in sheared granular materials. *Nature Physics*, *7*(7), 554–557. https://doi.org/10.1038/nphys1957
Dieterich, J. H. (1992). Earthquake nucleation on faults with rate-and state-dependent strength. *Tectonophysics*, *211*(1), 115–134. https://doi.org/10.1016/0040-1951(92)90055-B
Dorostkar, O., Guyer, R. A., Johnson, P. A., Marone, C., & Carmeliet, J. (2017). On the micromechanics of slip events in sheared, fluid-saturated fault gouge. *Geophysical Research Letters*, *44*(12), 6101–6108. https://doi.org/10.1002/2017GL073768
Dratt, M., & Katterfeld, A. (2017). Coupling of FEM and DEM simulations to consider dynamic deformations under particle load. *Granular Matter*, *19*(3), 49. https://doi.org/10.1007/s10035-017-0728-3
Frank, W. B. (2016). Slow slip hidden in the noise: The intermittence of tectonic release. *Geophysical Research Letters*, *43*(19), 10–125. https://doi.org/10.1002/2016gl069537
Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, *38*(4), 367–378. https://doi.org/10.1016/s0167-9473(01)00065-2
Gao, K., Euser, B. J., Rougier, E., Guyer, R. A., Lei, Z., Knight, E. E., et al. (2018). Modeling of stick-slip behavior in sheared granular fault gouge using the combined finite-discrete element method. *Journal of Geophysical Research: Solid Earth*, *123*(7), 5774–5792. https://doi.org/10.1029/2018jb015668
Geller, D. A., Ecke, R. E., Dahmen, K. A., & Backhaus, S. (2015). Stick-slip behavior in a continuum-granular experiment. *Physical Review E*, *92*(6), 060201. https://doi.org/10.1103/physreve.92.060201
Guérin-Marthe, S., Nielsen, S., Bird, R., Giani, S., & Di Toro, G. (2019). Earthquake nucleation size: Evidence of loading rate dependence in laboratory faults. *Journal of Geophysical Research: Solid Earth*, *124*(1), 689–708. https://doi.org/10.1029/2018jb016803
Hulbert, C., Rouet-Leduc, B., Johnson, P. A., Ren, C. X., Rivière, J., Bolton, D. C., & Marone, C. (2019). Similarity of fast and slow earthquakes illuminated by machine learning. *Nature Geoscience*, *12*(1), 69–74. https://doi.org/10.1038/s41561-018-0272-8
Kato, A., Obara, K., Igarashi, T., Tsuruoka, H., Nakagawa, S., & Hirata, N. (2012). Propagation of slow slip leading up to the 2011 M w 9.0 Tohoku-Oki earthquake. *Science*, *335*(6069), 705–708. https://doi.org/10.1126/science.1215141
Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., et al. (2017). Lightgbm: A highly efficient gradient boosting decision tree. In *Advances in neural information processing systems* (Vol. 30).
Lei, Z., Rougier, E., Knight, E. E., Munjiza, A. A., & Viswanathan, H. (2016). A generalized anisotropic deformation formulation for geomaterials. *Computational Particle Mechanics*, *3*(2), 215–228. https://doi.org/10.1007/s40571-015-0079-y
Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*.
Ma, G., Zhou, W., Chang, X.-L., & Chen, M.-X. (2016). A hybrid approach for modeling of breakable granular materials using combined finite-discrete element method. *Granular Matter*, *18*, 1–17. https://doi.org/10.1007/s10035-016-0615-3
Meroz, Y., & Meade, B. J. (2017). Intermittent granular dynamics at a seismogenic plate boundary. *Physical Review Letters*, *119*(13), 138501. https://doi.org/10.1103/physrevlett.119.138501
Mousavi, S. M., & Beroza, G. C. (2020). A machine-learning approach for earthquake magnitude estimation. *Geophysical Research Letters*, *47*(1), e2019GL085976. https://doi.org/10.1029/2019gl085976
Munjiza, A., & Andrews, K. (1998). NBS contact detection algorithm for bodies of similar size. *International Journal for Numerical Methods in Engineering*, *43*(1), 131–149. https://doi.org/10.1002/(sici)1097-0207(19980915)43:1<131::aid-nme447>3.0.co;2-s
Munjiza, A., Rougier, E., & John, N. W. M. (2006). MR linear contact detection algorithm. *International Journal for Numerical Methods in Engineering*, *66*(1), 46–71. https://doi.org/10.1002/nme.1538
Munjiza, A. A. (1992). *Discrete elements in transient dynamics of fractured media* (PhD Thesis). Swansea University.
Munjiza, A. A. (2004). *The combined finite-discrete element method*. John Wiley & Sons.
Munjiza, A. A., Knight, E. E., & Rougier, E. (2011). *Computational mechanics of discontinua*. John Wiley & Sons.
Munjiza, A. A., Rougier, E., & Knight, E. E. (2014). *Large strain finite element method: A practical course*. John Wiley & Sons.

Obara, K., & Kato, A. (2016). Connecting slow earthquakes to huge earthquakes. *Science*, *353*(6296), 253–257. https://doi.org/10.1126/science.aaf1512

Ohnaka, M. (1996). Nonuniformity of the constitutive law parameters for shear rupture and quasistatic nucleation to dynamic rupture: A physical model of earthquake generation processes. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(9), 3795–3802. https://doi.org/10.1073/pnas.93.9.3795

Ren, C. X., Dorostkar, O., Rouet-Leduc, B., Hulbert, C., Strebel, D., Guyer, R. A., et al. (2019). Machine learning reveals the state of intermittent frictional dynamics in a sheared granular fault. *Geophysical Research Letters*, *46*(13), 7395–7403. https://doi.org/10.1029/2019gl082706

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, 2016). "Why should i trust you?" Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144. https://doi.org/10.1145/2939672.2939778

Rivière, J., Lv, Z., Johnson, P. A., & Marone, C. (2018). Evolution of b-value during the seismic cycle: Insights from laboratory experiments on simulated faults. *Earth and Planetary Science Letters*, *482*, 407–413. https://doi.org/10.1016/j.epsl.2017.11.036

Rouet-Leduc, B., Hulbert, C., Lubbers, N., Barros, K., Humphreys, C. J., & Johnson, P. A. (2017). Machine learning predicts laboratory earthquakes. *Geophysical Research Letters*, *44*(18), 9276–9282. https://doi.org/10.1002/2017gl074677

Ruiz, S., Aden-Antoniow, F., Baez, J. C., Otarola, C., Potin, B., Del Campo, F., et al. (2017). Nucleation phase and dynamic inversion of the Mw 6.9 Valparaíso 2017 earthquake in Central Chile. *Geophysical Research Letters*, *44*(20), 10–290. https://doi.org/10.1002/2017gl075675

Ruiz, S., Metois, M., Fuenzalida, A., Ruiz, J., Leyton, F., Grandin, R., et al. (2014). Intense foreshocks and a slow slip event preceded the 2014 Iquique M w 8.1 earthquake. *Science*, *345*(6201), 1165–1169. https://doi.org/10.1126/science.1256074

Socquet, A., Valdes, J. P., Jara, J., Cotton, F., Walpersdorf, A., Cotte, N., et al. (2017). An 8 month slow slip event triggers progressive nucleation of the 2014 Chile megathrust. *Geophysical Research Letters*, *44*(9), 4046–4053. https://doi.org/10.1002/2017gl073023

Tsai, J. C., Voth, G. A., & Gollub, J. P. (2003). Internal granular dynamics, shear-induced crystallization, and compaction steps. *Physical Review Letters*, *91*(6), 064301. https://doi.org/10.1103/physrevlett.91.064301

Wang, K., Johnson, C. W., Bennett, K. C., & Johnson, P. A. (2021). Predicting fault slip via transfer learning. *Nature Communications*, *12*(1), 7319. https://doi.org/10.1038/s41467-021-27553-5

Wech, A. G., Creager, K. C., Houston, H., & Vidale, J. E. (2010). An earthquake-like magnitude-frequency distribution of slow slip in northern Cascadia. *Geophysical Research Letters*, *37*(22). https://doi.org/10.1029/2010gl044881

Wei, M. (2024). Python code for Machine learning predicts the slip duration and friction drop of laboratory earthquakes in sheared granular fault [code]. https://doi.org/10.5281/zenodo.13845904

Wei, M., & Gao, K. (2024). Supporting data for Machine learning predicts the slip duration and friction drop of laboratory earthquakes in sheared granular fault [dataset]. https://doi.org/10.5281/zenodo.12803380

References From the Supporting Information

Dorostkar, O., Guyer, R. A., Johnson, P. A., Marone, C., & Carmeliet, J. (2017). On the role of fluids in stick-slip dynamics of saturated granular fault gouge using a coupled computational fluid dynamics-discrete element approach. *Journal of Geophysical Research: Solid Earth*, *122*(5), 3689–3700. https://doi.org/10.1002/2017JB014099

Euser, B., Rougier, E., Lei, Z., Knight, E. E., Frash, L. P., Carey, J. W., et al. (2019). Simulation of fracture coalescence in granite via the combined finite–discrete element method. *Rock Mechanics and Rock Engineering*, *52*(9), 3213–3227. https://doi.org/10.1007/s00603-019-01773-0

Ferdowsi, B. (2014). Discrete element modeling of triggered slip in faults with granular gouge (Ph.D Thesis). In *Application to dynamic earthquake triggering*.

Gao, K., Euser, B. J., Rougier, E., Guyer, R. A., Lei, Z., Knight, E. E., et al. (2018). Modeling of stick-slip behavior in sheared granular fault gouge using the combined finite-discrete element method. *Journal of Geophysical Research: Solid Earth*, *123*(7), 5774–5792. https://doi.org/10.1029/2018JB015668

Lei, Q., & Gao, K. (2018). Correlation between fracture network properties and stress variability in geological media. *Geophysical Research Letters*, *45*(9), 3994–4006. https://doi.org/10.1002/2018GL077548

Lei, Z., Rougier, E., Munjiza, A., Viswanathan, H., & Knight, E. E. (2019). Simulation of discrete cracks driven by nearly incompressible fluid via 2D combined finite-discrete element method. *International Journal for Numerical and Analytical Methods in Geomechanics*, *43*(9), 1724–1743. https://doi.org/10.1002/nag.2929

MiDi, G. (2004). On dense granular flows. *The European Physical Journal E*, *14*(4), 341–365. https://doi.org/10.1140/epje/i2003-10153-0

Okubo, K., Bhat, H. S., Rougier, E., Marty, S., Schubnel, A., Lei, Z., et al. (2019). Dynamics, radiation and overall energy budget of earthquake rupture with coseismic off-fault damage. *Journal of Geophysical Research: Solid Earth*, *124*(11), 11771–11801. https://doi.org/10.1029/2019jb017304

Rougier, E., Munjiza, A., Lei, Z., Chau, V. T., Knight, E. E., Hunter, A., & Srinivasan, G. (2019). The combined plastic and discrete fracture deformation framework for FDEM. *International Journal for Numerical Methods in Engineering*, *121*(5), 1020–1035. https://doi.org/10.1002/nme.6255

Tatone, B. S. A., & Grasselli, G. (2015). A calibration procedure for two-dimensional laboratory-scale hybrid finite–discrete element simulations. *International Journal of Rock Mechanics and Mining Sciences*, *75*(Supplement C), 56–72. https://doi.org/10.1016/j.ijrmms.2015.01.011