Estimation of Physical Coefficients for CO_2 Sequestration using Deep Generative Priors based Inverse Modeling Framework

Jiawei Shen [1] **Jonghyun Lee** [1] **Hongkyu Yoon** [2]

Abstract

Estimation of permeability fields in the subsurface plays a crucial role in forecasting and risk evaluation of geologic carbon storage operations. In real-world scenarios, direct measurements of permeability and CO_2 plume extent are typically sparse due to the high cost and limited direct measurement methods. Although inverse modeling approaches allow us to estimate the subsurface properties including permeability using observations of other indirect data such as pressure, saturation, and measurements from geophysics, it suffers from expensive computation for large-scale problems with relatively high uncertainty. In this work, we test a deep generative prior to sample 3D permeability realizations from a low-dimensional latent space. Then we incorporate the constructed deep generative model to the inverse modeling framework and use observations of CO_2 saturation to reconstruct the permeability field.

1. Introduction

The increase of greenhouse gas carbon dioxide (CO_2) is recognized as a crucial contributor to climate warming. Carbon dioxide capture and geological storage (CCS) has emerged as a significant strategy to reduce the concentration of carbon dioxide in the atmosphere. However, large-scale injection of CO_2 into the subsurface system may cause leakage from CCS, which then potentially degrades the role of the subsurface as a carbon mitigation option. To reduce leakage risk and make the operation more feasible, a comprehensive analysis of the subsurface system and modeling of CO_2 storage is necessary.

The analysis of the subsurface system involves certain partial differential equations (PDEs) that describe multi-phase fluid flow as well as geomechanical deformation and geochemical reactions (Nordbotten & Celia, 2011). Analytical and numerical approaches for solving these equations usually rely on being aware of the subsurface properties, e.g., spatial permeability field that represent complex subsurface structures, appeared as the PDE coefficients. Therefore, characterization of subsurface properties including permeability and porosity plays a crucial role in modeling of CO_2 plume migration and potential leakage.

In practice, however, there is often sparse direct measurements of such properties due to the expensive cost of drilling in the deep subsurface. Inverse modeling/data assimilation approaches offer an option to inversely estimate the unknown subsurface properties from the observation data of primary variables through the governing equations (Forghani et al., 2022). But when it comes to large-scale fields, the computation of inverse modeling is expensive due to the computation of Jacobian and large matrix inversion.

Our focus in this work is to predict the large-scale permeability field using sparse saturation observation data with a reduced-order model (ROM). Specifically, we use a Bayesian method to update the probability density function (pdf) of unknown properties from the prior pdf. For the prior modeling, instead of widely used Gaussian models, deep generative models are used to better account for the subsurface structures with potential faults and fractures; for our purpose, a Variational Autoencoder (VAE) with 3D Convolutional Neural Network architecture is used to explain potential subsurface structures with a low-dimensional latent representation of the permeability field. Then multiphase flow simulation for pressure and CO_2 saturation is also approximated with deep-learning based ROM (Yoon et al., 2022) for the faster equation of the aquifer states. Bayesian data assimilation approaches are then performed to estimate the latent representation of the permeability field using saturation observation data. We apply the proposed method to characterize the 3D permeability field in the Illinois Basin Decatur Project (IBDP), the first CCS site in the United States that injected commercial volumes of CO_2 captured from a biofuel plant (Finley, 2014).

[1]University of Hawaii at Manoa, Honolulu, HI, USA [2]Sandia National Laboratories, Albuquerque, NM, USA. Correspondence to: Jonghyun Lee <jonghyun.harry.lee@hawaii.edu>.

Accepted after peer-review at the 1st workshop on Synergy of Scientific and Machine Learning Modeling, SynS & ML ICML, Honolulu, Hawaii, USA. July, 2023.

2. Methods

2.1. Forward Simulation

In this study, we focus on the modeling of CO_2 storage with highly heterogeneous permeability field. The CO_2-brine flow system involves multiple fluid phases that follow the mass conservation principle. The governing equation for immiscible flow of multiple fluids in porous media is

$$\frac{\partial(\phi \rho_\alpha S_\alpha)}{\partial t} = -\nabla \cdot (\rho_\alpha \boldsymbol{u}_\alpha) + q_\alpha \tag{1}$$

where α denotes phase, ϕ is the porosity, S_α is saturation, $\boldsymbol{u}_\alpha$ is the phase velocity, ρ_α is the density, q_α denotes the source/sink terms in each phase. The relationship between $\boldsymbol{u}_\alpha$ and primary variables can be described by multi-phase version of Darcy' law.

$$\boldsymbol{u}_\alpha = -\frac{k_{r,\alpha}\boldsymbol{k}}{\mu_\alpha}(\nabla p_\alpha - \rho_\alpha \boldsymbol{g}) \tag{2}$$

where $\boldsymbol{k}$ is absolute permeability, $k_{r,\alpha}$ is relative permeability, μ_α is the fluid viscosity, p_α is the pressure, and $\boldsymbol{g}$ is the gravity. Note that the pressure and saturation are the primary variables in the governing equation above.

The governing equation can be solved using numerical methods, e.g., finite difference (FD) and finite element method (FEM) (Binning & Celia, 1999; Kukreti & Rajapaksa, 1989). In this work, we use a deep learning-based surrogate model (Yoon et al., 2022), which is a convolutional neural network-long short-term memory (CNN-LSTM) model, to perform the forward simulation for its fast computation. The deep learning-based surrogate model uses permeability, porosity, and injection rates as inputs to forecast CO_2 saturation and pressure. Hence, it can be considered as a ML-based surrogate model of the governing equation.

2.2. Deep Generative Model

Solving the governing equation involves knowing the PDE coefficients (i.e., permeability and porosity). Given the physical coefficients, we can obtain primary variables (i.e., saturation and pressure) by solving the mass conservation equations of the CO_2-brine flow system. In this work, we trained a deep generative model that can generate samples from a low dimensional latent space so that we can accelerate the inversion. Our application is the Illinois Basin Decatur Project (IBDP) site (Finley, 2014). In this preliminary work, among many deep generative models such as generative adversarial networks, normalizing flow, or score-based models, we use Variational Autoencoder (VAE) with 3D CNN architecture to generate 3D permeability field realizations as shown in Figure 1. The training data were generated from different Gaussian models with fault inclusion within the geological formations where the entire 3D field has 11 geological formation and our preliminary choice of VAE is mainly due to its simplicity and faster training.

Autoencoder is a neural network designed to reconstruct high-dimensional variables from low-dimensional latent space. It consists of encoder and decoder. The encoder network maps the high-dimensional variables into low-dimensional latent space in order to achieve dimensionality reduction, while the decoder network offers the capability to transform the latent representation back to the original high-dimensional variables. VAE imposes a prior distribution of the latent representation by introducing a regularization term in the loss function (Kingma & Welling, 2013). The objective of VAE is to minimize the reconstruction error as well as the prior regularization term, which is the Kullback-Leibler (KL) divergence between the prior normal distribution $p(z)$ and the conditional distribution $q(z|x)$:

$$\mathcal{L}_{\text{VAE}} = -\mathbf{E}_{q(z|x)}[\log p(x|z)] + \mathbf{D}_{\text{KL}}(q(z|x)||p(z)) \tag{3}$$

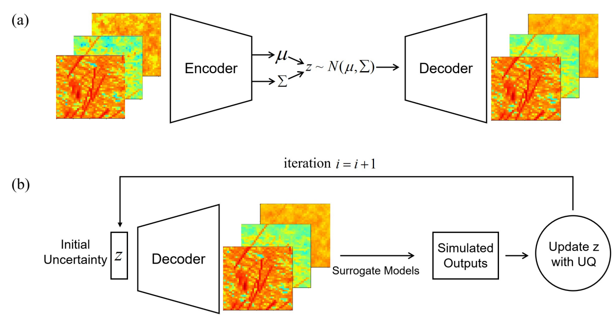

Figure 1. (a) Permeability generative model. (b) Flow diagram of latent space variational data assimilation

2.3. Inverse Modeling

After the VAE model is trained, we employed a Bayesian approach as the inverse problem solver to estimate the latent representations of permeability coefficients from the saturation observations. The forward problem can be defined as

$$y = h(G(z)) + \epsilon \tag{4}$$

where z is the latent representations of the unknown variable (e.g., permeability), G is the generative model (e.g., the decoder of VAE model), h is the forward map, and ϵ is the observation and model uncertainty noise, $\epsilon \sim \mathcal{N}(0, R)$, where R is the observation error matrix.Since we impose a Gaussian prior regularization in VAE, the prior distribution of z is assumed to follow Gaussian distribution, i.e., $z \sim \mathcal{N}(\mu, \Sigma)$. The Bayes' rule allows us to a posterior distribution of z via

$$
\begin{aligned}
p(z|y) &\propto p(y|z)p'(z) \\
&\propto \exp(-(y-h(G(z))))^T R^{-1}(y-h(G(z))))\exp(-z^T\Sigma^{-1}z)
\end{aligned}
\tag{5}
$$

Then, the maximum a posterior (MAP) estimate is

$$
z_{\text{map}} = \arg\max(-(y-h(G(z))))^T R^{-1}(y-h(G(z))) + -z^T\Sigma^{-1}z)
\tag{6}
$$

The MAP estimate can be approximated by the Gauss-Newton approach with iterative linearizations (Forghani et al., 2022; Lee & Kitanidis, 2014). For iteration count l and step size α, we have

$$
\begin{aligned}
\mathbf{z}^{l+1} = &\quad \mathbf{z}^l + \alpha\left(\boldsymbol{\Sigma}^{-1} + \mathbf{J}_l^T\mathbf{R}^{-1}\mathbf{J}_l\right)^{-1}\left(\boldsymbol{\Sigma}^{-1}z^l + \mathbf{J}_l^T\mathbf{R}^{-1}\left(\mathbf{y} - h(G(\mathbf{z}^l))\right)\right) \\
= &\quad (1-\alpha)\mathbf{z}^l + \alpha\boldsymbol{\Sigma}\mathbf{J}_l\left(\mathbf{J}_l\boldsymbol{\Sigma}\mathbf{J}_l^T + \mathbf{R}\right)^{-1}\left(\mathbf{y} - h(G(\mathbf{z}^l)) + \mathbf{J}_l\mathbf{z}^l\right)
\end{aligned}
\tag{7}
$$

where J_l is the Jacobian of the forward map from the latent space to permeability at the l-th iteration. J_l can be evaluated using the following formula:

$$
\mathbf{J}_l = \left.\frac{\partial h(G(\mathbf{z}))}{\partial \mathbf{z}}\right|_{\mathbf{z}=\mathbf{z}_l} = \left.\frac{\partial h}{\partial \mathbf{s}}\right|_{\mathbf{s}=G(\mathbf{z}_l)} \left.\frac{\partial \mathbf{s}}{\partial \mathbf{z}}\right|_{\mathbf{z}=\mathbf{z}_l} = \left.\mathbf{J}_h\right|_{\mathbf{s}=G(\mathbf{z}_l)} \left.\mathbf{J}_G\right|_{\mathbf{z}=\mathbf{z}_l}
\tag{8}
$$

3. Preliminary Results

In this section, we provide the results of our deep generative prior based inversion framework to estimate the highly heterogeneous permeability field. In section 3.1, we provide the results of applying the VAE model to reconstruct the 3D permeability data. In section 3.2, we present the inversion results of using sparse saturation measurement.

3.1. Performance of VAE for Permeability Reconstruction

In this section, we present the result of applying VAE model to the 3D domain. The domain dimension is $nx = 38, ny = 42, nz = 90$, which is the central region of the entire domain ($nx = 126, ny = 125, nz = 110$). The permeability in this case is anisotropic, i.e., different spatial fields k_{xx}, k_{yy}, and k_{zz}. We distinguish the horizontal and vertical permeability, as indicated by x and z, respectively. We take the logarithm of permeability and then normalized it to [0, 1]. We have currently available 100 permeability fields and simulated saturation fields every month for 3 yrs injection and 1 yr post-injection period in total offered from the IBDP site. 90 cases of the entire 100 cases are taken as training set and the remaining 10 is for validation. The major hyper-parameters of the model are present in table 1.

Figure 2 shows several layers of reconstructed permeability realizations on the test set and the fitting plot of permeability data with the root mean square error (RMSE) calculated on

Table 1. Major hyper-parameters

HYPER-PARAMETER	VALUE
LATENT DIMENSION	100
WEIGHT OF KL DIVERGENCE	10^{-6}
OPTIMIZER	ADAM
LEARNING RATE	10^{-5}
BATCHSIZE	16

the single layer. The realization for visualization is chosen randomly from the test set and its 74-th, 79-th, 84-th, and 89-th model layers are selected for visualization, which are representative of the subsurface structure. The RMSE values of permeability on the training set and test set are 0.0464 and 0.0491, respectively, indicating a reasonable construction of the deep prior model for the IBDP model.

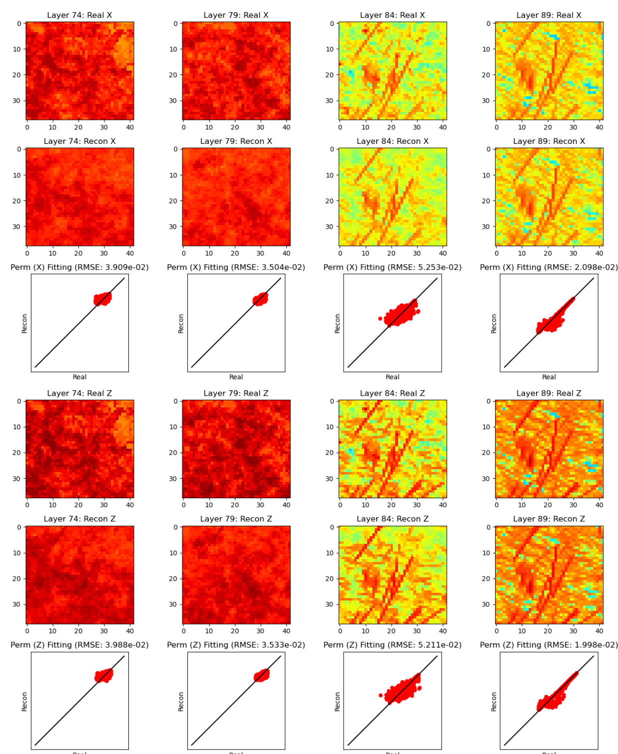

Figure 2. Reconstruction of permeability on the test set. Rows from top to bottom are real, reconstructed permeability and the data fitting plot in the horizontal and vertical directions, respectively. Columns from left to right are layers 74, 79, 84, and 89.

3.2. Performance of Inverse Modeling

In this section, we show the results of our deep generative prior-based inversion approach using sparse saturation measurement. We use the decoder of VAE as the generative model to produce 3D permeability field. The dimension of latent representation z is set to 100.

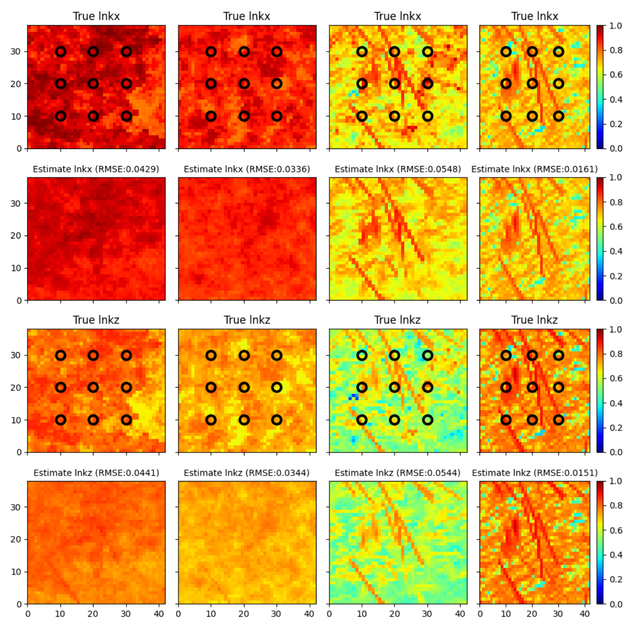

Figure 3. The generated permeability using the updated latent representation z. The first two rows are real and reconstructed permeability in x coordinate respectively, while the last tow rows are real and reconstructed permeability in z coordinate respectively. Columns are layers 74, 79, 84, 89 from left to right.

Once the permeability realizations are generated, we perform forward simulation using the pre-trained deep learning surrogate CNN-LSTM model to obtain saturation values. We chose 9 observation wells whose location is shown in Figure 3. Gauss-Newton approach is performed to update the latent representation z. Figure 3 shows generated permeability by the decoder of VAE with the estimated z. In Figure 4, we plot the observed saturation against simulated saturation, which shows a reasonable fitting of the CO_2 saturation.

4. Concluding Remarks

We implemented a deep generative model-based inversion approach to perform inversion for the IBDP CCS site and presented reasonable inversion results. The proposed method uses deep generative prior and reduced-order saturation prediction model so that one expect a great computational gain in the CO_2 injection site characterization. In the preliminary application, the estimated permeability fields captured important subsurface structures, i.e., faults and host rock permeability distribution, due to the informative prior used in the training.

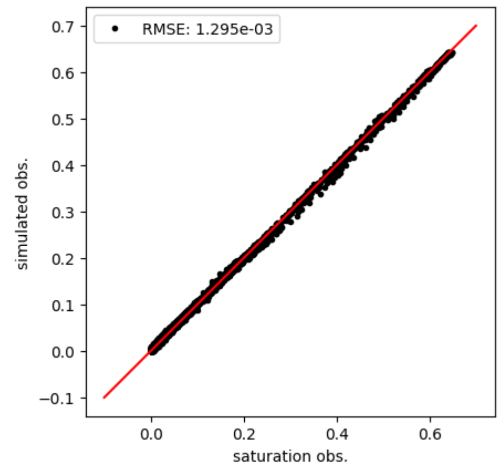

Figure 4. Saturation data fitting.

5. Broader Impact

Understanding the distribution of subsurface properties such as permeability is crucial for reliable carbon storage management. Our work aims to develop an efficient and fast inverse modeling framework to estimate key parameters of subsurface carbon storage sites in order to better forecast the migration and leakage of the carbon.

Acknowledgments

This work was supported by US Department of Energy Office of Fossil Energy and Carbon Management,the Science-informed Machine Learning for Accelerating Real-Time Decisions in Subsurface Applications (SMART) Initiative. Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC (NTESS), a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration (DOE/NNSA) under contract DE-NA0003525. This written work is authored by an employee of NTESS. The employee, not NTESS, owns the right, title and interest in and to the written work and is responsible for its contents. Any subjective views or opinions that might be expressed in the written work do not necessarily represent the views of the U.S. Government. The publisher acknowledges that the U.S. Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this written work or allow others to do so, for U.S. Government purposes. The DOE will provide public access to results of federally sponsored research in accordance with the DOE Public Access Plan.

References

Binning, P. and Celia, M. A. Practical implementation of the fractional flow approach to multi-phase flow simulation. *Advances in water resources*, 22(5):461–478, 1999.

Finley, R. J. An overview of the illinois basin–decatur project. *Greenhouse Gases: Science and Technology*, 4 (5):571–579, 2014.

Forghani, M., Qian, Y., Lee, J., Farthing, M., Hesser, T., Kitanidis, P. K., and Darve, E. F. Variational encoder geostatistical analysis (vegas) with an application to large scale riverine bathymetry. *Advances in Water Resources*, 170:104323, 2022.

Kingma, D. P. and Welling, M. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

Kukreti, A. R. and Rajapaksa, Y. A numerical model for simulating two-phase flow through porous media. *Applied Mathematical Modelling*, 13(5):268–281, 1989.

Lee, J. and Kitanidis, P. K. Large-scale hydraulic tomography and joint inversion of head and tracer data using the principal component geostatistical approach (pcga). *Water Resources Research*, 50(7):5410–5427, 2014.

Nordbotten, J. M. and Celia, M. A. Geological storage of co2: modeling approaches for large-scale simulation. In *Geological Storage of CO 2: Modeling Approaches for Large-Scale Simulation*. John Wiley and Sons, 2011.

Yoon, H., Lee, J., and Kadeethum, T. Deep learning-based data assimilation in the latent space for real-time forecasting of geologic carbon storage. *Available at SSRN 4294901*, 2022. doi: Doi10.2139/ssrn.4294901.