JGR Solid Earth

RESEARCH ARTICLE
10.1029/2022JB024144

Special Section:
Machine learning for Solid Earth observation, modeling and understanding

Characterization of Acoustic Emissions From Analogue Rocks Using Sparse Regression-DMDc

C. Fieseler[1], C. A. Mitchell[2], L. J. Pyrak-Nolte[2,3,4], and J. N. Kutz[5,6]

[1]Department of Neurobiology, University of Vienna, Vienna, Austria, [2]Department of Earth, Atmospheric, and Planetary Sciences, Purdue University, West Lafayette, IN, USA, [3]Department of Physics and Astronomy, Purdue University, West Lafayette, IN, USA, [4]Lyles School of Civil Engineering, Purdue University, West Lafayette, IN, USA, [5]Department of Applied Mathematics, University of Washington, Seattle, WA, USA, [6]Department of Electrical Engineering and Physics, University of Washington, Seattle, WA, USA

Abstract Moisture loss in rock is known to generate acoustic emissions (AE). Phenomena that result in AE during drying are related to the movement of fluids through the pores and induced-cracks that arise from differential mineral shrinkage, especially in clay-bearing rock. AE from the movement of fluids occurs from the reconfiguration of fluid interfaces during drying, while AE from mineral shrinkage involves the debonding within or between minerals. Here, analogue rock samples were used to examine the differences in the AE signatures when one or both AE source-types are present. An unsupervised sparse regression model, Dynamic Mode Decomposition with control, that extends Dynamic Mode Decomposition is used to characterize the AE signals recorded during the drying of porous analogue rock samples fabricated with ordinary Portland cement, with and without clay. This method can effectively and accurately reconstruct acoustic signals emitted from samples that only experience moisture loss without cracking. However, the method struggles to reconstruct signals from samples with intricate crack networks that formed during drying because AE generating mechanisms can emit contemporaneously, and the resulting waves propagate through drying-induced cracks that can lead to multiple internal reflections. Thus, the differential reconstruction accuracy of time series generated by different underlying physical processes provides a robust filter for reducing large data catalogs. In general, both dynamics and sparse initiating events are learned directly from data and this method exposes a data hierarchy based on the complexity of the intrinsic dynamics.

Plain Language Summary Moisture loss from porous media, in particular clay-rich rocks, can generate sound known as acoustic emissions (AE), as water moves through the pore structure and cracking caused by mineral shrinkage from dehydration occurs. Unique information about processes giving rise to these sounds and the changes in the material can be extracted from the AE data. To examine the differences in these signals, distinct sample types were fabricated with ordinary Portland cement, and in some instances with clay. A machine learning method requiring no prior information that is informed by the data was used to sort through and identify the AE signals. This method works best for AE signals emitted from samples that only experience movement of fluids with no cracking. When both processes occur, the AE signals contain overlapping information that simultaneously are related to drying and cracking. While this method does not perform well on AE data that contain information from many sources, it does provide a strong filter for large volumes of data which are quite challenging to examine and extract relevant information from. In general, this method is most robust to simultaneously learn the initiating action generated by the single sources and characteristics of the AE.

Key Points:

- Application of an extension of a dimensionality reduction technique, Dynamic Mode Decomposition
- Sparse Regression can be used to simultaneously discover intrinsic dynamics and external initiating events
- Acoustic emissions events related to the infiltration of the drying front are well captured by a linear dynamical system and a single initiating event

Correspondence to:
C. A. Mitchell,
mitch240@purdue.edu

Citation:
Fieseler, C., Mitchell, C. A., Pyrak-Nolte, L. J., & Kutz, J. N. (2022). Characterization of acoustic emissions from analogue rocks using sparse regression-DMDc. *Journal of Geophysical Research: Solid Earth*, *127*, e2022JB024144. https://doi.org/10.1029/2022JB024144

Received 31 JAN 2022
Accepted 7 JUN 2022

Author Contributions:
Conceptualization: C. Fieseler, C. A. Mitchell, L. J. Pyrak-Nolte, J. N. Kutz
Data curation: C. A. Mitchell, L. J. Pyrak-Nolte
Funding acquisition: L. J. Pyrak-Nolte, J. N. Kutz
Investigation: C. Fieseler, C. A. Mitchell
Methodology: C. Fieseler, C. A. Mitchell, L. J. Pyrak-Nolte, J. N. Kutz
Resources: L. J. Pyrak-Nolte
Software: C. Fieseler, J. N. Kutz

1. Introduction

The measurement and analysis of acoustic emissions is a non-destructive, passive-monitoring technique that is used to determine changes in the physical and chemical conditions of a rock. An *acoustic emission* (AE) is defined as a transient elastic wave generated by the rapid release of energy within a material (Lockner, 1993; Scruby, 1987). After energy release, AE emanates from the location, or zone, of abrupt and localized mechanical and interfacial energy that triggered the generation of the elastic waves (Scruby, 1987). In geophysics, acoustic emissions methods have been used to monitor crack initiation, coalescence, and propagation (Lockner, 1993), shearing behavior along fractures (Bolton et al., 2020; Rouet-Leduc et al., 2018), and the evolution of drying fronts (Moebius et al., 2012). More recently, acoustic emissions from transportable sources have been used to

Supervision: L. J. Pyrak-Nolte, J. N. Kutz
Visualization: C. Fieseler, C. A. Mitchell
Writing – original draft: C. Fieseler, C. A. Mitchell
Writing – review & editing: C. Fieseler, C. A. Mitchell, L. J. Pyrak-Nolte, J. N. Kutz

track flow paths through fractures (Pyrak-Nolte et al., 2020) and changes in saturation in a set of parallel fractures (Nolte & Pyrak-Nolte, 2022). AE can also be emitted during the movement of a drying (drainage) front in a porous medium (Moebius et al., 2012). In clay-rich rock, movement of a drying front also enables the formation of cracks if the mineral components that compose a rock exhibit differential volumetric shrinkage from the loss or gain of fluids as shown by Andreani et al. (2008) and Espinoza and Santamarina (2011). The ubiquity and diversity of AEs motivates our development of data-driven machine-learning algorithms that can characterize the rich time-series dynamics associated with transient elastic waves generation in various porous materials.

Acoustic emission signals carry information (energy, frequency, amplitude, coda wave patterns, source locations) that can be used to quantitatively evaluate and distinguish discontinuities and damage mechanisms in space and time. AE monitoring has been applied to gain insights into intrinsic phenomena and spatiotemporal variations in rock properties relevant to failure or damage occurring in rocks (e.g., Lei et al. (2000); Wang et al. (2013); Bunger et al. (2015); Moradian et al. (2016); Rück et al. (2017)) and pre-existing discontinuities (e.g., Labuz et al. (2001); Selvadurai and Glaser (2012); Johnson et al. (2013); Shi et al. (2019); Trugman et al. (2020)). It has been applied in studies to provide information and track crack paths in experiments related to hydraulic fracturing (Bunger et al., 2014; Ishida, 2001; Li et al., 2018), activities in mining (Sellers et al., 2003; Li et al., 2019), and geothermal applications (Kong et al., 2018; Wang, Bian, et al., 2021). This convenient and effective non-destructive technique has also been used to investigate the underlying physics of earthquake phenomena (McLaskey & Lockner, 2014).

With the many applications mentioned above, there are naturally an array of processes that lead to the internal rapid release of accumulated energy. In partially saturated porous media, e.g., particularly during moisture loss, many AE generating mechanism exist. As fluid front displacement occurs, Moebius et al. (2012) showed several potential mechanisms that generate acoustic emissions which included (a) rapid interfacial invasion into pores (Haines jumps or rheons), (b) air entrainment and oscillating bubbles, (c) liquid bridge rupture, (d) inter-facial snap-off, and (e) capillary-induced grain rearrangement and collisions. Probable AE generating mechanism influenced by the unique properties of clay minerals (such as ion-substitution and the ability to shrink or swell in response to changes in water content and/chemistry), may occur during moisture loss in partially-saturated clay-rich porous media. The local state of stress in clay-bearing porous media during drying can cause cracks to be induced because micro-scale processes in expandable clay minerals can induce macroscopically observed volumetric variations (Ilgen et al., 2017). This damage includes but is not limited to the formation, propagation, and coalescence of micro- and macro-cracks, and the emergence of shearing planes and other modes of rapid macroscopic failure. Such grain scale mechanical interactions generate high frequency (kHz) elastic waves that typically occur in geologic granular media (Michlmayr et al., 2012).

A key goal that drives the vast application of the non-destructive acoustic emission (AE) monitoring technique, is the assumption that AE signals are related to spatio-temporal responses and variations in the properties of the microstructure of materials, and contain features that can be used to identify the underlying generating mechanism(s). The ability to identify an infiltrating or escaping fluid front from damage mechanisms or zones of weakness using AE waveforms has significant advantages especially for predictive modeling and the recognition of unanticipated behaviors of a system. In consideration of these benefits of AE, machine learning (ML) techniques play a vital role in the acquisition of additional information from the data. ML techniques, particularly data-driven methods, interpret data based on the features and intrinsic patterns which are learned from the history of the system or applied processes. Application of supervised or unsupervised ML methods can greatly enhance the scientific discovery process by rapidly sifting through massive data sets to identify similarities and differences among signals (Bergen et al., 2019). The application of ML methods to differentiate AE waveforms is an invaluable tool for analysis and interpretation of data, owing to the ability to separate large datasets into smaller classes that are reflective of the intrinsic dynamics or physics that generate the emissions. For many experiments where AE data are recorded, ML techniques may help unravel the initiating source mechanism(s) especially if multiple sources exists. This would be particularly advantageous in systems where AE is generated from crack nucleation, fracture propagation, interfacial debonding, delamination, fracture coalescence, creep, cavitation, or advancing fluid fronts. Exploiting the intrinsic feature recognition ability of many supervised or unsupervised ML techniques can enhance event association, linking the whole or distinct characteristics of an AE waveform to a particular generating mechanism.

Several ML techniques have been explored to improve the accuracy of data analysis methods fundamental to earthquake science, microseismic and acoustic emission studies. Methods that have been expressly used to

improve the accuracy of AE event location include artificial neural networks (ANN) and support vector machines (SVM) (Zhao & Glaser, 2020), SVM classifier coupled with probabilistic output (Ince et al., 2010), Gaussian process regression with radial basis function (RBF) Kernel (Hensman et al., 2010) and cascaded region-based convolutional neural networks (CNN) (Wu et al., 2018). Several studies have tested ML techniques such as CNN (Guo et al., 2021), CNN coupled with K-means clustering (Chen et al., 2019), time delay neural network (King et al., 2021), and deep recurrent neural networks using long short-term memory architecture (Zheng, Lu, et al., 2018) to improve arrival time picking algorithms (auto-selection of the P- and S-wave arrival times). A random forest (RF) regression (supervised learning) has been employed to extract features of creep (Biswas et al., 2020) and to predict the onset of laboratory earthquakes (Rouet-Leduc et al., 2017) using information from continuously recorded acoustic emissions. A decision tree ensemble method, gradient boosted trees (the XGBoost Implementation) has been used to estimate fault friction from the instantaneous statistical characteristics of the AE signals, Rouet-Leduc et al. (2018) and Hulbert et al. (2019). ML models like multivariate Gaussian distribution (Wang, Hou, et al., 2021), wavelet transform coupled with ANNs (Liu et al., 2015), CNN (Huang et al., 2021), CNN with an antinoise architecture (Chen et al., 2019), twin neural networks (Nolte & Pyrak-Nolte, 2022) and RF compared with SVM (Lee et al., 2021), have been used to explore and unravel, features and patterns in recorded acoustic emissions, and classify and cluster them based on the mechanism that generated the release of energy (e.g., fracturing, infiltration of a drying/fluid front or evolving damage). Many of the ML models mentioned above involve some permutation of neural networks. Neural networks offer several advantages that include the ease of use and implementation with minimal statistical training, ability to implicitly detect complex nonlinear relationships between dependent and independent variables, ability to detect all possible interactions between predictor variables, and the availability of multiple training algorithms (Tu, 1996). Neural networks are however essentially a black box, present a great computational burden, are prone to over-fitting of the data, accuracy is dependent on the skills of the implementer, and the model development is characteristically empirical (Tu, 1996).

Emerging dimensionality reduction techniques, such as the *dynamic mode decomposition* (DMD), have garnered much attention since these methods simultaneously (a) reveal low-rank spatio-temporal patterns of activity (DMD modes), (b) discover dynamics in this low-rank subspace, and (c) provide approximations in terms of linear dynamical systems, which are amenable to simple analysis techniques (Brunton & Kutz, 2020). The model applied here, dynamic mode decomposition with control (DMDc), has theoretical connections to Koopman theory (Brunton & Kutz, 2020), and incorporates control theory through the addition of external input signals (Proctor et al., 2016). A recent extension of this DMDc method has relaxed the need for known input (supervised learning), instead allowing unsupervised learning of controls signals from the data (Fieseler et al., 2020).

In this paper we present the application of the novel unsupervised data-driven method of DMDc to AE signals recorded over 6 days during the drying of 3 types of saturated analogue rocks. DMDc is used to characterize the AE waveforms based on the initiating mechanism(s) and to examine the reconstructed patterns of recorded elastic waves. With DMDc, the intrinsic dynamics embedded in the AE waveforms are captured and the meaningful features of the signal data are efficiently reconstructed. DMDc groups the signals by the levels of complexity (i.e., linear vs. nonlinear data) and according to its relevance to probable AE generating mechanism and spatiotemporal interactions in the system. In the absence of detailed ground truth knowledge of the initiating mechanism of each acoustic emission, DMDc provides a valuable method to identify the different type of sample structures from recorded AE.

2. Experimental Methods

2.1. Experimental Sample Preparation

Analogue rock samples were fabricated to study the signatures of acoustic emission mechanisms from rock undergoing dehydration. These synthetic rocks minimize the complexities that can be encountered when evaluating data obtained from complex natural materials where sample-to-sample heterogeneity may prevent repeatable testing of hypotheses. Samples were fabricated with different internal structures with external dimensions of 38.1 mm $\pm$ 0.5 mm in diameter by 76.2 mm $\pm$ 1 mm in length. The synthetic samples were composed of ordinary Type II Portland cement (OPC from LaFargeHolcim:North-America:Inc, 2018), Ottawa sand (maximum particle size of 250 μm (U.S.-Silica, 2019), and montmorillonite K10 clay (product number: 69866 Sigma Aldrich). The montmorillonite used in this study is a commercially available clay with a reported surface area range of

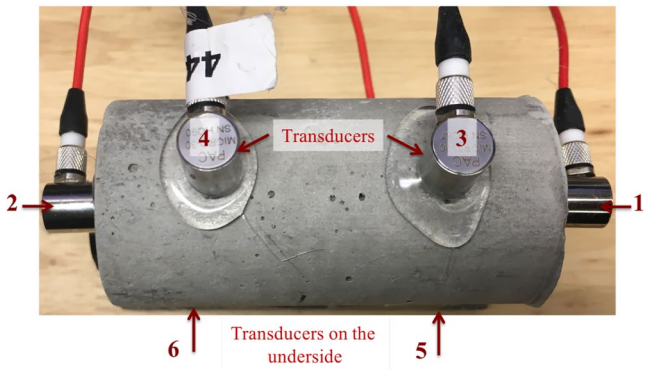


Figure 1. Acoustic emissions Monitoring transducer placement on Mortar only sample.

220–270 m^2/g (Sigma-Aldrich, 2019). The three types of samples were: (a) clay-free reference samples composed only of mortar (i.e., mixture of OPC and Ottawa sand) and referred to as "Mortar"; (b) a mortar samples with localized spherical assemblages of clay particles and referred to as "Localized"; and (c) a mortar sample with clay particles randomly distributed throughout the mortar matrix and referred to as "Distributed." The fabricated synthetic rocks are rich in $CaCO_3$ and SiO_2. Due to this composition the distribution of clay into such a matrix results in a rock analogue that bears a similarity to argillaceous rocks such as marls which are clay-bearing carbonate rich media. Marlstones are calcium carbonate or lime-rich mudstones with variable amounts of clays and silt (Pettijohn et al., 2012).

The reference matrix of the samples was made with a mixture ratio of 1:2.5 cement to sand, and an average water to cement ratio of 0.87. The high water/cement ratio for the matrix is used to obtain the same water/cement ratio used for the clay-containing samples. To fabricate a Mortar sample, the dry ingredients were first mixed together using a kitchen grade food processor for 3–5 min. Using a kitchen grade stand mixer [5-Qt KitchenAid stand mixer], deionized water was added to the combined dry ingredients and mixed at speeds between 2 (95 revolutions per minute (rpm)) and 4 (135 rpm) for no more than 5 min. A small square dental lab vibrator model oscillator with voltage outputs of 110 V (60 Hz) ± 10% and 230 V (50 Hz) ± 10%, and 3,000–3,600 cycles per minute was used to compact the wet blended mixture into the cylindrical molds and stimulate the removal of larger pockets of air. After vibration, each mold was covered with a plastic cylinder to minimize water loss during hardening and setting, and were stored within a covered chamber for 24 hr. The durability and strength of the fabricated samples depend on the hydration of cement, which is affected by the temperature and availability of moisture. After 24 hr the samples were de-molded, and placed into zip-locked plastic bags filled with deionized water, and each sample type was placed in a sample specific curing tank. All samples were then subjected to submerged curing in a plastic tank at a temperature of 30°C ± 1°C for a period of 7 days. Upon completion of curing, the samples were removed from the water bath for analysis. For the samples with clay material, the clay powders were pre-soaked for at least 48 hr prior to fabrication to ensure that the clay particles were fully saturated before adding the clay slurry to the OPC mortar mixture. This step facilitates good cement hydration by minimizing the amount of water the clay may divert from the matrix during the process of cement hydration. To fabricate the localized clay sample, the spherical assemblage of clay particles was formed by freezing a portion of the pre-soaked clay slurry for 24 hr prior to fabrication at approximately 5°C. Immediately post curing the uniaxial compressive strength (UCS) of a minimum of 3 representative saturated samples for each synthetic rock type was obtained using an Instron loading system where the load was applied in the axial direction at a rate of 0.08 mm per minute. The UCS is estimated using the maximum load recorded prior to failure and the cross-sectional area of the sample (mm^2). The UCS range for the clay-free rock analogue was 9.5–12.2 and 6 Mpa for the localized clay sample. A range of UCS values, 3.5–5.5 Mpa, was measured for the distributed clay synthetic rock; and in comparison, water saturated UCS values for Upper and Middle Miocene marlstones as noted by Erguler and Ulusay (2009) were 4.5 and 3.7 Mpa respectively. To prepare the sample for AE monitoring, before affixing transducers to the surface of a rock, each sample was allowed to sit for 10–15 min to minimize surface water which could hinder the secure coupling of the acoustic emission transducers.

2.2. AE Data Acquisition Setup

Drying experiments were performed on the unconfined samples under ambient temperatures and relative humidity for six (6) days. Six (6) Mistras Group - Physical Acoustics Corporation Micro 30 transducers were affixed to each sample using hot gorilla glue as a couplant (see Figure 1) to monitor the behavior of the unbounded clay-free and clay-rich analogue rocks during moisture loss. These μ30 sensors have a flat frequency response over the range of 100–600 kHz, and a resonant frequency at 125 kHz. In all experiments the sensors were attached to a Mistras Express 24 Channel AE System through a single-ended differential pre-amplifier set to a gain of 60 dB. Ambient noise was filtered using a threshold of 25 dB. Acoustic emissions were recorded with the Mistras AEWin Windows based software (Physical-Acoustics:Mistras-Group, 2018) when the signal amplitude exceeded this prescribed threshold. The threshold was fixed on the low end to accommodate AE with lower amplitudes that

Table 1
Number of Acoustic Emissions Recorded per Specimen Type and for Each Sensor Instrumented on the Each Synthetic Rock, and Image Analysis Results Regarding Clay Volume, Damage to the Microstructure, and Total Porosity (Void Volume), Where TL = Top Left, BL = Bottom Left, TR = Top Right, and Br = Bottom (See Figure 1 Right)

Specimen type	Distributed	Localized	Mortar
Acoustic Emission Data			
Sensor 1 [Top]	10,045	147	39
Sensor 2 [Base]	2,615	0	498
Sensor 3 [TL]	5,956	140	64
Sensor 4 [BL]	3,006	207	17
Sensor 5 [TR]	8,200	905	153
Sensor 6 [BR]	4,405	380	21
X-ray Microscopy Data			
Clay volume (%)	20	7	0
Damage volume (%)	95	7	0
Segmented void volume (%)	10	9	3

were affected by the expected attenuation of waveforms which is observed as porous mediums become more damaged with fractures. Appropriate AE acquisition parameters to be used for monitoring and capturing events across all samples were established with preliminary AE monitoring test. These parameters were defined in the AEWin software as the peak definition time which was set to 200 μS, the hit definition time set to 800 μS, the hit lock out time set at 1,000 μS and the maximum duration which was 1,000 μS. AE data were recorded throughout the drying experiment at a sampling frequency of 4 MHz for a record length with 2,048 points. Continuous recording was not performed to minimize recording noise in the data. Hundreds of acoustic emissions were produced resulting in a combined catalog of events that contained 36,793 AE waveforms, representing a potential array of generating mechanisms caused by interfacial processes (e.g., evaporation, imbibition, drainage or bubble collapse) and chemo-mechanical microscopic processes in expandable clay (e.g., volumetric changes that cause interfacial debonding, crack nucleation, etc.). Data are provide in Table 1 for the AE recorded per channel per sample type. There are notable differences in the number of AE recorded per transducer. This is attributed to the placement of the transducer relative to the event (which affects the collection angle), the coupling of the transducer to the sample, and environmental conditions. Some transducers will have difficulty recording a potential event if for example, the arrival amplitude is below the prescribed threshold of 25 dB. An AE occurring near one transducer then traversing the sample propagating through pre-existing cracks may be more difficult to record at transducers located farther away from the source or closer to a highly damaged zone.

2.3. AE Sensor Calibration

Calibration experiments were performed utilizing the methods described in McLaskey and Glaser (2010); McLaskey and Glaser (2012); McLaskey et al. (2015) and Wu and McLaskey (2018) for a ball impact source. We leave it up to the reader to extract the intricacies of this transducer calibration process which is outlined in the above mentioned references. For the calibration setup shown in Figure 2a, the sensors were coupled to the aluminum plate with hot glue. Several sources of different materials and diameters were tested, but data are only shown for a titanium ball source 0.8 mm in diameter with a corner frequency of 648 kHz. The ball source was vertically dropped three times from a drop height (DH) of 914 and 305 mm inside a transparent polycarbonate tube, onto a 102 mm thick (3-direction) Aluminum plate that was 610 mm in length and width (1 and 2 directions). The propagation speed of the longitudinal and shear waves in the Aluminum plate were 6,355 m/s and 3,805 m/s respectively. The calibration of the transducers were performed to compute the instrument response $I(\omega)$ which is obtained by deconvolving the theoretical displacement from the recorded signal (see Wu and McLaskey (2018)). Since DMDc is performed on data in the time domain, the calculated $I(\omega)$ is converted from the frequency domain to the time domain using the inverse Fourier transform, and the DMDc method is applied to these signals. The reconstruction accuracy for all transducer locations per DH for a titanium source of 0.8 mm diameter is shown in Figure 2b with inset for the transducer located at x_4, a representative waveform of $I(\omega)$ is shown in Figure 2c and the waveform reconstructed from the learned control signal in d for the transducer located at x_4.

2.4. Data Acquisition for Damage Characterization

At the end of the experimental drying and AE monitoring period, 3D X-ray tomographic reconstructions of the state of the internal structure of each fabricated rock analogue were obtained with a 3D X-ray microscope (Zeiss Xradia 510 Versa) with a 0.4X objective, with varied exposure times and voltages (see Table 2), power 10 W, Zeiss proprietary filters, and voxel edge length resolution of approximately 40 μm. The parameters used in scanning recipes designed for the 3D X-ray Microscope are chosen to gain the most optimal data with comparable grayscale X-ray intensity histograms which are used to perform thresholding segmentation. To improve the signal-to-noise ratio all 3D X-ray microscopy is performed at bin 2 which is a process that combines the charges

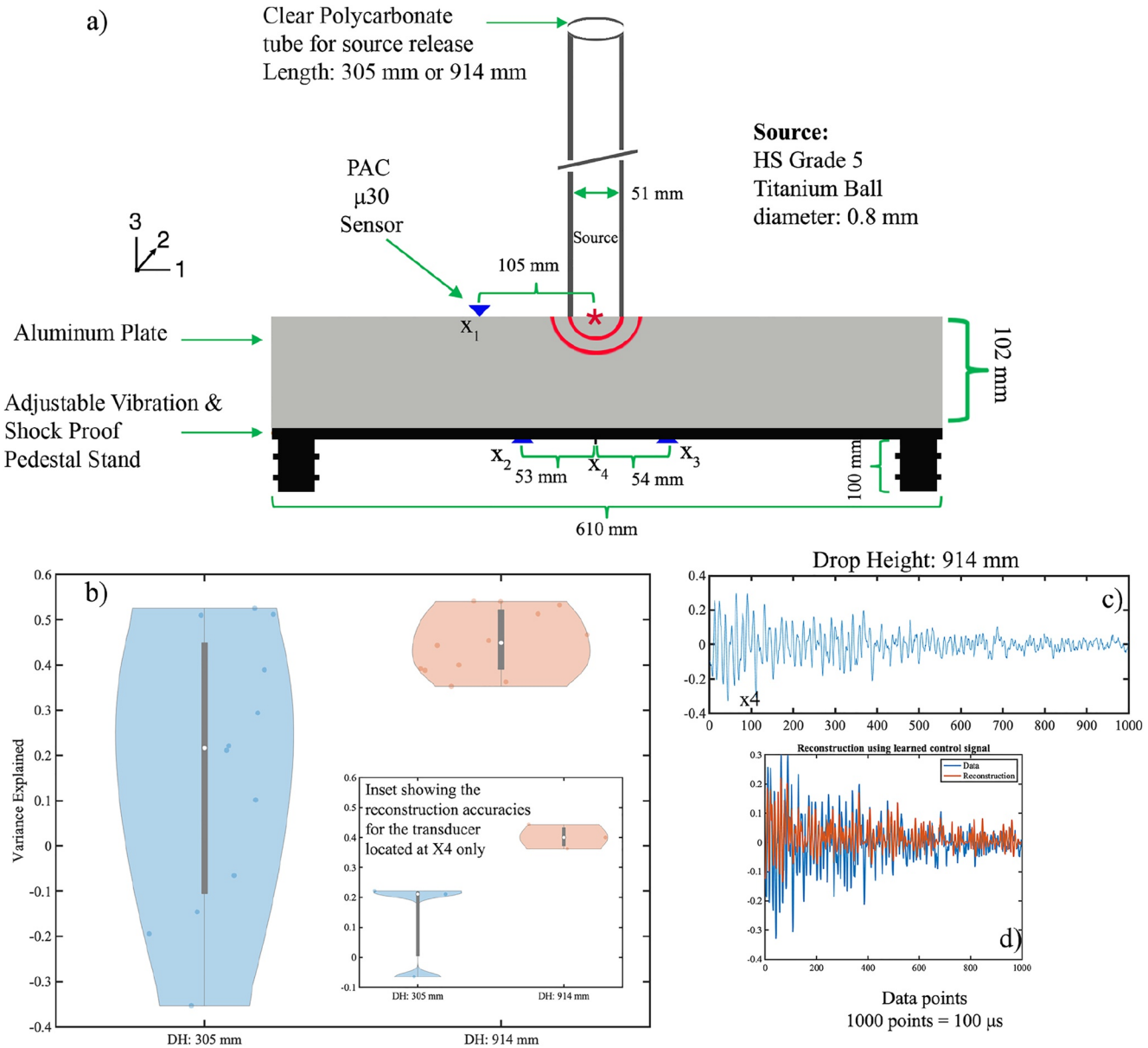


Figure 2. (a)Experimental setup for acoustic emissions Physical Acoustics Corporation (PAC) Micro-30 transducer calibration, (b) reconstruction accuracies for I(ω), (c) a representative waveform calculated for a drop height of 914 mm, and (d) the reconstruction of the data using the learned control signal.

Table 2
Zeiss Xradia 510 Versa X-Ray Microscope Settings for Imaging the Mortar, Localized and Distributed Synthetic Rock Samples at the 0.4X Objective and Bin 2

Sample type	Pixel resolution μm	Voltage (kV)	Power (W)	Exposure (secs)	Filter
Mortar	40.14	160	10	2	HE2
Localized Clay	39.81	120	10	2	HE1
Distributed Clay	39.43	140	10	1.1	HE1

from adjacent pixels to form a super-pixel. At bin 2 four (4) native pixels of the charge-coupled device are summed.

2.5. Image Data Analysis for Damage Characterization

All image processing which includes segmentation, data analysis, and visualization were performed using Object Research Systems (ORS) Dragonfly image processing software (ORS-Inc, 2020) and results are provided in Table 1. From the detailed 3D image data sets of the rock analogs, observations of the state of the microstructure and physical mechanisms responsible for macroscopic damage evolution and the generation of AE are examined which is crucial for the interpretation of the recorded AE data. The pore and fracture networks are quantified using segmentation procedures that rely on

the measured continuous grayscale which is a function of the material properties. All pixels are composed of a gray shade that varies from black (0) to white (255) and the value of each pixel carries information regarding the x-ray intensities which relate to the bulk density (relative density if no calibration is performed) of the material at that particular location. Darker areas are locations representative of material with lower bulk density, black areas signify voids, and brighter intensities (brighter greys or white) identity locations of higher bulk density. Features falling within similar gray bandwidths of the data measured intensity histogram, are apportioned into discrete groups which are defined as the undisturbed synthetic rock material, unconnected pores (e.g., round pores), or connected voids spaces (e.g., cracks). During refinement of each segmentation group, only features greater than 5-voxels are considered. The percent volume of voids per sample (Table 1) which is a measure of the effective total porosity was calculated using the total bulk volume of the sample and the segmented regions of interest (ROI) that represented the total void volume (connected and unconnected voids). The percent volume of material damaged during the advancement of the drying front also reported in Table 1, is estimated using an ROI of segmented rock that contain fractures and the bulk volume of the synthetic rock.

3. Machine Learning Model

3.1. Data Driven Dynamic Mode Decomposition With Control (DMDC)

The data-driven strategy presented in this paper is based on *dynamic mode decomposition* (DMD). DMD provides a linear model for the dynamics of the state space $\mathbf{x}_j = \mathbf{x}(t_j)$. It specifically finds the best fit linear model:

$$\mathbf{X}' = \mathbf{A}\mathbf{X} \tag{1}$$

where $\mathbf{X} = [\mathbf{x}_1\ \mathbf{x}_2\ \ldots\ \mathbf{x}_{m-1}]$ and $\mathbf{X}' = [\mathbf{x}_2\ \mathbf{x}_3\ \ldots\ \mathbf{x}_m]$ are temporal snapshots of the system that are offset by one time step. There are a number of variants for computing $\mathbf{A}$ (Kutz et al., 2016), with the *exact DMD* simply positing $\mathbf{A} = \mathbf{X}'\mathbf{X}^{\dagger}$ where † denotes the Moore-Penrose pseudo-inverse. However, the optimized-DMD (Askham & Kutz, 2018) (opt-DMD) and bagging-optimized DMD (Sashidhar & Kutz, 2021) (BOP-DMD) provide algorithms that provide substantial performance gains when considering noisy data. However, such improved algorithms have yet to be incorporated with control architetures.

Dynamic mode decomposition with control (DMDc) (Proctor et al., 2016), capitalizes on all of the advantages of DMD and provides the additional innovation of being able to disambiguate between the underlying dynamics and actuation signal $\mathbf{u}_j = \mathbf{u}(t_j)$. For a matrix of input signals $\mathbf{U} = [\mathbf{u}_1\ \mathbf{u}_2\ \ldots\ \mathbf{u}_{m-1}]$, DMDc will regress instead to the linear control system

$$\mathbf{X}' = \mathbf{A}\mathbf{X} + \mathbf{B}\mathbf{U}. \tag{2}$$

DMDc only utilizes snapshots in time of the state space and control input, making it compelling for systems whose governing equations are unknown. The governing matrices ($\mathbf{A}$ and $\mathbf{B}$) along with the control signal ($\mathbf{U}$) produce a predictive model, such that the state of the system far in the future can be predicted. For instance, the third time step can be estimated from the first via:

$$\mathbf{x}_3 = \mathbf{A}(\mathbf{A}\mathbf{x}_1 + \mathbf{B}\mathbf{u}_1) + \mathbf{B}\mathbf{u}_2 \tag{3}$$

DMDc can also be improved by enriching the observations of the dynamics in order to produce a more accurate Koopman (linear) operator (Proctor et al., 2018).

3.2. Learning Control Signals via Sparse Optimization

DMDc (*eq*.2) can be thought of as an error minimization problem over the dynamics and actuation matrices, $\mathbf{A}$ and $\mathbf{B}$. If the control signal is unknown, the minimization must be extended to the control signal $\mathbf{U}$ itself. However, there is now a trivial solution where the control signal dominates the model: $\mathbf{X}_2 = \mathbf{B}\mathbf{U}$ with $\mathbf{A} = 0$. For this reason, an assumption must be made about the control signals. In this case, the statement that these signals are sparse is directly interpretable, and means that initiating events should be rare. This "sparsity constraint" can be expressed in a mathematically precise way using the ℓ_0 norm:

$$\min_{\mathbf{A},\mathbf{B},\mathbf{U}} \left[||\mathbf{A}\mathbf{X}_1 + \mathbf{B}\mathbf{U} - \mathbf{X}_2||_2 + \lambda\, ||\mathbf{U}||_0 \right] \tag{4}$$

Directly solving this optimization problem is extremely difficult, although there are efficient algorithms in certain cases (Jewell & Witten, 2018). More recently, a convex relaxation of the ℓ_0 to an ℓ_1 norm is often solved (Donoho, 2006), though this has been recently shown to lead to errors in its selection pathway (Su et al., 2017). We use a different approximation, the sequential least squares thresholding algorithm as described in (Brunton et al., 2016a), which has been shown to converge to the minima of the original ℓ_0 problem (Zhang & Schaeffer, 2018; Zheng, Askham, et al., 2018). The code uses a modified form of Akaike Information Criteria (AIC) (Akaike, 1973; Sakamoto et al., 1986) to choose the best iteration of the algorithm. This modification reduces the importance of each element of the matrix **U**, because each entry is not a "global parameter" as in the original derivation of AIC; more information and derivation is given in the supplement. The matrix **U** in this algorithm is additionally constrained to be positive, for better interpretability as "on" transition signals. Note that the elements of **B** can be negative, and thus the action of control signals may be positive or negative.

3.3. Data Preparation for DMDc

Application of DMDc to acoustic emission data is advantageous because it is an unsupervised learning method that requires no training data and can be applied to the raw acoustic emission full waveforms. Although there are some theoretical studies on noise in the original DMD algorithm (Dawson et al., 2016), these have not been extended to the control setting and it is not trivial to do so. In addition, because this method is initialized using the residual as calculated by setting $\mathbf{U} = 0$ in Equation 4, the results are sensitive to filtering. In particular, the assumption that **U** is sparse in Equation 4 is equivalent to sharp discontinuities in the original time series, which should not be filtered away.

DMD requires multiple dimensions in order to produce oscillatory dynamics, and will otherwise produce only exponential decay or grow. However, time-delay embedding can be used in the single-dimensional case to increase the dimensionality, as in related works (Brunton et al., 2017; Champion et al., 2019). This procedure involves stacking the data, such that the new data matrix contains vectors that are offset by one time step:

$$\mathbf{X} = \left[\vec{x}_1 \ \vec{x}_2 \ \dots \ \vec{x}_{m-1}\right] \tag{5}$$

where **X** is the final matrix "embedded" with dimension m, and with final shape $(n-m)xm$. $\vec{x}_i$ is the original time series of shape $nx1$ with index starting at 1 and ending at $n-m$. The embedding dimension m is a free parameter.

No other normalization or amplitude modulation was applied to the data.

3.4. Quantification of Reconstruction Accuracy

Reconstruction accuracy (variance explained) (shown in Figures 7 and 9) is quantified via the following equation:

$$Accuracy = 1 - \frac{|\mathbf{X} - \mathbf{X}_{reconstructed}|_2}{|\mathbf{X}|_2} \tag{6}$$

Where **X** is the recorded data, $\mathbf{X}_{\text{reconstructed}}$ is the reconstructed data obtained using Equation 3. If the residual is greater than the norm of the original data, that is, if the reconstruction of the model is worse than a flat line at 0, then the value of the accuracy will be <0.

4. Results

4.1. Characterization of Damage in Fabricated Rock Analogs

The drying experiments on the analogue rock samples were designed to measure AE signals from samples where the AE are generated only from mechanisms that occur during fluid movement (Mortar) and from samples (Localized & Distributed) where additional AE generating mechanisms exist in addition to those emanating from the advancing drying front, namely cracking from volumetric changes in the clay during dehydration. Significant differences are observed in the internal structure of the different sample from X-ray images taken at the end of the experimental drying period. Figures 3a–3c provide a comparison of 2D images taken from 3D X-ray tomographic reconstructions of the 3 sample types after 6 days of drying. No cracks are visible in the Mortar sample at a resolution of 40 μm (Figure 3c). The Mortar sample was composed of OPC and Ottawa sand only (Figure 3c), and

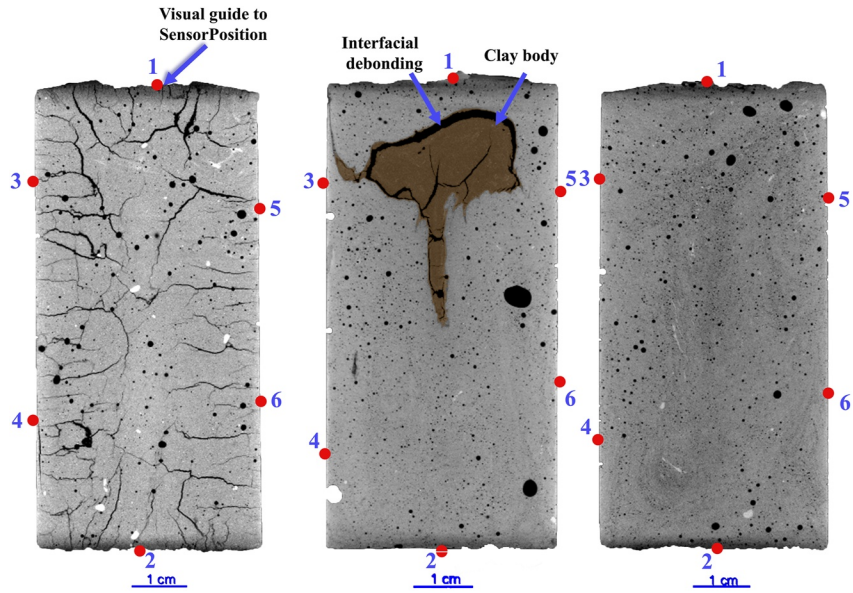


Figure 3. 2D image slices from the center of the (a) distributed clay, (b) localized clay, and (c) clay-free (mortar) samples with approximate sensor placement, marked by red dots to guide the eye.

the internal structure is mainly dominated by pores, fairly spherical voids, with a pore volume of 3%. The round pores are the result of remnants of air entrainment that occurs during mixing, setting and vibration compaction. All sample types contain these pores to some degree.

The crack distribution in the Localized and Distributed samples differs. In the Localized sample, cracks are observed internal to (Figure 3b) and around the perimeter (Figures 4d–4f) of the clay structure but not within the matrix. In Figures 4d and 4e, the body of clay and the discontinuities are represented in brown (clay) and red (fractures), respectively. The damage in the localized sample formed because of interfacial debonding at the clay-cement interface that occurred from shrinkage of the clay in response to moisture loss, and are confined to regions where the swelling clay material was located. These discontinuities evolved as the localized clay assemblage shrank (Figure 3b), and the fractures are only observable within or surrounding the zones of swelling clay (Figures 4d, 4e, 4f, and 3b). However, in the Distributed sample, an extensive crack network is observed (Figures 3a and 4a–4c). The segmented fracture volume is approximately 10% (percent ratio of the volume of voids to the volume of solid material), and the volume of the damaged zone 95% (Table 1.) In Figures 4a and 4b, the cracks are highlighted in red. A key question is whether the AE signals differ in terms of frequency content, phases, and decay given the different distribution of induced cracks from clay shrinkage and the fact that the Mortar sample contained no cracks.

4.2. Conventional Analysis of Acoustic Emission Data

All sample types generated acoustic emissions during moisture loss, with the highest number of acoustic emissions occurring in the distributed clay sample (34,227 AE). The number of acoustic emissions recorded per specimen type and sensor location are given in Table 1, and the approximate sensor locations are marked with red dots in Figure 3. The low count number for Sensor 2 on the Localized sample is attributed to sensor decoupling from the sample surface. The distribution of events count (number of hits per every 2 hours) over the duration of the monitoring period for the Mortar and Localized samples are very similar and quite different from the event distribution from the Distributed sample Figure 5. Acoustic emissions were generated immediately from the onset of drying. During the initial 2 hr of drying, 0.12%, 28.08% and 35.53% of the total number of events were measured on the Distributed, Localized and Mortar, respectively. Within the first 10-hr, the Mortar sample

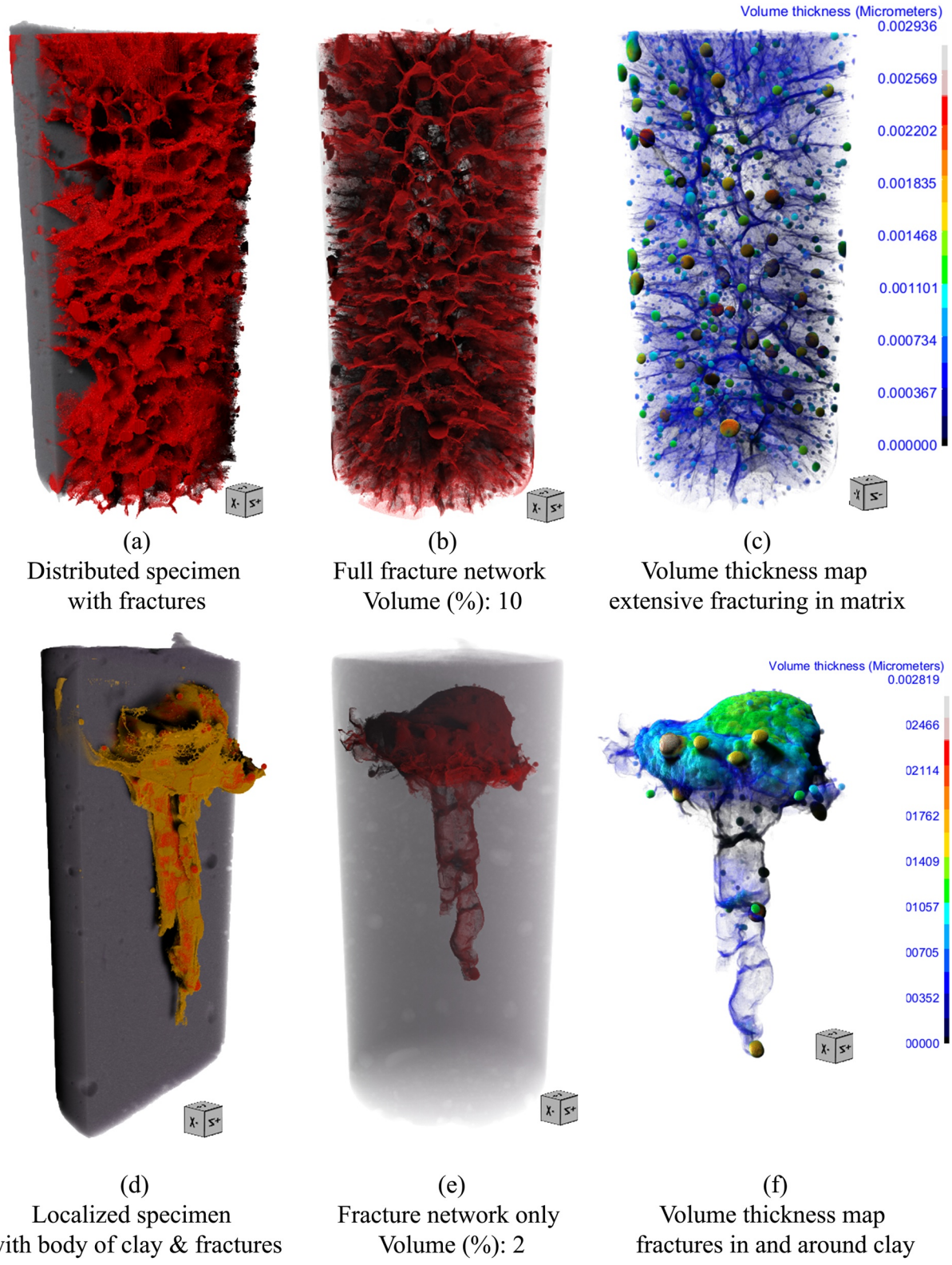


Figure 4. 3D visualizations of the interior of the Distributed (a)-(c) and Localized (d)-(f) specimens showing fracture damage (a and d), the full crack network (b and e), and the volume thickness of the segmented crack network within the distributed sample (c) and within and surrounding the body of clay (f). The volume thickness is a color-coded map generated from segmented regions of interest (ROI) that provide referential values of the local thickness between boundary points. The volume percent of clay, damage, and discontinuities (voids) are reported in Table 1.

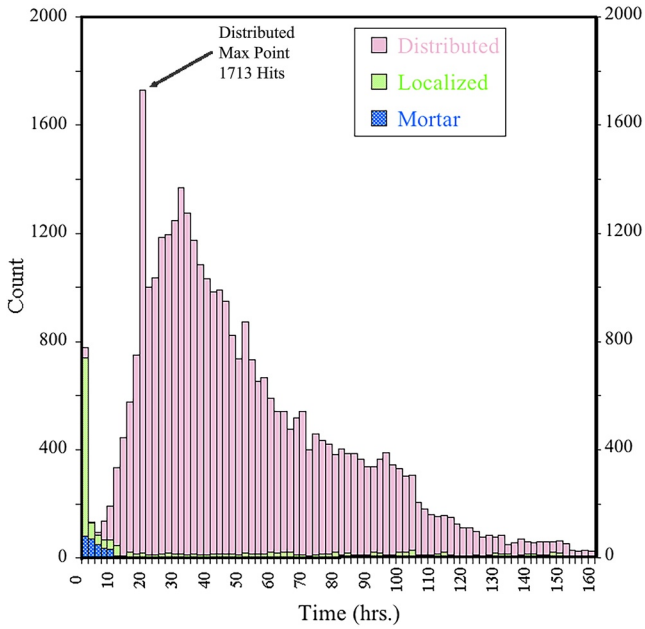


Figure 5. Stacked histogram of the total number of acoustic emissions recorded per hour over the duration of 160 hr of drying under ambient conditions for all samples, distributed clay, localized clay, and clay-free. Each bar represents 2-hr of drying time.

emitted 92.81% (271 AE hits) of the total number of events recorded for this specimen. The Localized sample emitted 44.08% (815), while the Distributed sample generated <1% (247) of its total AE count. A steep decline in the number of AE is observed for the Mortar and Localized specimens, and the event distribution is log-normal. For the Distributed sample, 11% (3,982 hits) of the total number of AE for this specimen was recorded after 20-hr of drying. This sample continued to generate AE at an average rate of 550 hits/hour for the next 20 hours. The Mortar specimen generated no events after 20 hr (Figure 5), while the Localized sample generated <1% of its total AE hits every 2 hours.

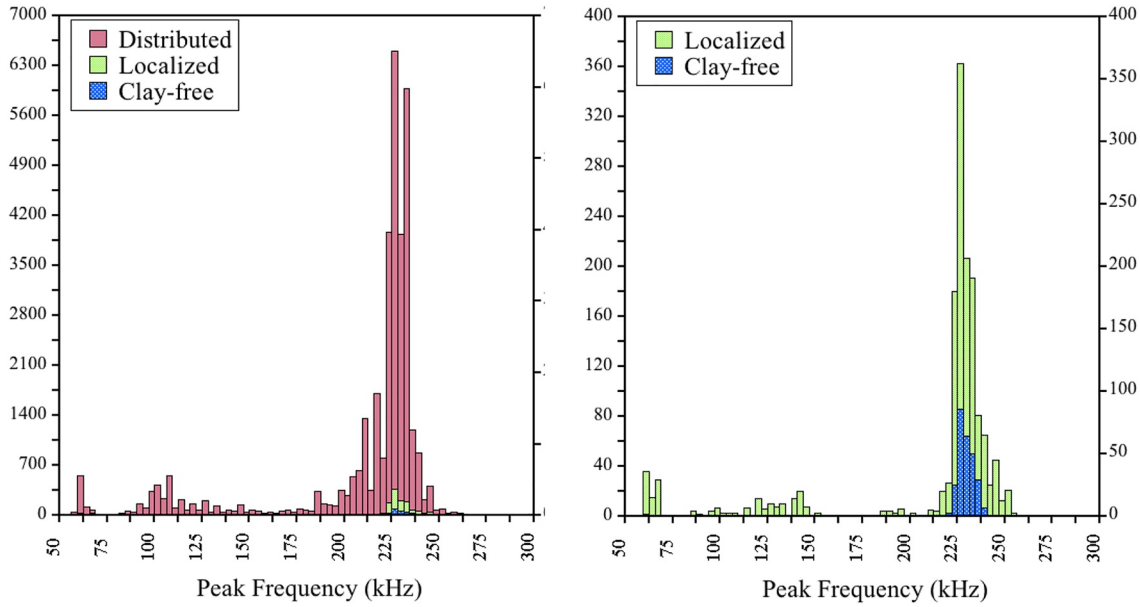
The traditional characteristics (amplitude, frequencies, etc.) of the acoustic emission signals were evaluated. In Figure 6, the distribution of the peak frequency (kHZ) is plotted against the event count (Figures 6a–6b) and amplitude (Figure 6c) for the 3 sample types. These plots are used to explore the data space and are an analysis tool that can indicate if the frequency distributions are similar across the samples, thereby illuminating if one particular synthetic rock type is dominated by a higher quantity of low peak frequency or low amplitude events (which may relate to the occurrence of damage), and/or if the majority of AE for a sample occur within a distinct frequency range. Such is observed in the AE data recorded for the Mortar sample where the peak frequency of the AE occurred only in the 225–235 kHz band (Figures 6b and 8a). However, the data from the Distributed and Localized samples fell into three different frequency bands (55–75, 100–150 and 210–265 kHz) (Figure 6a). The broadness of the amplitude-peak frequency distribution for the Distributed sample which had the most complicated crack network (Figure 4a) shows the complexity of the signal characteristics from samples with both fluid movement and induced-cracking, and potentially from signal interactions with multiple cracks along the propagation path.

4.3. Learning Dynamical Models With DMDc to Differentiate Data

Here we use sparse regression to discover control (onset) signals, and then apply DMDc to determine if the AE signals from the 3 types of samples exhibit similar or different dynamic responses by reconstructing the signals. The acoustic emission waveforms recorded in the Mortar samples are robustly reconstructed by the linear dynamics and control input (Equation 4). As shown in Figure 3c this sample type sustained no visible damage within its microstructure. These data when reconstructed with DMDc are characterized by higher reconstruction accuracy (RA) (Figure 7a). This demonstrates that the dynamical system corresponding to an AE (linear dynamics) can correspond to the underlying physical process (drying mechanism).

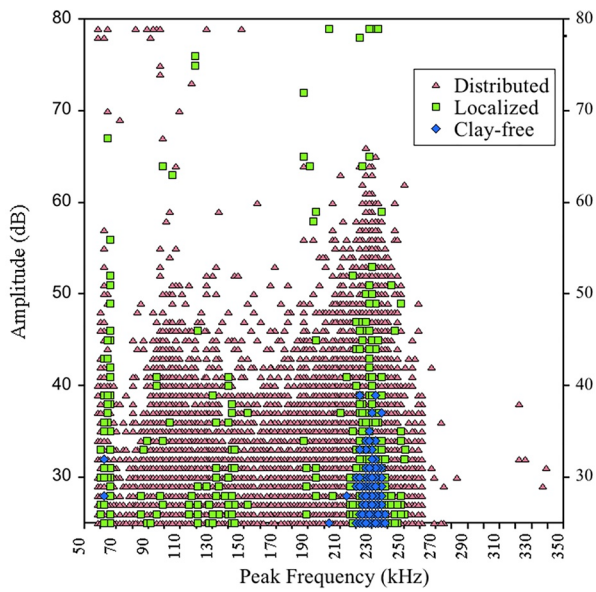
As mentioned in the experimental results this sample is uniquely confined to a small distinct frequency band Figure 8a clay-free, when compared to the other sample structures defined by increasing complexity, that is, Figure 8a-localized and -distributed clay specimens. Examining representative waveforms at different levels of accuracy highlight an existing similarity in the underlying intrinsic dynamics across all sample structures. In Figure 7b Level I, the representative waveforms shown are similar for all three sample structures, and on Level two for the clay-free and localized clay structures in particular. As the reconstruction accuracy drops below 0.3, the representative AE waveforms are no longer efficiently reconstructed and cannot be sufficiently described as a linear dynamical system (Figure 7b Level III and IV). Essentially, Figure 7 is an example of the simplest dynamical system model to be obtained from acoustic responses that occur in the synthetic rocks, and with linear dynamics as presented in Equation 1, a subset of signatures in the AE waveforms can be explained. The values of each level (I - IV) are not classification boundaries, but rather show example time series with different qualitative behavior.

Figure 7 additionally shows that the initiating spontaneous events can be discovered directly from data. Representative acoustic emission waveforms are plotted at levels of peak frequencies from high (level I) to low (level



(a)
Peak Frequency Count
All Samples

(b)
Peak Frequency Count
Localized + Mortar



(c)
Peak Frequency vs. Amplitude
All Samples

Figure 6. (a) Peak frequency (kHz) versus event count for all specimen types and (b) for Localized clay and clay-free samples, along with the peak frequencies (kHz) versus amplitudes (dB) which are displayed in (c) for all sample types the distributed clay, localized clay, and clay-free samples.

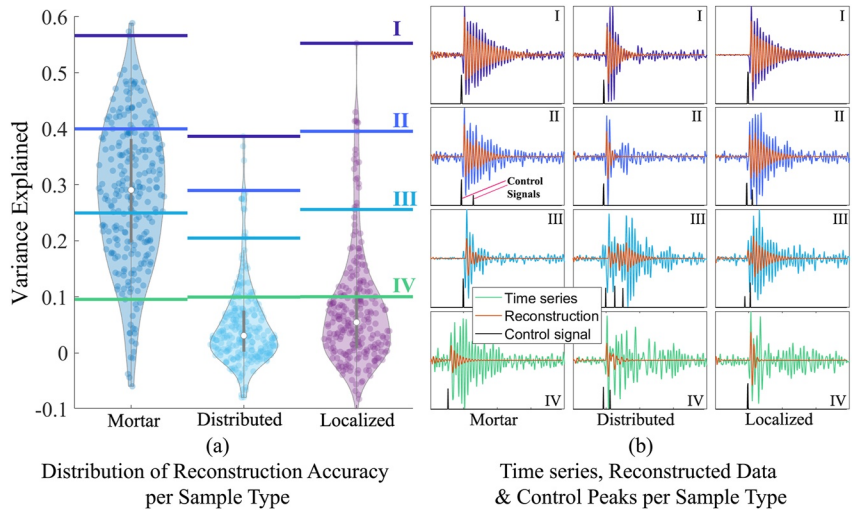


Figure 7. A subset of waveforms displaying linear dynamics. (a) Distribution of the variance in reconstruction accuracy per sample type, where distinct clusters are observed, and (b) representative time-series, reconstructed data (red) and, control signals for each sample type at the respective variance line. The values of each level (I – IV) are not classification boundaries, but rather show example time series with different qualitative behavior.

IV which is dependent on the sample structure) in Figure 8b, here it is quickly observed that AE recorded for the different sample structures do not have a clear signature in the frequency space. Moreover, peak frequency clusters do not clearly correspond to a distinct difference in underlying physical processes.

4.4. Differentiating Acoustic Emission Type With Reconstruction Accuracy (RA) and Peak Frequency

Clustering the AE obtained from monitoring the experimental rocks using the peak frequency only (Figure 9) is not a sufficient method to appropriately distinguish the different sample structures and phenomena that produce AE. Examining the reconstruction accuracy obtained from DMDc (Figure 7) results in an explicit differentiation between the different synthetic rock structures, their respective level of complexity and acoustic emission waveforms. Additionally, comparing the RA against the peak frequency, as presented in Figure 9a, provides some insight regarding the AE generating mechanisms. As the contours overlap, relationships between the analogue rocks and the AE generated by them are exposed. A fraction of the AE data obtained for the clay-bearing structures occur at lower RA and experience no overlap with the AE obtained for the clay-free structure. Interrelated AE are observed for all samples within the confines of the contours representative of reconstructed data learned from data recorded for the clay-free sample.

5. Discussion

Porous materials attract and hold water molecules in quantities that depend directly on the ambient relative humidity (Kallel et al., 1993). Upon drying, especially in clay-bearing porous media many individual or coupled mechanisms occur that result in the release of energy. An important unsolved problem is classifying the acoustic emission according to the generating mechanism using only AE monitoring techniques that non-invasively investigate the internal dynamics of materials. Here two non-destructive methods, X-ray microscopy and acoustic emission monitoring, are used to analyze the materials and mechanism occurring during drying. To fully exploit the acoustic emission data, an unsupervised ML algorithm, an extension of DMDc, is applied as a modeling tool for analyzing the spatial-temporal spread, reconstructing relevant aspects of the waveforms, and providing a means to partially cluster the AE according to underlying physical generating mechanism.

Analysis of post-drying 3D reconstructed tomographic data (2D Figure 3 and 3D Figure 4) reveal clear differences among the analogue rocks. Alterations in the state of the porous media (from high moisture content to low moisture content) and changes within the matrix of the internal stress state are expected to generate acoustic emissions. Since the clay-free (Figure 3c) sample experiences no observable damage in the form of discontinuities it

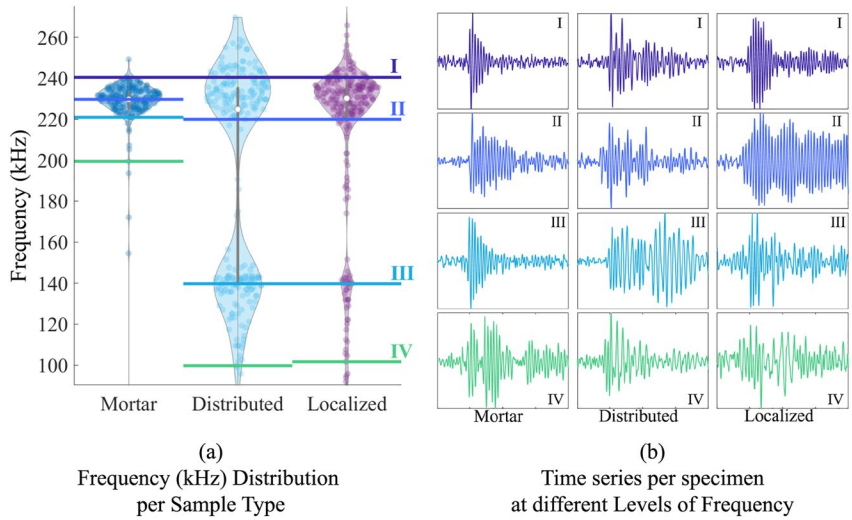


Figure 8. (a) Peak frequency (kHz) distribution per specimen type and, (b) Representative time-series for each analogue rock type at approximately corresponding levels frequency where possible.

is simply the transformation from a wet porous structure to a dry porous medium that effectuated the recorded acoustic emissions (Figure 5). These events were therefore attributed to the mechanisms induced by the infiltration of the drying front as air intrudes the sample structure, such as those linked with interfacial processes induced by evaporation, imbibition, and drainage. Waveforms from these AE events are comparatively simple and can be reconstructed and described using a linear dynamical system (Figure 7b). Thusly, all waveforms of similar nature and high DMDc reconstruction accuracy are attributed to the removal of moisture and infiltration of air. Importantly, reconstruction accuracy is not a trivial function of simpler properties of the time series, like dominant (peak) frequency, making it a novel way to reveal underlying mechanistic information.

In this study, we additionally consider internally evolving stress states of the more complex localized and distributed clay samples. Damage within these structures is observed in the 3D X-ray microscopy data (Figures 3 and 4). In the background reference matrix of the localized clay sample no damage is observed (Figure 3b), instead all damage are localized and only exist in and around the assemblage of clay. Much of the observed fracturing is

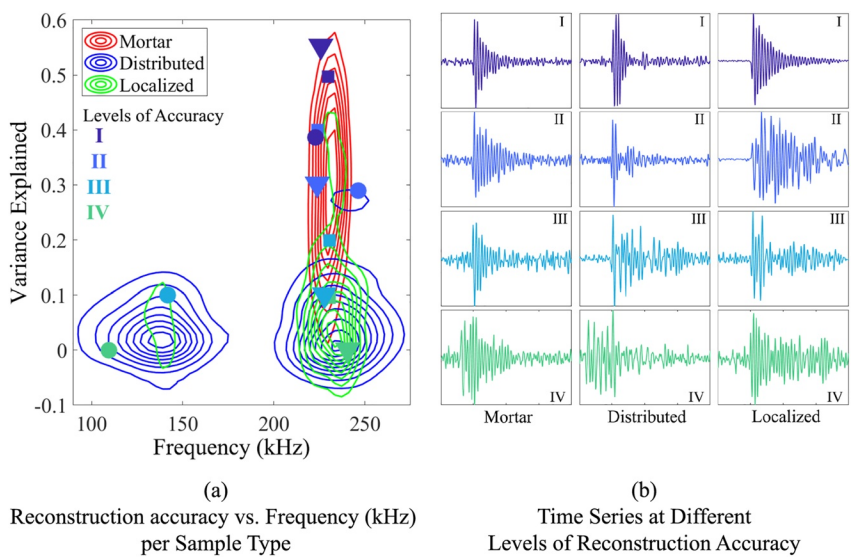


Figure 9. Peak frequency clusters. (a) Contour plot of the distribution of reconstruction accuracy versus peak frequency (kHz) per sample type and, (b) example traces extracted at the different reconstruction accuracy levels where the Mortar (square), Distributed (circle), and Localized (triangle) occur.

related to the unique properties of the swelling clay material that allow them to swell or shrink depending on the presence and chemistry of the in-situ fluid, and the conditions of the local environment. Besides being related to the mechanisms that occur as the drying front infiltrates the sample, the acoustic emissions generated here are also the result of macroscopic processes such as interfacial debonding, crack nucleation, propagation, and coalescence which are influenced by the microscopic chemo-mechanics of the shrinking clay particles. This also applies to the distributed clay sample, extensive fracturing was observed within the matrix, where fractures extended from the exterior all the way to the center of the sample (Figures 3a, 4a, 4b, and 4c). Multiple dominant frequencies are also present in these synthetic rocks (Figure 6c). Although the grouping of frequencies observed in Figures 6c and 8a and event counts (Figure 6b) seemingly cluster the data, dissimilarities in the waveforms in these respective groups are readily apparent (Figure 8b) thus indicating that the AE in these clusters may not share similar dynamics. As with the localized clay sample, dominant frequency ranges (Figures 6c and 8a), groupings by event counts (Figure 6a) and dissimilarities in waveforms extracted from the same ranges are observed (Figure 8b). High reconstruction accuracies (with the exception of one signal) were observed for $I(\omega)$ obtained through the ball impact sensor calibration method. Since DMDc is a linear method, and the calibration framework models $I(\omega)$ as a linear time-invariant system this output supports the abilities of DMDc in its current form. This begs the question of, are these well reconstructed, high accuracy waveforms only a representation of the sensor response? The recorded signal from an event is a convolution of the acoustic emission and the sensor response. As long as the acoustic emission is broadband, that is, is not a delta function, the measured signal contains information from the original emitted signal. If it was merely the sensor response, different frequencies would not be observed. The waveform contains information regarding the in-situ process, the state of the material, and its structural integrity. Furthermore, drying, that is, fluid movement is a dynamic process which can involve several simultaneously occurring phenomena and have been observed to emit AE by other authors (see DiCarlo et al. (2003); Chotard et al. (2006); Michlmayr et al. (2012); Moebius et al. (2012); Grapsas and Shokri (2014)). These waveforms are not solely a representation of the instrument response (which includes the cables, digitizer, coupling, recording instruments, etc). Future work with more detailed knowledge of the generating mechanism of individual waveforms which will include expanding on the calibration methods, will aid in efforts to assign a meaning to these putative clusters.

Assumptions and Limitations of Classic DMD and DMD with control (DMDc) which are linear methods, but many generating processes in the real world obey nonlinear dynamics. Specifically, DMD without control can only capture and reconstruct dynamics that are exponential growth, exponential decay, or oscillations. In any experimental settings that involves the recording of waveforms, the number of dynamical modes that can be captured is limited by the number of simultaneous measurement channels, with real data (no imaginary component) requiring a minimum of two channels to capture a single oscillation frequency. The addition of control to the DMD method enhances the complexity of the dynamics that can be captured. In principle, arbitrarily complex time series can be produced (or fit) by this method, as long as all non-linearities are included in the control signal. The method in this paper extends classical DMD to discover the control signal as well as the dynamics. Thus there exists a trivial solution: the entire time series is control, and nothing relates to the dynamics, in Equation 1, $\mathbf{A} = 0$ and $\mathbf{U} = \mathbf{X}$. Regularization can be used to penalize this trivial solution, and proper initialization away from this is key in practice. This work and the associated code (Fieseler, 2019) uses a sparsity penalty, which is interpretable as control signals that are rare in time, for example, sparse initiating cracking events. However, different regularization could be used, for example, penalizing the total variation of the control signal, which is interpretable as an input which is piecewise-smooth in time. As DMDc is an unsupervised ML method, the data are not trained, and no training process or other fitting is necessary to apply this method to the various time series. Each dynamical mode, that is, frequency of oscillation, growth, or decay time scale, and learned control signal is independent, except for the sharing hyper-parameters which are consistent across all datasets. Thus, differences in the number of time series has no effect on the reconstruction accuracies or other results.

6. Conclusion

In geophysics there has always been a great need for understanding and prediction of material behavior and responses of system from the laboratory to the large field scale. A major goal that drives many geophysical problems is the ability to model and predict subsurface materials and processes at high temporal and spatial resolutions. This requires a vigorous understanding of material properties, material behaviors, and the response

of the recording equipment and the materials when subjected to individual or coupled processes in different environments (laboratory controlled conditions, at the field scale in geological reservoirs or on a larger global scale), for example, in-situ static and/or evolving environments at different scales. To achieve and improve fundamental understanding of processes and their relationship to the host geologic environment, researchers often use arrays of acquisition equipment (e.g., seismometers, geophones, transducers) to passively monitor and detect signatures of different process, for example, failure, damage propagation, and fluid flow processes, which can result in big data sets. The application of data-driven methods like DMDc provide new algorithms to leverage large quantities of data for scientific insight and understanding of geophysical and geo-technical materials and processes. The sparse regression method presented in this paper (DMDc) extends previous algorithms (DMD) and provides a new feature space to extract AE waveforms where the reconstruction accuracies are significantly higher. Attractive features of the presented ML approach lies in the fact that it is an accessible unsupervised data-driven method that can be used on large datasets obtained through non-destructive methodologies (e.g., passive seismic monitoring in the field or acoustic emission monitoring in a laboratory environment), requires no training data, and is purposed to learn and extract the underlying dynamics.

Here DMDc makes use of a sparsity penalty while learning and collectively discovering the underlying dynamics and control signals. The approach discussed in this paper simultaneously and robustly characterizes the dynamics and onset of AE waveforms that are characteristic to a subsection of the data. This algorithm also successfully reconstructs a subset of recorded waveforms that are believed to be representative of fluid front movement as these particular waveforms are the only AE recorded during the drying of the clay-free, crack-free (mortar) samples. Such AE associated with mechanisms of the advancing drying front that are defined by similarities in their underlying dynamics, are recorded in all fabricated rock samples. The phenomena that occur during the advancement of the drying front that produces these AE are closer to impulsive sources than what is observed for the clay- and crack-bearing physically complex distributed sample, and the reconstruction of such waveforms incorporates a simple impulse ring down. The outcome of the application of DMDc to these data indicate that these particular AE do not originate from mechanisms that will lead to long-term damage and such impulses are learned without a priori information. Moreover as the impulse ring down fails to capture the information embedded in waveforms when clay is distributed within the relatively homogeneous reference medium (mortar) it is indicative that the mechanisms that lead to damage, for example, cracking and debonding, coupled in time with an advancing drying front has a more complicated impulse than a simple ring down. Essentially this shows that the elastic wave energy generated from the advancing drying front is different from that generated during cracking as shown by a deterioration of the reconstruction accuracies for data representative of the latter. The insufficient understanding of these complex waveforms is essentially the result of several coupled mechanisms that overlap in the waveform spatially and temporally. Cracking and debonding, for example, which occur during the drying of clay-bearing analogue rocks involve tensile and shear forces that may manifest at different times or simultaneously during damage progression and may spatially coincide with the AE generating mechanisms of the advancing drying front.

While the AE waveforms with single, clear initiating incidences may be of interest in many areas of geophysics (e.g., research involving non-turbulent fluid movement in porous materials), DMDc should be improved to leverage the capabilities of this method to extensively explore waveforms that record overlapping information that possibly relate to the stages of cracking and progression of damage in a porous medium. As shown herein, the complex waveforms which are exclusive to the clay-bearing samples that contain damage in the form of cracks, fall outside the scope of the DMDc method in its present form. Incorporating additional information to the method will aid in the reconstruction of the functional form of the crack induced waves. Obtaining a robust representation of such waveforms can point to underlying physics which may not be currently evident. These waveforms should be further analyzed to improve the method to not only robustly reconstruct the AE waveforms but to also extract the finer scale differences embedded and overlapping in such waveforms. In follow-up work, nonlinear methods, for example, neural networks, can be employed to find a basis in which the dynamics are approximately linear (Lusch et al., 2018). Alternatively, the basis functions can be expanded by assembling more domain knowledge of the expected differential equations of the waveforms (Brunton, Brunton, et al., 2016; Brunton et al., 2016b) and such enhancements can lead to better reconstruction of different classes of complex signals. Developing a two tier ML model by coupling DMDc with methods that can extract characteristics of the non-linear dynamics will allow for further clustering of the AE waveforms into sections where data are related to fracture nucleation and propagation or interfacial debonding. As non-destructive sensor technologies that capture acoustic emissions evolve,

analysis of the recorded time-series data can be coupled with modern ML methods to exhaustively explore the structure and composition of porous materials. This would expand the possibility of predicting the state of the system at some level of accuracy and the physical or acoustic mechanisms that generate AE thus allowing users to fully exploit the scientifically relevant information embedded in time-series geophysical data.

Currently there are several emerging methods that allow for a broader viewpoint of building models directly from noisy data, recent innovations in DMD, for instance, have shown that statistical bagging methods can greatly increase the discovery of robust, accurate and stable linear models (Askham & Kutz, 2018; Sashidhar & Kutz, 2021). This has motivated the use of ensembling and bagging for building nonlinear, parsimonious dynamic models (Fasel et al., 2022; Hirsh et al., 2021). Such non-linear dynamic models have the potential to unravel the contemporaneously overlapping information that are representative of AE or seismic events generated by different processes, for example, fracturing, slip on faults, chemical reactions, or the various phenomena that transpire during fluid movement through porous media. This would promote better understanding of such complex waveforms, and shed light on difficult questions such as; Can the source of an acoustic emission occurring in the subsurface be predicted or identified? Can we discriminate events related to individual or coupled process, such as the emergence of damage as a result of cracking induced by fluid movement and changes in the water content during interactions with subsurface clay-bearing porous mediums (e.g., at the interface of natural rock and engineered barriers in nuclear waste disposal repositories)? The value of extending DMDc can be captured by the fact that even when the correct variables are unknown, such data-driven methods can be used to discover advantageous latent representations that are often directly related to unknown and not directly measured physics (Bakarji et al., 2022; Champion et al., 2019; Chen et al., 2019). Several methods are available in open source python notebooks (Demo et al., 2018; Kaptanoglu et al., 2021), making their use and implementation in a discovery pipeline efficient and advantageous. All these methods, which are aimed at providing interpretable models, can also be used in combination with deep learning (Bakarji et al., 2022; Champion et al., 2019; Chen et al., 2019; Gin et al., 2021; Lusch et al., 2018). Thus data-driven modeling (Brunton et al., 2016a), like DMDc, offer flexible and robust mathematical framework for advancing scientific efforts and unraveling scientific questions.

Data Availability Statement

Software: The publicly accessible MATLAB (MATLAB, 2021) toolbox, Geophysics Control version, prepared by CSF that was used to perform the analysis of acoustic waveforms with the DMDc method is openly available for download from GitHub (Fieseler, 2019). url: https://github.com/Charles-Fieseler-Vienna/Geophysics_control Licensed under MIT License. *Data*: The acoustic emission waveforms collected during the monitoring of moisture loss (under ambient laboratory conditions), in synthetic rock samples composed of OPC and sand, and OPC, sand and distributed or localized masses of montmorillonite clay are provided by CAM and LJPN, and is hosted by Purdue University Research Repository (PURR) Publications. The publicly accessible version 1.0 publication bundle of the ”Data for Characterization of Acoustic Emissions from Analogue Rocks using Sparse Regression-DMDc” can be found with the following information: Publication repository: Data for Characterization of Acoustic Emissions from Analogue Rocks using Sparse Regression-DMDc v. 1.0 (#4077): https://doi.org/10.4231/4K64-4818 url: https://purr.purdue.edu/publications/4077/1 Licensed under CC0 1.0 Universal.

Acknowledgments
LJPN and CAM acknowledge support of the sample fabrication, acoustic emission measurements and 3D X-ray microscopy portion of this work by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Geosciences Research Program under Award Number (DE-FG02-09ER16022). The work of CSF was supported in part by the Air Force Office of Scientific Research MURI FA9550-19-1-0386. The work of JNK was supported in part by the US National Science Foundation (NSF) AI Institute for Dynamical Systems (dynamicsai.org), grant 2112085.

References

Akaike, H. (1973). Theory and an extension of the maximum likelihood principal. In *International symposium on information theory*. Akademiai kaiado.

Andreani, M., Gouze, P., Luquot, L., & Jouanna, P. (2008). Changes in seal capacity of fractured claystone caprocks induced by dissolved and gaseous CO_2 seepage. *Geophysical Research Letters*, *35*(14), L14404. https://doi.org/10.1029/2008gl034467

Askham, T., & Kutz, J. N. (2018). Variable projection methods for an optimized dynamic mode decomposition. *SIAM Journal on Applied Dynamical Systems*, *17*(1), 380–416. https://doi.org/10.1137/m1124176

Bakarji, J., Champion, K., Kutz, J. N., & Brunton, S. L. (2022). Discovering governing equations from partial measurements with deep delay autoencoders. arXiv preprint arXiv:2201.05136.

Bergen, K. J., Johnson, P. A., de Hoop, M. V., & Beroza, G. C. (2019). Machine learning for data-driven discovery in solid Earth geoscience. *Science*, *363*(6433). https://doi.org/10.1126/science.aau0323

Biswas, S., Castellanos, D. F., & Zaiser, M. (2020). Prediction of creep failure time using machine learning. *Scientific Reports*, *10*(1), 1–11. https://doi.org/10.1038/s41598-020-72969-6

Bolton, D. C., Shreedharan, S., Rivière, J., & Marone, C. (2020). Acoustic energy release during the laboratory seismic cycle: Insights on laboratory earthquake precursors and prediction. *Journal of Geophysical Research: Solid Earth*, *125*(8), e2019JB018975. https://doi.org/10.1029/2019jb018975

Brunton, S. L., Brunton, B. W., Proctor, J. L., Kaiser, E., & Kutz, J. N. (2017). Chaos as an intermittently forced linear system. *Nature Communications*, *8*(1), 1–9. https://doi.org/10.1038/s41467-017-00030-8

Brunton, S. L., Brunton, B. W., Proctor, J. L., & Kutz, J. N. (2016). Koopman invariant subspaces and finite linear representations of nonlinear dynamical systems for control. *PLoS One*, *11*(2), e0150171. https://doi.org/10.1371/journal.pone.0150171

Brunton, S. L., & Kutz, J. N. (2020). 7 data-driven methods for reduced-order modeling. *Snapshot-Based Methods and Algorithms*, *307*.

Brunton, S. L., Proctor, J. L., & Kutz, J. N. (2016a). Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, *113*(15), 3932–3937. https://doi.org/10.1073/pnas.1517384113

Brunton, S. L., Proctor, J. L., & Kutz, J. N. (2016b). Sparse identification of nonlinear dynamics with control (SINDYc). *IFAC-PapersOnLine*, *49*(18), 710–715. https://doi.org/10.1016/j.ifacol.2016.10.249

Bunger, A., Kear, J., Dyskin, A., & Pasternak, E. (2014). Interpreting post-injection acoustic emission in laboratory hydraulic fracturing experiments. In *48th us rock mechanics/geomechanics symposium*.

Bunger, A. P., Kear, J., Dyskin, A. V., & Pasternak, E. (2015). Sustained acoustic emissions following tensile crack propagation in a crystalline rock. *International Journal of Fracture*, *193*(1), 87–98. https://doi.org/10.1007/s10704-015-0020-7

Champion, K., Lusch, B., Kutz, J. N., & Brunton, S. L. (2019). Data-driven discovery of coordinates and governing equations. *Proceedings of the National Academy of Sciences*, *116*(45), 22445–22451. https://doi.org/10.1073/pnas.1906995116

Chen, Y., Zhang, G., Bai, M., Zu, S., Guan, Z., & Zhang, M. (2019). Automatic waveform classification and arrival picking based on convolutional neural network. *Earth and Space Science*, *6*(7), 1244–1261. https://doi.org/10.1029/2018ea000466

Chotard, T., Quet, A., Ersen, A., & Smith, A. (2006). Application of the acoustic emission technique to characterise liquid transfer in a porous ceramic during drying. *Journal of the European Ceramic Society*, *26*(7), 1075–1084. https://doi.org/10.1016/j.jeurceramsoc.2005.01.048

Dawson, S. T., Hemati, M. S., Williams, M. O., & Rowley, C. W. (2016). Characterizing and correcting for the effect of sensor noise in the dynamic mode decomposition. *Experiments in Fluids*, *57*(3), 42. https://doi.org/10.1007/s00348-016-2127-7

Demo, N., Tezzele, M., & Rozza, G. (2018). PyDMD: Python dynamic mode decomposition. *Journal of Open Source Software*, *3*(22), 530. https://doi.org/10.21105/joss.00530

DiCarlo, D. A., Cidoncha, J. I., & Hickey, C. (2003). Acoustic measurements of pore-scale displacements. *Geophysical Research Letters*, *30*(17), 1901. https://doi.org/10.1029/2003gl017811

Donoho, D. L. (2006). For most large underdetermined systems of linear equations the minimal l1-norm solution is also the sparsest solution. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, *59*(6), 797–829. https://doi.org/10.1002/cpa.20132

Erguler, Z., & Ulusay, R. (2009). Water-induced variations in mechanical properties of clay-bearing rocks. *International Journal of Rock Mechanics and Mining Sciences*, *46*(2), 355–370. https://doi.org/10.1016/j.ijrmms.2008.07.002

Espinoza, D., & Santamarina, J. (2011). Clay interaction with liquid and supercritical CO2: The relevance of electrical and capillary forces. *International Journal of Greenhouse Gas Control*, *10*, 351–362. https://doi.org/10.1016/j.ijggc.2012.06.020

Fasel, U., Kutz, J. N., Brunton, B. W., & Brunton, S. L. (2022). Ensemble-sindy: Robust sparse model discovery in the low-data, high-noise limit, with active learning and control. *Proceedings of the Royal Society A*, *478*(2260), 20210904. https://doi.org/10.1098/rspa.2021.0904

Fieseler, C. (2019). Geophysics control. Retrieved from https://github.com/Charles-Fieseler-Vienna/Geophysics_control

Fieseler, C., Zimmer, M., & Kutz, J. N. (2020). Unsupervised learning of control signals and their encodings in caenorhabditis elegans whole-brain recordings. *Journal of The Royal Society Interface*, *17*(173), 20200459. https://doi.org/10.1098/rsif.2020.0459

Gin, C., Lusch, B., Brunton, S. L., & Kutz, J. N. (2021). Deep learning models for global coordinate transformations that linearise PDEs. *European Journal of Applied Mathematics*, *32*(3), 515–539. https://doi.org/10.1017/s0956792520000327

Grapsas, N., & Shokri, N. (2014). Acoustic characteristics of fluid interface displacement in drying porous media. *International Journal of Multiphase Flow*, *62*, 30–36. https://doi.org/10.1016/j.ijmultiphaseflow.2014.01.011

Guo, C., Zhu, T., tao Gao, Y., Wu, S., & Sun, J. (2021). Aenet: Automatic picking of p-wave first arrivals using deep learning. *IEEE Transactions on Geoscience and Remote Sensing*, *59*(6), 5293–5303. https://doi.org/10.1109/tgrs.2020.3010541

Hensman, J., Mills, R., Pierce, S., Worden, K., & Eaton, M. (2010). Locating acoustic emission sources in complex structures using Gaussian processes. *Mechanical Systems and Signal Processing*, *24*(1), 211–223. https://doi.org/10.1016/j.ymssp.2009.05.018

Hirsh, S. M., Barajas-Solano, D. A., & Kutz, J. N. (2021). Sparsifying priors for Bayesian uncertainty quantification in model discovery. arXiv preprint arXiv:2107.02107.

Huang, J., Hu, Q., Song, Z., Zhang, G., Qin, C.-Z., Wu, M., & Wang, X. (2021). Classification of cracking sources of different engineering media via machine learning. *Fatigue and Fracture of Engineering Materials and Structures*, *44*(9), 2475–2488. https://doi.org/10.1111/ffe.13528

Hulbert, C., Rouet-Leduc, B., Johnson, P. A., Ren, C. X., Rivière, J., Bolton, D. C., & Marone, C. (2019). Similarity of fast and slow earthquakes illuminated by machine learning. *Nature Geoscience*, *12*(1), 69–74. https://doi.org/10.1038/s41561-018-0272-8

Ilgen, A. G., Heath, J. E., Akkutlu, I. Y., Bryndzia, L. T., Cole, D. R., Kharaka, Y. K., et al. (2017). Shales at all scales: Exploring coupled processes in mudrocks. *Earth-Science Reviews*, *166*, 132–152. https://doi.org/10.1016/j.earscirev.2016.12.013

Ince, N. F., Kao, C.-S., Kaveh, M., Tewfik, A., & Labuz, J. F. (2010). A machine learning approach for locating acoustic emission. *EURASIP Journal on Applied Signal Processing*, 1–14. https://doi.org/10.1155/2010/895486

Ishida, T. (2001). Acoustic emission monitoring of hydraulic fracturing in laboratory and field. *Construction and Building Materials*, *15*(5–6), 283–295. https://doi.org/10.1016/s0950-0618(00)00077-5

Jewell, S., & Witten, D. (2018). Exact spike train inference via l0 optimization. *Annals of Applied Statistics*, *12*(4), 2457. https://doi.org/10.1214/18-aoas1162

Johnson, P. A., Ferdowsi, B., Kaproth, B. M., Scuderi, M., Griffa, M., Carmeliet, J., et al. (2013). Acoustic emission and microslip precursors to stick-slip failure in sheared granular material. *Geophysical Research Letters*, *40*(21), 5627–5631. https://doi.org/10.1002/2013gl057848

Kallel, F., Galanis, N., Perrin, B., & Javelas, R. (1993). Effects of moisture on temperature during drying of consolidated porous materials. *Journal of Heat Transfer*, *115*(3), 724–733. https://doi.org/10.1115/1.2910744

Kaptanoglu, A. A., de Silva, B. M., Fasel, U., Kaheman, K., Callaham, J. L., Delahunt, C. B., et al. (2021). Pysindy: A comprehensive python package for robust sparse system identification. arXiv preprint arXiv:2111.08481.

King, T., Benson, P., De Siena, L., & Vinciguerra, S. (2021). Acoustic emission waveform picking with time delay neural networks during rock deformation laboratory experiments. *Seismological Society of America*, *92*(2A), 923–932. https://doi.org/10.1785/0220200188

Kong, B., Li, Z., & Wang, E. (2018). Fine characterization rock thermal damage by acoustic emission technique. *Journal of Geophysics and Engineering*, *15*(1), 1–12. https://doi.org/10.1088/1742-2140/aa9a54

Kutz, J. N., Brunton, S. L., Brunton, B. W., & Proctor, J. L. (2016). *Dynamic mode decomposition: Data-driven modeling of complex systems* (Vol. 149). SIAM.

Labuz, J. F., Cattaneo, S., & Chen, L.-H. (2001). Acoustic emission at failure in quasi-brittle materials. *Construction and Building Materials*, *15*(5–6), 225–233. https://doi.org/10.1016/s0950-0618(00)00072-6

LaFargeHolcim:North-America:Inc. (2018). *Lafarge portland cement (cement): Safety data sheet* (Vol. 77). LaFargeHolcim.

Lee, H.-L., Kim, J.-S., Hong, C.-H., & Cho, D.-K. (2021). Ensemble learning approach for the prediction of quantitative rock damage using various acoustic emission parameters. *Applied Sciences*, *11*(9), 4008. https://doi.org/10.3390/app11094008

Lei, X.-L., Kusunose, K., Nishizawa, O., Cho, A., & Satoh, T. (2000). On the spatio-temporal distribution of acoustic emissions in two granitic rocks under triaxial compression: The role of pre-existing cracks. *Geophysical Research Letters*, *27*(13), 1997–2000. https://doi.org/10.1029/1999gl011190

Li, N., Zhang, S., Zou, Y., Ma, X., Zhang, Z., Li, S., et al. (2018). Acoustic emission response of laboratory hydraulic fracturing in layered shale. *Rock Mechanics and Rock Engineering*, *51*(11), 3395–3406. https://doi.org/10.1007/s00603-018-1547-5

Li, P., Ren, F.-h., Cai, M.-f., Guo, Q.-f., Wang, H.-f., & Liu, K. (2019). Investigating the mechanical and acoustic emission characteristics of brittle failure around a circular opening under uniaxial loading. *International Journal of Minerals, Metallurgy, and Materials*, *26*(10), 1217–1230. https://doi.org/10.1007/s12613-019-1887-5

Liu, X., Liang, Z., Zhang, Y., Wu, X., & Liao, Z. (2015). Acoustic emission signal recognition of different rocks using wavelet transform and artificial neural network. *Shock and Vibration*, 1–14. https://doi.org/10.1155/2015/846308

Lockner, D. (1993). The role of acoustic emission in the study of rock fracture. *International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts*, *30*(7), 883–899. https://doi.org/10.1016/0148-9062(93)90041-b

Lusch, B., Kutz, J. N., & Brunton, S. L. (2018). Deep learning for universal linear embeddings of nonlinear dynamics. *Nature Communications*, *9*(1), 1–10. https://doi.org/10.1038/s41467-018-07210-0

MATLAB. (2021). *MATLAB version 9.8.0.13*. The MathWorks Inc.

McLaskey, G. C., & Glaser, S. D. (2010). Hertzian impact: Experimental study of the force pulse and resulting stress waves. *Journal of the Acoustical Society of America*, *128*(3), 1087–1096. https://doi.org/10.1121/1.3466847

McLaskey, G. C., & Glaser, S. D. (2012). Acoustic emission sensor calibration for absolute source measurements. *Journal of Nondestructive Evaluation*, *31*(2), 157–168. https://doi.org/10.1007/s10921-012-0131-2

McLaskey, G. C., & Lockner, D. A. (2014). Preslip and cascade processes initiating laboratory stick slip. *Journal of Geophysical Research: Solid Earth*, *119*(8), 6323–6336. https://doi.org/10.1002/2014jb011220

McLaskey, G. C., Lockner, D. A., Kilgore, B. D., & Beeler, N. M. (2015). A robust calibration technique for acoustic emission systems based on momentum transfer from a ball drop. *Bulletin of the Seismological Society of America*, *105*(1), 257–271. https://doi.org/10.1785/0120140170

Michlmayr, G., Cohen, D., & Or, D. (2012). Sources and characteristics of acoustic emissions from mechanically stressed geologic granular media—A review. *Earth-Science Reviews*, *112*(3–4), 97–114. https://doi.org/10.1016/j.earscirev.2012.02.009

Moebius, F., Canone, D., & Or, D. (2012). Characteristics of acoustic emissions induced by fluid front displacement in porous media. *Water Resources Research*, *48*(11), W11507. https://doi.org/10.1029/2012wr012525

Moradian, Z., Einstein, H. H., & Ballivy, G. (2016). Detection of cracking levels in brittle rocks by parametric analysis of the acoustic emission signals. *Rock Mechanics and Rock Engineering*, *49*(3), 785–800. https://doi.org/10.1007/s00603-015-0775-1

Nolte, D., & Pyrak-Nolte, L. (2022). Monitoring fracture saturation with internal transportable seismic sources and twin neural network. *Journal of Geophysical Research: Solid Earth*, *127*, e2021JB023005.

ORS-Inc, O. R. S. (2020). Dragonfly pro software, version 2020.2 for [windows] from object research systems (ors) inc. montreal, Canada. ORS, Retrieved from http://www.theobjects.com/dragonfly

Pettijohn, F. J., Potter, P. E., & Siever, R. (2012). *Sand and sandstone*. Springer Science & Business Media.

Physical-Acoustics:Mistras-Group. (2018). Aewin software [windows] for real-time simultaneous acoustic emission feature and waveform processing, physical acoustics—Mistras group inc. Mistras, Retrieved from http://www.mistrasgroup.com

Proctor, J. L., Brunton, S. L., & Kutz, J. N. (2016). Dynamic mode decomposition with control. *SIAM Journal on Applied Dynamical Systems*, *15*(1), 142–161. https://doi.org/10.1137/15m1013857

Proctor, J. L., Brunton, S. L., & Kutz, J. N. (2018). Generalizing Koopman theory to allow for inputs and control. *SIAM Journal on Applied Dynamical Systems*, *17*(1), 909–930. https://doi.org/10.1137/16m1062296

Pyrak-Nolte, L., Braverman, W., Nolte, N., Wright, A., & Nolte, D. D. (2020). Probing complex geophysical geometries with chattering dust. *Nature Communications*, *11*(1), 5282. https://doi.org/10.1038/s41467-020-19087-z

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

Rück, M., Rahner, R., Sone, H., & Dresen, G. (2017). Initiation and propagation of mixed mode fractures in granite and sandstone. *Tectonophysics*, *717*, 270–283. https://doi.org/10.1016/j.tecto.2017.08.004

Sakamoto, Y., Ishiguro, M., & Kitagawa, G. (1986). *Akaike information criterion statistics*. D. Reidel.

Sashidhar, D., & Kutz, J. N. (2021). Bagging, optimized dynamic mode decomposition (BOP-DMD) for robust, stable forecasting with spatial and temporal uncertainty-quantification. arXiv preprint arXiv:2107.10878.

Scruby, C. B. (1987). An introduction to acoustic emission. *Journal of Physics E: Scientific Instruments*, *20*(8), 946–953. https://doi.org/10.1088/0022-3735/20/8/001

Sellers, E. J., Kataka, M. O., & Linzer, L. M. (2003). Source parameters of acoustic emission events and scaling with mining-induced seismicity. *Journal of Geophysical Research*, *108*(B9), 2418. https://doi.org/10.1029/2001jb000670

Selvadurai, P., & Glaser, S. (2012). Direct measurement of contact area and seismic stress along a sliding interface. In *46th us rock mechanics/geomechanics symposium*.

Shi, G., Yang, X., Yu, H., & Zhu, C. (2019). Acoustic emission characteristics of creep fracture evolution in double-fracture fine sandstone under uniaxial compression. *Engineering Fracture Mechanics*, *210*, 13–28. https://doi.org/10.1016/j.engfracmech.2018.09.004

Sigma-Aldrich. (2019). Specification sheet. In *Sigma Aldrich: Millipore Sigma* (Vol. 2021). Retrieved from https://www.sigmaaldrich.com/US/en/specification-sheet/ALDRICH/69866

Su, W., Bogdan, M., & Candes, E. (2017). False discoveries occur early on the lasso path. *Annals of Statistics*, *45*(5), 2133–2150. https://doi.org/10.1214/16-aos1521

Trugman, D. T., McBrearty, I. W., Bolton, D. C., Guyer, R. A., Marone, C., & Johnson, P. A. (2020). The spatiotemporal evolution of granular microslip precursors to laboratory earthquakes. *Geophysical Research Letters*, *47*(16), e2020GL088404. https://doi.org/10.1029/2020gl088404

Tu, J. V. (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of Clinical Epidemiology*, *49*(11), 1225–1231. https://doi.org/10.1016/s0895-4356(96)00002-9

U.S.-Silica. (2019). Sil-co-sil: Bright white, high-purity ground silica (Vol. *2021*). Retrieved from https://www.ussilica.com/products/sil-co-silr

Wang, C., Hou, X., & Liu, Y. (2021). Three-dimensional crack recognition by unsupervised machine learning. *Rock Mechanics and Rock Engineering*, *54*(2), 893–903. https://doi.org/10.1007/s00603-020-02287-w

Wang, D., Bian, X., Qin, H., Sun, D., & Yu, B. (2021). Experimental investigation of mechanical properties and failure behavior of fluid-saturated hot dry rocks. *Natural Resources Research*, *30*(1), 289–305. https://doi.org/10.1007/s11053-020-09760-x

Wang, S., Huang, R., Ni, P., Gamage, R. P., & Zhang, M. (2013). Fracture behavior of intact rock using acoustic emission: Experimental observation and realistic modeling. *Geotechnical Testing Journal*, *36*(6), 903–914. https://doi.org/10.1520/gtj20120086

Wu, B. S., & McLaskey, G. C. (2018). Broadband calibration of acoustic emission and ultrasonic sensors from generalized ray theory and finite element models. *Journal of Nondestructive Evaluation*, *37*(1), 1–16. https://doi.org/10.1007/s10921-018-0462-8

Wu, Y., Lin, Y., Zhou, Z., Bolton, D. C., Liu, J., & Johnson, P. (2018). Deepdetect: A cascaded region-based densely connected network for seismic event detection. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(1), 62–75. https://doi.org/10.1109/tgrs.2018.2852302

Zhang, L., & Schaeffer, H. (2018). On the convergence of the sindy algorithm. arXiv preprint arXiv:1805.06445.

Zhao, Q., & Glaser, S. D. (2020). Relocating acoustic emission in rocks with unknown velocity structure with machine learning. *Rock Mechanics and Rock Engineering*, *53*(5), 2053–2061. https://doi.org/10.1007/s00603-019-02028-8

Zheng, J., Lu, J., Peng, S., & Jiang, T. (2018). An automatic microseismic or acoustic emission arrival identification scheme with deep recurrent neural networks. *Geophysical Journal International*, *212*(2), 1389–1397. https://doi.org/10.1093/gji/ggx487

Zheng, P., Askham, T., Brunton, S. L., Kutz, J. N., & Aravkin, A. Y. (2018). A unified framework for sparse relaxed regularized regression: Sr3. *IEEE Access*, *7*, 1404–1423. https://doi.org/10.1109/access.2018.2886528