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
The University of Texas at Austin
Presentation Outline

- Technical Status
  - Background and system design
  - Online anomaly detection using machine learning
  - Predictive modeling
- Accomplishments to Date
- Lessons Learned
- Synergy Opportunities
- Project Summary
Task 2: Sensor data schema development and provisioning (Y1)
Task 3: Development of CEP, machine learning (Y1-3)
Task 4: Coupled modeling, UQ, and data assimilation (Y1-4)
Task 5: System integration and demonstration (Y1-4)
Task 3. Complex Event Processing

Sensor Feeds -> DB -> Complex Event Processing Engine

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

Notification

From raw data to structured data

Data acquisition -> Data transformation -> Feature extraction -> Feature alignment

Feature engineering
Data-Driven, ML-Enabled Anomaly Detection

• Machine learning (ML) is suitable for
  – Continuous monitoring
  – When physical process is not fully understood
  – Automated anomaly detection

• Requirements
  – Effective online ML algorithms
  – Labeled training data and expert insights!
  – High-performance, integrated computing infrastructure
Loosely Coupled Real-Time Data Processing System

- Streaming Platform
- Data archive
- Real-time
- DB
- ML Detectors
- Coupled Models
- Dashboard
- Kafka connectors
Anomaly Detection Case Study

Cranfield, MS experiments

Dataset include **Pressure and Temperature** measurements from
- Base pulse testing experiments (no known leaks)
- Controlled CO₂ release experiments (artificial leaks)
Anomaly Detection in Pressure Data

Use IsolationForest to perform unsupervised classification

- Randomly and recursively partition of data (trees)
- Calculate path length from root node to leaf node
- Use **tree path length** as a measure of normality
- Outliers correspond to shorter trees
Anomaly Detection in DTS Data

DTS Data Training Testing

Calculate $T^2$ statistic

- Form normalized training sample matrix
- Perform PCA (data compression) on the sample matrix
- Calculate $T^2$ statistic for a user-defined significance level

Sun et al., Env Modeling & Software, 2019
Zhong et al., 2019b
DIAL-GCS 1.0

Design 1.0:
- Web GIS
- Time series management
- A lot custom coding
DIAL-GCS 2.0

Design 2.0:
- Kafka-based
- Flexible
Task 4. Deep Learning Based Surrogate Modeling

- Deep learning (DL) is a very powerful tool for pattern recognition.
- In geosciences, there’s a lot of hype on DL but also many questions
- We developed an innovative DL pipeline for combining DL with physics-based models

A generic simulation/inversion framework
Deep learning based surrogate models for CO₂ plume prediction

Input: Permeability field
Output:
\( S_g \): Actual CMG-GEM result
\( \hat{S}_g \): Surrogate model results

Once trained, our model can be used to predict CO2 plume for given input properties and at any time.
Validation: Comparison with Monte Carlo Simulation
Dynamic mapping: Prediction of CO$_2$ Plume Movement

Zhong et al., 2019a
Combining with Geophysics Data

Saturation ↔ Acoustic Impedance

Zhong et al., under review
Web-Based Monitoring Planning

Metamodeling
- Risk Portfolio Development
- Model Development
- Uncertainty Characterization
- Validate metamodel
- Metamodel Creation
  - Generate samples
  - Generate Launcher script
  - HPC
  - Model training
- Deployment

Decision Support
- Apache-Django
  - Metamodel creation
  - Cost analysis
  - Monte Carlo simulation
  - Visualization
- Admin
  - Job management
  - Cost data
  - Model parameter management

Platform-as-a-service

Types of metamodeling supported:
- Gaussian process regression
- Sparse grid
Data-Space Inversion (DSI)

What is DSI?
- A new paradigm for long-term prediction and UQ without using history matching
- Prior knowledge is used to generate possible scenarios, but not to calibrate model
- DSI combines physically-based model with ML

Jeong et al., 2018a, A learning-based data-driven forecast approach for predicting future reservoir performance. AWR.
**Optimal Monitoring Network Design**

**Objective Function**
- Well cost = CAPEX($/well) + OPEX($/well/day) + Intervention($/well)
- Leakage cost = Brine($/ton) + CO₂($/ton)

**Optimization toolbox**
- Binary Integer Programming
  - Linear problem
  - Convex

**Constraints**
- # of monitoring wells ≤ Nₘₐₓ
- CO₂ leakage ≤ M% of total injected CO₂
- ΔP at tₗₑᵃ❦ₜₑ₍₆ₜₜ detection ≥ ΔPₜʰʳᵉ₉ᵈₑ₉₈"Š

---

Our tool maximizes NPV by considering
• High uncertainty in geologic models
• Monitoring budget
• Leakage damage cost
• Carbon credit: *45Q Tax Incentives for CCUS*

3D model site scale models

Optimization Toolbox for Pressure Monitoring Network
Lessons Learned

– We are developing an intelligent monitoring system to help extract intelligent information. Our applications include
  • Web-based monitoring planning
  • Pressure-based monitoring network design
  • Data space inversion
  • Deep learning tools

– Data-driven machine-learning models are suitable for continuous monitoring and anomaly detection and can be used together with physics-based models for surrogate modeling

– A viable approach is to combine prior information, expert knowledge, and state-of-the-art machine learning tools for knowledge discovery and representation
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
  • Year 3-4: Updated the existing platform for usability
– Task 4: Coupled modeling / data assimilation
  • Year 2: Implemented workflow for automating data assimilation. Demonstrated Web-based modeling approaches
  • Year 3-4: Focused on ML and DL tool development
– Task 5: Integration and demonstration
  • Year 1-4: Experimented with a large number of web-based technologies for making the system more user friendly
Synergy Opportunities

– DIAL-GCS is an intelligent monitoring system designed for anomaly detection, monitoring network design, physics-based machine learning, leakage cost estimation

– Most tools are web-based, or can be readily converted to web-based, for CCS decision support needs
Project Summary

– Developed and improved DIAL system

– All tasks are on revised schedule

– Next steps
  • Formalize data transformation and work flow
  • Provide deep learning based web service
  • Integrate different analytic modules and disseminate results
  • Wrap the project in the next year
Acknowledgements

- DOE/NETL PM: Bruce Brown

- University of Texas
  - Bureau of Economic Geology: Sue Hovorka, Katherine Romanak, Hoonyoung Jeong, Zhong Zhi
  - Texas Advanced Computing Center

- LBNL: Barry Freifeld (provided DTS data)
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
# Gantt Chart

## Table 1. Revised Project Gantt chart
(Numbers in table rows indicate milestones.)

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

– Peer-Review Manuscripts


• Sun, A. Y., 2018, Discovering state-parameter mappings in subsurface models using generative adversarial networks, Geophysical Research Letters, 45(20), 11,137-11,146.


– Presentations

• Development of anomaly detection models for deep subsurface monitoring, presented at the fall meeting of American Geophysical Union, New Orleans, LA, December, 2017
The optimal monitoring well locations are different because heterogeneous permeability affects

- Spatial pressure distribution
- Leakage detection time

<table>
<thead>
<tr>
<th>Geologic model</th>
<th>( C_{\text{brine}} )</th>
<th>( C_{\text{CO}_2} )</th>
<th>( C_{\text{brine}} )</th>
<th>( C_{\text{CO}_2} )</th>
<th>( C_{\text{brine}} )</th>
<th>( C_{\text{CO}_2} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \log_{10} k ) (md)</td>
<td>$10/t$</td>
<td>$10/t$</td>
<td>$10/t$</td>
<td>$1,000/t$</td>
<td>$100/t$</td>
<td>$10/t$</td>
</tr>
<tr>
<td>( \log_{10} k ) (md)</td>
<td>$8.76 \text{ MM}$</td>
<td>$9.63 \text{ MM}$</td>
<td>$29.75 \text{ MM}$</td>
<td>$9.16 \text{ MM}$</td>
<td>$9.99 \text{ MM}$</td>
<td>$31.37 \text{ MM}$</td>
</tr>
</tbody>
</table>

Black: leaky well
Green: injector
Magenta: monitoring well