

Probabilistic assessment of scalar transport under hydrodynamically unstable flows in heterogeneous porous media

Alessandra Bonazzi^a, Xiaoshu Zeng^a, Roger Ghanem^a, Birendra Jha^b, Felipe P.J. de Barros^{a,*}

^a Sonny Astani Department of Civil and Environmental Engineering, University of Southern California, 3620 S. Vermont Avenue, Los Angeles, 90089, CA, USA

^b Mork Family Department of Chemical Engineering and Materials Science, 925 Bloom Walk, Los Angeles, 90089, CA, USA

ARTICLE INFO

Keywords:

Uncertainty quantification
Heterogeneity
Viscous fingering
Stochastic hydrogeology
Probabilistic risk analysis
Hydrodynamic instability

ABSTRACT

Quantitative predictions of scalar transport in natural porous media is a nontrivial task given the presence of multi-scale spatial heterogeneity in the permeability field. Due to data scarcity, the structural map of the permeability field is subject to uncertainty and therefore, model predictions are uncertain. For such reasons, probabilistic models of flow and transport in natural porous media are required in risk assessment and to provide reliable decision making under uncertainty. Further complexities arise when the viscosity of the injected solute differs from that of the ambient fluid. Under the presence of viscosity contrast, hydrodynamic instabilities give rise to viscous fingering, which induces additional disorder in both velocity and solute concentration fields. This work examines the combined role of viscous fingering and permeability heterogeneity in the probabilistic description of transport predictions. In particular, we focus on metrics that are important for risk analysis, such as the solute plume's early arrival times and the maximum concentration observed at a given location. We propose to use the Projection Pursuit Adaptation (PPA) method in the Polynomial Chaos Expansion (PCE) framework to quantify uncertainty in transport model predictions. The PPA method is a data-driven approach that optimally represents a given quantity of interest in a low-dimensional manifold. Unlike other dimension reduction techniques in uncertainty quantification, the PPA method utilizes non-linear information of the quantity of interest to identify the low-dimensional manifold, thereby increasing the likelihood of finding a more accurate lower-dimensional space. Moreover, the PPA model converges to the physical solution in a mean squared sense with respect to the polynomial order, enabling the construction of a converged model even with limited available data. Then, the PPA results are compared to Monte Carlo simulations using the same amount of data. This comparison illustrates that while Monte Carlo simulations are able to capture low-order statistics, they struggle to represent more detailed probability density functions. Our results show how the combined effect of permeability heterogeneity and viscosity contrast can enhance the mobility of the solute plume.

1. Introduction

Obtaining deterministic predictions of the spatiotemporal dynamics of a solute body in the subsurface environment is an elusive goal given limited site characterization data. Due to the high costs associated with data acquisition, in-field measurements are scarce, and thus a detailed characterization of the spatial heterogeneity of hydrogeological properties (such as the permeability) is impractical. Because of this lack of information on all relevant scales, flow and transport models are subjected to uncertainty and the stochastic paradigm is invoked. Under the stochastic approach, the permeability field is conceptualized as a random space function (RSF) (Rubin, 2003; Christakos, 2005) and environmental performance metrics relevant for evaluating risks from contaminants migrating in groundwater, such as solute travel times and

concentrations, are characterized in terms of their probability density function (PDF).

Strategies for propagating the uncertainty in the permeability field to solute travel time and concentration predictions have been discussed in the literature (for an extensive review, see Fiori et al. 2015, Linde et al. 2017 and references therein). Perturbation theory has been applied to compute the statistical moments of solute travel time (Rubin and Dagan, 1992; Cvetkovic et al., 1992; Sanchez-Vila and Guadagnini, 2005) and resident concentration (Kapoor and Kitanidis, 1998; Andrićević, 1998; Fiori and Dagan, 2000; Tonina and Bellin, 2008; de Barros et al., 2022). The (semi)-analytical solutions derived in these studies are typically limited to low levels of heterogeneity in the log-permeability field, e.g., log-permeability variance $\sigma_Y^2 \lesssim 1$. To estimate

* Corresponding author.

E-mail address: fbarros@usc.edu (F.P.J. de Barros).

<https://doi.org/10.1016/j.advwatres.2024.104706>

Received 8 September 2023; Received in revised form 12 April 2024; Accepted 17 April 2024

Available online 19 April 2024

0309-1708/© 2024 Elsevier Ltd. All rights reserved.

uncertainty in solute transport in aquifers displaying high heterogeneity $\sigma_Y^2 > 1$, the governing equations need to be solved numerically and the Monte Carlo framework is often employed (Moslehi and de Barros, 2017; Libera et al., 2019). Recently, an open-source Python package for simulating uncertainty (through Monte Carlo) in hydrogeological systems has been proposed (Morvillo et al., 2022). Semi-analytical methods, such as the Multi-Indicator Model–Self Consistent Approach, enable to obtain an estimate of solute arrival times distributions in highly heterogeneous formations (Cvetkovic et al., 2016; de Barros et al., 2016; Fiori et al., 2017). Other studies, based on perturbation expansions or numerical Monte Carlo approaches, focused on the estimation of the one-point concentration PDF as a function of the RSF parameters of the log-permeability field (Dentz, 2012; Cirpka et al., 2011; de Barros and Fiori, 2014; Boso and Tartakovsky, 2016). In the absence of diffusion and local-scale dispersion, Dagan (1982) showed that the concentration PDF is binary, that is, either zero or equal to the initial concentration at the solute source location. Subsequent studies showed that the concentration PDF becomes unimodal if the sampling volume is increased and the effects of local-scale dispersion are taken into consideration (Bellin et al., 1994; Caroni and Fiorotto, 2005; Schwede et al., 2008; de Barros and Fiori, 2014). For low levels of heterogeneity of the permeability field (i.e., $\sigma_Y^2 \lesssim 1$), the concentration PDF can be approximated by the beta distribution model as originally suggested in Fiori (2001) and verified by others (Fiorotto and Caroni, 2002; Bellin and Tonina, 2007; Bonazzi et al., 2023). A joint velocity-concentration PDF method for computing the concentration statistics for higher levels of heterogeneity is also available in Meyer et al. (2010). Within the broader context of multi-physics, a multi-modular uncertainty quantification open-source Python package capable of handling multi-scale problems is also available in the literature (Olivier et al., 2020).

The above-mentioned studies (amongst many others) have provided insight into the propagation of uncertainty, with respect to scales of fluctuation of the random permeability, on solute travel times and concentration. One key assumption in all these studies is that both the dissolved solute and the ambient fluid (e.g. groundwater) have the same viscosity. Some exceptions exist, such as the works by Welty and Gelhar (1991) and Talon et al. (2004), where the authors employ stochastic methods to estimate the asymptotic longitudinal macrodispersivity. Bonazzi (2023) used Monte Carlo simulations to show how contrast in viscosity controlled the spatial maps of the concentration (ensemble) variance at different times. In many applications, groundwater contaminants (i.e., nonhalogenated semivolatile compounds and jet fuel) are more viscous than water. Flowers and Hunt (2007) provide a series of examples in environmental engineering and hydrology where viscosity contrast between the ambient groundwater and a contaminant can be found. However, incorporating the effects of viscosity contrast in numerical simulation models can be challenging due to high computational costs associated with the non-linearity of the governing equations. Viscosity contrast between the ambient fluid and the solute can lead to the hydrodynamic instability known as the Saffman–Taylor instability (Saffman, 1986), with formation of viscous fingers at the unstable interface. Due to the viscosity dependence on solute concentration, physical properties of the fluids vary in space and time, thus affecting the flow field (Jha et al., 2011a). In the presence of permeability heterogeneity, viscous fingering can enhance the mobility of a solute plume and its mixing rates (Bonazzi et al., 2021). Despite the extensive literature on the effects of viscous fingering on mixing in porous media (Tran and Jha, 2020; Van der Meer, 1993; Christie, 1989; Tchelepi and Orr Jr., 1994; Tan and Homsy, 1988; De Wit et al., 2005; Nicolaides et al., 2015; Bonazzi et al., 2021), additional efforts are needed to investigate how viscosity contrast impacts the uncertainty estimates of transport models when the flow field is spatially random. Due to the non-linearity of the flow and transport processes when modeling mixing of fluids with distinct viscosities, traditional methods for

uncertainty estimation, such as Monte Carlo, become unfeasible due to the high computational costs needed to achieve statistical convergence.

For high-dimensional uncertainty quantification, the Monte Carlo-based method, despite its advantageous feature of dimension-independent convergence, exhibits slow convergence overall, with the rate of convergence dependent on the complexity of the solution being sought (Ballio and Guadagnini, 2004; Leube et al., 2013; Moslehi et al., 2015). Consequently, numerous evaluations of the physical model are required, imposing a significant computational burden. In response, various techniques have emerged in engineering and science to alleviate this burden. These include Polynomial Chaos Expansion (PCE) (Ghanem and Spanos, 2012; Ghanem, 1999), Generalized Polynomial Chaos Expansion (gPCE) (Xiu and Karniadakis, 2002, 2003), Gaussian processes (MacKay et al., 1998; Seeger, 2004; Bilonis and Zabarar, 2012), Bayesian networks (Darwiche, 2009; Ghahramani, 2006; Zeng et al., 2023b), Probability Density Evolution Methods (PDEM) (Li and Chen, 2009; Zeng et al., 2017; Chen et al., 2017), Manifold Methods (Soize and Ghanem, 2021; Zhang et al., 2021; Giovanis and Shields, 2020), Wiener Path Integral Technique (Kougioumtzoglou and Spanos, 2012; Psaros et al., 2019), and others. In the context of computationally expensive porous media flow models (i.e. non-linear formulations), PCE-based methods were employed to estimate uncertainty in flow through the unsaturated zone (Li et al., 2009) and CO₂ storage (Oladyshkin et al., 2011, 2013). Methods such as Gaussian process modeling have been employed to estimate uncertainty in a diverse range of porous media flows application (Crevillen-Garcia et al., 2017; Tian et al., 2017), however handling high dimension problems can be challenging. PCE represents the Quantity of Interest (QoI) using a set of multivariate Hermite polynomial bases, where the approximation converges to the physical solution in a mean-square sense with respect to the polynomial orders. Despite its rigorous mathematical foundation, good convergence rate (Ghanem and Spanos, 2012), and high accuracy, PCE faces the challenge of the curse of dimensionality, for example, in applications involving a heterogeneous flow field within high-dimensional input spaces. Addressing the high dimensionality challenge, Least Angle Regression (Blatman and Sudret, 2010, 2011) and Compressive Sensing (Doostan and Owhadi, 2011; Sargsyan et al., 2014; Tsilifis et al., 2019) methods reduce computation by approximating the QoI on a selection of dominant polynomial bases. Active Subspace (Constantine et al., 2014a; Constantine, 2015) represents the QoI along active directions selected based on the gradient of the physical systems. Basis adaptation, within the PCE framework, exploits the insight that while the physical system may be complex, the sought solution typically corresponds to a simpler physical property. This method involves rotating the input parameters, allowing the QoI to be represented in a low-dimensional subspace formulated by the rotated variables. Applications and extended research (Thimmisetty et al., 2017; Tsilifis and Ghanem, 2017; Zeng et al., 2021; Ghauch et al., 2019; Zeng et al., 2023a) establish basis adaptation as a promising candidate for dimension reduction in uncertainty quantification. The efficiency of the basis adaptation method is enhanced when sparse quadrature rules are employed in constructing the PCE model. In contrast to Principal Component Analysis (PCA), which identifies components based on their importance ranking, basis adaptation incorporates QoI information into the dimension reduction process. Consequently, it is more tailored to the sought solution, enabling the discovery of even low-dimensional subspaces.

In a more recent work (Zeng and Ghanem, 2023), the authors proposed the *Projection Pursuit Adaptation* (PPA) method, which has the basis adaptation ideas embedded for dimension reductions. PPA is a data-driven method that learns optimal projections of the input parameters, as well as an optimal PCE that establishes the mapping from the projected parameters to the quantities of interest. This method is promising due to its advantages over other uncertainty quantification techniques. It reduces dimensions and utilizes a priori-specified data for optimal PCE representation. The achievement of optimality

is attained through the determination of least-squares solutions for the quantities of interest. This implies that the PPA model is constructed to minimize the squared residual between the data and the model. The PPA approach is also able to capture nonlinearities, thus identifying a lower-dimensional space compared to techniques like Gaussian adaptation (Tipereddy and Ghanem, 2014; Zeng et al., 2021) and active subspace (Constantine et al., 2014b), as has been shown in Zeng and Ghanem (2023). Finally, in contrast to classical PCE methods where sparse quadrature rules are typically applied to reduce the number of samples, the proposed PPA method utilizes independent samples as training data, offering increased flexibility. Furthermore, it possesses the capability to address multiple-output surrogate modeling. This means that the same training dataset can be repurposed to learn optimal PCEs for different quantities of interest.

The purpose of this work is twofold. The first one is to examine how fluid properties, such as viscosity, can affect the uncertainty associated with transport predictions that are relevant to probabilistic risk analysis in contaminated groundwater systems. Given the importance of viscosity contrast on solute concentrations (Jha et al., 2011b,a; Bonazzi et al., 2021) and arrival times (Nicolaides et al., 2015; Bonazzi et al., 2021), we quantify the uncertainty of these quantities that originate from the random fluctuations of the permeability field. The second goal is to propose an approach to compute the uncertainty associated with transport predictions in heterogeneous subsurface systems. Due to high computational costs in Monte Carlo simulations, we demonstrate how the PPA in the PCE framework estimates PDFs for these quantities.

The structure of the paper is as follows: in Section 2 we describe the mathematical formulation of the flow and transport model adopted in this study; in Section 3 we introduce the stochastic model underlying the structure of the permeability field (Sections 3.1 and 3.2), and the PPA method (Section 3.3); Section 4 describes the numerical implementation of the flow and transport model and the parameters adopted; Section 5 presents the computational results obtained and Section 6 provides a summary of this study.

2. Mathematical formulation of flow and transport model

This section introduces the flow and scalar transport model adopted in this study. In the following, we provide all governing equations in the dimensionless form. The dimensionless groups are reported in Appendix. We start by considering a two-dimensional (2D) porous medium described by the Cartesian coordinate system $\mathbf{x} = (x_1, x_2)$. The flow domain D is rectangular with dimensions $L_1 \times L_2 = \{(x_1, x_2) | x_1 \in [0, L_1] \text{ and } x_2 \in [0, L_2]\}$. The porous medium is characterized by a locally isotropic and spatially heterogeneous permeability field, $k(\mathbf{x})$. Because the variations in porosity are relatively small when compared to the variability in k , we adopt the common assumption of uniform porosity ϕ .

Assuming incompressible flow in the absence of sink/source terms, the governing dimensionless equation for the flow field is:

$$\nabla \cdot \mathbf{q}(\mathbf{x}, t) = 0, \quad (1)$$

where $\mathbf{q}(\mathbf{x}, t)$ is the specific discharge which can be computed through Darcy's law:

$$\mathbf{q}(\mathbf{x}, t) = -\frac{k(\mathbf{x})}{\mu(c(\mathbf{x}, t))} \nabla p(\mathbf{x}, t), \quad (2)$$

in which $p(\mathbf{x}, t)$ is the pressure field and $\mu(c(\mathbf{x}, t))$ is the viscosity as a function of the solute concentration, $c(\mathbf{x}, t)$, in the ambient fluid. The following model is adopted to relate the viscosity of the mixture to the solute concentration:

$$\mu(c(\mathbf{x}, t)) = e^{-Rc(\mathbf{x}, t)}, \quad (3)$$

where $R = \ln(\mu_0/\mu_s)$ is the log-viscosity ratio between the ambient fluid viscosity (μ_0) and the viscosity of the solute (μ_s). We employ permeameter-like boundary conditions for the flow Eq. (1),

$$p(0, x_2) = p_1; \quad p(L_1, x_2) = p_1; \quad \frac{\partial p(\mathbf{x})}{\partial x_2} \Big|_{x_2=0} = 0; \quad \frac{\partial p(\mathbf{x})}{\partial x_2} \Big|_{x_2=L_2} = 0 \quad (4)$$

Under these boundary conditions, the spatially heterogeneous flow field is driven by a uniform-in-the-mean pressure gradient. The spatially heterogeneous velocity field is given by $\mathbf{v}(\mathbf{x}, t) = \mathbf{q}(\mathbf{x}, t)/\phi$.

Transport of a non-reactive, dissolved tracer is assumed to be controlled by two mechanisms: (i) the advective movement due to random fluctuations of the (Darcy) velocity $\mathbf{v}$ and (ii) local-scale dispersion. Therefore, the local concentration c satisfies the Eulerian transport equation:

$$\frac{\partial c(\mathbf{x}, t)}{\partial t} + \mathbf{v}(\mathbf{x}, t) \cdot \nabla c(\mathbf{x}, t) = \frac{1}{\text{Pe}} \nabla^2 c(\mathbf{x}, t), \quad (5)$$

where Pe represents the Péclet number, defined as $\text{Pe} \equiv U\lambda/D$, in which λ represents the correlation length scale of the log-permeability field, U is the longitudinal mean velocity and D is the local-scale dispersion coefficient assumed to be constant (Scheidegger, 1961; Bear, 1988). The boundary conditions for the transport problem are as follows

$$\begin{aligned} \frac{\partial c(\mathbf{x})}{\partial x_1} \Big|_{x_1=0} &= 0; \\ [\mathbf{v}(\mathbf{x}, t)c(\mathbf{x}, t)]|_{x_1=L_1} &= J_{out}(L_1, x_2, t); \\ \frac{\partial c(\mathbf{x}, t)}{\partial x_2} \Big|_{x_2=0} &= 0; \\ \frac{\partial c(\mathbf{x}, t)}{\partial x_2} \Big|_{x_2=L_2} &= 0. \end{aligned} \quad (6)$$

Here, J_{out} is the solute mass flux exiting the domain D , i.e. a natural outflow boundary condition where the solute mixture is allowed to exit the domain through advection only (in the direction of the average flow velocity). The solute is instantaneously injected over an area $\mathcal{A}_s$ and the initial dimensionless concentration is set to be equal to unit value. The initial condition is expressed as

$$c(\mathbf{x}, 0) = \begin{cases} 1, & \text{if } \mathbf{x} \in \mathcal{A}_s; \\ 0, & \text{otherwise,} \end{cases} \quad (7)$$

where $\mathcal{A}_s = \ell_1 \times \ell_2$ with ℓ_1 and ℓ_2 denoting the longitudinal and transverse dimensions of the source zone respectively (see Fig. 1). In order to allow for the development of the hydrodynamic instability caused by the viscosity contrast between the solute and the ambient fluid, we apply a small perturbation on the edges of the source zone.

Due to the random nature of k , both the velocity $\mathbf{v}$ and solute concentrations, c are also random functions. As a consequence, model predictions associated to the flow field and solute concentration distribution (in both space and time) are quantified in terms of their ensemble statistics. The statistics of $\mathbf{v}$ have been related to the statistics of k , both analytically using first-order approximations (Rubin, 1990; Butera and Tanda, 1999), and numerically using Monte Carlo simulations (Bellin et al., 1992). Similarly, the relationship between the statistics of c and the statistics of k are also analytically reported through first-order approximations (Fiori and Dagan, 2000; Tonina and Bellin, 2008; de Barros and Fiori, 2014).

In this work, we will focus on characterizing the uncertainty of two quantities of interest. The first one is the early arrival times of a solute plume originating from a source zone, $\mathcal{A}_s$, to a control plane located at $x_1 = L_{cp}$ (see Fig. 1). The early arrival time, denoted here by t_e , is defined as the time when the dimensionless solute concentration (averaged over the control plane) is equal to 0.01. The spatially averaged (over the control plane) concentration is

$$\bar{c}(t) = \frac{1}{L_2} \int_0^{L_2} c(L_{cp}, x_2) dx_2, \quad (8)$$

where L_2 is the transverse dimension of the flow domain D (i.e., the control plane has dimensions of L_2). The second quantity of interest is the maximum concentration observed (in time) at the control plane:

$$c_{max} = \max_{t \in \mathcal{T}} \bar{c}(t), \quad (9)$$

where $\mathcal{T}$ is the total residence time of the solute plume crossing the control plane at $x_1 = L_{cp}$. The maximum concentration is a quantity of particular interest for risk analysis (de Barros and Fiori, 2021).

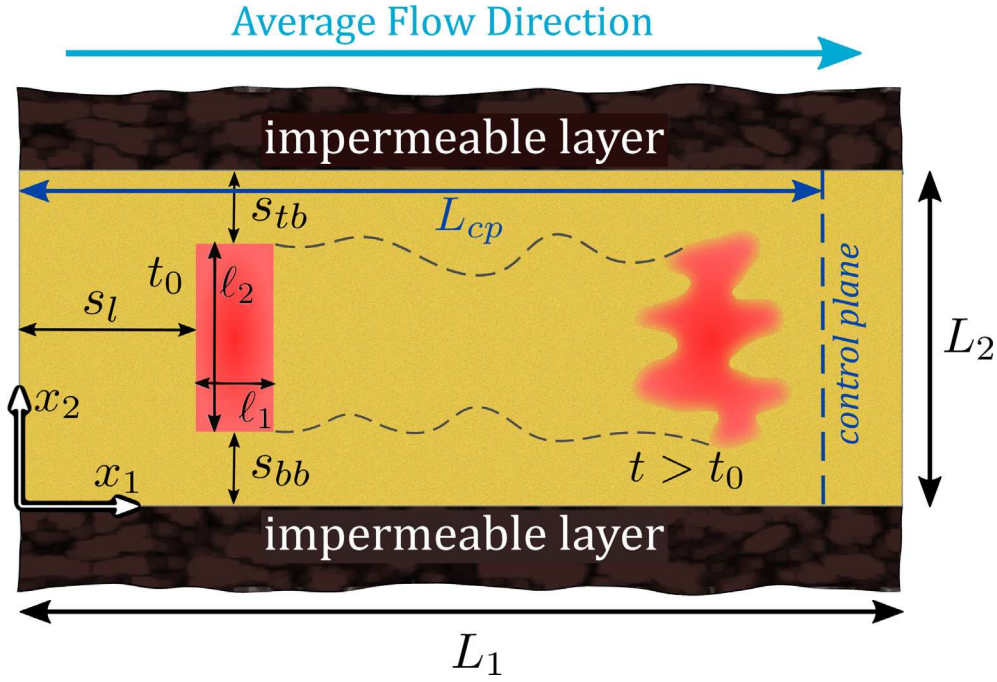


Fig. 1. Schematic representation of the conceptual model and the computational domain. A solute is instantaneously injected within an areal source zone. The porous medium is two-dimensional and its permeability field is spatially heterogeneous.

3. Stochastic model

3.1. Multi-Gaussian log-permeability field

The structure of the log-permeability field, denoted here by $Y(\mathbf{x}, \omega)$ $\equiv \ln[k(\mathbf{x}, \omega)]$, is modeled through a RSF (Kitanidis, 1997; Rubin, 2003) with $\mathbf{x} \in D$ and $\omega \in \Omega$ (the probability space). For the purpose of the current study, $Y(\mathbf{x}, \omega)$ is represented as a multi-Gaussian spatial random field and assumed to be statistically stationary such that it can be fully characterized by its mean $\langle Y(\mathbf{x}) \rangle$ and the correlation function of its fluctuations $Y'(\mathbf{x}, \omega) = Y(\mathbf{x}, \omega) - \langle Y(\mathbf{x}, \omega) \rangle$, which is provided by $C_Y(\mathbf{x}, \mathbf{x}') = \langle Y'(\mathbf{x}, \omega)Y'(\mathbf{x}', \omega) \rangle$. Here, the angled brackets $\langle \cdot \rangle$ denote ensemble expectation. The variance of the Y -field is defined as $C_Y(\mathbf{0}) = \sigma_Y^2$. In this work, we adopt an isotropic exponential covariance model for Y :

$$C_Y(\mathbf{x}, \mathbf{x}') = \sigma_Y^2 \exp \left[-\frac{\|\mathbf{x} - \mathbf{x}'\|}{\lambda} \right], \quad (10)$$

where λ represents the correlation length. Therefore, the RSF model for the permeability field is based on two-point spatial statistics. Furthermore, we assume that σ_Y^2 and λ are deterministic and we neglect model uncertainty in the covariance shape, namely C_Y .

3.2. Representation through the Karhunen–Loeve expansion

Here, we describe the procedure to represent, discretize, and generate samples from the random field Y . The Karhunen–Loeve (KL) expansion for the random variable Y consists of a spectral expansion of its covariance function $C_Y(\mathbf{x}, \mathbf{x}')$ (Ghanem and Spanos, 2012; Christakos, 2005). Due to the fact that the covariance function is non-negative definite, symmetric, and with finite trace, it can be decomposed as follows,

$$C_Y(\mathbf{x}, \mathbf{x}') = \sum_{n=1}^{\infty} \lambda_n f_n(\mathbf{x}) f_n(\mathbf{x}'), \quad (11)$$

where the eigenvalues $\{\lambda_n\}_{n=1}^{\infty}$ and eigenfunctions $\{f_n\}_{n=1}^{\infty}$ can be computed as the solution to the Fredholm integral equation (Ghanem and Spanos, 2012)

$$\int_D C_Y(\mathbf{x}, \mathbf{x}') f(\mathbf{x}) d\mathbf{x} = \lambda f(\mathbf{x}'). \quad (12)$$

The log-permeability field can then be constructed through the KL expansion:

$$Y(\mathbf{x}, \omega) = \langle Y(\mathbf{x}) \rangle + \sum_{n=1}^{\infty} \xi_n(\omega) \sqrt{\lambda_n} f_n(\mathbf{x}), \quad (13)$$

where $\{\xi_n(\omega)\}_{n=1}^{\infty}$ are orthogonal Gaussian random variables, obtained by sampling the standard Gaussian distribution $\mathcal{N}(0, 1)$.

It can be shown that the trace of the sum of the eigenvalues λ_i is equal to the variance of the process Y . Thus, a common and reasonable truncation argument of the KL expansion is to retain enough terms such that the cumulative sum of the eigenvalues surpasses 99% of the total sum. When the truncation is performed, the KL expansion becomes KL decomposition. While the KL expansion significantly reduces the random dimension of the permeability field, the number of terms of the KL expansion that are not truncated can still be substantial, leading to slow convergence rates in PCE methods. To address this issue and alleviate the associated computational burden, we will utilize the *Projection Pursuit Adaptation* method to reduce even further the dimension of the permeability field (Zeng and Ghanem, 2023).

3.3. Projection pursuit adaptation in polynomial chaos expansions

The Projection Pursuit Adaptation (PPA) method, introduced by Zeng and Ghanem (2023), offers a data-driven approach for reducing the stochastic dimension and addressing uncertainty quantification and surrogate modeling in high-dimensional problems.

The PPA method involves projecting the original randomized parameters onto a low-dimensional space, where the projected variables are adapted to the quantity of interest Q and encompass sufficient information to accurately represent it. Subsequently, a Polynomial Chaos Expansion (PCE) (Ghanem and Spanos, 2012) is established for Q on this space, serving as a surrogate model for uncertainty quantification and predictive analysis. Notably, the method incorporates a new concept (Zeng and Ghanem, 2023), enabling it to simultaneously determine the optimal projections and the optimal PCE on the projected variables, leveraging available data. In this section, we present an overview of the fundamental principles behind the PPA method. As mentioned in Section 2, the quantity of interest (generically denoted by Q) would represent t_e or c_{max} .

3.3.1. Polynomial chaos expansion

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is the sample space, $\mathcal{F}$ is the σ -algebra of events, and $\mathbb{P}$ is a probability measure. Let $\xi = (\xi_1, \dots, \xi_d) \in \mathbb{R}^d$ be a collection of independent Gaussian variables defined on Ω . We assume that $\mathcal{H}$ denotes the Hilbert space generated by the linear span of ξ . Then, we can represent any random variable $Q \in L^2(\Omega, \mathcal{F}(\mathcal{H}), \mathbb{P})$ by a PCE that can be further truncated at a maximum polynomial order of b as

$$Q(\xi) = \sum_{\alpha \in J_d} Q_\alpha \psi_\alpha(\xi) \approx \sum_{\alpha \in J_{d,b}} Q_\alpha \psi_\alpha(\xi). \quad (14)$$

For example, the Q here can be t_e or c_{max} . In the above equation, $J_d := (\mathbb{N}_0)^d$ represents the set of d -dimensional multi-indices. Then, $\alpha = (\alpha_1, \dots, \alpha_d) \in J_d$ denotes a d -dimensional multi-index with a length given by $|\alpha| = \sum_{i=1}^d \alpha_i$. Moreover, $J_{d,b}$ is a subset of d -dimensional multi-indices, contained within J_d , that have a maximum length of b (where b is the maximum polynomial order of the truncated PCE) such that $|\alpha| \leq b$. The set $\{\psi_\alpha\}_{\alpha \in J_d}$ consists of orthogonal Hermite polynomials and forms a complete basis in the space $L^2(\Omega, \mathcal{F}(\mathcal{H}), \mathbb{P})$. The PCE coefficient associated with the basis ψ_α is denoted as Q_α . As $b \rightarrow \infty$, the PCE expressed in Eq. (14) converges to Q in a mean-squared sense. The convergence analysis with respect to polynomial orders is typically carried out by incrementally increasing the order until the discrepancy between two consecutive expansions falls within tolerance. The number of PCE terms in (14) is

$$P = \binom{d+b}{b} = \frac{(d+b)!}{d!b!}, \quad (15)$$

which grows factorially with d and b . Additionally, the computation of the PCE coefficients involves a d -dimensional integral. Thus, in high-dimensional scenarios with a large value of d , the number of samples required to calculate the PCE coefficients accurately increases at a super-linear rate. This phenomenon, known as the curse of dimensionality, leads to computationally expensive numerical estimations. To overcome this challenge, the PPA method is proposed.

3.3.2. Projection pursuit adaptation

Basis adaptation (Tiptredy and Ghanem, 2014; Zeng et al., 2021) is a dimension reduction technique developed specifically to attenuate the computational burden for high-dimensional polynomial chaos representations. Building on the basis adaptation concept, a data-driven PPA method was proposed in Zeng and Ghanem (2023), which will be employed in this work to propagate randomness from the permeability field to Q , constrained by the governing system of non-linear partial differential equations, see Eqs. (1) and (5).

The main idea is to initially identify a low-dimensional space constructed by projections on a Gaussian subspace. Subsequently, the PCE of Q can be constructed within this low-dimensional space, rather than the original high-dimensional space. These two steps are performed optimally using the provided data to ensure utilization of the lowest achievable low-dimensional manifold. Here, we will only illustrate the basic ideas, with a more detailed description available elsewhere (Zeng and Ghanem, 2023).

The PPA method employs multivariate PCE to approximate Q . Suppose we have a given dataset $\{(\xi^i, q^i)\}_{i=1}^N$, where $\xi^i \in \mathbb{R}^d$ and $q^i \in \mathbb{R}$. This dataset is obtained from running the complex computational model, serving as our training data set. We aim to utilize this data to develop a regression model that predicts the QoI based on input parameters. Let $\mathbf{A}_r = [\mathbf{a}_1 \dots \mathbf{a}_r]^T$, where $\{\mathbf{a}_j\}$ denotes the mutually orthonormal projections of size d . Let $\mathbf{Z}$ represent the projected variables, defined as $\mathbf{Z} = \mathbf{A}_r \xi$. The PPA model represents Q , as

$$Q = g_r(\mathbf{Z}). \quad (16)$$

Here, $g_r(\cdot)$ is an r -dimensional PCE with definition given in Eq. (14). The matrix $\mathbf{A}_r$ acts as a rotation matrix applied to the input parameters. In the present case, $\mathbf{A}_r$ is employed to transform the orthogonal Gaussian random variables produced from the KL expansion, while the

function g_r corresponds to the PCE functions we aim to develop for the variables t_e and c_{max} . The optimal projections $\mathbf{A}_r$ and the PCE function g_r need to be computed. This is achieved by minimizing the squared residual between the data, $\{(\xi^i, q^i)\}_{i=1}^N$, where q^i represent the data of Q , and the PCE representation above. That is, the following optimization problem is to be solved

$$(g_r, \mathbf{A}_r) = \arg \min_{(\hat{g}_r, \hat{\mathbf{A}}_r)} \sum_{i=1}^N [q^i - \hat{g}_r(\mathbf{z}^i)]^2 = \arg \min_{(\hat{g}_r, \hat{\mathbf{A}}_r)} \sum_{i=1}^N [q^i - \hat{g}_r(\hat{\mathbf{A}}_r \xi^i)]^2. \quad (17)$$

A forward stage-wise strategy is utilized to determine $(g_r, \mathbf{a}_r)$ at each stage. During each stage, a projection is integrated, and the process concludes when incorporating further information no longer significantly enhances model fitting. That is, in each stage, a new projection corresponds to a new dimension of the transformed parameters used to fit the PCE. Notably, this convergence pertains to the PCE model, which, given an appropriate selection of the polynomial order, converges to the physical model. Consequently, the convergence observed in the PCE model implies the convergence of the PDF of the quantity of interest (t_e or c_{max}), establishing a more robust property than mere convergence in the lower moments.

The optimization problem is solved in an alternating manner. Given $\mathbf{a}_r$, $\mathbf{A}_r$ is constructed, and the objective becomes finding the optimal g_r that minimizes the squared residual. This can be achieved through least-square regression.

On the contrary, when g_r is given, our goal is to minimize the squared residual with respect to $\mathbf{a}_r$. To avoid excessive computation, we update only $\mathbf{a}_r$ at each stage. As proposed in Zeng and Ghanem (2023), we use a Gauss–Newton search algorithm to solve the optimization problem with the given g_r . By keeping the projections of $\mathbf{a}_1, \dots, \mathbf{a}_{r-1}$ unchanged and fixing g_r , we need to solve the following optimization problem:

$$\mathbf{a}_r = \arg \min_{\hat{\mathbf{a}}_r} \sum_{i=1}^N [q^i - g_r(\mathbf{z}^i)]^2 = \arg \min_{\hat{\mathbf{a}}_r} \sum_{i=1}^N [q^i - g_r([\mathbf{a}_1, \dots, \mathbf{a}_{r-1}, \hat{\mathbf{a}}_r]^T \xi^i)]^2. \quad (18)$$

With the Gauss–Newton method, the optimization problem becomes (Zeng and Ghanem, 2023)

$$\mathbf{a}_r = \arg \min_{\hat{\mathbf{a}}_r} \sum_{i=1}^N \left(\frac{\partial g_r(\mathbf{z}_{old}^i)}{\partial \mathbf{z}_{r,old}^i} \right)^2 \left[\left(\mathbf{a}_{r,old}^T \xi^i + \frac{q^i - g_r(\mathbf{z}_{old}^i)}{\partial g_r(\mathbf{z}_{old}^i) / \partial \mathbf{z}_{r,old}^i} \right) - \hat{\mathbf{a}}_r^T \xi^i \right]^2. \quad (19)$$

Here, $\mathbf{a}_{r,old}$ is the current estimate of $\mathbf{a}_r$; $\mathbf{z}_{old}^i = [\mathbf{a}_1 \dots \mathbf{a}_{r-1} \mathbf{a}_{r,old}]^T \xi^i$; $\mathbf{z}_{r,old}^i = \mathbf{a}_{r,old}^T \xi^i$. The optimization problem (19) can be solved by

the weighted least-square regression with target $\hat{\mathbf{u}} = \frac{q^i - g_r(\mathbf{z}_{old}^i)}{\partial g_r(\mathbf{z}_{old}^i) / \partial \mathbf{z}_{r,old}^i}$,

regressors ξ^i , weight $\hat{\mathbf{w}} = \left(\partial g_r(\mathbf{z}_{old}^i) / \partial \mathbf{z}_{r,old}^i \right)^2$, and no bias term. Let $\hat{\mathbf{W}} \in \mathbb{R}^{N \times N}$ be a diagonal matrix with weight $\hat{\mathbf{w}}$ as its entries. Also let $\hat{\mathbf{u}} \in \mathbb{R}^N$ be a vector with entries of $\hat{\mathbf{u}}$, and Ξ the data matrix of $\{\xi^i\}$. Then, $\mathbf{a}_r$ can be computed by the following

$$\mathbf{a}_r = \arg \min_{\hat{\mathbf{a}}_r} \|\hat{\mathbf{u}} - \Xi \hat{\mathbf{a}}_r\|_{\hat{\mathbf{W}}}^2 = (\Xi^T \hat{\mathbf{W}} \Xi)^{-1} \Xi^T \hat{\mathbf{W}} \hat{\mathbf{u}}. \quad (20)$$

After computing $\mathbf{a}_r$, the Gram–Schmidt procedure is utilized on $\mathbf{A}_r$ to ensure orthogonality is maintained. Additionally, with the updated $\mathbf{a}_r$, g_r is refitted to minimize the residual. Subsequently, $\mathbf{a}_r$ is updated again using the new g_r function. This alternating procedure is repeated until convergence is achieved for $(g_r, \mathbf{a}_r)$.

We note that the PPA model inherits the mean-square convergence property with respect to the polynomial orders of the PCE. This implies that one can systematically elevate the order to assess whether higher-order information contributes to the QoI and to verify the convergence of the PCE model. Achieving a PCE model that converges with its

Table 1

Input parameters used in the simulations. See Fig. 1 for details regarding the model configuration. Dimensional values are provided in Appendix.

Parameter	Symbol	Value
Geometric mean of the Y -field	$\exp(Y)$	1
Variance of the Y -field considered	σ_Y^2	0.5, 1.5
Domain length in x_1	L_1	12
Domain length in x_2	L_2	10
Mesh size in x_1 x_2	Δ	1/12
Source's distance from top boundary	s_{tb}	3
Source's distance from bottom boundary	s_{bb}	3
Source's distance from left boundary	s_l	2
Length of source in x_1	ℓ_1	1
Length of source in x_2	ℓ_2	4
Control plane's distance from left boundary	L_{cp}	11

physical counterpart involves progressively increasing the polynomial orders. Subsequently, PPA models are constructed based on the selected order. The process continues until the difference between two successive PPA models with consecutive orders falls within the specified tolerance.

In summary, for any given polynomial order, the PPA method employs two layers of operations to identify the optimal projections and an optimal PCE on the projected (or adapted) variables; see below.

Operation layer 1. A forward stage-wise strategy is used to determine the number of projections. Specifically, we add one projection at each stage and stop the procedure if adding a new projection cannot improve model fitting.

Operation layer 2. At each stage, the optimal (a_r, g_r) is discovered in an alternating manner. We first fix a_r and compute the associated optimal g_r . We then fix g_r and compute the associated optimal a_r which updates its current value. These two steps are repeated until convergence.

4. Numerical implementation and input parameters

In this section we describe the methods employed to solve the governing equations and provide the input parameters used in the simulations. The flow Eq. (1) is solved numerically following the details provided in Bonazzi et al. (2021) which are briefly summarized in this section. Eq. (1) is solved for cell-centered pressures using a finite-volume method with second-order accuracy. The Darcy flux in Eq. (2) is linearized using the two-point flux approximation. As for the solute transport, we employ an explicit scheme to compute the concentration field in Eq. (5). In particular, we use a second-order finite difference method, and compute the concentration at the next time step in terms of the advective and diffusive mass fluxes.

In this study, we consider two different values of variance of the Y -field discussed in Section 3.1 in order to compare the effect of high and low heterogeneity: $\sigma_Y^2 = 0.5$ and $\sigma_Y^2 = 1.5$. We also consider two different levels of viscosity contrast, defined in terms of the log-viscosity ratio presented in Eq. (3): $R = 0$ (i.e. absence of viscosity contrast), and $R = 1.5$. We thus consider a total of four cases, given by all the possible combinations in heterogeneity level and viscosity contrast considered.

We discretized the domain with a two-dimensional Cartesian mesh with grid blocks of equal (dimensionless) size in both x_1 and x_2 directions, i.e. $\Delta x_1 = \Delta x_2 = \Delta$. Details regarding the dimensionless parameters used in our computational analysis are reported in Table 1. All lengths are normalized by the correlation length in the x_1 and x_2 -direction, $\lambda = 5$ m. Please refer to Fig. 1 for additional details regarding the computational domain and other relevant information.

5. Computational results

Given the randomness of $Y(\mathbf{x}, \omega) = \ln[k(\mathbf{x}, \omega)]$, with $\mathbf{x} \in D$ and $\omega \in \Omega$, model outputs such as the maximum concentration observed at

the control plane and early arrival times will be quantified statistically through their moments and probability density functions (PDF) or cumulative density functions (CDF). These model outputs are considered environmental performance metrics given their importance in a series of applications in probabilistic risk analysis in hydrology and environmental engineering (de Barros et al., 2012; Libera et al., 2019).

As mentioned in Section 3.3, we will use the PPA method to quantify the uncertainty in c_{max} and t_e . We will compare the results of the PPA method with the results obtained through traditional Monte Carlo simulations applied to the computationally expensive physics-based model. For the Monte Carlo simulations relative to the physics-based model, we randomly generate a large number (N_r) of permeability fields using the KL approach described in Section 3. For all considered cases, we truncated the KL expansion after the first 130 terms, whose cumulative sum was enough to reach 99% of the total eigenvalue summation, as explained in Section 3.2. For each realization of the permeability field, we numerically solve the flow and transport equations, see Eqs. (1)–(5), in order to compute t_e and c_{max} . We then compute the statistics of t_e and c_{max} over the ensemble of size N_r . In our work, we set $N_r = 7000$.

Figs. 2 and 3 show the first four statistical moments of t_e and c_{max} as a function of N_r for different R and σ_Y^2 , obtained from the Monte Carlo simulations applied to the physics-based model. The results highlight how the physics of the problem control the convergence of the statistical moments of the quantities of interest. The first four statistical moments of a quantity of interest Q , i.e. mean $\mu_Q^{(1)}$, variance $\mu_Q^{(2)}$, skewness $\mu_Q^{(3)}$ and kurtosis $\mu_Q^{(4)}$, of a sample of size N_r , are calculated as follows:

$$\mu_Q^{(1)} = \frac{1}{N_r} \sum_{i=1}^{N_r} Q_i \quad (21)$$

$$\mu_Q^{(2)} = \frac{1}{(N_r - 1)} \sum_{i=1}^{N_r} [Q_i - \mu_Q^{(1)}]^2 \quad (22)$$

$$\mu_Q^{(3)} = \frac{1}{[\mu_Q^{(2)}]^{3/2}} \sum_{i=1}^{N_r} \frac{1}{N_r} [Q_i - \mu_Q^{(1)}]^3 \quad (23)$$

$$\mu_Q^{(4)} = \frac{1}{[\mu_Q^{(2)}]^2} \sum_{i=1}^{N_r} \frac{1}{N_r} [Q_i - \mu_Q^{(1)}]^4 \quad (24)$$

As depicted in Figs. 2 and 3, N_r was sufficient to achieve statistical convergence of the mean of t_e and c_{max} . On one hand, we observe that the mean t_e at the control plane (Fig. 2, top left) stabilizes for $N_r \lesssim 2000$ for all R and σ_Y^2 values considered. On the other hand, when we consider the variance of the early arrival times, $\mu_{t_e}^{(2)}$ (Fig. 2, top right), there is a distinction in the number of realization required to reach convergence, and that distinction is based on the level of heterogeneity of the log-permeability field. For the lowest level of permeability heterogeneity considered ($\sigma_Y^2 = 0.5$), for both values of R considered, convergence is reached for $N_r \lesssim 3000$. However, an increase of heterogeneity level ($\sigma_Y^2 = 1.5$) required a higher number of realizations to reach statistical convergence: $N_r \approx 5000$ for $R = 1.5$, and $N_r \gtrsim 6000$ for $R = 0$. In particular, we note that the high heterogeneity case with no viscosity contrast ($\sigma_Y^2 = 1.5$, $R = 0$) requires additional simulations to reach full convergence of $\mu_{t_e}^{(2)}$. This can be explained considering that the presence of a viscosity contrast ($R > 0$) generates fingering at the front of the plume and enhances the mobility of the plume, thus leading to earlier early arrival times at the control plane and reducing the sample-to-sample fluctuations of t_e . Fig. 2 reveals that the higher statistical moments of t_e have not reached convergence at $N_r = 7000$ for the highest level of permeability heterogeneity considered ($\sigma_Y^2 = 1.5$).

As for the maximum concentration at the control plane, c_{max} , we can see from Fig. 3 that the mean value, $\mu_{c_{max}}^{(1)}$, reaches convergence for $N_r \lesssim 3000$. Convergence in the mean c_{max} is reached faster in the presence of viscosity contrast. This is due to the fact that the presence of a viscosity contrast (i.e., $R > 0$) augments mixing rates of the

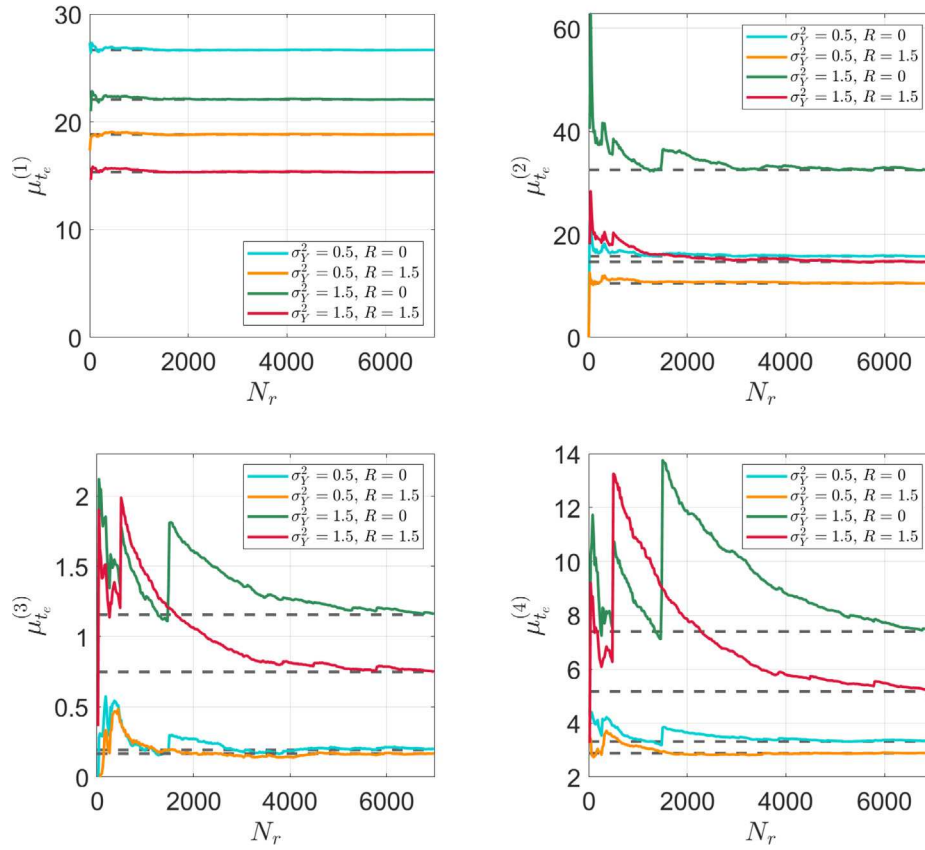


Fig. 2. Convergence analysis of the Monte Carlo simulations for the first four statistical moments of the plume's early arrival time at the control plane for the four cases considered in this study. The results depict the mean $\mu_{t_e}^{(1)}$ (top left), variance $\mu_{t_e}^{(2)}$ (top right), skewness $\mu_{t_e}^{(3)}$ (bottom left), and kurtosis $\mu_{t_e}^{(4)}$ (bottom right). The dashed gray lines represent the value of each statistical moment obtained from the Monte Carlo simulations for an ensemble of size $N_r = 7000$.

solute plume thus reducing the sample-to-sample fluctuations of the ensemble (Bonazzi et al., 2021). Notice that, while the second order moment $\mu_{c_{max}}^{(2)}$ seems to be reaching convergence for $N_r \approx 7000$, the higher moments of c_{max} have not reached convergence yet at $N_r = 7000$.

Based on the previous results, it is evident that the Monte Carlo method applied to the physics-based model is proficient at obtaining converged low-order statistics. However, when it comes to more intricate probabilistic information, particularly detailed PDFs of the QoI, its accuracy may be lacking. In the forthcoming sections, we will present results derived from the PCE-based PPA method and draw comparisons with the Monte Carlo method. The primary objective of this comparison is to highlight the limitations of the Monte Carlo method applied to the physics-based model in capturing detailed probabilistic information. As discussed in Section 3.3, the PPA method inherits the mean-square convergence property of the PCE concerning the polynomial order. To ensure the convergence of the PPA model, we systematically increase the polynomial orders in the PPA models. We stop this procedure when the PPA models with two successive maximum polynomial orders (e.g., order 3 and order 4 in Fig. 4) exhibit a difference within a predefined tolerance. In this study, our emphasis is on uncertainty quantification rather than point-wise prediction. Therefore, we use the comparison of PDFs as a criterion to assess convergence with respect to the polynomial orders.

It is important to note that the ideal approach to confirm the convergence of the PPA model would involve generating a significantly larger number of Monte Carlo samples and utilizing the Monte Carlo model as a reference. However, practical constraints, particularly computational limitations, make this nearly impossible in many applications, especially when dealing with complex underlying physics, as is the case in the application presented in this paper. In a recent work (Zeng and

Ghanem, 2023), this approach was demonstrated in two applications, where the computational burden of conducting an extensive number of physical model evaluations remained manageable. The results revealed that the converged PPA model, in terms of polynomial orders, exhibited very good accuracy when compared to a reference model.

In the following, we report the PDF and CDFs of t_e and c_{max} obtained by the PPA method for different values of R and σ_Y^2 . We compare the PPA-based uncertainty estimates with the ones obtained through the Monte Carlo simulations. For the PPA method, we randomly choose 6000 of the 7000 Monte Carlo samples as the training data set, while the remaining 1000 data are employed as test data set. We computed the relative L2 norm error to compare the PPA model prediction of the test set with the outputs from the flow and transport simulations, for all QoIs and all four scenarios considered in this study; the order of magnitude of the relative L2 norm error was 10^{-1} . The input of the PPA model are the KL terms of the permeability field, which has a dimension $D = 130$ (i.e. the number of terms considered from the KL expansion). Fig. 4 illustrates the convergence of the early-time PDF with respect to the PCE order. As depicted in Fig. 4, the findings suggest that the PDFs of the third-order PCEs closely align with those of the fourth-order PCEs. This implies that higher-order information has minimal impact on the QoI starting from order four. Consequently, a fourth-order PCE proves to be sufficient for achieving convergence in the PPA-based PDFs. Furthermore, it is noteworthy that the PPA dimensions for all these cases are discovered to be 5. This signifies that, despite the input dimension being D , the PPA method is capable of identifying a 5-dimensional space that accurately represents the QoI using fourth-order PCEs. The dimension reduction achieved here is crucial as it alleviates the need for an extensive number of samples. In this particular application, the convergence of the PPA

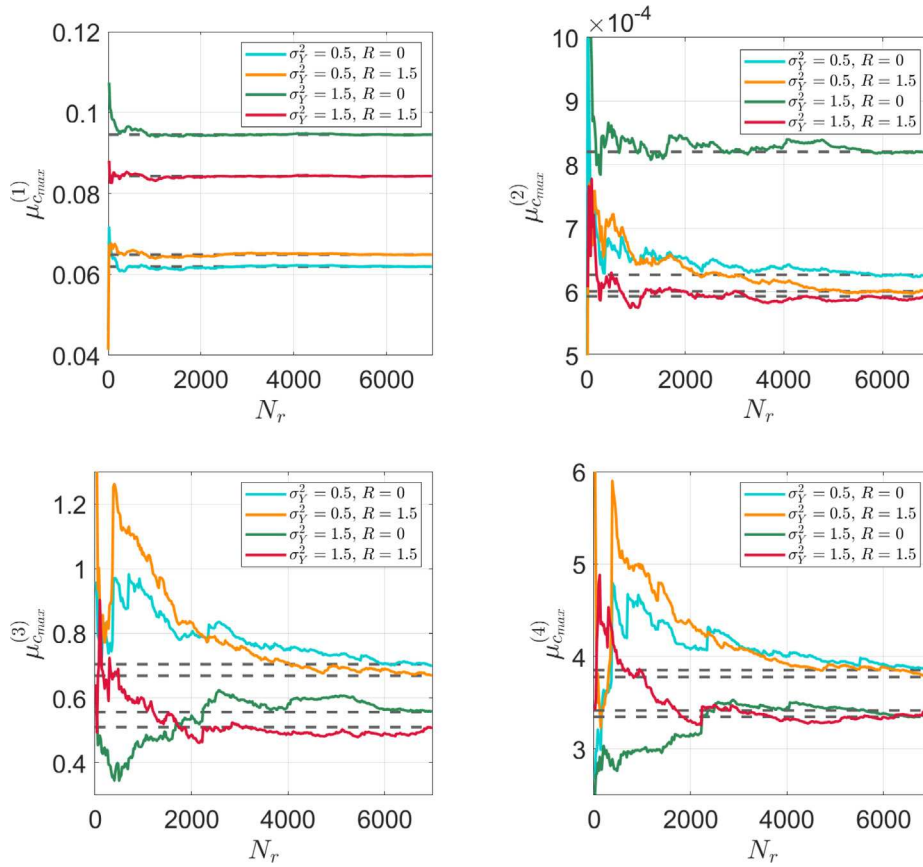


Fig. 3. Convergence analysis of the Monte Carlo simulations for the first four statistical moments of the maximum concentration at the control plane for the four cases considered in this study. The results depict the mean $\mu_{C_{max}}^{(1)}$ (top left), the variance $\mu_{C_{max}}^{(2)}$ (top right), skewness $\mu_{C_{max}}^{(3)}$ (bottom left), and kurtosis $\mu_{C_{max}}^{(4)}$ (bottom right). The dashed gray lines represent the value of each statistical moment obtained from the Monte Carlo simulations for an ensemble of size $N_r = 7000$.

models demonstrates that 6000 samples prove to be adequate for the 5-dimensional representations.

Fig. 5 reports the PDFs of t_e for $R = 0$ (in blue) and $R = 1.5$ (in yellow), for two levels of log-permeability heterogeneity, namely $\sigma_Y^2 = 0.5$ (Fig. 5, left) and $\sigma_Y^2 = 1.5$ (Fig. 5, right). The results in Fig. 5 depict how viscous fingering enhances, in a statistical sense, the mobility of the leading edge of the solute plume. Close inspection of Fig. 5 shows that there is higher probability of observing earlier arrival times when $R = 1.5$ when compared to the case of $R = 0$. When viscous fingering is present, the solute reaches the control plane earlier. This is expected since viscous fingering enhances the mobility of the solute plume, hence leading to earlier first arrival times (Nicolaides et al., 2015; Bonazzi et al., 2021). The presence of viscous fingering also causes a slight reduction in the uncertainty associated with t_e , especially for the case of higher level of permeability heterogeneity (compare $\sigma_Y^2 = 0.5$ to 1.5 cases), as visible in Fig. 5 (right), where the tails of the PDF for $R = 0$ span a larger t_e interval when compared to the PDF for $R = 1.5$. Fig. 5 also shows the differences between the PPA-based PDF and the PDF obtained through Monte Carlo simulations. The observed differences are mainly due to the lack of convergence of the higher order moments of t_e (see Fig. 2). Table 2 reports the differences between the statistics of the PDFs reported in Fig. 5 obtained via the PPA-method and the Monte Carlo simulations.

The differences between the Monte Carlo and the PPA results can be better quantified by analyzing the CDF of t_e (Fig. 6). With the exception of the case for $\sigma_Y^2 = 0.5$ and $R = 0$, differences are observed between the Monte Carlo (colored lines) and the PPA (black lines) CDFs. From the figure, it is possible to see that the Monte Carlo method lacks in accuracy when it comes to capturing the tails of the distribution. Similar to the PDF analysis (Fig. 5), the results depicted in Fig. 6 highlight

the significant role of viscous fingering in enhancing plume mobility when compared to the results obtained with $R = 0$. As discussed in Fig. 5, the probability of observing earlier breakthroughs at the control plane is higher when both high permeability heterogeneity and viscous fingering are present (red continuous line), as both sources of disorder lead to an increase of fluid channeling effects, thus allowing the front of the plume to travel faster. Lowering the permeability heterogeneity (dashed orange line) results in a lower probability $\mathbb{P}[t < \tau]$, where τ can represent a critical travel time established by a regulatory agency (e.g. for safety assessment of waste repositories). For example, if we look at $\mathbb{P}[t_e < 15]$ (i.e., the probability that the plume will reach the control plane earlier than $t = 15$) for the case with $\sigma_Y^2 = 1.5$ and $R = 1.5$ (red line), we have $\mathbb{P}[t_e < 15] \approx 0.5$. On the other hand, for the case with lower permeability heterogeneity ($\sigma_Y^2 = 0.5$ and $R = 1.5$, orange line) we have $\mathbb{P}[t_e < 15] \approx 0.1$. This means that the probability that the plume will reach a target location earlier than $t = 15$ is five times higher in the case of higher heterogeneity when compared to the case with lower heterogeneity, for the same level of viscosity contrast ($R = 1.5$).

Interestingly, by comparing the case where viscosity contrast is absent and higher permeability heterogeneity is considered ($\sigma_Y^2 = 1.5$ and $R = 0$, solid green line) with the case with lower permeability heterogeneity and viscosity contrast ($\sigma_Y^2 = 0.5$ and $R = 1.5$, dashed orange line) we can see that the probability of non-exceedance, i.e. $\mathbb{P}[t_e < \tau]$, is always higher for the case where a viscosity contrast is present, even if the level of permeability heterogeneity is lower (dashed orange line), for any considered value of τ . This suggests that the role of viscous fingering in enhancing connectivity might be predominant over the role of permeability heterogeneity, at least for the values of σ_Y^2 and R here considered in this work. As discussed in Bonazzi et al. (2021), the significance of viscous fingering in enhancing plume mobility, when

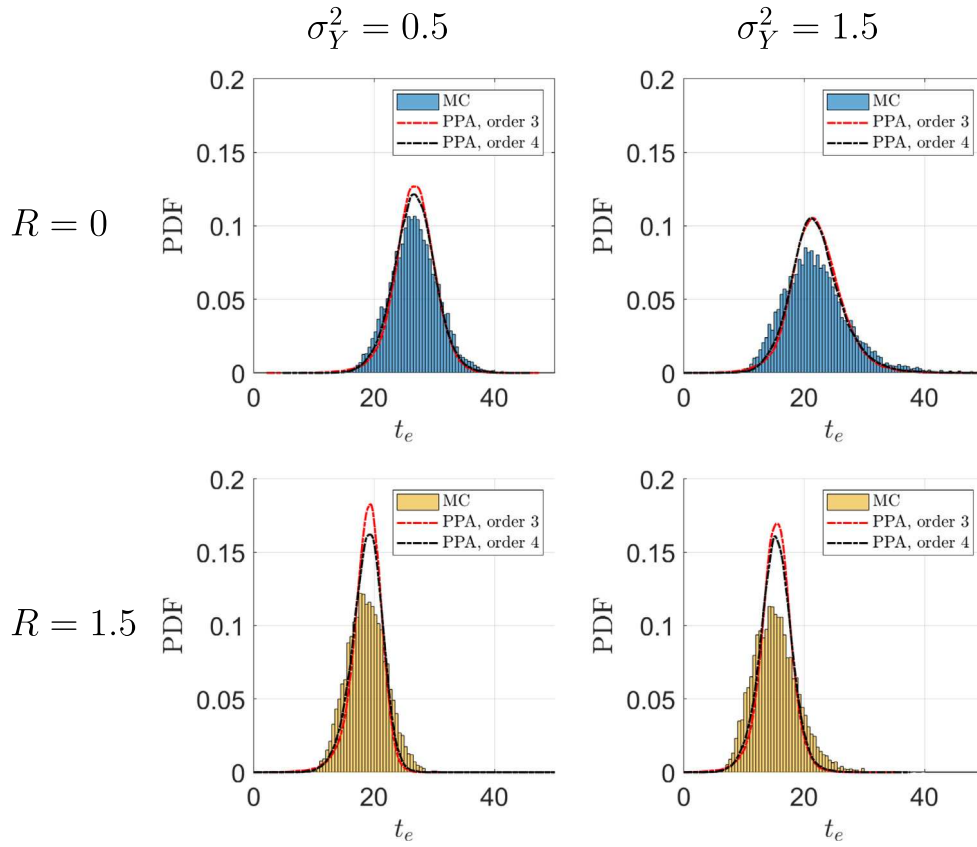


Fig. 4. Convergence analysis of the PDF obtained with PPA of the plume's early arrival time at the control plane for the four cases considered in this study.

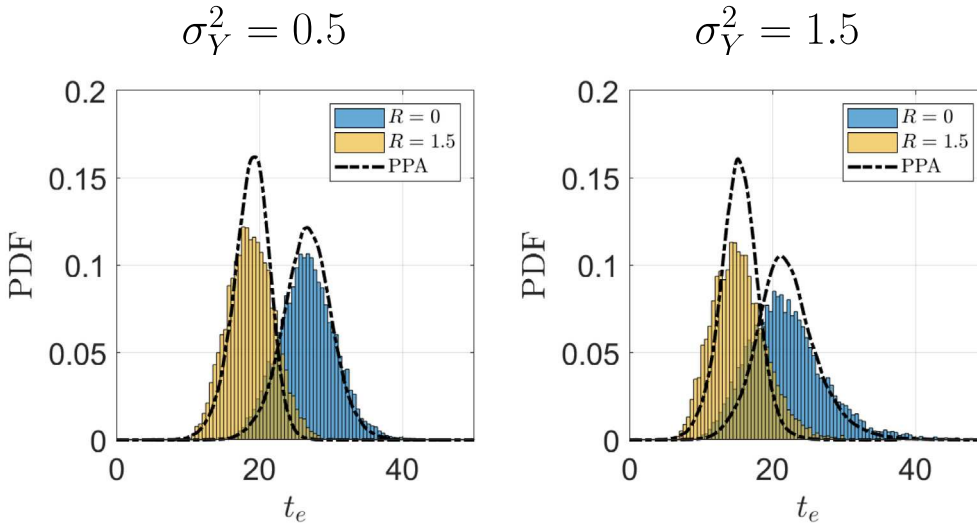


Fig. 5. PDFs of early arrival times t_e for the low permeability heterogeneity cases ($\sigma_Y^2 = 0.5$, left) and the high permeability heterogeneity cases ($\sigma_Y^2 = 1.5$, right). The colored results correspond to the Monte Carlo results and the dashed black lines represent the PDF calculated with the PPA method (fourth-order PCE). The PPA curves were generated using 10^5 samples.

compared to permeability heterogeneity, will depend on the distance between the source and the control plane.

Next, we analyze the PDFs of the maximum concentration observed at the control plane (Fig. 7). As depicted in Fig. 7, an increase in R from 0 to 1.5 leads to steeper and narrower PDFs for c_{max} . Viscous fingering enhances mixing between the two fluids which contributes to a reduction of the sample-to-sample variability of the plume's concentration field within the ensemble. The mixing between both fluids is further augmented when the heterogeneity level is increased from

$\sigma_Y^2 = 0.5$ to 1.5. This is because higher permeability heterogeneity leads to increased plume spreading, thus larger inter-facial area between the plume and the surrounding fluid therefore enhancing diffusive mass transfer (de Barros et al., 2015). On average, we observe a smaller c_{max} for the case of $\sigma_Y^2 = 1.5$ when compared to the PDFs obtained for $\sigma_Y^2 = 0.5$.

Fig. 8 reports the results in terms of CDFs. The probability of non-exceedance for the maximum concentration registered at the control plane, defined as $\mathbb{P}[c_{max} < c^*]$, is always higher for the cases with

Table 2

Comparison between the statistics of the early arrival times obtained through the PPA method and Monte Carlo simulations. The first four statistical moments of early arrival times are reported, as well as the ratios of the statistical moments between Monte Carlo and PPA. The subscript “MC” denotes the moments obtained from the Monte Carlo simulations whereas the subscript “PPA” correspond to the moments computed through the Projection Pursuit Adaptation method.

Simulation	Mean			Variance		
	$\mu_{MC}^{(1)}$	$\mu_{PPA}^{(1)}$	$\mu_{MC}^{(1)}/\mu_{PPA}^{(1)}$	$\mu_{MC}^{(2)}$	$\mu_{PPA}^{(2)}$	$\mu_{MC}^{(2)}/\mu_{PPA}^{(2)}$
$\sigma_Y^2 = 0.5, R = 0$	26.66	26.68	1.00	15.75	11.98	1.31
$\sigma_Y^2 = 0.5, R = 1.5$	18.83	18.83	1.00	10.50	6.64	1.58
$\sigma_Y^2 = 1.5, R = 0$	22.06	22.08	1.00	32.53	19.69	1.65
$\sigma_Y^2 = 1.5, R = 1.5$	15.33	15.31	1.00	14.67	7.29	2.01
Simulation	Skewness			Kurtosis		
	$\mu_{MC}^{(3)}$	$\mu_{PPA}^{(3)}$	$\mu_{MC}^{(3)}/\mu_{PPA}^{(3)}$	$\mu_{MC}^{(4)}$	$\mu_{PPA}^{(4)}$	$\mu_{MC}^{(4)}/\mu_{PPA}^{(4)}$
$\sigma_Y^2 = 0.5, R = 0$	0.19	-0.002	-89.97	3.32	3.45	0.96
$\sigma_Y^2 = 0.5, R = 1.5$	0.17	-0.37	-0.45	2.89	4.26	0.68
$\sigma_Y^2 = 1.5, R = 0$	1.16	0.89	1.31	7.40	8.44	0.88
$\sigma_Y^2 = 1.5, R = 1.5$	0.75	0.02	36.00	5.18	3.89	1.33

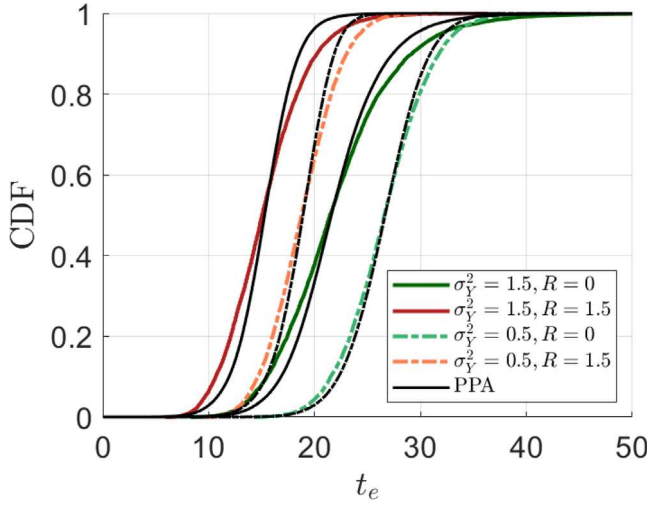


Fig. 6. CDFs of early arrival times t_e for the four combinations of R and σ_Y^2 considered. The black lines represent the CDFs calculated with the PPA method (fourth-order PCE) whereas the colored lines correspond to Monte Carlo-based CDFs. The PPA curves were generated using 10^5 samples.

$\sigma_Y^2 = 1.5$ compared to the cases with $\sigma_Y^2 = 0.5$ (assuming the same level of viscosity contrast), thanks to the increased dilution caused by higher levels of heterogeneity. For example, in absence of viscosity contrast between the fluids (Fig. 8, left) the Monte Carlo approach provides $\mathbb{P}[c_{max} < 0.1] \approx 0.9$ for the higher heterogeneity case ($\sigma_Y^2 = 1.5$, solid line), while $\mathbb{P}[c_{max} < 0.1] \approx 0.6$ for the lower heterogeneity case ($\sigma_Y^2 = 0.5$, dashed line). Considering the CDF provided by the PPA method, the difference in probability of non-exceedance between the two levels of heterogeneity is even more marked, with $\mathbb{P}[c_{max} < 0.1] \approx 1$ and $\mathbb{P}[c_{max} < 0.1] \approx 0.6$ for the higher and lower levels of heterogeneity, respectively. When we consider the presence of a viscosity contrast between the fluids (Fig. 8, right) the Monte Carlo approach gives $\mathbb{P}[c_{max} < 0.1] \approx 0.9$ for the higher heterogeneity case ($\sigma_Y^2 = 1.5$, solid line), while $\mathbb{P}[c_{max} < 0.1] \approx 0.75$ for the lower heterogeneity case ($\sigma_Y^2 = 0.5$, dashed line). On the other hand, the PPA method provides $\mathbb{P}[c_{max} < 0.1] \approx 1$ and $\mathbb{P}[c_{max} < 0.1] \approx 0.9$ for the higher and lower levels of heterogeneity, respectively.

Finally, both Figs. 7 and 8 highlight the inability of a computationally expensive Monte Carlo simulation (with 7000 realizations) to capture the uncertainty of the maximum concentration at the control plane. This is particularly evident for the case of $R = 1.5$, when non-linear effects are present in the governing flow and transport equations. The differences between the PPA-based method and the Monte Carlo

simulation are also reported in Table 3. Table 3 reports a quantitative comparison between the statistical moments of the maximum concentration at the control plane obtained from the PDFs reported in Fig. 7, for both the Monte Carlo simulations and the proposed PPA method.

6. Summary

We investigated how uncertainty related to the spatial structure of the permeability field affects the maximum concentration and early arrival times at a target location when viscous fingering is present. We consider two different levels of permeability heterogeneity (random field characterized by $\sigma_Y^2 = 0.5$ and $\sigma_Y^2 = 1.5$), and viscosity contrast ($R = 0$, i.e. no viscosity contrast, and $R = 1.5$), for a total of four possible combinations of level of permeability heterogeneity and viscosity contrast. Estimating uncertainty in highly non-linear problems, such as the one addressed in our work, through traditional approaches (such as Monte Carlo simulations) can be computationally expensive. Especially when the application at hand requires the estimation of the probability density function of a given quantity of interest. This is particularly important in probabilistic risk analysis where decision makers are interested in the probability of extreme events. In order to quantify the uncertainty associated with these two quantities of interest, we propose employing the PPA method, which offers several appealing advantages over other methods. It is a data-driven approach that optimally represents a given quantity of interest in a low-dimensional manifold. Unlike other dimension reduction techniques in uncertainty quantification, such as basis adaptation and active subspace, PPA utilizes non-linear information of the quantity of interest to identify the low-dimensional manifold, thereby increasing the likelihood of finding a more accurate lower-dimensional space. Additionally, PPA uses independent samples for all computations, which imposes fewer restrictions on the training dataset. Lastly, the PPA method possesses multi-quantity of interest capability, thus allowing the construction of surrogates for various quantities of interest using the same dataset. To the best of our knowledge, our work represents one of the first applications of the PPA method to highly non-linear physical formulation describing fluid flow and scalar transport in heterogeneous porous media.

Our results highlight the role of viscous fingering in enhancing the mobility of the solute plume being transported in a heterogeneous permeability field. Results show how viscous fingering increases the probability of observing earlier breakthroughs at a target location. We found that the presence of a viscosity contrast reduces the uncertainty associated with early arrival times t_e and maximum concentration c_{max} .

Notably, the number of Monte Carlo realizations considered in this study, $N_r = 7000$, is not sufficient to reach convergence of the PDFs when compared to the results obtained through the PPA method. The discrepancies between the PPA method and the Monte Carlo results are more evident when c_{max} is the quantity of interest. These results

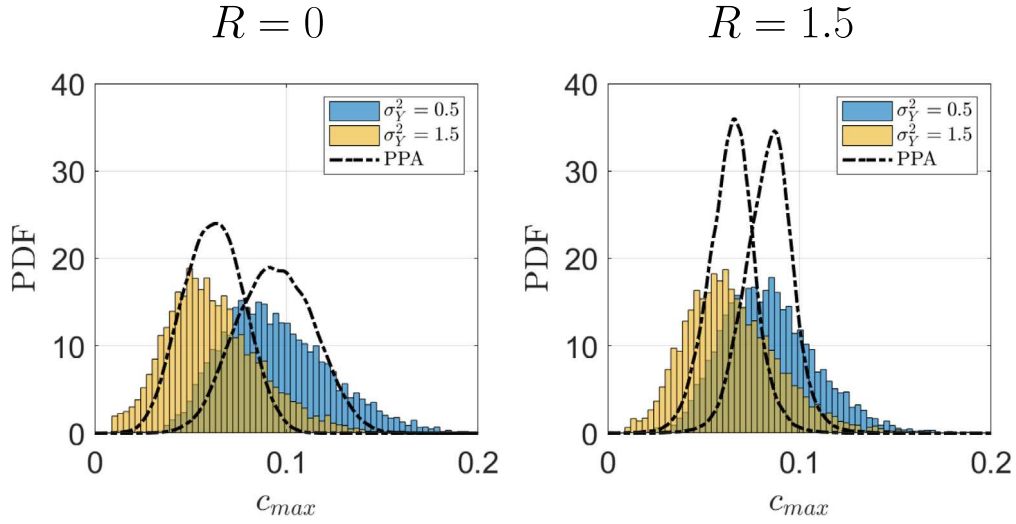


Fig. 7. PDFs of maximum concentration registered at the control plane for $R = 0$ (left) and $R = 1.5$ (right) for two distinct levels of heterogeneity ($\sigma_Y^2 = 0.5$ and 1.5). The black lines represent the PDFs calculated with the PPA method (fourth-order PCE), whereas the colored PDFs correspond to Monte Carlo simulations. The PPA curves were generated using 10^5 samples.

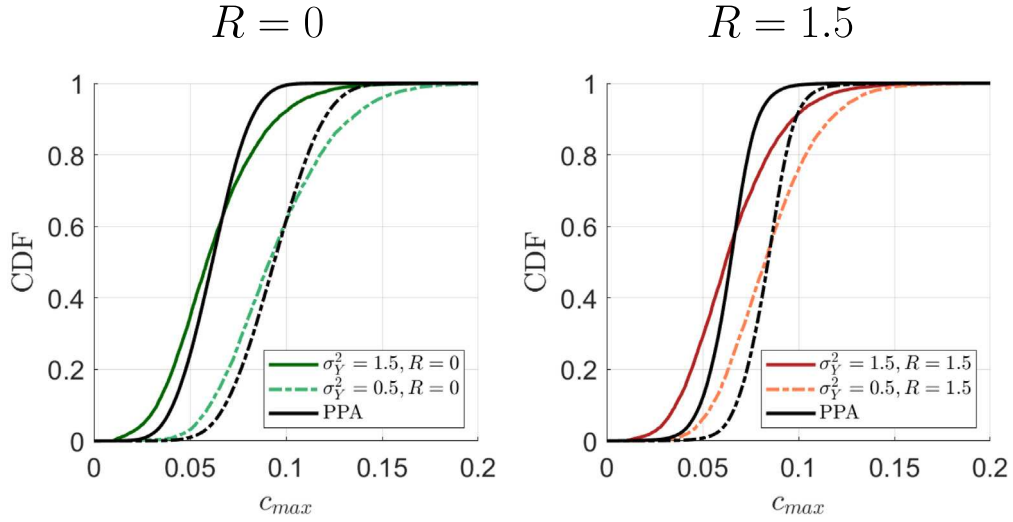


Fig. 8. CDFs of maximum concentration registered at the control plane for $R = 0$ (left) and $R = 1.5$ (right) for two distinct levels of permeability heterogeneity. The black lines represent the results calculated with the PPA method (fourth-order PCE) whereas the colored lines correspond to the results obtained via Monte Carlo simulations. The PPA curves were generated using 10^5 samples.

Table 3

Comparison between the statistics of the maximum concentration obtained through the PPA method and Monte Carlo simulations. The first four statistical moments of maximum concentration are reported, as well as the ratios of the statistical moments between Monte Carlo and PPA. The subscript “MC” denotes the moments obtained from the Monte Carlo simulations whereas the subscript “PPA” correspond to the moments computed through the Projection Pursuit Adaptation method.

Simulation	Mean			Variance		
	$\mu_{MC}^{(1)}$	$\mu_{PPA}^{(1)}$	$\mu_{MC}^{(1)}/\mu_{PPA}^{(1)}$	$\mu_{MC}^{(2)}$	$\mu_{PPA}^{(2)}$	$\mu_{MC}^{(2)}/\mu_{PPA}^{(2)}$
$\sigma_Y^2 = 0.5, R = 0$	0.095	0.094	1.00	$8.2 \cdot 10^{-4}$	$3.9 \cdot 10^{-4}$	2.08
$\sigma_Y^2 = 0.5, R = 1.5$	0.084	0.084	1.00	$5.9 \cdot 10^{-4}$	$1.7 \cdot 10^{-4}$	3.47
$\sigma_Y^2 = 1.5, R = 0$	0.062	0.061	1.00	$6.2 \cdot 10^{-4}$	$2.5 \cdot 10^{-4}$	2.54
$\sigma_Y^2 = 1.5, R = 1.5$	0.065	0.065	1.00	$6.0 \cdot 10^{-4}$	$1.6 \cdot 10^{-4}$	3.86
Simulation	Skewness			Kurtosis		
	$\mu_{MC}^{(3)}$	$\mu_{PPA}^{(3)}$	$\mu_{MC}^{(3)}/\mu_{PPA}^{(3)}$	$\mu_{MC}^{(4)}$	$\mu_{PPA}^{(4)}$	$\mu_{MC}^{(4)}/\mu_{PPA}^{(4)}$
$\sigma_Y^2 = 0.5, R = 0$	0.56	0.06	9.17	3.35	3.12	1.07
$\sigma_Y^2 = 0.5, R = 1.5$	0.51	-0.01	-51.72	3.42	8.10	0.42
$\sigma_Y^2 = 1.5, R = 0$	0.70	0.05	13.88	3.85	3.05	1.26
$\sigma_Y^2 = 1.5, R = 1.5$	0.67	-0.12	-5.66	3.78	5.21	0.73

highlight the importance of having alternative methods, such as the PPA method, for risk assessment of contaminated subsurface sites and reliable decision making under uncertainty.

Aside from the uncertainty stemming from the permeability field, other sources of uncertainty can play a role. Further research is needed to assess the impact of model uncertainty associated with the relationship between the dissolved solute concentration and fluid viscosity. Although the heterogeneous log-permeability fields investigated in this work are multi-Gaussian, the proposed uncertainty estimation method can also be used to address non-Gaussian fields by making use of copula-based geostatistical tools or specific transformations (i.e., Nataf transformation). The results in this study are limited to a 2D configuration. As shown in the literature (Bonazzi et al., 2023), the solute mixing dynamics in a 3D heterogeneous flow can be quite different when compared to a 2D case. Therefore, additional research efforts should be allocated to explore how the uncertainty of first arrival times and maximum concentration are affected by viscous fingering in 3D flows. Finally, the results in this work are limited to $R \geq 0$. Similar analysis should be carried out for $R < 0$.

CRediT authorship contribution statement

Alessandra Bonazzi: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis. **Xiaoshu Zeng:** Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Formal analysis. **Roger Ghanem:** Writing – review & editing, Supervision, Methodology. **Birendra Jha:** Writing – review & editing, Supervision, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Felipe P.J. de Barros:** Writing – review & editing, Writing – original draft, Resources, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

A. Bonazzi, B. Jha and F.P.J. de Barros acknowledge the financial support from the National Science Foundation NSF Award 2025285. The authors are grateful for the constructive comments raised by the Editor and the anonymous Reviewers.

Appendix. Dimensionless groups

In the following, we present the dimensionless group adopted in this study. All dimensional quantities are denoted with a $\hat{\cdot}$ symbol. Eqs. (1)–(5), see Section 2, are based on the following dimensionless groups:

$$\mathbf{x} = \frac{\hat{\mathbf{x}}}{\lambda}, \mathbf{u} = \frac{\hat{\mathbf{u}}}{U}, t = \frac{\hat{t}}{\tau_u}, c = \frac{\hat{c}}{c_o}, \mu = \frac{\hat{\mu}}{\mu_o}, p = \frac{\hat{p}}{p_c}, k = \frac{\hat{k}}{h_c} \quad (\text{A.1})$$

where τ_u is the advective time scale, U is the mean longitudinal velocity, c_o is the initial concentration of the conservative tracer, μ_o is the viscosity of the ambient fluid, k_c is a characteristic (mean) permeability and p_c is the characteristic pressure. Here, we have $k_c = \exp[\langle Y \rangle] = 1$ m (i.e., geometric mean), $p_c = (\mu_o U) / (\lambda k_c)$ and $\tau_u = \lambda / U$. Another dimensionless group present in Eq. (5) is the Péclet number, $Pe = U \lambda / D$. All lengths are normalized by the correlation length λ . For this work, we set $\lambda = 5$ m, $\hat{L}_1 = 60$ m, $\hat{L}_2 = 50$ m, $\hat{D} = 0.42$ m, $\hat{s}_{ib} = 15$ m, $\hat{s}_{bb} = 15$ m, $\hat{s}_l = 10$ m, $\hat{\ell}_1 = 5$ m, $\hat{\ell}_2 = 20$ m and $\hat{L}_{cp} = 55$ m. Table 1 contains these parameters in dimensionless values.

References

- Andričević, R., 1998. Effects of local dispersion and sampling volume on the evolution of concentration fluctuations in aquifers. *Water Resour. Res.* 34 (5), 1115–1129.
- Ballio, F., Guadagnini, A., 2004. Convergence assessment of numerical Monte Carlo simulations in groundwater hydrology. *Water Resour. Res.* 40 (4).
- Bear, J., 1988. *Dynamics of Fluids in Porous Media*. Dover Publications.
- Bellin, A., Rubin, Y., Rinaldo, A., 1994. Eulerian-Lagrangian approach for modeling of flow and transport in heterogeneous geological formations. *Water Resour. Res.* 30 (11), 2913–2924.
- Bellin, A., Salandini, P., Rinaldo, A., 1992. Simulation of dispersion in heterogeneous porous formations: Statistics, first-order theories, convergence of computations. *Water Resour. Res.* 28 (9), 2211–2227.
- Bellin, A., Tonina, D., 2007. Probability density function of non-reactive solute concentration in heterogeneous porous formations. *J. Contam. Hydrol.* 94 (1–2), 109–125.
- Bilionis, I., Zabarar, N., 2012. Multi-output local Gaussian process regression: Applications to uncertainty quantification. *J. Comput. Phys.* 231 (17), 5718–5746.
- Blatman, G., Sudret, B., 2010. An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis. *Probab. Eng. Mech.* 25 (2), 183–197.
- Blatman, G., Sudret, B., 2011. Adaptive sparse polynomial chaos expansion based on least angle regression. *J. Comput. Phys.* 230 (6), 2345–2367.
- Bonazzi, A., 2023. On the Transport Dynamics of Miscible Fluids in Porous Media Under Different Sources of Disorder (Ph.D. thesis). University of Southern California.
- Bonazzi, A., Dentz, M., de Barros, F.P.J., 2023. Mixing in multidimensional porous media: a numerical study of the effects of source configuration and heterogeneity. *Transp. Porous Media* 146 (1–2), 369–393.
- Bonazzi, A., Morvillo, M., Im, J., Jha, B., de Barros, F.P.J., 2021. Relative impacts of permeability heterogeneity and viscosity contrast on solute mixing. *Phys. Rev. Fluids* 6 (6), 064501.
- Boso, F., Tartakovsky, D.M., 2016. The method of distributions for dispersive transport in porous media with uncertain hydraulic properties. *Water Resour. Res.* 52 (6), 4700–4712.
- Butera, I., Tanda, M.G., 1999. Solute transport analysis through heterogeneous media in nonuniform in the average flow by a stochastic approach. *Transp. Porous Media* 36, 255–291.
- Caroni, E., Fiorotto, V., 2005. Analysis of concentration as sampled in natural aquifers. *Transp. Porous Media* 59, 19–45.
- Chen, J., Zeng, X., Peng, Y., 2017. Probabilistic analysis of wind-induced vibration mitigation of structures by fluid viscous dampers. *J. Sound Vib.* 409, 287–305.
- Christakos, G., 2005. *Random Field Models in Earth Sciences*. Dover Publications.
- Christie, M., 1989. High-resolution simulation of unstable flows in porous media. *SPE Reservoir Eng.* 4 (03), 297–303.
- Cirpka, O.A., de Barros, F.P.J., Chiogna, G., Nowak, W., 2011. Probability density function of steady state concentration in two-dimensional heterogeneous porous media. *Water Resour. Res.* 47 (11).
- Constantine, P.G., 2015. *Active Subspaces: Emerging Ideas for Dimension Reduction in Parameter Studies*. vol. 2, SIAM.
- Constantine, P.G., Dow, E., Wang, Q., 2014a. Active subspace methods in theory and practice: applications to kriging surfaces. *SIAM J. Sci. Comput.* 36 (4), A1500–A1524.
- Constantine, P.G., Dow, E., Wang, Q., 2014b. Active subspace methods in theory and practice: applications to kriging surfaces. *SIAM J. Sci. Comput.* 36 (4), A1500–A1524.
- Crevillen-Garcia, D., Wilkinson, R., Shah, A., Power, H., 2017. Gaussian process modelling for uncertainty quantification in convectively-enhanced dissolution processes in porous media. *Adv. Water Resour.* 99, 1–14.
- Cvetkovic, V., Fiori, A., Dagan, G., 2016. Tracer travel and residence time distributions in highly heterogeneous aquifers: Coupled effect of flow variability and mass transfer. *J. Hydrol.* 543, 101–108.
- Cvetkovic, V., Shapiro, A., Dagan, G., 1992. A solute flux approach to transport in heterogeneous formations: 2. uncertainty analysis. *Water Resour. Res.* 28 (5), 1377–1388.
- Dagan, G., 1982. Stochastic modeling of groundwater flow by unconditional and conditional probabilities: 2. The solute transport. *Water Resour. Res.* 18 (4), 835–848.
- Darwiche, A., 2009. *Modeling and Reasoning with Bayesian Networks*. Cambridge University Press.
- de Barros, F.P.J., Bellin, A., Cvetkovic, V., Dagan, G., Fiori, A., 2016. Aquifer heterogeneity controls on adverse human health effects and the concept of the hazard attenuation factor. *Water Resour. Res.* 52 (8), 5911–5922.
- de Barros, F.P.J., Ezzedine, S., Rubin, Y., 2012. Impact of hydrogeological data on measures of uncertainty, site characterization and environmental performance metrics. *Adv. Water Resour.* 36, 51–63.
- de Barros, F.P.J., Fiori, A., 2014. First-order based cumulative distribution function for solute concentration in heterogeneous aquifers: Theoretical analysis and implications for human health risk assessment. *Water Resour. Res.* 50 (5), 4018–4037.

- de Barros, F.P.J., Fiori, A., 2021. On the maximum concentration of contaminants in natural aquifers. *Transp. Porous Media* 140 (1), 273–290.
- de Barros, F.P.J., Fiori, A., Boso, F., Bellin, A., 2015. A theoretical framework for modeling dilution enhancement of non-reactive solutes in heterogeneous porous media. *J. Contam. Hydrol.* 175, 72–83.
- de Barros, F.P.J., Guadagnini, A., Riva, M., 2022. Features of transport in non-Gaussian random porous systems. *Int. J. Heat Mass Transfer* 184, 122244.
- De Wit, A., Bertho, Y., Martin, M., 2005. Viscous fingering of miscible slices. *Phys. Fluids* 17 (5), 054114.
- Dentz, M., 2012. Concentration statistics for transport in heterogeneous media due to stochastic fluctuations of the center of mass velocity. *Adv. Water Resour.* 36, 11–22.
- Doostan, A., Owhadi, H., 2011. A non-adapted sparse approximation of PDEs with stochastic inputs. *J. Comput. Phys.* 230 (8), 3015–3034.
- Fiori, A., 2001. The Lagrangian concentration approach for determining dilution in aquifer transport: Theoretical analysis and comparison with field experiments. *Water Resour. Res.* 37 (12), 3105–3114.
- Fiori, A., Bellin, A., Cvetkovic, V., de Barros, F.P.J., Dagan, G., 2015. Stochastic modeling of solute transport in aquifers: From heterogeneity characterization to risk analysis. *Water Resour. Res.* 51 (8), 6622–6648.
- Fiori, A., Dagan, G., 2000. Concentration fluctuations in aquifer transport: A rigorous first-order solution and applications. *J. Contam. Hydrol.* 45 (1–2), 139–163.
- Fiori, A., Zarlanga, A., Jankovic, I., Dagan, G., 2017. Solute transport in aquifers: The comeback of the advection dispersion equation and the first order approximation. *Adv. Water Resour.* 110, 349–359.
- Fiorotto, V., Caroni, E., 2002. Solute concentration statistics in heterogeneous aquifers for finite Peclet values. *Transp. Porous Media* 48 (3), 331–351.
- Flowers, T.C., Hunt, J.R., 2007. Viscous and gravitational contributions to mixing during vertical brine transport in water-saturated porous media. *Water Resour. Res.* 43 (1).
- Ghahramani, Z., 2006. Learning dynamic Bayesian networks. In: *Adaptive Processing of Sequences and Data Structures: International Summer School on Neural Networks “ER Caianiello” Vietri sul Mare, Salerno, Italy September 6–13, 1997 Tutorial Lectures*. Springer, pp. 168–197.
- Ghanem, R., 1999. Ingredients for a general purpose stochastic finite elements implementation. *Comput. Methods Appl. Mech. Engrg.* 168, 19–34.
- Ghanem, R.G., Spanos, P.D., 2012. *Stochastic Finite Elements: A Spectral Approach*. Dover Publications.
- Ghauch, Z.G., Aitharaju, V., Rodgers, W.R., Pasupuleti, P., Dereims, A., Ghanem, R.G., 2019. Integrated stochastic analysis of fiber composites manufacturing using adapted polynomial chaos expansions. *Composites A* 118, 179–193.
- Giovanis, D.G., Shields, M.D., 2020. Data-driven surrogates for high dimensional models using Gaussian process regression on the Grassmann manifold. *Comput. Methods Appl. Mech. Engrg.* 370, 113269.
- Jha, B., Cueto-Felgueroso, L., Juanes, R., 2011a. Fluid mixing from viscous fingering. *Phys. Rev. Lett.* 106.
- Jha, B., Cueto-Felgueroso, L., Juanes, R., 2011b. Quantifying mixing in viscously unstable porous media flows. *Phys. Rev. E* 84 (6), 066312.
- Kapoor, V., Kitanidis, P.K., 1998. Concentration fluctuations and dilution in aquifers. *Water Resour. Res.* 34 (5), 1181–1193.
- Kitanidis, P.K., 1997. *Introduction to Geostatistics: Applications in Hydrogeology*. Cambridge University Press.
- Kougioumtzoglou, I., Spanos, P., 2012. An analytical Wiener path integral technique for non-stationary response determination of nonlinear oscillators. *Probab. Eng. Mech.* 28, 125–131.
- Leube, P.C., de Barros, F.P.J., Nowak, W., Rajagopal, R., 2013. Towards optimal allocation of computer resources: Trade-offs between uncertainty quantification, discretization and model reduction. *Environ. Model. Softw.* 50, 97–107.
- Li, J., Chen, J., 2009. *Stochastic Dynamics of Structures*. John Wiley & Sons.
- Li, W., Lu, Z., Zhang, D., 2009. Stochastic analysis of unsaturated flow with probabilistic collocation method. *Water Resour. Res.* 45 (8).
- Libera, A., Henri, C., de Barros, F.P.J., 2019. Hydraulic conductivity and porosity heterogeneity controls on environmental performance metrics: Implications in probabilistic risk analysis. *Adv. Water Resour.* 127, 1–12.
- Linde, N., Ginsbourger, D., Irving, J., Nobile, F., Doucet, A., 2017. On uncertainty quantification in hydrogeology and hydrogeophysics. *Adv. Water Resour.* 110, 166–181.
- MacKay, D.J., et al., 1998. *Introduction to Gaussian processes*. NATO ASI series F Comput. Syst. Sci. 168, 133–166.
- Meyer, D.W., Jenny, P., Tchelepi, H.A., 2010. A joint velocity-concentration PDF method for tracer flow in heterogeneous porous media. *Water Resour. Res.* 46 (12).
- Morvillo, M., Im, J., de Barros, F.P.J., 2022. VisU-Hydra: a computational toolbox for groundwater contaminant transport to support risk-based decision making. *Front. Earth Sci.* 10, 916198.
- Moslehi, M., de Barros, F.P.J., 2017. Uncertainty quantification of environmental performance metrics in heterogeneous aquifers with long-range correlations. *J. Contam. Hydrol.* 196, 21–29.
- Moslehi, M., Rajagopal, R., de Barros, F.P.J., 2015. Optimal allocation of computational resources in hydrogeological models under uncertainty. *Adv. Water Resour.* 83, 299–309.
- Nicolaides, C., Jha, B., Cueto-Felgueroso, L., Juanes, R., 2015. Impact of viscous fingering and permeability heterogeneity on fluid mixing in porous media. *Water Resour. Res.* 51 (4), 2634–2647.
- Oladyshkin, S., Class, H., Helmig, R., Nowak, W., 2011. A concept for data-driven uncertainty quantification and its application to carbon dioxide storage in geological formations. *Adv. Water Resour.* 34 (11), 1508–1518.
- Oladyshkin, S., Class, H., Nowak, W., 2013. Bayesian updating via bootstrap filtering combined with data-driven polynomial chaos expansions: methodology and application to history matching for carbon dioxide storage in geological formations. *Comput. Geosci.* 17, 671–687.
- Olivier, A., Giovanis, D.G., Aakash, B.S., Chauhan, M., Vandanapu, L., Shields, M.D., 2020. UQpy: A general purpose Python package and development environment for uncertainty quantification. *J. Comput. Sci.* 47, 101204.
- Psaros, A.F., Petromichelakis, I., Kougioumtzoglou, I.A., 2019. Wiener path integrals and multi-dimensional global bases for non-stationary stochastic response determination of structural systems. *Mech. Syst. Signal Process.* 128, 551–571.
- Rubin, Y., 1990. Stochastic modeling of macrodispersion in heterogeneous porous media. *Water Resour. Res.* 26 (1), 133–141.
- Rubin, Y., 2003. *Applied Stochastic Hydrogeology*. Oxford University Press.
- Rubin, Y., Dagan, G., 1992. Conditional estimation of solute travel time in heterogeneous formations: Impact of transmissivity measurements. *Water Resour. Res.* 28 (4), 1033–1040.
- Saffman, P., 1986. Viscous fingering in Hele-Shaw cells. *J. Fluid Mech.* 173, 73–94.
- Sanchez-Vila, X., Guadagnini, A., 2005. Travel time and trajectory moments of conservative solutes in three dimensional heterogeneous porous media under mean uniform flow. *Adv. Water Resour.* 28 (5), 429–439.
- Sargsyan, K., Safta, C., Najm, H.N., Debusschere, B.J., Ricciuto, D., Thornton, P., 2014. Dimensionality reduction for complex models via Bayesian compressive sensing. *Int. J. Uncertain. Quantif.* 4 (1), 63–93.
- Scheidegger, A.E., 1961. General theory of dispersion in porous media. *J. Geophys. Res.* 66 (10), 3273–3278.
- Schwede, R.L., Cirpka, O.A., Nowak, W., Neuweiler, I., 2008. Impact of sampling volume on the probability density function of steady state concentration. *Water Resour. Res.* 44 (12).
- Seeger, M., 2004. Gaussian processes for machine learning. *Int. J. Neural Syst.* 14 (02), 69–106.
- Soize, C., Ghanem, R., 2021. Probabilistic learning on manifolds constrained by nonlinear partial differential equations for small datasets. *Comput. Methods Appl. Mech. Engrg.* 380, 113777.
- Talon, L., Martin, J., Rakotomalala, N., Salin, D., 2004. Stabilizing viscosity contrast effect on miscible displacement in heterogeneous porous media, using lattice Bhatnagar–Gross–Krook simulations. *Phys. Fluids* 16 (12), 4408–4411.
- Tan, C., Homsy, G., 1988. Simulation of nonlinear viscous fingering in miscible displacement. *Phys. Fluids* 31 (6), 1330–1338.
- Tchelepi, H., Orr Jr., F., 1994. Interaction of viscous fingering, permeability heterogeneity, and gravity segregation in three dimensions. *SPE Reservoir Eng.* 9 (04), 266–271.
- Thimmisetty, C., Tsilifis, P., Ghanem, R., 2017. Homogeneous chaos basis adaptation for design optimization under uncertainty: Application to the oil well placement problem. *Ai Edam* 31 (3), 265–276.
- Tian, L., Wilkinson, R., Yang, Z., Power, H., Fagerlund, F., Niemi, A., 2017. Gaussian process emulators for quantifying uncertainty in CO₂ spreading predictions in heterogeneous media. *Comput. Geosci.* 105, 113–119.
- Tipireddy, R., Ghanem, R., 2014. Basis adaptation in homogeneous chaos spaces. *J. Comput. Phys.* 259, 304–317.
- Tonina, D., Bellin, A., 2008. Effects of pore-scale dispersion, degree of heterogeneity, sampling size, and source volume on the concentration moments of conservative solutes in heterogeneous formations. *Adv. Water Resour.* 31 (2), 339–354.
- Tran, M., Jha, B., 2020. Coupling between transport and geomechanics affects spreading and mixing during viscous fingering in deformable aquifers. *Adv. Water Resour.* 136, 103485.
- Tsilifis, P., Ghanem, R.G., 2017. Reduced Wiener chaos representation of random fields via basis adaptation and projection. *J. Comput. Phys.* 341, 102–120.
- Tsilifis, P., Huan, X., Safta, C., Sargsyan, K., Lacaze, G., Oefelein, J.C., Najm, H.N., Ghanem, R.G., 2019. Compressive sensing adaptation for polynomial chaos expansions. *J. Comput. Phys.* 380, 29–47.
- Van der Meer, L., 1993. The conditions limiting CO₂ storage in aquifers. *Energy Convers. Manage.* 34 (9–11), 959–966.
- Welty, C., Gelhar, L., 1991. Stochastic analysis of the effects of fluid density and viscosity variability on macrodispersion in heterogeneous porous media. *Water Resour. Res.* 27 (8), 2061–2075.
- Xiu, D., Karniadakis, G.E., 2002. The Wiener–Askey polynomial chaos for stochastic differential equations. *SIAM J. Sci. Comput.* 24, 619–644.
- Xiu, D., Karniadakis, G.E., 2003. Modeling uncertainty in flow simulations via generalized polynomial chaos. *J. Comput. Phys.* 187, 137–167.
- Zeng, X., Geraci, G., Eldred, M.S., Jakeman, J.D., Gorodetsky, A.A., Ghanem, R., 2023a. Multifidelity uncertainty quantification with models based on dissimilar parameters. *Comput. Methods Appl. Mech. Engrg.* 415, 116205.
- Zeng, X., Geraci, G., Gorodetsky, A., Jakeman, J., Eldred, M.S., Ghanem, R.G., 2023b. Improving Bayesian networks multifidelity surrogate construction with basis adaptation. In: *AIAA SCITECH 2023 Forum*. p. 0917.

- Zeng, X., Ghanem, R., 2023. Projection pursuit adaptation on polynomial chaos expansions. *Comput. Methods Appl. Mech. Engrg.* 405, 115845.
- Zeng, X., Peng, Y., Chen, J., 2017. Serviceability-based damping optimization of randomly wind-excited high-rise buildings. In: *The Structural Design of Tall and Special Buildings*. Vol. 26, (11), Wiley Online Library, e1371.
- Zeng, X., Red-Horse, J., Ghanem, R., 2021. Accelerated basis adaptation in homogeneous chaos spaces. *Comput. Methods Appl. Mech. Engrg.* 386, 114109.
- Zhang, H., Guilleminot, J., Gomez, L.J., 2021. Stochastic modeling of geometrical uncertainties on complex domains, with application to additive manufacturing and brain interface geometries. *Comput. Methods Appl. Mech. Engrg.* 385, 114014.