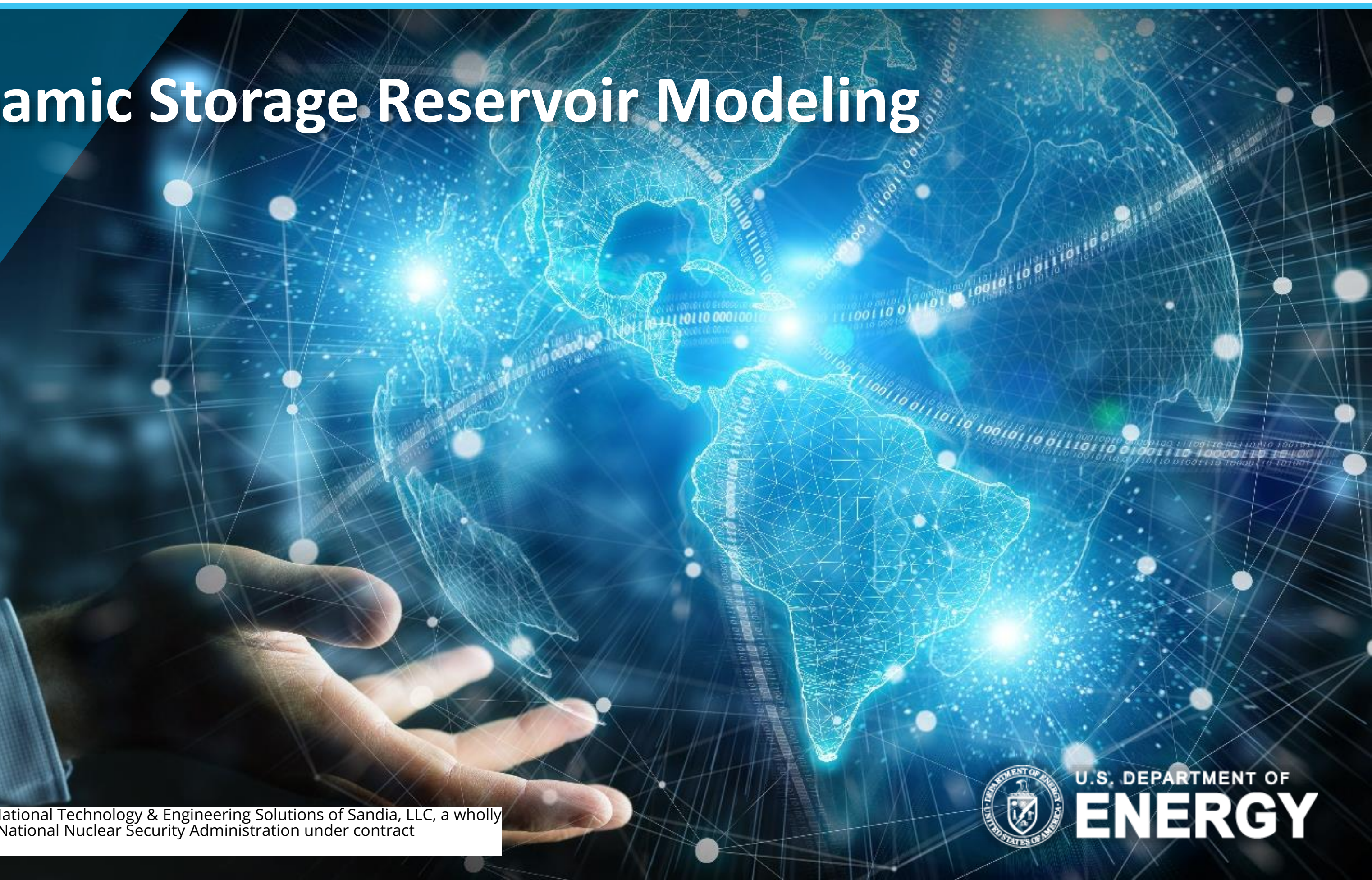
SAND2023-08438C

SMART-CS Initiative

Science-informed Machine Learning to Accelerate Real Time (SMART) Decisions in Subsurface Applications

Task 5: Field Deployment – Dynamic Storage Reservoir Modeling



FECM Project Review Meeting
August 28 – September 1, 2023

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell Inter Energ Admi NA0003525.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

U.S. DEPARTMENT OF ENERGY

Task 5 –Dynamic Storage Reservoir Modeling

Goal: Provide real-time modeling, data assimilation and forecasting to support:

- Field management, to maximize storage while minimizing pressure buildup
- Induced seismicity risk assessment

Task 5 –Dynamic Storage Reservoir Modeling

Goal: Provide real-time modeling, data assimilation and forecasting to support:

- Field management, to maximize storage while minimizing pressure buildup
- Induced seismicity risk assessment

5.1: Unified simulation platform and data generation

5.2: Rapid physics-based predictive models for flow and geomechanics

5.3: Machine learning surrogate models

5.4: Rapid data assimilation and history matching

5.5: Optimization of field Parameters

Task 5 –Dynamic Storage Reservoir Modeling

Goal: Provide real-time modeling, data assimilation and forecasting to support:

- Field management, to maximize storage while minimizing pressure buildup
- Induced seismicity risk assessment

Today

5.1: Unified simulation platform and data generation ← Part 1

5.2: Rapid physics-based predictive models for flow and geomechanics

5.3: Machine learning surrogate models ← Part 2

5.4: Rapid data assimilation and history matching ← Part 3

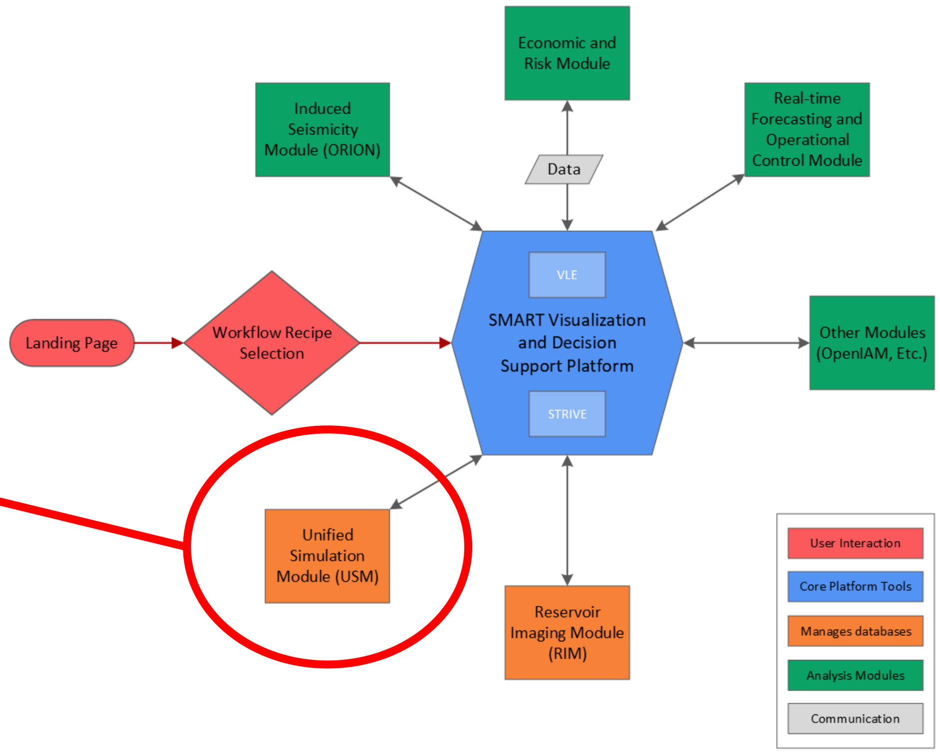
5.5: Optimization of field Parameters

Part 1: Unified Simulation Module

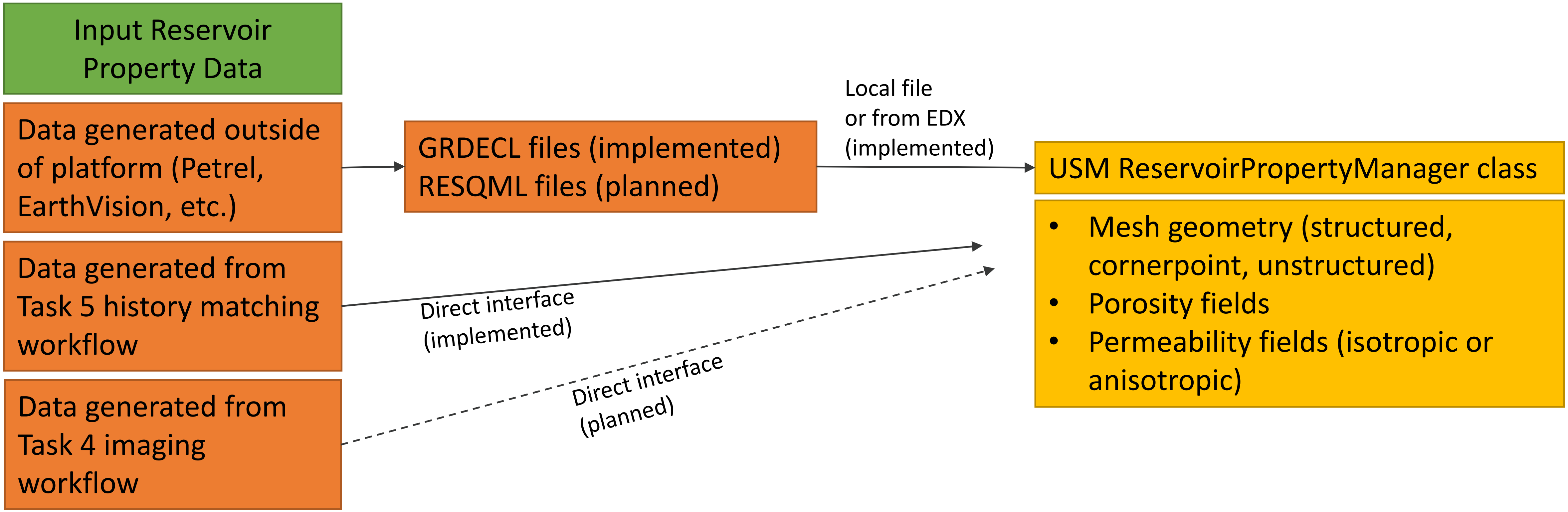
Unified Simulation Module

Objectives:

- Provide a unified and flexible way for a user to interact with reservoir simulation data
- Read in data from commonly-used formats (GRDECL, RESQML), or from other SMART workflows (Task 4 imaging)
- Convert data into formats needed by ML models (Numpy arrays)
- Convert output of ML models into formats needed for visualization and analysis in the VLE and STRIVE



Unified Simulation Module



Unified Simulation Module

Input Injection and Monitoring Well Data

Data generated outside of platform → *.csv, *.xlsx → Local file or from EDX (implemented) → USM OperationalScenarioManager class

Data generated from Task 5 history matching or VLE → Direct interface (implemented but not tested) →

- Well location
- Perforation depths
- Injection rate
- Pressure

Unified Simulation Module

USM ReservoirPropertyManager

USM OperationalScenarioManager

USM ForwardModel classes

- Base class that defines the interface, making all forward models interchangeable
- Derived classes implement specific forward models
- Directly use ReservoirPropertyManager and OperationalScenarioManager objects
- Outputs ReservoirStateManagerObject

USM ReservoirStateManager

- Mesh geometry
- Pressure field time history
- Saturation field time history
- Stress field time history (planned)
- Strain field time history (planned)

Unified Simulation Module

NATIONAL ENERGY TECHNOLOGY LABORATORY

USM ReservoirStateManager

- Mesh geometry
- Pressure field time history
- Saturation field time history
- Stress field time history (planned)
- Strain field time history (planned)

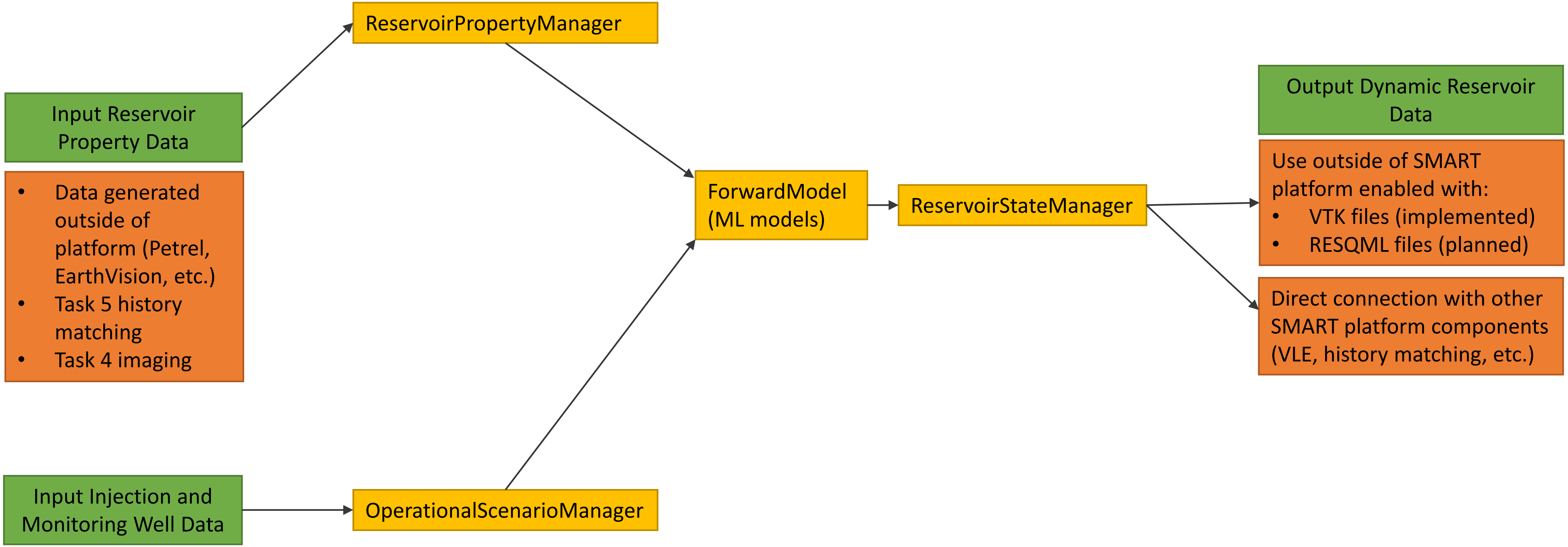
Output Dynamic Reservoir Data

Use outside of SMART platform enabled with:

- VTK files (implemented)
- RESQML files (planned)

Direct connection with other SMART platform components (VLE, history matching, etc.)

Unified Simulation Module Data Flow



Unified Simulation Module

NATIONAL ENERGY TECHNOLOGY LABORATORY

Quality control and documentation

- USM code hosted on GitLab
- Installable Python package makes it easy to use
- Automated unit testing suite tests every commit pushed
- Standardized code formatting and style
- Sphinx documentation is automatically built
- Issue and milestone tracking on GitLab

Testing pipeline

Status	Pipeline	Triggerer	Stages
passed 00:00:28 1 month ago	forgot to include new DataManagerBase sourc... #930552655 feature/pickle a0bec5e8 latest		
passed 00:00:28 1 month ago	applied yapf #930551570 feature/pickle 55c1784d		
passed 00:06:27 1 month ago	Merge branch 'petrel-support' into 'main' #929529305 main 60dbd821 latest		
passed 00:03:08 1 month ago	switched tests to using a small 1D 2-cell grdec... #929526526 6 e640725e latest merge request		
passed 00:00:28 1 month ago	switched tests to using a small 1D 2-cell grdec... #929526505 petrel-support e640725e		

Issue tracking

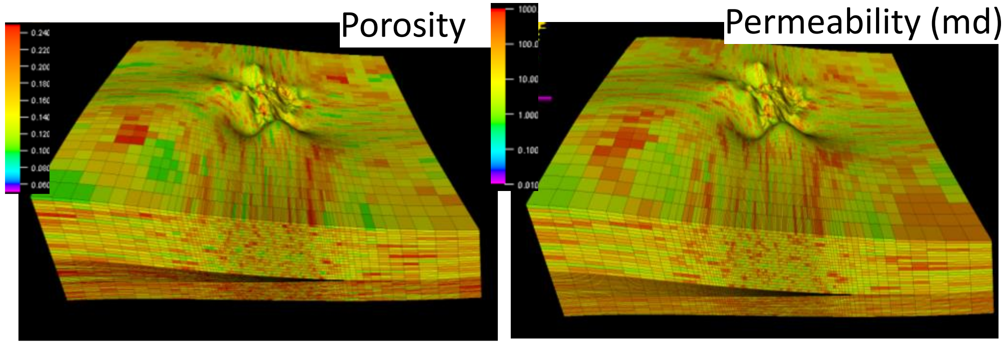
- **Create methods to access pickled objects from EDX** — 0
#9 · created 1 month ago by Jeffrey Burghardt USM Deployment — updated 1 month ago
Enhancement
- **fix data scaling in UTBEG model** — 0
#8 · created 1 month ago by Jeffrey Burghardt
- **STRIVE Integration** — 1
#7 · created 2 months ago by Christopher Sherman USM Deployment — updated 1 month ago
priority: high
- **Add sphinx documentation for mesh, property reshaping** — 0
#6 · created 2 months ago by Christopher Sherman Initial USM Development
priority: medium type: documentation
- **Code restructure related to handling of field attributes** — 0
#5 · created 2 months ago by Veronika Vasylkivska
Enhancement type:optimization
- **Convenience functions for slicing mesh** — 0
#3 · created 2 months ago by Kayla Kroll
Enhancement

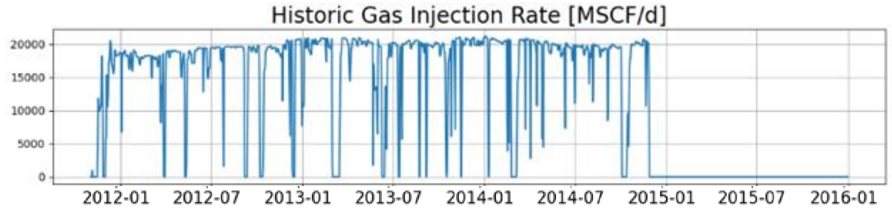
Part 2: ML Surrogate Modeling

ML Input Data

- Monthly pressure and saturation distributions at IBDP Site at 1.73M cells in 100 realizations of permeability and porosity fields with actual CO_2 injection rates (1 M tons for 3 years)

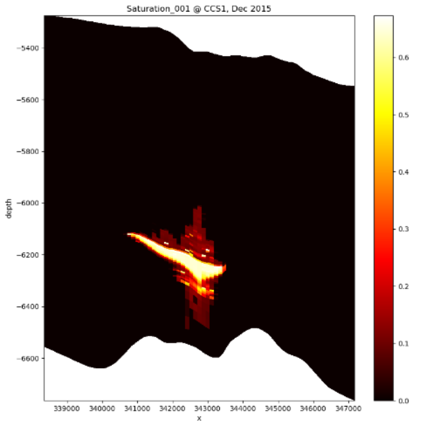
- Training (90 cases) and testing (10 cases)
- Input data
  - Injection rate: (100, 50)
  - Permeability: (100, 126, 125, 110, 3)
  - Porosity: (100, 126, 125, 110)
  - Topology: (100, 126, 125, 110)
- Output data
  - Pressure: (, 50, 126, 125, 110)
  - Saturation: (, 50, 40, 44, 94)
- Well data
  - Injection rates: three perforation zones
  - Monitoring: 6 multi-depth sensors



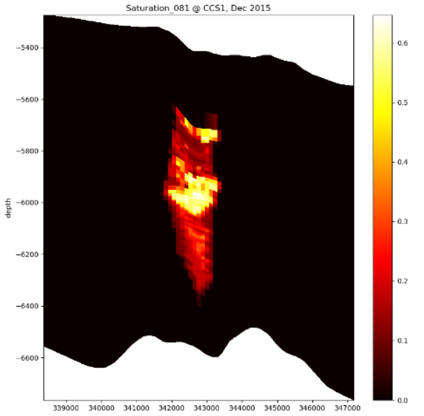


Example of porosity, permeability, and injection rates (input to ML models) & examples of CO_2 saturation distribution at 1 year after the end of injection (Eclipse)

80 cases with open fault horizontally



20 cases with closed fault horizontally



ML Models

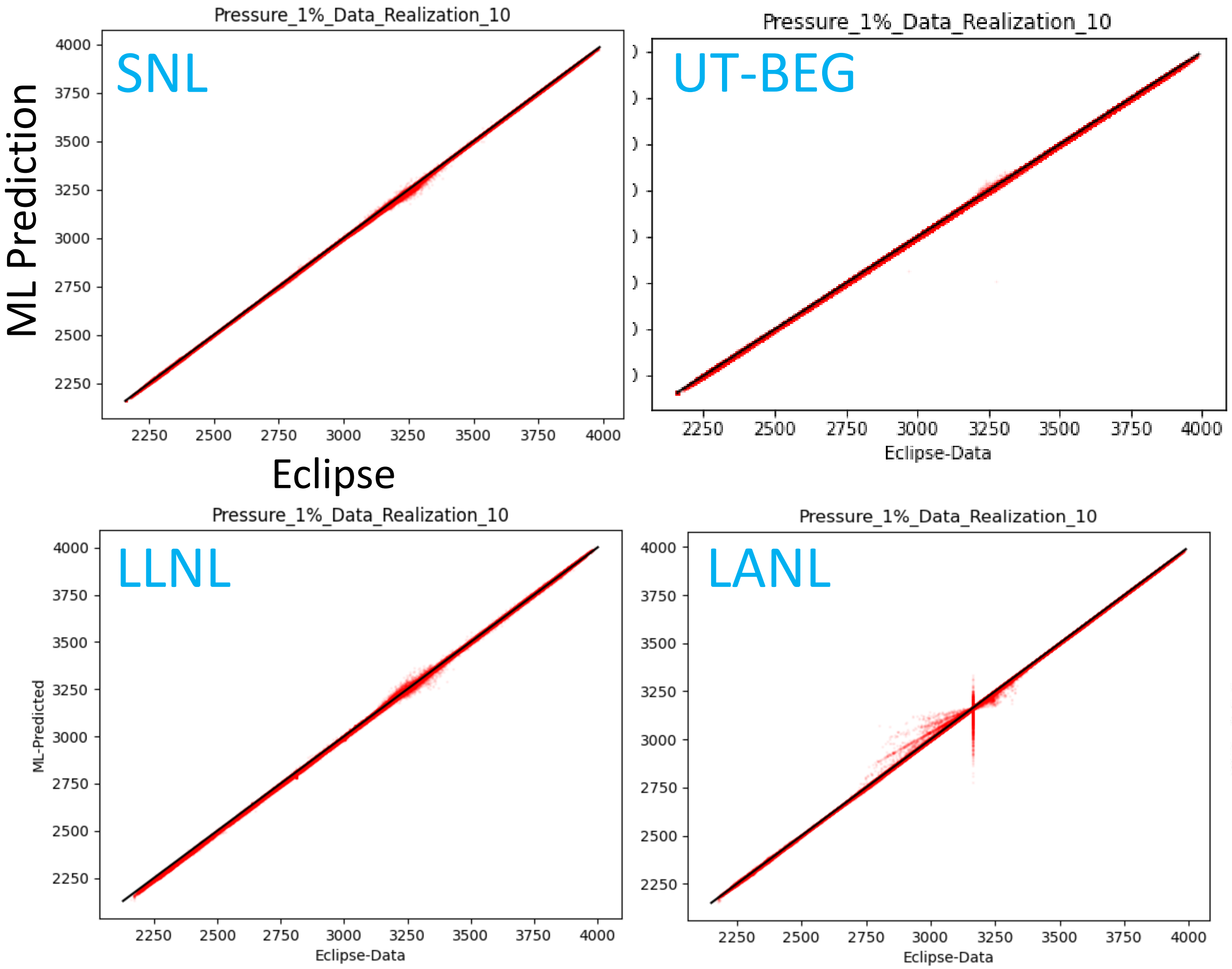
- Computational efficiency to handle real Illinois Decatur Basin Project (IBDP) data
- Prediction accuracy
- Flexibility associated with input, output, portability, and potentially transfer learning

ORG	ML Method	Pressure RMSE (psi)	Saturation RMSE (-)	Note
LANL*	Fourier Neural Operator	~5	~0.015	2D input due to data size on single GPU
LLNL	Fourier Neural Operator	~4	~0.015	32 GPUs for ML training with 3D data (2 & 1 hrs for P &S)
ORNL	Autoencoder-MLP	~20-25	~0.018	Latent space based approach, 2D slice model for pressure
SNL	Modified DeepONet with subsampling	~2	~0.018 (0.015#)	Subsampling for computational efficiency (~ 1hr training on 1 GPU & 2.2M parameters), handling full IBDP data
UIUC	Karhunen-Loeve (KL)-Deep Neural Network	<2	~0.02	Domain needs to be coarsened in both space and time to handle data.
UT-BEG	UNet-MLP	<2	~0.016	Relatively big model (122M parameters, 23.6 hr training on 2 GPUs), handling full IBDP data
WVU	Smart Proxy Modeling – Spatiotemporal DCNN			Not completed over time and space. In progress
NETL	Committee Machine			See figure
BMI	Library look-up model	Not so good	Not so good	Not ML model, but as baseline

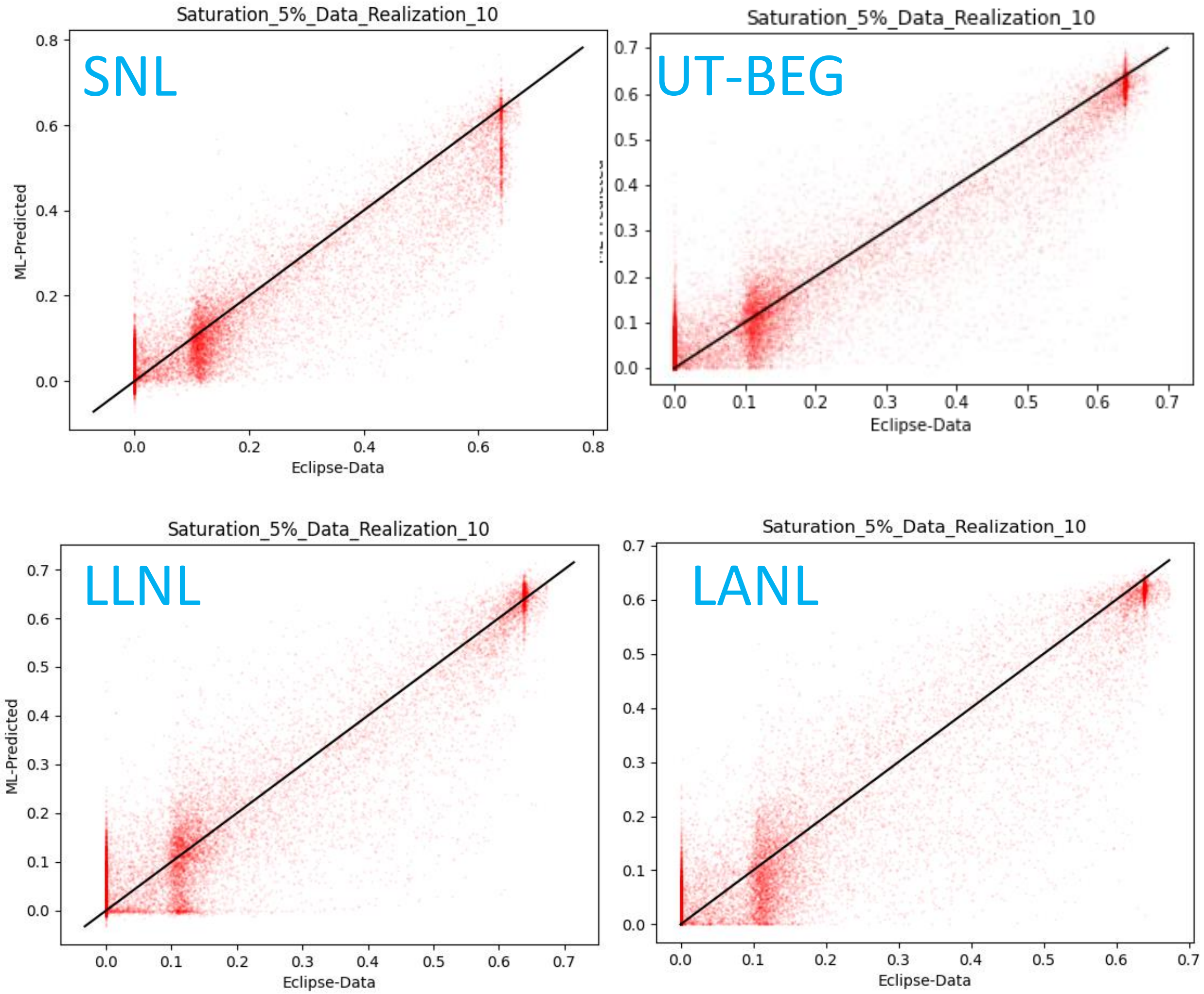
*UNet was also evaluated. Not reported here. # A simple CNN-LSTM model

Pressure & Saturation Prediction (realization 10)

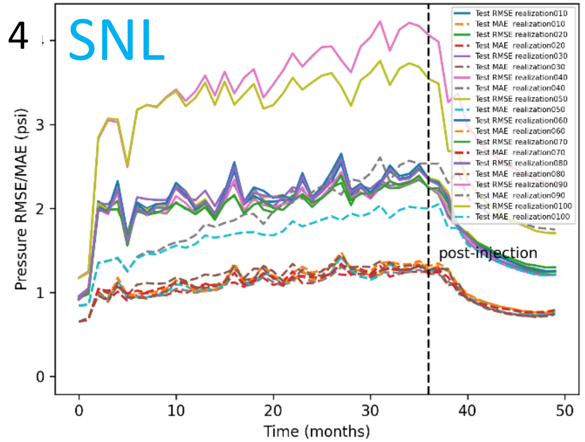
Pressure (1% data)

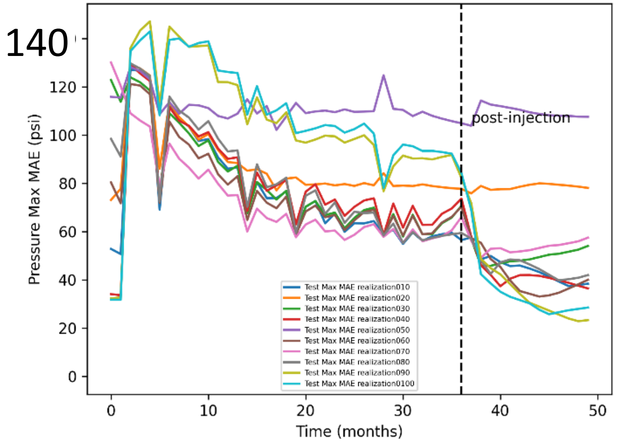


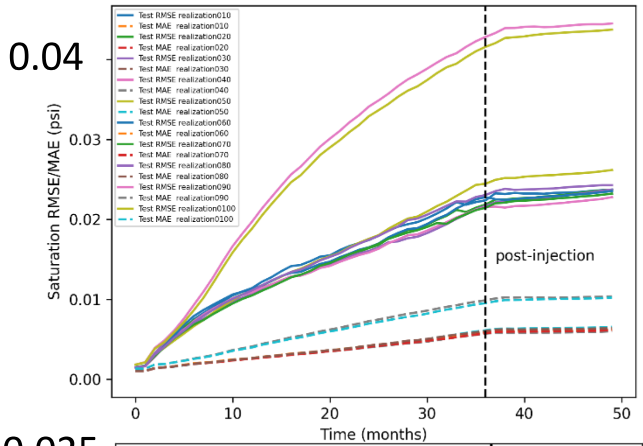
Saturation (5% data)

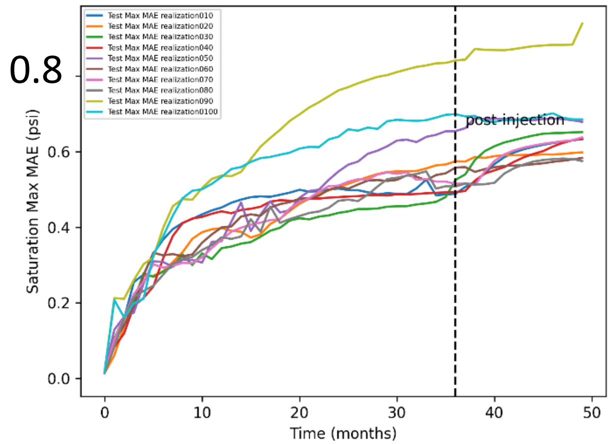


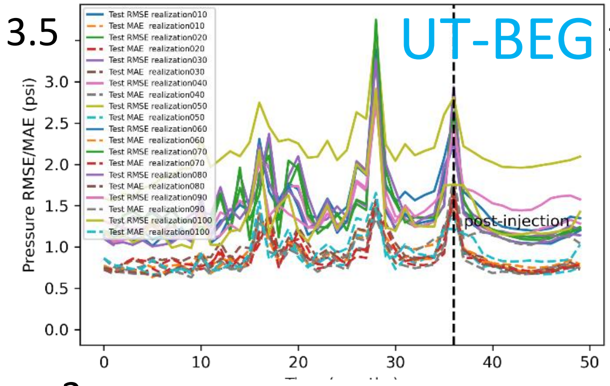
Monthly RMSE/MAE and Max MAE

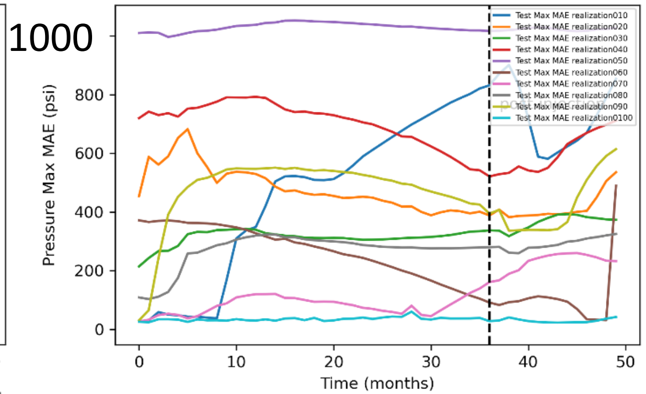


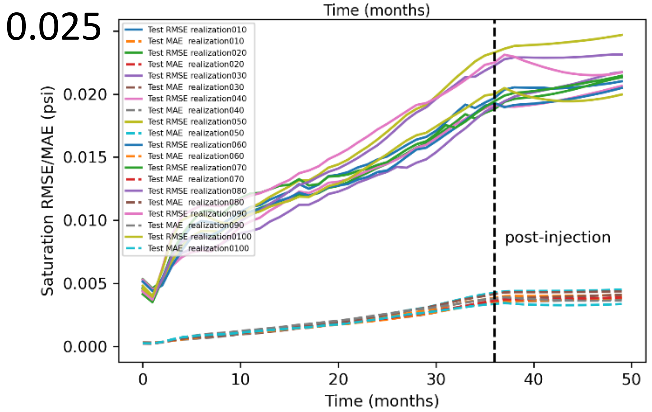


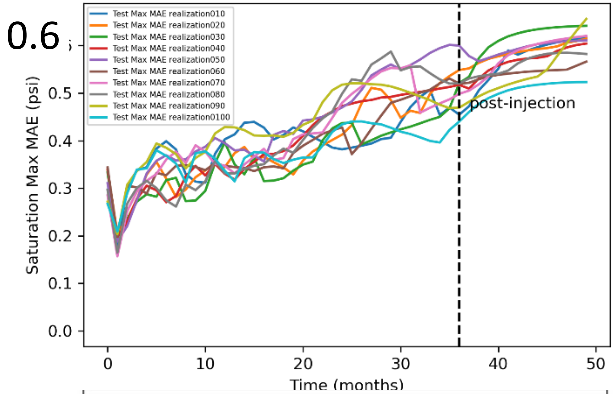


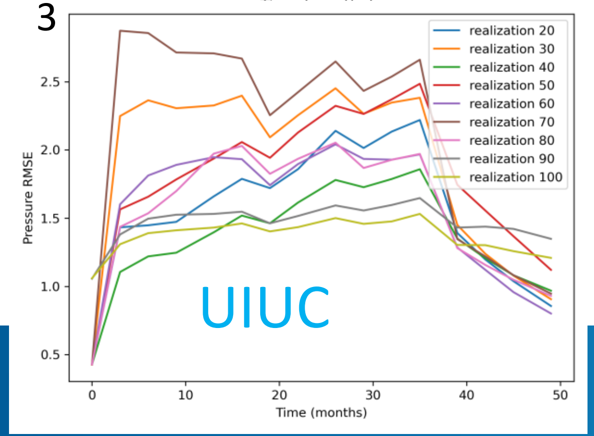


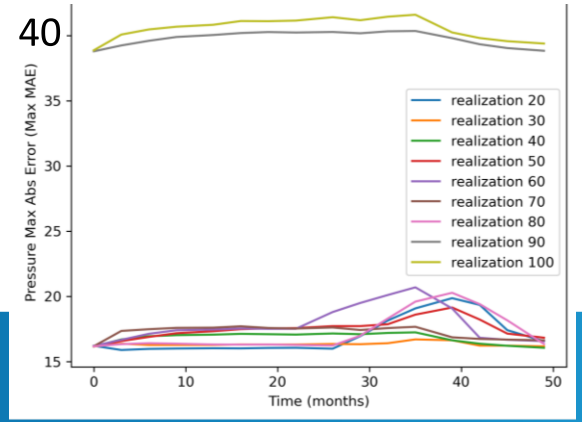


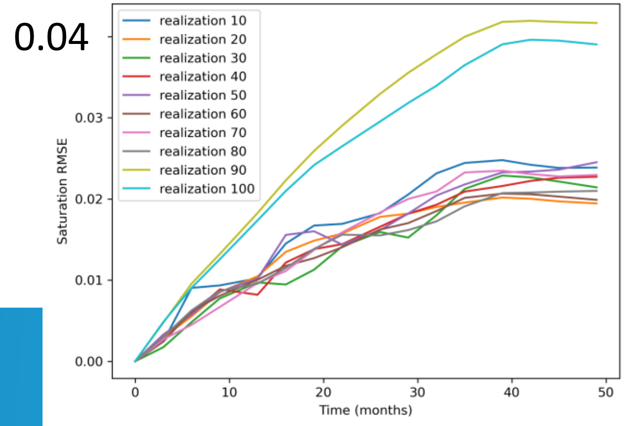


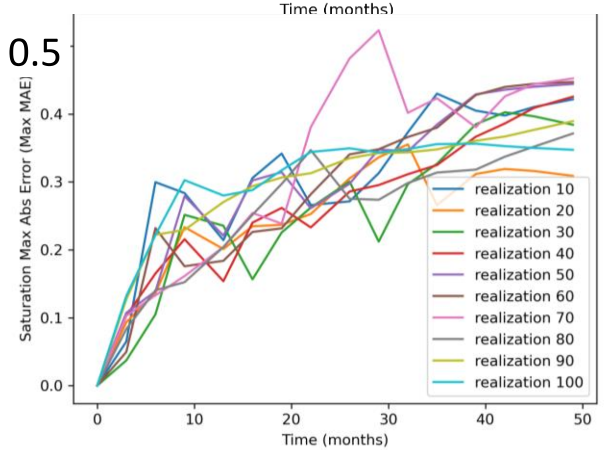




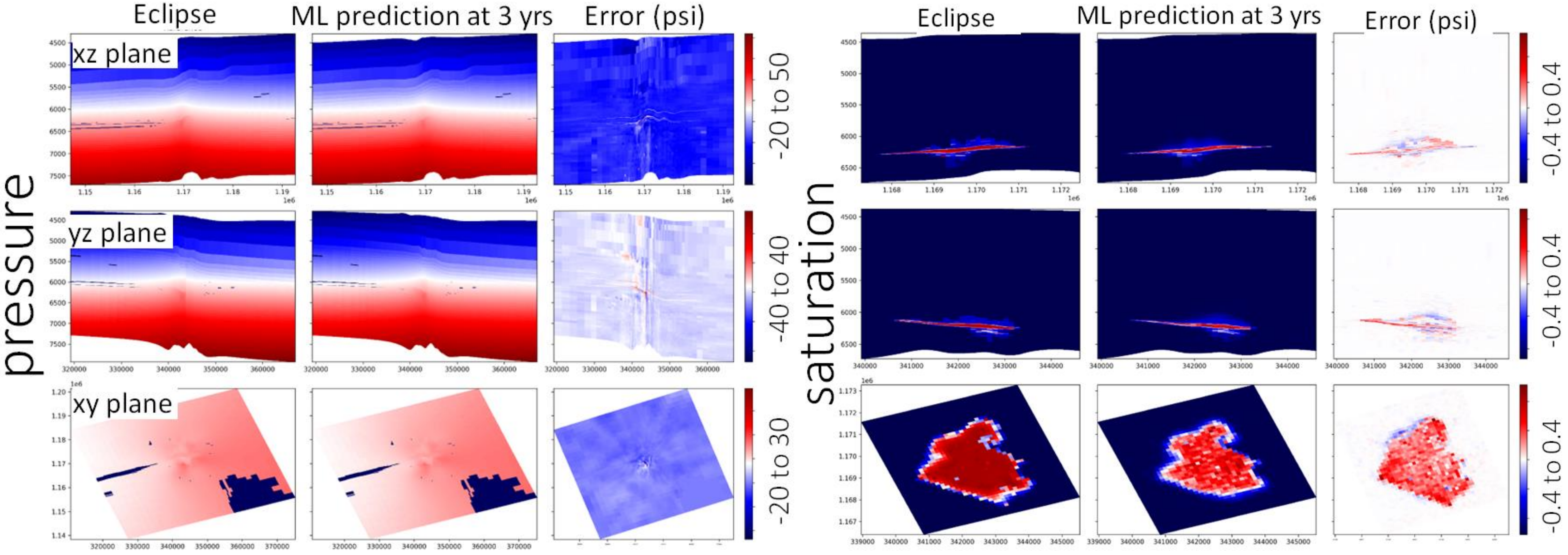




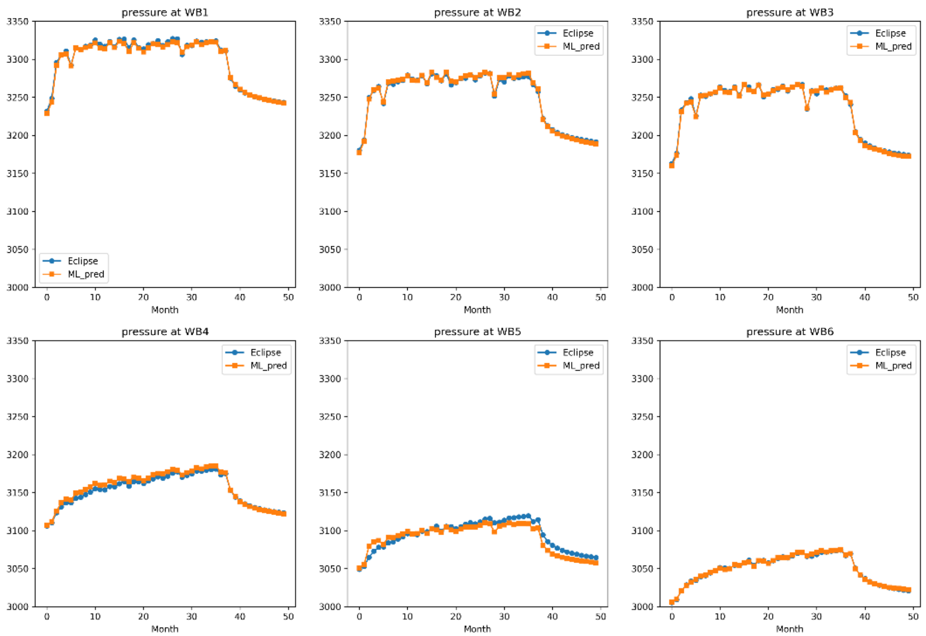


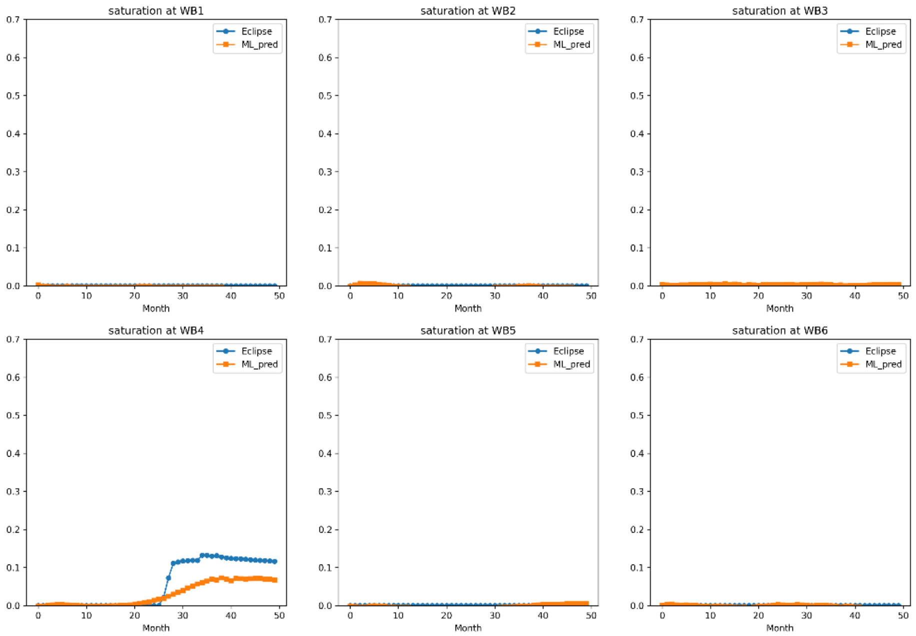


Snapshots of Pressure and Saturation

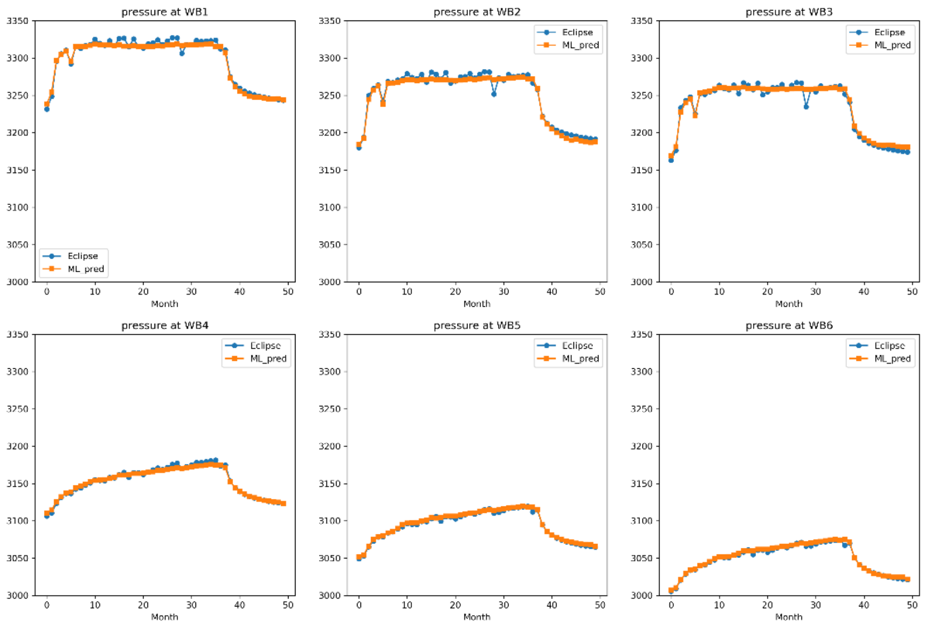


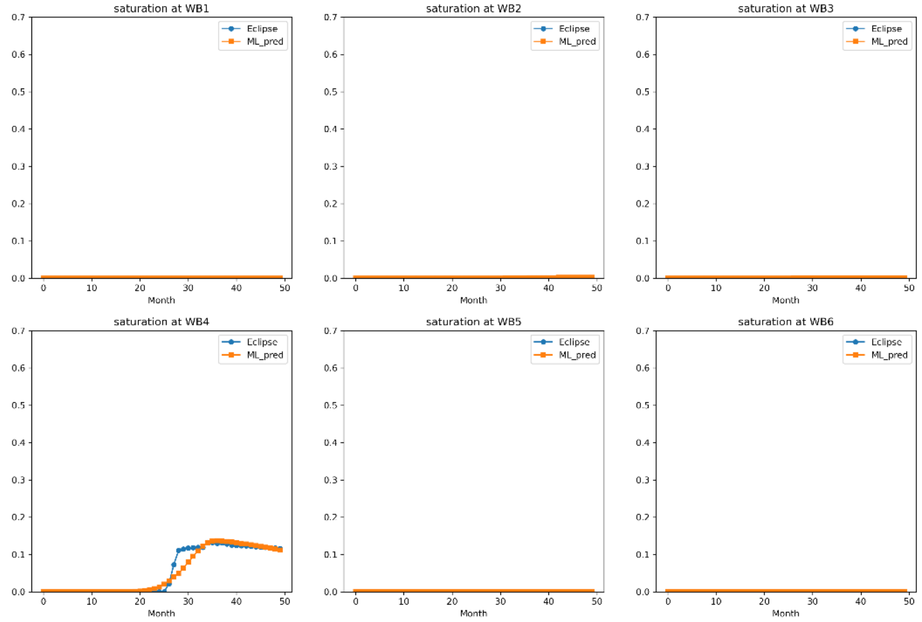
Pressure (psi) & Saturation (-) at six different depths in monitoring well (realization 10)





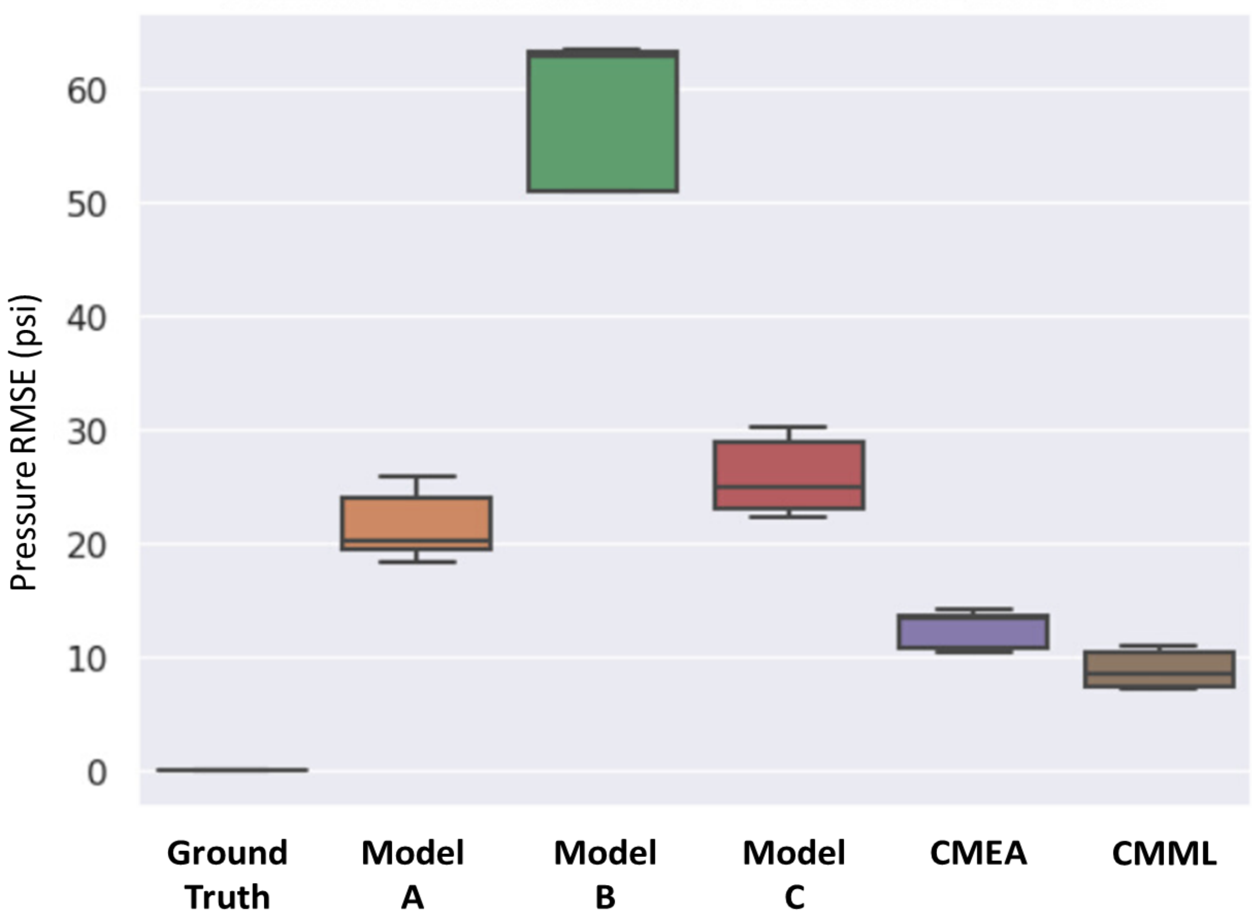
SNL





UT-BEG

ML Committee Machine



Part 3: Accelerate history matching through ML and transfer learning

Acknowledgements

Funding:

- This work was funded by the SMART Initiative through the DOE Office of Fossil Energy and Carbon Management and the Bipartisan Infrastructure Law.

Auspices:

- Portions of this work were performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07-NA27344.

- Portions of this work were performed under the auspices of the U.S. Department of Energy by Pacific Northwest National Laboratory under contract DE-AC05-76RL01830.

- Portions of this work were performed under the auspices of the U.S. Department of Energy by Sandia National Laboratories under contract DE-NA0003525.

- Portions of this research were executed through the NETL Research and Innovation Center's Carbon Storage Program

SMART-CS Initiative

Science-informed Machine Learning to Accelerate Real Time (SMART) Decisions in Subsurface Applications

Task 4

FECM Project Review Meeting
August 28 – September 1, 2023

U.S. DEPARTMENT OF ENERGY

Fault imaging through event detection and source location estimation

- Integrated ML approaches of event detection and source location estimation
- Data pre-processing of raw continuous microseismic data & event detection
- Data augmentation using WGAN (Wasserstein Generative Adversarial Network)
- PhaseNet used to downselect generated event data with high quality
- CNN model with multi-modal input for source location estimation of events

Data processing of raw waveform continuous data → **Event detection & arrival time: CNN & U-Net models** → **Synthetic data generation: SeismoML (WGAN)** → **Multi modal CNN for source locations of newly detected events** → **Event clustering & construct faults**

- 10s raw continuous waveform data (e.g., 4 to 3 channels)

- Train a CNN model for event detection with 3 channel & energy feature as input and retrain PhaseNet for arrival time

- WGAN model Input: source locations and distance of ~400 events from catalog and output of waveform
- Apply trained seismoML model to generate synthetic waveforms of each channel (H1,H2,V)
- Screen generated waveform data by phase arrival times (PhaseNet)

- Construct synthetic waveform data over a range of source locations & distance
- Train a multi modal CNN with spectrograms and P&S arrival (binary) of each channel data

- Event clustering using NMF-HMM model to construct planar faults