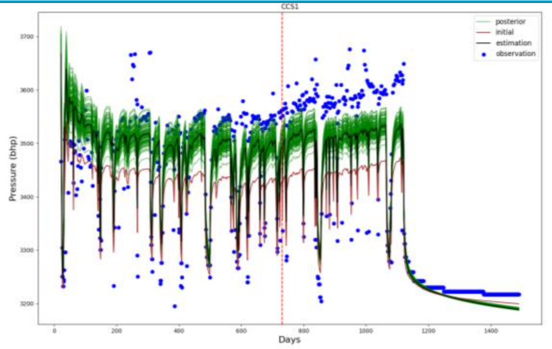
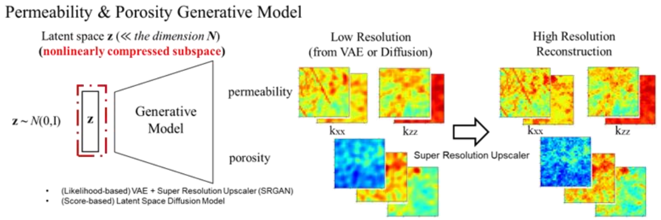
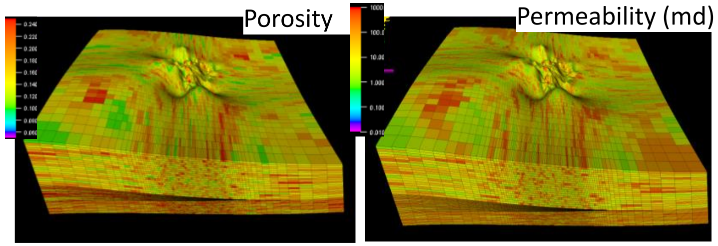
# Rapid model prediction and history matching using scientific machine learning for geologic carbon storage at the Illinois Basin-Decatur Project site




**Hongkyu Yoon**

Climate Change Security Center, Sandia National Laboratories#, NM, USA

Co-authors: Jonghyun Lee*, Jiawei Shen*, Teeratorn Kadeethum#

*University of Hawaii at Manoa

**GHGT 2024**

This work was supported by DOE Office of Fossil Energy and Carbon Management project -Science-informed Machine Learning to Accelerate Real Time (SMART) Decisions in Subsurface Applications.

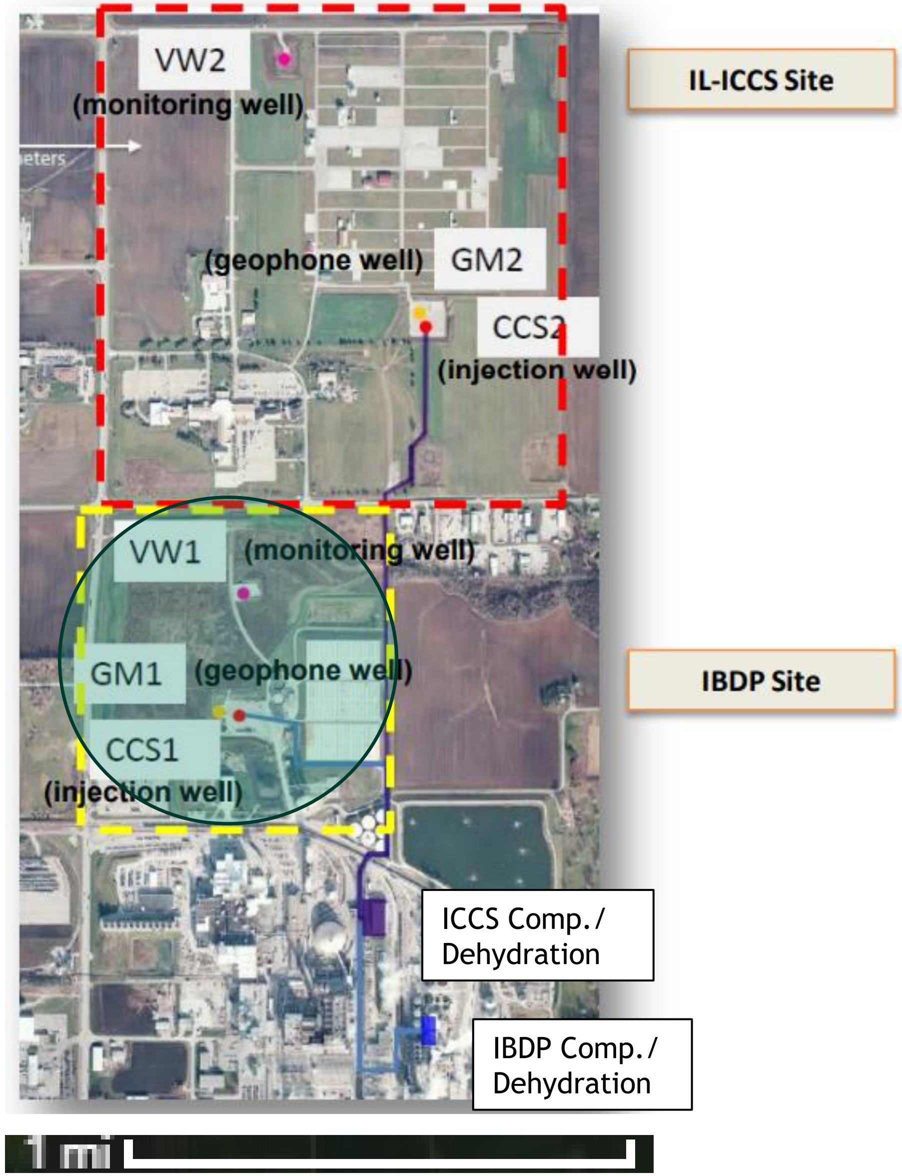
**SAND2024-XXXXXC**

# Illinois Basin Decatur Project Site

- Carbon capture, utilization, and storage (CCUS) can play a role in reducing net carbon emissions

## Illinois Basin Decatur Project (IBDP)

- Inject and store 0.33 million tonnes per year / 1 million total
- First $CO_2$ injection demonstration project from 2011 to 2014
- Novelty: First 1 million tonne biofuel CCS project in U.S. & microseismicity monitoring
- Variety of geologic, operational and monitoring data being collected and interpreted
- [Q] Can (near) real-time feedback be provided for operational control and optimization?
- How can ML-assisted workflows improve understanding of GCS system?

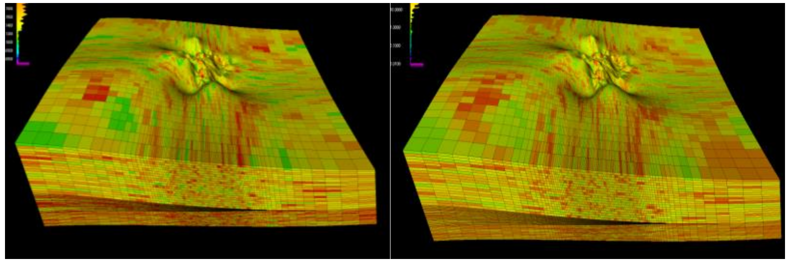


# Objective & ML training data

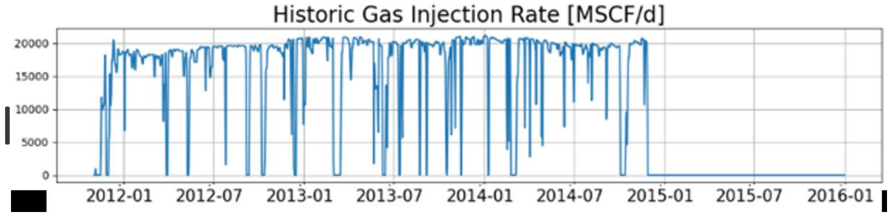
Objective: Machine learning-driven $CO_2$ modeling by combining fast ML-based forward modeling with latent space data assimilation (LSDA), resulting in real-time history matching of $CO_2$ operations and forecasting $CO_2$ and pressure plume development

## ML training data at the IBDP site

- 100 Sets of 3D permeabiltiy (x,y,z), porosity, transmissivity (x,y,z) fields at 126 x 125 x 110 tartan grid (1.73M cells)
- Eclipse simulation was performed to generate training data
- Well pressure daily at injection well and monitoring well (6 depths)
- Pressure and $CO_2$ saturation prediction every 1 month
- $CO_2$ injection for 3 years + 1 yr shut in
- All input data (e.g., injection rates and locations) reflects real historical $CO_2$ injection data

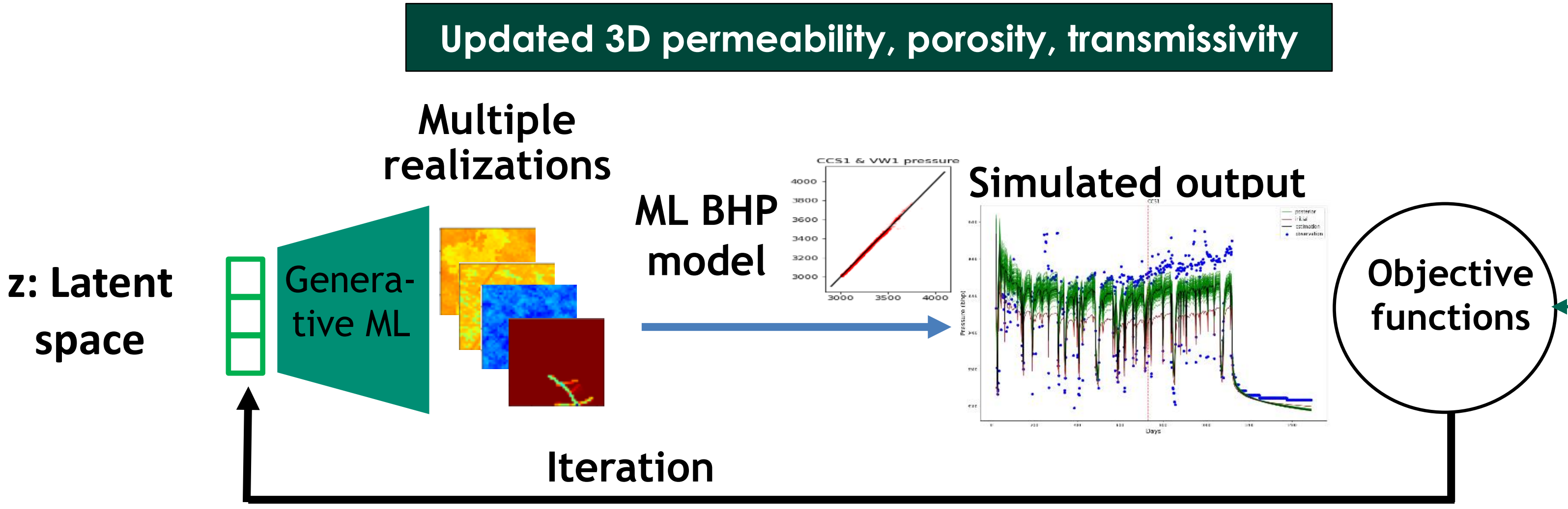


Permeability and porosity field



# Latent Space-based DA (LSDA) with generative priors

- Data assimilation in **low dimensional latent space of unknown parameters with dim(z)**
- Forward model executions can be significantly reduced with ML-based models
- Flexible modular structure and Bayesian inference or ensemble-based methods are available




**Observed data at IBDP**
- **Bottom hole pressures (BHPs) at injection and monitoring wells**
- Saturation & Temperature at well(s)
- MIcroseismicity
- $CO_2$ saturation plume

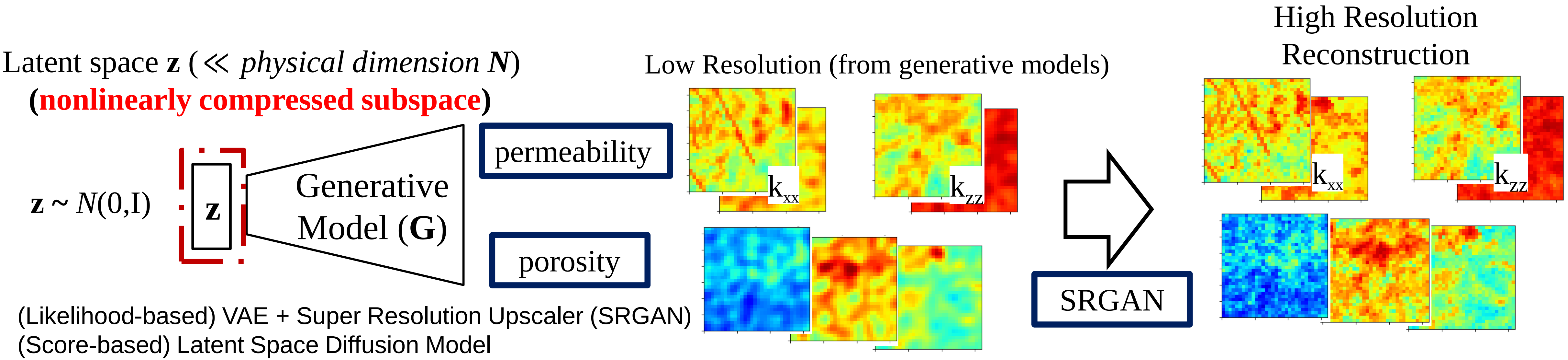
**DA & HM Methods**
- Latent space-based DA
- Ensemble-Smoother with Multiple DA (ES-MDA)

Latent space "**z**" obtained by generative ML models (e.g., VAE, WGAN, DM) is updated in (Ensemble or Variational) DA-based methods with various measured data

Bao et al. (Geoenergy Sci Eng, 2024),
Yoon et al. (GHGT 2022 & in prep)

# Generative Models

- Three generative models are constructed to generate ensemble samples from latent vector
- Variational autoencoder (VAE), Wasserstein generative adversarial network (WGAN-GP,) and diffusion model (DM) are all available



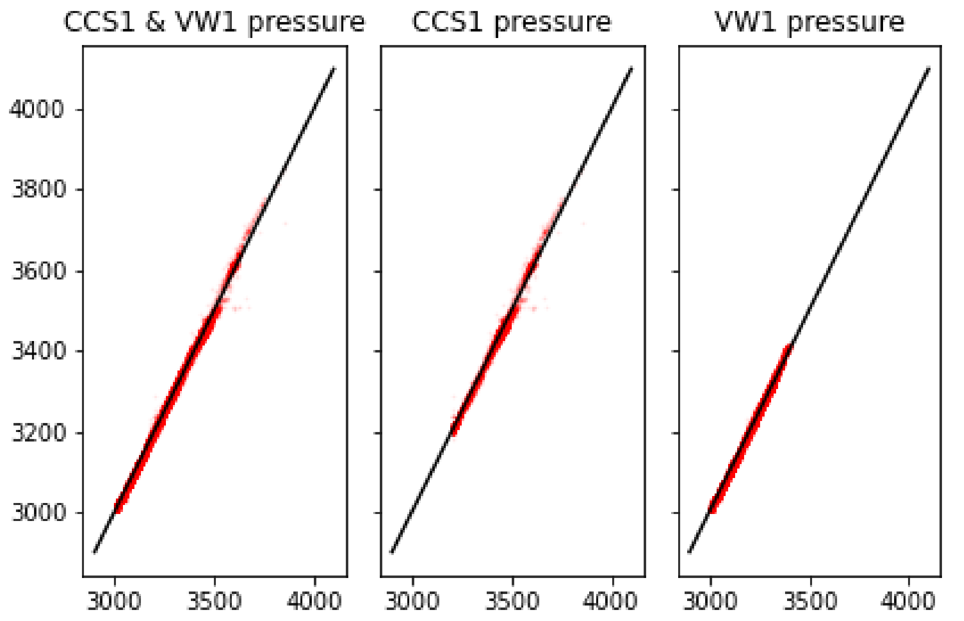

- (Likelihood-based) VAE + Super Resolution Upscaler (SRGAN)
- (Score-based) Latent Space Diffusion Model

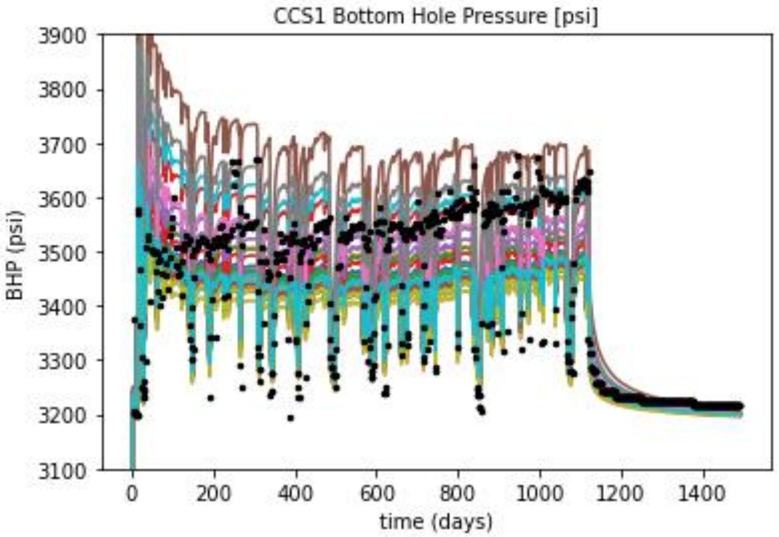
Kang et al. (WRR 2022), Bao et al. (Geoenergy Sci Eng, 2024), Kim et al. (JCH 2021,Sci. Rep. 2021)

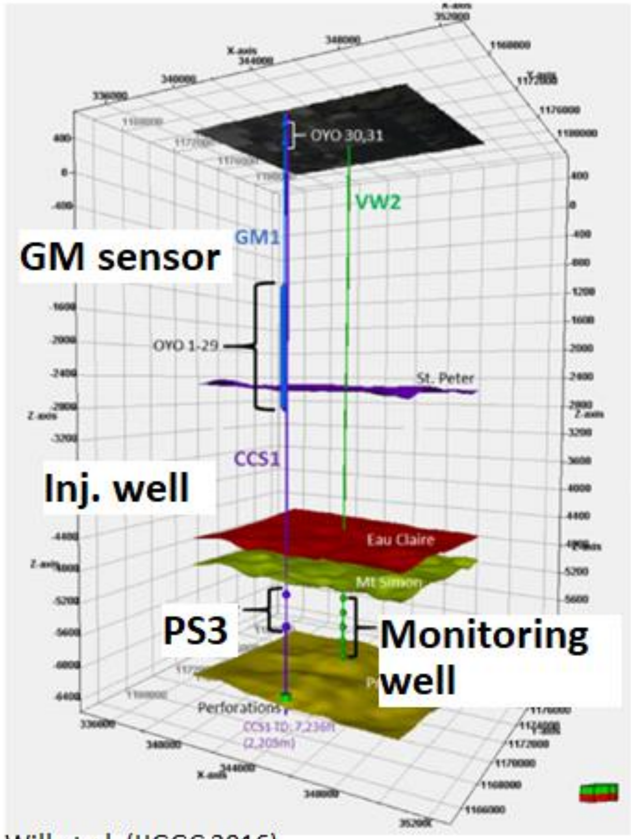
# ML surrogate model – Well Pressure

- For demonstration purpose, a sub-domain (40x44x94 out of 126x125x110) is used
- ML model: CNN-LSTM (static and dynamic data are concatenated after feature vectors are constructed by repeating static feature vector into dynamic data over time)
- Inputs:
  - Static fields (7 features): 3D permeabiltiy (x,y,z), porosity, transmissivity (x,y,z) fields
  - Dynamic data: Time, daily injection rates, cumulative injection volu
- Output: well pressures at seven sensor locations (injection (CCS1) and

Testing results RMSE = 4.4psi



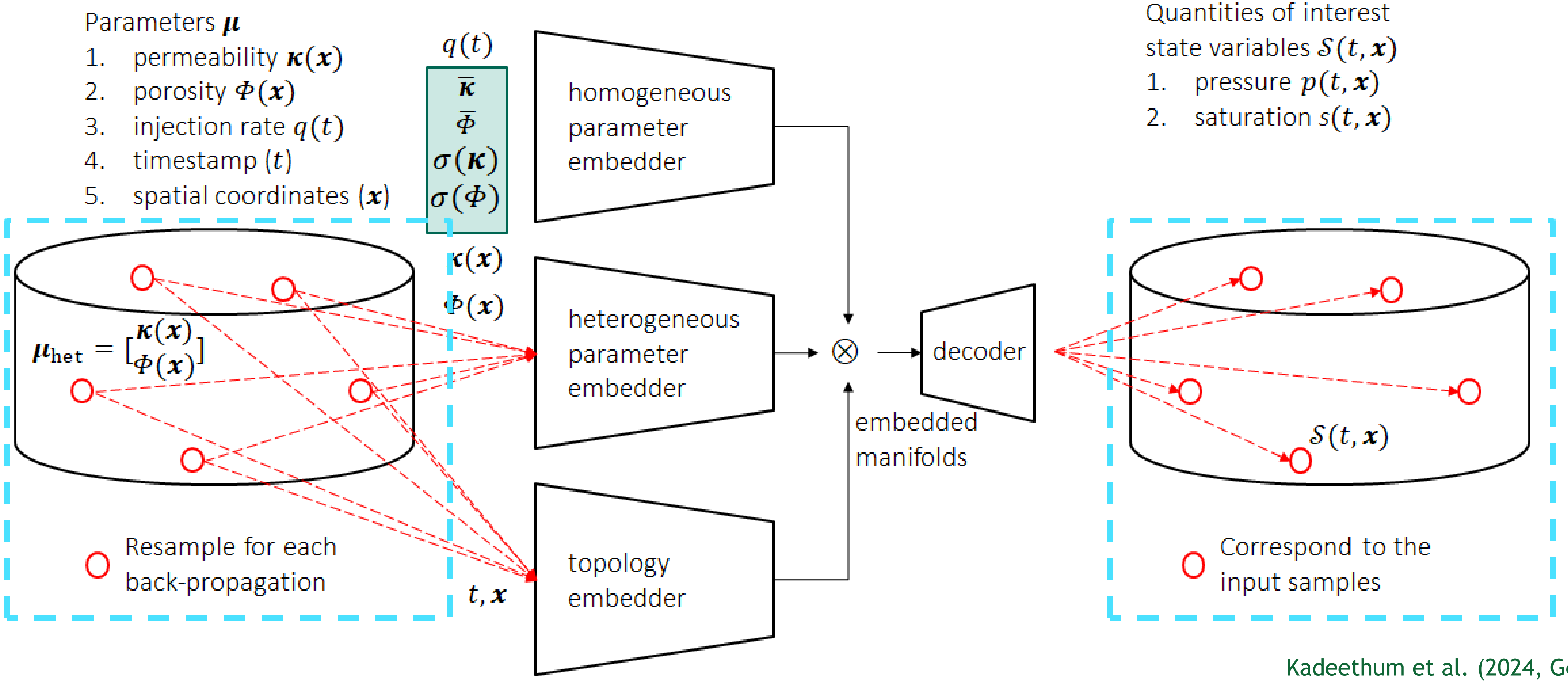




Will et al. (IJGGC 2016)

**Fig. 1.** Subsurface array configuration. Distance units are feet, Z axis is referenced to mean sea level.

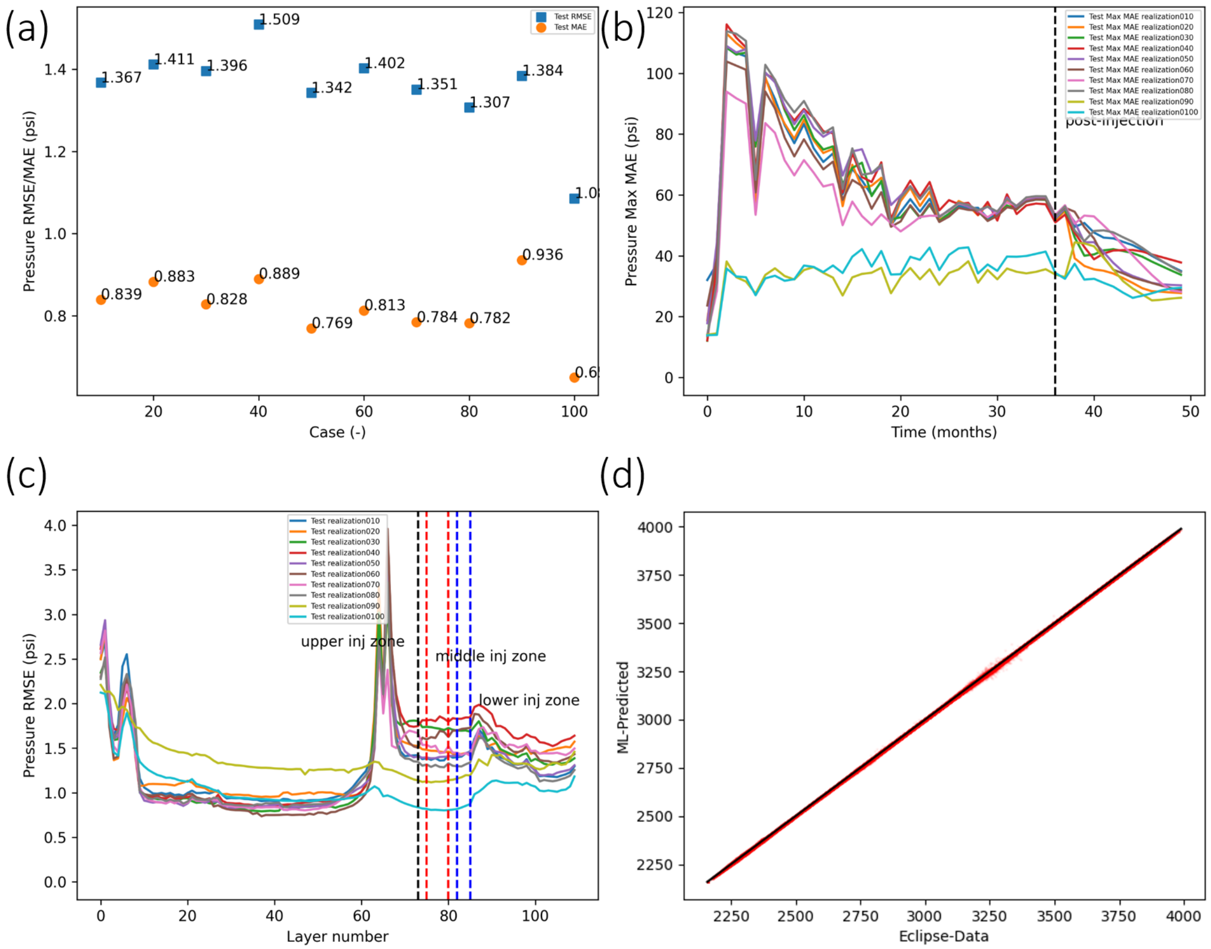
# Improved Neural Operator (INO): Pressure & Saturation at grid scale

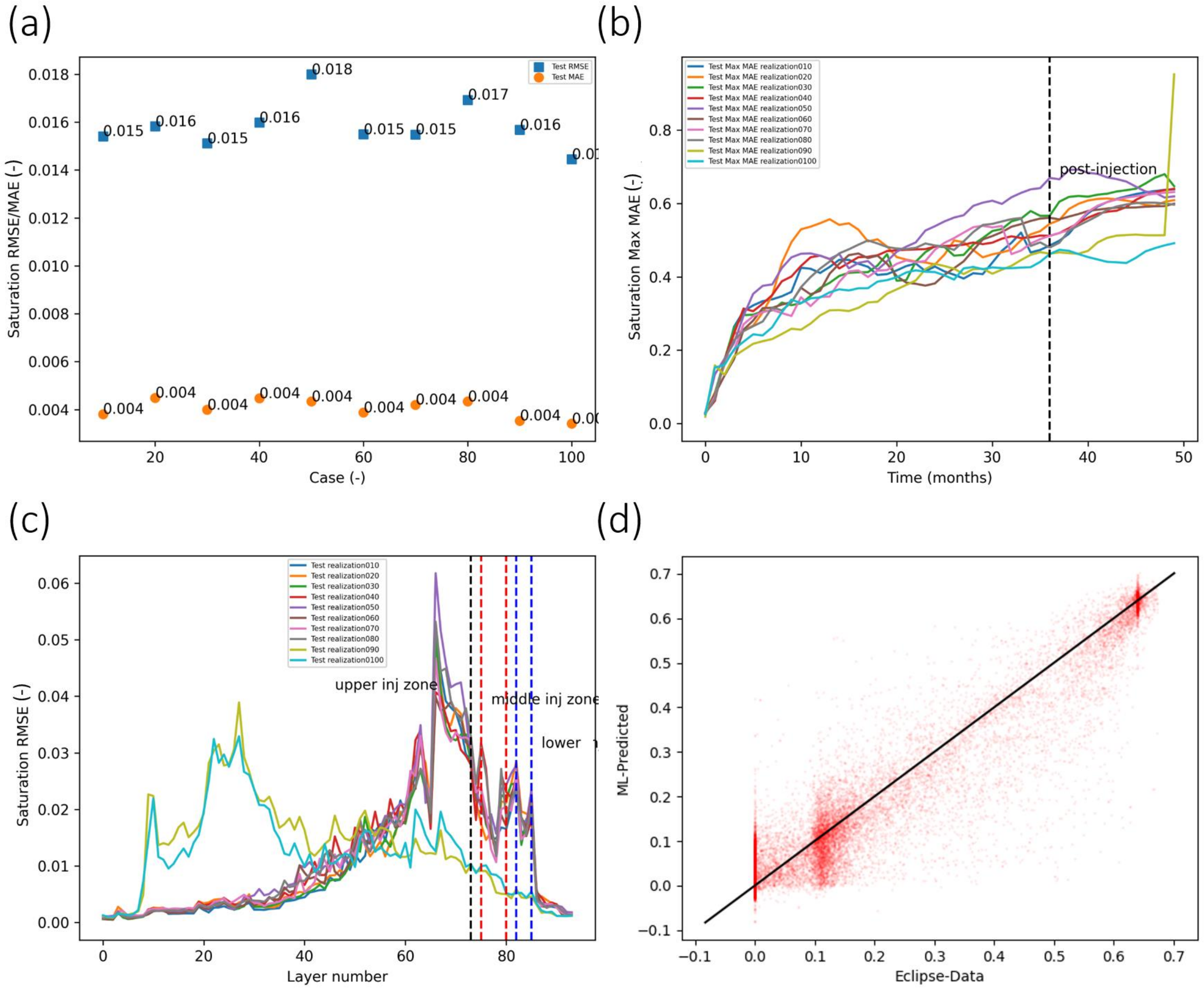
- Dramatically improved in computational efficiency with good accuracy through subsampling
- Inputs:
  - Static fields (7 features): 3D permeabiltiy (x,y,z), porosity, transmissivity (x,y,z) fields
  - Dynamic data: Time, daily injection rates, cumulative injection volume
- Output: Pressure or saturation at the grid scale




Kadeethum et al. (2024, Geoenergy Sci and Eng)

# INO Results for Pressure and Saturation

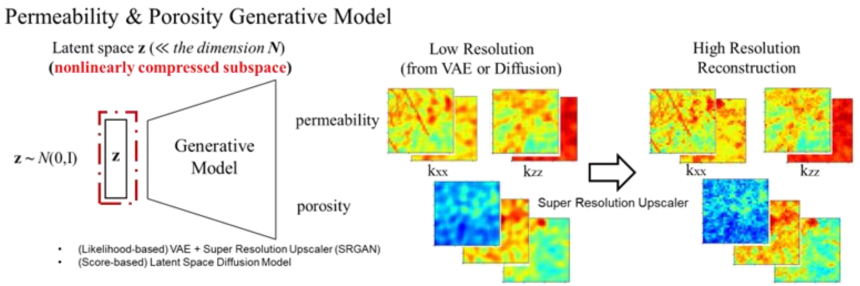


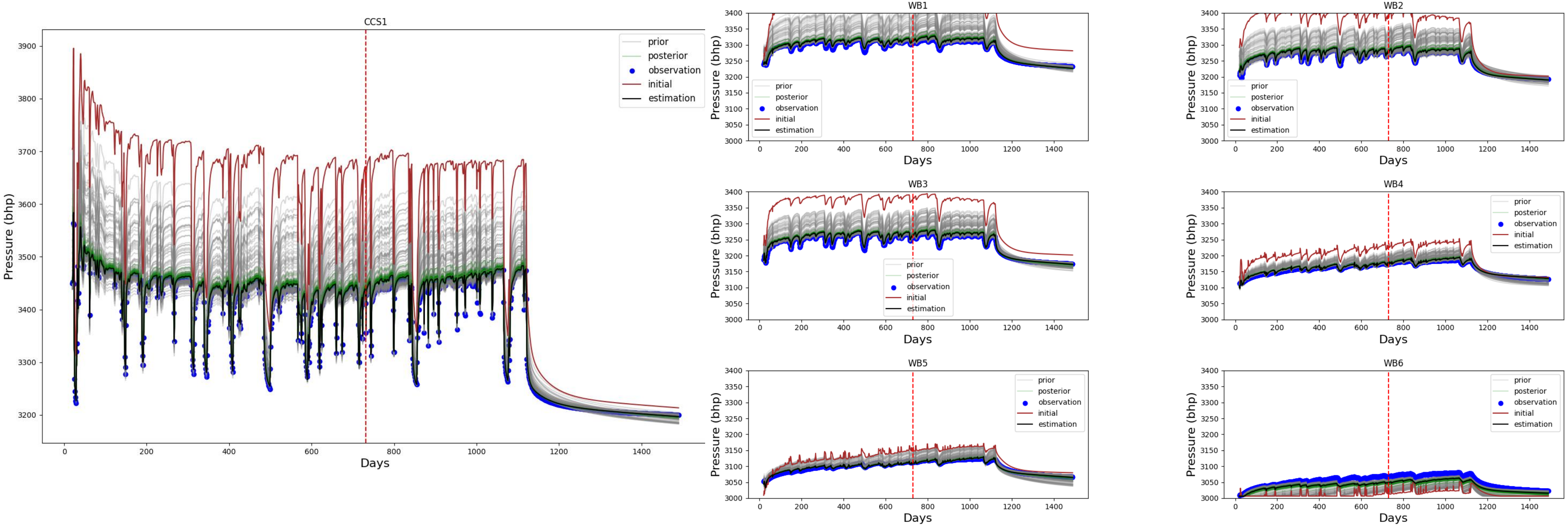




Kadeethum et al. (2024, Geoenergy Sci and Eng)

# History matching with synthetic observed pressures at wells

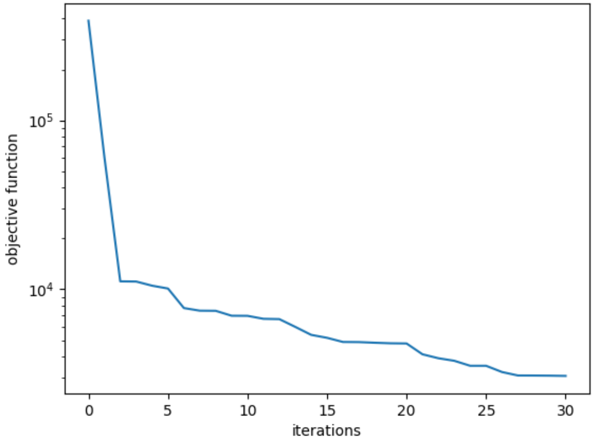
- Calibration with the first 2 yrs data and blind test with the rest period (VAE as generative model)
- With pre-trained ML models, only takes ~10min for HM



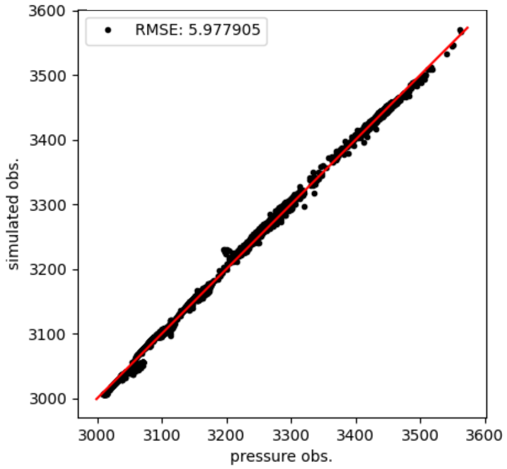



# History matching with synthetic observed pressures at wells

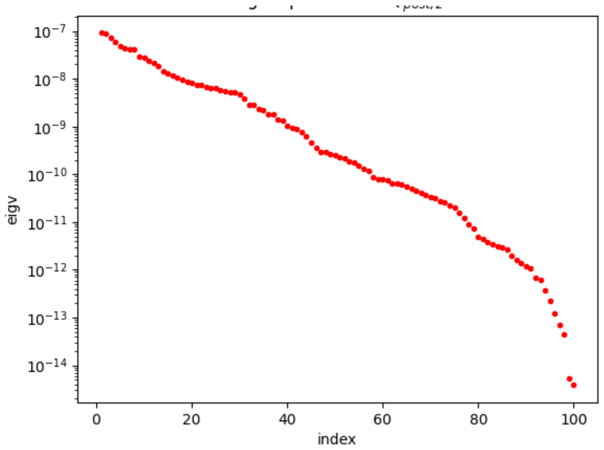
## Obj. function



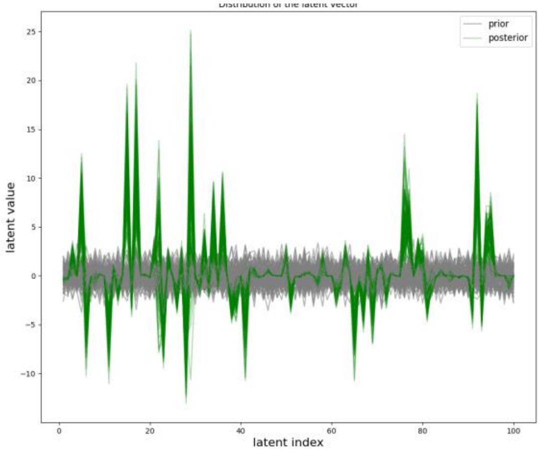

## Data fitting



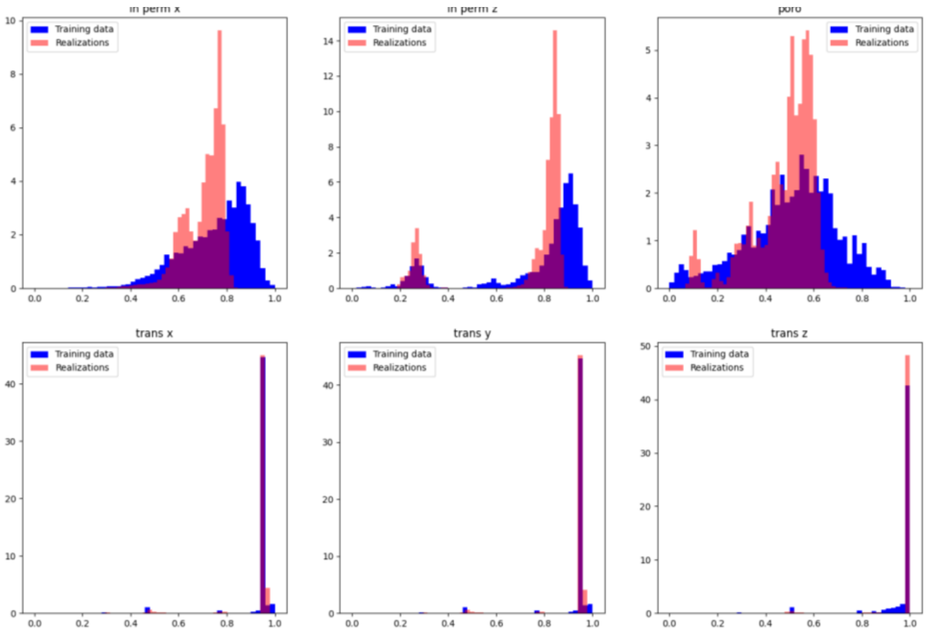

## Eigen Spectrum



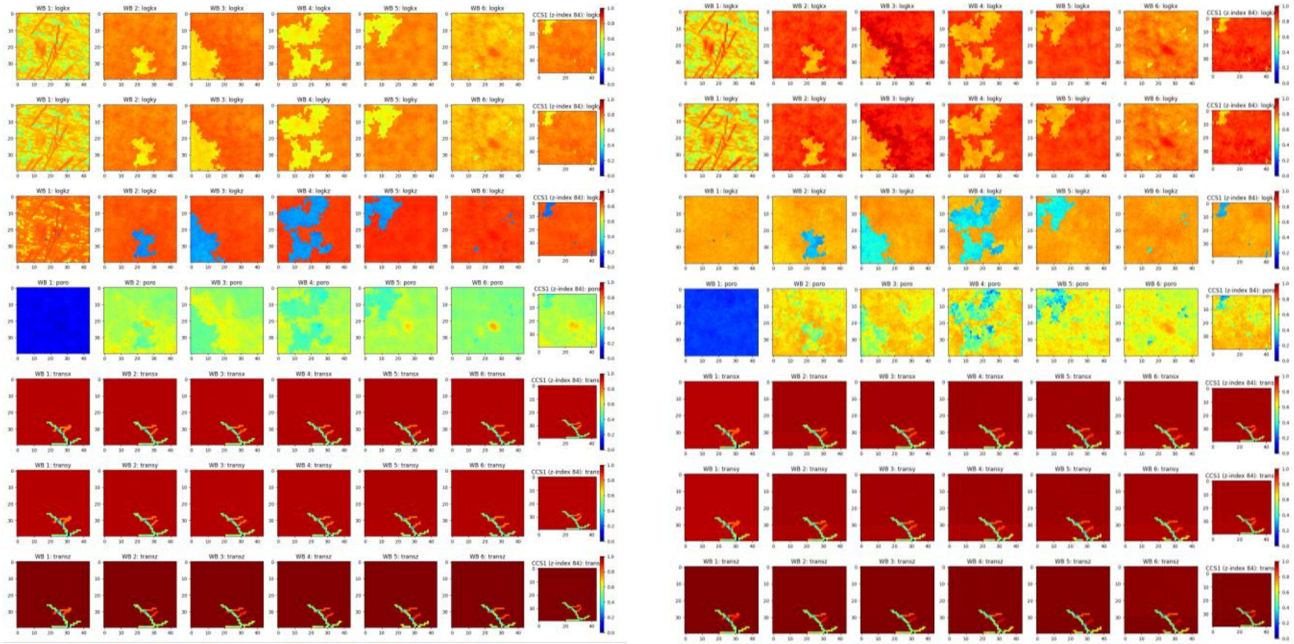

## Latent space in VAE



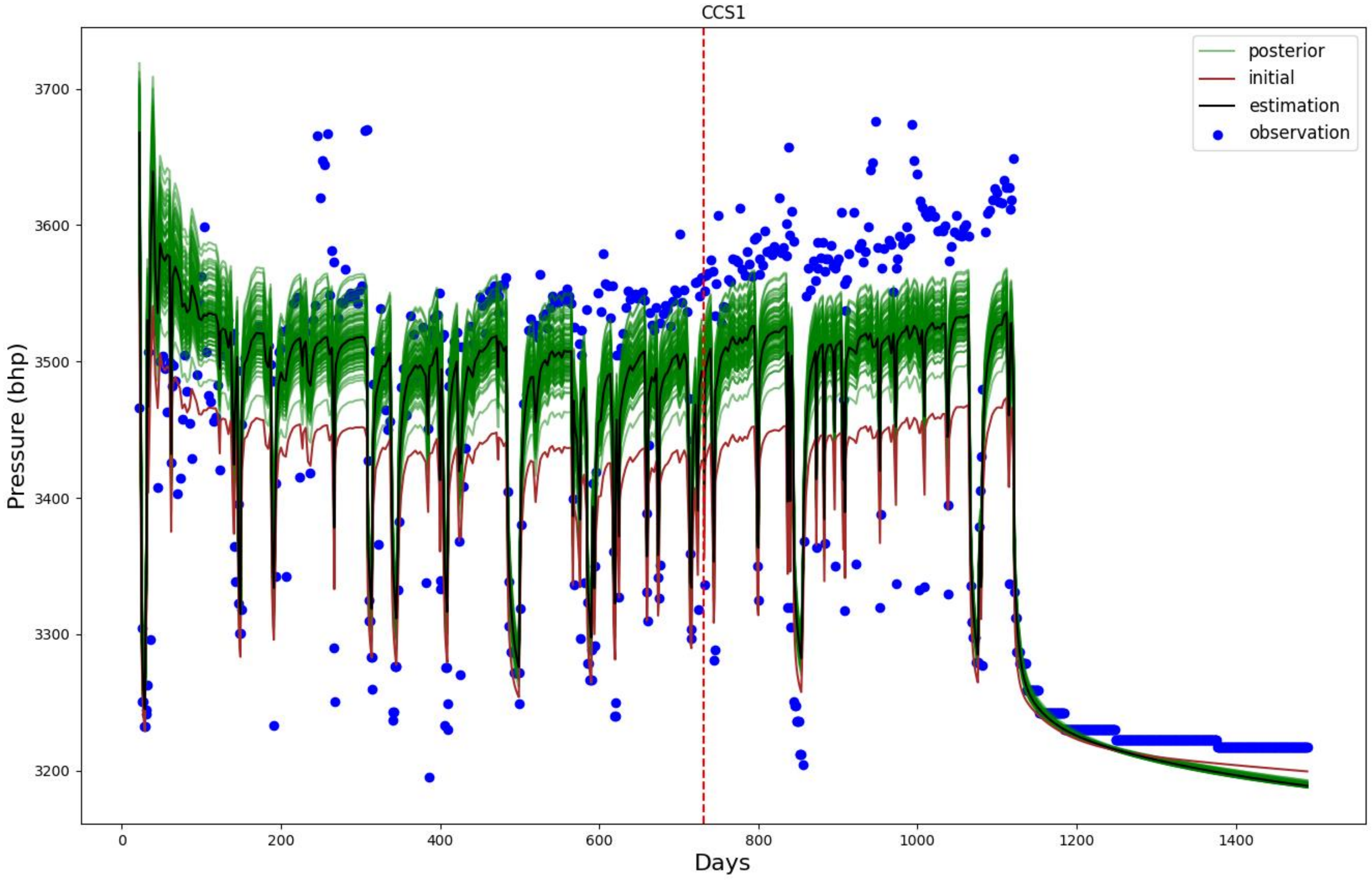

## Distribution of parameters (prior and MAP)




## Examples of 7 input parameters from initial and final estimates

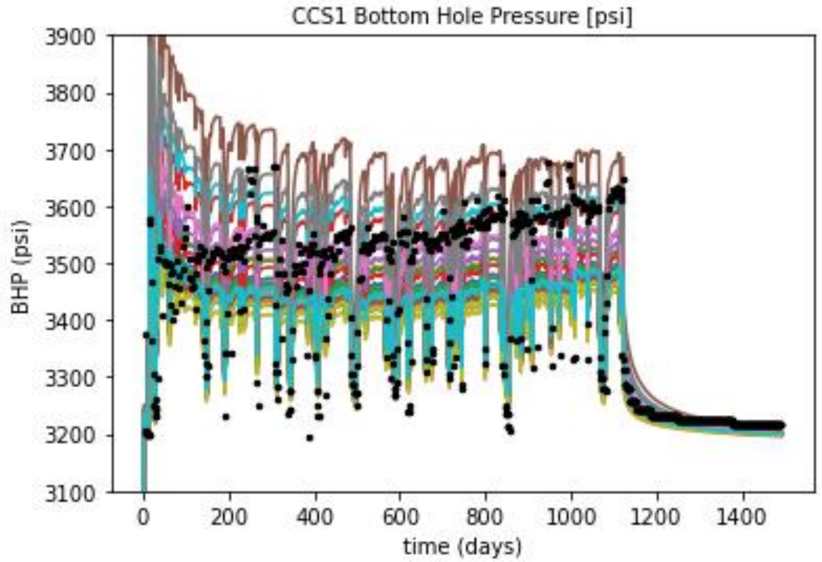


# History matching with REAL observed pressures at wells



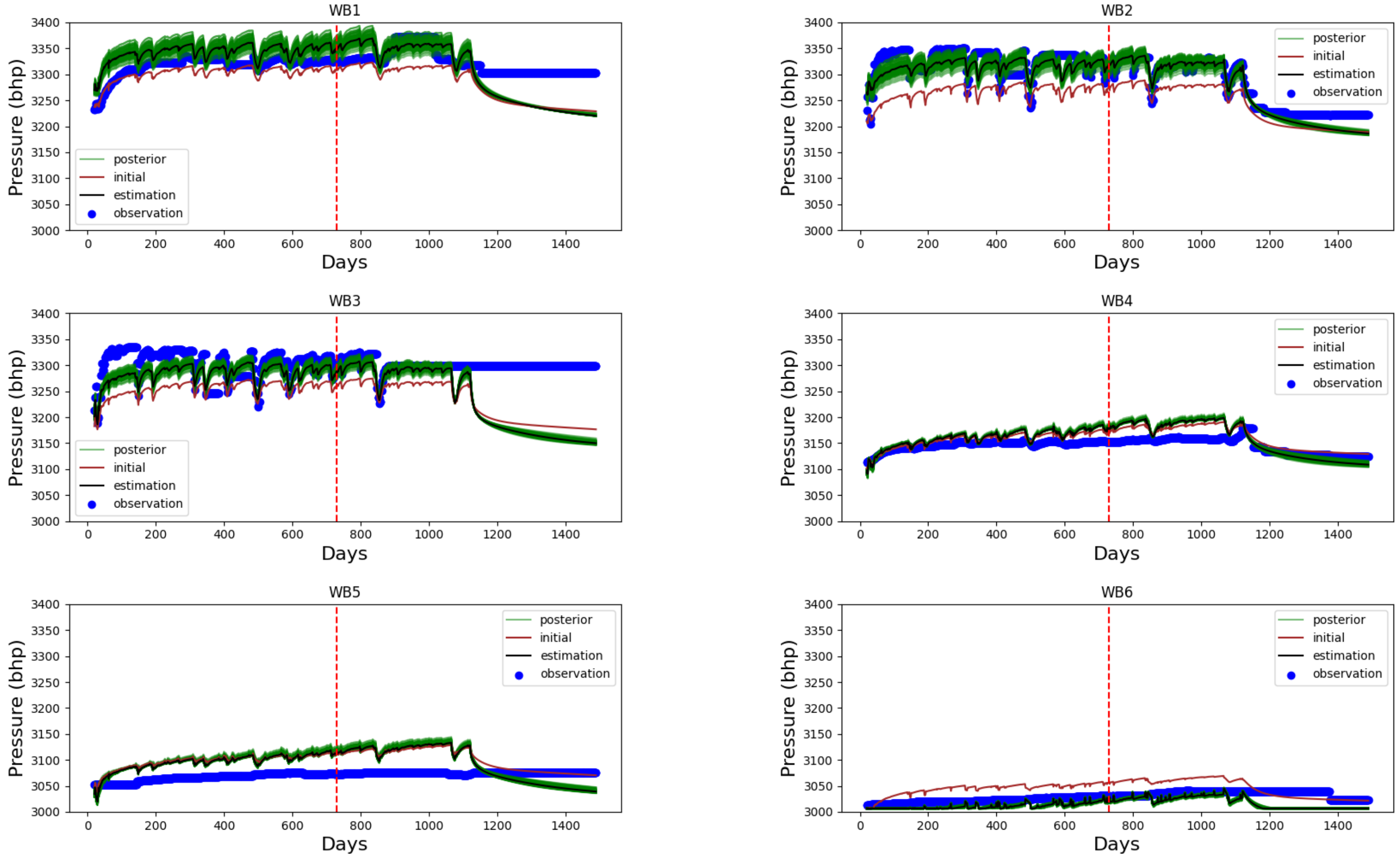

- BHP at CCS1 tends to increase after ~ 2yrs -> challenge to match after 2 yrs calibration
- Fluctuations due to frequent shut-ins pose a challenge to balance off between general BHP trend and low and high fluctuations during calibration

Training data

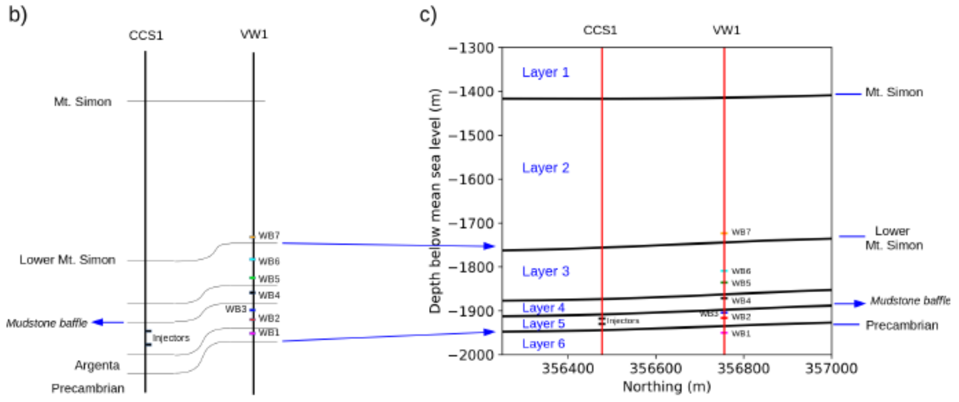


# History matching with REAL observed pressures at wells

Blue dots: Observation; black: the best estimate




- From observation data there may be a stronger barrier (e.g., baffles) than training data used
- Higher estimates at WB 1 (Basement) are compensated with lower estimates at WB2-3 (reservoir units)
- Lower uncertainty bounds may indicate the limit on information gained though calibration using VW1 P data



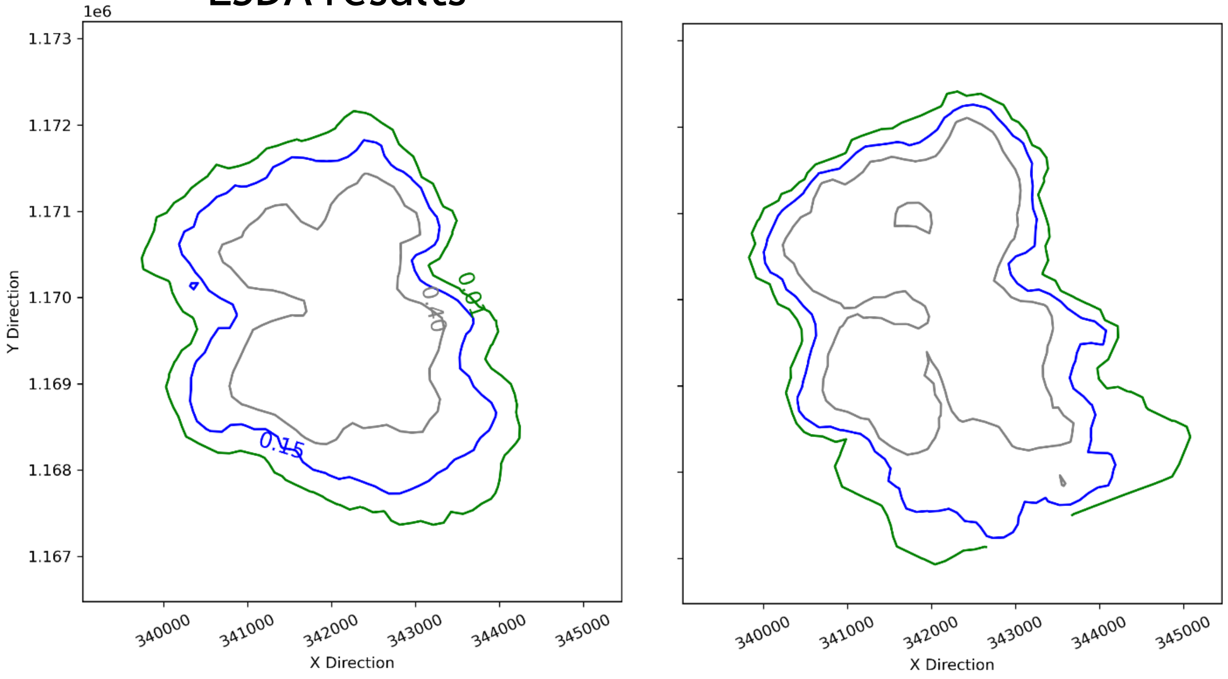
# Comparison of $CO_2$ saturation

- Updated reservoir fields from LSDA are used with the INO model to predict $CO_2$ saturation at the grid scale
- Traditional HM results using high fidelity Eclipse runs

Maximum spreading (Area of Review, AoR) projected on the horizontal plane

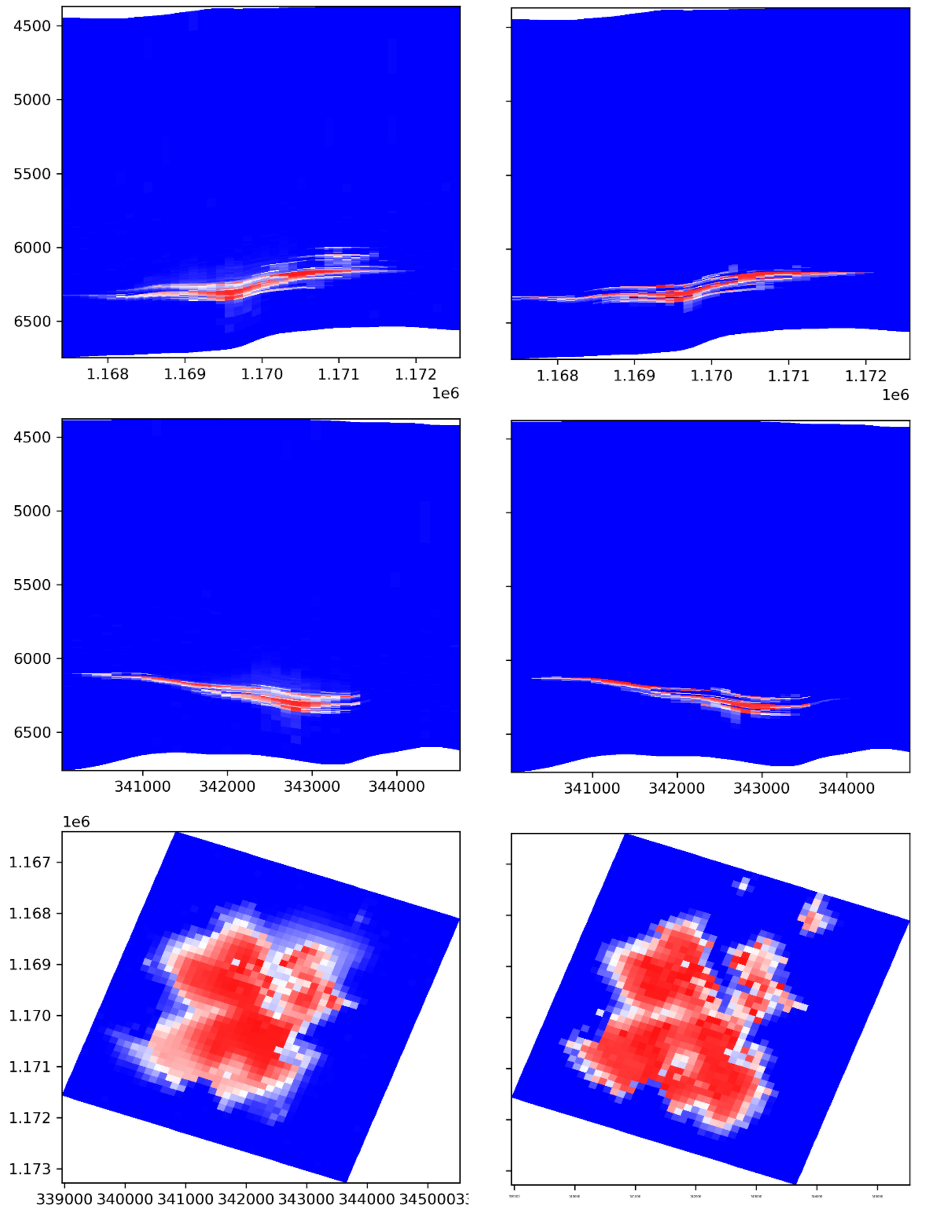
**$CO_2$ AoR**

LSDA results | Traditional HM results




**3D $CO_2$ plume**

LSDA results | Traditional HM results



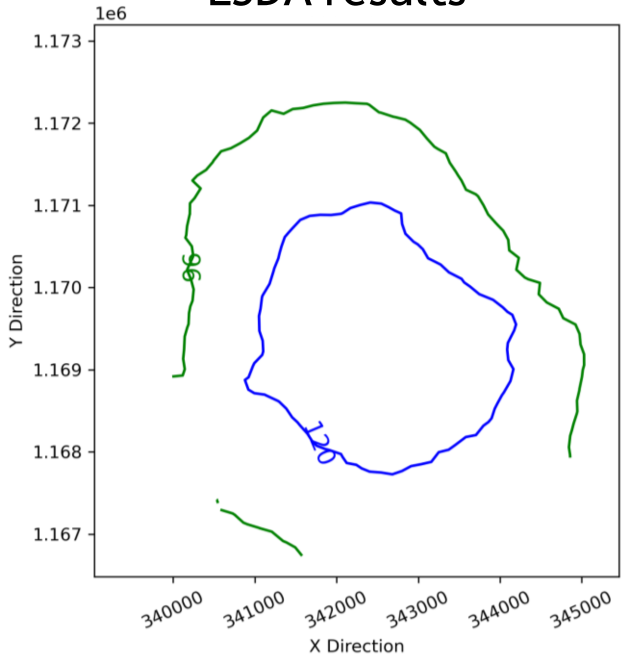
# Comparison of Pressure

- Updated reservoir fields from LSDA are used with the INO model to predict pressure at the grid scale
- Traditional HM results using high fidelity Eclipse runs

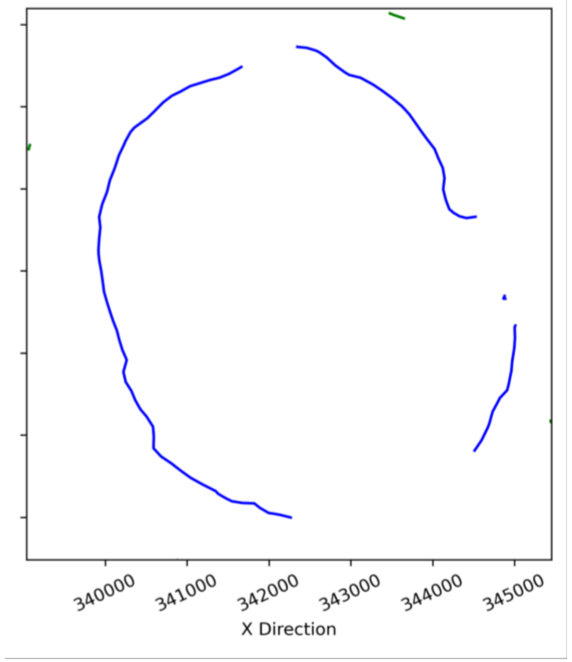
Pressure profile (Area of Review, AoR) projected on the horizontal plane

**Pressure AoR**

LSDA results

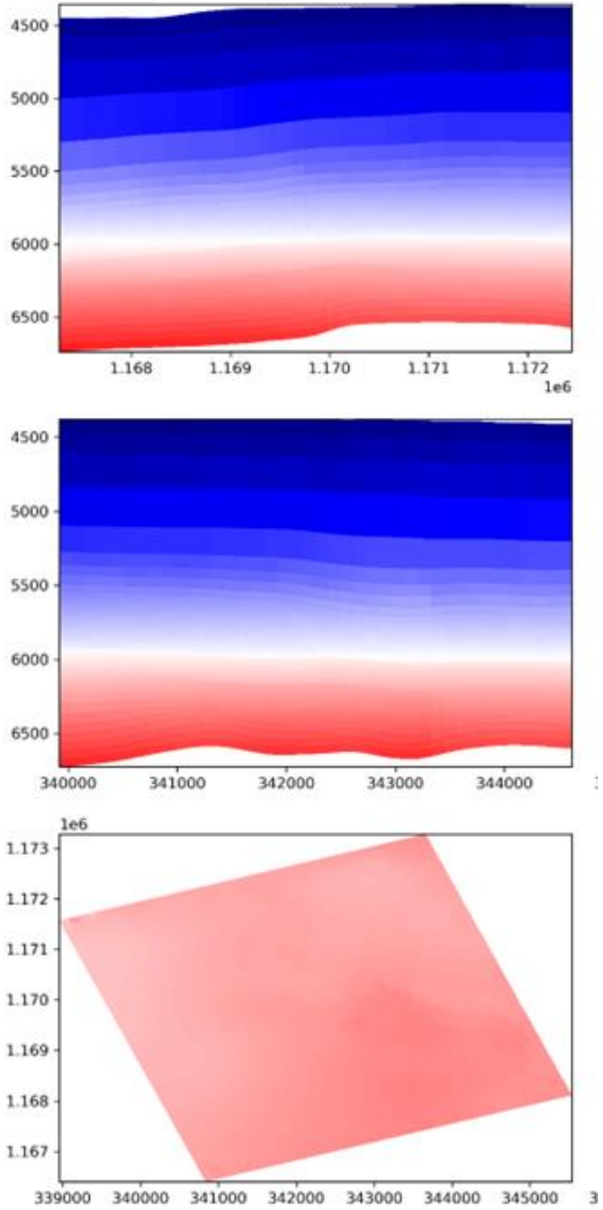



Traditional HM results




**3D Pressure field**

LSDA results



# Summary

- Data assimilation/History matching in the latent space with deep learning methods (VAE, WGAN, DM) and fast deep learning-based forward modeling can achieve real-time history matching of $CO_2$ operations and forecasting pressure plume development.

- Latent space optimization including optimal choice of the nonlinear dimension reduction requires further study with more realistic and various types of observed data.

- ML/DL with domain knowledge can lead to dramatic improvement in spatio-temporal data analytics and decision making for mitigating potential risks and optimal monitoring system development.

# Summary Slide

Machine learning-driven $CO_2$ modeling by combining fast ML-based forward modeling with latent space-based data assimilation (LSDA), resulting in real-time history matching (HM) of $CO_2$ operations and forecasting $CO_2$ and pressure plume development in an end-to-end fashion at the Illinois Basin-Decatur Project site, Decatur, IL, USA.

## Latent Space-based DA (LSDA)