Deriving three-dimensional properties of fracture networks from two-dimensional observations in rocks approaching failure under triaxial compression: Implications for fluid flow

Jessica McBeck[1] and François Renard[2]

[1]University of Oslo
[2]The Njord Centre, Departments of Geosciences and Physics

November 24, 2022

Abstract

Approximating the three-dimensional structure of a fault network at depth in the subsurface is key for robust estimates of fluid flow. However, only observations of two-dimensional outcrops are often available. To shed light on the relationship between two- and three-dimensional measurements of fracture networks, we examine data from a unique set of eleven X-ray synchrotron triaxial compression experiments that reveal the evolving three-dimensional fracture network throughout loading. Using machine learning, we derive relationships between the two- and three-dimensional measurements of three properties that control fluid flow: the porosity, and volume and tortuosity of the largest fracture at a particular differential stress step. The models predict the porosity and volume of the largest fracture with R^2 scores of >0.99, but predict the tortuosity with maximum R^2 scores of 0.68. To test the assumption that different rock types may require different equations between the two- and three-dimensional properties, we develop models for both individual rock types (granite, monzonite, marble, sandstone) and all of the experiments. Models developed using all of the experiments perform better than models developed for individual rock types, suggesting fundamental similarities between fracture networks in rocks often analyzed separately. Models developed with several parallel two-dimensional observations perform similarly to models developed with several perpendicular two-dimensional observations. When the models are developed with statistics of the two-dimensional observations, the models primarily depend on the mean and median when they predict the porosity, and minimum when they predict the volume and tortuosity.

Deriving three-dimensional properties of fracture networks from two-dimensional observations in rocks approaching failure under triaxial compression: Implications for fluid flow

J. McBeck[1] and F. Renard[1,2]

[1]The Njord Centre, Departments of Geosciences and Physics, University of Oslo, Oslo, Norway.

[2]ISTerre, Univ. Grenoble Alpes, Grenoble INP, Univ. Savoie Mont Blanc, CNRS, IRD, Univ. Gustave Eiffel, 38000, Grenoble, France.

Corresponding author: Jess McBeck (j.a.mcbeck@geo.uio.no)

Key Points:

- Machine learning predicts three-dimensional fracture properties from two-dimensional measurements.
- Model performance does not depend on the orientation of the two-dimensional observations relative to the maximum compression direction.
- Models developed with several rock types perform better than models developed from individual rock types.

Abstract

Approximating the three-dimensional structure of a fault network at depth in the subsurface is key for robust estimates of fluid flow. However, only observations of two-dimensional outcrops are often available. To shed light on the relationship between two- and three-dimensional measurements of fracture networks, we examine data from a unique set of eleven X-ray synchrotron triaxial compression experiments that reveal the evolving three-dimensional fracture network throughout loading. Using machine learning, we derive relationships between the two- and three-dimensional measurements of three properties that control fluid flow: the porosity, and volume and tortuosity of the largest fracture at a particular differential stress step. The models predict the porosity and volume of the largest fracture with R^2 scores of >0.99, but predict the tortuosity with maximum R^2 scores of 0.68. To test the assumption that different rock types may require different equations between the two- and three-dimensional properties, we develop models for both individual rock types (granite, monzonite, marble, sandstone) and all of the experiments. Models developed using all of the experiments perform better than models developed for individual rock types, suggesting fundamental similarities between fracture networks in rocks often analyzed separately. Models developed with several parallel two-dimensional observations perform similarly to models developed with several perpendicular two-dimensional observations. When the models are developed with statistics of the two-dimensional observations, the models primarily depend on the mean and median when they predict the porosity, and minimum when they predict the volume and tortuosity.

Plain Language Summary

A fundamental problem in geoscience is extrapolating two-dimensional observations to three-dimensional systems. For example, we may directly observe the length and width of a fracture where it intersects the Earth's surface. Occasionally we may be able to find multiple two-dimensional observations of the same fracture network. Attempts to simulate three-dimensional fluid flow using two-dimensional systems have provided incorrect estimates of the true permeability. Thus, to estimate the dynamics of fluid flow through a fault network we must interpolate and/or extrapolate from two-dimensional observations to approximations of three-dimensional systems. Here, we use machine learning to estimate three-dimensional characteristics of fracture networks that control fluid flow from two-dimensional observations. In situ X-ray tomography triaxial compression experiments provide unparalleled access to both the three- and two-dimensional observations as the rock experiences increasing differential stress, and develops more fractures. The work provides insight into the fundamental similarities between fracture network development in different rock types, into the feasibility of developing equations between two- and three-dimensional properties that control fluid flow, and into the statistics of the two-dimensional property that are most beneficial to predicting the three-dimensional properties.

Keywords: fracture network; fluid flow; triaxial compression; granite; sandstone; machine learning

1 Introduction

Estimating characteristics of the three-dimensional structure of a tectonic system or fault network from sparse, two-dimensional data is a key aspect of many field analyses (e.g., Moore et al., 1990; Gueting et al., 2018), scaled physical experiments (e.g., Sassi et al., 1993; Dominguez et al., 2000; Tong et al., 2014), and laboratory deformation experiments (e.g., Bobet & Einstein, 1998; Cao et al., 2020). Constraining the geometry of fracture networks in three-dimensions in the subsurface is critical to robust assessments of seismic hazard (e.g., Kozłowska et al., 2018), fluid flow (e.g., Auradou et al., 2005), and thus potential for CO_2 storage (e.g., Iding & Ringrose, 2010; Luhmann et al., 2017). However, the process of reconstructing three-dimensional fracture geometries from two-dimensional data is often qualitative, and rarely formalized with benchmarks to assess the accuracy of such reconstructions (e.g., Caumon et al., 2009; Lei et al., 2017).

To reconstruct three-dimensional data from two-dimensional measurements, previous analyses have relied on serial sectioning (e.g., Wei et al., 2019) and sub-sampling two-dimensional slices (Karimpouli & Tahmasebi, 2016). Other analyses have used analytical and statistical methods to estimate three-dimensional values (e.g., Roberts et al., 1997; Yeong & Torquato, 1998; Manwart & Hilfer, 1999; Keehm et al., 2004; Lei et al., 2015; Saxena & Mavko, 2016). Such statistical methods measure properties, such as porosity, in two-dimensional images, and then generate three-dimensional systems so that the statistics of the given property (e.g., porosity) match the measured three-dimensional statistics. While these methods can provide reasonable estimates of some properties, they struggle to accurately capture the connectivity of pore and fracture networks in three-dimensions (Hazlett, 1997; Manwart et al., 2000; Øren & Bakke, 2002). Process-based reconstruction provides an alternative method of reconstruction that can more closely approximate the connectivity of the pore network in sandstone than statistical reconstruction (e.g., Bakke & Øren, 1997; Øren & Bakke, 2002). However, this method simulates the packing of grains and subsequent processes, such as compaction and diagenesis, that produce sandstones, for example, and thus cannot be applied to rocks that do not form with this process, such as granite (e.g., Dong & Blunt, 2009).

Due to the importance of fracture networks on fluid flow, recent studies have used machine learning to predict the permeability of synthetic porous media (Tian et al., 2020; Santos et al., 2020), and of natural rock cores, including sandstone, carbonate, and limestone (Sudakov et al., 2019; Kamrava et al., 2020; Elmorsy et al., 2022). This work has produced models that can predict the permeability of granular, porous rocks with strong positive correlations between the measured and predicted permeability. However, this work did not attempt to predict the permeability of rocks with more heterogeneous fracture networks, such as granite. It may be more difficult to estimate the permeability of rocks that contain more heterogeneous fracture networks, with a wider range of fracture lengths and spacing between fractures, than more homogeneous rocks. Moreover, previous work focused on nominally intact rocks, rather than rocks that had undergone some differential stress loading. In the field, the volumes of crust with the largest permeabilities may tend to be those that experienced some inelastic deformation, such as the highly fractured damage zone adjacent to the principal slip zone of faults (e.g., Mitchell & Faulkner, 2012). Consequently, estimating the permeability of volumes of the crust with fracture networks that developed due to increasing differential stress is critical for robust predictions of fluid flow.

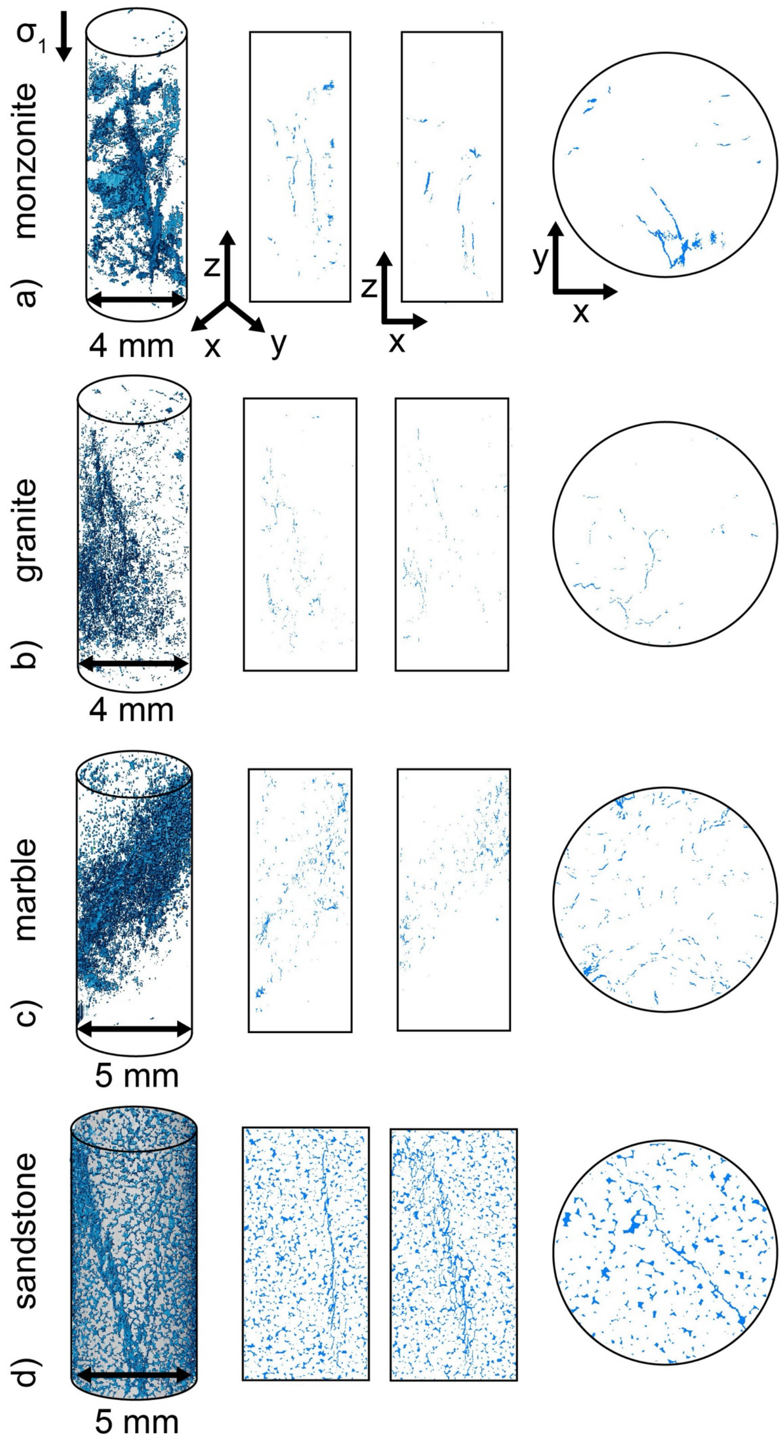

Figure 1. Example images of the fracture and pore networks in three-dimensional volumes (first column), two-dimensional vertical slices (second and third columns), and two-dimensional horizontal slices (fourth column) in four experiments: a) monzonite #5, b) granite #4, c) marble #2, and d) sandstone #2. These scans were acquired immediately preceding macroscopic failure, after the initially intact rock was loaded to failure. The key goal of the present study is to predict the three-dimensional properties of fracture networks from two-dimensional observations.

In order to shed light on the relationships between two- and three-dimensional measurements of the fracture networks within homogeneous and heterogeneous rocks subjected to increasing differential stress, we examine data from eleven synchrotron X-ray microtomography triaxial compression experiments performed on Fontainebleau sandstone, Westerly granite, quartz monzonite, and Carrara marble (**Figure** 1). This data set is perhaps the most comprehensive accumulation of two- and three-dimensional observations of fracture networks that developed during triaxial compression deformation experiments with a wide range of rock types. In these experiments, we systematically increase the differential stress on the rock cores until the rocks fail, and acquire X-ray tomograms (i.e., scans) after each increase of differential stress (**Figure** 2). We derive relationships between the two- and three-dimensional measurements of three properties that control the fluid flow: the porosity, and the volume and tortuosity of the largest fracture in the network at a particular differential stress step. We build these relationships using two machine learning algorithms: gradient boosting and linear regression. In order to constrain the feasibility of deriving an equation between these two- and three-dimensional properties, we compare the ability of the machine learning models to predict their values. We systematically examine how much data (e.g., number of two-dimensional slices) is required to make accurate estimates of the three-dimensional property, and whether the orientation of the two-dimensional slices controls the model success. Comparing the performance of the models developed with slices of various orientations sheds light on the most appropriate orientation of thin-sections in natural rock cores with varying mechanical structures, including low porosity crystalline rocks and porous granular rocks, and the best type of two-dimensional observations to gather in the field. We develop models for both individual rock types (granite, monzonite, marble, and sandstone) and all of the experiments combined in order to test the assumption that different rock types may require different equations between the two- and three-dimensional property. We then develop models using the statistics of sets of two-dimensional observations in order to identify the statistics that may be the most useful when predicting the three-dimensional property. This work thus provides insight into 1) the predictability of different fracture network properties, 2) the amount and type of data required for successful estimates, and 3) the similarities of the relationships between two- and three-dimensional properties in different rock types.

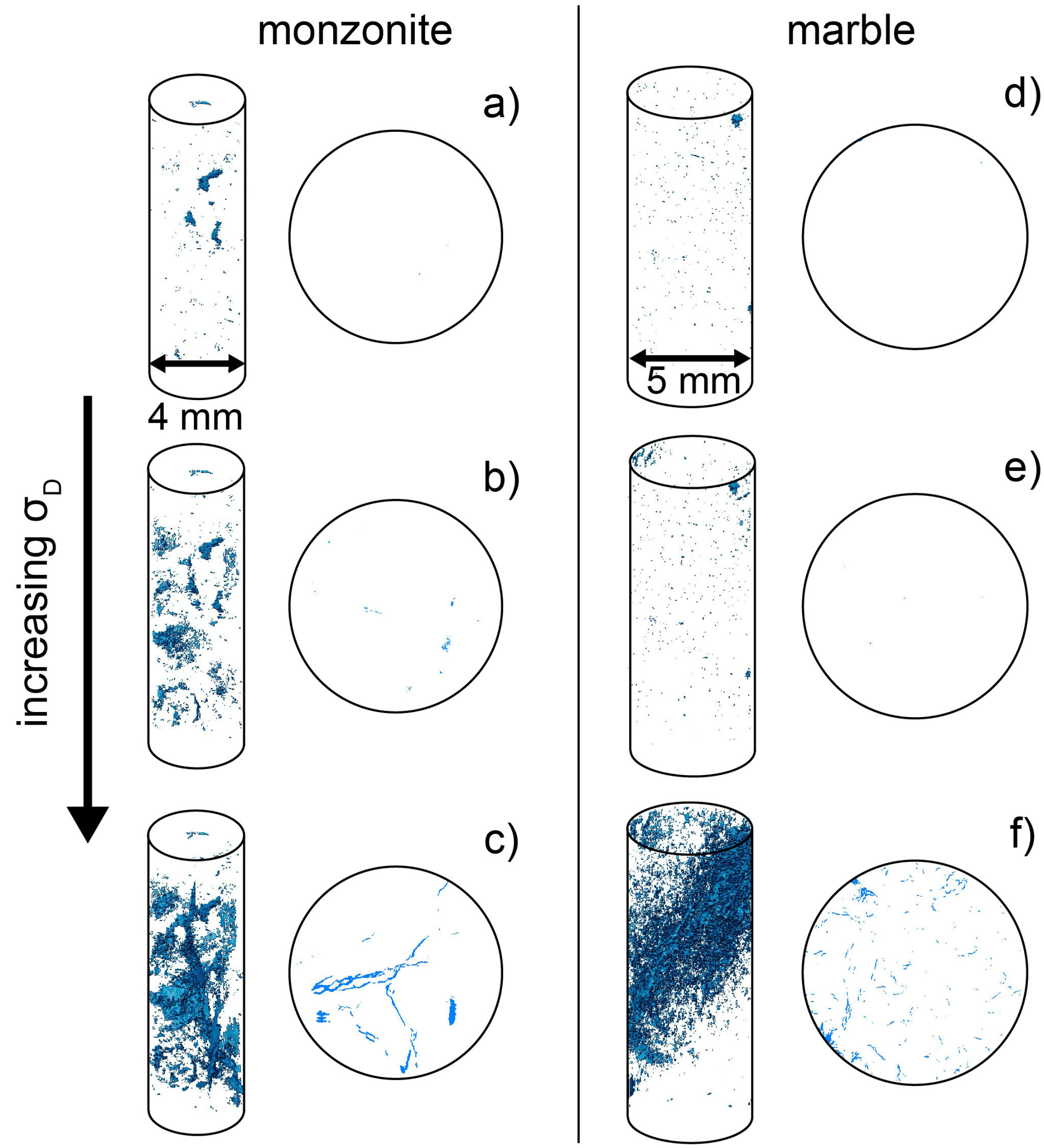

Figure 2. Example images of the three-dimensional and two-dimensional fracture networks from early in loading until immediately preceding macroscopic failure in two experiments: monzonite #5 (a-c), and marble #2 (d-f). The differential stress acting on the rocks, σ_D, increases from the top to the bottom of the figure.

2 Methods

2.1 Experimental conditions

We performed eleven triaxial compression experiments at beamline ID19 at the European Synchrotron and Radiation Facility, Grenoble, France. In these experiments, we insert one 10 mm tall and 4-5 mm diameter cylinder rock core in the Hades triaxial compression apparatus (Renard et al., 2016) installed on the beamline. The rock cylinders have different diameters so that the cores fail before the applied axial stress reaches the limit of the Hades apparatus (200 MPa for 5 mm diameter samples, and 312 MPa for 4 mm diameter samples). We then impose a confining stress (5-35 MPa) using pressurized oil against the jacket surrounding each core sample (**Table** 1), and increase the axial stress in steps of 0.5-5 MPa, with smaller steps closer to failure, until the rock fails in a sudden stress drop (**Figure** S1), at ambient temperature conditions. After each increase in axial stress, we acquire an X-ray scan within 1.5 minutes while the rock is under load inside the Hades apparatus. Thus, the total number of X-ray scans acquired in an experiment depends on the chosen set of axial stress steps, and the stress conditions at which a rock fails, producing 38-136 scans for a given experiment (**Table** 1).

We deformed four rock types: Westerly granite, quartz monzonite, Carrara marble, and Fontainebleau sandstone. We analyze the relationship between the two- and three-dimensional properties of the fracture networks of these rocks because they represent endmembers of fracture network properties. Westerly granite and monzonite are low-porosity crystalline rocks dominated by interlocking quartz, feldspar, and mica crystals. The initial porosity of the monzonite and granite is lower than 1%. Carrara marble is a low porosity metamorphic rock that consists of calcite grains. Carrara marble has an initial porosity of about 0.2%, and grain sizes from 100-200 µm (e.g., Rutter, 1972; Malaga-Starzec et al., 2002). Fontainebleau sandstone is comprised of cemented quartz grains. These Fontainebleau sandstone cores have a mean grain size of 250 µm and higher initial porosity than the marble: 5.5-7% (measured using the X-ray tomography images), and 6±1% (measured using imbibition with water) (Renard et al., 2018).

Experiment	Core diameter (mm)	Confining stress (MPa)	Number of scans	Experiment abbreviation
Westerly granite #1	4	5	38	WG01
Westerly granite #2	4	5	27	WG02
Westerly granite #4	4	10	53	WG04
monzonite #3	4	20	61	MONZ03
monzonite #4	4	35	62	MONZ04
monzonite #5	4	25	76	MONZ05
Carrara marble #1	5	20	39	M8_1
Carrara marble #2	5	25	44	M8_2
Fontainebleau sandstone #1	5	20	136	FBL01
Fontainebleau sandstone #2	5	10	43	FBL02
Fontainebleau sandstone #3	5	10	51	FBL03

Table 1. Experimental conditions of the eleven experiments. The number of scans is the total number of scans acquired throughout each complete experiment.

2.2. Data extraction

This analysis uses properties of the fracture networks identified in the X-ray tomograms. Following the experiment, we reconstruct the acquired radiographs into three-dimensional volumes that are 1600x1600x1600 voxels. One side length of a voxel is 6.5 µm. During reconstruction, we apply several corrections to remove noise, such as ring artefacts. We then remove remaining noise in the three-dimensional data using the image analysis software Avizo3D™, including the application of a non-local means filter (Buades et al., 2005). Using these three-dimensional volumes, we segment the solid rock from the pores and fractures. We use an algorithm similar to Otsu's thresholding technique to identify a global threshold between the solid material and the fractures and pores (McBeck et al., 2021). This thresholding technique is robust to noise (McBeck et al., 2021) and produces segmented scans with porosity similar to measured values (Renard et al., 2018).

Because we aim to predict three-dimensional properties from two-dimensional measurements, we only consider a subset of the data so that the rounded edges of the rock cylinders do not influence the two-dimensional measurements. We only extract data within a rectangular prism at the center of the core, shown by the black rectangles in **Figure** 3. The size of the base of the prism depends on the width of the rock core. We set the positions of the sides of the base to $\frac{w}{2}-\frac{w}{4}$ to $\frac{w}{2}+\frac{w}{4}$, where w is the width of the core.

We test the influence of extracting the two-dimensional data along different orientations in order to constrain the best method of extracting cross sections of natural rock cores. We test four different methods of extracting two-dimensional slices: 1) acquiring horizontal slices along the vertical axis (z-axis and parallel to the maximum compression direction) (**Figure** 3a), 2) acquiring vertical slices along one of the horizontal axes (**Figure** 3b), 3) acquiring orthogonal vertical slices along both horizontal axes (**Figure** 3c), and 4) acquiring horizontal and vertical slices. Consequently, the area of the rock core captured in each method of slice extraction differs. We chose to compare slices of different areas in order to mirror the method of cutting natural rock cores.

We also test the influence of the amount of data provided to the models on the predictability of the three-dimensional properties. We vary the number of slices provided to the models from one to twenty slices, as well as all of the slices within the black rectangles shown in **Figure 3**. For the results of models trained on one slice, the methods of extraction that use multiple planes use one slice along each orientation, or two in total. When we take one slice, we use the slice within the center of the core, in either the horizontal or the vertical direction. When we take multiple slices, the slices are equally spaced throughout the rectangular prism (**Figure** 3).

For each slice, and corresponding rectangular prism, at each differential stress step of each experiment, we calculate three properties. We focus on properties that control fluid flow within fracture networks: the porosity, the volume/area of the largest fracture, and the geometric tortuosity of the largest fracture. Note, we identify the largest fracture in the three-dimensional data, and the largest fracture in each slice of the two-dimensional data separately. Consequently, the largest fracture identified in the two- and three-dimensional data at a given stress step of a given experiment may not be the same fracture. The geometric tortuosity is the ratio of the length of the true path between end points to the linear distance between end points (**Figure** S2). Here, we identify the end points of fractures as the locations of the fractures at the maximum and minimum *z*-coordinates. Thus, our tortuosity measurements reflect the path that a fluid must travel parallel to the maximum compression direction, and the long axis of the rock core. Because it is non-trivial to calculate tortuosity, we benefit from the Matlab function *bwdistgeodesic*, which calculates the distance of the true path between two points in two- and three-dimensions.

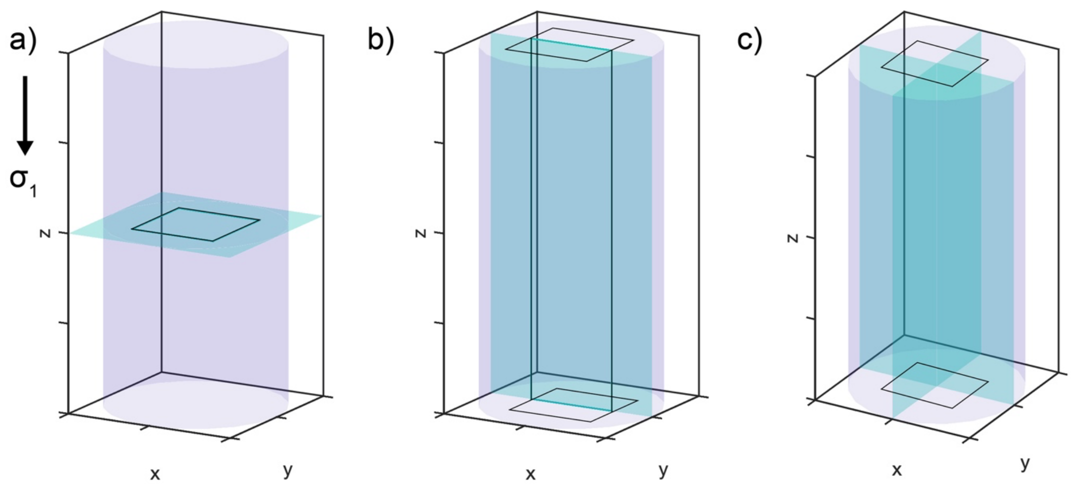

Figure 3. Sketch of the method of extracting the two-dimensional slices: horizontal slices within the *x-y* plane (a), vertical slices within the *z-x* plane (b), and orthogonal vertical slices, within the *z-x* plane and *z-y* plane (c). The *z*-axis is parallel to the maximum compression direction and vertical. We select the slices within a rectangular region (black) so that the cylindrical boundary of the rock core does not influence the calculated properties. When we take multiple slices, the slices are equally spaced throughout the core.

While the porosity considers all of the fractures in a particular slice or volume, the other two properties only consider the largest (most volumetric) fracture in the data. We focus on only the largest fracture for these properties, rather than all the fractures, because the volume and tortuosity of the largest fracture controls fluid flow. Following these calculations, we attain three datasets for each experiment corresponding to the three fracture network properties. Each of these datasets includes one three-dimensional measurement at each differential stress step of the experiment, and several hundred two-dimensional measurements: all of the horizontal slices along the vertical axis, and all of the vertical slices along both horizontal axes. Consequently, the data is arranged as a table where each row represents one differential stress step of a particular experiment, and the columns include the three-dimensional measurement, and all of the two-dimensional measurements.

2.3. Machine learning analysis

We design the machine learning models to predict three-dimensional properties from two-dimensional measurements of the porosity, area of the largest fracture, and tortuosity of the largest fracture. We test the influence of combining different rock types on the predictability of the three-dimensional values. To combine data from multiple experiments, we append the data sets described in the previous section. For example, one column in the data table contains the two-dimensional value of the property at the horizontal slice at the position 600 voxels above the base of the rock core throughout all of the scans acquired in all of the experiments included in the data. Due to the different number of scans acquired in each experiment, the dataset of each experiment contains different numbers of rows, or samples (**Figure** S3). To account for these varying number of scans, and ensure that the properties of one experiment do not dominate the other experiments, we extract the same numbers of samples of each experiment for the models that include combinations of experiments. Thus, for the models that use data from all of the experiments, we only take the number of samples from each experiment that match that lowest number of samples (the granite #1 experiment) (**Table** 1, **Figure** S3). Because the fracture networks grow with increasing loading, we select data from the end of the experiment.

We divide the data of the combined experiments into training and testing data sets such that none of the data in the training set includes data in the testing set. We divide the data by rows, so the training and testing datasets are unique in time (differential stress steps) for each experiment. Because the performance of the models varies depending on how we split the training and testing data sets, we report the performance of ten models that only differ in how the training and testing data sets are randomly split. We use 30% of the data for testing, and 70% of the data for training. Dividing the data by continuous blocks of time into training and testing datasets does not produce lower model performance than dividing the data randomly by time.

We use both gradient boosting models (XGBoost) and linear regression models. We chose XGBoost regression models because of its efficiency and accuracy (e.g., Friedman, 2001; Bühlmann & Yu, 2003; Chen & Guestrin, 2016). We scale the features using the RobustScaler of SciKit learn, which scales the data using the 25th and 75th quantile of the data (Pedregosa et al., 2011). We perform a grid search over the hyperparameter-space to find the best set of hyperparameters (Lundberg & Lee, 2017).

We also apply a simpler algorithm, linear regression, because of the interpretability of the parameters of the trained models. Training linear regression models produces a set of coefficients with one coefficient for each feature, or two-dimensional slice in this analysis, so that the three-dimensional property is estimated as a linear combination of the features weighted by each coefficient. From these coefficients, we aim to derive a function between the two- and three-dimensional properties. For the models that predict the fracture volume and porosity, we force the y-intercept of the regression model to intersect the origin because we suspect that two-dimensional measurements of volume or porosity that are near zero should produce three-dimensional values of zero. For these models, we do not scale the features so that the coefficients of the linear regression models are more interpretable.

We then develop gradient boosting models using the statistics of sets of two-dimensional observations. In particular, we subdivide each scan into sets of 20 vertical two-dimensional slices, and then calculate a range of statistics on the two-dimensional values measured in each slice, thereby producing a set of statistics for each group of 20 vertical slices of each scan. These statistics include: the mean, standard deviation, coefficient of variation, skewness, kurtosis, minimum, maximum, and the 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th} percentile of the dataset. Changing the number of slices within each group from 10 to 50 does not influence the key results. We only focus on the results of models developed with sets of vertical slices because our analyses indicate that the slice orientation does not systematically control the model performance, as described in the Results section. Similar to the other analyses, we vary the amount of data provided to the models from 10% to 90%. We randomly select ten different portions of the data with the given percentage and then develop ten different models with these varying parts of the data.

Because in this analysis we aim to determine which statistic of the two-dimensional observations provides the most useful information, we examine the impact of each feature (e.g., statistic) on the predictions of the model using a widely used metric: Shapely Additive Explanations (SHAP) (e.g., Lundberg & Lee, 2017). We compare the mean absolute value of the SHAP (mean |SHAP|) of each feature across all of the samples, and so focus on the overall influence of that feature on the model prediction. Similar to the other machine learning analyses, we divide the data into training and testing datasets by time such that none of the scans that occur in the training dataset occur in the testing dataset. Because we develop ten different models for each model of varying amounts of data, we calculate a normalized importance of the mean |SHAP| value, s, as $s/\max(s)$ for each model, $\hat{s}$. We weight this normalized importance by the R^2 score of the model and then find the mean of these values across all the models, $s_w = \sum(R^2\hat{s})/n$, where n is the number of models, so that more accurate models (with higher R^2) will have a greater influence on the results than less accurate models. The distribution of s_w thus indicates the relative importance of each feature on the model predictions across several models.

3 Results

3.1. Predicting three-dimensional properties with gradient boosting models

First, we compare the performance of the gradient boosting models (XGBoost) developed for all of the experiments combined together. To assess model performance, we compare the R^2 scores of the models, which represent the correlation coefficients between the observed and predicted three-dimensional values (**Figure** 4). High positive R^2 scores (0.8-1.0) indicate strong correlations between the observed and predicted values, while scores 0.4-0.7 indicate moderate correlations, and scores <0.4 indicate weak or non-existent correlations. To identify the best approach to rock core thin-section extraction and field analyses of fracture networks, we compare the influence of extracting slices from different orientations relative to the maximum compression direction, including horizontal and vertical slices, orthogonal vertical slices, and both horizontal and vertical slices (e.g., **Figure** 3). To assess the influence of the number of two-dimensional observations on the predictions, we also compare the performance of models developed with varying amounts of data, from one to twenty slices in increments of two slices, and also all of the slices. **Figure** 4 shows the mean ± standard deviation of the R^2 scores of the ten models developed for each fracture network property, and number and type of data, which only differ in how the training and testing are split.

The models that predict the porosity and volume perform better than the models that predict the tortuosity (**Figure** 4a-d). For example, the mean R^2 score of the models developed with one horizontal and one vertical slice are: 1) 0.99 (porosity), 2) 0.94 (volume), and 3) 0.28 (tortuosity). Even when the models have access to all of the slices, they can only predict the tortuosity with a mean R^2 score of 0.68, compared to the R^2 scores near 1.0 for the other properties (**Figure** 4, **Figure** 5). This trend holds regardless of the orientation of the extracted slices. This result suggests that it may be more difficult to derive a function between the two- and three-dimensional tortuosity compared to the other properties.

Increasing the amount of data generally increases the performance of the models, as expected (**Figure** 4, **Figure** 5). However, the models that predict the porosity using all of the rock types perform exceptionally well using only one slice (**Figure** 4). Similarly, the models that predict the volume of the largest fracture perform very well using only about four slices, regardless of the method of slice extraction. Consequently, the difference in the model performance when using 20 slices or one slice (porosity) to four slices (volume) is minor (<0.005 of the R^2 score). The influence of the amount of data on the model performance is most significant for the property that the models struggle to predict: the tortuosity. Thus, extracting many two-dimensional slices of a system is most beneficial when estimating the tortuosity and related properties, such as the permeability. In contrast, when estimating the porosity or volume of the largest fracture, only one to four two-dimensional measurements may be required for reasonably accurate estimates.

In contrast to expectations, the different methods of slice extraction produce models that perform similarly to each other when they use more than one slice of the data (**Figure** 4, **Figure** 5a-c). We expected that vertical slices (parallel to the maximum compression direction) may produce more accurate models than horizontal slices because they sample a larger area, and therefore produce stronger positive correlations between the two- and three-dimensional properties (**Text** S1, **Figure** S4, **Figure** S5, **Figure** S6). In agreement with these expectations, when the models use one slice, horizontal slices produce less successful models than vertical slices for the models that predict the porosity and volume. Both the horizontal and vertical slices produce poorly performing models with 0.2-0.3 R^2 scores when the models predict the tortuosity (**Figure** 5c). We expected that models developed with two orthogonal vertical slices or both vertical and horizontal slices may perform better than models developed with only horizontal or vertical slices because they sample more area, and at perpendicular orientations. In contrast to this idea, the models that use orthogonal vertical slices or both horizontal and vertical slices do not perform significantly better than the other models. The comparable performance of the models developed with more than one slice at perpendicular orientations to each other and series of parallel slices indicates that it may not be critical to obtain field observations and thin-sections of rock cores at perpendicular orientations in order to make robust estimates of the fracture network geometry, and thus fluid flow. Instead, several parallel observations may be sufficient for accurate estimates.

Next, we compare the performance of models developed with all of the rock types to models developed for individual rock types. We expect that models developed for specific rock types may perform better than models developed for all of the experiments because different relationships may exist between the two- and three-dimensional data in different rock types. For example, the equation that relates the two-dimensional porosity to the three-dimensional porosity may differ between the sandstone and granite because sandstone hosts many quasi-spherical pores, whereas the granite contains fractures and pores with shapes closer to cigars, rather than spheres (e.g., Renard et al., 2018). We focus on the models developed using both horizontal and vertical slices here (**Figure** 5d-f).

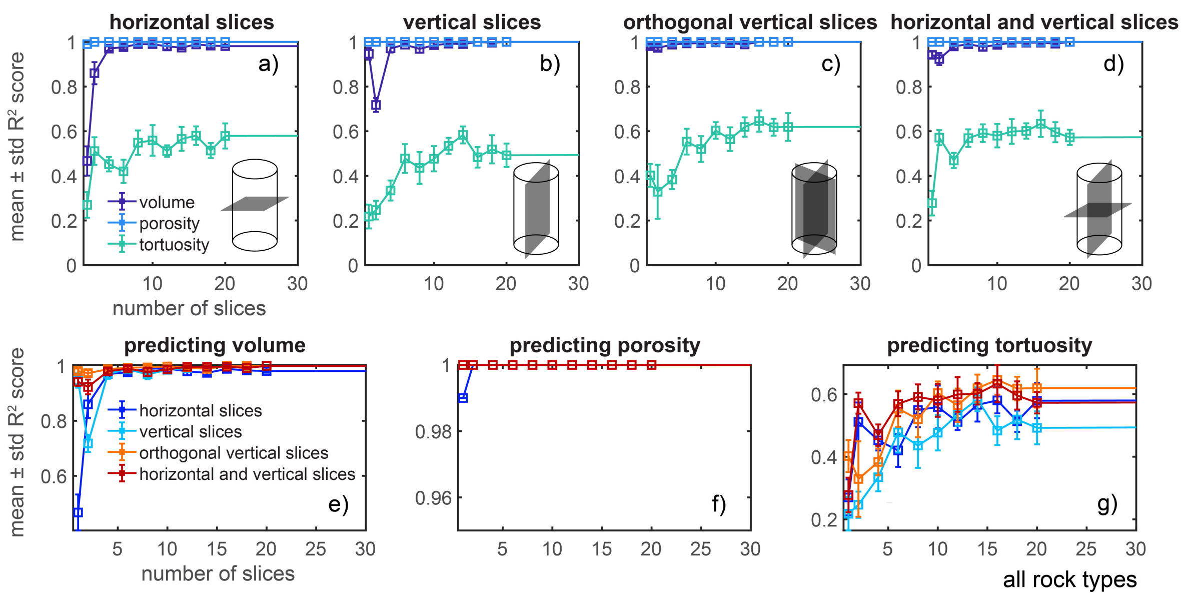

Figure 4. Performance of the gradient boosting models developed for all of the rock types using horizontal slices (a), vertical slices (b), orthogonal vertical slices (c), and both horizontal and vertical slices (d) to predict the volume (e), porosity (f) and tortuosity (g). For the models developed with the orthogonal vertical slices, and the horizontal and vertical slices, the score reported for one slice indicates the score of models developed with either one slice along both orthogonal horizontal axes, or one slice along the vertical axis and a horizontal axis. The models struggle to predict the tortuosity, but predict the volume and porosity with strong correlations between the predicted and observed values. The orientation of the extracted slices does not systematically influence the model performance.

Consistent with the trend of the correlation coefficients (**Text** S1, **Figure** S6), the models developed for all of the rock types do not perform worse than the models developed for the individual rock types (**Figure** 5). Instead, some of the models developed for individual rock types perform worse than the models developed with all of the data (**Figure** 5d, e). The models that predict the tortuosity differ somewhat from this trend. The performance of the models that predict the tortuosity using one slice of the data of all the rock types perform worse than models developed with only the granite data, and models developed with only the sandstone data.

However, the models developed with all of the rock types perform better than models developed with only the monzonite data, and models developed with only the marble data. Consequently, the models developed using all of the rock types perform similarly successfully as the models developed for the individual rock types when they predict the volume and porosity, and perform better than half of the rock types when they predict the tortuosity. This result contradicts the expectation that different relationships link the two- and three-dimensional properties in different rock types, such as sandstone and granite.

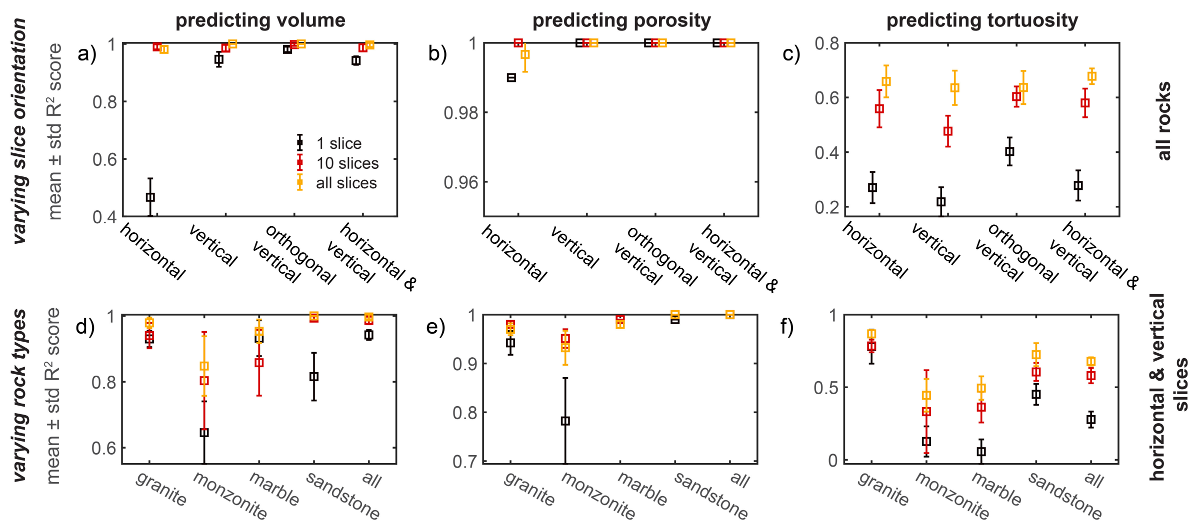

Figure 5. Performance of the gradient boosting models developed for all of the rock types using the four methods of slice extraction (a-c), and each individual rock type and all of the rock types using both horizontal and vertical slices (d-f). The performance of models developed with one slice (black), 10 slices (red), or all of the slices (yellow) are shown with different colors. The model performance does not systematically depend on the method of slice extraction (a-c). The models developed using all of the rock types perform better than models developed for individual rock types (d-f).

3.2. Deriving a relationship between two-dimensional and three-dimensional properties

The strong performance of the gradient boosting models suggest that a less complex algorithm may be able to predict the three-dimensional properties from the two-dimensional observations with success. Here, we use linear regression models to predict the three-dimensional properties because these models enable identifying a function between the two-dimensional measurements and the three-dimensional properties. In particular, a trained linear regression model includes a suite of coefficients, one for each slice in this analysis, c_i, that provides an equation between the two-dimensional measurements (i.e., features), f_i^{2D}, and the three-dimensional property, p^{3D}, (Pedregosa et al., 2011):

$$p^{3D} = \sum_{i=1}^{n} c_i f_i^{2D} \qquad \text{Eq. 1}$$

where n is the number of features: the number of two-dimensional slices in this analysis. Consequently, the three-dimensional property is estimated as a linear combination of the two-dimensional property measured at each slice multiplied by the associated coefficient, c_i. Due to the simplicity of these types of models, we expect that the performance of the models will be worse than the performance of the gradient boosting models.

Indeed, the linear regression models perform with lower R^2 scores than the boosting models (**Figure** 6). However, the linear regression models perform reasonably well when they predict the volume and porosity using only one slice along both the vertical and horizontal axes, with moderate-strong correlations between the observed and predicted values (**Figure** 6a-b). Similar to the boosting models, the linear regression models do not predict the tortuosity with strong correlations between the predicted and observed values (**Figure** 6c). The linear regression models that use data from all of the experiments perform better than the models that use data from individual rock types (**Figure** S7), similar to the boosting models. This result lends further support to the idea that the relationships between two- and three-dimensional fracture network properties may be similar in porous granular rocks (sandstone) and lower porosity crystalline rocks (granite).

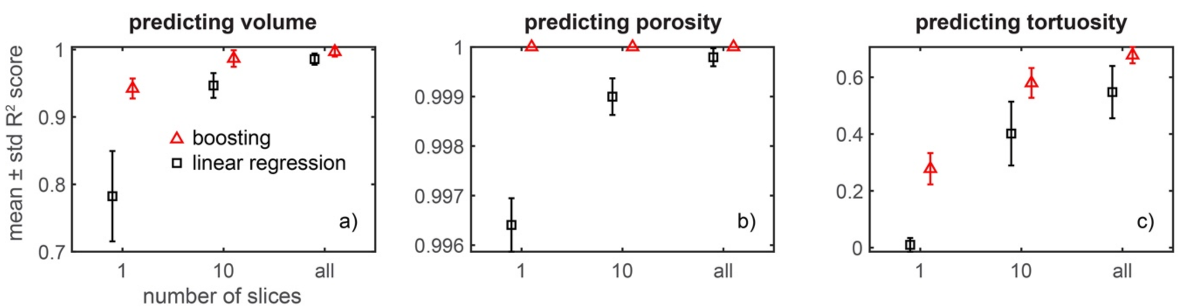

Figure 6. Performance of the gradient boosting models and linear regression models when predicting the volume (a), porosity (b), and tortuosity (c) using one, ten, and all of the slices, for data from all of the rock types, using slices along the horizontal and vertical axes. The boosting models perform better than the linear regression models. The linear regression models predict the volume of the largest fracture and porosity with strong correlations between the predicted and observed values.

The strong performance of the linear regression models that predict the porosity and volume suggest that we may examine their coefficients in order to constrain the equations that relate the two- and three-dimensional porosity and fracture volume. We expect that these two- and three-dimensional properties will be positively correlated to each other: higher two-dimensional porosity or fracture volume should indicate higher three-dimensional porosity or fracture volume.

To examine how the coefficients of the models relate to the two-dimensional properties, we focus on the results of the linear regression models that use one slice of the data, in either the vertical or horizontal direction, with data from all the experiments (**Figure** 7, **Figure** 8). As described in the Methods section, the models that use one vertical slice from one experiment sample one slice position in the rock core across all of the differential stress steps, and when the models use data from several experiments, they sample one position in all of the rock cores and all the associated stress steps. Because we expect that the value of the two-dimensional property will influence the magnitude of the coefficient, we compare the coefficient and mean of the two-dimensional value across all of the slices at one particular location, and throughout all the differential stress steps and experiments.

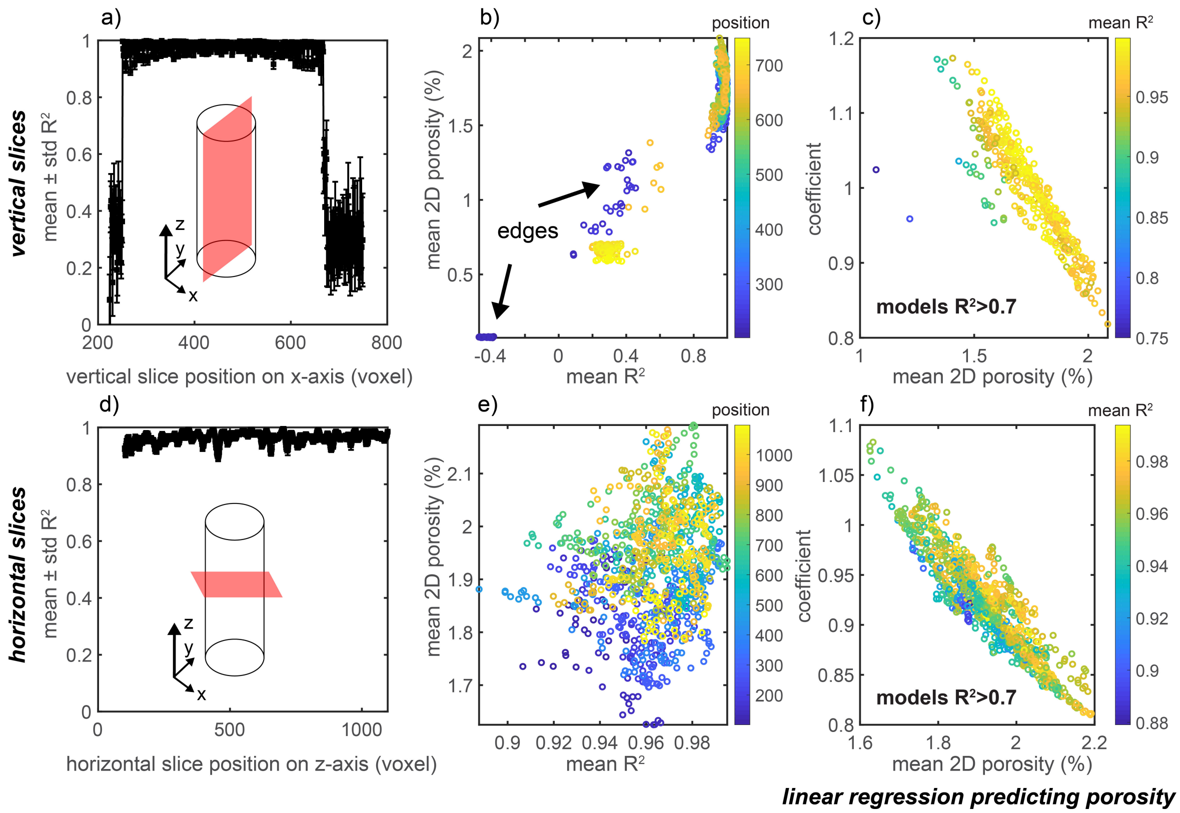

Figure 7. Relationships between the model performance, two-dimensional porosity and coefficient of the linear regression models developed with data from all the experiments, with vertical (a-c) and horizontal (d-f) slices. a, d) Mean ± one standard deviation of the R^2 of models developed from vertical slices (a) and horizontal slices (d) relative to their position along the x- or z-axis. These models only differ in how the training and testing data are split. b, e) The mean R^2 score of the models, and the corresponding mean two-dimensional porosity at the slice position used to develop the model. The color of the symbol indicates the position along the *x*- or *z*-axis of the model and two-dimensional measurement. The arrows in (b) highlight regions of the core at the edges that produce models with lower R^2 scores. c, f) The mean two-dimensional porosity and mean coefficient of the corresponding set of models for models with higher R^2 scores (>0.7). The color of the symbol indicates the R^2 score of the model.

The position of the vertical slice within the cores influences the ability of the models to predict the porosity (**Figure** 7a). When the models use data at the edges of the cores, the R^2 scores range from -0.4 to 0.5. However, when they use data within the central portion of the cores, the R^2 scores are >0.9. The lower porosity at the edges of the cores may explain the lower model performance (**Figure** 7b). Our method of accumulating the data from the different experiments produces these apparent regions of lower porosity. Because the marble and sandstone cores are wider than the granite and monzonite cores, the database that includes data from all of the experiments have columns in which no porosity is reported for the thinner cores, and in which non-zero porosity is reported for the wider cores. Thus, the models developed for these portions of the cores will likely be prone to low performance. However, this edge effect does not emerge when the models predict the volume of the largest fracture (**Figure** 8a). When the models predict the largest fracture, the position of the vertical slices that produce the most incorrect models include the edge of the cores, as well as positions within the center of the cores, near 400 and 600 voxels from the edge of the cores (**Figure** 8a). In contrast to the strong influence of position on the model performance when they predict the porosity using vertical slices, the position of the horizontal slices does not lead to systematic changes in the R^2 scores, for both the models that predict the porosity (**Figure** 7d) and volume (**Figure** 8d).

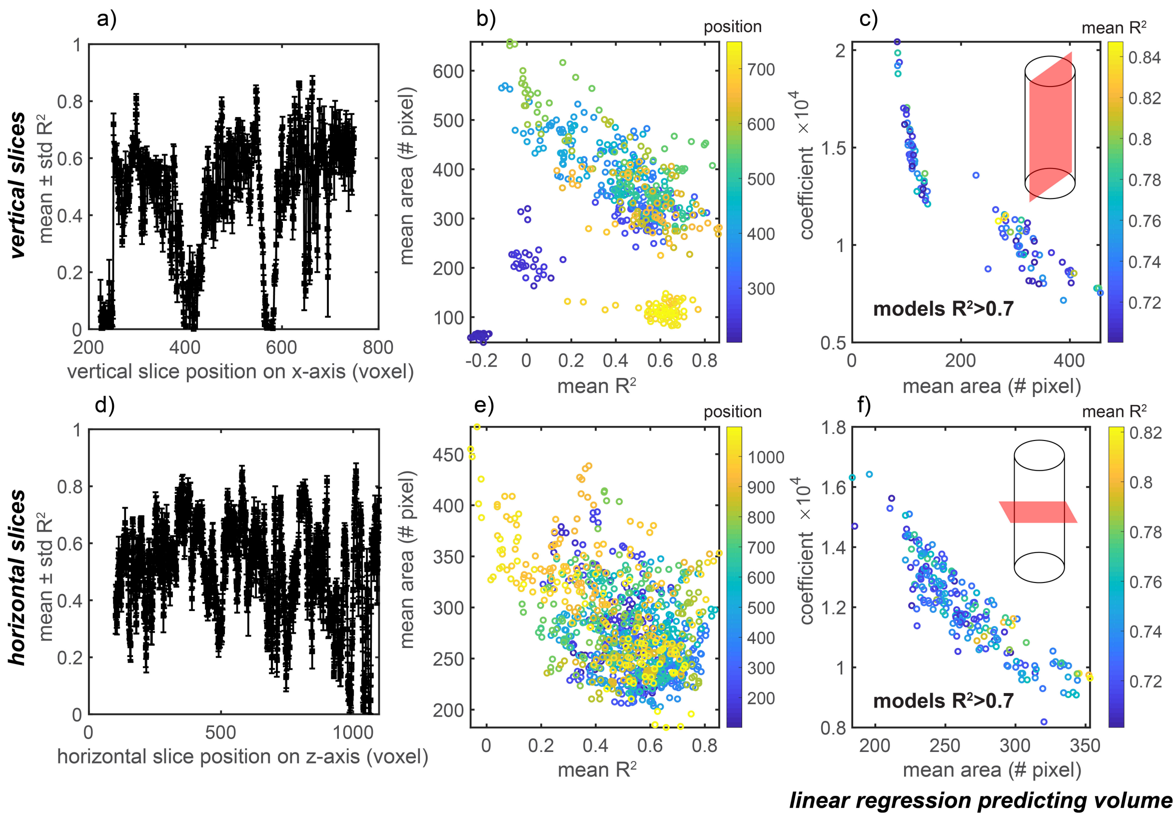

Figure 8. Relationships between the model performance, two-dimensional volume and coefficient of the linear regression models developed with data from all the experiments, with vertical (a-c) and horizontal (d-f) slices. The format of the figure is the same as **Figure** 7.

Because the values of the coefficients c_i (Eq. 1) derived from the linear regression models may depend on the magnitude of the fracture network property at a particular slice, we now compare the coefficient and mean of the two-dimensional values across all of the slices at one particular location, and throughout all the differential stress steps and experiments. For the models that predict the porosity with R^2 scores greater than 0.7, the coefficients range from about 0.8 to 1.1 (**Figure** 7c, f). The ranges of the 25th-75th percentile of the coefficients for the models developed with horizontal slices and vertical slices are 0.89-0.95 and 0.95-1.05, respectively (**Figure** S8). For the models that predict the volume of the largest fracture, the coefficients range from about 0.8-2.0$\cdot 10^4$ (voxels/pixels) (**Figure** 8c, f). The units of these coefficients are the number of voxels/pixels because they use the fracture area (pixels) to estimate the fracture volume (voxels) with the equation: fracture volume (voxels)=c(voxels/pixels)·fracture area (pixels), where c is the coefficient. The ranges of the 25th-75th percentile of the coefficients for the models developed with horizontal slices and vertical slices to predict the fracture volume are $11 \cdot 10^3$ to $13 \cdot 10^3$ (voxels/pixels) and $9 \cdot 10^3$ to $14 \cdot 10^3$ (voxels/pixels), respectively (**Figure** S8). These ranges are consistent with our expectation that a positive correlation exists between the two- and three-dimensional porosity and fracture volume. The precise values of the coefficients for the models that predict the porosity suggest that horizontal slices of rock cores may tend to overestimate the three-dimensional porosity as they require coefficients less than one, with mean values of 0.92 for the models with R^2 scores greater than 0.7. In contrast, vertical slices can provide a close approximation of the three-dimensional porosity, requiring coefficients close to one, with mean values of precisely 1.00.

There is a negative correlation between the two-dimensional porosity and the model coefficient, for both the vertical (**Figure** 7c) and horizontal slices (**Figure** 7f). Thus, higher two-dimensional porosity leads to lower coefficients for the models that predict the porosity. Similar to these models, the models that predict the volume also host a negative correlation between the two-dimensional property and the coefficient (**Figure** 8c, f). This negative correlation arises because slices that produce anomalously higher values of porosity and fracture area require a lower coefficient to approximate the three-dimensional value than slices that produce lower values.

3.3. Comparing machine learning algorithms to simple statistics

Examining the relationship between the three-dimensional porosity, and the two-dimensional measurements (e.g., **Figure** S4-S6) suggest that calculating simple statistics of the population of two-dimensional slices may closely approximate the three-dimensional porosity. In particular, the correlation between the mean of the vertical two-dimensional slices at each differential stress step and the three-dimensional porosity of a given step is strong, with correlation coefficients near one (e.g., **Figure** 9a-c). Due to the ability of this simple statistic to approximate the three-dimensional porosity, we identify when the machine learning models predict the three-dimensional porosity with greater success than calculating the mean, given the same amount of data. In particular, we find the number of vertical slices for which the mean of these slices is within 5% of the three-dimensional value at each stress step of each experiment, n_m (e.g., **Figure** S9, **Figure** 9). We randomly select a given number of slices throughout the rock, calculate the mean of those slices, and then determine if the two-dimensional mean falls within 5% of the three-dimensional value. If the two-dimensional mean does not match the three-dimensional value, then we increase the number of slices and repeat the process. To account for heterogeneity in the rock sample, we repeat this selection process 10,000 times, thereby sampling a different population of slices. We then compare the mean ± one standard deviation of n_m throughout each experiment to the number of vertical slices required for the machine learning algorithms to produce R^2 scores >0.9, n_a (**Figure** 9d). In order to compare these values, we use the models developed for the individual rock types with the vertical slices, for both the gradient boosting models and the linear regression models.

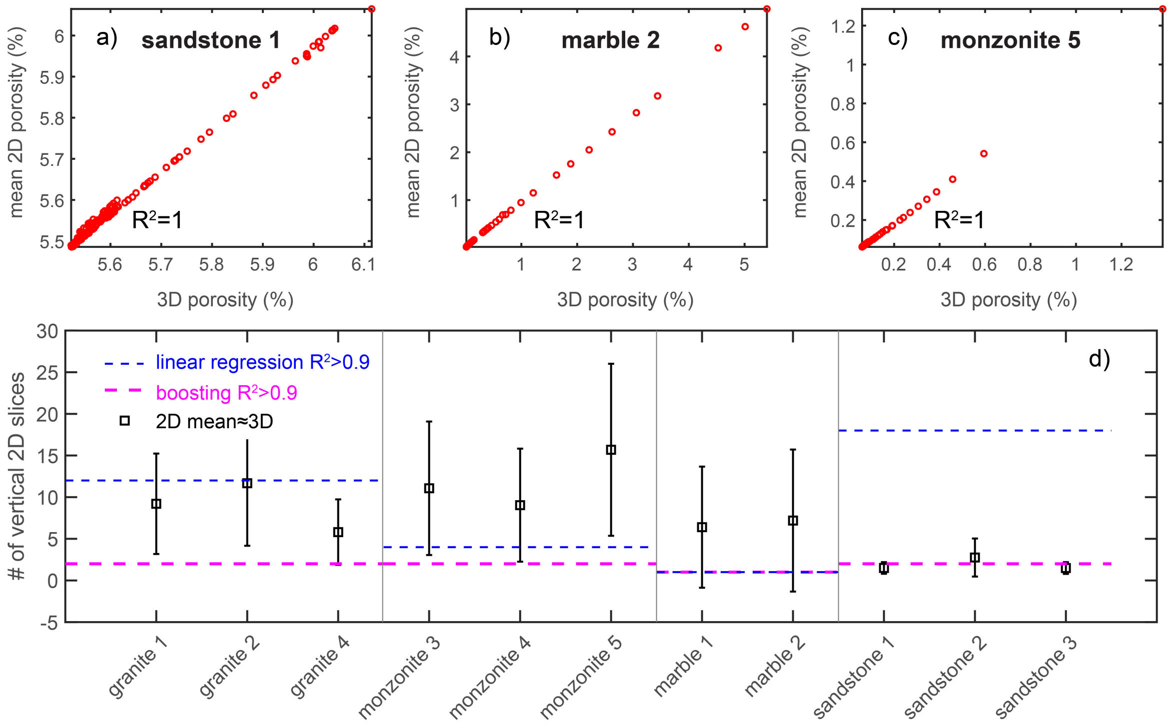

Figure 9. Comparing the performance of a simple statistic (the mean) to the machine learning algorithms. a-c) Relationship between the three-dimensional porosity and the mean two-dimensional porosity calculated from the vertical slices in three example experiments: a) sandstone #1, b) marble #2, and c) monzonite #5. The correlation coefficients between these values are 1.0 in each of these experiments. d) Comparing the number of vertical slices for which the mean of the slices is within 5% of the three-dimensional mean at each stress step (scan) of each experiment, and the number of slices required by both machine learning algorithms to predict the porosity with R^2 scores greater than 0.9. The blue and pink dashed lines indicate the linear regression and boosting results, respectively. Calculating the mean provides similarly accurate estimates as the gradient boosting models for the most homogeneous rock types (sandstone), given the same amount of data. For the more heterogeneous rocks, the machine learning algorithms provide more accurate predictions of the three-dimensional porosity than calculating the mean.

The heterogeneity of the rock determines if calculating the mean requires less data than the machine learning models in order to make successful predictions. For the boosting models, the n_a is generally smaller than the n_m, except for the sandstone experiments, indicating the superiority of the machine learning algorithms for all the experiments except the sandstone experiments (**Figure** 9d). The sandstone cores require a similar n_m and n_a. For the linear regression models, n_m is lower than n_a for two rock types: the granite and sandstone, indicating the superiority of calculating the mean rather than using linear regression for these rocks. The lower average and standard deviation of the n_m of the sandstone cores indicates that the sandstone develops the most homogeneous pore and fracture networks of the rocks examined here. These homogeneous networks then enable simple statistics to provide more accurate estimates of the three-dimensional property than the linear regression models for the same amount of data. For the more heterogeneous rocks, although they host strong correlations between the three-dimensional porosity and mean two-dimensional porosity calculated with all of the vertical slices (**Figure** 9a-c), the machine learning algorithms require less information to successfully predict the three-dimensional property than the mean (**Figure** 9d). This analysis thus reveals the rock types for which simple statistical approaches may provide reasonably accurate estimates of the three-dimensional porosity.

3.4. Identifying the statistics that predict the three-dimensional properties

The analysis of the previous section indicates that statistical approaches can help build equations between the two- and three-dimensional porosity at least in homogeneous rocks, such as sandstone. However, we cannot apply this approach to determine the usefulness of statistics for the other two properties: the volume and tortuosity of the largest fracture. Instead, we may use machine learning to identify the statistics that are most beneficial to predictions of the three-dimensional volume and tortuosity. To identify these statistics, we reformat our dataset of two-dimensional observations so that the models use the statistics of several two-dimensional observations to predict the three-dimensional value. We focus on the results of models developed using all of the rock types and vertical slices because our previous analyses indicate that combining several rock types can produce more accurate models, and that the orientation of the observation does not systematically control the model performance (**Figure** 5).

Comparing the performance of models developed from statistics of the two-dimensional observations indicates that the models predict the porosity and volume of the largest fracture with success, but struggle to predict the tortuosity (**Figure** 10a), similar to the models developed from the raw slices. In particular, the mean R^2 scores for models developed with varying amounts of data are 0.99 (porosity), 0.96 (volume) and 0.59 (tortuosity). An additional similarity is that the amount of data provided to the models does not lead to a continuous improvement of the model performance. In particular, varying the amount of data provided to the models developed using statistics from 10% to 90% of the available data does not change the R^2 score by more than 0.03 for the models that predict the porosity and volume of the largest fracture. For the models that predict the tortuosity, the mean R^2 score in fact decreases by about 0.08 when the amount of data increase from 10 to 90%, in contrast to the idea that more data invariably leads to better model performance. Because we randomly select the data provided to the models, the decrease in model performance does not arise from selecting data earlier or later in loading, for example.

To identify the statistics that are most useful to the predictions, we focus on the models developed with 90% of the data. Comparing the features that produce the highest R^2-weighted |SHAP| value, s_w, for each set of models indicates that the models that predict the porosity primarily depend on the mean and 50th percentile (**Figure** 10b). The dependence of the porosity models on these statistics agrees with the ability of the mean of the two-dimensional porosity to estimate the three-dimensional porosity using less slices than the machine learning models for the sandstone experiments (**Figure** 9). Moreover, the dependence of the porosity models on these statistics suggests that when the data from multiple rock types are combined together, the average and median of the two-dimensional porosity measurements may provide better estimates of the three-dimensional porosity than the other tested statistics.

In contrast to the models that predict the porosity, the models that predict the other two properties do not strongly depend on the mean (**Figure** 10b). Surprisingly, the models that predict the volume and tortuosity of the largest fracture depend primarily on the minimum area, or minimum tortuosity, of the largest fracture identified in the group of two-dimensional slices. This result suggest that the extreme values of the population of the largest fracture identified in a particular two-dimensional slice (the minimum value) provides the most useful information about the properties of the largest fracture identified in the full three-dimensional scan.

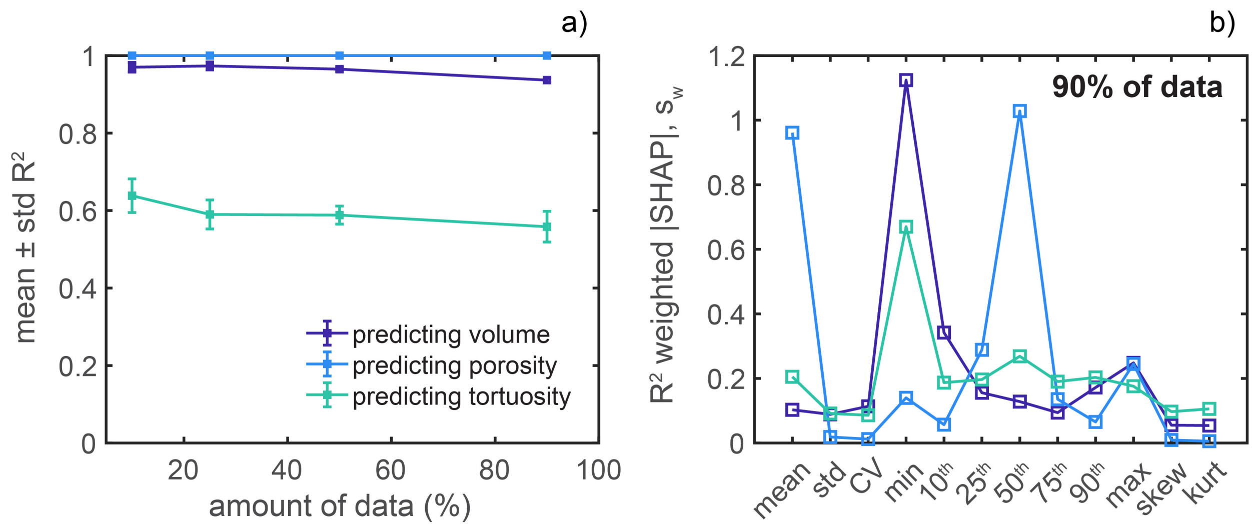

Figure 10. Score (a) and R^2 weighted |SHAP| value distribution (b) of models developed using the statistics of groups of vertical two-dimensional measurements for all of the rock types combined: the results for models with all of the tested amounts of data (a), and the results for only the models with 90% of the data (b). The statistics used as features are the mean, standard deviation (std), coefficient of variation (CV), minimum, 10^{th}-90^{th} percentile, maximum, skewness (skew), and kurtosis (kurt). a) The models that predict the volume and porosity perform very well, and better than the models that predict the tortuosity. Varying the amount of data from 10-90% does not lead to large changes in the mean R^2 score. b) The models that predict the porosity primarily depend on the mean and median value. In contrast, the models that predict the properties of the largest fracture primarily depend on the minimum value in a group of two-dimensional observations.

4 Discussion

4.1. Predictability of three-dimensional properties of fracture networks

The present study enables a direct comparison of the predictability of several fracture network characteristics that control fluid flow, and thus the potential ability to derive an equation between the two- and three-dimensional measurements. The results indicate that it may be more difficult to derive a function between the two- and three-dimensional tortuosity than the other properties (**Figure** 4). This difficulty of predicting the three-dimensional tortuosity is consistent with previous work that found that using two-dimensional estimates of fluid flow leads to erroneous estimates of the three-dimensional flow properties, such as the permeability (e.g., Li et al., 2005; Duda et al., 2011; Mostaghimi et al., 2013; Lang et al., 2014; Marafini et al., 2020). This result is also consistent with work that compared the size of the representative elementary volume (REV) for permeability, porosity, and specific surface area (Mostaghimi et al., 2013). The REV is the minimum volume for which the property of interest (i.e., permeability) varies less than some threshold from the property calculated in larger volumes (e.g., Bear, 1988; Zhang et al., 2000). Mostaghimi et al. (2013) found that the REV for the permeability was up to twice as large as the REV for the porosity and specific surface area in granular rocks, including carbonate and sandstone. The larger REV suggests that the permeability distribution was more heterogeneous than the porosity and the specific surface area. Because the permeability depends on the tortuosity of the fracture network, the larger REV of the permeability relative to the porosity (Mostaghimi et al., 2013) agrees with the excellent performance of the machine learning models in the present analysis when they predict the porosity and volume of the largest fracture, and mediocre performance of the models when they predict the tortuosity (**Figure** 4). The more heterogeneous and anisotropic distribution of the tortuosity of the fracture network compared to the other properties, and the importance of considering connectivity in three-dimensions, produce both results.

Tortuosity, and therefore permeability, are thus more difficult to predict in the laboratory than other properties of the fracture network. Similarly, estimates of the subsurface permeability can depend on the scale of the measurement. In particular, estimates of permeability derived from core samples tend to be lower than the estimates of permeability derived from pumping tests (e.g., Rovey & Cherkauer, 1995; Sánchez-Villa et al., 1996; Raghavan, 2006). These studies suggest that permeability increases with spatial scale. The presence of stratification, layering and other heterogeneities with a spatial dimension larger than the typical core sample produces this spatial dependence (e.g., Raghavan, 2006). Laboratory data further supports the idea that permeability increases with spatial scale (e.g., Schulze-Makuch et al., 1999). However, simulations indicate that permeability can both increase and decrease with spatial scale (e.g., Sahimi et al., 1986; Nordahl & Ringrose, 2008; Esmaeilpour et al., 2021; Ghanbarian, 2022). Consequently, this previous work demonstrates that varying the spatial scale in three-dimensions can change estimates of permeability, and the present analysis demonstrates that it is similarly difficult to estimate the three-dimensional tortuosity (and thus permeability) from two-dimensional observations. The dimensional contraction from three to two dimensions thus produces a similar effect as varying the spatial length scale in three dimensions.

The greater predictability of the porosity suggests that we may be able to estimate the three-dimensional mechanical properties of rocks that depend on the porosity (i.e., Young's modulus) with more success than the properties that depend on the tortuosity (i.e., permeability). Consequently, estimates of the permeability may require direct numerical computation of the flow within the three-dimensional system, such as solving for Stokes flow or performing lattice Boltzmann simulations (e.g., Bultreys et al., 2016). Moreover, the relationship between the two- and three-dimensional estimates of elastic moduli may be easier to constrain than the corresponding relationship for the permeability. Indeed, recent work has successfully estimated three-dimensional elastic moduli using the aspect ratio of the pores reconstructed from three orthogonal two-dimensional slices of Berea sandstone and Grosmont carbonate (Karimpouli et al., 2018). In summary, because two-dimensional measurements of porosity can be closely linked to the three-dimensional porosity, properties that depend on the porosity, such as the elastic moduli, may be similarly predictable.

4.2. Amount and type of data required for successful predictions

The analysis helps constrain the amount and type of data required for successful estimates of the three-dimensional properties. One may expect that increasing the amount of data provided to the models would increase the model performance. Our results agree with this expectation: higher numbers of two-dimensional estimates generally produce larger R^2 scores (**Figure** 4, **Figure** 5, **Figure** 6). However, the performance of the models does not continually increase with the number of two-dimensional slices (**Figure** 4). Instead, the model performance increases relatively rapidly over at most ten slices, and then does not change significantly between ten slices, and all of the available slices (e.g., **Figure** 5). Only one to four slice positions are required for reasonably accurate estimates of the porosity and volume of the largest fracture.

This analysis enables direct comparison of the usefulness of examining two-dimensional slices that are oriented parallel (vertical) or perpendicular (horizontal) to the maximum compression direction, as well as combinations of these orientations. One may expect that the models developed with both horizontal and vertical slices would provide the most accurate estimates. However, the results do not indicate that one method systematically produces higher model performance (**Figure** 4, **Figure** 5a-c). This result suggests that the orientation of thin-sections or field measurements may not exert a significant control on estimates of the three-dimensional property. Moreover, several parallel two-dimensional observations may provide comparable estimates of the three-dimensional system as several perpendicular two-dimensional observations. The orientation of the measurement could be more significant for properties in rocks that are more anisotropic than those analyzed here, such as layered sedimentary rock.

This analysis provides further insight into the influence of rock type on the predictability of the fracture network characteristics. One may expect that different relationships can develop between the two- and three-dimensional data in different rock types. For example, the equation that relates the two-dimensional porosity to the three-dimensional porosity may differ between the sandstone and crystalline rocks because sandstone hosts many quasi-spherical pores (e.g., Dong & Blunt, 2009), whereas the granite contains fractures and pores with more anisotropic shapes (e.g., Renard et al., 2018). If such different relationships exist, one would expect that the performance of the models developed for individual rock types would be higher than the models developed for a combination of rock types. In contrast, the results indicate that the models developed using all of the experimental data perform better than or similarly to the models developed using individual rock types (**Figure** 5d-f). This trend may arise in part from the larger amount of data provided to the models developed for all the rock types compared to the models developed for individual rock types. However, the higher values of the correlation coefficients for the data accumulated from all of the experiments compared to the data from individual experiments (**Figure** S6) suggest that the content of the data also influences the model performance, and not only the amount of data. Previous analyses that link two- and three-dimensional properties have tended to focus on specific rock types (e.g., Karimpouli et al., 2018), rather than accumulating data from both low porosity crystalline rocks and granular rocks. Moreover, previous analyses have identified specific failure criteria that are applicable to porous granular rock, such as sandstone, but not to low porosity crystalline rock with interlocking minerals (e.g., Wilshaw, 1971; Zhang et al., 1990). Similarly, theories from linear fracture mechanics that relate the geometric properties of individual fractures, including the length and orientation, to their propensity for propagation using the stress intensity factor do not explicitly account for the presence of nearby quasi-spherical pores (e.g., Paterson & Wong, 2005), and therefore may only be applicable to low porosity rocks, such as granite, before neighboring fractures begin to perturb each other's local stress field. The present analysis suggests that future work may benefit from considering data from a variety of rock types, including both sandstone and granite, when deriving equations between the two- and three-dimensional properties.

4.3. Comparing complex machine learning algorithms to simple algorithms and statistics

Consistent with the performance of the gradient boosting models (**Figure** 4), and the correlation coefficients between the two- and three-dimensional porosity, and fracture area and volume (**Figure** S6), linear regression models are able to perform with moderate-strong correlations between the observed and predicted values (**Figure** 6a-b). The strong performance of these models allowed close examination of the coefficients of the models (**Figure** 7, **Figure** 9), which help provide constraints on equations that relate the three-dimensional and two-dimensional properties. For models that predict the porosity with R^2 scores greater than 0.7, the range of the value of the coefficients (**Figure** S8) suggests that the horizontal slices of rock cores tend to overestimate the three-dimensional porosity as they require coefficients less than one. In contrast, the vertical slices can provide a close approximation of the three-dimensional porosity, requiring coefficients close to one. One explanation for this stronger correlation could be that the presence of vertically-trending fractures produces larger porosities and areas measured along the vertical orientation. However, measuring the orientation of individual fractures in a subset of these experiments, which includes all but the sandstone experiments, does not reveal a clear preferred orientation of the fractures (McBeck et al., 2022). The preferred explanation of this result is that the vertical slices provide more information about the system than the horizontal slices because they sample a larger area. Indeed, when the models only have access to one slice, the models developed with one vertical slice perform better than the models developed with one horizontal slice (e.g., **Figure** 5a-b) when they predict the volume and porosity using data from all of the rock types.

Examining the coefficients of the linear regression models that predict the volume of the largest fracture reveals that the complex geometry of the largest fractures produces a discrepancy between the mathematically-expected relationship between the fracture area and volume, and the coefficients determined from machine learning (**Text** S2). **Text** S2 describes the derivation of the value of the coefficients expected from the intersection of a plane with ellipsoids of varying shape anisotropy. The geometric complexity of the large, system-spanning fractures can cause the most volumetric fracture identified in the entire three-dimensional system to not be the largest fracture, with the highest area, identified in a particular two-dimensional slice. The machine learning models then must use the area of the largest fracture identified in a particular slice to predict the volume of the largest fracture found throughout the system, which may not be the same fracture. The high values of the coefficients highlight that accurate estimates of the volume of the largest fracture in a three-dimensional system from two-dimensional measurements of the area require a larger multiplicative factor than expected mathematically from the intersection of a plane with an ellipsoid.

Due to the strong performance of the linear regression models and strong correlations between the mean of the two-dimensional porosity measurements and three-dimensional values (**Figure** 9a-c), we examined the ability of simple statistics to estimate the three-dimensional porosity. For the rocks with the most isotropic and homogeneous fracture networks (sandstone), calculating the mean of the two-dimensional measurements requires about the same number of slices for successful estimates as the gradient boosting models, and a lower number of slices compared to the linear regression models (**Figure** 9d). However, for the more heterogeneous rocks (granite, monzonite, marble), the gradient boosting models require a lower number of slices, and the linear regression models require the same or a lower number of slices than calculating the mean for all of the rocks.

The homogeneous pore and fracture networks of the sandstone thus enable simple statistics to provide more accurate estimates of the three-dimensional property than the linear regression models, and similarly accurate estimates as the gradient boosting models, for the same amount of data. This result is consistent with previous work that estimated the three-dimensional permeability using two-dimensional measurements in sandstone with success, but was not able to derive accurate estimates in more heterogeneous carbonate rocks (Saxena et al., 2017). The heterogeneity of the fracture and pore network controls the potential accuracy of estimates of the three-dimensional property because a random two-dimensional slice will likely be less representative of the full three-dimensional system in a heterogeneous rock compared to a homogeneous rock. Thus, homogeneous and heterogeneous rocks may require different approaches to approximating the influence of fracture networks on fluid flow. An equivalent porous medium approach, in which representative elementary volumes approximate the distribution of hydrogeologic properties (e.g., Shaik et al., 2011), may be appropriate for homogeneous rock. In contrast, such continuum approaches may not be appropriate for more heterogeneous rocks, which instead may require modelling with a discrete fracture network (e.g., Cacas et al., 1990). In this type of modelling, a population of fractures is stochastically generated using probability density functions that determine their geometric properties, including lengths, apertures, and orientations (e.g., Lei et al., 2017). Previous work has used observations of natural fracture networks to build two-dimensional models with discrete fracture networks (e.g., Belayneh & Cosgrove, 2004). However, because there is no established method of extrapolating fracture geometries from two- to three-dimensions, few studies have extended such natural observations into a three-dimensional discrete fracture network model (Lei et al., 2017). The present analysis provides insight into how to robustly perform this extrapolation.

Similar to the apparent effectiveness of continuum approaches for the homogeneous rocks (e.g., **Figure** 9d), models developed with all of the rock types to predict the porosity using a set of statistics of the two-dimensional measurements primarily depend on the mean and median (**Figure** 10). However, the models that predict the volume and tortuosity of the largest fracture do not strongly depend on the mean, and instead depend primarily on the minimum area, or minimum tortuosity, of the largest fracture identified in the group of two-dimensional slices. The minimum values may be the most useful to the model predictions because they control the ability of the fracture network to achieve the percolation threshold, when the fracture network traverses the system. Indeed, in synthetic isotropic fracture networks with power law sizedistributed fractures near the percolation threshold, the two-dimensional permeability tends to underestimate the three-dimensional permeability by three orders of magnitude (Lang et al., 2014). Percolation thus may occur in three-dimensions, but not in two-dimensions when the fracture density is low (Lang et al., 2014). Consequently, the minimum value of the fracture area and tortuosity measured in two-dimensions should have a significant impact on the overall three-dimensional volume and tortuosity. Thus, efforts to predict the geometry and resulting connectivity of the largest fracture in a system should focus on the smallest values in a population of two-dimensional observations rather than the mean or maximum values.

5 Conclusions

Data from eleven in situ X-ray tomography experiments during triaxial deformation provide unique insights into the relationship between two- and three-dimensional measurements of properties that control fluid flow in both homogeneous and heterogeneous rocks subjected to differential stress loading until macroscopic failure. The machine learning models that predict the porosity and volume of the largest fracture perform with strong correlations between the predicted and observed values, and better than models that predict the tortuosity of the largest fracture. This result highlights the difficulty of successfully estimating the tortuosity, and related properties such as the permeability, from two-dimensional measurements or simulations, but suggests that two-dimensional approaches may provide robust insights for analyses that focus only on the volume of the largest fracture or porosity.

The analysis enables close examination of the amount and type of two-dimensional data required for successful estimates of the three-dimensional property. The models presented here can achieve accurate estimates of the porosity and volume of the largest fracture using only one to four two-dimensional slices. Moreover, the method of slice extraction does not systematically influence the model performance. In addition, models developed using data from all of the experiments perform better than models developed for individual rock types. Consequently, based on our dataset of triaxial compression experiments: 1) dense sampling of the subsurface or rock core may not be required for successful estimates of some three-dimensional properties (porosity, fracture volume); 2) the orientation of field measurements may not exert a significant control on estimates of the three-dimensional property; and 3) including data from a variety of rock types may lead to more successful estimates of the three-dimensional property than only focusing on one rock type, in contrast to previous work that developed separate failure criteria for separate rock types (e.g., Zhang et al., 1990; Paterson & Wong, 2005).

Comparing the amount of data required for more complex machine learning algorithms (gradient boosting) to estimate the porosity to the amount required for simple statistics highlights the benefit of machine learning for heterogeneous rocks. When rocks contain homogeneous and isotropic pore and fracture networks, calculating the mean of the two-dimensional slices requires a similar amount of data to successfully estimate the three-dimensional values as the gradient boosting models, and less data than the linear regression models. Similarly, models developed using a set of statistics of the two-dimensional measurements of all of the rock types indicate that the mean and median of the two-dimensional value provide the most useful information when the models predict the porosity. These results suggest that equivalent porous medium approaches (e.g., Shaik et al., 2011) may be appropriate for homogeneous rocks, such as sandstone. However, when the models predict the tortuosity and volume of the largest fracture, the models primarily depend on the minimum value of the two-dimensional measurement in a set of slices. Thus, efforts to reconstruct the geometry and connectivity of the largest fracture in a system should focus on the smallest values in a population of local two-dimensional observations, rather than the mean. Heterogeneous rocks that include fracture populations with a wide range of lengths may require modelling using discrete fracture networks (e.g., Lei et al., 2017), rather than continuum approaches.

Acknowledgments

The study was funded by the Norwegian Research Council (grant 300435 to JM), UNINETT Sigma2 AS (project NN9806K), and the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (grant agreement No. 101019628 BREAK to FR). The experimental data is available on the Sigma2/NIRD/Norstore repository (Renard, 2017, 2018, 2021).

References

Auradou, H., Drazer, G., Hulin, J. P., & Koplik, J. (2005). Permeability anisotropy induced by the shear displacement of rough fracture walls. *Water Resources Research*, 41(9).

Bakke, S., & Øren, P. E. (1997). 3-D pore-scale modelling of sandstones and flow simulations in the pore networks. *SPE Journal*, 2(02), 136-149.

Bear, J. (1988) Dynamics of fluids in porous media. Elsevier, New York.

Belayneh, M., & Cosgrove, J. W. (2004). Fracture-pattern variations around a major fold and their implications regarding fracture prediction using limited data: an example from the Bristol Channel Basin. *Geological Society, London, Special Publications*, 231(1), 89-102.

Bobet, A., & Einstein, H. H. (1998). Fracture coalescence in rock-type materials under uniaxial and biaxial compression. *International Journal of Rock Mechanics and Mining Sciences*, 35(7), 863-888.

Buades, A., Coll, B., & Morel, J. M. (2005). A non-local algorithm for image denoising. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (CVPR'05) (Vol. 2, pp. 60-65). IEEE.

Bultreys, T., De Boever, W., & Cnudde, V. (2016). Imaging and image-based fluid transport modeling at the pore scale in geological materials: A practical introduction to the current state-of-the-art. *Earth-Science Reviews*, 155, 93-128.

Bühlmann, P., & Yu, B. (2003). Boosting with the L2 loss: regression and classification. *Journal of the American Statistical Association*, 98(462), 324-339.

Cacas, M. C., Ledoux, E., de Marsily, G., Tillie, B., Barbreau, A., Durand, E., ... & Peaudecerf, P. (1990). Modeling fracture flow with a stochastic discrete fracture network: calibration and validation: 1. The flow model. *Water Resources Research*, 26(3), 479-489.

Cao, R., Yao, R., Meng, J., Lin, Q., Lin, H., & Li, S. (2020). Failure mechanism of non-persistent jointed rock-like specimens under uniaxial loading: laboratory testing. *International Journal of Rock Mechanics and Mining Sciences*, 132, 104341.

Caumon, G., Collon-Drouaillet, P. L. C. D., Le Carlier de Veslud, C., Viseur, S., & Sausse, J. (2009). Surface-based 3D modeling of geological structures. *Mathematical Geosciences*, 41(8), 927-945.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKKD international conference on knowledge discovery and data mining*, 785-794.

Dominguez, S., Malavieille, J., & Lallemand, S. E. (2000). Deformation of accretionary wedges in response to seamount subduction: Insights from sandbox experiments. *Tectonics*, 19(1), 182-196.

Dong, H., & Blunt, M. J. (2009). Pore-network extraction from micro-computerized-tomography images. *Physical Review E*, 80(3), 036307.

Duda, A., Koza, Z., & Matyka, M. (2011). Hydraulic tortuosity in arbitrary porous media flow. *Physical Review E*, 84(3), 036319.

Elmorsy, M., El-Dakhakhni, W., & Zhao, B. (2022). Generalizable Permeability Prediction of Digital Porous Media via a Novel Multi-scale 3D Convolutional Neural Network. *Water Resources Research*, e2021WR031454.

Esmaeilpour, M., Ghanbarian, B., Liang, F., & Liu, H. H. (2021). Scale-dependent permeability and formation factor in porous media: Applications of percolation theory. *Fuel*, 301, 121090.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 29(5), 1189-1232.

Ghanbarian, B. (2022). Estimating the scale dependence of permeability at pore and core scales: Incorporating effects of porosity and finite size. *Advances in Water Resources*, 104123.

Gueting, N., Caers, J., Comunian, A., Vanderborght, J., & Englert, A. (2018). Reconstruction of three-dimensional aquifer heterogeneity from two-dimensional geophysical data. *Mathematical Geosciences*, 50(1), 53-75.

Hazlett, R. D. (1997). Statistical characterization and stochastic modeling of pore networks in relation to fluid flow. *Mathematical Geology*, 29(6), 801-822.

Iding, M., & Ringrose, P. (2010). Evaluating the impact of fractures on the performance of the In Salah CO2 storage site. *International Journal of Greenhouse Gas Control*, 4(2), 242-248.

Kamrava, S., Tahmasebi, P., & Sahimi, M. (2020). Linking morphology of porous media to their macroscopic permeability by deep learning. *Transport in Porous Media*, 131(2), 427-448.

Karimpouli, S., & Tahmasebi, P. (2016). Conditional reconstruction: An alternative strategy in digital rock physics. *Geophysics*, 81(4), D465-D477.

Karimpouli, S., Tahmasebi, P., & Saenger, E. H. (2018). Estimating 3D elastic moduli of rock from 2D thin-section images using differential effective medium theory. *Geophysics*, 83(4), MR211-MR219.

Keehm, Y., Mukerji, T., & Nur, A. (2004). Permeability prediction from thin sections: 3D reconstruction and Lattice-Boltzmann flow simulation. *Geophysical Research Letters*, 31(4).

Kozłowska, M., Brudzinski, M. R., Friberg, P., Skoumal, R. J., Baxter, N. D., & Currie, B. S. (2018). Maturity of nearby faults influences seismic hazard from hydraulic fracturing. *Proceedings of the National Academy of Sciences*, 115(8), E1720-E1729.

Lang, P. S., Paluszny, A., & Zimmerman, R. W. (2014). Permeability tensor of three-dimensional fractured porous rock and a comparison to trace map predictions. *Journal of Geophysical Research: Solid Earth*, 119(8), 6288-6307.

Lei, Q., Latham, J. P., Tsang, C. F., Xiang, J., & Lang, P. (2015). A new approach to upscaling fracture network models while preserving geostatistical and geomechanical characteristics. *Journal of Geophysical Research: Solid Earth*, 120(7), 4784-4807.

Lei, Q., Latham, J. P., & Tsang, C. F. (2017). The use of discrete fracture networks for modelling coupled geomechanical and hydrological behaviour of fractured rocks. *Computers and Geotechnics*, 85, 151-176.

Li, Y., LeBoeuf, E. J., Basu, P. K., & Mahadevan, S. (2005). Stochastic modeling of the permeability of randomly generated porous media. *Advances in Water Resources*, 28(8), 835-844.

Luhmann, A. J., Tutolo, B. M., Bagley, B. C., Mildner, D. F., Seyfried Jr, W. E., & Saar, M. O. (2017). Permeability, porosity, and mineral surface area changes in basalt cores induced by reactive transport of CO2-rich brine. *Water Resources Research*, 53(3), 1908-1927.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 4765-4774.

Manwart, C., & Hilfer, R. (1999). Reconstruction of random media using Monte Carlo methods. *Physical Review E*, 59(5), 5596.

Manwart, C., Torquato, S., & Hilfer, R. (2000). Stochastic reconstruction of sandstones. *Physical Review E*, 62(1), 893

Marafini, E., La Rocca, M., Fiori, A., Battiato, I., & Prestininzi, P. (2020). Suitability of 2D modelling to evaluate flow properties in 3D porous media. *Transport in Porous Media*, 134(2), 315-329.

McBeck, J., Aiken, J. M., Cordonnier, B., Ben-Zion, Y., & Renard, F. (2022). Predicting fracture network development in crystalline rocks. *Pure and Applied Geophysics*, 179(1), 275-299.

McBeck, J. A., Cordonnier, B., & Renard, F. (2021). The influence of spatial resolution and noise on fracture network properties calculated from X-ray microtomography data. *International Journal of Rock Mechanics and Mining Sciences*, 147, 104922.

Mitchell, T. M., & Faulkner, D. R. (2012). Towards quantifying the matrix permeability of fault damage zones in low porosity rocks. *Earth and Planetary Science Letters*, 339, 24-31

Moore, G. F., Shipley, T. H., Stoffa, P. L., Karig, D. E., Taira, A., Kuramoto, S., Tokuyama, H., & Suyehiro, K. (1990). Structure of the Nankai Trough accretionary zone from multichannel seismic reflection data. *Journal of Geophysical Research: Solid Earth*, 95(B6), 8753-8765.

Mostaghimi, P., Blunt, M. J., & Bijeljic, B. (2013). Computations of absolute permeability on micro-CT images. *Mathematical Geosciences*, 45(1), 103-125.

Nordahl, K., & Ringrose, P. S. (2008). Identifying the representative elementary volume for permeability in heterolithic deposits using numerical rock models. *Mathematical Geosciences*, 40(7), 753-771.

Paterson, M. S., & Wong, T. F. (2005). Experimental rock deformation-the brittle field. *Springer Science & Business Media*.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplass, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, 12, 2825-2830.

Raghavan, R. (2006). Some observations on the scale dependence of permeability by pumping tests. *Water Resources Research*, 42(7)

Renard, F. (2017). Critical evolution of damage towards system size failure in a crystalline rock [Data set]. Norstore. doi:10.11582/2017.00025.

Renard, F. (2018). Volumetric and shear processes in crystalline rock during the approach to faulting [Data set]. Norstore. doi:10.11582/2018.00023.

Renard, F. (2021). X-ray tomography data of Westerley granite [Data set]. Norstore. doi:10.11582/2021.00002.

Renard, F., Cordonnier, B., Dysthe, D. K., Boller, E., Tafforeau, P., & Rack, A. (2016). A deformation rig for synchrotron microtomography studies of geomaterials under conditions down to 10 km depth in the Earth. *Journal of Synchrotron Radiation*, 23(4), 1030-1034.

Renard, F., McBeck, J., Cordonnier, B., Zheng, X., Kandula, N., Sanchez, J. R., & Dysthe, D. K. (2019). Dynamic in situ three-dimensional imaging and digital volume correlation analysis to quantify strain localization and fracture coalescence in sandstone. *Pure and Applied Geophysics*, 176(3), 1083-1115.

Renard, F., Weiss, J., Mathiesen, J., Ben-Zion, Y., Kandula, N., & Cordonnier, B. (2018). Critical evolution of damage toward system-size failure in crystalline rock. *Journal of Geophysical Research: Solid Earth*, 123(2), 1969-1986.

Roberts, A. P. (1997). Statistical reconstruction of three-dimensional porous media from two-dimensional images. *Physical Review E*, 56(3), 3203.

Rovey, C. W., & Cherkauer, D. S. (1995). Scale dependency of hydraulic conductivity measurements. *Groundwater*, 33(5), 769-780.

Rutter, E. H. (1972). The influence of interstitial water on the rheological behaviour of calcite rocks. *Tectonophysics*, 14(1), 13-33.

Sahimi, M., Hughes, B. D., Scriven, L. E., & Davis, H. T. (1986). Dispersion in flow through porous media—I. One-phase flow. *Chemical Engineering Science*, 41(8), 2103-2122.

Santos, J. E., Xu, D., Jo, H., Landry, C. J., Prodanović, M., & Pyrcz, M. J. (2020). PoreFlow-Net: A 3D convolutional neural network to predict fluid flow through porous media. *Advances in Water Resources*, 138, 103539.

Sánchez-Vila, X., Carrera, J., & Girardi, J. P. (1996). Scale effects in transmissivity. *Journal of Hydrology*, 183(1-2), 1-22.

Sassi, W., Colletta, B., Balé, P., & Paquereau, T. (1993). Modelling of structural complexity in sedimentary basins: the role of pre-existing faults in thrust tectonics. *Tectonophysics*, 226(1-4), 97-112.

Saxena, N., & Mavko, G. (2016). Estimating elastic moduli of rocks from thin sections: Digital rock study of 3D properties from 2D images. *Computers & Geosciences*, 88, 9-21.

Saxena, N., Mavko, G., Hofmann, R., & Srisutthiyakorn, N. (2017). Estimating permeability from thin sections without reconstruction: Digital rock study of 3D properties from 2D images. *Computers & Geosciences*, 102, 79-99.

Schulze-Makuch, D., Carlson, D. A., Cherkauer, D. S., & Malik, P. (1999). Scale dependency of hydraulic conductivity in heterogeneous media. *Groundwater*, 37(6), 904-919.

Shaik, A. R., Rahman, S. S., Tran, N. H., & Tran, T. (2011). Numerical simulation of fluid-rock coupling heat transfer in naturally fractured geothermal system. *Applied Thermal Engineering*, 31(10), 1600-1606.

Sudakov, O., Burnaev, E., & Koroteev, D. (2019). Driving digital rock towards machine learning: Predicting permeability with gradient boosting and deep neural networks. *Computational Geosciences*, 127, 91–98.

Tian, J., Qi, C., Sun, Y., Yaseen, Z. M., & Pham, B. T. (2021). Permeability prediction of porous media using a combination of computational fluid dynamics and hybrid machine learning methods. *Engineering with Computers*, 37(4), 3455-3471.

Tong, H., Koyi, H., Huang, S., & Zhao, H. (2014). The effect of multiple pre-existing weaknesses on formation and evolution of faults in extended sandbox models. *Tectonophysics*, 626, 197-212.

Wei, T., Fan, W., Yu, N., & Wei, Y. N. (2019). Three-dimensional microstructure characterization of loess based on a serial sectioning technique. *Engineering Geology*, 261, 105265.

Wilshaw, T. R. (1971). The Hertzian fracture test. *Journal of Physics D: Applied Physics*, 4(10), 1567.

Yeong, C. L. Y., & Torquato, S. (1998). Reconstructing random media. *Physical Review E*, 57(1), 495.

Zhang, J., Wong, T. F., & Davis, D. M. (1990). Micromechanics of pressure-induced grain crushing in porous rocks. *Journal of Geophysical Research: Solid Earth*, 95(B1), 341-352.

Zhang, D., Zhang, R., Chen, S., & Soll, W. E. (2000). Pore scale study of flow in porous media: Scale dependency, REV, and statistical REV. *Geophysical Research Letters*, 27(8), 1195-1198.

Øren, P. E., & Bakke, S. (2002). Process based reconstruction of sandstones and prediction of transport properties. *Transport in Porous Media*, 46(2), 311-343.

Øren, P. E., & Bakke, S. (2003). Reconstruction of Berea sandstone and pore-scale modelling of wettability effects. *Journal of Petroleum Science and Engineering*, 39(3-4), 177-199.