Uncertainty analyses of CO2 plume expansion subsequent to wellbore CO2 leakage into aquifers

Zhangshuan Hou a,*, Diana H. Bacon a, Dave W. Engel b, Guang Lin b, Yilin Fang a, Huiying Ren a, Zhufeng Fang a

a Earth Systems Science Division, Pacific Northwest National Laboratory, P.O. Box 999, Richland, WA 99352, United States
b Computational Science & Mathematics Division, Pacific Northwest National Laboratory, P.O. Box 999, Richland, WA 99352, United States

A R T I C L E   I N F O

Article history:
Received 4 March 2013
Received in revised form 23 April 2014
Accepted 14 May 2014
Available online 5 June 2014

Keywords:
Carbon sequestration
Uncertainty quantification
Wellbore leakage
Response surface
Adaptive sampling

A B S T R A C T

This study focused on CO2 plume expansion subsequent to wellbore CO2 leakage into a shallow unconfined aquifer post-CO2 injection. The target response variables included CO2 plume size, as well as flux to the atmosphere. Many processes contribute to CO2 plume expansion in the aquifer; here we considered process and model parameters including those affecting the abandoned well leak rate, aquifer hydraulic properties, and aquifer geochemistry. In order to identify the significant factors affecting leakage, we adopted an uncertainty quantification framework to quantify input uncertainty, generate exploratory samples effectively, perform scalable numerical simulations, visualize output uncertainty, and evaluate input–output relationships. We combined quasi-Monte Carlo and adaptive sampling approaches to reduce the number of forward calculations while fully exploring the input parameter space and quantifying the output uncertainty. The CO2 migration was simulated with STOMP-CO2 (water–salt–CO2 module). Response surfaces of model outputs were built with respect to input parameters to determine the individual and combined effects. Four most significant parameters were identified to be dominating the CO2 plume expansion process subsequent to wellbore CO2 leakage: distance between the leaky and injection wells, maximum leak rate, porosity, and hydraulic conductivity.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

The success of geological carbon sequestration relies on the efficacy and security of CO2 storage and on public acceptance, which requires the modelers to use reliable simulators of CO2 migration, make reasonable assumptions of simplifications, and provide risk analyses on natural, wellbore, and/or fault leakage and potential contamination of groundwater. Gaining a better understanding and quantification of these risks is needed to make sure that the sequestration practice will comply with safety standards during and after injection. There could be CO2 leakage through the seal into the atmosphere or groundwater aquifers, which either results in failure to meet the target of CO2 containment or causes groundwater pollution. Induced seismicity is another concern, as injection might result in fracture initiation and fault reactivation as pressure builds up around the injection location. Damen et al. (2006) summarized the risks of CO2 storage in a geological reservoir and divided them into five categories: CO2 leakage (CO2 migration out of the reservoir to other formations, from where it may escape into the atmosphere); CH4 leakage (CO2 injection might cause CH4 present in the reservoir to migrate out of the reservoir to other formations and possibly into the atmosphere); seismicity (the occurrence of (micro) earth tremors caused by CO2 injection); ground movement (subsidence or uplift of the earth’s surface as a consequence of pressure changes induced by CO2 injection); and displacement of brine caused by injection of CO2 in open aquifers.

Efforts have been made to assess the risks associated with geological sequestration of CO2 with underground injection through laboratory and field experiments, model development, and deployment of a risk assessment framework. Bachu (2003) developed approaches for the assessment and ranking of sedimentary basins in terms of their suitability for CO2 sequestration. Wilson et al. (2003) reviewed the risks and regulatory history of deep underground injection on the U.S. mainland and surrounding continental shelf. Oldenburg (2008) developed a screening and ranking framework to evaluate potential geologic storage sites on the basis of health, safety, and environmental risk arising from CO2 leakage. Many studies focused on wellbore leakage and aquifer groundwater quality issues; for example, Viswanathan et al. (2008) assessed the fate of CO2 via wellbore release and the resulting impacts of...
CO₂ on a shallow aquifer and release to the atmosphere, using a system level model CO₂-PENS. Nordbotten et al. (2005) developed a semi-analytical solution framework that allows simple and efficient prediction of leakage rates for the case of injection of supercritical CO₂ into a brine-saturated deep aquifer. In Lu et al. (2010), a laboratory batch experiment was conducted to explore the range of CO₂ impact on groundwater quality of a spectrum of representative aquifers in the Gulf Coast region, USA. Results show that CO₂ elevated the concentrations of many cations within hours or days. Crow et al. (2010) determined that the potential for wellbore leakage depends in part on the quality of the original construction as well as geochemical and geomechanical stresses that occur over its life cycle. Choi et al. (2013) identified wellbore integrity as a significant risk factor and discussed the effect of the quality of cement work on corrosion of casing steel, possible pathways in cement that permit corrosive fluid to attack casing, and corrosion behavior of carbon steel under CO₂ geologic storage conditions. Kharaka et al. (2010) evaluated changes in the chemistry of shallow groundwater related to the 2008 injection of CO₂ at the Zero Emission Research and Technology Field site, Bozeman, Montana. Trautz et al. (2013) studied the effect of dissolved CO₂ on a shallow groundwater system using controlled-release field experiments. Yang et al. (2013) conducted a single-well push-pull test for assessing potential impacts of CO₂ leakage on groundwater quality in a shallow Gulf Coast aquifer in Cranfield, Mississippi. Zheng et al. (2013) studied the potential impacts of CO₂ sequestration on shallow groundwater by looking at transport of organics and co-injected H₂S by supercritical CO₂ to shallow aquifers. Pan et al. (2011) developed a coupled wellbore and reservoir model for simulating the dynamics of CO₂ injection and leakage through wellbores, and applied the model to situations relevant to geologic CO₂ storage involving upward flow (e.g., leakage) and downward flow (injection). Wellbores in general have been identified as the most likely conduit for CO₂ and brine leakage from geologic carbon sequestration sites, especially those in sedimentary basins with historical hydrocarbon production.

Migration, storage, and leakage of CO₂ and its potential interaction with groundwater bodies involve complex physical and chemical processes that we cannot completely comprehend or fully couple in the models; great uncertainty stems from site characterization, site capacity, injection rate, caprock integrity, and initial and boundary conditions (Pruess, 2011). Uncertainty quantification (UQ) approaches have recently been used for understanding the underlying mechanisms of CO₂ migration and leakage, with respect to various factors including material (geological, hydrological, elastic, geochemical) properties and process parameters (Bao et al., 2013; Hou et al., 2012b, 2013; Hu et al., 2012; Liu and Zhang, 2011; Lu et al., 2012; Rohmer and Bouc, 2010; Sun et al., 2013; Wainwright et al., 2013). Uncertainty analysis yields an indication of which parameters have the most impact on carbon sequestration and provides guidance for the design, injection, regulation, and future data collection of proposed sequestration sites. For example, some studies (Eccles et al., 2009; Oldenburg and Unger, 2003) indicate that subsurface permeability and heterogeneity may have a large effect on predicting surface CO₂ leakage, implying that careful characterization and selection of a sequestration site are important. Some studies examined the impacts of uncertainties in reservoir and caprock properties on the migration, injectivity, and leakage of CO₂ (Hou et al., 2012b).

In this study, we focused on determining the effects of various physical and geochemical properties and CO₂ injection design parameters on CO₂ plume expansion behavior subsequent to wellbore leakage and quantifying the associated uncertainty with CO₂ plume predictions through UQ and risk analyses. In our approach, after the parameterization was determined and information about the parameters was collected, different procedures were used to perform uncertainty and sensitivity analyses. A sampling-based sensitivity is one in which the model is executed repeatedly for combinations of values sampled from the distribution (assumed known) of the input factors. Once the sample is generated, several strategies (including simple input–output scatterplots) can be used to derive sensitivity measures for the factors. One challenge is the computational time and demand related to the numerous simulations required to perform stable sensitivity analyses and/or parameter calibrations. Therefore, choosing an efficient and exploratory sampling approach is critical to a UQ framework. A low-discrepancy sampling approach such as quasi-Monte Carlo (QMC) can be used to derive reliable response curves and response surfaces. Adaptive sampling has been used previously (Dawson and Hall, 2006; Givens and Raftery, 1996), although most techniques did not use a global search algorithm or try to minimize the response variance, which are essential for the adaptive-response modeling (ARM) process. In adaptive sampling, one can identify strategic locations for new samples in the overall parameter space for simulations and produce comparable results with fewer samples than would be possible with traditional Monte Carlo sampling methods. It is also possible to combine adaptive sampling with QMC to obtain reliable response surfaces, yet with much fewer forward model calculations. Previous approaches, such as CO₂-PENS (Stauffer et al., 2009) and PSUADE (PSUADE, 2013), have dealt with major issues associated with reliability and accuracy of input UQ and development of physically and statistically reasonable reduced-order models (ROMs). In this study, we integrate efficient sampling and response-surface-based adaptive sampling for high efficiency with a greatly reduced number of necessary numerical simulations. Capabilities include a built-in approach to test parameter significance and develop ROMs with stepwise backward removal using the Akaike information criterion (AIC; Akaike, 1974).

We adopted the UQ framework to understand the mechanisms of CO₂ plume expansion with respect to major physical and geochemical factors and quantified the associated uncertainty and risks subsequent to wellbore leakage. In the next sections, we describe the model setup, parameterization, sampling, numerical simulations, and data analyses and visualizations.

2. Methodology

2.1. Site description

We studied the impact of possible leakage from saline reservoirs on water quality in a carbonate aquifer, because carbonate aquifers are an important source of water in the United States (Lindsey et al., 2009). To guide our numerical model construction, we selected a particularly well-characterized example: the carbonate Edwards Aquifer located in south-central Texas (Fig. 1). This aquifer covers an area of more than 10⁵ km² (Painter et al., 2007). We focused on an unconfined portion of the aquifer near San Antonio. The San Antonio segment is one of the most productive carbonate aquifers in the world, and is the sole source aquifer for more than two million people (Musgrove et al., 2010). Water levels and groundwater chemistry data from U.S. Geological Survey reports for the San Antonio area were used to establish the local hydrologic gradient and background chemistry (Lindgren et al., 2005; Musgrove et al., 2010). The aquifer is composed of carbonate rocks of the Georgetown Formation and the Edwards Group (or their stratigraphic equivalents), which range in thickness from 121 to 152 m. In the United States, the U.S. Environmental Protection Agency (EPA) Class VI Rule (EPA, 2012) sets a minimum federal technical criterion that injection of supercritical CO₂ in geologic reservoirs be protective of overlying underground sources of drinking water that have less than 10,000 mg.L⁻¹ dissolved solids; the average dissolved solids concentration in the shallow, urban, unconfined
portion of the Edwards Aquifer is 330 mg L\(^{-1}\) (Musgrove et al., 2010).

2.2. Numerical simulator

In this study, STOMP-CO2-R (White et al., 2012) was used for the aquifer simulations. The STOMP suite of simulators (White and Oostrom, 2006) comprises integrated finite-volume codes that use structured orthogonal grids for spatial discretization and a fully implicit formulation for temporal discretization. STOMP-CO2-R simulates multiphase flow of CO\(_2\), water and salt, coupled with equilibrium/kinetic aqueous geochemical reactions via the ECKECHEM (Equilibrium-Conservation-Kinetic-Equation Chemistry) module (White and McGrail, 2005). CO\(_2\) properties are calculated using the Span and Wagner (1996) equation of state. The dominant nonlinear functions within the STOMP simulator are the relative permeability–saturation–capillary pressure (\(k–s–p\)) relations. The STOMP simulator allows the user to specify these relations through a large variety of popular and classic functions. Two-phase (gas-aqueous) \(k–s–p\) relations can be specified with hysteretic or nonhysteretic functions or nonhysteretic tabular data. Phase equilibria calculations in STOMP-CO2 use the formulations of Spycher et al. (2003) for temperatures below 100 °C and Spycher and Pruess (2010) for temperatures above 100 °C, with corrections for dissolved salt provided in Spycher and Pruess (2010).

2.3. CO\(_2\) injection scenario setup, parameterization, field simulation

A three-dimensional, homogeneous model of groundwater flow and reactive transport in the Edwards Aquifer was created. Wellbore leakage into the aquifer was represented as a point source at the bottom of the aquifer. The wellbore leakage model is a simple function with two input parameters: distance from the injection well to the leaky well (m) and maximum leak rate (kg/s). Injection into the target formation was assumed to last for 50 years. An error function curve was fit to resulting wellbore leak predictions prior to 50 years, and an exponential decay function fit to predictions after 50 years. The distance from the injection well to the leaky well is assumed to vary log-uniformly from 10 m to 4300 m, and the maximum leakage rate is assumed to vary log-uniformly from 0.01 to 1 kg/s. Wellbore leak rate vs. time for average distance (207 m) and maximum rate (0.1 kg/s) is shown in Fig. 2.

The unconfined aquifer was assumed to extend 150 m below the ground surface and 19 km × 23 km in area. The model grid was refined around the wellbore leak point, located at \(x=0, y=20\) km, and \(z=0\). The model domain was 9.5 km wide, one-half as wide as the aquifer in the \(x\)-direction, assuming an axis of symmetry at \(x=0\). A vertical hydrostatic pressure gradient of \(9.79 \times 10^{-3}\) MPa/m was assumed in the \(z\) direction, and a horizontal pressure gradient of \(8.5 \times 10^{-6}\) MPa/m was assumed in the \(y\)-direction, based on observed heads (Hutchison and Hill, 2011). The average porosity of the aquifer was assumed to be 18% (Lindgren, 2007), and the assumed average hydraulic conductivity was 2.03 m/day (Hutchison and Hill, 2011). For two-phase flow gas and water, the aqueous saturation as a function of capillary pressure is calculated using the van Genuchten (1980) correlation function assuming an inverse air entry suction parameter value of 4.69 cm\(^{-1}\), a pore size distribution parameter value of 2.57, and a residual saturation of 0. These parameters are consistent with an easily drained porous

![Fig. 1. Location of unconfined portion of the Edwards Aquifer in south-central Texas (Musgrove et al., 2010).](image)

![Fig. 2. Wellbore leakage rate (kg CO\(_2\)/s) vs. time, assuming distance between the injection and leaky wells is 207 m and maximum leak rate is 0.1 kg/s.](image)
media in order to represent flow through the vugs and fractures of the Edwards Limestone. The Mualem (1976) porosity distribution function, using the aforementioned \( \eta \) and residual saturation values, was used to calculate aqueous and gas relative permeability.

The southern boundary \((y = 0)\) was an outflow boundary. The eastern \((x = 9.5 \text{ km})\) and western \((x = 0)\) boundaries were no-flow. The bottom boundary was no-flow, except for the CO\(_2\) leakage at \(x = 0 \text{ km}, y = 20 \text{ km}\). Because the aquifer was assumed to be homogeneous, flow was symmetric along the \(y\)-axis at \(x = 0\). The upper boundary was the water table, a no-flow boundary for the liquid phase but a constant pressure boundary of 0.101325 MPa (atmospheric pressure) for the gas phase.

A shallow groundwater sample was used as the initial condition for the model (Table 1). The aquifer is assumed to be composed of calcite and dolomite in varying proportions, with an average ratio of 50% calcite and 50% dolomite. The parameters for the Transition State Theory kinetic reactions (Table 2) were taken from Palandri and Kharaka (2004). As calcite and dolomite progressively dissolve in the groundwater, the equilibrium reactions shown in Table 3 are predicted to be significant by preliminary simulations done using the EQ3/6 v. 8.0 software (Wolery and Jarek, 2003).

Previous studies also show that porosity and permeability/conductivity are significant factors dominating hydrogeological processes such as CO\(_2\) migration in aquifers and sequestration reservoirs (Eccles et al., 2009; Oldenburg and Unger, 2003; Hou et al., 2012b, 2013; Sun et al., 2013; Wainwright et al., 2013). In order to evaluate whether geochemistry has comparable impacts with these process or physical parameters, geochemical parameters (e.g., surface area, equilibrium constant, dissolution rate) are also included given the dominant presence of calcite and dolomite in the formation. The parameters and their prior information are summarized in Table 4. Samples were generated according to the probability distributions (also shown in Table 4) of these parameters, and each combined parameter sample set was fed into the CO\(_2\) multiphase flow simulator. Given each parameter sample set of inputs, we simulated the transient shapes of CO\(_2\) plumes and wellbore leakage rate.

### 2.4. Uncertainty quantification framework

Fig. 3 illustrates the UQ pipeline framework we used for the case study of wellbore leakage. The framework has the capability to quantify input uncertainty, generate exploratory samples effectively, perform scalable numerical simulations, visualize output uncertainty, and evaluate input–output relationships. The UQ framework includes low-discrepancy sampling (e.g., QMC), ARM, response surface analyses, parameter ranking/screening, and ROM development.

#### 2.4.1. Low-discrepancy and adaptive sampling

Sampling technique was an important choice in our inversion because the success of a numerical approach hinges upon evaluating all possibilities defined by the model space. Given a large number of dimensions, systematic sampling techniques such as by Simpson’s rule are not sufficient (Tarantola, 2005). Therefore, we turned to QMC techniques, which incorporate deterministic sequences to guarantee good dispersion between sample points (Caflisch, 1998). The QMC approach has been used successfully for exploring parameter space as well as for stochastic inversion (Dick and Pillichshammer, 2010; Hou et al., 2006, 2012a,b).

### Table 1

<table>
<thead>
<tr>
<th>Species</th>
<th>Concentration, mg L(^{-1})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Br(^{-})</td>
<td>0.08</td>
</tr>
<tr>
<td>Ca(^{2+})</td>
<td>88.87</td>
</tr>
<tr>
<td>Cl(^{-})</td>
<td>16.075</td>
</tr>
<tr>
<td>F(^{-})</td>
<td>0.13</td>
</tr>
<tr>
<td>pH</td>
<td>6.94</td>
</tr>
<tr>
<td>HCO(_3)(^{-})</td>
<td>370.47</td>
</tr>
<tr>
<td>K(^{+})</td>
<td>0.81</td>
</tr>
<tr>
<td>Mg(^{2+})</td>
<td>11.73</td>
</tr>
<tr>
<td>Na(^{+})</td>
<td>6.22</td>
</tr>
<tr>
<td>SO(_4)(^{2-})</td>
<td>12.97</td>
</tr>
<tr>
<td>SiO(_2)(aq)</td>
<td>11.13</td>
</tr>
<tr>
<td>O(_2)(aq)</td>
<td>10.69</td>
</tr>
</tbody>
</table>

### Table 2


<table>
<thead>
<tr>
<th>Kinetic reaction</th>
<th>Specific surface area, cm(^2)/g</th>
<th>Dissolution rate, mol/m(^2)/s at 25 °C</th>
<th>Activation energy, kJ/mol</th>
<th>Equilibrium coefficient, log</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calcite + H(^+) + Ca(^{2+}) + HCO(_3)(^{-})</td>
<td>0.1</td>
<td>1.5 × 10(^{-6})</td>
<td>23.5</td>
<td>–5.82</td>
</tr>
<tr>
<td>Dolomite-dis + 2H(^+) + Ca(^{2+}) + Mg(^{2+}) + 2HCO(_3)(^{-})</td>
<td>0.1</td>
<td>2.9 × 10(^{-8})</td>
<td>52.2</td>
<td>–7.54</td>
</tr>
</tbody>
</table>

---

Fig. 3. The overall uncertainty quantification pipeline.
QMC requires a choice regarding input of a low-discrepancy sequence. It is widely acknowledged that Sobol sequences (Sobol, 1967) perform well for problems of greater than six dimensions and avoid degradation effects observed in many other low-discrepancy sequences (Atanassov et al., 2010; Sobol and Shukhman, 2007; Wang, 2009).

Adaptive sampling identifies new sample points based on previous results. The ARM method uses the response surface fit to the model results to identify new points along the surface. One method for identifying new points is to develop uncertainty estimates along the response surface and identify those points or regions where the uncertainty is high. The overall flow is described in Engel et al. (2004).

First, we selected a relatively small initial sample of design points using a space-filling design or a similar design based on a joint probability distribution of the design variables. We then ran the program representing the computer model to obtain the model response at each design point in the sample. Next, we applied a response approximation algorithm to the existing responses to develop a surrogate model, together with a measure of goodness of fit to the true response (uncertainty estimate). If the fit was adequate, or a budget for computing resources was exhausted, the process ended, and we implemented the final surrogate model together with associated predictive uncertainty estimates. Otherwise, another algorithm identified a set of the most informative new design points based on a number of possible criteria. New model evaluations at the additional design points then augmented the previous set of responses, and the approximation algorithm was applied to the augmented set of responses to refine the surrogate model.

### 2.4.2. Response curve and response surface analyses

The method we implemented for estimating the response surface is the multivariate adaptive regression splines (MARS) method (Friedman, 1991). The MARS method falls into the category of machine learning techniques and is essentially a nonparametric piecewise polynomial regression model designed to perform well for high-dimensional data. The MARS predictor consists of a sum of weighted-spline basis-functions and is a piecewise linear approximation to the response surface of the computer model.

We used the following approach to create a proxy for prediction uncertainty at any given design point. In addition to the MARS predictor $\hat{y}$, we computed a predictor $\hat{y}^i$ that excludes the $i$th observed response, for each of the $n$ observed responses. We defined the prediction interval proxy to be the sample variance $s^2(m)$ of these leave-one-out predictors at any design point $m$. That is,

$$s^2(m) = \frac{1}{n-1} \sum_{i=1}^{n} \left[ \hat{y}^i(m) - \frac{1}{n} \sum_{j=1}^{n} \hat{y}^j(m) \right]^2$$

(1)

This measure has the same qualitative shape around each linear subsection of the MARS predictor as that of a standard linear regression prediction interval.

For the analysis presented in this paper, a large portion of the parameter set was simulated up front (using the QMC sampling techniques). The ARM process was adapted to use the samples previously generated. This was accomplished by using the sample point closest to each identified adaptive sample location, without replacement (i.e., once a sample point was selected by the ARM, it was removed for later checking). The goal of using the ARM process was to produce results (responses) using a reduced number of model evaluations (compared to a quasi-random approach) while producing similar results. To measure how well the process worked, we used the root mean squared error (RMSE) metric to compare the response surface from the reduced simulation results to the “true” surface. Because the “true” surface was not known, it was approximated by fitting a response surface to the responses from all of the simulated parameter sets.

### 2.4.3. Statistical tests and reduced-order models

In the framework, we used multivariate generalized linear model (GLM) analyses and significance statistical tests to rank the significance of input parameters and develop relationships (ROMs) between inputs and outputs using response surface plots. We performed GLM analysis, which can be used to evaluate the statistical significance of explanatory terms (e.g., linear, quadratic, and interaction) through statistical $F$-tests. For each response variable, we selected a subset of models with only the most explanatory terms and then, by evaluating the fitted variance of each term, identified the one parameter/factor/term that was the least important. Then the AIC (Akaike, 1974) was adopted to evaluate the goodness of fit of a model and the model complexity to decide whether to keep a term in a fitted GLM based on its contribution. Terms kept in the GLM were those significant enough to compensate for the penalty due to increase in model complexity. We removed the least significant terms one by one using a procedure called AIC-based backward removal. Eventually, we obtained a model with minimum AIC, and no more terms could be dropped.

### Table 4: Input parameters and their ranges for uncertainty quantification.

<table>
<thead>
<tr>
<th>Number</th>
<th>Parameter</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>p1</td>
<td>Distance from injection well to leaky well, m</td>
<td>10</td>
<td>4300</td>
<td>Log uniform</td>
</tr>
<tr>
<td>p2</td>
<td>Maximum leak rate, kg/s</td>
<td>0.01</td>
<td>1</td>
<td>Log uniform</td>
</tr>
<tr>
<td>p3</td>
<td>Porosity, %</td>
<td>13</td>
<td>23</td>
<td>Normal</td>
</tr>
<tr>
<td>p4</td>
<td>Hydraulic conductivity, m/day</td>
<td>0.09</td>
<td>4.6</td>
<td>Log normal</td>
</tr>
<tr>
<td>p5</td>
<td>Calcite, solid phase, volumetric %</td>
<td>0.1</td>
<td>0.99</td>
<td>Normal</td>
</tr>
<tr>
<td>p6</td>
<td>Calcite dissolution rate at 25 °C, mol/m²/s</td>
<td>$1.5 \times 10^{-7}$</td>
<td>$1.5 \times 10^{-5}$</td>
<td>Log normal</td>
</tr>
<tr>
<td>p7</td>
<td>Dolomite dissolution rate at 25 °C, mol/m²/s</td>
<td>$2.9 \times 10^{-9}$</td>
<td>$2.9 \times 10^{-7}$</td>
<td>Log normal</td>
</tr>
<tr>
<td>p8</td>
<td>Surface area, cm²/g</td>
<td>0.01</td>
<td>1</td>
<td>Log normal</td>
</tr>
<tr>
<td>p9</td>
<td>Calcite equilibrium constant at 25 °C, log</td>
<td>1.35</td>
<td>2.35</td>
<td>Normal</td>
</tr>
<tr>
<td>p10</td>
<td>Dolomite equilibrium constant at 25 °C, log</td>
<td>2.01</td>
<td>3.01</td>
<td>Normal</td>
</tr>
</tbody>
</table>
3. Output and response surface analysis

3.1. Output uncertainty due to combined impacts of input parameters

Quasi-Monte-Carlo samples were generated for the 10 input parameters shown in Table 4, given their previous distributions and ranges. In QMC sampling, we were able to obtain a series of samples with controlled deterministic inputs instead of random inputs, thus alleviating the clumping issue of the traditional Monte Carlo methods. The well-dispersed QMC samples enabled more efficient exploration of multidimensional parameter spaces (see Fig. 4).

For each combination of parameter samples, we performed numerical simulations of CO$_2$ plume migration in the aquifer due to wellbore leakage. The output variables included length, width, height, and volume of the pH < 6.5 plume in the aquifer as well as the CO$_2$ flux to the atmosphere. The size of the groundwater plume with pH < 6.5 has been adopted as the performance measure of the model, as an indication of the size of the CO$_2$ plume, because the EPA has set a minimum pH of 6.5 as a secondary drinking water standard (EPA, 2009). Also, the flux of CO$_2$ to the atmosphere is of concern, because the goal of geologic carbon sequestration is to prevent excess CO$_2$ from reaching the atmosphere by sequestering it in deep geological formations.

The resulting uncertainty of output variables (e.g., pH<6.5 volume in the aquifer, rate of CO$_2$ flux to the atmosphere) is represented using histograms shown in Fig. 5. The distributions of pH<6.5 volume at different post-injection observation times are close to Gaussian, particularly at later times, while the CO$_2$ flux has log-normal distributions. The output variables that have nonsymmetrical distributions should be transformed (e.g., logarithmically) before evaluating their predictive/confidence intervals or performing further analysis of input–output relationships. Fig. 5 also shows that the average pH<6.5 plume volume does not increase much after the major injection period, and the uncertainty ranges associated with the early and late post-injection observation times are similar.

3.2. Response curve analysis

The marginal impacts of the parameters are shown as boxplots in Figs. 6 and 7. The boxplots show the relationships (linear or nonlinear, positive or negative) between input parameters and output variables, as well as how output uncertainty varies with respect to the input parameters.

For each output variable, a different set of input parameters play important roles. For pH<6.5 volume, both the maximum leak rate and hydraulic conductivity have a positive and nonlinear relationship with pH<6.5 plume volume at higher values of those parameters. The maximum leak rate influences the total mass of CO$_2$ leaked, and hence the total volume of impacted aquifer, given that the density of CO$_2$ is constant at the assumed aquifer temperature.
of 25 °C. The hydraulic conductivity has an increasing impact with time: because the hydraulic gradient is held constant, the horizontal flow rate, and therefore the amount of plume dispersion with time, is positively related to the hydraulic conductivity. For CO₂ flux, the maximum leakage rate is the dominant parameter. Very little CO₂ dissolves in the groundwater, and it is less dense than the groundwater, so most of the leaked CO₂ rises buoyantly to the water table and escapes to the atmosphere.

The analyses also demonstrated the uncertainty associated with the simulated output of pH < 6.5 plume volume and leakage flux rate. For example, uncertainty in the CO₂ flux to the atmosphere is not affected by log K, but it increases with maximum leakage rate. In contrast, uncertainty in simulated pH < 6.5 plume volume decreases with both log K and maximum leakage rate. The larger the maximum leak rate, the larger the plume, and the more heterogeneity the plume encounters as it grows, leading to greater uncertainty in the plume volume.

To rank the significance of the parameters and their interactions, we adopted GLM analysis to do statistical tests of the linear (the first-order terms), interaction (cross-products), and quadratic terms (the second-order terms) of these input parameters (Fig. 8). This analysis starts with a quadratic GLM regression model that takes the following form:

\[ Y_i = \theta_0 + \sum_j \theta_j p_{ij} + \sum_{jk} \theta_{jk} p_{ij} p_{ik} + \sum_j \gamma_j p_{ij}^2 + \epsilon_i, \quad i = 1, \ldots, n \]

where \( p_{ij} \) and \( p_{ik} \) represent the ith realization of the jth and kth parameters, respectively; \( \theta_0 \) is the intercept term; \( \theta_j \), \( \theta_{jk} \), and \( \gamma_j \) are coefficients of the linear, interaction, and quadratic terms, respectively; \( Y_i \) denotes the ith realization of the response variable and \( \epsilon_i \) denotes the corresponding fitted residual. The model is fitted using \( n \) sets of response and explanatory variables. Then the unnecessary terms are dropped one by one using the stepwise backward removal approach based on the AIC (Akaike, 1974).

Through the above GLM analyses of linear and quadratic terms, we ranked the significance of the input parameters according to the output (response variable) variability that can be interpreted/fitted by each parameter. Table 5 shows the remaining significant terms in the quadratic GLM final model for pH < 6.5 plume volume, and the corresponding regression coefficients and P-values of the statistical tests. Five significant parameters (p1, p2, p3, p4, and p9) were identified, among which p2 (i.e., maximum leakage rate) has a strong linear effect on the pH < 6.5 plume volume, while p3 and p4 (i.e., porosity and hydraulic conductivity) have strong linear and quadratic impacts. The three parameters together contributed more than 85% of the total output variability in the ensemble simulations of pH < 6.5 plume volume, and they have clear interactive effects with each other. The importance of the maximum leakage rate and permeability were discussed previously. In a homogeneous aquifer, the porosity would have a simple linear relationship to the volume of CO₂, and hence the pH > 6.5 plume volume, but heterogeneity complicates this relationship. The process parameter p1 (i.e., distance between the leaky and injection wells) has nonlinear impacts on plume expansion and is not interacting with
other parameters as expected. The distance between the leaky and injection wells mainly influences the time delay before the start of CO₂ leakage into the aquifer; the greater the distance between the injection and leaky wells, the greater the lag time before the onset of leakage into the aquifer. The parameter p9 (i.e., calcite equilibrium constant) mainly has a nonlinear effect and interacts with hydraulic conductivity. Interestingly, the other geochemical parameter p10 (i.e., dolomite equilibrium constant) is not significant, which is partially related to its relatively narrow prior range. Calcite dissolution is a relatively fast process, and so the kinetic parameters related to calcite dissolution (p6, p8) are not important. The result that the dolomite equilibrium and kinetic parameters (p7, p10) are not important indicates that equilibrium with calcite is the dominant geochemical process controlling the pH. This holds true even at low volume fractions of calcite, so that parameter p5 is also not important. The final model (also called the reduced-order model), can be effectively used to predict pH < 6.5 plume expansion behavior subsequent to wellbore leakage after collecting information about major parameters/factors such as hydraulic conductivity, porosity, maximum leakage rate, and distance between the leaky and injection wells. Since the plume is defined by pH < 6.5 (the U.S. EPA drinking water standard), the prediction results, together with uncertainty estimates, can provide guidance on subsequent water quality and risk analysis at similar sites.

### 3.3. Response surface analysis with parameter subset and adaptive sampling

Response surface analysis was performed on the 10-parameter dataset to evaluate and compare sampling strategies. The performance of the iterative ARM is compared to the QMC samples, with the RMSE as the metric of comparison. For each sampling method, we fit a response surface, as explained in Section 2.4.2, to a subset of the sample results (32, 64, 96, and 128 samples) and then estimated the response value at the 512 locations (parameter space). The RMSE was then calculated from the difference in the estimated response and the actual calculated response.

The first step of the analysis was to perform a sensitivity analysis to see how sensitive the response variables were to a variation in each input parameter. The results of this analysis for the (log₁₀) volume are shown in Table 6. For this analysis, the F-statistic was calculated for each input–output variable combination at four different times. The F-statistic calculates the significance of the parameter coefficient (i.e., slope), assuming a linear dependence.

The results of our validation tests using the log₁₀ volume response variable at 100 years are shown in Table 7. The left side of

---

**Table 5**

Fitted coefficients and their t/P-values for the finalized quadratic GLM model for pH<6.5 plume volume.

| Coefficient | Estimate  | Std. error | t-Value | P>|t| |
|-------------|-----------|------------|---------|-----|
| β₀          | 2.05      | 5.41 × 10⁻³ | 3.79 × 10⁴ | <2.00 × 10⁻³ |
| β₁          | 1.41 × 10⁻² | 1.08 × 10⁻³ | 1.31 × 10⁴ | <2.00 × 10⁻³ |
| β₂          | 3.63 × 10⁻² | 1.39 × 10⁻³ | 2.61 × 10⁴ | <2.00 × 10⁻³ |
| β₃          | -4.35 × 10⁻³ | 5.86 × 10⁻³ | -7.42 | 5.14 × 10⁻¹ |
| β₄          | 1.56 × 10⁻¹ | 7.42 × 10⁻³ | 2.10 × 10⁴ | <2.00 × 10⁻³ |
| β₅          | -3.86 × 10⁻¹ | 2.31 × 10⁻⁴ | -1.67 × 10⁴ | <2.00 × 10⁻³ |
| β₆          | 9.26 × 10⁻¹ | 1.61 × 10⁻¹ | 5.74 | 1.68 × 10⁻⁸ |
| β₇          | 4.70 × 10⁻³ | 6.00 × 10⁻³ | 7.84 × 10⁴ | <2.00 × 10⁻³ |
| β₈          | -3.87 × 10⁻² | 1.33 × 10⁻⁴ | -2.90 × 10⁴ | <2.00 × 10⁻³ |
| β₉          | 4.99 × 10⁻² | 7.58 × 10⁻³ | 6.59 | 1.14 × 10⁻⁸ |
| β₁₀         | -5.84 × 10⁻² | 1.40 × 10⁻³ | -4.17 × 10⁴ | <2.00 × 10⁻³ |
| β₁₁         | -5.84 × 10⁻³ | 2.82 × 10⁻² | -2.07 × 10⁴ | <2.00 × 10⁻³ |
| β₁₂         | -7.95 × 10⁻³ | 8.44 × 10⁻⁴ | -9.42 | <2.00 × 10⁻³ |

---

Fig. 6. Marginal impacts of four selected input parameters on pH plume volume in the aquifer.
Table 6
Estimates of the influence of each parameter in the modeling of pH < 6.5 plume volume, based on the F-statistic.

<table>
<thead>
<tr>
<th>Number</th>
<th>Parameter</th>
<th>Years post-injection</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>50</td>
</tr>
<tr>
<td>p1</td>
<td>Distance from injection well to leaky well, m</td>
<td>16</td>
</tr>
<tr>
<td>p2</td>
<td>Maximum leak rate, kg/s</td>
<td>1069</td>
</tr>
<tr>
<td>p3</td>
<td>Porosity, %</td>
<td>22</td>
</tr>
<tr>
<td>p4</td>
<td>Hydraulic conductivity, m/day</td>
<td>60</td>
</tr>
<tr>
<td>p5</td>
<td>Calcite, solid phase volumetric %</td>
<td>0</td>
</tr>
<tr>
<td>p6</td>
<td>Calcite dissolution rate at 25 °C, mol/m²/s</td>
<td>0</td>
</tr>
<tr>
<td>p7</td>
<td>Dolomite dissolution rate at 25 °C, mol/m²/s</td>
<td>0</td>
</tr>
<tr>
<td>p8</td>
<td>Surface area, cm²/g</td>
<td>0</td>
</tr>
<tr>
<td>p9</td>
<td>Calcite equilibrium constant at 25 °C, log</td>
<td>6</td>
</tr>
<tr>
<td>p10</td>
<td>Dolomite equilibrium constant at 25 °C, log</td>
<td>0</td>
</tr>
</tbody>
</table>

the table shows the comparison results using the RMSE when using all 10 input parameters to fit the response surface. The right side shows the results when using only the three most sensitive parameters in the construction of the response surface. The column with the “%” header represents how much better/worse the ARM method performed compared to the QMC method using the same number of samples (% = 100 – 100 × ARM/QMC). These results indicate that the two sampling methods are very similar when only the most influential parameters are used to construct the response surface (right side of Table 7). When more parameters (e.g., all 10 parameters) are included, the ARM method seems to construct a response surface that is closer to the true simulated response values, particularly when sample sizes are small (e.g., around 64 samples). In addition, when we constructed the response surface, both methods better estimated the “true” surface when all 10 parameters were used, as seen by comparing the RMSE. Similar results were observed when the different response variables (length, width, flux) and different post-injection times (50, 100, 150, 200 years) were used.

Table 7
Comparison (RMSE) of sampling techniques for estimating log₁₀ volume.

<table>
<thead>
<tr>
<th>No. samples</th>
<th>All 10 parameters</th>
<th>Top 3 sensitive parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ARM</td>
<td>QMC</td>
</tr>
<tr>
<td>64</td>
<td>0.0036</td>
<td>0.0047</td>
</tr>
<tr>
<td>96</td>
<td>0.0031</td>
<td>0.0031</td>
</tr>
<tr>
<td>128</td>
<td>0.0029</td>
<td>0.0031</td>
</tr>
</tbody>
</table>

Fig. 7. Marginal impacts of four selected input parameters on CO₂ flux to the aquifer.
4. Discussion and conclusions

The parameter ranking and screening studies helped identify the most significant parameters/factors affecting the CO₂ plume expansion process due to wellbore leakage at the Edwards Aquifer. The significance of the process and model parameters were ranked based on statistical tests and response surface analysis. The most significant parameters included distance to injection well, maximum leakage rate, porosity, and hydraulic conductivity.

With the UQ framework, we evaluated how the input uncertainty propagates and analyzed uncertainty ranges of the output variables (e.g., CO₂ plume volume, width, height, flux rate). These results can be used for subsequent analysis of contamination risk due to expansion of a CO₂ plume in the groundwater body at the Edwards Aquifer or a similar site.

To facilitate the risk analysis, ROMs of CO₂ plume (volume, width, height, flux rate) with respect to the dominant parameters/factors can be used, such that the CO₂ plume can be predicted without numerical simulations, given information about distance between the leaking wellbore and injection well, the maximum leakage rate, porosity, hydraulic conductivity, and so on. The limit of the developed ROMs is that in practice, we are not able to explore all the factors controlling CO₂ migration behavior after injection and evaluate their relative significance; the developed ROMs linking input parameters/factors and output variables of interest could be specific to the problem or site. Therefore, although the UQ approach can be applied for broad risk assessment studies, the conclusions regarding parameter significance may change as different processes/parameters are considered for different practical studies.

Despite the efforts to deal with the computational demand using effective sampling approaches (QMC) and effective computing techniques, a tremendous amount of time might still be required to obtain reliable output statistics if the output response variables are not sensitive to the input parameters. The adaptive sampling method showed some promising results when its response surface was compared to a surrogate "true" response surface. The advantage of this method, particularly when the computational demand is high, is to develop a response surface (and ultimately an uncertainty analysis) that is as close to the "true" response surface with as few samples (i.e., STOMP runs) as possible.
Acknowledgements

Funding for this research was provided by the National Risk Assessment Partnership (NRPAP) Project within the U.S. Department of Energy Office of Fossil Energy’s Carbon Sequestration Program, and by the Pacific Northwest National Laboratory (PNNL) Carbon Sequestration Initiative. The study was conducted at the Pacific Northwest National Laboratory, which is operated by Battelle Memorial Institute for the U.S. Department of Energy under Contract DE-AC05-76RL01830. The research was performed using PNNL Institutional Computing at Pacific Northwest National Laboratory.

References


