Applied Energy 225 (2018) 332–345

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

Geologic CO_2 sequestration monitoring design: A machine learning and uncertainty quantification based approach

Bailian Chen*, Dylan R. Harp, Youzuo Lin, Elizabeth H. Keating, Rajesh J. Pawar

Earth and Environmental Sciences Division, Los Alamos National Laboratory, Los Alamos, NM 87544, United States

HIGHLIGHTS

- Filtering-based data assimilation method is developed to perform monitoring design.
- Machine learning is used to reduce computational cost of data assimilation process.
- Uncertainty reduction is chosen as the metric to quantify the VOI of monitoring data.

ARTICLE INFO

Keywords:
Geologic carbon sequestration
Monitoring design
Machine learning
Reduced order model
Data assimilation
Uncertainty reduction

ABSTRACT

Monitoring is a crucial aspect of geologic carbon dioxide (CO_2) sequestration risk management. Effective monitoring is critical to ensure CO_2 is safely and permanently stored throughout the life-cycle of a geologic CO_2 sequestration project. Effective monitoring involves deciding: (i) where is the optimal location to place the monitoring well(s), and (ii) what type of data (pressure, temperature, CO_2 saturation, etc.) should be measured taking into consideration the uncertainties at geologic sequestration sites. We have developed a filtering-based data assimilation procedure to design effective monitoring approaches. To reduce the computational cost of the filtering-based data assimilation process, a machine-learning algorithm: Multivariate Adaptive Regression Splines is used to derive computationally efficient reduced order models from results of full-physics numerical simulations of CO_2 injection in saline aquifer and subsequent multi-phase fluid flow. We use example scenarios of CO_2 leakage through legacy wellbore and demonstrate a monitoring strategy can be selected with the aim of reducing uncertainty in metrics related to CO_2 leakage. We demonstrate the proposed framework with two synthetic examples: a simple validation case and a more complicated case including multiple monitoring wells. The examples demonstrate that the proposed approach can be effective in developing monitoring approaches that take into consideration uncertainties.

1. Introduction

Geologic CO_2 sequestration (GCS) is being considered as an important technology to reduce anthropogenic greenhouse gas emissions to the atmosphere [1–9]. Many potential reservoirs have been proposed to store anthropogenic CO_2 emissions, such as depleted oil or gas reservoirs, coal beds, deep oceans and deep saline formations [10–13]. The isolation of CO_2 from the environment is imperative for a GCS project not only for the project to successfully store CO_2, but also due to the fact that CO_2 leakage is a threat to the environment, the groundwater resources and human health [14–16]. CO_2 leakage may occur through improperly plugged and abandoned wellbores or through natural fractures or faults [1,17–19]. Given that depleted oil or gas reservoirs with significant numbers of abandoned wellbores are attractive locations for GCS [10,20–27], potential leakage through abandoned wellbores becomes a primary concern.

To ensure that large-scale GCS is safe and effective, a risk management strategy is generally used to minimize and mitigate risks during CO_2 injection and post-injection periods of a storage site [28–30]. Monitoring is an essential aspect of GCS risk management. To effectively monitor for CO_2 leakage, several monitoring technologies have been developed, including near-surface measurements of soil CO_2 flux and tracer [31,32], pressure monitoring [33–35], shallow groundwater chemistry monitoring [36,37], and micro-seismic and cross-well seismic survey [32,38]. A few studies have been conducted to evaluate the performance of different monitoring strategies or perform monitoring optimization for CO_2 storage sites. Next, we provide a discussion of some of the most relevant work in order to put the methodology introduced in this paper in context.

* Corresponding author.
E-mail address: chenbailian719@gmail.com (B. Chen).

https://doi.org/10.1016/j.apenergy.2018.05.044
Received 20 December 2017; Received in revised form 6 May 2018; Accepted 7 May 2018
Available online 15 May 2018
0306-2619/ Published by Elsevier Ltd.

Nomenclature

Symbols

d_i	ith individual data point that would be obtained if the monitoring design were implemented
$\widetilde{d}^j$	jth data vector corresponding to each $\widetilde{m}^j$
$\widehat{d}^k$	kth data vector corresponding to each $\widehat{m}^k$
D	data that would be measured for a particular monitoring program
D^j	jth realization of D
e^j	jth realization of the vector of measurement errors
E_d	expectation with respect to all realizations of D
l_d	number of data realizations
l_{mc}	number of Monte Carlo samples
m	uncertain input parameter
$\widetilde{m}^j$	jth model realization generated from prior probability density function of m
$\widehat{m}^k$	kth Monte Carlo sample
M_c	cumulative CO_2 leakage
n_d	number of data points in the data vector D
$O(m)$	ROM vector for data
$P(M_c)$	prior probability density function of M_c
$P(M_c\|D^j)$	posterior probability density function of M_c
P_{10}	10th percentile
P_{90}	90th percentile
P_{acc}	acceptance probability
U	amount of uncertainty
U_R	uncertainty reduction
τ	threshold for maximum absolute error

Acronyms

AZMI	above zone monitoring intervals
BIPP	binary integer programming problem
EnKF	ensemble Kalman Filter
ES-MDA	ensemble smoother with multiple data assimilation
FEHM	Finite Element Heat and Mass transfer
GCS	Geologic Carbon Sequestration
LHS	Latin Hypercube Sampling
MAE	maximum absolute error
MARS	Multivariate Adaptive Regression Splines
MCMC	Markov chain Monte Carlo
PCKF	probabilistic collocation based Kalman Filter
PDF	probability density function
ROMS	reduced-order-models
VOI	value of information

Yang et al. [39] provided a probabilistic method for predicting the performance of different monitoring networks at GCS sites. The objective of their method is to estimate the probability that a monitoring network will detect CO_2 leakage. Yang et al. [40] developed a risk-based monitoring assessment methodology, which was an extended work of Yang et al. [39], to incorporate background data for monitoring design. In this method, the detection probability defined as the probability that a measured signal will be above the preselected threshold at the monitoring location is first calculated, and then the detection probability is used to estimate the monitoring well density and response time for both known and unknown wellbore leakage locations.

Seto and McRae [41] presented a model-based framework for integrated monitoring design that can provide a quantitative understanding of the trade-offs between operational costs and risks in potential monitoring strategies. In addition, the challenges, risks and design considerations of large scale CO_2 storage were comprehensively reviewed and discussed in their work. In the work of Seto and McRae [42], a method for CO_2 detection based on a Bayesian model selection framework was introduced and applied to distinguish whether detected CO_2 is from a leak or from background fluctuations. The limits of different monitoring technologies, the challenges of leak detection and what are acceptable rates of leakage were thoroughly investigated in their research.

Sun et al. [43] proposed an approach to optimize monitoring networks under geological uncertainty. A binary integer programming problem (BIPP) formulated in their work was demonstrated for effectively selecting optimal monitoring locations in both homogeneous and heterogeneous formations. However, the proposed method requires running a forward full-scale model many times for different design options, which makes the BIPP computationally demanding. Sun et al. [43] suggested that a reduced order model or surrogate model can be used to speed up the BIPP optimization process. Cameron [44] demonstrated an approach to optimize sensor locations for CO_2 plume monitoring at a GCS site under geological uncertainty by minimizing the expected prediction error of the CO_2 plume size.

Dai et al. [45] proposed a data-worth analysis approach using probabilistic collocation based Kalman Filter (PCKF) to optimize the surveillance operation in a GCS project. In this approach, surrogate models are developed using polynomial chaos expansion to replace the original flow models. Thereafter, the expected variance reduction of field cumulative CO_2 leakage is assessed via data-worth analysis. An optimal monitoring operation scheme is selected by comparing the value of data-worth for different monitoring strategies. Dai et al. [46] applied the data-worth analysis and PCKF based approach to quantify the uncertainty reduction associated with the characterization of a migrating contaminant plume for different monitoring networks in groundwater systems.

Chen et al. [47] proposed an approach based on the Markov chain Monte Carlo (MCMC) method and reduced order models (ROMs) to quantify the uncertainty reduction of different pilot designs in an oil field. Note that a pilot refers to small-scale test and data collection operations prior to a full field development. In the proposed method, multiple realizations of monitoring data from a pilot test are generated, and probabilistic history matching (data assimilation) based on an MCMC method is performed for each data realization to obtain the corresponding posterior distribution. Though the MCMC based history matching is accomplished with the help of ROMs, the computational demand is still high due to the fact that for each data realization, MCMC often takes hundreds of thousands of runs to converge [48–50]. As mentioned in the work of Chen et al. [47], one limitation of their proposed framework is that with the increase of the uncertain model parameters, it would be difficult for the Markov chain to converge to the target distribution.

In this work, we build off the previously mentioned monitoring design algorithms to capture full physics simulations in a computationally efficient manner by using reduced order models and determine the optimal monitoring design for GCS sites based on many potential leakage scenarios that could occur. To quantify the uncertainty reduction, instead of using an MCMC method as in the work of Chen et al. [47], we apply a filtering based data assimilation method [51] in this study, which overcomes the convergence issue inherent in the MCMC method. In addition, a popular machine learning technique, Multivariate Adaptive Regression Splines (MARS) [52], is used to construct the ROMs or proxy models to reduce the computational cost when the filtering based data assimilation method is applied.

2. Methodology

2.1. Uncertainty quantification

In this study, the goal of monitoring design is to evaluate the value of data, where the value of data is quantified by the amount of uncertainty that is reduced in one of the metrics for CO_2 leakage risk assessment, cumulative CO_2 leakage. The cumulative CO_2 leakage over a specified time period is designated as the objective function here and denoted as M_c. The prior probability density function (PDF) of M_c is denoted as $P(M_c)$. Note that "prior" refers to a probability distribution before a monitoring program is implemented. The distribution of potential monitoring data that could be measured for a particular monitoring program and leakage scenario is denoted by D; that is, $D = [d_1,d_2,\cdots,d_{n_d}]$, where $d_i, i = 1,2,\cdots,n_d$ are the individual monitoring data points that would be obtained if the monitoring design were implemented given a particular leakage scenario and n_d is the number of monitoring data points in D. Here, d_i represents the ith monthly pressure, CO_2 saturation, or temperature values at a monitoring well. We denote D^j to be the jth realization from D. For each D^j, a posterior PDF of M_c denoted by $P(M_c|D^j)$ can be calculated by an ensemble-based history matching or data assimilation procedure, where $P(M_c|D^j)$ is the probability that the cumulative CO_2 leakage is M_c given that the data are D^j.

In conventional history-matching, D would represent the actual measurements and the simulation model(s) would be calibrated so that they match or honor these measurements [53–56]. In this study, D corresponds to a potential monitoring dataset from a distribution of potential monitoring datasets that are all considered equally based on available information. The objective of monitoring optimization is to quantify the value of information (VOI) estimated from a distribution of potential monitoring datasets. Quantifying the VOI of a distribution of potential monitoring data sets allows one to choose a CO_2 monitoring design that will perform well for a distribution of potential leakage scenarios.

As illustrated in the work of Chen et al. [47], the VOI of monitoring data, similar to the concept of data-worth [45], can be quantified by the uncertainty reduction in an objective function. In this study, uncertainty reduction is also applied to quantify the VOI of potential monitoring data. We denote the amount of uncertainty in a cumulative CO_2 leakage distribution $P(M_c)$ as $U(P(M_c))$, which, as in Le and Reynolds [57], is defined as

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

where $P_{10}(P(M_c))$ and $P_{90}(P(M_c))$, respectively, are the 10th and 90th percentiles of the distribution $P(M_c)$. Note that in our approach, we don't expect that the distribution of cumulative CO_2 leakage, $P(M_c)$, is measured in field practices, but simulated in groundwater flow models. The distribution of cumulative CO_2 leakage is due to the uncertainty in the model parameters which lead to the monitoring data realizations D^j. Selecting a monitoring design that reduces the uncertainty in M_c ensures that the monitoring design will function effectively under many potential leakage scenarios. The expected posterior uncertainty in M_c given D is given by

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

where E_d represents the expectation with respect to all realizations of D, and l_d is the number of data realizations. The expected uncertainty reduction, U_R, due to data acquisition from a potential CO_2 monitoring operation, is defined as the difference between the prior uncertainty and the expected posterior uncertainty in the cumulative CO_2 leakage, which is given by

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

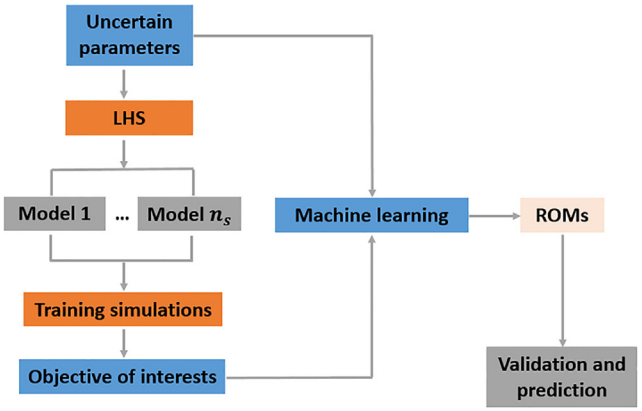


Fig. 1. Workflow diagram for ROM development.

The uncertainty reduction, U_R, quantifies the effectiveness of a particular monitoring design. The higher reduction in uncertainty in the cumulative CO_2 leakage, the higher the VOI contained in the monitoring data that would be obtained from the monitoring design.

2.2. Reduced order model development

To make the filtering approach computationally feasible, reduced order models (ROM) are developed in this study. The workflow for the ROM development is presented in Fig. 1. In this section, we give a brief summary of the main steps of this workflow.

Step 1: *Perform experimental design:* Staring from a set of uncertain parameters, we generate n_s training samples using Latin Hypercube Sampling (LHS) [58]. In this study, the uncertain parameters include the permeability multiplier of the storage reservoir and the permeabilities of all potentially leaky pathways.
Step 2: *Perform training simulations:* A numerical simulation of CO_2 injection and subsequent migration is performed with each of the n_s training samples. The Finite Element Heat and Mass transfer code (FEHM simulator) [59] is used to perform the simulations.
Step 3: *Collect inputs and outputs from the training simulations:* For each simulation, the uncertain parameters, monitoring datasets, and cumulative CO_2 leakage are collected. Note that in Fig. 1, the uncertain parameters are the inputs and the objectives of interest are the outputs from training simulations. For the purposes of ROM building, the objectives of interest can be cumulative CO_2 leakage and different monitoring data.
Step 4: *Generate ROMs for the objectives of interest:* A reduced order model can be used to approximate relations between training simulation input parameters and outputs. Note that we build one ROM for each objective of interest. The objectives of interest include cumulative CO_2 leakage (M_c) and simulated monitoring data (D) at specified times. A Multivariate Adaptive Regression Splines (MARS) algorithm [52], which has been successfully applied to the ROM development in GCS related research [60,18,61], is implemented to build the ROMs.
Step 5: *Validate the fidelity of ROMs to numerical simulator:* In this study, the fidelity of the ROMs to the numerical simulation results is tested with 10-fold cross-validation ([64]). Note that the cross-validation does not check whether the ROMs can predict the cases that it is trained with, but it can predict the cases that were not used for training. The validated ROMs can be used for further prediction instead of computationally intensive numerical simulations.

2.3. Workflow for monitoring design

In this section, a filtering and ROM based workflow for CO_2

monitoring design is presented. The flow diagram is shown in Fig. 2. The main steps of this workflow are summarized below.

Step 1. *Develop ROMs for the objective function, M_c, and predicted monitoring data, D:* The details of the ROM development were presented in the previous section. Note that we build one ROM for each monitoring data point in each data vector D^j). The vector of ROM simulated monitoring data is denoted as $O(m) = [O_1(m),\cdots,O_{n_d}(m)]^T$, where m denotes the vector of uncertain input parameters. The ROMs can then be used to replace the FEHM simulator to predict the objectives of interest for new input parameters that are not in the training ensemble.

Step 2. *Generate plausible realizations of the monitoring data, D:* First, l_d realizations are sampled from the prior PDF of m, and are denoted by $\widetilde{m}^j, j = 1,2,\cdots,l_d$. Then the monitoring data $\widetilde{d}^j_{\text{obs}}$ corresponding to each $\widetilde{m}^j$ is generated by the following equation;

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

where $O(\widetilde{m}^j)$ is calculated from the ROMs for the n_d predicted monitoring data points, i.e., $O(\widetilde{m}^j) = [O_1(\widetilde{m}^j),\cdots,O_{n_d}(\widetilde{m}^j)]^T$ and the vector e denotes a realization of the vector of measurement errors. Note that the elements of e follow a Gaussian distribution.

Step 3. *Generate Monte Carlo samples and calculate the prior uncertainty:* A large number of Monte Carlo samples $\widehat{m}^k, k = 1,2,\cdots,l_{mc}$ are generated from the prior distribution of m, and these Monte Carlo samples are used to construct the prior distribution. Then, the amount of uncertainty in the prior distribution of m can be calculated using Eq. (1).

Step 4. *Filter the Monte Carlo samples and compute the expected posterior uncertainty:* We use the filtering method [51] to construct a posterior distribution of m conditional to each $\widetilde{d}^j_{\text{obs}}, j = 1,2,\cdots,l_d$. The filtering method (or rejection sampling) with ROMs is performed as follows. First, using the Monte Carlo samples, $\widehat{m}^k$ generated in Step 3, we simulate the corresponding monitoring data $\widehat{d}^k$ using the ROMs constructed in Step 1, i.e., $\widehat{d}^k = O(\widehat{m}^k)$. The data vector $\widehat{d}^k$ can be expressed as $\widehat{d}^k = [\widehat{d}^k_1,\widehat{d}^k_2,\cdots,\widehat{d}^k_{n_d}]$ and represents a realization from the distribution of potential monitoring datasets that capture the many potential leakage scenarios that may occur given the uncertainty in the input parameters. We define the data assimilation error of the Monte Carlo samples, $\widehat{m}^k$, as the maximum absolute error (MAE), which is expressed as

$$\text{MAE}(\widehat{d}^k, d^j_{\text{obs}}) = \max_{1\leqslant i\leqslant n_d} \left|\widehat{d}^k_i - \widetilde{d}^j_{\text{obs},i}\right|. \tag{5}$$

$\widehat{m}^k$ is accepted as a legitimate realization of the posterior distribution according to the following acceptance probability:

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

This means the Monte Carlo sample $\widehat{m}^k$ will be accepted only if the corresponding MAE is smaller than the threshold τ. The threshold τ should be chosen based on engineering judgment that takes into consideration the measurement error and the modeling error. In other words, $\widehat{m}^k$ is accepted if it is deemed sufficiently consistent with the monitoring data realization. The acceptance probability in Eq. (6) is evaluated for all the Monte Carlo samples, and the accepted cases constitute the posterior distribution of m conditional to monitoring data realization $\widetilde{d}^j_{\text{obs}}$. This procedure is repeated for all the data realizations so that l_d posterior distributions of m are obtained. We compute the amount of uncertainty for each posterior distribution of m using Eq. (1) and then calculate the expected posterior uncertainty using Eq. (2).

Step 5. *Calculate the expected amount of uncertainty reduction U_R:* The expected uncertainty reduction U_R is calculated by comparing the uncertainty of the prior distribution and the expectation of the uncertainty for the posterior distributions of the cumulative CO_2 leakage using Eq. (3).

With this workflow, the expected uncertainty reduction in the cumulative CO_2 leakage for each of the potential monitoring operations can be computed, and the optimal monitoring design can be determined based on choosing the monitoring design that reduces the uncertainty in the simulated amount of CO_2 leakage.

3. Model description

To validate the introduced workflow for CO_2 monitoring design, we consider a synthetic GCS model consisting of a heterogeneous storage reservoir, a homogeneous caprock layer and a homogeneous aquifer. Fig. 3 presents the schematic of the base model. The thicknesses of the reservoir, caprock layer and aquifer are all 30 m. The model is 1000 m wide in the horizontal dimensions. The depth from the top of the model to the ground surface is 1000 m. It has a CO_2 injection well located in the center of the reservoir and multiple potentially leaky pathways in the caprock. CO_2 injected into the storage reservoir can potentially leak to the aquifer through the leaky pathways in the caprock layer. Note that only one leak location is shown in Fig. 3 while we have considered scenarios with multiple leak points. We consider a homogeneous permeability distribution for both the caprock layer and the aquifer with the permeability equal to 1×10^{-19} and 1×10^{-13} m^2, respectively. Note 10^{-15} m^2 is approximately equal to 1 mD. The heterogeneous permeability distribution for the storage reservoir of the base model is generated assuming a spherical anisotropic variogram model [62] with a major correlation length of 680 m and a minor correlation length of 280 m. The direction of maximum spatial correlation continuity is equal to 45 degrees measured from the positive x-axis. The mean of the permeability field of the reservoir is 1×10^{-13} m^2. We assume that the storage reservoir permeability is uncertain. To represent the uncertainty in reservoir permeability, we use a permeability multiplier to multiply the base permeability distribution described earlier. The value of the permeability multiplier is varied randomly to generate multiple realizations of permeability distributions. The lower and upper bounds for the multiplier are shown in Table 1. The other uncertain parameters we consider are the permeabilities of the potentially leaky pathways.

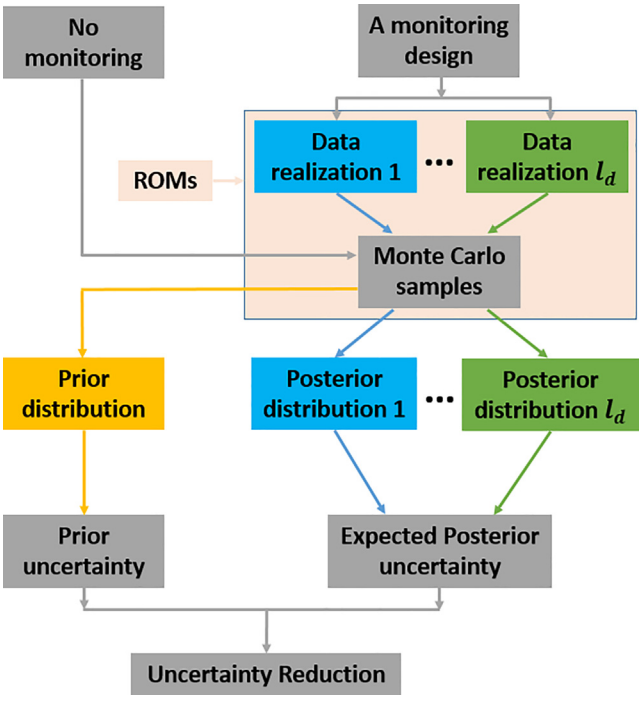


Fig. 2. Filtering and ROMs based workflow for CO_2 monitoring design.

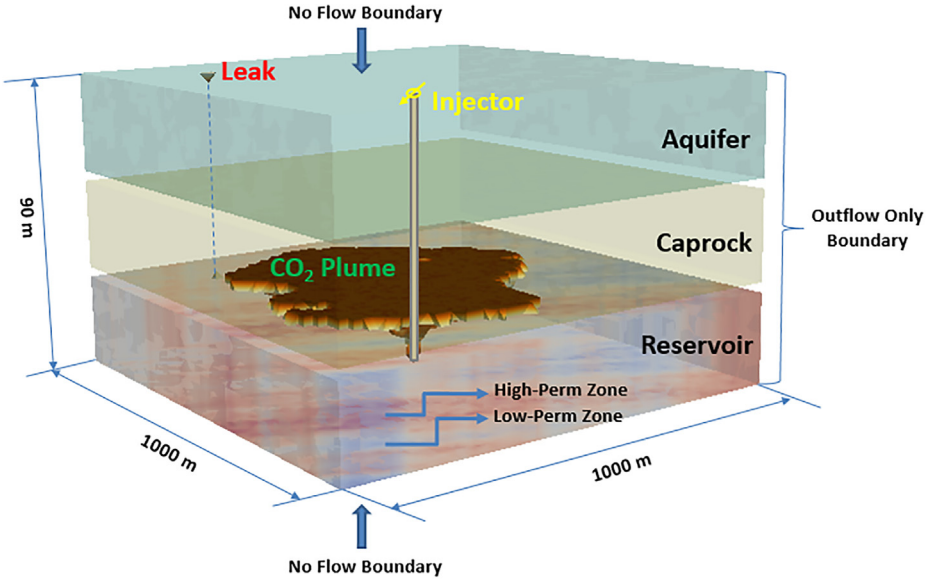


Fig. 3. Schematic of the model, in which storage reservoir and aquifer are separated by caprock. A CO_2 injection well is located in the center of the storage reservoir. The vertical axis is exaggerated 7 times.

Table 1
Uncertain parameters and their lower and upper bounds.

Uncertain parameters	Lower bound	Upper bound	Unit
Reservoir permeability multiplier	0.5	2	–
Permeability of leaky pathway(s)	−19	−14	$\log_{10}$ [m^2]
	(0.0001)	(10)	mD

The values of these are assumed to vary between the lower and upper bounds listed in Table 1.

The numerical mesh for the reservoir simulation model were generated using a grid generation toolkit LaGriT (http://lagrit.lanl.gov) [63]. The numerical mesh has 51 nodes in both x and y directions and 31 nodes in the z direction. The distance between two nodes in x and y directions is 20 m, and the distance between two nodes in z direction is 3 m. The number of nodes used for reservoir, cap rock and aquifer, respectively, is 26010, 26010 and 28611. The total number of nodes for simulation is 80631. 3D multi-phase simulations are performed by the Finite Element Heat and Mass transfer code (FEHM simulator) [59]. The sides of the reservoir, caprock layer and aquifer are Dirichlet boundaries allowing outflow of CO_2 and water at pressures above hydrostatic. Inflow is not allowed at these boundaries. The top and bottom of the model are no-flow boundaries. The model is initialized to a geothermal gradient of 0.03 °C/m with a temperature of 20 °C along the top. Pressures are initialized to a gradient of 9.81×10^{-3} MPa/m with a pressure of 0.2 MPa along the top. In this study, we consider a five-year CO_2 injection period with the injection rate equal to 0.1 million tons/year.

4. Results and analysis

4.1. Workflow validation

We validate the workflow for CO_2 monitoring design using a simple example. Fig. 4 shows the log-permeability distribution for the base model, the three potentially leaky locations (in yellow) and two optional locations (M1 and M2) for a single monitoring well design. All the monitoring data in this study are collected in the aquifer, similar to monitoring at the above zone monitoring intervals (AZMI) in the work of Sun et al. [43]. In this example, the monitoring data used is pressure. The monitoring frequency is once per month for the duration of 5 years, resulting in 60 monitoring data points. The objective function, M_c, is cumulative CO_2 leakage at the end of 5 years. In the model, we set up three material zones corresponding to three formations (i.e., Reservoir, Caprock and Aquifer). The cumulative CO_2 in different material zones can be output from the simulator FEHM. The cumulative CO_2 leakage is computed by summing up the CO_2 mass of in the aquifer and caprock. Note that our approach for monitoring design is based on quantifying the uncertainty reduction by monitoring pressure, temperature and/or CO_2 saturation at potential monitoring well(s). The pressure, temperature and saturation for a certain location (corresponding to a node in the simulator) at a particular time can also be computed in the simulator.

The data assimilation error tolerance, τ from Eq. (6), for pressure data is set equal to 0.002 MPa. Note that the choice of τ is site and case specific and should be based on engineering judgment that takes into consideration the measurement and modeling error.

Four uncertain model parameters are considered: the permeability multiplier for storage reservoir and the three permeability values for

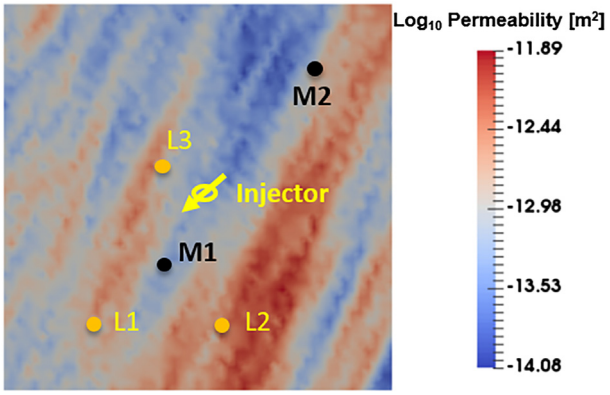


Fig. 4. Log-permeability distribution for the base model and the locations of potential leakages, potential monitoring wells and injector. The unit for permeability is $\log_{10}$ [m^2]. The darkest blue color corresponds to the lowest permeability, while the darkest red color corresponds to the highest permeability. Yellow dots indicate the potentially leaky points; black dots indicate the potential monitoring location(s); yellow circle with an arrow indicates the injector. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the three potentially leaky pathways. The lower and upper bounds for these uncertain parameters can be found in Table 1. We first run 500 training simulations generated by LHS sampling with the four uncertain parameters. It takes about 22 min for each simulation. In this study, 8 simulations are run concurrently. It takes around 23 h to finish all 500 training simulations.

Here, we choose one simulation from 500 training simulations to show when leakage occurs. The values of different parameters for the chosen model are shown in Table 2. Fig. 5 shows the cumulative CO_2 leakage over time computed from one chosen training simulation. It can be seen from Fig. 5 that CO_2 leakage happens after about 210 days' CO_2 injection. The leaked CO_2 saturation distribution at the top of the aquifer is shown in Fig. 6. We can see that the CO_2 is leaked through the potentially leaky pathway (L3) which is 141.4 m away from the injector, while no CO_2 leakage is observed within the 5 years' CO_2 injection at the potentially leaky pathways (L1 & L2). It should be noted that the permeability of L3 is higher than that of L1 and L2 for this specific example.

Based on the training simulations, ROMs for the monitoring data and the cumulative CO_2 leakage are constructed using MARS. Fig. 7 shows the quality of the ROMs tested by 10-fold cross validation [64]. Fig. 7(a) and (b), respectively, show the quality of ROMs from the cross validation for the cumulative CO_2 leakage and for an example data point pressure at M1 at year two. The correlation coefficients between the true values from simulations and the ROM predictions for the objective function and monitoring data point at year two, respectively, are 0.996 and 0.995, which indicates that the fidelity of the ROMs to the numerical simulations is high in this case and also in the other ROMs which are not shown here.

For the example introduced above, it is obvious that monitoring at M1 would obtain more information (i.e., higher uncertainty reduction in objective function) than monitoring at M2 since M1 is much closer to the potentially leaky points than M2; that is, simple intuition indicates that monitoring at M1 is a better choice than monitoring at M2. The purpose of this simple example is to demonstrate that our proposed workflow can successfully capture this intuitive result.

With the proposed workflow, the expected uncertainty reduction of the cumulative CO_2 leakage can be computed for the two monitoring data sets, one obtained from M1 and the other one obtained from M2. For each data set, 200 plausible realizations of monitoring data are generated following Step 2 described in Section 2.3. Fig. 8 shows one plausible monitoring data realization obtained from potential monitoring locations M1 and M2. It can be seen that the monitoring data predicted by ROM matches well with the simulation results. To calculate 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 plausible monitoring data realization D^j should be computed. Fig. 9 shows how these values, i.e., the prior and posterior uncertainties, are obtained. For simplicity, only one posterior CDF (obtained from specific data realization used for this example) for each potential monitoring location is shown in Fig. 9. In Fig. 9, the difference between the 10th and 90th percentiles of a cumulative density function (CDF) is the amount of uncertainty in a distribution. $U(P(M_c))$ denotes the amount of uncertainty in a cumulative CO_2 leakage distribution $P(M_c)$, i.e., prior uncertainty; $U(P(M_c|D^j))$ denotes the amount of uncertainty in $P(M_c)$ conditional to plausible monitoring data realization D^j, i.e., one posterior uncertainty. 200 plausible data realizations generate 200 posterior CDFs which produce 200 posterior uncertainties corresponding to each data realization. The expected posterior uncertainty which is the average value of the 200 posterior uncertainties can then be computed by Eq. (2).

With the computed prior and expected posterior uncertainties, the uncertainty reduction is calculated by Eq. (3). Fig. 10 shows the uncertainty reduction at the end of five years given the distribution of potential monitoring data sets that may be obtained at M1 and M2. In Fig. 10, the red[1] diamond indicates the prior uncertainty which is equal to 39.24 thousand tons for this example; the blue dash line indicates the expected posterior uncertainty which is the average value of the 200 posterior uncertainties corresponding to each data realization. It can be

[1] For interpretation of color in Fig. 10, the reader is referred to the web version of this article.

Table 2
The parameters for one chosen model from 500 training models.

Parameters	Value
CO_2 injection rate	3.17 kg/s
Thickness of caprock layer	30 m
Permeability of 1st potentially leaky pathway	2.19×10^{-17} m^2 (0.0219 mD)
Permeability of 2nd potentially leaky pathway	3.37×10^{-17} m^2 (0.0337 mD)
Permeability of 3rd potentially leaky pathway	2.97×10^{-16} m^2 (0.297 mD)
Distance between injector and 1st potentially leaky pathway	424.3 m
Distance between injector and 2nd potentially leaky pathway	360.6 m
Distance between injector and 3rd potentially leaky pathway	141.4 m
Permeability for aquifer layer	1×10^{-13} m^2 (100 mD)
Permeability for caprock layer	1×10^{-19} m^2 (0.0001 mD)
Reservoir permeability multiplier	1.88

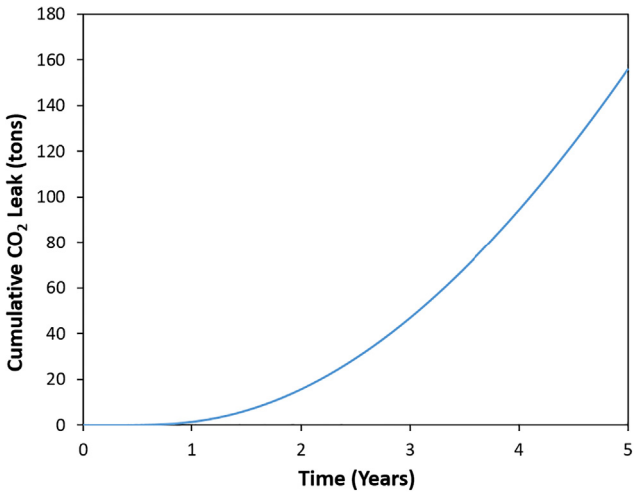


Fig. 5. Cumulative CO_2 leakage over time computed from one chosen training simulation.

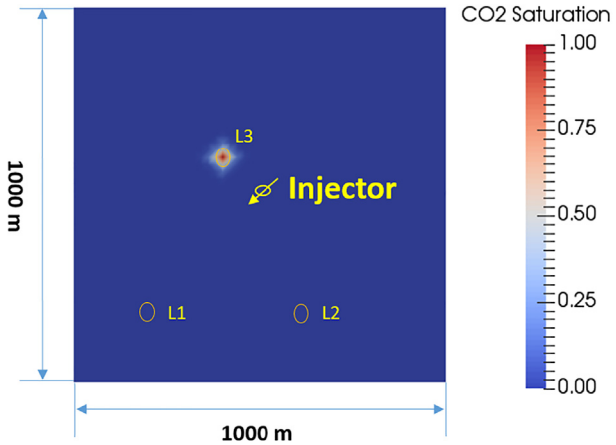
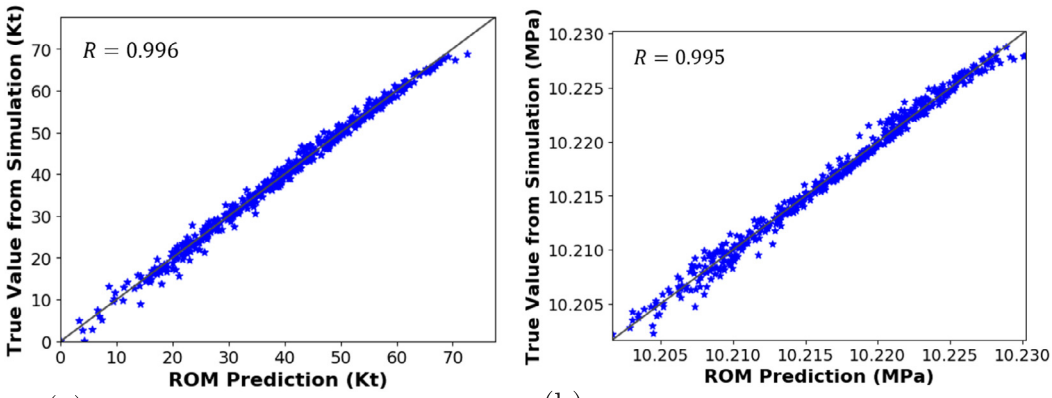


Fig. 6. Plan view (top of the aquifer) of CO_2 leakage at the end of five years based on the chosen training simulation. Yellow circles indicate the potentially leaky locations. Unit for CO_2 saturation is fraction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



(a) ROM quality for objective function (b) ROM quality for pressure at year two, M1

Fig. 7. Quality of the ROMs tested by 10-fold cross validation. R indicates the correlation coefficient between the true values from simulations and the ROM predictions for the objective function.

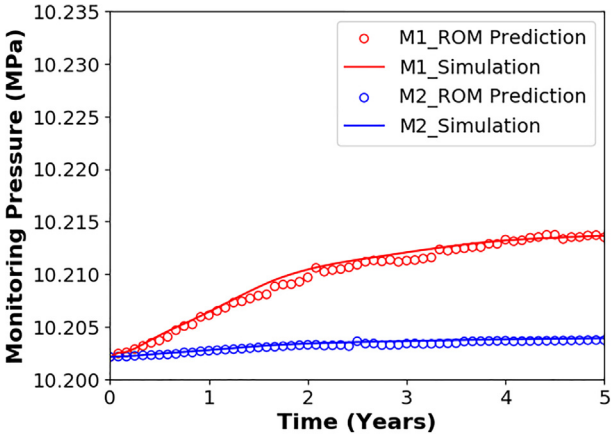


Fig. 8. One plausible monitoring data realization obtained from potential monitoring locations M1 and M2.

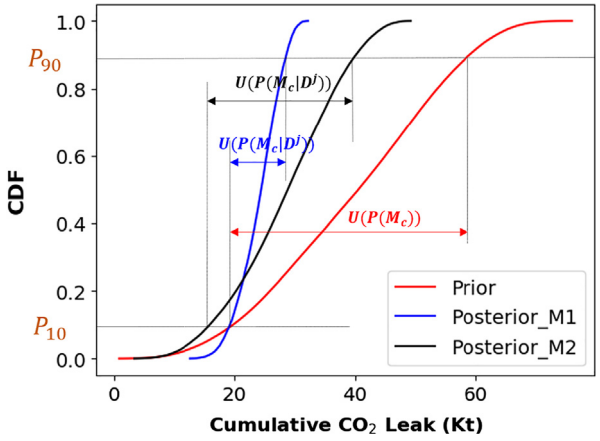


Fig. 9. CDF of cumulative CO_2 leakage for the prior distribution and posterior distributions conditional to plausible monitoring data D^j obtained from potential monitoring locations M1 and M2.

seen that monitoring at location M1 gives much higher uncertainty reduction than M2. Higher uncertainty reduction of the objective function indicates greater VOI in the monitoring data set to be obtained from location M1, thereby demonstrating our assertion. Through this example, we can see that our proposed workflow can be effectively used to determine the optimal CO_2 monitoring design from a set of alternative monitoring designs.

The evolution of the expected posterior uncertainty in the cumulative CO_2 leakage over time can be evaluated to understand the VOI of the monitoring data over time. Fig. 11 shows the expected posterior uncertainty in cumulative CO_2 leakage over time for two monitoring data sets obtained from monitoring locations M1 and M2, respectively. It can be observed that the pressure data set obtained from M1 is expected to reduce the uncertainty more than the pressure data obtained from M2 within the first 1 year. This is due to the fact that we can observe more significant pressure changes at the monitoring location M1 than M2 at year one as can be seen in Fig. 12. Almost no uncertainty can be reduced with the pressure data obtained from monitoring location M2 after 2 years, this is because significant pressure changes do not occur after 2 years at M2 for all the possible pressure realizations; see Fig. 12(b).

Sensitivity analysis is performed using Pearson correlation [65]. The plot in Fig. 13 shows the correlation coefficients between the four uncertain model parameters and the maximum pressure data in each pressure data realization as shown in Fig. 12. Recall that the four uncertain model parameters are the permeabilities of the three potentially leaky pathways (K_L1, K_L2 and K_L3) and the reservoir permeability multiplier (K_Reservoir). It can be observed from Fig. 13 that the pressure data have positive sensitivity to the permeabilities of the three potentially leaky pathways; that is to say the higher the permeability of the potentially leaky pathways, the higher the pressure that could be measured at locations M1 or M2. Moreover, the pressure data are most sensitive to the permeability of the third potentially leaky pathway (L3), due to the fact that the third potentially leaky pathway is closer to the injector than the other two potentially leaky pathways (see Fig. 4). It can also be seen from Fig. 13 that the pressure data have negative sensitivity to the reservoir permeability multiplier (K_Reservoir). This is because the higher the permeability of the reservoir, the faster the CO_2 migrates towards to the boundaries of the storage site, which results in less CO_2 leaking through the potentially leaky pathways and the strength of pressure signals will be reduced accordingly. We can also observe from Fig. 13 that the pressure data monitored at M1 are about twice as sensitive to the permeability of the first potentially leaky pathways (L1) than the pressure data monitored at M2. This is due to the fact that M1 is much closer to L1 than M2 (see Fig. 4).

Fig. 14 shows box plot of the samples remaining after the filtering approach for each monitoring data realization (200 realizations in total) of the pressure data obtained from monitoring locations M1 and M2. It can be seen from Fig. 14 that for any monitoring data realization, the number of remaining samples to construct the posterior distribution is over 1800, which is greater than the minimum required samples (30 samples) to construct a PDF to calculate P90 and P10 [47].

Fig. 15 shows the histograms for the posterior distributions of the objective function obtained from data realizations 1 and 100 as well as

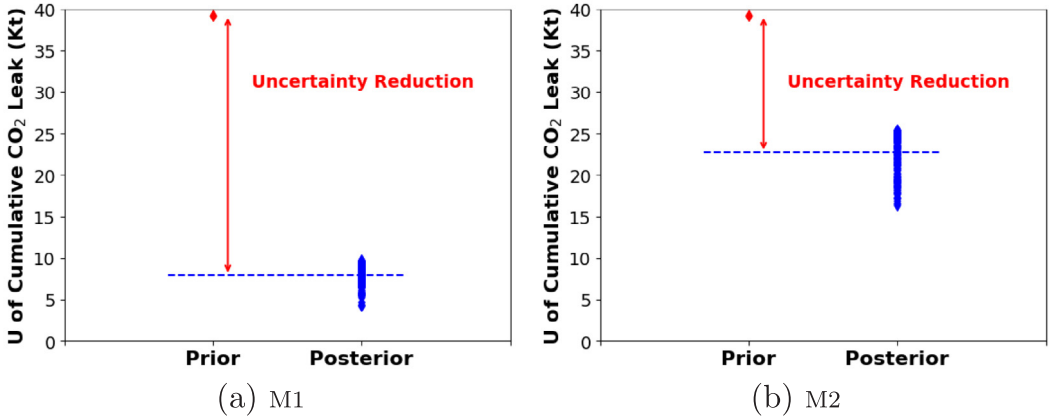


Fig. 10. The U value (uncertainty) of the prior and posterior distributions for the objective function.

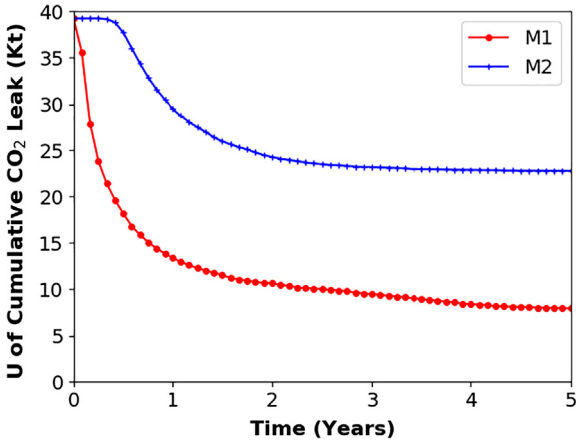


Fig. 11. Expected posterior uncertainty in objective function over time.

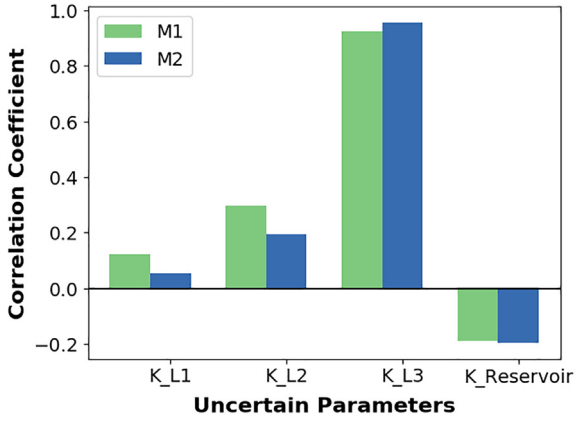


Fig. 13. Correlation coefficients between the uncertain model parameters and the maximum pressure data to be monitored.

the histogram for the prior distribution of the objective function. Note that the prior distribution of the objective function is generated by using the Latin hypercube sampling of the uncertain input parameters with uniform distributions and calculating the cumulative CO_2 leakage using the ROMs. We can observe from Fig. 15 that the variances of the posterior distributions calculated from data realizations 1 and 100 as well as the variances of the posterior distributions calculated from other data realizations which are not shown here are significantly decreased compared to the variance of prior distribution, and the variances of posterior distributions for M1 are much smaller than those for M2, which results in significantly higher uncertainty reduction in cumulative CO_2 leakage for M1 than for M2.

4.2. *Effect of measurement type and location in vertical direction*

Next, we will investigate the effect of different types of measurements (i.e., pressure, CO_2 saturation and temperature) as well as the monitoring location on the uncertainty reduction in the cumulative CO_2 leakage. We explore two different monitoring locations in the vertical direction at M1. The data measurement frequency is still once per month for five years. The data assimilation error tolerances, τ, for pressure, CO_2 saturation and temperature, respectively, are 0.002 MPa, 0.05 and 0.002 °C. All the other parameters that are mentioned here are the same as those in the first case.

Fig. 16 shows the expected posterior uncertainty in the objective

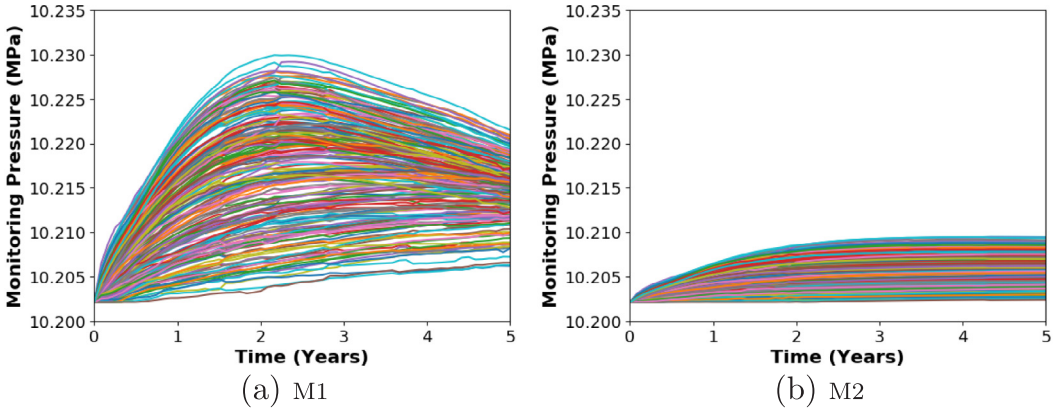


Fig. 12. Pressure for different data realizations obtained from monitoring locations M1 and M2.

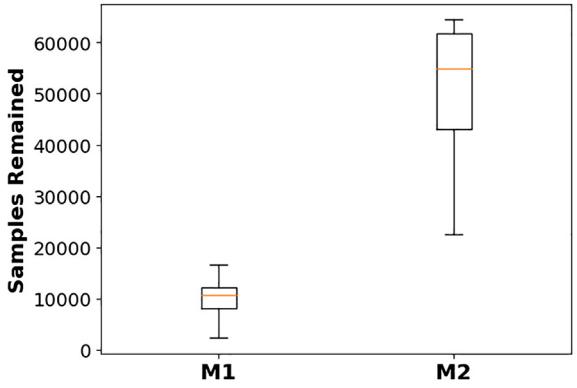


Fig. 14. Samples remaining after the filtering approach. The number of prior samples is 100,000.

function changing over time for different type of measurements and different vertical monitoring locations (at the bottom or top of the aquifer). The values of uncertainty reduction for different monitoring designs are recorded in Table 3. From Fig. 16 and Table 3, it can be observed that pressures (P) observations lead to higher uncertainty reduction compared to CO_2 saturation (S) and temperature (T). Additionally, pressure observations measured at the bottom of the aquifer give slightly higher uncertainty reduction than those measured at the top of the aquifer. It can also be observed that CO_2 observations measured at the bottom of the aquifer do not lead to any reduction in uncertainty. This is due to the fact that because of the buoyancy, CO_2 primarily pools towards the top of the aquifer which results in very little to no CO_2 saturation observed at the bottom of the aquifer; see Fig. 17(c). Since no CO_2 is observed during the first year at the monitoring location (see Fig. 17(d)), no uncertainty reduction is observed at the beginning of the first year; see Fig. 16(d). However, after one year, CO_2 saturation at the monitoring location starts changing, which leads to significant uncertainty reduction. It is important to note that the lag of the uncertainty reduction with CO_2 saturation data set compared to the uncertainty reduction with pressure data set is due to the fact that pressure signals travel much faster than the signals for CO_2 saturation in porous media. From Fig. 17(e), we cannot observe any significant temperature change over time, resulting in no uncertainty reduction when the temperature data were measured at the bottom of the aquifer. However, in Fig. 17(f), we can observe the temperature change due to the leaked CO_2 accumulated at the top of the aquifer, but the difference in the temperature for different realizations at a particular time is negligible for the first two and half years (i.e., the temperature data are not sensitive to the uncertain parameters for the first two and half years), which results in no uncertainty reduction in the first two and half years. After two and half years, the temperature data measured at the top of aquifer do result in a small amount of uncertainty reduction.

We also computed the effect of joint measurements of multiple types of data on reduction in uncertainty. The results are summarized in Table 3. A combination of pressure and CO_2 saturation data sets leads to a slightly higher uncertainty reduction (1.09 thousand tons) than that due to pressure data set alone. On the other hand, combined measurement of pressure and temperature data sets leads to only 0.17 thousand tons higher uncertainty reduction compared to the pressure data set alone. Furthermore, only 0.01 thousand tons more uncertainty can be reduced by adding the temperature data set to the combined pressure and CO_2 saturation data sets. The above analysis indicates that one can benefit more from monitoring both pressure and CO_2 saturation data than monitoring only pressure data. However, monitoring pressure, CO_2 saturation, and temperature data sets together will not add any VOI. The prior and expected posterior uncertainties for different types of monitoring strategies (monitoring at the top of the aquifer) are presented in Fig. 18, which clearly shows the distributions for posterior uncertainties computed from different type of monitoring strategies.

4.3. Effect of the number of monitoring wells

To investigate the effect of the number of monitoring wells on the decision of monitoring strategy, we develop a second example as shown in Fig. 19. More potentially leaky points (6 leaky points) are considered in this example than those in the first example. 500 training samples are generated by LHS sampling with the seven uncertain parameters, i.e., 6 uncertain parameters for the permeability of 6 potentially leaky points and one uncertain parameter for the reservoir permeability multiplier. All the training samples are simulated by the FEHM simulator. The type of data to be measured is pressure. All the other parameters for the example set-up that are not mentioned here are the same as described in the first example.

Based on the inputs and outputs of training simulations, we build the ROMs for both objective function (cumulative CO_2 leakage) and the 60 data measurements. Fig. 20 shows the quality of the ROMs tested by 10-fold cross validation. The correlation coefficients between the simulated values and the ROMs predicted values for the objective function and the example monitoring data point at year one, respectively, are 0.995 and 0.994, which indicates the constructed ROMs as well as the other ROMs not shown here have high fidelity to the numerical simulations.

The specifications of the four schemes, according to the different number of the monitoring locations and the corresponding uncertainty reductions in the objective function, are listed in Table 4. The expected posterior uncertainty in the objective function over time for different schemes is shown in Fig. 21. From Table 4 and Fig. 21, it can be

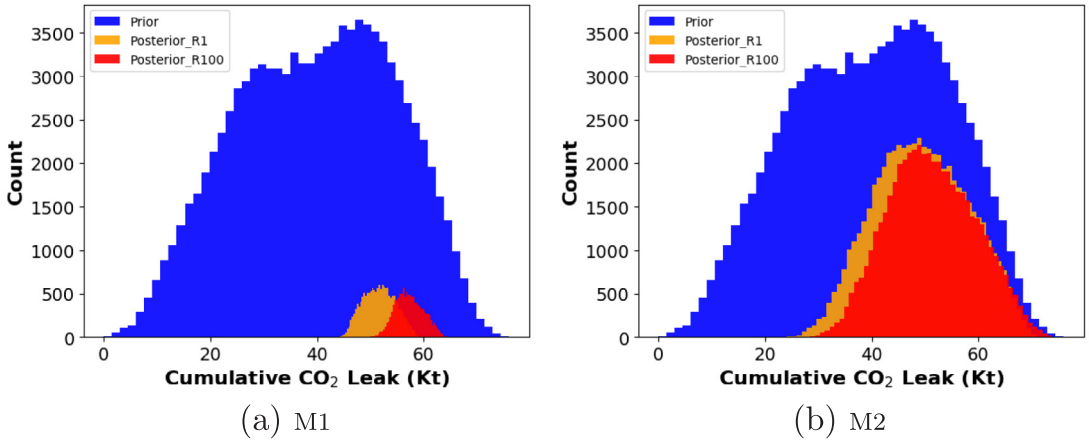


Fig. 15. The histograms for the posterior distributions obtained from the data realizations 1 and 100 as well as the histogram for the prior distribution.

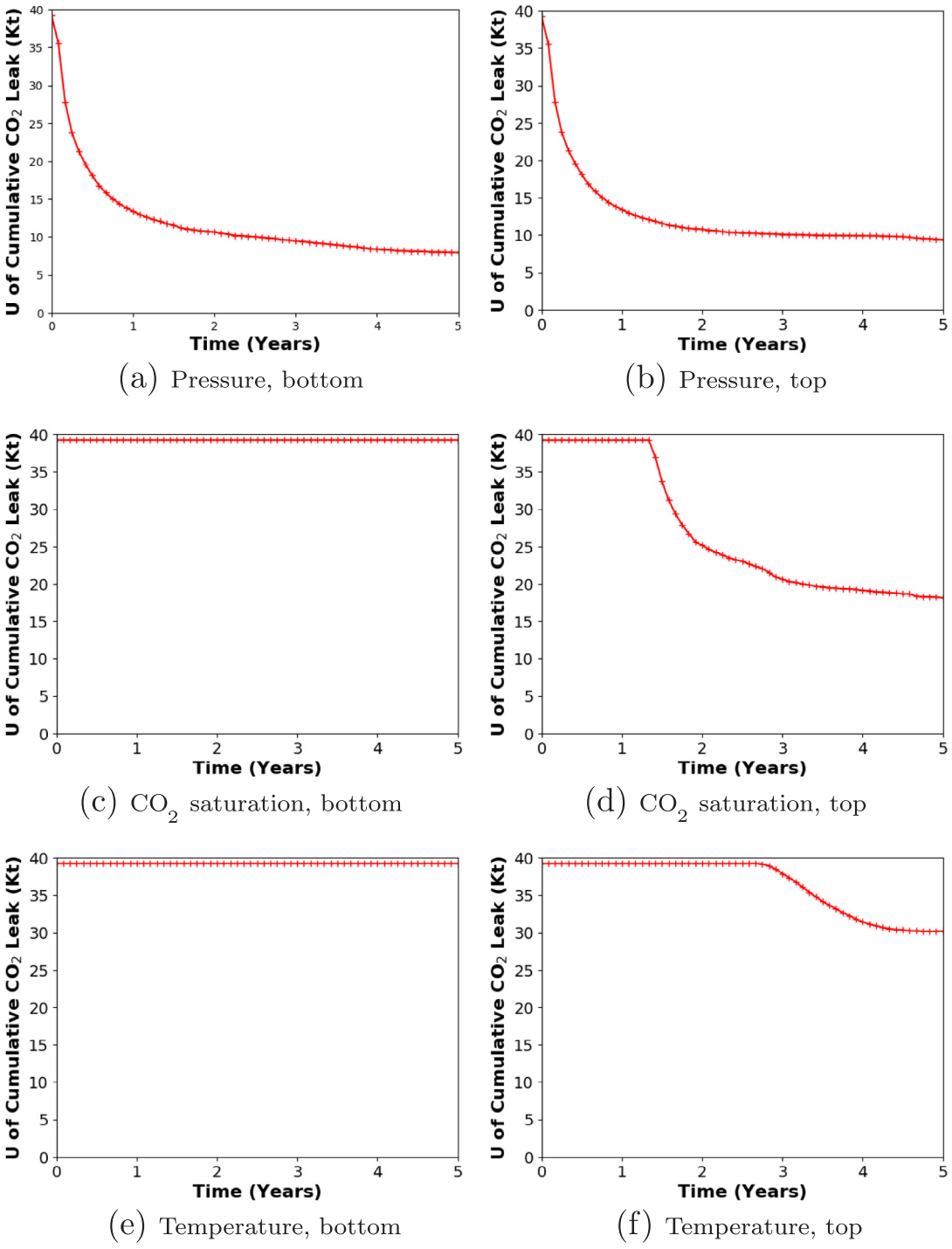


Fig. 16. Expected posterior uncertainty in objective function over time.

observed that measurements at M2 would give a 24.3% higher uncertainty reduction in cumulative CO_2 leakage than measurements at M1. A combination of the measurements at M1 and M2 would give a much higher uncertainty reduction than each individual measurement. For the scheme 4, addition of a third monitoring well at location M3 leads to only a slight increase in the uncertain reduction in the objective function. The result indicates that the additional costs of adding a third monitoring location M3 will not may not be justified given the insignificant effect on uncertainty reduction resulting from the data collected at that location.

5. Discussion

The high cost of performing monitoring operations often requires evaluating the potential value of monitoring data (e.g., the potential value of pressure, temperature and/or CO_2 saturation data) before the actual monitoring strategy takes place. The workflow proposed here can be used to select a monitoring design that is optimal under many potential leakage scenarios. Even though the examples used in our study demonstrated how monitoring data from a shallow aquifer can be used, the proposed workflow can be applied to monitoring data

Table 3
Uncertainty reduction under different type of measurements at end of five years.

Measurement	Uncertainty reduction (Kt), top	Uncertainty reduction (Kt), bottom
P	29.88	31.29
S	21.10	0
T	9.05	0
P + S	30.97	\
P + T	30.05	\
S + T	21.40	\
P + S + T	30.98	\

collected at any location within the CO_2 storage complex ranging from storage reservoir to shallow aquifer to atmosphere. The potential value of monitoring data can be evaluated by the presented workflow.

At a CO_2 storage field operation, an optimal monitoring schedule based on the VOI approach described here can be used to collect monitoring data. The monitoring data can be assimilated to calibrate the uncertain field model(s) using the traditional history matching or data assimilation methods, such as ensemble Kalman filter (EnKF) [66] and ensemble smoother with multiple data assimilation (ES-MDA) [67] methods. The calibrated model(s) can subsequently be used to improve accuracy in predictions of the long-term fate of stored CO_2.

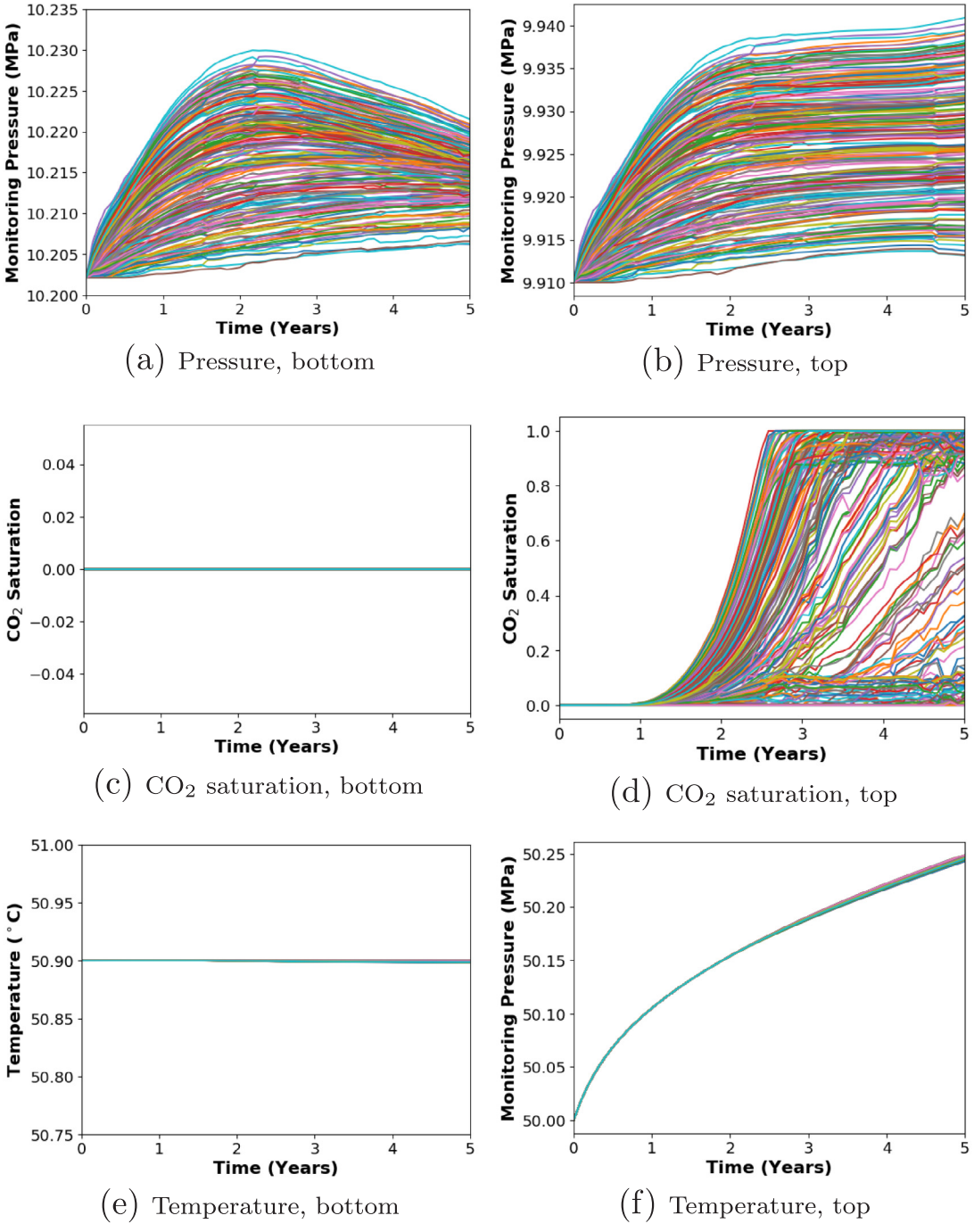


Fig. 17. Data realizations obtained from different measurements. Unit for CO_2 saturation is fraction.

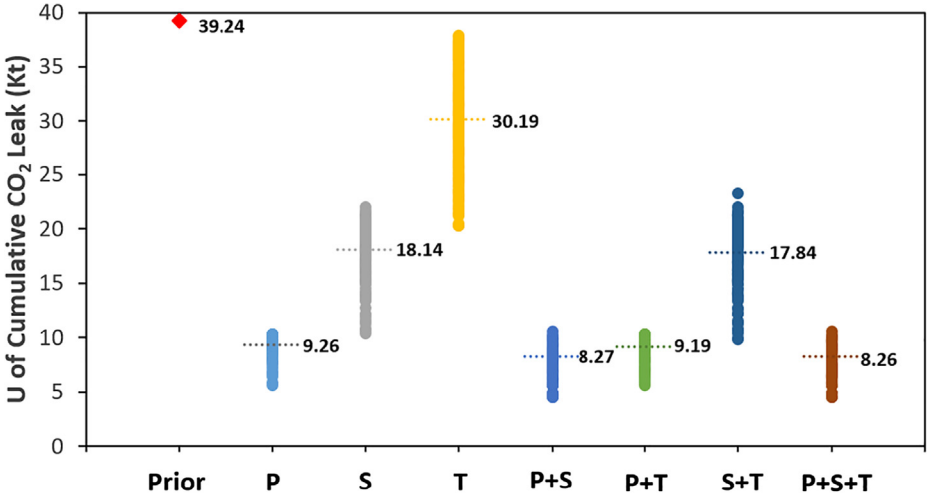


Fig. 18. The prior and posterior uncertainties for different types of monitoring strategies (monitoring at the top of the aquifer). The dash lines in this plot indicate the expected (average) posterior uncertainties for different monitoring strategies.

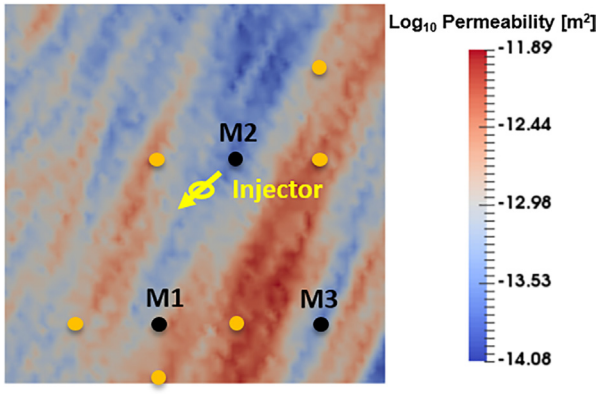
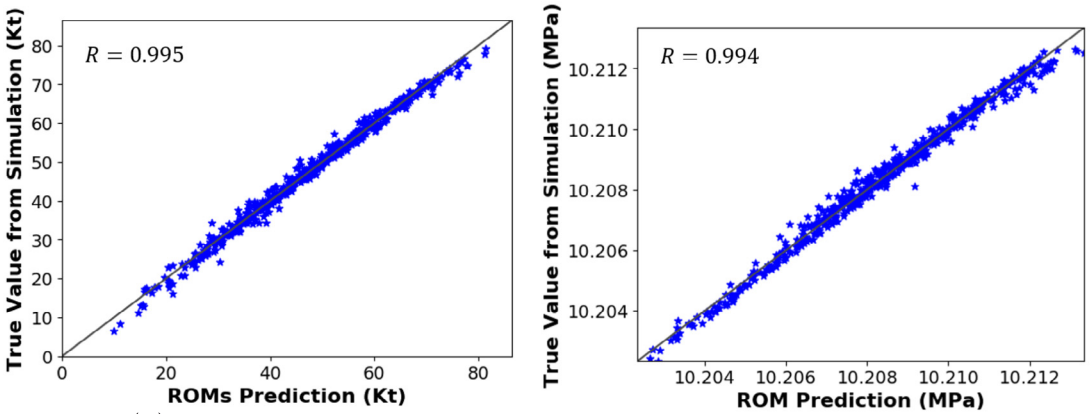


Fig. 19. Locations of potential leaky points, potential monitoring wells and injector, example 2. The darkest blue color corresponds to the lowest permeability, while the darkest red color corresponds to the highest permeability. The unit for permeability is $\log_{10}$ [m^2]. Yellow dots indicate potential leaky locations; black dots indicate potential monitoring well locations; yellow circle with an arrow indicates the injector. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

6. Conclusions

In this study, a workflow based on a machine learning technique and uncertainty quantification method is proposed for geologic CO_2 sequestration monitoring design. We use the uncertainty reduction in simulated cumulative CO_2 leakage as a metric to quantify the potential value of yet-to-be measured monitoring 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 utilizing different monitoring designs and has been demonstrated using a synthetic CO_2 storage site.
(2) The effect of different types of measurements (pressure, CO_2 saturation and temperature) and the effect of monitoring locations (at the top or bottom of the aquifer) on the choice of monitoring design is investigated. It is observed that pressure data has more value of information than CO_2 saturation and temperature data, while the temperature data has the least value of information.
(3) Simultaneous use of multiple types of monitoring signals can reduce the uncertainty more than using a single type of signal though the relative benefits of utilizing different types of signals can vary.
(4) The incremental reduction in uncertainty in the cumulative CO_2 leakage may not increase proportional to the number of monitoring locations.



(a) ROM quality for obj (b) ROM quality for pressure at year one, M1

Fig. 20. Quality of the ROMs tested by 10-fold cross validation, example 2. *R* indicates the correlation coefficient between the true values from simulations and the ROM predictions for the objective function.

Table 4
Uncertainty reduction under different type of measurements.

Scheme	Measurement locations	Uncertainty reduction (Kt)
1	M1	21.89
2	M2	27.22
3	M1 + M2	31.96
4	M1 + M2 + M3	32.01

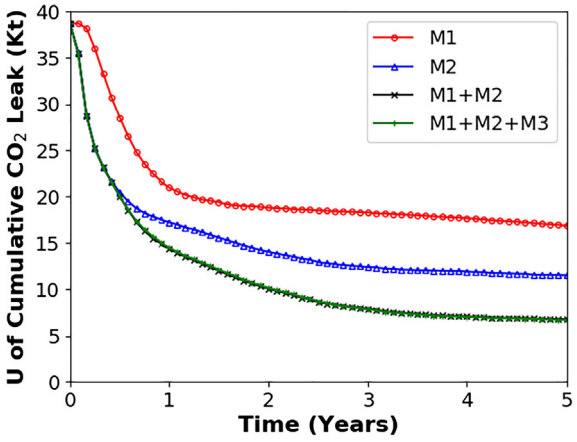


Fig. 21. Expected posterior uncertainty in objective function over time, example 2. The green curve almost overlaps with black curve. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Acknowledgments

This work 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

[1] Metz B. Carbon dioxide capture and storage: special report of the intergovernmental panel on climate change. Cambridge University Press; 2005.
[2] Michael K, Golab A, Shulakova V, Ennis-King J, Allinson G, Sharma S, et al. Geological storage of CO_2 in saline aquifers—a review of the experience from existing storage operations. Int J Greenhouse Gas Control 2010;4(4):659–67.
[3] Sharma SS. Determinants of carbon dioxide emissions: empirical evidence from 69 countries. Appl Energy 2011;88(1):376–82.
[4] Viebahn P, Vallentin D, Höller S. Prospects of carbon capture and storage (CCS) in India's power sector—an integrated assessment. Appl Energy 2014;117:62–75.
[5] Viebahn P, Vallentin D, Höller S. Prospects of carbon capture and storage (CCS) in China's power sector—an integrated assessment. Appl Energy 2015;157:229–44.
[6] Zhou W, Wang T, Yu Y, Chen D, Zhu B. Scenario analysis of CO_2 emissions from China's civil aviation industry through 2030. Appl Energy 2016;175:100–8.
[7] Aminu MD, Nabavi SA, Rochelle CA, Manovic V. A review of developments in carbon dioxide storage. Appl Energy 2017;208:1389–419.
[8] Chen B, Harp D, Lin Y, Keating E, Pawar R. A machine-learning and filtering based data assimilation framework for geologic carbon sequestration monitoring optimization. In: AGU fall meeting abstracts; 2017a.
[9] Lin Y, Harp D, Chen B, Pawar R. Geologic carbon sequestration leakage detection: a physics-guided machine learning approach. In: AGU fall meeting abstracts; 2017.
[10] Dai Z, Middleton R, Viswanathan H, Fessenden-Rahn J, Bauman J, Pawar R, et al. An integrated framework for optimizing CO_2 sequestration and enhanced oil recovery. Environ Sci Technol Lett 2013;1(1):49–54.
[11] Zhao X, Liao X, He L. The evaluation methods for CO_2 storage in coal beds, in China. J Energy Inst 2016;89(3):389–99.
[12] Herzog H, Golomb D, Zemba S. Feasibility, modeling and economics of sequestering power plant CO_2 emissions in the deep ocean. Environ Prog Sustain Energy 1991;10(1):64–74.
[13] Shabani B, Vilcáez J. A fast and robust TOUGH2 module to simulate geological CO_2 storage in saline aquifers. Comput Geosci 2018;111:58–66.
[14] Benson SM, Myer L. Monitoring to ensure safe and effective geologic sequestration of carbon dioxide. In: Workshop on carbon dioxide capture and storage; 2003.
[15] Wilkin RT, DiGiulio DC. Geochemical impacts to groundwater from geologic carbon sequestration: controls on pH and inorganic carbon concentrations from reaction path and kinetic modeling. Environ Sci Technol 2010;44(12):4821–7.
[16] Keating E, Bacon D, Carroll S, Mansoor K, Sun Y, Zheng L, et al. Applicability of aquifer impact models to support decisions at CO_2 sequestration sites. Int J Greenhouse Gas Control 2016;52:319–30.
[17] Song J, Zhang D. Comprehensive review of caprock-sealing mechanisms for geologic carbon sequestration. Environ Sci Technol 2012;47(1):9–22.
[18] Harp DR, Pawar R, Carey JW, Gable CW. Reduced order models of transient CO_2 and brine leakage along abandoned wellbores from geologic carbon sequestration reservoirs. Int J Greenhouse Gas Control 2016;45:150–62.
[19] Jia B, Tsau J-S, Barati R. Role of molecular diffusion in heterogeneous, naturally fractured shale reservoirs during CO_2 huff-n-puff. J Petrol Sci Eng 2018;164:31–42.
[20] Godec M, Koperna G, Petrusak R, Oudinot A. Enhanced gas recovery and CO_2 storage in gas shales: a summary review of its status and potential. Energy Procedia 2014;63:5849–57.
[21] Tapia JFD, Lee J-Y, Ooi RE, Foo DC, Tan RR. Optimal CO_2 allocation and scheduling in enhanced oil recovery (EOR) operations. Appl Energy 2016;184:337–45.
[22] Ampomah W, Balch R, Cather M, Rose-Coss D, Dai Z, Heath J, et al. Evaluation of CO_2 storage mechanisms in CO_2 enhanced oil recovery sites: application to Morrow sandstone reservoir. Energy Fuels 2016;30(10):8545–55.
[23] Ampomah W, Balch R, Cather M, Will R, Gunda D, Dai Z, et al. Optimum design of CO_2 storage and oil recovery under geological uncertainty. Appl Energy 2017;195:80–92.
[24] Welkenhuysen K, Rupert J, Compernolle T, Ramirez A, Swennen R, Piessens K. Considering economic and geological uncertainty in the simulation of realistic investment decisions for CO_2-EOR projects in the North Sea. Appl Energy 2017;185:745–61.
[25] Francu J, Pereszlényi M, Riis F, Prokop O, Jurenka L, Hladík V, et al. 3D geological model of potential CO_2 storage: abandoned oil and gas field LBr-1in the Vienna Basin. Energy Procedia 2017;114:2772–80.
[26] Jin L, Hawthorne S, Sorensen J, Pekot L, Kurz B, Smith S, et al. Advancing CO_2 enhanced oil recovery and storage in unconventional oil play – experimental studies on Bakken shales. Appl Energy 2017;208:171–83.
[27] Chen B, Reynolds AC. CO_2 water-alternating-gas injection for enhanced oil recovery: optimal well controls and half-cycle lengths. Comput Chem Eng 2018;113:44–56.
[28] Pawar R, Bromhal G, Dilmore R, Foxall B, Jones E, Oldenburg C, et al. Quantification of risk profiles and impacts of uncertainties as part of US DOE's National Risk Assessment Partnership (NRAP). Energy Procedia 2013;37:4765–73.
[29] Pawar R, Bromhal G, Carroll S, Chu S, Dilmore R, Gastelum J, et al. Quantification of key long-term risks at CO_2 sequestration sites: latest results from US DOE's National Risk Assessment Partnership (NRAP) Project. Energy Procedia 2014;63:4816–23.
[30] Pawar RJ, Bromhal GS, Carey JW, Foxall W, Korre A, Ringrose PS, et al. Recent advances in risk assessment and risk management of geologic CO_2 storage. Int J Greenhouse Gas Control 2015;40:292–311.
[31] Yang Y-M, Small MJ, Ogretim EO, Gray DD, Wells AW, Bromhal GS, et al. A Bayesian belief network (BBN) for combining evidence from multiple CO_2 leak detection technologies. Greenhouse Gases: Sci Technol 2012;2(3):185–99.
[32] Ren B, Ren S, Zhang L, Chen G, Zhang H. Monitoring on CO_2 migration in a tight oil reservoir during CCS-EOR in Jilin Oilfield China. Energy 2016;98:108–21.
[33] Keating E, Dai Z, Dempsey D, Pawar R. Effective detection of CO_2 leakage: a comparison of groundwater sampling and pressure monitoring. Energy Procedia 2014;63:4163–71.
[34] Wang Z, Small MJ. A Bayesian approach to CO_2 leakage detection at saline sequestration sites using pressure measurements. Int J Greenhouse Gas Control 2014;30:188–96.
[35] Azzolina NA, Small MJ, Nakles DV, Bromhal GS. Effectiveness of subsurface pressure monitoring for brine leakage detection in an uncertain CO_2 sequestration system. Stochast Environ Res Risk Assess 2014;28(4):895–909.
[36] Dai Z, Keating E, Bacon D, Viswanathan H, Stauffer P, Jordan A, et al. Probabilistic evaluation of shallow groundwater resources at a hypothetical carbon sequestration site. Sci Rep 2014;4:4006.
[37] Yang C, Hovorka SD, Treviño RH, Delgado-Alonso J. Integrated framework for assessing impacts of CO_2 leakage on groundwater quality and monitoring-network efficiency: case study at a CO_2 enhanced oil recovery site. Environ Sci Technol 2015;49(14):8887–98.
[38] Zhang L, Ren B, Huang H, Li Y, Ren S, Chen G, et al. CO_2 EOR and storage in Jilin oilfield China: monitoring program and preliminary results. J Petrol Sci Eng 2015;125:1–12.
[39] Yang Y-M, Small MJ, Ogretim EO, Gray DD, Bromhal GS, Strazisar BR, et al. Probabilistic design of a near-surface CO_2 leak detection system. Environ Sci Technol 2011;45(15):6380–7.
[40] Yang Y-M, Dilmore R, Mansoor K, Carroll S, Bromhal G, Small M. Risk-based monitoring network design for geologic carbon storage sites. In: 13th International conference on greenhouse gas control technologies, 14–18 November, Lausanne, Switzerland; 2016.
[41] Seto C, McRae G. Reducing risk in basin scale CO_2 sequestration: a framework for integrated monitoring design. Environ Sci Technol 2011;45(3):845–59.
[42] Seto C, McRae G. Reducing risk in basin scale sequestration: a Bayesian model selection framework for improving detection. Energy Procedia 2011;4:4199–206.
[43] Sun AY, Nicot J-P, Zhang X. Optimal design of pressure-based, leakage detection monitoring networks for geologic carbon sequestration repositories. Int J Greenhouse Gas Control 2013;19:251–61.
[44] Cameron D. Optimization and monitoring of geological carbon storage operations [Ph.D. thesis]. Stanford (California, USA): Stanford University; 2013.

[45] Dai C, Li H, Zhang D, Xue L. Efficient data-worth analysis for the selection of surveillance operation in a geologic CO_2 sequestration system. Greenhouse Gases: Sci Technol 2015;5(5):513–29.
[46] Dai C, Xue L, Zhang D, Guadagnini A. Data-worth analysis through probabilistic collocation-based Ensemble Kalman Filter. J Hydrol 2016;540:488–503.
[47] Chen B, He J, Wen X, Chen W, Reynolds A. Uncertainty quantification and value of information assessment using proxies and Markov chain Monte Carlo method for a pilot project. J Petrol Sci Eng 2017;157:328–39.
[48] Liu N, Oliver DS, et al. Evaluation of Monte Carlo methods for assessing uncertainty. SPE J 2003;8(02):188–95.
[49] Emerick AA, Reynolds AC, et al. Combining the ensemble Kalman filter with Markov-Chain Monte Carlo for improved history matching and uncertainty characterization. SPE J 2012;17(02):418–40.
[50] Chen B, He J, Wen X, Chen W, Reynolds A. Pilot design analysis using proxies and Markov chain Monte Carlo method. In: ECMOR XV-15th European conference on the mathematics of oil recovery; 2016.
[51] Caers J. Modeling uncertainty in the earth sciences. John Wiley & Sons; 2011.
[52] Friedman JH. Multivariate adaptive regression splines. Ann Stat 1991:1–67.
[53] Oliver DS, Reynolds AC, Liu N. Inverse theory for petroleum reservoir characterization and history matching. Cambridge University Press; 2008.
[54] Zhang K, Zhang X, Zhang L, Li L, Sun H, Huang Z, et al. Assisted history matching for the inversion of fractures based on discrete fracture-matrix model with different combinations of inversion parameters. Comput Geosci 2017;21(5–6):1365–83.
[55] Zhang K, Ma X, Li Y, Wu H, Cui C, Zhang X, et al. Parameter prediction of hydraulic fracture for tight reservoir based on micro-seismic and history matching. Fractals 2018:1840009.
[56] Guo Z, Reynolds AC, Zhao H, et al. A physics-based data-driven model for history matching, prediction, and characterization of waterflooding performance. SPE J 2018;23(02):367–95.
[57] Le DH, Reynolds AC. Optimal choice of a surveillance operation using information theory. Comput Geosci 2014;18(3–4):505–18.
[58] Iman RL. Latin hypercube sampling. Wiley Online Library; 2008.
[59] Zyvoloski G. FEHM: a control volume finite element code for simulating subsurface multi-fluid heat and mass transfer. Los Alamos unclassified Report LA-UR-07-3359.
[60] Jordan AB, Stauffer PH, Harp D, Carey JW, Pawar RJ. A response surface model to predict CO_2 and brine leakage along cemented wellbores. Int J Greenhouse Gas Control 2015;33:27–39.
[61] Keating EH, Harp DH, Dai Z, Pawar RJ. Reduced order models for assessing CO_2 impacts in shallow unconfined aquifers. Int J Greenhouse Gas Control 2016;46:187–96.
[62] Chen B, Reynolds AC. Optimal control of ICV's and well operating conditions for the water-alternating-gas injection process. J Petrol Sci Eng 2017;149:623–40.
[63] George D, Kuprat A, Carlson N, Gable C. LaGriT - Los Alamos Grid Toolbox. < http://lagrit.lanl.gov/ > .
[64] Geisser S. Predictive inference vol. 55. CRC Press; 1993.
[65] Pearson K. Note on regression and inheritance in the case of two parents. Proc Roy Soc Lond 1895;58:240–2.
[66] Evensen G. Data assimilation: the ensemble Kalman filter. Springer Science & Business Media; 2009.
[67] Emerick AA, Reynolds AC. Ensemble smoother with multiple data assimilation. Comput Geosci 2013;55:3–15.