Three-Dimensional Permeability Inversion Using Convolutional Neural Networks and Positron Emission Tomography

Zitong Huang[1], Takeshi Kurotori[2], Ronny Pini[3], Sally Benson[2], and Christopher Zahasky[1]

[1]University of Wisconsin-Madison
[2]Stanford University
[3]Imperial College London

November 22, 2022

Abstract

Quantification of heterogeneous multiscale permeability in geologic porous media is key for understanding and predicting flow and transport processes in the subsurface. Recent utilization of in situ imaging, specifically positron emission tomography (PET), enables the measurement of three-dimensional (3-D) time-lapse radiotracer solute transport in geologic media. However, accurate and computationally efficient characterization of the permeability distribution that controls the solute transport process remains challenging. Leveraging the relationship between local permeability variation and solute advection rates, an encoder-decoder based convolutional neural network (CNN) is implemented as a permeability inversion scheme using a single PET scan of a radiotracer pulse injection experiment as input. The CNN consists of Densely Connected Neural Networks that can accurately capture the 3-D spatial correlation between the permeability and the radiotracer solute arrival time difference maps in geologic cores. We first test the inversion accuracy using 500 synthetic test datasets. We then use a suite of experimental PET imaging datasets acquired on four different geologic cores. The network-inverted permeability maps from the geologic cores are used to parameterize forward numerical models that are directly compared with the experimental PET imaging datasets. The results indicate that a single trained network can generate robust, denoised 3-D permeability inversion maps in seconds. Numerical models parameterized with these permeability maps closely capture the experimental solute arrival time behavior. This approach presents an unprecedented improvement for efficiently characterizing multiscale permeability heterogeneity in complex geologic materials.

Three-Dimensional Permeability Inversion Using Convolutional Neural Networks and Positron Emission Tomography

Zitong Huang[1], **Takeshi Kurotori**[2,3], **Ronny Pini**[3], **Sally M. Benson**[2], **Christopher Zahasky**[1]

[1]Department of Geoscience, University of Wisconsin-Madison, Madison, WI, USA
[2]Department of Energy Resources Engineering, Stanford University, Stanford, CA, USA
[3]Department of Chemical Engineering, Imperial College London, London, UK

Key Points:

- Positron emission tomography (PET) quantifies spatially-resolved solute transport that provides input data for permeability map inversion.
- A deep encoding–decoding convolutional neural network (CNN) is developed for permeability map inversion from PET image-based data.
- The inverted permeability map of large experimental datasets are used to parameterize forward numerical models to validate CNN predictions.

Corresponding author: Christopher Zahasky, `czahasky@wisc.edu`

Abstract

Quantification of heterogeneous multiscale permeability in geologic porous media is key for understanding and predicting flow and transport processes in the subsurface. Recent utilization of in situ imaging, specifically positron emission tomography (PET), enables the measurement of three-dimensional (3-D) time-lapse radiotracer solute transport in geologic media. However, accurate and computationally efficient characterization of the permeability distribution that controls the solute transport process remains challenging. Leveraging the relationship between local permeability variation and solute advection rates, an encoder-decoder based convolutional neural network (CNN) is implemented as a permeability inversion scheme using a single PET scan of a radiotracer pulse injection experiment as input. The CNN consists of Densely Connected Neural Networks that can accurately capture the 3-D spatial correlation between the permeability and the radiotracer solute arrival time difference maps in geologic cores. We first test the inversion accuracy using 500 synthetic test datasets. We then use a suite of experimental PET imaging datasets acquired on four different geologic cores. The network-inverted permeability maps from the geologic cores are used to parameterize forward numerical models that are directly compared with the experimental PET imaging datasets. The results indicate that a single trained network can generate robust, denoised 3-D permeability inversion maps in seconds. Numerical models parameterized with these permeability maps closely capture the experimental solute arrival time behavior. This approach presents an unprecedented improvement for efficiently characterizing multiscale permeability heterogeneity in complex geologic materials.

1 Keywords

Convolutional Neural Network, Deep Learning, Machine Learning, Permeability Inversion, Positron Emission Tomography, X-ray Computed Tomography

2 Plain Language Summary

The first step in understanding how water and contaminants are flowing in the subsurface is to describe the ease at which fluid can flow–this property is termed permeability. Variation in permeability is an intrinsic property of geologic materials that arises due to differences in the underlying geologic processes that generated the materials. The use of medical imaging techniques in the field of hydrogeology enables scientists to better understand how water and contaminants flow through geologic porous media. This study leverages these imaging techniques combined with recent advances in deep learning to develop a new way for measuring permeability variation in geologic materials. In this study, we use a deep learning network to perform 3-D permeability prediction. This network is first trained on a diverse set of synthetic permeability maps and corresponding mathematical models of fluid flow through these permeability maps. The training is done by guiding the network to identify the characteristics in the flow data that provide insights on permeability distribution. Compared to traditional mathematical modeling approaches, the trained deep learning network significantly reduces the computational cost while accurately predicting the 3-D permeability distributions in real geologic materials.

3 Introduction

Understanding flow and transport in porous media is crucial for understanding complex hydrogeologic systems, designing contaminant remediation strategies, and utilizing subsurface energy resources. To improve the applicability and accuracy of subsurface flow and transport models, 3-D characterization of hydrogeologic properties that govern these processes—such as intrinsic permeability—is required. Despite this necessity, approaches for non-destructive experimental measurement of multi-scale permeability variation in geologic materials remains a critical challenge. Current approaches for measuring spatially variable permeability are experimentally challenging, computationally expensive, and typically rely on sample-specific porosity-permeability or capillary pressure scaling relationships.

Medical, industrial, and synchrotron-based imaging methods applied to problems in the field of hydrogeology have revolutionized our understanding of physical processes from the nanometer to the meter scale (Akin & Kovscek, 2003; Blunt et al., 2013; Armstrong et al., 2014; Crandall et al., 2017; Zahasky et al., 2019). Photon transmission imaging techniques such as X-ray computed tomography (X-ray CT) excel at characterizing materials with different electron densities. As a result, at the micron scale, X-ray CT is ideal for mapping pore geometry and fluid interfaces (Garing et al., 2017; Zahasky et al., 2019; Garfi et al., 2020). At the continuum scale—the scale at which Darcy's Law can be used to describe flow in a porous medium—X-ray CT can map the spatial distribution of fluids of different densities or variations in porosity (Akin & Kovscek, 2003; Vega et al., 2014; Glatz et al., 2016; Minto et al., 2017).

A range of methods have been developed to approximate spatially-variable permeability using X-ray CT measurements of porosity and fluid saturation (Krause et al., 2013; Krause, 2012; Rabinovich, 2017). The approach developed by Krause et al. (2013) utilizes multiphase core-flooding experiments, mercury injection capillary pressure data, and Leverett-J scaling to estimate sub-core permeability variation. This scaling approach has been validated in sandstone rocks that have intra-sample pore size distribution similarity. More commonly, measurements of porosity are implemented directly into empirical relationships (Chilingar, 1964; Chilingarian, 1991) to estimate local permeability. While strong correlations between porosity and permeability often exist in geologic materials, the empirical form of these correlations depends on rock type, extent of lithification, and sedimentological properties of the rock. For instance, in the model of Chilingar (1964), the same porosity in coarse sand could correspond to two different permeability values that differ by 300%. This discrepancy is due to the geology-specific nature of these relations and is difficult to quantify when the composition and lithification of the geologic materials are unknown.

In carbonates, multi-scale heterogeneity often generates large variation in both permeability and porosity distributions within a sample. Previous studies have shown that variance in porosity-permeability relationship increases with decreasing sample volume for carbonate materials (Vik et al., 2013). In many carbonates, a significant portion of inter-particle porosity are characterized as vug—pores larger than the typical grain size (Lucia, 1983). Depending on the connectivity of vugs, the porosity-permeability relationship can vary significantly and thus be challenging to characterize or generalize. For example, the presence of isolated vugs significantly increases the porosity but it does not lead to proportional increase in permeability. Alternatively, permeability is often disproportionately high for inter-connected vugs (Lucia, 1983). These characteristics pose unique challenges to applying traditional experiment-based permeability inversion methods in carbonates.

While the most widely used imaging tool in hydrogeology is X-ray CT, other imaging approaches that can provide complementary dynamic quantification of continuum-scale transport processes—such as positron emission tomography (PET)—are emerging.

Emission tomography methods are used to detect and reconstruct images-based on photons emitted from radiolabeled fluids in otherwise opaque materials. This difference in image acquisition and reconstruction provides complementary approaches for quantifying different properties of solute transport in geologic materials (Zahasky et al., 2020). By radiolabeling and imaging the solutes directly, PET imaging excels at obtaining fast, time-lapse, high signal-to-noise images of solute concentration in geologic materials. This has opened up new opportunities to understand fundamental aspects of flow and transport processes, such as solute tailing driven by diffusion into microporous carbonates (Kurotori et al., 2019), flow path alteration in fractured carbonates (Brattekas & Seright, 2017), herbicide transport in soil columns (Kulenkampff et al., 2018), multiphase flow (Ferno et al., 2015), multi-scale dispersion (Zahasky & Benson, 2018), and the impact of very strong heterogeneity created by structural features such as deformation bands (Romano et al., 2020).

Positron emission tomography generates multiple 3-D solute concentration maps at user defined time steps. A PET image at a single time step often consists of over ten thousand concentration measurements throughout a sediment column or geologic core; an entire PET scan may consist of over a million concentration measurements. These massive time-lapse datasets are the result of the millimeter-scale discretization of PET images, termed voxels. The application of these imaging methods enables the generation of massive volumes of data not typically available from traditional hydrogeologic laboratory or field approaches. These datasets thus provide orders of magnitude more measurements for heterogeneity characterization than even the most heavily instrumented field sites (Mackay et al., 1986; Boggs & Adams, 1992). These image-based observations combined with recently developed deep learning tools provides a unique opportunity to advance understanding of multi-scale transport processes in heterogeneous geologic materials.

Convolution neural networks (CNNs) are a subcategory of deep learning models that are designed for processing data that has grid-like topology to extract multi-scale features from high-dimensional input (Goodfellow et al., 2016). By connecting each convolutional layer with all its subsequent layers, Densely Connected Neural Networks (DenseNet) fully leverage the hierarchical advantages of CNNs by encouraging feature propagation, sharing, and reuse among all the layers (G. Huang et al., 2017). To further solve the vanishing-gradient problem for the gradient-based learning methods, while diversifying the learned features, a residual-in-residual structure can be applied to all the DenseNet blocks (Wang et al., 2018; Zhang et al., 2018). Built from the residual-in-residual dense block, the architecture of the encoder–decoder based CNN is defined by hyperparameters such as kernel size, stride, padding, and the number and growth rate of layers. Once the model architecture has been defined, the model is then trained—a process requiring additional hyperparameters such as batch size, learning rate, and optimizer selection—to learn the relationship between the input data space (e.g. imaging data) and desired model output data space (e.g. permeability). Using a subset of the input data, termed the training dataset, the network predictions are compared against the training targets through loss functions. The loss is minimized by back propagating and updating the network weights using a different subset of input data, termed the validation dataset. Finally, an unbiased evaluation of the trained network is performed on a third subset of data, termed the test dataset.

In recent years, CNNs have been applied across a range of hydrogeologic applications including parametrizing hydrogeological properties in highly complex digital rock images (Sudakov et al., 2019; Tian et al., 2020; Kamrava et al., 2021), groundwater inventory maps (Panahi et al., 2020), and synthetic hydrogeological parameter maps (Canchumuni et al., 2019; Mo et al., 2019c). A deep dense convolution encoder-decoder network was developed (Zhu & Zabaras, 2018) and expanded (Mo et al., 2019a, 2019b; Zhong et al., 2019; Tang et al., 2021; Wen et al., 2021) to provide a surrogate model to replace full-physics forward models. These methods have successfully replicated forward model results with dramatic reductions in computational cost, but have not been applied directly to sample-specific permeability inversion tasks. At the pore scale, CNNs have been used to determine the average permeability or dispersion of a geologic sample from a pore-scale digital rock image (Sudakov et al., 2019; Tian et al., 2020; Kamrava et al., 2021). These digital workflows are a promising avenue for experiment-free parameterization of flow and transport properties in geologic materials; however, they require repeated discrete analysis to characterize permeability spatial variation at the continuum scale.

In this study, we first trained an encoder–decoder based CNN to determine the 3-D permeability map of geologic core samples based on PET imaging-derived solute transport data. This approach of using a CNN for parameter inversion is fundamentally different from traditional geostatistical inversion approaches because rather than iterating a simulation model to fit a specific geologic sample, the encoder–decoder based CNN is trained to estimate the permeability of any geologic sample within the parameter space represented by the training data. The model was trained and tested on a large synthetically generated dataset and then further tested with PET imaging datasets from one sandstone and three carbonate rock cores. A second CNN was then constructed that utilizes X-ray CT data as an additional input channel to determine the value of rock structure information in predicting 3-D permeability. Predicted permeability maps from the trained network were fed into a forward flow and transport numerical model. These modeled solute transport data were then directly compared with the experimental measurements to validate the applicability of a single trained CNN for permeability inversion using image-based datasets in sedimentary rocks.

4 Methods

4.1 Experimental Positron Emission Tomography Data Acquisition

Four different geologic cores with a range of lithologies and permeability structures were used to provide robust experimental datasets to test the encoder–decoder based CNN inversion algorithm. The samples include a laminated Berea sandstone (Zahasky & Benson, 2018, 2019), an Indiana limestone, an Edwards Brown limestone (Kurotori et al., 2020), and a Ketton limestone (Kurotori et al., 2019, 2020). All of the samples are 5.04 cm in diameter, between 10–10.3 cm long, and have a core-average permeability between 23 mD and 1920 mD. See the referenced studies and Table S1 in the Supporting Information for additional details of the core sample properties.

A detailed description of the PET data acquisition, imaging system, and experimental platform can be found in Zahasky and Benson (2018) or Zahasky et al. (2019). Briefly, the cores were loaded into a flow-through coreholder that enabled the application of confining pressure and thus no-flow boundary conditions on the cylindrical faces of the samples. Samples were saturated with water by first flushing the sample with low pressure CO_2 and then injecting water into the inlet face of the sample while applying a backpressure at the outlet face to prevent gravity-driven desaturation. The differential pressure was monitored, and steady state conditions were determined to have been reached when the differential pressure stabilized. All of the presented experiments were performed at a flow rate of 2 mL/min.

To begin the imaging experiments, a positron-emitting radiotracer—Fludeoxyglucose (^{18}F-FDG)—was diluted in water to reach the optimal radioactivity concentration for minimizing imaging noise (Zahasky et al., 2019). Fludeoxyglucose is a commercially available conservative tracer with a half-life of 109.7 minutes. The PET scans were performed using a Siemens pre-clinical Inveon DPET scanner. Once a scan was started, pulses of radiotracer—between 0.02–0.10 pore volumes—were injected into the samples and displaced with water containing no ^{18}F-FDG. Images of the radiotracer distributions at two different times in the four rock cores are illustrated in Figure 1. This figure highlights the significant variation in transport behavior and the multiscale permeability heterogeneity present in each of the cores used in this study.

4.2 Arrival Time Analysis

Arrival time analysis was used to efficiently summarize the impact of spatial permeability variation on radiotracer transport while reducing the time-lapse experimental PET datasets from four dimensions (x, y, z, t) to three dimensions (x, y, z). This dimension reduction was performed by calculating the quantile arrival time for every voxel in the core.

$$Q(\tau) = \frac{\int_0^\tau C_i(t)dt}{\int_0^\infty C_i(t)dt} \tag{1}$$

Here $C_i(t)$ is the concentration of voxel i within a reconstructed 3-D PET image as a function of time (t) and τ is the time when $Q(\tau)$ reaches the quantile q. The 0.5 quantile, corresponding to the time when half of the solute has passed through the voxel, was used in this study. Using the discrete form of Equation 1, the arrival time were calculated for every voxel location in the imaged sample. The quantiles were calculated based on the numerical interpolation and integration of the breakthrough curve in every voxel in the core samples. An example 3-D arrival time map is illustrated for the Berea sandstone sample in Figure 2.

In addition to dimension reduction, utilization of quantile-based arrival time rather than the time-lapse radiotracer concentration data has several key advantages for inversion applications. First, arrival time values are independent of solute pulse volume and initial concentration, enabling the comparison of experiments with different pulse volumes and different starting concentrations. Second, the arrival time is insulated from variations in hydrodynamic and numerical dispersion. This is particularly important for

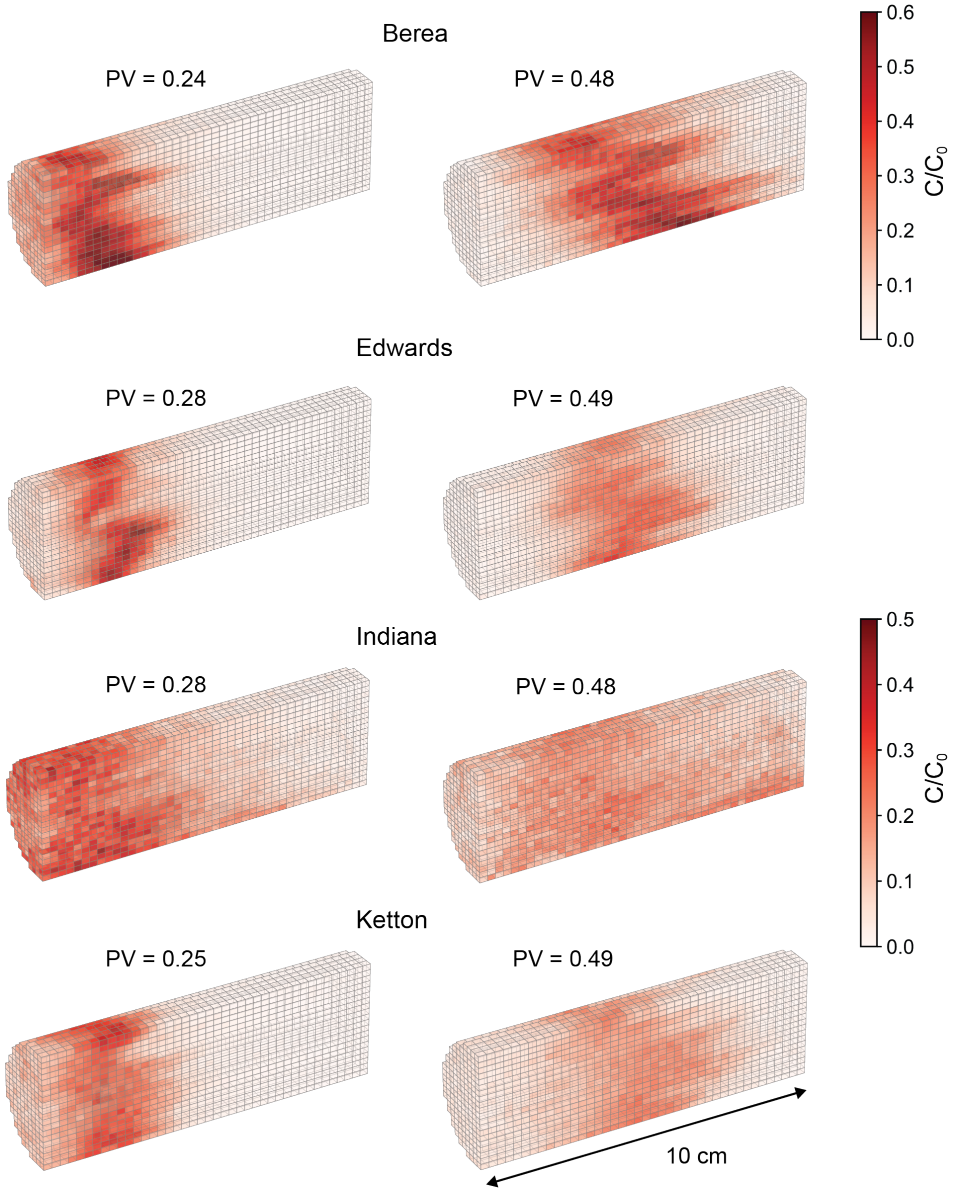

Figure 1: Example PET imaging time frames from each of the four cores used in this study. The pore volumes injected (PV) is indicated for each image and is referenced from the start of tracer injection. Note that the top sandstone core has a slightly larger colorbar scale because the pulse volume of tracer injected was 4 mL as opposed to the three limestone cores that had a pulse volume of 2 mL. The voxel size dimensions for all models are 0.2329 cm $\times$0.2329 cm $\times$0.2388 cm. These images highlight the local sub-core permeability heterogeneity present in all four cores.

the generation of numerical-generated neural network training data as it allows for comparison with experimental data without knowledge of experimental dispersion behavior and without needing to account for the potential impacts of numerical dispersion. Third, application of the quantile-based arrival time is especially advantageous when working with experimental data because the integration of the breakthrough curves averages out much of the imaging measurement error (Harvey & Gorelick, 1995). Furthermore, the quantile-based arrival time is less susceptible to solute tailing and background measurement noise than the normalized first moment because the first moment is a time-weighted integration of the voxel breakthrough curves. An example of this comparison for two different voxels of the PET data in the Berea sandstone and the Ketton limestone is shown in Figure S1 in the Supporting Information.

Computer vision tasks benefit from shared underlying structure (Isola et al., 2017; Zhu & Zabaras, 2018). However, the calculated arrival times include the underlying linear trend due to the flow from the inlet to the outlet of the samples. This linear trend can mask arrival time variation and is fundamentally different from the underlying permeability structure of the samples as illustrated in the upper left plot in Figure 2. To increase the structural similarity and amplify the signal of subtle differences in arrival times, the arrival time was first normalized to nondimensional units of pore volumes injected (upper right image in Figure 2). The nondimensionalized data was then subtracted from the linear trend, resulting in what we call an arrival time difference map as shown in the bottom plot of Figure 2. This representation of arrival times more closely reflects the underlying permeability structure. Greener voxels in Figure 2 have arrival times faster than the core average as a result of higher permeability zones. Pinker voxels in Figure 2 have arrival times slower than the core average, thus are likely corresponding to regions of lower permeability. These arrival time difference maps were used as input for the CNN inversion workflow.

4.3 Experimental Porosity Map Calculation

The traditional approach for measuring porosity maps in geologic materials is to use X-ray CT (Akin & Kovscek, 2003). The 3-D porosity map (Φ) is calculated via the linear scaling expression in Equation 2. This scaling requires a scan of the sample when it is dry (X_a), and a second scan when the sample is fully saturated with water (X_w). The difference between these scans is then scaled by using the difference between pure

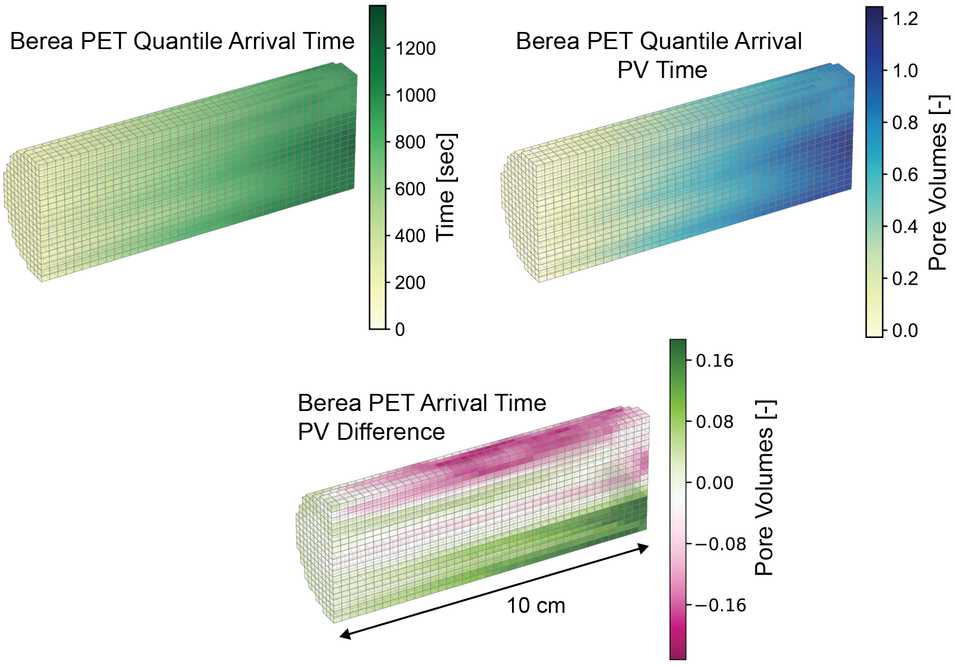

Figure 2: (Upper left) Quantile (0.5) arrival time map collected in the Berea core using the PET data illustrated in the top of Figure 1. (Upper right) Quantile arrival time map in normalized units of pore volumes of water injected since the start of tracer injection. (Bottom) Quantile arrival time difference map in units of pore volumes.

air and water phase Hounsfield X-ray CT numbers ($\Delta_{a,w} = 1000$). An illustration of the porosity in the Berea sandstone calculated with Equation 2 is illustrated in the left plot of Figure 3.

$$\Phi = \frac{X_w - X_a}{\Delta_{a,w}} \tag{2}$$

For application to permeability inversion with a neural network, it is the spatial structure of the porosity map—as opposed to the actual values of porosity—that may provide information to improve the 3-D permeability map prediction. The true values of porosity may not be useful because the network was trained on datasets that lack a specific porosity-permeability relationship, as will be described in the following section. Therefore, the inversion workflow was also tested using a single dry X-ray CT scan, where the Hounsfield values have been scaled to a typical porosity range. This simplification has the advantage of reduced scanning costs and experimental data collection times. In addition, a single or average set of dry scans can also have less measurement noise due to the lack of registration errors that may arise when collecting X-ray CT scans over the course of an experiment. The numerical subtraction of CT data in Equation 2 leads to an amplification of these potential registration errors. Furthermore, since the density of dry air is much less than water, a dry X-ray CT scan provides a higher contrast between the pore spaces and geologic material; thus, highlight the spatial structure of the porosity map. However, a risk of using scaled X-ray CT scans is that they are more susceptible to X-ray CT imaging artifacts such as beam hardening that are reduced or removed during porosity linear scaling calculations (Akin & Kovscek, 2003). In addition, the lack of measured porosity when using scaled X-ray CT maps requires the use of core-average porosity for numerical model parameterization.

To test the network with single X-ray CT scan data, dry scans were normalized and then scaled to have a range from 0.15-0.25 using Equation 3, similar to typical porosity ranges in consolidated rocks.

$$\tilde{\Phi} = 0.10 \cdot \|X_a\| + 0.15 \tag{3}$$

An illustration of the rescaled dry X-ray CT scan in the Berea sandstone calculated with Equation 3 is illustrated in the right plot in Figure 3. All PET and X-ray CT datasets described in this study are provided in the repository referenced in the Acknowledgments.

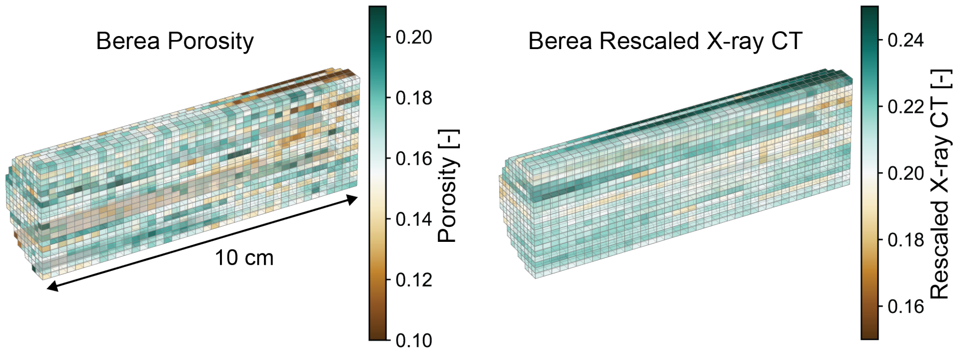

Figure 3: (Left) Porosity map of Berea sandstone calculated using linear scaling with Equation 2. (Right) Air-saturated X-ray CT scan of Berea sandstone scaled to a typical porosity range using Equation 3.

4.4 Synthetic Training Dataset Generation

Two different synthetic datasets were generated to train and test the neural network for 3-D permeability inversion from image-based datasets. The first dataset is composed of arrival time difference maps calculated from numerical solute transport simulations on synthetically generated permeability maps with homogeneous porosity. The second dataset is composed of arrival time difference maps with the same synthetically generated permeability but with the addition of a corresponding heterogeneous porosity map.

4.4.1 Training Dataset Without Porosity

Permeability maps were generated using the exponential covariance random field generation algorithm and open source Python codes from Müller and Schüler (2021). Latin hypercube sampling (Deutsch & Deutsch, 2012; Tartakovsky et al., 2020) was used to generate 26,000 permeability maps that varied in mean permeability from 10 mD–20 D; $\log_{10}$ standard deviation from -1.7–9.9 mD; spatial correlation length from 0.25–12.5 cm in the x, y, and z directions; rotation from 0 to 90 degrees in each of the x, y, and z planes; and 0–2 dummy slices added to the model inlet face. This range of training dataset properties spans the range of consolidated and unconsolidated geologic materials that are typically found in unfractured aquifers and conventional reservoirs.

The solute arrival time in all grid cells was determined by running numerical steady state flow simulations on the synthetic 3-D permeability maps using MODFLOW 2005 (Harbaugh, 2005) and MT3DMS (Bedekar et al., 2016) scripted in FloPy (Bakker et al., 2016). To mimic the experimental settings, the flow simulation was done on synthetic cylindrical cores with a radius of 2.5 cm and length of 10 cm. To replicate this cylindrical shape with a no-flow boundary, permeability and porosity values outside the cylindrical profile were set to zero. The flow rate was set to 2 mL/min and back-pressure was assigned to 70 kPa for simulating the fluid pressure condition below the water table. The simulated 3-D permeability and arrival time difference maps were all represented with dimensions of $20\times20\times40$, which was nearly the same as the dimension of the 3-D PET arrival time images obtained from experiments discussed in Section 4.1. The grid cells for all models have dimensions of 0.233 cm $\times$0.233 cm$\times$0.25 cm. Dummy slices were added at the inlet and outlet of the model to replicate the conditions of the coreholder faces. The width of the dummy slices was varied randomly in the training data to reflect the imperfect inlet solute boundary conditions that occur during the experiments. The width was varied by adding up to three 0.25 cm slices. The strength of these boundary effects has been observed in other in situ transport imaging experiments and is difficult to predict a priori (Lehoux et al., 2016). The solute transport model results were used to calculate 3-D arrival time maps using the same quantile calculation, pore volume normalization, and differencing procedure described in Section 4.2.

Experimental PET data contain Gaussian distributed noise due to the measurement and reconstruction errors (Zahasky et al., 2020). This noise varies between experiments depending on background radiation in the scanner room, instrument error, and the number of coincidence detection events used in a given image reconstruction—as determined by time step size and quantity of positron-emitting radiotracer in the scanner. To replicate this noise in the training data, all of the simulated arrival time difference maps were corrupted with Gaussian white noise prior to loading into the neural network. To account for variation in dataset noise while ensuring that all datasets experience some noise, the noise applied to the input arrival time difference maps was assigned with a Gaussian distribution. The distribution had a mean of zero and a standard deviation that was scaled to 1/70 of the arrival time range for each training set. This value was determined both from quantification of numerical measurement error and hyperparameter tuning during network training.

An additional physical constraint available from routine experimental measurements is the sample average permeability. For each training dataset, the average permeability of each core ($\bar{k}$) was numerically calculated using Darcy's Law solved for $\bar{k}$.

$$\bar{k} = \frac{Q}{A} \cdot \mu \cdot \frac{L}{\Delta P} \tag{4}$$

The flow rate (Q) through the synthetic core was set equal to the model flow rate of 2 mL/min. The cross-sectional area A was based on the modeled core cross-sectional area and the length of the model core L was 10 cm, nearly identical to the experimental datasets. The variable μ is the viscosity of water and ΔP was the pressure drop calculated by subtracting the average pore pressure at the outlet slice minus the average pore pressure at the inlet slice in the steady state MODFLOW model. The calculated average permeability of the core was then represented by a 20×20 tensor padding at the left boundary of the simulated arrival time difference map. The final dimension of every input dataset was then 20×20×41. Adding the average permeability as a boundary condition to the inversion process is key to preserving the uniqueness of the arrival time difference-permeability relationship.

4.4.2 Training Dataset With Porosity

A second training dataset was constructed to explore the impact of porosity heterogeneity and porosity structure information on permeability inversion in geologic cores. There are two potential advantages to incorporating porosity as an additional input. First, geometric information associated with porosity map in geologic cores can be accurately characterized through X-ray CT (Akin & Kovscek, 2003; Vega et al., 2014; Glatz et al., 2016; Minto et al., 2017). Second, core-averaged porosity has been shown to have a geometric correlation with permeability (Chilingar, 1964; Chilingarian, 1991). By using both the normalized solute arrival time and porosity maps as the inputs for the inversion process, this second network aimed to improve the accuracy of permeability map inversion by gaining insights on the geometric distribution heterogeneity in the core. In this dataset, the same permeability training data realizations as the first training set were used but synthetic 3-D porosity maps corresponding to each permeability map were added as an additional input channel. The porosity-permeability relationship was varied with each training data realization because porosity-permeability correlations vary across geologic settings.

The synthetic porosity maps were generated based on the corresponding permeability map utilizing an empirical porosity-permeability function given by Equation 5.

$$\phi_n = \frac{\frac{\ln(k_n)}{a} + b}{100} \tag{5}$$

Here ϕ_n is the porosity of a given grid cell in training set n, k_n is the permeability in milliDarcy of a given grid cell in training set n, a is a constant ranging from 0.25–1, and b is another constant ranging from 5–20. These empirical parameters varied with each training set realization and were sampled by including them with the Latin hypercube sample of the permeability map characteristics (e.g. mean, standard deviation spatial correlation length, etc.). Varying the constants a and b in each training realization enables the generation of a porosity map corresponding to a wide range of sedimentary rocks types (Chilingar, 1964; Chilingarian, 1991). An illustration of the variation in porosity-permeability relationships is illustrated in Figure S2 in the Supporting Information by plotting the porosity-permeability relationship of all 500 test set realizations. Each synthetic porosity map was then concatenated to its corresponding arrival time map as an additional input channel. To maintain consistent input channel sizes, the average permeability of the core ($\bar{k}$) was also padded at the left boundary of the 3-D porosity data resulting in a dimension of 20×20×41. Two different randomly selected training datasets generated with the above workflow are illustrated in 3-D plots in Figure S3 in the Supporting Information. The Python codes used for training data generation and the full compilation of training data are available in the data repository cited in the Acknowledgements.

4.5 Network Construction and Training

4.5.1 Convolutional Neural Network

Convolutional neural networks (CNNs) are used to analyze, interpret, or classify image-based data. A convolutional layer contains a sequence of filters/kernel, each representing an abstract feature of the input image channels. A convolutional layer extracts features from input images through: $\boldsymbol{x}^{(l+1)} = f_{l+1}(\boldsymbol{W}^{(l+1)}\boldsymbol{x}^{(l)} + \boldsymbol{b}^{(l+1)})$, where $\boldsymbol{W}^{(l+1)}$ is the weight matrix (or kernel), $\boldsymbol{b}^{(l+1)}$ is the bias vector, and f_{l+1} is the nonlinear activation function that maps the input map $\boldsymbol{x}^{(l)}$ to a corresponding output map $\boldsymbol{x}^{(l+1)}$. In a convolutional layer, every neuron is linked to a receptive field, a region in the input that represents a particular feature. As the number of connected convolutional layers increases, the input spatial information gets selected and refined through encoding. The accumulated receptive fields of shallower (or earlier) layers makes the region exposed to the neurons in the deeper (or later) layers larger. This enables CNNs to capture smaller scale features in the shallower layers and the more global information in the deeper layers (Gu et al., 2018). For the networks in this study, 3-D convolutional layers were utilized, allowing the network to learn the 3-D spatial correlations within and among feature maps.

4.5.2 Residual-in-residual Dense Network

The number of parameters in a network increases as a network grows deeper, theoretically improving the performance of the network. However, gradients among parameters experience loss during the back-propagation process due to repeated multiplication, and the loss generally increases as the networks get deeper. To solve the gradient-vanishing problem, Densely Connected Neural Networks (DenseNet) were developed to connect all layers—with matching feature map sizes—directly with all their subsequent layers (G. Huang et al., 2017). The direct connections are established by using the outputs of all preceding layers as the inputs of the current layer, so the current layer can obtain and concatenate all the preceding input feature maps and then generate its own feature maps to all subsequent layers (G. Huang et al., 2017). The growth rate of a dense block refers the number of new feature maps concatenated at each layer. In addition to alleviating the gradient-vanishing problem, the densely connected structure also strengthens feature propagation and reuse, further reducing the parameters of the networks (G. Huang et al., 2017). In a dense block, after receiving the concatenated feature maps as inputs, each layer carries out the batch normalization (BN) (Ioffe & Szegedy, 2015) and the ReLU (Rectified Linear Unit: $\mathrm{ReLU}(x) = \max(0, x)$) nonlinear activation. Finally, the main features of the activated prediction are captured by a convolution layer and then passed to all subsequent layers.

To further increase the depth of the networks without the gradient-vanishing or gradient-exploding problem, a residual learning framework (He et al., 2016) was adopted to connect the dense blocks in the networks (Zhang et al., 2018). Residual-in-residual dense block (RRDB) has been successfully applied in image super-resolution (Wang et al., 2018) and geologic features parameterization (Mo et al., 2019c). Based on these previous models, the networks built here contain a residual dense block that consisted of five dense blocks with each RRDB contained three residual dense blocks. The growth rate of the dense block was set to 48 and the residual scaling factor β was set to 0.2 (Wang et al., 2018; Mo et al., 2019c). An illustration and additional descriptions of the components in each residual-in-residual dense block is given in Figure S4 in the Supporting Information.

4.5.3 Network Architecture

The 3-D encoder–decoder based CNN extracts high-level features of the input and output data through the convolutional blocks and refines the extracted features through the residual-in-residual dense blocks. A detailed illustration of the overall network is presented in the upper portion of Figure 4. The convolutional block consists of a single 3-D convolutional layer—indicated by blue blocks in Figure 4. The residual-in-residual dense block consists of fifteen dense blocks— indicated by green blocks in Figure 4. During the training, features selection through compression and reconstruction was achieved through the pooling and up–sampling blocks—yellow blocks in Figure 4. Each pooling block halved the dimension of the input feature maps through a combination of batch normalization (BN), ReLU activation, and average pooling layers. Each up-sampling block doubled the dimension of the input feature maps through a combination of batch normalization, ReLU activation, and Conv-Transpose layers. In total, the entire network contains forty-eight 3-D convolutional layers, two average pooling layers, and two Conv-Transpose layers with a total 8,570,690 trainable parameters. Both the networks trained with and without additional porosity maps have the same architecture with the only difference being the number of input channels. The network without porosity has one input channel and the network that takes into account porosity has two input channels.

4.5.4 Network Training

For the network trained with homogeneous porosity, the network training was a supervised process with 3-D image tensors containing the arrival time difference maps, corrupted with noise as described in Section 4.4, as the inputs and the permeability maps of the corresponding synthetic geologic core as the labeling data. To evaluate how porosity information improves the permeability prediction, a second network was trained with the porosity maps of the synthetic geologic cores as additional inputs. During the training process, the encoder first extracted and parameterized the high-level features of the

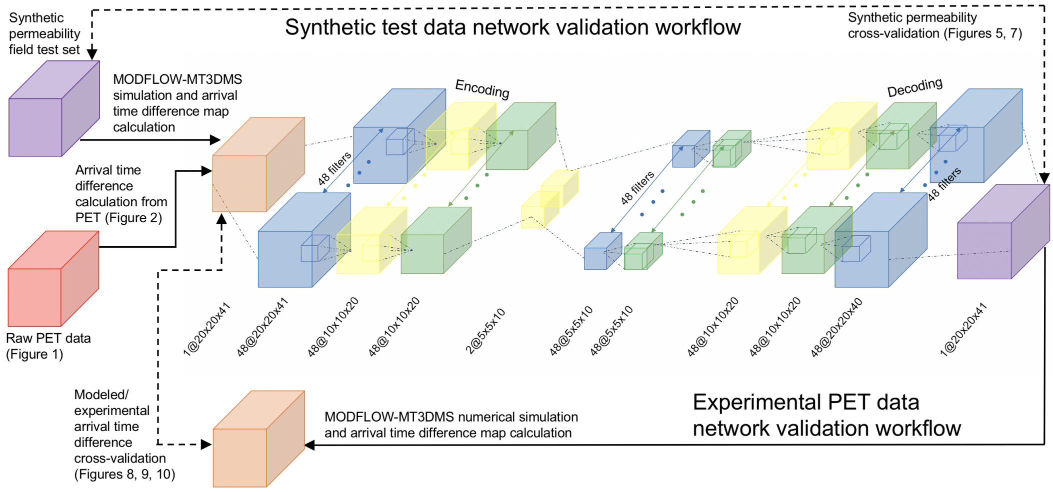

Figure 4: Schematic illustration of the inversion-validation workflow using both synthetic (top loop) and experimental PET data (bottom loop). Figure includes the network's encoding-decoding architecture, MODFLOW-MT3DMS numerical forward flow simulation, and cross-validation. The purple blocks correspond to synthetic/predicted permeability maps, the red block is the PET data, the orange blocks are experimental and modeled arrival time difference maps. The CNN components include convolutional blocks (blue), up/down-sampling block (yellow), and residual-in-residual dense blocks (green).

input data. The compressed high-level features map, referred to as the latent space, having a dimension of 5×5×10. The decoder then constructed the labeling permeability based on the extracted high-level features in the latent space. The predicted permeability maps by the decoder had a final dimension of 20×20 ×40, the same as the dimension of the labeling synthetic permeability maps. The predicted permeability maps were then compared with the labeling synthetic permeability maps through loss functions. The loss function used in this study was a combination of L1 loss (Equation 6) and KL-Divergence loss (Equation 7). L1 loss measures the absolute distance between the labeling ($p(x)$) and predicted ($q(x)$) permeability maps.

$$D(p(x)\|q(x))_{L1} = \sum_{i=1}^{n} |p(x_i) - q(x_i)| \tag{6}$$

KL-Divergence loss measures the differences in probability distributions between the labelling and predicted permeability maps in all three dimensions.

$$D(p(x)\|q(x))_{KL} = \sum_{i=1}^{n} p(x_i) \cdot \log\left(\frac{p(x_i)}{q(x_i)}\right) \tag{7}$$

Generally, small loss indicates less difference and large loss indicates less similarly between the ground truth and prediction. The loss propagation was monitored through observing the gradient and minimum of the loss curve for the predictions on synthetic permeability maps in the validation set. To monitor and examine the performance of the network, the total 26,000 numerically simulated data were divided into 20,000 for training, 5,500 for validation, and 500 for test sets.

Adaptive Moment Estimation (Adam) algorithm was adopted to back-propagate the differentiable activation functions through stochastic gradient descent on a series of mini-batches. The purpose of adopting Adam optimizer was to save the memory usage while efficiently propagating the sparse gradients caused by the high complexity of the imagery data (Kingma & Ba, 2014). The initial learning rate for the Adam optimizer was set to 0.005 with a batch size of 32. During the training process, over-fitting, when the validation loss stagnates at a relatively high value while the training loss is still steadily decreasing, was often observed. To address the over-fitting issue, a learning rate scheduler was adopted with a weight decay factor of 0.5 for every plateau or increase in validation loss over 15 epochs. In addition, a 3-D dropout layer (Hinton et al., 2012) was added after the ReLU activation layer in every dense blocks to simulate a sparser activation that further reduce the network's propensity to overfit.

Training accuracy was evaluated on the test set by comparing the synthetic permeability maps with the network predicted permeability maps. For the experimental data, the network was evaluated by comparing the experimental arrival time difference maps and the numerically simulated arrival time difference maps based on the network permeability map prediction. The root-mean-squared error (RMSE in Equation 8) and coefficient of determination (R^2 in Equation 9) statistical indicators were used to evaluate the accuracy of permeability predictions.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{n}(y_i - y_i^*)^2} \tag{8}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - y_i^*)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{9}$$

Here N is the number of voxels in a core, y_i and y_i^* are the real and predicted value in each voxel, respectively. The variable $\bar{y}_i$ is the core-averaged real value.

Each training run of 300 epochs generally took 26 to 36 hours to complete on a Nvidia GeForce GTX980 GPU at the University of Wisconsin-Madison Center for High Throughput Computing (CHTC), and the trained parameters for the network were stored in a path file after the training. The Python codes for the neural networks used in this study and the trained models are available in the repository referenced in the Acknowledgements.

4.5.5 PET Data Inversion-Validation Workflow

After the encoder–decoder based CNNs were fully trained, a set of experimental 3-D arrival time difference maps obtained from the PET imaging methods discussed in Sections 4.1 and 4.2 were used to generate permeability map predictions. In the second network, both 3-D porosity maps and scaled X-ray CT scans were tested as additional inputs to the arrival time data. Using the algorithms discussed in Section 4.4.1, arrival time difference maps were then generated using the inverted 3-D permeability maps as numerical model input. The modeled arrival time difference maps were then directly compared with the experimental measurements to validate the accuracy of the network permeability map predictions using the experimental data input. An illustration of the overall workflow including permeability inversion, forward numerical flow and transport modeling, and cross validation is presented in Figure 4.

5 Results

5.1 Network Results with Synthetic Test Data and Homogeneous Porosity

The accuracy of the trained encoder–decoder based CNNs was evaluated by comparing the average RMSE accuracy of the 500 3-D permeability predictions in the test dataset. The RMSE accuracy was calculated by comparing the predicted permeability (y^* in Equation 8) with the synthetic permeability maps (y in Equation 8) that were used to generate the arrival time difference input data for the network. The arrival time difference maps in this testing all included Gaussian noise. The grid cell-average RMSE of all of the $\log_{10}$-permeability maps in the test set is 0.057, or 1.1 mD. To illustrate the grid cell-level network performance, fifteen sample permeability map predictions from the test set were randomly chosen and are presented in Figure 5. These grid cell 3-D permeability predictions were directly compared against the corresponding grid cells in the synthetic permeability maps that were used to generate the input arrival time difference maps.

To better describe the overall uncertainty of the network predictions given the wide range of test set realization mean permeability, it is useful to calculate the RMSE relative to the mean of each permeability map. Figure 6 illustrates the relative uncertainty for the 500 test set realizations. For each synthetic permeability map in the test set, the input relative uncertainty was calculated using the average of the added Gaussian random noise divided by the mean arrival time difference; the output relative uncertainty was calculated using the average RMSE prediction accuracy divided by the mean of the synthetic permeability map. The average relative input uncertainty of all of the arrival time difference maps in the test set is 0.063 and the average relative output uncertainty of all of the $\log_{10}$-permeability maps in the test set is 0.032. The relative uncertainty of all the permeability predictions consistently a range between 0.01 to 0.26—lower than the range of relative uncertainties (noise level) for all the network input data (0.025 to 0.35). Figure 6 not only highlights the quality of the parameter inversion by the network but also the ability of the network to denoise the data. The denoising is apparent from the overall lower relative uncertainty in the output permeability predictions compared to the relative uncertainty in the input arrival difference maps. These results highlight

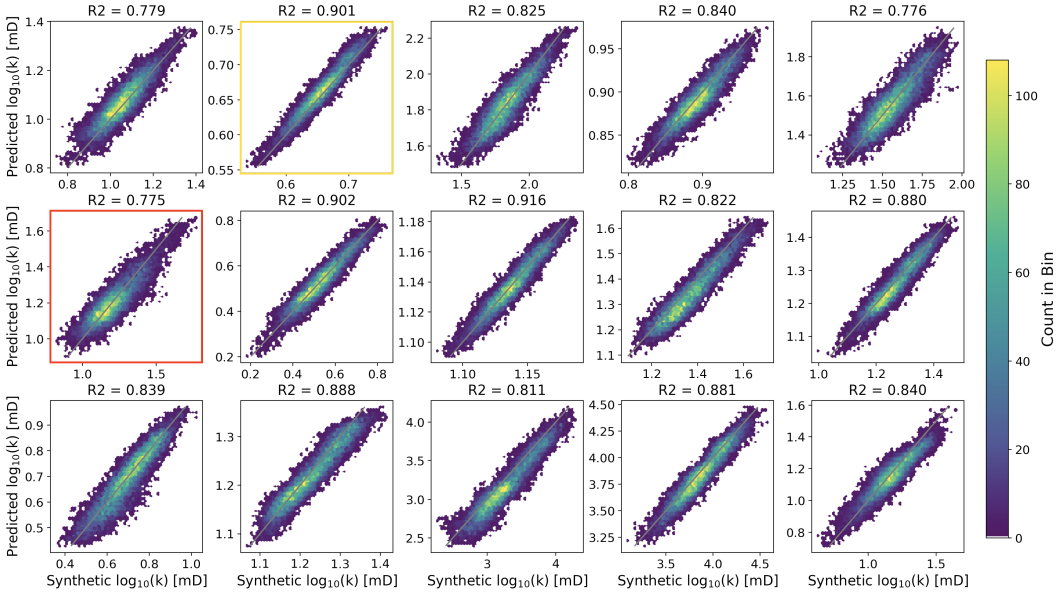

Figure 5: Fifteen randomly chose samples of permeability map prediction using arrival time difference maps from the test set that included Gaussian noise. For each subplot, the x-axis represents the grid cell-level synthetic permeability associated with the test set arrival time, and the y-axis represents the corresponding grid cell-level predicted permeability. To illustrate the density of the correlations, the cross-plots are colored by the number of points in a given bin or local region of the cross-plot. The plots with the gold and red outlines correspond to the top and bottom rows of plots in Figure 7, respectively.

the capability of the network to distinguish useful features in parameter maps from systematic and/or random errors.

An example of two spatially resolved permeability inversion results are plotted in Figure 7. The top plot of Figure 7 provides a 3-D example of 90^{th} percentile permeability map prediction (with a R^2 score of 0.901) and the bottom plot of Figure 7 provides a 3-D example of a 10^{th} percentile permeability map prediction (with a R^2 score of 0.775). Each set includes the arrival time difference map, the predicted permeability map, and the corresponding synthetic permeability map. Based on this multilevel analysis, the trained encoder–decoder based CNN is able to learn the key features of the arrival time difference map and the relationship with the corresponding heterogeneous permeability map.

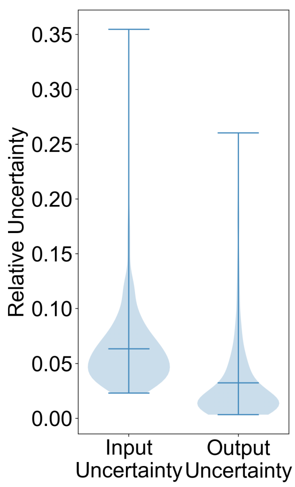

Figure 6: The distribution of relative uncertainty of permeability predictions utilizing the network trained without heterogeneous porosity maps evaluated on the 500 test set realizations. For each realization in the test set, the input relative uncertainty is calculated using the average of the added Gaussian noise divided by the mean arrival time difference. The output relative uncertainty is calculated using the average RMSE prediction accuracy divided by the mean of the permeability map. The average relative input uncertainty of all of the arrival time difference maps in the test set is 0.063 and the average relative output uncertainty of all of the $\log_{10}$-permeability maps in the test set is 0.032.

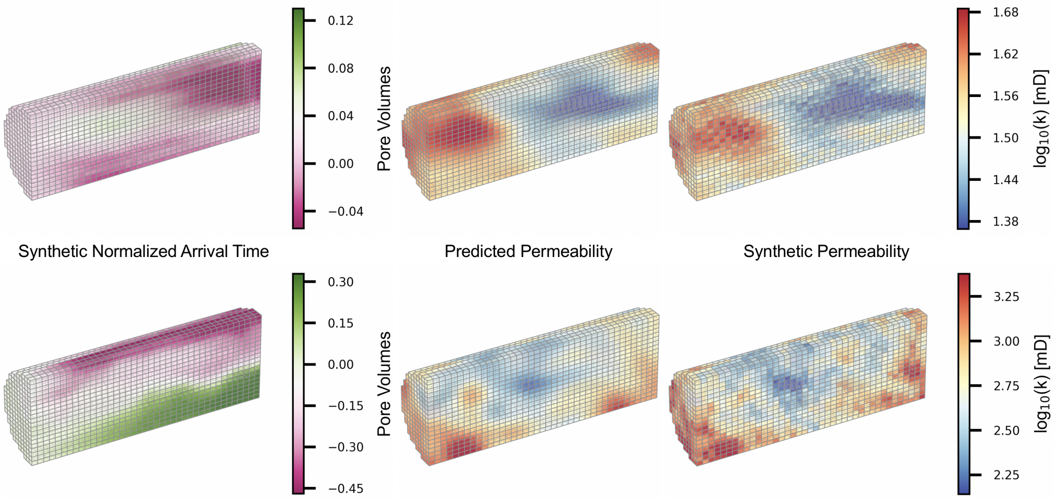

Figure 7: Illustration of two numerically calculated arrival time difference maps using MODFLOW-MT3DMS (left column) based on the corresponding synthetically generated permeability maps (right column). The arrival time difference maps—plotted here without the Gaussian noise—were the input data used to generate the corresponding predicted permeability maps (middle column) by the network trained under with homogeneous porosity. In terms of the R^2 accuracy, the top row corresponds to a 90^{th} percentile quality prediction. This dataset is also shown in the scatter plot in Figure 5 marked with the gold box. The bottom row corresponds to a 10^{th} percentile quality prediction and is given by the scatter plot in Figure 5 marked with the red box. The grid cells for all models are 0.233 cm $\times$0.233 cm $\times$0.25 cm.

5.2 Network Results with Experimental PET Data

Following the network evaluation with synthetic test set data, permeability predictions were generated on the experimental arrival time difference datasets collected from four geologic cores using the PET imaging methods illustrated in Figure 2 and described in Section 4.1. Figure 8 and Figure 9 show the 3-D experimental arrival time difference map calculated from the PET data, the predicted permeability map from the network, and the simulated arrival time difference map based from the MODFLOW-MT3DMS model parameterized with the predicted permeability map. Grid cell-level comparison of the arrival time data is shown in the top row of cross-plots in Figure 10. In Figure 10 the experimental grid cell-level arrival time difference is given on the x-axis and modeled grid cell-level arrival time difference—based on the network permeability map prediction—is given on the y-axis. For the experimental data, the arrival time difference map predictions have an R^2 accuracy ranging from 0.756 (Ketton limestone) to 0.831 (laminated Berea sandstone), verifying the capability and robustness of a single trained network to predict the 3-D permeability map of geologic samples.

5.3 Results of the Model Trained with Heterogeneous Porosity

A second network was trained assuming a spatial correlation between porosity and permeability maps. This was done to test if additional structural information provided by the porosity maps improved the permeability prediction. This network was trained with an additional input channel of the porosity map as described in Section 4.4.2. Similar to the first network trained with homogeneous porosity, the training and validation loss curves of the second network also display a clear downward trend. The training performance of this second network with heterogeneous porosity is slightly better than the first network with homogeneous porosity as illustrated by both the lower training and validation loss, and lower overall testing root mean square error in Figure 11. For this second network, the average RMSE accuracy of all the $\log_{10}$-permeability maps in the test set is 0.047, a slight improvement relative to the network with no porosity data that has an RMSE of 0.057. The improved performance on the synthetic data is attributed to the strong spatial correlations between the synthetic permeability and porosity maps as illustrated in Figure S3 in the Supporting Information.

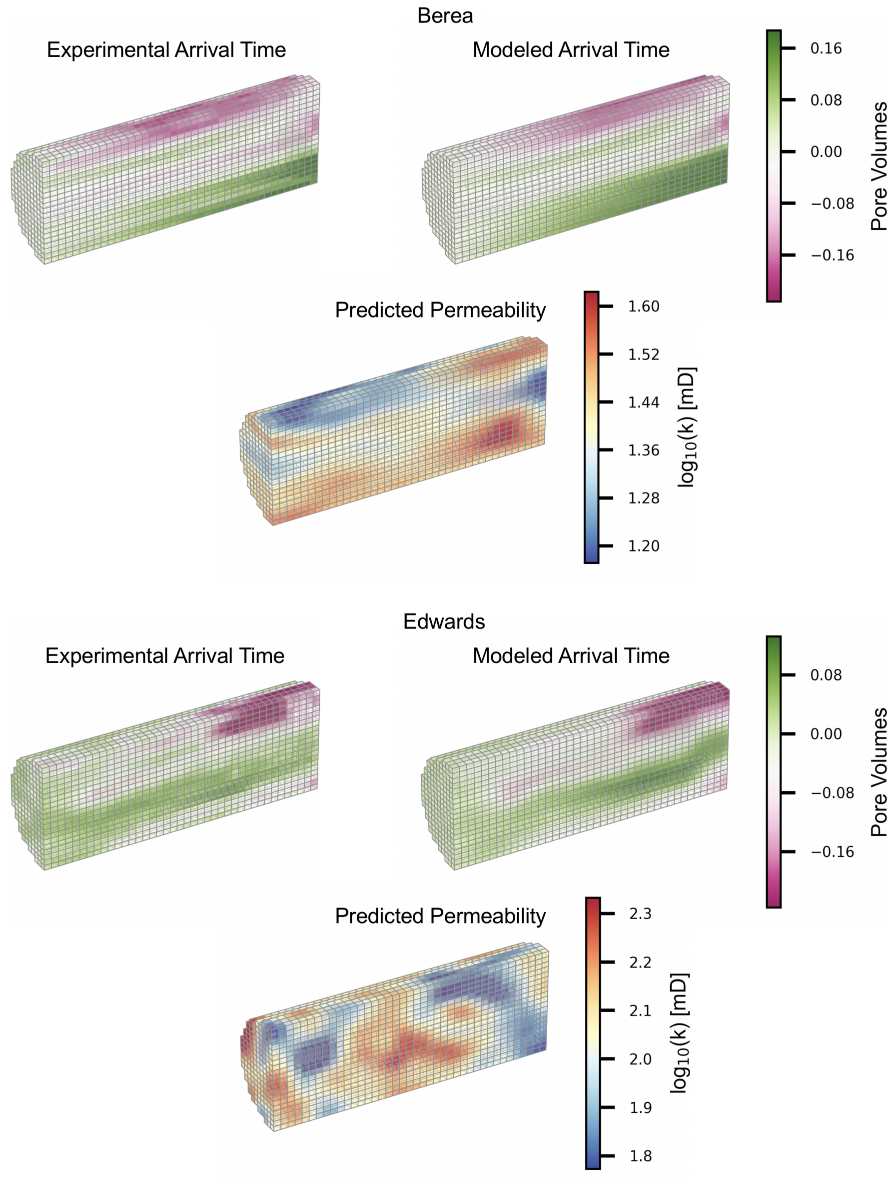

Figure 8: Cross-comparison of the network trained with homogeneous porosity using experimental arrival time difference maps measured with PET on a laminated Berea sandstone (top three subplots) and an Edwards Brown limestone (bottom three subplots). The upper left subplots show the arrival time difference map calculated from the PET imaging data, the lower plot shows the predicted permeability map by the network, and the upper right shows the numerically simulated arrival time difference map based on the predicted permeability map. Note that the experimental and modeled arrival times are plotted on the same colorscale.

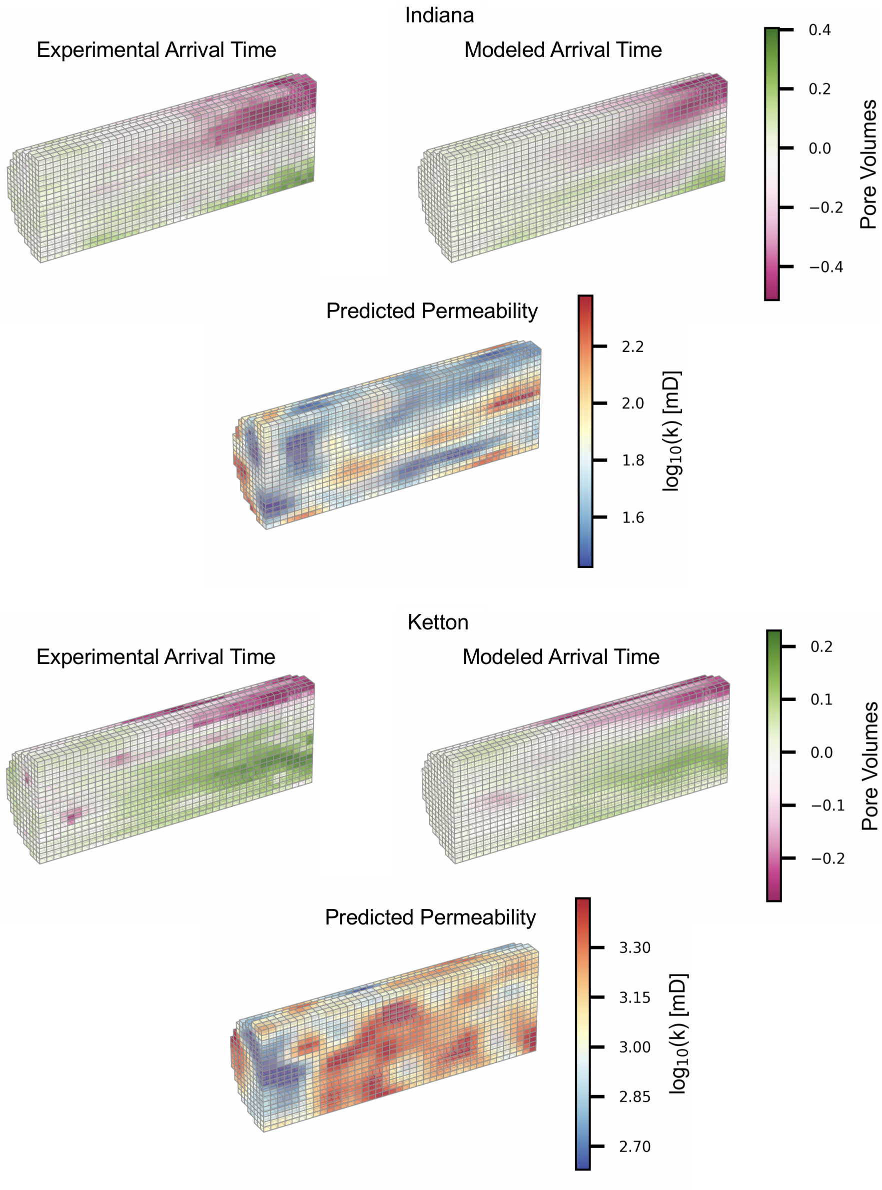

Figure 9: Cross-comparison of the network trained with homogeneous porosity using experimental arrival time difference data measured with PET collected from an Indiana limestone (top three subplots) and a Ketton limestone (bottom three subplots). The upper left subplots show the arrival time difference map calculated from the PET imaging data, the lower plot shows the predicted permeability by the network, and the upper right shows the numerically simulated arrival time difference map based on the predicted permeability map. Note that the experimental and modeled arrival times are plotted on the same colorscale.

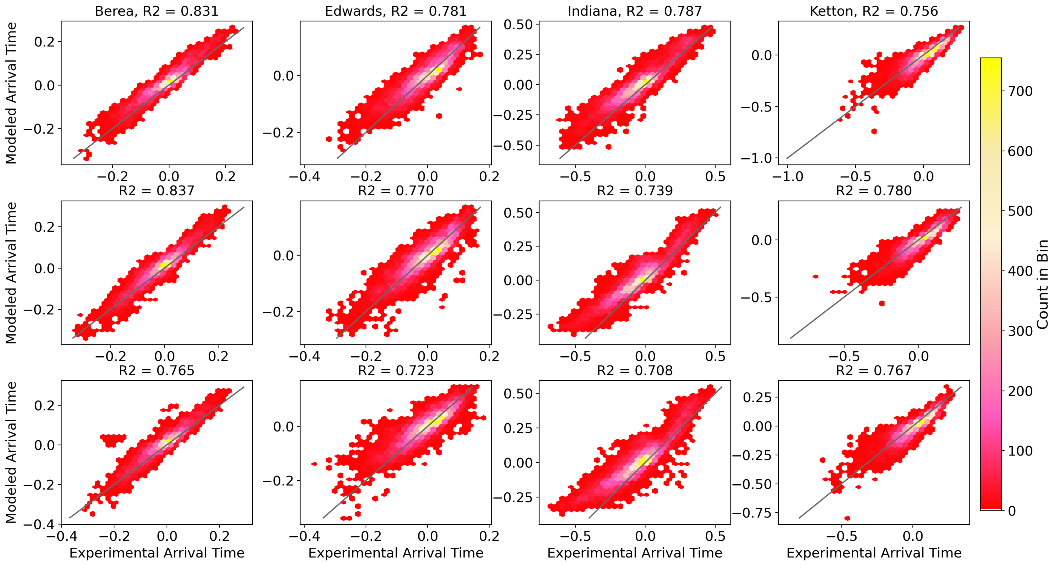

Figure 10: Cross-plot of experimental arrival time difference data (x-axis) and modeled arrival time difference from network permeability map prediction for the four geologic cores (from left to right): Berea sandstone, Edwards Brown limestone, Indiana limestone, and Ketton limestone. The top row of plots show the results using the arrival time difference map as the only network input channel; the middle row of plots show the results using the scaled dry X-ray CT scan as the second input channel; the bottom row of plots show the results using the X-ray CT-measured porosity map as the second input channel. To illustrate the density of the correlations, the cross-plot is colored by the number of points in a given bin or local region of the cross-plot. These results indicate that the additional of X-ray CT-derived data provides very little or no improvement in permeability map prediction.

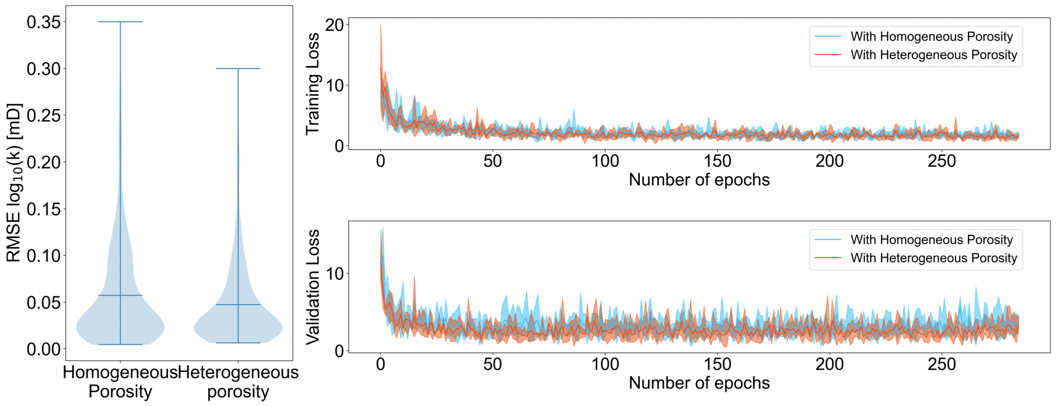

Figure 11: Training and testing performance of both networks trained with and without heterogeneous porosity maps, including training loss (top right plot), validation loss (bottom right plot), and the distribution of the RMSE accuracy of the permeability map predictions based on the test set data (left plot). For the test set, the average RMSE of all of the predicted $\log_{10}$-permeability maps using the network trained without heterogeneous porosity maps is 0.057, and the average RMSE of all of the predicted $\log_{10}$-permeability maps using the network trained with heterogeneous porosity maps is 0.047.

Despite the slightly better performance on synthetic data, a contradictory phenomenon was observed regarding the experimental data. For the four geologic cores and PET datasets presented in Figures 8 and 9, both traditional X-ray CT-measured porosity maps and scaled dry X-ray CT scans were tested as the additional inputs for permeability map prediction (see the full description of this data in Section 4.3). Figure 10 illustrates the results of the modeled arrival time analysis compared against the experimental arrival time measurements using the same experimental cross-comparison process as the previous network. The network trained with heterogeneous porosity maps generally under-performed the network trained with only the arrival time difference data. This is illustrated by the consistent reduction in the R^2 accuracy in the middle and bottom row of plots in Figure 10. The only instances of higher R^2 accuracy relative to the network using only arrival time data are the Ketton core with both scaled X-ray CT data and X-ray CT porosity and the Berea core with scaled X-ray CT data. In all cases the R^2 accuracy improved by less than three percent with the addition of X-ray CT-derived input data.

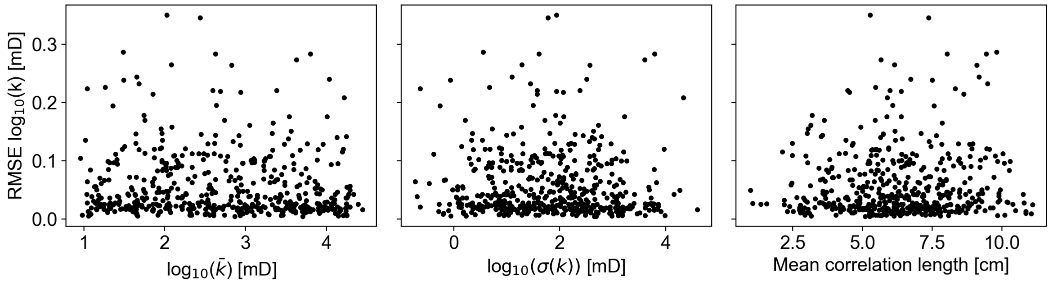

Figure 12: Summary of network RMSE of all test set data plotted against the corresponding mean permeability (left plot), permeability field standard deviation $\sigma(k)$ (center plot), and mean correlation length of the three principle axes pre-rotation (left plot).

6 Discussion

The results illustrate that the network can accurately determine the local patterns and magnitudes of permeability variations from both noisy synthetic and experimentally measured arrival time difference maps. High permeability areas generally have more rapid arrival times and thus more positive arrival time differences whereas low permeability areas generally have slower arrival times and thus more negative arrival time differences. However, in many cases the structure of the permeability variation can distort obvious relationships with arrival times as indicated by Figures 8 and 9.

Statistical analysis of the inversion results summarized in the left plot of Figure 11 indicates that the RMSE of the network predicted permeability relative to the original synthetic permeability field is consistently low across the entire range of 500 test set permeability fields. Analysis of RMSE as a function of mean permeability, permeability field standard deviation, and mean correlation length indicates that there is no correlation between RMSE and permeability field characteristics as illustrated in Figure 12. The lack of correlation between test set RMSE and mean correlation length of the 3-D permeability field indicates that there is minimal feature loss resulting from feature smoothing during the encoding and decoding process. This verifies that using an encoding-decoding network significantly reduces network training computational cost while maintaining the robustness of permeability inversion.

In addition to computational cost, a key challenge of determining the 3-D permeability distribution from 3-D time lapse solute transport measurements is isolating the transport characteristics that are permeability dependent. Convolutional neural networks excel at finding spatial correlations between distinct high frequency features such as contours or edges of distributions. Therefore, it is crucial to minimize the high frequency experimental noise—distinct features that are unrelated with permeability distribution—in the input data. The quantile-based arrival time analysis emphasizes the advective transport that is directly influenced by permeability and minimizes the effects of hydrodynamic and numerical dispersion, experimental imaging noise, variation in initial solute concentration, and solute tailing behavior. While flow rate dependencies are known to exist in complex carbonate materials (Kurotori et al., 2019), the quantile threshold can be adjusted to minimize the influence of these effects on the permeability inversion process. The normalization of the arrival time map is thus able to reduce the influence of experimental conditions such as flow rate and variation in sample dimensions. This preprocessing and dimension reduction using classic transport analysis methods converts the raw 4-D datasets down to a 3-D maps of arrival time information. This constrains the domain of the inversion problem while minimizing the complexity, leading to a more unique and computationally efficient permeability prediction.

Porosity-permeability relationships are likely to exist in structured sedimentary rocks such as sandstones, while these relationships often breakdown in carbonates. The accuracy of the permeability predictions in the second network that included correlations between porosity and permeability was marginally improved in the synthetic data as illustrated in Figure 11. However there was minimal improvement or even worse predictions in the experimental data inversion as illustrated in Figure 10. This highlights the importance of validating deep learning methods on experimental or field data as deep learning model efficacy can be hampered by the intrinsic oversimplification of synthetic training datasets.

The results summarizing the experimental data inversion in Figure 10 generally found higher R^2 scores for the permeability map predictions using scaled X-ray CT scans as opposed to porosity map data. The network using scaled X-ray CT scans as inputs slightly outperformed the results without X-ray CT data for permeability predictions on geologic cores with distinct structural features—such as the clear lamination in the Berea sandstone. However, scaled X-ray CT scans suffer from the same uncertainty in the strength of a single porosity-permeability relationship for a given sample volume. Extensive hyperparameter exploration was performed on the porosity-permeability relationships by adding different levels of noise to the porosity data, thus weakening the underlying porosity-permeability relationships in the training data. Nevertheless, these results indicate that the porosity-permeability framework adopted in this study is likely not universal enough for spanning all geologic materials with a single trained network. Thus, using only PET-derived arrival time difference maps provides the best general performance for 3-D permeability inversion. Moreover, the validation results suggest that the presented method is rigorous because spatial permeability distributions can be accurately predicted from PET datasets alone, without the need to obtain structural information on the geologic cores.

7 Implications

This study demonstrates a new permeability inversion strategy by applying a deep convolutional encoder-decoder neural network—utilizing multilevel residual learning strategy and the dense connection structure—to massive image-based datasets. The network accurately predicts the local patterns and magnitude of the 3-D permeability maps using local arrival time difference maps generated from PET scans and routine mean permeability measurements on four different geologic core samples. Although the initial network training process is computationally intensive, the trained network is able to invert for the permeability map of nearly any unfractured geologic core sample in a matter of seconds. Furthermore, each path file that contains the trained parameters for the entire encoder-decoder network is only tens of megabytes. An equivalent numerical inversion approach would typically require repeated flow and transport simulations on an ensemble of 100's of models to generate a permeability map of a single rock sample.

The orders of magnitude reduction in multiscale permeability inversion time provides an opportunity for a paradigm shift in core scale analysis and characterization methods. This workflow generates an accurate experimentally derived 3-D permeability map of a geologic sample rather than a single sample-average permeability measurement. This type of rapid characterization is key for building more accurate models of subsurface flow and transport processes.

Acknowledgements

Python scripts for training data generation, data analysis, CNN operation, and trained network parameters are permanently available at (Z. Huang & Zahasky, 2021). The full training datasets and experimental data are permanently available at (Z. Huang et al., 2022). This work was supported as part of the Center for Mechanistic Control of Water-Hydrocarbon-Rock Interactions in Unconventional and Tight Oil Formations (CMC-UF), an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science under DOE (BES) Award DE-SC0019165. Further support for this research was provided by the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison with funding from the Wisconsin Alumni Research Foundation and the University of Wisconsin-Madison Hilldale Undergraduate/Faculty Research Fellowship.

References

Akin, S., & Kovscek, A. (2003, 01). Computed tomography in petroleum engineering research. *Geological Society, London, Special Publications*, *215*, 23–38. doi: 10.1144/GSL.SP.2003.215.01.03

Armstrong, R. T., Georgiadis, A., Ott, H., Klemin, D., & Berg, S. (2014). Critical capillary number: Desaturation studied with fast X-ray computed microtomography. *Geophysical Research Letters*, *41*(1), 55–60. doi: 10.1002/2013GL058075

Bakker, M., Post, V., Langevin, C. D., Hughes, J. D., White, J. T., Starn, J. J., & Fienen, M. N. (2016). Scripting MODFLOW Model Development Using Python and FloPy. *Groundwater*, *54*(5), 733–739. doi: 10.1111/gwat.12413

Bedekar, V., Morway, E., Langevin, C., & Tonkin, M. (2016). *MT3D-USGS version 1: A U.S. Geological Survey release of MT3DMS updated with new and expanded transport capabilities for use with MODFLOW* (Tech. Rep.). Retrieved from `http://pubs.er.usgs.gov/publication/tm6A53` doi: 10.3133/tm6A53

Blunt, M. J., Bijeljic, B., Dong, H., Gharbi, O., Iglauer, S., Mostaghimi, P., ... Pentland, C. (2013). Pore-scale imaging and modelling. *Advances in Water Resources*, *51*, 197–216. Retrieved from `http://dx.doi.org/10.1016/j.advwatres.2012.03.003` doi: 10.1016/j.advwatres.2012.03.003

Boggs, J. M., & Adams, E. E. (1992). Field study of dispersion in a heterogeneous aquifer: 4. Investigation of adsorption and sampling bias. *Water Resources Research*, *28*(12), 3325–3336. doi: 10.1029/92WR01759

Brattekas, B., & Seright, R. S. (2017). Implications for improved polymer gel conformance control during low-salinity chase-floods in fractured carbonates. *Journal of Petroleum Science and Engineering*. doi: 10.1016/j.petrol.2017.10.033.This

Canchumuni, S. W. A., Emerick, A., & Pacheco, M. (2019, 04). Towards a robust parameterization for conditioning facies models using deep variational autoencoders and ensemble smoother. *Computers & Geosciences*, *128*. doi: 10.1016/j.cageo.2019.04.006

Chilingar, G. (1964). Relationship between porosity, permeability, and grain-size distribution of sands and sandstones. *Developments in sedimentology*, *1*, 71-75.

Chilingarian, G. (1991). Empirical expression of permeability in terms of other petrophysical properties. In (p. 49-55). Springer Science+Business Media New York. doi: 10.1007/978-1-4899-0617-5_5

Crandall, D., Moore, J., Gill, M., & Stadelman, M. (2017). CT scanning and flow measurements of shale fractures after multiple shearing events. *International Journal of Rock Mechanics and Mining Sciences*, *100*(November 2016), 177–187. Retrieved from `https://doi.org/10.1016/j.ijrmms.2017.10.016` doi: 10.1016/j.ijrmms.2017.10.016

Deutsch, J. L., & Deutsch, C. V. (2012). Latin hypercube sampling with multidimensional uniformity. *Journal of Statistical Planning and Inference*, *142*(3), 763–772. Retrieved from `http://dx.doi.org/10.1016/j.jspi.2011.09.016` doi: 10.1016/j.jspi.2011.09.016

Ferno, M. A., Hauge, L. P., Uno Rognmo, A., Gauteplass, J., Graue, A., Fernø, M. A., … Graue, A. (2015). Flow visualization of CO2 in tight shale formations at reservoir conditions. *Geophysical Research Letters*, *42*(18), 7414–7419. Retrieved from `http://doi.wiley.com/10.1002/2015GL065100` doi: 10.1002/2015GL065100

Garfi, G., John, C. M., Lin, Q., Berg, S., & Krevor, S. (2020). Fluid Surface Coverage Showing the Controls of Rock Mineralogy on the Wetting State. *Geophysical Research Letters*, *47*(8), 1–9. doi: 10.1029/2019gl086380

Garing, C., de Chalendar, J. A., Voltolini, M., Ajo-Franklin, J. B., & Benson, S. M. (2017). Pore-scale capillary pressure analysis using multi-scale X-ray micromotography. *Advances in Water Resources*, *104*, 223–241. Retrieved from `http://dx.doi.org/10.1016/j.advwatres.2017.04.006` doi:

10.1016/j.advwatres.2017.04.006

Glatz, G., Castanier, L., & Kovscek, A. (2016, 09). Visualization and quantification of thermally induced porosity alteration of immature source rock using x-ray computed tomography. *Energy & Fuels*, *30*. doi: 10.1021/acs.energyfuels.6b01430

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., ... Wang, G. (2018). Recent advances in convolutional neural networks. *Pattern Recognition*, *77*, pp 354–377. doi: 10.1016/j.patcog.2017.10.013

Harbaugh, A. W. (2005). *Modflow-2005, the u.s. geological survey modular groundwater model - the groundwater flow process* (Tech. Rep. No. 6-A16). U.S. Geological Survey.

Harvey, C. F., & Gorelick, S. M. (1995). Mapping hydraulic conductivity: Sequential conditioning with measurements of solute arrival time, hydraulic head, and local conductivity. , *31*(7), 1615–1626.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the ieee conference on computer vision and pattern recognition* (Vol. 7, pp. 770–778).

Hinton, G., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint*, *arXiv*.

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *2017 ieee conference on computer vision and pattern recognition (cvpr)* (pp. 2261–2269). doi: 10.1109/CVPR.2017.243

Huang, Z., Kurotori, T., Pini, R., Benson, S. M., & Zahasky, C. (2022). *Dynamic three-dimensional maps of solute concentration and solute arrival times in synthetic and geologic porous media.* Stanford Digital Repository. doi: 10.25740/gz610dt4642

Huang, Z., & Zahasky, C. (2021). *Neural_network_inversion: Public release with data separated (v1.1.1).* Zenodo. doi: 10.5281/zenodo.5644094

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd international conference on machine learning* (Vol. 37, pp. 448–456).

Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, *2017-Janua*, 5967–5976. doi: 10.1109/CVPR.2017.632

Kamrava, S., Im, J., Barros, F. P., & Sahimi, M. (2021). Estimating Dispersion Coefficient in Flow Through Heterogeneous Porous Media by a Deep Convolutional Neural Network. *Geophysical Research Letters*. doi: 10.1029/2021gl094443

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. Retrieved from `http://arxiv.org/abs/1412.6980`

Krause, M. (2012). Modeling and investigation of the influence of capillary heterogeneity on relative permeability. *SPE Annual Technical Conference and Exhibition*(October), 8–10. Retrieved from `https://www.onepetro.org/conference-paper/SPE-160909-STU`

Krause, M., Krevor, S., & Benson, S. (2013, 07). A procedure for the accurate determination of sub-core scale permeability distributions with error quantification. *Transport in Porous Media*, *98*, 565–588. doi: 10.1007/s11242-013-0161-y

Kulenkampff, J., Sto, M., Gründig, M., Manse, A., Lippmann-pipke, J., & Kersten, M. (2018). Time-lapse 3D imaging by positron emission tomography of Cu mobilized in a soil column by the herbicide MCPA. *Scientific Reports*, *8*(7091), 1–9. doi: 10.1038/s41598-018-25413-9

Kurotori, T., Zahasky, C., Benson, S. M., & Pini, R. (2020). Description of Chemical Transport in Laboratory Rock Cores Using the Continuous Random Walk Formalism. *Water Resources Research*, *56*(9), 1–19. doi: 10.1029/2020wr027511

Kurotori, T., Zahasky, C., Hosseinzadeh Hejazi, S. A., Shah, S. M., Benson, S. M., & Pini, R. (2019). Measuring, imaging and modelling solute transport in a microporous limestone. *Chemical Engineering Science*, *196*, 366–383. Retrieved from `https://doi.org/10.1016/j.ces.2018.11.001` doi: 10.1016/j.ces.2018.11.001

Lehoux, A. P., Rodts, S., Faure, P., Michel, E., Courtier-Murias, D., & Coussot, P. (2016). Magnetic resonance imaging measurements evidence weak dispersion in homogeneous porous media. *Physical Review E*, *94*(5), 1–9. doi:

10.1103/PhysRevE.94.053107

Lucia, F. J. (1983, 3). Petrophysical parameters estimated from visual descriptions of carbonate rocks: a field classification of carbonate pore space. *J. Pet. Technol.; (United States)*, *35*(3). Retrieved from https://www.osti.gov/biblio/5887137 doi: 10.2118/10073-PA

Mackay, D. M., Freyberg, D. L., Roberts, P. V., & Cherry, J. A. (1986). A natural gradient experiment on solute transport in a sand aquifer: 1. Approach and overview of plume movement. *Water Resources Research*, *22*(13), 2017–2029. doi: 10.1029/WR022i013p02017

Minto, J., Hingerl, F., Benson, S., & Lunn, R. (2017, 09). X-ray ct and multiphase flow characterization of a 'bio-grouted' sandstone core: The effect of dissolution on seal longevity. *International Journal of Greenhouse Gas Control*, *64*, 152-162. doi: 10.1016/j.ijggc.2017.07.007

Mo, S., Zabaras, N., Shi, X., & Wu, J. (2019a). Deep autoregressive neural networks for high-dimensional inverse problems in groundwater contaminant source identification. *Water Resources Research*, *55*, 3856–3881. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018WR024638 doi: 10.1029/2018WR024638

Mo, S., Zabaras, N., Shi, X., & Wu, J. (2019b). Deep convolutional encoder-decoder networks for uncertainty quantification of dynamic multiphase flow in heterogeneous media. *Water Resources Research*, *55*, 703–728. doi: 10.1029/2018WR023528

Mo, S., Zabaras, N., Shi, X., & Wu, J. (2019c). Integration of Adversarial Autoencoders With Residual Dense Convolutional Networks for Estimation of Non-Gaussian Hydraulic Conductivities. *Water Resources Research*, *56*(2), 1–24. doi: 10.1029/2019WR026082

Müller, S., & Schüler, L. (2021). *GeoStat-Framework/GSTools: v1.3.0 'Pure Pink'.* Zenodo. doi: 10.5281/zenodo.4687075

Panahi, M., Sadhasivam, N., Pourghasemi, H., Rezaie, F., & Lee, S. (2020, 05). Spatial prediction of groundwater potential mapping based on convolutional neural network (cnn) and support vector regression (svr). *Journal of Hydrology*. doi: 10.1016/j.jhydrol.2020.125033

Rabinovich, A. (2017). Estimation of sub-core permeability statistical properties from coreflooding data. *Advances in Water Resources*, *108*, 113–124. Retrieved from `https://doi.org/10.1016/j.advwatres.2017.07.012` doi: 10.1016/j.advwatres.2017.07.012

Romano, C. R., Zahasky, C., Garing, C., Minto, J. M., Benson, S. M., Shipton, Z. K., & Lunn, R. J. (2020). Sub-core scale fluid flow behavior in a sandstone with cataclastic deformation bands. *Water Resources Research*, 1–16. doi: 10.1029/2019wr026715

Sudakov, O., Burnaev, E., & Koroteev, D. (2019, 02). Driving digital rock towards machine learning: predicting permeability with gradient boosting and deep neural networks. *Computers & Geosciences*, *127*. doi: 10.1016/j.cageo.2019.02.002

Tang, M., Ju, X., & Durlofsky, L. (2021). Deep-learning-based coupled flow-geomechanics surrogate model for co_2 sequestration. *Computer Methods in Applied Mechanics and Engineering*, *376*.

Tartakovsky, A., Marrero, C. O., Perdikaris, P., Tartakovsky, G., & Barajas-Solano, D. (2020). Physics-Informed Deep Neural Networks for Learning Parameters and Constitutive Relationships in Subsurface Flow Problems. *Water Resources Research*, e2019WR026731. doi: 10.1029/2019wr026731

Tian, J., Qi, C., Sun, Y., & Yaseen, Z. (2020, 07). Surrogate permeability modelling of low-permeable rocks using convolutional neural networks. *Computer Methods in Applied Mechanics and Engineering*, *366*, 113103. doi: 10.1016/j.cma.2020.113103

Vega, B., Dutta, A., & Kovscek, A. (2014, 10). Ct imaging of low-permeability, dual-porosity systems using high x-ray contrast gas. *Transport in Porous Media*, *101*. doi: 10.1007/s11242-013-0232-0

Vik, B., Bastesen, E., & Skauge, A. (2013, 12). Evaluation of representative elementary volume for a vuggy carbonate rock—part: Porosity, permeability, and dispersivity. *Journal of Petroleum Science and Engineering*, *112*, 36–47. doi: 10.1016/j.petrol.2013.03.029

Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., … Tang, X. (2018, 09). Esrgan: Enhanced super-resolution generative adversarial networks. In *European conference on computer vision (eccv).*

Wen, G., Hay, C., & Benson, S. (2021). Ccsnet: a deep learning modeling suite for

co_2 storage. *In Review*.

Zahasky, C., & Benson, S. M. (2018). Micro-Positron Emission Tomography for Measuring Sub-core Scale Single and Multiphase Transport Parameters in Porous Media. *Advances in Water Resources*, *115*, 1–16. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S030917081731182X doi: 10.1016/j.advwatres.2018.03.002

Zahasky, C., & Benson, S. M. (2019, nov). Spatial and temporal quantification of spontaneous imbibition. *Geophysical Research Letters*, *46*(21), 11972–11982. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1029/2019GL084532 doi: 10.1029/2019GL084532

Zahasky, C., Jackson, S. J., Lin, Q., & Krevor, S. (2020). Pore network model predictions of Darcy-scale multiphase flow heterogeneity validated by experiments. *Water Resources Research*, 1–16. doi: 10.1029/2019wr026708

Zahasky, C., Kurotori, T., Pini, R., & Benson, S. M. (2019). Positron emission tomography in water resources and subsurface energy resources engineering research. *Advances in Water Resources*, *127*(March), 39–52. Retrieved from https://doi.org/10.1016/j.advwatres.2019.03.003 doi: 10.1016/j.advwatres.2019.03.003

Zhang, Y., Tian, Y., Kong, Y., Zhong, B., & Fu, Y. (2018, 06). Residual dense network for image super-resolution. In *Conference on computer vision and pattern recognition* (p. 2472-2481). doi: 10.1109/CVPR.2018.00262

Zhong, Z., Sun, A., & Jeong, H. (2019, 06). Predicting co 2 plume migration in heterogeneous formations using conditional deep convolutional generative adversarial network. *Water Resources Research*, *55*. doi: 10.1029/2018WR024592

Zhu, Y., & Zabaras, N. (2018). Bayesian deep convolutional encoder–decoder networks for surrogate modeling and uncertainty quantification. *Journal of Computational Physics*, *366*, 415 - 447. Retrieved from http://www.sciencedirect.com/science/article/pii/S0021999118302341 doi: https://doi.org/10.1016/j.jcp.2018.04.018

Supporting Information for "Three-Dimensional Permeability Inversion Using Convolutional Neural Networks and Positron Emission Tomography"

Zitong Huang[1], Takeshi Kurotori[2,3], Ronny Pini[3], Sally M. Benson[2],

Christopher Zahasky[1]

[1]Department of Geoscience, University of Wisconsin-Madison, Madison, WI, USA

[2]Department of Energy Resources Engineering, Stanford University, Stanford, CA, USA

[3]Department of Chemical Engineering, Imperial College London, London, UK

Contents of this file

1. Table S1 summarizing rock cores used for gathering experimental PET imaging datasets.

2. Figure S1 with illustration of arrival time calculation.

3. Figure S2 with illustration of synthetic porosity-permeability training data relationship.

4. Figure S4 and description of Residual-in-Residual Dense Block.

Residual-in-Residual Dense Block

To increase the depth of the networks without the gradient-vanishing or gradient-exploding problem, a residual learning framework was adopted to connect the Dense Blocks in the networks. Instead of directly learning the unreferenced original mapping, the residual connection adopts a skip-connection between blocks that learn residual functions with reference to the layer inputs (He et al., 2016). Suppose x is the input for the current layer and let x denotes the residual. Let $F(x)$ denote the optimal mapping of the current layer and let $R(x)$ denotes the original mapping (or the residual function) of the current layer, and let $F(x)\ R(x)\ +\ x$. The $F(x)$ is then passed to the next layer, so if the original $R(x)$ of the current layer enlarges the error, the next layer could always refer back to the residual x, which could be considered as skipping the layer that enlarges the error. To the other extreme, if the original mapping $R(x)$ is optimal, the residual x will be set to zero. Therefore, the deeper layer would produce no higher error than the upper layer (He et al., 2016). Compare to the original mapping, it is easier to optimize the residual mapping. The residual-in-residual dense block (RRDB) are composed of a stack of residual dense blocks connected in another residual structure (Wang et al., 2018; Mo et al., 2019). Therefore, the residual learning was used in two levels, resulting in a residual-in-residual structure. For both of the two levels, the desired output is actually denoted as $F(x)\beta \times R(x) + x$, where $\beta \in (0, 1]$ is the residual scaling factor (Wang et al., 2018).

References

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the ieee conference on computer vision and pattern recognition* (Vol. 7, pp. 770–778).

Mo, S., Zabaras, N., Shi, X., & Wu, J. (2019). Integration of Adversarial Autoencoders With Residual Dense Convolutional Networks for Estimation of Non-Gaussian Hydraulic Conductivities. *Water Resources Research*, *56*(2), 1–24. doi: 10.1029/2019WR026082

Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., ... Tang, X. (2018, 09). Esrgan: Enhanced super-resolution generative adversarial networks. In *European conference on computer vision (eccv).*

Sample Name	Core length [cm]	Average Permeability [mD]	Average Porosity [-]
Berea Sandstone	10.0	23	0.20
Indiana limestone	10.3	98	0.17
Edwards Brown limestone	10.3	132	0.41
Ketton limestone	10.0	1920	0.23

Table S1. Table summarizing rock cores used for gathering experimental PET imaging datasets.

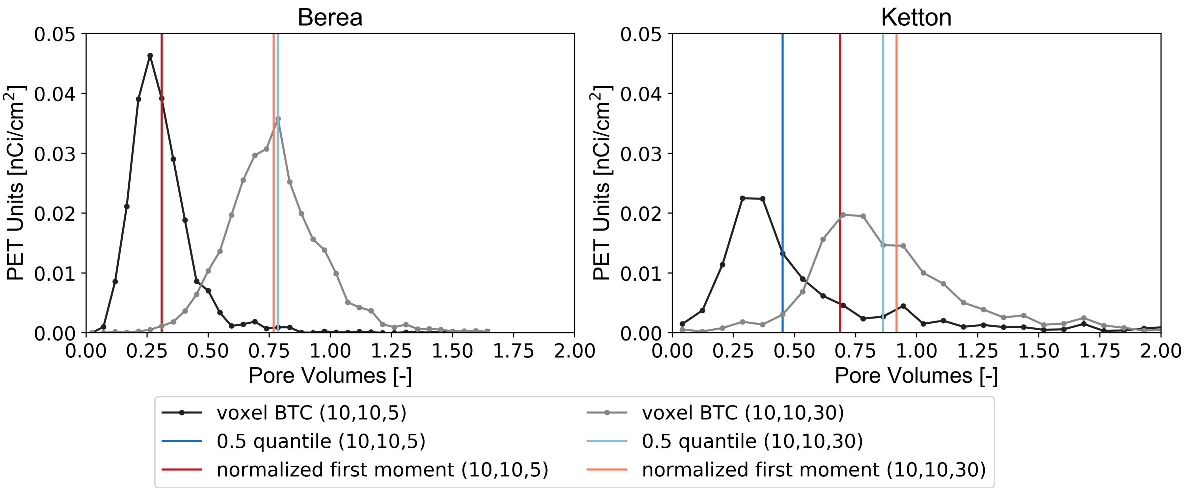

Figure S1. Example breakthrough curves derived from different voxels of the PET scans of the Berea sandstone (left plot) and the Ketton limestone (right plot) experiments. The black lines and darker blue and red lines are for the voxel near the central axis of the core and 1.19 cm from the inlet (voxel coordinate: 10,10,5). The lighter colors correspond to the voxel near the central axis of the core and 7.16 cm from the inlet (voxel coordinate: 10,10,30). While normalized first moments and 0.5 quantiles are very similar in a Berea sandstone, the significant microporosity and resulting solute tailing in the Ketton limestone generates significant delay of the normalized first moment location relative to the 0.5 quantile.

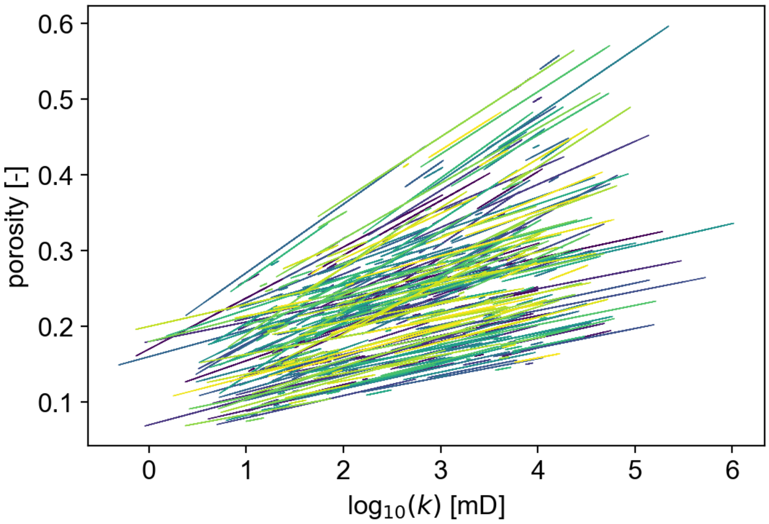

Figure S2. Porosity-permeability relationships for every grid cell of each test set realization using Equation 5 with randomly sampled a and b parameters. Each test set has a different line color and all 500 test datasets are plotted.

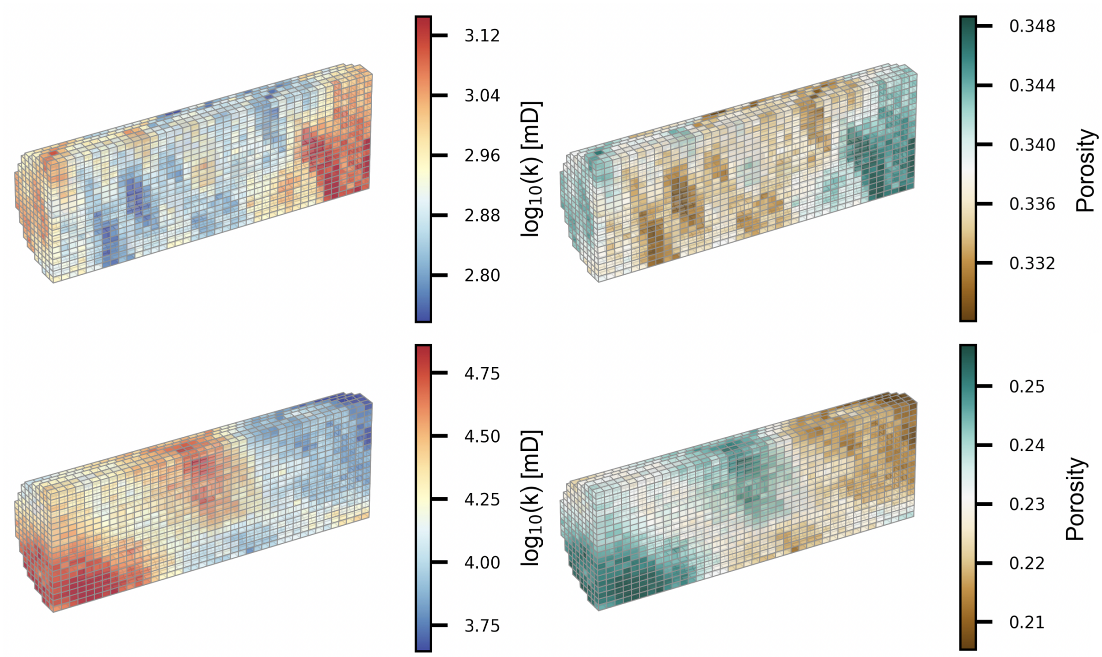

Figure S3. Two test sets of synthetic permeability and corresponding porosity maps used in the second neural network that incorporates heterogeneous porosity. The top row illustrates sample 430 from the test set; the corresponding porosity is generated with $a = 0.4565$ and $b = 19$ using Equation 5. The bottom row illustrates sample 404 from the test set; the corresponding porosity is generated with $a = 0.5407$ and $b = 5$. The grid cells for all models are 0.233 cm $\times$0.233 cm $\times$0.25 cm.

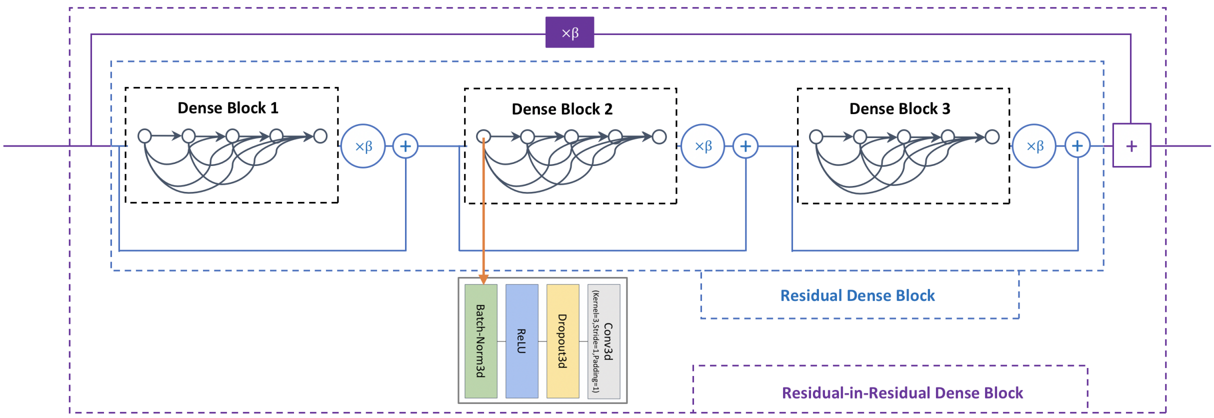

Figure S4. Architecture of the residual-in-residual dense block (Wang et al., 2018; Mo et al., 2019)