Induced microearthquakes predict permeability creation in the brittle crust

Ziyan Li (✉ liziyan1992@gmail.com)

The Pennsylvania state university

Derek Elsworth

Pennsylvania State University

Chaoyi Wang

The Pennsylvania state university

Article

Keywords:

Posted Date: January 13th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1244014/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Induced microearthquakes predict permeability creation in the brittle crust

Ziyan Li^{1,3*}, Derek Elsworth¹⁻³, Chaoyi Wang^{1,3*}, EGS-Collab⁴

¹Department of Energy and Mineral Engineering, The Pennsylvania State University, University Park, PA, USA

²Department of Geosciences, The Pennsylvania State University, University Park, PA, USA

³G3 Center and EMS Energy Institute, The Pennsylvania State University, University Park, PA, USA

⁴EGS-Collab Team – Author list appended at the end of the manuscript.

Abstract

Fracturing controls rates of mass, chemical and energy cycling within the crust. We use observed locations and magnitudes of microearthquakes (MEQs) to illuminate the evolving architecture of fractures reactivated and created in the otherwise opaque subsurface. We quantitatively link seismic moments of laboratory MEQs to the creation of porosity and permeability at field scale. MEQ magnitudes scale to the slipping patch size of remanent fractures reactivated in shear - with scale-invariant roughnesses defining permeability evolution across nine decades of spatial volumes – from centimeter to decameter scale. This physics-inspired seismicity-permeability linkage enables hybrid machine learning (ML) to constrain in-situ permeability evolution at verifiable field-scales (~10 m). The ML model is trained on early injection and MEQ data to predict the dynamic evolution of permeability from MEQ magnitudes and locations, alone. The resulting permeability maps define and quantify flow paths verified against ground truths of permeability.

1 **Introduction**

2 Enhanced geothermal systems (EGS) offer the possibility to provide plentiful and continuous
3 carbon-free energy for the increasing global demand of electric power (Kneafsey et al., 2018&
4 2019; Williams et al., 2008) and have significant potential to shift the current dependence on
5 fossil fuels (Fridleifsson et., al 2018). The net recoverable energy and power depend on reservoir
6 temperature, utilizable reservoir volume (Bauer et al., 2019) and rates of fluid mass and heat
7 transfer, which are in turn controlled by reservoir-scale petrophysical parameters (Laubach et al.,
8 2009; Bauer et al., 2017; Kushnir et al., 2018). Thus, the ability to create a reservoir and to
9 predict permeability evolution before, during and after stimulation at reservoir scale and at the
10 resolution of individual fractures is essential to accurately estimate fluid mass transfer and heat
11 recovery from the reservoir.

12 A key challenge is in defining reservoir-scale response in the opaque subsurface where
13 observation is limited to the acoustic signal from the micro-earthquakes (MEQs) that typically
14 accompany the stimulation. This is exacerbated by the highly stress- and pressure-sensitive
15 controls in creating new fractures and reactivating existing fractures. This resulting and
16 improved fluid transmission and heat transfer response may be evaluated if the geometric
17 evolution and connectivity of the fracture network is known/predictable. However, the
18 complexity of this nonlinear response and reinforcing feedbacks linking changes in reservoir
19 pressure to changes in permeability renders such problems intrinsically ill-constrained and ill-
20 defined. This situation is compounded by the reality that initial conditions are also poorly
21 constrained and perhaps unknowable. Additionally, these models rely on the integration of
22 multiple geologic, geochemical, and geophysical observations collected during and after well
23 stimulation. Thus, this innate complexity renders accurate predictions of permeability evolution
24 during well stimulation nigh impossible – limiting the utility of traditional forecasting methods
25 in light of such complexity.

26 The application of machine learning (ML) models potentially eases this difficulty by
27 accommodating prior learning from multiple prior experiences in previously-probed and self-
28 similar reservoirs with common nonlinear interactions and feedbacks. Machine learning (ML),
29 deep learning (DL) and hybrid methods combined with other techniques offer promise in
30 predicting static permeability structure in reservoirs. This includes permeability structure

recovered from key petrophysical features (e.g. density, grain or pore diameter and porosity) recovered from core data (Al Khalifah et al., 2019; Erofeev et al., 2019) using ML methods and from well logging data (Gholami et al., 2012; Eriavbe et al., 2019; Ahmadi et al., 2019; Khanet al., 2019; Karimpouli et al., 2020) using DL. Combining 3D seismic and magnetotelluric (MT) datasets (Matzel, et al., 2021) or pore structure recovered from X-ray micro-CT (Wu et al., 2018; Tian et al., 2020; Tembely et al., 2021) have replicated permeability structure using hybrid and DL models. These methods provide valuable insights into applying ML or DL methods to predict permeability. However, lacking are ML methods to predict permeability evolution where stimulation at reservoir scale results in significant changes in permeability.

Figure 1. Logic chart defining steps to recover “ground-truth” mean permeabilities and to develop a permeability map.

We present a novel hybrid ML method to map 3D permeability evolution using microearthquake (MEQ) data recovered concurrently with fluid injection history during well stimulation. This strategy combines an ML model to predict average permeability evolution in the reservoir with a physical model connecting fracture permeability MEQ magnitude. We use experimental data from a unique intermediate-scale (~10m) hydraulic stimulation experiment conducted at depths and stresses representative of real reservoirs (Kneafsey et al., 2017&2018) – the highly constrained EGS-Collab project. Step-rate injection is applied to create fractures by hydraulic stimulation with well head pressures and flow rates concurrently measured together with continuous monitoring of active and passive seismics (CASSM) (Fu et al., 2018, Schoenball et al. 2020). We analyze the data from five stimulation episodes, where the location, timing, and

relative magnitudes of MEQs are defined concurrently with time histories of injection (Schoenball et al. 2020; Fu et al.,2021). The analysis is completed in three steps (Figure 1).

In Step 1: We evaluate injectivity directly from the ratio of wellhead pressure and flow rate histories. These are defined over the five stimulation episodes of the experiment, punctuated by halts and shut-ins. Then we convert injectivity to a mean permeability by incorporating an approximated flow geometry. This defines the “ground truth” – a deterministic measure of both injectivity and average permeability of the system.

In Step 2: We use the injectivity and MEQ data from stimulation over the first three episodes (#1-#3) to build a supervised ML model based on a gradient boosting algorithm (XGBoost, Chen et al.,2016). We use these trained data predict injectivity from MEQs alone over the final two episodes (#4). We then convert this history of injectivity into a mean permeability – using the same moving geometric conversion used in Step 1. The predicted injectivity from the ML model is then compared with the ground truth value of injectivity recovered from Step 1 to evaluate the accuracy of the ML model. Furthermore, this predicted injectivity is converted to an average permeability and compared to the deterministic estimate of permeability from Step 1.

In Step 3: We use magnitudes and locations of MEQs to independently constrain the spatial (fracture-by-fracture) permeability distribution using an empirical physical model that links permeability evolution to MEQ moment. The detailed derivation of this empirical relation can be found in Methods. This spatially distributed fracture-by-fracture permeability is then combined to yield an average permeability and compared with the ground truth for validation.

Results

Deterministic Evaluation of Permeability [Step 1]. Time histories of injection pressure (red) and flow rate (blue) for the five stimulation episodes are shown in Figure 2 a1-a5, including the distance from injection to the spawned MEQs (orange circles scaling with event magnitude). Figure 2 b1-b5 shows these processed data that define injectivity as the ratio of flow rate to pressure (light green line) and cumulative number of seismic events (orange line) for each episode. This deterministic magnitude of injectivity is converted to an average reservoir permeability by assuming that: (1) fluid flow across the formation quasi-radial from the injection

wellbore to an external pressure boundary, (2) this boundary migrates outwards with the cloud of MEQs, and (3) the MEQs are partitioned between those resulting from changes in effective stress (80% assumed) and those beyond this region resulting from changes in total stress (20% assumed). Thus, the activated reservoir is confined to a volume that only contains fractures (MEQs) reactivated by fluid percolating directly from the injection (80% of the events) and that this stimulated volume grows with time as the external cylindrical boundary contour migrates as more distant MEQs are recorded. Thus, this migrating cylindrical envelope is estimated from the cumulative frequency of MEQs-with-radius and capped where cumulative frequency is 0.8 (Figure 2 b1-b5).

Figure 2. Seismic and injection data for the EGS-Collab stimulation experiment (a1-a5) show the injection history of the five stimulation episodes (blue line shows wellhead pressure and red line shows flow rate) and distance to individual MEQs from the injection well (orange symbols). Symbol size proportional to MEQ magnitude. (b1-b5) show these data converted to injectivity history (ratio of flow rate and pressure; light green lines) and cumulative number of MEQs (orange line) for each of the five stimulation episodes. Dashed line delimits the periods from the first recorded MEQ to well shut-in during Episodes #4 (b4) and #5 (b5).

Figure 3 (a) shows the individual and cumulative number of MEQs with their distance from the injection well for episode #4 at time T . The radius to the migrating outer pressure boundary is defined at this time T where cumulative frequency is 0.8 (grey line). However, the migrating flow radius is a continuous function of elapsed time as shown for episodes #4 and #5 in Figures 3(b) and 3(c). Clear from this is that the effective radius of flow migrates monotonically outwards during these two latter experiments. The radius is zero before any MEQ and grows to ~ 7.5 m at the end of episode #4 and ~ 12 m for episode #5, suggesting a continuous flow pathway between injection and production wells.

118

119 We estimate average permeability from the injectivity by presuming radial flow from the 10m
 120 long borehole with interior and exterior pressure boundary conditions fixed to those of the
 121 injection and production well, respectively, as, $k = \frac{I\mu \ln(r/r_0)}{2\pi h}$, where k is the average
 122 permeability (m^2), μ is the viscosity of water ($8.9 \times 10^{-4} \text{ Pa}\cdot\text{s}$), r is the migrating radius of the
 123 cylindrical volume (m), which is derived from the migration of MEQs during the experiment. r_0
 124 is the interior/injection wellbore radius (0.048m). This represents an approximation of the
 125 geometric correction for flow that is likely heterogeneous with azimuth but is here considered
 126 uniform. The deterministic estimate of permeability for episodes #4 and #5 is shown as the light
 127 green line in Figure. 3(d) and 3(e). Figure 3(f) shows well pressure (black line) and flow rate
 128 (blue line) histories for each of the injection (top) and production wells (bottom) in episode #4.
 129 Flow rate was observed to spike for approximately one minute at the production well,
 130 highlighted in the blue shaded area, indicating that fluid had propagated to the production well.

131 The permeability during this short period can be estimated using Darcy's Law, $k = \frac{dq\mu L}{Adp}$, where
132 dq (1.68 LPM) is the net average flowrate between the two wells, L is the distance between
133 injection and production wells ($\sim 10\text{m}$), A is the cross-sectional area ($\pi \cdot (\frac{10}{2})^2 m^2$), dp (25.14 MPa)
134 is the average pressure difference between two wells. The resulting permeability is $1.2 \times$
135 $10^{-16} m^2$, which is close to our estimation of $2 \times 10^{-16} m^2$ at the same time (T) based on the
136 cylindrical flow geometry. This further confirms that the assumption of $\sim 80\%$ of seismic events
137 directly induced by effective stress (direct contact with elevated fluid pressures) is reasonable.

138 **ML predicted average permeability in the reservoir [Step 2].** The ML model is built using
139 data from the first three injection episodes #1-#3 with ~ 450 MEQs and ~ 2.5 hours of well
140 stimulation (training data). This comprises nine statistical features of the seismic data calculated
141 over a small moving time window and labeled by injectivity over the same time window. Each
142 time window is one minute in duration. The first 80% of the dataset is used to train the model
143 with the remaining 20% used to quantify whether the model is overfitting. Hyperparameters are
144 set by five-fold cross-validation and the optimal model is selected using root mean squared error
145 ($RMSE$) as an evaluation metric. A step-by-step description of both the data analysis and
146 machine learning methods is documented in the Methods section. Note that the training data may
147 be any single episode, or multiple episodes, that precede the predicted episode. For example, to
148 estimate injectivity in episode #4 (testing data), the training data can be episode #3 only, or all
149 three episodes from #1 to #3. We run six tests in total to determine the optimal training dataset
150 and input features to predict injectivity for episodes #4 and #5. The nine statistical features of
151 the MEQs produce a total of 511 possible combinations. Each combination (of features) is
152 applied to build the ML model for each test. The best model for each test is defined with R^2
153 closet to unity using the difference between the deterministic and predicted in each episode.
154 These models are selected for each of the best with its input features, R^2 , $RMSE$ and
155 hyperparameters shown in Table 1.

156 Table 1. Input features and hyperparameter performance for various training and testing datasets.

Test no	Train Dataset	Test Dataset	R2	RMSE	Input Features	Max_Depth	Min Child Weight	Eta	Subsample	Colsample bytree
1	#1 -#3	#4	0.00	7.58E-04	4,9	9	1	0.2	0.9	0.9
2	#3	#4	0.63	5.60E-02	3,8	9	1	0.5	0.9	0.9
3	#1 -#3	#5	0.00	4.89E-04	4,9	9	1	0.2	0.9	0.9
4	#3	#5	0.69	1.71E-02	3,5,7,8	2	8	0.4	0.9	0.9
5	#1 -#4	#5	0.43	2.01E-02	5,6,8	5	8	0.3	0.9	0.9
6	#3 -#4	#5	0.13	6.00E-03	3,5,7,8	9	1	0.1	0.9	0.9
7	#4	#5	0.35	1.87E-02	1,2,3,4,7	2	7	0.4	0.9	0.9

input features No: 1.R 2.Dmean 3.Dmax 4.Dmin 5.Mw_mean 6.Mw_max 7.Mw_min 8.N 9.Ntot1

158 We show the most accurate predictions of injectivity from the ML model (red line) for episode
159 #4 (Figure.4a) and episode #5 (Figure.5a) and compare them with the deterministic injectivities
160 (blue line). ML is applied to predict the injectivity during the period beginning from the first
161 MEQ to well shut-in (the time between the two dashed lines in Figure 2 b4 and b5). Scatter plots
162 in Figures 4b and 5b shows the ground truth of the average injectivity versus the ML estimations
163 using only the statistical characteristics of the MEQs. A perfect model would follow the grey
164 dashed line. *RMSE* and *R2* are calculated for each episode, with *RMSE*=0.056 and *R2*=0.63 for
165 episode #4 and *RMSE*=0.017, *R2*=0.69 for episode #5. This, together with the general
166 concurrence in the time histories, confirms the excellent agreement between ground truth and
167 ML predicted injectivities. Importantly, this injectivity is predicted purely based on training of
168 the ML algorithm against injectivity-vs-MEQ histories over episode #3 to predict injectivity over
169 episodes #4 and #5 from the MEQs-alone. These predictions are derived purely from the
170 statistical features of the seismic events corresponding to a history of injectivity that the model
171 has never seen. We emphasize that there is no past nor future information considered when
172 making such a prediction. Each prediction uses only the statistical features of seismic events
173 within a single moving window. Thus, we can predict the corresponding history of injectivity in
174 episodes #4 & #5 from the seismic catalog, alone.

176 Figure 4. **Predicted injectivity during episode #4.** (a) Average injectivity within a moving time window
 177 (blue line) and predicted injectivity (red line) from the ML model for episode #4. (b) Input features used
 178 to construct the ML model. (c) Average permeability vs. predicted permeability for episode #4. (d)
 179 Feature importance of input features.

181 Figure 5. **Predicted injectivity during episode #5.** (a) Average injectivity within a moving time window
 182 (blue line) and predicted injectivity (red line) from the ML model for episode #5. (b) Input features used

to construct the ML model. (c) Average permeability vs. predicted permeability for episode #5. (d) Feature importance of input features.

The trained XGBoost model automatically evaluates the relative importance on the predictive features in honoring the training data and therefore provides a nascent physical understanding for the problem. The F-score defines the relative importance of each input feature, which infers the relative importance of that corresponding feature in controlling response of the physical system. The evolution of the predictive key features for episodes #4 and #5 are shown in Figures 4c and 5c, together with their feature importance in Figures 4d and 5d. For estimates of injectivities, the top two features correspond to the maximum separation of the evolving MEQs cloud from the injection well and the number of MEQs occurring within a given time window. Both of these indicate the overriding importance of MEQs in generating connected porosity that is then reflected in the evolving permeability architecture.

We then convert the predicted injectivity to a mean permeability by incorporating an approximated flow geometry and estimated magnitude of the migrating flow radius as in Step 1. The evolution of the predicted average permeability for each of episodes #4 and #5 is shown in Figure 2d and 2e for ready comparison. We observe that both the deterministic and ML predicted average permeabilities gradually increase to a maximum until well shut-in. The excellent agreement between deterministic and ML predicted average permeability curves confirms the accuracy of ML model with $RMSE=2.35 \times 10^{-17}$ and $R2=0.72$ for episode #4 and $RMSE=2.18 \times 10^{-17}$ and $R2=0.81$ for episode #5.

3D maps of in situ permeability evolution and permeability architecture [Step 3]. A physics-inspired model of MEQs is applied to define the spatial evolution of permeability on individual fractures remobilized by MEQs. This model assumes that the roughness of individual fractures is self-similar and fractal and may be represented by a power law (Eq. 2). Thus, observed seismic moment is linearly related to remobilized shear displacement (Eq. 2) that in turn may be related to a change in permeability (Eq. 3) for slip over an equivalent length of the fracture (Eq. 2). This allows change in permeability to be related to the seismic moment— and since the fractures are self-similar over many decades of scale, laboratory measurements on a small sample may be directly scaled to MEQ patch dimensions of the remobilized fractures (Eq. 10).

Thus, we determine local fracture permeability created by MEQs from equation (10) and represented by the data-point cloud. The matrix is effectively impermeable ($\sim 10^{-19} \text{ m}^2$; Neupane et al., 2019). Figures 6 a & b show the evolution of permeability created on individual fractures by MEQs from initiation to termination of the respective single injection episodes #4 and #5. The predicted fracture permeability is shown spatially and indexed by color intensity. Symbol size reflects the magnitude of the related seismic events. Fracture permeability created by all MEQs during the entire stimulation episode are shown in Figure 6. c (from #1-#4) and Figure 6. d (#1-#5). The cylindrical outer contour of the moving pressure boundary migrates with the MEQs with the color indicating average reservoir permeability predicted from the ML model. The stimulated zone of enhanced fracture permeability migrates radially outwards from the injection well (red line) preferentially towards the production wells (green line). This suggest that (re)activation of new and existing fractures propagates from the injection well towards the production well that is azimuthally focused towards the production well. A plan view of the 3D MEQ point cloud for all episodes (Figure 7) shows the form of this fracture propagation from the injection to production well and identifies possible fluid pathways. Symbols shaded from light to dark green shows the temporal evolution of permeability from episode #1 to #5 with symbol size indicating the MEQ magnitude. Reactivation migrates from injection to production well (Figure.7a) and propagates from shallow to deep within the reservoir (Figure.7b). The 3D in-situ permeability map shown in Figure 6 (c)&(d) integrates both average reservoir permeability and fracture permeability distribution. Importantly, the map defines both the spatial distribution of permeability (symbol color) and the magnitude and distribution of heat transfer area (symbol size proportional to magnitude) – a key factor in defining the efficiency of heat exchange within a prospective geothermal reservoir. The greatest permeability enhancement is represented by the denser zone of fractures. The insights gained through application of these methods can be incorporated into conceptual models and utilized for planning exploration and development strategies in operating geothermal fields.

241 **Figure 6. Spatial distribution of evolving in situ permeability** Spatial distribution of permeability on
 242 individual fractures (MEQs) for episodes (a) #4 and (b) #5 only. Local fracture permeability accumulated
 243 from episodes (c) #1-#4 and (d) #1-#5. The cylindrical constant fluid pressure boundary migrates to
 244 include of 80% of the MEQs at any time with color indicating average reservoir permeability predicted
 245 from the ML model at the end of episodes #4 and #5. Red and blue line represent injection and production
 246 well, respectively.

248 **Figure 7. Point cloud viewed from (a) top and (b) side showing distribution of MEQs** originating
 249 from each stimulation episode to visualize the sequence of fracture propagation and development of
 250 potential fluid paths. (I) Comparison between the ground truth of average permeability (Step 1) and
 251 average permeability estimated from MEQ-predicted spatial permeability variation (Step 3). Times as
 252 referenced in Figures 3 & 4.

253

254 Table 2. Cross validation of permeability by different methods.

Episode No	permeability between two wells	deterministic permeability (Step 1)	ML model (Step2)	Physic-inspired model (step3)
1	1.20E-16	2.03E-16	2.28E-16	8.03E-16
2	-	2.21E-16	2.55E-16	9.06E-16

Episode No	Permeability Between Two Wells	Deterministic Permeability (Step 1)	ML Model (Step2)	Physics-inspired Model (step3)
1	1.20E-16	2.03E-16	2.28E-16	8.03E-16
2	-	2.21E-16	2.55E-16	9.06E-16

257 We average the evolving fracture-by-fracture spatial distribution of permeability previously
 258 recovered from equation (10) (Step 3) to compare against the average permeability recovered
 259 from Steps 1 & 2. We mesh a 20×20×10m cube that contains the 10m radius cylinder and
 260 discretize the volume using 0.1 m grid. Each calculated fracture permeability value is then
 261 assigned into the mesh block based on its location. If several fractures are in the same mesh
 262 block, then the largest (most dominant) of the fracture permeabilities is selected. Mesh blocks
 263 without fractures are assigned a matrix permeability of 10^{-19}m^2 . We arithmetically average and
 264 volume-weight the permeability across all mesh blocks. The cross comparison of average
 265 permeability by different methods at the end of episodes #4 and #5 are shown in Table.2.
 266 Permeability between two wells for episode#4 is calculated by Darcy's law as shown in Figure
 267 3e-insert. No data of flow rate and pressure at the production well for episode #5 are collected,
 268 thus permeability estimated using this method is missing. It is observed that average permeability
 269 estimated by the physics-inspired model (equation 10) is only slightly higher than the other
 270 methods. The deterministic calculation of permeabilities for episodes #4 and #5, recovered from
 271 the injectivity ground truth (Step 1), are of the same magnitude as the average permeabilities
 272 estimated from the fracture permeability maps (Step 3) - indicating that fracture permeabilities
 273 may be directly constrained from MEQs and based on moment magnitude.

274 **Discussion**

We apply deterministic ground truthing (Step 1), hybrid ML methods (Step 2) and physics-inspired MEQ-permeability constraints (Step 3) to cross correlate permeability evolution in an unusually tightly constrained field experiment. Based on the ML methods, injectivity is predicted from the statistical features of the seismic events corresponding to a history of injectivity that the model has never seen. This is accomplished by training the ML model in early episodes of observation and predicting over the later (final two) episodes. Average permeability of the fractured reservoir is then calculated from injectivity, using a geometric correction of the steady state. This causality between injectivity/permeability and MEQs is then used to develop a physics-inspired model linking permeability evolution to MEQ magnitude. This relies on the parallel plate analog linking permeability to fracture aperture then dilation and then supplemented by laboratory observations scaling dilation to equivalent MEQ magnitude. These data are then scaled to link individual MEQs of known magnitude and location to define incremental changes in permeability in both space and in time. This pointwise distribution of permeability is then used to define a map of permeability structure and then averaged to recover the mean permeability evolution. This prediction closely matches the measured ground truth of ensemble permeability recovered from this unusually well-constrained injection experiment. The promising approach to quantify evolution of permeability as a function of observable MEQs provides a valuable tool in the characterization of reservoirs with the potential to visualize the evolution of fluid flow paths, evaluate reactive/heat-transfer surface area and capable of further constraint using other superimposed signals, say of tracer response. Challenges for the application of this method to larger scale include dealing with much noisier data, applying longer injection histories and operated under various conditions need to be collected as a database to build ML models and to evaluate and improve accuracy.

Domain knowledge plays an essential role in choosing the key statistical features that condition response for ML models. We select nine statistical features in the model to estimate the dependence of injectivity on (1) number of MEQs, (2) magnitude of MEQs, and (3) the distance of MEQs from injection well in the regression problem. The results illustrate that the distance to MEQs from the injection well is the dominant feature in predicting injectivity, followed by the number of MEQs and then the magnitude of MEQs. It should also be noted that the success of supervised ML approaches are heavily dependent on the training data used. Our results show that using the data solely from episode #3 to predict response in episodes #4 or #5 is much more

accurate than using all three episodes #1-#3. This is because flow rates following episode #3 were more than an order of magnitude larger than those used in prior stimulations (episode #1-2) with similar injection pressure, resulting in a factor of ten increase in injectivity. If we choose episodes #1-#3 instead of episode #3 alone, this will bias the training with ~78% (110 out of 140 minutes) of low injectivity data conforming to only ~33% (150 out of 450 seismic events) of MEQs - data irrelevant for the higher injectivities observed during the later episodes #4 and #5. Using the most recent and relevant data, alone, undoubtedly increases the accuracy in the prediction. Therefore, choosing a dataset with similar characteristics delivers better prediction than using a larger dataset including less relevant performance data.

Methods

Data analysis and machine learning methods.

(1) Data arrangement. In this project, well injection data is continuously recorded every second, and seismic catalogue is recorded for each seismic event, which is not continuous in time like well data. We start by creating files of well injection history and seismic catalogue with a common time base. We then scan both well data and seismic data using the same moving time windows, and compute statistic features of seismic data and average injectivity for each of these windows. We define 9 features (the statistic features of seismic data) for ML model including maximum of seismic magnitude (M_{wmax}), minimum of seismic magnitude (M_{wmin}), average seismic magnitude (M_{wmean}) within each window, maximum value (D_{max}), minimum value (D_{min}), and average value (D_{mean}) of distance from seismic events to injection well; estimated flow radius (R); the number of seismic events within each window (N), and the cumulative number of seismic events from the beginning to current window (N_{total}). Thus, each line of the database contains the variables and the corresponding average injectivity during the same time. The time series of variables is then labelled with the injectivity correspondingly. The size of time moving window is picked for a minute, and each time window moves forward for a second. We build variables and the corresponding labels, and then add the resulting list of labelled features to the database at each time increment. Finally, each line of dataset i is a list $\{x_i^1, x_i^2, x_i^3 \dots x_i^n, y_i\}$, with x_i^n the n th feature of the i th time window of seismic data, n is the total number of features, and y_i the average injectivity during the same time moving window.

(2) Train-validate-test split. We aim to use ML model to predict injectivity in episodes #4 and #5. The training data can be the dataset of one single episode or the combination of few episodes that occurred before the predicted episode. For example, to estimate injectivity in episode #4 (testing data), the training data can be episode #3 only, or episodes #1 to #3. The detail of training data and testing data is given in Table1. As the database has a time series features of the seismic data and corresponding average injectivity, the train–validate split must be of two continuous datasets in time, not a random split. Here we use the first 80% of the training data to construct the ML model and rest 20% of the training data for validating the model.

(3) Choose the features. The dataset is now composed of 9 features describing the seismic data versus time moving window. Our goal is to predict the injectivity based on the given features. The features can be any combinations of these 9 features, which results in 511 possibilities.

(4) Training and tuning the ML model. We build trail models using both the train and test dataset with default hyperparameters and evaluate the model performance by root mean square error (RMSE). Then hyperparameters are tuned on the train dataset to minimize the RMSE using cross-validation. We tune hyperparameters in the model, and all combinations of hyperparameters with specific values in a range are tested to guarantee the optimal solution (max depth [1,10], min child weight [1,10], eta [0.005,0.5], subsample [0,1], colsample by tree [0,1]. Five-fold cross-validation is applied for each combination of hyperparameters, where train dataset is split into 5 folds and iteratively keeps one of the folds for test purposes and returns a RMSE score. The optimum hyperparameters are chosen with the minimum RMSE. Then the model is checked by validate dataset to make sure this model is not overfitting. The final model is then created using train and validate dataset with updated hyperparameters.

(5) Make prediction. We can now make predictions using the best model (final model). The combination of features will be selected and then input into the best model to predict injectivity. The predicted data will be compared with ground truth of injection to assess the performance of the model.

(6) Feature importance. we can look for the best features identified by our model, to try to understand how the model reached its estimations. Here we apply feature importance type as ‘gain’, which implies the relative contribution of the corresponding feature to the model calculated by taking each feature’s contribution for each tree in the model. A higher value of this metric when compared to another feature implies it is more important for generating a prediction.

Linking seismic magnitude with fracture permeability using empirical equation. Here we develop a new method to link MEQs and fracture permeability change during hydraulic stimulation of a fractured reservoir. First, a simple configuration of hydro-shear failure is assumed to be generated by the microearthquakes (MEQs) during stimulation, with a shear displacement of u (m) uniformly applied to a square fracture plane of fracture length L (m). The fracture area A (m^2) is then simply defined as

$$A = L^2 \quad (1)$$

It has been demonstrated that fracture length can be estimated by aperture using an exponent relationship as (Ishibashi, 2016).

$$b = \alpha L^\beta \quad (2)$$

where b is the fracture aperture (m) and α and β are the coefficient that link the fracture aperture and length. Permeability k_f (m^2) can be estimated from the fracture aperture using the cubic law as

$$k_f = \frac{b^2}{12} \quad (3)$$

Based on the concept of the seismic moment M_0 (Scholz, 2005).

$$M_0 = AGu \quad (4)$$

Where G is shear modulus of the reservoir rock, usually of the order of 30 GPa (McGarr, 2014)

Seismic moment can be converted to an MEQ magnitude M_w using

$$\log M_0 = 1.5M_w + 9.1 \quad (5) \quad (\text{Hanks and Kanamori, 1979})$$

Seismic moment relates to fracture area can be expressed as

$$\log M_0 = 1.5 \log(A) + 6.09 \quad (6) \quad (\text{Leonard, 2010})$$

If we combine equation (1-3) and (5-6), we can successfully link fracture permeability with seismic magnitude.

$$k_f = \frac{\alpha^2}{12} 10^{\beta(0.5M_w+1)} \quad (7)$$

Assuming $\alpha_1 = \frac{\alpha^2}{12}$, and taking logs of both sides of equation (7) yields

$$\log(k_f) = \beta(0.5M_w + 1) + \log(\alpha_1) \quad (8)$$

To determine the coefficients α_1 and β for the reservoir injection here (EGS-Collab project) we use experimental data that concurrently measured aperture and shear displacement during slip on laboratory faults as pore pressure was incremented due to fluid injection (Fig. 1a of Li et al.,

2021). These are the same reservoir rocks as in the field experiment. Seismic moment is then calculated from measured reactivation displacement (equation (4)) and permeability is evaluated from measured aperture (equation (3)) – both independently. Substitution of multiple correlated values of permeability k_f and moment M_w into equation (8) then yields the coefficients α_1 and β . The results for change in fracture permeability k_f as a function of seismic moment M_w are shown in Figure 7 with regression defining the fitting parameters as,

$$\log(k_f) = 0.8972M_w - 7.57 \quad (9)$$

and therefore, from eqn (8) we obtain $\alpha_1 = 4.6 \times 10^{-10}$ and $\beta = 1.7584$.

Figure 8 (a). Log-log plot of fracture permeability versus seismic moment magnitude. Data from Li et al. (2021). (b). Fracture length distribution for all fractures reactivated by the MEQs. MEQs are then linked to change in fracture permeability during the hydraulic stimulation as

$$k = 4.6 \times 10^{-10} \times 10^{0.8972M_w + 1.7584} \quad (10)$$

References

- Bauer, Johanna F., et al. "A numerical sensitivity study of how permeability, porosity, geological structure, and hydraulic gradient control the lifetime of a geothermal reservoir." *Solid Earth* 10.6 (2019): 2115-2135.
- Bauer, Johanna F., et al. "Predictability of properties of a fractured geothermal reservoir: the opportunities and limitations of an outcrop analogue study." *Geothermal Energy* 5.1 (2017): 1-27.
- Cheng, Qinglu, Xiaonan Wang, and Ahmad Ghassemi. "Numerical simulation of reservoir stimulation with reference to the Newberry EGS." *Geothermics* 77 (2019): 327-343.
- Dobson, Patrick, et al. "An introduction to the EGS Collab Project." Geothermal Resources Council 41st Annual Meeting-Geothermal Energy: Power To Do More, GRC 2017. *Geothermal Resources Council*, 2017.
- Elith, Jane, John R. Leathwick, and Trevor Hastie. "A working guide to boosted regression trees." *Journal of Animal Ecology* 77.4 (2008): 802-813.
- Fang, Yi, Derek Elsworth, and Trenton T. Cladouhos. "Reservoir permeability mapping using microearthquake data." *Geothermics* 72 (2018): 83-100.
- Frash, L. P., J. P. Morris, and EGS Collab Team. "Stochastic prediction of multi-well fracture connectivity and application to EGS Collab experiment 2." *53rd US Rock Mechanics/Geomechanics Symposium. American Rock Mechanics Association*, 2019.
- Frash, Luke P., J. William Carey, and Nathan J. Welch. "EGS Collab Experiment 1 Geomechanical and Hydrological Properties by Triaxial Direct Shear." *44th Workshop on Geothermal Reservoir Engineering*. 2019.
- Fridleifsson, Ingvar B., et al. "The possible role and contribution of geothermal energy to the mitigation of climate change." *IPCC scoping meeting on renewable energy sources, proceedings, Luebeck, Germany*. Vol. 20. No. 25. Citeseer, 2008.
- Fu, Pengcheng, et al. "Predicting hydraulic fracture trajectory under the influence of a mine drift in EGS Collab experiment I." *PROCEEDINGS, 43rd Workshop on Geothermal Reservoir Engineering*. 2018.
- Fu, Pengcheng, et al. *Microseismic Signatures of Hydraulic Fracturing: A Preliminary Interpretation of Intermediate-Scale Data from the EGS Collab Experiment 1*. No. LLNL-CONF-767444. Lawrence Livermore National Lab.(LLNL), Livermore, CA (United States), 2019.
- Hanks, Thomas C., and Hiroo Kanamori. "A moment magnitude scale." *Journal of Geophysical Research: Solid Earth* 84.B5 (1979): 2348-2350.
- Ishibashi, Takuya, et al. "Linking microearthquakes to fracture permeability change: The role of surface roughness." *Geophysical Research Letters* 43.14 (2016): 7486-7493.
- Ishibashi, Takuya, et al. "Linking microearthquakes to fracture permeability change: The role of surface roughness." *Geophysical Research Letters* 43.14 (2016): 7486-7493.

- Kneafsey, Timothy J., et al. "An overview of the EGS Collab project: field validation of coupled process modeling of fracturing and fluid flow at the Sanford Underground Research Facility, Lead, SD." *43rd Workshop on Geothermal Reservoir Engineering, Stanford University*, paper SGP-TR-213. Preprint at, <https://pangea.stanford.edu/ERE/pdf/IGAstandard/SGW/2018/Kneafsey.pdf>. 2018.
- Kneafsey, Timothy J., et al. "EGS Collab project: Status and progress." *Proceedings 44th Workshop on Geothermal Reservoir Engineering, Stanford University*. 2019.
- Kushnir, Alexandra RL, Michael J. Heap, and Patrick Baud. "Assessing the role of fractures on the permeability of the Permo-Triassic sandstones at the Soultz-sous-Forêts (France) geothermal site." *Geothermics* 74 (2018): 181-189.
- Laubach, Stephen E., Jon E. Olson, and Michael R. Gross. "Mechanical and fracture stratigraphy." *AAPG bulletin* 93.11 (2009): 1413-1426.
- Li, Ziyang, Derek Elsworth, and Chaoyi Wang. "Constraining maximum event magnitude during injection-triggered seismicity." *Nature communications* 12.1 (2021): 1-9.
- Lu, Jianrong, and Ahmad Ghassemi. "Coupled Thermo–Hydro–Mechanical–Seismic Modeling of EGS Collab Experiment 1." *Energies* 14.2 (2021): 446.
- McGarr, Arthur. "Maximum magnitude earthquakes induced by fluid injection." *Journal of Geophysical Research: solid earth* 119.2 (2014): 1008-1019.
- Morris, Joseph P., et al. "Experimental design for hydrofracturing and fluid flow at the DOE EGS collab testbed." *52nd US Rock Mechanics/Geomechanics Symposium. American Rock Mechanics Association*, 2018.
- Neupane, Ghanashyam Hari, et al. EGS Collab Earth Modeling: Integrated 3D Model of the Testbed. No. INL/CON-19-53232-Rev002. Idaho National Lab.(INL), Idaho Falls, ID (United States), 2019.
- Schoenball, Martin, et al. "Creation of a Mixed-Mode Fracture Network at Mesoscale Through Hydraulic Fracturing and Shear Stimulation." *Journal of Geophysical Research: Solid Earth* 125.12 (2020): e2020JB019807.
- Scholz, Christopher H. *The mechanics of earthquakes and faulting*. Cambridge university press, 2019.
- Szilard Pafka, 2015. "Benchmarking Random Forest Implementations".
- Tezuka, Kazuhiko, and Kimio Watanabe. "Fracture network modeling of Hijiori hot dry rock reservoir by deterministic and stochastic crack network simulator (D/SC)." *Proc. World Geotherm. Cong* (2000): 3933-3938.
- Tezuka, Kazuhiko, Tetsuya Tamagawa, and Kimio Watanabe. "Numerical simulation of hydraulic shearing in fractured reservoir." *Proceedings World Geothermal Congress*. 2005.
- Ulrich, C., et al. "The distribution, orientation, and characteristics of natural fractures for Experiment 1 of the EGS Collab Project, Sanford Underground Research Facility." *52nd US Rock Mechanics/Geomechanics Symposium. American Rock Mechanics Association*, 2018.

- Warren, J. E., and P. J. Root. "The behavior of naturally fractured reservoirs." *Society of Petroleum Engineers Journal* 3.03 (1963): 245-255.
- White, Mark D., et al. "Numerical simulation applications in the design of EGS Collab Experiment 1." *PROCEEDINGS, 43rd Workshop on Geothermal Reservoir Engineering*. 2018.
- Williams, Colin F., et al. Assessment of moderate-and high-temperature geothermal resources of the United States. No. 2008-3082. *Geological Survey (US)*, 2008.

Acknowledgements This work is a partial result of support from the DOE EGS Collab Project: Stimulation Investigations for Geothermal Modeling, Analysis and Validation and from DOE award DE-EE0008763. This support is gratefully acknowledged.

Author contributions DE conceived the study, ZL performed experiments, data analyses, and prepared the manuscript and figures. CW contributed to the interpretation of all results. All authors discussed the results and implications and revision of the manuscript.

Data availability. Data are available based on personal request.

EGS Collab Team:

J. Ajo-Franklin, T. Baumgartner, K. Beckers, D. Blankenship, A. Bonneville, L. Boyd, S. Brown, J.A. Burghardt, C. Chai, A. Chakravarty, T. Chen, Y. Chen, B. Chi, K. Condon, P.J. Cook, D. Crandall, P.F. Dobson, T. Doe, C.A. Doughty, D. Elsworth, J. Feldman, Z. Feng, A. Foris, L.P. Frash, Z. Frone, P. Fu, K. Gao, A. Ghassemi, Y. Guglielmi, B. Haimson, A. Hawkins, J. Heise, C. Hopp, M. Horn, R.N. Horne, J. Horner, M. Hu, H. Huang, L. Huang, K.J. Im, M. Ingraham, E. Jafarov, R.S. Jayne, T.C. Johnson, S.E. Johnson, B. Johnston, S. Karra, K. Kim, D.K. King, T. Kneafsey, H. Knox, J. Knox, D. Kumar, K. Kutun, M. Lee, D. Li, J. Li, K. Li, Z. Li, M. Maceira, P. Mackey, N. Makedonska, C.J. Marone, E. Mattson, M.W. McClure, J. McLennan, T. McLing, C. Medler, R.J. Mellors, E. Metcalfe, J. Miskimins, J. Moore, C.E. Morency, J.P. Morris, T. Myers, S. Nakagawa, G. Neupane, G. Newman, A. Nieto, T. Paronish, R. Pawar, P. Petrov, B. Pietzyk, R. Podgorney, Y. Polsky, J. Pope, S. Porse, J.C. Primo, C. Reimers, B.Q. Roberts, M. Robertson, V. Rodriguez-Tribaldos, W. Roggenthen, J. Rutqvist, D. Rynders, M. Schoenball, P. Schwering, V. Sesetty, C.S. Sherman, A. Singh, M.M. Smith, H. Sone, E.L. Sonnenthal, F.A. Soom, D.P. Sprinkle, S. Sprinkle, C.E. Strickland, J. Su, D. Templeton, J.N. Thomle, C. Ulrich, N. Uzunlar, A. Vachaparampil, C.A. Valladao, W. Vandermeer, G. Vandine, D. Vardiman, V.R. Vermeul, J.L. Wagoner, H.F. Wang, J. Weers, N. Welch, J. White, M.D. White, P. Winterfeld, T. Wood, S. Workman, H. Wu, Y.S. Wu, E.C. Yildirim, Y. Zhang, Y.Q. Zhang, Q. Zhou, M.D. Zoback

Team members by organization

DOE

L. Boyd, Z. Frone, E. Metcalfe, A. Nieto, S. Porse, W. Vandermeer

INL

R. Podgorney, G. Neupane,

LBNL

J. Ajo-Franklin, P.J. Cook, P.F. Dobson, C.A. Doughty, Y. Guglielmi, M. Hu, R.S. Jayne, K. Kim, T. Kneafsey, S. Nakagawa, G. Newman, P. Petrov, M. Robertson, J. Rutqvist, M. Schoenball, E.L. Sonnenthal, F.A. Soom, C. Ulrich, C.A. Valladao, T. Wood, Y.Q. Zhang, Q. Zhou

LANL

L. Huang, Y. Chen, B. Chi, Z. Feng, L.P. Frash, K. Gao, E. Jafarov, S. Karra, N. Makedonska, W. Pan, R. Pawar, N. Welch

LLNL

P. Fu, R.J. Mellors, J.P. Morris, M.M. Smith, D. Templeton, H. Wu

NETL – J. Moore, S. Brown, D. Crandall, P. Mackey, T. Paronish, S. Workman

NREL - B. Johnston, K. Beckers, J. Weers

ORNL - Y. Polsky, Monica Maceira, Chengping Chai

PNNL - A. Bonneville, J.A. Burghardt, J. Horner, T.C. Johnson, H. Knox, J. Knox, B.Q. Roberts, P. Sprinkle, C.E. Strickland, J.N. Thomle, V.R. Vermeul, M.D. White

SNL

D. Blankenship, M. Ingraham, J. Pope, P. Schwering, A. Foris, DK King, J. Feldman, M. Lee, J. Su

SURF

T. Baumgartner, J. Heise, M. Horn, B. Pietzyk, D. Rynders, G. Vandine, D. Vardiman,

Subs:

- Thomas Doe, TDoeGeo Rock Fracture Consulting; Golder Associates Inc;
- J. McLennan, University of Utah
- CSM – Y.S. Wu, J. Miskimins, P. Winterfeld, K. Kutun
- Stanford – M.D. Zoback - A. Singh,
- Stanford – R.N. Horne, K. Li, A. Hawkins, Y. Zhang
- Mattson Hydrology LLC: E. Mattson
- Penn State – D. Elsworth, K.J. Im, Z. Li, C.J. Marone, E.C. Yildirim

- Rice University – Jonathan Ajo-Franklin
- The University of Oklahoma - A. Ghassemi, Dharmendra Kumar, Varahanaresh Sesetty, Alex Vachaparampil
- University of Wisconsin -Madison– H.F. Wang, Hiroki Sone, Kate Condon, and Bezalel Haimson.
- South Dakota School of Mines and Technology – W. Roggenthen, C. Medler, N. Uzunlar, Carson Reimers
- ResFrac - M.W. McClure