Comparison of Three-Dimensional Permeability Inversion from Positron Emission Tomography Experimental Data Using Convolutional Neural Networks and Ensemble Kalman Filter

Zitong Huang[1], and *Christopher* Zahasky[1,*]

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

Abstract. In the field of core analysis, parameterization of permeability variation is the foundation for quantifying the role of geologic heterogeneity on a wide range of conservative, reactive, and colloidal transport processes. Recent utilization of *in situ* imaging, specifically positron emission tomography (PET), enables the measurement of three-dimensional (3-D) time-lapse radiotracer transport in geologic porous media. This experimental tool provides unprecedented spatial and temporal data that can capture millimeter-scale spatial variation in transport phenomena that enables the quantification of heterogeneous multiscale permeability in geologic cores. Using spatially and temporally resolved concentration measurements, an ensemble Kalman filter (EnKF) is first implemented for inverting for the subcore scale permeability in a sandstone core. In addition to this traditional inversion approach, an encoder-decoder based convolutional neural network (CNN) is trained to predict the 3-D subcore scale permeability map from the same PET experimental data. Through forward numerical models, the permeability inversion accuracy of the trained CNN is compared with the EnKF on an experimental PET imaging dataset acquired using a heterogeneous Berea sandstone core. The results indicate that, compared to the EnKF, a single trained CNN can capture the variable solute arrival time behavior in a core-flooding experiment with orders of magnitude improvement in computational efficiency. Finally, mild permeability perturbations are added to the CNN inverted permeability map as an approach for generating the initial EnKF ensemble to further examine the optimal balance between the inversion accuracy and computational efficiency. This experimental approach combined with deep learning and numerical workflows provides an unprecedented improvement in 3-D multiscale permeability determination in heterogeneous geologic core samples.

1 Introduction

Multiscale permeability heterogeneity is a fundamental challenge to quantitative analysis and interpretation of flow and transport observations during core-scale experiments. Permeability heterogeneity is present in all geologic materials, even the most uniform and well-characterized formations, such as Berea sandstone [1] and Bentheimer sandstone [2]. Improved understanding of multiphase flow processes, fines migration, and enhanced recovery technologies first requires the estimation of initial permeability conditions, potential permeability evolution, and spatial variation in permeability. Without rapid methods for 3-D permeability map determination, advanced understanding of these core-scale processes may be limited.

Medical, industrial, and synchrotron-based imaging methods applied to problems in the field of petroleum engineering and hydrogeology have revolutionized the understanding of physical processes from the nanometer to the meter scale [3–6]. Photon transmission imaging techniques such as X-ray computed tomography (X-ray CT) across a range of scales excel at characterizing materials with different electron densities. As a result, at the micrometer scale, X-ray CT is ideal for mapping pore geometry and fluid interfaces [7, 8]. 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 [3, 9, 10]. More recently, positron emission tomography (PET) has emerged as a non-destructive imaging technique for quantifying conservative tracer and colloid transport in geologic materials [5, 11]. Using PET, it is possible to map 3-D time-lapse tracer distributions collected during core-flooding experiments. This has enabled studies of fundamental flow and transport processes such as solute tailing driven by diffusion into microporous carbonates [12], flow path alteration in fractured carbonates [13], fluid saturation [14, 15], fines migration in reservoir formations [11], and the rock compartmentalization driven by the presence of deformation bands [16].

Despite the aid provided by the high-resolution characterization of tracer and colloid transport, robust inversion of flow and transport parameters from image-based data remains challenging due to the intrinsic ill-posedness and high computational cost associated with the inversion problems. The non-linearity in the subsurface transport processes and the heterogeneity in the geologic materials

* Corresponding author: czahasky@wisc.edu

make subsurface image inversion problems often ill-posed [17]. Because of the ill-posedness, small uncertainties in the input observations due to imaging artifacts or measurement error can cause large uncertainties in the inversion results, and it is difficult for the inversion to converge to a unique maximum likelihood solution [18]. To reduce the ill-posedness of the numerical inversion problems, iterative regularization techniques based on certain prior information are explicitly imposed to only preserve the optimal set of features (or basis vectors) and prevent the amplification of the uncertainties in observations [17, 19]. However, previous studies have shown that the explicitly imposed regularization terms are either case-specific for over-simplified datasets or unable to incorporate all the prior information numerically for realistic datasets [20].

A PET image at a single timestep is often composed of over 10,000 voxel-level concentration measurements in a geologic core, and therefore the computational cost of utilizing all the information for 3-D permeability inversion can be high. Specifically, the high computational cost associated with the numerous forward modeling at every iteration can make the inversion of high-dimensional and geologically realistic permeability maps often unfeasible [21]. The need for a large number of forward models makes ensemble smoother and ensemble Kalman filter (EnKF) based inversion methods very computationally expensive [22–26]. For a 2-D inversion problem with less than 2,000 grid cells, the ensemble-based inversion methods would typically involve repeated flow and transport simulations on an ensemble of $O(10^2)$ models to get one complete inversion [27]. Development of high precision sub-core scale permeability inversion algorithms involving iterative J-Function characteristic calculations and coreflooding simulations is another core-scale permeability inversion approach [29]. Given the accurate porosity, saturation, and reasonable initial permeability maps, a convergence in sub-core scale permeability distribution generally took four to nine full iterations. Therefore, the reduction in computational time is often a trade-off with the reduction in experimental data collection time and resources.

In contrast, through a combination of weight decay, batch normalization, drop out, and early stopping, deep learning inversion methods can enforce the sparsity and smoothness in the solution spaces during the weight optimization process. Consequently, the encoder–decoder based convolutional neural networks (CNN) has been shown to address the ill-posedness problem by reducing the effects of noise [30], capturing features at a finer resolution [19], and learning a generalizable regularizer based on the training dataset without explicitly enforcing any constraints [20]. Moreover, instead of utilizing limited conditioning data, CNNs can be trained with large ensembles of flow and transport field characteristics (i.e. permeability) that span a diverse set of geologic materials in typical unfractured aquifers using classic methods such as Latin hypercube sampling. The representative set of training data allows CNNs to identify less significant features and capture the maximum amount of variability with the minimum number of features [17], thus alleviating the ill-posedness of the inverse problem.

In this study, we first perform PET data dimension reduction, quantile-based analysis, and normalization procedures [31] to emphasize transport characteristic information that is most important for permeability distribution determination. This processing reduces the influence of experimental imaging noise, solute tailing behavior, variation in initial solute concentration, and solute dispersion. We then use this data to compare the inversion performance between the deep learning approach, specifically the recently developed encoder-decoder based CNN [31], and a traditional numerical inversion approach, specifically EnKF, in terms of computational cost and inversion accuracy. To further enhance these methods, we also demonstrate the use of a hybrid CNN with EnKF tuning to maximize the permeability inversion accuracy while minimizing the associated computational cost.

2 Methods

2.1 Experimental Data Acquisition

The details of the PET data acquisition, imaging system, and experimental platform can be found in previous work [5]. For this study, a 10 cm long Berea sandstone sample was loaded into a custom aluminium coreholder that enabled the application of confining pressure and thus no-flow boundary conditions on the cylindrical faces of the samples. The core was saturated with low-pressure CO_2 and then injected with water while applying backpressure at the outlet face to prevent gravity-driven desaturation during horizontal fluid injection. The differential pressure was monitored, and steady-state conditions were determined to have been reached when the differential pressure stabilized. Core-average permeability was calculated to be 23 mD using Darcy's Law based on measurements of steady-state differential pressure, core geometry, and fluid injection rate.

For the imaging experiment, the water injection rate was set to 2 mL/min (0.245 cm/min). The positron-emitting radiotracer—fludeoxyglucose (^{18}F-FDG) —was diluted in water to a radio-concentration of 0.323 mCi/mL to reach the optimal radioactivity injected for minimizing imaging noise [5]. Fludeoxyglucose is a commercially available conservative tracer with a half-life of 109.7 minutes with high radioactive yields. High radioactive yields enable high signal radiotracer concentrations without altering aqueous fluid properties such as density—an advantage over X-ray computed tomography-based tracer approaches [5]. The PET scans were performed using a Siemens pre-clinical Inveon DPET scanner. Shortly after the scan started a 4 mL pulse of radiotracer (1.292 mCi) was injected into the core and displaced with water containing no ^{18}F -FDG. An illustration of four frames of the time-lapse reconstructed PET image is shown in Figure 1.

The reconstructed PET data was further processed to reduce the dimensionality of the data for use in the permeability inversion workflows. In short, quantile analysis was performed on every voxel of the PET images to calculate the 0.5 quantile, equivalent to the median arrival time and therefore insensitive to potential tailing effects in more heterogeneous cores. This quantile calculation enables the reduction of the 4D PET data down to a 3D image that provides the necessary information of permeability inversion with added benefits of reducing image noise and reducing the influence of sample-specific dispersion. The resulting quantile arrival time map for the Berea sandstone core was then normalized by converting from time to pore volumes injected. Additional details of this dimension reduction can be found in previous work [31, 32]. The final data processing step was to remove the linear trend in quantile arrival times that results from the bulk 1D flow from one end face of the core to the other. This linear trend was removed from each voxel by subtracting the quantile arrival time that would occur in a homogeneous core based on the distance of each voxel from the inlet of the core. The result after this trend removal is termed the arrival time difference map. These difference maps highlight areas of faster advection, with values greater than zero, and areas of slower advection, with values less than zero. The experimental arrival time difference map serves as input into the inversion workflow. A plot of the arrival time maps is provided in the results section.

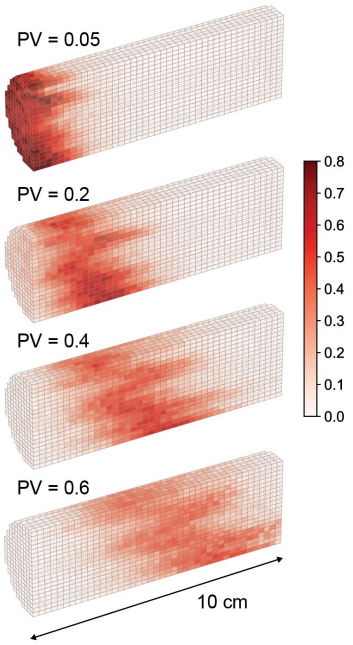

Fig. 1. Positron emission tomography time frames measuring radiotracer concentration distribution in the Berea sandstone core that was used in this study. The pore volumes injected (PV) are indicated above each image and fluid is injected from the left face of the core. The radiotracer was [18F]-FDG. The voxel size dimensions are 0.2329 x 0.2329 x 0.2388 cm^3. These images highlight the permeability heterogeneity of the Berea sample that resulted in faster tracer breakthrough in the upper middle and lower portion of the core. Corresponding X-ray CT images confirm horizontal planar laminations in the lower portion of the core with subtle cross-bedding containing higher clay content in the upper portion of the core.

2.2 Training data generation

An encoder-decoder based CNN was constructed and trained on 16,000 pairs of synthetic permeability and arrival time difference maps. Specifically, 10,000 pairs were used as the training dataset—for providing learnable features, 5,500 pairs were used as the validation datasets—for monitoring and guiding the CNN during the training, and 500 pairs were used as the testing set—for verifying the CNN after the training. To account for the instrumental error during the PET imaging process, a Gaussian white noise with a mean of zero and a standard deviation of 1/70 of the range is applied to every arrival time map. A similar level of noise is observed in experimental arrival time maps.

The 3-D synthetic permeability maps were generated via the random field generation algorithm using the exponential covariance model using open-source package developed by [33]. The grid cells for all the synthetic permeability maps have dimensions of 0.25 × 0.25 × 0.25 cm^3, and dimensions of 20 x 20 x 40 cells such that the model dimensions are nearly identical to the core samples and corresponding PET images. The synthetic permeability maps were sampled from parameters space with the mean ranging from 1 mD–20 D, log10 variance ranging from −1.7 to 9.9. In each x, y, and z directions, the correlation length of all synthetic permeability maps spans from 0.25 to 12.5 cm—spanning sedimentological lengths scales contained in core-scale samples—and the angle of rotation spans from 0 to 90°. Three example permeability fields are shown in Figure 2. By using the Latin hypercube sampling algorithm [34], all the training, validating, and testing data were well-balanced and representative of this parameter space. Three example permeability fields are illustrated in the left column of Figure 2.

Using MODFLOW 2005 with MT3DMS [35], a set of arrival time maps were numerically simulated for each synthetic permeability map. The simulated arrival time map also has a dimension of 20 × 20 × 40, and it underwent the same pre-processing and dimension reduction procedures as the 3-D experimental PET data for calculating arrival time difference maps. Three example arrival time difference maps calculated from synthetically generated permeability fields are illustrated in the right column of Figure 2.

To further guide the training and preserve the uniqueness of the inversion, the average permeability ($\underline{k}$) for every synthetic core was calculated by Darcy's Law based on modeled pressure drop data for each synthetic permeability map.

$$\underline{k} = \frac{Q_t}{A} \cdot \frac{L}{\Delta P} \cdot \mu \qquad (1)$$

Identical to the experimental conditions, all the synthetic cores had a length (L) of 10 cm and a cross-sectional area (A) of 78.5 cm^2. The flow rate through the synthetic core (Q_t) was set to 2 mL/min, and the pressure drop across the synthetic cores (ΔP) was calculated by subtracting the simulated

average pore pressure difference between the outlet slice and the inlet slice. The average permeability was expanded to a 1 × 20 × 20 tensor and then concatenated to the inlet of the input arrival time map. Therefore, the overall dimension of the input is 20 × 20 × 41.

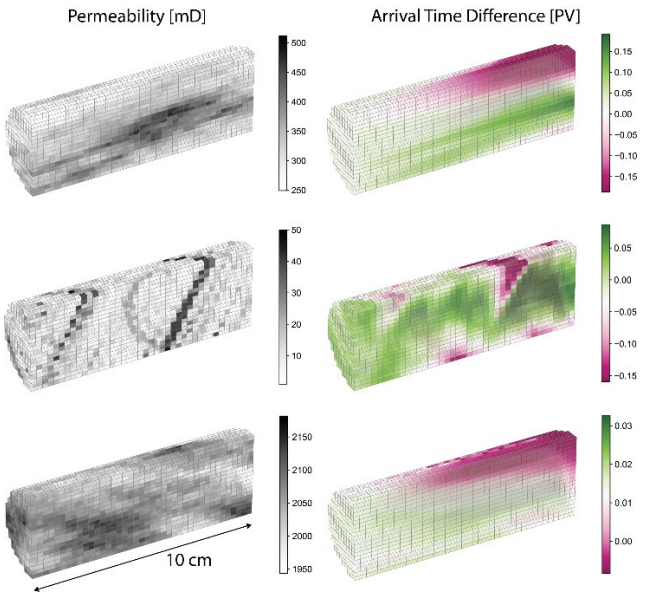

Fig. 2. Three example training data used for CNN training and test. The left column shows synthetically generated permeability field and the right column shows the corresponding arrival time difference map calculated from numerical simulation output.

2.3 Convolutional Neural Network Architecture

Convolution, sampling, and regression are the three main components of the encoder-decoder based CNN. The convolution blocks are composed of convolutional layers, that are used to gain an overall parameterization of the image. Sampling blocks are used to extract the high-frequency features of the input through downsampling and verify the learning through upsampling. The regression blocks are composed of a series of dense blocks connected in a residual-in-residual structure, which are used to perform deep regression over the sampled images.

A convolutional layer extracts features from input images through: $\mathbf{x}^{(l+1)} = \mathbf{f}_{l+1}(\mathbf{W}^{(l+1)}\mathbf{x}^{(l)} + \mathbf{b}^{(l+1)})$, where $\mathbf{W}^{(l+1)}$ is the weight matrix (or kernel), $\mathbf{b}^{(l+1)}$ is the bias vector, and $\mathbf{f}_{l+1}$ is the nonlinear activation function that maps the input map $\mathbf{x}^{(l)}$ to a corresponding output map $\mathbf{x}^{(l+1)}$. In a convolutional layer, every neuron is linked to a receptive field, a region in the input that represents a particular feature. As the result, CNN captures smaller-scale features in the shallower layers and the more global information in the deeper layers. There are two main reasons for adopting CNN in the image-to-image regression problems. First, the parameter sharing features and sparse connectivity of CNNs require fewer parameters than the traditional neural networks [36], which reduces the computational cost of processing large datasets. Decompressing the input to latent spaces while learning further reduces the computational cost. Secondly, CNN is highly capable of finding spatial correlations within a map by enforcing a local connectivity pattern between neurons of adjacent layers [36], which enables high accuracy in voxel-wise inversion.

By connecting each convolutional layer with all its subsequent layers, Densely Connected Convolutional Networks (DenseNet or dense block) fully leverage the hierarchical advantages of CNNs by encouraging feature propagation, sharing, and reuse among all the layers [37]. 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 input. 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. The residual-in-residual dense block (RRDB) are composed of a stack of residual dense blocks connected in another residual structure [38, 39]. 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)** := β × **R(x)** + **x**, where β ∈ (0, 1] is the residual scaling factor [38].

The inverted permeability map $\hat{x}$ is then compared against the synthetic permeability map (or the ground truth) $\boldsymbol{x}$ through loss functions, and the loss is going to be minimized by back-propagating the weights and bias through gradient descent. The overall loss function (Equation 2) is a combination of L1 loss (Equation 3) and KL-Divergence loss (Equation 4) weighted by α.

$$L_{total} = L_{L1} + \alpha L_{KL} \tag{2}$$

The L1 loss evaluates the absolute error between the synthetic (or ground truth) permeability map ($\boldsymbol{x}$) and inverted permeability map ($\hat{x}$).

$$L_{L1} = \| \boldsymbol{x} - \hat{x} \|_1 \tag{3}$$

The KL-Divergence loss measures the similarity between the synthetic (or ground truth) permeability distribution ($P(\boldsymbol{x})$) and inverted permeability distribution ($Q(\hat{x})$).

$$L_{KL} = D(P(\boldsymbol{x}) \| Q(\hat{x}))_{KL} = P(\boldsymbol{x}) \cdot \log \frac{P(x)}{Q(\hat{x})} \tag{4}$$

The encoder of the CNN extracts the high-frequency features of the input through the convolutional block while compressing the input map down through the sampling block.

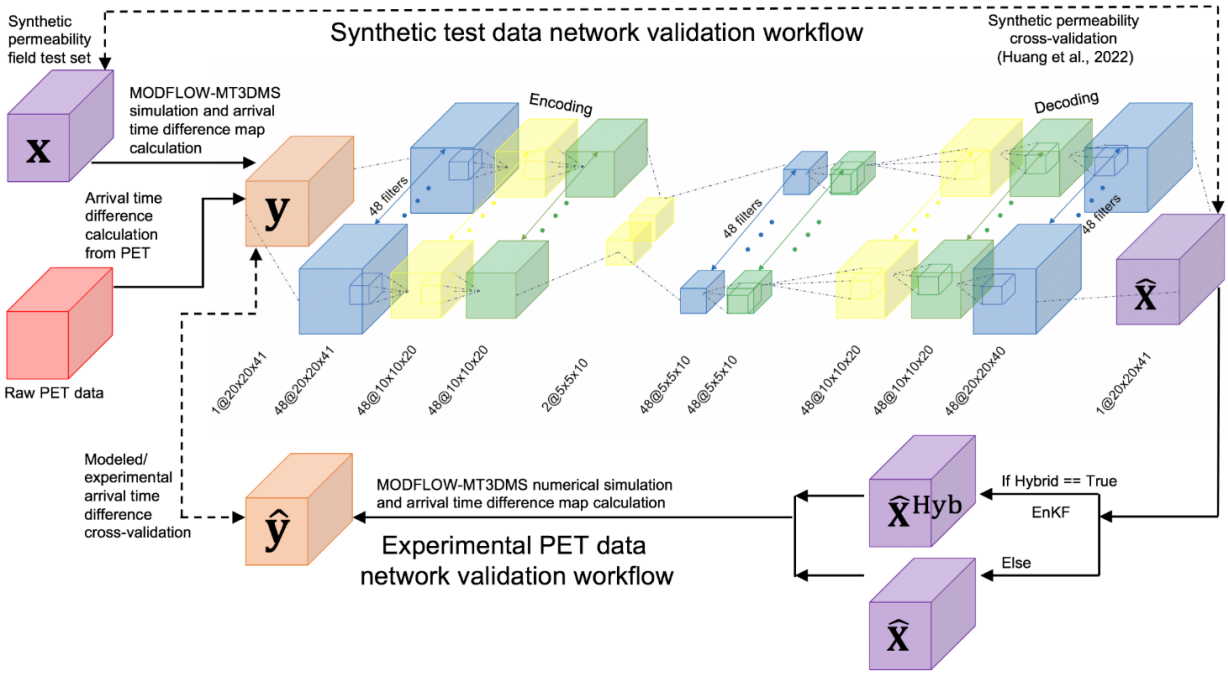

Fig. 3 Schematic of the overall inversion workflow. The top loop demonstrates the validation of the trained encoder-decoder based CNN using synthetic permeability maps [31]. The middle chain represents the CNN, including convolutional blocks (blue), up/down-sampling block (yellow), and residual-in-residual dense blocks (green). The bottom loop demonstrates the validation of the CNN using an experimental arrival time map collected from the Berea sandstone core. Note that the purple blocks correspond to synthetic ($\mathbf{x}$) or predicted ($\hat{\mathbf{x}}$) or EnKF tuned ($\hat{\mathbf{x}}^{Hyb}$) permeability maps, the red block is the PET data, and the orange blocks correspond to experimental ($\mathbf{y}$) and modeled ($\hat{\mathbf{y}}$) arrival time maps.

The compressed high-level features map is referred to as the latent space of the input map, which has a dimension of 5 × 5 × 10. The decoder then up transposes the latent map based on the extracted high-frequency features. The inverted (or decoded) permeability map has a dimension of 20 × 20 × 40.

To further validate the trained CNN performance under experimental settings, we run MODFLOW-MT3DMS flow simulations on the inverted permeability map $\hat{x}$ to generate modeled arrival time map for direct comparison with the experimental arrival time map. The R2 score is used to evaluate the similarity between the modeled and experimental arrival time maps (R2 = 1 indicates perfect inversion). A detailed illustration of the overall training and validation workflow is illustrated in Figure 3. Additional details of the CNN architecture, training, and convergence performance can be found in [31].

2.4 Ensemble Kalman Filter

To compare the encoder-decoder based CNN with the most computationally efficient numerical inversion algorithm, an Ensemble Kalman filter [23] has been adopted and constructed for this study. The EnKF is the Monte Carlo approximation to Kalman filtering for sequential data assimilation. An advantage of the EnKF for this application is that uncertainty characterization of EnKF on the ensemble of inverted outcomes is computationally efficient [40] and the EnKF only requires the ensemble predicted error statistics to update the inverted parameters. The computational efficiency and inversion robustness of EnKF have been demonstrated by various hydrogeological data assimilation tasks, such as contaminant source identification [40] and history matching [41]. In this study, we adopted the standard EnKF to characterize geologic core heterogeneity through continuum-scale permeability inversion.

To begin a standard EnKF inversion, an initial ensemble of augmented state vectors ($\boldsymbol{X}_{init} = [\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(np)}]$) needs to be generated. In this study we performed EnKF inversion with three different types of ensembles. The first was generated by slightly perturbing the permeability prediction of the CNN permeability prediction as described in the following section. The second and third EnKF inversion were performed using a spatially correlated ensemble of permeability field based on the expected correlation in a layered sandstone (i.e., low-uncertainty ensemble) and using an ensemble of permeability fields that are generated assume less information is available (i.e., high-uncertainty ensemble). The ensemble of permeability fields was generated using the same algorithm as the CNN training data described in Section 2.2. A summary of the ensemble parameter space is given in Table 1.

According to previous studies, an assumption of the EnKF is that the true permeability map of the sample with unknown permeability lies in the linear span of the initial ensemble space [42]. If $\boldsymbol{X}$ evolves with time, each $\boldsymbol{X}$ contains the state vector ($\mathbf{S}$) with a dimension $n_s \times 1$ and a parameter vector ($\mathbf{P}$) with a dimension $n_p \times 1$. Through state operators (e.g., numerical model), the state vector (e.g., fluid pressure distribution) will vary with time based on the inverted parameter vector (e.g., permeability) at the corresponding time step. However, since $\boldsymbol{X}$ in this study (i.e., the permeability map) is not varying with time and the observation operator (i.e., MODFLOW-MT3DMS model) does not require a state vector as input, $\boldsymbol{X}$ only contains the

parameter vector with a dimension of $n_p \times 1$, where n_p equals the size of the 1-D flattened permeability map plus the average permeability calculated by Equation 1. Based on the corresponding state vectors, an ensemble of initial observation vectors ($\boldsymbol{Y}_{init} = [\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(no)}]$), each with a dimension $n_o \times 1$, were simulated via Equation 5, where **f** is the observation operator. All the dimension parameters used in this study are presented in Table 2.

$$\boldsymbol{Y}_{init} = \mathbf{f}(\boldsymbol{X}_{init}) \tag{5}$$

At iteration i, the standard EnKF involves two main steps. The first step is to compute the Kalman gain matrix G for updating the parameter vector ensemble:

$$\mathbf{G} = \mathbf{P}_{xy} \cdot (\mathbf{P}_{yy} + \mathbf{R})^{-1} \tag{6}$$

Here $\mathbf{P}_{xy}$ is the $n_p \times n_o$ cross-covariance matrix between the inverted permeability and simulated arrival time from the previous iteration (Equation 7), $\mathbf{P}_{yy}$ is the $n_o \times n_o$ covariance matrix of the simulated arrival time from the previous iteration (Equation 8), and **R** is the diagonal measurement error matrix created by projecting the measurement error at every voxel (or observation point) onto a $n_o \times n_o$ identity matrix.

$$\mathbf{P}_{xy} = (X_{i-1} - \underline{X_{i-1}}) \cdot (Y_{i-1} - \underline{Y_{i-1}})^T \tag{7}$$

$$\mathbf{P}_{yy} = (Y_{i-1} - \underline{Y_{i-1}}) \cdot (Y_{i-1} - \underline{Y_{i-1}})^T \tag{8}$$

Here, $\underline{X_{i-1}}$ is a $n_p \times n_e$ matrix with each column containing the average of one vector in the parameter ensemble. Similarly, $\underline{Y_{i-1}}$ is a $n_p \times n_e$ matrix with each column containing the average of one vector in the observation ensemble. Then, every parameter vector in the previous X_{i-1} ensemble is updated with the Kalman gain matrix G via Equation 9:

$$X_i = X_{i-1} + \mathbf{G} \cdot (Y_{i-1} - Y_{true}) \tag{9}$$

Here $\boldsymbol{Y}_{true}$ is the 1-D flattened experimental arrival time map measured from the Berea core. The vector ($\boldsymbol{Y}_{i-1} - \boldsymbol{Y}_{true}$) is used to quantify the new observational error. The combination of the Kalman gain matrix and observation error vector helps the Kalman filter to assimilate the observations for the new iteration of parameter updates.

The second step is to simulate the new observation ensemble $\boldsymbol{Y}_i$ (i.e., arrival time maps calculated from MODFLOW-MT3DMS model output) based on the updated parameter ensemble $\boldsymbol{X}_i$ via the observation operator for the new error statistics. By the end of every iteration, the ensemble average of all parameter vectors was generated as the permeability inversion result. The ensemble standard deviation provides an estimation of the inversion uncertainty.

2.5 Hybrid CNN with EnKF tuning

In the final permeability inversion approach, the CNN inversion results were used to generate an initial ensemble for final tuning of the 3-D permeability prediction using the EnKF. The ensemble generation was performed by multiplying the CNN inverted permeability map by an ensemble of spatially correlated exponential covariance fields with a mean of 1 and variance ranging from 0.001 to 0.05; the correlation length and rotation angle of the fields were within the same range as the low-uncertainty parameter space presented in Table 1.

Table 1. The parameter space boundaries for generating the initial Berea EnKF permeability ensemble. Assuming no exact knowledge of the spatial information of the Berea permeability map, the permeability maps in the low-uncertainty ensemble are with a well-defined range of spatial information constrained by prior knowledge obtained from geologic knowledge and X-ray CT images on the core. In contrast, the permeability maps in the high-uncertainty ensemble are sampled from a larger parameter space. The mean permeability for both ensembles were calculated by Equation 1 using the data from core-flooding experiments.

Parameter Name	Low-uncertainty ensemble	High-uncertainty ensemble
Mean Permeability	23.2 mD	23.2 mD
log10 exponential variance	-3.0 - -2.0	-7.0 - -1.0
Correlation length in x, y directions	12.5 - 37.5 cm	0.25 - 25 cm
Correlation length in z directions	0 cm	0 cm
Rotation angle in x, y, z directions	0 Rad	0 - 3.14 Rad

Table 2. The dimension parameters used for EnKF.

Dimensional Parameter Name	Number of parameters
Number of ensemble (n_e)	400
Number of observation (n_o)	16,000
Number of parameters (n_p)	16,001

3 Results

For the encoder-decoder based CNN, each training took around 30 hours to complete on a Nvidia GeForce GTX980 GPU. All the 8,570,690 trained parameters were then stored in two separate path files (a 10.8 MB path file for the encoder model and a 22.1 MB path file for the decoder model). A summary of the training and validation loss is illustrated in Figure 4.

The EnKF inversion was an iterative process, and the time required to stabilize inversion accuracy heavily depends on the quality of the initial parameter ensembles. For the inversion via hybrid CNN with EnKF, the range of initial ensemble is very well-defined and constrained, and the

accuracy stabilization generally requires 3 to 4 hours. When the initial ensemble is less constrained, the range for the initial ensemble expands. The accuracy stabilization time was between 5 hours and 9 hours when sampling from low or high uncertainty ensemble parameter space.

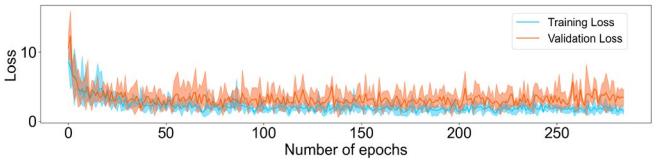

Fig. 4. An illustration of training and validation loss curves from five independent CNN training trails. At every epoch, the upper and lower bounds of the loss curve area indicate the maxima and the minima of the five loss curves. The bold middle line within the loss range represents the average loss of the five trials at every epoch.

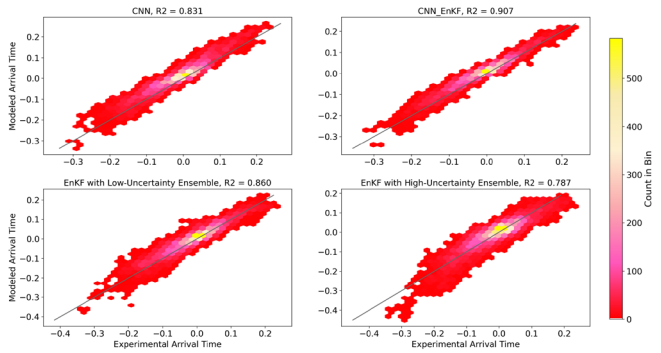

Fig. 5. Cross-plot of flattened experimental arrival time data (*x*-axis) and modeled arrival time difference based on the inverted permeability maps from CNN (top left), hybrid CNN with EnKF tuning (top right), EnKF with the low-uncertainty initial ensemble (bottom left), and EnKF with the high-uncertainty initial ensemble (bottom right). To illustrate the density of the correlations, the cross-plot is colored by the number of points in a local region of the cross-plot. The R2 correlation results indicate that the EnKF can bring significant improvement in permeability map prediction and the quality of the initial ensemble for the EnKF has a profound impact on the inversion accuracy.

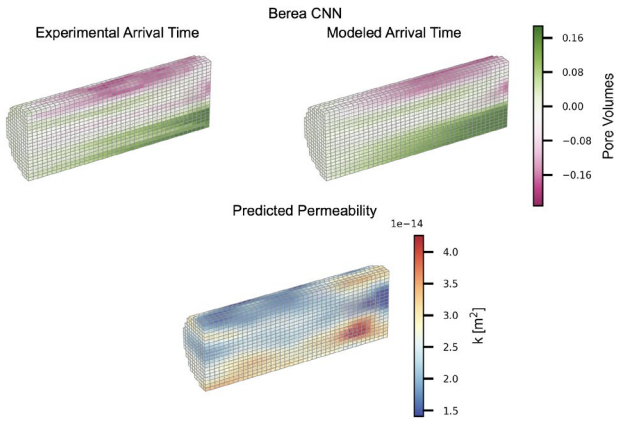

Fig. 6. Qualitative validation of the trained encoder-decoder based CNN using experimentally measured arrival time data measured with PET collected from the Berea sandstone core. The upper left subplots show the arrival time map calculated from the PET imaging data, the lower plot shows the predicted permeability by the network, and the upper right shows the modeled arrival time map based on the predicted permeability map. Note that the experimental and modeled arrival times have the same colorscale.

Three sets of initial ensembles with increasing levels of uncertainity (i.e., CNN-based, low uncertainty, and high uncertainty) were tested as the input for the EnKF inversion algorithm. Similar to the CNN validation workflow, modeled arrival time maps were generated based on the EnKF inverted permeability maps through MODFLOW-MT3DMS numerical simulation. As illustrated in Figure 5, the R2 correlation between the experimental and modeled arrival time decreases as the precision of the initial ensemble decreases. With an R2 accuracy of 0.891, the hybrid CNN with EnKF tuning excels both in terms of computational cost and inversion accuracy. As demonstrated in Figure 6 and 7, the EnKF algorithm captures more discrete heterogeneity that is more geologically realistic than the smoothed permeability map inversion produced by the CNN, resulting in higher inversion accuracy.

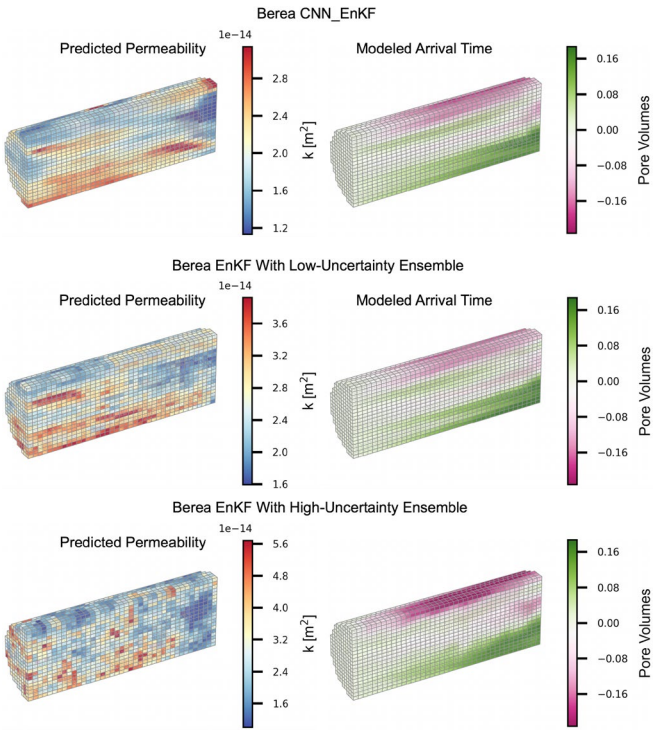

Fig. 7. A quantitative comparison of the inversion results for the hybrid CNN with EnKF tuning (top), the EnkF with the low-uncertainty initial ensemble (middle), and the EnkF with the high-uncertainty initial ensemble (bottom). The left subplots show the inverted 3-D permeability maps and the right subplots show the modeled arrival time maps.

4 Discussion

The high R2 correlation between the experimental and modeled arrival time maps illustrates that the trained encoder-decoder based CNN is capable of accurately determining the overall magnitude and local variations of 3-D permeability maps. With a reasonable initial ensemble, the Ensemble Kalman filter can achieve a similar or higher level of inversion accuracy as the encoder-decoder based CNN. As noted in previous inversion studies, the quality of the initial ensemble has a considerable impact on both the accuracy and computational efficiency of the inversion.

Trained CNNs are capable of describing the input imagery (e.g., PET data) in terms of dense and generalizable features, resulting in a significant reduction in computational time and memory requirements [43, 44]. Although the initial CNN training process is computationally intensive, the trained network can invert the permeability map of nearly any unfractured geologic core sample within 10 seconds. Furthermore, each path file that contains the trained parameters for the entire encoder-decoder network is only tens of megabytes.

Training of the encoder–decoder based CNN is purely data-driven, relying on statistical relationships instead of any iterative forward operator involved in the inversion process. Our work illustrates that the encoder–decoder based CNN is capable of learning a generalizable mapping function between the arrival time difference and permeability maps. The mapping function served as a pseudo-inverse operator [18], and it eliminates the need for numerically defining a forward operator, running iterative forward simulations, and computing covariance matrices. This significantly reduces computational time if inverting the permeability maps for multiple core samples.

As indicated by the inversion accuracy stabilization time, the EnKF requires high computational time and memory because of the iterative forward simulations and covariance matrices computations involved in the inversion process. For a system with $n = n_o = n_p$, the computational cost of the standard EnKF is at least in the order of $O(n_e^2 \cdot n)$ [45]. To mitigate this issue, both the ensemble size and the number of iterations need to be reduced by a well-constrained initial ensemble estimation.

Similar to CNN training data parameter space definition, the geologically reasonable EnKF ensemble definition is essential for accurate inversion results. Guided by the geologic knowledge and sedimentary structures present in X-ray CT core image obtained from previous experiments, the EnKF initialized with low-uncertainty permeability fields accurately capture the magnitude and variation of the Berea core's permeability map. Furthermore, the CNN inverted permeability map provides additional structural information that is difficult to defined manually to the EnKF inversion, resulting in the highest R2 inversion accuracy. In contrast, results of the EnKF initialized with the high-uncertainty permeability fields illustrate a lack of geologic realism.

In addition, if proper regularization techniques are not explicitly imposed, small uncertainties in the input observations arising from measurement error can result in large uncertainties in the inversion results due to the ill-posedness of the nonlinear inversion problem [17, 19]. For instance, as the uncertainty in the input ensemble increases, a clear reduction in the inversion accuracy is observed. This issue can be attenuated by using CNN to guide the generation of an initial ensemble with low uncertainty.

5 Conclusion

This study demonstrates the capability of both CNNs and traditional numerical inversion methods for determining subcore permeability maps from PET imaging data. The CNN can learn a generalizable relationship between the PET arrival time and permeability maps that allows a single trained network to invert for the permeability of any geologic sample represented by the training data parameter space. This enables an accurate 3-D permeability characterization of geologic cores. The data-driven nature of the CNN inversion approach has been shown to lead to an unprecedented reduction in computational time and memory requirements such that a single inversion can be run in seconds.

The EnKF numerical inversion algorithm requires more computationally intensive covariance calculations and observation operation that must be run for each sample-specific permeability field inversion. Similar to the CNN training data parameter space definition, the EnKF ensemble parameter space has a strong influence on inversion accuracy and computational efficiency. With a well-defined initial ensemble, the Ensemble Kalman filter may outperform the CNN results in terms of inversion accuracy. However, due to the variation and spatial complexity of different permeability fields in different reservoir rocks, prior knowledge about a specific geologic core will often be limited. As a solution, the CNN inverted permeability map with subtle perturbations that span the inversion uncertainty can serve as the starting point for generating a geologically realistic and sample specific ensemble with much lower uncertainty than traditional geostatistical ensemble generation methods. The hybrid CNN with EnKF tuning not only aids the EnKF to further improve its inversion accuracy but also broadens its capability at a larger scale or a finer resolution. Combined, this approach provides a transformational approach for rapid sub-core permeability characterization in complex geologic reservoir core samples.

Acknowledgments: Python scripts for training data generation, data analysis, CNN operation, and trained network parameters are permanently available at https://doi.org/10.5281/zenodo.6354768. The full training datasets and experimental data are permanently available at https://doi.org/10.25740/gz610dt4642. 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

[1] C. Zahasky and S. M. Benson, "Micro-Positron Emission Tomography for Measuring Sub-core Scale Single and Multiphase Transport Parameters in Porous Media," *Adv. Water Resour.*, vol. 115, pp. 1–16, 2018.

[2] C. Zahasky, S. J. Jackson, Q. Lin, and S. Krevor, "Pore network model predictions of Darcy-scale multiphase flow heterogeneity validated by experiments," *Water Resour. Res.*, pp. 1–16, 2020.

[3] S. Akin and A. R. Kovscek, "Computed Tomography in Petroleum Engineering Research," *Appl. X-ray Comput. Tomogr. Geosci.*, vol. 215, pp. 23–38, 2003.

[4] M. J. Blunt *et al.*, "Pore-scale imaging and modelling," *Adv. Water Resour.*, vol. 51, pp. 197–216, 2013.

[5] C. Zahasky, T. Kurotori, R. Pini, and S. M. Benson, "Positron emission tomography in water resources and subsurface energy resources engineering research," *Adv. Water Resour.*, vol. 127, no. March, pp. 39–52, 2019.

[6] L. Leu, S. Berg, F. Enzmann, R. T. Armstrong, and M. Kersten, "Fast X-ray Micro-Tomography of Multiphase Flow in Berea Sandstone: A Sensitivity Study on Image Processing," *Transp. Porous Media*, vol. 105, no. 2, pp. 451–469, 2014.

[7] R. T. Armstrong, M. L. Porter, and D. Wildenschild, "Linking pore-scale interfacial curvature to column-scale capillary pressure," *Adv. Water Resour.*, vol. 46, pp. 55–62, 2012.

[8] C. Garing, J. A. de Chalendar, M. Voltolini, J. B. Ajo-Franklin, and S. M. Benson, "Pore-scale capillary pressure analysis using multi-scale X-ray micromotography," *Adv. Water Resour.*, vol. 104, pp. 223–241, 2017.

[9] J. M. Minto, F. F. Hingerl, S. M. Benson, and R. J. Lunn, "X-ray CT and multiphase flow characterization of a 'bio-grouted' sandstone core: The effect of dissolution on seal longevity," *Int. J. Greenh. Gas Control*, vol. 64, no. September 2016, pp. 152–162, 2017.

[10] C. Zahasky and S. M. Benson, "Spatial and temporal quantification of spontaneous imbibition," *Geophys. Res. Lett.*, vol. 46, no. 21, pp. 11972–11982, Nov. 2019.

[11] C. Sutton and C. Zahasky, "An approach for image-based quantification of fines migration in geologic columns and core samples," in *SCA*, 2022.

[12] T. Kurotori, C. Zahasky, S. A. Hosseinzadeh Hejazi, S. M. Shah, S. M. Benson, and R. Pini, "Measuring, imaging and modelling solute transport in a microporous limestone," *Chem. Eng. Sci.*, vol. 196, pp. 366–383, 2019.

[13] B. Brattekas and R. S. Seright, "Implications for improved polymer gel conformance control during low-salinity chase-floods in fractured carbonates," *J. Pet. Sci. Eng.*, 2017.

[14] Y. Hu *et al.*, "Unsteady-State Coreflooding Monitored by Positron Emission Tomography and X-ray Computed Tomography," no. July 2018, pp. 1–11, 2019.

[15] M. A. Ferno *et al.*, "Combined positron emission tomography and computed tomography to visualize and quantify fluid flow in sedimentary rocks," *Water Resour. Res.*, vol. 51, no. 9, pp. 7811–7819, 2015.

[16] C. R. Romano *et al.*, "Sub-core scale fluid flow behavior in a sandstone with cataclastic deformation bands," *Water Resour. Res.*, pp. 1–16, 2020.

[17] E. K. Oware, S. M. J. Moysey, and T. Khan, "Physically based regularization of hydrogeophysical inverse problems for improved imaging of process-driven systems," *Water Resour. Res.*, vol. 49, no. 10, pp. 6238–6247, 2013.

[18] J. Adler and O. Öktem, "Solving ill-posed inverse problems using iterative deep neural networks," *Inverse Probl.*, vol. 33, no. 12, pp. 1–24, 2017.

[19] Y. Kim and N. Nakata, "Geophysical inversion versus machine learning in inverse problems," *Lead. Edge*, vol. 37, no. 12, pp. 894–901, 2018.

[20] M. T. Mccann, K. H. Jin, and M. Unser, "Deep CNN for Inverse Problems in Imaging," *IEEE Signal Process. Mag.*, vol. 26, no. 9, pp. 4509–4522, 2017.

[21] J. Fu and J. J. Gómez-Hernández, "A blocking markov chain Monte Carlo method for inverse stochastic hydrogeological modeling," *Math. Geosci.*, vol. 41, no. 2, pp. 105–128, 2009.

[22] P. J. van Leeuwen and G. Evensen, "Data Assimilation and Inverse Methods in Terms of a Probabilistic Formulation," *Mon. Weather Rev.*, vol. 124, no. 12, pp. 2898–2913, Dec. 1996.

[23] G. Evensen, "Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics," *J. Geophys. Res.*, vol. 99, no. C5, 1994.

[24] H. Zhou, J. J. Gómez-hernández, H. H. Franssen, and L. Li, "Advances in Water Resources An approach to handling non-Gaussianity of parameters and state variables in ensemble Kalman filtering," *Adv. Water Resour.*, vol. 34, no. 7, pp. 844–864, 2011.

[25] A. A. Emerick and A. C. Reynolds, "Ensemble smoother with multiple data assimilation," *Comput. Geosci.*, vol. 55, pp. 3–15, 2013.

[26] T. Xu and J. J. Gómez-Hernández, "Simultaneous identification of a contaminant source and hydraulic conductivity via the restart normal-score ensemble Kalman filter," *Adv. Water Resour.*, vol. 112, no. December 2017, pp. 106–123, 2018.

[27] S. W. A. Canchumuni, A. A. Emerick, and M. A. C. Pacheco, "Towards a robust parameterization for conditioning facies models using deep variational autoencoders and ensemble smoother," *Comput. Geosci.*, vol. 128, no. January, pp. 87–102, 2019.

[28] Q. Guo, X. Shi, X. Kang, S. Hao, L. Liu, and J. Wu, "Evaluation of the benefits of improved permeability estimation on high-resolution characterization of DNAPL distribution in aquifers with low-permeability lenses," *J. Hydrol.*, vol. 603, no. PB, p. 126955, 2021.

[29] M. Krause, S. Krevor, and S. M. Benson, "A Procedure for the Accurate Determination of Sub-Core Scale Permeability Distributions with Error Quantification," *Transp. Porous Media*, vol. 98, no. 3, pp. 565–588, 2013.

[30] G. Barbastathis, A. Ozcan, and G. Situ, "On the use of deep learning for computational imaging," *Optica*, vol. 6, no. 8, p. 921, Aug. 2019.

[31] Z. Huang, T. Kurotori, R. Pini, S. M. Benson, and C. Zahasky, "Three-Dimensional Permeability Inversion Using Convolutional Neural Networks and Positron Emission Tomography," *Water Resour. Res.*, vol. 58, no. 3, pp. 1–21, 2022.

[32] C. F. Harvey and S. M. Gorelick, "Mapping hydraulic conductivity: Sequential conditioning with measurements of solute arrival time, hydraulic head, and local conductivity," vol. 31, no. 7, pp. 1615–1626, 1995.

[33] S. Müller and L. Schüler, "GeoStat-Framework/GSTools: v1.3.0 'Pure Pink.'" Zenodo, 2021.

[34] J. L. Deutsch and C. V. Deutsch, "Latin hypercube sampling with multidimensional uniformity," *J. Stat. Plan. Inference*, vol. 142, no. 3, pp. 763–772, 2012.

[35] C. Zheng and P. P. Wang, "MT3DMS - A Modular Three-Dimensional Multispecies Transport Model," *Strateg. Environ. Res. Dev. Progr.*, pp. 1–40, 1999.

[36] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[37] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 2261–2269, 2017.

[38] X. Wang *et al.*, "ESRGAN: Enhanced super-resolution generative adversarial networks," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11133 LNCS, pp. 63–79, 2019.

[39] S. Mo, N. Zabaras, X. Shi, and J. Wu, "Deep Autoregressive Neural Networks for High-Dimensional Inverse Problems in Groundwater Contaminant Source Identification," *Water Resour. Res.*, vol. 55, no. 5, pp. 3856–3881, 2019.

[40] Z. Chen, J. J. Gómez-Hernández, T. Xu, and A. Zanini, "Joint identification of contaminant source and aquifer geometry in a sandbox experiment with the restart ensemble Kalman filter," *J. Hydrol.*, vol. 564, no. August, pp. 1074–1084, 2018.

[41] D. Erdal and O. A. Cirpka, "Preconditioning an ensemble Kalman filter for groundwater flow using environmental-tracer observations," *J. Hydrol.*, vol. 545, pp. 42–54, Feb. 2017.

[42] N. K. Chada, M. A. Iglesias, L. Roininen, and A. M. Stuart, "Parameterizations for Ensemble Kalman Inversion," *Inverse Problems*, vol. 34, no. 10, pp. 1361–1392, 2018.

[43] J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber, "Stacked convolutional auto-encoders for hierarchical feature extraction," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 6791 LNCS, no. PART 1, pp. 52–59, 2011.

[44] M. Norouzi, M. Ranjbar, and G. Mori, "Stacks of convolutional restricted boltzmann machines for shift-invariant feature learning," *2009 IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2009*, pp. 2735–2742, 2009.

[45] A. J. Majda and X. T. Tong, "Performance of Ensemble Kalman Filters in Large Dimensions," *Commun. Pure Appl. Math.*, vol. 71, no. 5, pp. 892–937, 2018.