Optimal sensor placement and monitoring design in geologic CO_2 sequestration: A machine learning and uncertainty quantification approach

Misael M. Morales[a,b,*], **Bailian Chen**[a], **Mohamed Mehana**[a,*]

(a) Earth and Environmental Sciences Division, Los Alamos National Laboratory

(b) Hildebrand Department of Petroleum and Geosystems Engineering, The University of Texas at Austin

*Corresponding author; email: misaelmorales@lanl.gov, mzm@lanl.gov

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

Geologic CO_2 sequestration (GCS) projects have large uncertainties in geological properties, and require optimal monitoring designs in order to quantify and manage risks. An effective monitoring design is crucial to ensure the safe and permanent storage of CO_2 throughout the life-cycle of a GCS project. Optimal monitoring design involve selecting: (i) what is the optimal placement of a monitoring well, and (ii) what is the optimal monitoring measurement data (pressure, CO_2 saturation, temperature, etc.). We have developed a filtering-based data assimilation approach to design an optimal GCS monitoring strategy for variable well placement and monitoring data design. To accelerate the optimization time and reduce computational costs, a machine-learning algorithm, namely Artificial Neural Networks, is used to derive computationally efficient reduced-order models from the results of full-physics numerical simulations of CO2 injection in saline aquifers. We validate our workflow with example scenarios of CO2 leakage through legacy or abandoned wellbores and show an optimal monitoring strategy can be selected with the aim of reducing the cumulative CO2 leakage in the GCS site. The examples demonstrate that the proposed approach is effective in developing optimal monitoring designs that take into consideration geologic uncertainties.

1. Introduction

Geologic CO_2 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 CO_2 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 CO_2 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 CO_2 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 CO_2 injection and post-injection periods (Placeholder1). Monitoring is thus an important aspect of GCS 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 CO_2 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 CO_2 plume and help mitigate potential leakage risks (Placeholder). In common GCS 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 data sets, 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, CO_2 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 CO_2 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 ES-MDA 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 CO_2-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 CO_2 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 CO_2 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 naive 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 CO_2 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 CO_2 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.

2. Methodology

2.1 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 CO_2 leakage, M_c, over the

duration of a GCS project. The prior probability density function (PDF) of the cumulative CO_2 leakage is denoted as $P(M_c)$. 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 $D = [d_1, d_2, \ldots, d_{n_d}]$, where $\{d_i\}_{i=1}^{n_d}$ are the individual monitoring data points obtained if a monitoring design were implemented in a particular leakage scenario and n_d is the total number of monitoring data points in D. In this study, monitoring data is sampled monthly, and can represent pressure, CO_2 saturation, or temperature values at the monitoring well. Thus, we denote D^j as the j^{th} realization of D. For each D^j, we obtain a posterior PDF denoted by $P(M_c|D^j)$, which can be calculated using a data assimilation procedure as the cumulative CO_2 leakage, M_c, for a given monitoring design data D^j. 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 CO_2 leakage distribution $P(M_c)$ as $U[P(M_c)]$, defined as:

$$U[P(M_c)] = P_{90}[P(M_c)] - P_{10}[P(M_c)] \tag{1}$$

where $P_{10}[\bullet]$ is the 10^{th} percentile of a distribution and $P_{90}[\bullet]$ is the 90^{th}. The distribution of cumulative CO_2 leakage can be attributed to the uncertainty in model parameters, in this case the number of and the vertical transmissibility of potential leaky pathways, k_v^ℓ, and the reservoir permeability multiplier, k_R. Therefore, selecting a monitoring design that reduces the uncertainty in M_c ensures that the monitoring design will function effectively under multiple possible potential leakage scenarios.

The expected posterior uncertainty distribution in M_c given D is given by:

$$E_d[U[P(M_c|D]] = \frac{1}{\ell_d} \sum_{j=1}^{\ell_d} U[P(M_c|D^j)] \tag{2}$$

where E_d is the expectation with respect to all realizations of D and l_d is the number of data realizations. The expected uncertainty reduction, U_R, 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 CO_2 leakage, as defined by:

$$U_R = U[P(M_c)] - E_d[U[P(M_c|D)]] \tag{3}$$

By selecting the optimal monitoring well placement and monitoring measurement type, the uncertainty

reduction, U_R, 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.

2.2 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:

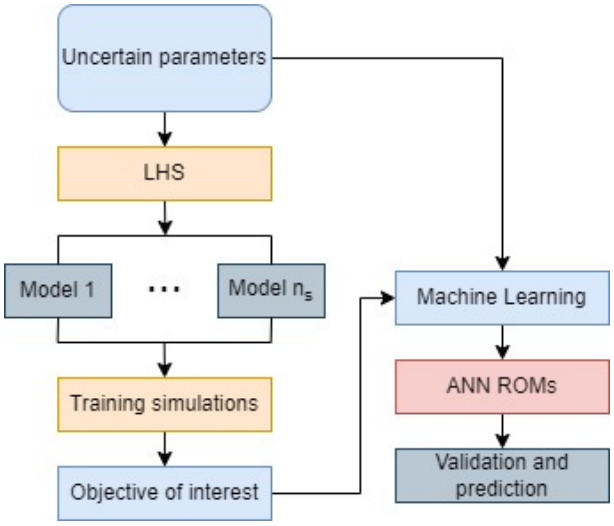

Figure 1: Workflow diagram for machine learning-based ROM development.

Step 1: *Experimental design*: Given a set of uncertain parameters $k_v^{\ell_d}$ and k_R, we generate n_s training samples using Latin Hypercube Sampling (LHS) (Placeholder).

Step 2: *Forward simulations*: Physics-based HFS of CO_2 injection and post-injection migration is performed with each of the n_s 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 CO_2 leakage are collected. In Fig.1, we see that the uncertain parameters are inputs for the ROM training and the objectives of interest (cumulative CO_2 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 CO_2 leakage (M_c) and the simulated monitoring data (D) 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 10-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.

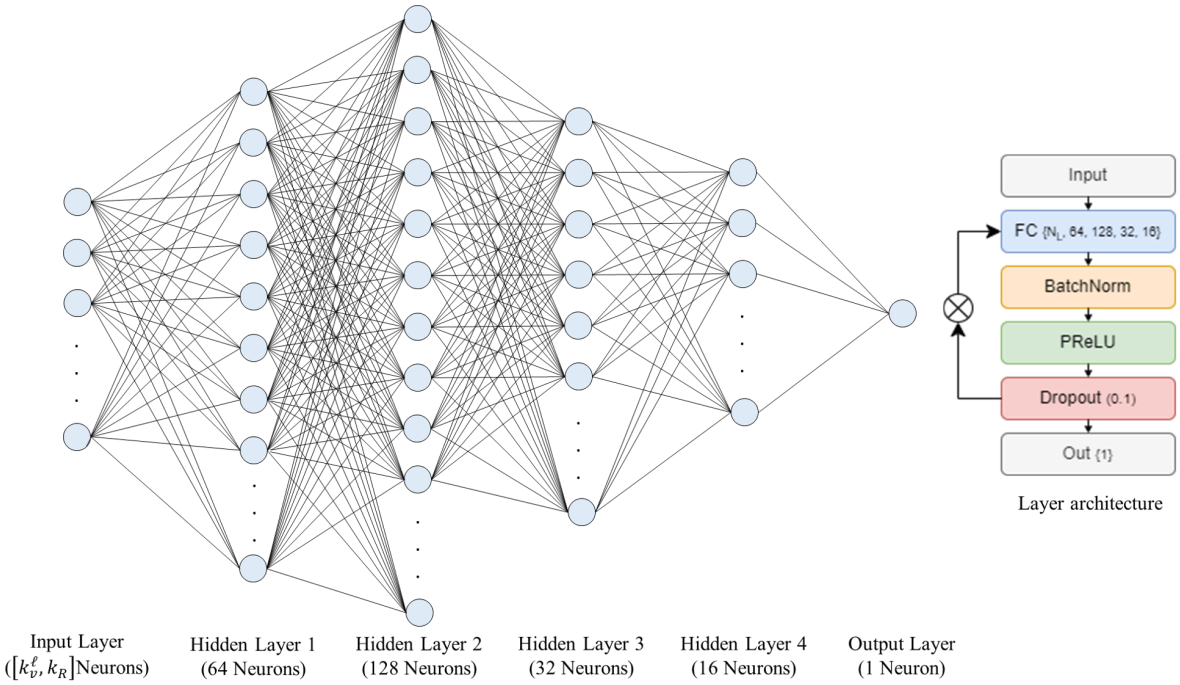

Figure 2: Artificial neural network ROM overall architecture (left), and per-layer architecture (right).

2.3 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 64, 128, 32, and 16, respectively, with a total number of parameters equal to $14,705$. A kernel regularizer is applied with the ℓ_1-norm, and dropout of 10% 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. The Adam optimizer (Placeholder) is used with a mean squared error (MSE) loss function. Training is performed on an NVIDIA RTX A6000 GPU in about 2 minutes for each ROM using 10-fold cross-validation. The average validation MSE is approximately 8.5×10^{-4} and the correlation coefficient (R^2) is approximately 0.98. The truth vs. prediction performance for a set of 500 realizations of uncertain parameters is shown in Fig.3.

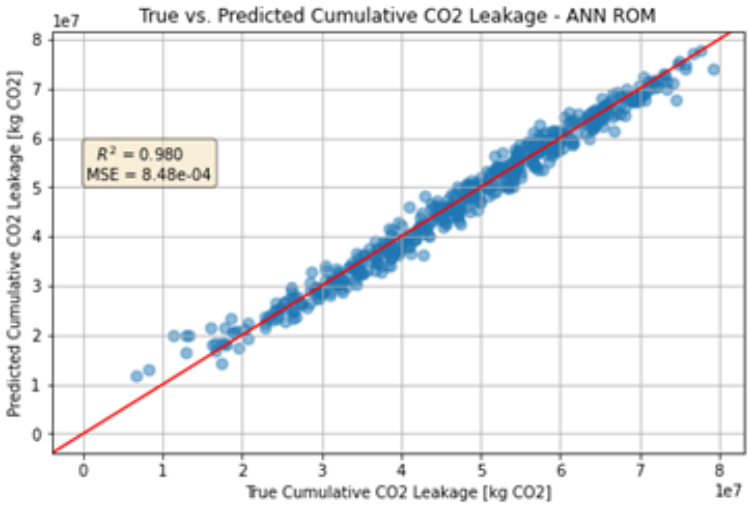

Figure 3: Cumulative CO_2 leakage prediction form ANN ROM vs true cumulative CO_2 leakage.

2.4 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.

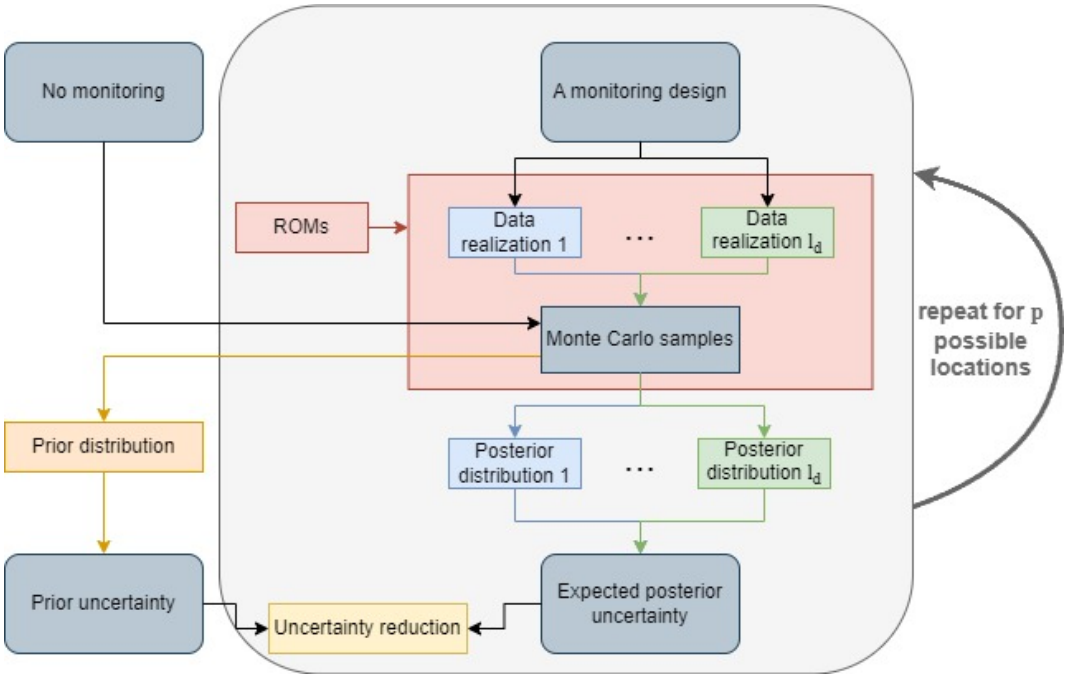

Figure 4: Workflow diagram for optimal monitoring design.

Step 1: *Develop ROMs for the objective function, M_c, and predict monitoring data, D*: A detailed

description of the ROM development workflow were presented in the previous section. We build one ROM for each monitoring data point, d_i, in each data vector D^j. The vector of predicted monitoring data is denoted as $O(m) = [O_1(m), O_2(m), \ldots, O_{n_d}(m)]^T$, where m is the vector of uncertain model input parameters, namely k_v^ℓ and k_R. 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, D*: Initially, l_d realizations are sampled from the prior PDF of m, and are denoted as $\{\widetilde{m}^j\}_{j=1}^{\ell_d}$. The corresponding monitoring data, $\widetilde{d}_{obs}^j$, for each $\widetilde{m}^j$ are given by:

$$\widetilde{d}_{obs}^j = O(\widetilde{m}^j) + e^j \tag{4}$$

where $O(\widetilde{m}^j$ is the ROM prediction for n_d monitoring data points and e^j denotes the j^{th} realization of measurement errors which follow a Gaussian distribution.

Step 3: *Generate Monte Carlo samples, and calculate prior uncertainty*: A large number $(50,000)$ Monte Carlo samples are generated from the prior distribution of m, and denotes as $\{\hat{m}^k\}_{k=1}^{\ell_{MC}}$. 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 m conditional to each $\widetilde{d}_{obs}^j$. First, using the Monte Carlo samples, $\hat{m}^k$, generated in Step 3, we simulate the corresponding monitoring data $\hat{d}^k$ with the ROMs generated in Step 1, such that $\hat{d}^k = O(\hat{m}^k)$. Here, $\hat{d}^k$ represents a realization from the distribution of potential monitoring data sets that capture potential CO_2 leakage scenarios given the uncertain input parameters k_v^ℓ and k_R. The data assimilation error is defined as the maximum absolute error (MAE) as follows:

$$MAE(d_{obs}^j) = \max_{1 \le i \le n_d} |\widetilde{d}_{obs,i}^j - \hat{d}_i^k| \tag{5}$$

Given a threshold value τ, the $\hat{m}^k$ sample is accepted as a legitimate realization of the posterior distribution according to the following acceptance probability:

$$P_{acc}(\hat{m}^k) = \begin{cases} 1, & \text{if } MAE < \tau \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

The threshold value, τ, is chosen based on engineering judgement and takes into consideration the measurement and modeling errors. Therefore, $\hat{m}^k$ 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 m conditional to the monitoring data realization $\widetilde{d}^j_{obs}$ such that ℓ_d posterior samples of m are obtained. The expected posterior uncertainty is calculated using Eq. (2).

Step 5: *Calculate the expected amount of uncertainty reduction U_R*: The expected amount of uncertainty reduction, U_R, 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, D^j. In order to accelerate the optimization procedure, we coarsen the simulation grid into a 4×4 subgrid, meaning there are 16 possible monitoring well locations. We calculate the expected amount of uncertainty reduction for each monitoring data type, D^j, for each possible monitoring well location $\{x^p\}_{p=1}^{16}$, and obtain the monitoring design that maximally reduces the uncertainty in cumulative CO_2 leakage (maximally reducing the uncertainty is equivalent to minimizing the negative expected uncertainty reduction), as shown in Eq. (7)

$$x_p^* = \min_{1 \leq p \leq 16} -U_R^{x_p} \tag{7}$$

This results in an exhaustive search in the subgrid to obtain the optimal well location, x_p^*, that yields the highest uncertainty reduction, defined by $U_R^{x_p}$ as follows:

$$U_R^{x_p} = U^{x_p}[P(M_c)] - E_d[U^{x_p}[P(M_c|D^j]] \tag{8}$$

With this optimal monitoring design workflow, the expected uncertainty reduction in cumulative CO_2 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 CO_2 leakage is obtained.

3. 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 30 m, and the model is 1 km wide in the horizontal dimensions. The depth from ground surface to the top of the model is 1000 m. A CO_2 injection well is placed at the center of the reservoir and multiple potential leakage pathways traverse the caprock, where CO_2 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 1×10^{-1} m^2 and 1×10^{-13} m^2, 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 680 m and 280 m, respectively, with a major direction of 45° from the positive x-axis.

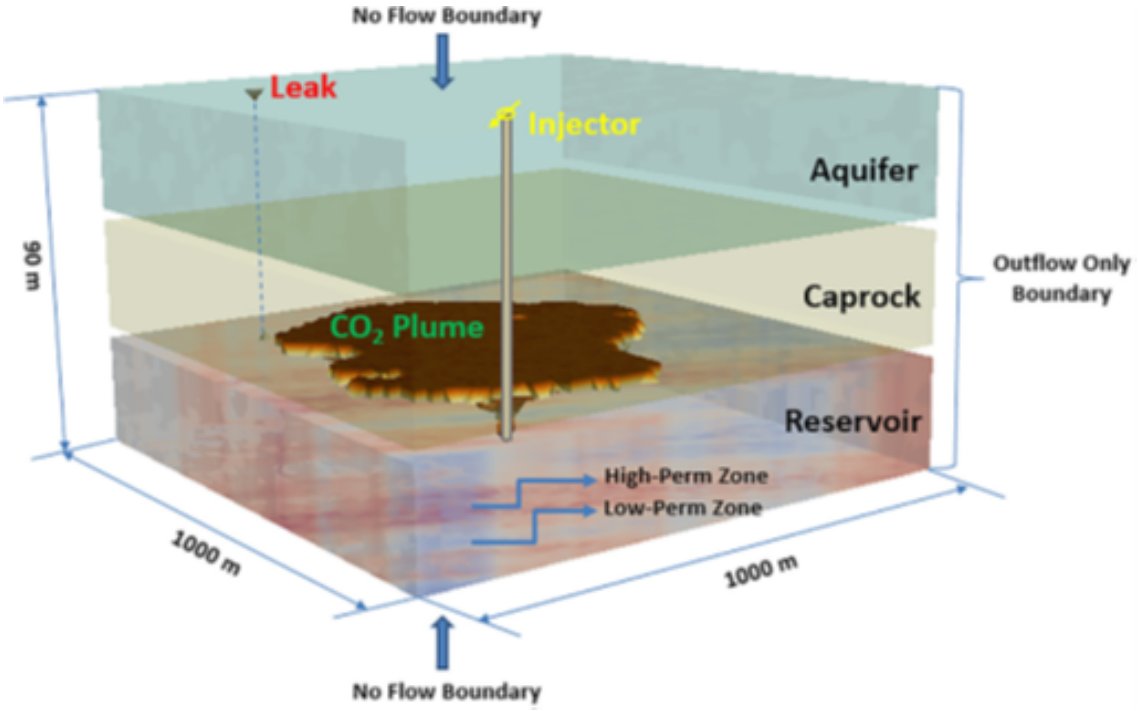

Figure 5: Schematic of the base model, in which a storage reservoir and aquifer are separated by a caprock. At the center is a CO_2 injection well. The vertical axis is exaggerated 7 times.

The mean of the permeability field is 1×10^{-13} m^2. For each realization, we assume that the reservoir permeability is uncertain, and to honor this uncertainty we use a permeability multiplier, k_R, to multiply the aforementioned base permeability distribution. The lower and upper bounds for the multiplier k_R and the potential leaky pathways k_v^ℓ are shown in Table 1.

Table 1: Uncertain parameters and their lower and upper bounds

Uncertain parameters	Symbol	Lower bound	Upper bound	Unit
Reservoir permeability multiplier	k_R	0.5	2	–
Permeability of leaky pathway(s)	k_v^ℓ	-19 0.001	-14 10	$log_{10}\ [m^2]$ mD

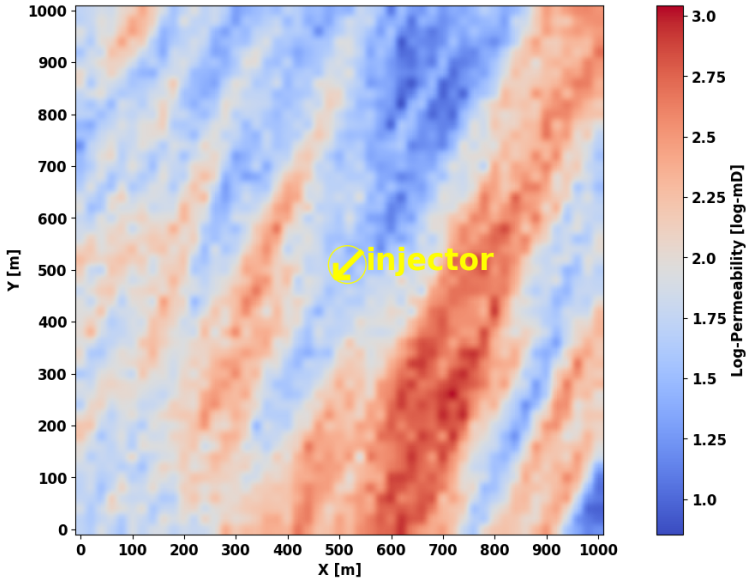

Figure 6: Log-permeability distribution of the base model. The darkest blue color corresponds to the lowest permeability, while the darkest red color corresponds to the highest. The yellow circle with an arrow indicates the CO_2 injection well.

A numerical mesh for the reservoir simulation is made using the grid generation toolkit *LaGriT* (Placeholder 63). The numerical mesh has 51 nodes in both the x- and y-directions, and 31 nodes in the z-direction. The distance between each node in the x- and y-directions is 20 m, and in the z-direction it is 3 m. The total number of nodes used in the simulation is 80,631, with 26,010 nodes in the reservoir and caprock, respectively, and 28,611 nodes in the aquifer. FEHM is used for 3D multi-phase flow simulations (Placeholder 59). The boundary conditions of the reservoir are defined as Dirichlet boundaries, allowing CO_2 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 0.03°C/m with a temperature of 20°C at the top. Pressure gradients are initialized at 9.81×10^{-3} MPa/m with a pressure of 0.2 MPa along the top. In this study, CO_2 is constantly injected in a five-year period, monitored monthly, with a constant injection rate of 0.1 million tons/year.

4. Results and Analysis

4.1 Workflow validation

We validate the workflow for optimal GCS monitoring design using a simple example. Fig. 6 shows the log-permeability distribution for the base model with a CO_2 injection well at the center, noted with a yellow circle and arrow. All the monitoring data in this study are collected in the aquifer zone, similar to monitoring at the above zone monitoring interval (AZMI) in the work of Sun et al. (Placeholder 43). The monitoring frequency is once per month for the duration of 5 years injection, resulting in 60 monitoring data points. The objective function, M_c, is the cumulative CO_2 leakage at the end of 5 years. In the model, we set up three material zones corresponding to the three adjacent formations, namely the storage reservoir, caprock, and aquifer. The cumulative CO_2 saturation in each zone can be output from the FEHM simulation results, and the cumulative leakage is computed by summing the CO_2 mass in the aquifer and caprock layers. Our approach for monitoring design involves quantifying the uncertainty reduction by monitoring pressure, CO_2 saturation, or temperature at each potential monitoring well location.

The data assimilation error tolerance, τ from Eq. (6), for pressure is set equal to 0.002 MPa, while for CO_2 saturation it is 0.05, and for temperature it is 0.002°C. Note that the choice of τ is site and case specific and is based on engineering judgement that takes into consideration the measurement and modeling error.

Two case studies are considered in this study: (1) GCS project with 3 potential leakage pathways, and (2) GCS project with 6 potential leakage pathways. The uncertain parameters are the permeability multiplier, k_R for the storage reservoir, and the ℓ permeability values for the ℓ potential leakage pathways, where $\ell = 3$ and $\ell = 6$, respectively. The total number of uncertain parameters, u^ℓ are 4 and 7, respectively. The lower and upper bounds for the uncertain parameters are shown in Table 1. For each case study, we run 500 training simulations generated by LHS with u^ℓ uncertain parameters. Each HFS requires approximately 22 minutes. We perform parallelization on an 8-node cluster, and the total simulation time is approximately 23 hours to finish all 500 training realizations. Fig. 7 shows the base model for Case 1 and Case 2 respectively.

Table 2: The parameters for one chosen model from the 500 training realizations in Case 1

Parameters	Value	Unit
CO_2 injection rate	3.17	kg/s
Thickness of caprock layer	30	m
Permeability of 1^st potential leakage pathway	$2.19 \times 10^{-}17$	m^2
Permeability of 2^nd potential leakage pathway	$3.37 \times 10^{-}17$	m^2
Permeability of 3^rd potential leakage pathway	$2.97 \times 10^{-}16$	m^2
Distance between injector and 1^st potential leakage pathway	424.3	m
Distance between injector and 2^nd potential leakage pathway	360.6	m
Distance between injector and 3^rd potential leakage pathway	141.4	m
Permeability for aquifer layer	1×10^{-13}	m^2
Permeability for caprock layer	1×10^{-19}	m^2
Reservoir permeability multiplier	1.88	–

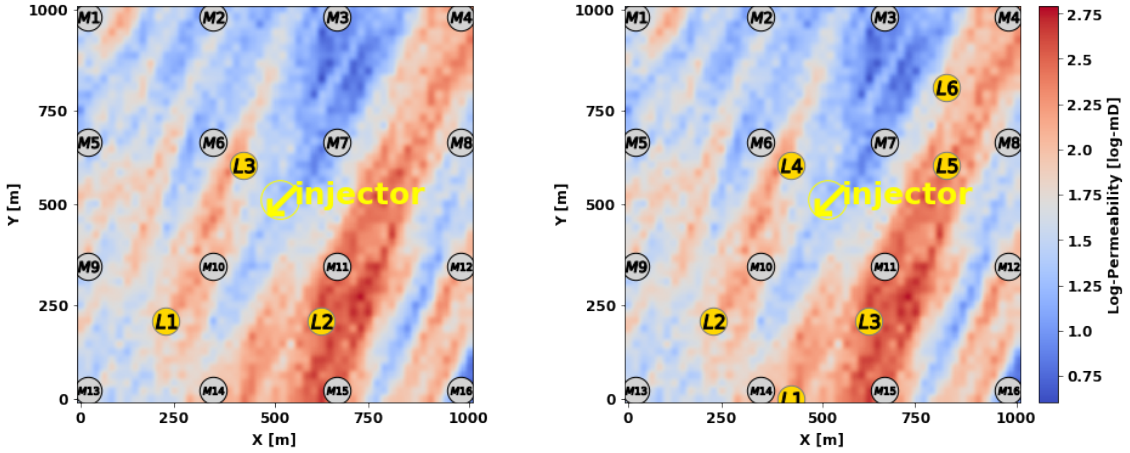

Figure 7: Log-permeability distribution of the base model for Case 1 (left) with 3 potential leaky pathways, and Case 2 (right) with 6 potential leaky pathways. The dark yellow circles labeled L_i represent the leakage pathways, light gray circles labels M_i are the possible monitoring well locations, and the yellow circle with an arrow is the CO_2 injection well.

We choose one simulation from the 500 training realizations in Case 1 to show when CO_2 leakage occurs. The values of the different parameters for the chosen model are shown in Table 2. The cumulative CO_2 leakage over the GCS project time is shown in Fig. 8. Figure 9 shows the leaked CO_2 saturation distribution at the top of the aquifer. It can be seen that CO_2 leakage occurs after about 210 days of injection. We observe that CO_2 is leaking through the potential pathway L_3, which is 141.4 m away from the injector, while no leakage occurs at potential pathways L_1 and L_2 after 5 years of injection. For this specific example, it is important to note that the permeability of L_3, k_v^3 is higher that that of L_1 and L_2.

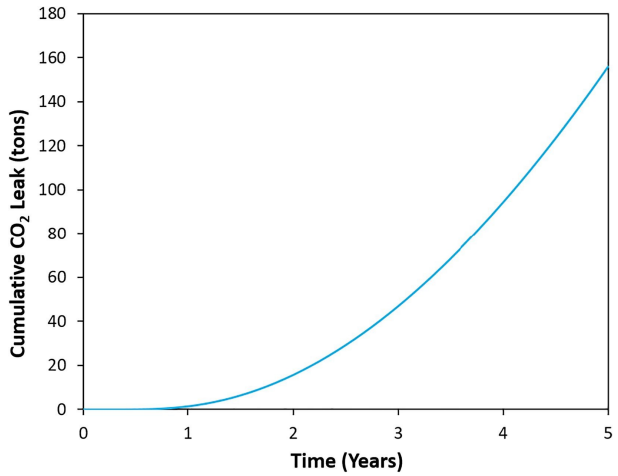

Figure 8: Cumulative CO_2 leakage over time computed for one chosen training realization in Case 1.

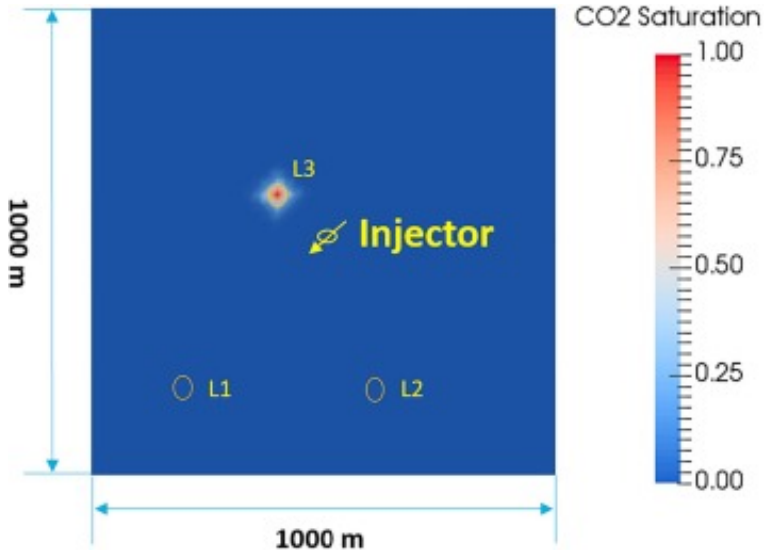

Figure 9: Plan view (top of the aquifer) of CO_2 leakage at the end of 5 years of injection based on one chosen training realization in Case 1. Yellow circles indicate the potential leakage pathways. Units for CO_2 saturation is fraction.

For each case, the 500 training realizations are used to train ROMs for the monitoring data and cumulative CO_2 leakage using the ANN architecture in Fig 2. Fig. 3 shows the quality of the ROMs tested by 10-fold cross-validation (Placeholder 64). The MSE and R^2 are 8.5×10^{-4} and 0.98, respectively. This proves that the fidelity of ROMs to the numerical simulations is high at the advantage of a much lower computational cost.

With the proposed workflow, the expected uncertainty reduction of the cumulative CO_2 leakage can be computed for each of the 16 possible monitoring well locations, for each monitoring measurement type. For each data set, 200 possible realizations of monitoring data are generated following Step 2 in Section 2.4.

To obtain the expected uncertainty reduction using Eq. (3), the prior uncertainty $U[P(M_c)]$ and posterior uncertainty $U[P(M_c|D^j)]$ corresponding to each possible monitoring data realization D^j for each possible well location x^p should be computed. Higher uncertainty reduction of the objective function indicates greater VOI in the monitoring data obtained from the optimal well location and monitoring measurement type. Through these examples, we can see that our proposed workflow can be effectively used to determine optimal CO_2 monitoring design from a set of alternative monitoring designs.

In Fig. 10 we observe the uncertainty reduction obtained at each possible monitoring well location and for each monitoring measurement type. We observe that monitoring for pressure provides the highest uncertainty reduction in general, followed by CO_2 saturation and lastly pressure. Fig. 11 shows a point-wise comparison of the uncertainty reduction at each monitoring well location for each measurement type. One can observe that placing a monitoring well at location 6 and assimilation the pressure measurements provides the highest uncertainty reduction possible in the monitoring design. The optimal monitoring design given by $(pressure, x^p = 6)$ yields an uncertainty reduction in the cumulative leakage of CO2 of approximately 29.42×10^6 tons (29.24 MT), while the optimal design for CO2 saturation and temperature monitoring yield an uncertainty reduction of approximately 19.34 MT and 17.71 MT, respectively

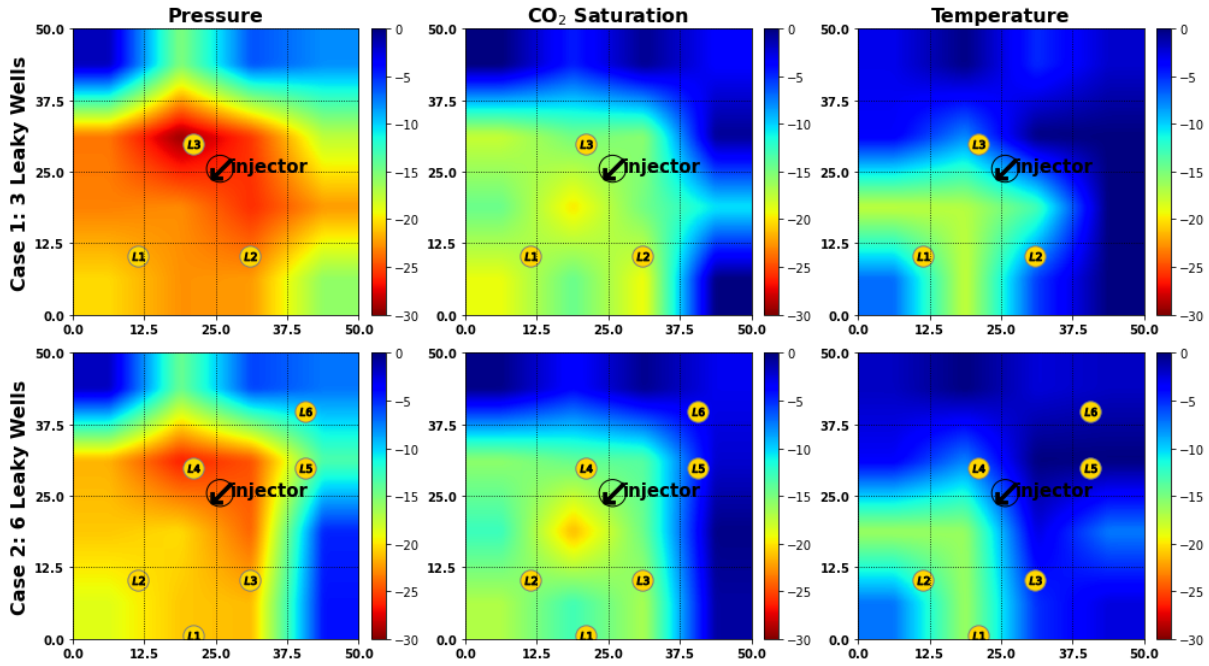

Figure 10: Plan view (top of the aquifer) of the uncertainty reduction obtained by all possible monitoring well locations. Top row represents Case 1 with 3 leakage pathways, and the bottom row represents Case 2 with 6 leakage pathways. Each column represents monitoring data for pressure, CO_2 saturation, and temperature, respectively.

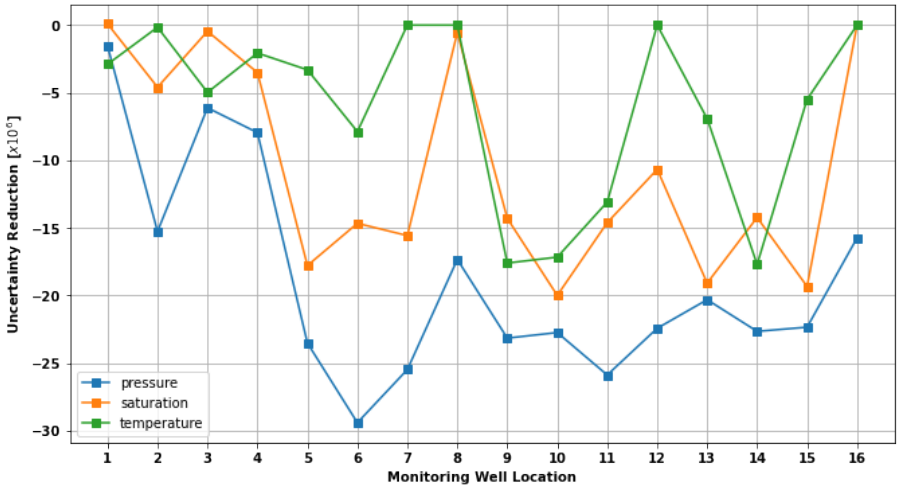

Figure 11: The point-wise calculated uncertainty reduction at each possible monitoring well location for each measurement type.

Fig. 12 shows the histograms for the prior and posterior distributions of the objective function obtained from the data realizations 1 and 100 for Case 1 and 2, respectively. The prior distribution is generated using LHS from the set of uncertain input parameters, k_V^ℓ and k_R, with a uniform distribution and calculating the cumulative CO_2 leakage using the ROMs. The variances of the posterior distributions calculated show significant reduction in uncertainty of cumulative CO_2 leakage compared to the priors.

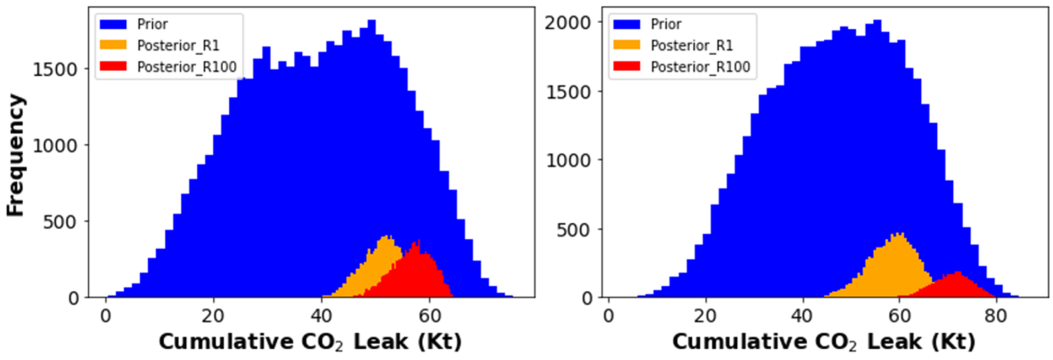

Figure 12: The histograms for the prior (blue) and posterior distributions obtained at the optimal monitoring design from the data realizations 1 (orange) and 100 (red) for Case 1 (left) and Case 2 (right), respectively.

5. Discussion

GCS monitoring operations require detailed data processing and interpretation in order to accurately quantify and potentially minimize leakage risks. Associated costs of performing monitoring operations requires evaluating the potential value of monitoring measurement type, and optimal monitoring well location, be-

fore the actual monitoring strategy takes place in the field. The workflow proposed can be used to select an optimal monitoring design that is robust under multiple potential leakage scenarios. Even though the examples used in our study to demonstrate how monitoring data from a shallow aquifer can be used, the proposed workflow can be extended and applied to monitoring data collected at any location and time within the GCS project. The potential value of such monitoring data can be evaluated by the presented workflow. Furthermore, placing several monitoring wells can provide a slight advantage compared to a single injector-monitor pair, but is impractical in field applications. Moreover, using several monitoring measurement types simultaneously provides little to no advantage compared to pressure monitoring. Refer to (Placeholder, Chen 2018) for further details.

At a CO_2 storage field operation, an optimal monitoring schedule and location based on the VOI described in this work can be used to collect the best possible monitoring data. The monitoring data can be assimilated to calibrate the uncertain model parameters using traditional data assimilation methods such as EnKF (Placeholder 66) or ES-MDA (Placeholder 67). The calibrated models can be used to improve the accuracy in prediction for future and long-term behavior of the injected CO_2.

6. Conclusions

In this study, a workflow based on a machine learning reduced-order modeling technique and uncertainty quantification method within an optimization loop is proposed for geologic CO_2 sequestration monitoring design. We use the uncertainty reduction in cumulative CO_2 leakage as the quantity of interest to measure the potential value of monitoring measurement data. The following conclusions have been drawn from this research:

1. The proposed workflow can generate reasonable values of uncertainty reduction in different risk metrics at CO_2 storage site, including cumulative CO_2 leakage by utilizing different monitoring designs and has been demonstrated using a synthetic GCS project.

2. The effect of different types of measurements (pressure, CO_2 saturation, and temperature) and the effect of monitoring well location on the choice of monitoring design is investigated. It is observed that pressure data has more value of information compared to CO_2 saturation, while temperature has the least value of information.

3. Well placement optimization is important to maximize the value of information for the monitoring design. Typical operations include pairs of one monitoring well for each injection well, partly due

to the cost of drilling and data acquisition. Determination of the best location provides significant benefits in reducing the uncertainty of cumulative CO_2 leakage.

4. The incremental reduction in uncertainty in the cumulative CO_2 leakage may not increase proportional to the distance from the injection well, and is a strong function of the reservoir permeability heterogeneity. Thus, an optimal monitoring well placement and measurement type is important to minimize potential risks.

Declaration of Competing Interest

The authors declare that they have no competing interests.

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

References