Optimal sensor placement and monitoring design in geologic CO2 sequestration: A machine learning and uncertainty quantification approach

Misael M. Moralesa,b,[[1]](#footnote-1), Bailian Chena, Mohamed Mehanaa

1. Earth and Environmental Sciences Division, Los Alamos National Laboratory, USA
2. Hildebrand Department of Petroleum and Geosystems Engineering, The University of Texas at Austin, USA

Highlights:

* Filtering-based data assimilation method is developed to perform monitoring design.
* Machine learning reduced-order model is used to reduce computational cost of data assimilation process.
* Monitoring well placement optimization is performed to reduce uncertainty and minimize leakage risks.

Keywords:

* Geologic carbon sequestration
* Monitoring design optimization
* Machine learning
* Reduced-order modeling
* Data assimilation
* Uncertainty quantification

Abstract

*Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nunc dapibus turpis sed mollis egestas. Sed pulvinar accumsan est, nec ornare erat dictum sit amet. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia curae; Sed lacinia quam at dignissim facilisis. Pellentesque et dictum orci, eu pellentesque mi. Aliquam erat volutpat. Sed odio ante, varius et tellus sed, interdum ultricies tellus. Vestibulum vestibulum mollis ex in pharetra. Morbi tempus dolor in diam molestie tincidunt. Duis fermentum urna vel rhoncus tincidunt. Nulla fringilla, enim ac ornare congue, risus libero consequat risus, vel faucibus felis sem id orci. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nunc dapibus turpis sed mollis egestas. Sed pulvinar accumsan est, nec ornare erat dictum sit amet. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia curae; Sed lacinia quam at dignissim facilisis. Pellentesque et dictum orci, eu pellentesque mi. Aliquam erat volutpat. Sed odio ante, varius et tellus sed, interdum ultricies tellus. Vestibulum vestibulum mollis ex in pharetra. Morbi tempus dolor in diam molestie tincidunt. Duis fermentum urna vel rhoncus tincidunt. Nulla fringilla, enim ac ornare congue, risus libero consequat risus, vel faucibus felis sem id orci. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nunc dapibus turpis sed mollis egestas. Sed pulvinar accumsan est, nec ornare erat dictum sit amet. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia curae; Sed lacinia quam at dignissim facilisis. Pellentesque et dictum orci, eu pellentesque mi. Aliquam erat volutpat. Sed odio ante, varius et tellus sed, interdum ultricies tellus. Vestibulum vestibulum mollis ex in pharetra. Morbi tempus dolor in diam molestie tincidunt. Duis fermentum urna vel rhoncus tincidunt. Nulla fringilla, enim ac ornare congue, risus libero consequat risus, vel faucibus felis sem id orci.*

1. Introduction

Geologic CO2 sequestration (GCS) has emerged as an important technology to reduce anthropogenic greenhouse gas emissions to the atmosphere (Metz, 2005; Michael, et al., 2010). Different types of underground formations have been proposed to store CO2 emissions including oil and gas reservoirs, coal beds and seams, and deep saline aquifers (Placeholder1). The main concern in GCS projects is potential leakage of the CO2 through leakage pathways, such as improperly abandoned wells, faults, and fractures (Placeholder1). Such risks can pose a major threat to overlying resources (e.g., groundwater resources, oil and gas reservoirs, etc.) and human health (Placeholder1). Monitoring and verifying CO2 behavior within the subsurface reservoir are crucial for detecting potential leakage, assessing storage capacity, and evaluating environmental impacts (Placeholder1).

To ensure safe and efficient operations in a large-scale GCS site, risk management techniques are used to minimize and mitigate potential risks during CO2 injection and post-injection periods (Placeholder1). Monitoring is thus an important aspect of CGS risk management, and one of the main goals of the Department of Energy (DOE) Office of Fossil Energy National Risk Assessment Partnership (NRAP). For this goal, several monitoring techniques have been developed, including near surface CO2 flux and tracer measurements (Placeholder1), groundwater chemistry monitoring (Placeholder1), seismic surveying (Placeholder1), and pressure monitoring (Placeholder1).

Optimal sensor placement and monitoring design play a critical role in achieving accurate and efficient monitoring in GCS projects. Depending on the reservoir properties and heterogeneity, the placement of monitoring wells can provide a more accurate measurement of the injected CO2 plume and help mitigate potential leakage risks (Placeholder). In common CGS operations, each injection well is paired with one monitoring well, though large-scale projects often incorporate a larger number of monitoring wells (Placeholder). Moreover, the selection of monitoring measurement plays an important role in reducing uncertainties and quantifying risks in GCS operations (Placeholder). Therefore, it is crucial to define an optimal monitoring strategy in terms of both well placement and monitoring measurement type.

Recent advancement in monitoring systems such as smart or intelligent wells are capable of providing large amounts of data in terms of volume, velocity, variety, value, and veracity (Placeholder). Classical techniques in data processing and forecasting are sometimes hindered by big data, therefore machine learning provides a promising approach to enhance data-driven subsurface energy resource systems (Placeholder). By analyzing extensive datasets, machine learning algorithms can uncover complex latent patterns and relationships that may not be discernible through traditional methods (Placeholder). Machine learning approaches, when combined with reduced-order modeling (ROM) techniques, enable efficient and accurate prediction of key parameters, including pressure distribution, CO2 plume migration, and reservoir behavior (Placeholder). These insights facilitate the optimization of sensor placement and monitoring strategies, enabling better decision making and forecasting in GCS projects.

Accurately quantifying uncertainties is vital for the reliability of predictions and optimizing monitoring design under uncertain conditions (Placeholder). Uncertainty quantification is particularly important in GCS due to inherent complexities and variabilities associated with subsurface conditions, fluid flow, and measurement errors (Placeholder). Several approaches for history matching or data assimilation have been applied to GCS, including Markov Chain Monte Carlo (MCMC), randomized maximum likelihood (RML), rejection sampling (RS), ensemble Kalman filtering (EnKF) and ensemble smoother with multiple data assimilation (ES-MDA) (Placeholder). Filter-based approaches provide a robust framework for characterizing uncertainties associated with reservoir properties, operating conditions, and measurement errors (Placeholder). Leveraging data assimilation techniques allows for informed risk assessment, ensuring the safety and efficiency of GCS projects.

Numerous research endeavors have been dedicated to addressing monitoring design, sensor placement, and uncertainty quantification in GCS. Previous studies have explored various modeling techniques, simulation frameworks, and optimization algorithms to enhance monitoring strategies and improve forecasting (Placeholder). These investigations have focused on different aspects, such as multi-objective optimization (Placeholder), real-time monitoring (Placeholder), and adaptive sampling strategies (Placeholder).

Pawar et al. (2022) provide a robust framework for quantitative risk assessment of leakage in GCS. Utilizing the NRAP-open-IAM (Integrated Assessment Model) tool, they are able to quantify the leakage risk through legacy or abandoned wells in large-scale GCS projects. This framework can then be used to support permit applications for GCS projects. Yonkofski et al. (2016) use a simulated annealing (SA) global optimization approach to obtain the optimal monitoring measurement design in a GCS project. Their objective is to minimize the estimated time to first detection (ETFD) by iteratively mutating potential monitoring designs. Sun et al. (2013) propose an approach to optimize monitoring well location based on pressure measurements for GCS under geologic uncertainty. Using binary integer programming problem (BIPP) formulation, they effectively select optimal monitoring locations for homogeneous and fluvial heterogeneous reservoirs. However, their method requires a large number of forward simulations, which can be computationally costly and time consuming. Oladyshkin et al. (2013) propose a polynomial chaos expansion (PCE) and bootstrap filtering approach for assimilating pressure data into reservoir models and quantifying the uncertainty reduction in CO2 leakage rate at a GCS site. Jia et al. (2018) propose a Bayesian model average and Monte Carlo simulation to quantify parameter uncertainty based on a PCE ROM. However, Monte Carlo strategies require a very large number of realizations and can be extremely computationally inefficient. Chen et al. (2020) propose a risk assessment approach using ESMDA with geometric inflation factors (ES-MDA-GEO) to quantify the uncertainty monitoring data and calibrate the prior uncertain geologic models. Their work leverages continuous data assimilation as new monitoring data becomes available in GCS projects to improve the underlying model and reduce uncertainties. Mehana et al. (2022) provide a ROM-based approach to quantify wellbore leakage from depleted reservoirs in CO2-EOR operations. They compare the performance of different machine learning-based ROMs for prediction of cumulative leakage and quantify the uncertainty using Monte Carlo simulations. Sun and Durlofsky (2019) use a data-space inversion (DSI) approach to optimize the monitoring well locations in a GCS project with a genetic algorithm (GA) global optimization. Using principal component analysis (PCA) as a model reduction strategy, they reduce the uncertainty in CO2 saturation plume using a RML approach. In this approach, posterior geological models are not generated in the DSI method, which is different from traditional ensemble-based data assimilation approaches. Liu and Grana (2020) propose a deep convolutional autoencoder as a ROM strategy to assimilate seismic monitoring data in GCS. Their method requires HFS to obtain CO2 saturation plume predictions from an ensemble of prior models, which is then used to calculate the seismic response. The autoencoder is used to project the observed monitoring measurements into latent space, where ES-MDA is used to update the model parameters and quantify the uncertainty in predictions.

In this paper, we build upon the work of Chen et al. (2018) to systematically design an optimal monitoring placement and measurement strategy for large-scale GCS beyond naïve monitoring well placement and monitoring design. We propose a method for optimal GCS monitoring design based on well placement optimization and monitoring measurement selection. We develop an artificial neural network ROM to predict cumulative CO2 leakage from a prior ensemble of uncertain model parameters, and implement a filter-based data assimilation approach to select the most informative monitoring well location and measurement type in order to reduce uncertainties and CO2 leakage risks.

The structure of this paper is as follows: Section 2 present our methodology, Section 3 presents the results of our approach for two synthetic cases, and Section 4 summarizes our findings, discusses their implications, and outlines potential avenues for future research in the field of GCS.

1. Methodology
   1. Uncertainty quantification

The goal of this study is to evaluate the value of data in GCS monitoring design. The value of data is quantified by the amount of uncertainty that is reduced in the cumulative CO2 leakage, , over the duration of a GCS project. The prior probability density function (PDF) of the cumulative CO2 leakage is denoted as . In this study, prior refers to the probability distribution before a monitoring program is implemented. The distribution of potential monitoring data that could be measured at the monitoring wells is denoted as , where are the individual monitoring data points obtained if a monitoring design were implemented in a particular leakage scenario and is the total number of monitoring data points in . In this study, monitoring data is sampled monthly, and can represent pressure, CO2 saturation, or temperature values at the monitoring well. Thus, we denote as the realization of . For each , we obtain a posterior PDF denoted by , which can be calculated using a data assimilation procedure as the cumulative CO2 leakage, , for a given monitoring design data . The objective is to quantify the value of information (VOI) estimated from a distribution of potential monitoring design, allowing us to choose an optimal monitoring well placement and monitoring measurement type to minimize the uncertainty in potential leakage scenarios.

Following Chen et al. (2017, 2018) and Le and Reynolds (2014), the VOI is quantified by the uncertainty reduction in the objective function. We denote the amount of uncertainty in cumulative CO2 leakage distribution as , defined as:

(1)

where is the percentile of a distribution, and the . The distribution of cumulative CO2 leakage can be attributed to the uncertainty in model parameters, in this case the number of and vertical transmissibility of potential leaky pathways, , and the reservoir permeability multiplier, . Therefore, selecting a monitoring design that reduces the uncertainty in ensures that the monitoring design will function effectively under multiple possible potential leakage scenarios.

The expected posterior uncertainty distribution in given is given by:

(2)

where is the expectation with respect to all realizations of and is the number of data realizations. The expected uncertainty reduction, , as a result of data acquisition from a potential monitoring design is given by the difference between the prior uncertainty and the expected posterior uncertainty in cumulative CO2 leakage, as defined by:

(3)

By selecting the optimal monitoring well placement and monitoring measurement type, the uncertainty reduction, , quantifies the effectiveness of the particular GCS monitoring design, where the higher the uncertainty reduction the higher the VOI in the monitoring data obtained in the monitoring design.

* 1. Reduced order model development

Given the computational cost of traditional filter-based data assimilation, a reduced-order model is developed in this study. The workflow for the ROM development is illustrated in Fig. 1. This section provides a summary of the main steps in the ROM development workflow:

**Step 1:** *Experimental design:* Given a set of uncertain parameters and , we generate training samples using Latin Hypercube Sampling (LHS) (Placeholder).

**Step 2:** *Forward simulations*: physics-based HFS of CO2 injection and post-injection migration is performed with each of the training samples using the Finite Element Heat and Mass Transfer (FEHM) simulator (Placeholder).

**Step 3:** *Collect training data*: For each training realization, the set of uncertain parameters, monitoring data, and cumulative CO2 leakage are collected. In Fig. 1, we see that the uncertain parameters are inputs for the ROM training and the objectives of interest (cumulative CO2 leakage and monitoring data) are the corresponding outputs.

**Step 4:** *Train ROMs for the objectives of interest*: a reduced-order model is used to map the relationship between the training parameters inputs and outputs. We build an ensemble of ROMs, one for each objective of interest, namely the cumulative CO2 leakage () and the simulated monitoring data () at each specified timestep. A fully-connected artificial neural network (ANN) is implemented to build the ROMs. Fig. 2 shows the architecture of the ANN.

**Step 5:** *Validate the ROMs against the HFS*: using -fold cross-validation (Placeholder), we test the predictions from the ROMs against the HFS results in order to perform hyper-parameter tuning and obtain robust ROMs that can be used for further predictions.

* 1. Artificial neural network ROM training and performance

Using the Python TensorFlow/Keras package (Placeholder), we develop a fully-connected ANN architecture to build the ROMs. Each ANN consists of four hidden layers with sizes , , , and , respectively, with a total number of parameters equal to . A kernel regularizer is applied with the -norm, and dropout of is used on each hidden layer. The activation function is the parametric rectified linear unit (PReLU), which learns the negative slope for each batch in each epoch. Training is performed on an NVIDIA RTX A6000 GPU in about 2 minutes for each ROM using -fold cross-validation. The average validation mean squared error (MSE) is approximately and the correlation coefficient () is approximately 0.98. The truth vs. prediction performance for a set of 500 realizations of uncertain parameters is shown in Fig. 3.

* 1. Workflow for optimal monitoring design

In this section we present a filtering and ROM based workflow for optimal monitoring design of GCS. The workflow diagram is shown in Fig. 4. The main steps for the optimal monitoring design workflow are summarized below.

**Step 1:** *Develop ROMs for the objective function, , and predict monitoring data, :* A detailed description of the ROM development workflow were presented in the previous section. We build one ROM for each monitoring data point, , in each data vector . The vector of predicted monitoring data is denoted as , where is the vector of uncertain model input parameters, namely and . The ROMs are used to replace FEHM physics-based simulations and to predict the objectives of interest for a set of new input parameters not in the training data.

**Step 2:** *Generate an ensemble of realizations of monitoring data,* : initially, realizations are sampled from the prior PDF of , and are denoted as . The corresponding monitoring data, , for each are given by:

(4)

where is the ROM prediction for monitoring data points and denotes the realization of measurement errors which follow a Gaussian distribution.

**Step 3:** *Generate Monte Carlo samples and calculate prior uncertainty:* A large number () Monte Carlo samples are generated from the prior distribution of , and denotes as . The Monte Carlo samples are used to calculate the prior PDF and the amount of uncertainty in the prior can be computed using Eq. (1).

**Step 4:** *Filter the Monte Carlo samples, and compute expected posterior uncertainty*: Using a filtering-based method (Placeholder 51), also known as rejection sampling, we construct a posterior distribution of conditional to each . First, using the Monte Carlo samples, , generated in Step 3, we simulate the corresponding monitoring data with the ROMs generated in Step 1, such that . Here, represents a realization from the distribution of potential monitoring datasets that capture potential CO2 leakage scenarios given the uncertain input parameters and . The data assimilation error is defined as the maximum absolute error (MAE) as follows:

(5)

Given a threshold value , the sample is accepted as a legitimate realization of the posterior distribution according to the following acceptance probability:

(6)

The threshold value, , is chosen based on engineering judgement and takes into consideration the measurement and modeling errors. Therefore, is accepted if it is deemed sufficiently consistent with the true monitoring data realization. Every Monte Carlo sample is evaluated using Eq. (6) and the accepted samples constitute the posterior distribution of conditional to the monitoring data realization such that posterior samples of are obtained. The expected posterior uncertainty is calculated using Eq. (2).

**Step 5:** *Calculate the expected amount of uncertainty reduction,* : the expected amount of uncertainty reduction, , is calculated by comparing the uncertainty in the prior distribution and the expected value of the uncertainty in the posterior distribution using Eq. (3).

**Step 6:** *Monitoring well placement optimization*: We repeat Steps 1-5 for every possible monitoring well location in the GCS area of review (AOR), conditional to the data for each possible measurement type, . In order to accelerate the optimization procedure, we coarsen the simulation grid into a subgrid, meaning there are 16 possible monitoring well locations. We calculate the expected amount of uncertainty reduction for each monitoring data type, , for each possible monitoring well location , and obtain the monitoring design that maximally reduces the uncertainty in cumulative CO2 leakage (maximally reducing the uncertainty is equivalent to minimizing the negative expected uncertainty reduction), as shown in Eq. (7)

(7)

This results in an exhaustive search in the subgrid to obtain the optimal well location, , that yields the highest uncertainty reduction, defined by as follows:

(8)

With this optimal monitoring design workflow, the expected uncertainty reduction in cumulative CO2 leakage for each potential monitoring measurement and each potential monitoring well location can be computed, and the optimal monitoring design that reduces the uncertainty in the simulated amount of CO2 leakage is obtained.

1. Model description

We implement the optimal monitoring design workflow on a synthetic GCS model consisting of a heterogeneous storage reservoir, a homogeneous caprock layer and a homogeneous aquifer, as shown in the schematic of the base model in Fig. 5. The thickness of each of the three layers is , and the model is wide in the horizontal dimensions. The depth from ground surface to the top of the model is . A CO2 injection well is placed at the center of the reservoir and multiple potential leakage pathways traverse the caprock, where CO2 could potentially leak into the aquifer. Note that only one possible leakage pathway is shown in Fig. 5, while we have considered several scenarios with multiple potential leakage pathways. The caprock and aquifer layers have a homogeneous permeability distribution equal to and , respectively. The storage reservoir has a heterogeneous permeability distribution, as shown in Fig. 6. The base model is generated using a spherical variogram model (Placeholder 62) with major and minor correlation lengths of and , respectively, with a major direction of from the positive -axis.

The mean of the permeability field is . For each realization, we assume that the reservoir permeability is uncertain, and to honor this uncertainty we use a permeability multiplier, , to multiply the aforementioned base permeability distribution. The lower and upper bounds for the multiplier and the potential leaky pathways are shown in Table 1.

A numerical mesh for the reservoir simulation is made using the grid generation toolkit LaGriT (Placeholder 63). The numerical mesh has nodes in both the - and -directions, and nodes in the -direction. The distance between each node in the - and -directions is , and in the -direction it is . The total number of nodes used in the simulation is , with nodes in the reservoir and caprock, respectively, and nodes in the aquifer. FEHM is used for 3D multiphase flow simulations (Placeholder 59). The boundary conditions of the reservoir are defined as Dirichlet boundaries, allowing CO2 to flow out but not in, and water pressure above hydrostatic. The top and bottom boundary conditions of the simulation model are no-flow boundaries. The thermal conditions of the model are initialized using a geothermal gradient of with a temperature of at the top. Pressure gradients are initialized at with a pressure of along the top. In this study, CO2 is constantly injected in a five-year period, monitored monthly, with a constant injection rate of million tons/year.

1. Conclusions

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nunc dapibus turpis sed mollis egestas. Sed pulvinar accumsan est, nec ornare erat dictum sit amet. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia curae; Sed lacinia quam at dignissim facilisis. Pellentesque et dictum orci, eu pellentesque mi. Aliquam erat volutpat. Sed odio ante, varius et tellus sed, interdum ultricies tellus. Vestibulum vestibulum mollis ex in pharetra. Morbi tempus dolor in diam molestie tincidunt. Duis fermentum urna vel rhoncus tincidunt. Nulla fringilla, enim ac ornare congue, risus libero consequat risus, vel faucibus felis sem id orci.

Acknowledgement

This project was funded by the US DOE’s Fossil Energy Office through the National Risk Assessment Partnership (NRAP) managed by the National Energy Technology Laboratory (NETL). Numerical simulations were performed on Los Alamos National Laboratory clusters supported by the High Performance Computing Division

1. Corresponding author

   *E-mail address:* [misaelmorales@lanl.gov](mailto:misaelmorales@lanl.gov) (M. M. Morales) [↑](#footnote-ref-1)