1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

A probabilistic collocation Eulerian-Lagrangian localized adjoint method on sparse grids for assessing CO₂ leakage through wells in randomly heterogeneous porous media

Hong Wang^{a,b} Yongqiang Ren^a Jinhong Jia^a Michael A. Celia^c

^a*School of Mathematics, Shandong University, Jinan, Shandong 250100, China*

^b*Corresponding author. Department of Mathematics, University of South Carolina,
Columbia, South Carolina 29208, USA. Phone: 803-777-4321, Fax: 803-777-6527.*

Email: hwang@math.sc.edu

^c*Department of Civil and Environmental Engineering, Princeton University, Princeton,
New Jersey 08544, USA*

Abstract

We develop a probabilistic collocation Eulerian-Lagrangian localized adjoint method on sparse grids for assessing CO₂ leakage through wells in randomly heterogeneous porous media, by utilizing the intrinsic mathematical, numerical, and physical properties of the mathematical model. We model the process in which CO₂ is injected into a deep aquifer, spreads within the aquifer and, upon reaching a leaky well, rises up to a shallower aquifer, to quantify the leakage rate, which depends on the pressure build-up in the aquifer due to injection and the buoyancy of CO₂. The underlying Eulerian-Lagrangian framework has high potential to improve the efficiency and accuracy for the numerical simulation of complex flow and transport processes in CO₂ sequestration. The sparse grid probabilistic collocation framework adds computationally efficient uncertainty quantification functionality onto pre-existing Eulerian-Lagrangian methods in a nonintrusive manner. It also provides a scalable framework to consider uncertainty in a straightforward parallel manner. Preliminary numerical experiments show the feasibility and potential of the method.

Keywords: CO₂ sequestration and leakage, Eulerian-Lagrangian method, heterogeneous porous media, probabilistic collocation method, risk assessment, uncertainty quantification

Preprint submitted to Computer Methods Appl. Mech. Engrg.

November 10, 2014

1. Introduction

Anthropogenic emissions of carbon dioxide continue to cause increases in the atmospheric concentration of carbon dioxide, which are leading to global warming of the earth with wide-ranging environmental implications [39, 40]. The major source of anthropogenic emissions of CO₂ is fossil fuel combustion such as petroleum hydrocarbon and coals. The ideal solution to reduce the amount of fossil fuel burned is to replace fossil fuels with clean renewables. However, this is infeasible in the near and intermediate future, since carbon-free energy sources are not available at the scale required for modern economies. This leaves carbon capture and storage (CCS) of captured CO₂ in geological materials as the only large-scale technology to allow both the continued use of coal and petroleum hydrocarbon and a solution to the carbon problem [33, 38, 39, 48, 57, 59, 62]. The amount of CO₂ that needs to be injected in order to have a significant impact on the climate system is very large - on the order of gigatonnes (Gt) of carbon per year. Injection of such large quantities will create subsurface perturbations on a scale that has not heretofore been seen, and puts unprecedented demands on the mathematical and computational analyses that must accompany any serious evaluation of this technology. Of particular concern is the possibility that CO₂, or the brine displaced during the CO₂ injection operation, might leak into shallow subsurface zones and thereby contaminate underground drinking water supplies or other subsurface resources, or leak back to the atmosphere, thereby negating the objective of the injection. Proper analysis of the risks associated with large-scale CO₂ injection requires identification and analysis of leakage pathways, and application of a range of modeling tools that are suited to answer the relevant questions. All of this must be placed into a risk-assessment framework, with appropriate inclusion of probabilistic and stochastic analyses.

The idea of CCS is to inject captured CO₂ into geological formations that are deep enough to have pressures and temperatures beyond the critical point for CO₂, which is approximately 7.4 Mega-Pascal (MPa) and 31°C (see Figure 1). Under typical geothermal gradients, the critical point is reached at a depth around 800 meters. Supercritical CO₂ is slightly miscible in brine, up to a few percent by volume [39]. So a small fraction of the injected CO₂ will dissolve into the brine, while most of the CO₂ will form a separate fluid

Figure 1: Phase diagram for carbon dioxide. Green line follows the indicated temperature and pressure gradients with depth. From Bruant (pers. comm.)

Figure 2: Worldwide density of existing oil and gas wells [39]

phase, thereby leading to a multiphase flow system. A small fraction (less than one percent) of the water can also evaporate into the CO₂, producing “wet” CO₂ from the injected dry CO₂. The dissolved CO₂ leads to reductions in pH in the aqueous phase, which can drive different geochemical reactions. Similarly, wet CO₂ is much more corrosive to pipelines and well casing than dry CO₂. So the interphase mass transfer leads to a compositional system that drives a range of (geo)chemical reactions that may have important consequences. These kinds of reactions can be especially important along critical leakage pathways such as fault zones and old oil and gas wells. For example, cements used in wells may degrade in the presence of low-pH brines, while well casings can be corroded by wet CO₂ [11, 12, 58]. This significantly complicates the problem because, in addition to more standard geochemical reactions within the formation, material reactions along concentrated leakage pathways, at scales of millimeters or smaller, may also need to be resolved appropriately. Millimeter-scale multiphase flow and reactive transport must thus be simulated, perhaps along thousands of such pathways (see Figures 1 and 3), and these calculations must be embedded into an overall field-scale model that can cover thousands of square kilometers and includes all important geological heterogeneities. Hence the extreme computational challenge associated with this problem.

In addition to these complexities, equations of state can play an important role. While depth of injection coupled with specific geothermal gradients give a range of densities and viscosities [51], the supercritical CO₂ will always be less dense than the resident brine by a factor of 0.25 to 0.75. This gives a very strong buoyant drive, with gravity segregation playing an important role in the fluid dynamics. In addition, the viscosity of the injected CO₂ will be one to two orders of magnitude less than the viscosity of the resident brine [51], leading to viscous instabilities as well. Finally, brine with CO₂ dissolved is slightly more dense than brine without CO₂, and that leads to important density impacts in the aqueous transport problem over longer time scales.

2. Governing equations

The description of the porous medium flow system involving CO₂ and brine, while relatively simple in concept, can become extremely complex mathematically. The mathematical description involves equations of mass, momentum, and energy balances for the fluid and solid phases, and the components within the phases. These are augmented by material-specific proper-

Figure 3: Location around Wabamun Lake showing 50km $\times$ 50km region and the location of existing oil and gas wells [7]

ties including thermodynamic relationships for the fluids, phase partitioning behavior for the components, material conductances for fluid and heat transport, and other related properties.

As the first step in the development of a probabilistic collocation Eulerian-Lagrangian localized adjoint method for assessing CO₂ leakage problem, we consider a mathematical model for describing the displacement process of resident brine by injected dry CO₂ through randomly heterogeneous porous media with leaky wells. In this context, balance equations usually have a standard structure

$$\frac{\partial}{\partial t}(\phi s^\alpha \rho^\alpha) + \nabla \cdot (\mathbf{u}^\alpha \rho^\alpha) = \rho^\alpha q^\alpha. \quad (1)$$

Here $\alpha = c$ refers to the carbon-rich (typically nonwetting) phase and $\alpha = b$ the brine phase. In addition, ρ^α is the density of phase α ; ϕ is porosity; s^α denotes fluid saturation; $\mathbf{u}^\alpha$ is the volumetric flux vector for phase α ; and q^α is the volumetric flow rate per unit bulk volume that accounts for external supplies of phase α .

Another set of equations derives from the multiphase extension of Darcy's

law

$$\mathbf{u}^\alpha = -\frac{k^{r\alpha}}{\mu^\alpha} K (\nabla p^\alpha - \rho^\alpha \mathbf{g}). \quad (2)$$

Here K is the intrinsic permeability; $k^{r\alpha}$ is the relative permeability taken to be a function of the saturation s^α ; μ^α is the viscosity of phase α ; p^α is the pressure of fluid phase α ; and $\mathbf{g}$ is the gravitational vector. The simple geometric observation that the two fluid phases must fill all of the void space gives a saturation constraint $s^c + s^b = 1$. Finally, from a large number of experiments, and from arguments associated with the interfaces formed between the two fluids at the pore scale, a standard (nonlinear) functional relationship $s^\alpha = s^\alpha(p_{cap})$ is usually provided between the phase saturation s^α and the difference between the phase pressures, referred to as the capillary pressure $p_{cap} = p^c - p^b$ [25]. The saturation constraint and the saturation and capillary-pressure relationship constitute two independent equations, which together with (1) and (2) give a total of 10 equations for the 10 unknowns p^α , $\mathbf{u}^\alpha$, and s^α with $\alpha = c, b$.

Since pressure is the major driving force for the CO₂ sequestration problem, it should be chosen as a primary variable. There are two general approaches for selecting a pressure variable. One approach is to select a phase pressure (say, p^c) as a primary variable then add up equations (1) with $\alpha = c$ and b , respectively, and then to combine p^c with the capillary pressure p_{cap} to replace the other phase pressure p^b . Next one solves the resulting pressure equation for p^c and uses the capillary pressure p_{cap} to recover the other phase pressure p^b [1, 37]. Finally, equations (2) are used to generate phase velocities $\mathbf{u}^c$ and $\mathbf{u}^b$, which are needed in the solution of the saturation equation. Phase pressures and phase velocities have salient physical significance and are computed directly. However, when the corresponding phase vanishes, the other phase pressure has to be used as the primary variable. This switching could potentially complicate the numerical procedure.

In this paper we follow Chavent and Jaffré [15] to introduce a global pressure

$$p = \frac{1}{2}(p^c + p^b) + \frac{1}{2} \int_{s_r}^s (f^c(\xi) - f^b(\theta)) p'_c(\theta) d\theta. \quad (3)$$

Here s_r is the residual saturation of brine phase which satisfies $p_{cap}(s_r) = 0$, f^α is the fractional flow function of phase α defined by

$$\lambda^\alpha = \frac{k^{r\alpha}}{\mu^\alpha}, \quad \lambda = \lambda^c + \lambda^b, \quad f^\alpha = \frac{\lambda^\alpha}{\lambda} \quad (4)$$

with λ^α and λ being the phase mobility of phase α and the total mobility, respectively. Then the phase pressures p^c and p^b can be expressed in terms of the global pressure p as follows

$$p^\alpha = p \pm \frac{1}{2}p_{cap} - \frac{1}{2} \int_{s_r}^s (f^c(\theta) - f^b(\theta))p'_{cap}(\theta)d\theta. \quad (5)$$

Here $\alpha = c$ and $\alpha = b$ correspond to the “+” and “−” signs in $\pm$, respectively. Let $\mathbf{u} = \mathbf{u}^c + \mathbf{u}^b$ be the total velocity. We write equations (1) and (2) for $\alpha = b$ and c and further assume that the fluid phases are incompressible to obtain the following pressure equation expressed in terms of the total velocity $\mathbf{u}$ and global pressure p

$$\begin{aligned} \nabla \cdot \mathbf{u} &= q^c + q^b =: q, \\ \mathbf{u} &= -\lambda K [\nabla p - (f^c \rho^c + f^b \rho^b) \mathbf{g}]. \end{aligned} \quad (6)$$

The phase velocity $\mathbf{u}^c$ can be expressed in terms of the total velocity $\mathbf{u}$ as follows:

$$\mathbf{u}^c = f^c \mathbf{u} - f^c f^b \lambda K (\nabla p_{cap} + (\rho^b - \rho^c) \mathbf{g}). \quad (7)$$

In addition, we incorporate the Darcy’s law (2) with $\alpha = c$ into the saturation equation (1) with $\alpha = c$ to rewrite the saturation equation in terms of $s = s^c$ as follows

$$\phi \frac{\partial s}{\partial t} + \nabla \cdot (f^c(s) \mathbf{u} - f^c f^b \lambda K p'_{cap}(s) \nabla s) = q^c + f^c f^b (\rho^b - \rho^c) \lambda K \mathbf{g}. \quad (8)$$

Note that $p'_{cap}(s)$ is positive, $f^c(0) = 0$ and $f^b(1) = 0$. Hence, equation (8) is a nonlinear advection-diffusion equation with a degenerate diffusion. This set of equations (6) and (8) models the CO₂-brine system and assumes only 5 unknowns: the saturation s of the CO₂ phase, the global pressure p , and the total velocity $\mathbf{u}$. All other variables in equations (6) and (8), including densities, viscosities, porosity, permeability, capillary pressure and relative permeability functions, and source/sink terms, are assumed to be known. These in turn imply that temperature is known, fluid composition is known, and there are no important geomechanical or geochemical processes to change the parameters (like permeability or porosity) in the system. Mathematically, this set of equations (6) and (8) does not involve any degeneracy and so does not need to switch between different phase pressures. Nevertheless, this set of equations requires the relative permeability $k^{r\alpha}$ and the capillary

pressure p_{cap} to be a function of saturation only, so it is difficult to be applied when different subdomains have different capillary pressure functions. We emphasize that the Eulerian-Lagrangian localized adjoint formulation works well with both the phase pressure approach and the global pressure-total velocity approach.

3. Stochastic presentation of randomly heterogeneous porous media and leaking wells

Due to the high spatial variability of geological properties of porous media and the scarcity of available data, the hydrological parameters describing continuous or macroscopic properties of porous media, such as the intrinsic permeability and porosity, cannot be accurately characterized in detail and are often modeled as spatially correlated random fields [19, 34, 75].

Additional uncertainty in the context of CCS results from the abandoned or poorly treated old oil and gas wells. Leakage along these wells is particularly important in North America, with many millions of oil and gas wells having been drilled over the last century (see Figures 2 and 3). Quantitative characterization of these wells is difficult due to severe lack of data, thereby leading to very significant parameter uncertainty. The concentrated leakage along these human-induced wells appears to be the most significant risk in CCS. In this paper we follow the idea in [13] to take advantage of the simple geometry of the wells to assign probability distributions for the effective permeability of each well segment in the domain. Recall that a typical CCS site could have hundreds to thousands of existing wells, each of which could go through several geological layers in the caprock formations.

Let $D \subset \mathbb{R}^3$ be a bounded domain with the boundary $\Gamma = \partial D$. Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a complete probability space, with Ω being the sample space, $\mathcal{F}$ a σ -algebra of subsets of Ω , and $\mathcal{P}$ a probability measure. Then $K = K(\mathbf{x}; \omega)$ in (6) represents the random intrinsic permeability of the porous media if $\mathbf{x}$ lies in the formation or that of human-induced wells if $\mathbf{x}$ happens to fall into them, where ω represents the random event in the sample space Ω . The strong impact of the randomness of the intrinsic permeability $K(\mathbf{x}; \omega)$ renders the set of equations (6) and (8) stochastic. Ideally, one would like to find the probability density functions (PDF) for the saturation $s(\mathbf{x}, t; \omega)$ of the CO₂-rich phase, the global pressure $p(\mathbf{x}, t; \omega)$, and the total velocity $\mathbf{u}(\mathbf{x}, t; \omega)$ of the stochastic flow and transport equations (6) and (8). In general, these PDFs are non-Gaussian and characterized by an infinite number of statistical

moments, and hence are deemed to be numerically impossible to compute. It is common in applications to compute only the first two statistical moments, the mean and (co)variance. The mean provides an unbiased estimate of the variables and the variance quantifies the uncertainty associated with this estimate. Finally, we close the set of stochastic flow equations (6) and (8) with appropriate initial and boundary conditions such that the system is well-posed $\mathcal{P} - a.e. \omega \in \Omega$.

We assume that $Y(\mathbf{x}; \omega) = \ln K(\mathbf{x}; \omega)$ is a second-order continuous stationary random process with multivariate Gaussian distribution. The Karhunen-Loève expansion, which is based on the spectral decomposition of the covariance function of the input random process, is one of the most widely used techniques for dimension reduction in representing a random process. It fits naturally with the probabilistic collocation approach [26] and provides the best linear approximation in the mean square sense for a second-order stationary random process [35, 47]. Assume that

$$\mathbb{E}[|Y(\mathbf{y}; \omega)|^2] < +\infty \quad \text{and} \quad \lim_{\mathbf{y} \rightarrow \mathbf{x}} \mathbb{E}[|Y(\mathbf{y}; \omega) - Y(\mathbf{x}; \omega)|^2] = 0.$$

Then $Y(\mathbf{x}, \omega)$ can be expanded as the Karhunen-Loève expansion (i.e., a generalized Fourier series in $L^2(D \times \Omega)$) [47]

$$Y(\mathbf{x}; \omega) = \bar{Y}(\mathbf{x}) + \sum_{i=1}^{\infty} \sqrt{\lambda_i} \xi_i(\omega) \psi_i(\mathbf{x}). \quad (9)$$

Here $\bar{Y}(\mathbf{x})$ is the mean of $Y(\mathbf{x}; \omega)$, $\{\lambda_i\}_{i=1}^{\infty}$ and $\{\psi_i(\mathbf{x})\}_{i=1}^{\infty}$ are the corresponding sets of eigenvalues and orthogonal eigenfunctions of the following eigenvalue problems

$$\int_D C_Y(\mathbf{x}, \mathbf{y}) \psi_i(\mathbf{y}) d\mathbf{y} = \lambda_i \psi_i(\mathbf{x}), \quad \mathbf{x} \in D. \quad (10)$$

$C_Y(\mathbf{x}, \mathbf{y}) = \text{cov}(Y(\mathbf{x}; \cdot), Y(\mathbf{y}; \cdot))$ is the symmetric and positive-definite covariance kernel. $\{\xi_i(\omega)\}_{i=1}^{\infty}$ is the set of mutually uncorrelated random variables with zero mean and unit variances given by

$$\xi_i(\omega) = \frac{1}{\sqrt{\lambda_i}} \int_D (Y(\mathbf{x}; \omega) - \bar{Y}(\mathbf{x})) \psi_i(\mathbf{x}) d\mathbf{x}, \quad 1 \leq i < \infty. \quad (11)$$

In this paper we assume that the physical domain D is three-dimensional rectangular $D = \{(x_1, x_2, x_3) : 0 \leq x \leq L_1; 0 \leq y \leq L_2; 0 \leq z \leq L_3\}$.

We further assume that the second-order stationary covariance kernel is of separable exponential form

$$C_Y(\mathbf{x}, \mathbf{y}) = \sigma_Y^2 \exp\left(-\frac{|x_1 - y_1|}{\eta_1} - \frac{|x_2 - y_2|}{\eta_2} - \frac{|x_3 - y_3|}{\eta_3}\right) \quad (12)$$

where η_1, η_2 and η_3 are the correlation length in the x_1, x_2 , and x_3 coordinates. In this context, the eigenvalues and corresponding eigenfunctions can be solved analytically [35]. For a three-dimensional domain with a nonrectangular shape or a general covariance kernel, the corresponding eigenvalue problems (10) have to be solved numerically and can be numerically expensive to solve.

In numerical simulations, the Karhunen-Loéve expansion (9) has to be truncated to a finite number of terms

$$Y_M(\mathbf{x}; \omega) = \bar{Y}(\mathbf{x}) + \sum_{i=1}^M \sqrt{\lambda_i} \xi_i(\omega) \psi_i(\mathbf{x}) = \bar{Y}(\mathbf{x}) + Y_c(\mathbf{x}; \boldsymbol{\xi}) \quad (13)$$

with

$$Y_c(\mathbf{x}; \boldsymbol{\xi}) = \sum_{i=1}^M \sqrt{\lambda_i} \xi_i(\omega) \psi_i(\mathbf{x})$$

being a mean zero random function such that the corresponding optimal mean-square truncation error

$$\int_D \mathbb{E}[|Y(\mathbf{x}; \omega) - Y_M(\mathbf{x}; \omega)|^2] d\mathbf{x} = \sum_{i=M+1}^{\infty} \lambda_i \quad (14)$$

is sufficiently small, which depends on the decay rate of the eigenvalues $\{\lambda_i\}_{i=1}^{\infty}$ that in turn depends on the regularity of the correlation function and its correlation length.

4. A probabilistic collocation ELLAM time-stepping procedure on sparse grids

We develop a probabilistic collocation ELLAM time-stepping procedure on sparse grids for the system of stochastic flow equations (6) and (8) for assessing the CO₂ leakage.

4.1. The probabilistic collocation method

Let $\rho_i(\xi_i) : \Theta_i \rightarrow \mathbb{R}^+$ denote the probability density function (PDF) of the i -th random variable $\xi_i (i = 1, \dots, M)$ in (13), where the images $\Theta_i := \xi_i(\Omega)$ are assumed to be finite or (one sided or two sided) infinite intervals in $\mathbb{R}$. If $\xi_i(\omega)$, $i = 1, \dots, M$, are independent random variables, then the joint PDF ρ of $\boldsymbol{\xi} := (\xi_1, \dots, \xi_M)$ can be expressed as

$$\rho(\boldsymbol{\xi}) = \prod_{i=1}^M \rho_i(\xi_i), \quad \boldsymbol{\xi} := (\xi_1, \dots, \xi_M) \in \boldsymbol{\Theta} := \prod_{i=1}^M \Theta_i$$

with the support $\boldsymbol{\Theta} \subset \mathbb{R}^M$.

Thus, via the truncated Karhunen-Loève expansion (13), one can express the system of stochastic flow equations (6) and (8) (up to the truncation errors incurred in (14)) as the following parametric deterministic problem on the domain $D \times [0, T] \times \boldsymbol{\Theta}$: Find the global pressure $p(\mathbf{x}, t; \boldsymbol{\xi})$, the total velocity $\mathbf{u}(\mathbf{x}, t; \boldsymbol{\xi})$, and the saturation $s(\mathbf{x}, t; \boldsymbol{\xi})$ of the brine-rich phase such that

$$\begin{aligned} \nabla \cdot \mathbf{u}(\mathbf{x}, t; \boldsymbol{\xi}) &= q, \\ \mathbf{u} &= -\lambda K_M(\mathbf{x}; \boldsymbol{\xi}) [\nabla p(\mathbf{x}, t; \boldsymbol{\xi}) - (f^c \rho^c + f^b \rho^b) \mathbf{g}], \\ \phi \frac{\partial s(\mathbf{x}, t; \boldsymbol{\xi})}{\partial t} &+ \nabla \cdot (f^c(s) \mathbf{u} - f^c f^b \lambda K_M(\mathbf{x}; \boldsymbol{\xi}) p'_{cap}(s) \nabla s) \\ &= q^c + f^c f^b (\rho^b - \rho^c) \lambda K_M(\mathbf{x}; \boldsymbol{\xi}) \mathbf{g}, \\ \mathbf{x} \in D, \quad t \in [0, T]; \quad \boldsymbol{\xi} \in \boldsymbol{\Theta}. \end{aligned} \tag{15}$$

Here the truncated intrinsic permeability $K_M(\mathbf{x}; \boldsymbol{\xi}) := \exp(Y_M(\mathbf{x}; \boldsymbol{\xi}))$ can be expressed as

$$K_M(\mathbf{x}; \boldsymbol{\xi}) := \bar{K}(\mathbf{x}) \times K_c(\mathbf{x}; \boldsymbol{\xi}) \tag{16}$$

with

$$\bar{K}(\mathbf{x}) := \exp(\bar{Y}(\mathbf{x})), \quad K_c(\mathbf{x}; \boldsymbol{\xi}) = \exp(Y_c(\mathbf{x}; \boldsymbol{\xi})).$$

The general procedure for the PCM is similar to that for the MC with a different choice of the quadrature points and weights.

1. Generate N_c collocation points $\{\boldsymbol{\xi}^k\}_{k=1}^{N_c}$ as independent random inputs based on a quadrature formula.

2. Solve the system (15) at each collocation point $\boldsymbol{\xi}^k$ for the global pressure $p(\mathbf{x}, t; \boldsymbol{\xi}^k)$, the total velocity $\mathbf{u}(\mathbf{x}, t; \boldsymbol{\xi}^k)$, and the saturation $s(\mathbf{x}, t; \boldsymbol{\xi}^k)$ of the brine-rich phase for $k = 1, 2, \dots, N_c$.
3. Estimate the solution statistics using the corresponding quadrature rule

$$\begin{aligned}
\bar{p}(\mathbf{x}, t) &= \int_{\Theta} p(\mathbf{x}, t; \boldsymbol{\xi}) \rho(\boldsymbol{\xi}) d\boldsymbol{\xi} \approx \sum_{i=1}^{N_c} w_i p(\mathbf{x}, t; \boldsymbol{\xi}_i), \\
\bar{\mathbf{u}}(\mathbf{x}, t) &= \int_{\Theta} \mathbf{u}(\mathbf{x}, t; \boldsymbol{\xi}) \rho(\boldsymbol{\xi}) d\boldsymbol{\xi} \approx \sum_{i=1}^{N_c} w_i \mathbf{u}(\mathbf{x}, t; \boldsymbol{\xi}_i), \\
\bar{s}(\mathbf{x}, t) &= \int_{\Theta} s(\mathbf{x}, t; \boldsymbol{\xi}) \rho(\boldsymbol{\xi}) d\boldsymbol{\xi} \approx \sum_{i=1}^{N_c} w_i s(\mathbf{x}, t; \boldsymbol{\xi}_i), \\
\sigma_p(\mathbf{x}, t) &= \left(\sum_{k=1}^{N_c} w_k p(\mathbf{x}, t; \boldsymbol{\xi}_k)^2 - \bar{p}(\mathbf{x}, t)^2 \right)^{1/2}, \\
\sigma_{\mathbf{u}}(\mathbf{x}, t) &= \left(\sum_{k=1}^{N_c} w_k \mathbf{u}(\mathbf{x}, t; \boldsymbol{\xi}_k)^2 - \bar{\mathbf{u}}(\mathbf{x}, t)^2 \right)^{1/2}, \\
\sigma_s(\mathbf{x}, t) &= \left(\sum_{k=1}^{N_c} w_k s(\mathbf{x}, t; \boldsymbol{\xi}_k)^2 - \bar{s}(\mathbf{x}, t)^2 \right)^{1/2}
\end{aligned} \tag{17}$$

where $\{w_k\}_{k=1}^{N_c}$ is a set of weights corresponding to the set of quadrature points $\{\boldsymbol{\xi}^k\}_{k=1}^{N_c}$.

In summary, in the PCM approach the deterministic and stochastic parts of the input data $Y(\mathbf{x}; \omega) = \ln K(\mathbf{x}; \omega)$, which is assumed to be a second-order continuous stationary random process, are separated by the Karhunen-Loéve expansion (9). Then the expansion is truncated after M terms in (13) based on the error criterion (14). With a change of measure from $d\mathcal{P}(\omega)$ to $\rho(\boldsymbol{\xi})d\boldsymbol{\xi}$ which is standard in modern probability theory, the original system of stochastic flow equations (6) and (8) that has random input data is converted to an $D + 1 + M$ -dimensional, parametric deterministic system (15) that is imposed on the product domain $D \times [0, T] \times \Theta$. Then system (15) is enforced at a sequence of collocation points $\boldsymbol{\xi}^k \in \Theta$ for $k = 1, 2, \dots, N_c$. Then at each collocation point $\boldsymbol{\xi}^k$ with $k = 1, 2, \dots, N_c$, the resulting system (15) has the same functional form as the underlying deterministic system. Hence, the PCM, like the MC, is non-intrusive. Furthermore, all the collocated systems (15) with $\boldsymbol{\xi} = \boldsymbol{\xi}^k$ for $k = 1, 2, \dots, N_c$ are fully decoupled in the parametric

space Θ and can be trivially parallelized. The overall computational cost of the PCM approach is the number of collocation points times the cost of solving the corresponding deterministic problem.

In general, the physical domain D may be decomposed as a union of N_p different subdomains $D^{(n)}$ for $n = 1, \dots, N_p$. In each subdomain $D^{(n)}$ the random system parameter $Y^{(n)}(\mathbf{x}; \omega)$ can be described, with a given degree of accuracy, by $M^{(n)}$ independent random variables $\xi_i^{(n)}(\omega)$ ($i = 1, \dots, M^{(n)}$) in accordance with (13) with M replaced by $M^{(n)}$. Then the total number of independent random variables $\xi_i^{(n)}(\omega)$ needed to describe $Y^{(n)}(\mathbf{x}; \omega)$ for all $\mathbf{x} \in D = \bigcup_{n=1}^{N_p} D^{(n)}$ is $M = \sum_{n=1}^{N_p} M^{(n)}$. For instance, in the current context of CO₂ sequestration, different $Y(\mathbf{x}; \omega)$ with different stochastic properties may be specified in the porous medium formation and in the leaky well, which is modeled as a porous medium with a higher permeability than the formation.

4.2. Selection of collocation points and sparse grids

The random dimension M in (13) that is needed to avoid erroneous oscillations for both mean and variance estimates could increase rapidly as the correlation length or the regularity of the covariance function decreases. In such high-dimensional context, the grid set based on tensor product of one-dimensional constructs of J collocation points yields a total number of $N_c = J^M$ collocation points that increases exponentially with M . This is the well known effect of the "the curse of dimensionality".

We follow the idea of [10, 45, 49, 60, 73, 74] to use the Smolyak formula [61] to evaluate the solution statistics in (17). The resulting sparse grid has a significantly smaller number of points than the full tensor product set and does not depend as strongly on the dimensionality of the random space and so is more suitable for applications with high-dimensional random inputs. We briefly describe the algorithm here. Without loss of generality, we assume that the random vectors ξ has a compact support $\Theta \subset \mathbb{R}^M$.

For each random dimension $i = 1, \dots, M$, construct a one-dimensional quadrature formula for a smooth function g :

$$\mathcal{U}_j(g) = \int_{\Theta_j} g(\xi) p_j(\xi_j) d\xi_j \approx \mathcal{U}_j^{i_j}(g) = \sum_{k_j=1}^{n_j} w_{k_j}^{i_j} g(\xi_{k_j}^{i_j}). \quad (18)$$

Here $i_j \in \mathbb{N}$ specifies "the degree of the integration" in the j -th random dimension, where n_j is the number of quadrature points used to approxi-

mate the integral and $\xi_{k_j}^{i_j}$ is the k_j -th quadrature point in the j -th random dimension with a weight $w_{k_j}^{i_j}$. This formula serves as a building block for the Smolyak formula. The latter approximates the M -dimensional integrals in (17) with a linear combination of tensor products as [10, 61]

$$A_{q,M}(g) \equiv \sum_{q-M+1 \leq |\mathbf{i}| \leq q} (-1)^{q-|\mathbf{i}|} \binom{M-1}{q-|\mathbf{i}|} (\mathcal{U}_1^{i_1} \otimes \cdots \otimes \mathcal{U}_M^{i_M})(g). \quad (19)$$

Here the sparseness parameter $q \geq M$ determines the order of the formula, $l := q - M$ is called the “level” of the Smolyak formula, and $|\mathbf{i}| = i_1 + i_2 + \cdots + i_M$. Next, we set $n_{i_j} (j = 1, \dots, M)$, the number of quadrature (collocation) points in (17), to $n_1 = 1$ and $n_l = 2^{l-1} + 1$ for $l > 1$. Then, to compute $A_{q,M}(g)$, we only need to evaluate g on the Smolyak “sparse grid”, i.e. on the collocation point set Ψ_M

$$\Psi_M := \bigcup_{q-M+1 \leq |\mathbf{i}| \leq q} (\Psi_1^{i_1} \times \cdots \times \Psi_M^{i_M}), \quad (20)$$

where $\Psi_j^{i_j}$ represents a one-dimensional collocation point set in the j -th random dimension.

4.3. A Eulerian-Lagrangian localized adjoint method time-stepping procedure for system (15)

At each sparse collocation point $\boldsymbol{\xi}^k$ with $k = 1, 2, \dots, N_c$, the system (15) reduces to a deterministic two-phase flow problem to which we present a Eulerian-Lagrangian localized adjoint method (ELLAM) based iterative time-stepping procedure. Other characteristic methods, such as the characteristic mixed finite element method [3] and the modified method of characteristics (with or without adjusted advection) [21, 22, 23, 24], can also be used in the time-stepping procedure. We define a temporal partition on the time interval $[0, T]$ by $0 = t_0 < t_1 < \cdots < t_n < \cdots < t_N = T$ with time steps $\Delta t_n = t_n - t_{n-1} (n = 1, 2, \dots, N)$ for a positive integer N . In the time-stepping procedure the saturation equation and the pressure equation are solved sequentially in an iterative manner at each time step t_n for $n = 1, 2, \dots, N$. Hence, we present an ELLAM formulation for the saturation equation in (15) with the parameter $\boldsymbol{\xi}$ dropped and assume that the total velocity $\mathbf{u}(\mathbf{x}, t_n)$ and the global pressure $p(\mathbf{x}, t_n)$ are known. For simplicity of presentation, we assume a noflow boundary condition is imposed and refer

the readers to [14, 66, 71] for the derivation of an ELLAM formulation for general boundary conditions and the corresponding analysis.

The ELLAM framework was originally presented in [14], which provides a general characteristic solution procedure for time-dependent advection-diffusion equations and a consistent framework for treating general boundary conditions and maintaining mass conservation. Thus, the ELLAM methodology overcomes the principal shortcomings of many characteristic methods while maintaining their numerical advantages. Let $z(\mathbf{x}, t)$ be any test function that is continuous and piecewise smooth on the space-time strip $D \times (t_{n-1}, t_n]$ except at time t_{n-1} , at which $z(\mathbf{x}, t)$ is left continuous and has right limit with respect to time t . We multiply the saturation equation in (15) by such space-time test functions $z(\mathbf{x}, t)$, integrate the resulting equation on strip $D \times (t_{n-1}, t_n]$, and incorporate the noflow boundary condition to obtain a space-time weak formulation

$$\begin{aligned}
& \int_D \phi(\mathbf{x}) s(\mathbf{x}, t_n) z(\mathbf{x}, t_n) d\mathbf{x} \\
& + \int_{t_{n-1}}^{t_n} \int_D f^c f^b \lambda K_M(\mathbf{x}) p'_{cap}(s) \nabla s(\mathbf{x}, t) \cdot \nabla z(\mathbf{x}, t) d\mathbf{x} dt \\
& - \int_{t_{n-1}}^{t_n} \int_D s(\mathbf{x}, t) \left[\phi \frac{\partial z(\mathbf{x}, t)}{\partial t} + \frac{f^c(s)}{s} \mathbf{u} \cdot \nabla z(\mathbf{x}, t) \right] d\mathbf{x} dt \quad (21) \\
& = \int_D \phi(\mathbf{x}) s(\mathbf{x}, t_n) z(\mathbf{x}, t_{n-1}^+) d\mathbf{x} \\
& + \int_{t_{n-1}}^{t_n} \int_D (q^c + f^c f^b (\rho^b - \rho^c) \lambda K_M(\mathbf{x}) \mathbf{g}) z(\mathbf{x}, t) d\mathbf{x} dt,
\end{aligned}$$

where $z(\mathbf{x}, t_{n-1}^+) = \lim_{t \rightarrow t_{n-1}^+} z(\mathbf{x}, t)$ takes into account for the fact that $z(\mathbf{x}, t)$ might be discontinuous at time t_{n-1} .

In the ELLAM framework [14, 66], the test function $z(\mathbf{x}, t)$ is chosen to satisfy the hyperbolic part of the adjoint equation of the saturation equation

$$\phi \frac{\partial z(\mathbf{x}, t)}{\partial t} + \frac{f^c(s)}{s} \mathbf{u} \cdot \nabla z(\mathbf{x}, t) = 0. \quad (22)$$

Thus, the test function $z(\mathbf{x}, t)$ should be constant along the characteristic curve $\mathbf{y} := \mathbf{r}(t; \mathbf{x}, t_n)$ defined by the initial-value problem of the ordinary differential equation

$$\frac{d\mathbf{r}(t; \mathbf{x}, t_n)}{dt} = \frac{f^c(s)}{s} \frac{\mathbf{u}}{\phi}, \quad t \in [t_{n-1}, t_n]. \quad (23)$$

Once it is specified in $\overline{D}$ at the time step t_n , the test function $z(\mathbf{x}, t)$ is determined completely in the space-time strip $\overline{D} \times [t_{n-1}, t_n]$ by a constant extension of $z(\mathbf{x}, t_n)$ along the characteristic curve $\mathbf{r}(t; \mathbf{x}, t_n)$ from the time step t_n to the time step t_{n-1} . Incorporating such test function into equation (21), applying a backward Euler quadrature to evaluate the temporal integral, and dropping the local truncation error yields an ELLAM numerical scheme

$$\begin{aligned}
& \int_D \phi(\mathbf{x}) s(\mathbf{x}, t_n) z(\mathbf{x}, t_n) d\mathbf{x} \\
& + \Delta t_n \int_D f^c f^b \lambda K_M(\mathbf{x}) p'_{cap}(s) \nabla s(\mathbf{x}, t_n) \cdot \nabla z(\mathbf{x}, t_n) d\mathbf{x} \\
& = \int_D \phi(\mathbf{x}) s(\mathbf{x}, t_n) z(\mathbf{x}, t_{n-1}^+) d\mathbf{x} \\
& + \Delta t_n \int_D (q^c + f^c f^b (\rho^b - \rho^c) \lambda K_M(\mathbf{x}) \mathbf{g}) z(\mathbf{x}, t_n) d\mathbf{x}.
\end{aligned} \tag{24}$$

Tracking the characteristic curve (22) for the two-phase flow process [22, 23] is more delicate than that for miscible displacement processes [21, 24, 30, 32, 65, 68]. Near a moving front in a strongly nonlinear multiphase flow process, additional care might need to be taken, such as operator-splitting or the use of certain fast Riemann solvers [2, 9, 20, 29, 43, 44]. In the ELLAM iterative time-stepping procedure, a mixed finite element method was used to solve the pressure equation and an iterative sequential decoupling technique is used to solve the coupled system [30, 32, 65, 68, 70].

5. Numerical experiments

In this section we conduct preliminary numerical experiments to investigate the performance of the probabilistic collocation ELLAM iterative time-stepping procedure for the system of CO₂-brine two-phase flow with random permeability. The test problem is constructed based on a benchmark leakage problem of injected CO₂ into overlying formations through a leaky well [18]. Figure 4 illustrates a vertical two-dimensional cross section of the three-dimensional porous medium formation under consideration. CO₂ is injected into a deep aquifer, spreads within the aquifer and, upon reaching a leaky well, rises up to a shallower aquifer. The goal of the simulation is to quantify the leakage rate, which depends on the pressure build-up in the aquifer due to injection and the buoyancy of CO₂. The leaky well and the injection well

are 100m away. Both aquifers are 30m thick and the separating aquitard is 100m thick. The leaky well is modeled as a porous medium with a higher permeability than the formation.

Parameter	Value
CO ₂ density, ρ_g	479kg/m ³
Brine density, ρ_w	1045 kg/m ³
CO ₂ viscosity, μ_g	3.950×10^{-5} Pa · s
Brine viscosity, μ_w	2.535×10^{-4} Pa · s
Aquifer permeability, K_A	2×10^{-14} m ²
Aquifer thickness	30m
Aquitard thickness	100m
Leaky well permeability, K_L	1×10^{-12} m ²
Porosity, ϕ	0.15
Leakage and injection well radius	0.15m
Injection rate	8.87kg/s(1600m ³ /d)
Distance between wells	100m
Dimensions of model domain	1000m×1000m×160m
Simulation time, t	1000days

Table 1: Simulation parameters

The following assumptions are made in the benchmark study [18]: Fluid properties such as density and viscosity are constant; all processes are isothermal; CO₂ and the brine are two separate and immiscible phases; mutual dissolution is neglected; the formation is isotropic; and the capillary pressure is negligible. On the lateral boundary the pressure is constant over time and obeys a hydrostatic pressure distribution which is dependent on the brine density and is 3.086×10^7 Pa at the bottom (at 3000m depth) of the lateral boundary. All other boundaries are noflow boundaries. The relative permeabilities are the linear functions of the saturation. CO₂ is injected at a constant rate of 8.87 kg/s, which corresponds to 1600m³/d. The aquifers are initially filled with brine. Simulation time is 1000 days. The relevant parameters used for the simulation are given in Table 1.

The numerical experiment focuses on the validation of the probabilistic collocation ELLAM method on the sparse grid by the Monte Carlo simulations. With this in mind, we consider a smaller version of the above benchmark problem. In this version of the problem, the dimensions of the domain is 300m × 300m × 160m. The injection well is located at $x=150$ m

Figure 4: Leakage scenario from Class et al. [18]

and $y=150\text{m}$ with z ranging from -3000m to -2970m . The leaky well is located at $x=250\text{m}$ and $y=150\text{m}$ with z ranging from -3000m to -2840m . The simulation time is 30 days.

In the numerical simulations, away from the neighborhood of the injection well and the leaky well, a uniform mesh of $dx = dy = 10\text{ m}$ and $dz = 6\text{m}$ is used. Within the neighborhood of the injection well or the leaky well, the grid size is gradually reduced to $dx = dy = 0.26\text{m}$ while dz is still equal to 6m . A time step size of 1 day is used. We plot the locally refined mesh in Figure 5.

5.1. A uniformly distributed log intrinsic permeability

In the numerical experiment, we consider the intrinsic permeability of the reservoir formation and that of the leaky well as the uncertain parameters. We assume that the log intrinsic permeability $Y(\mathbf{x};\omega)$ satisfies a uniform distribution on the interval $(-5, 5)$, but with different mean permeabilities, that is, $\bar{Y}_A = \ln K_A$ for the aquifer and $\bar{Y}_L = \ln K_L$ for the leaky well with the permeabilities K_A and K_L specified in Table 1. We also assume that the correlation length $\eta_1 = \eta_2 = \eta_3 = 1$ in (12). Four random dimensions (i.e., $M = 4$) are used in (13) [46]. We validate the probabilistic collocation ELLAM time-stepping procedure (PC-ELLAM) on sparse grids against the traditional Monte Carlo simulations of the same ELLAM (MC-ELLAM)

Figure 5: Illustration of numerical mesh and the leaky well

Figure 6: The mean saturation of the CO_2 phase at 30 days for the uniformly distributed permeability: (a) the PC-ELLAM simulation on sparse grids, (b) the MC-ELLAM simulation.

Figure 7: The standard deviation of the CO_2 phase at 30 days for the uniformly distributed permeability: (a) the PC-ELLAM simulation on sparse grids, (b) the MC-ELLAM simulation.

time-stepping procedure. We use a Monte Carlo simulation of the ELLAM time-stepping procedure with 4000 realizations as a reference solution. The accuracy of the PC-ELLAM simulation is evaluated via comparison to the MC-ELLAM simulation. In Figures 6 and 7 we present the three-dimensional plots of the mean $\bar{s}$ and the standard deviation σ_s of the saturation s obtained from the PC-ELLAM approach with 2025 points compared with the MC-ELLAM simulation with 4000 points.

In Figure 8 we present the mean $\bar{s}$ and the standard deviation σ_s of the saturation s along the line $y = 150\text{m}$, $z = -2973\text{m}$ obtained from the MC-ELLAM simulation with 4000 realizations as well as the PC-ELLAM simulations on sparse grids with 81, 513 and 2025 points (that correspond to 4 levels, 5 levels, and 6 levels in the Smolyak formula), for the purpose of investigating the convergence behavior of the PC-ELLAM scheme. In Figure 9 we present the mean $\bar{s}$ and the standard deviation σ_s of the saturation s of the estimated CO_2 leakage through the leaky well.

The numerical results show that the PC-ELLAM with 81 sparse collocation points generates satisfactory agreement with the MC-ELLAM simulation with 4000 realizations. Namely, the accuracy of the PC-ELLAM can be retained from far fewer collocation points than the MC-ELLAM samples. It is evident that the PC-ELLAM approach is computationally more efficient than the MC-ELLAM simulations.

Figure 8: (a) The mean $\bar{s}$ and (b) the standard deviation σ_s of the injected CO₂ at 30 days for the uniformly distributed permeability, along the line $y = 150$ m and $z = -2973$ m with x from 0 to 300m, obtained from the MC-ELLAM and the PC-ELLAM simulations.

Figure 9: (a) The mean and (b) the standard deviation of CO₂ leakage for the uniformly distributed permeability, obtained from the MC-ELLAM and the PC-ELLAM simulations.

Figure 10: The mean saturation of the CO_2 phase at 30 days for the lognormal permeability: (a) the PC-ELLAM simulation on sparse grids, (b) the MC-ELLAM simulation.

Figure 11: The standard deviation of the CO_2 phase at 30 days for the lognormal permeability: (a) the PC-ELLAM simulation on sparse grids, (b) the MC-ELLAM simulation.

Figure 12: (a) The mean $\bar{s}$ and (b) the standard deviation σ_s of the injected CO₂ at 30 days for the lognormal permeability, along the line $y = 150\text{m}$ and $z = -2973\text{m}$ with x from 0 to 300m, obtained from the MC-ELLAM and the PC-ELLAM simulations.

5.2. A lognormal intrinsic permeability

In the numerical experiments, we assume that the log intrinsic permeability $Y(\mathbf{x}; \omega)$ satisfies a normal distribution, with the same mean permeabilities for the aquifer and for the leaky well as in the previous subsection. All the data are still given in Table 1. We still use four random dimensions (i.e., $M = 4$) in (13) [46]. We validate the probabilistic collocation ELLAM time-stepping procedure (PC-ELLAM) on sparse grids against the traditional Monte Carlo simulations of the same ELLAM (MC-ELLAM) time-stepping procedure.

We use a Monte Carlo simulation of the ELLAM time-stepping procedure with 1000 realizations as a reference solution. We then use the PC-ELLAM scheme with 9 points, 45 points, and 165 points, respectively, in numerical simulations, compared to the reference solution. In Figures 10 and 11 we present the three-dimensional plots of the mean $\bar{s}$ and the standard deviation σ_s of the saturation s obtained from the PC-ELLAM approach with 165 points compared with the MC-ELLAM simulation with 1000 points.

In Figure 12 we present the mean $\bar{s}$ and the standard deviation σ_s of the saturation s along the line $y = 150\text{m}$, $z = -2973\text{m}$ obtained from the MC-ELLAM simulation with 1000 realizations as well as the PC-ELLAM simulations on sparse grids with 9, 45 and 165 points, for the purpose of investigating the convergence behavior of the PC-ELLAM scheme. In Figure 13 we present the mean $\bar{s}$ and the standard deviation σ_s of the saturation s of the estimated CO₂ leakage through the leaky well.

Figure 13: (a) The mean and (b) the standard deviation of CO₂ leakage for the lognormal permeability, obtained from the MC-ELLAM and the PC-ELLAM simulations.

We also carry out preliminary sensitivity analysis by varying the standard deviation σ and the correlation length $\eta_1 = \eta_2 = \eta_3$ in (12) to investigate their impact, and present the plots in Figures 14 and 15.

We observe from Figure 12 that the PC-ELLAM simulation with 9 points already generates very satisfactory results, which shows a greatly improved efficiency of the PC-ELLAM scheme. We also note from the sensitivity study that increasing the standard deviation of the log intrinsic permeability or the correlation length would result in an increase of the volatility of the saturation distribution.

6. Concluding remarks

We developed a probabilistic collocation Eulerian-Lagrangian localized adjoint method on sparse grids for assessing CO₂ leakage through wells in randomly heterogeneous porous media, by utilizing the intrinsic mathematical, numerical, and physical properties of the mathematical model. We model the process in which CO₂ is injected into a deep aquifer, spreads within the aquifer and, upon reaching a leaky well, rises up to a shallower aquifer, to quantify the leakage rate, which depends on the pressure build-up in the aquifer due to injection and the buoyancy of CO₂. The following assumptions were made on the mathematical model: all processes are isothermal; CO₂ and the brine are two separate and immiscible phases; mutual dissolution is neglected; and the leaky well is modeled as a porous medium with a higher permeability than the formation and both the formation and the well have the same probability distribution. The underlying Eulerian-Lagrangian

Figure 14: The mean of the saturation of the injected CO_2 phase at 30 days for lognormal permeability with different standard deviation and correlation length, along the line $y=150\text{m}$ and $z=-2973\text{m}$ with x from 0 to 300m, obtained from the PC-ELLAM simulations with 9 points.

Figure 15: The standard deviation of the saturation of the injected CO_2 phase at 30 days for lognormal permeability with different standard deviation and correlation length, along the line $y=150\text{m}$ and $z=-2973\text{m}$ with x from 0 to 300m, obtained from PC-ELLAM with 9 points.

framework has high potential to improve the efficiency and accuracy for the numerical simulation of complex flow and transport processes in CO₂ sequestration [2, 3, 9, 22, 30, 43, 44, 66]. The sparse grid probabilistic collocation framework adds computationally efficient uncertainty quantification functionality onto pre-existing Eulerian-Lagrangian methods in a nonintrusive manner. It also provides a scalable framework to consider uncertainty in a straightforward parallel manner. The preliminary numerical experiments in this paper demonstrated the feasibility and potential of the method.

This paper represents a first step in our effort to develop a probabilistic collocation Eulerian-Lagrangian localized adjoint method on sparse grids for assessing CO₂ leakage through wells in randomly heterogeneous porous media. Many very difficult issues remain to be resolved: (i) The slight miscibility of CO₂ into the aqueous phase, and of H₂O into the CO₂-rich phase, can play important roles because of their corrosive nature. Under acidic conditions, well cements (especially in abandoned wells) may be degraded by leakage of the acidic aqueous phase, which can in turn enhance the leakage along the well. Depending on the buffering capacity in the injection formation, and the initial state of a given well, this geochemical degradation may play an important role in leakage estimation and in the overall risk assessment. For the CO₂-brine system, the important components are CO₂ dissolved in the brine, water (H₂O) evaporated into the CO₂-rich phase, and the total dissolved salts in the brine. These may then be complemented by a much larger number of components that result from the carbonated brine - rock geochemical reactions, or the brine-cement reactions [4, 41, 42]. In addition, if the injected CO₂ is not pure but includes some fraction of other emission products like sulfur and nitrogen compounds, then these components need to be included in the analysis as well [75]. These problems will lead to a multiphase multicomponent compositional flow and transport process coupled with geochemical reaction and with random inputs due to the uncertainty of the porous medium basin and, even more importantly, the uncertainty of primarily human-reduced leaky wells. Very little work has been reported in the literature on the Eulerian-Lagrangian method for multiphase multicomponent compositional flow and transport [69]. (ii) A range of (geo)chemical reactions occurring in the compositional flow and transport processes can be especially important along critical leakage pathways such as fault zones and old oil and gas wells. For example, cements used in wells may degrade in the presence of low-pH brines, while well casings can be corroded by wet CO₂ [11, 12, 58]. This significantly complicates the problem because, in addition

to more standard geochemical reactions within the formation, material reactions along concentrated leakage pathways, at scales of millimeters or smaller, may also need to be resolved appropriately. Millimeter-scale multiphase flow and reactive transport must thus be simulated, perhaps along thousands of such pathways (see Figures 1 and 3), and these calculations must be embedded into an overall field-scale model that can cover thousands of square kilometers and includes all important geological heterogeneities. Finally, brine with CO_2 dissolved is slightly more dense than brine without CO_2 , and that leads to important density impacts in the aqueous transport problem over longer time scales. In other words, the problem of assessing CO_2 leakage is inherently multiscale in space and time, which requires further investigation.

(iii) While the probabilistic collocation Eulerian-Lagrangian localized adjoint method on sparse grids can be used to simulate complex CO_2 leakage problem, it is severely limited by computational constraints due to the extreme complexities of the problem. And while simplified models were proposed to solve the kinds of probabilistic leakage problems that are needed for practical decision making, they are limited by the severe set of assumptions needed to make them computationally tractable [50, 52, 53, 54, 55]. A feasible objective is to develop a new set of computational methods that can fill in the large gap between these two model types, with a specific focus on more efficient models to solve the complex set of equations, hybrid models that combine both the simplified and the complex in a kind of multi-scale framework, and a more efficient treatment that will allow the probabilistic treatment of the leakage problem to be accomplished much more efficiently. Certain upscaling or multiscale techniques might also need to be applied [17, 27, 28, 67].

(iv) We considered the critical leakage pathways like human-induced, abandoned wells and assumed that they are vertical. In reality some of these wells may be oblique. In addition, critical leakage pathways can also include faults. In this context, discontinuous Galerkin methods and mortar methods [36, 63, 64, 72] may be explored in the development of probabilistic collocation Eulerian-Lagrangian localized adjoint method.

Acknowledgements

This work was supported in part by the National Science Foundation under Grants EAR-0934722, EAR-0934747, and DMS-1216923, by the National Natural Science Foundation of China under Grants 91130010 and 11471194, and by the State Scholarship Fund from China Scholarship Council. The

authors would like to express their sincere thanks to the referees for their very helpful comments and suggestions, which greatly improved the quality of this paper.

References

- [1] G. Acs, S. Doleschall, and E. Farkas, General purpose compositional model, *Soc. Petrol. Engrg. J.*, 25:543–553, 1985.
- [2] T. Arbogast, Ch.-S. Huang, and T. F. Russell, A locally conservative Eulerian-Lagrangian method for a model two-phase flow problem in a one-dimensional porous medium, *SIAM J. Sci. Comput.*, 34 (2012), A1950-A1974.
- [3] T. Arbogast and M.F. Wheeler, A characteristic-mixed finite element method for advection-dominated transport problems, *SIAM J. Numer. Anal.*, 32:404–424, 1992.
- [4] Audigane, P., I. Gaus, I. Czernichowski-Lauriol, K. Priess, and T. Xu, Two-dimensional Reactive Transport Modeling of CO₂ Injection in a Saline Aquifer at the Sleipner Site, North Sea, *American Journal of Science*, Vol. 307, pp. 974-1008, 2007.
- [5] H. Aziz and A. Settari, *Petroleum Reservoir Simulation*, Applied Science Publishers, New York, 1979.
- [6] I. Babuška, F. Nobile, and R. Tempone, A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM J. Numer. Anal.*, 45:1005-1034, 2007.
- [7] Bachu, S. and M.A. Celia, Assessing the Potential for CO₂ Leakage, Particularly through Wells, from CO₂ Storage Sites”, to appear, *The Science and Technology of Carbon Sequestration*, AGU Monograph, 2009.
- [8] J. Bear, *Dynamics of Fluids in Porous Materials*, American Elsevier, New York, 1972.
- [9] P.J. Binning and M.A. Celia, A finite volume Eulerian-Lagrangian localized adjoint method for solution of the contaminant transport equations in two-dimensional multi-phase flow systems, *Water Resources Research*, 32, (1996) 103–114.

- [10] H.-J. Bungartz and M. Griebel, Sparse grids, *Acta Numerica*, 2004, 1–123.
- [11] Carey, J.W., Wigand, M., Chipera, S., WoldeGabriel, G., Pawar, R., Lichtner, P., Wehner, S., Raines, M., and G. D. Guthrie, Analysis and Performance of Oil Well Cement with 30 Years of CO₂ Exposure from the SACROC Unit, West Texas, USA, *International Journal of Greenhouse Gas Control*, 1, 75-85, 2007.
- [12] Carey, J.W., J. Zhang, P.C. Lichtner, R. Grigg, B. Svec, and R. Pawar, Reaction of CO₂ and Brine at the Interface between Portland Cement and Casing Steel: Application to CO₂ Sequestration, *American Geophysical Union*, Fall Meeting 2008, Abstract #H11J-08, 2008.
- [13] M.A. Celia, S. Bachu, J.M. Nordbotten, D. Kavetski, and S. Gasda, A risk assessment tool to quantify CO₂ leakage potential through wells in mature sedimentary basins. In *Proceedings of the 8th Conference on Greenhouse Gas Technologies*, 20-25. June 2006.
- [14] M.A. Celia, T.F. Russell, I. Herrera, and R.E. Ewing, An Eulerian-Lagrangian localized adjoint method for the advection-diffusion equation, *Adv. Water Resour.*, 13 (1990), 187–206.
- [15] G. Chavent and J. Jaffré, *Mathematical Models and Finite Elements for Reservoir Simulation*, North-Holland, Amsterdam, 1986.
- [16] Z. Chen, G. Huan, and Y. Ma, *Computational Methods for Multiphase Flows in Porous Media*, Computational Science and Engineering Series, 2, SIAM Philadelphia, 2006.
- [17] A. Cheng, K. Wang, and H. Wang, A preliminary study on multiscale ELLAM schemes for transient advection-diffusion equations, *Numer. Methods Partial Diff. Eqns.*, 26 (2010), 1405–1419.
- [18] Class, H., Ebigbo, A., Helmig, R., Dahle, H., Nordbotten, J., N., Celia, M., A., Audigane, P., Darcis., M., Ennis-King, J., Fan, Y., Flemisch, B., Gasda, S., Jin, M., Krug, S., Labregere; D., Naderi, A., Pawar, R., J., Sbair, A., Sunil, G.,T., Trenty, L., Wei, L., A benchmark study on problems related to CO₂ storage in geologic fomations, *Computational Geosciences*, 13(2009), 451-467.

- [19] G. Dagan, Flow and transport in porous formations. New York: Springer; 1989.
- [20] H.K. Dahle, R.E. Ewing, and T.F. Russell, Eulerian-Lagrangian localized adjoint methods for a nonlinear convection-diffusion equation, *Comput. Meth. Appl. Mech. Engrg.*, 122, (1995) 223–250.
- [21] C.N. Dawson, T.F. Russell, and M.F. Wheeler, Some improved error estimates for the modified method of characteristics, *SIAM Numer. Anal.*, 26 (1989), 1487–1512.
- [22] J. Douglas, Jr., F. Furtado, and F. Pereira, On the numerical simulation of waterflooding of heterogeneous petroleum reservoirs, *Computational Geoscience*, 1, (1997) 155–190.
- [23] J. Douglas, Jr., C.-S. Huang, and F. Pereira, The modified method of characteristics with adjusted advection. *Numer. Math.*, 83 (1999), 353–369.
- [24] J. Douglas, Jr. and T.F. Russell, Numerical methods for convection-dominated diffusion problems based on combining the method of characteristics with finite element or finite difference procedures, *SIAM Numer. Anal.*, 19 (1982), 871–885.
- [25] Dullien, F.A.L., *Porous Media: Fluid Transport and Pore Structure*, Second Edition, Academic Press, 1992.
- [26] A. Dur, On the optimality of the discrete KarhunenLove expansion, *SIAM J. Control Optim.*, 1998 36:1937-1949.
- [27] Y. Efendiev, J. Galvis, R. Lazarov, S. Margenov, and J. Ren, Robust two-level domain decomposition preconditioners for high-contrast anisotropic flows in multiscale media, *Computational Methods in Applied Mathematics*, 12 (4) (2012), 415–436.
- [28] Y. Efendiev, J. Galvis, R. Lazarov, M. Moon, M. Sarkis, Generalized Multiscale Finite Element Method. Symmetric Interior Penalty Coupling, *J. Comput. Physics*, 255 (2013) 1 – 15.
- [29] M.S. Espedal and R.E. Ewing, Characteristic Petrov-Galerkin subdomain methods for two-phase immiscible flow, *Comp. Meth. Appl. Mech. and Engrg.*, 64:113–135, 1987.

- [30] Ewing, R.E., Russell, T.F., Wheeler, M.F.: Simulation of miscible displacement using mixed methods and a modified method of characteristics, SPE 12241:71–81 (1983)
- [31] R.E. Ewing, (ed), *The Mathematics of Reservoir Simulation*, Research Frontiers in Applied Mathematics, 1, SIAM, Philadelphia, 1984.
- [32] R.E. Ewing, T.F. Russell, and M.F. Wheeler, Convergence analysis of an approximation of miscible displacement in porous media by mixed finite elements and a modified method of characteristics, *Comput. Meth. Appl. Mech. Eng.*, 47, (1984), 73–92.
- [33] Fernando, H., J. Venezia, C. Rigdon, and P. Verma, Capturing King Coal: Deploying Carbon Capture and Storage Systems in the U.S. at Scale, World Resources Institute, 2008.
- [34] L.W. Gelhar, Stochastic subsurface hydrology. Englewood Cliffs, NJ: Prentice-Hall; 1993.
- [35] Ghanem, R.G. and P. Spanos, *Stochastic Finite Elements: a Spectral Approach*, Springer-Verlag, 1991.
- [36] V. Girault, S. Sun, M. F. Wheeler and I. Yotov, "Coupling discontinuous Galerkin and mixed finite element discretizations using mortar finite elements", SIAM Journal on Numerical Analysis, 46(2), 949979, 2008.
- [37] R. Helmig, *Multiphase Flow and Transport Processes in the Subsurface*, Springer, Berlin, 1997.
- [38] International Energy Agency (IEA), Energy Technology Perspectives 2008, OECD/IEA, 2008.
- [39] Intergovernmental Panel on Climate Change (IPCC), Special Report on Carbon Dioxide Capture and Storage. Edited by B. Metz et al., Cambridge University Press, 431 pp., 2005
- [40] Intergovernmental Panel on Climate Change (IPCC), Climate Change 2007: Synthesis Report. Contributions of Working Groups I, II, and III to the Fourth Assessment Report of the IPCC [Core Writing team, Pachauri, R.K. and Reisinger, A. (eds.)]. IPCC, Geneva, Switzerland, 104 pp., 2007

- [41] Johnson, J.W., J.J. Nitao, and K.G. Knauss, Reactive Transport Modeling of CO₂ Storage in Saline Aquifers to Elucidate Fundamental Processes, Trapping Mechanisms, and Sequestration Partitioning”, In: Baines, S.J. and R.H. Worden (Eds.), *Geological Storage of Carbon Dioxide*, Geological Society of London, Special Publications, 233:107-128, 2004.
- [42] Johnson, J.W., J.J. Nitao, and J.P. Morris, Reactive Transport Modeling of Cap-rock Integrity during Natural and Engineered CO₂ Storage, Chapter 8, *Carbon Dioxide Capture for Storage in Deep Geologic Formations*, Vol. 2, D.C. Thomas and S.M. Benson (Eds.), Elsevier, 2005.
- [43] K.H. Karlsen, K. Brusdal, H.K. Dahle, and S. Evje, and K.-A. Lie, The corrected operator splitting approach applied to a nonlinear advection-diffusion problem, *Comput. Meth. Appl. Mech. Engrg.*, 167, (1998) 239–260.
- [44] K.H. Karlsen, K.-A. Lie, and N.H. Risebro, A fast marching method for reservoir simulation, *Comput. Geosci.* 4, (2000) 185–206.
- [45] Q. Liao and D. Zhang, Probabilistic collocation method for strongly nonlinear problems: 1. Transform by location, *Water Resour. Res.*, 49(2013), 7911-7928.
- [46] G. Lin, A.M. Tartakovsky, An efficient, high-order probabilistic collocation method on sparse grids for three-dimensional flow in random porous media, *Adv. Water Resour.* 32, (2009) 712–722.
- [47] M. Loève, *Probability theory*. 4th ed. New York: Springer-Verlag, 1977.
- [48] M.I.T. Report, The Future of Coal: Options for a Carbon-constrained World, M.I.T., 2007.
- [49] F. Muller, P. Jenny, and D.W. Meyer, Probabilistic collocation and Lagrangian sampling for advective tracer transport in randomly heterogeneous porous media, *Adv. Water. Resour.* , 34(2011), 1527-1538.
- [50] J.M. Nordbotten, M.A. Celia, S. Bachu, and H.K. Dahle, Semi-analytical solution for CO₂ leakage through an abandoned well, *Environmental Science and Technology*, 39:602–611, 2005.

- [51] Nordbotten, J.M., M.A. Celia, and S. Bachu, Injection and Storage of CO₂ in Deep Saline Aquifers: Analytical Solution for CO₂ Plume Evolution during Injection, *Transport in Porous Media*, 58, 339-360, 2005.
- [52] Nordbotten, J.M. and M.A. Celia: Similarity Solutions for Fluid Injection into Confined Aquifers, *Journal of Fluid Mechanics*, 561, 307-327, 2006.
- [53] Nordbotten, J.M. and M.A. Celia, An Improved Analytical Solution Interface Upconing around a Well", *Water Resources Research*, 42, (doi:10.1029/2005WR004738), 2006.
- [54] J.M. Nordbotten and M.A. Celia, Analysis of plume extent using analytical solutions for CO₂ storage. In *Proceedings of the 16th conference on Computational Methods in Water Resources*, edited by Philip J. Binning et al. Copenhagen, Denmark, June, 2006 (<http://proceedings.cmwr-xvi.org>), 2006.
- [55] Nordbotten, J.M., D. Kavetski, M.A. Celia, S. Bachu, A Semi-analytical Model Estimating Leakage associated with CO₂ Storage in Large-scale Multi-layered Geological Systems with Multiple Leaky Wells, published online 17 December 2008, *Environmental Science and Technology*, 2008.
- [56] Oladyshkin, S., Class, H., Helmig, R., Nowak, W., An integrative approach to robust design and probabilistic risk assessment for CO₂ storage in geological formations. *Computational Geosciences*, 15(2011), 565-577.
- [57] Pacala, S. and R. Socolow, Stabilization Wedges: Solving the Climate Problem for the Next 50 Years with Current Technologies, *Science*, 305: 968-972, 2004.
- [58] Scherer, G.W., M.A. Celia, J.H. Prevost, S. Bachu, R. Bruant, A. Duguid, R. Fuller, S.E. Gasda, M. Radonjic, and W. Vichit-Vadakan, Leakage of CO₂ through Abandoned Wells: Role of Corrosion of Cement, in *The CO₂ Capture and Storage Project (CCP)*, Volume II, D.C. Thomas and S.M. Benson (Eds.), 823-844, 2005.
- [59] Socolow, R.H., Can We Bury Global Warming? *Scientific American*, pages 49-55, July 2005.

- [60] Ch. Schwab and R.-A. Todor, Sparse finite elements for stochastic elliptic problems, *Numer. Math.* 95 (2003) 707–734.
- [61] Smolyak, S.A., Quadrature and interpolation formulas for tensor products of certain classes of functions, *Soviet Math. Dokl.*, 4:240–243, 1963.
- [62] Socolow, R.H. and S.W. Pacala: A Plan to keep Carbon in Check *Scientific American*, Vol. 295, pp. 50-57 (2006).
- [63] S. Sun and M. F. Wheeler, Anisotropic and dynamic mesh adaptation for discontinuous Galerkin methods applied to reactive transport, *Computer Methods Appl. Mech. Engrg.*, 195, 3382–3405, 2006.
- [64] S. Sun and M. F. Wheeler, Local problem-based a posteriori error estimators for discontinuous Galerkin approximations of reactive transport, *Comput. Geos.*, 11, 87–101, 2007.
- [65] H. Wang, An optimal-order error estimate for a family of ELLAM-MFEM approximations to porous medium flow, *SIAM J. Numer. Anal.*, 46 (2008), 2133–2152.
- [66] H. Wang, H.K. Dahle, R.E. Ewing, M.S. Espedal, R.C. Sharpley, and S. Man, An ELLAM Scheme for advection-diffusion equations in two dimensions, *SIAM J. Sci. Comput.*, 20 (1999) 2160–2194.
- [67] H. Wang, Y. Ding, K. Wang, R.E. Ewing, Y.R. Efendiev, A multiscale Eulerian-Lagrangian localized adjoint method for transient advection-diffusion equations with oscillatory coefficients, *Comput. Visual. Sci.*, 12 (2009), 63–70, 2009.
- [68] H. Wang, D. Liang, R.E. Ewing, S.L. Lyons, and G. Qin, An approximation to miscible fluid flows in porous media with point sources and sinks by an Eulerian-Lagrangian localized adjoint method and mixed finite element methods, *SIAM J. Sci. Comput.*, 22 (2000), 561–581.
- [69] H. Wang, W. Zhao, M.S. Espedal, and A.S. Telyakovskiy, A component-based Eulerian-Lagrangian formulation for multicomponent multiphase compositional flow and transport in porous media, *SIAM J. Sci. Comput.*, 35 (2013), B462-B486.

- [70] H. Wang, W. Zhao, R.E. Ewing, M. Al-Lawatia, M.S. Espedal, and A.S. Telyakovskiy, An Eulerian-Lagrangian solution technique for single-phase compositional flow in three-dimensional porous media, *Computers Math. Appl.*, 52 (2006), 607–624.
- [71] K. Wang and H. Wang, An optimal-order error estimate to ELLAM schemes for transient advection-diffusion equations on unstructured meshes, *SIAM J. Numer. Anal.*, 48 (2010), 681–707.
- [72] M. F. Wheeler, G. Xue, and I. Yotov, A Multipoint Flux Mixed Finite Element Method on Distorted Quadrilaterals and Hexahedra, *Numerische Mathematik*, 121 (2012), 165–204.
- [73] D. Xiu and J.S. Hesthaven, High-order collocation methods for differential equations with random inputs, *SIAM J. Sci. Comput.*, 27 (2005), 1118–1139.
- [74] D. Xiu and G.E. Karniadakis, Modeling uncertainty in flow simulations via generalized polynomial chaos, *J. Comput. Phys.*, 187:137–167, 2003.
- [75] D. Zhang, *Stochastic Methods for Flow in Porous Media, Coping with Uncertainties*, Academic Press, New York, 2002.,