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
0306-2619/ Published by Elsevier Ltd.

Nomenclature

Symbols

d_i	ith individual data point that would be obtained if the monitoring design were implemented
$\tilde{d}^j$	jth data vector corresponding to each $\tilde{m}^j$
$\hat{d}^k$	kth data vector corresponding to each $\hat{m}^k$
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
$\tilde{m}^j$	jth model realization generated from prior probability density function of m
$\hat{m}^k$	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.