CO_2 Storage Site Characterization using Deep Generative Models

Jichao Bao[1], Hongkyu Yoon[2], Jonghyun Harry Lee[1]

[1]Civil and Environmental Engineering, University of Hawaii at Manoa, Honolulu, HI, USA; [2]Geomechanics Department, Sandia National Laboratories, Albuquerque, NM, USA

Introduction

In this work, we present a fast Machine Learning (ML)-based inverse modeling for real-time subsurface history matching and forecasting. The work aims to support CO_2 related operations and forecasting CO_2 and pressure plume development for DOE Science-informed Machine Learning to Accelerate Real Time (SMART) initiative. The proposed method utilizes one of the current efficient data assimilation methods, i.e., *Hierarchical Bayesian with Gaussian Prior* (through power-transformation of unknowns). In Bayesian framework, the prior uncertainty of subsurface unknown properties, e.g., permeability field, is typically parameterized as log-normal with mean and covariance. However, there are two major challenges for high-dimensional applications:

- the number of (expensive) forward model simulations
- computational burdens for high-dimensional matrix-matrix computations.

To address these challenges, we use deep-generative model-based data assimilation methods for large-scale carbon storage site characterization and forecast. Variational AutoEncoder (VAE) and novel Wasserstein Generative Adversarial Networks are used to learn the approximate distribution from multipoint geostatistics-derived training images as prior and accelerated data assimilation is performed on the low-dimensional latent space in a Bayesian framework. Numerical examples with synthetic 2D permeability fields with fluvial channels confirm that our proposed method provides promising subsurface site characterization with reliable uncertainty quantification.

Model Reduction

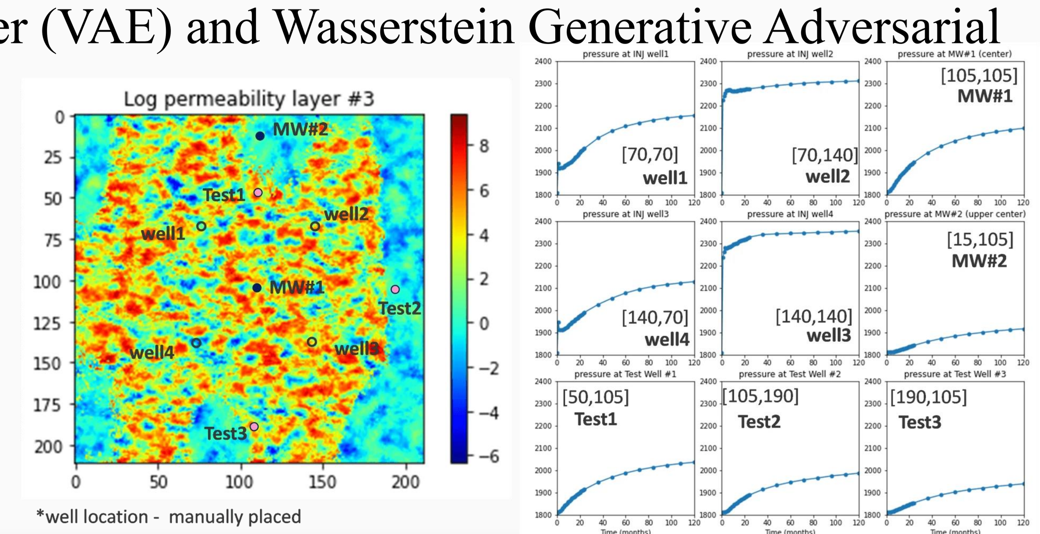
Forward modeling:

$$\mathbf{y} = \mathbf{F}(\mathbf{m}) \approx \mathbf{F}(\mathbf{G}(\mathbf{z}))$$

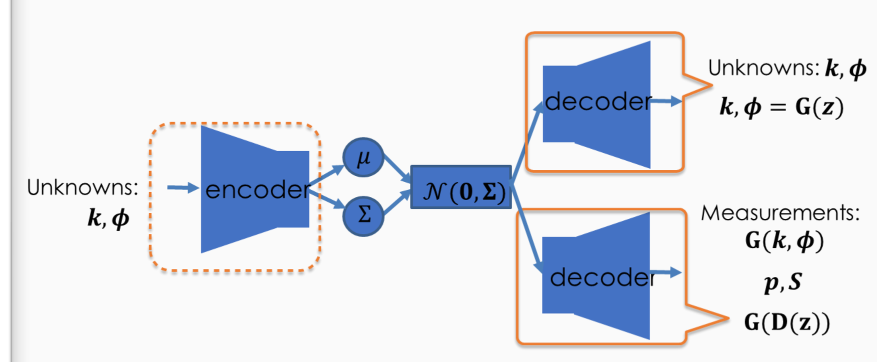
- $\mathbf{y}$ is a (nobs x 1) simulated observation vector,
- $\mathbf{F}$ is a flow model that produces outputs at obs. locations
- $\mathbf{m}$ is a (m x m) permeability matrix,
- $\mathbf{G}$ is a generator or deterministic map from z to s
- $\mathbf{z}$ is a (k x k) latent space matrix.

G(z) can be obtained from any generative models and here we used Variational AutoEncoder (VAE) and Wasserstein Generative Adversarial Networks (WGANs).

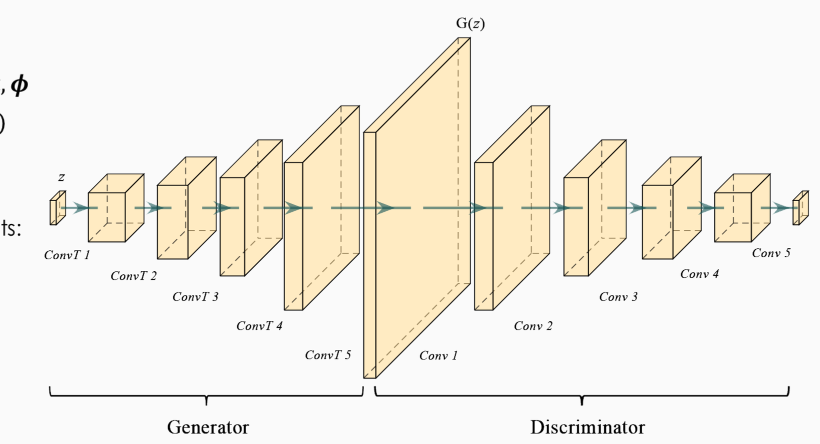
Clastic Shelf MODEL:



Deep Generative Model Architecture

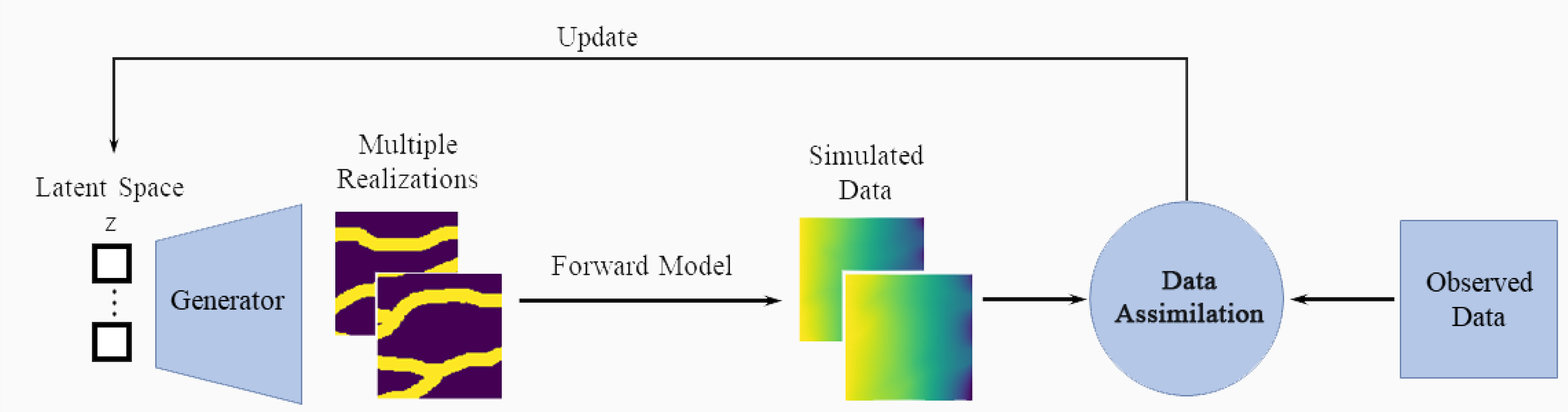


Encoder-Decoder/Generator Architecture



Generator-Discriminator Architecture

Data Assimilation with Deep Generative Models



Main Conceptual Idea

Forward problem: $\mathbf{y} = \mathbf{F}(\mathbf{m})$

with data assimilation framework:

$$\boldsymbol{m}^{i+1} = \boldsymbol{m}^{i} + \boldsymbol{C}_{MD}^{i}\left(\boldsymbol{C}_{DD}^{i} + \alpha_i \boldsymbol{C}_D\right)^{-1}\left(\boldsymbol{d}_{uc}^{i} - \boldsymbol{d}^{i}\right)$$

$$\boldsymbol{d}_{uc}^{i} = \boldsymbol{d}_{obs} + \sqrt{\alpha_i}\boldsymbol{C}_D^{1/2}\boldsymbol{I}_d$$

where $\boldsymbol{C}_{MD}^{i}$ is the cross-covariance between the model parameters and simulated data, $\boldsymbol{C}_{DD}^{i}$ is the auto-covariance of the simulated data, $\boldsymbol{C}_D$ is the observation error covariance, $\boldsymbol{d}^{i}$ is the simulated data, $\boldsymbol{d}_{uc}^{i}$ is the perturbed observation data, and α_i is an inflation factor.

with any (nonlinear) dimension reduction:

$$\mathbf{y} = \mathbf{F}(\mathbf{G}(\mathbf{z})),\ \dim(\mathbf{z}) << \dim(\mathbf{m})$$

$$\boldsymbol{z}^{i+1} = \boldsymbol{z}^{i} + \boldsymbol{C}_{ZD}^{i}\left(\boldsymbol{C}_{DD}^{i} + \alpha_i \boldsymbol{C}_D\right)^{-1}\left(\boldsymbol{d}_{uc}^{i} - \boldsymbol{d}^{i}\right)$$

VAE/WGAN is our choice since it constructs the prior $\boldsymbol{z} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{I})$

- Covariance can be directly computed with Adjoint method and Fast Linear Algebra (i.e., Variational Inference) or can be approximated with Ensemble via Ensemble Smoother.
- Depending on the choice (variational inference or Ensemble approximation), one would need a few hundreds of forward modeling evaluation

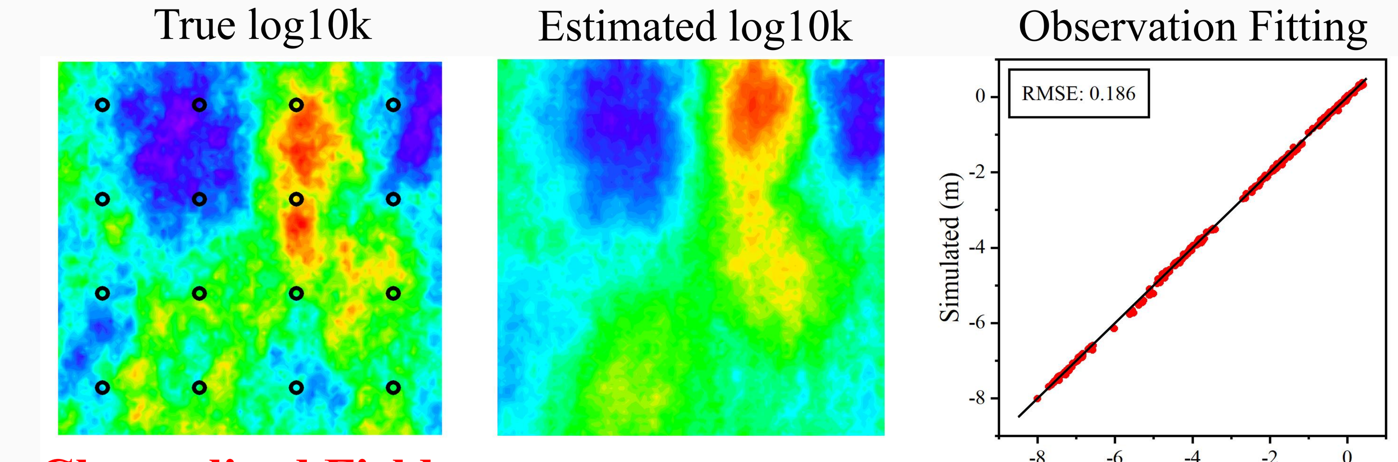
Experimental Results

Validation with Single Phase Flow Application

Here we used USGS MODFLOW as "full" physics single phase flow model. 96 x 96 unknown k field has been compressed to 3x3 latent dimension through WGAN. 16 wells hydraulic tomography was applied to obtain pressure data.

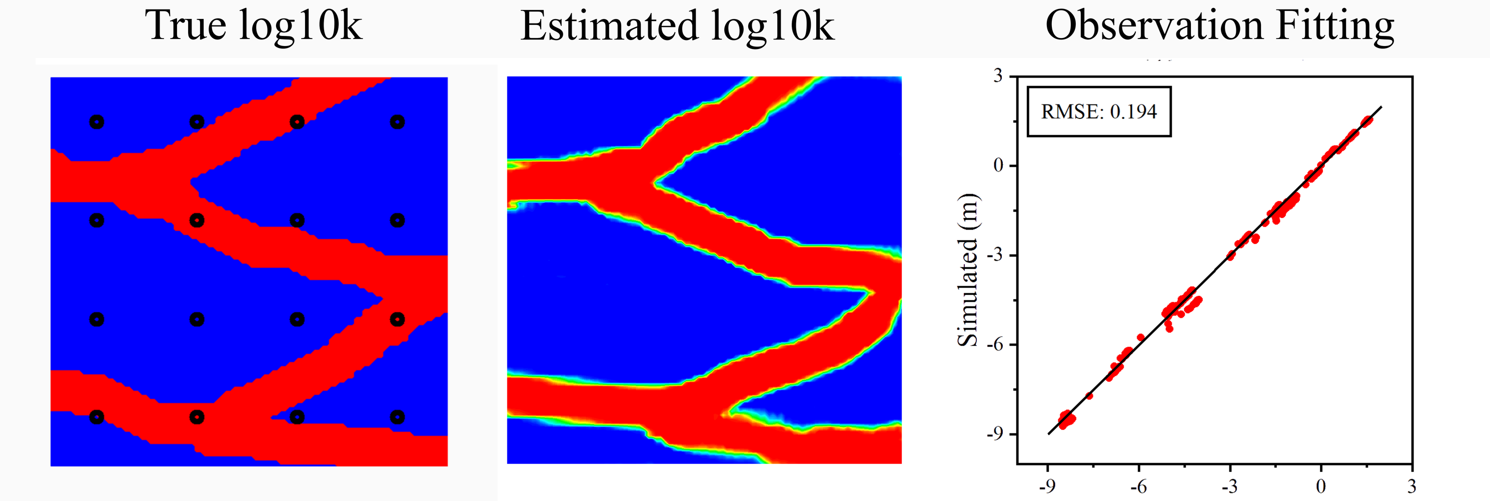
Gaussian Fields

True log10k | Estimated log10k | Observation Fitting



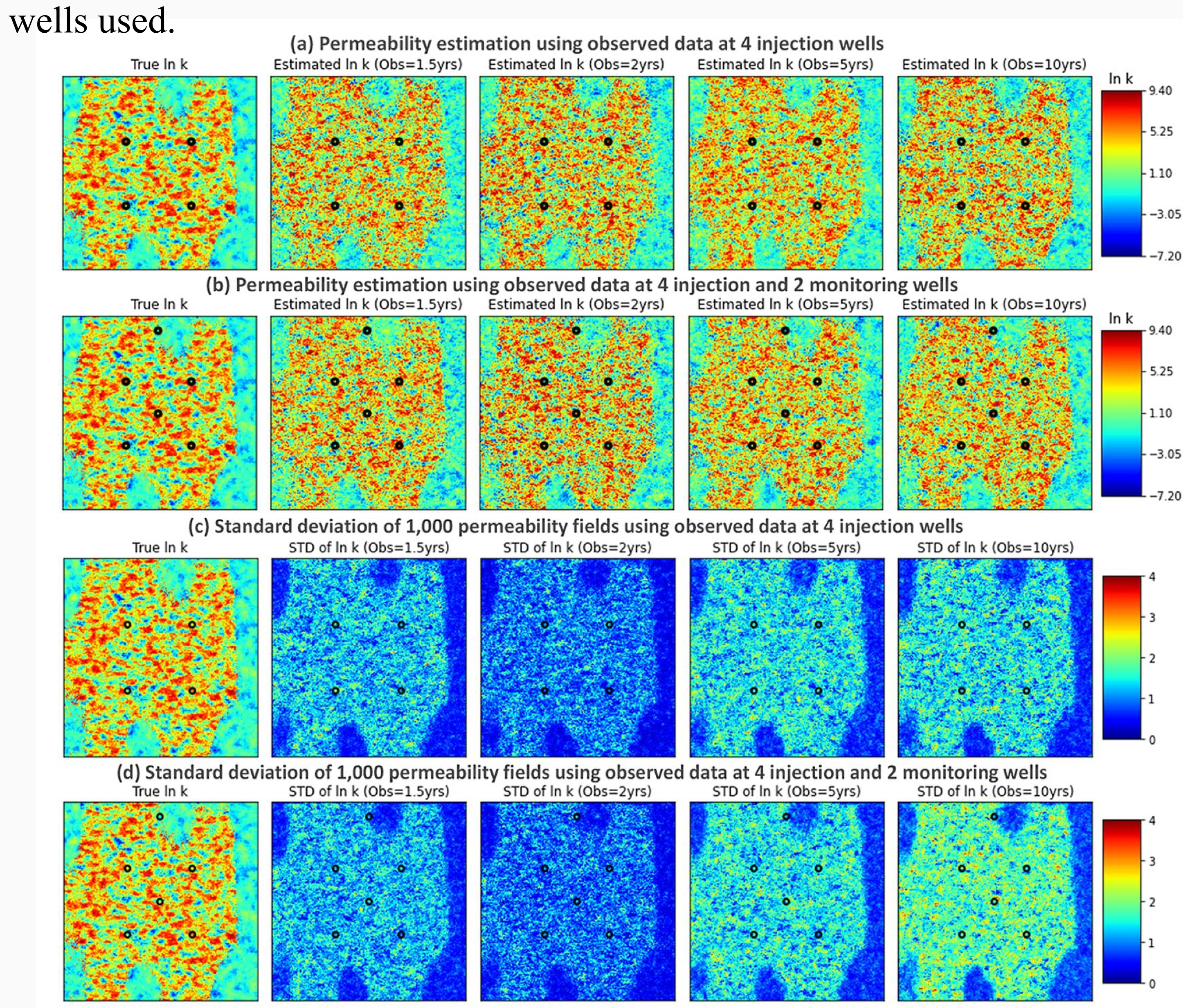
Channelized Fields

True log10k | Estimated log10k | Observation Fitting



Multi-Phase Flow Application using Clastic Shelf Data

Here we used CMG for CO2 injection simulation. 211 x 221 unknown k field has been compressed to 32 latent dimension through VAE. 4 injection/2 monitoring wells used.



Acknowledgement

This work is supported by the US Department of Fossil Energy and Carbon Management, Science-informed Machine Learning to Accelerate Real Time Decisions-Carbon Storage (SMART-CS) initiative.