Development of a Framework for Data Integration, Assimilation, and Learning for Geological Carbon Sequestration (DIAL-GCS)

Project #: DE-FE0026515

Alex Sun, Ph.D., P.E.

Bureau of Economic Geology

University of Texas at Austin

U.S. Department of Energy

National Energy Technology Laboratory

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

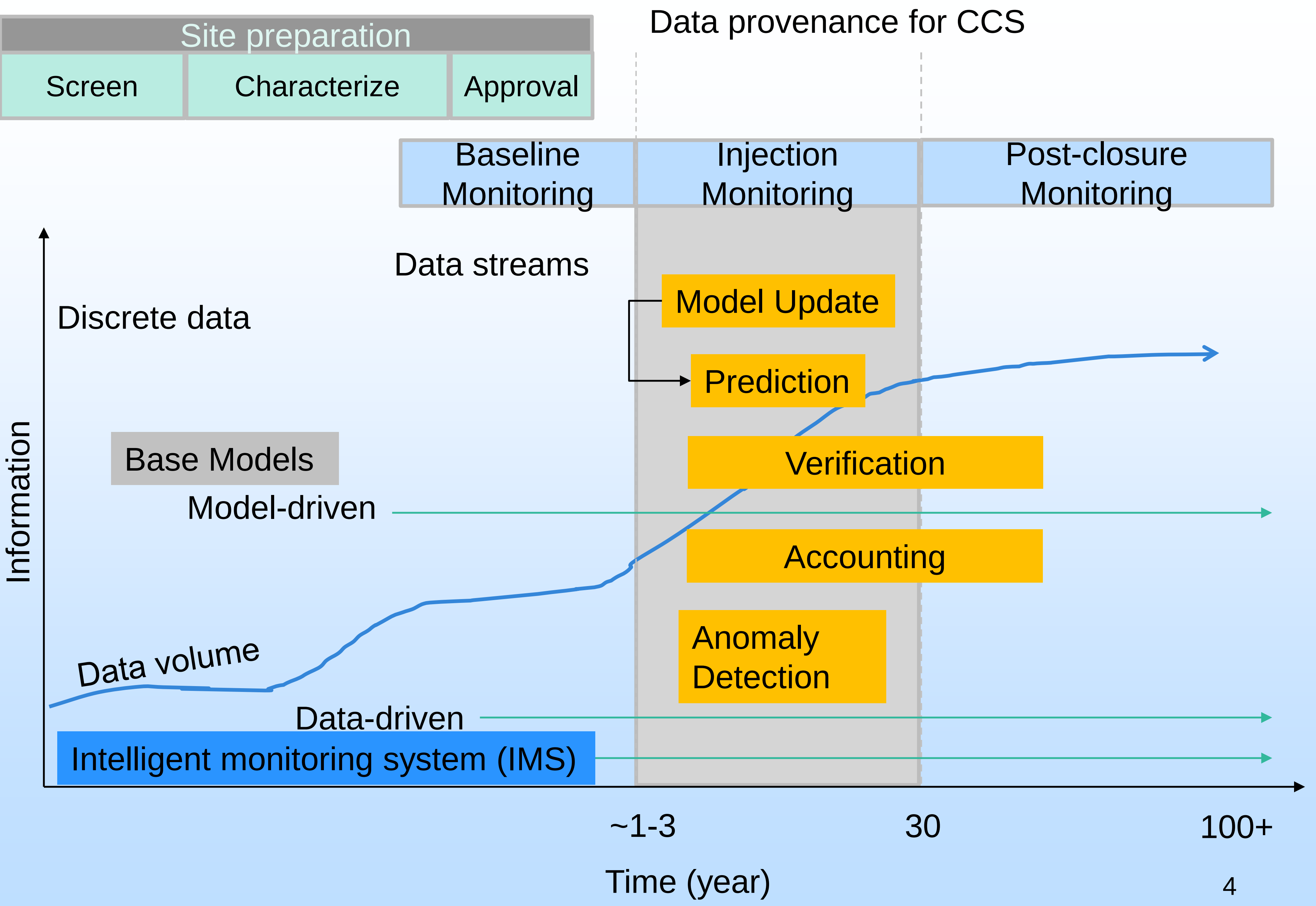
August 1-3, 2017

Presentation Outline

- Technical Status
- Accomplishments to Date
- Lessons Learned
- Synergy Opportunities
- Project Summary

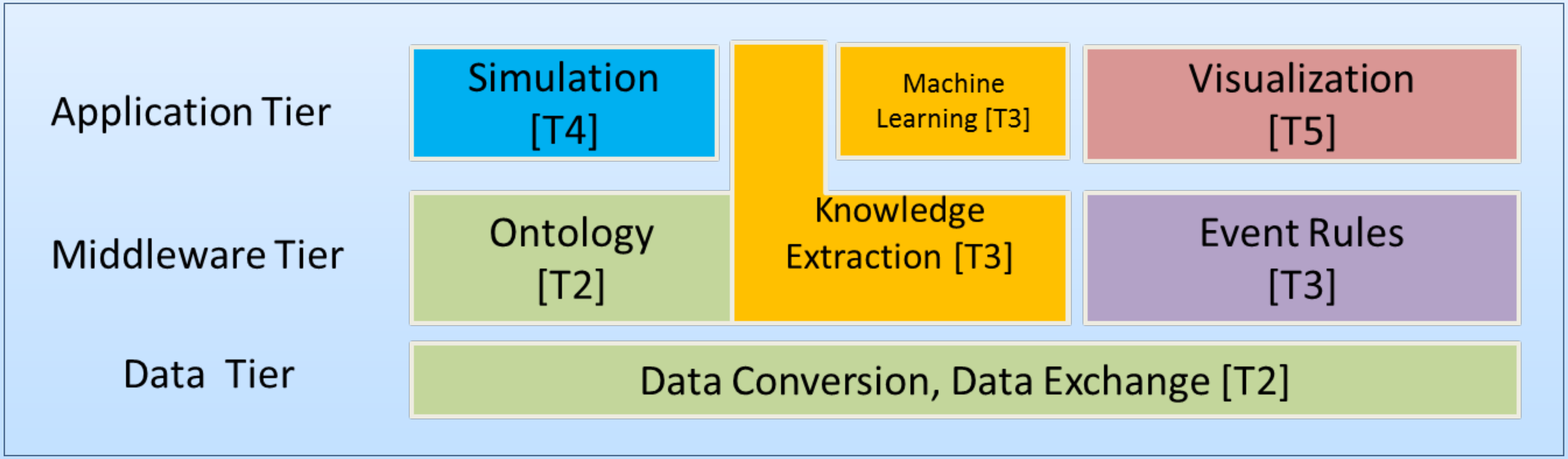
Technical Status

- Overview
- Data-driven modeling
- Process-based modeling
- Integration


Site preparation
Screen
Characterize
Approval
Data provenance for CCS
Baseline Monitoring
Injection Monitoring
Post-closure Monitoring
Data streams
Discrete data
Model Update
Prediction
Base Models
Verification
Model-driven
Accounting
Information
Data volume
Anomaly Detection
Data-driven
Intelligent monitoring system (IMS)
~1-3
30
100+
Time (year)

DIAL Project Overview

Task 2: Sensor data schema development and provisioning(Y1)
Task 3: Development of CEP, machine learning (Y1-2)
Task 4: Coupled modeling, UQ, and data assimilation (Y1-4)
Task 5: System integration and demonstration (Y1-4)

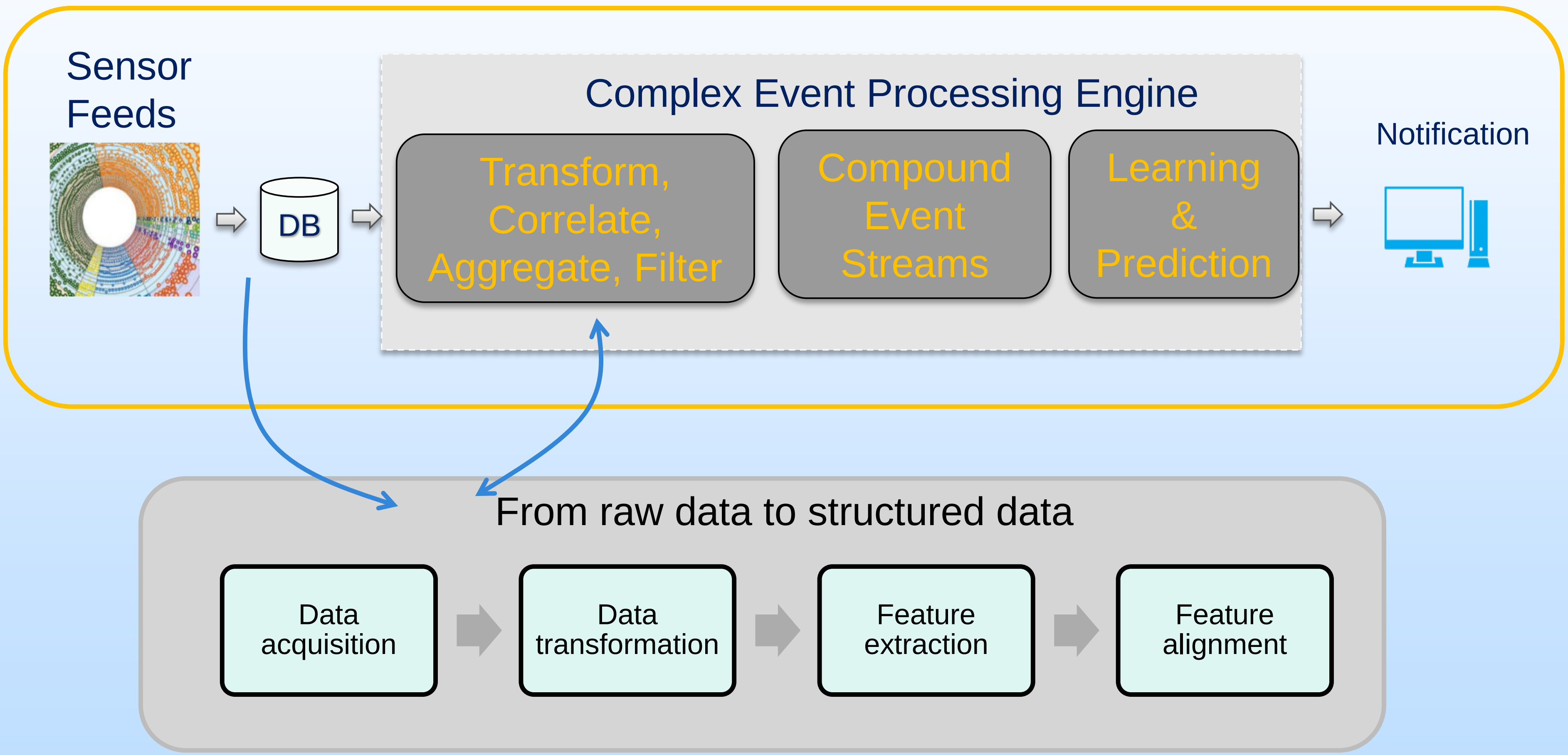


Top level architecture

Data-Driven Anomaly Detection

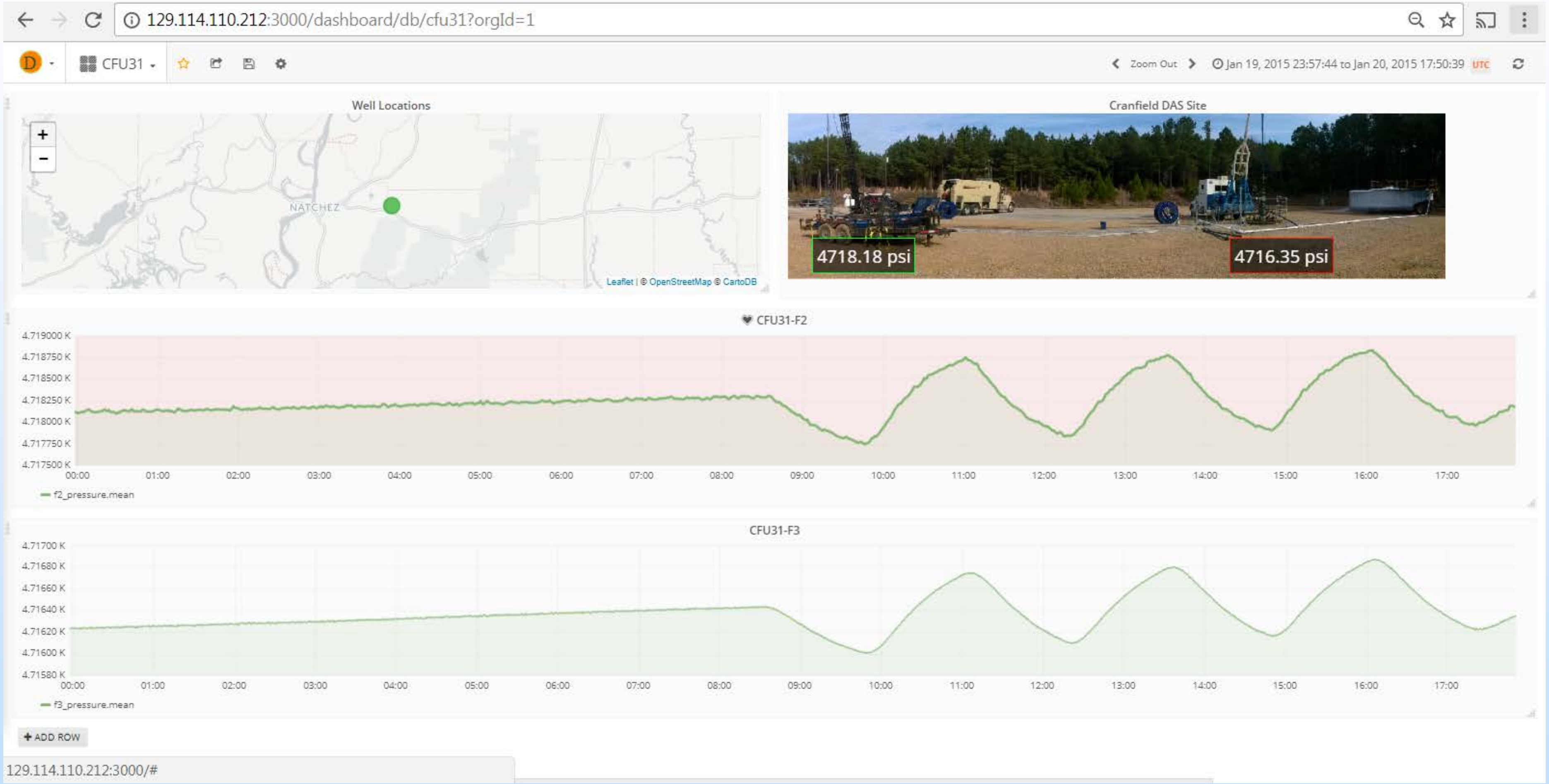
- Suitable for
  - Continuous monitoring
  - Automated anomaly detection
  - When high-fidelity process-level model is not available
- Requirements
  - High-performance, integrated computing infrastructure
  - Effective online machine learning algorithms
  - Training data and expert insights!

Sensor Data Processing



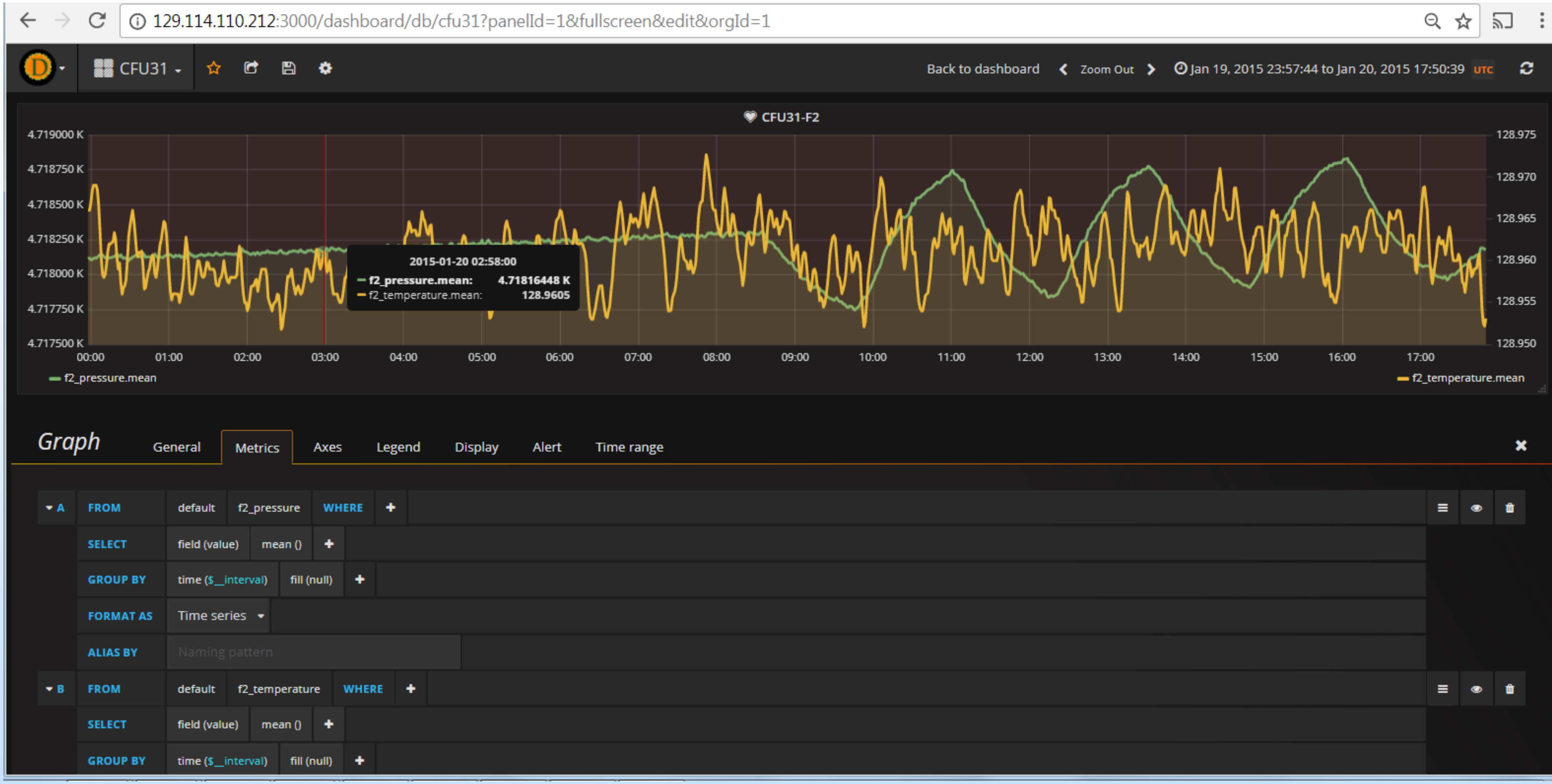
Built on top of high-performance database and dashboard

A configurable web dashboard



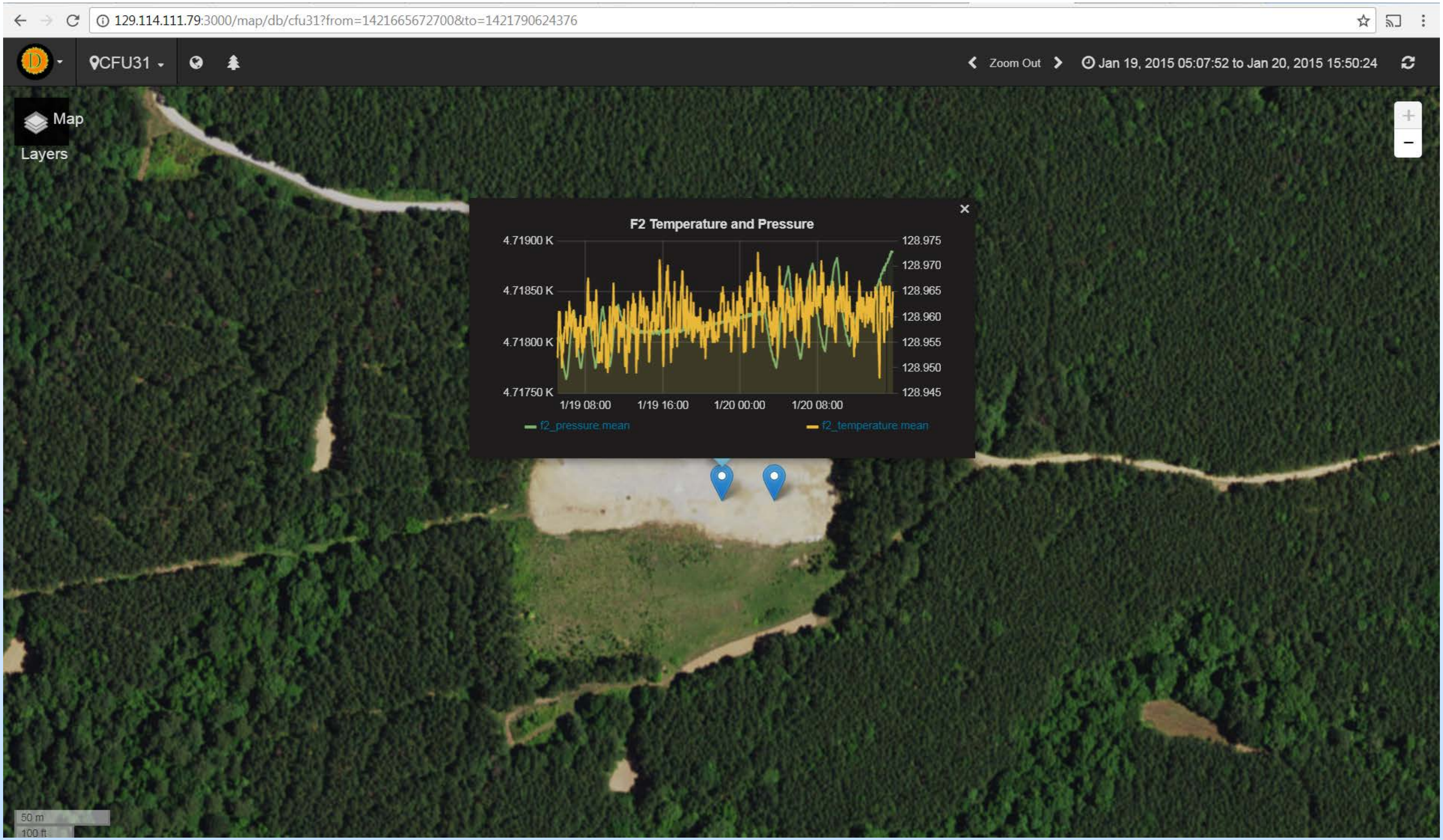
Data collected during simulated leakage field test at Cranfield, MS
Details of the experiment will be shown in my presentation in the afternoon

Time Series Graph Interface



Versatile time series plot helps to understand data better

Alternative Map View

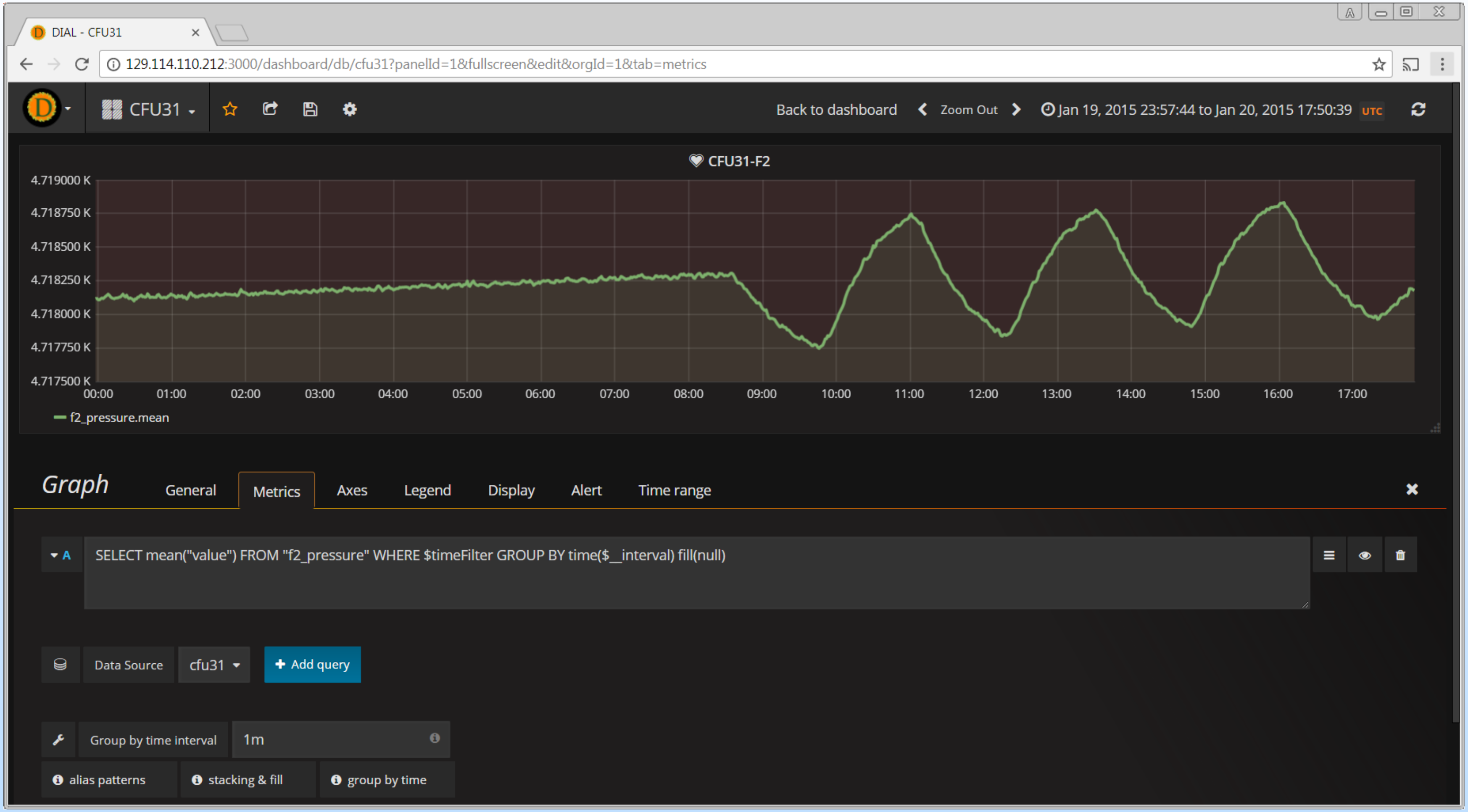


Defining Intelligence in DIAL

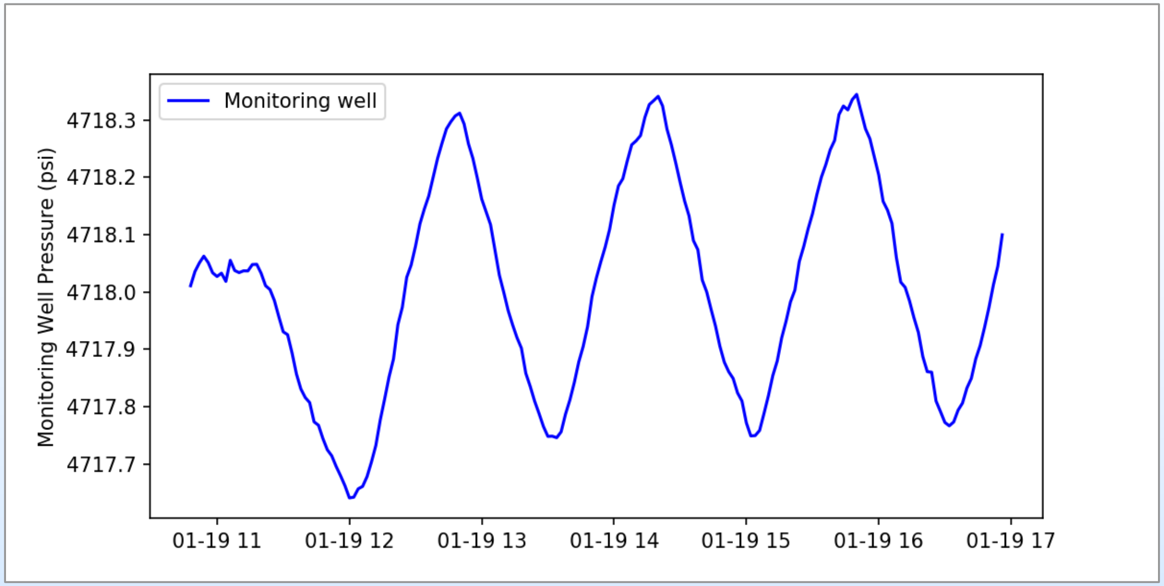
Types of problems

- Diagnosis (after the fact)
  - Credit card fraud
  - Medical imaging
  - Manufacturing process
  - Leakage
- Prognosis (before the fact)
  - Natural hazard (flooding, drought)
  - Cancer prevention procedures
- Categories of detection techniques
  - Unsupervised learning (outlier detection, training set unlabeled)
  - Semi-supervised learning (training set labeled for normal points)
  - Fully supervised (training set labeled for normal and anomalies)
- Efficacy of algorithms can be domain specific

Pressure-based monitoring: How do we locate anomalies?

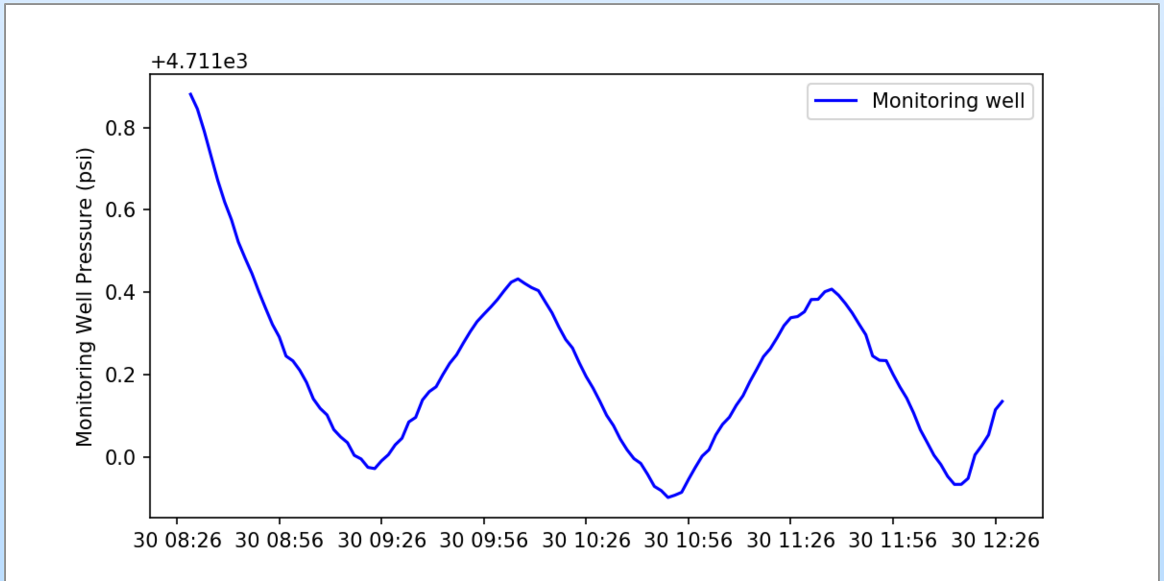


Single Sensor Data (BHP Only)



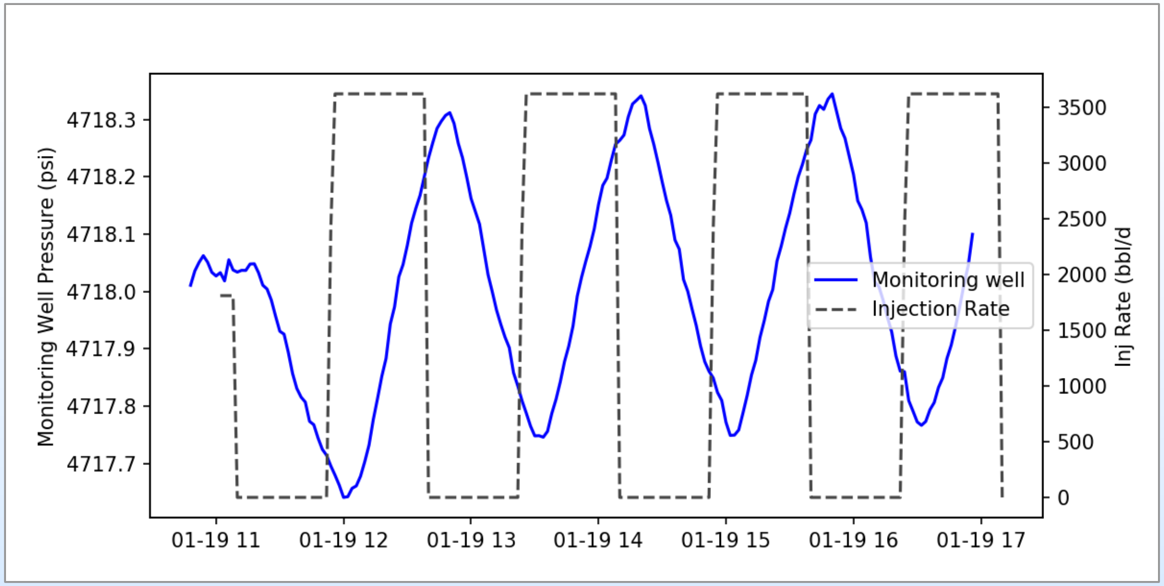
Baseline

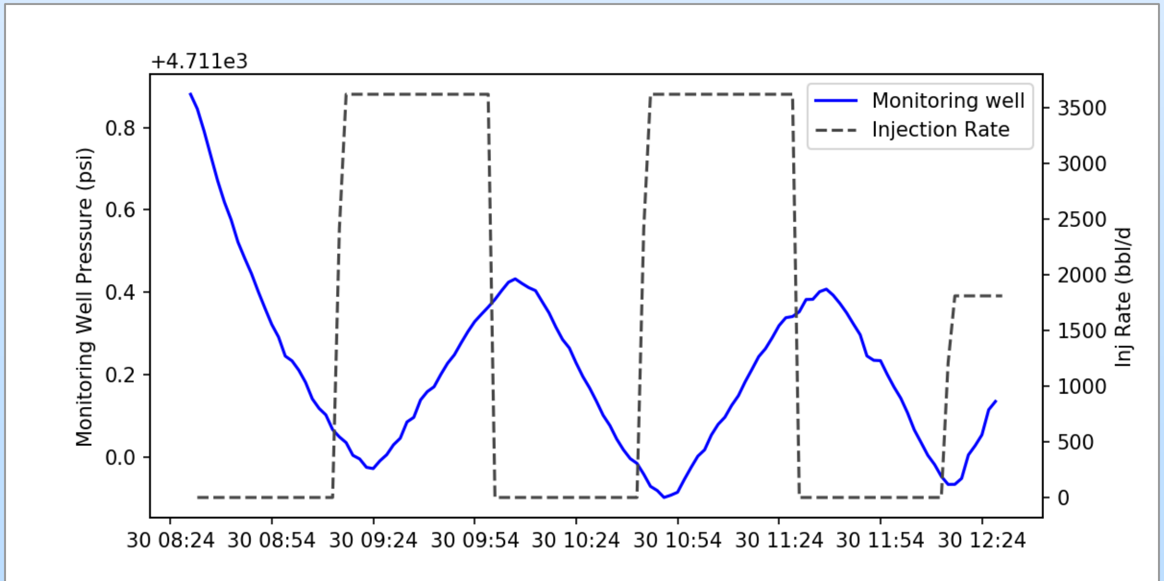
Can we see leaks?



Adding Covariant Information

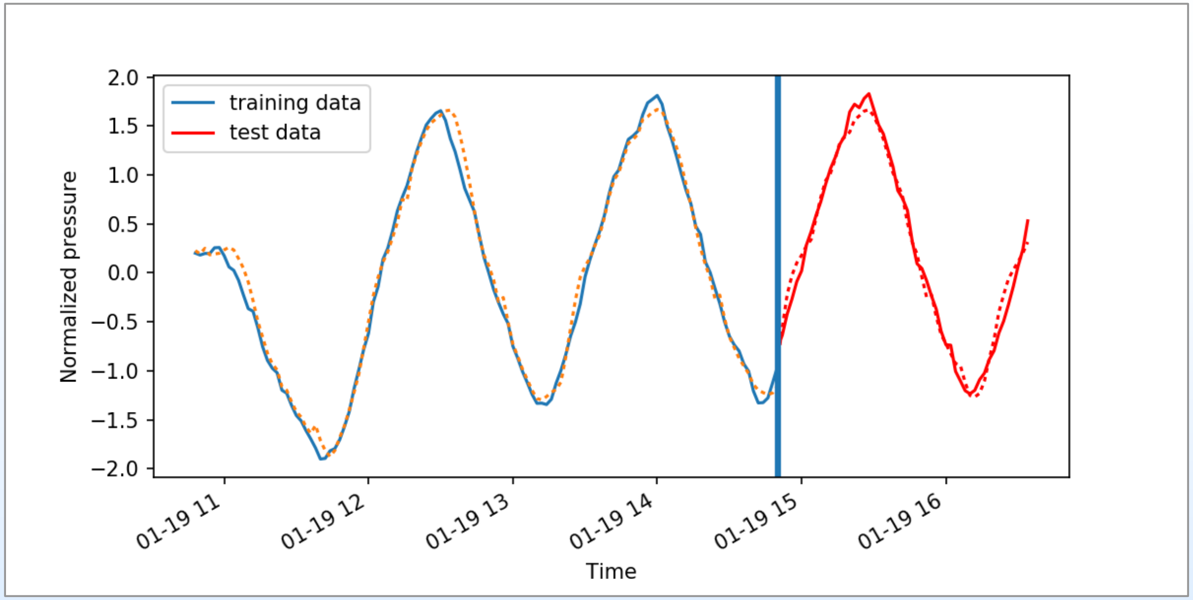
BHP &
Injection Rate





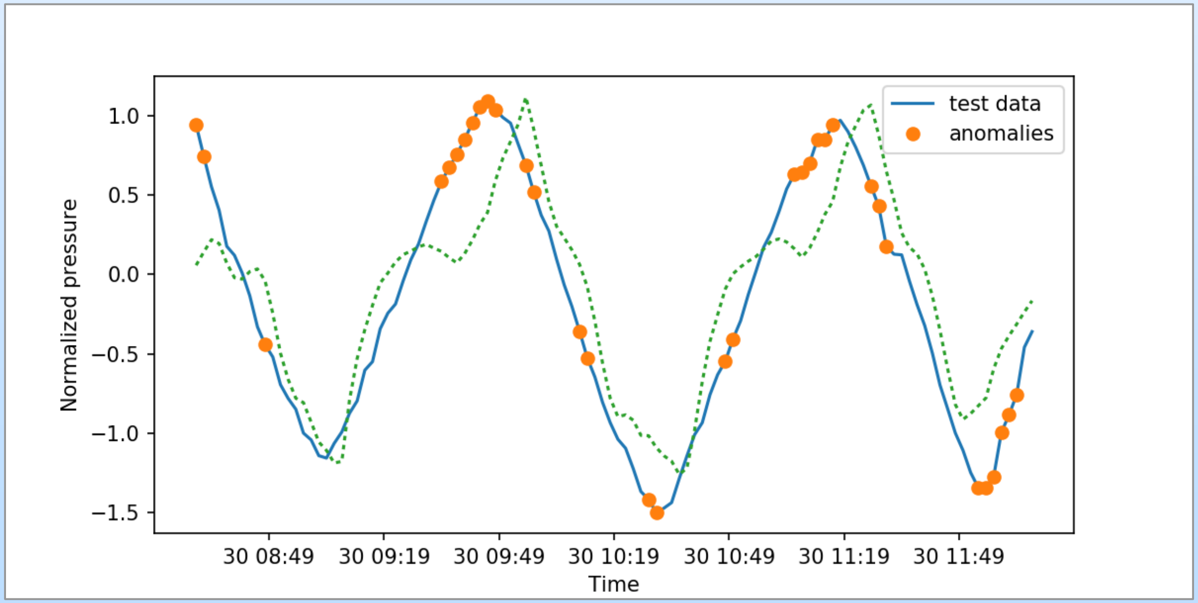
Learn the cause-effect relationship between injector and observer

Training of predictive model (dashed line) using data from baseline experiment



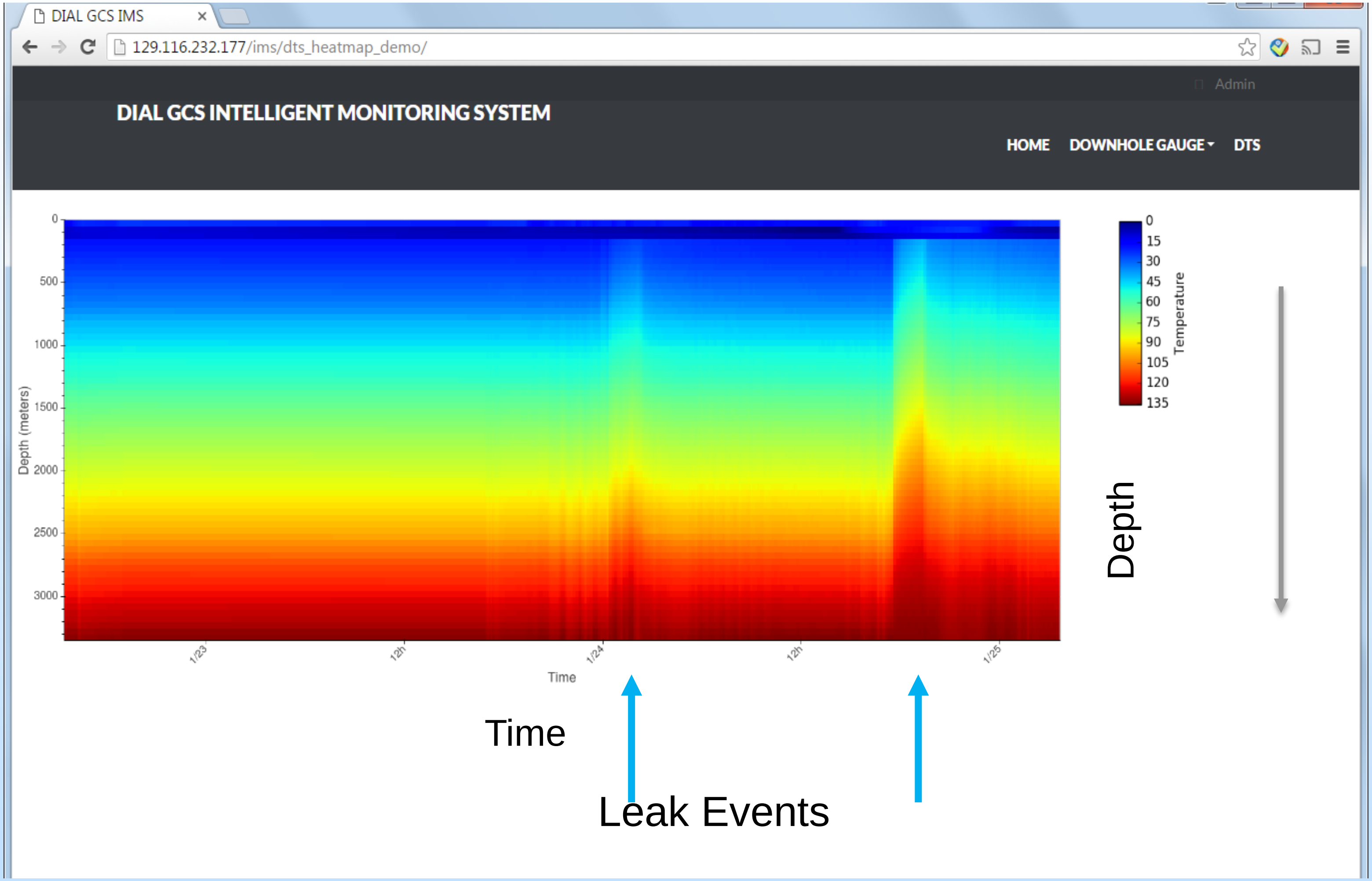
No anomalies

Prediction



Leaks caused phase and magnitude deviations from the predictive model

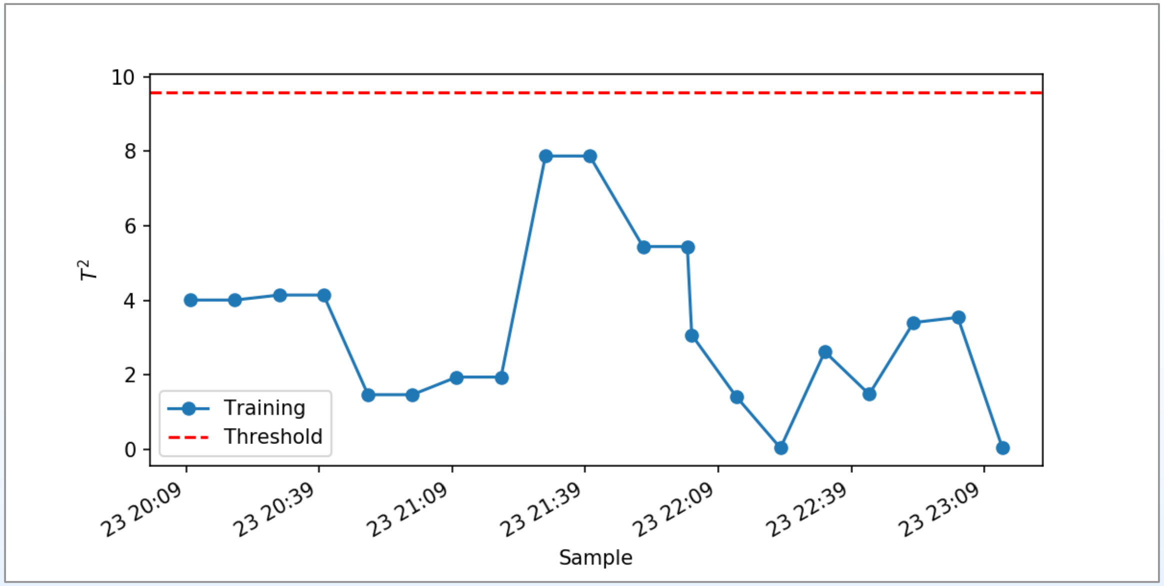
Distributed Sensor



Sensor data need to be compressed

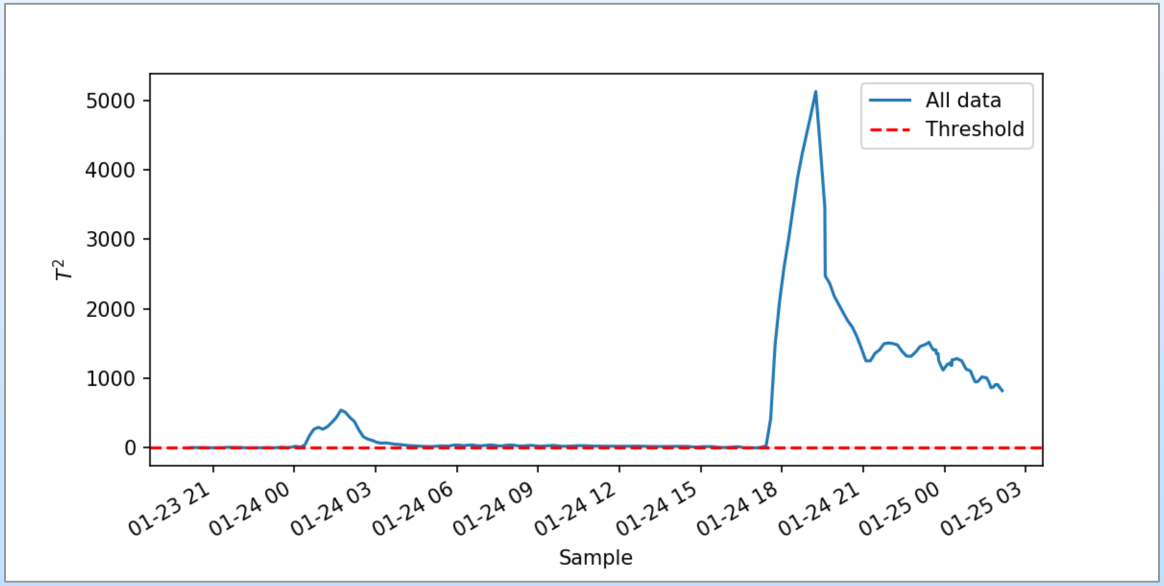
Data courtesy: Barry Freifeld, LBL

Training
a statistic



Using PCA to extract main features to calculate test statistics

Detection



Combing dimension reduction with anomaly detection

Threshold alpha = 0.05

Automating Anomaly Detection

Time

Task Creation

Collectors

Define sensors…

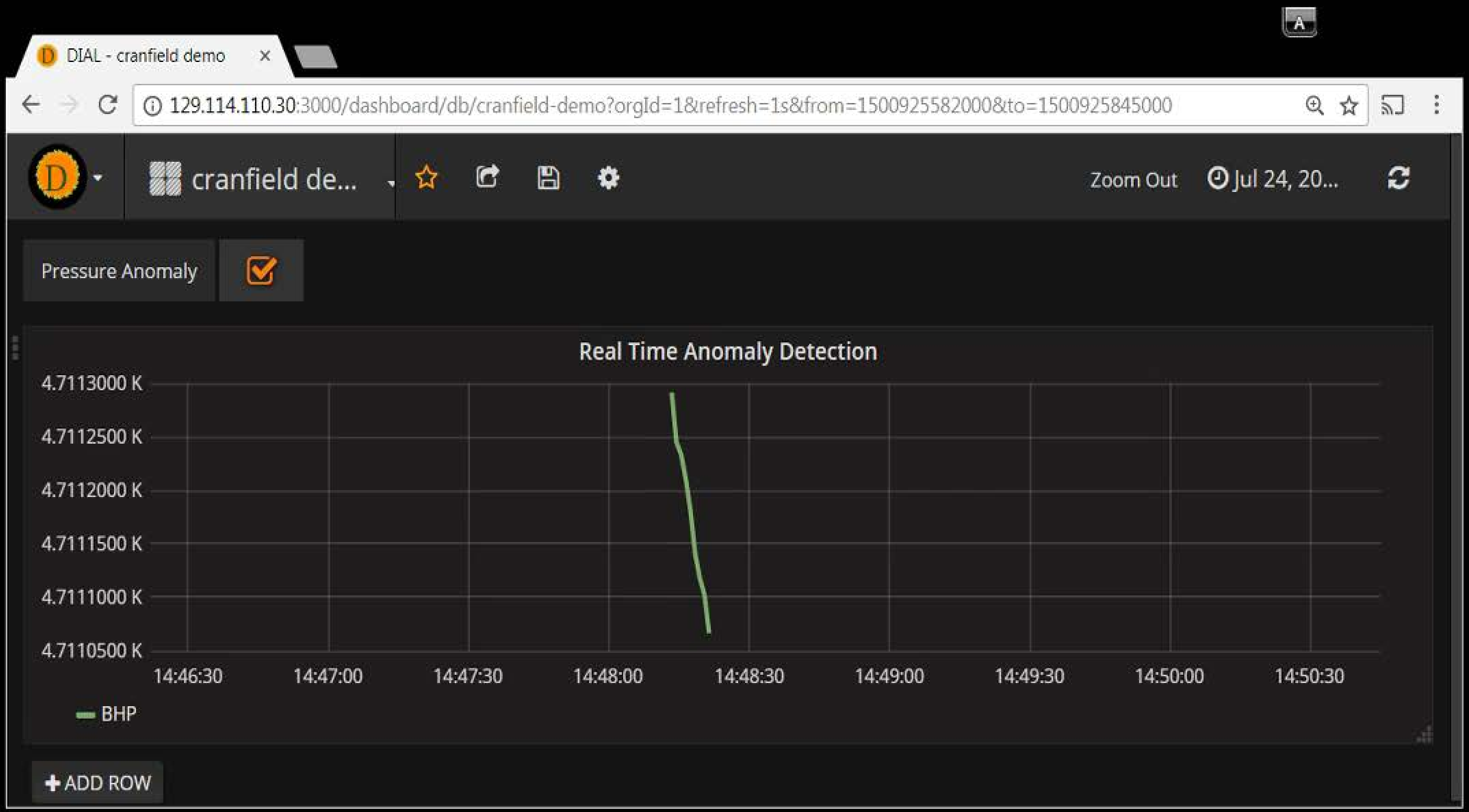
Processors

Define AD algorithm

Publishers

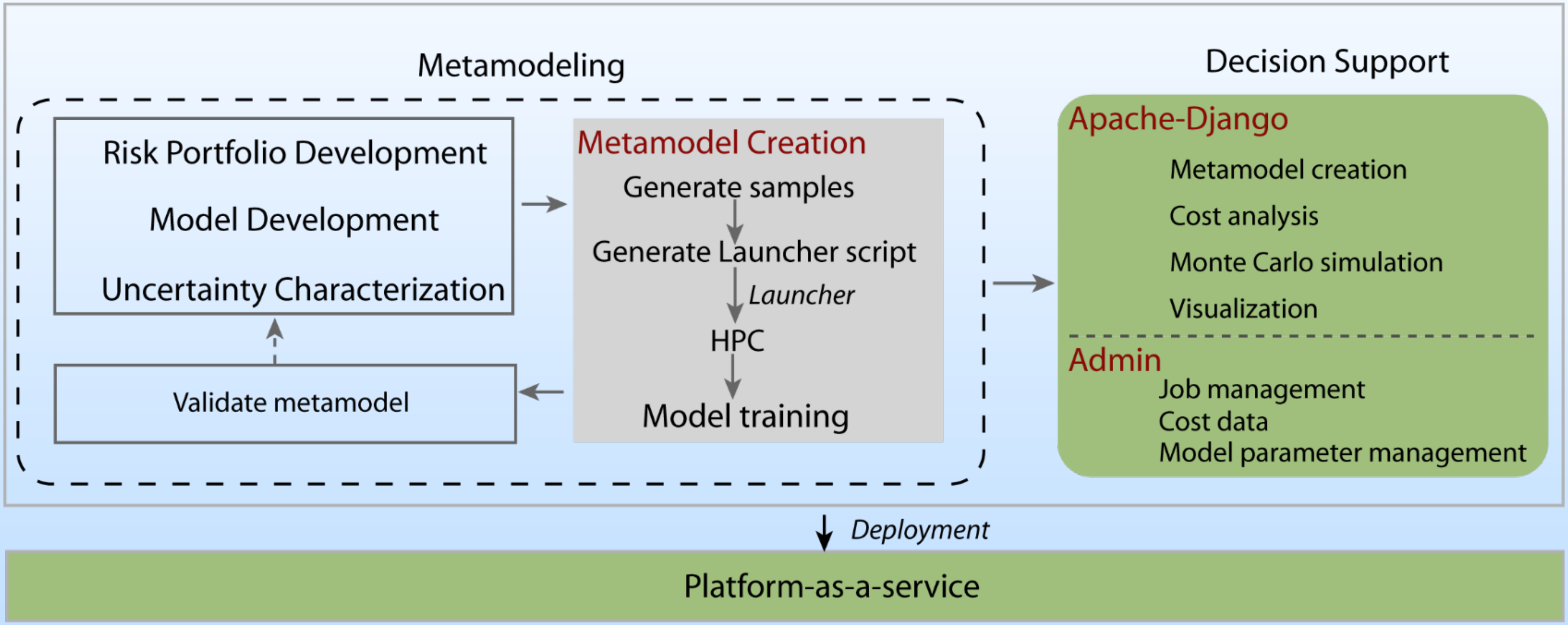
Define recipients, e.g., dashboard, subscribers

A scalable, distributed architecture


DIAL - cranfield demo
129.114.110.30:3000/dashboard/db/cranfield-demo?orgId=1&refresh=1s&from=1500925582000&to=1500925845000
cranfield de...
Zoom Out
Jul 24, 20...
Pressure Anomaly
Real Time Anomaly Detection
4.7113000 K
4.7112500 K
4.7112000 K
4.7111500 K
4.7111000 K
4.7110500 K
14:46:30
14:47:00
14:47:30
14:48:00
14:48:30
14:49:00
14:49:30
14:50:00
14:50:30
BHP
ADD ROW

Process Modeling

Case I: Assessing economic impact of leakage during planning



Combining physical modeling, metamodeling, and web decision support

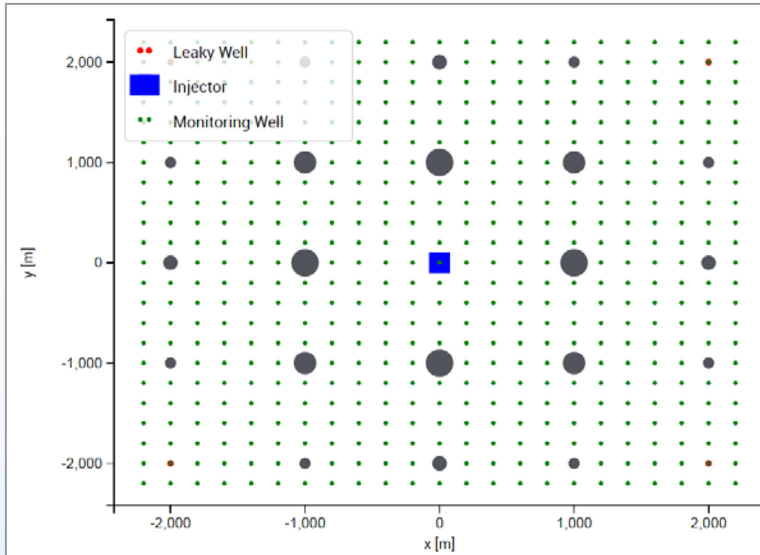
DIAL Project

Home | ReducedModel

Injection Zone Parameters

CO2 density (kg/m3):	479.0
Brine density (kg/m^3):	1045.0
CO2 viscosity (Pa*s):	3.95e-05
Brine viscosity (Pa*s):	0.0002535
Formation Compressibility (1/Pa):	4.2e-10
Residual brine saturation:	0.15
Endpoint co2 relative perm:	0.55
Thickness (m):	10.0
Permeability (m^2):	5e-14
Porosity (-):	0.1
Injection Rate [Mt/yr]:	1.0

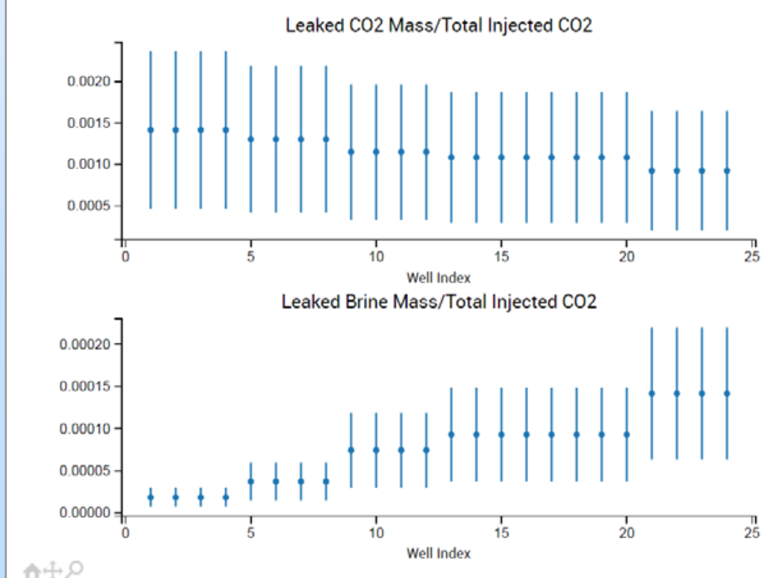
Above Zone Parameters



Total leaked CO2 = 0.61 Mt Total leaked Brine = 0.08 Mt

Estimated damage = $ 5.62MM

Monte Carlo Statistics



Case 2: Optimal Monitoring Network Design

Objective Function

Well cost =
CAPEX($/well)+
OPEX($/well/day)+
Intervention($/well)

Leakage cost =
Brine($/ton)+
CO_2($/ton)

Optimization toolbox

Binary Integer Programming

- Linear problem
- Convex

Optimize monitoring network

Constraints

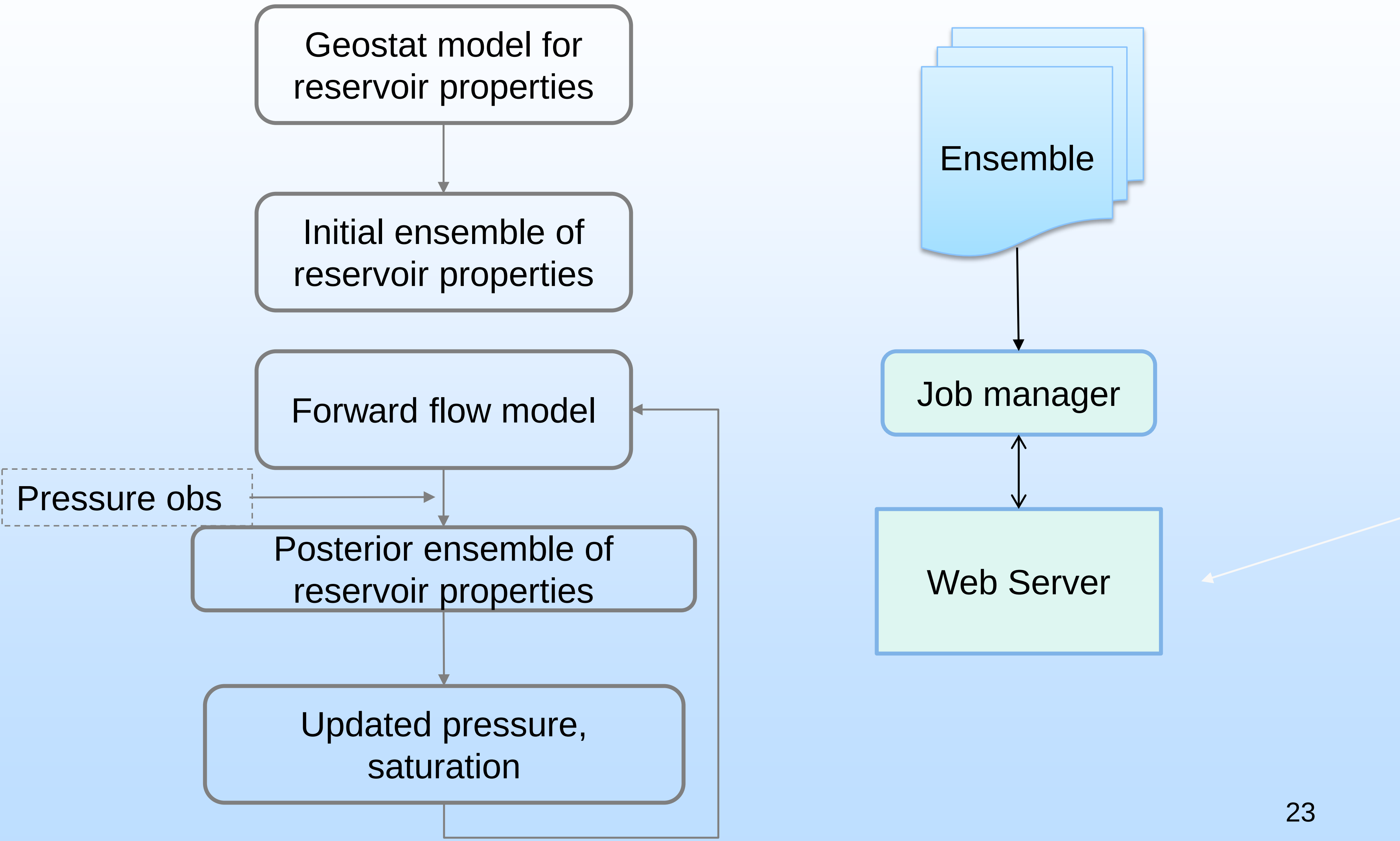
of monitoring wells ≤ N_{max}

CO_2 leakage ≤ M% of total injected CO_2

ΔP at $t_{leakage\ detection} \geq \Delta P_{threshold}$

Please visit Hoonyoung Jeong's poster

Case 3: Data Assimilation



Ensemble Data Assimilation @ 300 days

A pulse experiment was done in first 7 days to stimulate reservoir

Task 5 Integration and Deployment

- Integration is being continuously performed
- Test and production servers on UT’s Cloud platforms

Lessons Learned

- No off-the-shelf or fixed combination of algorithms will fit all possible GCS datasets
  - Best practice should be developed
  - Establish a repository of data for testing algorithms

- Data-driven models are suitable for predictive modeling if the predictors (causal factors) can be correctly identified

- Properly designed site experiments will be highly valuable for developing predictive models

Accomplishments to Date

- Task 2: Data management
  - Year 1: Developed schema and data adaptors for storing, exchanging information, and visualizing information
- Task 3: Complex event processing using machine learning
  - Year 2: Implemented predictive models on different test datasets
- Task 4: Coupled modeling / data assimilation
  - Year 2: Implemented workflow for automating data assimilation. Demonstrated Web-based modeling approaches
- Task 5: Integration and demonstration
  - Year 1-2: Experimented with a large number of web-based technologies for making the system more user friendly

Synergy Opportunities

- A flexible platform has been developed
- We are always interested in collaboration and testing different datasets

Project Summary

- Developed prototype of DIAL system

- Next Steps
  - Formalize data transformation and work flow
  - Enable web service for data assimilation and well placement studies
  - Experiment with different data-driven models and data types
  - Provide useful web services

Appendix

– These slides will not be discussed during the presentation, but are mandatory.

Benefit to the Program

- Carbon storage program goals being addressed

Develop and validate technologies to ensure 99 percent storage permanence

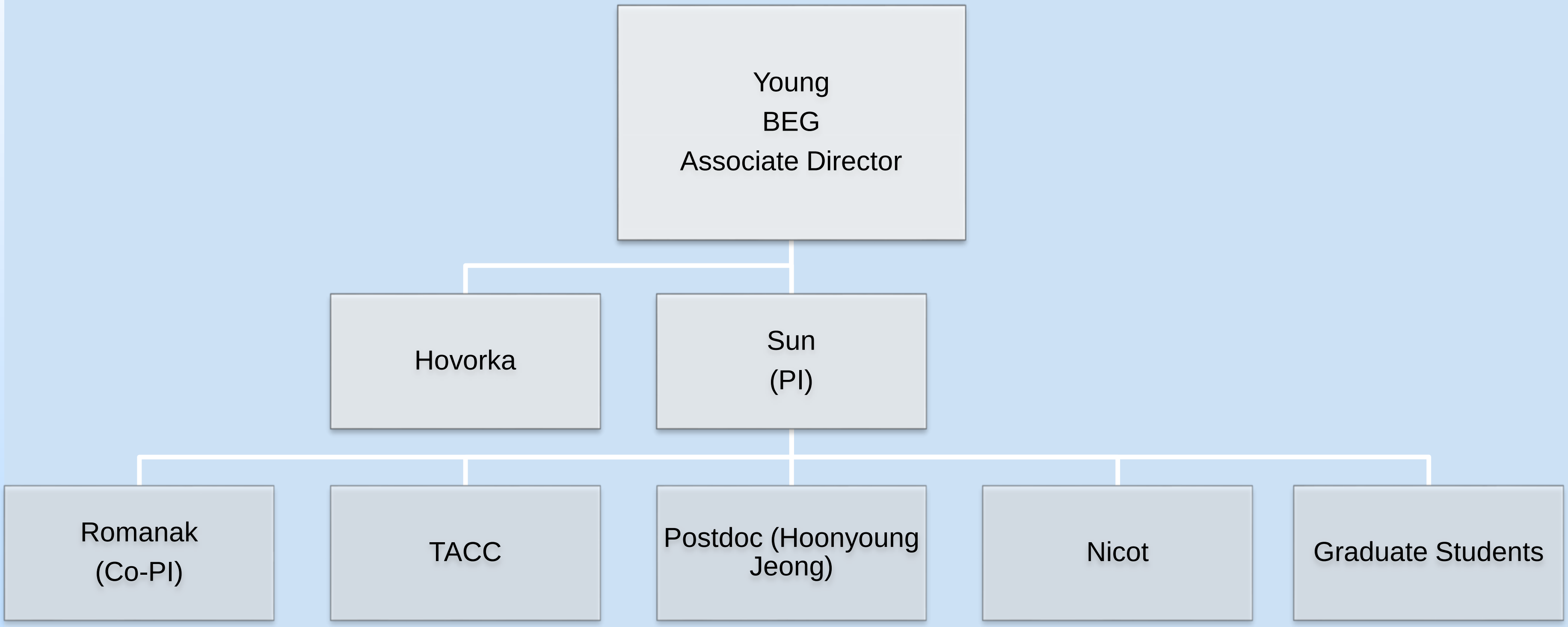
- Expected benefits of this IMS Project
  - Transform scientific knowledge to decision power and public knowledge
  - Promote data sharing and visual analytics
  - Better collaboration among team members
  - Public outreach
  - Streamline CCS data management and decisionmaking
  - Facilitate the optimal allocation of monitoring resources

Project Overview

Goals and Objectives

- Develop GCS data management module for storing, querying, exchanging, and visualizing GCS data from multiple sources and in heterogeneous formats
  - Success Criterion: Whether a flexible, user-friendly Web portal is set up for enabling data exchange and visual analytics
- Incorporate a complex event processing (CEP) engine for detecting abnormal situations by seamlessly combining expert knowledge, rule-based reasoning, and machine learning
  - Success Criterion: Whether a set of decision rules are developed for identifying abnormal signals in monitoring data
- Enable uncertainty quantification and predictive analytics using a combination of coupled-process modeling, data assimilation, and reduced-order modeling
  - Success Criterion: Whether a suite of computational tools are developed for UQ and predictive analytics
- Integrate and demonstrate the system's capabilities with both real and simulated data
  - Success Criterion: Whether the IMS tools developed under Goals A to C are integrated, streamlined, and demonstrated for a realistic GCS site

Organization Chart



Gantt Chart

Table 2. Project Gantt chart
(Numbers in table rows indicate milestones).
(Phase I ; Phase II)

Task	Description	Year 1				Year 2				Year 3			
		1	2	3	4	1	2	3	4	1	2	3	4
1	Update project management plan	1											
2	Sensor data management												
2.1	Ontology/schema development												
2.2	Sensor data adaptor development		2										
3	CEP Development												
3.1	Rule definition												
3.2	Reasoning and machine learning												
3.3	Testing					3							
4	Coupled modeling/Assimilation												
4.1	Coupled modeling												
4.2	Data assimilation						4						
5	Integration and demonstration												
5.1	Integration											5	
5.2	Demonstration												
6	Synthesis of results												
6.1	Dissemination of results												
6.2	Technology transfer												6

Bibliography

- Manuscripts
  - Jeong, H., A. Sun, and X. Zhang, Cost-optimal design of pressure-based monitoring networks for carbon sequestration projects, with consideration of geological uncertainty. Submitted to International Journal of Greenhouse Gas Control
  - Sun, A., H. Jeong, A. Gonzalez, T. Templeton, Metamodeling-based approach for risk assessment and cost estimation: Application to geological carbon sequestration. Submitted to Computers & Geosciences.
- Presentations
  - Development of an intelligent monitoring system for geological carbon sequestration projects, presented at the fall meeting of American Geophysical Union, San Francisco, CA, December, 2016

Needs for Intelligent Systems in Geosciences

Data collection

Data integration

Data analysis

Data processing

Visualization

Intelligent and Information Systems for Geosciences, NSF, 2015