

Unsupervised clustering of catalogue-driven features for characterizing temporal evolution of labquake stress

Sadegh Karimpouli,¹ Grzegorz Kwiatek,¹ Patricia Martínez-Garzón,¹ Georg Dresen^{1,2} and Marco Bohnhoff^{1,3}

¹Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, 14473 Potsdam, Germany. E-mail: sadegh.karimpouli@gfz-potsdam.de

²Institute of Earth and Environmental Sciences, Universität Potsdam, 14476 Potsdam, Germany

³Department of Earth Sciences, Free University Berlin, 14195 Berlin, Germany

Accepted 2024 February 27. Received 2024 February 27; in original form 2023 August 18

SUMMARY

Earthquake forecasting poses significant challenges, especially due to the elusive nature of stress states in fault systems. To tackle this problem, we use features derived from seismic catalogues obtained from acoustic emission (AE) signals recorded during triaxial stick-slip experiments on natural fractures in three Westerly granite samples. We extracted 47 physically explainable features from AE data that described spatio-temporal evolution of stress and damage in the vicinity of the fault surface. These features are then subjected to unsupervised clustering using the K-means method, revealing three distinct stages with a proper agreement with the temporal evolution of stress. The recovered stages correspond to the mechanical behaviour of the rock, characterized as initial stable (elastic) deformation, followed by a transitional stage leading to an unstable deformation prior to failure. Notably, AE rate, clustering-localization features, fractal dimension, *b*-value, interevent time distribution, and correlation integral are identified as significant features for the unsupervised clustering. The systematically evolving stages can provide valuable insights for characterizing preparatory processes preceding earthquake events associated with geothermal activities and waste-water injections. In order to address the upscaling issue, we propose to use the most important features and, in case of normalization challenge, removing non-universal features, such as AE rate. Our findings hold promise for advancing earthquake prediction methodologies based on laboratory experiments and catalogue-driven features.

Key words: Machine learning; Acoustic properties; Earthquake interaction, forecasting and prediction.

1 INTRODUCTION

The problem of earthquake forecasting holds significant importance for society, economy and geoscience. However, fault constitutive properties and stress evolution vary in space and time and are not directly observable in the field (Ben-Zion *et al.* 2003). This renders earthquake forecasting a notoriously difficult challenge. Across the seismogenic zone of a fault zone, the evolution of stress and strength is closely correlated with large earthquakes. Kato & Ben-Zion (2020) proposed that the occurrence of large earthquakes is associated with the progressive localization of shear deformation around a rupture zone, sometimes associated with the increase in seismicity rates and decrease in the Gutenberg–Richter *b*-value (Ben-Zion & Lyakhovsky 2002; Bowman & Sammis 2004; Mignan *et al.* 2011; Durand *et al.* 2020; Kwiatek *et al.* 2024). Eventually, this transitions into rapid loading (i.e. foreshocks) within a localized crustal volume near the hypocentre of an ensuing major dynamic

rupture. The critical phase usually occurs at some stress saturation level, the *intermittent criticality* phase (Sammis & Smith 1999; Ben-Zion *et al.* 2003), where stress fluctuates around a critical peak stress, and a large earthquake occurs as a triggered, statistical event. Following a stress drop, the seismic loading cycle resumes. Identifying distinct phases of stress evolution may provide a significant step forward in improving earthquake forecasting and seismic hazard assessment.

Recent advances in machine learning (ML) algorithms helped improving time to failure (TTF) prediction (Rouet-Leduc *et al.* 2017), shear stress estimation (Shokouhi *et al.* 2021; Shreedharan *et al.* 2021; Borate *et al.* 2023), preparatory phase identification (Picozzi & Iaccarino 2021) and clustering of stress temporal evolution (Bolton *et al.* 2019) using laboratory and field data. In these studies, seismic waveforms are the only available data. Provided seismic waveforms contain enough signatures of temporal evolution of the stress, informative patterns and high-level information,

known as seismic features, may be extracted computationally or using ML methods. Seismic features could be broadly categorized as: (1) waveform-driven features and (2) catalogue-driven features. Waveform-driven features are obtained either directly using various statistical measures (Rouet-Leduc *et al.* 2017) or indirectly by ML networks such as convolutional neural networks (Mousavi *et al.* 2019b). Some of the statistical features are the mean, variance, skewness and kurtosis of signal distribution, time correlation features using integrals of power spectra over narrow frequency bands and autocorrelation features. In the context of a double-friction shear test, Bolton *et al.* (2019) employed waveform-driven statistical features and demonstrated that the selected features exhibit temporal behaviour that allows for the detection of evolutionary phases of the stress curve using mean-shift and K-means clustering algorithms.

Catalogue-driven features are typically used to characterize seismic response to deformation, such as event/seismic moment or radiated energy rate, b -value, correlation integral, interevent time distribution features, fractal dimension, magnitude correlation, Shannon's entropy, clustering and localization features, energy index and features based on focal mechanisms and deviatoric stress tensor (Kwiatiek *et al.* 2024). Picozzi & Iaccarino (2021) employed a Recurrent Neural Network (RNN) to identify the preparatory phase of magnitude $M \sim 4$ induced earthquakes, showcasing the potential of catalogue-driven features and RNN for forecasting large induced seismic events. Similar patterns of damage evolution have been observed through the analysis of acoustic emissions (AEs) during triaxial tests on rock samples (Dresen *et al.* 2020). Karimpouli *et al.* (2023) computed catalogue-driven features at various time windows, and utilized them as inputs for an ensemble of long-short term memory (LSTM) networks to predict TTF in three stick-slip experiments on rough faults. Using an explainable ML approach, they found event rate, correlation integral, event proximity and focal mechanism-based features to be the most important features for TTF prediction.

These studies demonstrated that catalogue-driven features provide valuable information showing different patterns during each phase of stress evolution. However, the high number of features require an unsupervised clustering method for unbiased detection of data structures in an n -dimensional space of features (Beroza *et al.* 2021). Unsupervised deep-ML algorithms have been used for different aspects such as clustering events signals and background noises (Seydoux *et al.* 2020; Aden-Antoniów *et al.* 2022), discriminate different local and teleseismic seismic events (Mousavi *et al.* 2019a), repeating earthquake clustering (Huang *et al.* 2023), induced seismicity (Yoon *et al.* 2015; Holtzman *et al.* 2018) and volcanic tremor classification (Soubestre *et al.* 2018). To investigate preparatory processes, Shi *et al.* (2021) used continuous seismic data and hierarchical clustering of waveform-driven features. They found a systematic temporal evolution during the 2009 L'Aquila earthquake. Iaccarino & Picozzi (2023) used K-means clustering for induced seismicity at The Geysers geothermal site and Zali *et al.* (2024) employed deep-ML for volcanic tremors. Using an autoencoder network, they extracted high level features out of the waveform and unsupervisedly clustered them by K-means method. Results show evolutionary phases preceding earthquakes and volcanic eruptions.

The key motivation of this study is to investigate if an ensemble of physics-based features calculated from seismic data allows separating mechanical loading stages and help to constrain proximity to failure. In a laboratory-scale investigation, we employ catalogue-driven features developed in Kwiatek *et al.* (2024) and used in

Karimpouli *et al.* (2023) to assess the effectiveness of unsupervised clustering methods in identifying distinct loading phases concealed within the seismic features leading to failure. In contrast to previous studies, the unsupervised learning is applied to stick-slip experiments performed on rough faults, that bear more similarity with heterogeneous fault structures in nature, but make ML-based clustering more difficult (Johnson *et al.* 2021). Rough faults are more complicated yet closer to the (generally complex) faults observed in nature, whereas so far relatively simple planar faults were analysed in TTF studies, based on a multitude of repetitive stick slip cycles. We show that highly correlated features effectively construct the main trend of data similar to the stress trend allowing for a meaningful clustering. By leveraging the most significant features, we discuss the potential upscaling of the used features for characterization of preparatory processes before hazard-prone earthquakes related to exploitation of geothermal reservoirs.

2 EXPERIMENTAL DATA AND CATALOGUE-DRIVEN FEATURES

The input data originate from three triaxial stick-slip experiments WgN04, WgN05 and WgN07 performed on Westerly granite samples with a diameter of 40 mm and a length of 102–107 mm (Goebel *et al.* 2012). All specimens were prepared with a 1.5–2.2 cm deep notch inclined at an angle of 30° to control shear fracture formation. The specimens were fractured at a confining pressure of 75 MPa, resulting in the generation of rough fault planes. To conduct a series of subsequent stick-slip experiments, the confining pressure was increased to 150 MPa to lock the fault. The axial load was increased at a rate of 0.02 mm min⁻¹ or equivalently an axial strain rate of 3×10^{-6} s⁻¹. The experiments resulted in a heterogeneous and erratic slip pattern of large and small stress drops (see more details on experimental procedure and mechanical data overview in Goebel *et al.* (2012) and Kwiatek *et al.* (2024). The AE activity was recorded with 16 AE sensors (resonant frequency 2 MHz) glued to the sample surface providing optimal coverage and full waveform data at 10 MHz sampling rate and 16-bit resolution. The raw waveform data were then processed (see Kwiatek *et al.* (2024) and supporting text from Karimpouli *et al.* (2023) for details) resulting in the development of basic AE event catalogues with $n = 102\,540$, $n = 240\,328$ and $n = 199\,255$ events above the magnitude of completeness for WgN04, WgN05 and WgN07 experiments, respectively. The resulting basic AE catalogues contain information on origin times, hypocentral locations, AE magnitude and associated quality information. In addition, the full moment tensor inversion has been performed for a subset of events using hybridMT package (Kwiatek *et al.* 2016). The raw waveform data, AE sensor information and developed basic AE catalogues and seismic moment tensor catalogues are available in a separate data publication (Kwiatek & Goebel 2024). In the following, input basic AE catalogue and moment tensor catalogues were used to develop an ensemble of time series of 17 independent features calculated using running time windows of various lengths (Table 1). These features capture local and global damage and stress evolution during preparatory processes leading up to large slip events. The discussion of physical meaning of developed features is shown in Kwiatek *et al.* (2024), while a brief overview of them is presented in the Appendix A. For the overview, Fig. 1 is a visual representation of a few selected features for the WgN04 experiment (for WgN05 and WgN07 see Karimpouli *et al.* 2023 and Kwiatek *et al.* 2024, respectively). A scatterplot of independent features as well as TTF and stress values are illustrated in

Table 1. A summary of all catalogue-driven features used in this study.

Parameter		Notation	Time windows (s)	Dimension	Some references
	AE event rate	n	23, 45, 90, 180	Time	(Scholtz 1968; Dresen <i>et al.</i> 2020)
	GR law's b -value	b	10, 30, 90, 180	Time	(Gutenberg & Richter 1944)
	Correlation integral	c	90, 180 ($r = 5$ mm)* 45, 90 ($r = 20$ mm)	Space Time	(Kagan & Knopoff 1980; Henderson <i>et al.</i> 1999)
Interevent time distribution features	Ratio (deviation from uniform distribution at edges of the empirical distribution)	r	23, 45, 90, 180	Time	(Van Der Elst & Brodsky 2010; Davidsen <i>et al.</i> 2021; Kwiatek <i>et al.</i> 2022)
	Deviation from uniform distribution (χ^2 test) [§]	$rx2$			
	Fractal dimension	$d2$	45, 90, 180	Space Time	(Sadovskiy 1984; Hirata <i>et al.</i> 1987)
	Magnitude correlation	dm	90	Time	
Clustering and localization features	Product of T and R [§]	trp	25, 50, 100	Space Time Magnitude	(Baiesi & Paczuski 2005; Zaliapin <i>et al.</i> 2008; Zaliapin & Ben-Zion 2013; Martinez-Garzón <i>et al.</i> 2019)
	Quotient of T and R	trq			
	Proportion of foreshocks	pfo			
	Proportion of aftershocks	paf			
	Proportion of mainshocks	pma			
Focal mechanisms and deviatoric stress tensor based features	Median fault plane variability	vm	100, 200	Space Time	(Kwiatek <i>et al.</i> 2014, 2024; Martinez-Garzón <i>et al.</i> 2014; Vavryčuk 2014)
	Maximum principal stress plunge	sld	90, 180		
	Maximum principal stress variance	$svar$			
	Stress ratio	sr			

* r is a scale limit for spatial distance between coordination of two events.

[§] χ^2 is the Chi-squared distribution.

[§]'T' and 'R' denote the time and space components of the proximity formulation (for more details see the Appendix A).

Fig. A1 (in the Appendix A) to clarify the distribution of full data set.

Given that each feature is computed within a specific time window, the values of features estimated after slips are influenced by the data from the previous slip event. To ensure consistency and isolate individual labquake cycles, we eliminated a 150 s period following each slip, resulting in per-cycle features. Additionally, we treated every small and large slip as an independent labquake cycle. Subsequently, we normalized all features within each cycle using minimum and maximum of each feature individually and scaled between 0 and 1 (min-max normalization Karimpouli & Fattahi 2016). The rationale behind normalization is to address varying feature scales affecting distance computation between data points.

3 UNSUPERVISED CLUSTERING

The primary objective of employing unsupervised clustering methods is to organize data (features) into distinct classes based on their inherent characteristics or similarities. While humans excel at

clustering data in two or three dimensions, high-dimensional data need automated algorithms. The problem becomes more challenging when the number of clusters is unknown. Numerous algorithms have been developed to tackle the problem of clustering with unknown number of clusters (Jain 2010). To avoid this challenge, we arbitrarily decided to separate the data into three classes (we refer to them later as 'stages', emphasizing their transient, time-dependent occurrence), conceptually following the idea of 'traffic light system'. The rationale behind this clustering is that an alert could be issued based on the probability of occurrence of a specific stage and its temporal persistence before the main rupture, indicating the high seismic hazard.

Commonly used clustering methods with known class numbers include K-means clustering (Jain 2010), Gaussian mixture model (GMM; Dempster *et al.* 1977; McLachlan *et al.* 2019), hierarchical clustering (Ward 1963) and spectral clustering (Ng *et al.* 2001). Among them, K-means clustering has been proposed already more than 60 yr ago (Jain 2010) and remains widely used. Similar to this study, Bolton *et al.* (2019) applied K-means clustering algorithms to identify patterns in the AE activity during the laboratory

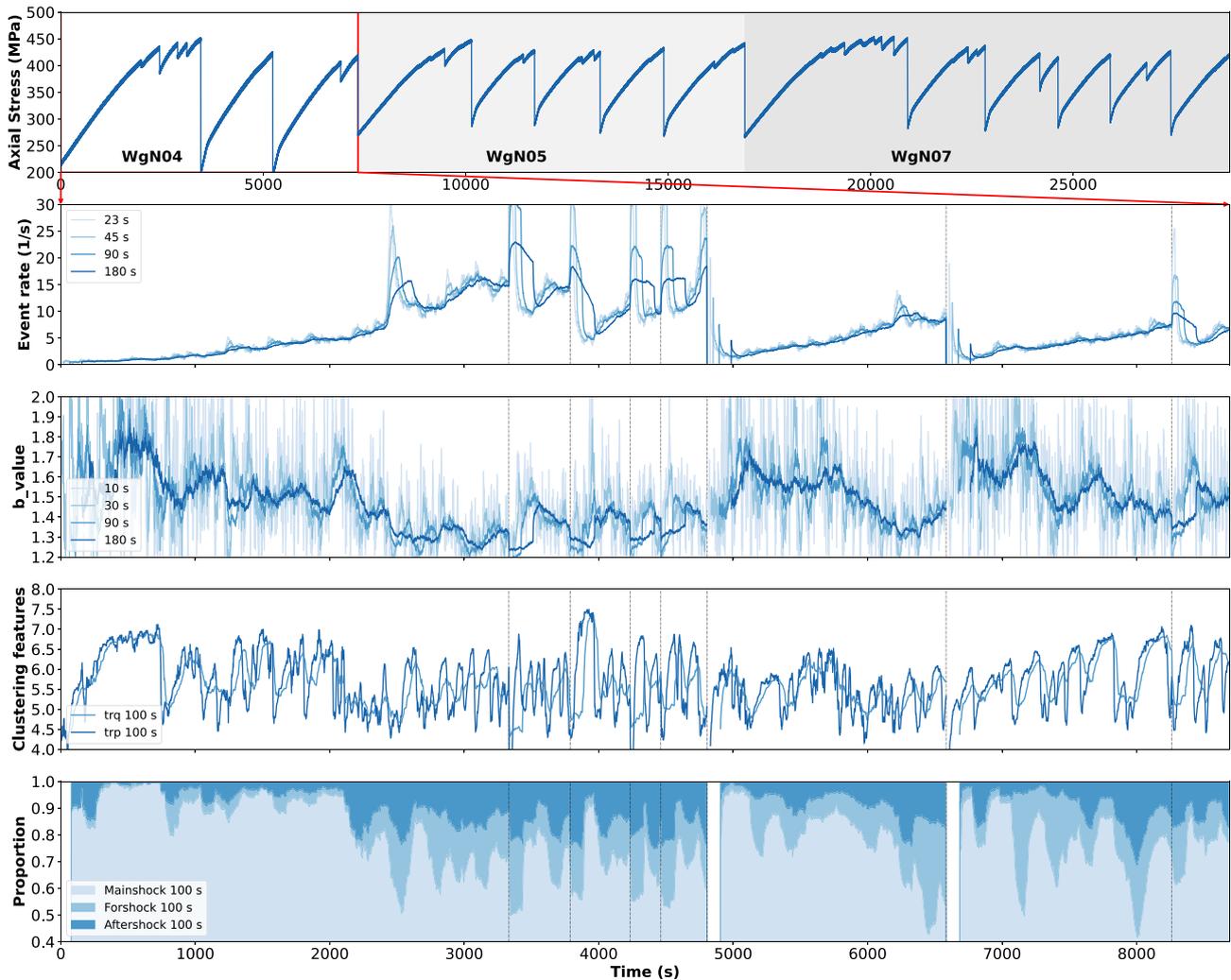


Figure 1. Temporal evolution of stress data of multiple loading cycles for all three specimens (a) following with representation of several features from sample WgN04 (b-e). Experiments have been conducted separately, but we put all data together representing here with cumulative time for illustration purposes. From top to bottom: (a) axial stress, (b) AE event rate, (c) b -value, (d) clustering features and (e) proportion of mainshocks, foreshocks and aftershocks. Each feature is computed taking the past data from a time window of selected length (indicated in s). Vertical dashed lines represent medium to large stress drops.

stick-slip cycle and precursors to labquakes. They reported consistent trends and evolutionary transitions throughout the seismic cycle. Thus, we first provide a brief overview of the fundamentals of the K-means algorithm and then other clustering methods for comparison.

1. K-means clustering: It is an iterative clustering algorithm designed to partition data into k clusters, where k is a predetermined number based on *a priori* knowledge. Let X be a set of n data points in a d -dimensional space, $X \subset \mathbb{R}^d$. The algorithm aims to select k centroids, denoted as C , that minimize the inertia or squared error:

$$\sum_{x \in X} \min_{c \in C} \|x - c\|^2. \quad (1)$$

The algorithm first calculates randomly initial k class centroids and assigns each data point to the nearest centroid. The centroids are then updated by computing the mean of the data points assigned to each cluster. This process continues iteratively until convergence, optimizing the within-class sum of squared distances. The convergence of the algorithm is heavily influenced by the

initial selection of class centroids. To address this issue, Arthur & Vassilvitskii (2007) proposed a *K-means++* variation, which initializes the centroids to be distant from each other. This approach yields improved and faster results compared to random initialization.

2. Gaussian mixture model (GMM) clustering: It is assumed that the data points are generated from a combination of k Gaussian distributions (Dempster *et al.* 1977). The algorithm estimates the parameters of the Gaussians (mean, covariance and mixing coefficients) using the Expectation–Maximization (EM) algorithm (Meng & Van Dyk 1997) and assigns probabilities to each data point belonging to each cluster.

3. Hierarchical clustering: The method builds a tree-like structure of clusters, known as a dendrogram. This can be performed in two ways: agglomerative (bottom-up) or divisive (top-down). Agglomerative clustering starts with each data point as a separate class and iteratively merges the closest pair of clusters based on a distance measure until all points belong to a single class. Divisive clustering

starts with all points in a single class and recursively splits them into smaller clusters.

4. Spectral clustering: This approach treats data points as nodes in a graph and uses the eigenvalues and eigenvectors of the graph's Laplacian matrix to partition the data. It first constructs an affinity matrix that captures pairwise similarities between data points. The eigenvectors corresponding to the smallest eigenvalues of the Laplacian matrix are then used to embed the data into a lower-dimensional space. Finally, traditional clustering algorithms like K-means are applied to the embedded space to obtain the final clusters. Spectral clustering is particularly useful for data with complex geometric structures and can handle non-convex clusters (Ng *et al.* 2001).

4 RESULTS

4.1 K-means clustering of catalogue-driven features

We started with the K-means clustering algorithm in a 47-D space using data points obtained from 47 catalogue-driven features. In this study, we utilize the *scikit-learn* library to implement all clustering algorithms in *Python*. For implementing the K-means algorithm, we used 3 numbers of clusters, with equal sample weight. To speed up the convergence, the '*k-means++*' method (Arthur & Vassilvitskii 2007) was selected as the initial class centroids. The algorithms stop when the difference between cluster centres of two consecutive iterations reaches 10^{-4} . Although the maximum number of iterations is set to 100, the convergence criterion is met at iteration 15–25, depending on the initial cluster centroids. To compute the probability of features belonging to a specific stage, we employed the inverse power of distance function with Euclidean distance and a power of 2.

In Fig. 2(a), the clustering results are illustrated overlaid on the stress curve along time for all three samples. Figs 2(b), (c) and (d) show part of the results of each individual sample together with the corresponding AE event magnitudes and the probability of each data point belonging to its respective cluster. From a rock mechanical point of view, this figure reveals that these clusters fit very well with the temporal stress evolution. This enables us to categorize them into three main stress stages, namely *stable*, *transitional* and *unstable*.

These stages include an initial stable stage, followed by a transitional stage and terminating with an unstable stage where loading cycles are finished. This sequence of stages is in excellent agreement with the different stages of the loading cycles as indicated by the stress–strain curves of the mechanical test and the stress evolution in earthquake models such as the intermittent criticality model (Ben-Zion *et al.* 2003). Accordingly, the observed stable stage may correspond to the initial elastic deformation in each cycle where a fault is dominantly locked. The transitional stage often corresponds to a stress level beyond the fault yield stress leading to some small but stable events and increasing AE activity. Finally, the unstable stage terminating each loading cycle corresponds to the final deformation stage around peak stress leading to a cascade of small and large slip events, where large AE events are associated with failure of asperities. The final slip events typically coincide with a maximum in AE activity (Figs 2b, c and d). It is also worth to be noted that we explored 4 and 5 classes, but no logical patterns among various stages were obtained, at least in our dataset. In fact, the complexity of data distribution in the feature space is much

more complex than to see similar patterns for a higher number of classes.

Overall, the unstable stage contains 28 per cent of a cycle in average for both large and small slips. For the large slips only, the average coverage of unstable stage is 29.7 per cent of the cycle with a minimum of 15.6 per cent. This means, for example, for a cycle with 1000 s length, the unstable stage is approximately detectable at least 150 s and in average 300 s before the main slip.

These results may be compared with other normalization methods and other clustering algorithms. For example, Lasocki (2014) proposed an equivalent dimension (ED) method which is a transformation of data metric into $[0, 1]$ using a non-parametric kernel-based cumulative distribution function. The main concept is that the lengths of parameter intervals are equivalent if the probability for events is the same (Lasocki 2014). We normalized all features following the ED approach and then clustered the data again using K-means algorithms. Figs 3(a) and (b) show the results for two different normalization methods: (a) min-max and (b) ED. As it is observed, the results are very similar except for some minor differences. For comparison among clustering algorithms, we use GMM, hierarchical and spectral clustering employing three clusters. For GMM, we selected all parameters similar to the K-means parameters. The algorithm stops in an epoch from 20 to 40, where the convergence tolerance is met before the maximum iteration of 100 is reached. For the hierarchical clustering, the 'ward' linkage (Ward 1963) with Euclidean distance was selected, which minimizes the variance of the clusters being merged. Since the number of clusters is known, the algorithm continues to build a full tree covering all 3 clusters. In spectral clustering, a radial basis function (RBF) kernel with Euclidean distance was used to construct the affinity matrix. Figs 3(c)–(e) shows the results of these algorithms, where the results obtained from the GMM and spectral clustering algorithms are not in agreement with the observed loading stages from stress–strain curves. In contrast, the 'Ward' hierarchical clustering algorithm, similar to K-means, yields the best agreement with the mechanical loading curves. Since the K-means algorithm is simpler to understand, widely used and computationally efficient, we use it in the remainder of the paper to investigate the feature importance.

4.2 The importance of features

The resolved stages are in good agreement with the observed mechanical stages. However, the computation of such a large number of features becomes impractical, especially when applied to field data. Hence, it is worthwhile to identify the most significant and informative features that capture the main trends within the 47-D feature space. To accomplish this, we employ a principal component analysis (PCA), which involves the eigenvalue decomposition of the covariance matrix. The first principal component (PC) accounts for only 35 per cent of the total variance in the data, and it requires at least 14 PCs to capture 85 per cent of the total variance (Fig. 4a). Since each PC represents a linear combination of all the features, the importance of each feature could be evaluated based on the feature coefficients constructing the main structure of data. Similar to regression models where the coefficients are examined to assess the contribution of each variable, we sort features within PC1 based on their coefficients (Fig. 4b). According to this figure, the 12 most significant features in PC1 are related to the AE-rate (*n*) and features characterizing clustering and localization (*trp*, *pma*, *paf*) computed in different time windows. The correlation coefficient matrix of all

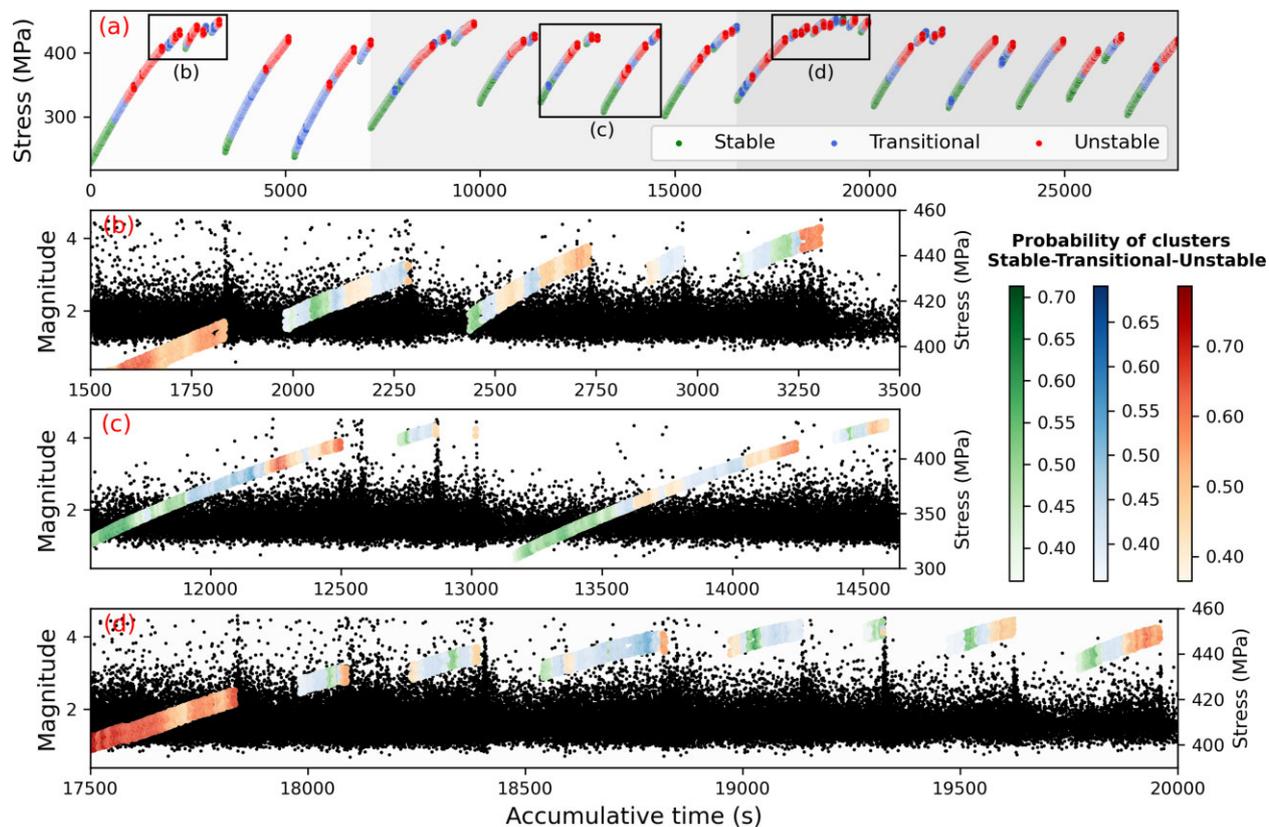


Figure 2. (a) Axial stress evolution of three samples indicated by different grey shading, colour-coded by K-means clustering results for three stages based on 47 catalogue-driven features. Temporal evolution of damage is separated into three distinct stages starting with stable (elastic) deformation, followed by a transitional stage leading to unstable deformation prior to failure. (b), (c) and (d) show three sections of individual samples, as indicated in (a) overlaid on AE magnitudes and colour coded by stage-wise probabilities.

features (Fig. 4c) reveals that these important features exhibit the highest correlations among themselves (see Fig. A1 for scatterplots of independent features). As a result, they construct the primary variation or trend within the entire 47-D feature space, which is captured by PC1. This implies that it may be possible to replicate similar results as shown in Fig. 2 by using only the most important features. Therefore, we repeat the same clustering computations, but with the 12 most important features (Fig. 5).

Comparing Figs 5 and 2 show that even with reduction in the number of features to the 12 most important ones, similar results and performance are obtained compared to the use of all features. To evaluate these clustering results with the results obtained by all 47-features, we compute the confusion matrix between them (Table 2A). Accordingly, the unstable stage is clustered accurately in 97 per cent cases even when utilizing only the 12 selected features. However, the stable and transitional stages are clustered almost interchangeably, with precisions of 71 and 74 per cent. If the goal of clustering is precisely detection of the unstable stage, such as forecasting a slip event, then using 12 features may be sufficient. Otherwise more features must be used.

4.3 Clustering using independent features

In the previous section, we showed that the 12 most important features are primarily derived from 4 independent features (n , trp , pma , paf), based on different time windows. It is worth noting that using small- and long-time windows can result in the generation

of features with different historical backgrounds (Kwiatk et al. 2024), which could be beneficial for prediction purposes (Karimpouli et al. 2023). However, in the context of clustering, it increases the influence of the highly correlated features, decreasing the effect of moderate to low correlated features (see Figs 4 and A1). Based on Fig. A1, these features are not distributed like as separate patches but as continuous trends. This means that using different time windows strengthens the role of highly correlated features in clustering leading to neglecting the effects of other constructive features such as b -value and fractal dimension (b , d_2). We now select 17 independent features according to their importance based on the coefficients in PC1, as illustrated in Fig. 4(b). According to this figure, regardless of the time window length the most important independent features, for example, trp_{100} , n_{90} , pma_{100} , paf_{100} , d_2_{180} , trq_{100} , b_{180} and so on are selected among all other features. We put this as the basis of our independent feature selection and, similar to previous section, conduct feature importance and clustering computations, illustrated in Fig. 6.

The results of the PCA (Fig. 6a) indicates that the first PC captures 36 per cent of the total variance in the data, and it still requires at least 7 additional PCs to account for 85 per cent of the variances. When sorting the features in PC1 based on their coefficients (Fig. 6b), we observe similar results to those obtained with all 47 features. Specifically, the AE-rate and clustering-localization features (n , trp , pma , paf) remain among the top 4 most important features. Furthermore, we note significant contributions of other

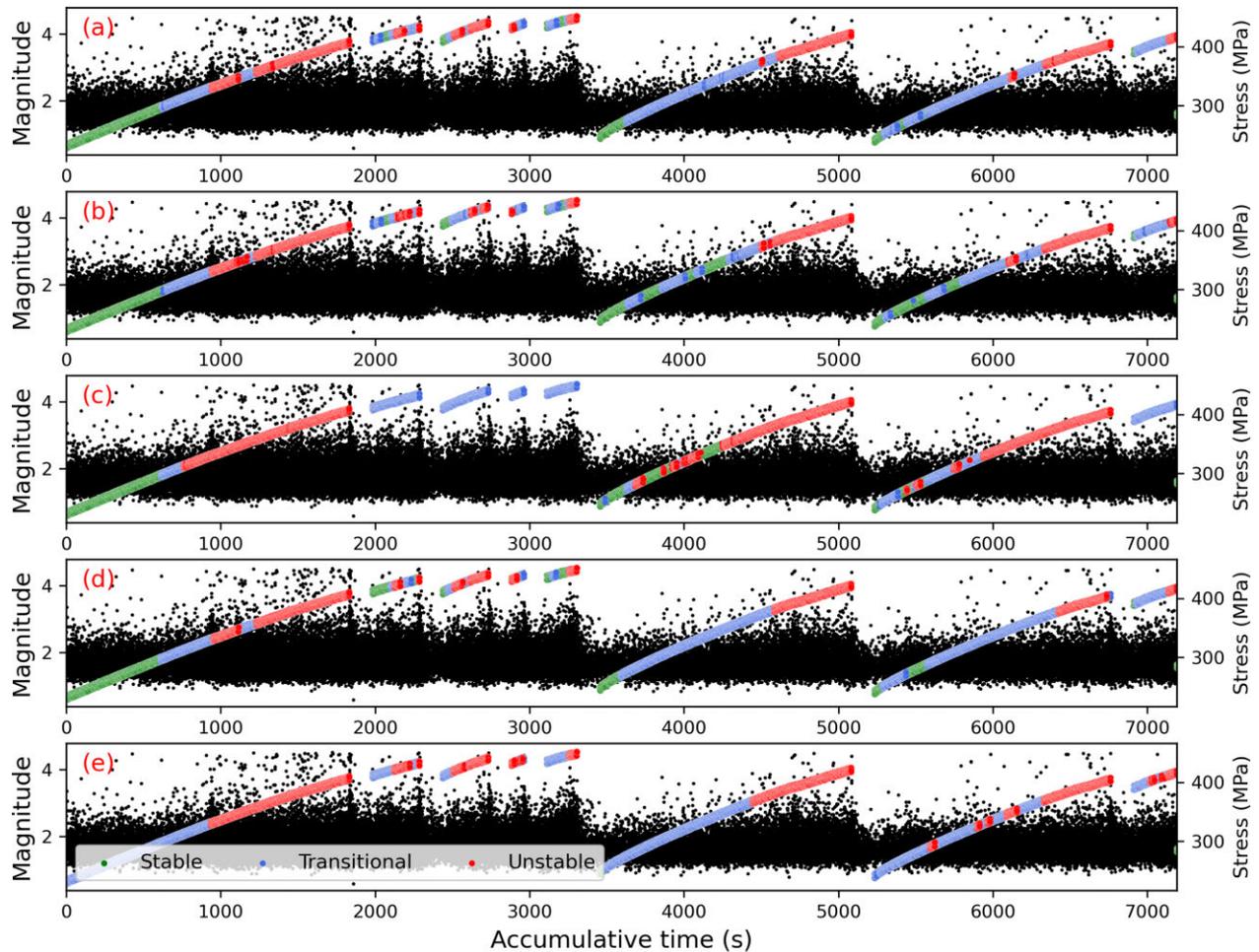


Figure 3. A comparison between K-means clustering with (a) min-max and (b) ED normalization methods. Also, these are compared with the results using other clustering algorithms such as (c) the GMM, (d) hierarchical clustering (Ward tree) and (e) spectral clustering.

features such as fractal dimension, b -value, interevent time distribution, and correlation integral (d_2 , b , rx_2 , c), as well as remaining clustering-localization features (trq and pfo). The correlation matrix of these features (Fig. 6c) reveals that by removing highly correlated features from different windows, a greater number of independent features emerge as important contributors. Fig. A1 is the counterpart of the correlation matrix showing the scatterplots of independent features.

Fig. 7 displays the K-means clustering results based on the 10 most significant independent features. Comparing these results with the clustering results obtained using all 47 features (compare Figs 7 and 2), the outcomes exhibit visual similarity, even better than the clustering outcomes achieved with the 12 most important features (Fig. 5). The corresponding confusion matrix (Table 2B) confirms that the first two stages (stable and transitional) are clustered more accurately (85 and 86 per cent of cases, respectively) compared to the previous combination (Table 2A). The correctly clustered unstable stage is only slightly reduced to 93 per cent cases.

5 DISCUSSION

Several studies have revealed that laboratory and natural earthquakes may be preceded by a preparatory process reflecting the stress and damage evolution (McLaskey & Lockner 2014; Goebel *et al.* 2017; Kato & Ben-Zion 2020; Picozzi *et al.* 2023). Monitoring

changes in various parameters associated with natural and induced earthquakes is crucial to improve forecasting of rupture events, potentially estimating TTF and/or the probability of a critical stage prior to an event. Waveform-driven and catalogue-driven features provide a more detailed view into the damage evolution process leading to dynamic failure, which is still only partly understood (Bolton *et al.* 2019; Kwiatek *et al.* 2024). Recently, ML algorithms have offered new insights in seismology and earthquake applications (see Johnson *et al.* (2021) and the reviews by Ren *et al.* (2020) and Mousavi & Beroza (2023)). In this study, we use unsupervised clustering algorithms to identify evolutionary stages signifying distinct phases of preparatory processes using 47 catalogue-driven physically explainable features. Since these features are signatures of the preparatory process, their combinations show characteristic changes allowing to relate the extracted stages to physical and mechanical characteristics of the evolving deformation process. Hence, our detailed analysis of seismic features may potentially help in upscaling applications such as improved traffic light warning systems.

Our results showed that the separation and clustering of features captures the general evolutionary steps in both large and small slip cycles (Fig. 2a). For all cycles, the stress increases from a relative minimum value up to a peak stress level and the data are normalized per-slip (see Section 2). However, in some cases the sequence of stages is mixed, and in some other cases the duration of the final unstable stage is very short relative to the other stages.

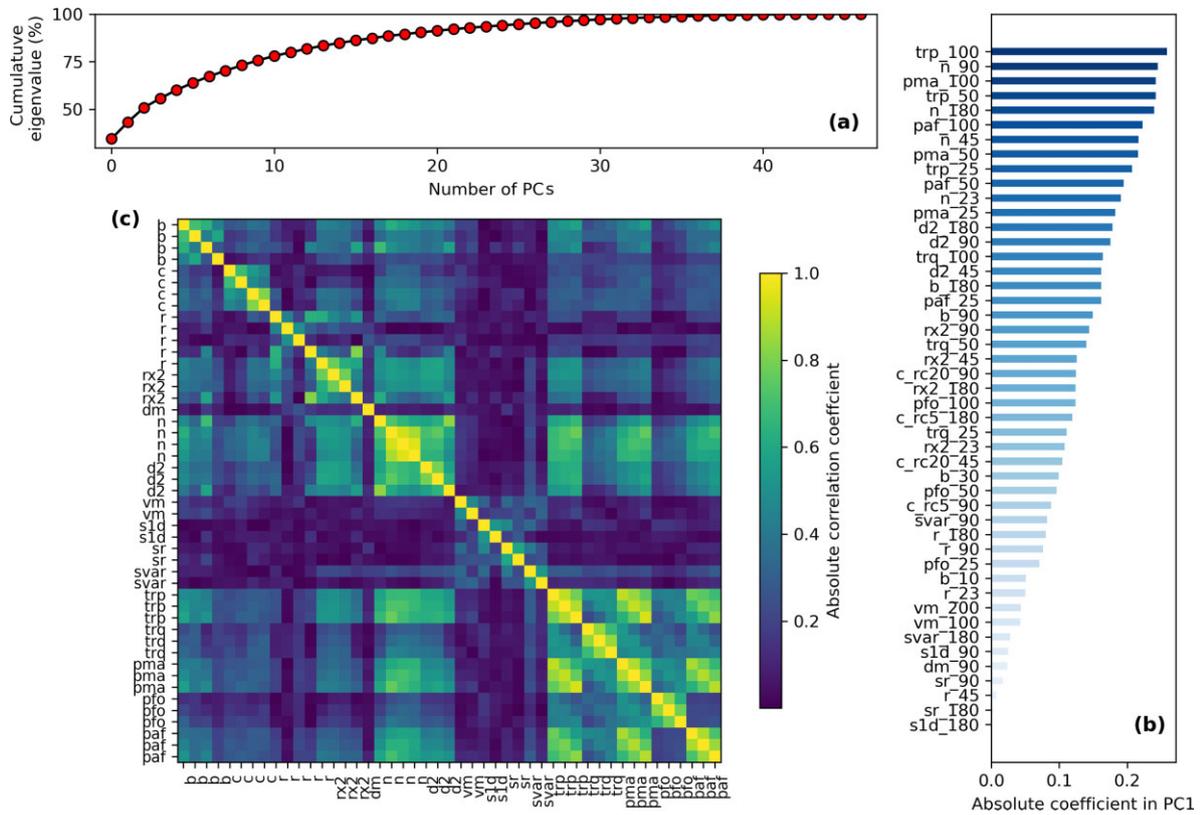


Figure 4. (a) Cumulative eigenvalue percentage versus number of PCs. At least 14 PCs are needed to cover about 85 per cent of the data variance. (b) Ranking of features based on their absolute coefficients to form PC1. (c) The correlation matrix of all 47 features. *trp* as the most important feature, shows the highest correlations with *n*, *pma* and *paf*, which are the next most important features.

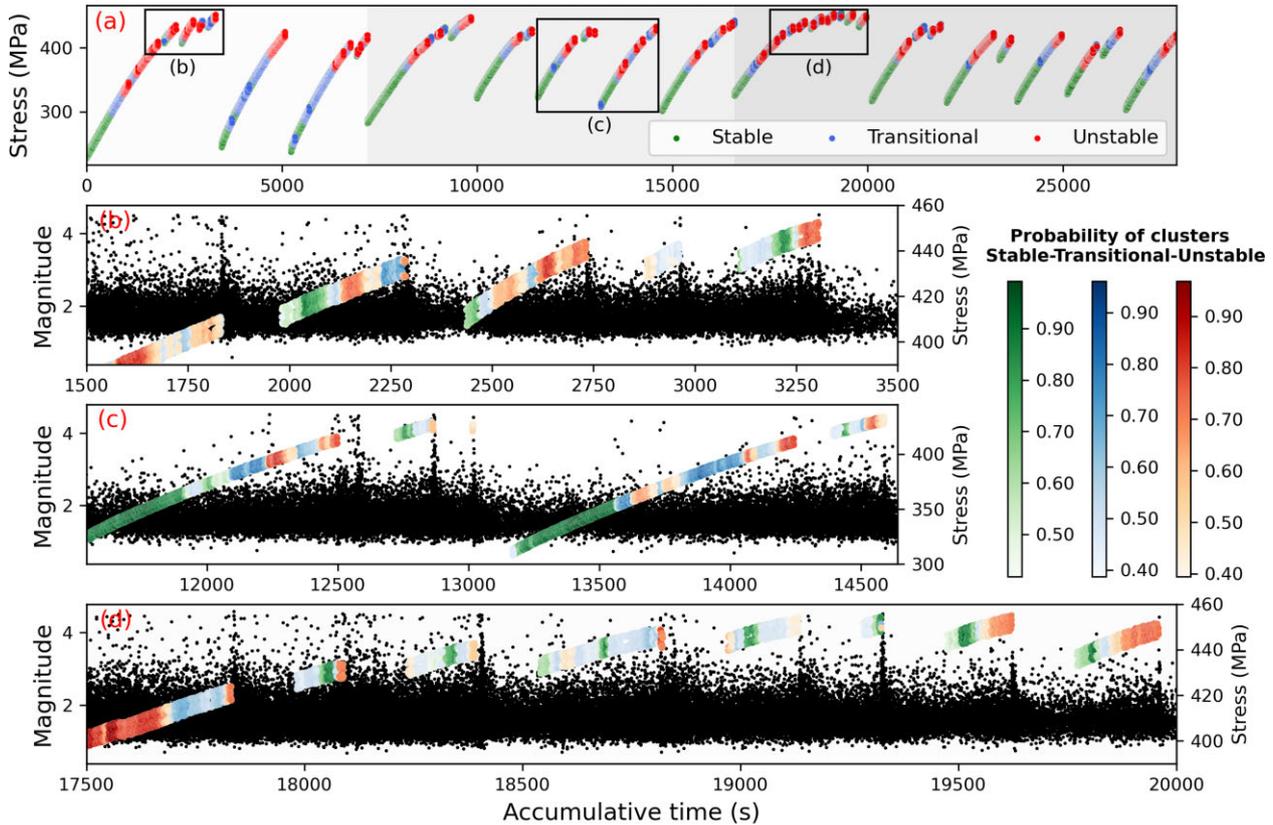


Figure 5. Same as Fig. 2, but based on the 12 most important features of the PCA (Fig. 4b).

Table 2. Confusion matrix between stages using all 47 features and other combinations. For each combination (i.e. 3×3 matrix), the i th row and j th column indicates the percentage of cases with the true label being i th stage and clustered label being j th stage.

Different combinations of all features		Stages	Clustering with 47 features		
			Stable	Transitional	Unstable
A(Fig. 5)	12 important features different time windows	Stable	0.71	0.013	0
		Transitional	0.29	0.74	0.028
		Unstable	0	0.24	0.97
B(Fig. 7)	10 important features unique time windows	Stable	0.85	0.039	0
		Transitional	0.15	0.86	0.072
		Unstable	0	0.098	0.93

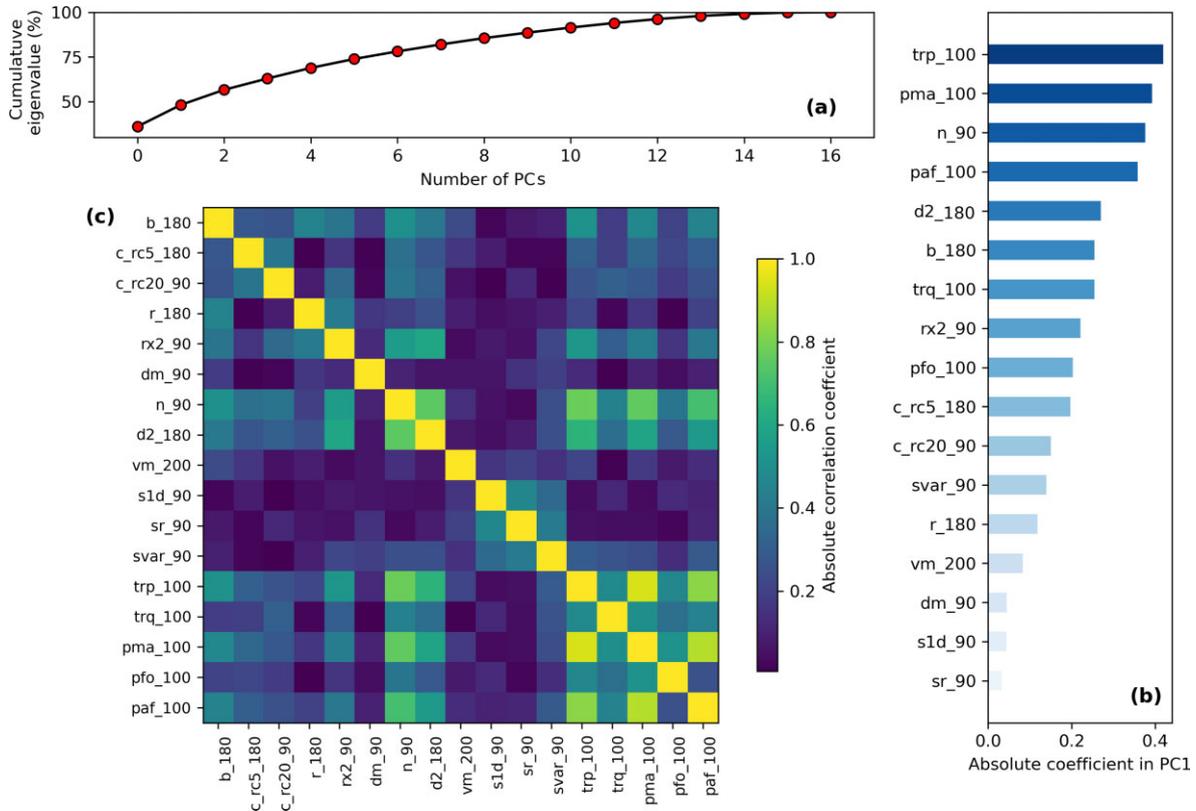


Figure 6. Similar to Fig. 4, but with 17 independent features. Windows are selected based on their priority on Fig. 4b.

For example, there is an unstable stage in the middle of a cycle (Fig. 2b around 2200 and 2500 s, Fig. 2c around 13 700 s and Fig. 2d around 18 600 s), a transitional stage in the final unstable stage (Fig. 2c around 12 400 s), mixed stable-transitional stages (Fig. 2b around 3200 s and Fig. 2d around 18 400 s) and also a short-length unstable stage (Fig. 2b around 2250 s, Fig. 2c around 12 900 s and Fig. 2d around 19 300 s). The occurrence of these short-time and transient temporal changes in clustering is interpreted as follows:

(i) Kwiatek *et al.* (2024) correlated short-lasting AE bursts with the small-scale slips confined in the sample volume that were mostly not reflected in the recorded external stress data due to their local appearance. These local Confined Slips Events (CSE) attributed to local asperity failures provide a significant AE footprint with localized AE activity forming distinct foreshock–main shock–aftershock sequences. Regarding the AE magnitude of events (M_{AE}), they are observed, for example, in Fig. 2(b) around 2200 and 2500 s, Fig. 2(c)

and 13 700 s, where CSEs with higher magnitude compared to surrounding events happen along a transition stage. To investigate the relationship between AE magnitudes and stage type, we calculated the number and proportion of events per second with a magnitude larger than an arbitrary magnitude (M_{AE}), for all stages, as illustrated in Fig. 8. As expected from the Gutenberg–Richter relation, the number of events with arbitrarily selected $M_{AE} > 2.5$ is much smaller than low magnitude events (Fig. 8a) regardless of the stage association. However, the proportional number of events among three stages (Fig. 8b) reveal that the stable and transitional stages are predominantly populated by AE events with small to medium magnitudes. Events with $M_{AE} > 2.7$ are dominantly found in stages formed during the unstable stage before failure. These stages could be related to occurrence of CSEs, as for example, in Fig. 2(d) around 18 350 and 18 600 s that are impossible to be detected even using stress measurements or by observing the temporal evolution of individual features. This shows that the proposed clustering method allows to identify even small footprints of CSEs using the combination of different features.

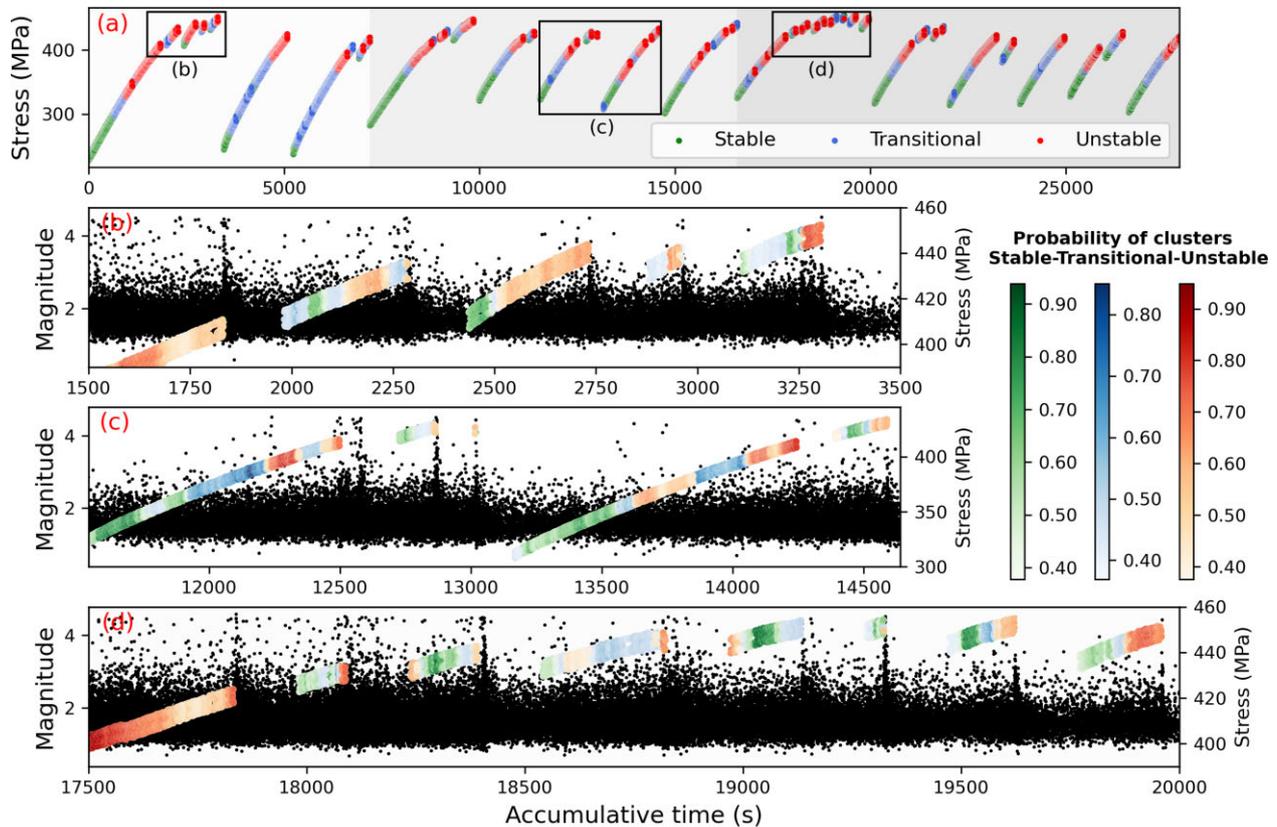


Figure 7. Same as Fig. 2, but with 10 most important independent features of the PCA (Fig. 6b).

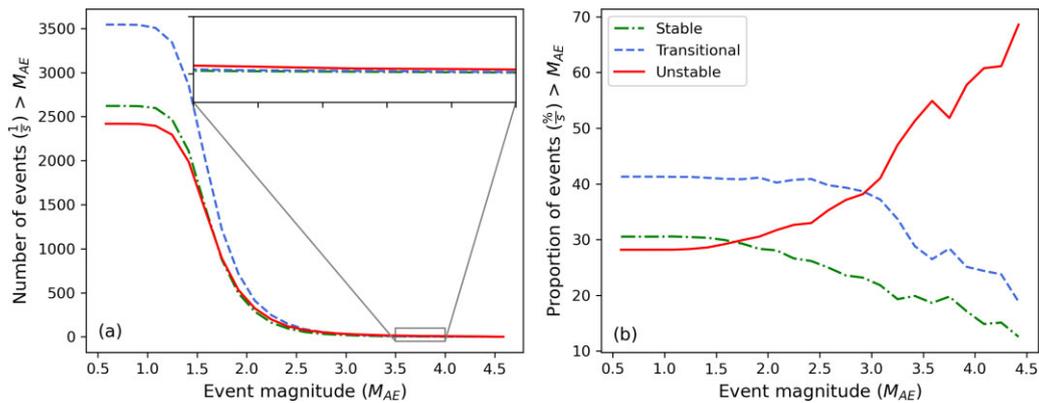


Figure 8. (a) The number of events (per second) with a $M > M_{AE}$ for stable, transitional and unstable stages. (b) Corresponding changes in proportion of events with $M > M_{AE}$ for the three stages, showing that large AEs dominate in the 'unstable' stage.

(ii) Close to the peak stress, the fault surface is in the *intermittent criticality* state (Ben-Zion *et al.* 2003). In this stage, long-range stress correlations are developed by occurrence of numerous small-scale slips (CSE) that lead to local stress drops. At the global scale, the sample retreats from failure temporarily, but ultimately the progressive smoothing out of the stress field enables a random stress perturbation to propagate across the fault surface (see discussion in Kwiatek *et al.* (2024) for this experiment). The retreat from critical state is reflected by AE-features and, thus, temporal changes are obtained in clustering occurring at different spatio-temporal scales. This is seen in Fig. 2(c) around 12 400 s and Fig. 2(b) around 3200 s and Fig. 2(d) around 18 400 s. The criticality could also be responsible for short-length unstable stages at the end of some small cycles. Since most of the small cycles are in the intermittent criticality

state, statistically triggering of some small asperities may not be preceded by local foreshocks. A sudden trigger in such asperities could lead to a short-length unstable stage as they are found in Fig. 2(b) around 2250 s, Fig. 2(c) around 12 900 s and Fig. 2(d) around 19 300 s.

(iii) Note that, in mismatched clustering data, the probability of data points belonging to a certain stage (colour codes) is lower than the other data. This indicates that certain data points have a similar distance to two or more stage centres leading to relatively low probabilities.

Unsupervised clustering algorithms have been successfully applied for preparatory phase detection (Shi *et al.* 2021; Iaccarino & Picozzi 2023; Zali *et al.* 2024). In a similar study on double-

shear tests performed on smooth faults at low normal stress, Bolton *et al.* (2019) used clustering methods on 43 waveform-driven features extracted from AE data. Similar to our results performed on the rough fault, they identified systematic temporal trends during the preparatory phase approaching fault slip. Bolton *et al.* (2019) demonstrated that 85 per cent of the total variance could be captured by just the first 2 PCs, while 14 PCs are needed in our data to reach such a variance (Fig. 4a). This difference highlights the highly complex nature of stick-slip processes occurring on naturally fractured rough faults at high pressures and, potentially, the role of distinct types of features employed (waveform-driven versus catalogue-driven).

K-means clustering has also been used in many earthquake-related studies for different purposes (Ji *et al.* 2020; Yuan 2021; Zali *et al.* 2024). However, the best clustering algorithm is not unique and the algorithm performance depends on the structure of data in the data space. For instance, Lasocki (2014) utilized the ‘Ward’ hierarchical clustering algorithm (Ward 1963) and successfully identified temporally and spatially closest earthquake neighbours and evaluated variations in earthquake clustering over time. Our results also confirm the high performance of this method (Figs 3a and d). Ross *et al.* (2020) used GMM for resolving modes of rupture directivity in large populations of earthquakes. In our data, this method did not result into a separation of stages that compare well with the stress evolution (Fig. 3c). Accordingly, all small slip cycles are attributed to a single unstable stage.

5.1 Applications and upscaling

Our study shows that catalogue-driven seismic features may serve to characterize different stages of the temporal stress evolution leading to failure. In particular, the approach provides critical stress states of faults and indicates a potential transition from stable fluid injection to run-away ruptures during geothermal reservoir stimulation and wastewater disposal. Similarly, such a clustering analysis may help to improve earthquake early warning (EEW) systems (Chin *et al.* 2020) and volcano monitoring (Retailleau *et al.* 2022) for alerting high level risk. To this end, for each online field data, the probabilities of all stages are estimated, which can be used for an alert system.

Seismic events induced by wellbore stimulation and production in geothermal reservoirs pose a seismic hazard that has repeatedly terminated new projects. Martínez-Garzón *et al.* (2020) classified some observations in reservoirs with high potential for unstable (runaway) earthquakes as, for example, noticeable drop in b -value, high seismic injection efficiency [defined as the ratio of cumulative seismic radiated energy to hydraulic energy, the time integral of the cumulative injected volume times the average reservoir pore pressure, see Martínez-Garzón *et al.* (2020)], predominance of double-couple stress components and abundant triggering and earthquake interaction. They suggested that high-frequency seismic monitoring and near-real time analysis of aforementioned parameters is essential to identify in due time the transition from stable, pressure-controlled state to the runaway state. Our study attempts to provide the methodological blueprint and trial set of reliable features that, concurrently with hydraulic information, will contribute to assess potential runaway behaviour.

There are numerous challenges to upscale the methodology to field applications. First, computing such a large number of features as used in this study may be impractical due to the increased data dimensionality and computational expenses involved. The PCA

identified the most crucial independent features that form the fundamental structure of the data space, including AE event rate (n), clustering-localization features (pma , paf , trp , trq), fractal dimension ($d2$), b -value (b), interevent time distribution ($rx2$) and correlation integral (c). From the results presented in Fig. 7 and Table 2B, it is evident that all stages can be detected using just 10 independent features with reasonable accuracy.

Secondly, at least several earthquake cycles should be involved in the dataset. In most volcanic monitoring and EEW systems applications, historical data are available and could be used. However, for geothermal application, for example, in a new project, this may be problematic. Therefore, it is essential to train a more general clustering and finetune it on several geothermal fields, which may be beneficial for new projects.

Third, all input features have to be available and normalized. The former requires the availability of a high-resolution seismic catalogue. The latter requires careful processing of input data so the physical information is not lost. Frequently, preserving the observable parameter range is a priority from the perspective of physical understanding of processes. For example, in geothermal applications, seismic injection efficiency is an important parameter potentially characterizing the propensity of the reservoir to develop runaway conditions. However, the injection efficiency may vary over 10 orders of magnitude (Goodfellow *et al.* 2015) posing a challenge for normalization and training ML models. Likewise, in this study we use the AE-rate as one of the input parameters. However, this parameter varies between the different tectonic settings and is depending on the magnitude of completeness. In Appendix B, we propose to either remove this feature or to replace it with exponentially normalized feature. Results show that even with removing of AE rate similar results as Fig. 2 are obtained.

6 CONCLUSIONS

In this study, we applied unsupervised clustering methods to identify distinct stages of stress evolution leading to failure in laboratory experiments on rough faults. We used 47 catalogue-driven features derived from acoustic emission data and found that even with an unsupervised clustering feature space is divided into stable, transitional and unstable deformation stages. The clustering results showed a good agreement with the temporal evolution of stress and the mechanical behaviour of the rock, revealing systematic patterns of damage accumulation and stress localization before failure. Our investigations demonstrate that:

- (i) The min-max and ED normalization methods, in conjunction with K-means and hierarchical clustering algorithms, successfully identify different stages of loading in a series of small to large stick-slip cycles during triaxial compression tests using catalogue-driven features.
- (ii) Although catalogue-driven features are computed over different time windows, we observed more compelling clustering results when employing independent features. This improvement can be attributed to the reduction of data dimensionality and the mitigated influence of highly correlated similar features.
- (iii) The most important features that capture the main trends in the data are AE event rate, clustering-localization features, fractal dimension, b -value, interevent time distribution and correlation integral. These features reflect the physical processes of faulting, such as seismicity rate, spatial distribution, magnitude distribution, temporal clustering and complexity of the rupture process.

(iv) These stages allow detection of CSEs even though they are not easily observable in the stress curve.

(v) Even if a normalization problem limits the use of non-universal features like AE-rate, omitting such a feature is the simplest solution with a minor effect on clustering results. This is because other highly correlated features could compensate for the lack of other features.

ACKNOWLEDGMENTS

This work was performed in the frame of EU HORIZON DT-GEO project (HORIZON-INFRA-2021-TECH-01, project number 101058129). P.M.G. acknowledges funding from the Helmholtz Association in the frame of the Young Investigators Group VH-NG-1232 (SAIDAN) and ERC Starting Grant 101076119 (QUAKE-HUNTER). The authorship contribution is as follow: Sadeh Karimpouli: conceptualization, software, methodology, formal analysis, investigation, writing—original draft, writing—review and editing, visualization. Grzegorz Kwiatek: conceptualization, resources, data curation, writing—review & editing, supervision, project administration. Patricia Martínez-Garzón: conceptualization, resources, data curation, writing—review and editing. Georg Dresen: conceptualization, validation, writing—review & editing. Marco Bohnhoff: validation, writing—review and editing.

DATA AVAILABILITY

Seismic catalogues, moment tensor catalogues, raw waveform data, geomechanical data and associated information related to stick-slip experiments analysed in this study are available at GFZ Data Services via separate data publication (CC-BY 4.0 license): Kwiatek & Goebel (2024); <https://dataservices.gfz-potsdam.de/panmetaworks/showshort.php?id=5a8a92b0-1cb2-11ee-95b8-f851ad6d1e4b>.

REFERENCES

- Aden-Antoniów, F., Frank, W.B. & Seydoux, L., 2022. An adaptable random forest model for the declustering of earthquake catalogs. *J. geophys. Res.*, **127**, e2021JB023254, doi:10.1029/2021JB023254.
- Arthur, D. & Vassilvitskii, S., 2007. K-means++ the advantages of careful seeding, in *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, New Orleans, Louisiana, 7–9 January 2007, pp. 1027–1035, Society for Industrial and Applied Mathematics.
- Baiesi, M. & Paczuski, M., 2005. Complex networks of earthquakes and aftershocks. *Nonlin. Process Geophys.*, **12**, 1–11.
- Ben-Zion, Y. & Lyakhovskiy, V., 2002. Accelerated seismic release and related aspects of seismicity patterns on earthquake faults, in *Earthquake Processes: Physical Modelling, Numerical Simulation and Data Analysis Part II*, Pageoph Topical Volumes, Vol. **159**, pp. 2385–2412, eds Matsu'ura, M., Mora, P., Donnellan, A. & Yin, Xc., Birkhäuser.
- Ben-Zion, Y., Eneva, M. & Liu, Y., 2003. Large earthquake cycles and intermittent criticality on heterogeneous faults due to evolving stress and seismicity. *J. geophys. Res.*, **108**, doi:10.1029/2002JB002121.
- Beroza, G.C., Segou, M. & Mostafa Mousavi, S., 2021. Machine learning and earthquake forecasting—next steps. *Nat. Commun.*, **12**, 1–3.
- Bolton, D.C., Marone, C., Shokouhi, P., Rivière, J., Rouet-Leduc, B., Hulbert, C. & Johnson, P.A., 2019. Characterizing acoustic signals and searching for precursors during the laboratory seismic cycle using unsupervised machine learning. *Seismol. Res. Lett.*, **90**, 1088–1098.
- Borate, P., Rivière, J., Marone, C., Mali, A., Kifer, D. & Shokouhi, P., 2023. Using a physics-informed neural network and fault zone acoustic monitoring to predict lab earthquakes. *Nat. Commun.*, **14**, 1–12.
- Bowman, D.D. & Sammis, C.G., 2004. Intermittent criticality and the Gutenberg-Richter distribution, in *Computational Earthquake Science Part I*, Pageoph Topical Volumes, pp. 1945–1956, eds Donnellan, A., Mora, P., Matsu'ura, M. & Yin, Xc., Birkhäuser.
- Bridle, J.S., 1990. Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition, in *Neurocomputing*, Vol. **68**, pp. 227–236, eds Soulié, F.F. & Héroult, J., Springer.
- Chin, T.L., Chen, K.Y., Chen, D.Y. & Lin, D.E., 2020. Intelligent real-time earthquake detection by recurrent neural networks. *IEEE Trans. Geosci. Remote Sens.*, **58**, 5440–5449.
- Davidson, J., Kwiatek, G. & Dresen, G., 2012. No evidence of magnitude clustering in an aftershock sequence of nano-and picoseismicity, *Phys. Rev. Lett.*, **108**, doi:10.1103/PHYSREVLETT.108.038501/FIGURES/1/MEDIUM.
- Davidson, J., Goebel, T., Kwiatek, G., Stanchits, S., Baró, J. & Dresen, G., 2021. What controls the presence and characteristics of aftershocks in rock fracture in the lab? *J. geophys. Res.*, **126**, e2021JB022539, doi:10.1029/2021JB022539.
- Dempster, A.P., Laird, N.M. & Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc., B*, **39**, 1–22.
- Dresen, G., Kwiatek, G., Goebel, T. & Ben-Zion, Y., 2020. Seismic and aseismic preparatory processes before large stick-slip failure. *Pure appl. Geophys.*, **177**, 5741–5760.
- Durand, V. et al., 2020. A two-scale preparation phase preceded an Mw 5.8 earthquake in the Sea of Marmara Offshore Istanbul, Turkey. *Seismol. Res. Lett.*, **91**, 3139–3147.
- Goebel, T.H.W., Becker, T.W., Schorlemmer, D., Stanchits, S., Sammis, C., Rybacki, E. & Dresen, G., 2012. Identifying fault heterogeneity through mapping spatial anomalies in acoustic emission statistics. *J. geophys. Res.*, **117**, doi:10.1029/2011JB008763.
- Goebel, T.H.W., Kwiatek, G., Becker, T.W., Brodsky, E.E. & Dresen, G., 2017. What allows seismic events to grow big?: Insights from b-value and fault roughness analysis in laboratory stick-slip experiments. *Geology*, **45**, 815–818.
- Goodfellow, S.D., Nasser, M.H.B., Maxwell, S.C. & Young, R.P., 2015. Hydraulic fracture energy budget: insights from the laboratory. *Geophys. Res. Lett.*, **42**, 3179–3187.
- Gutenberg, B. & Richter, C.F., 1944. Frequency of earthquakes in California*. *Bull. seism. Soc. Am.*, **34**, 185–188.
- Henderson, J.R., Barton, D.J. & Foulger, G.R., 1999. Fractal clustering of induced seismicity in The Geysers geothermal area, California. *Geophys. J. Int.*, **139**, 317–324.
- Hirata, T., Satoh, T. & Ito, K., 1987. Fractal structure of spatial distribution of microfracturing in rock. *Geophys. J. Int.*, **90**, 369–374.
- Holtzman, B.K., Paté, A., Paisley, J., Waldhauser, F. & Repetto, D., 2018. Machine learning reveals cyclic changes in seismic source spectra in Geysers geothermal field. *Sci. Adv.*, **4**, doi:10.1126/SCIADV.AAO2929/SUPPL_FILE/AAO2929_SM.PDF.
- Huang, Y., Li, H., Ma, Y. & Ma, J., 2023. Long-term spatial-temporal evolution of seismicity of the 2010 Ms 7.1 Yushu, Qinghai, China Earthquake. *IEEE Trans. Geosci. Remote Sens.*, **61**, doi:10.1109/TGRS.2022.3231878.
- Iaccarino, A.G. & Picozzi, M., 2023. Detecting the preparatory phase of induced earthquakes at The Geysers (California) using K-means clustering. *J. geophys. Res.*, **128**(10), doi:10.1029/2023JB026429.
- Jain, A.K., 2010. Data clustering: 50 years beyond K-means. *Pattern Recognit. Lett.*, **31**, 651–666.
- Ji, K., Wen, R., Ren, Y. & Dhakal, Y.P., 2020. Nonlinear seismic site response classification using K-means clustering algorithm: case study of the September 6, 2018 Mw6.6 Hokkaido Iburi-Tobu earthquake, Japan. *Soil Dyn. Earthq. Eng.*, **128**, doi:10.1016/J.SOILDYN.2019.105907.
- Johnson, P.A. et al., 2021. Laboratory earthquake forecasting: a machine learning competition. *Proc. Natl. Acad. Sci.*, **118**, e2011362118, doi:10.1073/PNAS.2011362118.
- Kagan, Y.Y., 1991. Likelihood analysis of earthquake catalogues. *Geophys. J. Int.*, **106**, 135–148.
- Kagan, Y.Y. & Knopoff, L., 1980. Spatial distribution of earthquakes: the two-point correlation function. *Geophys. J. Int.*, **62**, 303–320.

- Karimpouli, S. & Fattahi, H., 2016. Estimation of P- and S-wave impedances using Bayesian inversion and adaptive neuro-fuzzy inference system from a carbonate reservoir in Iran. *Neural. Comput. Appl.*, **29**, 1059–1072.
- Karimpouli, S. *et al.*, 2023. Explainable machine learning for labquake prediction using catalog-driven features. *Earth planet. Sci. Lett.*, **622**, doi:10.1016/J.EPSL.2023.118383.
- Kato, A. & Ben-Zion, Y., 2020. The generation of large earthquakes. *Nat. Rev. Earth Environ.*, **2**(1), 26–39.
- 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., Charalampidou, E.M., Dresen, G. & Stanchits, S., 2014. An improved method for seismic moment tensor inversion of acoustic emissions through assessment of sensor coupling and sensitivity to incidence angle. *Int. J. Rock Mech. Min. Sci.*, **65**, 153–161.
- Kwiatek, G., Martínez-Garzón, P. & Bohnhoff, M., 2016. HybridMT: a MATLAB/shell environment package for seismic moment tensor inversion and refinement. *Seismol. Res. Lett.*, **87**, 964–976.
- Kwiatek, G., Martínez-Garzón, P., Davidsen, J., Malin, P., Karjalainen, A., Bohnhoff, M. & Dresen, G., 2022. Limited earthquake interaction during a geothermal hydraulic stimulation in Helsinki, Finland. *J. geophys. Res.*, **127**, e2022JB024354, doi:10.1029/2022JB024354.
- Kwiatek, G., Martínez-Garzón, P., Goebel, T., Bohnhoff, M., Ben-Zion, Y. & Dresen, G., 2024. Intermittent criticality multiscale processes leading to large slip eventson rough laboratory faults. *Journal of Geophysical Research: Solid Earth*, **129**, e2023JB028411. doi: 10.1029/2023JB028411.
- Lasocki, S., 2014. Transformation to equivalent dimensions—a new methodology to study earthquake clustering. *Geophys. J. Int.*, **197**, 1224–1235.
- Lei, X. & Ma, S., 2014. Laboratory acoustic emission study for earthquake generation process. *Earthq. Sci.*, **27**, 627–646.
- Martínez-Garzón, P., Kwiatek, G., Ickrath, M. & Bohnhoff, M., 2014. MSATSI: a MATLAB package for stress inversion combining solid elastic methodology, a new simplified user-handling, and a visualization tool. *Seismol. Res. Lett.*, **85**, 896–904.
- Martínez-Garzón, P., Ben-Zion, Y., Zaliapin, I. & Bohnhoff, M., 2019. Seismic clustering in the Sea of Marmara: implications for monitoring earthquake processes. *Tectonophysics*, **768**, doi:10.1016/J.TECTO.2019.228176.
- Martínez-Garzón, P., Kwiatek, G., Bentz, S., Bohnhoff, M. & Dresen, G., 2020. Induced earthquake potential in geothermal reservoirs: insights from The Geysers, California. *Leading Edge*, **39**, 873–882.
- McLachlan, G.J., Lee, S.X. & Rathnayake, S.I., 2019. Finite mixture models. *Annu. Rev. Stat. Appl.*, **6**, 355–378.
- McLaskey, G.C. & Lockner, D.A., 2014. Preslip and cascade processes initiating laboratory stick slip. *J. geophys. Res.*, **119**, 6323–6336.
- Meng, X.L. & Van Dyk, D., 1997. The EM algorithm—an old folk-song sung to a fast new tune. *J. R. Stat. Soc., B*, **59**, 511–567.
- Mignan, A., Werner, M.J., Wiemer, S., Chen, C.C. & Wu, Y.M., 2011. Bayesian estimation of the spatially varying completeness magnitude of earthquake catalogs. *Bull. seism. Soc. Am.*, **101**, 1371–1385.
- Mousavi, S.M. & Beroza, G.C., 2023. Machine learning in earthquake seismology. *Annu. Rev. Earth planet. Sci.*, **51**, 105–129.
- Mousavi, S.M., Zhu, W., Ellsworth, W. & Beroza, G., 2019a. Unsupervised clustering of seismic signals using deep convolutional autoencoders. *IEEE Geosci. Remote Sens. Lett.*, **16**, 1693–1697.
- Mousavi, S.M., Zhu, W., Sheng, Y. & Beroza, G.C., 2019b. CRED: a deep residual network of convolutional and recurrent units for earthquake signal detection. *Sci. Rep.*, **9**, 1–14.
- Ng, A., Jordan, M. & Weiss, Y., 2001. On spectral clustering: analysis and an algorithm, in *Advances in Neural Information Processing Systems 14 (NIPS 2001)*, eds Dietterich, T., Becker, S. & Ghahramani, Z., NIPS.
- Picozzi, M. & Iaccarino, A.G., 2021. Forecasting the preparatory phase of induced earthquakes by recurrent neural network. *Forecasting*, **3**, 17–36.
- Picozzi, M., Iaccarino, A.G., Spallarossa, D. & Bindi, D., 2023. On catching the preparatory phase of damaging earthquakes: an example from central Italy. *Sci. Rep.*, **13**, doi:10.1038/s41598-023-41625-0.
- Ren, C.X., Hulbert, C., Johnson, P.A. & Rouet-Leduc, B., 2020. Machine learning and fault rupture: a review. *Adv. Geophys.*, **61**, 57–107.
- Retailleau, L. *et al.*, 2022. Automatic detection for a comprehensive view of Mayotte seismicity. *Compt. Rend.—Geosci.*, **354**, 1–18.
- Ross, Z.E., Trugman, D.T., Azizzadenesheli, K. & Anandkumar, A., 2020. Directivity modes of earthquake populations with unsupervised learning. *J. geophys. Res.*, **125**, e2019JB018299, doi:10.1029/2019JB018299.
- Rouet-Leduc, B., Hulbert, C., Lubbers, N., Barros, K., Humphreys, C.J. & Johnson, P.A., 2017. Machine learning predicts laboratory earthquakes. *Geophys. Res. Lett.*, **44**, 9276–9282.
- Sadovskiy, M.A., 1984. Characteristic dimensions of rock and hierarchical properties of seismicity. *Izvest., Earth Phys.*, **20**, 87–96.
- 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*, Pageoph Topical Volumes, pp. 307–334, eds Wyss, M., Shimazaki, K. & Ito, A., Birkhäuser.
- Scholtz, R.A., 1968. Detection, estimation, and modulation theory. Part I. *IEEE Trans. Inf. Theory*, **14**, 612–613.
- Seydoux, L., Balestrieri, R., Poli, P., Hoop, M.d., Campillo, M. & Baraniuk, R., 2020. Clustering earthquake signals and background noises in continuous seismic data with unsupervised deep learning. *Nat. Commun.*, **11**, 1–12.
- Shi, P., Seydoux, L. & Poli, P., 2021. Unsupervised learning of seismic wavefield features: clustering continuous array seismic data during the 2009 L'Aquila earthquake. *J. geophys. Res.*, **126**, e2020JB020506, doi: 10.1029/2020JB020506.
- 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. *Geophys. Res. Lett.*, **48**, e2021GL093187, doi: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. *J. geophys. Res.*, **126**, e2020JB021588, doi:10.1029/2020JB021588.
- Snedecor, G.W. & Cochran, W.G., 1989. *Statistical Methods*, 8th edn, Iowa State Univ. Press.
- Soubestre, J. *et al.*, 2018. Network-based detection and classification of seismovolcanic tremors: example from the Klyuchevskoy Volcanic Group in Kamchatka. *J. geophys. Res.*, **123**, 564–582.
- Van Der Elst, N.J., 2021. B-positive: a robust estimator of aftershock magnitude distribution in transiently incomplete catalogs. *J. geophys. Res.*, **126**, e2020JB021027, doi:10.1029/2020JB021027.
- Van Der Elst, N.J. & Brodsky, E.E., 2010. Connecting near-field and far-field earthquake triggering to dynamic strain. *J. geophys. Res.*, **115**, doi: 10.1029/2009JB006681.
- Vavryčuk, V., 2014. Iterative joint inversion for stress and fault orientations from focal mechanisms. *Geophys. J. Int.*, **199**, 69–77.
- Ward, J.H., 1963. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.*, **58**, 236–244.
- Yoon, C.E., O'Reilly, O., Bergen, K.J. & Beroza, G.C., 2015. Earthquake detection through computationally efficient similarity search. *Sci. Adv.*, **1**, doi:10.1126/SCIADV.1501057/SUPPL_FILE/1501057_SM.PDF.
- Yuan, R., 2021. An improved K-means clustering algorithm for global earthquake catalogs and earthquake magnitude prediction. *J. Seismol.*, **25**, 1005–1020.
- Zali, Z., Mousavi, S.M., Ohrnberger, M., Eibl, E.P.S. & Cotton, F., 2024. Tremor clustering reveals precursors and evolution of the 2021 Geldingadalir eruption. *Commun. Earth Environ.*, **5**, doi:10.1038/s43247-023-01166-w.
- Zaliapin, I. & Ben-Zion, Y., 2013. Earthquake clusters in southern California. I: identification and stability. *J. geophys. Res.*, **118**, 2847–2864.
- Zaliapin, I., Gabriellov, A., Keilis-Borok, V. & Wong, H., 2008. Clustering analysis of seismicity and aftershock identification. *Phys. Rev. Lett.*, **101**, doi:10.1103/PHYSREVLETT.101.018501/FIGURES/5/MEDIUM.

APPENDIX A: CATALOGUE-DRIVEN FEATURES

AE event rate (n): For all events with a magnitude bigger than the magnitude of completeness ($M_{AE} > M_c$), the AE rate in a time window of ΔT is:

$$n = \frac{N_{AE}}{\Delta T} \quad (A1)$$

where N_{AE} is the number of AE events.

GR law's b -value (b): Following to the Gutenberg–Richter law:

$$\log_{10}(N_{AE}) = a - bM \quad (A2)$$

where N_{AE} is the number of AE events, a and b are productivity and relative event size distribution (b -value) and M is event magnitude (M_{AE}). In this work, we used the b -positive method (van der Elst 2021).

Correlation integral (c): The correlation integral computes probability of two events with separation distance of r , relative to all events as:

$$c = \lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{i=0}^n \sum_{j=0}^n H(r - |x_i - x_j|) \quad (A3)$$

where n indicates the number of data in the analysis time window, x the hypocentre coordinates, and H the Heaviside step function:

$$H = \begin{cases} 1 & x > 0 \\ 0 & x \leq 0 \end{cases} \quad (A4)$$

Interevent time distribution ($r1$, $rx2$): Interevent time ratio is computed as:

$$R = [R_i] = (T_{i+1} - T_i) / (T_{i+1} - T_{i-1}), \quad (A5)$$

where T_i is the origin time of the i^{th} event of the time-ordered sequence of earthquakes above the magnitude of completeness. The deviation of distribution of interevent time ratio $P(R)$ from uniform distribution at the distribution edges is calculated as:

$$r1 = \int_0^{0.1} P(R) dr + \int_{0.9}^1 P(R) dr - 0.2, \quad (A6)$$

The amount of $r1$ values more than 0 are representatives of deviation level from the uniform distribution. The deviation of $P(R)$ could also be computed across the whole distribution using chi-squared goodness-of-fit measure (Snedecor & Cochran 1989):

$$rx2 = \int_0^1 (P(r) - u(r))^2 / u(r) dr \quad (A7)$$

where $u(r)$ is the uniform distribution.

Fractal dimension ($d2$): Using the catalogue data and 3-D location of all events, the fractal dimension could be computed by the box-counting method (Hirata *et al.* 1987). However, it is also computed using the pair correlation integral, c , with separation distance, r , as (Lei & Ma 2014) as:

$$d2 = \lim_{r \rightarrow 0} \frac{\log(c)}{\log(r)} \quad (A8)$$

Magnitude correlation (dm): Following to the method by (Davidson *et al.* 2012), for each time window, we start from the vector of magnitudes ordered in the time sequence:

$$\Delta M = [\Delta M_i] = [M_{i+1} - M_i] \quad (A9)$$

The magnitudes are correlated (i.e. they do not behave as randomly drawn from GR distribution) if the PDF is built over the observed magnitude differences, $P(\Delta M)$ is statistically significantly different from a distribution $P(\Delta M^*)$ built over the uncorrelated vector of magnitudes ΔM^* . Multiple realizations of distribution $P(\Delta M^*)$ can be developed by reshuffling the original catalogue of magnitudes to destroy any potential magnitude correlations. In the following, we calculate the deviation between the observed and multiple realizations of the reshuffled cumulative distribution functions:

$$\delta P(\Delta m) = P(\Delta M < \Delta m) - P(\Delta M^* < \Delta m) \quad (A10)$$

A vector of input magnitudes can be considered uncorrelated if the resulting $\delta P(\Delta m)$ is not significantly deviating from 0 for all considered magnitude intervals Δm at the assumed level of significance.

Clustering and localization features (trp , trq , pma , pfo , paf): For each event i and j , magnitude-normalized time, T_{ij} and space, R_{ij} , components are computed as (Zaliapin *et al.* 2008):

$$T_{ij} = t_{ij} 10^{-qb m_i} \quad (A11)$$

$$R_{ij} = (r_{ij})^d 10^{-(1-q) b m_i} \quad (A12)$$

where $t_{ij} = t_j - t_i$ ($t_j > t_i$) and r_{ij} are the temporal and spatial distances between the events i and j , b is the b -value, d is the fractal dimension and m_i is the magnitude of the earlier event in time. Product (proximity parameter) and ration of normalized time and space components are defined as:

$$trp = T_{ij} \cdot R_{ij} \quad (A13)$$

$$trq = T_{ij} / R_{ij} \quad (A14)$$

In our study, we identify acoustic emission (AE) clusters that display proximity distances smaller than an estimated threshold. We classify each AE event that is linked to the parent cluster by a distance longer than the threshold as a *background* event, which initiates a new cluster. We define a *single* cluster as one that contains only one background event without any *foreshocks* or *aftershocks*. Clusters with multiple events are referred to as *families*. The most significant event within each cluster is labelled as the *main shock*, while all the events occurring before or after the main shock are classified as *foreshocks* or *aftershocks*. The temporal evolution of the median proximity parameter in a sliding time window is computed as:

$$\hat{\eta} = \text{median}\{trq_i\} \quad (A15)$$

Finally, the fraction of *aftershocks* (paf), *foreshocks* (pfo) and *main shocks* (pma) in each examined time window is computed from all clusters conditioned that:

$$paf + pfo + pma = 1 \quad (A16)$$

Focal mechanisms and deviatoric stress tensor based features (vm , sld , sr , $svar$): For each time window, we first calculate median Kagan angles (Kagan 1991) between each AE focal mechanism and 20 nearest focal mechanisms (Goebel *et al.* 2017). This provides information on local in space variability of the focal mechanism over the fault plane, vm , for a particular time window. In the following, vm is calculated as a median from spatially calculated focal mechanism variabilities (i.e. it is averaged over the whole fault surface).

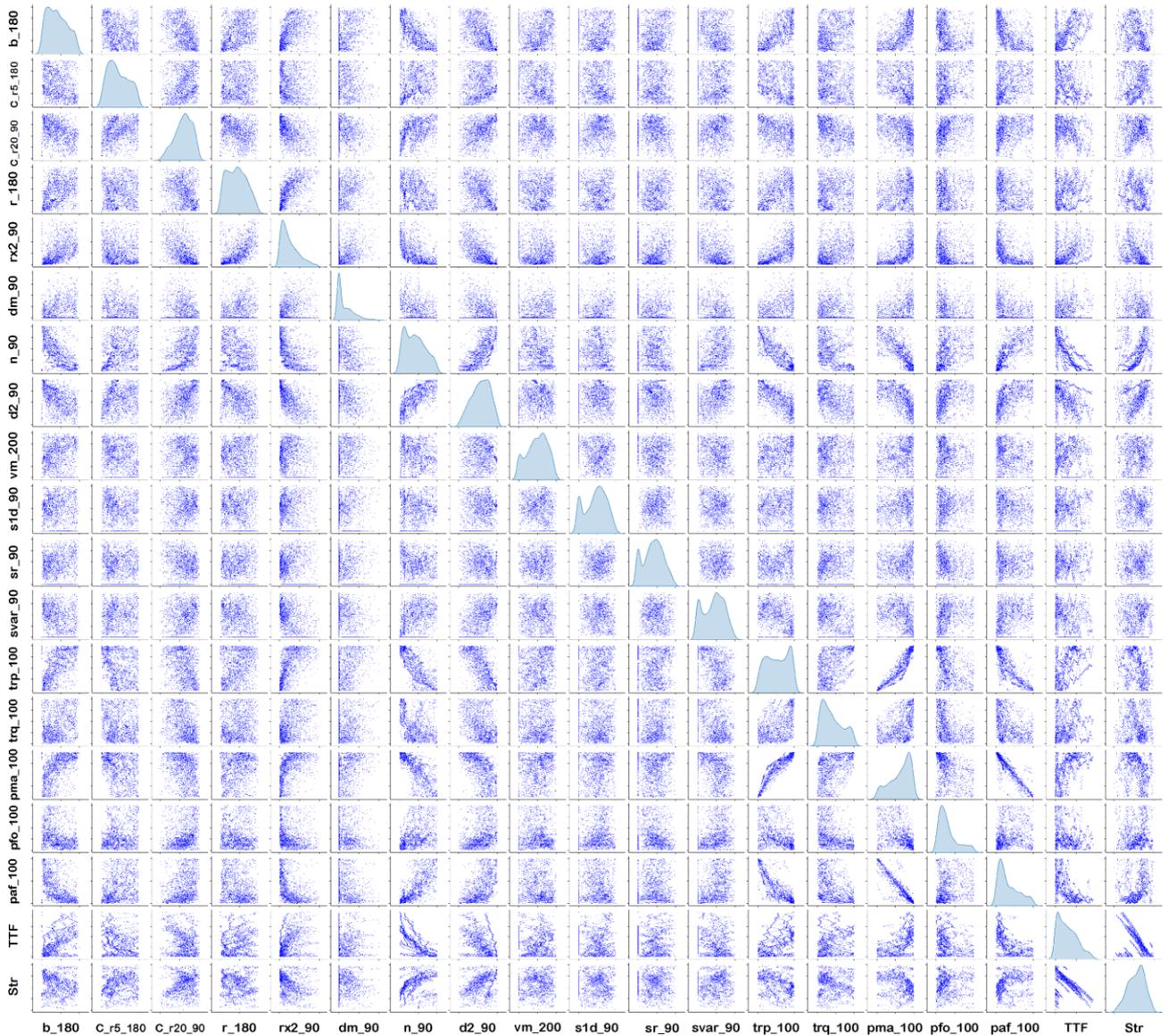


Figure A1. Scatter plot of 17 independent features (Table 1) in addition to TTF and stress value (Str). For a quantitative comparison see correlation matrix in Fig. 6(c). All axes are scaled into [0, 1].

Stress tensor inversion has been performed using the STRESSINVERSE package (Vavryčuk 2014). Local stress tensors have been estimated from at least 40 AE focal mechanisms using the same nearest-neighbour procedure as employed for the vm parameter. In the following, from available local stress tensors distributed over the fault surface, we calculated the median plunge of the maximum principal stress ($s1d$), median stress shape ratio (sr) and median rotation between principal stress axes ($svar$). The calculated parameters indicate local stress concentrations, and variability in orientation of the principal stresses, respectively. Since all of these values are computed per event, the median of locally computed values is used.

All of these features have been computed in different time windows (see the text), while Fig. A1 illustrates the distribution of 17 independent features as well as TTF and stress.

APPENDIX B: STRATEGIES FOR CHALLENGING PARAMETERS ON THE CLUSTERING PROCESS

Clustering without AE-rate

Clustering operates by capturing the primary patterns within the data. In theory, if two features exhibit a strong correlation, the clustering outcome will remain the same even if one of the features is removed. The AE-rate n is the first highly correlated feature relative to trp , indicating that its removal should not significantly impact the final clustering outcome.

We performed K-means clustering using 9 important independent features excluding the AE-rate. The resulting confusion matrix in Table B1-C compares the clustering results with those obtained using all 47 features. The table shows that clustering still performs

Table B1. Similar to Table 2.

Different combinations of all features		Stages	Clustering with 47 features		
			Stable	Transitional	Unstable
C	9 important featuresunique time windowsremoved AE-n	Stable	0.81	0.046	0.0003
		Transitional	0.19	0.81	0.083
		Unstable	0	0.14	0.92
D	10 important featuresunique time windowsconverted AE-n to EXP(n)	Stable	0.81	0.044	0.0002
		Transitional	0.19	0.80	0.071
		Unstable	0	0.15	0.93

well, reaching 81, 81 and 92 per cent of cases for the stable, transitional and unstable stages, respectively. The impact of removing AE-rate can be observed by comparing the confusion matrices before (Table 2B) and after (Table B1-C) its removal. As anticipated, the correct clustering decreased by 4, 5 and 1 per cent for each stage, respectively, revealing minor effect of AE-rate removal.

Clustering with a generally normalized AE-rate

In order to incorporate the AE-rate in clustering while ensuring independence from the data boundaries, a normalization method is required that considers the statistical distribution of the data. Such methods can be based on either statistical parameter, such as Z-score normalization, or exponential functions, such as the softmax function (Bridle 1990). In this study, we propose a reformulated version of the exponential function to define the normalized AE-rate (n_{norm}). The most important point is that rather than AE-rate, the difference between AE-rate and the background rate or seismicity (BS) of the respective geological setting should be considered. Therefore, we firstly need *a priori* information of BS (eq. B1), which is defined as an average AE-rate of the first part of each labquake or earthquake cycle. This is because all features are normalized per cycle. The exponent is then defined as a scale to compensate difference of the AE-rate and BS as eq. (B1), where a is the scaling factor. In our case, we use $a = 10$ for all three samples assuming that BC is a non-zero positive value (eq. B2). Accordingly, when the AE-rate tends to 0, the n_{norm} will be 0 and when the AE-rate tends to infinity, the n_{norm} will be 1 (eq. B3).

$$BS = \frac{1}{N} \sum_{i=1}^N n_i, \quad N = \text{first part of each cycle} \quad (\text{B1})$$

$$n_{\text{norm}} = e^{-\left(\frac{a \times BS}{n - BS}\right)} \quad (\text{B2})$$

$$n_{\text{norm}} = \begin{cases} 1, & n \rightarrow \infty \\ 0, & n \rightarrow 0 \end{cases} \quad (\text{B3})$$

Fig. B1 illustrates both AE-rate n from our experiments and the normalized rate n_{norm} (scaled-up for visual comparison). For sample WgN04, both rates are computed for different time windows (23, 45, 90 and 180 s). For samples WgN05, and WgN07, the rates are shown just for the 180 s time window. The normalized rates n_{norm} follow the same trend as n except for the beginning and the end of the cycles.

By substituting the original AE-rates with the corresponding normalized rates, we recalculated the clustering results using the K-means algorithm. The resulting confusion matrix is presented in Table B1-D. A comparison between these results (Table B1-D) and the results by using original AE-rates (Table 2B) shows that the correct clustering cases of the first two stages has decreased by approximately 4 and 6 per cent, while for the unstable stage remains unchanged. This reduction may be attributed to using an exponential function at the boundaries of the data. Specifically, during the initial and final portions of each cycle, the n_{norm} values fluctuate at a slower rate compared to the original values. Consequently, this leads to a different distribution of the n_{norm} relative to the primary distribution established by the highly correlated features. Note that the correlation coefficient between trq and n , decreased from 77 to 39 per cent for trq and n_{norm} , confirming a different distribution.

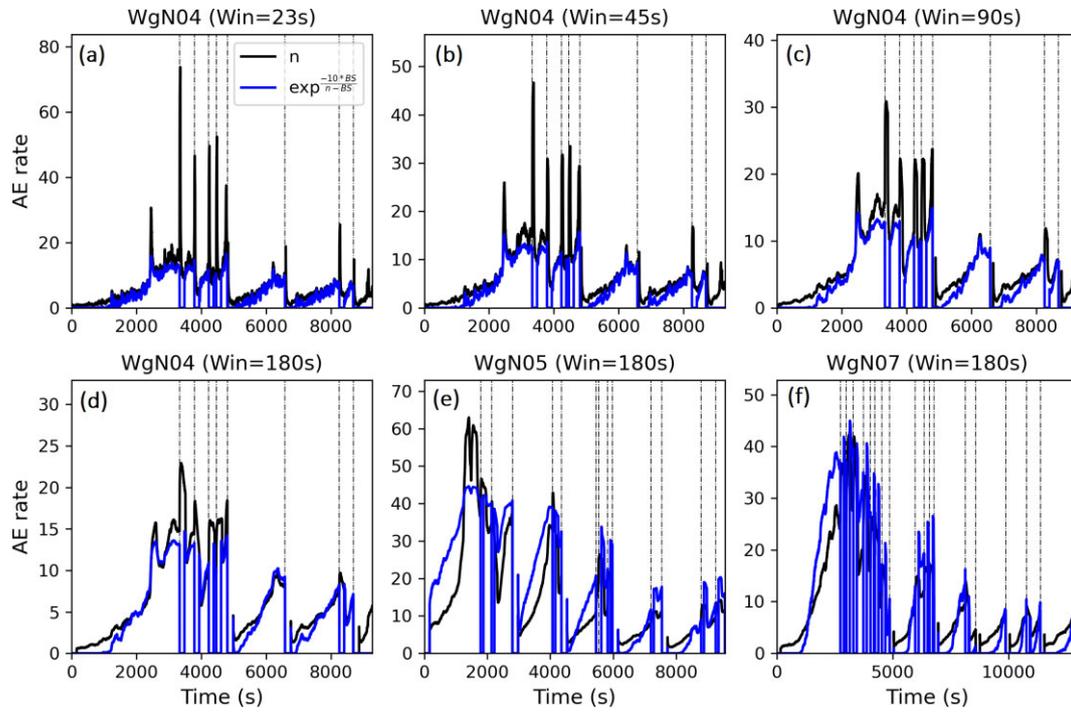


Figure B1. Comparison between original AE-rate (black) and the proposed normalized version (eq. B2, blue) for (a–d) different time windows in sample WgN04 and for (e–f) the longest time window for the samples WgN05 and WgN07. The normalized values are re-scaled for comparison purpose.