A Transfer Learning-Based Surrogate Model for Geological Carbon Storage with Multi-Fidelity Training Data

Su Jiang
Department of Energy Resources Engineering
Stanford University
Stanford, CA 94305
sujiang@stanford.edu

Hewei Tang
Atmospheric, Earth, and Energy Division
Lawrence Livermore National Laboratory
Livermore, CA 94550
tang39@llnl.gov

Pengcheng Fu
Atmospheric, Earth, and Energy Division
Lawrence Livermore National Laboratory
Livermore, CA 94550
fu4@llnl.gov

Honggeun Jo
Department of Petroleum and Geosystems
Engineering, The University of Texas
Austin, TX 78712
honggeun.jo@utexas.edu

Abstract

Geologic carbon storage (GCS) entails injecting large volumes of carbon dioxide (CO_2) in deep geologic formations to prevent its release to the atmosphere. Reservoir simulation is widely used in GCS applications to predict subsurface pressure and CO_2 saturation. High fidelity numerical models are prohibitively expensive for data assimilation and uncertainty quantification, which require a large number of simulation runs. Deep learning-based surrogate models have shown a great promise to alleviate the high computational cost. However, the training cost is high as thousands of high-fidelity simulations are often necessary for generating the training data. In this work, we explore the use of a transfer learning approach to reduce the training cost. Compared with the surrogate model trained with high-fidelity simulations, our new transfer learning-based model shows comparable accuracy but reduces the training cost by 80%.

1 Introduction

We can prevent 90% of industry emissions from reaching the atmosphere by capturing the CO_2 at source and then permanently store them underground. The storage process is known as geologic carbon storage (GCS). Reservoir simulation is widely used in GCS reservoir management to predict subsurface pressure and CO_2 saturation. However, when used in the context of data assimilation and uncertainty quantification, high-fidelity numerical models are prohibitively expensive due to the large number of simulations required. Deep-learning-based surrogate models are emerging as an effective and efficient alternative to conventional reservoir simulators for GCS applications (Mo et al. [2019], Tang et al. [2021a], Tang et al. [2021b], Wen et al. [2021]).

To train a neural network that predicts reservoir responses on a grid consisting of $\sim$100,000 cells, thousands of reservoir simulations are usually required, incurring a high training cost. In this work, we explore the possibility of reducing the training cost by using multi-fidelity data for the training. In other words, we first use an ensemble of low-fidelity simulations to train the network. The low-fidelity ensemble embodies the same physics as the high-fidelity models but adopts a lower grid resolution to reduce the per-simulation computational cost by an order of magnitude. A small number of high-

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2021.

fidelity simulations with the original resolution are then employed to train and fine-tune the network. Note that "high/low fidelity" in this work is used interchangeably with "high/low resolution".

Such a transfer learning-based approach has been recently applied to subsurface surrogate model development such as in [Song and Tartakovsky, 2021] for a 2D Gaussian case. The challenges posed by our target application include the complexity of 3D two-facies geological modeling and simulation, and the complex well control for CO_2 injection. In this work, we apply transfer learning on 3D recurrent R-U-Net, developed by Tang et al. [2021c], to train a surrogate model for pressure prediction of a 3D GCS case with bimodal distribution.

2 Surrogate Model with Transfer Learning

Forward Simulation. The GCS problem is simulated as a multi-phase flow problem in porous media. Mass conservation for each phase and Darcy law are applied to solve the pressure and saturation. Finite volume formation is applied to solve the problem. In this work, we apply GEOSX (http://www.geosx.org/) to perform forward simulation. We provide the geological parameters (permeability, porosity, and etc.) and other modeling parameters (well controls). GEOSX is applied to generate pressure and CO_2 saturation predictions.

3D Recurrent R-U-Net. In this work, we apply a 3D recurrent R-U-Net developed by Tang et al. [2021c] as surrogate model. 3D recurrent R-U-Net is composed of residual U-Net (R-U-Net) and convolutional long-short term memory network (3D convLSTM) as shown in Figure 1. The R-U-Net is applied to capture the spatial correlation between simulation input $\mathbf{m}$ (permeability, porosity, etc.) and simulation output $\mathbf{y}$ (pressure, saturation, etc.). 3D convLSTM is applied to capture the temporal correlation between simulation output in different time steps. The latent features from R-U-Net are the input for the convLSTM. The feature maps after convLSTM will feed into the decoding net to generate temporal predictions of reservoir response. We use $\mathbf{f}$ to denote the surrogate model, the prediction can be generated through $\hat{\mathbf{y}} = \mathbf{f}(\mathbf{m})$.

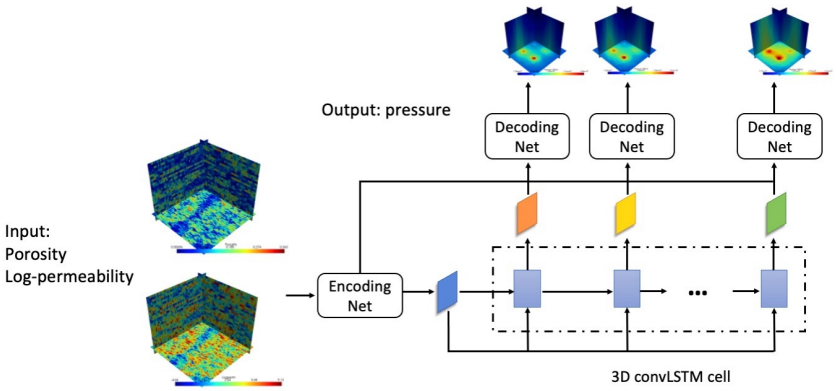


Figure 1: Architecture of Recurrent R-U-Net

Workflow for Training with Transfer Learning. The 3D recurrent R-U-Net surrogate model requires a large number (O(2000)) of high-fidelity simulations (HFS). We apply the idea of transfer learning to train the network based on the multi-scale simulations. We generate N_r^{LFS} (O(2000)) low-fidelity simulations (LFS) and N_r^{HFS} (O(100)) HFS. The input for the surrogate model is fine-scale geological properties $\mathbf{m} \in \mathbb{R}^{N_m \times 1}$, where N_m denotes the number of model parameters.

The process of surrogate model training includes three steps. First, a 3D recurrent R-U-Net is constructed as base model $\mathbf{f}_{\text{base}}$ with weights $\mathbf{w}_{\text{base}}$. The input is high-fidelity geological parameters $\mathbf{m} \in \mathbb{R}^{N_m \times 1}$, and output is $\mathbf{y}_{\text{base}}$. An additional output layer is added after the decoding net of the base model $\mathbf{f}_{\text{base}}$ to map the base model output $\mathbf{y}_{\text{base}}$ to LFS data $\mathbf{y}_{\text{LFS}} \in \mathbb{R}^{N_{\text{output}}^{\text{LFS}} \times N_t}$. The weight for the additonal layer is $\mathbf{w}_{\text{output}}^{\text{LFS}}$. A set of N_r^{LFS} LFS samples are applied to train the network $\mathbf{f}_{\text{LFS}}$ with weights $\{\mathbf{w}_{\text{base}}, \mathbf{w}_{\text{output}}^{\text{LFS}}\}$. The second step is replacing the output layer (weights $\mathbf{w}_{\text{output}}^{\text{LFS}}$) with a new output layer (weights $\mathbf{w}_{\text{output}}^{\text{HFS}}$). The new output layer maps the base model output $\mathbf{y}_{\text{base}}$ to HFS data $\mathbf{y}_{\text{HFS}} \in \mathbb{R}^{N_{\text{output}}^{\text{HFS}} \times N_t}$. The trained base model weights $\mathbf{w}_{\text{base}}$ from step 1 are applied but fixed in this

step. N_r^{HFS} HFS samples are used in this step to train the output layer (weights $\mathbf{w}_{\text{output}}^{\text{HFS}}$) of the new 3D recurrent R-U-Net $\mathbf{f}_{\text{HFS}}$. The last step is fine-tuning all the weights $\{\mathbf{w}_{\text{base}}, \mathbf{w}_{\text{output}}^{\text{HFS}}\}$ of the surrogate model $\mathbf{f}_{\text{HFS}}$. We will apply the training process to generate 3D surrogate model for CO_2 storage case.

3 Surrogate Results

Problem Setup. In this work, we consider a model includes two rock types, sand and shale. The fine-scale (high-fidelity) model is defined on a $64 \times 64 \times 28$ grid. The 3D facies model is generated with sequential indicator simulation conditioned to the hard data at nine well locations in Fig. 2. 3D Gaussian realizations are generated with sequential Gaussian simulation. The porosity and log-permeability are assigned to each block according to the facies and Gaussian realizations using cookie-cutter method. Figure 2 presents one realization of geomodel, with facies shown in left, porosity in middle, and log-permeability in right. The low-fidelity model is defined on $32 \times 32 \times 28$ grid. The porosity and permeability are upscaled with the distance-weighted mean value. There are $N_{\text{inj}} = 4$ injectors in the field. CO_2 is injected through these four injectors with a constant total injection rate (2 million metric tons per year) for the whole field. The simulation period is 10 years. We simulate $N_r^{\text{LFS}} = 1800$ LFS runs and $N_r^{\text{HFS}} = 100$ HFS runs as the training data. In this work, we train the surrogate model for pressure predictions.

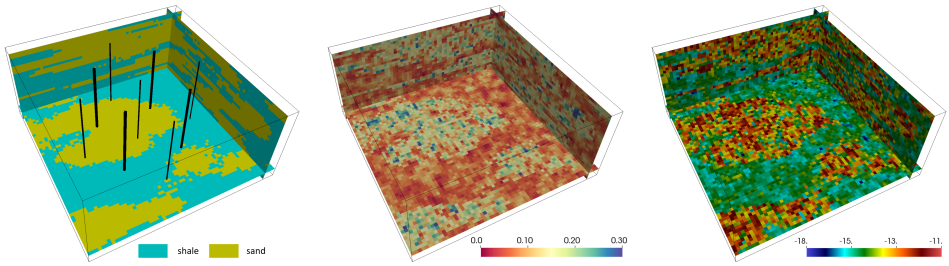


Figure 2: A representative geological realization with facies (left), porosity (middle), log-permeability (right). The nine wells used to constrain the model are shown in the facies plot.

Training Process. The input for recurrent R-U-Net is composed of porosity and log-permeability field (two input channel). The dimension of input is $N_m = N_x \times N_y \times N_z \times 2 = 68 \times 68 \times 28 \times 2 = 258,944$. We normalize the porosity and log-permeability by the maximum value. The output include pressure for 10 time steps (simulation data is sampled per year). We apply the min-max normalization to transfer data to 0-1 range. The output dimension for LFS surrogate model is $N_{\text{output}}^{\text{LFS}} \times N_t = 32 \times 32 \times 28 \times 10 = 286,720$. The output dimension for HFS model is $N_{\text{output}}^{\text{HFS}} \times N_t = 64 \times 64 \times 28 \times 10 = 1,146,880$, where $N_{\text{output}}^{\text{HFS}}$ is the number of grid block of HFS.

In the training process, we train the low-fidelity surrogate model $\mathbf{f}_{\text{LFS}}$ with N_r^{LFS} LFS samples, and fine tune the high-fidelity model $\mathbf{f}_{\text{HFS}}$ with N_r^{LFS} LFS samples. The optimization is

$$[\mathbf{w}_{\text{base}}^*, (\mathbf{w}_{\text{output}}^k)^*] = \underset{\mathbf{w}_{\text{base}}, \mathbf{w}_{\text{output}}^k}{\text{argmin}} \frac{1}{N_r^k} \frac{1}{N_t} \sum_{i=1}^{N_r^k} \sum_{t=1}^{N_t} (||\hat{\mathbf{y}}_i^{k,t} - \mathbf{y}_i^{k,t}||_2^2 + \lambda \frac{1}{N_{\text{inj}}} \sum_{j=1}^{N_{\text{inj}}} ||\hat{\mathbf{y}}_{i,j}^{k,t} - \mathbf{y}_{i,j}^{k,t}||_2^2), \quad (1)$$

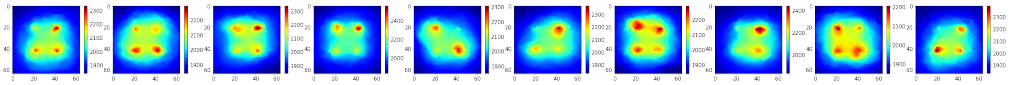
for $k = \text{LFS}, \text{HFS}$, where λ denotes the weights for loss in the well locations. The training loss includes the reconstruction loss and the hard data loss for well locations. The mismatch is calculated with L^2 norm. The network is trained with adaptive moment estimation (ADAM) optimizer [Kingma and Ba, 2014]. The batch size is set be to 8, and 200 epochs are used to the LFS model. 100 epochs are used to train the output layer in step 2 and fine tune the model in step 3.

Pressure Fields. We generate a set of 100 new realizations at high fidelity to evaluate the pressure prediction from surrogate model. We compute the relative error for pressure prediction of each realization i, written as

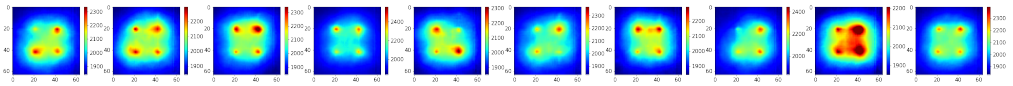
$$e_i = \frac{1}{N_{\text{output}}^{\text{HFS}}} \frac{1}{N_t} \sum_{j=1}^{N_{\text{output}}^{\text{HFS}}} \sum_{t=1}^{N_t} \frac{|\hat{\mathbf{y}}_{i,j}^t - \mathbf{y}_{i,j}^t|}{(\mathbf{y}_{i,j}^t)_{\max} - (\mathbf{y}_{i,j}^t)_{\min}}, \quad (2)$$

for $i = 1, 2, \ldots, 100$. The relative error is normalized by the minimum and maximum value of pressure at each grid block j, time step t for test realization i.

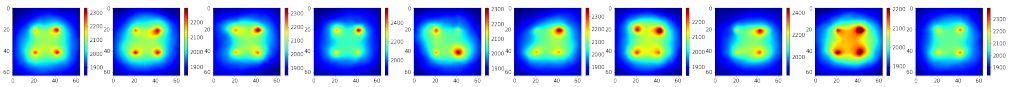
We first train the network with 1800 HFS samples and use the HFS surrogate model $\mathbf{f}_{\text{HFS}}^{\text{ref}}$ as reference. Figure 3 presents the pressure fields of 10 realizations for layer 28 at $t = 10$ year. The upper row shows the HFS simulation results, the middle show presents the surrogate results from multi-fidelity training. The bottom row shows the surrogate results from high-fidelity training. We rank these 100 realizations from low relative error to high relative error, and sample 10 realizations in sequence. For the surrogate results with multi-fidelity training, the relative error for the first realization is 1.62% (the lowest value). The largest relative error is 3.62%. For the reference surrogate results with high-fidelity training, the errors are 1.50% and 4.05%. Both results show close agreement with HFS simulations. These 10 realizations are representative and present a large variation. The surrogate predictions with multi-fidelity training presents reasonable accuracy.



(a) Simulation pressure predictions (psi)



(b) Surrogate pressure predictions with multi-fidelity training (psi)



(c) Surrogate pressure predictions with high-fidelity training (psi)

Figure 3: Pressure maps for HFS (upper row) surrogate results with multi-fidelity training (middle row), surrogate results with high-fidelity training (bottom row) for 10 test cases at layer 28 at 10 years

Computational Cost. The computational time of building surrogate model includes two parts, simulation time and network training time. Table 1 summarizes the computational cost of generating reference surrogate model $\mathbf{f}_{\text{HFS}}^{\text{ref}}$ and generating surrogate model $\mathbf{f}_{\text{HFS}}$ with transfer learning. For the reference surrogate model, 1800 high-fidelity simulations are used to train the model. The total simulation time for high-fidelity training is 3600 CPU hours. While for multi-fidelity training, only 100 HFS samples and 1800 low-fidelity simulations are required. The total simulation time for multi-fidelity training is 740 CPU hours, which is around 20% of the simulation time of high-fidelity ensemble. The multi-fidelity framework significantly improve the computational efficiency of the training dataset.

Table 1: Computational cost for high-fidelity training and multi-fidelity training

	High-fidelity training	Multi-fidelity training
GEOSX simulation (64×64×28)	2 core hours×1800 = 3600 core hours	2 core hours×100 = 200 core hours
GEOSX simulation (32×32×28)	N/A	0.3 core hours×1800 = 540 core hours
GPU training time (Nvidia Tesla V100)	11.5 hours	9.3 hours (step 1) + 0.1 hours (step 2)+ 0.3 hours (step 3) = 9.7 hours

4 Conclusions and Discussion

In this study, we used a transfer learning approach to train a 3D recurrent R-U-Net surrogate model for predicting pressure responses of a geologic carbon storage (GCS) reservoir subjected to CO_2 injection. The full model has an original resolution of $64 \times 64 \times 28$. We trained the neural network with 1,800 low-fidelity simulations embodying the same physics but with low resolution. The

model is then fine-tuned with 100 full-resolution simulations. The resultant surrogate model exhibits satisfactory accuracy while reduces the training cost by 80%. This suggests that the multi-fidelity training framework can be applied to surrogate modeling of field-scale GCS projects.

Acknowledgments and Disclosure of Funding

This manuscript has been authored by Lawrence Livermore National Security, LLC under Contract No. DE-AC52-07NA2 7344 with the US. Department of Energy (DOE). The work was completed as part of the Science-informed Machine learning to Accelerate Real Time decision making for Carbon Storage (SMART- CS) Initiative.

References

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Shaoxing Mo, Yinhao Zhu, Nicholas Zabaras, Xiaoqing Shi, and Jichun Wu. Deep convolutional encoder-decoder networks for uncertainty quantification of dynamic multiphase flow in heterogeneous media. *Water Resources Research*, 55(1):703–728, 2019.

Dong H Song and Daniel M Tartakovsky. Transfer learning on multi-fidelity data. *arXiv preprint arXiv:2105.00856*, 2021.

Hewei Tang, Pengcheng Fu, Christopher S Sherman, Jize Zhang, Xin Ju, François Hamon, Nicholas A Azzolina, Matthew Burton-Kelly, and Joseph P Morris. A deep learning-accelerated data assimilation and forecasting workflow for commercial-scale geologic carbon storage. *arXiv preprint arXiv:2105.09468*, 2021a.

Meng Tang, Xin Ju, and Louis J Durlofsky. Deep-learning-based coupled flow-geomechanics surrogate model for CO_2 sequestration. *arXiv preprint arXiv:2105.01334*, 2021b.

Meng Tang, Yimin Liu, and Louis J Durlofsky. Deep-learning-based surrogate flow modeling and geological parameterization for data assimilation in 3D subsurface flow. *Computer Methods in Applied Mechanics and Engineering*, 376:113636, 2021c.

Gege Wen, Meng Tang, and Sally M Benson. Towards a predictor for CO_2 plume migration using deep neural networks. *International Journal of Greenhouse Gas Control*, 105:103223, 2021.