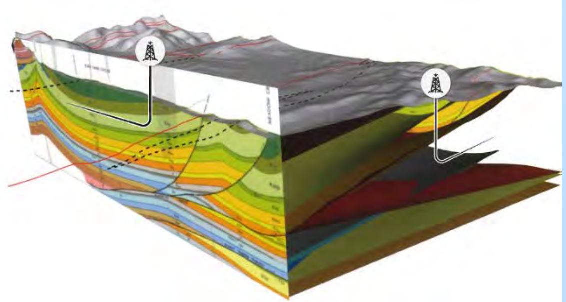
Machine Learning, an informal discussion

William Harbert,

University of Pittsburgh

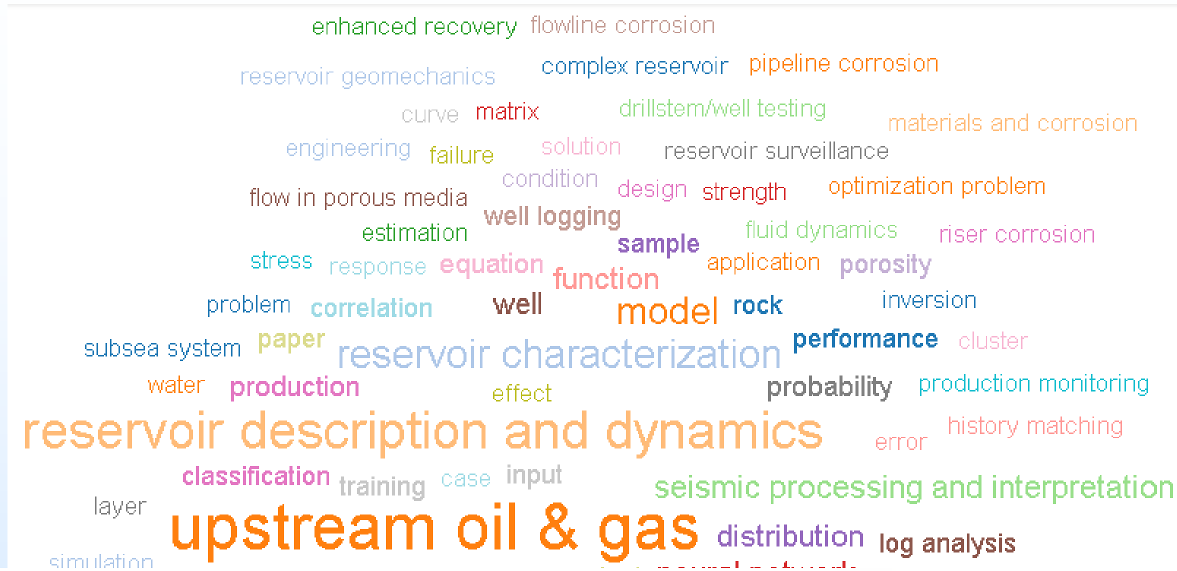


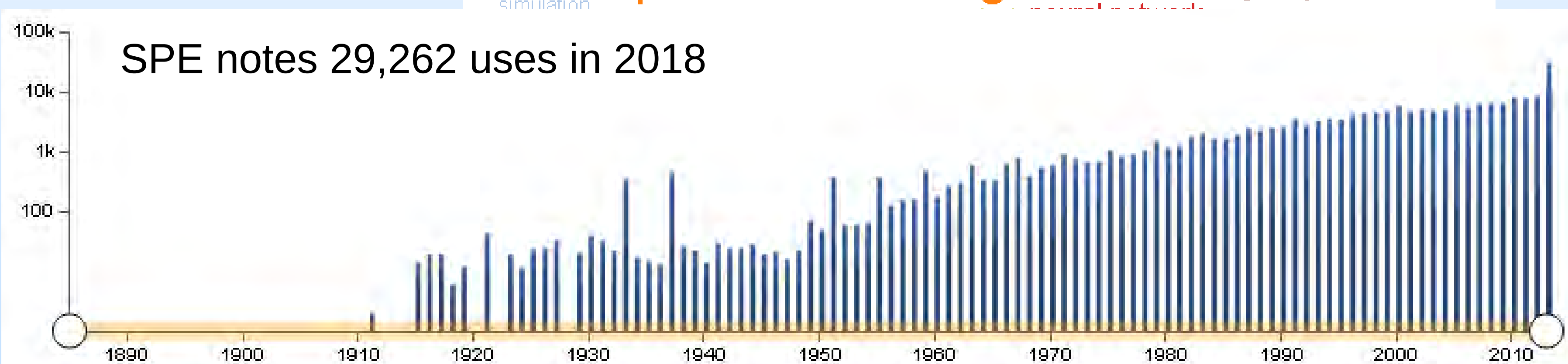
2018 Mastering the Subsurface Through Technology Innovation, Partnerships and Collaboration: Carbon Storage and Oil and Natural Gas Technologies Review Meeting

An approaching tsunami

- Several technology and science based advanced have coalesced in the technology of Machine Learning or ML.
- ML is presently being used in geophysical applications focused on applications such as hydrocarbon exploration, production activities, and anomaly detection using potential fields.
- The promise is that high value targets can be confidently identified using ML.
- The potential application areas within the geosciences are very broad.

SPE Word cloud for “Machine Learning”





Introduction to AI and Machine Learning

by Phil Bording

Machine Learning and AI are being applied to geophysics interpretation and processing problems and the SEG workforce needs education on how these methods work, and when to use them.

Duration

Two days

Intended Audience

Intermediate level

Prerequisites (Knowledge/Experience/Education Required)

Seismic processing and interpretation, basic computer programming and scripting, reasonable linear algebra skills, an understanding of mathematical optimization methods

Course Description

AI introduction including rule based systems, Bayesian processes, game playing search trees with breadth first and depth first data structures. Symbolic manipulation and textual search methods. Linear algebra for machine learning, non-negative and singular value decomposition of matrices. Optimization methods, stochastic gradients as applied to convolutional neural nets

This course is being offered as part of the SEG 2018 Annual Meeting in Anaheim, CA on 13-14 October.

Register now

Learn more

Complex

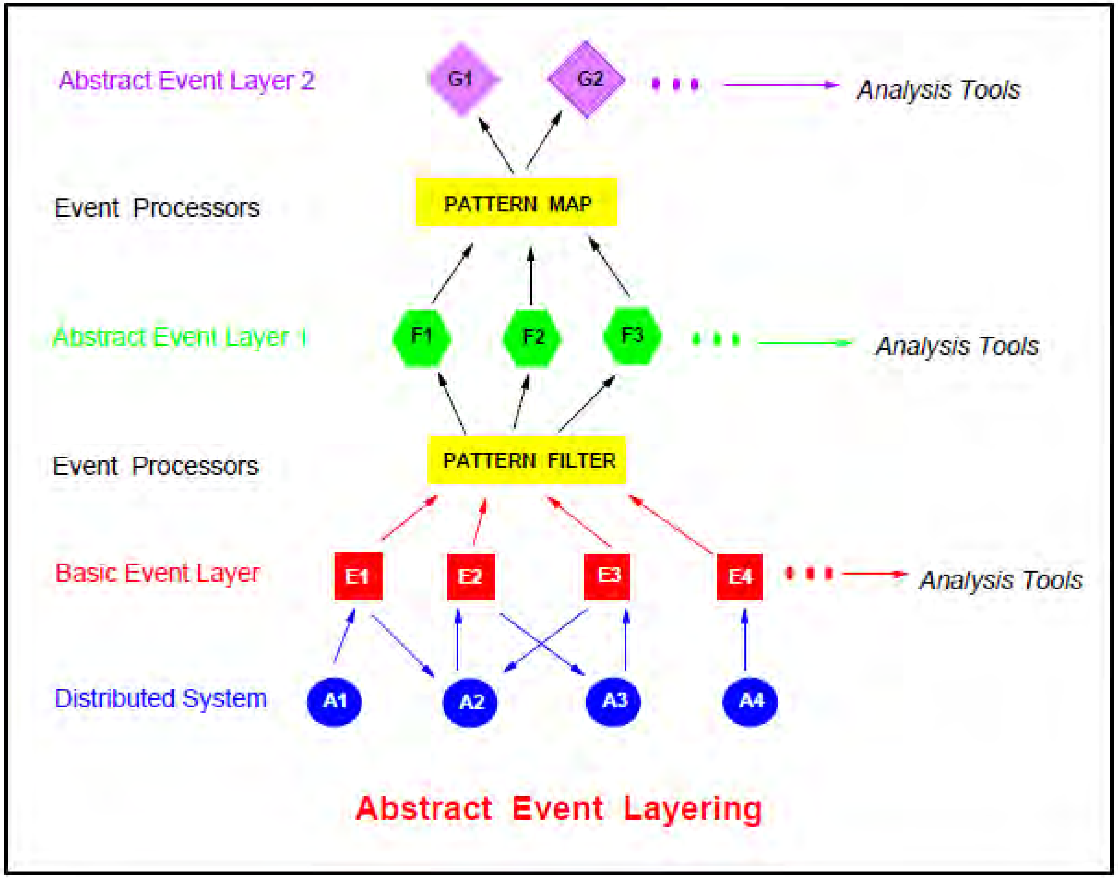
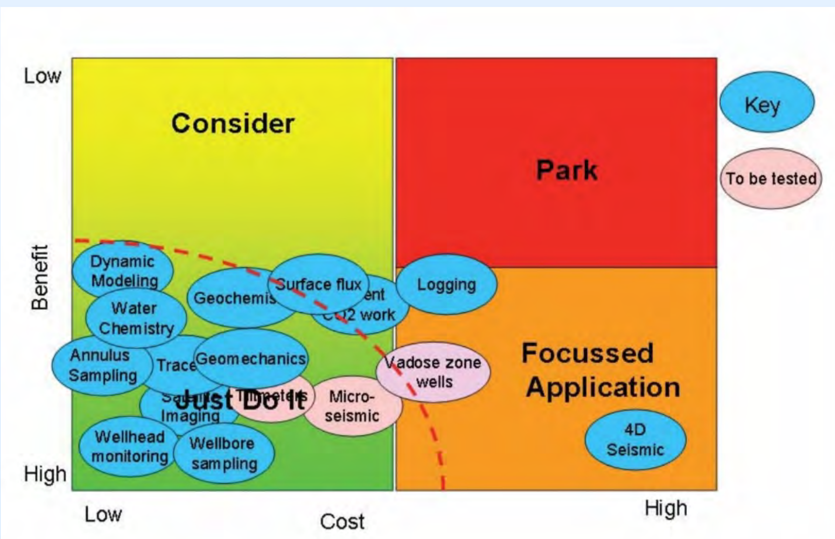


Figure 4: Hierarchical organization of event processing objects

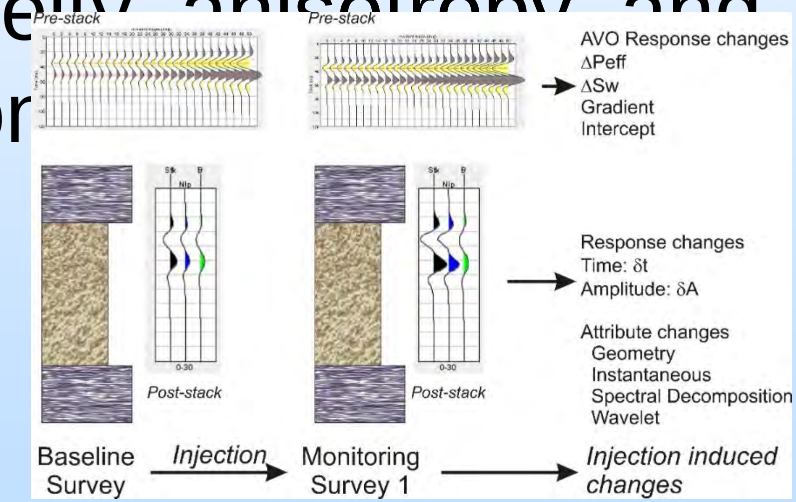
Complex Event Processing in Distributed Systems, Luckham and Frasca, Stanford, 1998

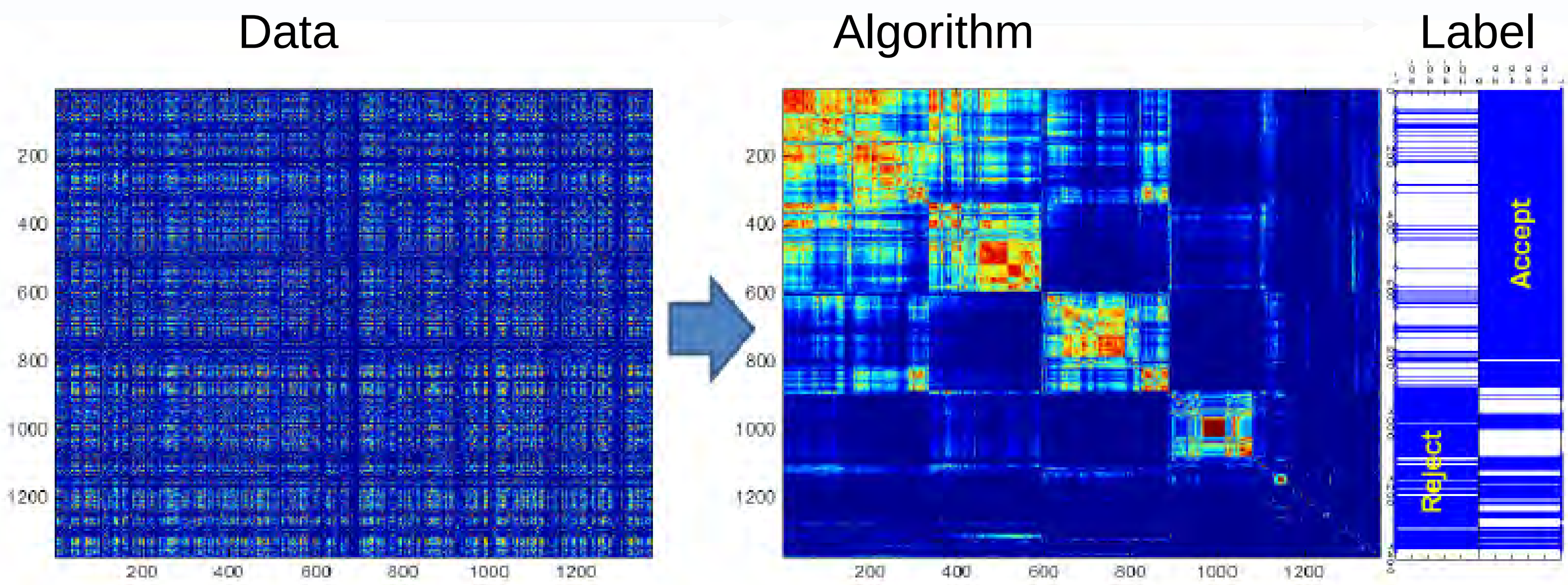
Additional factors relevant to subsurface applications

- Scales of monitoring resolution, both temporal and spatial.
- Scales of relevant physical processes.
- neity, anisotropy, and
  or



Modified after Mathieson et al., 2010





5a. Similarity matrix (data in chronological order)

5b. Rearranged similarity matrix (with expert labels)

Figure 5. Comparison of similarity matrices before and after rearrangement to visual data structure. The similarity matrix is shown as a heat map (red indicates a high value of similarity and blue indicates a low value). The stem plot to the right of Figure 5b shows the known labels for each test (row).

SPE-167839-MS

Advanced Machine Learning Methods for Production Data Pattern Recognition

Niranjan Subrahmanya, ExxonMobil Research and Engineering Company, Peng Xu, ExxonMobil Upstream Research Company, Amr El-Bakry, ExxonMobil Production Company, Carmon Reynolds, ExxonMobil Information Technology

parameter estimation

– What will a pre-curso

event look like

– When and where will a pre-cursor to an

Breaking Cascadia's Silence: Machine Learning Reveals the Constant Chatter of the Megathrust

[Claudia Hulbert](), [Paul A. Johnson]()

(*Submitted*)

Tectonic faults slip in various manners, ranging from ordinary earthquakes to slow slip events to aseismic fault creep. The frequent occurrence of slow earthquakes and their sensitivity to stress make them a promising probe of the neighboring locked zone where megaquakes take place. This relationship, however, remains poorly understood. We show that the Cascadia megathrust is continuously broadcasting a tremor-like signal that precisely informs of fault displacement rate throughout the slow slip cycle. We posit that this signal provides indirect, real-time access to physical properties of the megathrust and may ultimately reveal a connection between slow slip and megaquakes.

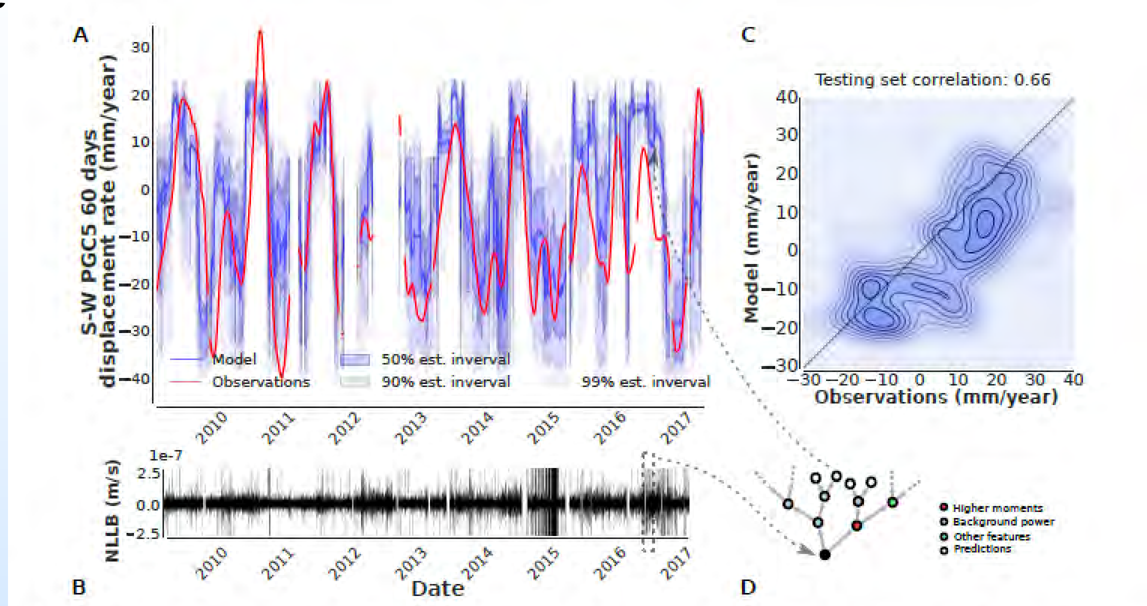


Figure 2: **Estimating the GPS displacement rate from the continuous seismic data.** (**A**) The red line shows the actual, smoothed GPS displacement rate (60 day, from the PGC5 GPS station). The blue bold curve is an estimate from the ML model (with estimation intervals noted with shades of blue) using characteristics of the full continuous seismic data as input. The figure shows the testing set, for which the algorithm *only* has access to the seismic data (e.g., (B)). The data gaps indicate missing (GPS and/or seismic) data. (**B**) Continuous seismic data from station NLLB over the same time interval. (**C**) Distribution of observed versus predicted displacement rates, with contours showing empirical iso-density, from 10 to 90%. The Pearson correlation coefficient between estimates and actual displacement rate is of 0.66, showing that continuous seismic waves contain rich information about the fault's state, apparently at all times. (**D**) Once the full continuous seismic data has been turned into a database of statistical features, it is fed to an ensemble of decision trees (schematic) that partitions the data to build a model of GPS displacement rate as a function of statistics of the seismic data.

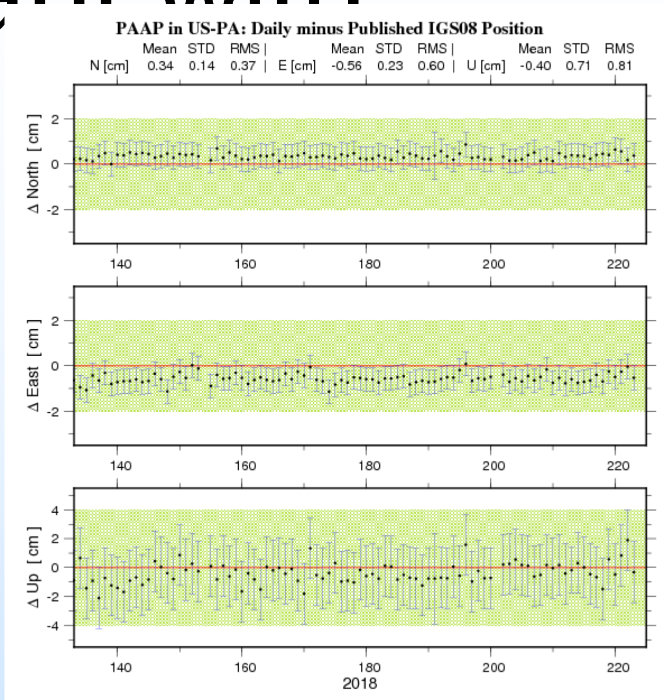
Cited as: [arXiv:1805.06689]() [physics.geo-ph]

Some ML details

- Data formats conversion: Opportunity for Data *Standardization*.
- Metadata, velocity of data, chain of processing and ownership. Opportunity for *Metadata Standardization*.
- Robust training datasets. These determine the impact of ML. Opportunity— *Deep ML education and training using real-world datasets*.

Developments relevant with ML

- Continuous monitoring
- DAS
- Dark Fiber
- Modular Borehole Monitoring (MBM)
- Smart hydrocarbon fields
- Opportunity: *Noise as Signal*, GPS CORS example. Use GPS noise to determine atmo
- LiDAR on a dim
- Internet of thing



https://www.spar3d.com/news/lidar/mits-10-lidar-chip-will-change-3d-scanning-know/

https://ocw.mit.edu/resources/res-ll-003-build-a-small-radar-system-capable-of-sensing-range-doppler-and-synthetic-aperture-radar-imaging-januar
2011/

Potential new partnerships

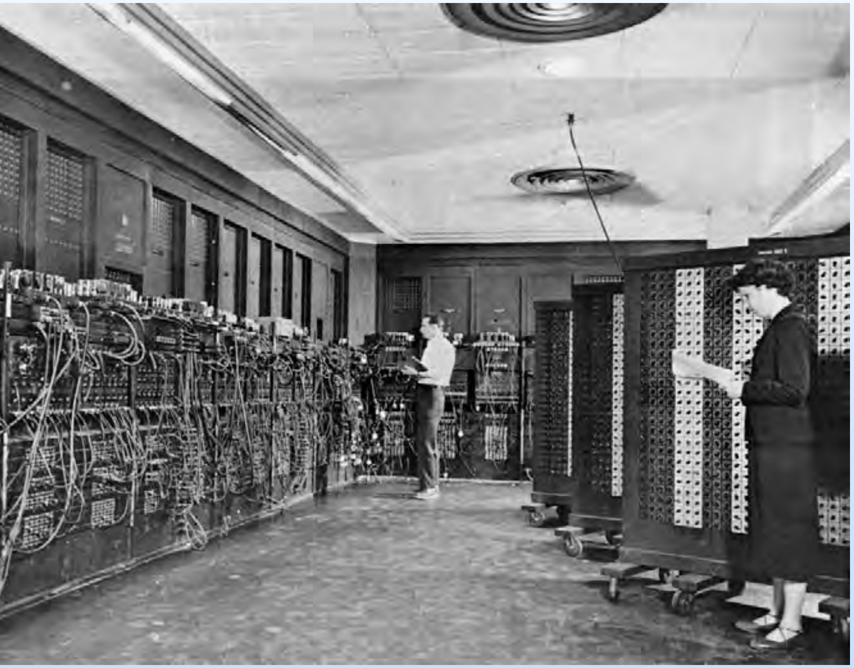
momacs

MODELING AND MANAGING
COMPLICATED SYSTEMS

- The MOMACS Institute was announced by Pitt in May, 2018. It is associated with the School of Computing and Information, Pitt's first new school since 1995. According to Founding Dean Paul Cohen, the mission of the school is to develop technologies to help humans model and manage hugely complicated, interacting systems.

- The MOMACS Institute does applied research with stakeholders in industry, nonprofits, government, the DoD and intelligence communities. It draws on the expertise of Pitt faculty in many disciplines to solve problems associated with complicated systems, including the brain, energy systems, financial systems and others.

Thank you!



Data Analytics and Machine Learning Panel

Real-Time Decision Making for the Subsurface Workshop

- **July 17-18, 2018**
- **Attendees:**
  - DOE and National Labs: ~45%
  - Academics: ~33%
  - Industry: ~22%
- **Two technical areas:**
  - Carbon Storage
  - Unconventional Oil and Gas
- **Three main breakout themes:**
  - Resource recovery and utilization
  - Autonomous monitoring
  - Seismicity and dynamic stress state

Hosted by: Carnegie Mellon University
Wilton E. Scott Institute for Energy Innovation

Another Workshop hosted by USEA
Held July 12
Focus on clean coal and carbon management
Applications of Big Data and Machine Learning

Key industry-relevant use cases

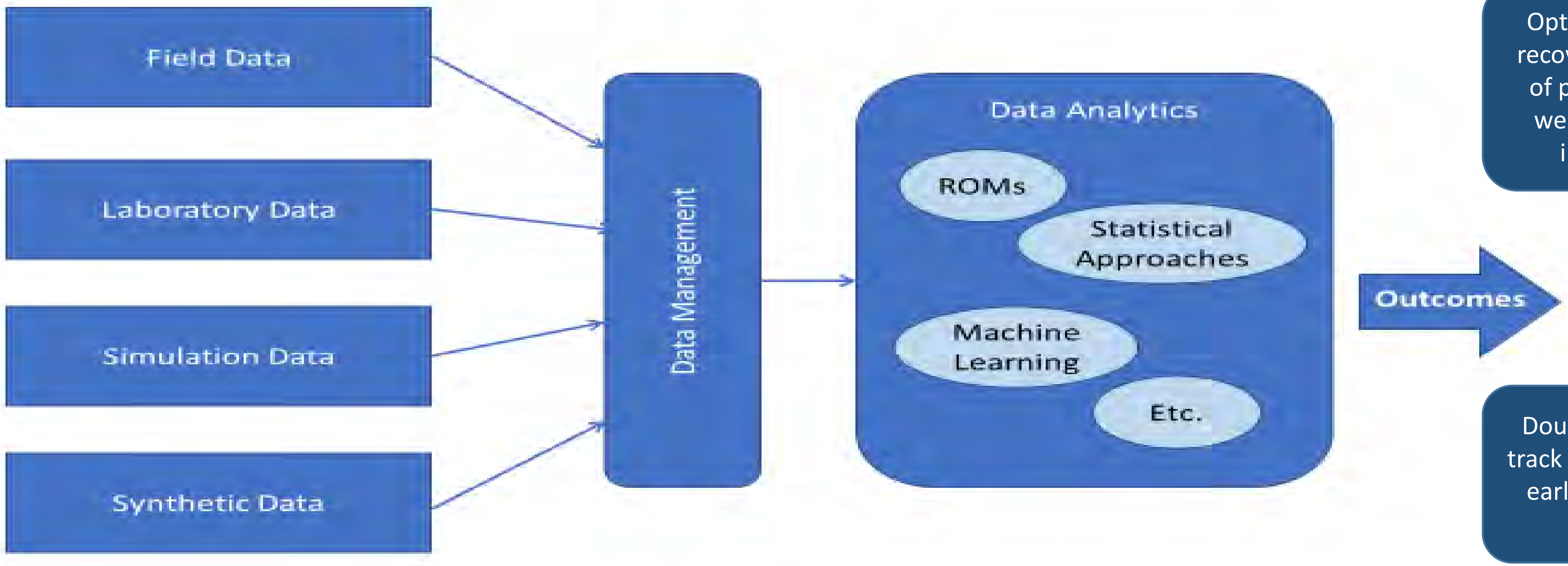
- **Resource recovery and utilization**
  - Completion optimization
  - Reservoir operations
  - Drilling/geosteering
- **Autonomous monitoring**
  - Safety—decisions in seconds to hours
  - Decisions at the well—decisions in hours to days
  - Reservoir management—decisions in days to months
- **Seismicity and dynamic stress state**
  - Prevent damaging seismicity
  - Improve reservoir characterization and monitoring

Needs/barriers identified

- **Need for complete data sets, shared**
- **Need systems to integrate large data sets of different types**
  - Multiple sensors
  - Distributed, point, continuous, discrete data
- **Reduce data/distribute analysis**
  - Compression/reduction of data
  - Edge computing (Raspberry Pi)
- **Minimize data biases**
- **Quantify uncertainty**
- **Insufficient resources (personnel, computational)**

Move beyond traditional data analytics approaches to overcome barriers

- Bring subject matter experts together with data scientists
- Enable “human-in-the-loop” and eventually autonomous systems



Optimize resource recovery; Reduce #s of poor and failed wells; Reduce HF interference

Double CO_2 stored; track CO_2, ΔP plumes; early warning leak detection

Thank you!

Some needs for future subsurface operations…

Learning prior to field observations/experience…

More knowledge from the subsurface
(quicker, cheaper, higher relevance)…

Higher efficiency, reliability, etc. for reservoir management…

Some needs for future subsurface operations…
…that could be addressed by (physics+) machine learning

Learning prior to field observations/experience…

…virtual learning in a variable/uncertain subsurface

via rapid emulation of multi-scale, complex, nonlinear systems

Some needs for subsurface operations…
…that could be addressed by (physics+) machine learning

Learning prior to field observations/experience…
…virtual learning in a variable/uncertain subsurface

More knowledge from the subsurface
(quicker, cheaper, higher relevance)…

…knowledge from noise
e.g., more from existing data

…autonomous monitoring
i.e., Data → Analysis + Visualization → Decision by Person
"human in the loop"

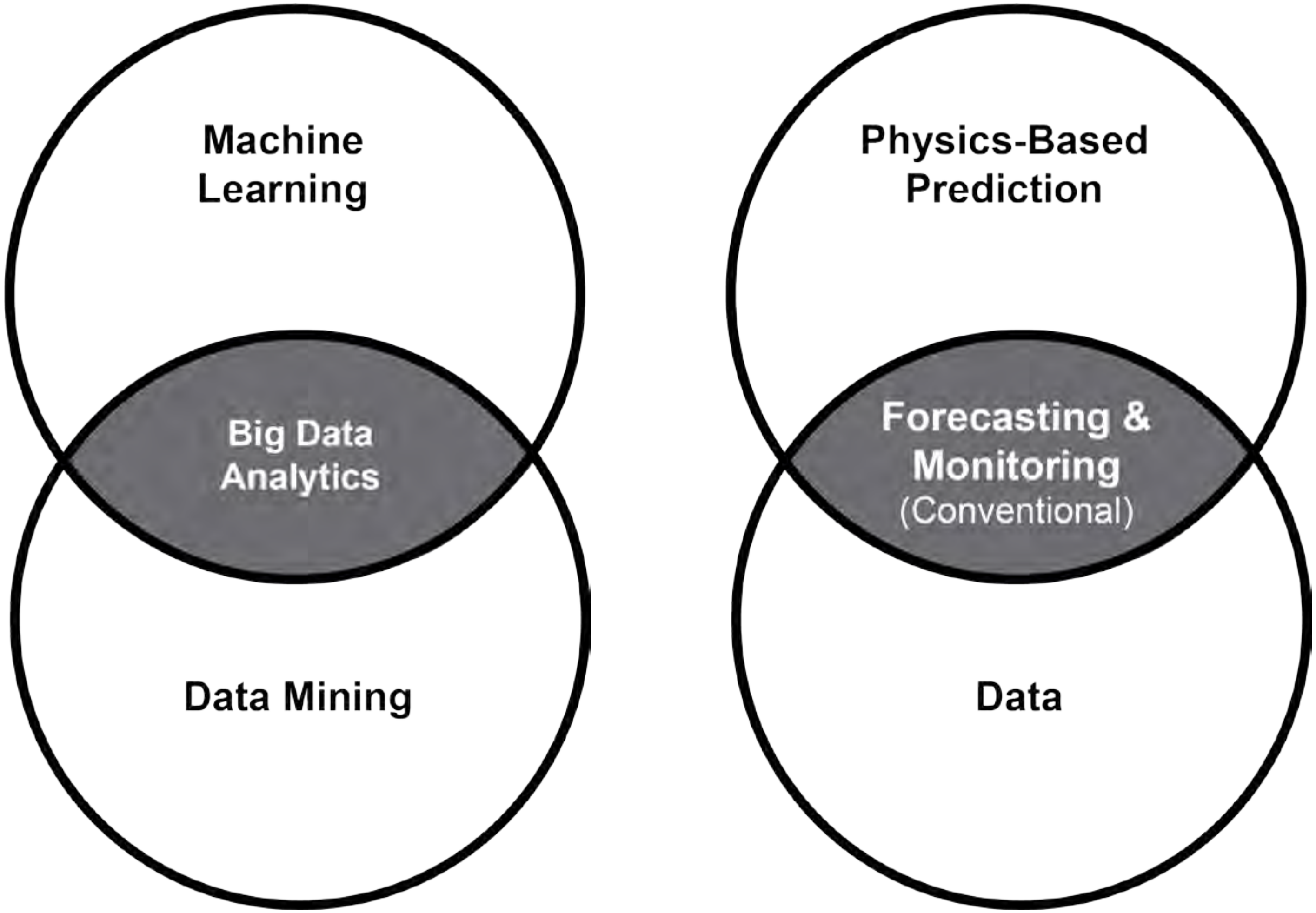
Some needs for subsurface operations…
…could be addressed by (physics+) machine learning

Learning prior to field observations/experience…
…virtual learning in a variable/uncertain subsurface

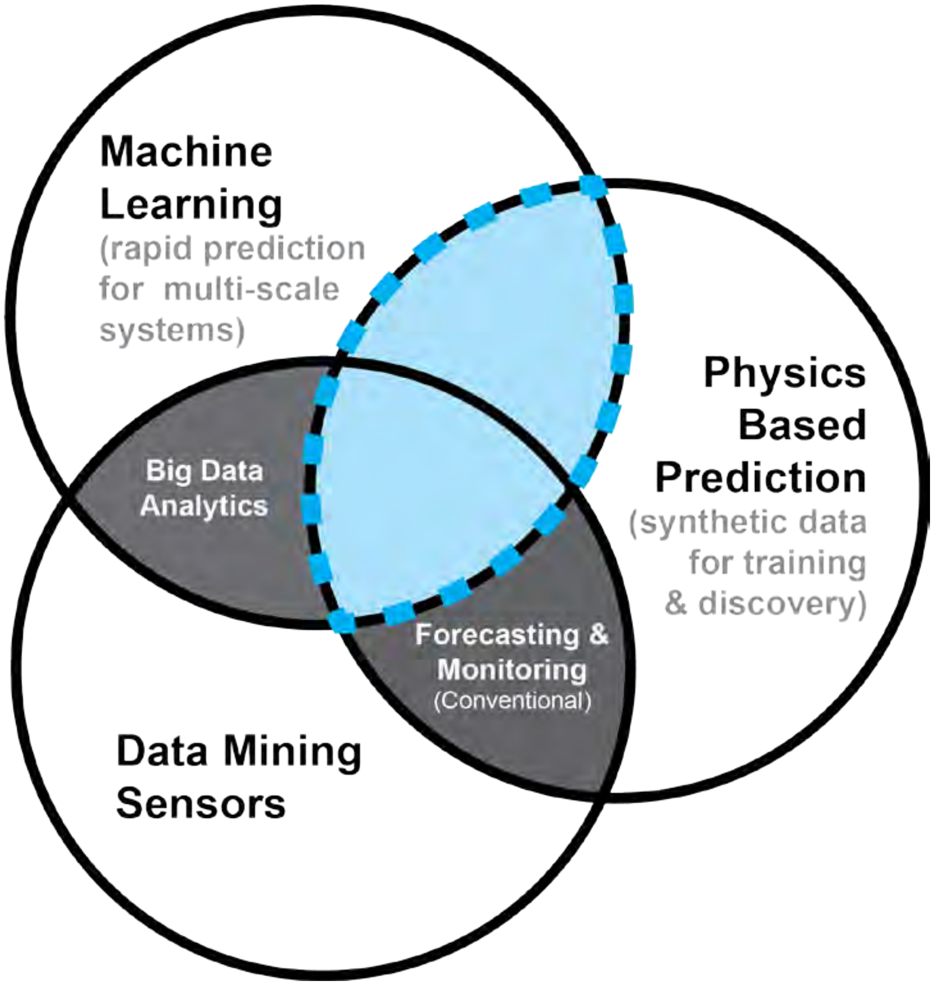
More knowledge from the subsurface
(quicker, cheaper, higher relevance)…
…knowledge from noise; autonomous monitoring

Higher efficiency, reliability, etc. for reservoir management…
…autonomous monitoring and control

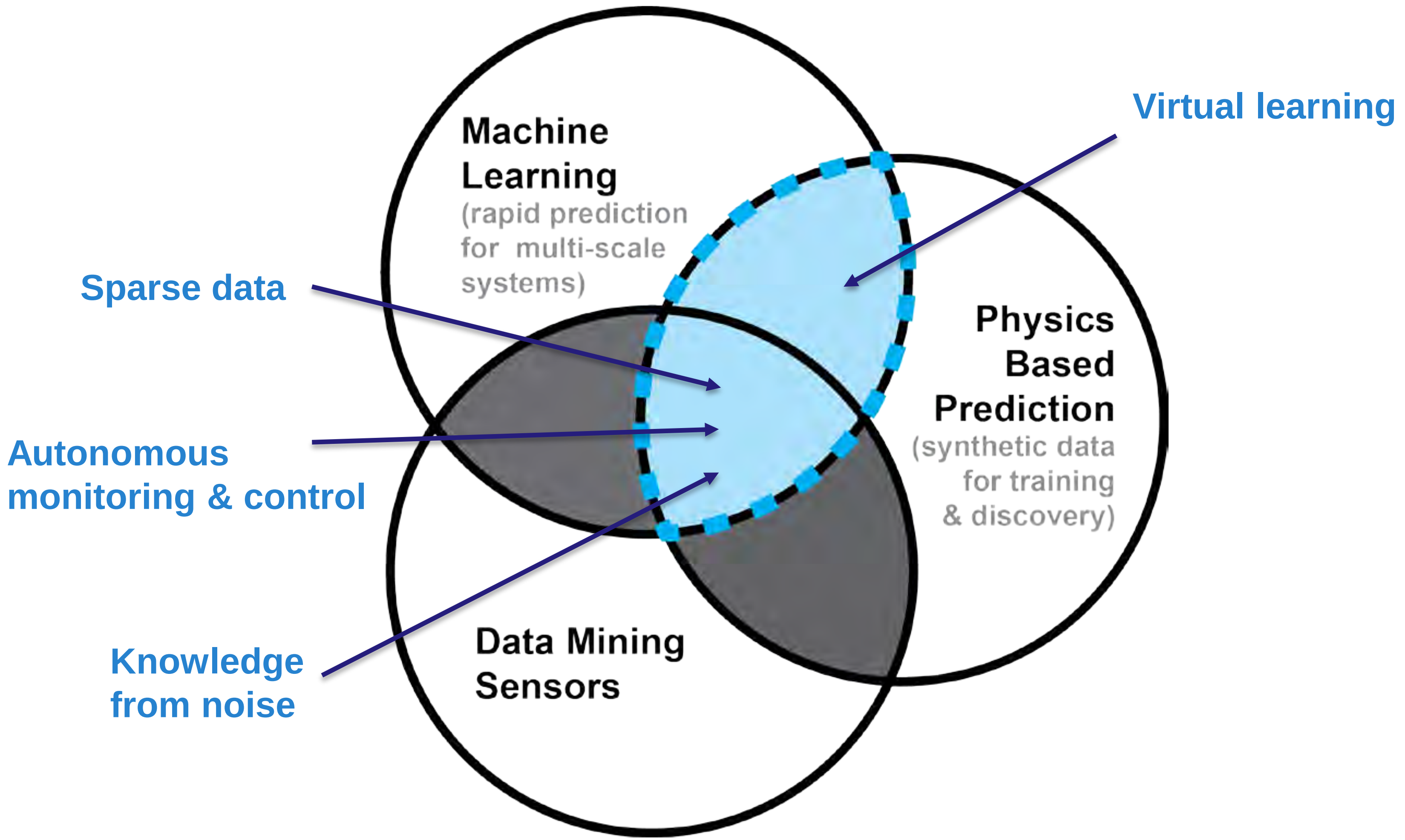
i.e., Data → Analysis + Visualization → Decision by Machine


Machine
Learning
Big Data
Analytics
Data Mining
Physics-Based
Prediction
Forecasting &
Monitoring
(Conventional)
Data

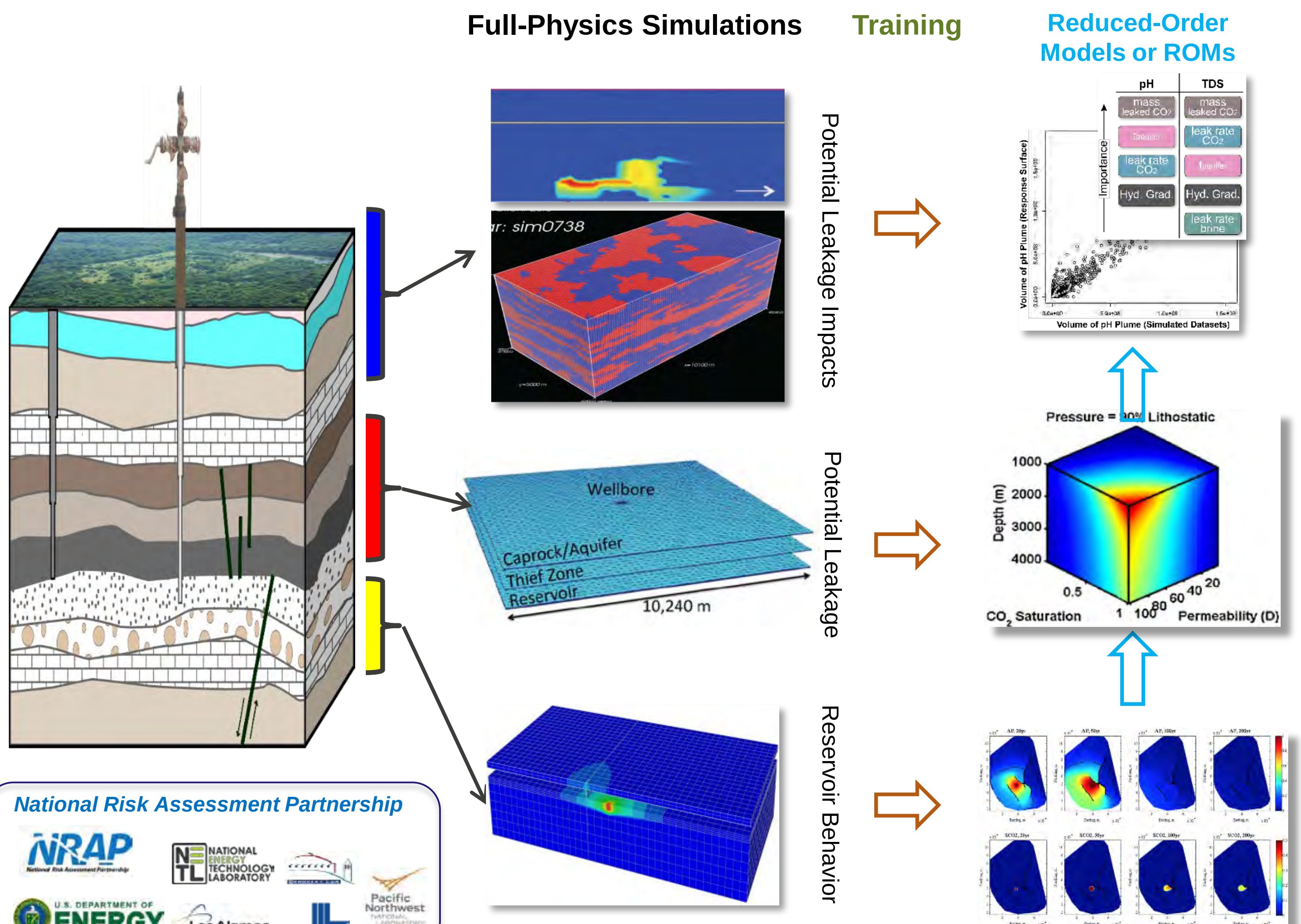
Intersection of ML and physics can enable applications that neither alone can address.



Intersection of ML and physics can enable applications that neither alone can address.

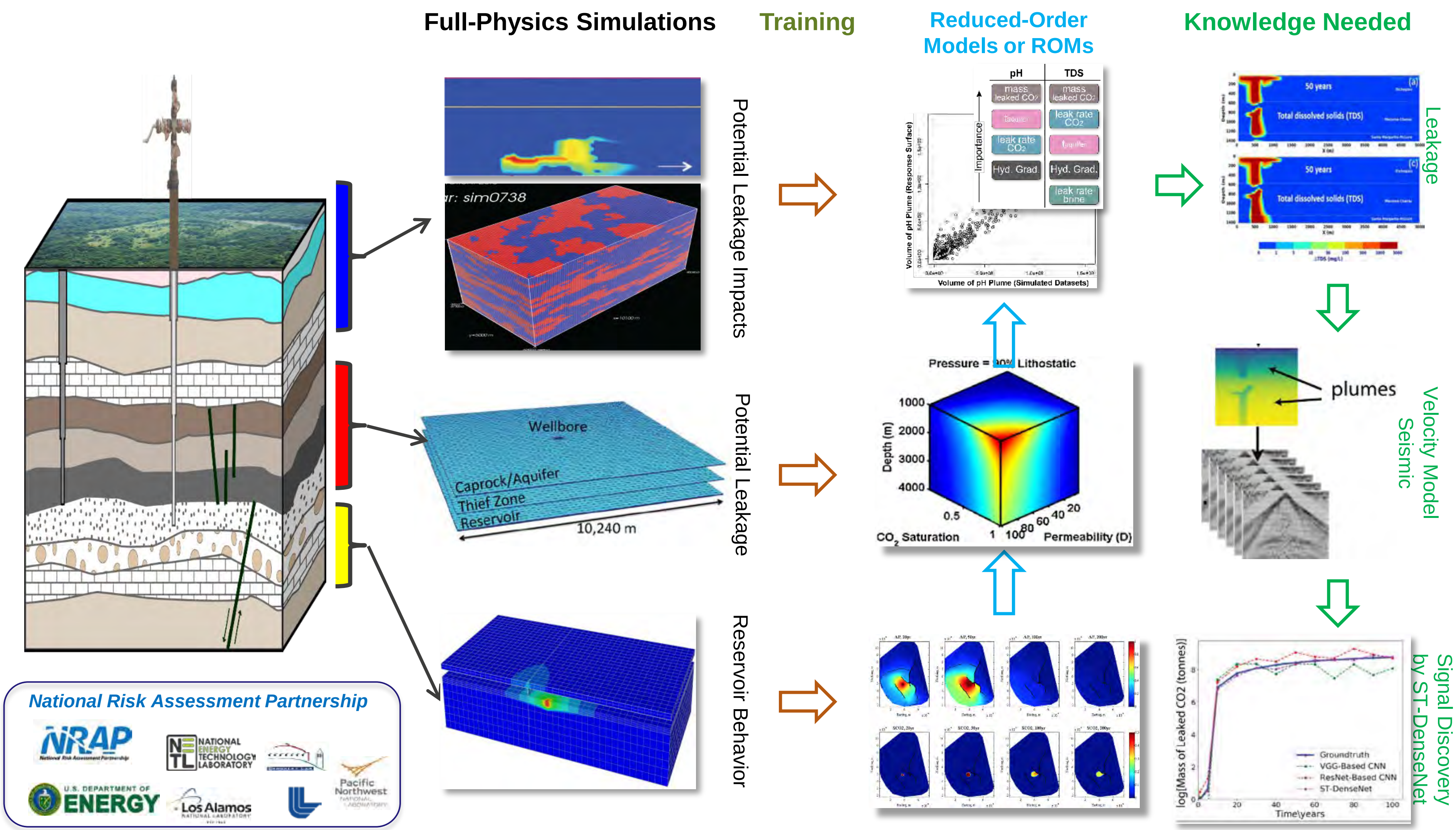


NRAP's approach is to fuse physic-based prediction with empirical models to describe complex system behavior…

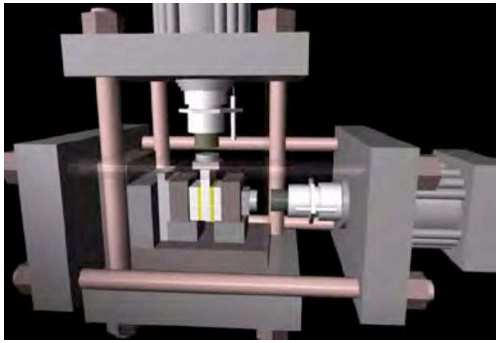


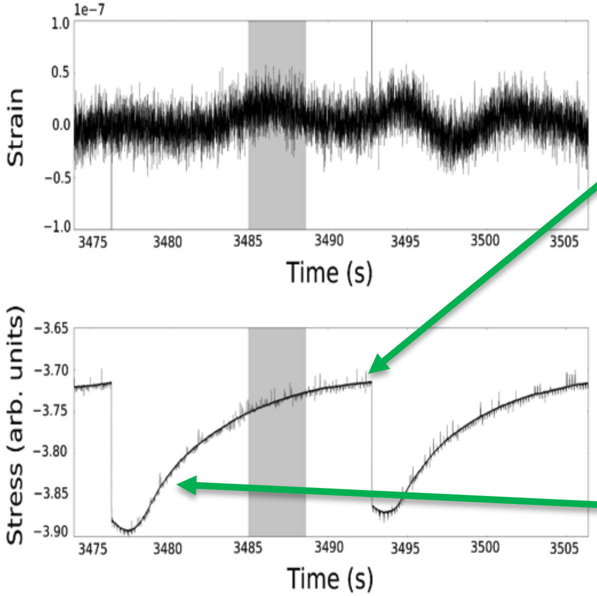
National Risk Assessment Partnership

…and we can now use system behavior to discover signals that can be used to monitor for leaks.



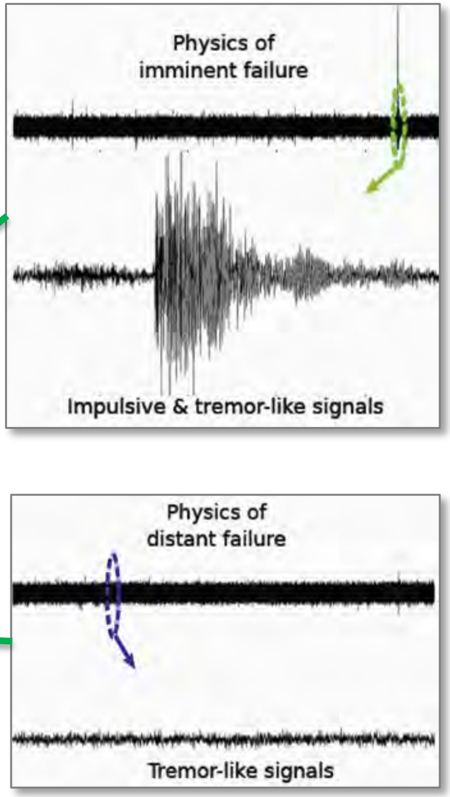
"Earthquake machine" is being used to probe for predictive signatures on state of stress using random forest methods.



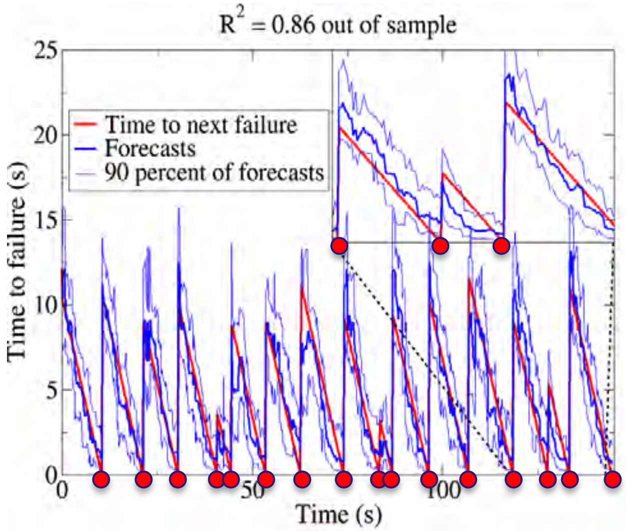


Experimental Data on Slip

Acoustic Emission Signal



Time to Failure Forcasted from Acoustic Emissions



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

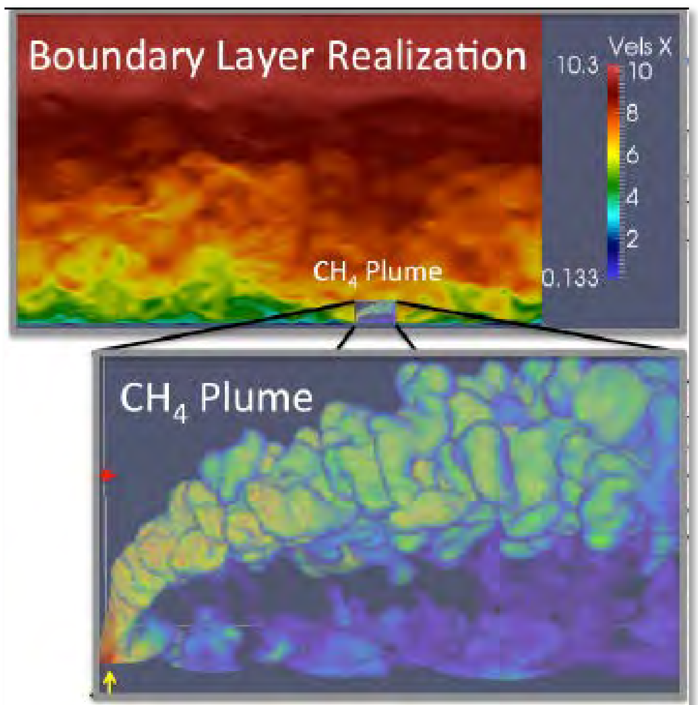
Rouet-LeDuc, B., C. Hulbert , N. Lubbers , K. Barros, and P. Johnson, Learning the Physics of Failure, *Geophys. Res. Lett.*

PennState

CFD simulations can be used to pre-train a neural net to recognize a signal prior to direct field experience.

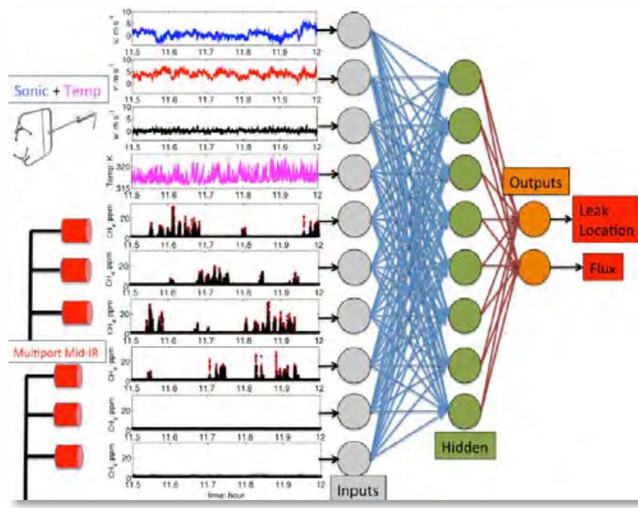
Example: Using computational fluid dynamics simulations to pre-train an artificial neural network (ANN) coupled to a CH_4 sensor and a meteorological tower for detection of NG leak.

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)

Dependent variables: Leak location; NG flux

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

Intersection of ML and physics can enable applications that neither alone can address.