Predicting Fault Slip via Transfer Learning

Kun Wang
Los Alamos National Laboratory
Christopher Johnson
Los Alamos National Laboratory
Kane Bennett
Los Alamos National Laboratory https://orcid.org/0000-0002-0822-6288
Paul Johnson (paj@lanl.gov)
Los Alamos National Laboratory https://orcid.org/0000-0002-0927-4003

Article

Keywords: earth science, seismology, fault slip prediction

Posted Date: July 22nd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-700852/v1

Version of Record: A version of this preprint was published at Nature Communications on December 1st, 2021. See the published version at https://doi.org/10.1038/s41467-021-27553-5.

Predicting Fault Slip via Transfer Learning

Kun Wang[1,2], Christopher W. Johnson[1], Kane C. Bennett[1], and Paul A. Johnson[1,*]

[1]Geophysics Group, Earth and Environmental Sciences Division, Los Alamos National Laboratory, Los Alamos, NM 87545 USA
[2]Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM, 87545, USA
*paj@lanl.gov

ABSTRACT

Data-driven machine-learning for predicting instantaneous and future fault-slip in laboratory experiments has recently progressed markedly due to large training data sets. In Earth however, earthquake interevent times range from 10's-100's of years and geophysical data typically exist for only a portion of an earthquake cycle. Sparse data presents a serious challenge to training machine learning models. Here we describe a transfer learning approach using numerical simulations to train a convolutional encoder-decoder that predicts fault-slip behavior in laboratory experiments. The model learns a mapping between acoustic emission histories and fault-slip from numerical simulations, and generalizes to produce accurate results using laboratory data. Notably slip-predictions markedly improve using the simulation-data trained-model and training the latent space using a portion of a single laboratory earthquake-cycle. The transfer learning results elucidate the potential of using models trained on numerical simulations and fine-tuned with small geophysical data sets for potential applications to faults in Earth.

* <paj@lanl.gov> 2015-02-09T12:07:31.197Z:

Introduction

In Earth, predicting instantaneous and future characteristics of fault slip has long been a fundamental goal of geoscientists from an earthquake hazards perspective, but also to improve the basic understanding of fault mechanics[1]. Recent progress towards these goals has been achieved by applying a variety of machine learning (ML) approaches[2,4] in the laboratory using shear experimental data to describe physical properties[5–11] and in the Earth using geophysical data to characterize episodic slow-slip that occurs in subduction zones[8,12] as well as transform faults[13]. In shear experiments, earthquakes or "labquakes", generated during a single experiment produce a sufficiently large data set for training ML models. However, on natural faults the repeat cycles for all but the smallest earthquakes can span timescales on the order of decades to hundreds of years. Thus, *in-situ* geophysical measurements as input for data-driven ML models are generally not available or sufficiently complete for more than a portion of a single earthquake cycle. In particular this problem exists for large magnitude (M>7) earthquakes that produce strong, damaging ground motions. This conundrum presents a serious challenge if the goal is to use data-driven modeling techniques to characterize the physics of fault slip throughout the complete earthquake cycle and to advance earthquake hazards assessment.

A type of model generalization known as transfer learning[15,16] is one potential solution to overcome the problem of data sparsity. Generalizing ML models using transfer learning has been explored in a number of areas in geophysics; for instance in seismology applications, transfer learning has been used to improve nonlinear and ill-posed inverse problems associated with seismic imaging or subsurface feature classification[17–20]. To our knowledge, no attempt has been made to apply transfer learning using data from numerical simulations to make quantitative predictions of fault slip in laboratory experiments or Earth observations.

In this work we develop a deep learning convolutional encoder-decoder (CED) model that employs a time-frequency representation of acoustic emissions (AE) from numerical simulations and laboratory shearing experiments. The model is a U-shape design that encodes the salient features to a latent space that is then decoded to estimate the friction coefficient measured in the experiment. The model is initially trained using numerical fault-slip simulation data to learn the mapping between the AE and fault slip. The latent space is then trained using only a small fraction of the laboratory experimental data, and the resulting cross-trained CED model is applied to unseen laboratory experimental data (see Figure 1). We describe results from the CED model transfer learning and show the successful application of this technique for multiple data sets.

The long term goal is to develop a ML approach that characterizes seismogenic faults. In Earth we generally record only a portion of a slip cycle on a fault and one could apply a similar approach of transfer learning and cross-training. If such a procedure works at the laboratory scale, it may be applicable to Earth, despite the dramatic changes in scale and complexity of seismogenic faults.

Results

Transfer learning from numerical simulations to laboratory shear experiments

The laboratory data[10,23] is from a bi-axial shear device that consists of a slider-block bounded by fault gouge and external blocks through which a confining load is applied (Figure 1). A constant shear velocity is applied and when the system approaches steady state conditions, repetitive stick-slip motion occurs (see Methods and Figure 2b). The bi-axial device simultaneously measures acoustic emission (AE) and the normal and shear stresses required to calculate the bulk friction coefficient.

The numerical simulation[24] applies a combined Finite-Discrete Element Method (FDEM) model of a fault-shear apparatus resembling the bi-axial device used in the laboratory experiments described here[14] (see Methods and Figure 1). The input training data to the CED model from simulation is the kinetic energy, which is a proxy for the measured continuous AE signal in the bi-axial shear experiment. Changes in seismic moment are reflected in variations observed in the kinetic energy; therefore, the kinetic energy represents the kinematic behavior of the granular fault system (see Methods). The CED label data is the bulk friction coefficient between the sliding blocks.

In summary, the data from these sources, numerical simulation and laboratory experiment (Figure 2), are used by the CED model (Figure 3), which is described in greater detail in the Methods section. The supervised learning method is a regression procedure, using the AE from experiment or the kinetic energy from simulation to predict the instantaneous characteristics of slip, specifically the coefficient of friction.

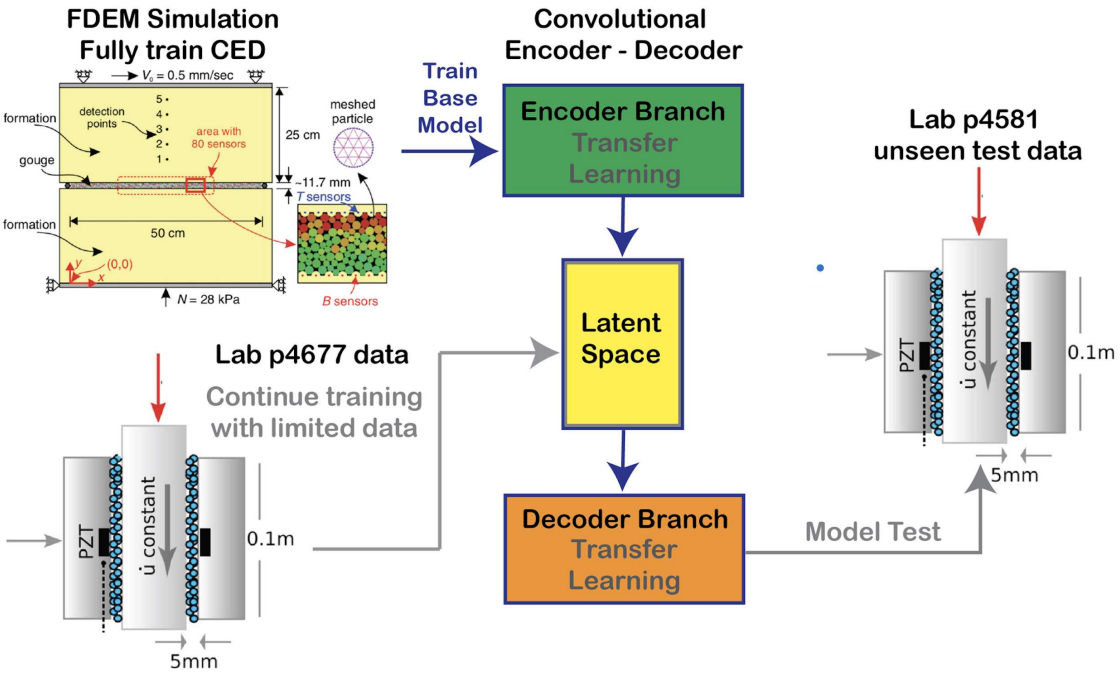

Figure 1. Workflow with numerical simulation and experimental configurations for the transfer learning analysis. See methods for simulation and experimental details. During simulation and experiment, characteristics are recorded including shear displacement, shear stress, normal stress, bulk friction and kinetic energy (simulation) and acoustic emission (experiment). The simulation kinetic energy, a proxy for acoustic emission, is used to train the model to predict the instantaneous bulk friction coefficient at all times in the slip cycle. The model's latent space *only* is then further trained using acoustic emission data from the laboratory experiment (number p4677). The simulation-trained encoder and decoder are left unchanged. The new model is then used to predict the instantaneous friction for experimental data the model has not previously seen, from a different laboratory experiment (p4581).

As a point of reference for the transfer learning approach, the results shown in Figure 4 are produced by training, validating, and testing entirely on the numerical simulation data. The scalogram represents the average performance among all trained models (see Methods section and Supplementary Material for details). The predicted friction coefficient captures the general slip trends including many frictional failures. However, the results are modest as reported by the Mean Absolute Percentage Error (MAPE) of 4.237% for the numerical simulation data.

The same procedure is followed using only the laboratory AE and friction data to train a separate CED model. The first 20% of the AE signal (0-60 s) is used for training data. The friction predictions from the testing data produce a MAPE of 1.137%

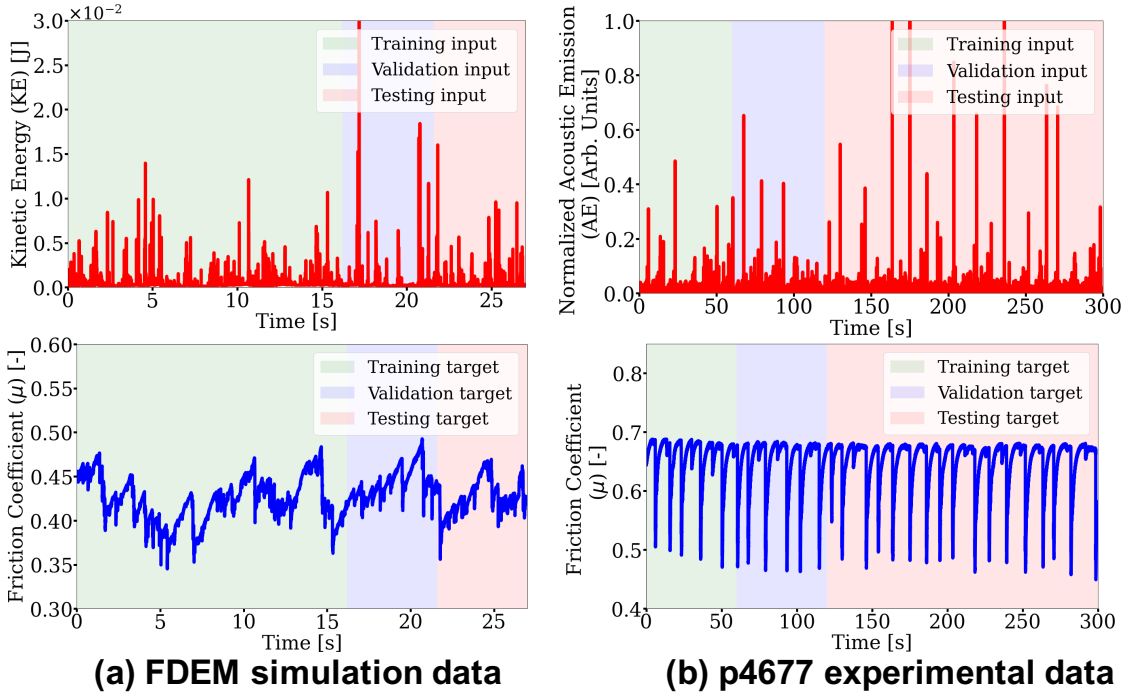

Figure 2. Numerical simulation and experimental data used in the transfer learning analysis. The top row shows the deep learning model input signal as the kinetic energy and acoustic emissions, respectively, and the bottom row shows the target friction coefficient. (a) FDEM time series are split into training/validation/testing segments (60/20/20%) shown in green, blue, and pink shades, respectively. The convolutional encoder-decoder is fully trained and tested using these data. (b) The experimental data (p4677) are split into training/validation/testing segments (20/20/60%) to include 6 cycles of stick-slip events for training the model latent space.

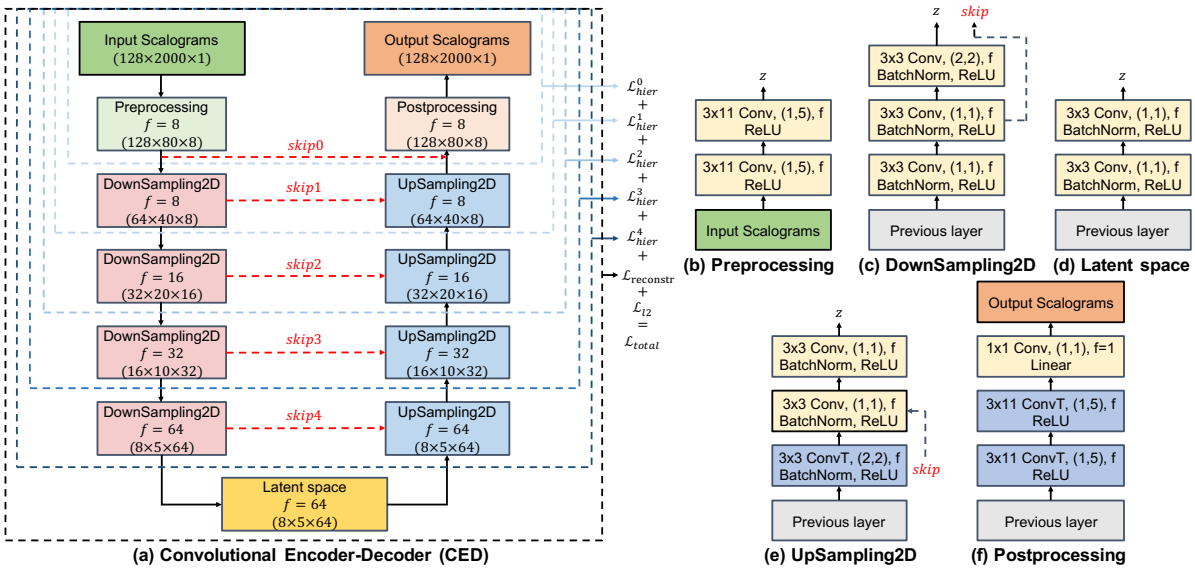

Figure 3. The convolutional encoder-decoder (CED) model for predicting fault friction from scalograms, obtained from kinetic energy (simulation) and acoustic emission (experiment). (a) CED model architecture. The encoder branch contains a (b) Preprocessing block and (c) four DownSampling 2D blocks that populate the (d) latent space. The decoder branch reverses the procedure using (e) four UpSampling 2D blocks and a (f) Postprocessing block. The encoder and decoder models are connected by skip connections (dashed lines) between the downsampling and upsampling blocks as shown in (a). The number of filters (f) for each block are shown in (a). The image size after each layer block are provided in parentheses. The blue dashed lines indicate the sub-models used when computing the hierarchical components[44] and the associated training loss function to obtain the total loss (L_{total}). The model layer notations are Conv (convolutional layer), ConvT (convolutional transpose layer), BatchNorm (batch normalization layer), Linear (linear connected layer), ReLU (rectified linear unit).

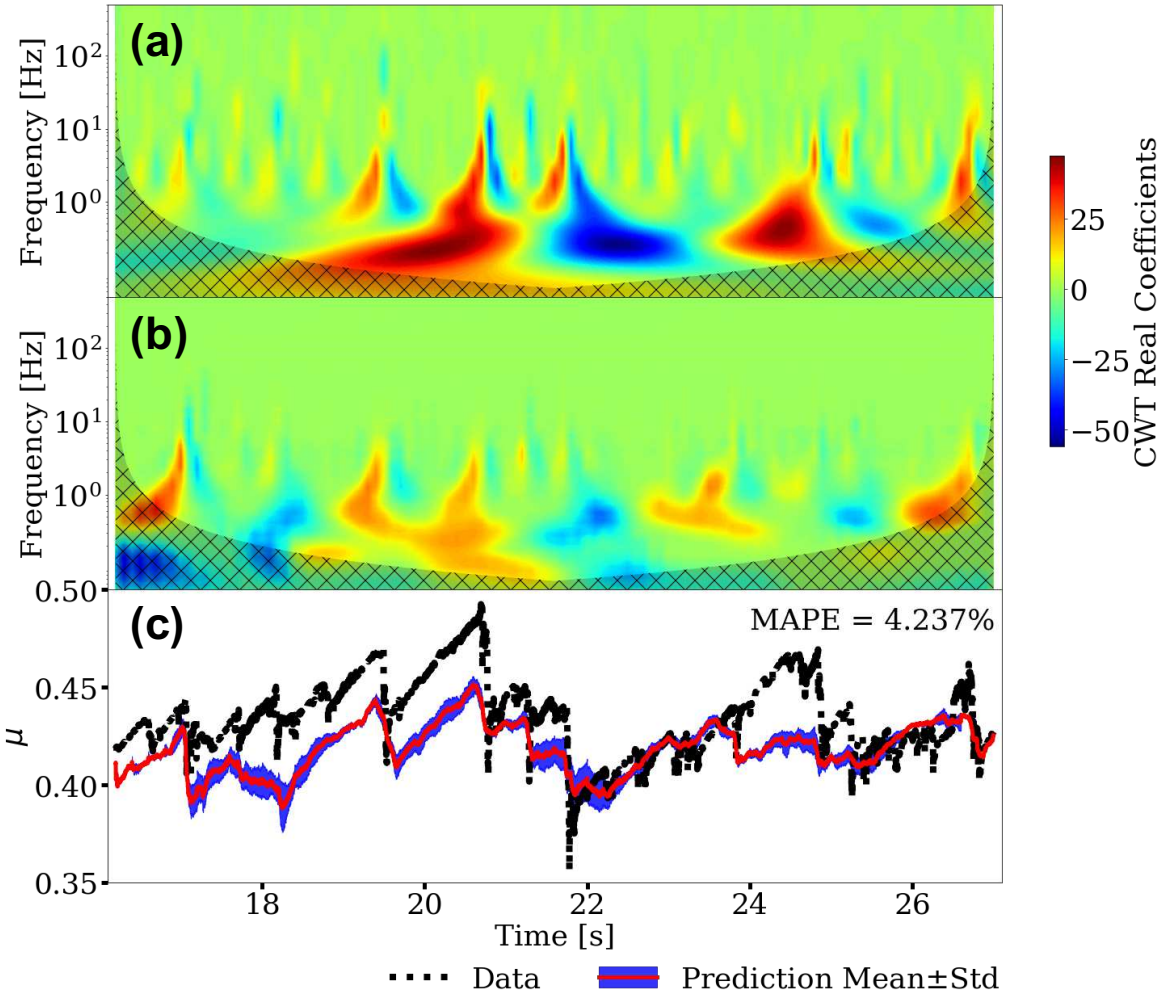

Figure 4. Instantaneous frictional coefficient prediction from the CED model trained on FDEM simulation data. The (a) input and (b) predicted scalograms are shown. The cross-hatched region in (b) indicates the cone of influence where edge effects are important. The predictions from the CED are made applying sliding windows with 2 s length and step size of 0.2 s. The predicted scalogram is the average of all sliding windows. (c) Ground truth versus prediction: the numerical simulation data (black line) and model-predicted friction coefficient from the inverse of the scalogram (red line) is shown with the blue region indicating 1-standard deviation for the predictions in the overlapping windows. The Mean Absolute Percentage Error (MAPE) is listed for the numerical simulation and predicted values.

(Figure 5a). The model performs very well estimating the variations in friction coefficients predicting the frictional failures associated with slip events.

For the first transfer learning exercise, we use the model trained on simulation and apply it to predict the friction in the laboratory experiment. The trained model uses the experiment AE as input and the label is the friction from experiment. We emphasize that the CED model never sees experimental data during training with the simulation data. The predictions show a decrease in performance with a MAPE of 4.232% (Figure 5b) when compared to the model trained solely on the laboratory data. The maximum friction drop, which has an equivalence to event moment (moment = GAu; where G is the gouge shear modulus, A is slip area and u is the fault displacement), is consistently less than that measured from the experiment. Under-prediction of the event moment is a common problem with many ML models when applied to the bi-axial shear data[13], without considering transfer learning. Nonetheless, we find the timing and scale of the predictions are surprisingly good considering the significant differences between the FDEM numerical simulation and the laboratory shear experiment.

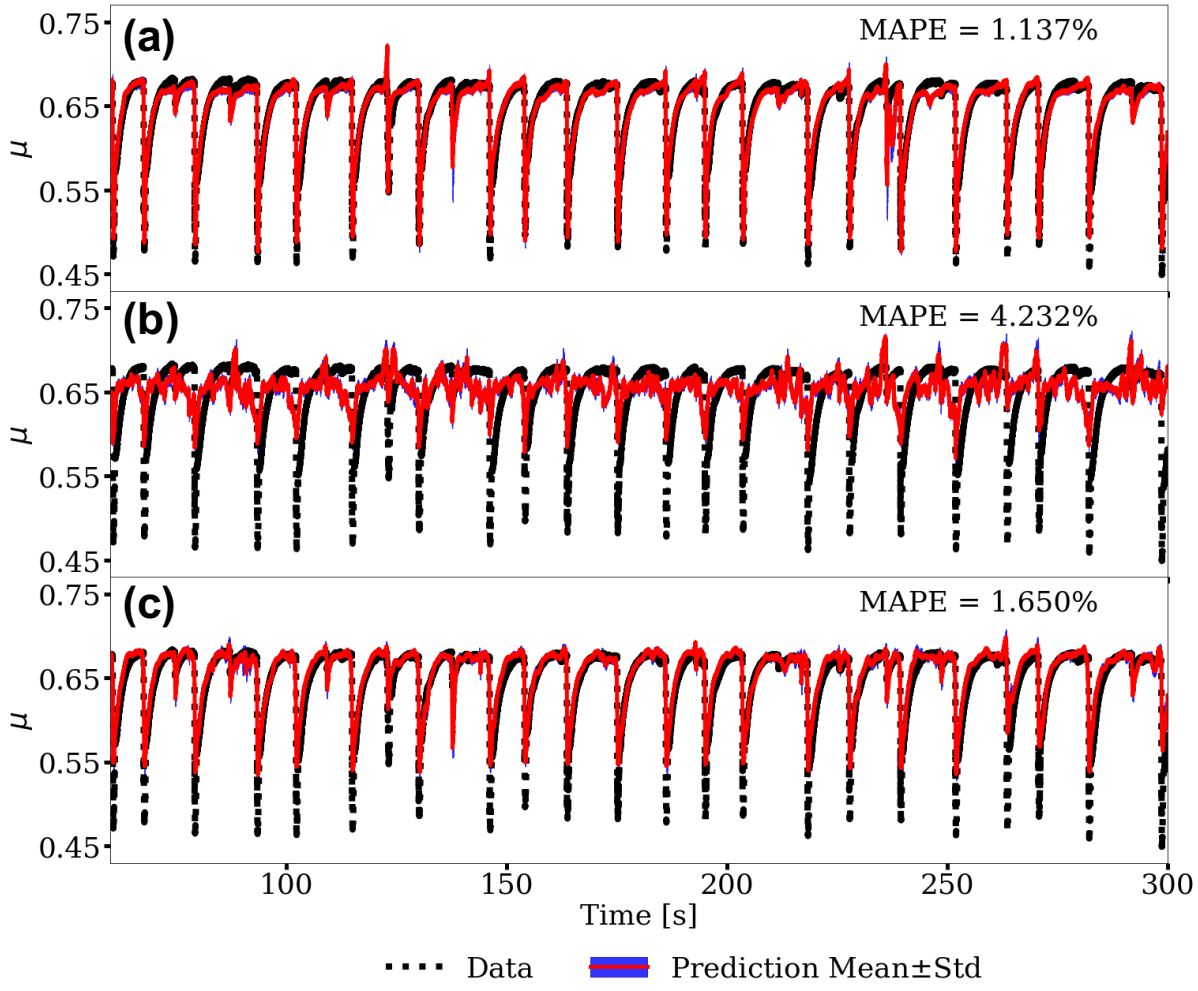

Figure 5. Predictions from CED models. (a) Model trained and tested on the laboratory data. (b) Model trained on simulations and tested on laboratory data. The first 20% of the AE signal (0 s to 60 s) was used for training to construct the model. (c) Cross-trained model. Model trained on the simulations, then fixing the encoder and decoder layers, with the model then additionally trained on the bottleneck (latent space) applying a portion of the laboratory (experiment p4677) data. See Methods for details.

With an eye to faults in Earth where obtaining sufficient training data is a challenge, we introduce transfer learning by cross-training the model. Here we allow the latent space of the CED to be trained on limited laboratory data, while fixing the encoder and decoder layers that are trained using only the simulation. This approach is an extension of an established transfer learning technique used in image classification tasks, e.g.,[26,27] (As applied to image classification, the convolutional layers of a model are pre-trained on a large database (e.g., ImageNet[28,29] and then specific convolutional layers are extracted and held

constant, then merged with an additional fully-connected classification layer that is trained with data specific to the problem of interest.).

Here, we apply a transfer learning approach in this same spirit. We emphasize that alternatively here, though analogously, it is the encoder and decoder layers of the CAE that are directly extracted, and only the latent space weights are updated. For the ML models previously trained on the numerical simulation data, all the parameters in the encoder and decoder are rendered non-trainable, while the parameters in the latent space are updated and fit to the training data from laboratory experiments.

The resulting predictions shown in Figure 5c and are very good with MAPE=1.650%, which is a significant improvement from the MAPE of 4.232% before cross-training. The model predictions are now comparable to the MAPE of 1.137% obtained when training directly on p4677 laboratory data.

As a more rigorous test on how well the cross-trained CED model predicts the laboratory experiments, we apply the identical model to a different laboratory experiment conducted in the bi-axial apparatus. These experimental data are never seen during training of the latent space. This second experiment was conducted over a range of confining loads (normal stress) from 3-8 MPa (Figure 6). The only information applied from the different confining load levels are the mean and standard deviation statistics used to normalize the AE and μ signals when producing the input scalograms to the model and when reconstructing the output scalograms (see Methods).

The predictions applying the cross-trained model to the second experiment are shown in (Figure 7). The results are remarkably good as indicated by the MAPE's. The 3 MPa data exhibits the best MAPE, presumably because the confining load is close to the 2.5 MPa value in the p4677 experiment that was used to train the latent space (Figure 7a). The model predictions as manifest by the MAPE increase with increasing normal loads. The prediction errors appear to be due primarily to the poor predictions of the frictional failure magnitudes. Nonetheless, the instantaneous slip-event times are captured at all load levels, as are the stress buildups during inter-event periods.

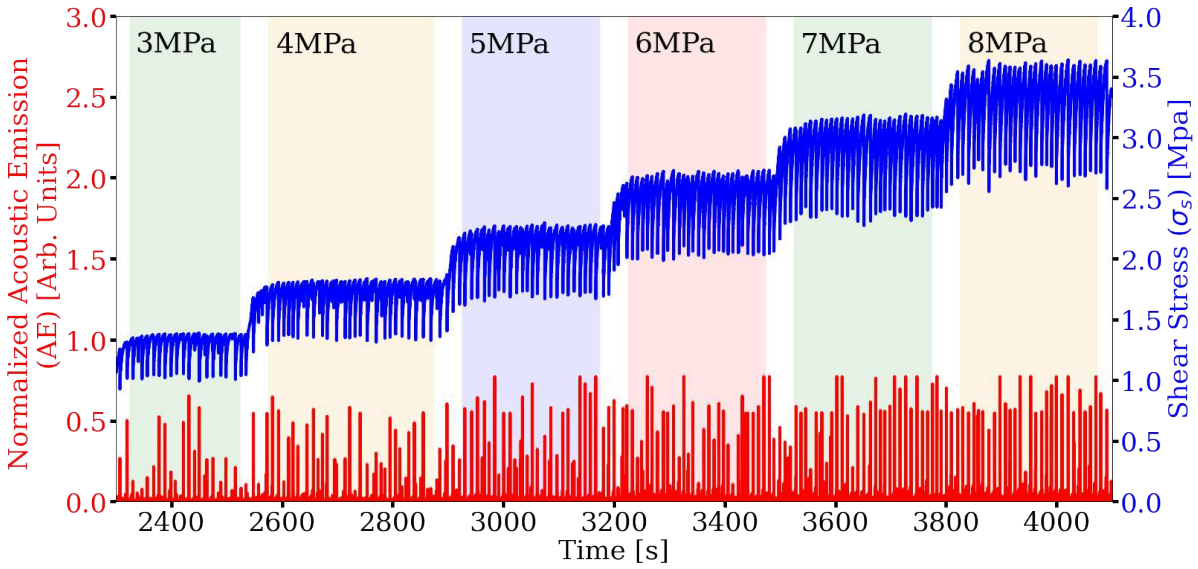

Figure 6. CED Model generalization. The cross-trained model is rigorously validated using a second laboratory experiment (p4581) as an independent data set. The input signal is the acoustic emission (AE, in red) and the target signal is the friction derived from the shear stress (σ_s, in blue) at progressively increasing applied normal loads (3-8 MPa), shown in sequence and delineated by different shading. Model predictions are shown in Figure 7.

Transfer learning with extremely limited data in laboratory experiments

Because slip cycles in Earth are so long (decades to 100's of years) and we rarely have more than a portion of associated seismicity within a full seismic cycle, we conduct a cross training exercise that mimics this data-poor circumstance. We do so by using only limited portions of a single slip cycle from the laboratory experiment for training the model latent space. Specifically, we train the latent space by applying only the *post-failure* or the *pre-failure* μ signals from experiment p4677 data (Figure 8). The post-failure comprises the time-period when the shear stress is increasing relatively rapidly following the previous slip event. The pre-failure period comprises the period when the fault is late in the cycle, near-critical state, and beginning to nucleate[23]. The model encoder and decoder trained with the simulation data again remain unchanged. The model is trained and validated using 90% and 10% of the data, respectively, for the pre-failure and post-failure analysis. Because the

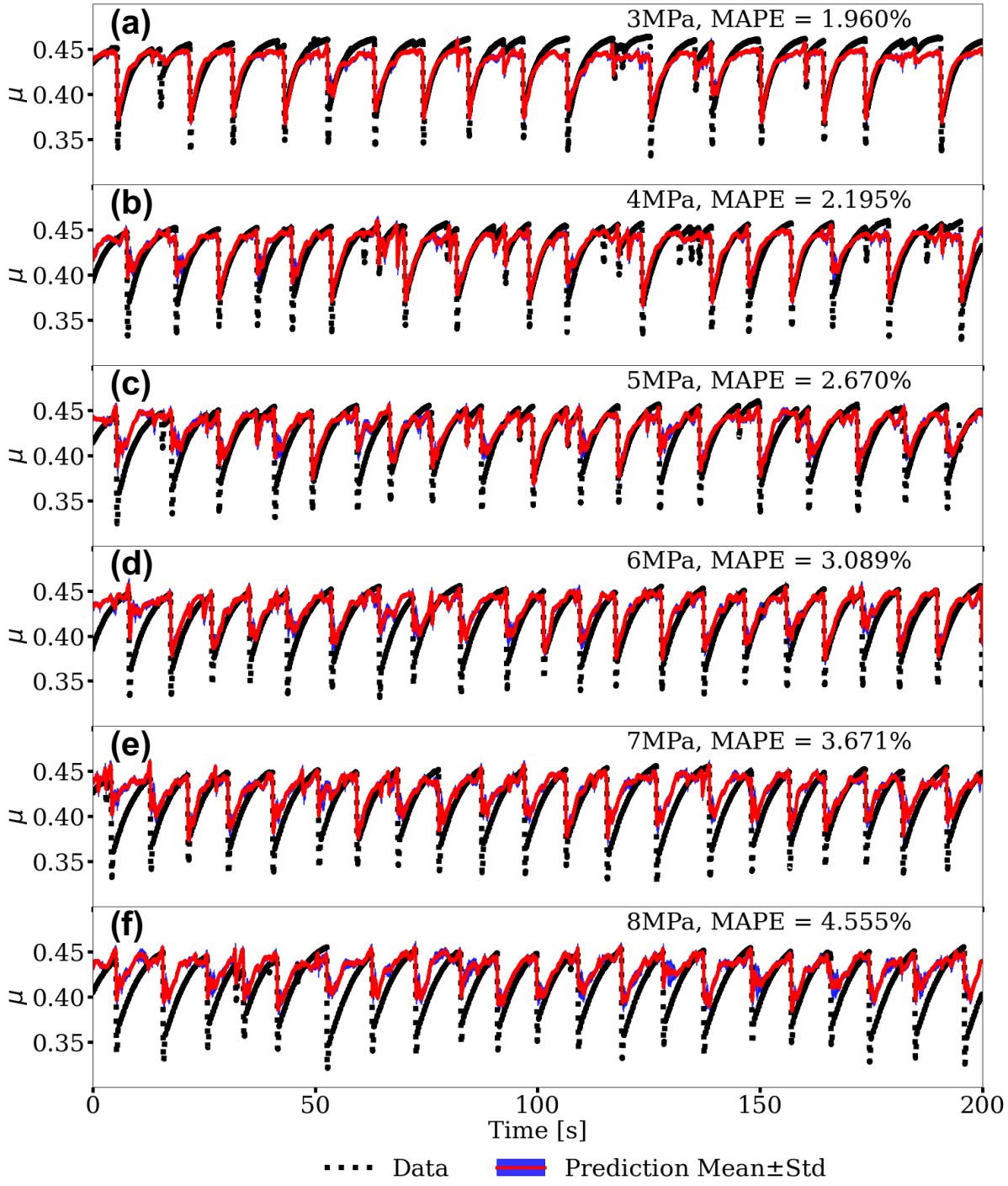

Figure 7. Transfer learning applied to an independent experiment. Shown are predictions from the cross-trained CED model (experiment p4581) with normal loads that progressively increase from 3MPa to 8MPa (see Figure 6). Each load level is predicted independently using the cross-trained model from simulation (the encoder and decoder) and data from experiment (p4677) conducted at 2.5 MPa. The predictions as manifest by the MAPE progressively decrease with increasing load level. Nevertheless, the results show that the transfer learning approach with cross-training of the latent space, which accounts for only 20% of the total CED model parameters, is a powerful approach to predicting the frictional state of the experimental fault.

available data only spans a short time interval, the size of the sliding windows is reduced from 2s to 0.4s and the step size is reduced from 0.2s to 0.1s (the sizes have no significant impact on the model performances, see Supplementary Material). The training is terminated when the validation does not reduce for 100 epochs to prevent over-fitting.

After cross-training the latent space using the two data sets from experiment p4677 (post-failure and pre-failure) to produce two separate CED models, the models are used to predict the friction in the second experiment, p4581, for 3MPa, 5MPa and 7MPa applied load, on the post-failure and pre-failure signals. The results are shown in Figure 9. As before, the magnitude of the frictional failures are not well predicted—otherwise the trained models perform surprisingly well in both cases. The result applying the post-failure data is slightly better than that from the pre-failure training data. This suggests the model has learned more frictional states during latent-space training. As anticipated, the model using experiment p4677 and trained applying 6 full cycles provides the most robust result; however, the model results with extremely limited training cross-training data are very encouraging.

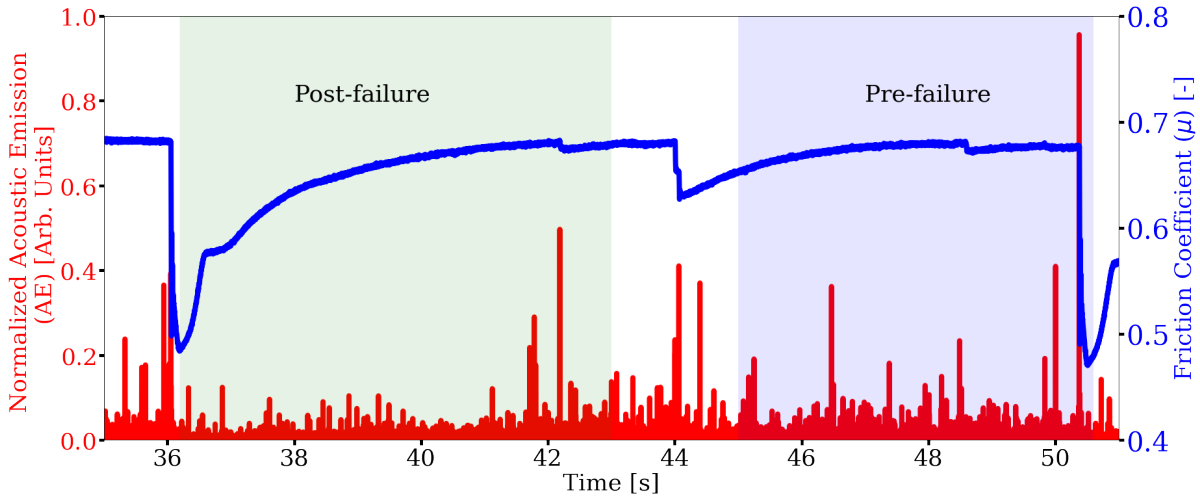

Figure 8. Testing the limits of the transfer learning: CED model with cross-training of the latent space applying limited portions of AE data from a single slip cycle. The model input signal is acoustic emission (AE), and the output signal is the friction coefficient (μ). The green region shows the *post-failure* data used for the transfer learning exercise, and the blue region shows the *pre-failure* data (see Figure 9 for model results).

Transfer learning predicting time to failure (TTF) in laboratory experiments

Transfer learning can also be applied to predicting other output time series. Here we showcase the predictions on the signals of time-to-failure (TTF). Failure times are defined as when the time derivative of the μ signal is below -10 $/s$. The raw AE signal is used as input, just as for the instantaneous predictions of the friction coefficient. The encoder and decoder models are again directly applied from the CED model trained on the numerical simulation data for the task of predicting the μ signal (There is no need to train applying the FDEM TTF data.). Next, the latent space is trained to fit the TTF training data from experiment p4677 (the first 20% of the signal, from 0s to 60s, including 6 stick-slip cycles). Data from p4581 are again used for testing purposes only. The predictions are illustrated in Figure 10 for 3MPa, 5MPa and 7MPa load levels. The predictions are good, if not perfect considering the task, as underscored by their respective MAPE scores. Indeed they are notable considering the they are obtained from cross-training and transfer learning.

Discussion

The predictions of the instantaneous friction obtained applying transfer learning from FDEM simulations to laboratory data from the bi-axial shear device are surprisingly good. When model cross-training is then applied, the predictions improve significantly. Further, when we apply the same cross-trained model to the second experiment conducted at multiple applied loads, the model predictions are still surprisingly good—there exists a larger misfit with increasing load, but the timing of the event is accurate regardless of the under-prediction in friction failure magnitude. The results are even more remarkable considering the FDEM simulation was not meant to directly simulate the experiment—material properties and dimensions of the fault gouge and shear-blocks were considerably different. Indeed, the results indicate the simulations contain an AE (kinectic energy) evolution captured in the spectral characteristics that can predict the actual AE in experiment. The results suggest the simulations, despite the differences, provide a sufficient range of behaviors for the models to learn and reproduce

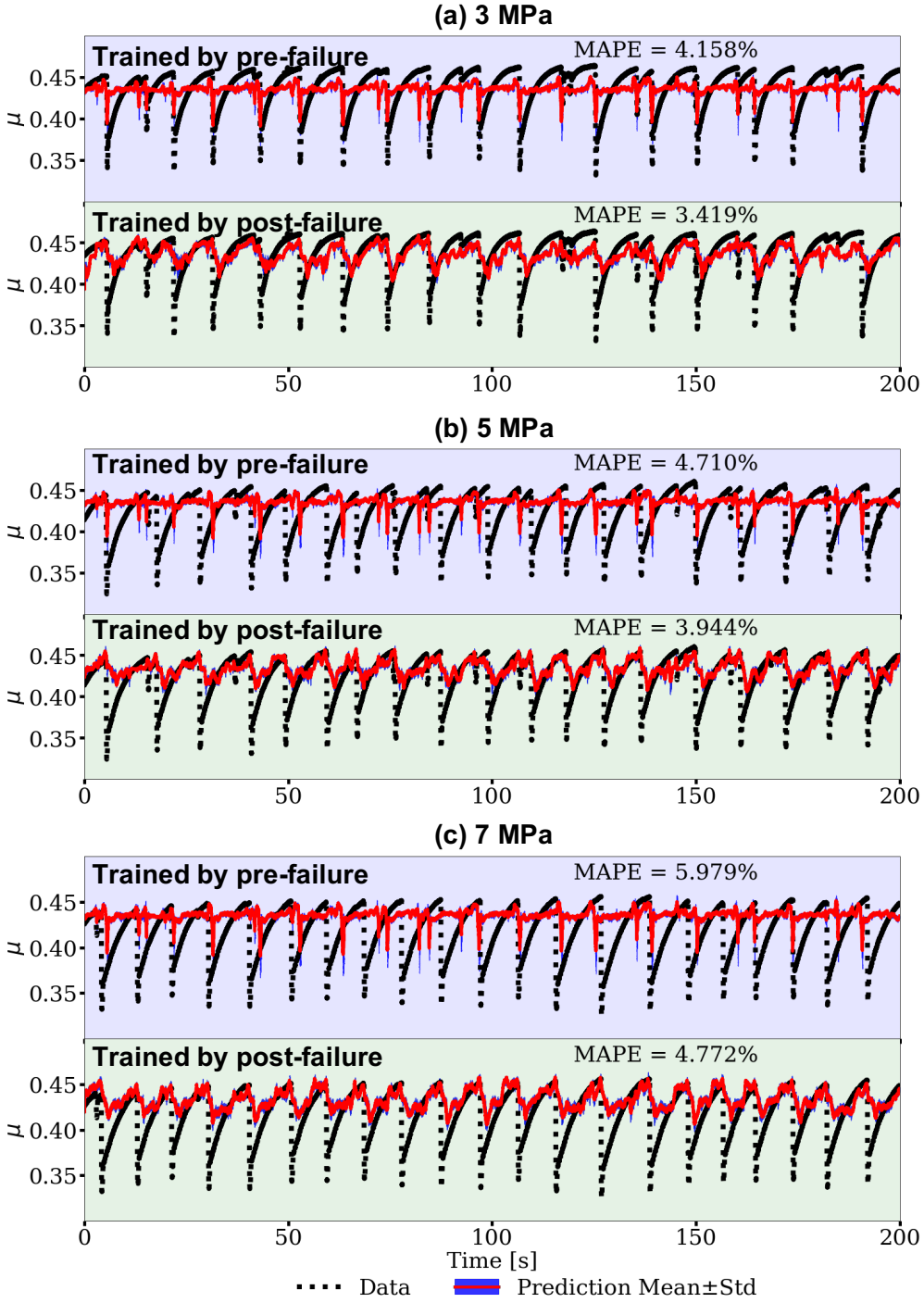

Figure 9. Model cross-training applying limited portions of the experimental slip cycle. Two models are trained. In both cases only the latent space is trained using data from experiment p4677 applying the following (see Figure 8. One model is trained applying AE data from the '*post-failure* portion of the slip cycle, comprising the time-period when the shear stress is increasing relatively rapidly. The second model is trained applying AE data from the *pre-failure* period comprising the period the fault is in a critical state and nucleating. The model encoder and decoder trained applying the simulation data remain unchanged. Friction predictions on data from the experiment p4581 testing data set for pre-failure at applied loads of 3,5, and 7 are shown in each top row of (a, b, c) and post-failure in each bottom row of (a, b, c). The color in the panels correspond to the colored training data in Figure 8.

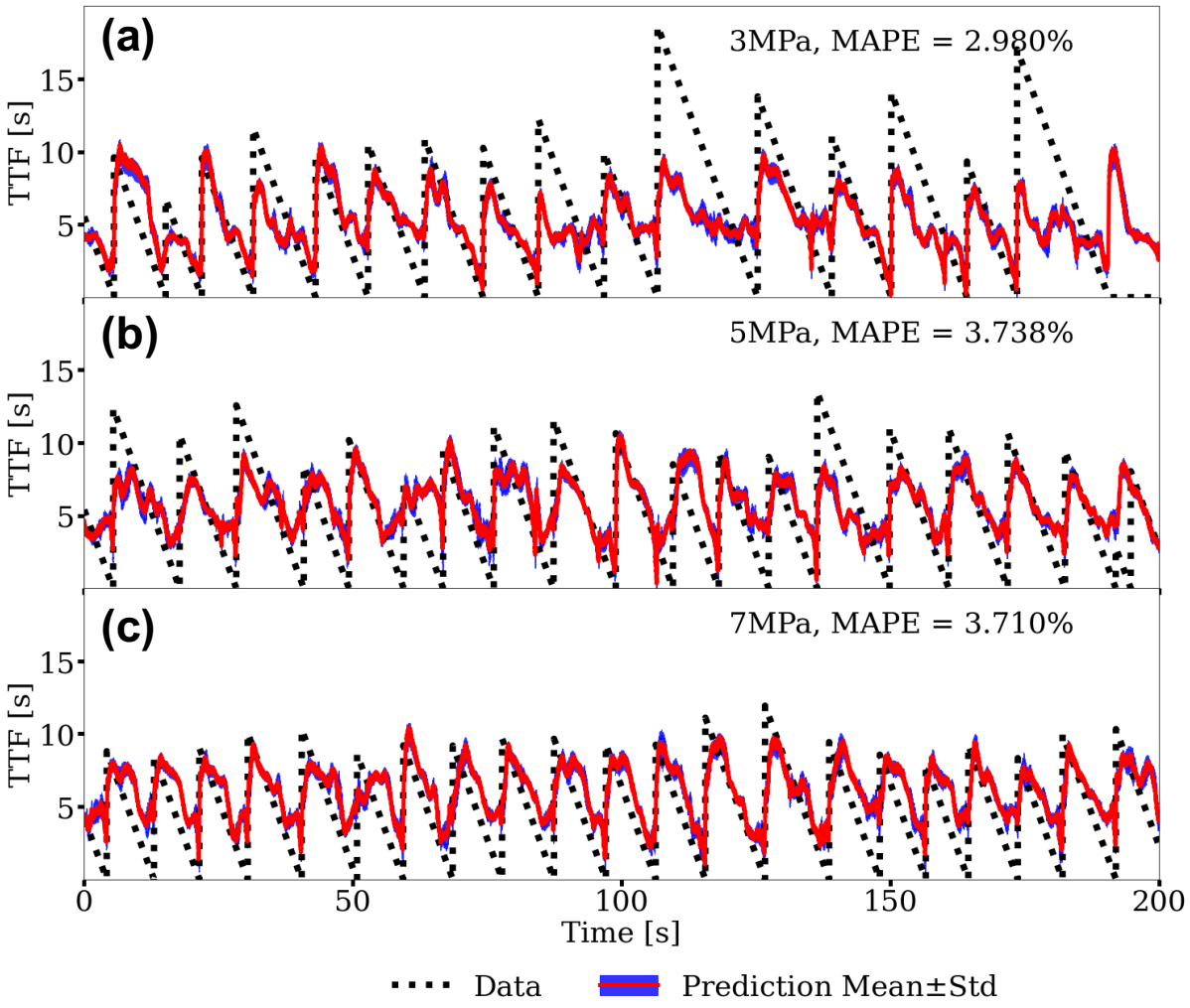

Figure 10. Predictions of time-to-failure (TTF) applying the transfer learning, cross-trained model. Predictions of TTF for laboratory p4581 data at (a) 3 MPa, (b) 5 MPa, and (c) 7 MPa confining loads using transfer learning of the CED model trained on FDEM data. Only the latent space is trained on laboratory p4677 data for the TTF predictions. The blue dashed line shows the ground truth TTF derived from the experiment. The red curve shows the model predicted TTF.

the laboratory behavior—the FDEM simulation exhibits more complex slip behaviors than the experiments in regards to the range of interevent times. Consequently, the trained model is able to predict the simpler behavior with more quasi-periodic interevent times exhibited by the experimental data. The slip frictional failure magnitude predictions are less accurate than the timing—the full range of frictional failures observed during sliding is under-predicted. Knowing this, one could conduct simulations that produce larger frictional failures to determine if this improves the laboratory failure predictions.

To summarize the significant highlights, when the model latent space is additionally trained applying long segments of the experimental AE data, the predictions improve markedly. As the latent space training data set is decreased to a portion of a single cycle, the predictions are still reliable, even when the trained model is applied to a different experiment and at varying applied load levels. Further, the time-to-failure predictions applying the cross-trained model are also surprisingly good.

Overall, the results are very promising and suggest further improvements are possible. For real-world seismic applications, the stick-slip repeat cycles can be on timescales ranging from several decades to centuries, and generally the available field recordings only cover a partial slip event. We imagine a scenario in Earth in which we have a temporally limited set of observations used in tandem with fault simulations to train a similar type of model. This may be of great value as we address evolving fault slip and earthquake hazards in the real Earth.

Methods

Numerical simulation and laboratory experiments

Numerical simulations of a laboratory experiment performed by Gao et al.[24] were obtained by applying the combined finite-discrete element method (FDEM) using the Hybrid Optimization Software Suite package (HOSS)[35] (Figure 1). The FDEM was originally developed by Munjiza[3] to simulate continuum to discontinuum transitional material behavior. FDEM combines the algorithmic advantages of the discrete element method with those of the finite element method. In FDEM, each discrete element is comprised of a subset of finite elements that are allowed to deform according to the applied load, which is particularly useful in capturing deformations in the fault gouge material as well as at the gouge particle–plate boundary.

The FDEM model was applied to simulate a two-dimensional, photoelastic shear laboratory experiment conducted by Geller and others[14]. Two-dimensional plane stress conditions were assumed and the model comprised 2,817 circular particles confined between two identical plates. Three thousand bi-dimensional particles with diameters of 1.2 and 1.6 mm were used, respectively (1,500 of each). The plates had dimensions of 570 $\times$ 250 mm. At the plate interfaces semi-circular shaped 'teeth' were placed to increase friction between plates. The particles had Young's modulus and Poisson's coefficient of 0.4 GPa and 0.4, respectively, while the plates had Young's modulus and Poisson's coefficient of 2.5 MPa and 0.49, respectively, far smaller than those used in the bi-axial shear experiment described below. Shearing velocity was 0.5 mm/s.

The laboratory data[23,30–33] were obtained from a bi-axial shear apparatus (Figure 1). Laboratory experiments fail in quasi-periodic cycles of stick and slip that mimic to some degree the seismic cycle of loading and failure on tectonic faults. The apparatus comprises a central steel block that is driven at fixed loading velocity of 10 µm/s for the experiment. This loading imparts shear stresses within two gouge layers that are 100 mm square with an initial thickness of 5 mm. The gouge layers are located on either side of the central driving block and confined by a second steel layer of 20 mm thickness. The gouge consists of monodisperse glass beads of 104–149 µm diameter with Young's modulus of 70 GPa and Poisson coefficient of 0.3; the steel blocks have Young's modulus of approximately 180 GPa and Poisson's coefficient of approximately 0.29. A load-feedback servo control system maintains a fixed normal stress of 2.5 MPa for experiment p4677, while measuring shear stress throughout the experiment. For experiment p4581, progressively larger loads were applied, and at each load level, steady state was achieved before a change to the successive load level. The shearing speed was 5 mm/s for both experiments. Mechanical data measured on the apparatus throughout the experiments included the shearing block velocity, the applied load, the gouge layer thickness, the shear stress, and the coefficient of friction. Continuous AE emissions from the fault zone seismic wave radiation were recorded with piezo-ceramics embedded inside blocks of the shear assembly[34].

We note that in the FDEM simulations considered here the AEs were not propagated in the model. We assume the kinetic energy obtained from the fault simulations as being equivalent to the AEs recorded on the experimental shear apparatus based on previous analysis[24]. Used here as an equivalent quantity to the AE is the kinetic energy (E_k) summed from the entire system. Since the plates and particles work together as an ensemble, it is the aggregate energy evolution that governs the stick-slip behavior in granular fault gouge. In the bi-axial experiment, the source of the AE signal is at the grain contact level[42]. Fault gouge contacts broadcast AE independently and/or simultaneously[42], and displace the sideblocks equivalently along the dimensions of the block due to the extreme stiffness of the steel, in analogy to the E_k behavior in the simulation. Thus, the E_k is approximately equivalent to the magnitude of the continuous AE time series (norm of the acoustic emission recorded by the two channels of the lab experiments), which is the source of elastic waves. We say approximately because there is a very modest amount of wave dissipation during wave propagation in the experiment from the fault gouge layer through the steel plates to the detectors.

Training, validation and testing data

The continuous time series signals (AE, kinetic energy, and friction coefficient) from the experiment and FDEM simulation are converted into scalograms using the Continuous Wavelet Transform (CWT; see reference[36] for a comprehensive description of the method) to utilize the time-frequency signal strength in the CED models. We adopt the real Ricker (Mexican-hat, DOG (m=2)) wavelet for the CWT, which is commonly used in analyzing seismic data[37]. For comparison we also tested the Morlet wavelet and found the Ricker to produce improved MAPE results. The reconstruction of the signal from the scalograms (inverse CWT) is the sum of the real part of the wavelet transform over all scales.

For the FDEM simulation, the CWT is performed on the training/validation/testing (60/20/20% split) segments of the kinetic energy (E_{kin}) and friction (μ) time series. Scalograms are calculated using moving windows with a size of 2 seconds and step of 0.2 seconds. The sliding window size does not impact the accuracy of the CED model, see Supplementary Material. For a sampling frequency of $f_s = 1000Hz$, each scalogram is 128×2000. The procedure produces 73 and 19 pairs of input E_{kin} and output μ scalograms for the training and validation, respectively.

The training data is augmented by producing additional noisy E_{kin} input signals. The procedure is as follows: (1) take the Fast Fourier Transform (FFT) of the original signal data, (2) shuffle the positive-frequency terms of the imaginary coefficients, (3) set the negative-frequency terms to the opposite of the shuffled terms, (4) and perform the inverse FFT to produce a new E_{kin} signal with the same amplitude spectrum and random phase. The procedure is repeated 3 times for the training signal and the final training data contains 292 scalogram pairs.

The CWT transform procedure is applied to the laboratory experiment p4677 acoustic emission (AE) and friction (μ) time series to produce training/validation/testing (20/20/60% split) data. The scalogram dimensions are the same as the numerical simulations. The final data set contains 292 pairs of input AE_{norm} and output μ scalograms for the training and validation data. Scalograms are calculated for the laboratory experiment p4581 and only used as testing data for experiments conducted at different normal stresses.

Before applying the CWT, all input and output signals are normalized by subtracting the mean and dividing by the standard deviation using the statistics extracted from the training signal data. For FDEM data, the statistics are 3.28E-4$\pm$5.00E-4 for the input E_k signals and 4.23E-1$\pm$2.52E-2 for the output μ signals. For transfer learning on the p4677 data, the statistics from the training signals (0s to 60s, including 6 stick-slip cycles) are 8.932$\pm$14.900 for the input AE signals and 0.657$\pm$0.0382 for the output μ. In the cases of limited sub-cycle data, the post-failure training signal has statistics of 7.712$\pm$10.667 for AE and 0.641$\pm$0.0440 for μ, and the pre-failure training signal has statistics of 10.205$\pm$20.137 for AE and 0.667$\pm$0.0377 for μ. When making predictions using the laboratory p4581 data with increasing normal loads, the statistics are extracted from the first 20% of the 3MPa signal (from 0s to 40s, including 5 stick-slip cycles) to obtain 17.776$\pm$46.700 for AE and 0.433$\pm$0.0230 for μ. For the TTF statistics the values are 4.816$\pm$3.257 on the p4677 data and 4.817$\pm$2.873 on the p4581 data, extracted from the same aforementioned signal segments.

Convolutional encoder-decoder model and transfer learning

The CED architecture is composed of an encoder branch containing the salient features that feeds to a latent space, and a decoder branch to construct the output variable. The input signal is passed to an encoding branch with a preprocessing block containing 2 convolutional layers and a rectified linear unit (ReLU) activation function (Figure 3). Preprocessing is used to reduce the image size in the time dimension by a factor of 25. This is passed through 4 downsampling blocks containing 3 convolutional layers, each with batch normalization, ReLU activation, and a skip connection. The latent space contains 2 convolutions and a ReLU activation. The decoding branch reverses the encoding using convolutional transpose layers. The postprocessing contains 2 convolutional transpose layers to obtain the original dimensions. The dimension of each layer, i.e., the filter size, depth, and skip connections are labeled in Figure 3. The model contains 5 *skip* connections[38] that directly link the weights from the downsample blocks in the encoder to the upsample blocks in the decoder at each level. The trainable weights are initialized using glorot_uniform and the biases are nontrainable and set to zero.

Loss functions are calculated hierarchically for each pair of encoder/decoder blocks. This type of hierarchical regularization was recently introduced by Wang et al.[44]to provide better interpretability and generalizability of CED models for learning fluid-flow patterns in complex rock pore-structures. This regularization is found to improve the MAPE accuracy in predicting FDEM test data by 1% and provides a similar level of MAPE for the transfer learning. The total loss is calculated as $L_{total} = L^0_{hier} + L^1_{hier} + L^2_{hier} + L^3_{hier} + L^4_{hier} + L_{reconstr} + L_{l2}$. Where L^i_{Hier} is the mean square error (MSE) between the target and predicted values for each sub-model linked with a *skip* connection (Figure 3 (a)). $L_{reconstr}$ is the reconstruction loss using the entire CED model. And L_{l2} is the loss associated to the L2 regularization using a penalty multiplier[43] set to 1E-5. After the initial training, the *skip* connections are deactivated so that information only passes down the encoder, through the latent space, and up the decoder for a prediction. The CED model contains 363,696 trainable parameters, with 73,984 in the latent space.

The model is trained using 292 pairs of scalograms with a batch size of 8, the Adam optimizer, and a learning rate of 1E-3. Validation is performed with 19 pairs of scalograms. The training is stopped when the reconstruction loss on the training data is

below 0.1 and the validation reconstruction loss does not diminish for 100 epochs. The model with the lowest validation loss are used as the final CED model.

Transfer learning is applied using the laboratory p4677 data. A new CED model is instantiated with the weights from the final model trained on the FDEM simulation data. All trainable weights, except the latent space, are rendered non-trainable and held constant while the latent space is further trained with the laboratory data. Since the encoder and decoder branches are non-trainable layers, the total loss is $L_{total} = L_{reconstr} + L_{l2}$ and the early stopping is the same.

Model training reproducibility

Due to random variable initialization and the stochastic nature of training a neural network, repeating the training procedure gives different results and variations in the overall performance. We trained models using the transfer learning work flow and tested three source of randomness: (1) initialization of the network weights, (2) shuffling of the FFT coefficients to produce noise for data augmentation, and (3) shuffling of training data for optimizer. The prediction accuracy on the validation and testing signals is shown with the MAPE in Tables S1, S2, S3 and Figure S1 in the Supplementary Material. The results show that the random initialization and shuffling of the batches produce discrepancies between the model predictions, and increasing the noise in the data reduces the variance and improves accuracy. The figures presented in this work come from the CED model trained in Run No. 8 of Table S3, which has the overall accuracy nearest to the mean performance of the 10 separate runs with random initial weights and random noisy data augmentation.

References

1. Scholz, C. H. The Mechanics of Earthquakes and Faulting. *Cambridge University Press* (2019).
2. Ren, C. X., Hulbert, C., Johnson, P. A., & Rouet-Leduc, B. Machine learning and fault rupture: a review. *Advances in Geophysics* **61**, 57–107 (2020).
3. Munjiza, A. The combined finite-discrete element method. *John Wiley & Sons, Ltd* (2004).
4. Bergen, K. J., Johnson, P. A., Maarten, V., & Beroza, G. C. Machine learning for data-driven discovery in solid Earth geoscience. *Science* **363**(6433), (2019).
5. Rouet-Leduc, B., Hulbert, C., Lubbers, N., Barros, K., Humphreys, C. J., & Johnson, P. A. Machine learning predicts laboratory earthquakes. *Geophysical Research Letters* **44**(18), 9276–9282 (2017).
6. Rouet-Leduc, B. et al. Estimating fault friction from seismic signals in the laboratory. *Geophysical Research Letters* **45**(3), 1321–1329 (2018).
7. Lubbers, N., Bolton, D. C., Mohd-Yusof, J., Marone, C., Barros, K., & Johnson, P. A. Earthquake catalog-based machine learning identification of laboratory fault states and the effects of magnitude of completeness. *Geophysical Research Letters* **45**(24), 13–269 (2018).
8. Hulbert, C., Rouet-Leduc, B., Johnson, P. A., Ren, C. X., Rivière, J., Bolton, D. C., & Marone, C. Similarity of fast and slow earthquakes illuminated by machine learning. *Nature Geoscience* **12**(1), 69–74 (2019).
9. Jasperson, H., Bolton, C., Johnson, P., Marone, C., & de Hoop, M. V. Unsupervised classification of acoustic emissions from catalogs and fault time-to-failure prediction. *arXiv preprint arXiv:1912.06087* (2019).
10. Bolton, D. C., Shokouhi, P., Rouet-Leduc, B., Hulbert, C., Rivière, J., Marone, C., & Johnson, P. A. Characterizing acoustic signals and searching for precursors during the laboratory seismic cycle using unsupervised machine learning. *Seismological Research Letters* **90**(3), 1088–1098 (2019).
11. Zhou, Z., Lin, Y., Zhang, Z., Wu, Y., & Johnson, P. Earthquake detection in 1D time-series data with feature selection and dictionary learning. *Seismological Research Letters* **90**(2A), 563–572 (2019).
12. Rouet-Leduc, B., Hulbert, C., McBrearty, I. W., & Johnson, P. A. Probing slow earthquakes with deep learning. *Geophysical research letters* **47**(4), e2019GL085870 (2020).
13. Johnson, C.W. & Johnson, P. A. Learning the low frequency earthquake daily intensity on the central San Andreas Fault. *Geophysical research letters* **48**, e2021GL092951 (2021).
14. Geller, D. A., Ecke, R. E., Dahmen, K. A., & Backhaus, S. Stick-slip behavior in a continuum-granular experiment. *Phys. Rev. E* **92**, 060201 (2015).
15. Pan, S. J., & Yang, Q. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering* **22**(10), 1345–1359 (2009).
16. Goodfellow, I., Bengio, Y., & Courville, A. Deep learning. *MIT press* (2016).

17. Chevitarese, D., Szwarcman, D., Silva, R. M. D., & Brazil, E. V. Transfer learning applied to seismic images classification. *AAPG Annual and Exhibition* (2018).

18. Siahkoohi, A., Louboutin, M., & Herrmann, F. J. The importance of transfer learning in seismic modeling and imaging. *Geophysics* **84**(6), A47–A52 (2019).

19. Cunha, A., Pochet, A., Lopes, H., & Gattass, M. Seismic fault detection in real data using transfer learning from a convolutional neural network pre-trained with synthetic seismic data. *Computers & Geosciences* **135**, 104344 (2020).

20. Zhang, Z., & Lin, Y. Data-driven seismic waveform inversion: A study on the robustness and generalization. *IEEE Transactions on Geoscience and Remote sensing* **58**(10), 6900–6913 (2020).

21. Rouet-Leduc, B., Jolivet, R., Dalaison, M., Johnson, P. A., & Hulbert, C. Autonomous Extraction of Millimeter-scale Deformation in InSAR Time Series Using Deep Learning. *arXiv preprint arXiv:2012.13849* (2020).

22. Sun, X., Zimmer, A., Mukherjee, S., Kottayil, N. K., Ghuman, P., & Cheng, I. DeepInSAR—A deep learning framework for SAR interferometric phase restoration and coherence estimation. *Remote Sensing* **12**(14), 2340 (2020).

23. Johnson, P. A. et al. Acoustic emission and microslip precursors to stick-slip failure in sheared granular material. *Geophysical Research Letters* **40**(21), 5627–5631 (2013).

24. Gao, K. et al. Modeling of stick-slip behavior in sheared granular fault gouge using the combined finite-discrete element method. *Journal of Geophysical Research: Solid Earth* **123**(7), 5774–5792 (2018).

25. Okubo, K., Rougier, E., Lei, Z., & Bhat, H. S. Modeling earthquakes with off-fault damage using the combined finite-discrete element method. *Computational Particle Mechanics* **7**, 1057–1072 (2020).

26. Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. How transferable are features in deep neural networks?. *arXiv preprint arXiv:1411.1792* (2014).

27. Shin, H. C. et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE transactions on medical imaging* **35**(5), 1285–1298 (2016).

28. Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. Imagenet: A large-scale hierarchical image database. *2009 IEEE conference on computer vision and pattern recognition*, 248–255 (2009).

29. Huh, M., Agrawal, P., & Efros, A. A. What makes ImageNet good for transfer learning?. *arXiv preprint arXiv:1608.08614* (2016).

30. Dieterich, J. H., & Conrad, G. Effect of humidity on time-and velocity-dependent friction in rocks. *Journal of Geophysical Research: Solid Earth* **89**(B6), 4196–4202 (1984).

31. Marone, C. Laboratory-derived friction laws and their application to seismic faulting. *Annual Review of Earth and Planetary Sciences* **26**(1), 643–696 (1998).

32. Niemeijer, A., Marone, C., & Elsworth, D. Frictional strength and strain weakening in simulated fault gouge: Competition between geometrical weakening and chemical strengthening. *Journal of Geophysical Research: Solid Earth* **115**(B10), (2010).

33. Scuderi, M. M., Marone, C., Tinti, E., Di Stefano, G., & Collettini, C. Precursory changes in seismic velocity for the spectrum of earthquake failure modes. *Nature geoscience* **9**(9), 695–700 (2016).

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

35. Knight, E. E. at al. HOSS: an implementation of the combined finite-discrete element method. *Computational Particle Mechanics* **7**(5), 765–787 (2020).

36. Torrence, C., & Compo, G. P. A practical guide to wavelet analysis. *Bulletin of the American Meteorological society* **79**(1), 61–78 (1998).

37. Gholamy, A., & Kreinovich, V. Why Ricker wavelets are successful in processing seismic data: Towards a theoretical explanation. *2014 IEEE Symposium on Computational Intelligence for Engineering Solutions (CIES)*, 11–16 (2014).

38. Ronneberger, O., Fischer, P., & Brox, T. U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*, 234–241 (2015).

39. Szegedy, C. et al. Going deeper with convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1–9 (2015).

40. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. Rethinking the inception architecture for computer vision. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2818–2826 (2016).

41. Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. Inception-v4, inception-resnet and the impact of residual connections on learning. *Proceedings of the AAAI Conference on Artificial Intelligence* **31**(1), (2017).

42. Trugman, D.T., McBrearty, I.W., Bolton, D.C., Guyer, R.A., Marone, C. & Johnson, P. A. The Spatiotemporal Evolution of Granular Microslip Precursors to Laboratory Earthquakes. *Geophysical Research Letters* **47** (16), (2020).

43. Chollet, F. Xception: Deep learning with depthwise separable convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1251–1258 (2017).

44. Wang, K., Chen, Y., Mehana, M., Lubbers, N., Bennett, K. C., Kang, Q., Viswanathan, H. S. & Germann, T. C. A physics-informed and hierarchically regularized data-driven model for predicting fluid flow through porous media. *Journal of Computational Physics*, 110526 (2021).

Acknowledgements

KW, KCB and PAJ acknowledge support by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Chemical Sciences, Geosciences, and Biosciences Division under grant 89233218CNA000001. KW acknowledges support by the Center for Nonlinear Studies (CNLS) at Los Alamos National Laboratory. CWJ acknowledges Institutional Support (Laboratory Directed Research and Development) at Los Alamos National Laboratory. The authors declare no competing interests. We thank Chris Marone for the laboratory data and Ke Gao for the numerical simulation data.

Author contributions statement

KW developed the CED model and conducted the machine learning analysis with input from CWJ and PAJ. KW and KCB conceived of the transfer learning analysis and workflow. PAJ conduced experiments with collaborators at the Pennsylvania State University. PAJ was involved in collecting the simulation and experimental data. All authors were involved in writing the manuscript.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TransferLearningsupportinfo.pdf