Water Resources Research

RESEARCH ARTICLE
10.1029/2021WR031554

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], and 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 maps of large experimental datasets are used to parameterize forward numerical models to validate CNN predictions

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

C. Zahasky,
czahasky@wisc.edu

Citation:

Huang, Z., Kurotori, T., Pini, R., Benson, S. M., & Zahasky, C. (2022). Three-dimensional permeability inversion using convolutional neural networks and positron emission tomography. *Water Resources Research*, *58*, e2021WR031554. https://doi.org/10.1029/2021WR031554

Received 4 NOV 2021
Accepted 7 MAR 2022

Author Contributions:

Conceptualization: Christopher Zahasky
Data curation: Takeshi Kurotori, Christopher Zahasky
Formal analysis: Zitong Huang, Takeshi Kurotori, Christopher Zahasky
Funding acquisition: Ronny Pini, Sally M. Benson, Christopher Zahasky
Investigation: Zitong Huang, Takeshi Kurotori, Christopher Zahasky
Methodology: Zitong Huang, Christopher Zahasky
Project Administration: Christopher Zahasky

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 behavior, 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 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 synthetic test datasets and then test the accuracy on a suite of experimental PET imaging datasets acquired on four different geologic cores. The network-predicted permeability maps from the geologic cores are used to parameterize forward numerical models that are directly compared with the experimental PET imaging data. The results indicate that a single trained network can generate robust 3-D permeability inversion maps in seconds. Numerical models parameterized with these permeability maps closely capture the experimentally observed solute arrival time behavior. This work provides an unprecedented approach for efficiently characterizing multiscale permeability heterogeneity in complex geologic samples.

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—a hydrogeologic property 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 model to perform 3-D permeability prediction. The model is trained by guiding the model to identify the characteristics in the transport data that provide insights on permeability distribution. Compared to traditional mathematical modeling approaches, the trained deep learning model significantly reduces the computational cost while accurately predicting the 3-D permeability distributions in real geologic materials.

1. Introduction

Understanding flow and transport in porous media is crucial for understanding complex hydrogeologic systems, designing contaminant remediation strategies, and utilizing subsurface energy resources. Mechanistic understanding of how geologic heterogeneity controls the associated reactive transport, colloidal transport, or multiphase flow processes first requires an accurate 3-D parameterization of multiscale intrinsic permeability. Despite this necessity, approaches for non-destructive experimental measurement of multi-scale permeability in geologic core samples remains a critical challenge. Current approaches for measuring spatially variable permeability are experimentally challenging or laborious (Krause et al., 2013), computationally expensive (Guo et al., 2021; Mostaghimi

Resources: Ronny Pini, Sally M. Benson, Christopher Zahasky
Software: Zitong Huang, Christopher Zahasky
Supervision: Ronny Pini, Christopher Zahasky
Validation: Zitong Huang, Christopher Zahasky
Visualization: Zitong Huang, Christopher Zahasky
Writing – original draft: Zitong Huang, Christopher Zahasky
Writing – review & editing: Zitong Huang, Takeshi Kurotori, Ronny Pini, Christopher Zahasky

et al., 2013), and typically rely on sample-specific porosity-permeability (Chilingar, 1964; Chilingarian, 1991; Han et al., 2019) or capillary pressure scaling assumptions (Krause et al., 2013; Rabinovich, 2017).

A range of methods have been developed to approximate spatially-variable permeability in geologic samples (Krause, 2012; Krause et al., 2013; Rabinovich, 2017). The approach developed by Krause et al. (2013) utilizes multiphase core-flooding experiments, X-ray computed tomography (X-ray CT), 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 similarities. 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, the 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 the 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 is due to the presence of vugs—pores larger than the typical grain size (Lucia, 1983). The porosity-permeability relationship can vary significantly depending on the connectivity of the vugs and therefore be challenging to characterize or generalize. For example, the presence of isolated vugs significantly increases the porosity but it does not lead to a 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 carbonate samples.

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 3-D images based on photons emitted from radiolabeled fluids. This difference in image acquisition and reconstruction provides a complementary approach 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 heterogeneity created by structural features such as deformation bands (Romano et al., 2020).

Positron emission tomography generates 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 (Boggs & Adams, 1992; Mackay et al., 1986). These image-based observations combined with recently developed deep learning tools provide 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 Convolutional Networks (DenseNet) fully leverage the hierarchical advantages of CNN by encouraging feature propagation, sharing, and reuse among all the layers (G. Huang et al., 2017). The number of parameters in a network increases as layers are added to a network, theoretically improving the potential performance of the network. To

overcome some of the challenges that arise in training deeper networks, a residual-in-residual structure can be applied to all the DenseNet blocks (X. 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 utilized for a range of image-based inversion tasks in fields ranging from medical imaging (Adler & Ozan, 2017; Barbastathis et al., 2019; McCann et al., 2017) to geophysics (Kim & Nakata, 2018). Using CNNs for inversion tasks has been shown to address issues of ill-posedness by reducing the effects of noise (Barbastathis et al., 2019), capturing finer resolution features (Kim & Nakata, 2018), and learning a generalizable regularizer during the training process (McCann et al., 2017). In addition to inversion tasks, CNNs have been constructed for a range of hydrogeologic applications including parametrizing hydrogeological properties in highly complex digital rock images (Kamrava et al., 2021; Sudakov et al., 2019; Tian et al., 2020), 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; Tang et al., 2021; Zhong et al., 2019) 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 experimentally constrained 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 (Kamrava et al., 2021; Sudakov et al., 2019; Tian et al., 2020). These digital workflows—reviewed in detail by Y. D. Wang et al. (2021)—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 numerical inversion approaches because rather than iterating a simulation model to fit a specific geologic sample, the 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 collected in sedimentary rocks.

2. Methods

2.1. 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 CNN inversion approach. 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 were 5.04 cm in diameter and between 10 and 10.3 cm long. 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 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 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 with pore water velocities spanning from 0.245 cm/min to 0.590 cm/min. Core-average permeability was calculated with Darcy's Law using the steady-state differential pressure, core geometry, and fluid injection rate. The cores had a sample-average permeability spanning from 23 mD to 1920 mD (Table S1 in the Supporting Information S1).

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 min. The PET scans were performed using a Siemens pre-clinical Inveon DPET scanner. Once a scan was started, pulses of radiotracer—between 0.02 and 0.10 pore volumes—were injected into the samples and displaced with water containing no ^{18}F-FDG. The images were reconstructed and processed with voxel dimensions of $0.2329 \times 0.2329 \times 0.2388$ cm^3. This voxel dimension size was determined as the smallest volume that captured the continuum scale heterogeneity of the four geologic cores used in this study. This determination was based on representative elementary volume (REV) analysis of continuum properties in similar geologic cores (Jackson et al., 2020; Zahasky et al., 2020) and theoretical metrics of the minimum permeability REV derived from Navier-Stokes simulations (Narsilio et al., 2009) and using previously measured sample grain size distributions (Kurotori, 2019). 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.

2.2. 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.

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

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_q is the time when $Q(\tau_q)$ reaches the qth quantile. The 0.5 quantile ($\tau_{0.5}$) is the median arrival time and corresponds to the time when half of the solute has passed through the voxel. Using the discrete form of Equation 1, the arrival time values 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. The $\tau_{0.5}$ arrival time was used in this study and an example of the resulting 3-D arrival time map for the Berea sandstone sample is illustrated 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 (Dagan et al., 1992; Harvey & Gorelick, 1995). 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 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). Third, the arrival time is insulated from variations in hydrodynamic and numerical dispersion, particularly at the 0.5 quantile (Dagan & Nguyen, 1989). This is important for numerically 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. This is also advantageous for the comparison of different geologic materials that exhibit strong variation in mechanical dispersion and tailing driven by microporosity (Kurotori et al., 2019). The quantile-based arrival time is less susceptible to solute tailing and background measurement noise than the normalized first moment because the

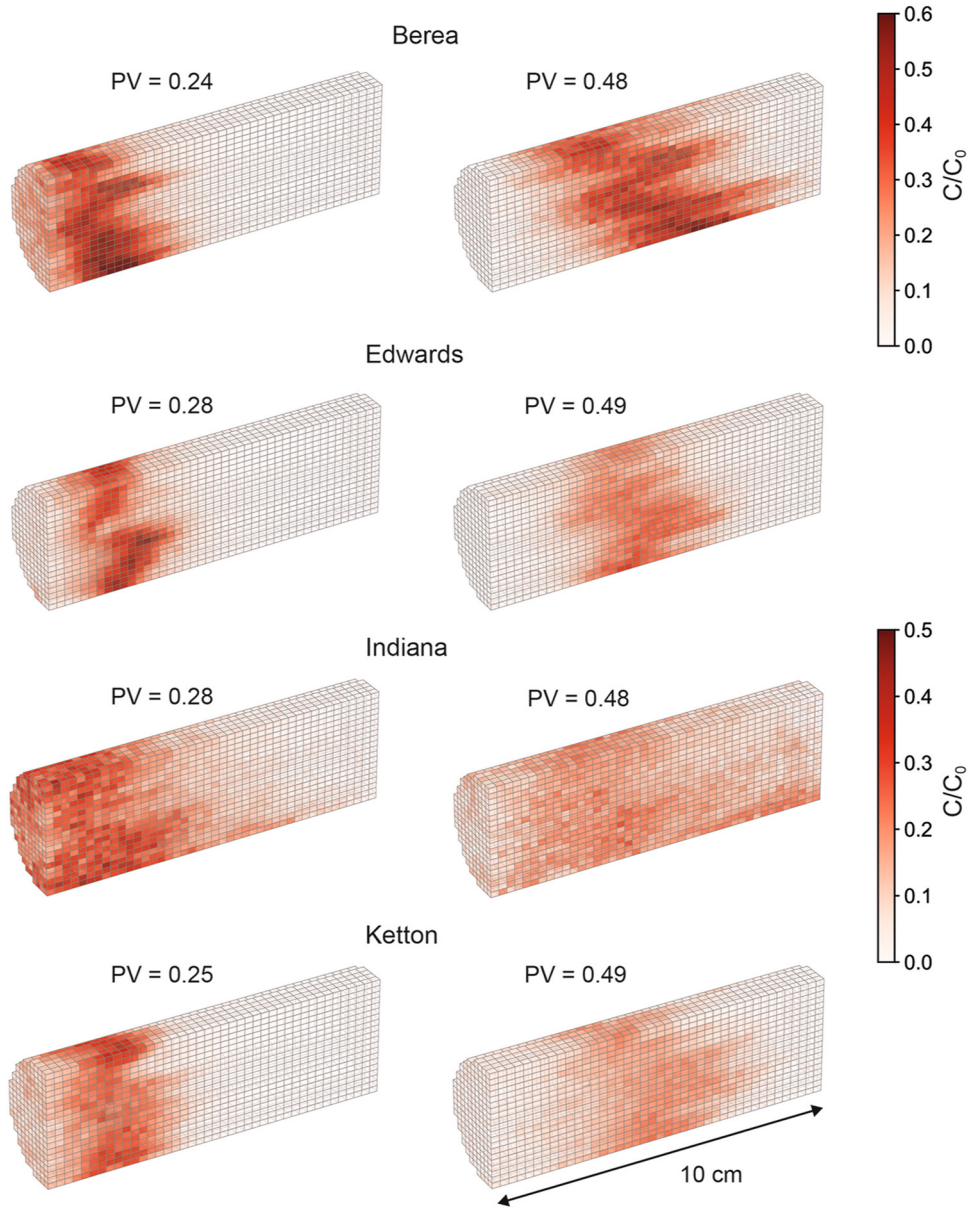

Figure 1. Example PET imaging time frames from each of the four cores used in this study. The pore volume injected (PV) is indicated for each image and are referenced from the start of tracer injection. The difference in the displayed pore volumes is a result of slight differences in temporal image reconstruction. The Berea sandstone core (displayed at the top row) has a slightly larger color bar scale because the pulse volume of tracer injected was 4 mL as opposed to the three limestone cores (displayed in rows two, three, and four) that had pulse volumes of 2 mL. The voxel size dimensions for all models are $0.2329 \times 0.2329 \times 0.2388$ cm^3. These images highlight the local sub-core permeability heterogeneity present in all four cores.

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 S1.

To accentuate the structural similarity and amplify the signal of subtle differences in arrival times, the underlying linear trend in the calculated arrival times—due to the bulk flow from the inlet to the outlet of the samples—was removed. This linear trend can mask arrival time variation and is fundamentally different from the underlying permeability structure of the samples as illustrated in Figure 2. The arrival time map was first normalized to nondimensional units of pore volumes injected (upper right image in Figure 2), and the nondimensionalized map was then subtracted from the linear trend, resulting in what we call an arrival time difference map as shown in

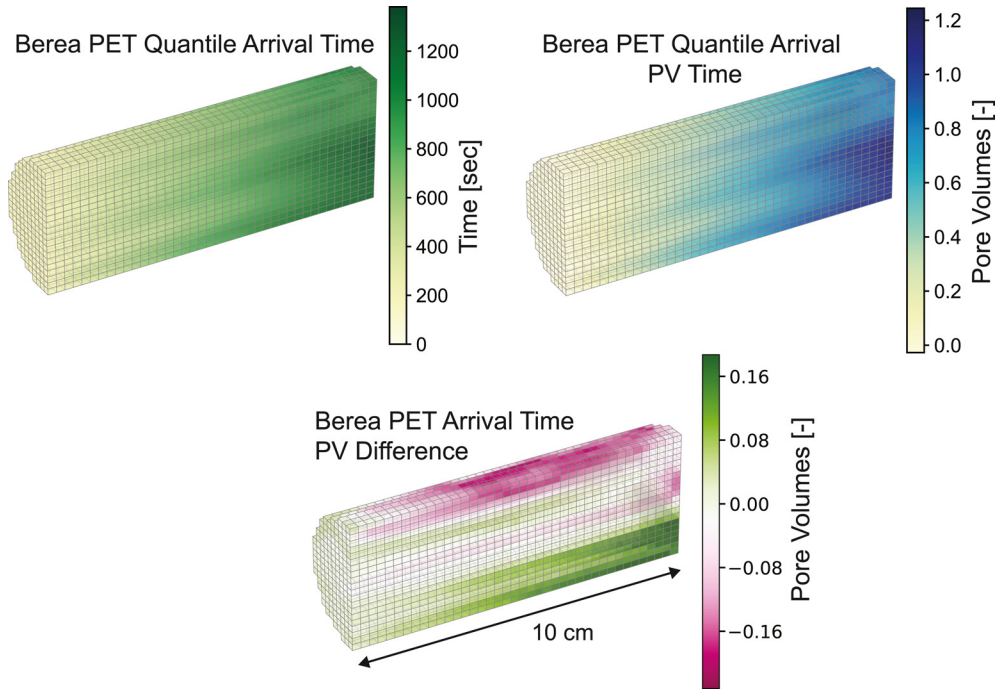

Figure 2. (Upper left) Quantile (0.5) arrival time map collected in the Berea core using the PET data illustrated in the top of Figure 1. (Upper right) Quantile arrival time map in normalized units of pore volumes of water injected since the start of tracer injection. (Bottom) Quantile arrival time difference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

the bottom plot of Figure 2. This representation of arrival times more closely reflects the underlying permeability structure. Greener voxels in the bottom plot of Figure 2 have arrival times faster than the core average as a result of higher permeability zones. Pinker voxels in the bottom plot of 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.

2.3. 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 ($\boldsymbol{\phi}$) is calculated via the linear scaling expression in Equation 2. This scaling requires a dry scan of the sample ($\boldsymbol{Z}_a$), and a second scan when the sample is fully saturated with water ($\boldsymbol{Z}_w$). The difference between these scans is then scaled by using the difference between pure 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.

$$\boldsymbol{\phi} = \frac{\boldsymbol{Z}_w - \boldsymbol{Z}_a}{\Delta_{a,w}} \tag{2}$$

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 the most valuable 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 were 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 throughout 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 high porosity and low porosity regions of a sample. A risk of using scaled X-ray CT

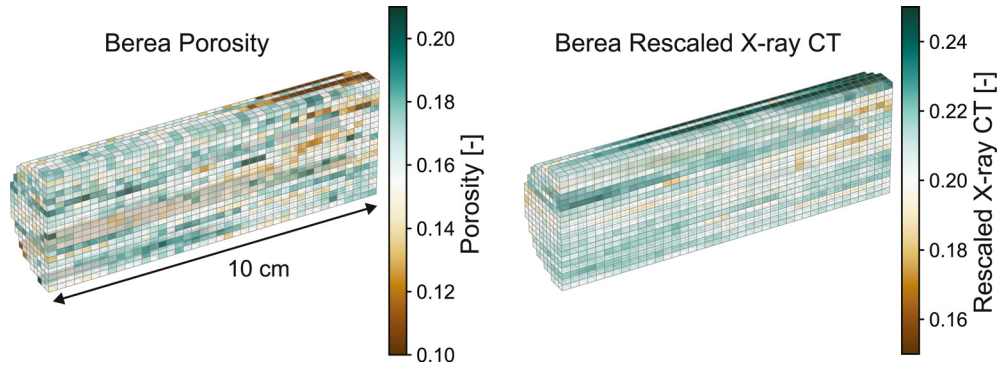

Figure 3. (Left) Porosity map of Berea sandstone calculated using linear scaling with Equation 2. (Right) Air-saturated X-ray CT scan of Berea sandstone scaled to a typical porosity range using Equation 3.

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 to 0.25 using Equation 3, similar to typical porosity ranges in consolidated rocks.

$$\tilde{\boldsymbol{\phi}} = 0.10 \cdot \|\boldsymbol{Z}_a\| + 0.15 \tag{3}$$

An illustration of the rescaled dry X-ray CT scan in the Berea sandstone calculated with Equation 3 is presented 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.

2.4. 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 was composed of arrival time difference maps calculated from numerical solute transport simulations on synthetically generated permeability maps with homogeneous porosity. The second dataset was composed of arrival time difference maps with the same synthetically generated permeability but with the addition of a corresponding heterogeneous porosity map. For each dataset, a total of 26,000 numerically simulated data pairs were generated.

2.4.1. Training Dataset Without Porosity

Permeability maps were generated with the exponential covariance random field generation algorithm using the open-source Python codes from Müller and Schüler (2021). Latin hypercube sampling (Deutsch & Deutsch, 2012; Mckay et al., 1979; Tartakovsky et al., 2020) was used to generate the permeability map parameter space defined by a mean permeability from 10 mD–20 D; $\log_{10}$ standard deviation from −1.7 to 9.9 mD; spatial correlation length from 0.25 to 12.5 cm in the x, y, and z directions; rotation from 0 to 90° in each of the x, y, and z planes. The minimum correlation length was chosen to equal the numerical model discretization such that very little permeability correlation was present in some of the permeability maps. The sampling also included a variable number (0–2) of heterogeneous dummy slices added to the model inlet face. The purpose of the variable number of heterogeneous dummy slices was to improve network generalization by creating subtle perturbations of solute distribution in the synthetic datasets that are intrinsically present in experimental inlet boundary conditions. The strength of these experimental boundary effects has been observed in other in situ transport imaging studies and are difficult to predict a priori (Lehoux et al., 2016). 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 steady-state flow simulations on the synthetic 3-D permeability maps using MODFLOW 2005 (Harbaugh, 2005) and MT3DMS (Zheng & Wang, 1999)

scripted in FloPy (Bakker et al., 2016). To mimic the experimental settings, the flow simulation was done on synthetic cylindrical cores with a diameter of 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. The synthetic 3-D permeability and modeled arrival time difference maps were all represented with dimensions of $20 \times 20 \times 40$, nearly identical to the discretization of the 3-D PET arrival time maps obtained from experiments discussed in Section 2.1. The grid cells for all models have dimensions of $0.233 \times 0.233 \times 0.25$ cm^3, such that they are nearly identical to the voxel size in the processed PET images. 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 2.2.

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

$$\bar{k} = \frac{Q_t}{A} \cdot \mu \cdot \frac{L}{\Delta p} \tag{4}$$

The flow rate (Q_t) 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 L was the length of the model core. The variable μ was defined by 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 $20 \times 20 \times 41$. 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.

2.4.2. Training Dataset With Porosity

A second training dataset was constructed to explore the impact of porosity heterogeneity and porosity structure information on the permeability inversion. 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; Glatz et al., 2016; Minto et al., 2017; Vega et al., 2014). Second, core-averaged porosity has been shown to have a geometric correlation with permeability (Chilingar, 1964; Chilingarian, 1991). By using both the arrival time difference map and porosity map as the inputs for the inversion process, this second network aimed to improve the accuracy of permeability map inversion by providing insights on the geometric distribution of permeability heterogeneity in the core. In this dataset, the same synthetic permeability maps 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 synthetic porosity maps were generated by utilizing an empirical porosity-permeability function given by Equation 5.

$$\phi_i = \frac{\frac{\ln(k_i)}{a} + b}{100} \tag{5}$$

Here ϕ_i is the porosity of a given grid cell and k_i is the permeability in millidarcy of grid cell i in a given training realization. The variable a is a constant ranging from 0.25 to 1, and b is another constant ranging from 5 to 20. These empirical parameters varied with each training set realization and were sampled with the Latin hypercube sampling of the permeability map characteristics described in the previous section. 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. An illustration of the variation in porosity-permeability relationships is illustrated in Figure S2 in the Supporting Information S1 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 $20 \times 20 \times 41$. Two different randomly selected training realizations generated with the above workflow are illustrated in the 3-D plots in Figure S3 in

the Supporting Information S1. The Python codes used for training data generation and the full compilation of training data are available in the data repository cited in the Acknowledgments.

2.5. Network Construction and Training

2.5.1. 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/kernels, each representing an abstract feature of the input image. As the number of connected convolutional layers increases, the input spatial information gets selected and refined. The accumulated receptive fields of shallower (or earlier) layers make the region exposed to the neurons in the deeper (or later) layers larger. This enables CNN 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.

2.5.2. Residual-In-Residual Dense Network

Stochastic gradient descent on the loss function (i.e., backpropagating the partial derivatives with respect to a selected subset of network parameters at every gradient step) is often implemented for the training of deep neural networks (Kingma & Ba, 2014). As the network grows deeper, gradient-based methods may suffer gradient-vanishing or gradient-exploding problems as the gradients accumulate during the training and backpropagating process. Specifically, gradient instabilities caused by factors such as repeated weights multiplication, saturating activation functions, or improper parameters initialization (Glorot & Bengio, 2010; Lu et al., 2020) are likely to be amplified during the backpropagating process, causing either early saturation or stagnation in the network learning process (Ioffe & Szegedy, 2015).

To solve the gradient-vanishing problem, DenseNet was adopted to connect all layers—with matching feature map sizes—directly with all their subsequent layers (G. Huang et al., 2017). The direct connections were 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 feature maps to all subsequent layers (G. Huang et al., 2017). The growth rate of a dense block refers to 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). After receiving the concatenated feature maps as inputs, each dense block carried out the batch normalization (BN; Ioffe & Szegedy, 2015) and the rectified linear unit (ReLU) nonlinear activation. Finally, the main features of the activated prediction were 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 implemented to connect the dense blocks in the networks (Zhang et al., 2018). Residual-in-residual dense blocks (RRDB) have been successfully applied in image super-resolution (X. Wang et al., 2018) and geologic features parameterization (Mo et al., 2019c). Based on these previous models, the RRDB built here contained three residual dense blocks that included five dense blocks connected in a residual learning framework. The growth rate of each dense block was set to 48 and the residual scaling factor β was set to 0.2.

2.5.3. Network Architecture

The 3-D encoder-decoder based CNN used an encoder to first extract and parameterize the high-level features of the input data. The compressed high-level features map, referred to as the latent space, had a dimension of $5 \times 5 \times 10$. The decoder then constructed the permeability maps based on the extraction of high-level features in the latent space. The predicted permeability maps had a final dimension of $20 \times 20 \times 40$, the same as the dimension of the arrival time different input maps generated with the forward model using the synthetic permeability maps as described in Section 2.4.1.

An illustration of the network architecture is presented in Figure 4. The convolutional block consisted of a single 3-D convolutional layer, the residual-in-residual dense block consisted of 15 dense blocks, and feature selection through compression and reconstruction was achieved through average pooling and up-sampling blocks. Each

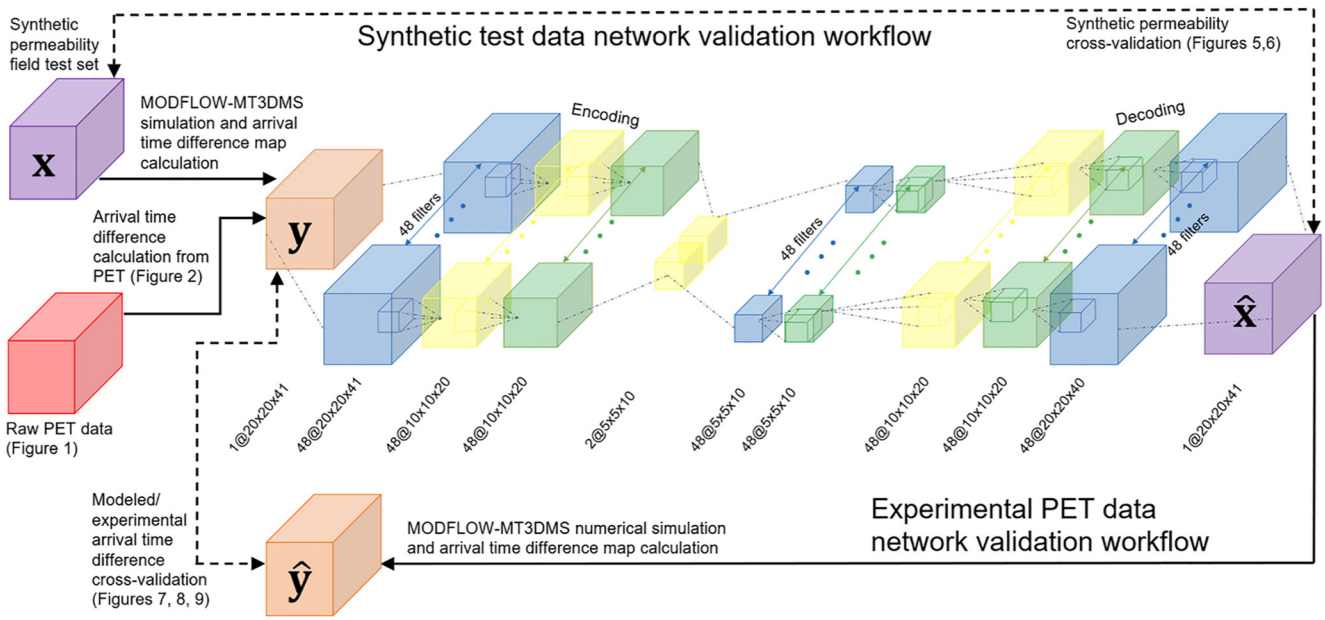

Figure 4. Schematic illustration of the inversion-validation workflow using both synthetic (top loop) and experimental PET data (bottom loop). The purple blocks correspond to synthetic/predicted permeability maps, the red block is the PET data, the orange blocks are experimental and modeled arrival time difference maps. The CNN components include convolutional blocks (blue), up/down-sampling block (yellow), and residual-in-residual dense blocks (green).

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 of 8,570,690 trainable parameters. Both the networks trained with and without porosity maps had the same architecture with the only difference being the number of input channels. The network trained without porosity had one input channel and the network trained with porosity had two input channels.

2.5.4. Network Training

The network training was a supervised process. The inputs were 3-D image tensors containing the arrival time difference maps ($\boldsymbol{y}$) and the permeability maps of the corresponding synthetic geologic core were the target data ($\boldsymbol{x}$). 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.

The addition of noise to the input data during network training is key for network application to noisy experimental data while improving network generalization and reducing overfitting (DeVries & Taylor, 2017; Noh et al., 2017). Image-based PET data contain Gaussian measurement noise that varies between experiments depending on the 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 (Zahasky et al., 2019). To replicate this noise in the training data, all of the simulated arrival time difference maps were corrupted with Gaussian white noise. To account for variation in dataset noise while ensuring that all of the datasets experienced some noise, the noise applied to the input arrival time difference maps was assigned with a distribution that 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 magnitude of added noise was determined both from quantification of PET measurement error and hyperparameter tuning during network training.

The predicted permeability maps ($\hat{\boldsymbol{x}}$) were compared with the target synthetic permeability maps ($\boldsymbol{x}$) through loss functions. The loss function ($\mathcal{L}_t$) used in this study is given in Equation 6. The selected loss function was a combination of $L1$ loss ($\mathcal{L}_{L1}$) and KL-Divergence loss ($\mathcal{L}_{KL}$) evaluated between the target log-permeability maps and the corresponding predicted log-permeability maps in the dataset.

$$\mathcal{L}_t = \mathcal{L}_{L1} + \alpha \mathcal{L}_{KL} \tag{6}$$

In Equation 6, the α term is the weight for the KL-Divergence loss. Equation 7 gives the $L1$ loss that measures the absolute distance between the target and predicted log-permeability maps.

$$\mathcal{L}_{L1} = \|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_1 \tag{7}$$

The KL-Divergence loss, given in Equation 8, measures the differences in probability distributions between the target ($P(\boldsymbol{x})$) and predicted log-permeability maps ($Q(\hat{\boldsymbol{x}})$).

$$\mathcal{L}_{KL} = D\left(P(\boldsymbol{x}) \| Q(\hat{\boldsymbol{x}})\right)_{KL} = P(\boldsymbol{x}) \cdot \log\left(\frac{P(\boldsymbol{x})}{Q(\hat{\boldsymbol{x}})}\right) \tag{8}$$

To accentuate the distribution in both the targets and predictions, all the log-permeability maps were normalized by the softmax function separately along the x, y, and z dimensions before performing the KL-divergence calculation.

By minimizing the discrepancy in probability distributions along all three dimensions, the use of KL-divergence loss alleviates the problem of criterion-dependent overfitting (e.g., generating a blurred permeability map with a small, averaged discrepancy in permeability but a large difference in local permeability). The KL-divergence loss is critical for training a network that not only captures the core average permeability but also the local voxel/grid cell permeability. Due to this emphasis on distribution differences, the KL-divergence loss has been widely implemented in restricting the spatial variation within 3-D images for auto-encoding tasks (Larsen et al., 2016; Wu et al., 2016).

The loss propagation was monitored by observing the gradient as a function of the training epoch and the minimum of the loss curve for the predictions on synthetic permeability maps in the validation set. The 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 the 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). A batch size of 32 was assigned to solve the poor generalization problem that originates from large batch size (Keskar et al., 2017) while maintaining high computational efficiency.

An initial learning rate of 0.01 was observed to be the highest learning rate capable of driving rapid training loss reduction while maintaining minimal divergence between the validation loss and the training loss (Bengio, 2012). Therefore, the initial learning rate for the training was set to 0.005 for preventing the early plateau of loss at suboptimal levels. During the training process, over-fitting was monitored as indicated by validation loss stagnation at a relatively high value while the training loss steadily decreases. 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 block to simulate a sparser activation that further reduced the potential for overfitting.

Training accuracy was evaluated on the test set by individually comparing each synthetic permeability map with the network predicted permeability map or by comparing the experimental arrival time difference maps ($\boldsymbol{y}$) and the numerically simulated arrival time difference maps ($\hat{\boldsymbol{y}}$) based on the network permeability map prediction. The root-mean-squared error (RMSE) and coefficient of determination (R^2) statistical indicators were used to evaluate the accuracy of a given network prediction. The form of these equations applied to the permeability maps are described in Equations 9 and 10.

$$\text{RMSE} = \sqrt{\frac{1}{V}\sum_{i=1}^{V}(x_i - \hat{x}_i)^2} \tag{9}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{V}(x_i - \hat{x}_i)^2}{\sum_{i=1}^{V}(x_i - \bar{x})^2} \tag{10}$$

Here, V is the number of voxels/grid cells in a sample, x_i is the target synthetic permeability value in a single voxels/grid cell, $\hat{x}_i$ is the predicted permeability value in the corresponding voxels/grid cell, and $\bar{x}$ is the sample-average target synthetic log permeability. Equation 10 had the same form when applied to arrival time difference maps.

The two synthetic datasets with and without porosity each had 20,000 arrival time difference map and synthetic permeability map pairs assigned to the training set, 5,500 pairs for the validation set, and 500 for the test set. Each training run of 300 epochs generally took 26–36 h to complete on an 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 two separate path files after the training (a 10.8 MB path file for the encoder model and a 22.1 MB path file for the decoder model). The Python codes for the neural networks used in this study and the trained models are available in the repository referenced in the Acknowledgments.

2.5.5. PET Data Inversion-Validation Workflow

After the encoder-decoder based CNN was fully trained, a set of experimental 3-D arrival time difference maps obtained from the PET imaging methods discussed in Sections 2.1 and 2.2 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 2.4.1, arrival time difference maps ($\hat{\boldsymbol{y}}$) were then generated using the predicted 3-D permeability maps ($\hat{\boldsymbol{x}}$) as forward numerical model input. The modeled arrival time difference maps were then directly compared with the experimental measured arrival time difference maps ($\boldsymbol{y}$) to validate the accuracy of the network permeability map predictions. A summary of the overall workflow including permeability inversion, forward numerical flow and transport modeling, and cross validation is presented in Figure 4.

3. Results

3.1. Network Results With Synthetic Test Data and Homogeneous Porosity

The accuracy of the trained encoder-decoder based CNN was evaluated by comparing the RMSE accuracy (Equation 9) of the 500 3-D permeability predictions in the test set. The arrival time difference maps used to generate these predictions all included Gaussian noise. The grid cell-average RMSE of all of the $\log_{10}$-permeability maps in the test set was 0.057, or 1.1 mD. Figure 5 illustrates the grid cell-level network performance of 15 randomly chosen sample permeability map predictions from the test set compared against the corresponding grid cells in the synthetic target permeability maps that were used to generate the input arrival time difference maps.

Examples of two spatially resolved permeability inversion results are plotted in Figure 6. The top plot of Figure 6 provides a 3-D example of 90th percentile permeability map prediction (with a R^2 score of 0.901) and the bottom plot of Figure 6 provides a 3-D example of a 10th 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 learned the key features of the arrival time difference map and the relationship with the corresponding heterogeneous permeability map.

3.2. Network Results With Experimental PET Data

Following the network evaluation with the synthetic test set data, permeability predictions were generated on the experimental arrival time difference datasets collected from four geologic cores using the PET imaging data illustrated in Figure 1 and described in Section 2.1. Figures 7 and 8 show the 3-D experimental arrival time difference maps calculated from the PET data, the predicted permeability maps from the network, and the simulated arrival time difference maps from the MODFLOW-MT3DMS model simulated on the predicted permeability maps. Grid cell-level comparison of the arrival time data is shown in the top row of cross-plots in Figure 9. In Figure 9, 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 had 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.

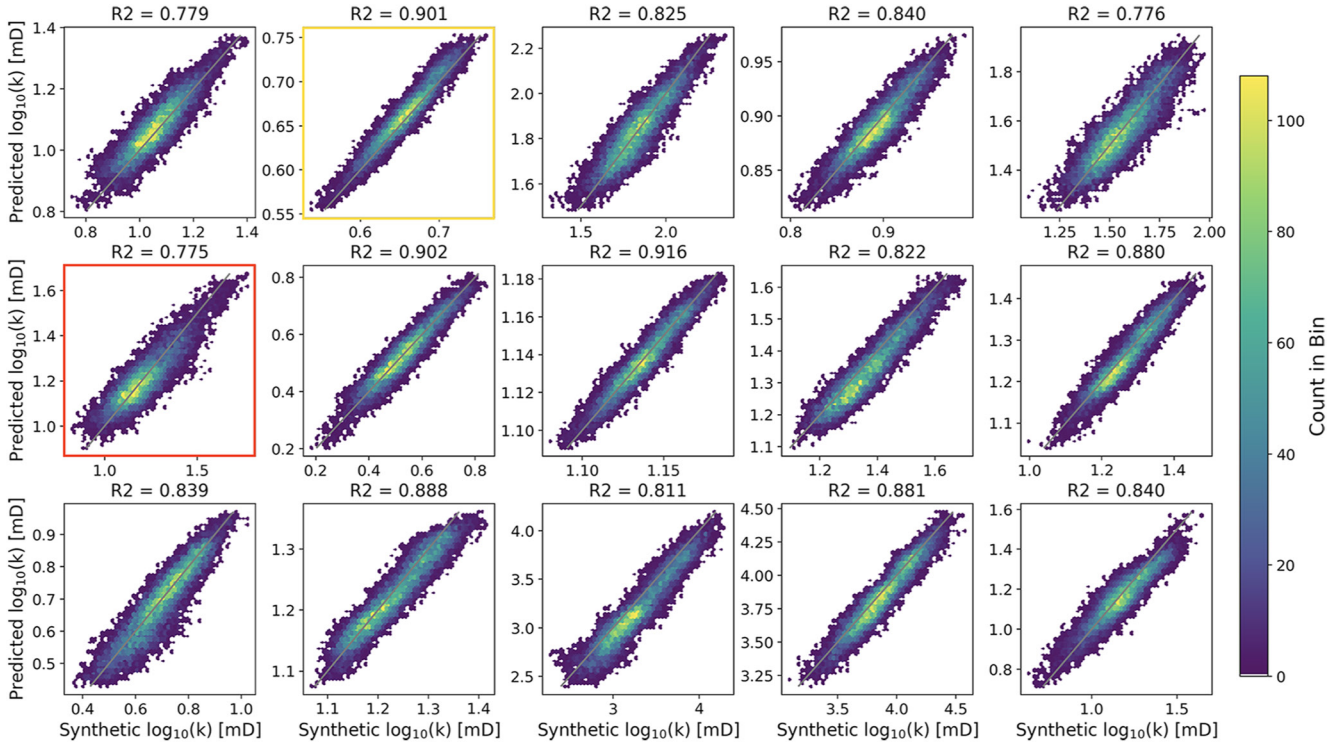

Figure 5. 15 randomly chose samples of permeability map prediction using arrival time difference maps from the test set that included Gaussian noise. For each subplot, the *x*-axis represents the grid cell-level synthetic permeability associated with the test set arrival time, and the *y*-axis represents 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 6, respectively.

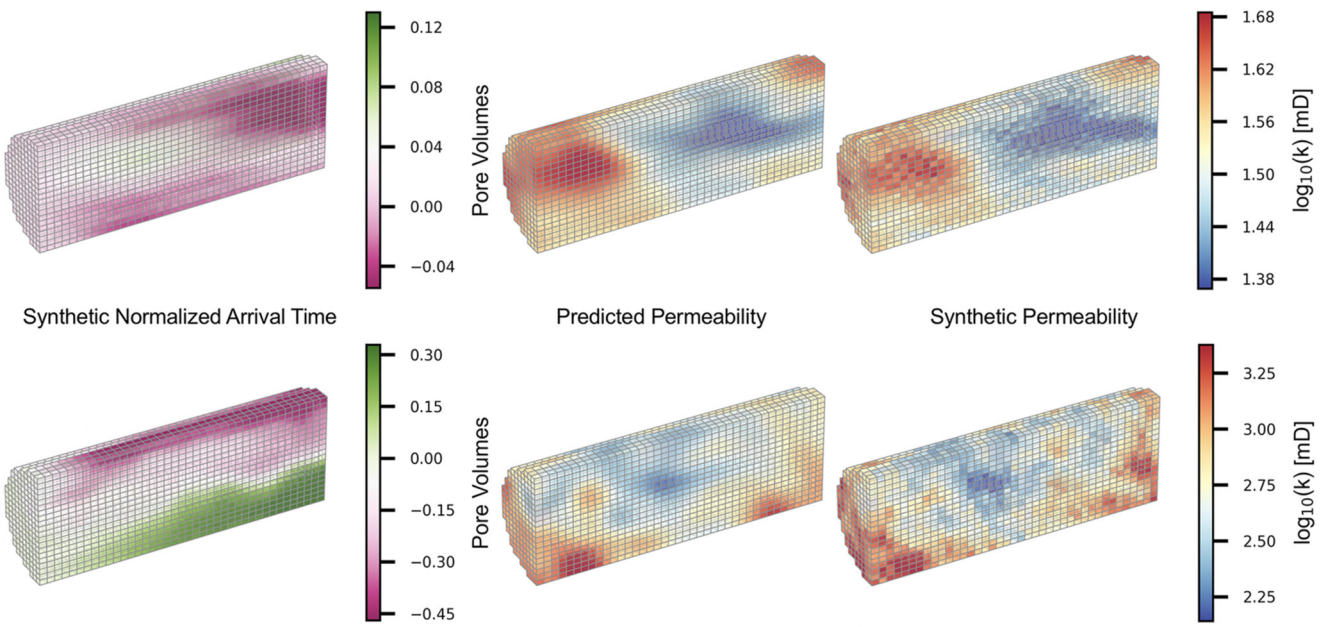

Figure 6. Illustration of two numerically calculated arrival time difference maps using MODFLOW-MT3DMS (left column) based on the corresponding synthetically generated permeability maps (right column). The arrival time difference maps—plotted here without the Gaussian noise—were the input data used to generate the corresponding predicted permeability maps (middle column) by the network trained with homogeneous porosity. In terms of the R^2 accuracy, the top row corresponds to a 90th 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 10th percentile quality prediction and is given by the scatter plot in Figure 5 marked with the red box. The dimension of the grid cells for all models are $0.233 \times 0.233 \times 0.25$ cm^3.

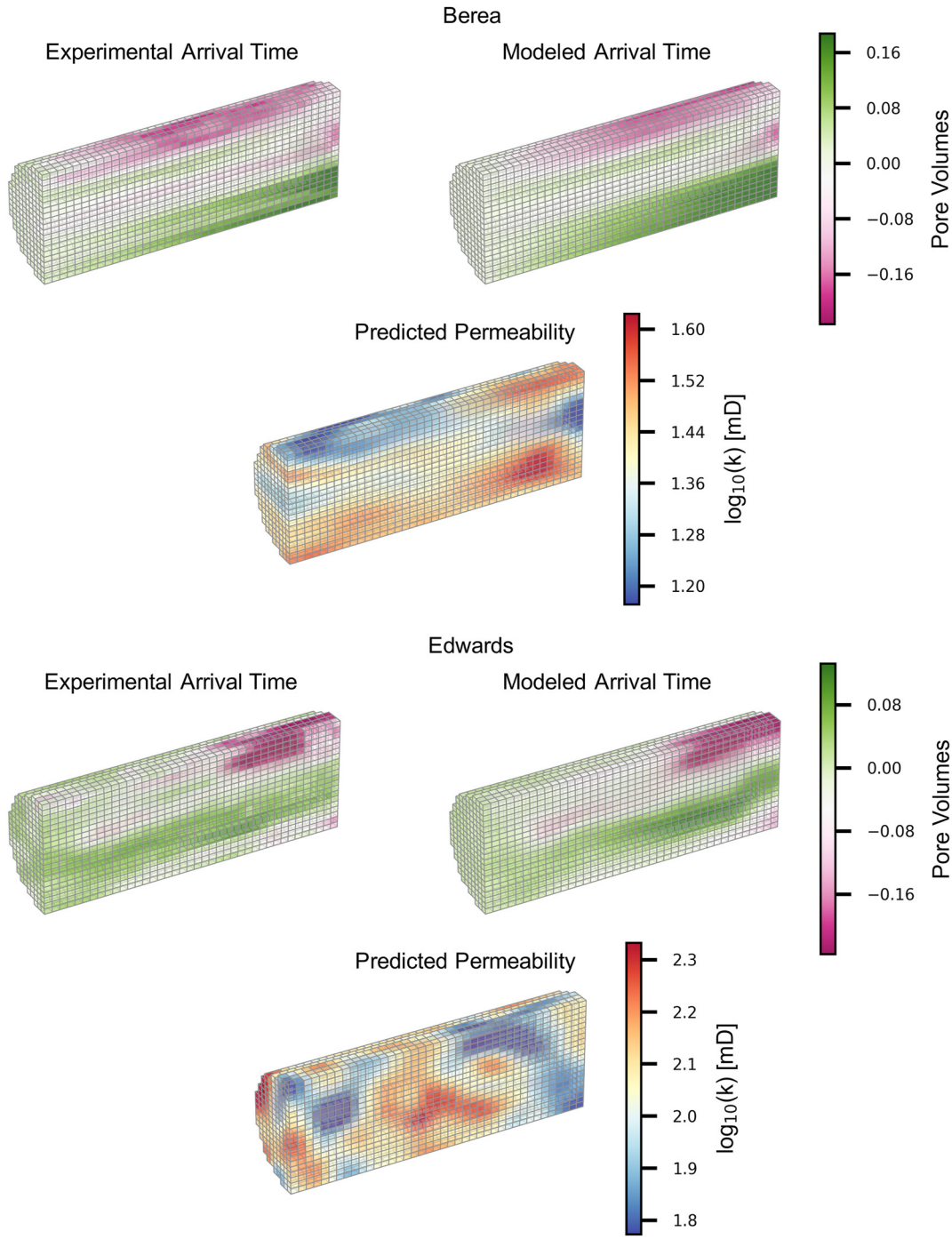

Figure 7. Cross-comparison of the network trained with homogeneous porosity using experimentally measured arrival time difference maps measured with PET on a laminated Berea sandstone (top three subplots) and an Edwards Brown limestone (bottom three subplots). The upper left subplots show the arrival 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 modeled arrival time difference map based on the predicted permeability map. Note that the experimental and modeled arrival times are plotted on the same colorscale.

3.3. 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

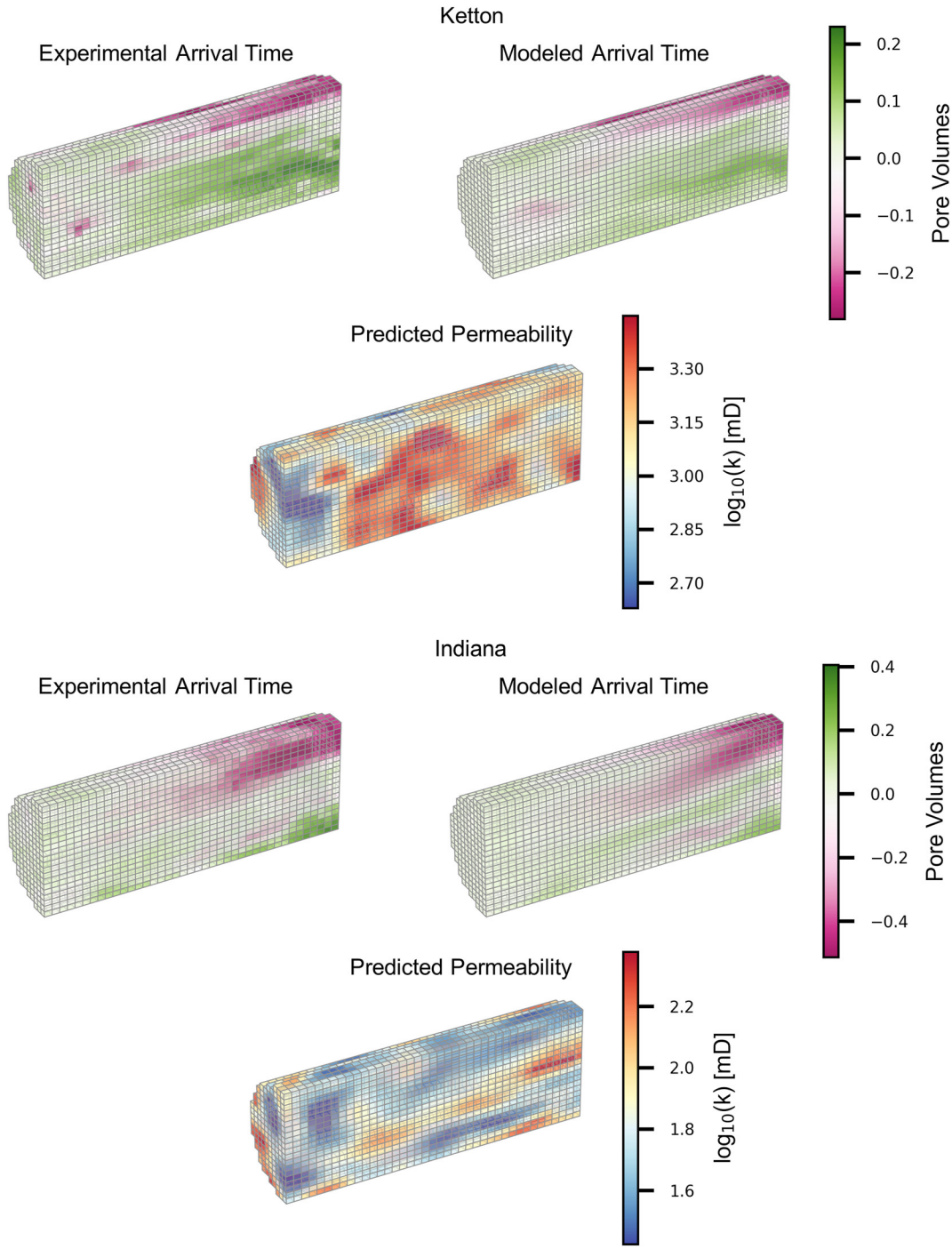

Figure 8. Cross-comparison of the network trained with homogeneous porosity using experimentally measured arrival time difference data measured with PET collected from a Ketton limestone (top three subplots) and a Indiana limestone (bottom three subplots). The upper left subplots show the arrival time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

prediction. This network was trained with an additional input channel of the porosity map as described in Section 2.4.2. 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

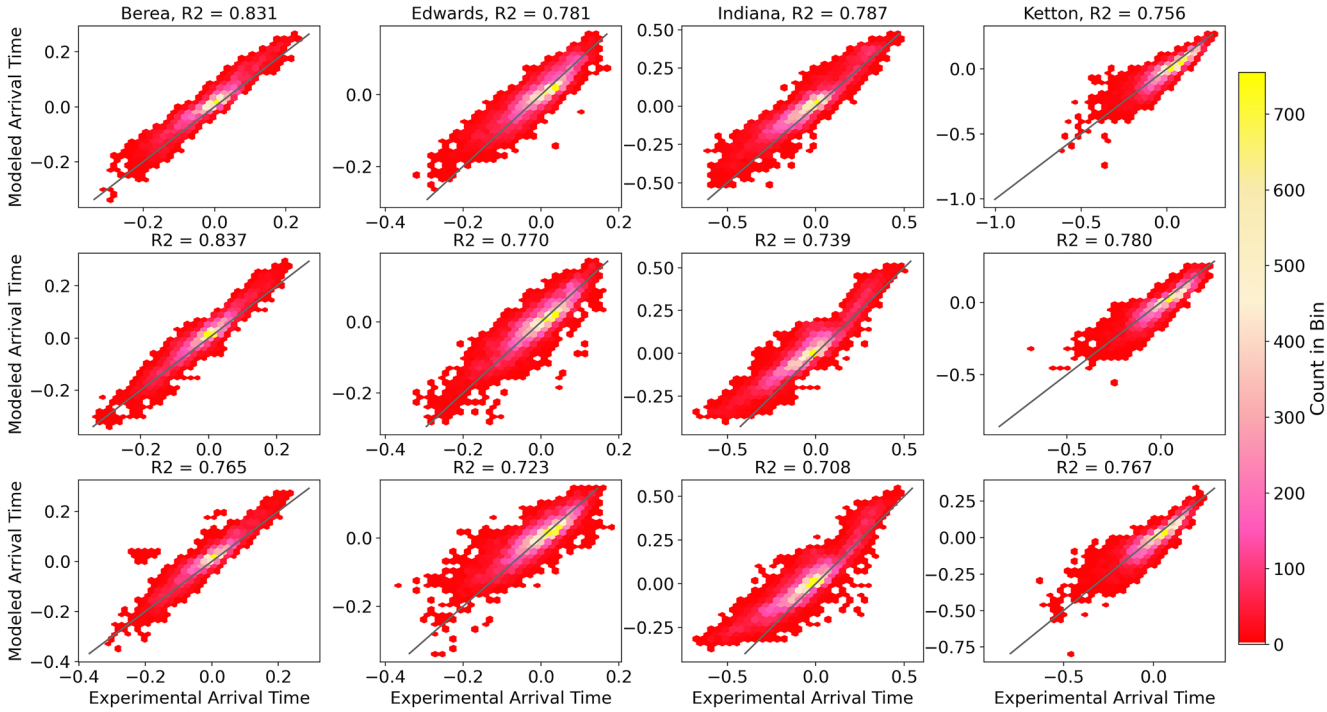

Figure 9. Cross-plot of experimental arrival time difference data (*x*-axis) and modeled arrival time difference from network permeability map prediction for the four geologic cores (from left to right): Berea sandstone, Edwards Brown limestone, Indiana limestone, and Ketton limestone. The top row of 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 addition of X-ray CT-derived data provides very little or no improvement in permeability map prediction.

network with heterogeneous porosity was 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 10. For this second network, the average RMSE accuracy of all the $\log_{10}$-permeability maps in the test set was 0.047, a slight improvement relative to the network with no porosity data that had an RMSE of 0.057. The improved performance on the synthetic data was attributed to the strong spatial correlations between the synthetic permeability and porosity maps as illustrated in Figure S3 in the Supporting Information S1.

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 7 and 8, 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 2.3). Figure 9 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 9. The only instances of higher R^2 accuracy relative to the network using only arrival time data were 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 3% with the addition of X-ray CT-derived input data.

4. Discussion

The results illustrate that the network accurately determines 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,

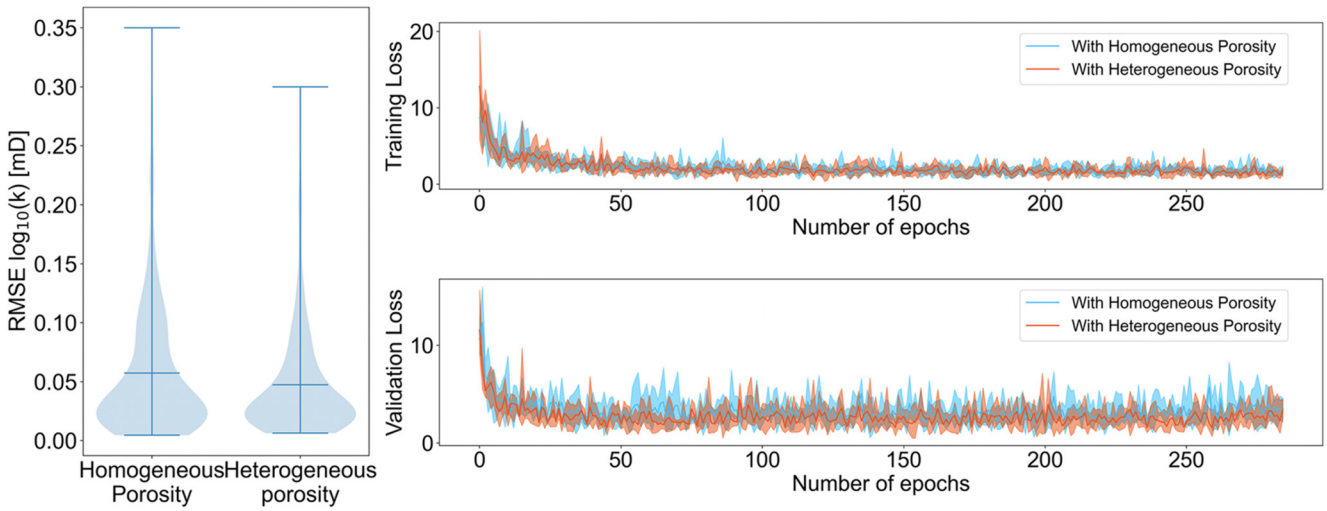

Figure 10. Training and testing performance of both networks trained with and without heterogeneous porosity maps, including training loss (top right plot), validation loss (bottom right plot), and the distribution of the RMSE accuracy of the permeability map predictions based on the test set data (left plot). The loss curves are the combination of five independent training/validation trials. The upper and lower bounds of the loss curve area at every epoch illustrate the maximum and the minimum loss values from the five trials, respectively. The bold line through the middle of the loss curve area illustrates the mean of the five training/validation trials. For the test set, the average RMSE of all of the predicted $\log_{10}$-permeability maps using the network trained without heterogeneous porosity maps was 0.057, and the average RMSE of all of the predicted $\log_{10}$-permeability maps using the network trained with heterogeneous porosity maps was 0.047. These plots illustrate the improved training and predictive performance of the network when a consistent relationship exists between porosity and permeability for a given sample—as is the case for the synthetic training datasets with heterogeneous porosity.

in many cases, the structure of the permeability variation can distort obvious relationships with arrival times as indicated by Figures 7 and 8.

Statistical analysis of the inversion results summarized in the left plot of Figure 10 indicates that the RMSE of the network-predicted permeability relative to the original synthetic permeability field was consistently low across the entire range of 500 test set permeability maps. Analysis of the test set RMSE as a function of mean permeability, permeability field standard deviation, and mean correlation length indicates that there was no correlation between RMSE and permeability field characteristics. Plots illustrating RMSE as a function of permeability field characteristics are provided in Figures S4, S5, and S6 in the Supporting Information S1. The lack of correlation between test set RMSE and 3-D permeability field correlation length statistics indicates that there was minimal feature loss resulting from feature smoothing during the encoding and decoding process. This verifies that using an encoding-decoding network reduces network training computational and memory requirements while maintaining the robustness of permeability inversion.

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 to the 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. 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 pre-processing and dimension reduction using classic transport analysis methods converts the raw 4-D datasets down to 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 10. However, there were minimal improvements or even worse predictions in the

experimental data inversion as illustrated in Figure 9. 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 9 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 with scaled X-ray CT scans input data 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.

5. Implications

This study demonstrates a new method for rapid 3-D permeability inversion of geologic core samples. This inversion was performed by applying a deep encoder-decoder based convolutional neural network—utilizing multilevel residual learning strategy and the dense connection structure—to massive image-based datasets. The network accurately predicted the local patterns and magnitude of the 3-D permeability maps using local arrival time difference maps. In addition to synthetic training data, the network was validated on experimentally measured data from PET scans and core-average permeability measurements on four different sandstone and carbonate core samples.

The trained network can generate a permeability map prediction—in less than ten seconds—of any geologic core sample with a permeability field defined by the training data parameter space. This rapid reduction in computational time and memory requirements arises from the data-driven nature of this approach (Kim & Nakata, 2018; Masci et al., 2011), where the trained encoder-decoder based CNN is capable of learning a generalizable mapping function between the input arrival time difference maps and output permeability maps. As long as the training data represent the data distribution for the inversion problem, the mapping function serves as a pseudo-inverse operator (Adler & Ozan, 2017). This approach avoids the computational complexity of common numerical optimization and nonlinear approaches (Carrera et al., 2005; Yeh, 1986) that would require iterative forward simulation to estimate the permeability map for a single core sample.

The orders of magnitude reduction in multiscale permeability inversion time provides an opportunity for a paradigm shift in core scale analysis and characterization. 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 3-D permeability characterization is a key starting point for disentangling the impacts of heterogeneity on multiphase flow observations and a range of reactive transport and colloidal transport processes.

Acknowledgments
Python scripts for training data generation, data analysis, CNN operation, and trained network parameters are permanently available at (Z. Huang & Zahasky, 2022). 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

Adler, J., & Ozan, k. (2017). Solving ill-posed inverse problems using iterative deep neural networks. *Inverse Problems*, *33*, 124007. https://doi.org/10.1088/1361-6420/aa9581

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

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. https://doi.org/10.1111/gwat.12413

Barbastathis, G., Ozcan, A., & Situ, G. (2019). On the use of deep learning for computational imaging. *Optica*, *6*, 921–943. https://doi.org/10.1364/OPTICA.6.000921

Bengio, Y. (2012). Practical recommendations for gradient-based training of deep architectures. In G.Montavon, G. B.Orr, & K.Müller (Eds.), *Neural Networks: Tricks of the trade* (2nd ed., Vol. 7700, pp. 437–478). Springer. https://doi.org/10.1007/978-3-642-35289-8_26

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. https://doi.org/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*. https://doi.org/10.1016/j.petrol.2017.10.033

Canchumuni, S. W. A., Emerick, A., & Pacheco, M. (2019). Towards a robust parameterization for conditioning facies models using deep variational autoencoders and ensemble smoother. *Computers & Geosciences*, *128*, 87–102. https://doi.org/10.1016/j.cageo.2019.04.006
Carrera, J., Alcolea, A., Medina, A., Hidalgo, J., & Slooten, L. J. (2005). Inverse problem in hydrogeology. *Hydrogeology Journal*, *13*(1), 206–222. https://doi.org/10.1007/s10040-004-0404-7
Chilingar, G. (1964). Relationship between porosity, permeability, and grain-size distribution of sands and sandstones. *Developments in Sedimentology*, *1*, 71–75. https://doi.org/10.1016/s0070-4571(08)70469-2
Chilingarian, G. (1991). *Empirical expression of permeability in terms of other petrophysical properties* (pp. 49–55). New York: Springer Science+Business Media. https://doi.org/10.1007/978-1-4899-0617-5_5
Dagan, G., Cvetkovic, V., & Shapiro, A. (1992). A solute flux approach to transport in heterogeneous formations 1. The general framework. *Water Resources Research*, *28*, 1369–1376. https://doi.org/10.1029/91WR03086
Dagan, G., & Nguyen, V. (1989). A comparison of travel time and concentration approaches to modeling transport by groundwater. *Journal of Contaminant Hydrology*, *4*(1), 79–91. https://doi.org/10.1016/0169-7722(89)90027-2
Deutsch, J. L., & Deutsch, C. V. (2012). Latin hypercube sampling with multidimensional uniformity. *Journal of Statistical Planning and Inference*, *142*(3), 763–772. https://doi.org/10.1016/j.jspi.2011.09.016
DeVries, T., & Taylor, G. W. (2017). *Improved regularization of convolutional neural networks with cutout*. Retrieved from http://arxiv.org/abs/1708.04552
Ferno, M. A., Hauge, L. P., Uno Rognmo, A., Gauteplass, J., Graue, A., Fernø, M. A., et al. (2015). Flow visualization of CO2 in tight shale formations at reservoir conditions. *Geophysical Research Letters*, *42*(18), 7414–7419. https://doi.org/10.1002/2015GL065100
Glatz, G., Castanier, L., & Kovscek, A. (2016). Visualization and quantification of thermally induced porosity alteration of immature source rock using x-ray computed tomography. *Energy & Fuels*, *30*, 8141–8149. https://doi.org/10.1021/acs.energyfuels.6b01430
Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In *The 13th international conference on artificial intelligence and statistics (AISTATS)* (Vol. 9, pp. 249–256).
Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. Retrieved from http://www.deeplearningbook.org
Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., et al. (2018). Recent advances in convolutional neural networks. *Pattern Recognition*, *77*, 354–377. https://doi.org/10.1016/j.patcog.2017.10.013
Guo, Q., Shi, X., Xueyuan, K., Hao, S., Liu, L., & Wu, J. (2021). Evaluation of the benefits of improved permeability estimation on high-resolution characterization of DNAPL distribution in aquifers with low-permeability lenses. *Journal of Hydrology*, *603*, 126955. https://doi.org/10.1016/j.jhydrol.2021.126955
Han, D., Wang, Z., Wang, Q., Wu, B., Yu, T., & Wang, D. (2019). Analysis of the Kozeny-Carman model based on pore networks. *Journal of Geophysics and Engineering*, *16*(6), 1191–1199. https://doi.org/10.1093/jge/gxz089
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. https://doi.org/10.3133/tm6a16
Harvey, C. F., & Gorelick, S. M. (1995). Mapping hydraulic conductivity: Sequential conditioning with measurements of solute arrival time, hydraulic head, and local conductivity, *Water Resources Research*, *31*(7), 1615–1626. https://doi.org/10.1029/95wr00547
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). https://doi.org/10.1109/cvpr.2016.90
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 *IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 2261–2269). https://doi.org/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*. https://doi.org/10.25740/gz610dt4642
Huang, Z., & Zahasky, C. (2022). Neural_network_inversion: Post-publication release (v1.2). *Zenodo*. https://doi.org/10.5281/zenodo.6354768
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).
Jackson, S., Lin, Q., & Krevor, S. (2020). Representative elementary volumes, hysteresis, and heterogeneity in multiphase flow from the pore to continuum scale. *Water Resources Research*, *56*. https://doi.org/10.1029/2019WR026396
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*. https://doi.org/10.1029/2021gl094443
Keskar, N. S., Mudigere, D., Nocedal, J., Smelyanskiy, M., & Tang, P. T. P. (2017). On large-batch training for deep learning: Generalization gap and sharp minima. In *5th international conference on learning representations, ICLR 2017, Toulon, France, April 24–26, 2017, conference track proceedings*. OpenReview.net. Retrieved from https://openreview.net/forum?id=H1oyRlYgg
Kim, Y., & Nakata, N. (2018). Geophysical inversion versus machine learning in inverse problems. *The Leading Edge*, *37*, 894–901. https://doi.org/10.1190/tle37120894.1
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. In *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). A procedure for the accurate determination of sub-core scale permeability distributions with error quantification. *Transport in Porous Media*, *98*, 565–588. https://doi.org/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. https://doi.org/10.1038/s41598-018-25413-9
Kurotori, T. (2019). Characterisation of solute transport in heterogeneous porous media by multidimensional imaging and modelling (Doctoral dissertation, Imperial College London). https://doi.org/10.25560/81654
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. https://doi.org/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. https://doi.org/10.1016/j.ces.2018.11.001
Larsen, A., Sønderby, S., & Winther, O. (2016). Autoencoding beyond pixels using a learned similarity metric. In *The 33rd international conference on machine learning* (Vol. 48).
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. https://doi.org/10.1103/PhysRevE.94.053107

Lu, L., Shin, Y., Su, Y., & Em Karniadakis, G. (2020). Dying ReLU and initialization: Theory and numerical examples. *Communications in Computational Physics*, *28*(5), 1671–1706. https://doi.org/10.4208/cicp.OA-2020-0165. Retrieved from http://global-sci.org/intro/article_detail/cicp/18393.html

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

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. https://doi.org/10.1029/WR022i013p02017

Masci, J., Meier, U., Ciresan, D., & Schmidhuber, J. (2011). Stacked convolutional auto-encoders for hierarchical feature extraction. In *ICANN* (pp. 52–59). https://doi.org/10.1007/978-3-642-21735-7_7

McCann, M. T., Jin, K. H., & Unser, M. A. (2017). Convolutional neural networks for inverse problems in imaging: A review. *IEEE Signal Processing Magazine*, *34*, 85–95. https://doi.org/10.1109/msp.2017.2739299

Mckay, M., Beckman, R., & Conover, W. (1979). A comparison of three methods for selecting vales of input variables in the analysis of output from a computer code. *Technometrics*, *21*, 239–245. https://doi.org/10.1080/00401706.1979.10489755

Minto, J., Hingerl, F., Benson, S., & Lunn, R. (2017). 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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/10.1029/2019WR026082

Mostaghimi, P., Blunt, M. J., & Bijeljic, B. (2013). Computations of absolute permeability on micro-CT images. *Mathematical Geosciences*, *45*(1), 103–125. https://doi.org/10.1007/s11004-012-9431-4

Müller, S., & Schüler, L. (2021). GeoStat-framework/GSTools: V1.3.0 'pure pink'. *Zenodo*. https://doi.org/10.5281/zenodo.4687075

Narsilio, G., Buzzi, O., Fityus, S., Yun, T., & Smith, D. (2009). Upscaling of Navier-Stokes equations in porous media: Theoretical, numerical and experimental approach. *Computers and Geotechnics*, *36*, 1200–1206. https://doi.org/10.1016/j.compgeo.2009.05.006

Noh, H., You, T., Mun, J., & Han, B. (2017). Regularizing deep neural networks by noise: Its interpretation and optimization. In *Advances in neural information processing systems, 2017-Decem(Nips)* (pp. 5110–5119).

Panahi, M., Sadhasivam, N., Pourghasemi, H., Rezaie, F., & Lee, S. (2020). Spatial prediction of groundwater potential mapping based on convolutional neural network (CNN) and support vector regression (SVR). *Journal of Hydrology*. https://doi.org/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. https://doi.org/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. https://doi.org/10.1029/2019wr026715

Sudakov, O., Burnaev, E., & Koroteev, D. (2019). Driving digital rock towards machine learning: Predicting permeability with gradient boosting and deep neural networks. *Computers & Geosciences*, *127*. https://doi.org/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*. https://doi.org/10.1016/j.cma.2020.113636

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. https://doi.org/10.1029/2019wr026731

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

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

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

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

Wang, Y. D., Blunt, M., Armstrong, R., & Mostaghimi, P. (2021). Deep learning in pore scale imaging and modeling. *Earth-Science Reviews*, *215*. https://doi.org/10.1016/j.earscirev.2021.103555

Wu, J., Zhang, C., Xue, T., Freeman, W., & Tenenbaum, J. (2016). Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling. In *2016 advances in neural information processing systems (NIPS)* (Vol. 29).

Yeh, W. W. (1986). Review of parameter identification procedures in groundwater hydrology: The inverse problem. *Water Resources Research*, *22*(2), 95–108. https://doi.org/10.1029/WR022i002p00095

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. https://doi.org/10.1016/j.advwatres.2018.03.002

Zahasky, C., & Benson, S. M. (2019). Spatial and temporal quantification of spontaneous imbibition. *Geophysical Research Letters*, *46*(21), 11972–11982. https://doi.org/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. https://doi.org/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. https://doi.org/10.1016/j.advwatres.2019.03.003

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

Zheng, C., & Wang, P. P. (1999). MT3DMS—A modular three-dimensional multispecies transport model. *Strategic Environmental Research and Development Program*, 1–40.

Zhong, Z., Sun, A., & Jeong, H. (2019). Predicting CO_2 plume migration in heterogeneous formations using conditional deep convolutional generative adversarial network. *Water Resources Research*, *55*. https://doi.org/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. https://doi.org/10.1016/j.jcp.2018.04.018. Retrieved from http://www.sciencedirect.com/science/article/pii/S0021999118302341