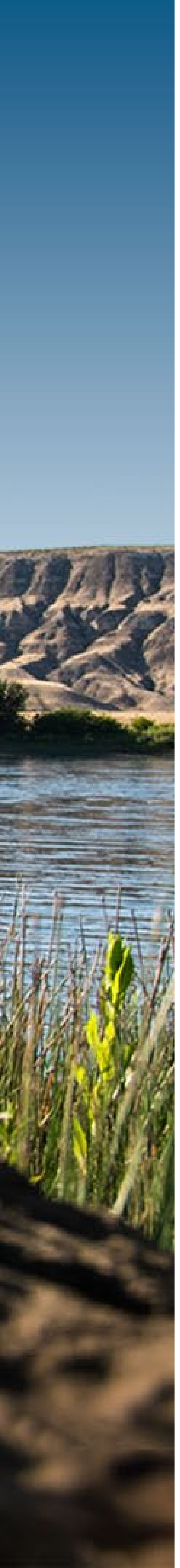
Simulator-Trained AI for Creating Subsurface Digital Twins using Time-Lapse Electrical Resistivity Tomography Data

Tim Johnson, Jose Hernandez-Mejia, Glenn Hammond, Piyoosh Jaysaval, Rob Mackley



Pacific Northwest
NATIONAL LABORATORY

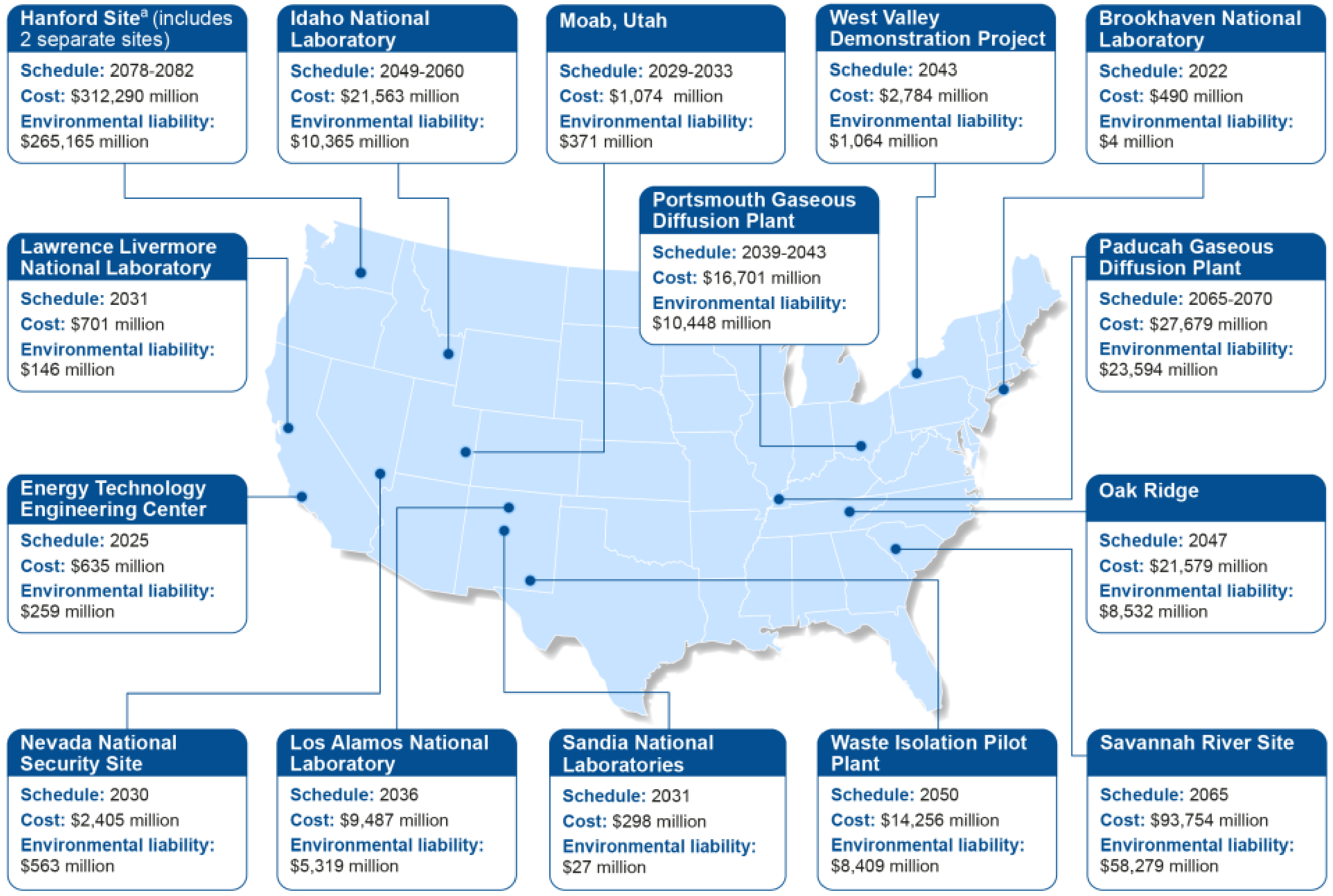
Motivation

What if we could significantly improve our capability to predict subsurface behavior?

- Reduced cost to taxpayers
- Reduced risk to human health and the environment
- Opportunity for scientific discovery

Note: We make predictions using numerical simulators.

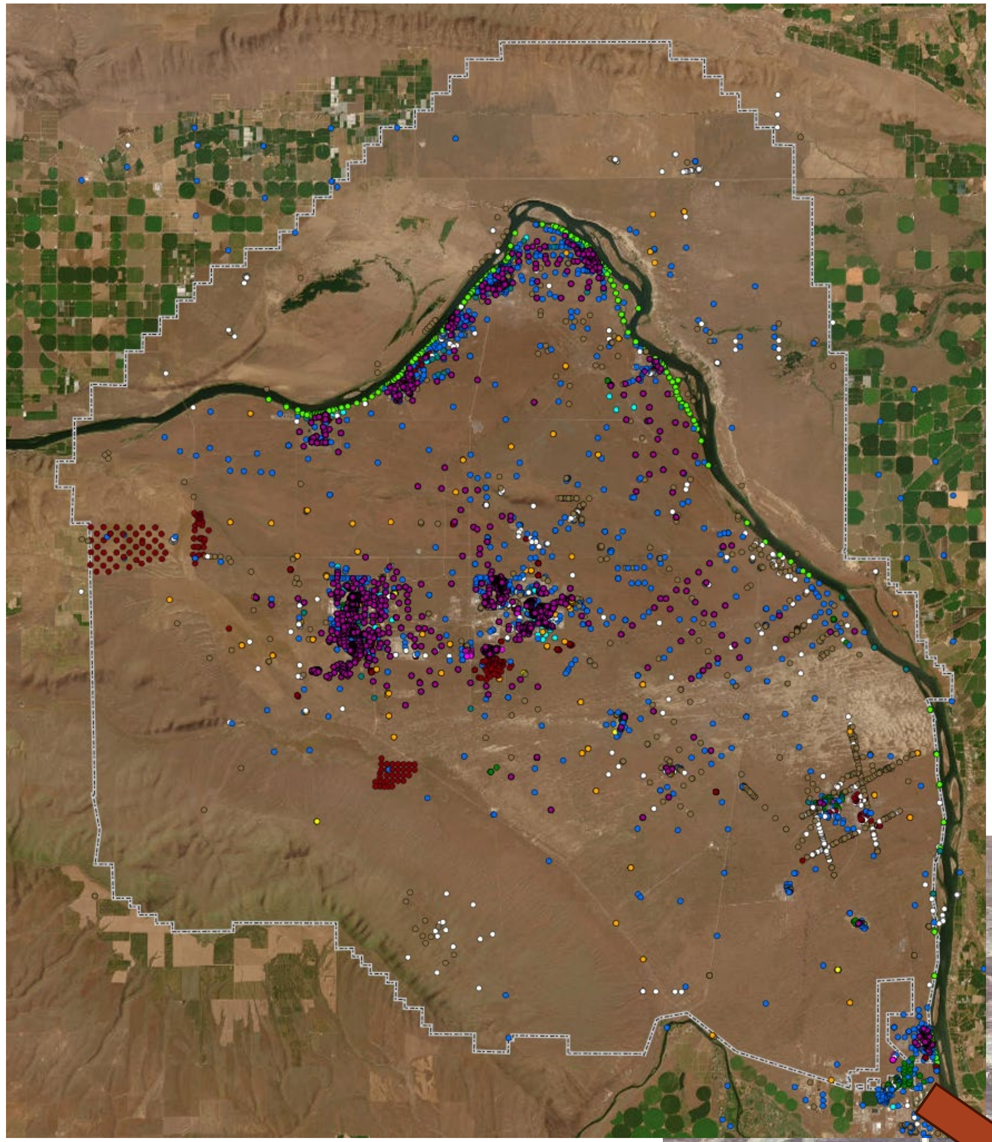
$525 Billion to Complete EM Complex Cleanup
$392 Billion in Environmental Liability
(GAO-22-104662)



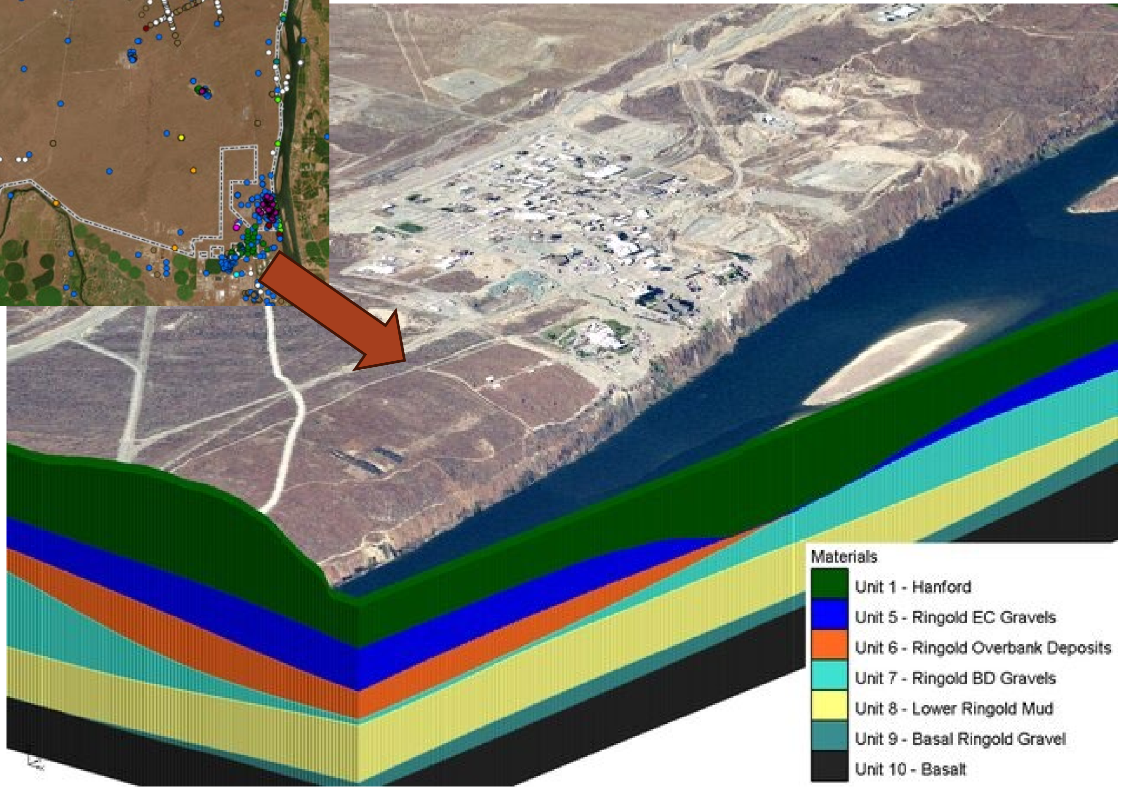
Sources: GAO analysis of Department of Energy information; Map Resources (map). | GAO-22-104662

Challenges

1. Subsurface complexity and heterogeneity
   - Data scarcity
   - Disparate data
   - Assimilation of disparate data (e.g. joint inversion)
2. Uncertainty Quantification
   - How accurate are our predictions?

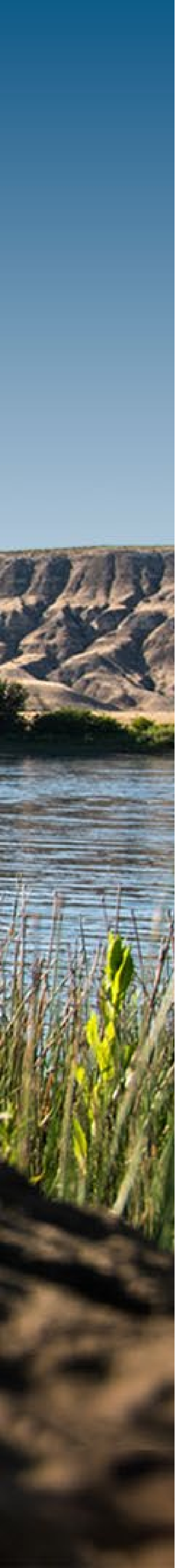


Hanford Site Borehole Locations



Hanford 300 Area Hydro-stratigraphic Units

(Hammond et al., 2007)



Pacific Northwest
NATIONAL LABORATORY

Text-to-Image Generation

Opportunities

Emerging generative AI methods are demonstrating ground-breaking capabilities in joint inversion.

- Assimilation of all available data = more accurate estimation of subsurface heterogeneity
- Generation of equally probably subsurface models, or ‘digital twins’ for uncertainty estimates.

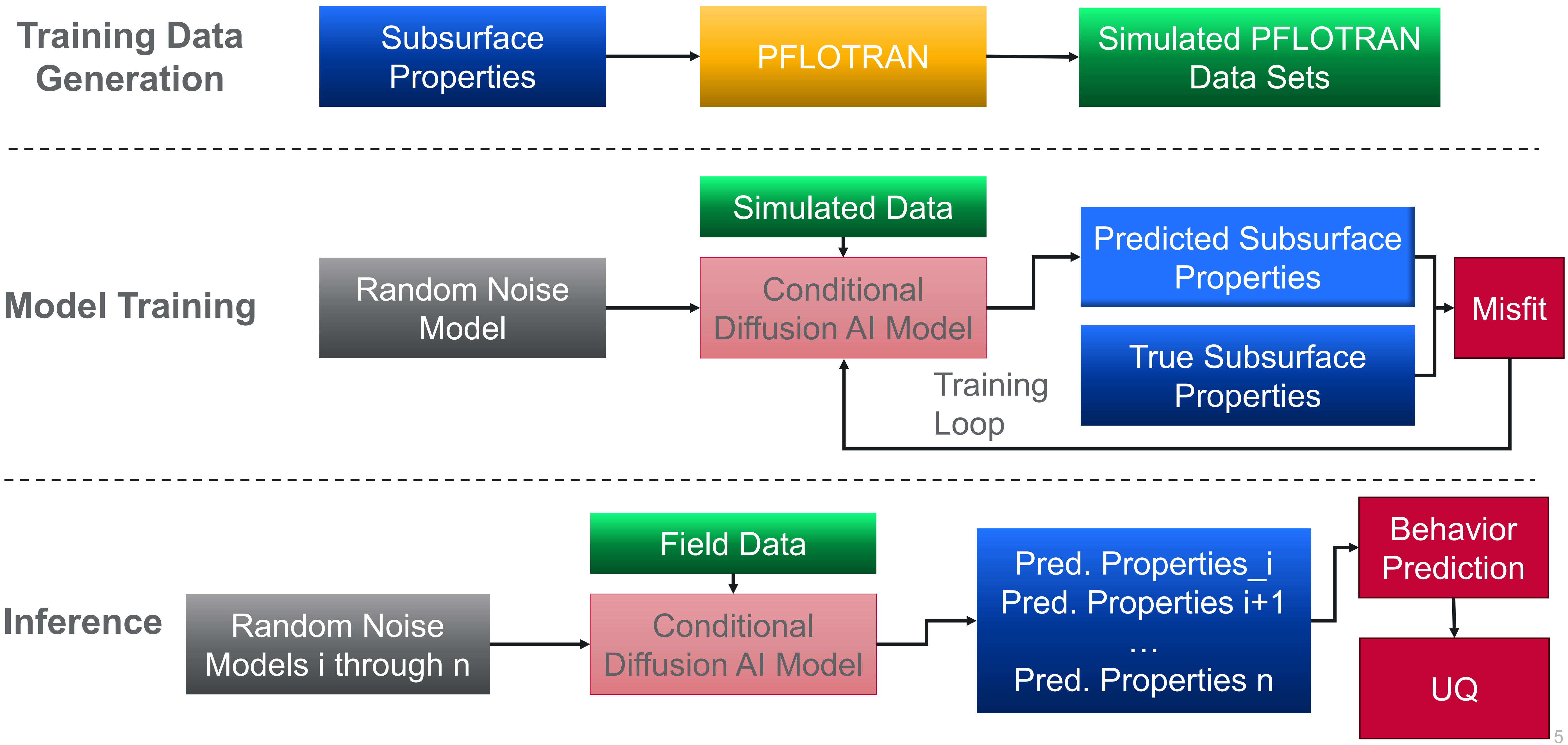
output

Conditional Diffusion Model

input

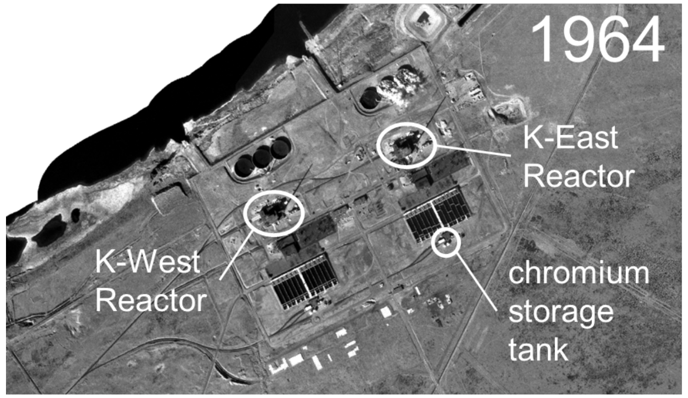
“Generate a realistic image of a bovine tumbleweed stampede with cowboys on horses in Richland Washington”

Conditional Diffusion Model Training and Application

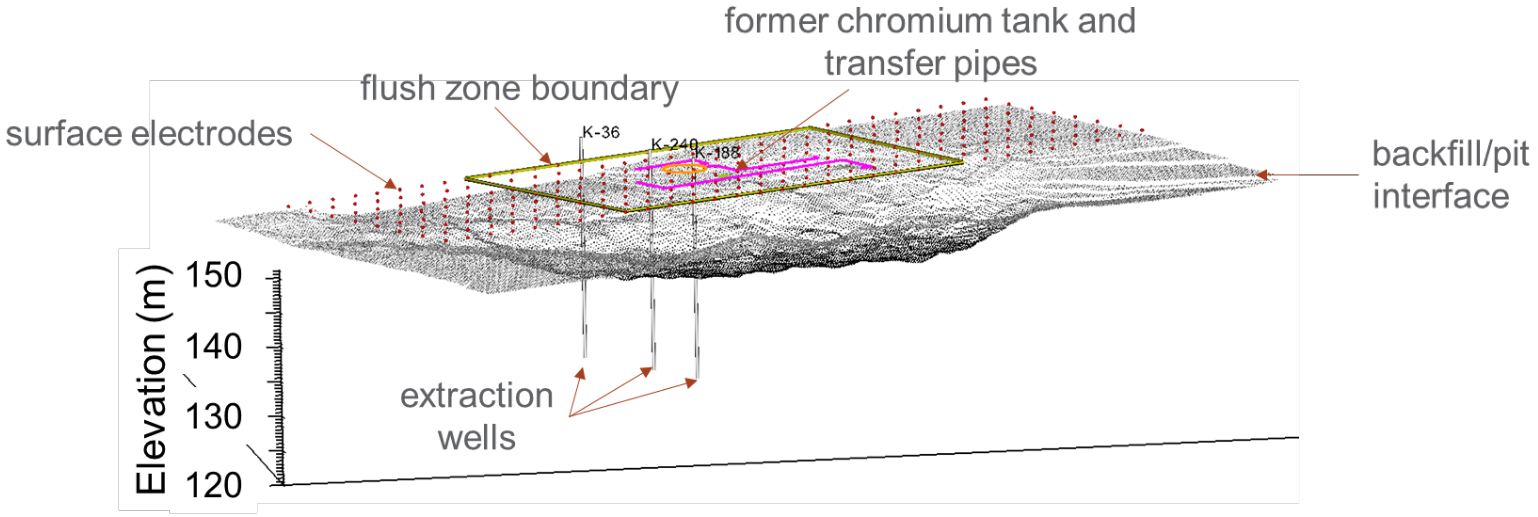


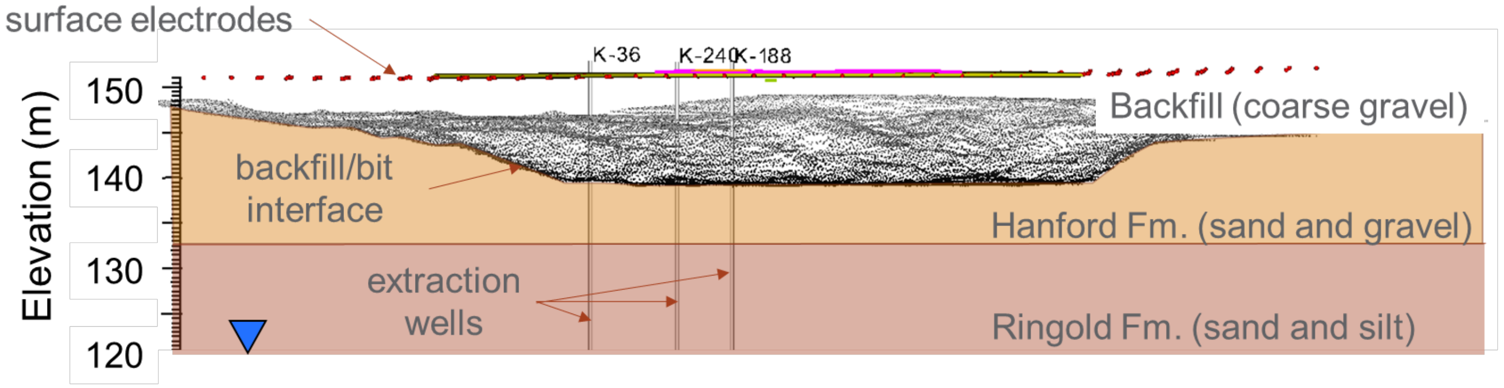
Test Case: Hanford 100-K Area Soil Flushing Test

Hanford 100-KE Reactor Area

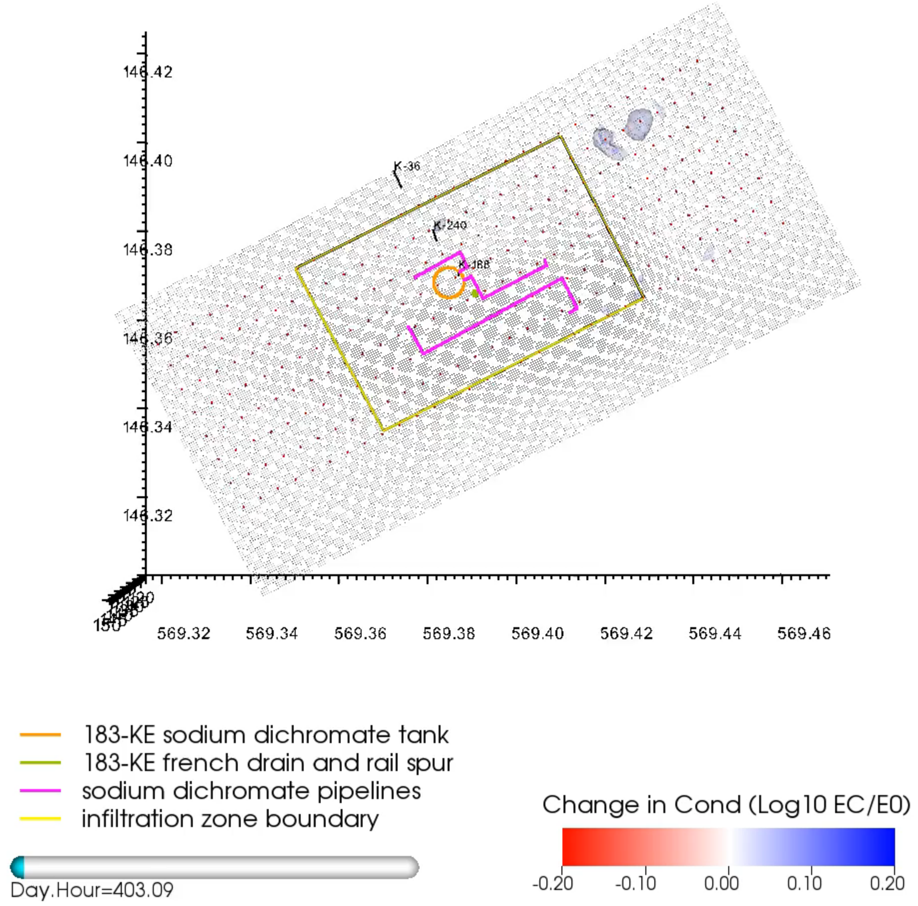


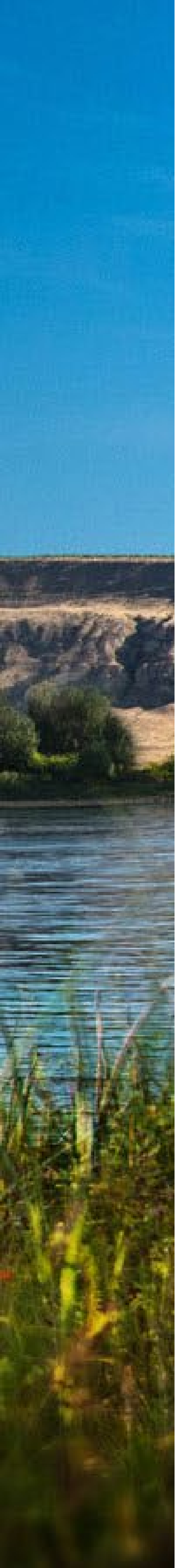
Soil Flushing Setup





Time-lapse ERT Images



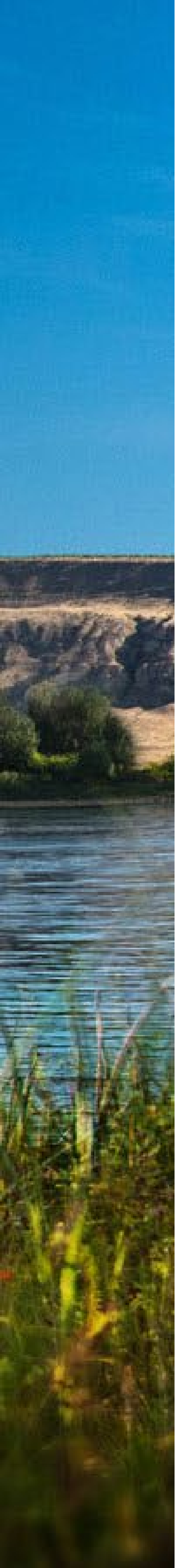


Pacific Northwest NATIONAL LABORATORY

Subsurface Properties

- Hydrogeologic, unsaturated flow, and petrophysical properties in each of the three units.
- Native pore water and flush water fluid conductivity
- Bounded by literature values
- 29 parameters total

Parameter	min value	max value
Hanford porosity	0.05	0.35
Hanford horizontal permeability [m^2]	2.00E-13	1.00E-07
Hanford vertical/horizontal permeability ratio	0.2	1
Hanford Archie's law cementation exponent	1.1	2.1
Hanford Archie's law saturation exponent	1.6	2.6
Hanford Archie's law tortuosity constant	0.8	1.2
Ringold unit porosity	0.2	0.5
Ringold horizontal permeability [m^2]	1.00E-12	1.00E-08
Ringold vertical/horizontal permeability ratio	0.2	1
Ringold Archie's law cementation exponent	1.1	2.1
Ringold Archie's law saturation exponent	1.6	2.6
Ringold Archie's law tortuosity constant	0.8	1.2
Pit porosity	0.05	0.35
Pit horizontal permeability [m^2]	2.00E-13	1.00E-07
Pit vertical/horizontal permeability ratio	0.2	1
Pit Archie's law cementation exponent	1.1	2.1
Pit Archie's law saturation exponent	1.6	2.6
Pit Archie's law tortuosity constant	0.8	1.2
Hanford and Pit VG-Alpha [1/m]	2.00E-05	0.009
Hanford and Pit VG-M	0.2	0.65
Hanford and Pit residual saturation	0.0055	0.24
Ringold VG-Alpha [1/m]	1.00E-05	0.008
Ringold VG-M	0.16	0.8
Ringold residual saturation	0.02	0.2
Handford surface electrical conductivity [S/m]	1.00E-05	0.01
Ringold surface electrical conductivity [S/m]	1.00E-05	0.01
Pit surface electrical conductivity [S/m]	1.00E-05	0.01
Native pore water conductivity [S/m]	0.005	0.1
Flush water conductivity [S/m]	0.005	0.1



Pacific Northwest
NATIONAL LABORATORY

Training Data Generation

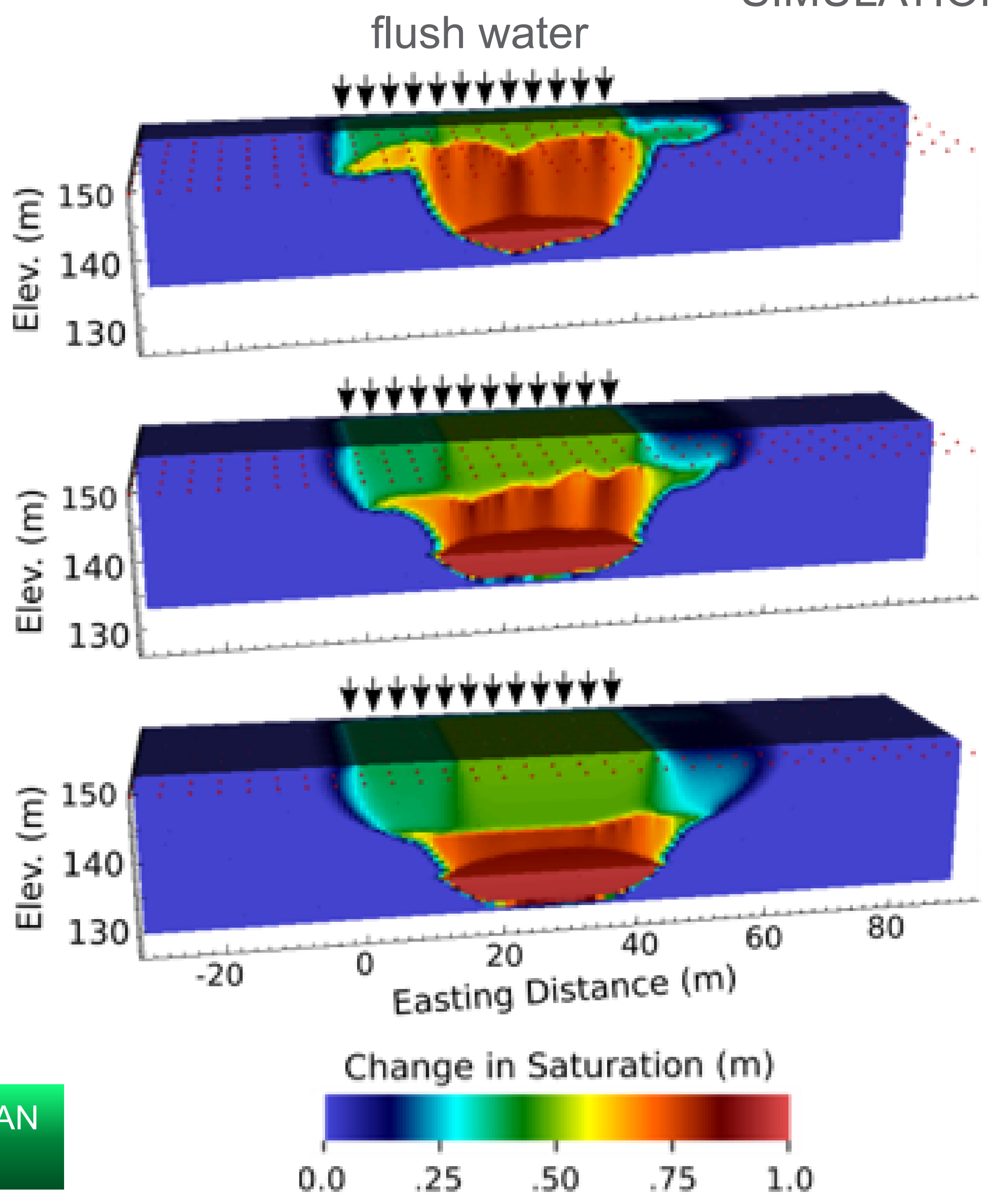
PFLOTRAN simulates field observations for many (2-3K) randomly chosen parameter sets (i.e. model inputs).

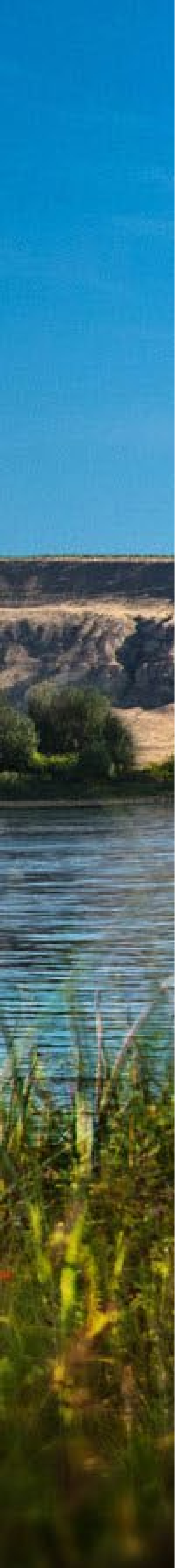
Simulations are executed on high-performance computing systems.

Computationally demanding, but tractable

Subsurface Properties → PFLOTRAN → Simulated PFLOTRAN Data Sets

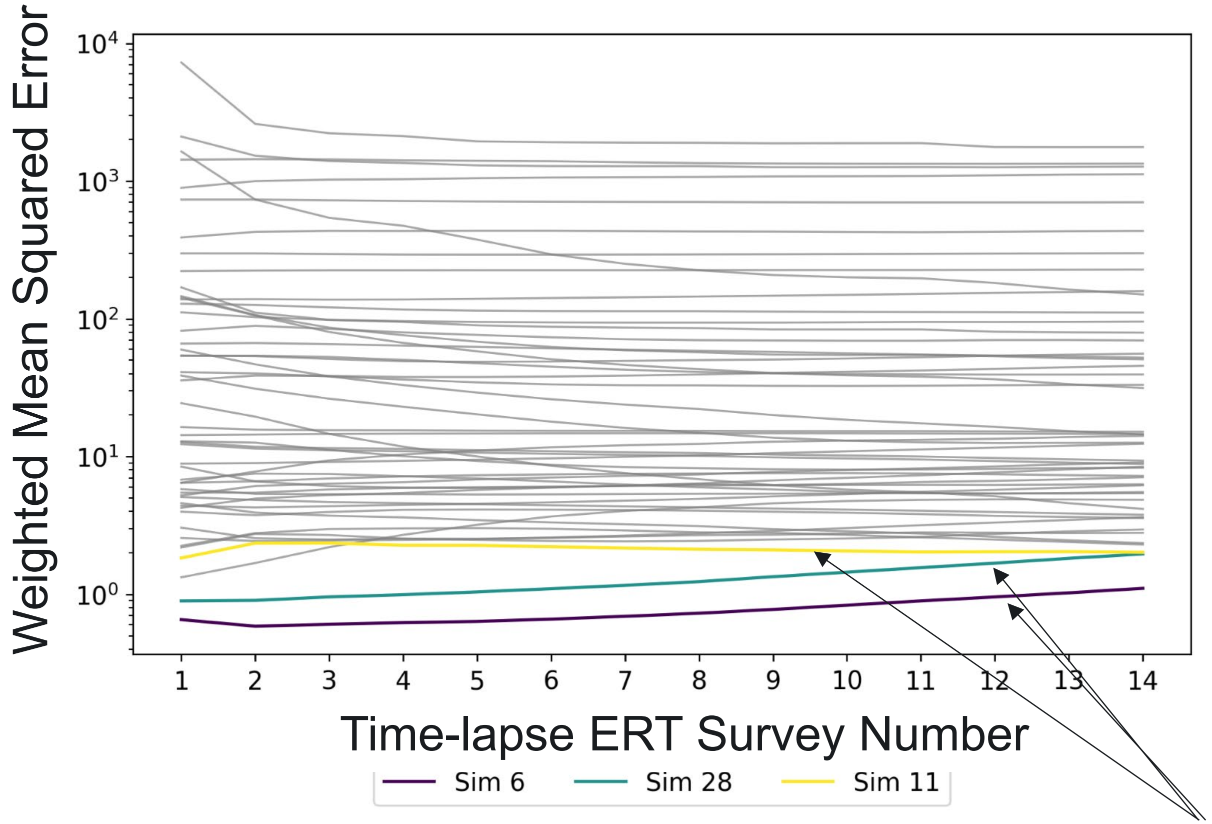
PFLOTRAN SIMULATION

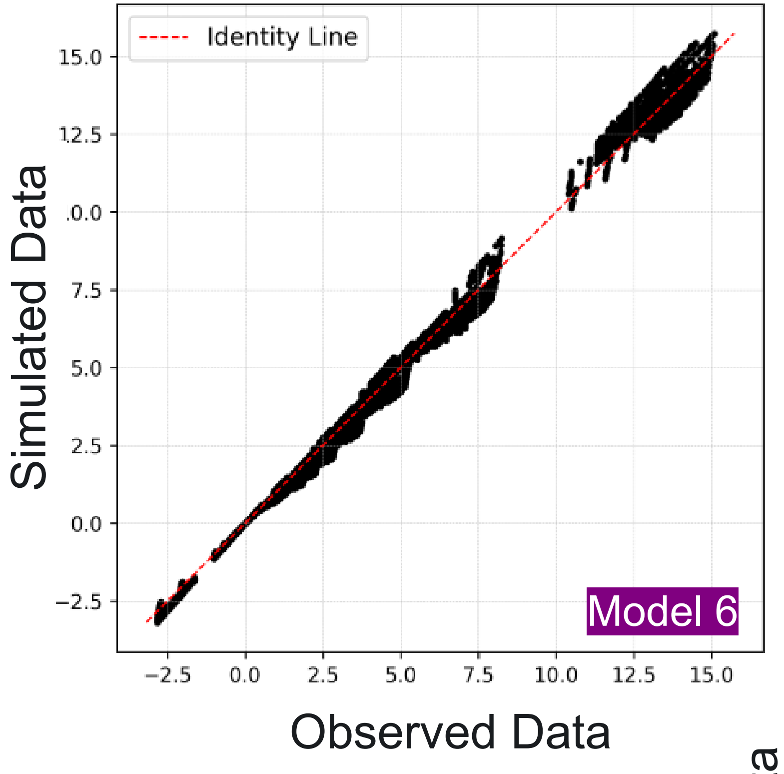


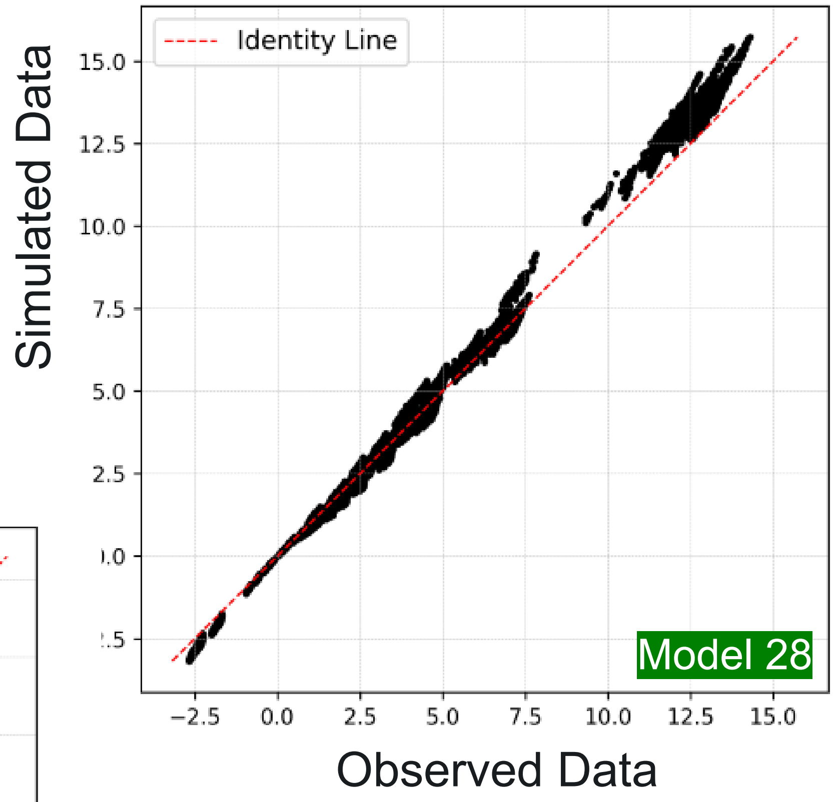


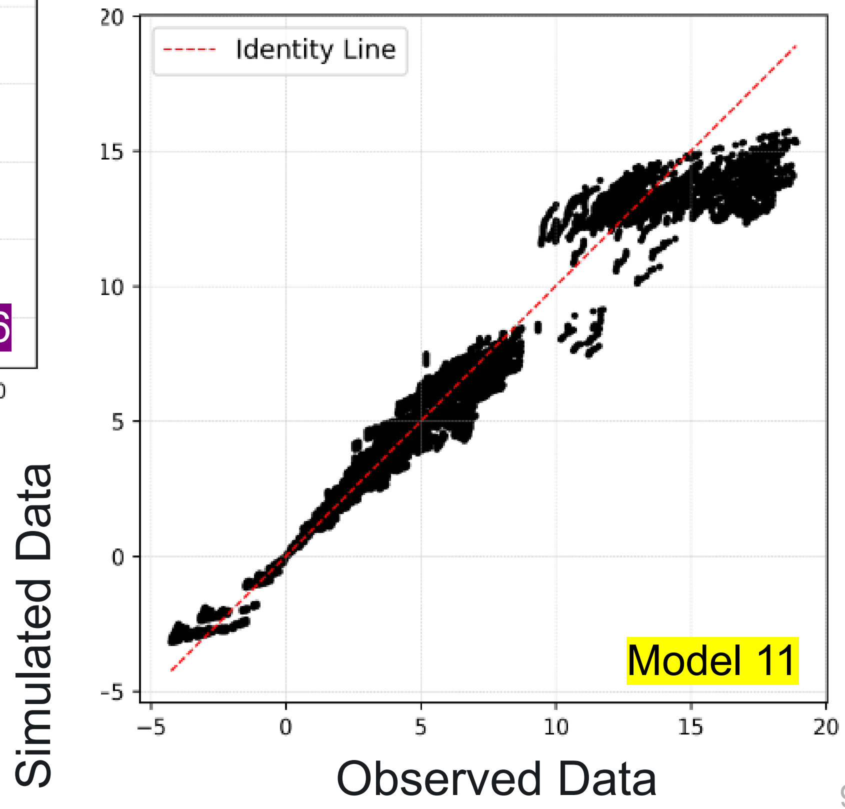
Condition Diffusion Model Performance

PFLOTRAN ERT Data Simulation Error vs. Time Lapse Survey Number for 50 AI Generated Models

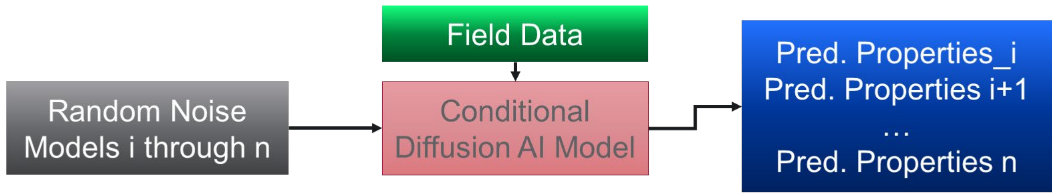


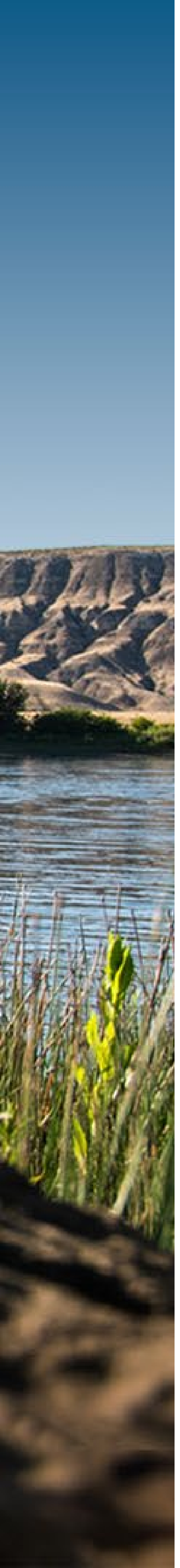






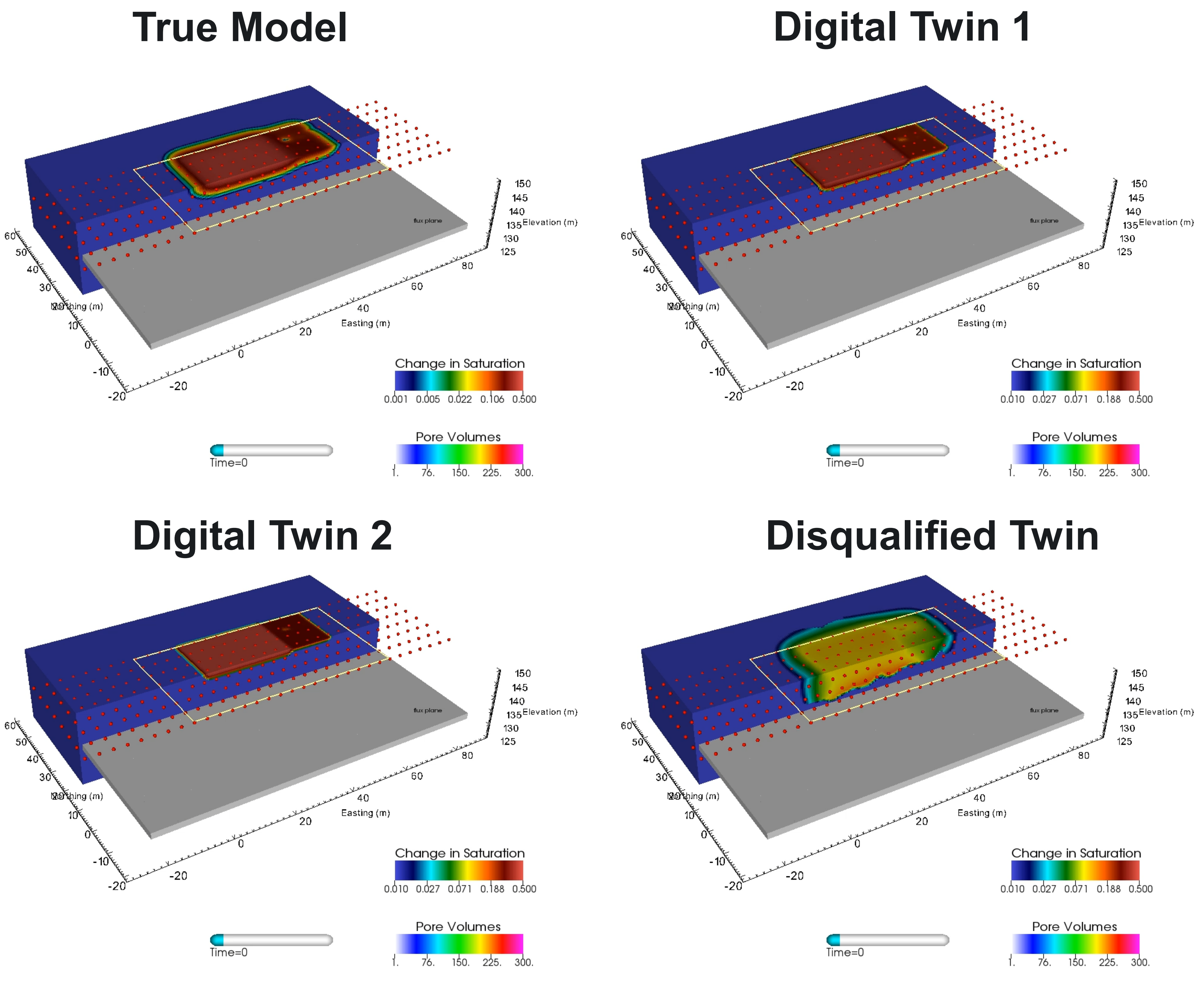
Three of the fifty generated models produced acceptable fits to the raw ERT data





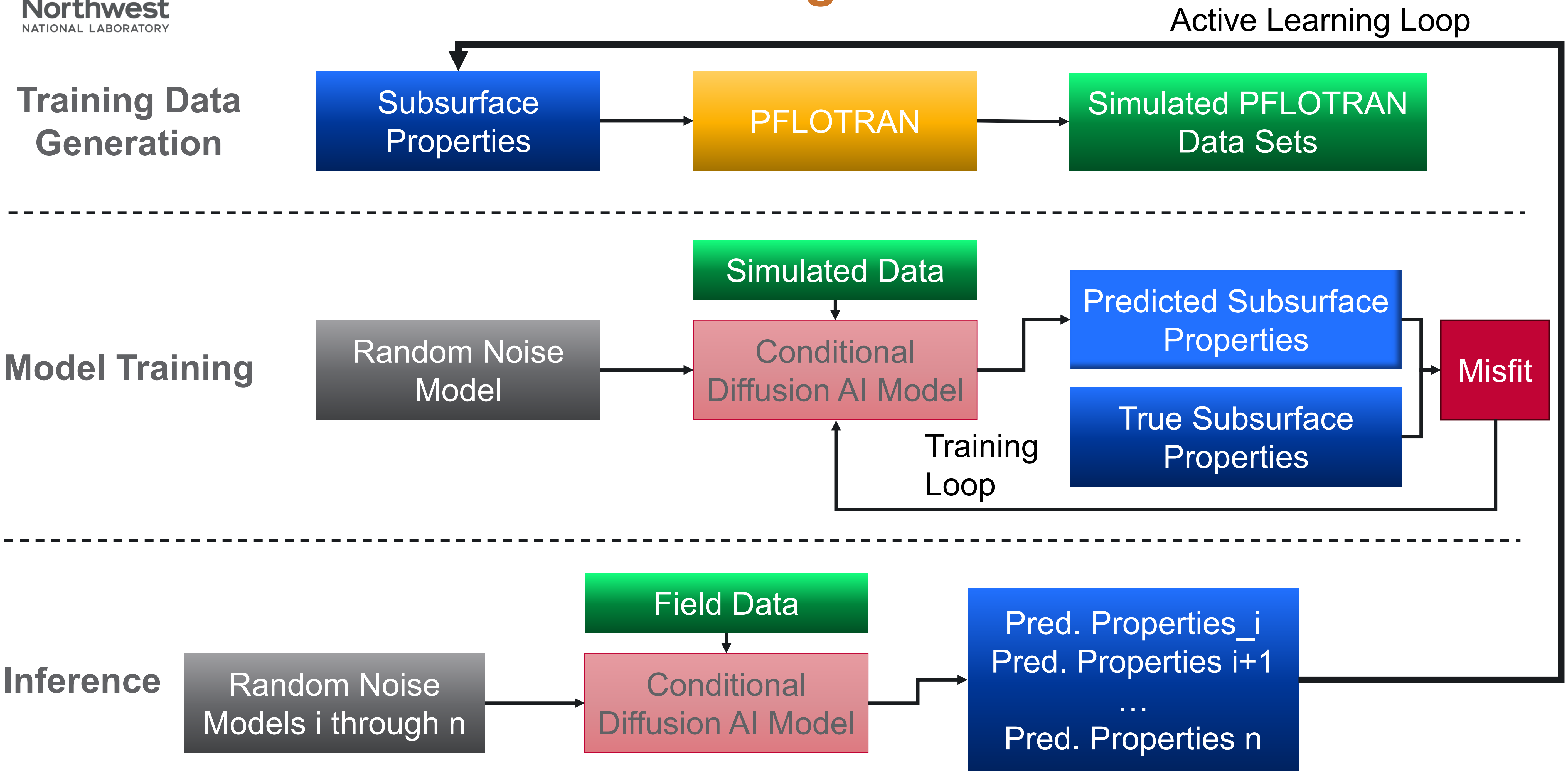
How well do the acceptable digital twins predict soil flushing?

- Twin 1 and Twin 2 honor the ERT data
- Disqualified twin does not
- Prediction performance based on # flush water pore volumes



animation

Reducing Training Data Requirements with Active Learning

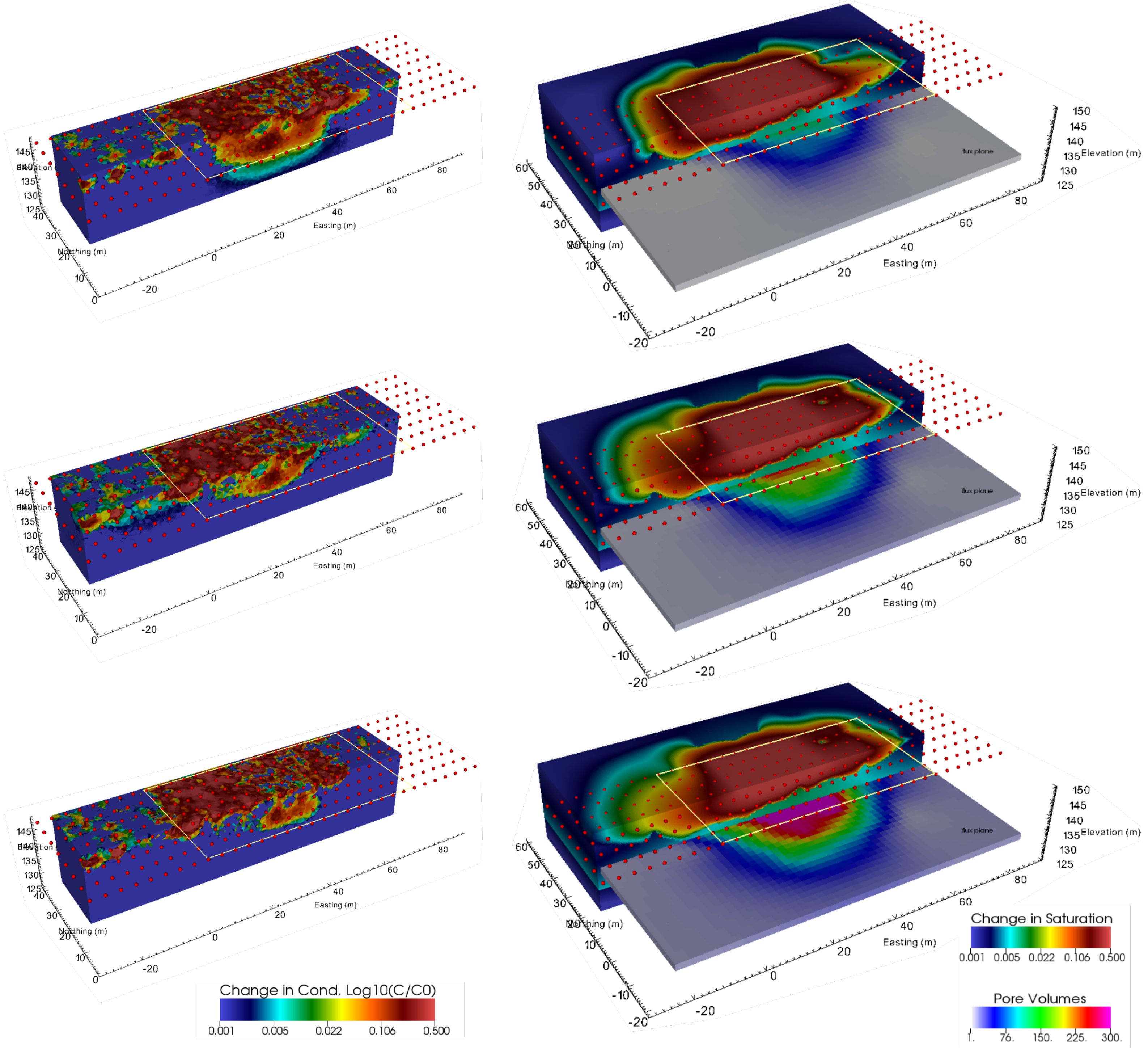


Summary

- AI/ML is enabling solutions to the multi-physics joint inversion problem (+ uncertainty estimation)
- Stands to significantly improve subsurface predictability and predictive uncertainty
- Work to be done …
  - training data generation
  - heterogeneity
- Outlook

ERT Analysis of Monitoring Data

AI Analysis of Monitoring Data



Pacific Northwest
NATIONAL LABORATORY

Thank you