Correspondence to:

S. Karimpouli,
sadegh.karimpouli@gfz-potsdam.de

Empowering Machine Learning Forecasting of Labquake Using Event-Based Features and Clustering Characteristics

Sadegh Karimpouli[1] , Grzegorz Kwiatek[1] , Yehuda Ben-Zion[2] , Patricia Martínez-Garzón[1] , Georg Dresen[1,3] , and Marco Bohnhoff[1,4]

[1]Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, Potsdam, Germany, [2]Department of Earth Sciences, Statewide California Earthquake Center, University of Southern California, Los Angeles, CA, USA, [3]Institute of Earth and Environmental Sciences, Universität Potsdam, Potsdam, Germany, [4]Department of Earth Sciences, Free University Berlin, Berlin, Germany

Abstract Following recent advances of machine learning (ML), we present a novel approach to extract spatiotemporal seismo-mechanical features from Acoustic Emission (AE) catalogs to empower ML-based forecasting. The AE data were recorded during laboratory stick-slip experiments on granite samples cut by rough faults. Based on the features computed for a past time window, a random forest (RF) classifier is used to forecast the occurrence of a large magnitude event ($M_{AE} > 3.5$) in the next time window. Event-based features allow us to associate informative time-space characteristics to each feature and nearest-neighbor clustering analysis enables us to separate background and clustered seismicity and train individual models. The results show that the separation of AEs enhances the forecasting accuracy from 73.2% for the entire catalog up to 82.1% and 89.0% if background and clustered events are used separately. The presented new approach may be upscaled for applications to forecast tectonic earthquakes.

Plain Language Summary It is widely discussed to use machine learning (ML) attempts to predict seismic events generated during rock deformation experiments in the laboratory. To improve these predictions, one needs to either refine the model or provide it with more detailed and informative physically understandable input data. In our study, we examine Acoustic Emission (AE) catalogs from laboratory experiments involving repetitive slips of the rough fault surfaces observed on three Westerly granite samples. By analyzing AE catalog data of past AE activity using a moving time window, we employ a random forest classifier, a type of ML tool, to forecast the likelihood of a significant earthquake happening in the future time period. What makes this approach different is that we introduce a new method to calculate features related to both location and timing of seismic activity. Our findings reveal that separating earthquake catalog into background and clustered seismicity is crucial for improving forecasting accuracy. Regarding an accuracy of 73.2% for the whole catalog, we achieved an enhanced accuracy of up to 82% for background events and 89% for clustered events. We discuss how the insights gained from this study can be scaled up for forecasting tectonic earthquakes in real time.

1. Introduction

Earthquakes are the second most devastating natural disasters after meteorological phenomena (e.g., flood, cyclone, etc.), causing human fatalities, injuries, and substantial financial losses. About half of the worldwide fatalities from natural disasters in the last 40 years were caused by earthquakes and associated cascading hazards such as landslides, fires, and tsunamis (see https://www.munichre.com/en/risks/natural-disasters/earthquakes.html). Predicting the time to location and magnitude of future large earthquakes is still impossible but critically needed to save lives and alleviate socioeconomic consequences. The runup to large earthquakes involves complex physical processes, especially in heterogeneous tectonic settings and fault zones that so far are not fully understood and near-impossible to monitor with sufficient resolution. Recently, machine learning (ML) methods opened a new avenue for time-to-failure (TTF) prediction using waveform data from laboratory double-shear experiments (Marone, 1998; Rouet-Leduc et al., 2017). Previous studies employed either supervised prediction of parameters such as shear stress, fault friction, displacement, and slip velocity (Lubbers et al., 2018; Rouet-Leduc et al., 2018) or unsupervised clustering (Bolton et al., 2019; Karimpouli et al., 2024). Following early promising results from a Kaggle competition based on laboratory earthquake prediction (Johnson et al., 2021), researchers focused on developing more general and accurate ML models for more complex and limited data sets

to forecast future earthquakes in real time. Further enhancements of the results were achieved by refining the "ML models" and/or enriching the "input data" as summarized below.

Jasperson et al. (2021) concluded that more *sophisticated ML architectures* such as ensembles of long-short-term-memory (LSTM) and attention networks could lead to more accurate shear stress and TTF forecasts over a wider variety of normal stress conditions in double-shear experiments. Subsequently, Laurenti et al. (2022) and Wang et al. (2022) used state-of-the-art transformer networks to forecast fault zone stress along with fault displacement and friction in laboratory experiments and slow slip events. Wang et al. (2021) used *transfer learning* from numerical simulations to improve the prediction of laboratory fault friction. Borate et al. (2023) used the governing physical equation in the training process to form *physics-informed neural networks* and showed that their network improves transfer learning for small training data sets.

Unlike these studies, which focused on refining the ML models, Karimpouli et al. (2023) used *catalog-driven physics-informed features*, developing informative input data for TTF prediction of labquakes. This approach was applied to a limited data set available from stick-slip experiments on rough fault samples. Extracting physics-informed features allowed the authors to explain the importance of each feature used for the predictive model. Other researchers *used/combined other sources of information* to enrich the input data for a better coverage of the complexity of the earthquake process. Shokouhi et al. (2021) and Shreedharan et al. (2021) showed that using active source ultrasonic monitoring has advantages for the prediction of timing and size of the labquakes. Corbi et al. (2019, 2020) used synthetic geodetic data in an experimental model and suggested that the imminence of earthquakes may be forecasted by complex motion recorded at subduction zones. Licciardi et al. (2022) used prompt elastogravity signals and showed that their model provides real-time access to the stress evolution of large earthquakes. Saad et al. (2023) showed that with a combination of features extracted from electromagnetic (EM) and geoacoustic (GA) sensors they were able to predict location and magnitude of the next large earthquake in southwestern China.

Although preparation processes of large earthquakes may involve spatio-temporal evolution and localization of low magnitude seismicity (Kato & Ben-Zion, 2020), most of the above studies did not include spatial information of seismic events. Methods based on catalog-driven features in time windows (Karimpouli et al., 2023; Picozzi & Iaccarino, 2021) partially use spatio-temporal information by providing some features such as correlation integral, fractal dimension, and clustering features. Zaliapin et al. (2008) showed that seismicity can be statistically divided into background and clustered events. Based on the space-time-magnitude properties of events, Zaliapin and Ben-Zion (2013a, 2013b) developed a data-driven nearest-neighbor approach to identify earthquake clusters, which may contain one (singles) or more (families) events. Topological properties of the families were used to separate them into classical burst-type sequences and swarms concentrated in separate regions reflecting different physical processes.

In this study, we propose a new approach of *event-based features* to empower ML-based earthquake forecasting by enriching the input data set. The main idea is to compute and assign the features based on two time and space windows for each event. Each feature incorporates an evolution spanning a defined time period in a 3D space. The advantages of this approach are (a) adding spatial information to conventional features, (b) relating multiscale spatiotemporal features to physical processes governing the earthquakes, (c) increasing the amount of data, which is desired for applying ML methods, and (d) allowing separating events with associated features into clustered and background populations. Since the clustering analysis identifies clusters of events when more than one event constitutes different families, we further introduce new family features. We show how clustering provides new additional features that increase the accuracy of forecasting of an upcoming large event in the next time window. We also account for the overall structure of evolving large event cycles corresponding to intermittent criticality (Ben-Zion, 2008; Ben-Zion et al., 2003; Kwiatek et al., 2024).

2. Experimental Setup and Acoustic Emission Data

The data set is obtained from triaxial stick-slip experiments carried out on three Westerly granite samples, namely WgN04, WgN05, and WgN07 (Goebel et al., 2012, 2013; Kwiatek & Goebel, 2024). Large blocks were cored to achieve cylindrical samples with 102–107 mm lengths and 40 mm diameter. To produce rough faults in intact specimens, two notches with 1.5–2.2 cm depth and 30° inclination to the specimen axis were cut in the samples (Figure 1). These specimens were first subjected to an isostatic confining pressure of 75 MPa, before axial load was increased until shear failure occurred forming a rough fault. Then, the confining pressure was elevated up to
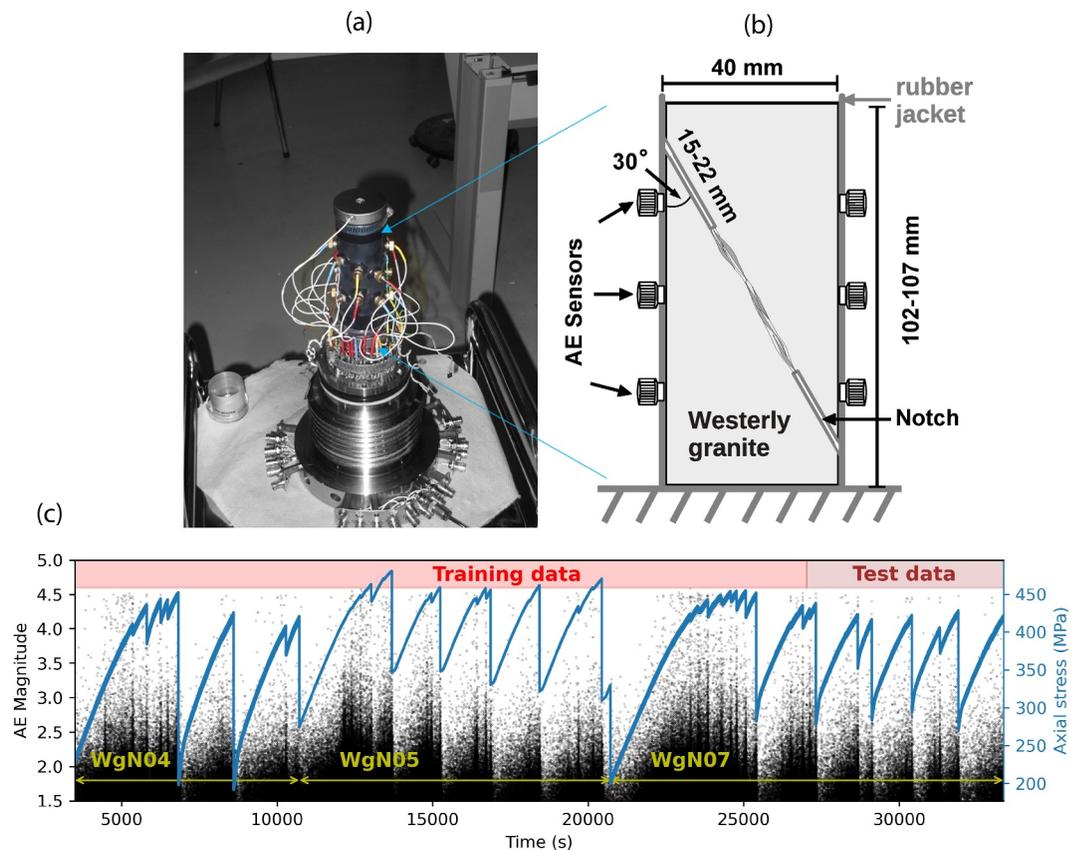
Figure 1. (a) Experiment setup. The sample is contained in a rubber jacket allowing the placement of 14 Acoustic Emission (AE) sensors, which are glued to the specimen surface, sealed in the jacket with epoxy and signals are transmitted through cabled seals in the bottom plug. (b) Schematic cross section of sample inside rubber jacket. The lines between two notches represent a rough fault (modified from Dresen et al., 2020; Goebel et al., 2013). (c) Axial stress (blue line, right axis) and AE events magnitude (black dots) recorded on three samples: WgN04, WgN05, and WgN07. Time axis is expressed as an accumulative form as all data from different experiments have been put together for a better visualization. Yellow arrows show the data obtained for each individual sample. Following such a configuration, the 75% first events were used as training data and the remaining 25% events as test data.

150 MPa to lock the faults. Subsequently, progressive axial loading at a constant displacement rate of 0.33 μm/s was performed to induce stick-slip events (Goebel et al., 2012, also see Text S1 in Supporting Information S1 for details). Figure 1c shows the axial stress curve measured for all three samples during the individual experiments with an accuracy level of ±0.05 MPa. To monitor Acoustic Emission (AE) activity, 14 piezoelectric AE sensors encased in brass housing were glued on the sample surface using a low-viscosity epoxy. Additional two sensors were mounted in steel spaces at the top and bottom of the sample. Using AE sensors, ultrasonic pulses were used to periodically update the changing velocity field. AE sensors have a resonant frequency of 1 MHz (Figure 1). Complete waveforms were captured with 16-bit resolution at a sampling rate of 10 MHz.

The fundamental processing steps involved extracting the time-dependent characteristics of AE data (Kwiatek et al., 2024) from triggered recordings. The AE waveform processing starts with the automatic identification of *P*-wave arrivals of AE events on the 14 AE sensors using a modified convolutional neural network scheme (Ross et al., 2018). Utilizing the continuously updated quasi-anisotropic velocity model derived from ultrasonic transmission measurements (Goebel et al., 2012), the hypocenter locations were determined through a grid search algorithm paired with the Coyote optimization algorithm (Pierezan & Dos Santos Coelho, 2018) with an estimated accuracy of approximately ±2 mm (more details in Dresen et al., 2020; Kwiatek et al., 2024). The relative AE magnitude was calculated from first P-wave amplitudes (Goebel et al., 2012). The obtained catalog includes information on origin time, location in the local Cartesian coordinate system, and AE magnitude. The completeness of AE magnitudes was estimated for all catalogs using a goodness-of-fit method (Wiemer &

Wyss, 2000), resulting in a magnitude of completeness M_C of 1.5. We used all events with magnitude above M_C (Figure 1c), which led to the seismicity catalogs for WgN04, WgN05, and WgN07 containing 75,810, 153,313, and 125,299 events, respectively. In this study, we combined all events and used the first 75% of these events (WgN04, WgN05, and the first part of WgN07, see Figure 1c) as training data and the following 25% of events (the last part of WgN07, see Figure 1c) as testing data. The data set is available as separate data publication (Kwiatek & Goebel, 2024).

3. Enhanced Seismic Features

Traditional feature computation is based on selecting a time window and computing the features on regular time intervals (Kwiatek et al., 2024; Picozzi et al., 2023). Subsequently, the computed features are assumed as time series and could be introduced into, for example, an LSTM network for either TTF prediction (Karimpouli et al., 2023; Laurenti et al., 2022) or foreshock-aftershock classification (Picozzi & Iaccarino, 2021). The main idea in this study is to enrich the features via including both temporal and spatial information for each event known as *event-based features*. Figure 2 illustrates how the selected 81 features in this analysis were computed for each AE event. When any AE event happens, three *immediate features* are defined (Figure 2a) consisting of the magnitude of the event (*m_ae*) and the separations in time and space between the current and previous event (*int-ev_time*, *int-ev_space*). To associate time-space characteristics to each feature, time and space windows are first defined relative to the origin time and the location of the considered event (Figure 2b). *Time-space features* are then computed based on the selected events included in the defined time and space windows (see Section 3.1 for details). However, time-space-magnitude distance (Baiesi & Paczuski, 2005) are computed for all events in time-lapse, while a clustering analysis (Zaliapin & Ben-Zion, 2013b, 2016; Zaliapin et al., 2008) divides the population into background and clustered events, which form clusters of one event (singles) or multiple events (families). *Family features* are thus defined based on the topological properties of these families (see Figure 2c and Section 3.2).

3.1. Time-Space Features

We compute time-space features individually for each AE event. We define a time window as well as a spherical space window relative to the origin time and location of the considered AE event (Figure 2). All AE events located inside the time-space window are used to compute a feature. We start from the feature pool developed and discussed in Kwiatek et al. (2024) and used in Karimpouli et al. (2023) for TTF prediction as well as additional features described in Picozzi et al. (2023). As the features are now computed per-event, we only use those for which the calculations can be performed reliably and efficiently while having limited input data (see Text S2 in Supporting Information S1 and Kwiatek et al., 2024). We ultimately use AE-rate (*n*: number of events per second), Gutenberg-Richter *b*-value (*b*: the slope of the magnitude-frequency distribution), scaled total magnitude (*stm*: sum of magnitude of all evens normalized to the volume of spatial distribution of events), correlation integral (*c(r)*: the probability of two points being separated by a spatial distance parameter "*r*" less than a certain value), and event proximity and clustering features (*trp, trq*: product and ratio of the magnitude-normalized time (*T*) and space (*R*) components as well as *pfo, pma, paf*: proportion of for-, main- and aftershocks). The fractal dimension that was used in Kwiatek et al. (2024) to characterize the roughness evolution is replaced with correlation integrals calculated over different distance parameters.

For temporal dimension, Karimpouli et al. (2023) found features computed over longer time windows important for an accurate prediction of TTF. However, some features with shorter time windows, such as AE-rate, were found to be even more important than others spanning longer time windows. Therefore, two time windows of 100 and 25 s are selected in this study (Figure 2b). As for the spatial dimension, we select three spherical space windows with a radius of 5, 10, and 20 mm (Figure 2b), which aim to capture physical processes at different length scales. Since the radius of the sample is 20 mm, the largest spatial window of 20 mm practically contains the whole volume around the fault plane. Consequently, the features calculated using the largest spatial window are effectively time-dependent (similar to that used by Karimpouli et al., 2023) and not space-and-time dependent.

Figures 3a–3c show the temporal changes in the values of the event-based AE-rate feature for sample WgN05 using different time-space windows. These features form cloud-like point distributions with a high variability of observable values. Since we calculate the features with at least 50 events available in the selected spatiotemporal window, these values were assumed to be robust. The different spreads for the various windows indicate that they
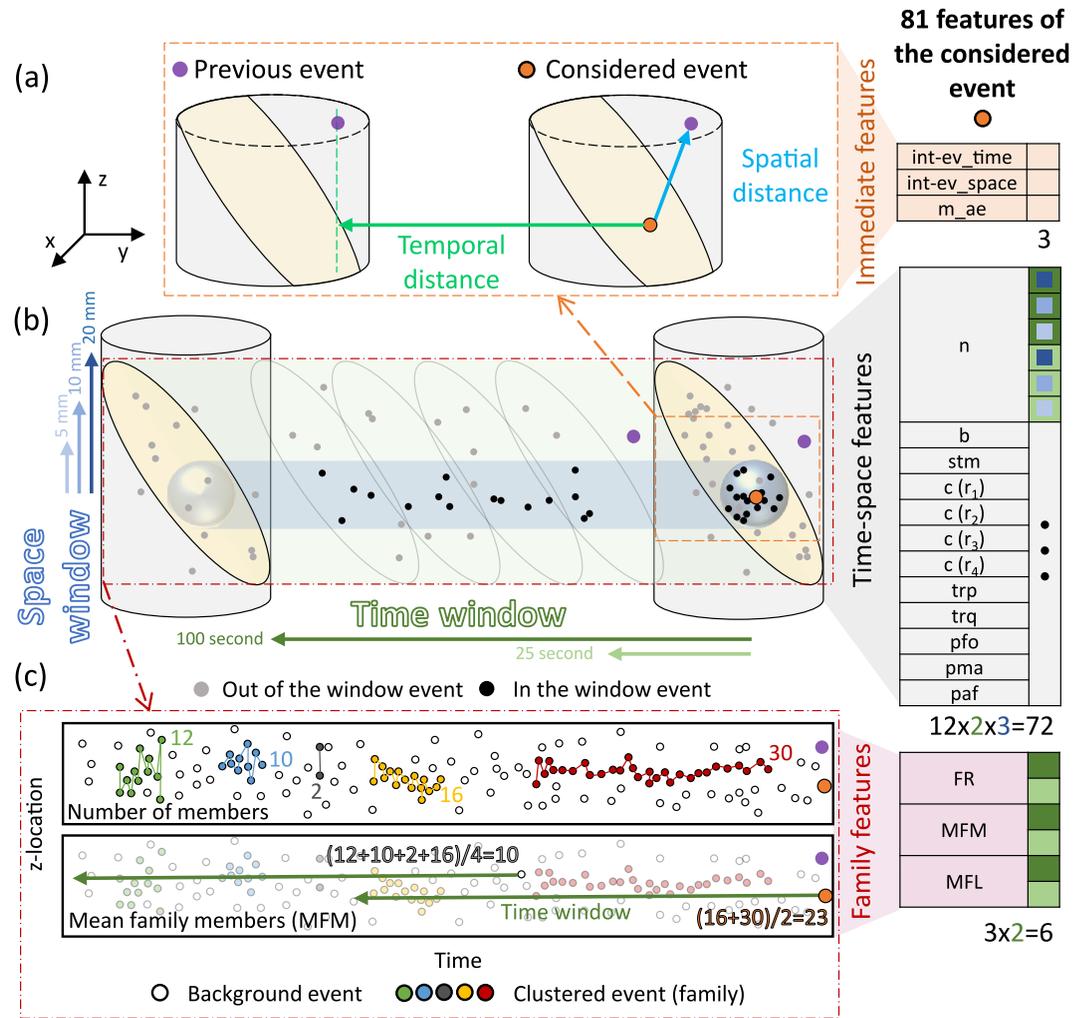
Figure 2. Illustration of the event-based computation of the 81 features used in this study. (a) Immediate features were based on the source characteristics of the considered Acoustic Emission (AE) event and its spatiotemporal relation to the previous AE event (3 features), (b) Time-space features computed based on the subsets of AE catalog constrained by time and space distance from the considered event (12 features × 2 time windows × 3 space windows = 72 features, See Section 3.1 for details), and (c) Family features based on topological characteristics of the event families extracted from clustering analysis in different time windows (3 features × 2 time windows = 6 features, see Section 3.2 for details). For example, in the upper plot, the number of members (events) per family is shown. In the lower plot, the mean family members (MFM) is computed for each individual event based on the mean value of all families existing in the time window (e.g., for the considered event, MFM = (16 + 30)/2, which is assigned to the corresponding event). For feature abbreviations and mathematical computations see Text S2 in Supporting Information S1.

provide information in different time-space scales. Features with larger spatial windows are less spread because they capture more general processes using more events available in the selected windows. Smaller windows involve events, which are more responsible to local processes, leading to more heterogeneity. A feature covering a small time-space window (e.g., Figure 3c) generally contains lower number of events and may not have enough events for a stable feature computation. Figures 3d–3f also show some of the other features from the same sample. In total, the different combinations of physical parameters and two spatial and three temporal windows yielded a pool of 72 time-space features, which are assigned to each event (Figure 2).

3.2. Family Features

Earthquake clustering is commonly observed in natural and induced seismicity. Following Zaliapin et al. (2008) and Zaliapin and Ben-Zion (2013b, 2016), we automatically separate events from the AE catalog into clustered
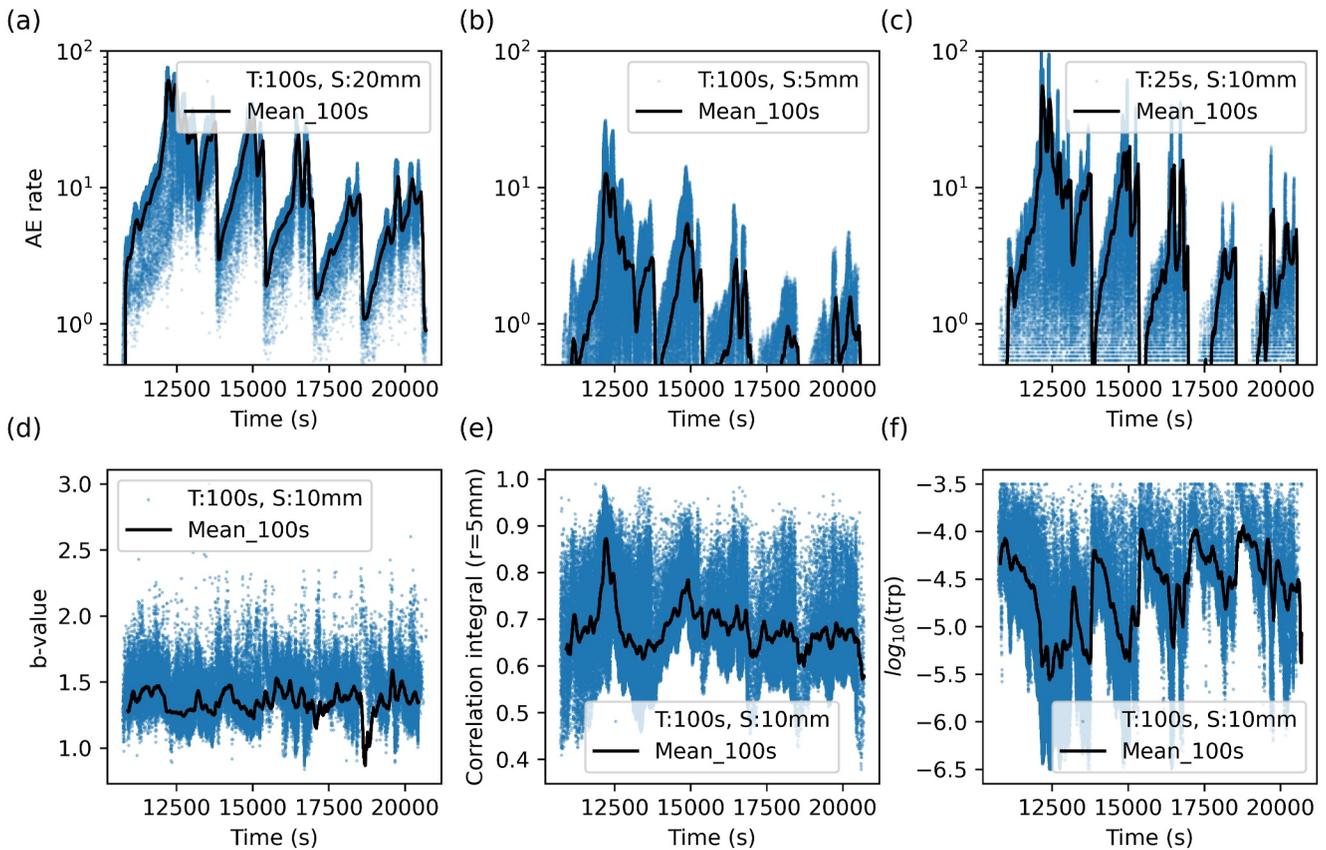
Figure 3. Variability of event-based features calculated for sample WgN05 (space-time windows used are marked in legends). Each dot represents the Acoustic Emission (AE) rate of each event. The black curve is the mean value of these dots in a 100 s moving window that indicates the overall trend (cf. Figures 2–4 in Kwiatek et al. (2024) for time features). (a)–(c) AE rates using different space-time windows (d) *b*-value, (e) correlation integral, and (f) product of T and R (*trp*) in the logarithmic scale.

and background events using the Gaussian mixture model (GMM) (Dempster et al., 1977). Figure 4a shows a representative distribution of nearest-neighbor distances for AE events of the sample WgN05. Clustered events show smaller distances in both normalized magnitude time (T) and space (R) components and tend to form a separate distribution from the background events. The relation between AE-rate values and earthquake clustering is shown in Figure 4b. Most events with a highly variable AE-rate belong to background seismicity, while clustered events display typically high rates.

Figure 4b illustrates families in different stress evolution stages from initial to critical stress states for the first labquake cycle of sample WgN05. In Figures 4c–4e, the clustered families of these stages are shown subdivided in 100 s long time windows. The resulting forests contain many background events and clusters of one event (singles) or multiple events (families) evolving through different stress states (Zaliapin & Ben-Zion, 2013b, 2016). With progressive loading causing a transition from initial to critical stress state (Ben-Zion et al., 2003), the number of families increases (Figures 4c–4e) and longer families (specially in time) with more members are formed (Kwiatek et al., 2024).

While general clustering features have been calculated in Sec. 3.1 (see Text S2 in Supporting Information S1), following Zaliapin and Ben-Zion (2013b, 2016), we also compute some features of families based on their topology. These include *family rate (FR)*: numbers of families (N_F) in a time window (t_{win}, here 100 s), *mean family members (MFM)*: average number of members per family (M_{F_i}) in a given t_{win}, and *mean family length (MFL)*: average time length per family (F_i) as a time difference between first (t_{first}) and last (t_{last}) member in a similar t_{win} (Equations 1–3).
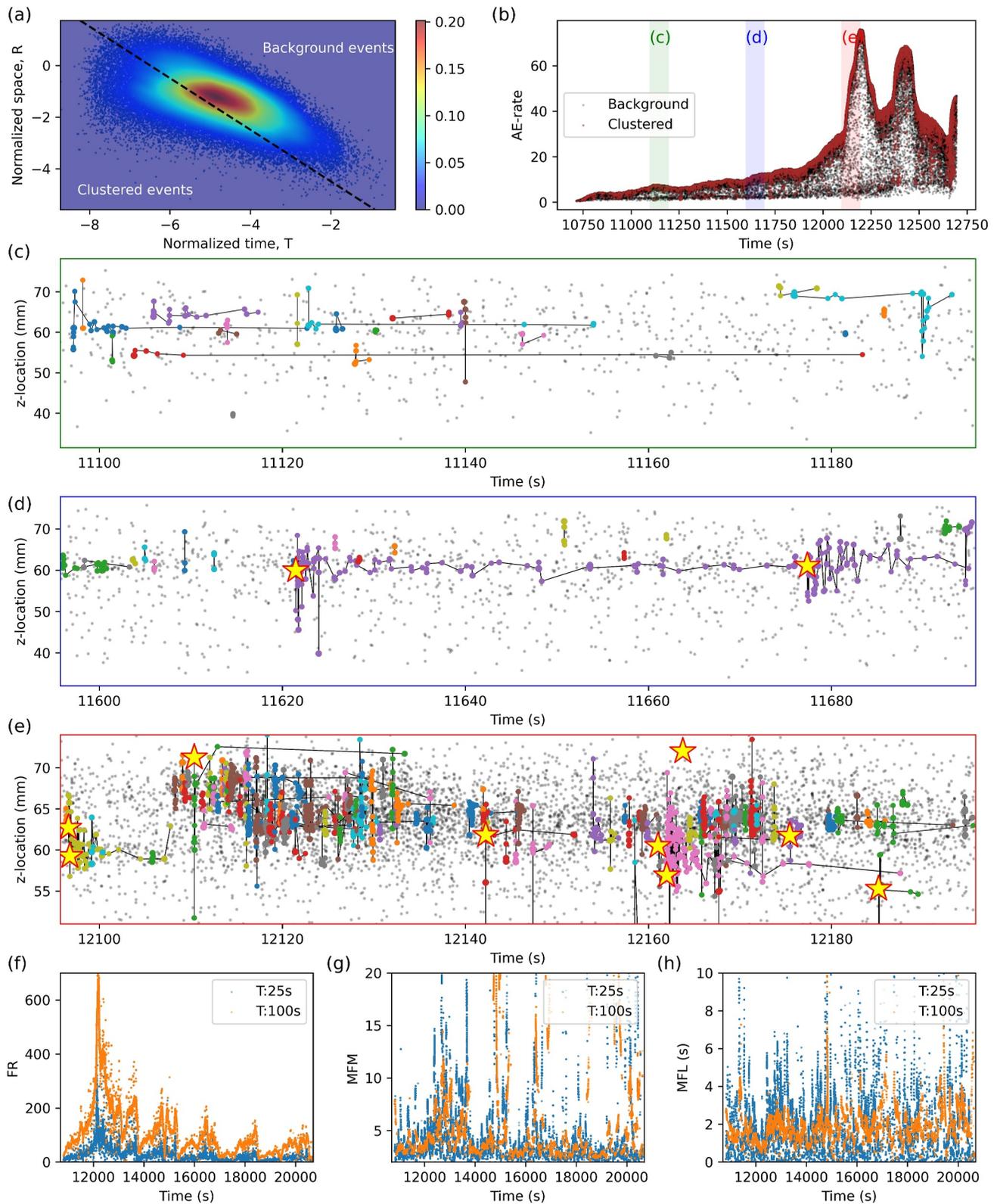
Figure 4.

$$FR = \frac{N_F}{t_{win}} \tag{1}$$

$$MFM = \frac{\sum\limits_{i=1}^{N_F} M_{F_i}}{N_F} \tag{2}$$

$$MFL = \frac{\sum\limits_{i=1}^{N_F} \left(t_{last} - t_{first}\right)_{F_i}}{N_F} \tag{3}$$

As illustrated in Figure 2c, for each current time window of the catalog data (100 s), we extract all families using a threshold value of nearest-neighbor proximity. Then, for each family, we calculate some characteristics such as the number of members and the time length of the family and assign them to the last event of each family. Following the idea of event-based features, the new family features should be assigned to all events representing either background or clustered seismicity. Specifically, we compute the family features (Equations 1–3) for each event using two previously defined moving average time windows of 100 and 25 s. These features are computed based on the family characteristics assigned to the last event of each family, leading to 6 new family features (cf. Figure 2c). Thus, although characteristics of families are computed using only clustered events, we attributed them to all AE events, regardless of whether they were classified as background or clustered (cf. Figure 2c). Figures 4f–4h show the variations of family features for sample WgN05.

Considering immediate, time-space, and family features, the feature vector for each AE event contains 81 (3 + 72 + 6) features serving as the input for the ML scheme.

4. ML Model and Results

4.1. Problem Definition

Application of ML-based methods for earthquake forecasting requires a precise definition of the problem. To define TTF (Corbi et al., 2019; Karimpouli et al., 2023; Rouet-Leduc et al., 2017), an earthquake cycle with exact start and end point in time should be known. In the laboratory data, relatively large stress drops define a cycle rather than large magnitude events. Since for earthquakes in nature the stress drop is not well known, the TTF definition is not easily applicable except for limited cases with cyclic behavior such as Cascadia (Hulbert et al., 2018). Importantly, tectonic earthquake cycles may not be redundant like in laboratory tests and even if they are, such long historical data sets may not be available. As an example of earthquake forecasting in real time, Saad et al. (2023) proposed a classification approach and used the data of a current time window (current week) to forecast if there is a large magnitude earthquake ($M_{AE} > 3.5$) in the next time window (next week). This is comparable to probabilistic seismic hazard assessment (PSHA), as defined by the probability of exceedance of specific magnitude in a specific time window (Cornell, 1968).

Having access to the catalog data of the previous 100 s, we define the current time window length (Figure 5). The length of the next time window for forecasting is arbitrarily set to 20 s (20% of the current window). We also use a threshold value of $M_{AE} = 3.5$ for AE-magnitude to define a large event. For each current event, the feature vector is used as the input of a random forest (RF) classifier to forecast the large event in the next time window. The RF model contains several trees, with each tree predicting a class of 0 or 1 based on the occurrence of a large event (no:0, yes:1). The final class is obtained using a soft voting, where each tree predicts the class probabilities, and then the ensemble is just weighted by their probabilities (see Text S3 and Figure S1 in Supporting Information S1 for details).

Figure 4. (a) Sample distribution of the normalized-magnitude distance T and R for WgN05 color-coded with sample normalized density. Dashed line is the separation line between background and clustered AEs. (b) AE-rate feature values as a function of time color-coded by the classification into background and clustered seismicity. (c)–(e) Evolution of clustered family trees in 100 s time windows for different stress stages of (c) stable, (d) transitional, and (e) critical along sample axis z (also shown in b). Each family consists of events with similar colors and a line connecting them. For better visualization, families with members more than or equal to 3, 3, and 5 are shown in plots c, d, and e. Large events ($M_{AE} > 3.5$) are shown with big yellow stars. (f)–(g) Examples of new family based features computed in 25 and 100 s windows, namely (f) FR: family rate, (g) MFM: mean family members, and (h) MFL: mean family length.
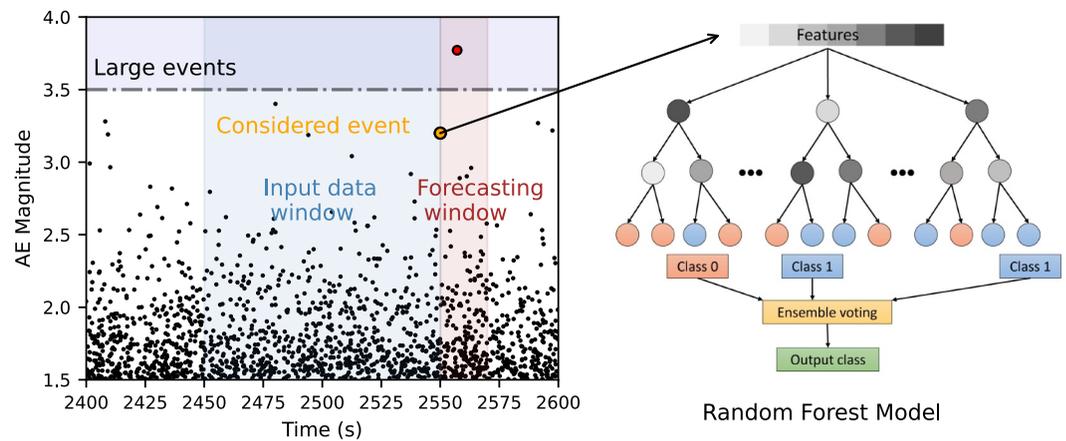
Figure 5. Problem definition and the proposed approach. We use all data from an input data window with 100 s length for computing features of the considered event (orange dot) and apply a random forest (RF) classifier to forecast the existence of a large event ($M_{AE} > 3.5$, red dot) in the next forecasting window with 20 s length. If yes, the output is 1, otherwise it is 0. For more details about the RF model and architecture of each tree, see Text S3 and Figure S1 in Supporting Information S1.

4.2. Random Forest Classification

We employed the RF algorithm, being a powerful ML classifier, to forecast the occurrence of a large AE event in the next time window. As illustrated in Figure 5, all 81 features of the current event are introduced as input values for the RF. We used the Scikit-Learn library in Python to construct, train, and test the RF model. A grid search on the number of individual trees and a maximum depth revealed that an optimum architecture for this problem is obtained with 50 trees and a maximum depth where there are less than two nodes in the tree to be split (see Text S3 in Supporting Information S1).

Our results reveal that the forecasting accuracy of the training data exceeds 99%, which indicates the RF model learned near perfectly the logic of such classification among the training data. However, the RF results on test data show accuracy, precision, and recall of 73.2%, 74.7%, and 80.3%, respectively (Table 1). Since RFs are resistant against overfitting (Breiman, 2001), a lower accuracy of test data means that a considerable amount of feature vectors from the test data is not contained in the trained model. These events would display a statistical distribution (i.e., in the feature space) out of the distribution of the training data (also reported by Karimpouli et al., 2023). To investigate further, we conduct a 10-fold cross-validation (CV) evaluation on the training data by dividing them into 10 folds. The model is trained based on 9 folds and validated by the remaining fold, and this procedure is repeated for all folds. The CV scores are 66.02%, 60.77%, 69.78%, 55.82%, 89.22%, 85.31%, 68.11%, 68.57%, 63.83%, and 92.82%. The mean value of CV score is 72.02% with a standard deviation of 11.96%, which is very close to what we obtained for the test data (with accuracy of 73.2%). This confirms the complexity contained in the experimental data set we analyze, even though all of them were recorded on a similar rock type with an identical experimental procedure.

We used clustering analysis once to compute family features; however, we here use it again to separate the background and clustered event of the entire catalog. A key result of our study is that the classification accuracy is enhanced considerably if the population of AE events is separated into background and clustered AEs. This requires considering event-based features where each event is attributed a feature vector composed of 81 features values. Separating background and clustered events allows training RF models individually for each population. For example, the accuracy of the classification increased from 73.2% while using the full catalog to 82.1% and 89.0% when only background or clustered events are considered, respectively. To evaluate the effect of new family features, all RF models were re-trained without family features, that is, with 75 (including immediate and time-space) features. Table 1 shows that the accuracy of the ML model excluding family based features is 73.1%, 80.5%, and 87.1%, when all background and clustered AE events are considered.

Table 1
Summary of Classification Results With RF Networks for All Events as Well as Separated Background and Clustered Events

RF network						Confusion matrix[b] (%)		
Population used	Family features Number of features	Accuracy (%)	Precision (%)	Recall (%)	ROC[a] (%)	PP	PN	
Entire catalog	Not included	73.1	73.6	82.3	80.4	72	26	P
	75					28	74	N
	Included	73.2	74.7	80.3	80.8	71	25	P
	81					29	75	N
Only background AEs	Not included	80.5	83.9	85.3	87.9	76	17	P
	75					24	83	N
	Included	82.1	84.8	86.6	89.1	77	15	P
	81					23	85	N
Only clustered AEs	Not included	87.1	88.6	95.2	91.7	83	11	P
	75					17	89	N
	Included	**89.0**	**90.0**	**96.1**	**92.8**	**85**	**10**	P
	81					**15**	**90**	N
Entire catalog	conventional time features[c]	54.7	41.7	15.4	62.4	80	58	P
	47					20	42	N

Note. For details about metrics see Text S4 in Supporting Information S1. Bold font indicates best result. [a]ROC: Area under curve (AUC) of receiver operating characteristic (ROC). [b]P: positive class, N: negative class, PN: predicted negative and PP: predicted positive. [c]The features are the same as Karimpouli et al. (2023). No family features are used.

5. Discussion

We present a new event-based approach to extract physics-informed seismo-mechanical features according to origin time and location of AE events generated during laboratory rock deformation stick-slip experiments on rough faults. We separate the input catalog data into clustered and background seismicity and include new event-based family features based on the topology of clustered event families. The association of feature vectors allocated to each AE event allows training individual RF models using background and clustered subcatalogs separately, which significantly improves the accuracy of labquake forecasting. A resulting suite of catalog-driven features allows capturing the spatiotemporal complexity of the evolving fault stress and damage on and in the vicinity of the fault surface with progressive loading. This constitutes a crucial improvement, since the preparation of large earthquakes is conditioned on the temporal and spatial evolution of the stress field and damage over the fault zone, as indicated in laboratory (Kwiatek et al., 2024) and field studies (Kato & Ben-Zion, 2020). In addition, it provides a useful framework, where feature vectors are associated with the individual events, enabling the use of more elaborated ML networks with sophisticated architecture. Finally, the enriched pool with family features evolving during the preparatory phase provide important information about the spatiotemporal evolution of clustered AE events related to the lifecycle of the cm-scale asperities (cf. discussion in Kwiatek et al. (2024).

Our results show a forecasting accuracy of 82.1% and 89.0% for background and clustered events, respectively. We compare our results with two similar studies for experimental lab-quakes and tectonic earthquakes. Using the same experimental data, Karimpouli et al. (2023) extracted 47 time series of various features in time and employed an ensemble long-short-term memory (LSTM) model for TTF prediction. To have a representative comparison, we used a new RF model based on the same 47 time-based features employed by Karimpouli et al. (2023) to forecast the existence of a large event ($M_{AE} > 3.5$) in the next forecasting window (similar to Figure 5). The results are reported in Table 1, where the model accuracy, precision, recall, and receiver operating characteristic (ROC) are 54.8%, 41.7%, 15.4%, and 62.4%. These metrics and the values of the confusion matrix clearly show that this is not a well-trained model where at least one class is detected randomly. This indicates that with our event-based features, the model is better trained due to a higher number of training data, and the fact that different aspects and scales of the labquake processes are more efficiently covered (in time and space), leading to
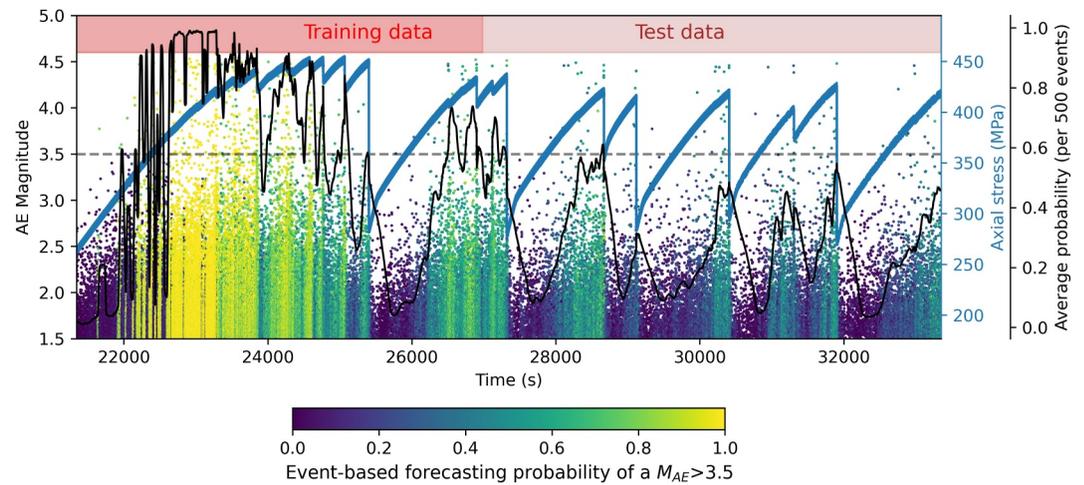
Figure 6. Temporal evolution of Acoustic Emission (AE) magnitude of AE event in sample WgN07 (see Figure 1) color coded by probability of a future large event (>3.5, dashed line) in the next time window (20 s) after each event generated by RF-models. The blue curve is the axial stress values and the black curve is the average probability per 500 events with 50 events step).

more reliable results. The other study is a real-time earthquake forecasting, where Saad et al. (2023) used EM and GA data of the current week and extracted 51 and 44 statistical features, respectively. Using a RF model, they obtained a 70% accuracy for forecasting large earthquakes in the next week. Statistical features from different sources of data (EM and GA) helped them to reach 70% accuracy in such a complex region. To separate earthquakes with different sources, Saad et al. (2023) divided the whole region into six subregions and for each subregion used two individual RF models for each type of data (i.e., EM and GA). The main idea behind this was to leverage individual models trained on a localized region with similar statistical behavior instead of having one general model for the entire region. This is technically very similar, in this study, to the separation of all events into background and cluster populations. Such a separation was also used by Picozzi et al. (2023) who found different patterns in the features relative to background and clustered seismicity that occurred prior to the L'Aquila earthquake on 6 April 2009 in central Italy.

To investigate the performance of the ML model, the probabilities of occurrence of a future large AE event ($M_{AE} > 3.5$) in the next 20 s are generated by two RF models (depending on the considered event being background or clustered) for each AE of sample WgN07 (Figure 6), which contains both training and test data. It is not surprising that the probability values for the training data are higher than the test data, as explained in Section 4.2. The probabilities of test data are high enough (>0.5) to be classified as 1 for forecasting a future large labquake. For tectonic earthquakes, these probabilities can be used to define a threshold for an alarming system, possibly with different threshold values for different subregions as in Saad et al. (2023). The probabilities indicate increasing trends within the emerging big events in a cycle. This is shown clearly by the black curve that gives the average probability in moving windows of 500 events. The average probability increases in the about first 2/3 of each large event cycle and then fluctuates at high values until the occurrence of a large event. The results are consistent with an *intermittent criticality* process (Ben-Zion et al., 2003; Sammis & Smith, 1999; see discussion in Kwiatek et al. (2024) for WgN05), where the stress (blue line) evolves to a critical state and then fluctuates until a large event occurs as a triggered statistical event (see also Figure 1). Since the model was trained only for forecasting of large AE events ($M_{AE} > 3.5$), we do not expect to see a clear trend for proximity to large or small stress drops similar to TTF prediction (Karimpouli et al., 2023). This is because large AE with M_{AE} exceeding 3.5 accompany both small slips activating a limited portion of the fault surface with barely visible stress drops as well as the slips over the entire surfaces leading to large stress drops (Kwiatek et al., 2024).

The improvement observed in the forecasting accuracy from separating background and clustered events suggests that including time-space features effectively isolates information from physical processes occurring at different temporal and spatial scales. AE events forming a background activity represent isolated events scattered across the fault surface. With increasing axial stress, the background seismicity rates steadily grow in response to the increased overall contact between the two faces of the fault. However, clustered AE activity is transient in time

and constrained in space and signifies the processes on cm-scale asperities. As presented in Kwiatek et al. (2024), the collective evolution and interaction of asperities at the high level of axial stress and depicted predominantly by clustered AE activity enable conditions for the major approaching failure.

We also computed feature importance in an RF model on clustered events (see Text S5 and Figure S2 in Supporting Information S1) to find which features, among all 81 features, are the keys for forecasting. Results revealed that all family features, AE-rate, scaled total magnitude, and *b*-value with different time-space windows are found to be the 10 most important features. During the preparatory phase preceding an earthquake, the seismicity rate (here AE-rate) and seismic energy increase in response to enhanced fault damage (Kwiatek et al., 2024). Similarly, as progressively more and larger magnitude events occur, the *b*-value decreases due to increased local stresses and coalescence of fractures (Main, 1991; Scholz, 2019). However, the RF model also uses family features revealing their importance in the performed classification. As the detailed spatiotemporal evolution of cm-scale asperities is reflected in these features and the asperities play an important role in the preparatory processes of large slip events (Kwiatek et al., 2024), we hypothesize that the new features contain important information about the preparation of the system-size slip. It should be noted that being important does not mean being sufficient. To explore the amount of information added by the family features, we re-train similar RF models excluding them from the training. Table 1 shows that family features do not add any information when whole catalog data are used. In the separated models, the accuracy of the classification increases, for example, up to about 2% for clustered events. A comparison reveals that the effect of event separation is much higher than the effect of family features.

Scaling-up of labquake studies to tectonic earthquakes is an important unresolved subject of continuing research, which can be discussed in different aspects (e.g., Ben-Zion & Dresen, 2022; McGarr et al., 2010). With regard to the application of our ML framework, we first address the *problem (output) definition*. We defined a classification problem with the aim of providing probability-like information on whether in the next forecasting window there will be a large AE event. This way of posing the problem allows not only to implement it in (near) real-time monitoring and forecasting but also to qualitatively compare the outcome of such forecasting with the classical probabilistic earthquake hazard assessment (PSHA) procedures. In the latter case, the time-and-space constrained catalog is used to calculate the probability of magnitude exceedance in the next time period assuming that the seismicity rates are approximately stationary. As for *model selection*, we find that RF models offer several advantages, including resilience to overfitting, scalability to high-dimensional data sets, fast training, and interpretability. These characteristics make the RF appropriate forecasting models that may potentially be applied to experimental (Rouet-Leduc et al., 2017) and real-time case studies (Saad et al., 2023). However, the performance of the developed RF-based forecasting framework and other forecasting approaches has yet to be tested with field data. Finally, it is important that the role of *input data* in the processes governing the runup to seismic events is understood. To this end, we employed an ensemble of physics-informed event-based features that are extractable from the seismicity catalog and could potentially provide an improved and data-driven information to forecast upcoming earthquakes. The proposed approach allows embedding new physics-based features that are attributed to each AE event. Importantly, the entire ML framework in this study is scalable and could be implemented for tectonic earthquakes. However, the applicability of the approach also depends on other aspects such as resolution of the seismicity catalog, availability of historical data, complexity of the tectonic region and so on.

A clear limitation of the proposed framework is the availability of a high-resolution seismicity catalog. Feature computation, especially in the time-space window requires a substantial number of events. However, current trends in ML-based earthquake processing (Mousavi et al., 2019) and high-resolution monitoring could make these approaches more applicable. The results of our study are based on experimental data, and the applicability of the proposed framework should be evaluated using other data sets at different scales. A general forecasting model needs to be trained on a large data set covering a full variability range of each feature for all possible scenarios and complexities of the earthquake in each region. This may be achievable for tectonic earthquakes if enough historical data are available and/or high-resolution data are available.

This study introduced a novel avenue to earthquake forecasting, and we assume it as a starting point which could be followed and highly investigated in the near future. For example, the background and clustered events separation is a statistical approach and we used GMM for such a purpose. However, Aden-Antoniów et al. (2022) introduced adaptive RF for background and cluster separation, which seems more efficient. The sensitivity of

earthquake forecasting to the separation methods is not known and could be potentially investigated. In addition, we defined and used three exemplary family based features. Other features such as branching index, average leaf depth, and many others, for example, introduced by Zaliapin and Ben-Zion (2013b) could also be applied.

6. Conclusions

We analyze AE data recorded during laboratory triaxial stick-slip experiments on three Westerly granite samples with rough faults. Unlike conventional methods, we present an event-based approach to extract numerous spatiotemporal seismo-mechanical and physics-informed features characterizing damage and stress evolution in the sample in the spatiotemporal windows attributed to each AE event. With the aim of near-real-time forecasting of the large labquakes, we reformulate the TTF prediction into a classification problem, where the computed features are used, through the RF classifier, to forecast the occurrence of a large magnitude event in the following short time window.

Separating the AE population into the background and clustered events allowed us to obtain a more accurate forecasting model while using the full catalog. Feature importance analysis revealed that not only AE-rate, seismic energy, and *b*-value are important but also family features developed from a topological cluster decomposition, playing a crucial role for labquake forecasting. The results suggest that event-based catalog-driven features, including clustering analysis provide informative input data, which can be used to improve forecasts of large tectonic earthquakes.

Data Availability Statement

Seismic catalogs, moment tensor catalogs, raw waveform data, geomechanical data, and associated information related to stick-slip experiments analyzed in this study are available at GFZ Data Services via separate data publication (CC-BY 4.0 license): Kwiatek and Goebel (2024) (https://doi.org/10.5880/GFZ.4.2.2023.003).

References

Aden-Antonióv, F., Frank, W. B., & Seydoux, L. (2022). An adaptable random forest model for the declustering of earthquake catalogs. *Journal of Geophysical Research: Solid Earth*, *127*(2), e2021JB023254. https://doi.org/10.1029/2021JB023254

Baiesi, M., & Paczuski, M. (2005). Complex networks of earthquakes and aftershocks. *Nonlinear Processes in Geophysics*, *12*(1), 1–11. https://doi.org/10.5194/NPG-12-1-2005

Ben-Zion, Y. (2008). Collective behavior of earthquakes and faults: Continuum-discrete transitions, progressive evolutionary changes, and different dynamic regimes. *Reviews of Geophysics*, *46*(4). https://doi.org/10.1029/2008RG000260

Ben-Zion, Y., & Dresen, G. (2022). A synthesis of fracture, friction and damage processes in earthquake rupture zones. *Pure and Applied Geophysics*, *179*(12), 4323–4339. https://doi.org/10.1007/S00024-022-03168-9

Ben-Zion, Y., Eneva, M., & Liu, Y. (2003). Large earthquake cycles and intermittent criticality on heterogeneous faults due to evolving stress and seismicity. *Journal of Geophysical Research*, *108*(B6), 2307. https://doi.org/10.1029/2002JB002121

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. *Seismological Research Letters*, *90*(3), 1088–1098. https://doi.org/10.1785/0220180367

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. *Nature Communications*, *14*(1), 1–12. https://doi.org/10.1038/s41467-023-39377-6

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324/METRICS

Corbi, F., Bedford, J., Sandri, L., Funiciello, F., Gualandi, A., & Rosenau, M. (2020). Predicting imminence of analog megathrust earthquakes with machine learning: Implications for monitoring subduction zones. *Geophysical Research Letters*, *47*(7), e2019GL086615. https://doi.org/10.1029/2019GL086615

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

Cornell, C. A. (1968). Engineering seismic risk analysis. *Bulletin of the Seismological Society of America*, *58*(5), 1583–1606. https://doi.org/10.1785/BSSA0580051583

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, *39*(1), 1–22. https://doi.org/10.1111/J.2517-6161.1977.TB01600.X

Dresen, G., Kwiatek, G., Goebel, T., & Ben-Zion, Y. (2020). Seismic and aseismic preparatory processes before large stick–slip failure. *Pure and Applied Geophysics*, *177*(12), 5741–5760. https://doi.org/10.1007/S00024-020-02605-X/FIGURES/8

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. *Journal of Geophysical Research*, *117*(B3), 3310. https://doi.org/10.1029/2011JB008763

Goebel, T. H. W., Schorlemmer, D., Becker, T. W., Dresen, G., & Sammis, C. G. (2013). Acoustic emissions document stress changes over many seismic cycles in stick-slip experiments. *Geophysical Research Letters*, *40*(10), 2049–2054. https://doi.org/10.1002/GRL.50507

Hulbert, C., Rouet-Leduc, B., Johnson, P. A., Ren, C. X., Rivière, J., Bolton, D. C., & Marone, C. (2018). 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

Jasperson, 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. *Journal of Geophysical Research: Solid Earth*, *126*(11), e2021JB022195. https://doi.org/10.1029/2021JB022195

Johnson, P. A., Rouet-Leduc, B., Pyrak-Nolte, L. J., Beroza, G. C., Marone, C. J., Hulbert, C., et al. (2021). Laboratory earthquake forecasting: A machine learning competition. *Proceedings of the National Academy of Sciences*, *118*(5), e2011362118. https://doi.org/10.1073/PNAS.2011362118

Karimpouli, S., Caus, D., Grover, H., Martínez-Garzón, P., Bohnhoff, M., Beroza, G. C., et al. (2023). Explainable machine learning for labquake prediction using catalog-driven features. *Earth and Planetary Science Letters*, *622*, 118383. https://doi.org/10.1016/J.EPSL.2023.118383

Karimpouli, S., Kwiatek, G., Martínez-Garzón, P., Dresen, G., & Bohnhoff, M. (2024). Unsupervised clustering of catalogue-driven features for characterizing temporal evolution of labquake stress. *Geophysical Journal International*, *237*(2), 755–771. https://doi.org/10.1093/GJI/GGAE071

Kato, A., & Ben-Zion, Y. (2020). The generation of large earthquakes. *Nature Reviews Earth & Environment*, *2*(1), 26–39. https://doi.org/10.1038/s43017-020-00108-w

Kwiatek, G., & Goebel, T. (2024). Acoustic Emission and Seismic moment tensor catalogs associated with the triaxial stick-slip experiment performed on the Westerly Granite Sample. *GFZ Data Services*. https://doi.org/10.5880/GFZ.4.2.2023.003

Kwiatek, G., Martínez-Garzón, P., Goebel, T., Bohnhoff, M., Ben-Zion, Y., & Dresen, G. (2024). Intermittent criticality multi-scale processes leading to large slip events on rough laboratory faults. *Journal of Geophysical Research: Solid Earth*, *129*(3), e2023JB028411. https://doi.org/10.1029/2023JB028411

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 and Planetary Science Letters*, *598*, 117825. https://doi.org/10.1016/J.EPSL.2022.117825

Licciardi, A., Bletery, Q., Rouet-Leduc, B., Ampuero, J. P., & Juhel, K. (2022). Instantaneous tracking of earthquake growth with elastogravity signals. *Nature 2022*, *606*(7913), 319–324. https://doi.org/10.1038/s41586-022-04672-7

Lubbers, N., Bolton, D. C., Mohd-Yusof, J., Marone, C., Barros, K., & Johnson, P. A. (2018). Earthquake catalog-based machine learning identification of laboratory fault states and the effects of magnitude of completeness. *Geophysical Research Letters*, *45*(24), 13269–13276. https://doi.org/10.1029/2018GL079712

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. *Geophysical Journal International*, *107*(2), 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. *Annual Review of Earth and Planetary Sciences*, *26*(1), 643–696. https://doi.org/10.1146/ANNUREV.EARTH.26.1.643

McGarr, A., Fletcher, J. B., Boettcher, M., Beeler, N., & Boatwright, J. (2010). Laboratory-based maximum slip rates in earthquake rupture zones and radiated energy. *Bulletin of the Seismological Society of America*, *100*(6), 3250–3260. https://doi.org/10.1785/0120100043

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. *Scientific Reports*, *9*(1), 10267. https://doi.org/10.1038/s41598-019-45748-1

Picozzi, M., & Iaccarino, A. G. (2021). Forecasting the preparatory phase of induced earthquakes by recurrent neural network. *Forecasting 2021*, *3*(1), 17–36. https://doi.org/10.3390/FORECAST3010002

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*. https://doi.org/10.1109/CEC.2018.8477769

Ross, Z. E., Meier, M. A., Hauksson, E., & Heaton, T. H. (2018). Generalized seismic phase detection with deep learning. *Bulletin of the Seismological Society of America*, *108*(5A), 2894–2901. https://doi.org/10.1785/0120180080

Rouet-Leduc, B., Hulbert, C., Bolton, D. C., Ren, C. X., Riviere, J., Marone, C., et al. (2018). Estimating Fault friction from seismic signals in the laboratory. *Geophysical Research Letters*, *45*(3), 1321–1329. https://doi.org/10.1002/2017GL076708

Rouet-Leduc, B., Hulbert, C., 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

Saad, O. M., Chen, Y., Savvaidis, A., Fomel, S., Jiang, X., Huang, D., et al. (2023). Earthquake forecasting using big data and artificial intelligence: A 30-week real-time case study in China. *Bulletin of the Seismological Society of America*, *113*(6), 2461–2478. https://doi.org/10.1785/0120230031

Sammis, C. G., & Smith, S. W. (1999). Seismic cycles and the evolution of stress correlation in cellular automaton models of finite fault networks. In *Seismicity patterns, their statistical significance and physical meaning* (pp. 307–334). https://doi.org/10.1007/978-3-0348-8677-2_6

Scholz, C. H. (2019). *The mechanics of earthquakes and faulting*. Cambridge University Press. Retrieved from https://books.google.de/books?hl=en&lr=&id=xa-BDwAAQBAJ&oi=fnd&pg=PR11&ots=DPY8MncSzv&sig=3-4S5SQB6j547PONZ0qQmhoO514&redir_esc=y#v=onepage&q&f=false

Shokouhi, P., Girkar, V., Rivière, J., Shreedharan, S., Marone, C., Giles, C. L., & Kifer, D. (2021). Deep learning can predict laboratory quakes from active source seismic data. *Geophysical Research Letters*, *48*(12), e2021GL093187. https://doi.org/10.1029/2021GL093187

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. *Journal of Geophysical Research: Solid Earth*, *126*(7), e2020JB021588. https://doi.org/10.1029/2020JB021588

Wang, K., Johnson, C. W., Bennett, K. C., & Johnson, P. A. (2021). Predicting fault slip via transfer learning. *Nature Communications*, *12*(1), 1–11. https://doi.org/10.1038/s41467-021-27553-5

Wang, K., Johnson, C. W., Bennett, K. C., & Johnson, P. A. (2022). Predicting future laboratory fault friction through deep learning transformer models. *Geophysical Research Letters*, *49*(19), e2022GL098233. https://doi.org/10.1029/2022GL098233

Wiemer, S., & Wyss, M. (2000). Minimum magnitude of completeness in earthquake catalogs: Examples from Alaska, the Western United States, and Japan. *Bulletin of the Seismological Society of America*, *90*(4), 859–869. https://doi.org/10.1785/0119990114

Zaliapin, I., & Ben-Zion, Y. (2013a). Earthquake clusters in southern California I: Identification and stability. *Journal of Geophysical Research: Solid Earth*, *118*(6), 2847–2864. https://doi.org/10.1002/JGRB.50179

Zaliapin, I., & Ben-Zion, Y. (2013b). Earthquake clusters in southern California II: Classification and relation to physical properties of the crust. *Journal of Geophysical Research: Solid Earth*, *118*(6), 2865–2877. https://doi.org/10.1002/JGRB.50178

Zaliapin, I., & Ben-Zion, Y. (2016). A global classification and characterization of earthquake clusters. *Geophysical Journal International*, *207*(1), 608–634. https://doi.org/10.1093/GJI/GGW300

Zaliapin, I., Gabrielov, A., Keilis-Borok, V., & Wong, H. (2008). Clustering analysis of seismicity and aftershock identification. *Physical Review Letters*, *101*(1), 018501. https://doi.org/10.1103/PHYSREVLETT.101.018501/FIGURES/5/MEDIUM

References From the Supporting Information

Hand, D. J., & Till, R. J. (2001). A simple generalisation of the Area under the ROC curve for multiple class classification problems. *Machine Learning*, *45*(2), 171–186. https://doi.org/10.1023/A:1010920819831/METRICS

Olson, D. L., & Delen, D. (2008). Advanced data mining techniques. In *Advanced data mining techniques* (pp. 1–180). https://doi.org/10.1007/978-3-540-76917-0/COVER

Patel, N., & Applications, S. U.-I. Journal of Computer, & 2012, Undefined. (2012). Study of various decision tree pruning methods with their empirical comparison in WEKA. *Citeseer*, *60*(12), 975–8887. Retrieved from https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=025b8c109c38dc115024e97eb0ede5ea873fffdb

van der Elst, N. J. (2021). B-positive: A robust estimator of aftershock magnitude distribution in transiently incomplete catalogs. *Journal of Geophysical Research: Solid Earth*, *126*(2), e2020JB021027. https://doi.org/10.1029/2020JB021027

Zaccagnino, D., Telesca, L., & Doglioni, C. (2023). Global versus local clustering of seismicity: Implications with earthquake prediction. *Chaos, Solitons & Fractals*, *170*, 113419. https://doi.org/10.1016/J.CHAOS.2023.113419