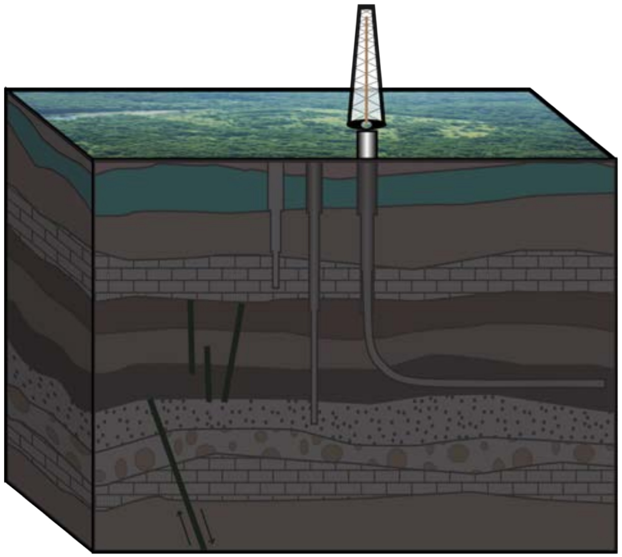
Potential Opportunities to Transform Utilization of Subsurface Resources via Machine Learning



George Guthrie

17 May 2018

Our ability to utilize the subsurface is limited by our lack of information, leading to uncertain decisions.

- **Virtual learning**
- **Signals from Noise**

LA-UR 18-24336

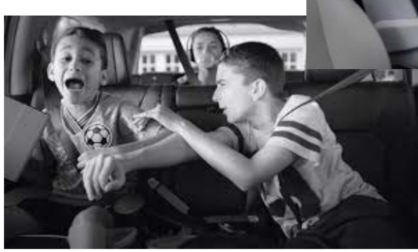
Consider machine learning & the evolution of driving safety

Autonomous Control

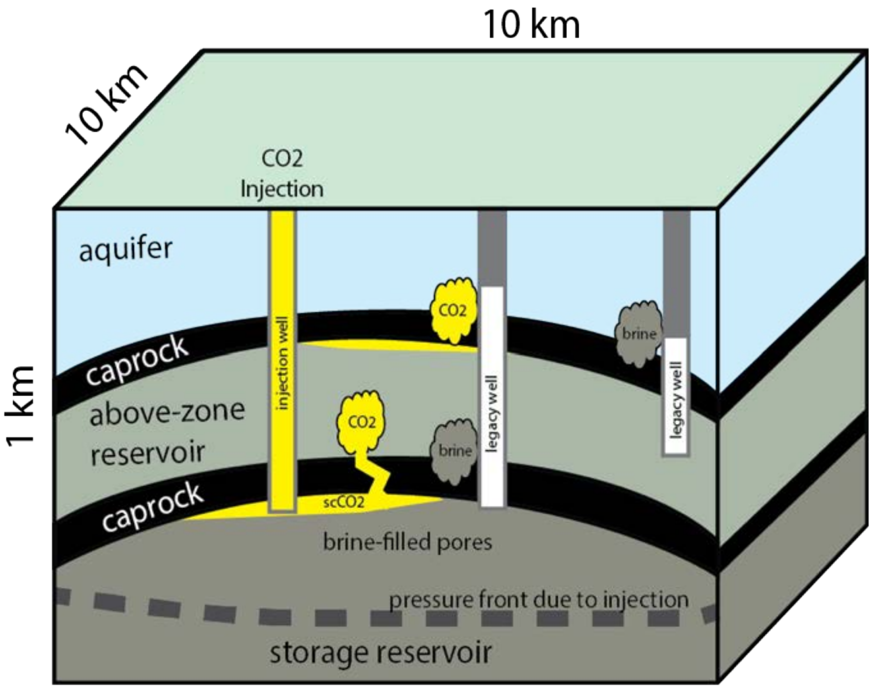
Real-time Data

Virtual Learning

Passive Systems

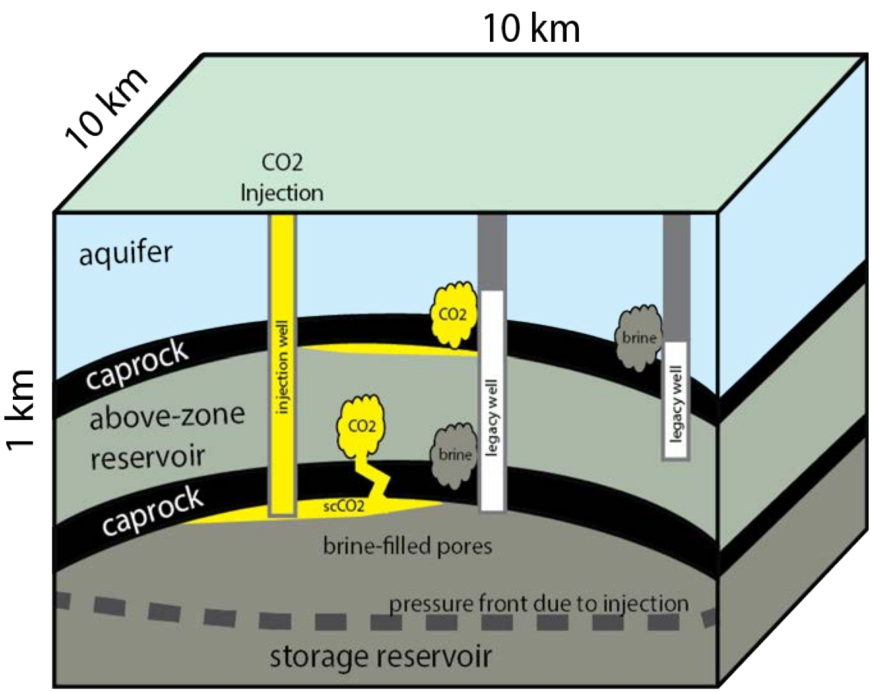


How do you design and test a monitoring system for a subsurface environment with limited real-world examples?



What leakage related signals occur, where/when do they occur, and can you detect/monitor them?

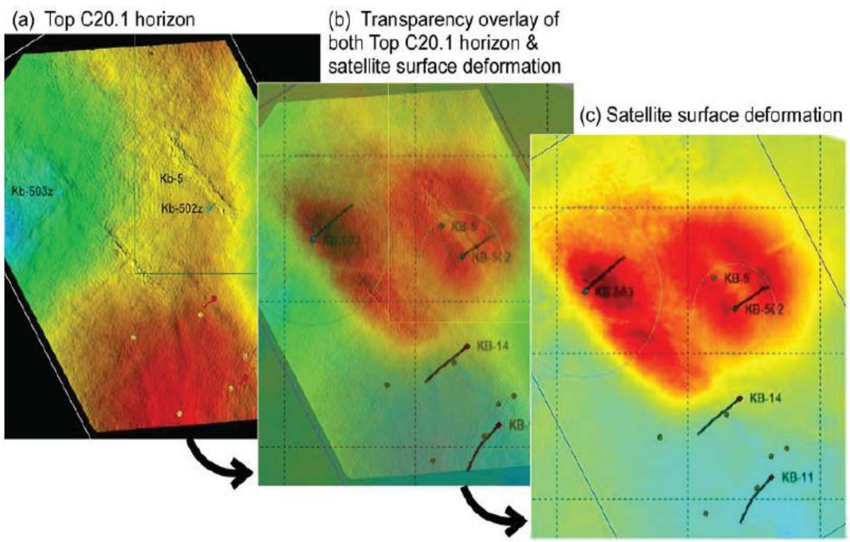
In the last decade, global research efforts have explored a plethora of approaches to monitor CO_2 storage sites…



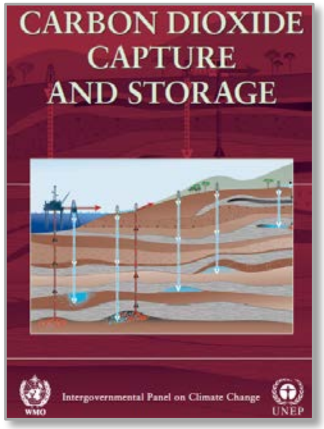
- InSAR monitoring and tilt-meters to monitor surface displacement
- Atmospheric monitoring using eddy-flux towers, LIDAR, perfluorocarbon tracers, …
- Groundwater monitoring using grab samples, electromagnetics, …
- Above-zone pressure monitoring with detailed physics-based simulation
- Data analytics and airborne magnetics to detect legacy wells
- Advanced seismic imaging to detect fractures in caprock
- 4D Seismic to quantify CO_2 in reservoir
- Borehole breakouts and imaging to detect fractures intersected by injection well
- …

…and it has become clear that conventional monitoring strategies are inadequate due to effectiveness & cost.

Seismic and InSAR Imaging at In Salah after 5 years of Injection



Ringrose et al. (2013) *Energy Procedia* ***37*:**6226–6236
Bond et al. (2013) *Geophys. Res. Let.* **40:**1284–1289

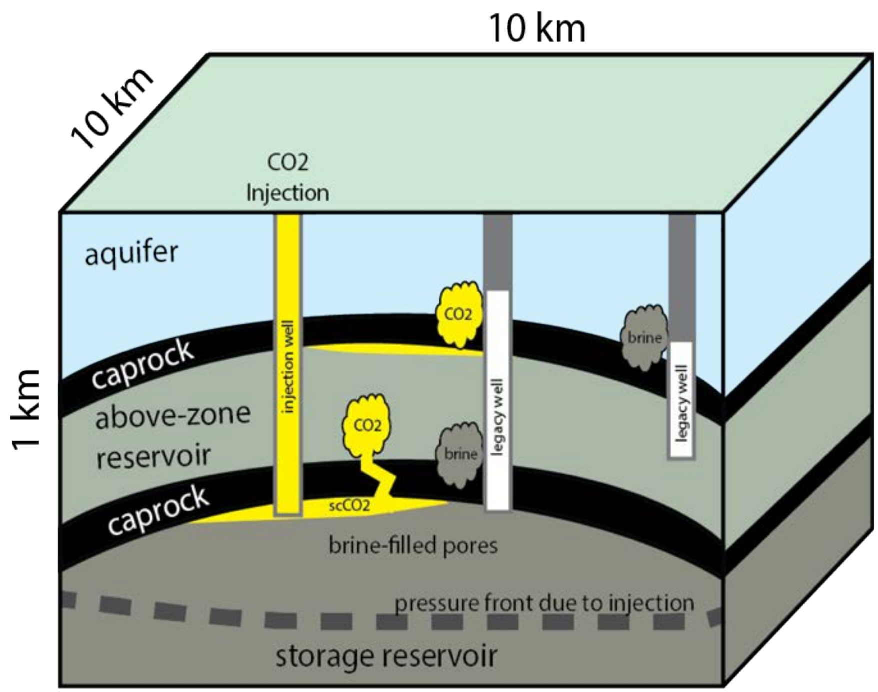


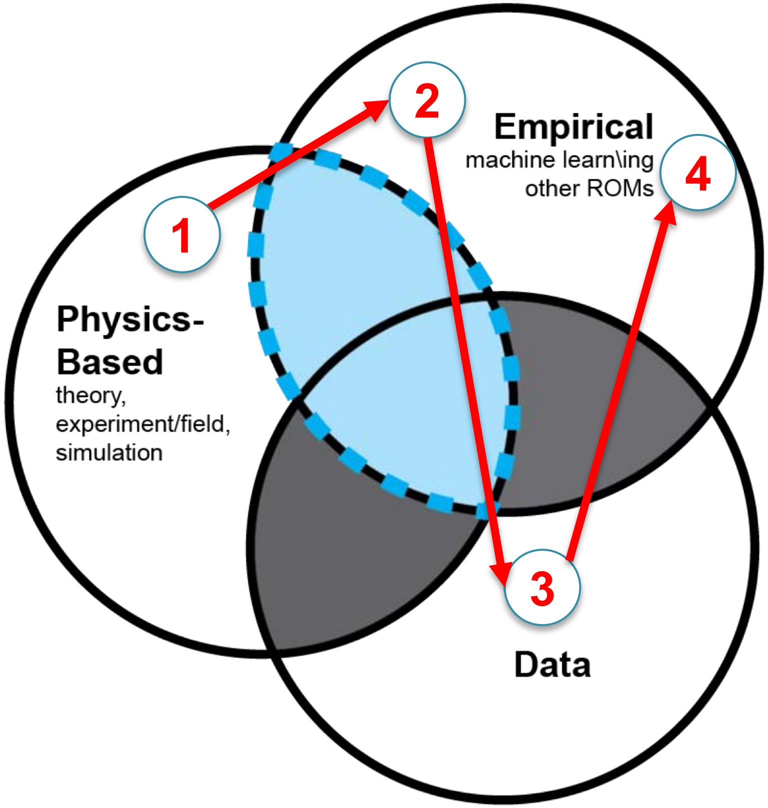
IPCC (2005)

NETL (2017)

- **IPCC (2005) estimated costs for onshore monitoring to be 0.1–0.3 USD/tCO2 (~1–2% of storage costs)**
- **NETL (2017) cost model estimates that to meet EPA Class VI regulations using <u>current technology</u> may be >10× higher (~50% of storage costs) due to…**
  - Large area (e.g., 10^2 km^2)
  - Long time frame (50 years)
  - Large battery of monitoring tools relying on conventional data analysis

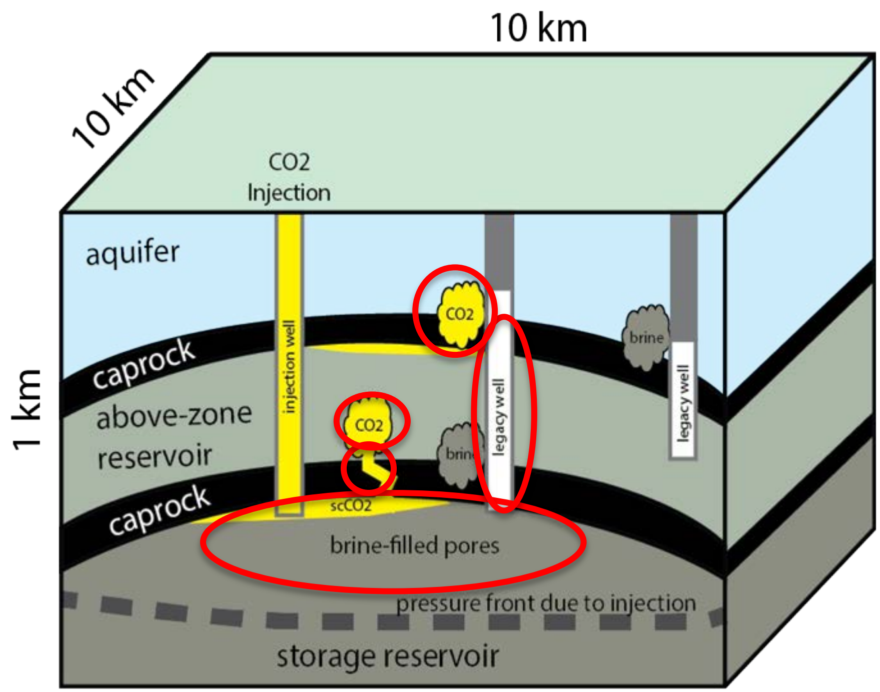
If you can predict the behavior of a system accurately, then you can create a virtual environment for learning.

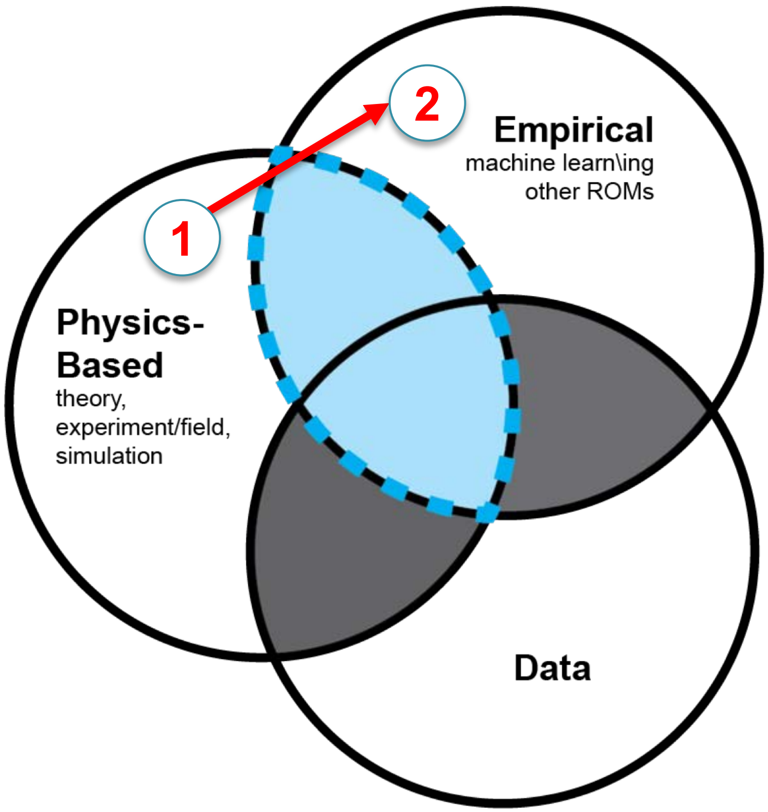




1. Develop predictive understanding of system
2. Use simulated behavior to train empirical models (ROMs; ML)
3. Use empirical models to create numerous simulated datasets (virtual environment)
4. Use machine learning methods to extract knowledge from virtual environment (new signatures and empirical relationships, leading to autonomous monitoring system)

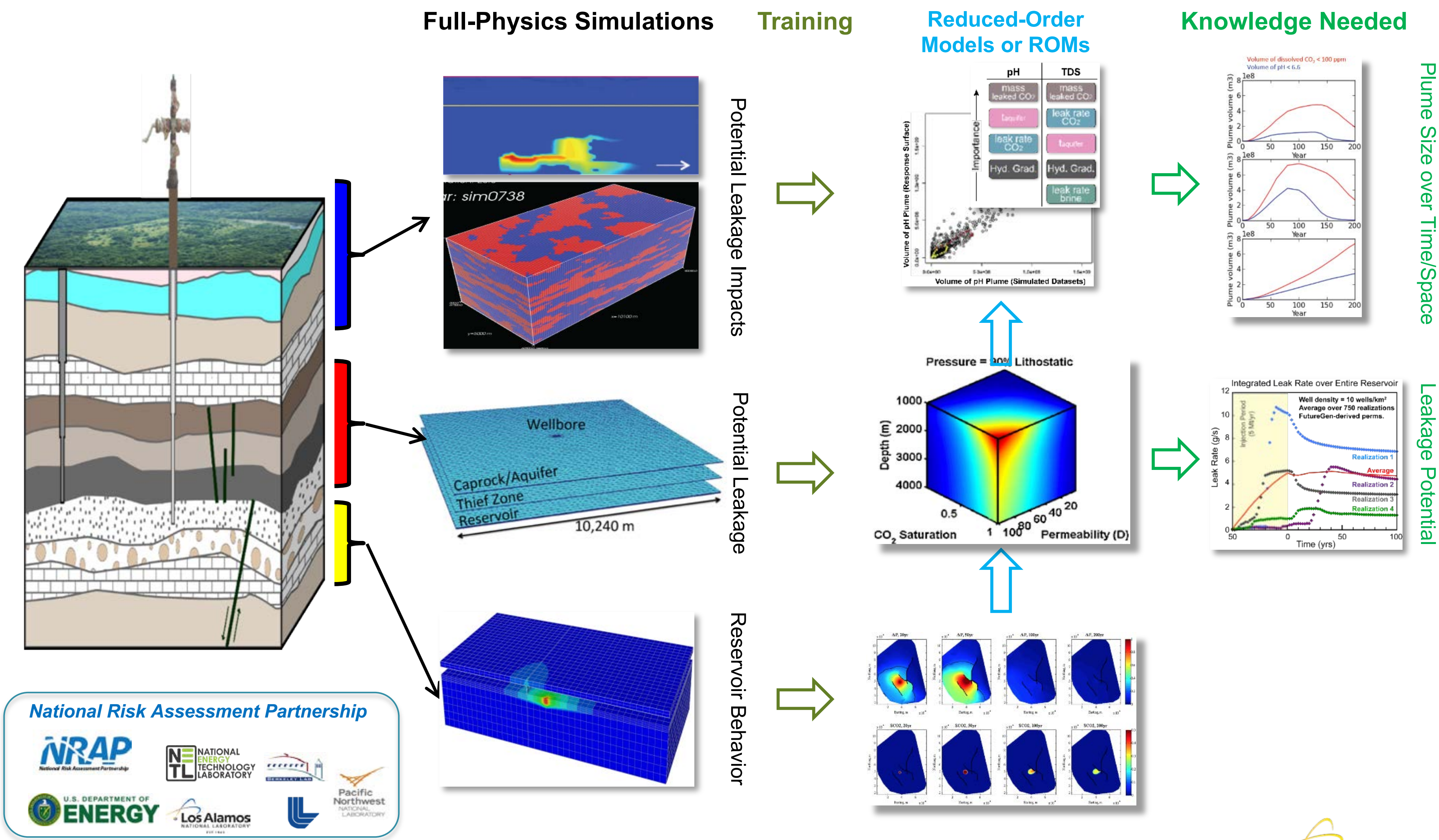
Predicting the behavior of a storage site cannot be done using a single high-fidelity simulation.



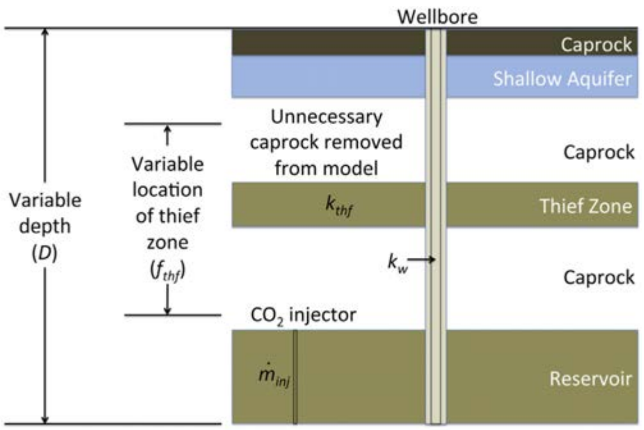


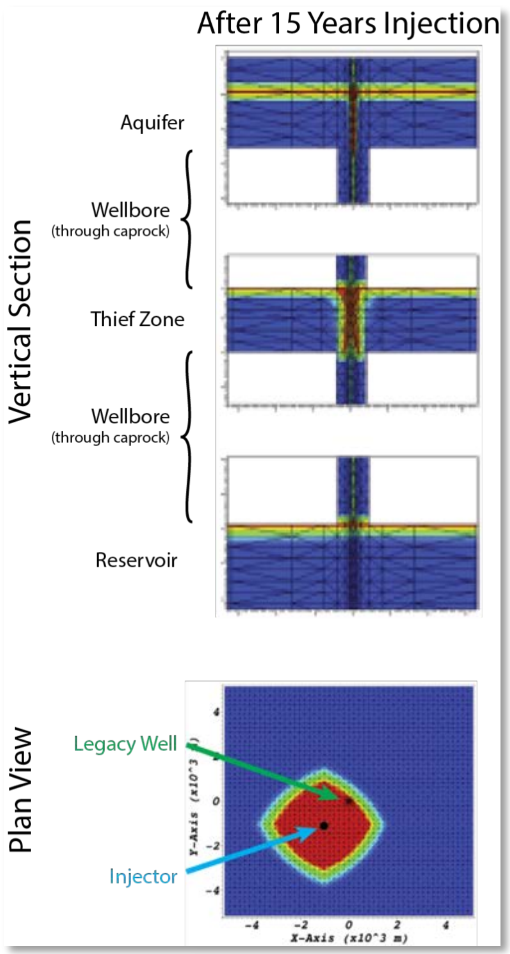
Need: Predict the evolution of the storage system from reservoir to receptors.

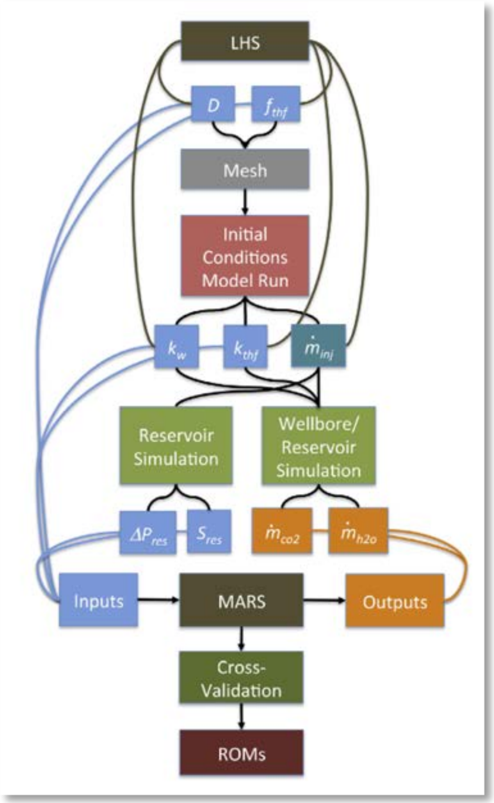
NRAP's approach has been to describe system behavior based on linked components.



NRAP developed wellbore-leakage ROMs based on full-physics simulations of entire system (reservoir to aquifer).







- Full-physics simulations generate virtual-reality data for multiple combinations of independent variables
  - permeabilities of wellbore, reservoir, thief zone, aquifer; ΔP in reservoir; depth; saturation; etc.
- Statistical methods identify key independent variables

Harp, D, Pawar, RJ, Carey, JW, Gable, CW (2016) Reduced order models of transient CO_2 and brine leakage along abandoned wellbores from geologic carbon sequestration reservoirs. Int. J. Greenhouse Gas Control 45 (2016) 150–162.

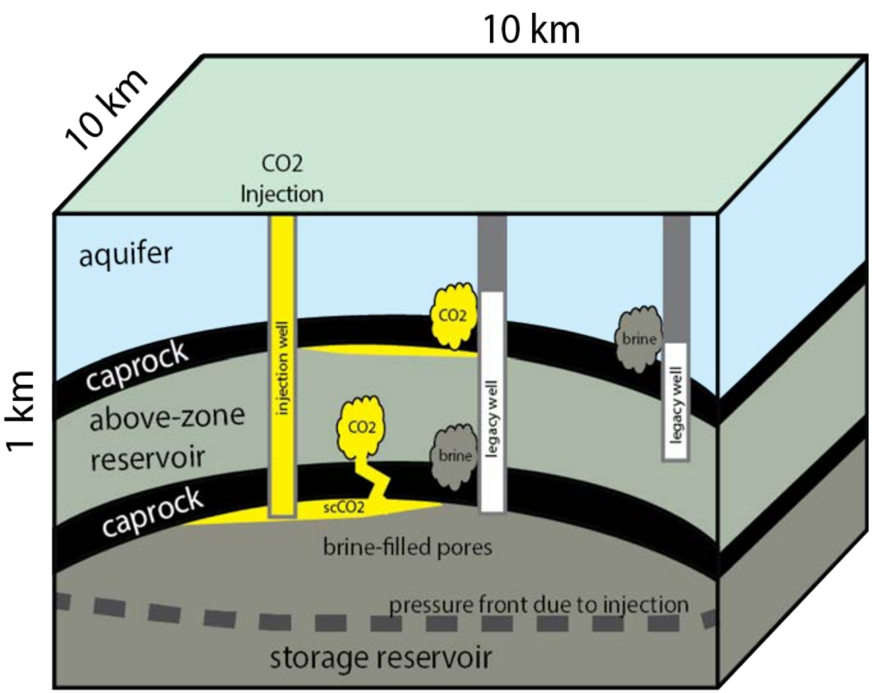
We are exploring various approaches to develop ROMs and are evaluating the range of ROM complexity needed.

- NRAP (FE-20)
- Geothermal (GTO)
- Oil & Gas FE-30; LDRD)
- Oil & Gas (ARPA-E)
- CO_2/Oil & Gas (FE-20)

complexity →

		CO_2 Storage Reservoir	Wellbores	Fractured Seal	Groundwater Aquifer	Geothermal Reservoir	Unconventional Reservoirs	Methane Leakage	ROZ CO_2 EOR/Storage
	Lookup Table (NRAP Team)	X	X	X					
	Analytical Model (NRAP Team)	X		X	X				
Response surfaces	LLNL's PSUADE (multiple RS types) (NRAP Team)	X			X				
Response surfaces	MARS (regression+cubic spline) (LANL)		X						X
Response surfaces	Polynomial Chaos Expansion (NETL/CMU)	X							
Response surfaces	Gaussian Regression (LBNL)	X							
Response surfaces	Surrogate Reservoir Model (NETL/WVU; LBNL)	X							
Response surfaces	Polynomial Non-linear Regression (LANL)					X			
Response surfaces	Artificial Neural Networks (LANL)							X	
Response surfaces	Support Vector Machine Learning (LANL)								X
	Graph theory (LANL)						X		

We now have the ability to predict many of the leakage-related behaviors of a complex, subsurface system.

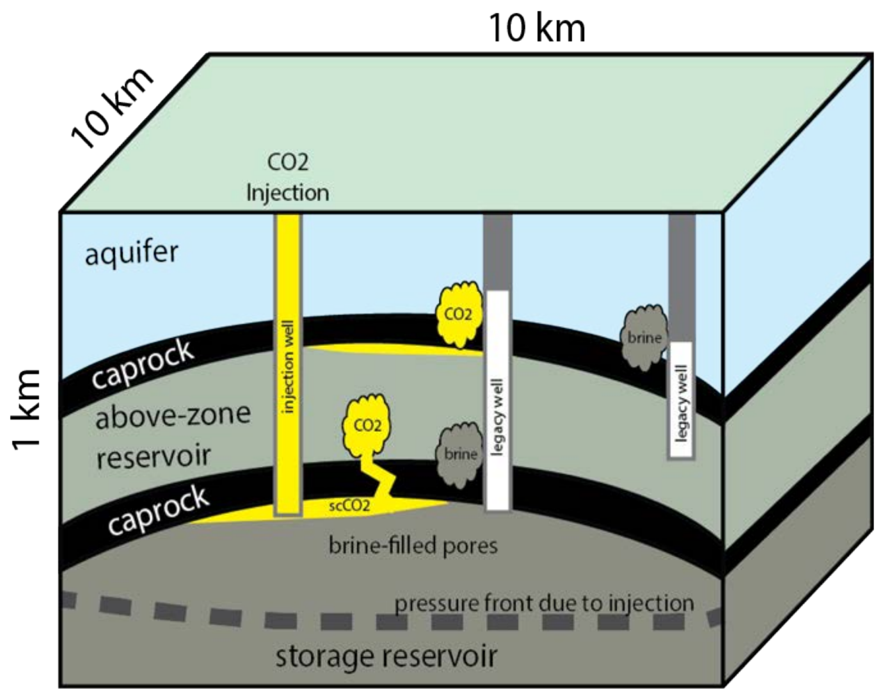


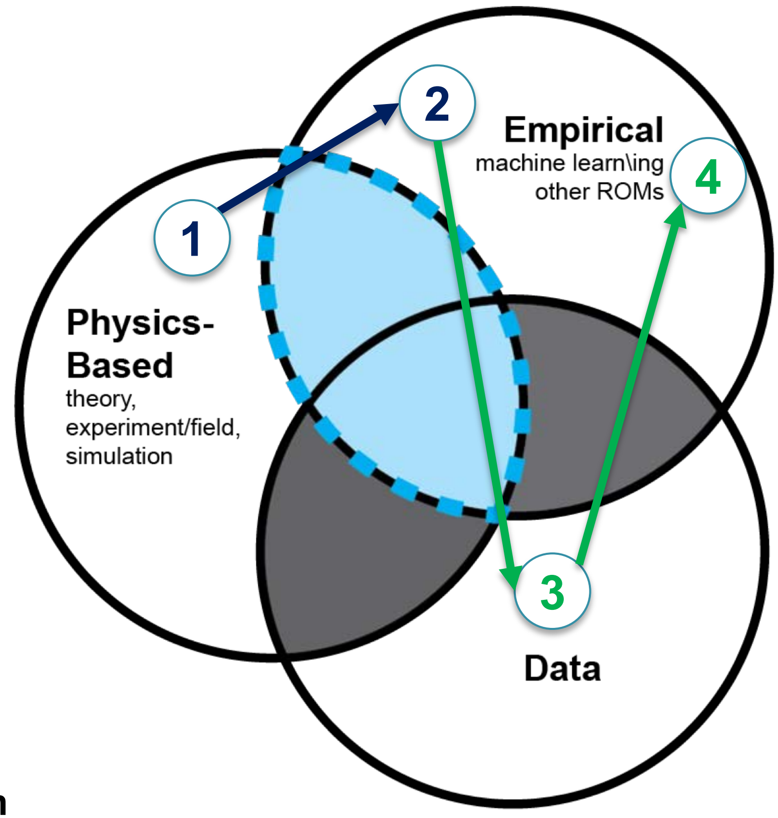
NRAP Toolset

Key gaps that have been filled by NRAP

- *How do you simulate the behavior of large, complex, and uncertain systems?*
  - ✓ Developed new stochastic approach based on physics-informed ROMs
- *How do you accurately and rapidly predict the movement of fluids in a fracture?*
  - ✓ Developed new methods for predicting fluid flow in wellbores & fractured shale
  - ✓ Collected new data on flow in fractures (permeability at conditions; self-sealing)
- *How do you accurately and rapidly predict the impact of leaked fluids on an aquifer?*
  - ✓ Developed new methods for predicting impacts of leaked fluids on aquifers
  - ✓ Collected new data on natural analogs

If you can predict the behavior of a system accurately, then you can create a virtual environment for learning.

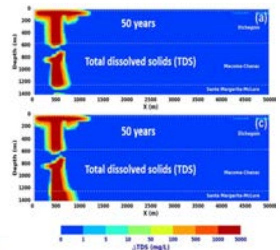




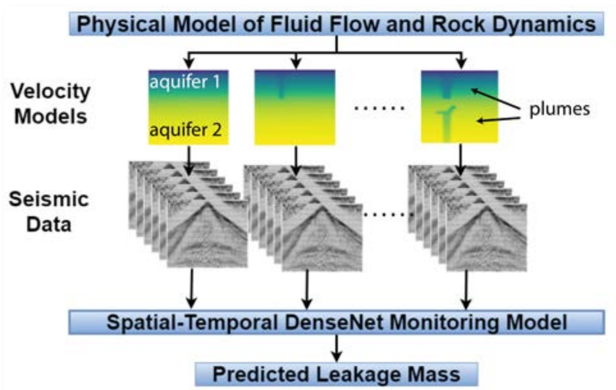
1. Develop predictive understanding of system
2. Use simulated behavior to train empirical models (ROMs; ML)
3. Use empirical models to create numerous simulated datasets (virtual environment)
4. Use machine learning methods to extract knowledge from virtual environment (new signatures and empirical relationships, leading to autonomous monitoring system)

We are beginning to explore for potentially measurable signatures using machine learning and virtual data.

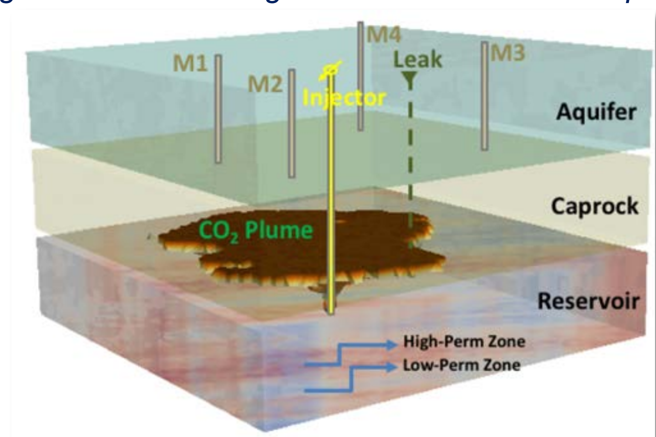
Schematic (below) Showing Use of Simulated Data to Probe for Observable Seismic Signatures based on Leakage Scenarios (right)



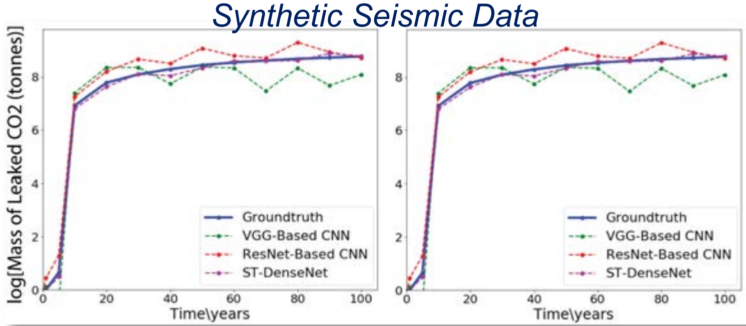
Buscheck et al. (2017)
LLNL-TR-731055



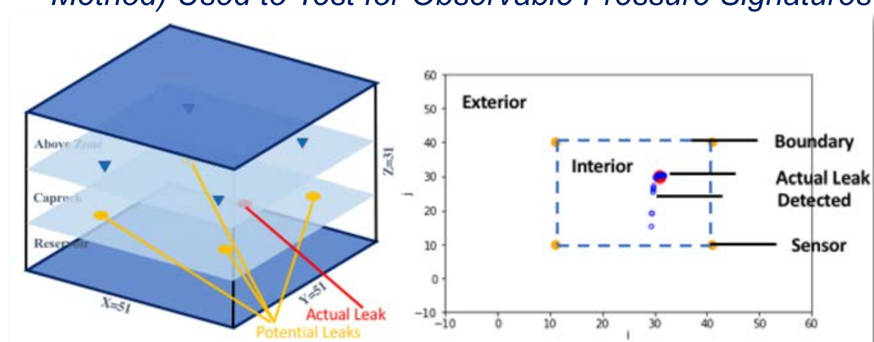
Virtual System Used to Probe for Observable Pressure Signatures for Leakage into an Above-Zone Aquifer



Evaluation of Ability of ML to Detect Plume using Synthetic Seismic Data



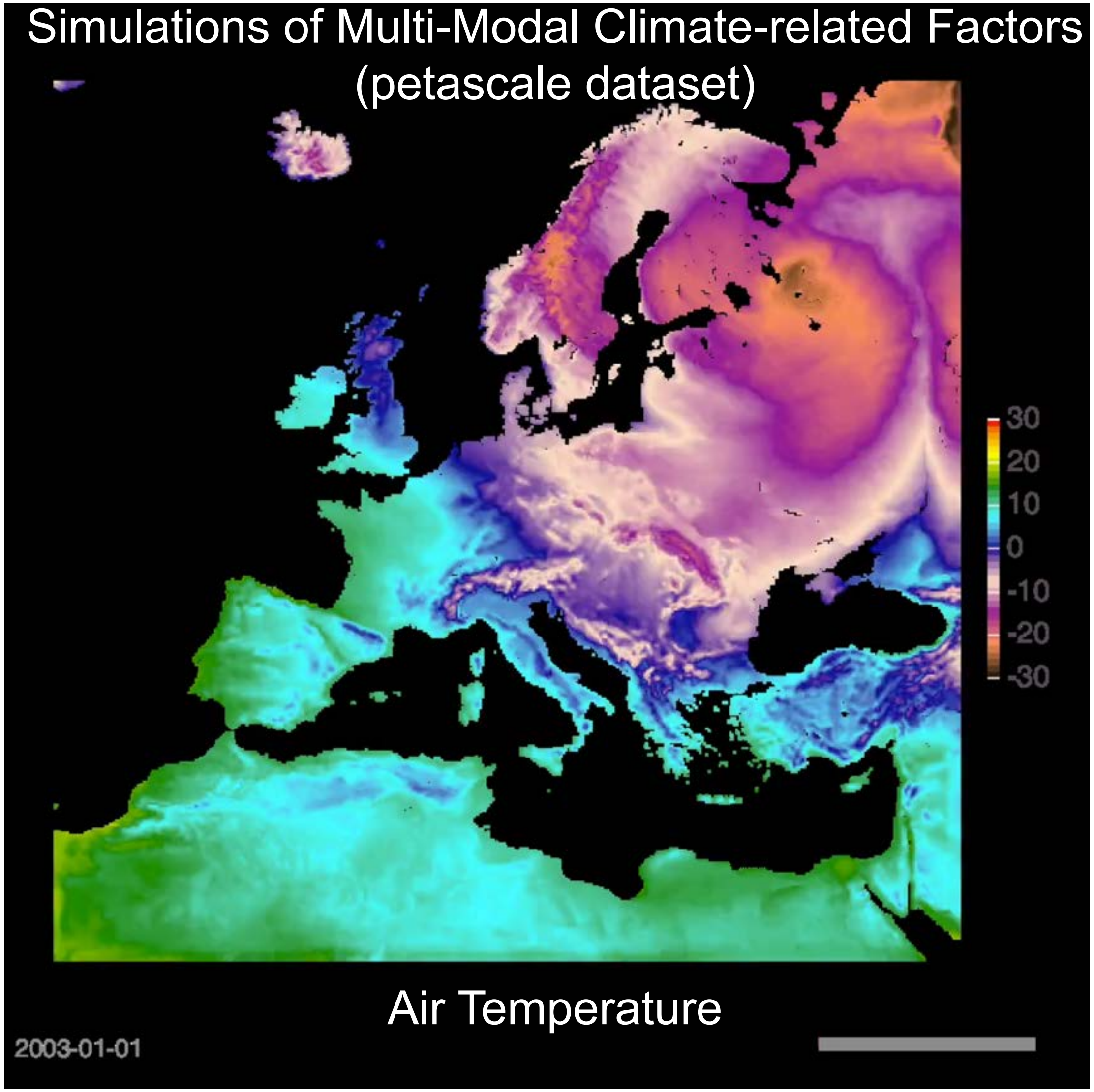
Schematic Showing Support-Vector-Machine Model (Kernel Method) Used to Test for Observable Pressure Signatures



Zhou, Z, Lin, Y, Wu, Y, Wang, Z, Dilmore, R (in review) J *Soc Exp. Geophys.*

Lin, Y, Harp, DR, Chen, B, Pawar, RP (in review) J *Soc Exp. Geophys.*

Internal investments have been developing non-negative tensor factorization as a robust unsupervised ML method.



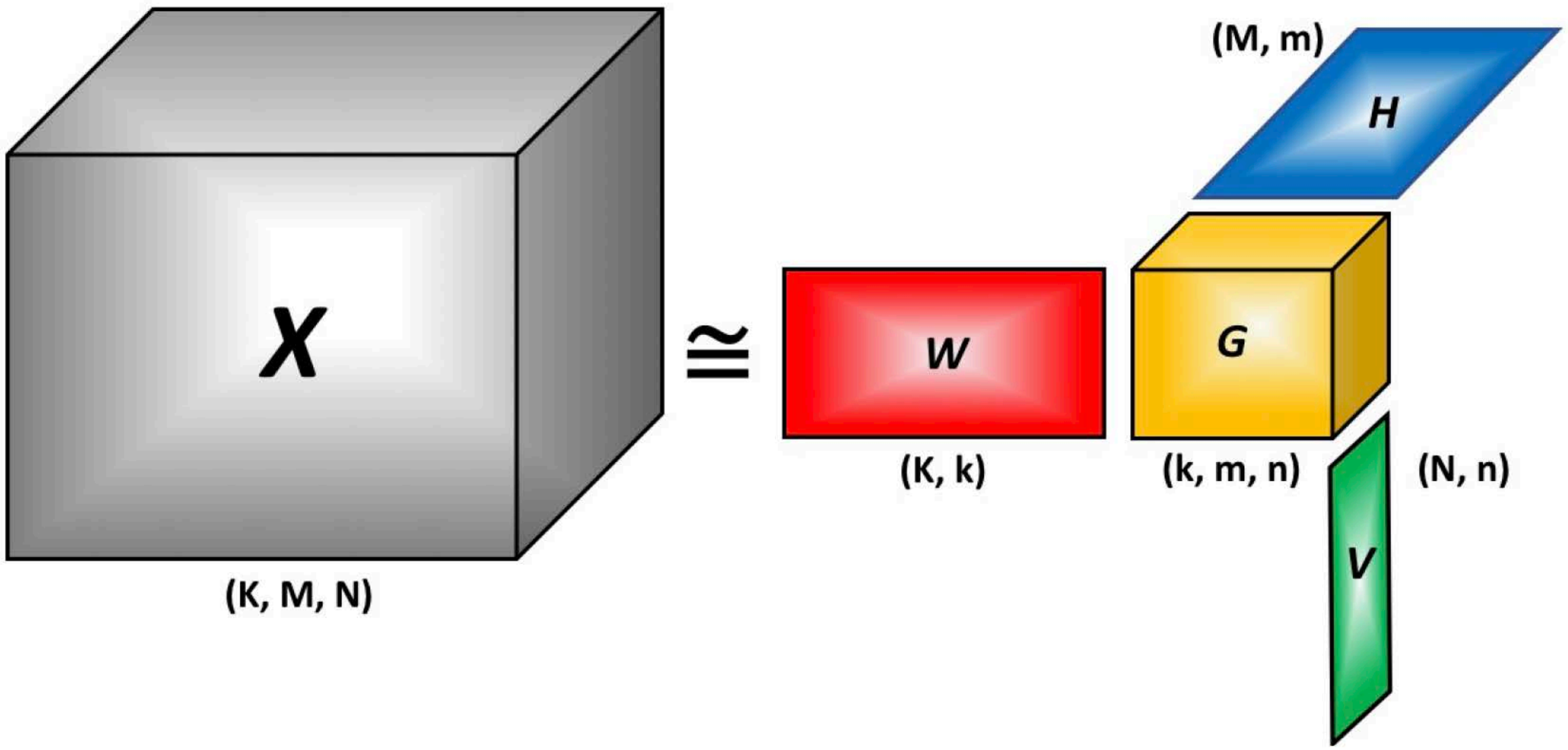
LDRD Team

Theory: Alexandrov (PI), Sandrasegaram, Manzini
Earth Sciences: Vesselinov (PI), O'Malley, Maccarthy
Computer Sciences: Djidjev (PI), Ahrens, Mniszewski, Patchett
Nonproliferation: Bauer, Fessenden, Triplett, Maskaly
UCSD: Ludmil Alexandrov

Current Developments

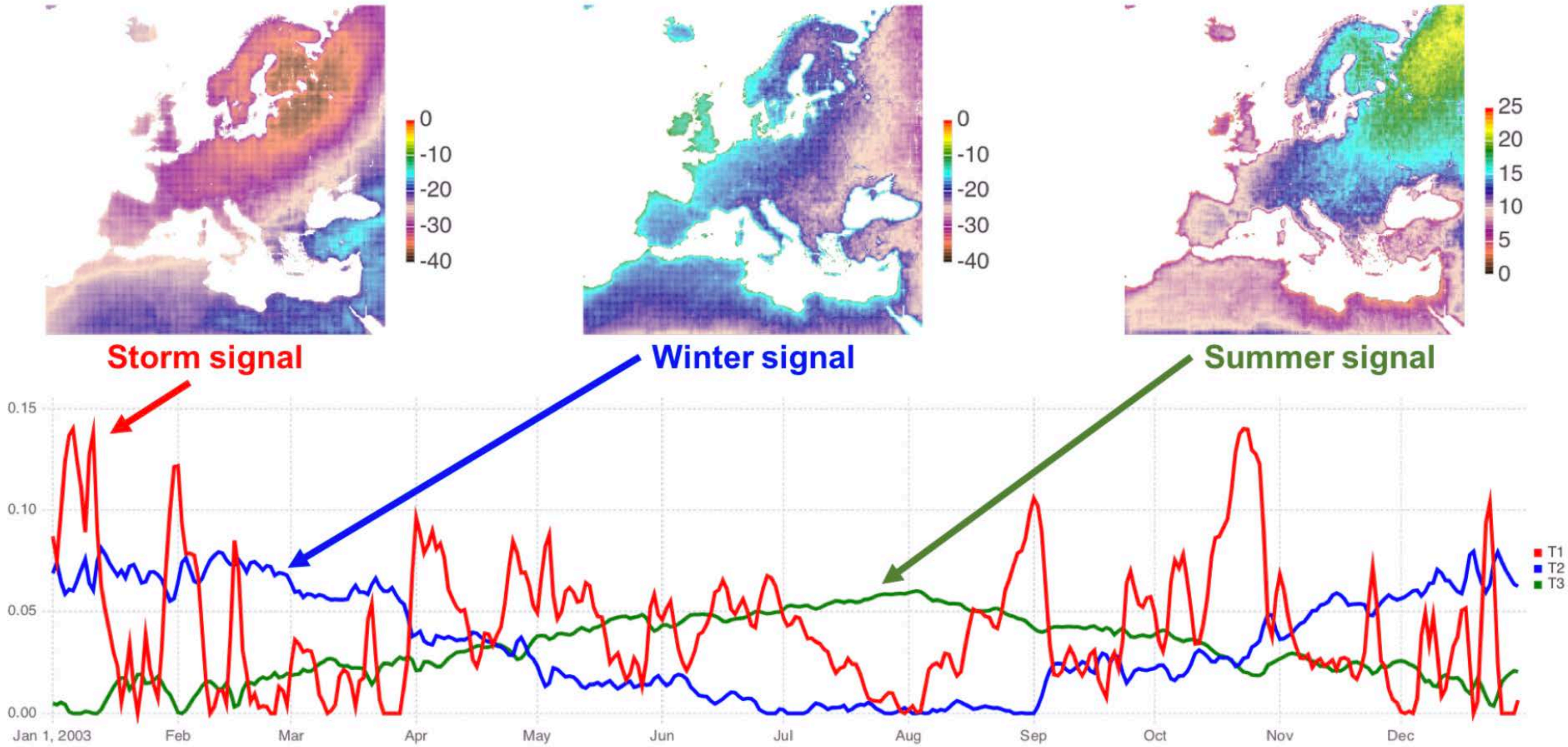
NMFk: Nonnegative Matrix Factorization (patent)
NBMFk: Nonnegative Binary Matrix Factorization (Quantum Computing; D-Wave)
NTFk: Nonnegative Tensor Factorization (copyright disclosure)

Tensor factorization extracts key relationships embedded in the full dataset.



Factorizing (compressing) in all 3 dimensions (K × N × N) → (k × m × n)

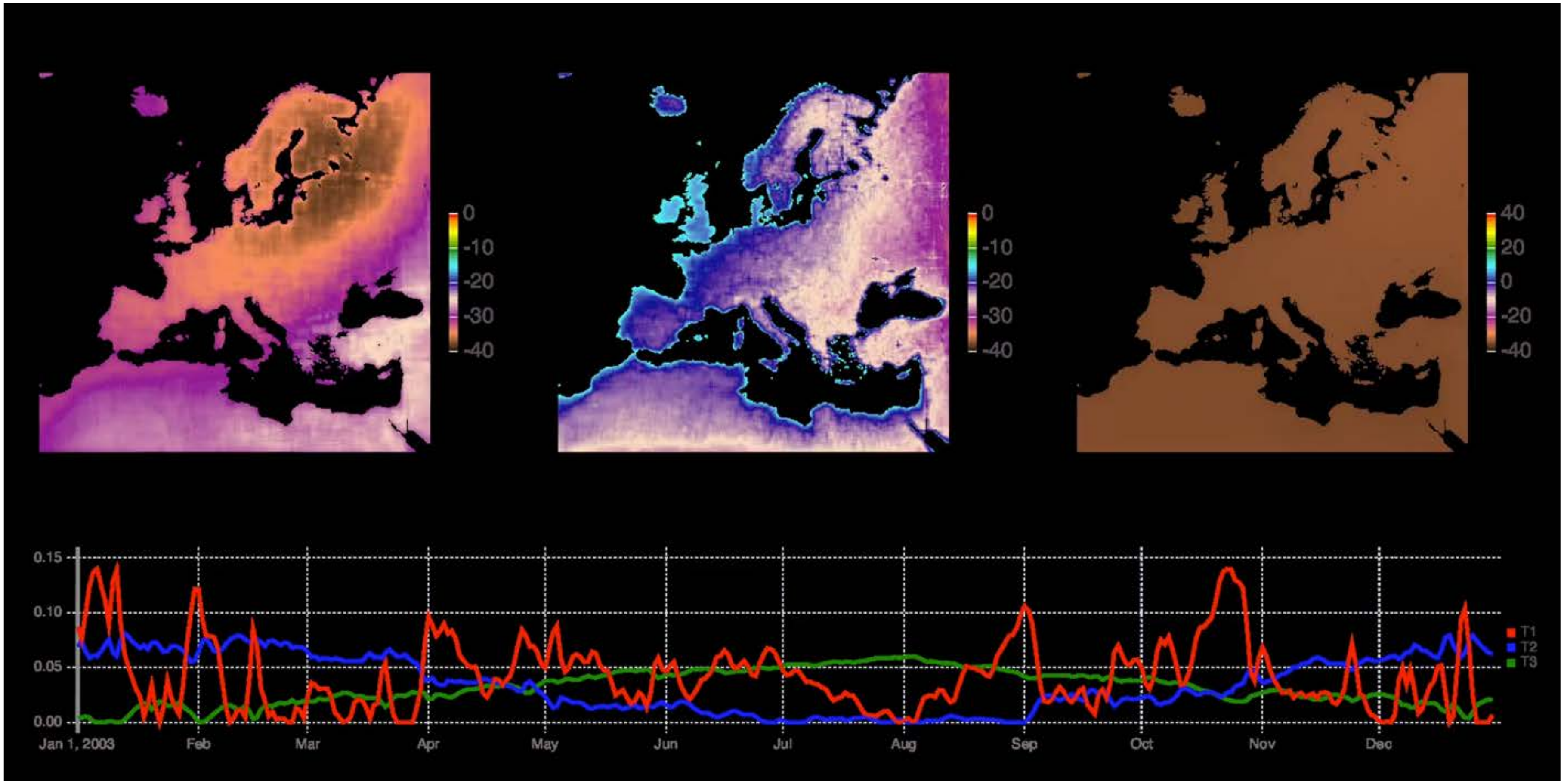
Example Application: Three factors (signals) were extracted from the air temperature data.



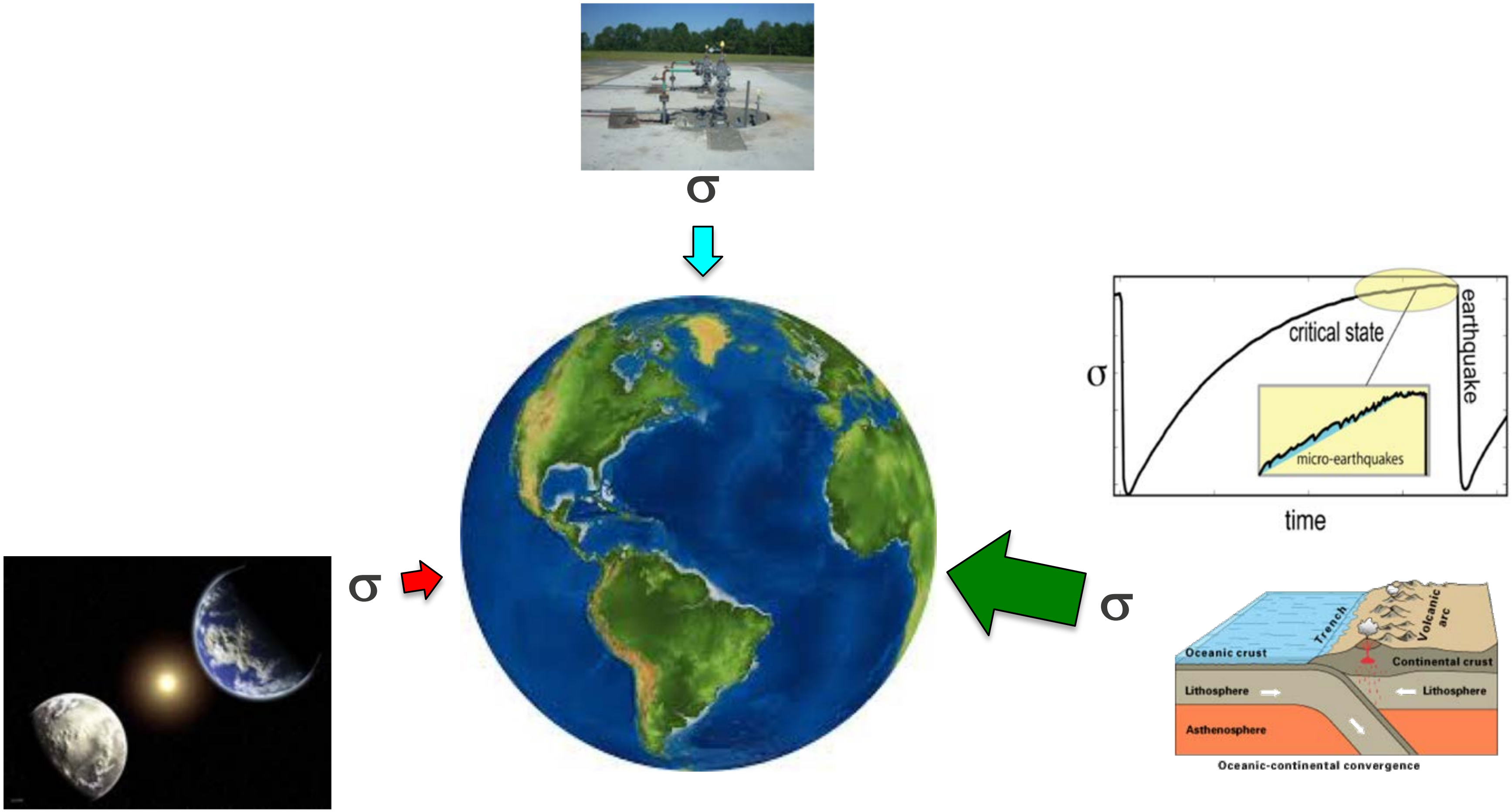
Spatially Averaged Values of Three Signals Embedded in Air Temperature Data

Example Application: Three factors (signals) were extracted from the air temperature data.

Spatial Variation of Three Signals Embedded in Air Temperature Data

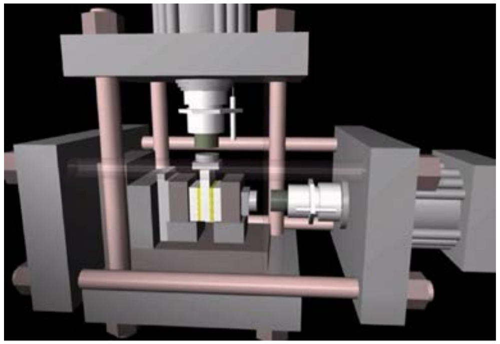


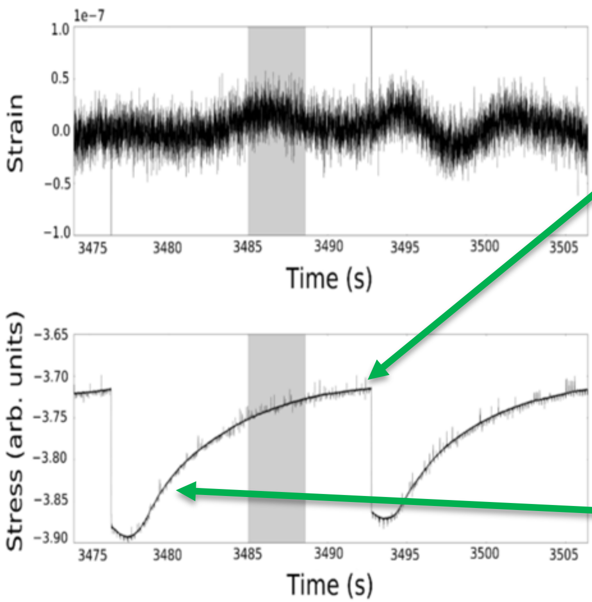
The Earth has a continuously changing state of stress due to tectonic forces, earth tides, injection/extraction, …



Can we “hear” when a fault is reaching a critical state of stress?

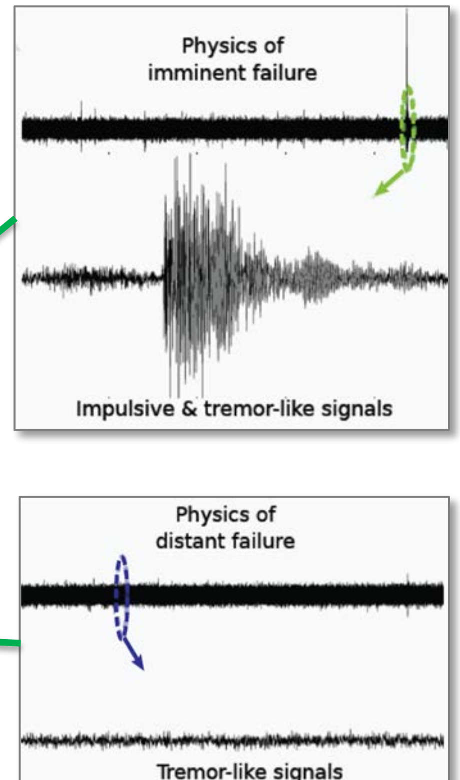
"Earthquake machine" is being used to probe for predictive signatures on state of stress using random forest methods.



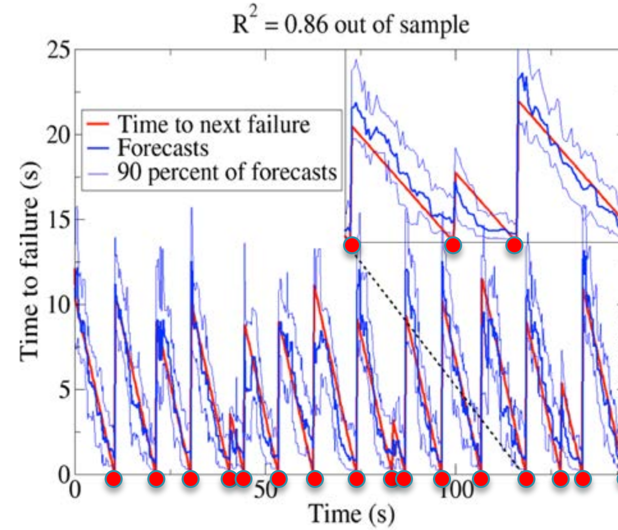


Experimental Data on Slip

Acoustic Emission Signal



Time to Failure Forcasted from Acoustic Emissions



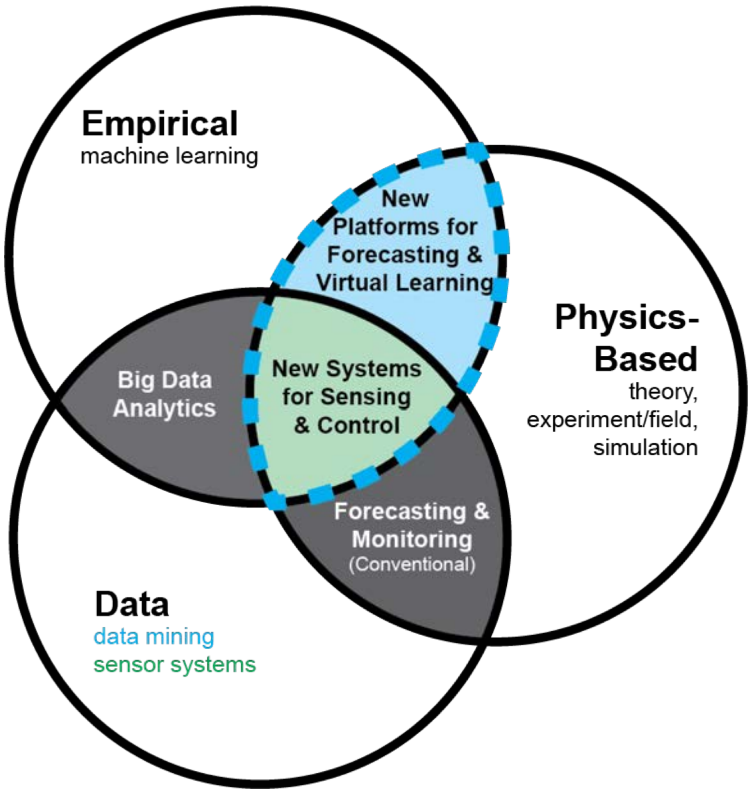
Time to failure is predicted with remarkable accuracy based only on acoustic emissions.

Rouet-LeDuc, B., C. Hulbert , N. Lubbers , K. Barros, and P. Johnson, Learning the Physics of Failure, *Geophys. Res. Lett.*

PennState

Los Alamos NATIONAL LABORATORY EST. 1943

Subsurface opportunities lie at the intersection of machine learning, physics, and data.



Machine learning can reveal controlling behaviors and signals in complex systems

- Large multidimensional datasets
- Signals from noise

Synthetic data can create virtual learning environment prior to field experience

- Testing of new engineering concepts
- Signature discovery

Fusion of synthetic and real data can help to constrain system behavior over real data alone

Autonomous Control

Real-time Data

Virtual Learning

Passive Systems

BACKUP

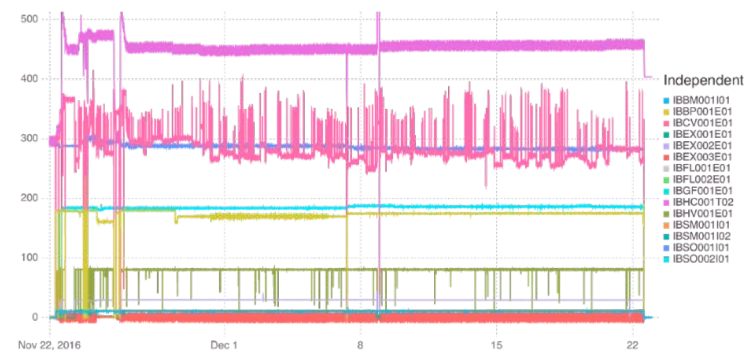
Non-negative matrix/tensor factorization can also apply to process optimization using disparate datasets.

Example: Forecasting (then optimizing) performance of linear accelerator at LANL's neutron scattering center (LANSCE)

- **Goal:**
  Optimal control of particle accelerator
- **Numerous independent variables:**
  factors impacting accelerator performance
- **Proof-of-concept with historical datasets**
  - ➢ Extracted/built a ROM to forecast performance based on a portion of dataset (e.g., green region)
  - ➢ ROM performance forecasts have high accuracy out for several days before degrading (i.e., requires dynamic training)

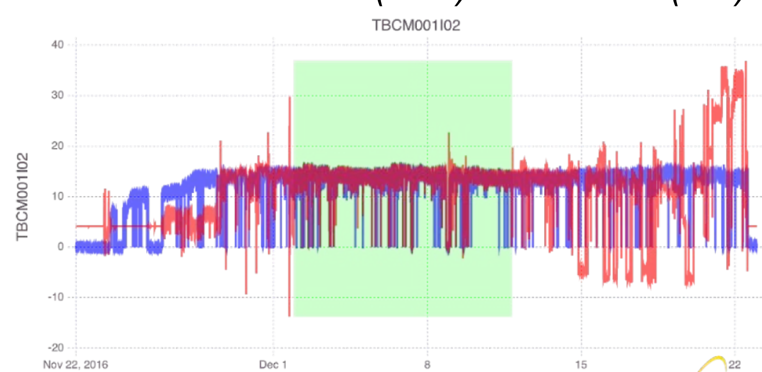
LANSCE Historical Data

Independent Variables



LANSCE Data (blue) & Forecast (red)

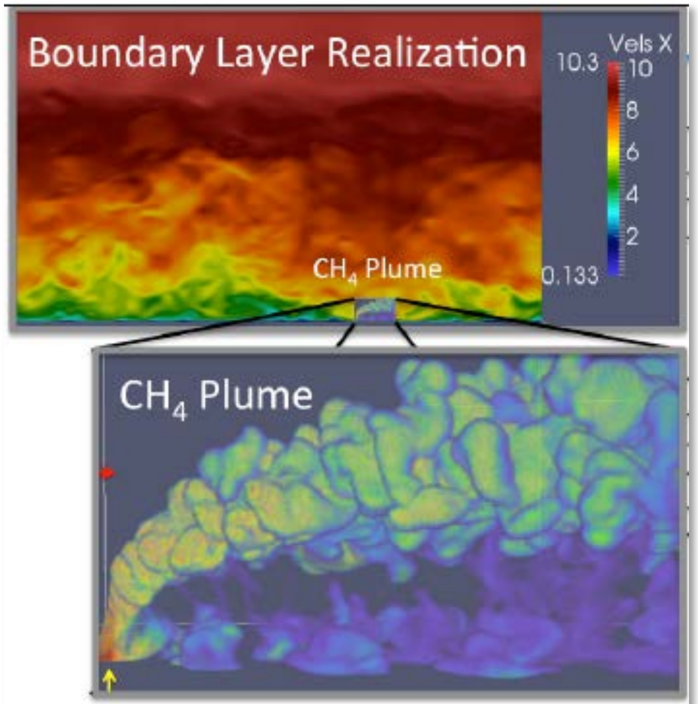
Accelerator Performance



Pre-training to Recognize a Signal prior to Field

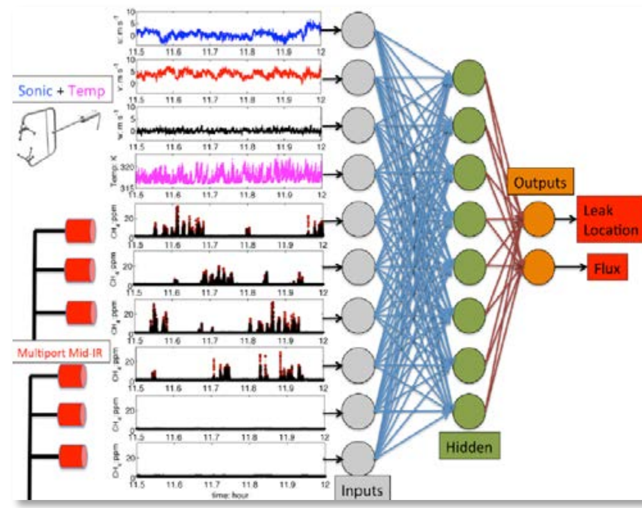
Example: Using computational fluid dynamics simulations to pre-train an artificial neural network (ANN) coupled to a CH\ sensor and a meteorological tower for detection of NG leak.

3D CFD Simulations



Simulations for pre-training, with site-specific field data to refine

ANN for dynamic signal analysis



Sauer, Travis, & Dubey (2017, LANL Copyright)

<u>Dependent variables:</u> Leak location; NG flux

<u>Independent variables:</u> wind speed/direction, temperature, conditions, terrain, time-series of CH_4 at sample stations