Advances in Water Resources 56 (2013) 49–60

Contents lists available at SciVerse ScienceDirect

Advances in Water Resources

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

Assessing leakage detectability at geologic CO_2 sequestration sites using the probabilistic collocation method

Alexander Y. Sun [a,*], Mehdi Zeidouni [a], Jean-Philippe Nicot [a], Zhiming Lu [b], Dongxiao Zhang [c]

[a] Bureau of Economic Geology, Jackson School of Geosciences, The University of Texas at Austin, TX, USA
[b] EES-16, Los Alamos National Laboratory, Los Alamos, NM, USA
[c] College of Engineering, Peking University, Beijing, China

ARTICLE INFO

Article history:
Received 5 June 2012
Received in revised form 20 November 2012
Accepted 25 November 2012
Available online 5 December 2012

Keywords:
Carbon sequestration and storage
Leakage detection
Probabilistic collocation method
Detectability
Signal-to-noise ratio
Uncertainty quantification

ABSTRACT

We present an efficient methodology for assessing leakage detectability at geologic carbon sequestration sites under parameter uncertainty. Uncertainty quantification (UQ) and risk assessment are integral and, in many countries, mandatory components of geologic carbon sequestration projects. A primary goal of risk assessment is to evaluate leakage potential from anthropogenic and natural features, which constitute one of the greatest threats to the integrity of carbon sequestration repositories. The backbone of our detectability assessment framework is the probability collocation method (PCM), an efficient, nonintrusive, uncertainty-quantification technique that can enable large-scale stochastic simulations that are based on results from only a small number of forward-model runs. The metric for detectability is expressed through an extended signal-to-noise ratio (SNR), which incorporates epistemic uncertainty associated with both reservoir and aquifer parameters. The spatially heterogeneous aquifer hydraulic conductivity is parameterized using Karhunen–Loève (KL) expansion. Our methodology is demonstrated numerically for generating probability maps of pressure anomalies and for calculating SNRs. Results indicate that the likelihood of detecting anomalies depends on the level of uncertainty and location of monitoring wells. A monitoring well located close to leaky locations may not always yield the strongest signal of leakage when the level of uncertainty is high. Therefore, our results highlight the need for closed-loop site characterization, monitoring network design, and leakage source detection.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Carbon capture and storage is an active R&D area being studied across the world for reducing greenhouse gas emission. A primary goal of geologic carbon sequestration is to ensure that injected CO_2 can be safely contained in host formations for extensive performance periods. Unintended CO_2 migration from storage formations can occur as a result of natural and anthropogenic impacts, such as overpressure in cap rocks and leakage through geologic faults and abandoned wells. Therefore, risk analysis and management play critical roles in all stages of CO_2 sequestration projects for mitigation of potential health and environmental risks. Many countries have or are in the process of mandating risk management for site license applications. For example, the European Union explicitly requires that a proposed CO_2 sequestration site shall "show no significant risk of leakage and no significant environmental or health risks under the proposed conditions of use" [1]. Under its Underground Injection Control program, the US Environmental Protection Agency (EPA) has proposed specific rules on CO_2 injection wells to protect underground sources of drinking water (USDW). In particular, the proposed EPA rules have extensive requirements to ensure that wells used for geologic sequestration are appropriately sited, constructed, tested, monitored, funded, and closed [2]. Recently, the US Department of Energy releases best practice manuals on risk analysis and management activities related to CO_2 storage projects [3,4].

Another important and closely related goal of risk management is to optimize monitoring networks for timely detection of CO_2 leakage under both model and data uncertainty. The first and foremost concern of regulators is how efficiently a proposed monitoring network can detect leakage signals when they first appear. *Detectability* refers to the capability of an observer or a piece of equipment to differentiate between *noise* and *signal plus noise* during an arbitrary observation interval or sampling event [5]. Assessment of detection probability invariably requires quantitative information on signal-to-noise ratios (SNRs). In a series of papers [6–8], Taguchi used SNR as a metric for optimal product design, which he referred to as "the operation of choosing settings for the design parameters of a product or manufacturing process to reduce sensitivity to noise". A now widely used Taguchi SNR for product design is the logarithm of the ratio between mean and

* Corresponding author. Tel.: +1 512 475 6190; fax: +1 512 471 0140.
E-mail address: alex.sun@beg.utexas.edu (A.Y. Sun).

0309-1708/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.advwatres.2012.11.017

standard deviation of observed signals; alternatively, the ratio can be used directly without taking logarithm. The optimization principles behind Taguchi's design are to (1) minimize product sensitivity to variations transmitted from components and (2) minimize product sensitivity to environmental fluctuations. We adopt the same principles for the design of CO_2 monitoring networks under parametric uncertainty.

In the context of the current work, noise is broadly interpreted as uncertainty caused by (1) natural variability in a system (aleatory uncertainty) and/or (2) lack of complete knowledge about system characteristics (epistemic uncertainty). Common examples of the latter are (1) hydrogeologic (or reservoir) properties of monitoring intervals, including both the injection zone itself and aquifers above the injection zone (i.e., the above-zone monitoring interval (AZMI)), and (2) the type of leakage source and its properties [9]. An inverse relationship generally exists between cost and sensitivity of leakage-detection techniques, which inherently depend on four macroscales pertaining to (1) the region needs to be monitored, (2) the region affected by leakage flux, (3) the main leakage zone (e.g., scale of sources), (4) the footprint of the monitoring equipment used [10,11], as well as on pore- and other microscales. Physical and chemical processes associated with these scales can all potentially affect SNR. Macroscales 1 and 4 are usually either known or can be practically considered deterministic. Macroscales 2 and 3, however, are rather dynamic and uncertain. The lack of information on system properties and length scales identified here, especially during the planning stage of many CO_2 sequestration projects, often leads to a situation in which uncertainty quantification (UQ) becomes the dominant question, overriding the influence of secondary physical processes [12].

The role of UQ in risk analysis and management is thus twofold. First, it helps in identifying the dominant system and environmental variables that contribute to system response variability, an analysis that is also important for subsequent activities such as data collection and monitoring network design. Second, UQ yields bounding scenarios for system outputs and, therefore, provides direct inputs to risk-informed performance assessment. UQ can become more powerful when coupled with data assimilation techniques and applied to site management adaptively. However, to be suitable for data fusion or real-time decision support, a UQ technique must be highly efficient. The purpose of this work is to investigate the use of the probabilistic collocation method (PCM), an efficient stochastic-response surface method for assessing detectability under parameter uncertainty.

Models have been used extensively to simulate the migration of injected CO_2 through various leakage pathways [13–17]. However, a number of challenges potentially exist when these models are applied to UQ using conventional techniques. For example, perturbation methods and stochastic-moment approaches require deriving and solving a set of coupled stochastic partial differential equations (PDEs) corresponding to various uncertain input variables, which is a nontrivial task for complex, nonlinear processes. In addition, the assumptions underlying these approaches largely restrict their applicability to small parameter variability [18]. Whereas Monte Carlo methods do not impose strong assumptions on the variability of uncertain variables and do not require modifications of existing codes (i.e., nonintrusive), they are computationally demanding and become intractable for large-scale problems without access to parallel or distributed computing facilities.

In recent years, a new breed of UQ techniques—the polynomial chaos expansion (PCE) method and stochastic collocation (SC) method—have received broad attention in engineering-reliability analyses [19–24]. Both UQ techniques belong to the so-called stochastic-response surface methods, and both represent parametric uncertainties as an expansion of orthogonal polynomials of independent random variables and propagate them to quantify model-output uncertainty. Exponential convergence rates can be achieved by both methods for a wide range of probabilistic analysis problems [21].

The classical PCE method, pioneered by Ghanem and Spanos [24], is based on the homogeneous chaos theory of Wiener [25]. The method starts with a spectral expansion of input uncertain variables, through which the variables are projected onto a stochastic space spanned by a set of complete orthogonal polynomials. A main consequence of this spectral expansion is that the uncertain variables are represented as a deterministic part (i.e., coefficients of expansion) and a stochastic part (i.e., polynomial chaos basis). Galerkin projection is then applied to each polynomial chaos basis, thereby replacing stochastic PDEs with a coupled deterministic system of equations, from which the coefficients of expansion can be solved for. Ghanem and Spanos [24] worked with Hermite polynomial chaos, which is optimal for Gaussian random variables. Xiu and Karniadakis [21] later introduced generalized polynomial chaos expansion using the Wiener–Askey scheme, such that a number of commonly used continuous and discontinuous probability distribution functions (PDFs) could be accommodated. The classical PCE method requires developing and solving a coupled system of deterministic, ordinary differential equations, a procedure that can be cumbersome and nontrivial when the problem at hand is complex and nonlinear.

The SC method, like its deterministic counterpart in the finite element method, seeks to construct a response surface using a prescribed set of collocation points, but in stochastic space rather than physical space. In the SC method, the expansion coefficients are nothing but model outputs corresponding to each of the collocation points. The quantities to be solved for are expansion polynomials, which themselves are based on Lagrange interpolation polynomials. The SC method is closely related to PCE because collocation points that offer high accuracy are also zeros of orthogonal polynomial bases used in PCE [20]. By design, the SC method is nonintrusive and leads naturally to uncoupled deterministic systems, as opposed to coupled system of equations resulting from Galerkin projection. The efficiency and accuracy of SC depend largely on the number and location of collocation points. Several methods exist for generating collocation points, such as the tensor product grid and Smolyak sparse grid. The tensor product method is suitable only for low-dimensional systems because of the large number of collocation points it generates. The sparse grid approach uses a subset of the points generated by the tensor product method and therefore can be significantly more efficient [19,26,27].

A variant of the SC method is the probabilistic collocation method (PCM), which combines features of both the classical PCE and SC methods. Like PCE, it prescribes a set of orthogonal polynomial bases. The expansion coefficients, however, are obtained by solving a linear system of equations, in which the right-hand-side data vector (or matrix for transient problem) consists of model responses at collocation points. The PCM, originally introduced by Tatang et al. [28], is attractive for engineering-reliability analyses because it is nonintrusive and its implementation is relatively straightforward. The PCM has been used to solve both single- and multiphase flow and mass transport problems in porous media [29–33]. Recently, Oladyshkin et al. [12] used PCM in risk assessment of a hypothetical CO_2 sequestration site, in which the authors considered the impact of uncertain parameters (reservoir porosity and permeability and leaky-well permeability) and design parameters (injection rate and size of screening interval) on predicted injection-zone variables (i.e., cap-rock pressure and CO_2 leakage rate). Oladyshkin et al. [12] assumed that all uncertain variables are independent and uniform in physical space; the uncertain variables were represented using Hermite polynomial chaos, which is not optimal for non-Gaussian random variables, as we have pointed out before. Walter et al. [34] later applied a similar approach to study

migration of brine into a multi-aquifer system by considering three uncertain parameters (injection zone permeability, aquifer anisotropy, and fault permeability). The authors assumed that the aquifer anisotropy is depth-independent. Although the assumption leads to reduction in random dimensions in their problem, it may not be realistic [35].

In this work, PCM is used to assess the effect of spatial heterogeneity on detectability of pressure anomalies in AZMI. In the literature, the effect of spatial heterogeneity has been investigated for the injection zone, for which the main concern is related mostly to CO_2 trapping mechanisms [36], shape of CO_2 plumes [37–39], and conceptual-model uncertainty [40]. Heterogeneity in the AZMI and connectivity of various geologic features can create preferential flow paths which, in turn, adversely affect a monitoring network's ability to detect pressure anomaly. Therefore, it is important to consider the effect of spatial heterogeneity when quantifying leakage probabilities.

We use Karhunen–Loève (KL) expansion, a commonly used dimension-reduction technique, to parameterize the spatially heterogeneous hydraulic-conductivity field. The resulting low-dimensional representation is then used in the PCM framework to propagate uncertainty. The analyses presented in this work reduce dimensionality of stochastic space and significantly improve the efficiency of stochastic simulation. Although only the pressure signal is considered here, our methodology can be readily adapted to quantifying uncertainties related to other types of signals (e.g., geochemical and thermal) and, thus, will provide valuable tools for assessing the detectability of CO_2 leakage in a model-driven framework. Note that we distinguish the UQ problem considered here from the pressure-anomaly inversion problem. The former is for risk assessment using given system (design) parameters and a prior information on parameter uncertainty, whereas the latter is for identifying leakage characteristics (location and leaking history) using observed pressure data. Although the two are closely related, the former problem focuses on model design in a forward sense, whereas the latter focuses more on remediation in an inversion sense and was addressed separately by Sun and Nicot [9], who showed that inversion of pressure-anomaly data can be a promising technology for quick identification of CO_2 leakage sources.

This paper is organized as follows. Section 2 presents the PCM approach. Section 3 demonstrates the use of PCM for assessing AZMI pressure-anomaly detectability through a number of numerical experiments. Finally, Section 4 summarizes the main findings of the current work and provides future research directions.

2. Technical approaches

2.1. Pressure anomaly detection in the AZMI

AZMI monitoring is recommended for CO_2 sequestration projects to mitigate risks associated with CO_2 leakage and minimize potential adverse impacts on USDW [41]. We introduce the formulation of the AZMI flow problem in this section. The approach to couple AZMI and reservoir flows will be discussed in Section 3.2. Let us consider an AZMI in which pressure is monitored for potential leakage from the underlying CO_2 injection zone. To be consistent with hydrogeologic terminology, we shall use *hydraulic head* in most of the following presentation, unless noted otherwise. The governing equation for saturated groundwater flow is

$$\begin{aligned} S_s(\mathbf{x})\tfrac{\partial h(\mathbf{x},t)}{\partial t} + \nabla\cdot\mathbf{q}(\mathbf{x},t) &= Q_w(\mathbf{x},t) \\ \mathbf{q}(\mathbf{x},t) &= -K_s(\mathbf{x})\nabla h(\mathbf{x},t) \end{aligned} \tag{1}$$

subject to initial and boundary conditions

$$h(\mathbf{x},0) = H_0(\mathbf{x}), \quad \mathbf{x}\in D \tag{2}$$

$$h(\mathbf{x},t) = H_b(\mathbf{x},t), \quad \mathbf{x}\in\Gamma_D \tag{3}$$

$$\mathbf{q}(\mathbf{x},t)\cdot\mathbf{n}(\mathbf{x}) = Q_b(\mathbf{x},t), \quad \mathbf{x}\in\Gamma_N \tag{4}$$

where h is hydraulic head; $\mathbf{q}$ is Darcy flux; S_s is specific storage; K_s is saturated hydraulic conductivity; Q_w is the sink/source term; H_0 is initial condition; D denotes the interior of the model domain; H_b and Q_b define the Dirichlet and Neumann boundary conditions on boundary segments Γ_D and Γ_N, respectively; and $\mathbf{n}(\mathbf{x})$ is an outward unit vector normal to the boundary.

As mentioned in the Introduction, predicting the extent and magnitude of head anomalies caused by CO_2 leakage is of great interest from a risk-assessment perspective because head perturbations travel faster than other signals [14,42–44] and their cost of acquisition is lower than that of other signals. However, detectability of head anomaly (e.g., SNR) can be subjected to interference by model and data uncertainties and, therefore, should be evaluated in a probabilistic sense. In the rest of this section, we show how PCM can be used to predict spatiotemporal variations in the hydraulic head field.

2.2. Probability collocation method (PCM)

Treating uncertain parameters (e.g., K_s) in Eq. (1) through (4) as random variables, we have a stochastic PDE, which can be denoted as

$$\mathcal{L}(\mathbf{x},t,\omega;h) = Q_w(\mathbf{x},t,\omega), \quad \mathbf{x}\in D, \quad \omega\in\Omega \tag{5}$$

where ω is a random event belonging to the outcome space, Ω, and $\mathcal{L}$ is a differential operator acting on h, which is itself a stochastic process, namely $h = h(\mathbf{x},t,\omega)$.

2.2.1. Generalized polynomial chaos

The starting point of PCM is to represent a response variable of interest using spectral expansion. Under Wiener's homogeneous chaos theory [24,25], a random variable is projected onto a stochastic space spanned by a set of Hermite polynomials, $\Psi(\xi)$, which, in turn, are functions of an infinite-dimensional random vector, ξ, consisting of independent standard Gaussian variables. Hermite polynomial chaos is orthogonal with respect to the Gaussian measure and converges in probability space to Gaussian distribution in a mean-square sense [20]. Although non-Gaussian variables may be represented using Gaussian variables through probability-space transformation, the convergence rate is typically degraded, and higher-order polynomial expansion is required. Xiu and Karniadakis [21] proposed generalized polynomial chaos expansion for non-Gaussian random variables, under which a general second-order, random process, $\mathcal{X}$, with finite variance can be represented by the following infinite expansion:

$$\mathcal{X}(\mathbf{x},t,\omega) = \sum_{i=0}^{\infty} a_i(\mathbf{x},t)\Psi_i(\xi(\omega)) \tag{6}$$

where ω is a random event; $\xi(\omega)$ is a vector of independent random variables; $\{a_i(\mathbf{x},t)\}$ are unknown expansion coefficients, which are also referred to as random modes of the system; and $\{\Psi_i(\xi)\}$ is a set of complete multivariate polynomials that satisfy the following orthogonality relation

$$\langle\Psi_i,\Psi_j\rangle = \left\langle\Psi_i^2\right\rangle\delta_{ij} \tag{7}$$

where δ_{ij} is Kronecker-delta function ($\delta_{ij}=1$ if $i=j$, and 0 otherwise), and the inner product operator, $\langle\cdot\rangle$, is defined as

$$\langle\Psi_i,\Psi_j\rangle = \int \Psi_i(\xi)\Psi_j(\xi)f(\xi)d\xi \tag{8}$$

with $f(\xi)$ the joint PDF of ξ. In practice, the infinite series given in Eq. (6) is truncated at finite terms

$$\mathcal{X}(\mathbf{x},t,\omega)=\sum_{i=0}^{M}a_i(\mathbf{x},t)\Psi_i(\xi(\omega)) \tag{9}$$

and the total number of terms is given by

$$M+1=\frac{(N+p)!}{N!p!} \tag{10}$$

where N denotes total random dimensions (i.e., the sum of dimensions of uncertain input variables) of the UQ problem and p is the order of polynomial expansion. The dimension of ξ is N, and the highest degree of each component of ξ is p. For example, the set of complete orthogonal bases for a second-order expansion involving two random dimensions consist of six terms

$$\begin{aligned}
\Psi_0(\boldsymbol{\xi})&=\psi_0(\xi_1)\psi_0(\xi_2)\\
\Psi_1(\boldsymbol{\xi})&=\psi_1(\xi_1)\psi_0(\xi_2)\\
\Psi_2(\boldsymbol{\xi})&=\psi_0(\xi_1)\psi_1(\xi_2)\\
\Psi_3(\boldsymbol{\xi})&=\psi_2(\xi_1)\psi_0(\xi_2)\\
\Psi_4(\boldsymbol{\xi})&=\psi_1(\xi_1)\psi_1(\xi_2)\\
\Psi_5(\boldsymbol{\xi})&=\psi_0(\xi_1)\psi_2(\xi_2)
\end{aligned} \tag{11}$$

where ψ_j $(j=1,2)$ are univariate polynomials and the dependence of ξ_i $(i=1,2)$ on ω is omitted for clarity. Eq. (11) indicates that $\Psi_i(i=0,\ldots,5)$ are simply products of univariate polynomials for each random dimension. As mentioned before, different PDFs have different optimal orthogonal polynomials under the framework of generalized polynomial chaos. In the example given in Eq. (11), the following set of univariate Hermite polynomials provide the optimal bases for Gaussian distribution for $p=2$:

$$\psi_0(\xi_i)=1,\quad \psi_1(\xi_i)=\xi_i,\quad \psi_2(\xi_i)=\xi_i^2-1,\quad i=1,2 \tag{12}$$

where ξ_i is standard Gaussian variable $\mathcal{N}(0,1)$. If a random variable follows uniform distribution instead, the optimal polynomial set will consist of Legendre polynomials,

$$\psi_0(\xi_i)=1,\quad \psi_1(\xi_i)=\xi_i,\quad \psi_2(\xi_i)=\frac{1}{2}\left(3\xi_i^2-1\right),\quad i=1,2 \tag{13}$$

and ξ_i becomes a uniform variable defined on interval $[-1,1]$, i.e., $\mathcal{U}(-1,1)$. A nice feature of the generalized polynomial chaos is that we can mix and match univariate polynomial bases for different distributions. Again, using the example in Eq. (11) and assuming the PDF of the first random variable is Gaussian and that of the second is uniform, we would then use cross-products of the univariate polynomials defined in Eqs. (12) and (13), and the six terms in Eq. (11) become

$$\begin{aligned}
&\Psi_0=1,\quad \Psi_1=\xi_1,\quad \Psi_2=\xi_2,\\
&\Psi_3=\xi_1^2-1,\quad \Psi_4=\xi_1\xi_2,\quad \Psi_5=\frac{1}{2}\left(3\xi_2^2-1\right)
\end{aligned} \tag{14}$$

The moments of Ψ_i are evaluated by performing integration using these standard random variables and their PDFs, which is straightforward for independent variables.

2.2.2. Formulation of the PCM system of equations

Applying the finite-sum series of Eq. (9) to approximate $h(\mathbf{x},t,\omega)$, we have

$$\hat{h}(\mathbf{x},t,\omega)=\sum_{i=0}^{M}a_i(\mathbf{x},t)\Psi_i(\xi) \tag{15}$$

Substitution of Eq. (15) into Eq. (5) results in a residual term R because of the approximation

$$R(\{a_i\},\xi)=\mathcal{L}(\hat{h})-Q_w \tag{16}$$

The main idea of the weighted-residual method in the traditional finite-element method is to force the residual in Eq. (16) to zero in an average sense, namely,

$$\int_{\xi}R(\{a_i\},\xi)W_j(\xi)f(\xi)d\xi=0,\quad i=0,\ldots,M,\quad j=0,\ldots,M \tag{17}$$

where $f(\xi)$ is the joint PDF of ξ and the number of weighting functions W_j is equal to the number of expansion coefficients a_i. The PCM chooses the Dirac delta function as weighting function [29], namely,

$$W_j(\xi)=\delta(\xi-\xi_j),\quad j=0,\ldots,M \tag{18}$$

where ξ_j is a point in random space and the set of all ξ_j is called *collocation points*. As a result, Eq. (17) becomes a set of uncoupled algebraic equations evaluated at $M+1$ collocation points and with $M+1$ expansion coefficients as unknowns

$$\mathbf{ZA}(\mathbf{x},t)=\mathbf{H}(\mathbf{x},t) \tag{19}$$

where elements of the $(M+1)\times(M+1)$ matrix, $\mathbf{Z}$, consist of $\Psi_i(\xi)$ evaluated at collocation points; columns of $\mathbf{A}(\mathbf{x},t)$ consist of expansion coefficients, a_i, for different simulation times; and columns of $\mathbf{H}(\mathbf{x},t)$ consist of model responses evaluated at collocation points.

In PCM, the collocation points are roots of univariate polynomial basis at one degree higher than the order of expansion [29]. For one-dimensional integration, it is well known that a Gaussian quadrature rule consisting N_p points can integrate polynomials up to degree $2N_p-1$ exactly [45]. For the second-order Hermite polynomial chaos considered in Eq. (12), for example, the roots are obtained from the third-order Hermite polynomial, $\xi^3-3\xi$, whereas for the second-order Legendre polynomial chaos in Eq. (13), the roots are zeros of the third-order Legendre polynomial, $\frac{1}{2}(5\xi^3-3\xi)$. Note that the number of all possible collocation points, $(p+1)^N$, is typically much greater than the number of unknowns, $M+1$. A subset of these collocation points should be selected in such a way that the resulting matrix $\mathbf{Z}$ in Eq. (19) maintains full rank. One way to generate these collocation points is to select roots corresponding to higher probability regions in each dimension first and then gradually move to other roots with lower probabilities [29]. Each time before a new collocation point is included, $\mathbf{Z}$ is tested for its row rank. If the full row-rank condition is not satisfied, the search process will continue until an $(M+1)$-rank $\mathbf{Z}$ is obtained. Using the full-rank $\mathbf{Z}$ we can solve Eq. (19) deterministically for a_i which, in turn, can be used to obtain the mean and variance of hydraulic head [29]

$$\begin{aligned}
\langle h\rangle&=\sum_{i=0}^{M}a_i\langle\Psi_i\rangle=a_0,\\
\sigma_h^2&=\sum_{i=1}^{M}a_i^2\left\langle\Psi_i^2\right\rangle
\end{aligned} \tag{20}$$

where the moments of Ψ_i are calculated as discussed in the text following Eq. (13).

2.2.3. Incorporation of spatial heterogeneity

The number of model runs required by PCM can still be significant when either the random dimensionality or the order of expansion, or both, is large as indicated by Eq. (10). Therefore, it is always desirable to apply some type of dimension reduction to reduce parameter dimensions before applying PCM, especially for spatially distributed parameters. A widely used parameter-dimension reduction technique is KL expansion [24,29,46]. In the current problem, KL expansion is used to parameterize the spatially heterogeneous hydraulic-conductivity field, which is modeled as a second-order stationary random process. Let $Y(\mathbf{x},\omega)=\ln K_s(\mathbf{x},\omega)$ and its spatial covariance be $C_Y(\mathbf{x},\mathbf{x}')$. The KL expansion of $Y(\mathbf{x},\omega)$ takes the form of infinite series

$$Y(\mathbf{x},\omega)=\overline{Y}(\mathbf{x})+\sum_{i=1}^{\infty}\sqrt{\lambda_i}\phi_i(\mathbf{x})\xi_i(\omega) \tag{21}$$

where $\overline{Y}(\mathbf{x})$ is the mean of $Y(\mathbf{x},\omega)$, $\xi_i(\omega)$ are independent standard Gaussian random variables, and λ_i and ϕ_i are eigenvalues and eigenfunctions of the following eigenvalue problem

$$\int_D C_Y(\mathbf{x},\mathbf{x}')\phi_i(\mathbf{x}')d\mathbf{x}' = \lambda_i\phi_i(\mathbf{x}) \tag{22}$$

which can be solved either analytically or numerically, depending on the specific form of C_Y. In practice, the number of terms that needs to be retained in (21) can be set by calculating the ratio between the sum of the first L largest eigenvalues and the sum of all eigenvalues, the latter of which represents the product of total variance and the area or volume of the domain. The truncation is done such that the percentage of retained variance is greater than a user-specified threshold [29,46].

The KL expansion is optimal in the sense that the mean-square error of the finite-term representation is minimized [22]. A direct consequence of KL expansion is that the spatial random process $Y(\mathbf{x},\omega)$ is now represented by a limited number of independent random variables and, therefore, can be readily incorporated into the PCM framework described herein. For demonstration purposes, we assume that the covariance of $Y(\mathbf{x},\omega)$ can be modeled using a separable exponential covariance that takes the following form for two-dimensional random fields [46]

$$C_Y(\mathbf{x},\mathbf{x}') = \sigma_Y^2 \exp\left(-|x_1 - x_1'|/\eta_1 - |x_2 - x_2'|/\eta_2\right) \tag{23}$$

where σ_Y is the standard deviation of $Y(\mathbf{x},\omega)$ and η_1 and η_2 are integral scales along x_1 and x_2, respectively. The covariance model in Eq. (22) is often used in the literature because analytical solutions of its eigenvalue problem exist for rectangular domains (see Appendix A in Zhang and Lu [46]). In general, the eigenvalue problem in Eq. (20) needs to be solved numerically to obtain eigenvalues and eigenfunctions.

3. Numerical experiments

3.1. Problem setup

We now use a series of numerical examples to show how PCM can be used to construct probability maps for assessing detectability of leakage in the AZMI. Fig. 1(a) illustrates the problem setup, in which our system consists of an injection zone (saline aquifer) and an AZMI (confined aquifer), separated by an aquitard. Albeit simple, such a layered system is representative of the basic structure of many geologic sequestration repositories and has been studied extensively in analytical and numerical studies [14,17,43]. The input uncertainty can be caused by a number of uncertain system characteristics, as discussed previously in the Introduction. A common strategy in system performance assessment is to divide the whole system into subsystems and perform UQ on each subsystem separately so that the required computation is manageable. Such a paradigm is behind, for example, the certification framework for CO_2 risk assessment [47], in which a subsystem is referred to as a *compartment*. We shall follow this paradigm in this work and model AZMI as a compartment.

Fig. 1(b) shows the plan view of the actual AZMI model domain, which has lateral dimensions of 1000×1000 m^2 and is 30 m thick. The leaky well is located at the origin of the domain. For the baseline scenario, the monitoring well is put at (100,100) m. Because there is only one source term, we have invoked the symmetry argument and model only a quadrant of the full AZMI domain. Although strictly speaking such symmetry argument is not valid when spatial heterogeneity is involved, our approach is appropriate in an ensemble averaging sense. Inclusion of additional features, such as large-scale faults or pumping wells, would require modeling the entire model domain and would increase the computational time for each realization. However, it does not invalidate our non-intrusive UQ approach. For simplicity, we assume two-dimensional horizontal flow in this work. Incorporating heterogeneity in the vertical dimension is straightforward using the KL expansion and has been considered, for example, by Li et al. [48]. We solved the confined-flow problem using the subsurface module of finite-element software, COMSOL (http://www.comsol.com), which automatically generates computational meshes and provides mesh refinement around wells.

3.2. Solution workflow

Fig. 2 illustrates the workflow for the numerical experiments. We use a widely adopted analytical solution developed by Nordbotten et al. [17] to generate pressure distribution at the top of injection zone, which is then used to calculate leakage flux into AZMI using Darcy's law [49]. The analytical solution of Nordbotten et al. assumes that (1) a sharp interface exists between CO_2 and brine, (2) phase saturations and fluid viscosities are constant within each zone, (3) the capillary effects are small, and (4) vertical equilibrium applies to the entire flow system. Uncertain reservoir parameters considered in this study are permeability and compressibility. Rock compressibility is uncertain due to incomplete knowledge of the pore structure and rock mechanical properties. In the numerical example, both reservoir parameters are assumed to have a triangular PDF, which is commonly used when uncertainty is presented in the form of expert opinions. Other types of PDFs can be easily accommodated if they are supported by site-characterization data.

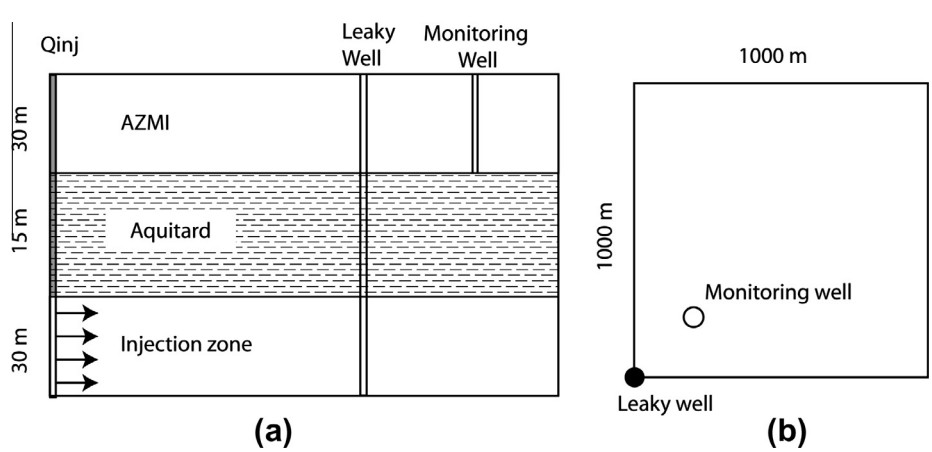

Fig. 1. Setup of demonstration problem: (a) profile view of full system, which includes injection zone, AZMI, and aquitard in between; (b) plan view of AZMI, which shows a leaky well and a monitoring well, at which detectability assessment is performed.

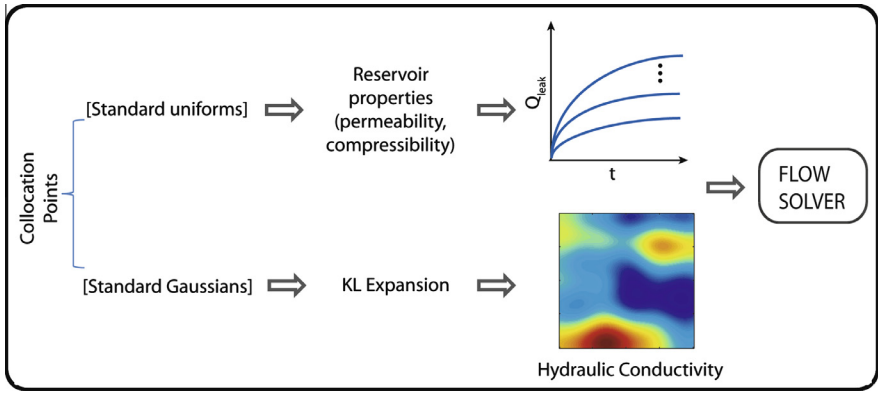

Fig. 2. Workflow diagram for constructing stochastic-response surfaces using PCM, where KL expansion is used to parameterize hydraulic conductivity and leakage fluxes are generated by sampling uncertain reservoir parameters.

Table 1 provides the PDF parameters used for reservoir permeability and compressibility. Table 2 lists all other variables needed by the analytical solution to calculate leakage flux from the injection zone. For risk-assessment purposes, location of the leaky well, required by the analytical solution, is assumed given and is at a distance of 500 m from the injection well in the injection zone. On average, the CO_2 plume reaches a radial distance of about 250 m after 100 days of injection for parameters listed in Table 2 and according to Eq. (6) in Nordbotten et al. [17]. Therefore, the head anomaly is caused mainly by brine leakage during our simulation period, and the solution of a single-phase flow is valid; otherwise, a two-phase problem involving both CO_2 and brine needs to be solved. For AZMI, realizations of log conductivity, Y, are generated through KL expansion using the separable exponential-covariance model (see Eq. (23)). The geometric mean of K_s is 0.07 m/d, which is approximately equal to the mean reservoir permeability for brine properties listed under Table 2. Bilinear interpolation was used to map K_s generated through KL expansion on a structured grid (5×5 m resolution) to the unstructured grid used by COMSOL. Because the grid resolution is fine relative to correlation length of Y, spatial interpolation is expected to have little impact on numerical results.

Table 1
Parameters of reservoir permeability and compressibility probability distributions, both are assumed triangular. Mode corresponds to the peak of triangular PDFs.

Parameter	Mode	Lower bound	Upper bound
Permeability(m^2)	2×10^{-14}	1×10^{-14}	2×10^{-13}
Compressibility (Pa^{-1})	2.1×10^{-10}	2.1×10^{-11}	2.1×10^{-9}

Table 2
All other parameters used in baseline scenario.

Parameter	Value
CO_2 density	479 kg/m^3
Brine density	1045 kg/m^3
CO_2 viscosity	3.95×10^{-5} Pa s
Brine viscosity	2.54×10^{-4} Pa s
Porosity[a]	0.15
Leaky-well permeability	2×10^{-11} m^2
Well radius[b]	0.15 m
CO_2 injection rate	947 m^3/d
Initial head[a]	1950 m
Injection-zone thickness	30 m
Aquitard thickness	15 m
AZMI thickness	30 m
AZMI hydraulic conductivity	0.07 m/d
Aquifer specific storage	1×10^{-6} m^{-1}
Leaky-well location (inj zone)	500 m

[a] Same for both injection zone and AZMI.
[b] Same for both injection well and leaky well.

Each collocation point ξ consists of a number of standard random variables, chosen to be optimal for the distribution of actual uncertain parameters. Two elements of ξ are uniformly distributed variables—one is for reservoir permeability and the other for reservoir compressibility. At runtime, the two uniform variables are transformed back to random samples from triangular distribution. Remaining elements of ξ are standard normal variables used for KL expansion. Each collocation point provides parameters for one COMSOL run. After the model is run $M + 1$ times, expansion coefficients corresponding to all time intervals are obtained simultaneously by solving Eq. (19). We then construct a stochastic response surface for approximating hydraulic head using the resulting expansion coefficients. For this example, the time required for a single COMSOL run is 25 s on a PC workstation.

3.3. Results and discussion

3.3.1. Baseline scenario

For the baseline scenario, Y is assumed to have an isotropic covariance structure with a correlation length of 400 m, and the variance of Y is 1.0. Although not exactly the same, the problem settings here resemble those of the geology in Texas Gulf Coast aquifers, in which the deep saline aquifer is characterized by thick sandstones that are laterally extensive and permeable and separated by regionally extensive shales deposited during marine transgressions [50]. The assumed correlation parameters are within the range found for alluvial aquifers ([51], Table 2.1). The number of terms retained in truncated KL expansion is seven, which has been found to be sufficient for the correlation-length to domain-size ratio considered here [29]. Note that the smaller the ratio between correlation length and domain size, the more KL terms would need to be included to get the same level of variance representation. The resulting total random dimension for the baseline is $N = 9$. The number of collocation points and, thus, model runs required are 55, 220, and 715, respectively, for $p = 2$–4. In this case, the computational time required to generate a full-rank $\mathbf{Z}$ grows significantly, from 1 s for $p = 2$ to 4400 s for $p = 4$.

Fig. 3 shows some realizations of leakage flux generated by random sampling of uncertain reservoir parameters. The leakage flux generally increases rapidly in the beginning and becomes flat later. If fitted to a log-normal distribution [52,53], log-transformed leakage fluxes have a coefficient of variation of 368 at $t = 1$ day and decrease to about 0.33 at 100 days. Fig. 4 shows two "realizations" of Y generated using the seven-term KL expansion ($p = 2$) and for the

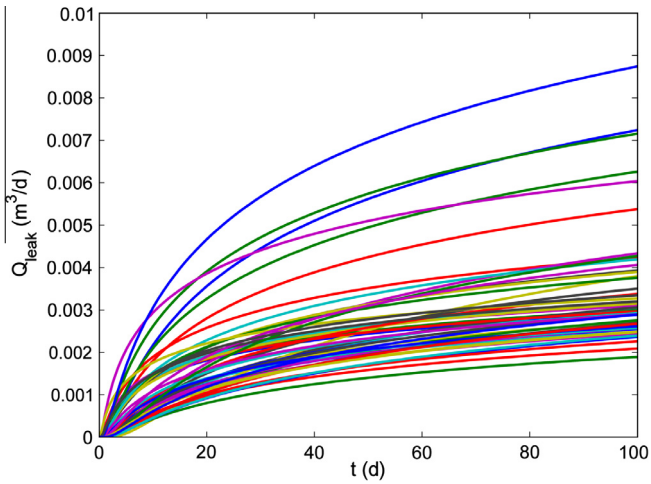

Fig. 3. Examples of leakage flux random realizations generated by sampling uncertain reservoir parameters.

second and third collocation points chosen for the PCM system of equations. The seemingly nonrandom pattern of realizations is due largely to the scheme used for selecting polynomial roots. Recall from Section 2.2 that our collocation-point selection process cycles through roots in high-probability regions first before moving to lower-probability regions. In this case, the collocation points used for generating Fig. 4(a) and (b) are $[-\sqrt{3},0,0,0,0,0,0]$ and $[0,-\sqrt{3},0,0,0,0,0]$, respectively.

Head anomalies, Δh, are computed by subtracting the initial head from model outputs. Fig. 5(a) and (b) show the mean and standard deviation of Δh calculated at the monitoring well location using Eq. (18) for $p = 2$, 3, and 4, respectively. We compared our results to those obtained from 5000 Monte Carlo simulations (open circles). The number of terms used for KL-expansion in the Monte Carlo runs is 60, and Latin Hypercube Sampling was used to generate parameters for each realization. In general, the match between PCM and Monte Carlo simulation improves as the order of chaos expansion increases. Li and Zhang [29] observed that the head variances computed by PCM match well with those calculated by Monte Carlo simulation for $p = 2$ and $\sigma_Y^2 = 1$. In our case, the PCM solutions obtained by the third and fourth-order expansion are close to each other and both are better than that obtained by the second-order expansion. The deviation of PCM from Monte Carlo simulation can be attributed to the large variance in leakage fluxes. For smaller variances, the match is expected to improve. Fig. 5(a) and (b) suggest that the third and fourth-order PCM solutions differ slightly in their curvature in the time period from 0 to 60 days. As we will show later in Fig. 8, this minor difference actually makes a significant difference when SNR is computed.

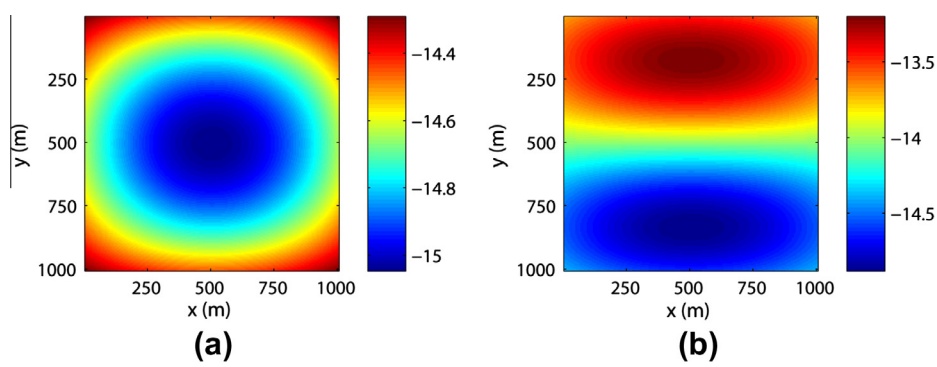

Fig. 4. Log-transformed hydraulic-conductivity fields corresponding to (a) second and (b) third collocation points for the case of second-order polynomial chaos expansion, $p = 2$.

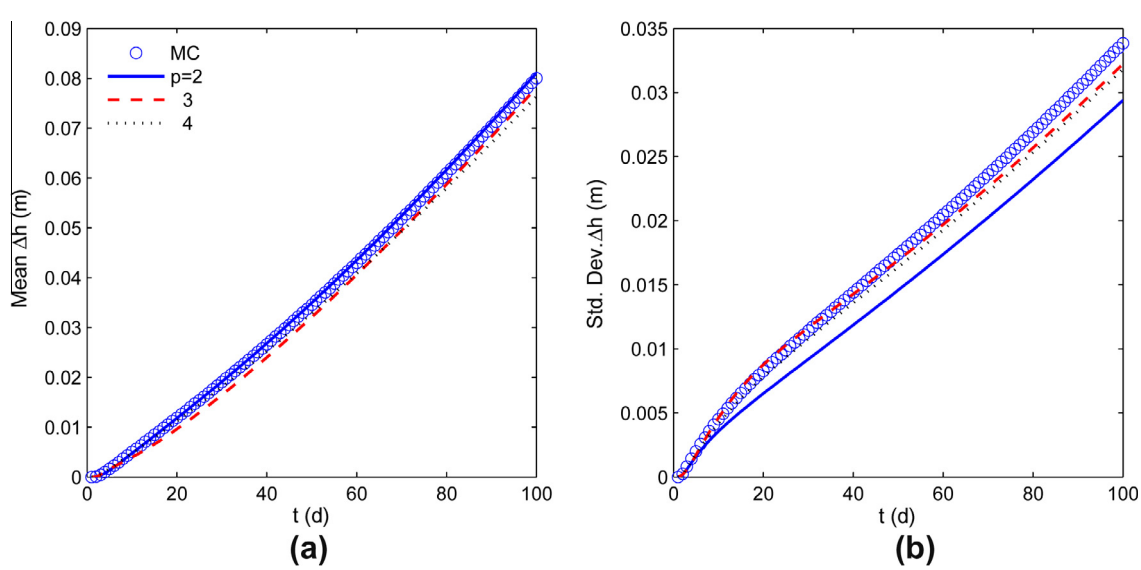

Fig. 5. Comparison between PCM and Monte Carlo simulation for (a) mean and (b) standard deviation of head anomalies, Δh.

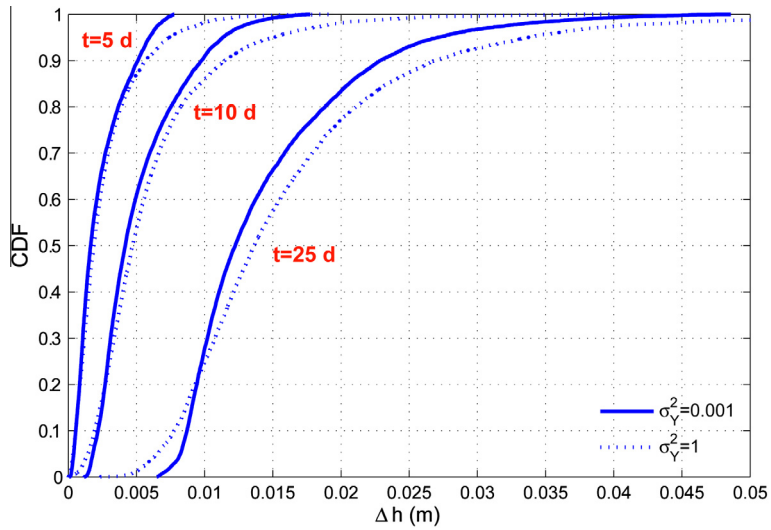

Fig. 6. Empirical CDF of Δh corresponding to 10, 50, and 100 days in base scenario for $\sigma_Y^2 = 0.001$ and 1, respectively. Head anomaly data were generated by running 10,000 stochastic simulations using PCM.

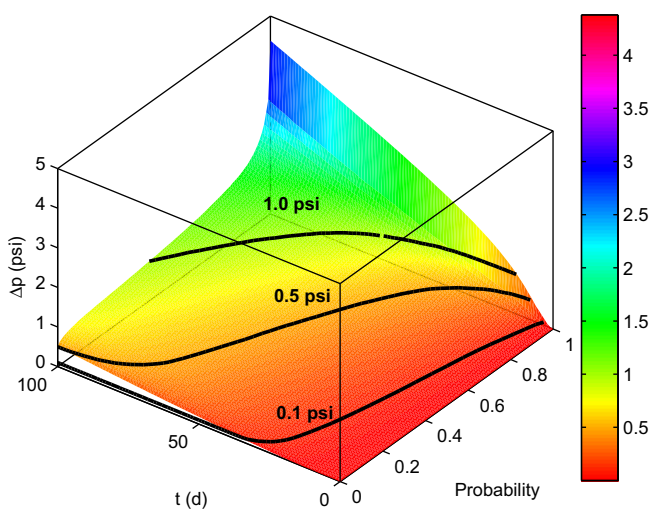

Fig. 7. Detection-probability map constructed by using empirical Δh CDF calculated for each time step. Dark lines are detection thresholds that separate map into detectable and nondetectable regions.

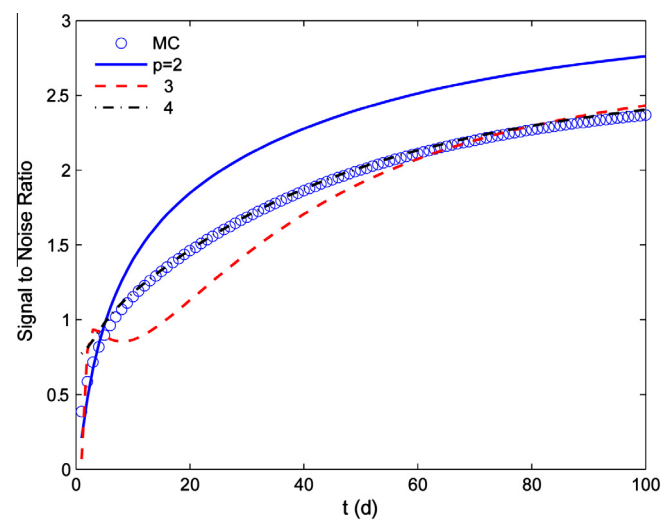

Fig. 8. SNR calculated by both Monte Carlo simulation and PCM.

In the next step, we performed 10,000 stochastic simulations using the PCM expansion coefficients obtained for $p = 4$ and then used the results to calculate the probability maps or empirical cumulative distribution functions (CDF). Fig. 6 shows the CDF plots for $t = 5$, 10, and 25 days, respectively (dotted lines). The Δh CDFs generally show a non-Gaussian pattern. For a given detection threshold Δh_D, and given time, we can use a plot like Fig. 6 to evaluate detectability, $1 - P(\Delta h < \Delta h_D)$. The effect of heterogeneity in the AZMI is contrasted by considering a case with little heterogeneity ($\sigma_Y^2 = 0.001$) while keeping all other parameters fixed. Comparing the two sets of results in Fig. 6, we see that the effect of AZMI heterogeneity, as parameterized here, is manifested by prolonging tails of the CDF, a well-known result for flow in heterogeneous porous media. For example, Nowak et al. [54] studied head variance in bounded, three-dimensional, heterogeneous porous media and showed that the head variance becomes increasingly skewed as σ_Y^2 is increased. Larger σ_Y^2 and longer correlation lengths create more continuous preferential flow paths, making it more challenging to determine monitoring well locations, as we will show below.

During monitoring, we need to tackle the issue of false positives that are caused by background fluctuations (which may be significant for shallow aquifers) and uncertainties associated with both AZMI and injection-zone properties. The former can be quantified via baseline measurements, while the latter should be dealt with probabilistically. In essence, we would like to assess the conditional probability of leakage, given the observed anomaly and the level of uncertainty. Using the same stochastic simulation results as those used to make Fig. 6, we can construct such a probability map to provide a holistic view of detectability. Fig. 7 shows the result for three different detection thresholds (thick, dark lines). To be consistent with pressure transducer readings, we have converted head anomalies to psi. Fig. 7 suggests that all anomalies will be greater than 0.1 psi, starting around $t = 30$ days for base-scenario parameters. Before that time, the monitoring well will probably fail to pick up pressure anomalies because of the instrument detection limit and the time required for the pressure anomaly to build up. To minimize the number of false positives, we may need to relax the detection thresholds. However, the adverse effect

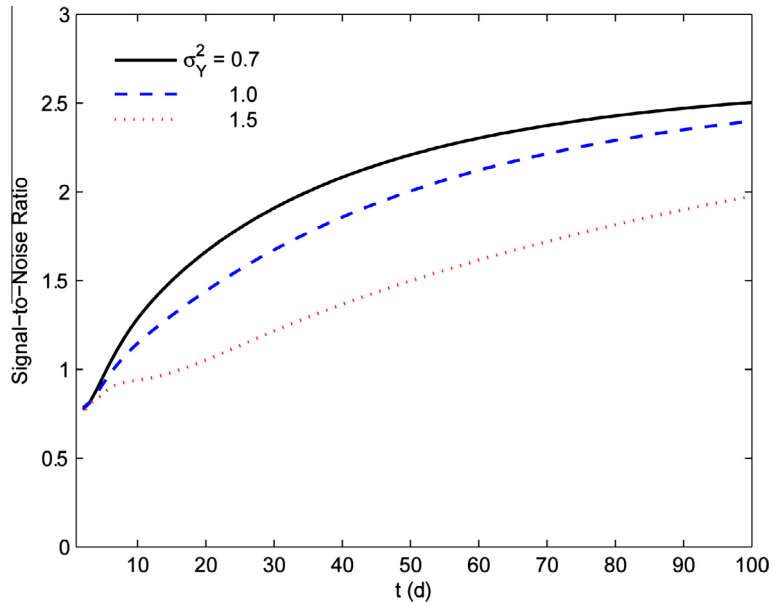

Fig. 9. Effect of variance of Y on predicted SNR.

of doing so is that the likelihood of missing an incident (i.e., false negative) is also increased.

Therefore, we suggest that the anomaly-detection process consist of a series of stages. The first stage is early warning, which is triggered when a monitoring network receives signals indicating possible leakage. The second stage is triage, during which an operator can use a model-based system like the one we have developed to quantify leakage probabilities rapidly for known problematic wells. The operator can continue to analyze signals to rule out the possibility of false positives. Statistical methods that have long been used in groundwater-compliance monitoring can be useful here. For instance, a *prediction limit* test can be conducted using background data so that new head observations are evaluated by determining whether they fall within a prediction interval derived from the background [55,56]:

$$PL = \overline{h}_b + \kappa s \tag{24}$$

where PL is prediction limit; $\overline{h}_b$ and s represent mean and standard deviation of background head data, respectively; and κ is a multiplier derived from Student's t-distribution for a given confidence level. An observed value greater than PL presents statistically significant evidence of a head increase over background. Field visits may also be initiated at this time to expedite the diagnosis process. The third stage is remediation and verification, for which the goal is to eliminate multiple leaky sources. Note that if an operator chooses to shut down injection immediately when anomalies arise, verification can be more challenging unless injection history is already incorporated into the simulation model.

So far, we have demonstrated detection assessment using absolute magnitudes of head anomalies. An alternative and probably more preferable way is to compute the SNR as a measure of detectability. Here, we define SNR as the ratio between mean, μ, and standard deviation, σ, of signal, which is nothing but the reciprocal of coefficient of variation,

$$\mathrm{SNR} = \frac{|\mu|}{\sigma} \tag{25}$$

The range of SNR defined in Eq. (25) is from 0 to ∞. Clearly, the larger the ratio, the greater the detectability.

Fig. 8 plots the SNR using essentially the same information as that presented in Figs. 5–7. The figure suggests that the SNR is low in early times and gradually rises to above 1.0 in about 7 days. It also indicates that the PCM solution with $p = 2$ captures the early trend obtained by Monte Carlo simulation well, although it overestimates its large time behavior. In comparison, PCM with $p = 4$ matches the Monte Carlo results well, except near the very beginning. Result obtained from the third-order expansion is significantly different from the fourth-order solution for $t < 60$ days, although the two look similar on Fig. 5 when the first and second moments are plotted separately. We, therefore, chose to use the fourth-order expansion for the remainder of this work.

Fig. 8 has delivered a conundrum and also a challenge. On the one hand, we would like to be able to detect leakage early; on the other hand, the SNR gets stronger only in late times. We think that such a conundrum can be dealt with only by embracing a strategy that can gradually reduce model uncertainty, such as data assimilation.

3.3.2. Effect of Y variance

The first sensitivity study concerns the effect of σ_Y^2 on SNR. Fig. 9 shows the results obtained by PCM ($p = 4$) for σ_Y^2 equal to 0.7, 1.0, and 1.5, respectively. As the level of variability increases, the predicted head variance at any monitoring location is greater, leading to smaller SNR. Larger variability in hydraulic conductivity translates directly to greater variability in shape and connectivity of flow paths and, therefore, leads to smaller detectability for a given location.

3.3.3. Effect of monitoring well locations

The second sensitivity study considers the impact of monitoring well distance. Fig. 10(a) and (b) show the mean and variance of head anomalies obtained by PCM ($p = 4$) for different monitoring well locations, (100,100), (150,150), and (400,400) m, respectively. The mean Δh increases when the monitoring well is closer to leaking locations, which is expected, and seems to suggest that we should put monitoring wells as close as possible to potential leaky locations to increase the likelihood of detection. However, Fig. 10(b) indicates that head variation and, thus, uncertainty, is greatest in the vicinity of a leaky location at a given point in time and under current boundary conditions. This observation echoes classical results in stochastic hydrogeology [18,57,58]. In Chapter 4 of Zhang [18], for example, the author obtained statistical

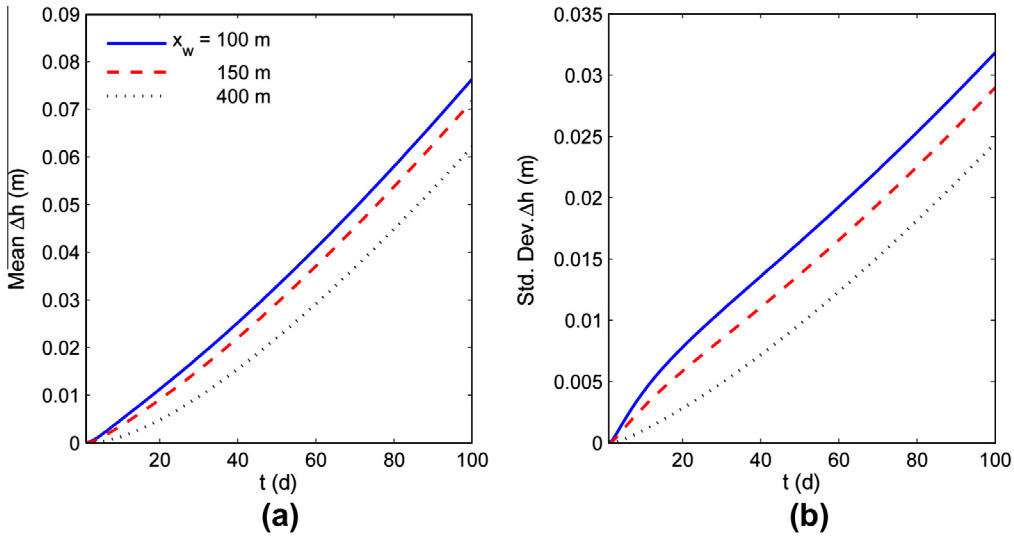

Fig. 10. Effect of monitoring-well location on predicted (a) mean and (b) standard deviation of Δh.

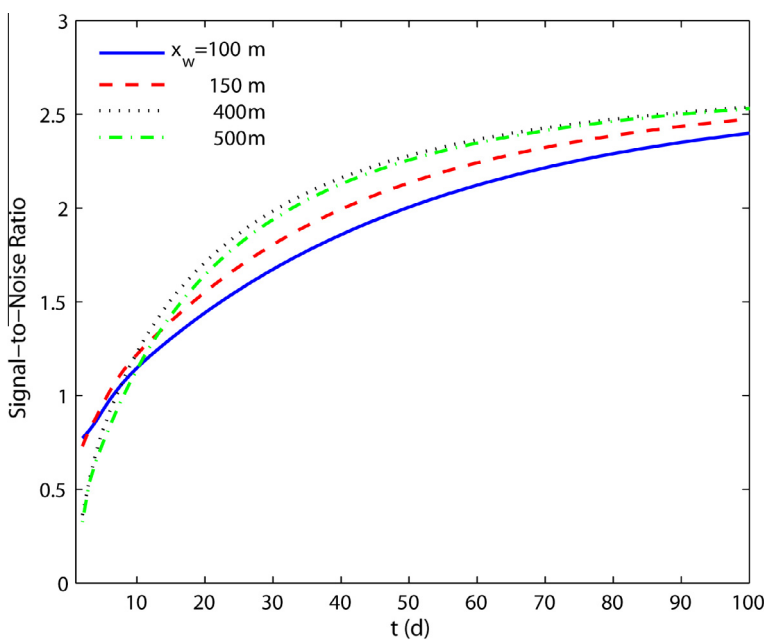

Fig. 11. SNRs calculated for monitoring wells located at (100,100), (150,150), (400,400), and (500,500) m.

moments of hydraulic head corresponding to pumping in the center of a two-dimensional, closed-boundary domain. The profile of head variation evolves with time. At early times, the standard deviation of head increases sharply in the vicinity of the pumping location and then slowly decreases toward the domain boundary. Later, the standard deviation of head flattens, except for a narrow region near the pumping well. The same observations apply to the current problem, except that we have a source instead of a sink term.

Fig. 11 compares SNR corresponding to the three monitoring locations. At early times, the well located at (100,100) m shows the largest SNR, whereas the well at (400,400) m has the smallest SNR because the anomaly has not propagated far enough yet. However, later (about 10 days in our case) the order of SNR is swapped—the well at (400,400) m has the largest SNR, whereas the well at (100,100) m has the lowest. With time, the three SNRs gradually converge and show similar values near 100 days. These observations have interesting implications for the design of monitoring networks, suggesting that an observation point placed closer to leaky locations may not necessarily always yield the most reliable signals when aquifer properties are uncertain. Recall that the correlation length of Y is 400 m. A monitoring well located at (400,400) m captures more flow paths on average than the one at (100,100) m does and, therefore, has a better chance of detecting head anomalies when spatial heterogeneity is involved. To further substantiate this latter point, we obtained the SNR for a monitoring well located at (500,500) m and the result is very close to that of the well at (400,400) m (see Fig. 11). Therefore, for problem settings considered in this study, our results suggest the optimal location for placing a monitoring well is at one correlation length away from the leaky source. Although the recommendation is less of concern for short Y correlations, it is important when Y is correlated over long distances. Without knowing the actual leaky well locations, a site operator may have to rely on information obtained from multiple monitoring wells. Then a plot like Fig. 11 is still

useful because it can be used to infer the proximity of monitoring wells to leaky locations.

4. Summary and conclusions

Leakage from anthropogenic and natural features represents one of the greatest risks to geologic CO_2 sequestration sites. Risk assessment, therefore, plays a critical role during site-feasibility studies and throughout the lifespan of storage projects. Interest is strong in reducing the computational burden associated with UQ, given the complexity of many site-scale simulation models and the number of runs typically required. In this work, we have implemented and applied a stochastic response surface method, PCM, to assessment of leakage detectability at geologic CO_2 sequestration sites. The methodology was demonstrated for different uncertain parameters. In particular, we parameterized the spatially heterogeneous hydraulic conductivity field using KL expansion. The advantages of PCM are (1) it is nonintrusive and can be applied to any existing model without modification of the code and (2) once constructed, PCM allows fast stochastic simulations at virtually no additional computational costs. PCM typically requires only a small fraction of the total number of runs required for a full-scale Monte Carlo simulation. We demonstrated how PCM can be used to construct probability maps for evaluating detectability of head anomalies, both spatially and temporally.

We found that PCM generally provides results that are comparable to Monte Carlo simulation. Although the second-order PCM did not perform well in this study because of the large variance of leakage flux, the fourth-order PCM did give satisfactory results. The total random dimension considered in this study is 9. For problems involving smaller correlation length to domain length ratios, more KL expansion terms need to be included in the PCM. The increase in random dimensions not only adds to the computational cost, but also makes it harder to formulate a full-rank matrix $\mathbf{Z}$ for the PCM system, Eq. (19). To circumvent this latter issue, one can generate a large number of collocation points through random sampling and solve an over-determined system of equations using least squares. Hosder et al. [59] suggested that at least $2(M+1)$ collocation points be used to ensure accuracy. Because all UQ methods will eventually break down in high-dimensional cases [60,61], a more practical approach is probably to perform *a priori* analyses to reduce the random dimensions as much as possible. We sampled two separate random variables to generate leakage fluxes in this work. A possible strategy for reducing random dimension is to treat the leakage flux as a single random variable and sample from its distribution, which is usually log normal [52,53,62]. Other common strategies include sensitivity analyses and metamodeling [63].

Detectability has been quantified using both absolute head anomalies and SNR. Under parameter uncertainty, we found that the likelihood of detecting anomalies depends on (1) level of uncertainty and (2) location of the monitoring well. A well located close to leaky locations may not always give the most robust indication of leakage when the level of uncertainty is high. Therefore, our research highlights the need for adaptive model-uncertainty reduction, monitoring-network design, and source-location identification. Because an operator can only install a limited number of deep AZMI monitoring wells, a risk assessment plan must also evaluate leakage detectability at shallower depths and in the vadose zone. In addition, the pressure-based technique needs to be supplemented by monitoring other leakage signals to increase the chance of detection.

Finally, the extension of the current methodology to three-dimensional and/or multiphase flow is straightforward. We have assumed statistically stationary, multi Gaussian log-conductivity random fields and used KL expansion as a means for parameterization and dimension reduction. With other parameterization techniques, the UQ methodology presented in this work may be extended to nonstationary or non-Gaussian random fields to accommodate multiple permeability zones [64] or multiple facies [65–67].

Acknowledgments

A. Sun, M. Zeidouni, and J.-P. Nicot are partly funded by EPA STAR Grant R834384. Publication authorized by the Director, Bureau of Economic Geology. The authors are grateful to the four anonymous reviewers and associated editor for their constructive comments.

References

[1] European Union. Directive 2009/31/EC on the geological storage of carbon dioxide and amending Council Directive 85/337/EEC, European Parliament and Council Directives 2000/60/EC, 2001/80/EC, 2004/35/. European Parliament and Council, 2009.
[2] US EPA. 40 CFR parts 144 and 146 federal requirements under the underground injection (UIC) program for carbon dioxide (CO_2) geologic sequestration (GS) wells; Proposed Rule, 2008. p. 43492–43541.
[3] NETL. Best practices for: monitoring, verification, and accounting of CO_2 stored in deep geologic formations. Available at <http://www.netl.doe.gov/technologies/carbon_seq/refshelf/MVA_Document.pdf>. Accessed on January 3, 2012. National Energy Technology Laboratory, Pittsburgh, PA, 2009.
[4] NETL. Best practices for: risk analysis and simulation for geologic storage of CO_2. Available at <http://www.netl.doe.gov/technologies/carbon_seq/refshelf/BPM-SiteScreening.pdf>. Accessed on January 3, National Energy Technology Laboratory, Pittsburgh, PA, 2011.
[5] Wickens TD. Elementary signal detection theory. Oxford, New York: Oxford University Press; 2002.
[6] Taguchi G. Introduction to quality engineering: designing quality into products and processes. White Plains, NY: Kraus International Publications; 1986.
[7] Taguchi G, Phadke MS. Quality engineering through design optimization. In: IEEE globecom 1984 conference, Institute of Electrical and Electronics Engineers: Atlanta, GA; 1984. p. 1106–1113.
[8] Taguchi G, Wu Y. Introduction to off-line quality control, Central Japan Quality Control Association, Nagaya, Japan, 1985.
[9] Sun AY, Nicot JP. Inversion of pressure anomaly data for detecting leakage at geologic carbon sequestration sites. Adv Water Resour 2012:20–9.
[10] Cortis A, Oldenburg CM, Benson SM. The role of optimality in characterizing CO_2 seepage from geologic carbon sequestration sites. Int J Greenh Gas Con 2008;2:640–52.
[11] Zeidouni M, Pooladi-Darvish M. Leakage characterization through above-zone pressure monitoring: 1 –Inversion approach. J Petrol Sci Eng 2012;98–99:95–106.
[12] Oladyshkin S, Class H, Helmig R, Nowak W. An integrative approach to robust design and probabilistic risk assessment for CO(2) storage in geological formations. Comput Geosci 2011;15:565–77.
[13] Class H, Ebigbo A, Helmig R, Dahle H, Nordbotten J, Celia M, et al. A benchmark study on problems related to CO_2 storage in geologic formations. Comput Geosci 2009;13:409–34.
[14] Ebigbo A, Class H, Helmig R. CO_2 leakage through an abandoned well: problem-oriented benchmarks. Comput Geosci 2007;11:103–15.
[15] Humez P, Audigane P, Lions J, Chiaberge C, Bellenfant G. Modeling of CO_2 leakage up through an abandoned well from deep saline aquifer to shallow fresh groundwaters. Transport Porous Med 2011;90:153–81.
[16] Pruess K. Integrated modeling of CO_2 storage and leakage scenarios including transitions between super- and subcritical conditions, and phase change between liquid and gaseous CO_2. Greenhouse Gases: Sci Technol 2011;1:237–47.
[17] Nordbotten JM, Celia MA, Bachu S. Injection and storage of CO2 in deep saline aquifers: analytical solution for CO_2 plume evolution during injection. Transport Porous Med 2005;58:339–60.
[18] Zhang D. Stochastic methods for flow in porous media: coping with uncertainties. Academic, San Diego, Calif. London, 2002.
[19] Ganapathysubramanian B, Zabaras N. Sparse grid collocation schemes for stochastic natural convection problems. J Comput Phys 2007;225:652–85.
[20] Xiu D. Numerical methods for stochastic computations: a spectral method approach. Princeton, NJ: Princeton University Press; 2010.
[21] Xiu D, Karniadakis GE. The Wiener–Askey polynomial chaos for stochastic differential equations. SIAM J Sci Comput 2002;24:619–44.
[22] Xiu D, Karniadakis GE. Modeling uncertainty in flow simulations via generalized polynomial chaos. J Comput Phys 2003;187:137–67.
[23] Eldred M. Recent advances in non-intrusive polynomial chaos and stochastic collocation methods for uncertainty analysis and design. In: 50th AIAA/ASME/ASCE/AHS/ASC structures structural dynamics and materials conference, Palm Springs, California, 2009.

[24] Ghanem R, Spanos PD. Stochastic finite elements: a spectral approach. New York: Springer-Verlag; 1991.
[25] Wiener N. The homogeneous chaos. Am J Math 1938;60:897–936.
[26] Nobile F, Tempone R, Webster CG. A sparse grid stochastic collocation method for partial differential equations with random input data. SIAM J Numer Anal 2008;46:2309–45.
[27] Babuska I, Nobile F, Tempone R. A stochastic collocation method for elliptic partial differential equations with random input data. SIAM J Numer Anal 2007;45:1005–34.
[28] Tatang MA, Pan W, Prinn RG, McRae GJ. An efficient method for parametric uncertainty analysis of numerical geophysical models. J Geophys Res 1997;102:21925–32.
[29] Li H, Zhang D. Probabilistic collocation method for flow in porous media: comparisons with other stochastic methods. Water Resour Res 2007;43:W09409.
[30] Lin G, Tartakovsky AM. An efficient, high-order probabilistic collocation method on sparse grids for three-dimensional flow and solute transport in randomly heterogeneous porous media. Adv Water Resour 2009;32:712–22.
[31] Shi L, Yang J, Zhang D, Li H. Probabilistic collocation method for unconfined flow in heterogeneous media. J Hydrol 2009;365:4–10.
[32] Ma X, Zabaras N. A stochastic mixed finite element heterogeneous multiscale method for flow in porous media. J Comput Phys 2011;230:4696–722.
[33] Zhang D, Shi L, Chang H, Yang J. A comparative study of numerical approaches to risk assessment of contaminant transport. Stoch Environ Res Risk Assess 2010;24:971–84.
[34] Walter L, Binning P, Oladyshkin S, Flemisch B, Class H. Brine migration resulting from CO2 injection into saline aquifers-an approach to risk estimation including various levels of uncertainty. In: International Journal of Greenhouse Gas Control, vol. 9, Elsevier, 2012. p. 495–506.
[35] Wang X-S, Jiang X-W, Wan L, Ge S, Li H. A new analytical solution of topography-driven flow in a drainage basin with depth-dependent anisotropy of permeability. Water Resour Res 2011;47:W09603.
[36] Szymkiewicz A, Helmig R, Kuhnke H. Two-phase flow in heterogeneous porous media with non-wetting phase trapping. Transport Porous Med 2011;86:27–47.
[37] Doughty C, Pruess K. Modeling supercritical carbon dioxide injection in heterogeneous porous media. Vadose Zone J 2004;3:837–47.
[38] Liu B, Zhang Y. CO_2 modeling in a deep saline aquifer: a predictive uncertainty analysis using design of experiment. Environ Sci Technol 2011;45:3504–10.
[39] Jahangiri HR, Zhang D. Effect of spatial heterogeneity on plume distribution and dilution during CO_2 sequestration. Int J Greenhouse Gas Control 2011;5:281–93.
[40] Li S, Zhang Y, Zhang X. A study of conceptual model uncertainty in large-scale CO_2 storage simulation. Water Resour Res 2011;47:W05534.
[41] DOE/NETL. Best management practices for risk analysis and simulation for geologic storage of CO_2. National Energy Technology Laboratory, Pittsburgh, PA, 2011.
[42] Birkholzer JT, Nicot JP, Oldenburg CM, Zhou Q, Kraemer S, Bandilla K. Brine flow up a well caused by pressure perturbation from geologic carbon sequestration: Static and dynamic evaluations. Int J Greenhouse Gas Control 2011;5:850–61.
[43] Nordbotten JM, Celia MA, Bachu S. Analytical solutions for leakage rates through abandoned wells. Water Resour Res 2004;40.
[44] Celia MA, Nordbotten JM, Court B, Dobossy M, Bachu S. Field-scale application of a semi-analytical model for estimation of CO_2 and brine leakage along old wells. Int J Greenhouse Gas Control 2011;5:257–69.
[45] Press WH. Numerical recipes in C++: the art of scientific computing. 2nd ed. UK, New York: Cambridge University Press, Cambridge; 2002.
[46] Zhang DX, Lu ZM. An efficient, high-order perturbation approach for flow in random porous media via Karhunen–Loeve and polynomial expansions. J Comput Phys 2004;194:773–94.
[47] Oldenburg CM, Bryant SL, Nicot JP. Certification framework based on effective trapping for geologic carbon sequestration. Int J Greenhouse Gas Control 2009;3:444–57.
[48] Li H, Chang H, Zhang D. Stochastic collocation methods for efficient and accurate quantification of uncertainty in multiphase reservoir simulations. In: SPE reservoir simulation symposium, 2–4 February 2009, Society of Petroleum Engineers the Woodlands, Texas, 2009.
[49] Zeidouni M, Pooladi-Darvish M, Keith DW. Analytical models for determining pressure change in an overlying aquifer due to leakage. Energy Proc 2011:3833–40.
[50] Hovorka SD, Doughty C, Benson SM, Pruess K, Knox PR. The impact of geologic heterogeneity on CO_2 storage in brine formations: a case study from the Texas Gulf Coast. In: Baines SJ, Worden RH, editors. Geological storage of carbon dioxide geological society, London, special publications, The Geological Society of London, London, 2004. p. 147–163.
[51] Rubin Y. Applied stochastic hydrogeology. Oxford, New York: Oxford University Press; 2003.
[52] Sun AY, Zhang D. A solute flux approach to transport through bounded, unsaturated heterogeneous porous media. Vadose Zone J 2004;3:513–26.
[53] Zhang D, Andricevic R, Sun AY, Hu X, He G. Solute flux approach to transport through spatially nonstationary flow in porous media. Water Resour Res 2000;36:2107–20.
[54] Nowak W, Schwede RL, Cirpka OA, Neuweiler I. Probability density functions of hydraulic head and velocity in three-dimensional heterogeneous porous media. Water Resour Res 2008;44:W08452.
[55] Gibbons RD, Bhaumik D, Aryal S. Statistical methods for groundwater monitoring. 2nd ed. Hoboken, NJ: Wiley; 2009.
[56] EPA. Statistical analysis of groundwater monitoring data at RCRA facilities: unified guidance. EPA 530/R-09-007. EPA Office of Resource Conservation and Recovery 2009.
[57] Tartakovsky DM, Neuman SP. Transient flow in bounded randomly heterogeneous domains: 1. Exact conditional moment equations and recursive approximations. Water Resour Res 1998;34:1–12.
[58] Zhang D, Sun AY. Stochastic analysis of transient saturated flow through heterogeneous fractured porous media: a double-permeability approach. Water Resour Res 2000;36:865–74.
[59] Hosder S, Perez R, Walters RW. A non-intrusive polynomial chaos method for uncertainty propagation in CFD simulations. In: 44th AIAA Aerospace Sciences Meeting and Exhibit, Reno, Nevada, 2006.
[60] Sun AY, Morris AP, Mohanty S. Sequential updating of multimodal hydrogeologic parameter fields using localization and clustering techniques. Water Resour Res 2009;45.
[61] Bellman R. Dynamic programming. Dover ed. Mineola, N.Y.: Dover Publications; 2003.
[62] Mitchell RL. Permanence of the log-normal distribution. J Opt Soc Am 1968;58:1267–72.
[63] Tolson BA, Shoemaker CA. Efficient prediction uncertainty approximation in the calibration of environmental simulation models. Water Resour Res 2008;44:W04411. http://dx.doi.org/10.1029/2007WR005869.
[64] Lu ZM, Zhang DX. Stochastic simulations for flow in nonstationary randomly heterogeneous porous media using a KL-based moment-equation approach. Multiscale Model Simul 2007;6:228–45.
[65] Caers J, Zhang T. Multiple-point geostatistics: a quantitative vehicle for integrating geologic analogs into multiple reservoir models. AAPG Mem 2004;20:383–94.
[66] Sarma P, Durlofsky L, Aziz K. Kernel principal component analysis for efficient, differentiable parameterization of multipoint geostatistics. Math Geosci 2008;40:3–32.
[67] Scheidt C, Caers J, Chen Y, Durlofsky L. A multi-resolution workflow to generate high-resolution models constrained to dynamic data. Comput Geosci 2011;15:545–63.