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
Presentation Outline

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

- Overview
- Data-driven modeling
- Process-based modeling
- Integration
Screen
Characterize
Approval

Site preparation

Baseline Monitoring
Injection Monitoring
Post-closure Monitoring

Data streams

Base Models
Model-driven

Data volume

Data-driven

Intelligent monitoring system (IMS)

Model Update
Prediction
Verification
Accounting
Anomaly Detection

Data provenance for CCS

Discrete data

Information

Time (year)

~1-3 30 100+ 100+

Screen Baseline Injection Post-closure
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

Sensor Feeds → DB → Complex Event Processing Engine

- Transform, Correlate, Aggregate, Filter
- Compound Event Streams
- Learning & Prediction

From raw data to structured data

Data acquisition → Data transformation → Feature extraction → Feature alignment

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?
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

Leaks caused phase and magnitude deviations from the predictive model
Distributed Sensor

Sensor data need to be compressed

Data courtesy: Barry Freifeld, LBL
Combing dimension reduction with anomaly detection

Threshold alpha = 0.05

Using PCA to extract main features to calculate test statistics
Automating Anomaly Detection

A scalable, distributed architecture
Case I: Assessing economic impact of leakage during planning

Combining physical modeling, metamodeling, and web decision support
Injection Zone Parameters

- **CO₂ density (kg/m³):** 479.0
- **Brine density (kg/m³):** 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 (m²):** 5e-14
- **Porosity (-):** 0.1
- **Injection Rate [Mt/yr]:** 1.0

Above Zone Parameters

Monte Carlo Statistics

- **Total leaked CO₂:** 0.61 Mt
- **Total leaked Brine:** 0.08 Mt
- **Estimated damage:** $6.62 MM
Case 2: Optimal Monitoring Network Design

Objective Function

Well cost = \text{CAPEX}($/\text{well}) + \text{OPEX}($/\text{well/day}) + \text{Intervention}($/\text{well})

Leakage cost = \text{Brine}($/\text{ton}) + \text{CO}_2($/\text{ton})

Optimization toolbox

Binary Integer Programming
- Linear problem
- Convex

Optimize monitoring network

Constraints

\# of monitoring wells \leq N_{\text{max}}

\text{CO}_2\text{ leakage} \leq M\%\text{ of total injected CO}_2

\Delta P \text{ at } t_{\text{leakage detection}} \geq \Delta P_{\text{threshold}}

Please visit Hoonyoung Jeong’s poster
Case 3: Data Assimilation

- Geostat model for reservoir properties
- Initial ensemble of reservoir properties
- Forward flow model
- Posterior ensemble of reservoir properties
- Updated pressure, saturation
- Pressure obs

- Ensemble
- Job manager
- Web Server
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

Young
BEG
Associate Director

Hovorka

Sun
(PI)

Romanak
(Co-PI)

TACC

Postdoc (Hoonyoung
Jeong)

Nicot

Graduate Students
# Gantt Chart

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

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

– 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