



A parametric acceleration of multilevel Monte Carlo convergence for nonlinear variably saturated flow

Prashant Kumar^{1,2} · Carmen Rodrigo³ · Francisco J. Gaspar³ · Cornelis W. Oosterlee^{1,4}

Received: 18 March 2019 / Accepted: 11 November 2019 / Published online: 20 December 2019
© The Author(s) 2019

Abstract

We present a multilevel Monte Carlo (MLMC) method for the uncertainty quantification of variably saturated porous media flow that is modeled using the Richards equation. We propose a stochastic extension for the empirical models that are typically employed to close the Richards equations. This is achieved by treating the soil parameters in these models as spatially correlated random fields with appropriately defined marginal distributions. As some of these parameters can only take values in a specific range, non-Gaussian models are utilized. The randomness in these parameters may result in path-wise highly nonlinear systems, so that a robust solver with respect to the random input is required. For this purpose, a solution method based on a combination of the modified Picard iteration and a cell-centered multigrid method for heterogeneous diffusion coefficients is utilized. Moreover, we propose a non-standard MLMC estimator to solve the resulting high-dimensional stochastic Richards equation. The improved efficiency of this multilevel estimator is achieved by parametric continuation that allows us to incorporate simpler nonlinear problems on coarser levels for variance reduction while the target strongly nonlinear problem is solved only on the finest level. Several numerical experiments are presented showing computational savings obtained by the new estimator compared with the original MC estimator.

Keywords UQ · Richards equation · MLMC · Modified Picard · Cell-centered multigrid

1 Introduction

Mass transport through a variably saturated porous medium can be accurately predicted using the Richards equation

✉ Prashant Kumar
pkumar@cwi.nl

Carmen Rodrigo
carmenr@unizar.es

Francisco J. Gaspar
fjgaspar@unizar.es

Cornelis W. Oosterlee
c.w.oosterlee@cwi.nl

- ¹ CWI - Centrum Wiskunde & Informatica, Amsterdam, The Netherlands
- ² Faculty of Aerospace Engineering, Delft University of Technology, Delft, The Netherlands
- ³ IUMA and Applied Mathematics Department, University of Zaragoza, Zaragoza, Spain
- ⁴ DIAM, Delft University of Technology, Delft, The Netherlands

[1]. This modeling approach is of critical importance for several physics and engineering problems, for instance, when studying aquifer recharge via rainfall infiltration, or for understanding the environmental impact of mining operations. When reliable measurements of the hydraulic properties are available, numerical solutions originating from the Richards equation have been reasonably successful for transport prediction in a broad range of soil types.

Different formulations for the Richards equation are available in the literature, along with well-established mathematical theory, such as the pressure head, the water content, or a mixed formulation, see, e.g., [2–7]. The aforementioned formulations contain nonlinearities due to a parametric dependence of the pressure head on the saturation and the relative hydraulic conductivity. Depending on the soil parameters, these nonlinearities range from mild to strong. The extreme sensitivity of the soil parameters on the output necessitates accurate measurements of the hydraulic properties. For many realistic problems, complete information of these quantities is however not available. In such scenarios, these parameters may be modeled in a probabilistic framework and the solution output may be expressed by means of a prediction

interval (with mean and variance), rather than as a single value. Such approaches are nowadays common in the case of saturated groundwater flow, where uncertainties are included when modeling the hydraulic conductivity as a spatially correlated lognormal random field [8–10]. The purpose of the present work is to develop and analyze a stochastic extension of the Richards equation, along with an efficient numerical method to solve the resulting nonlinear partial differential equation with random coefficients.

Previous work on the uncertainty quantification (UQ) of unsaturated flows was often based on an uncertain hydraulic conductivity [11–14]. In addition to that, in the present work, we introduce stochasticity in the so-called van Genuchten and Mualem model [15, 16], which is typically utilized to close the Richards equation. This model provides a closed-form analytic expression for the unsaturated hydraulic conductivity based on a *sigmoid* type function for the soil-water retention curve. This curve is defined by four independent parameters that are estimated by curve-fitting, based on field measurements. Typically, these parameters are fixed throughout the domain during numerical simulations, assuming the soil to be homogeneous. Realistic models should however also incorporate the intrinsic heterogeneity in the soil. Therefore, we model these soil parameters as random variables with a certain, specified probability distribution and spatial correlations. To assure the well-posedness of the Richards equation, these parameters should be within a certain range. Thus, the probability distributions for these parameters are chosen such that the random samples will be in the domain of validity for these parameters. A practical choice is to employ non-Gaussian random fields with marginal distributions from expert knowledge or from field measurements.

With the stochastic Richards equation formulated, an appropriate UQ technique is required to compute the statistical moments of the desired quantities of interest (QoIs). This choice primarily depends on the number of uncertainty dimensions. Other practical factors, such as ease of implementation and availability of an iterative solver which is robust with respect to the random input, also play a role in the selection of a suitable UQ technique. The proposed stochastic extension of the Richards equation results in a very high-dimensional problem, and the use of deterministic sampling approaches such as polynomial chaos expansion, stochastic collocation, or stochastic Galerkin is therefore limited. For these UQ methods, the cost grows exponentially with the number of random inputs. Furthermore, a deterministic sampling approach may not adequately represent those regions in the stochastic space where strong nonlinearity may be encountered. In previous works of Zhang [13, 17], the moments method was applied for the uncertainty quantification of solutions

of the Richards equation. The main disadvantage of a moment-based method is that it can only be reliably employed when the effect of uncertain inputs is mild and largely linear. For the proposed stochastic formulation of the Richards equation, Monte Carlo (MC)-based sampling approaches are the *methods of choice*, due to their dimension-independent convergence. Moreover, these MC-type methods can accurately represent the entire stochastic space given a sufficiently large number of samples. A well-known drawback of the plain MC method is its slow convergence of the sampling error, with $\mathcal{O}(1/\sqrt{N})$, where N is the number of samples, making it intractable for problems with a large cost per sample. Recently, efficient MC estimators based on the multilevel Monte Carlo (MLMC) method have been developed for a large class of problems, see, e.g., [18–21]. The efficiency of the MLMC estimation comes from solving the problem of interest on a coarse grid and subsequently adding corrections based on finer mesh resolutions. As these correction terms have smaller variances, they can be computed accurately using only a few samples. The estimates at different levels are then combined using a telescopic sum. The standard practice is to solve the PDE with random coefficients on a hierarchy of grids.

The original, grid-based MLMC estimator may be utilized to solve the stochastic Richards equation; however, this approach may not be the most efficient, especially not when strongly nonlinear problems need to be solved. Such problems require a very fine spatio-temporal mesh thereby restricting the use of coarse grids to improve the efficiency of the MLMC estimator. In this article, we utilize a non-standard MLMC estimator based on the *parametric continuation* technique. Continuation methods for solving nonlinear PDEs are very popular in engineering applications [22–27]. Within continuation methods, a nonlinearity dictating parameter Θ is introduced in the interval $\Theta_0 \leq \Theta \leq \Theta_*$ where the solution $p(\Theta_0)$ corresponds to a linear (or mildly nonlinear) problem and $p(\Theta_*)$ to the target strongly nonlinear. The key idea is to march from $p(\Theta_0)$ to $p(\Theta_*)$ in small steps of size $\delta\Theta$, where at each step we use the solution from the previous step as an initial guess. Usually, Θ is some physical parameter, for, e.g., the Reynolds number, and the Mach number. In the current work, we use parametric continuation to obtain variance reduction within the multilevel Monte Carlo framework. This is achieved by solving simpler nonlinear problems on coarser levels and the target strongly nonlinear problem is only solved on the finest level. This new estimator allows us to incorporate comparatively coarser spatio-temporal grids in the MLMC hierarchy and, as such, the computational cost of each estimator in the telescopic sum is greatly reduced. Here, we point out that in the context of the MLMC estimators, level-dependent simplifications of a

challenging target problem have already been explored in [28, 29]. In these articles, a level-dependent truncation of the Karhunen-Loève expansion of a random permeability field was employed in order to get a simplified representation of a heterogeneous permeability field on coarser levels.

We furthermore propose a solution method for Richards equation based on a combination of the modified Picard method [2] and a cell-centered multigrid, as proposed in [30]. We benchmark the performance of this combined solver in a probabilistic framework. A number of tests for a wide range of soil parameters and for hydraulic conductivities with different heterogeneity levels are performed.

The rest of the article is organized as follows. In Section 2, we briefly discuss the deterministic Richards equation along with the van Genuchten-Mualem parameterization. Section 3 describes the stochastic Richards equation as well as the modeling of various uncertain soil parameters. The description of the modified Picard method in combination with the cell-centered multigrid method is provided in Section 4. Also, in this section, we present some numerical experiments to assess the performance of the combined solver for an infiltration problem. The non-standard MLMC estimator is explained in Section 5 and its performance is analyzed in Section 6. Finally, some conclusions are drawn in Section 7.

2 Deterministic Richards equation

We begin by describing the deterministic version of the problem. The governing equations are defined in a bounded domain $\mathcal{D} \subset \mathbb{R}^n$, with the boundary $\partial\mathcal{D}$ and a finite time interval $\mathcal{T} = (0, T_{\text{final}}]$, with $T_{\text{final}} < \infty$. The classical Richards equation is a result of coupling the mass conservation equation of the water phase and the Darcian flow, i.e.,

$$\phi \frac{\partial S_w}{\partial t} + \nabla \cdot \mathbf{q} = f \quad \text{in } \mathcal{D} \times \mathcal{T}, \tag{2.1}$$

$$\mathbf{q} = -K_s K_{rw} (\nabla p + z) \quad \text{in } \mathcal{D} \times \mathcal{T}, \tag{2.2}$$

respectively, subject to boundary and initial conditions:

$$p = p_0 \quad \text{in } \mathcal{D}, \quad t = 0, \tag{2.3}$$

$$p = g_D \quad \text{in } \Gamma_D \times \mathcal{T}, \tag{2.4}$$

$$\mathbf{q} \cdot \mathbf{n} = g_N \quad \text{in } \Gamma_N \times \mathcal{T}, \tag{2.5}$$

where $\phi[L^3/L^3]$ is the porosity, $S_w[L^3/L^3]$ is the water-phase saturation; \mathbf{q} is the Darcy flux, which depends on the pressure head, $p[L]$, and the depth $z[L]$ in the vertical direction; $K_s[L/T]$ represents the saturated hydraulic conductivity field at saturation; K_{rw} is the relative conductivity of the water phase with respect to air and f is the source/sink term. The initial pressure head value is given by p_0 . The quantities g_D and g_N denote, respectively,

the Dirichlet and Neumann boundary conditions that are imposed at the boundaries Γ_D and Γ_N , respectively, with \mathbf{n} the unit normal vector to Γ_N .

The coupling of Eqs. 2.1 and 2.2 may result in different variants of the Richards equation, such as the pressure head, the moisture content, and the mixed formulation. The mixed formulation of the Richards equation is given by:

$$\frac{\partial \theta(p)}{\partial t} - \nabla \cdot (K_s K_{rw} (\nabla p + z)) = f \quad \text{in } \mathcal{D} \times \mathcal{T}. \tag{2.6}$$

It is obtained by substituting the moisture content, i.e. $\theta = \phi S_w(p)$. By using

$$\frac{\partial \theta(p)}{\partial t} = C(p) \frac{\partial p}{\partial t},$$

the above PDE can be reformulated into the pressure head formulation:

$$C(p) \frac{\partial p}{\partial t} - \nabla \cdot (K_s K_{rw} (\nabla p + z)) = f \quad \text{in } \mathcal{D} \times \mathcal{T}, \tag{2.7}$$

where $C(p) = \frac{\partial \theta}{\partial p}$ is the specific moisture capacity. It is well-known that numerical solutions originating from the pressure head formulation may give rise to a significant mass balance error, resulting in an inaccurate prediction of the infiltration depth.

Numerical methods based on the mixed form (using finite differences or mass-lumped finite elements) are popular as they result in mass conservation schemes [2]. Therefore, we will work with the mixed form Eq. 2.6 of the Richards equation.

2.1 Van Genuchten-Mualem model

To complete the PDE formulations, Eqs. 2.6 or 2.7, closure models for approximating K_{rw} and θ are required. A number of models have been presented in the literature and the most popular ones are by Brooks-Corey [31] and van Genuchten-Mualem [15, 16]. These two models employ nonlinear constitutive relations for K_{rw} and p , and for θ and p , respectively. We consider the parameterization introduced by van Genuchten and Mualem here. For the saturation, van Genuchten [15] proposed the following analytic formula:

$$S_w(p) = \frac{\theta(p) - \theta_r}{\theta_s - \theta_r} = \begin{cases} (1 + (|\alpha p|)^n)^{-m}, & p < 0, \\ 1, & p \geq 0, \end{cases} \tag{2.8}$$

where θ_s and θ_r are the saturated and residual water contents, respectively, and $\alpha[L^{-1}]$, n and $m = 1 - n^{-1}$ are obtained by fitting data characterizing the statistics of the soil. Specifically, the parameter α provides a measure of the average pore-size in the soil matrix and n is related to the pore-size distribution of the soil [32].

We may derive the specific moisture content, $C(p)$, analytically from Eq. 2.8, as

$$C(p) = \begin{cases} (\theta_s - \theta_r)\alpha mn(1 + |\alpha p|^n)^{-(m+1)}|\alpha p|^{n-1}, & p < 0, \\ 0, & p \geq 0. \end{cases} \quad (2.9)$$

In previous work, Mualem [16] derived a closed-form expression for K_{rw} , which is given by

$$K_{rw} = S_w^{1/2} \left[\frac{\int_0^{S_w} dS_w/p}{\int_0^1 dS_w/p} \right]^2. \quad (2.10)$$

Using Eq. 2.8, the above integral equation reduces to the following analytic expression:

$$K_{rw}(p) = \begin{cases} S_w(p)^{1/2} \left(1 - (1 - S_w(p)^{1/m})^m \right)^2, & p < 0, \\ 1, & p \geq 0. \end{cases} \quad (2.11)$$

The complexity of the numerical solution of the Richards equation depends on the values of the parameters n and α . For $n \in (1, 2)$ and $p \rightarrow 0$, the relative hydraulic conductivity $K_{rw}(p)$ is not Lipschitz continuous and the derivative $K'_{rw}(p)$ becomes infinite as p approaches zero [32, 33]. Moreover, for small values of n , a sharp K_{rw} vs. p profile is encountered. Similarly, for large values of the parameter α , the pressure head exhibits a transition behavior with a steep gradient from the saturated to the unsaturated region. In general, for a small n or for large α , strong nonlinearities are encountered, thus implying convergence issues for nonlinear iterative techniques such as the Newton or Picard methods.

3 Stochastic Richards model

Here, we describe a stochastic extension of the van Genuchten model. We assume that the unknown soil parameters belong to the probability space $(\Omega, \mathbb{F}, \mathbb{P})$, where Ω is the sample space with σ -field $\mathbb{F} \subset 2^\Omega$ as a set of events and the probability measure $\mathbb{P} : \mathbb{F} \rightarrow [0, 1]$.

The stochastic extension is based on modeling the soil parameters as spatially correlated random fields in order to incorporate spatial heterogeneity. For the saturated hydraulic conductivity, K_s , it is a standard practice to model it as a lognormal random field, as follows:

$$K_s(\mathbf{x}, \omega) = K_s^{(bl)}(\mathbf{x}) \exp(Z(\mathbf{x}, \omega)), \quad \mathbf{x} \in \mathcal{D}, \omega \in \Omega, \quad (3.1)$$

where $K_s^{(bl)}(\mathbf{x})$ is the baseline hydraulic conductivity and $Z(\mathbf{x}, \omega)$ is a zero mean Gaussian random field with a specified covariance kernel. So,

$$\mathbb{E}[Z(\mathbf{x}, \cdot)] = 0, \quad (3.2)$$

$$\text{Cov}(Z(\mathbf{x}_1, \cdot), Z(\mathbf{x}_2, \cdot)) = \mathbb{E}[Z(\mathbf{x}_1, \cdot)Z(\mathbf{x}_2, \cdot)], \quad \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}. \quad (3.3)$$

In the present work, we consider an anisotropic Matérn covariance function, C_Φ , defined as

$$\begin{cases} C_\Phi(\mathbf{x}_1, \mathbf{x}_2) = \sigma_c^2 \frac{2^{1-\nu_c}}{\Gamma(\nu_c)} (2\sqrt{\nu_c}\tilde{r})^{\nu_c} K_{\nu_c}(2\sqrt{\nu_c}\tilde{r}), \\ \tilde{r} = \sqrt{\frac{(x_1-x_2)^2}{\lambda_{cx}^2} + \frac{(z_1-z_2)^2}{\lambda_{cz}^2}}, \quad \text{with } \mathbf{x}_1 = (x_1, z_1), \mathbf{x}_2 = (x_2, z_2). \end{cases} \quad (3.4)$$

Here, we denote the gamma function by Γ and by K_{ν_c} the modified Bessel function of the second kind. The Matérn function is characterized by the parameter set $\Phi = \{\nu_c, \lambda_{cx}, \lambda_{cz}, \sigma_c^2\}$. Parameter $\nu_c \geq 0$ defines the differentiability of Z , $\sigma_c^2 > 0$ is the marginal variance, and λ_{cx} and λ_{cz} are correlation lengths along x - and z -coordinates, respectively. When $\nu_c = 1/2$, the Matérn function corresponds to an exponential covariance function and for $\nu_c \rightarrow \infty$ to a squared exponential covariance model. Simulating a Gaussian random field can be based on the Karhunen-Loève (KL) decomposition [34] of $Z(\mathbf{x}, \omega)$, i.e.,

$$Z(\mathbf{x}, \omega) = \sum_{j=1}^{\infty} \sqrt{\lambda_j} \Psi_j(\mathbf{x}) \xi_j, \quad \xi_j \sim \mathcal{N}(0, 1). \quad (3.5)$$

Here, λ_j and Ψ_j are eigenvalues and eigenfunctions of the covariance kernel $C_\Phi(\mathbf{x}_1, \mathbf{x}_2)$, obtained from the solution of the Fredholm integral,

$$\int_{\mathcal{D}} C_\Phi(\mathbf{x}_1, \mathbf{x}_2) \Psi(\mathbf{x}_1) d\mathbf{x}_1 = \lambda \Psi(\mathbf{x}_2). \quad (3.6)$$

The sum of Eq. 3.5 represents an infinite-dimensional uncertain field with a decaying contribution of the eigenmodes. The rate of decay typically depends on the smoothness and correlation length of the covariance function. The sum is truncated at a finite number of terms, M_{KL} , which is usually decided by balancing the KL-truncation error with other sources of error, like the discretization and sampling errors. For Gaussian processes with small correlation lengths and large variances, typically a large number of terms is needed to include the critical eigenmodes [34]. The evaluation of the eigenmodes in the KL-expansion is expensive as it requires solving the integral Eq. 3.6 for each mode. In the case of stationary covariance models, fast sampling of random fields can be achieved via spectral generators that employ the FFT (fast Fourier transform) [35, 36] for the factorization of the covariance matrix. Another advantage of using these spectral methods is that they are able to simulate random fields on the sampling mesh without any bias (for example, in the case of the KL-expansion). In this article, we use the fast Fourier transform moving average (FFT-MA) algorithm from [37] to sample Gaussian random fields, see Appendix for details.

3.1 Sampling of non-Gaussian random fields

For sampling the van Genuchten parameters, $\alpha(\mathbf{x}, \omega)$, $n(\mathbf{x}, \omega)$, $\theta_s(\mathbf{x}, \omega)$, and $\theta_r(\mathbf{x}, \omega)$ in Section 2.1, we employ random fields with *non-Gaussian marginal* distributions. This choice of distributions is practical as these parameters can only take values in a certain range, see, e.g, [33]. We introduce stochasticity in the parameters via an additive noise,

$$\alpha(\mathbf{x}, \omega) = \alpha^{(bl)}(\mathbf{x}) + \varepsilon_\alpha(\mathbf{x}, \omega), \quad (3.7)$$

Table 1 Two combinations of Matérn parameters $\Phi = (\nu_c, \lambda_{cx}, \lambda_{cz}, \sigma_c^2)$ corresponding to isotropic (Φ_1) and anisotropic (Φ_2) random fields

Φ_1	Φ_2
(1.0, 0.2, 0.2, 1)	(0.5, 0.1, 0.01, 1)

where $\alpha^{(bl)}(\mathbf{x})$ is the deterministic baseline value and $\varepsilon_\alpha(\mathbf{x}, \omega)$ is a random field with a non-Gaussian marginal distribution and covariance C_Φ . Notations are analogously for the other three van Genuchten parameters. Next, we describe a technique proposed in [38] for the point-wise transformation of a standard Gaussian random field to a non-Gaussian random field.

Non-Gaussian random fields are difficult to simulate as they are not uniquely determined by their mean and variance. There are however different techniques available

for simulating non-Gaussian fields, see, e.g., [38, 39]. In this work, we will follow a basic approach based on a generalized polynomial chaos (gPC) expansion [38], which approximates the non-Gaussian field in terms of a weighted combination of Hermite orthogonal polynomials of the standard Gaussian field,

$$Y(\mathbf{x}, \omega) \approx \sum_{j=0}^{N_{PC}} w_j \mathcal{H}_j(Z(\mathbf{x}, \omega)), \tag{3.8}$$

where $Y(\mathbf{x}, \omega)$ is the non-Gaussian random field (with a marginal distribution, e.g., the uniform distribution, gamma distribution, and truncated normal). $\mathcal{H}_j(Z)$ is the Hermite polynomial in Z of order j with weight w_j and N_{PC} is the order of the expansion. Hermite polynomials can be expressed as

$$\mathcal{H}_0(Z) = 1, \mathcal{H}_j(Z) = (-1)^j \exp(Z^2/2) \frac{d^j}{dx^j} \exp(-Z^2/2), \quad j \in \mathbb{N}. \tag{3.9}$$

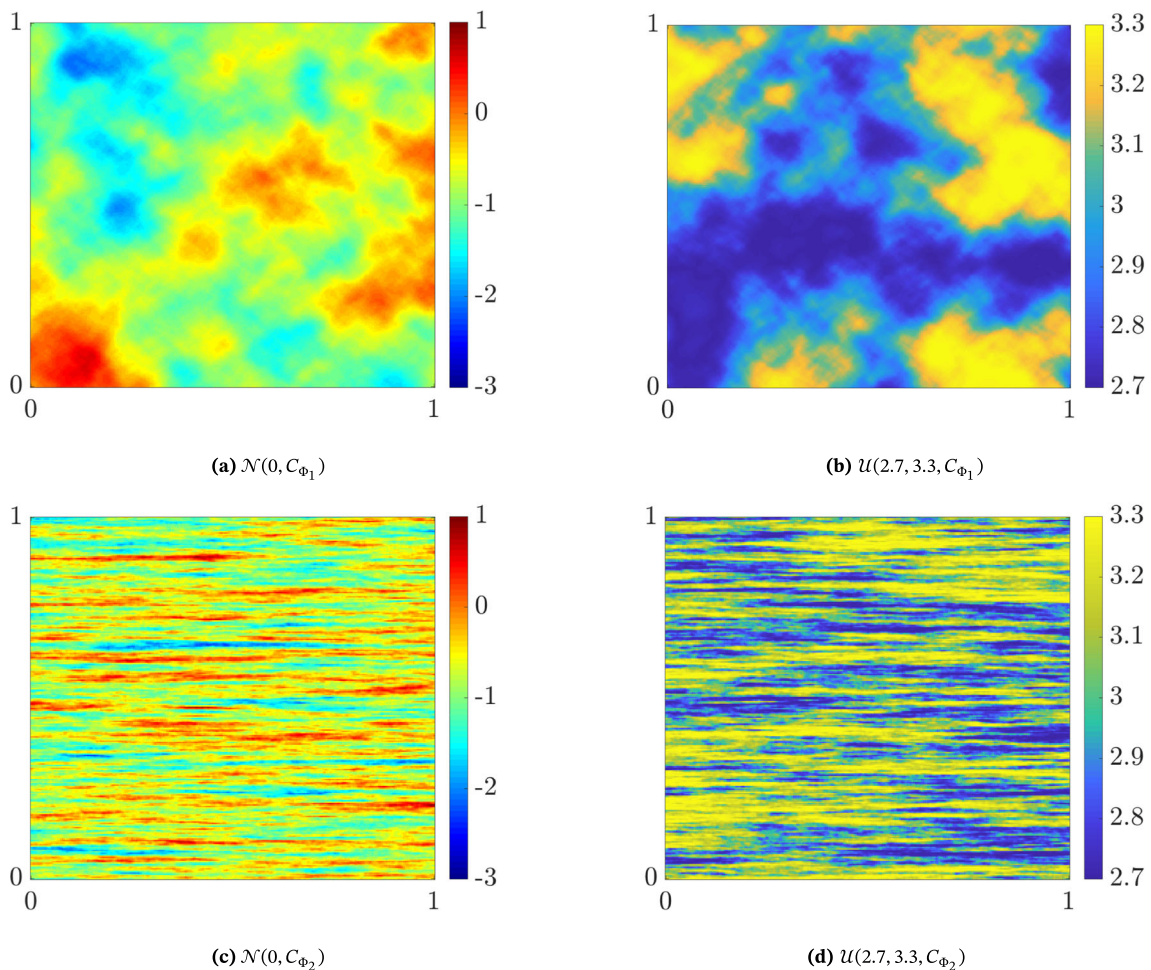


Fig. 1 a–d Samples of random fields generated using the isotropic Matérn parameter Φ_1 (top row) and the anisotropic parameter Φ_2 (bottom row); and standard normal marginal distribution (left column)

and uniform marginal distribution (right column). The notation $\mathcal{U}(2.7, 3.3, C_\Phi)$ represents a random field with uniform marginal distribution, $\mathcal{U}[2.7, 3.3]$, with spatial correlation defined by C_Φ

As Hermite polynomials are orthogonal with respect to the Gaussian measure, the weights can be evaluated using

$$w_j = \frac{\mathbb{E}[Y\mathcal{H}_j(Z)]}{\mathbb{E}[\mathcal{H}_j(Z)^2]} \tag{3.10}$$

Here, the denominator is basically an expectation of a polynomial of the Gaussian random variable, which has an analytic expression. As the dependence between Y and Z is unknown, the expectation in the numerator is not well-defined. Since the cumulative distribution for Y , defined as $F_Y(y) = \mathbb{P}(\text{Prob}(Y \leq y))$, is however known, one can utilize the relation $Y = F_Y^{-1}(F_Z(Z))$ to reformulate Eq. 3.10 as

$$w_j = \frac{1}{\mathbb{E}[\mathcal{H}_j(Z)^2]} \int_{I_Z} F_Y^{-1}[F_Z(z)]\mathcal{H}_j(z)dF_Z(z), \tag{3.11}$$

where I_Z is the support of Z in the range $(-\infty, \infty)$ and F_Y^{-1} representing the inverse of the distribution F_Y . Similarly, $F_Z(z) = \mathbb{P}(\text{Prob}(Z \leq z))$ is the cumulative distribution for standard Gaussian random variable Z . The integral Eq. 3.11 can be numerically computed using any conventional integration technique, for example, by using Monte Carlo quadrature. The weights only need to be computed once, so that the cost of sampling a non-Gaussian random field with a stationary covariance function is of the same order as that of a Gaussian random field.

We will experiment here with both isotropic and anisotropic Matérn covariance models. In Table 1, the two Matérn parameter sets are listed, Φ_1 corresponding to an isotropic model and Φ_2 to an anisotropic model. In Fig. 1, we present some samples of the random fields with a Gaussian and a uniform marginal distribution, for the two Matérn parameters. We use $N_{PC} = 6$ in Eq. 3.8 for generating the random fields with uniform marginal distribution. Due to a small correlation length and low spatial regularity, the numerical solutions of the PDE with random coefficients based on Φ_2 are comparatively more expensive to compute than those obtained with Φ_1 . A comprehensive study on the computational cost of solving elliptic PDEs with different Matérn parameters can be found in [30]. We will study the effect of covariance functions on the performance of the solver for the Richards equation.

4 Modified Picard iteration combined with the cell-centered multigrid method

Algorithms based on the modified Picard iteration from Celia et al. [2] are often employed as efficient iterative solution methods for the Richards equation. These methods are relatively easy to implement, as they do not require the computation of Jacobians and they also have low storage requirements. Within each modified Picard iteration, a diffusion equation with variable coefficients needs to be solved. For this, we propose to utilize the *cell-centered multigrid* (CCMG) for heterogeneous diffusion coefficients, as proposed in [30, 40, 41]. The CCMG algorithm is efficient as it is constructed with a simple set of transfer operators and it has been demonstrated to perform well for a large class of highly heterogeneous and also jumping diffusion coefficients [30].

4.1 Modified Picard iteration

We briefly recall the fully implicit Picard iteration for the mixed formulation of the Richards equation from [2]. With Δt the time step and for any integer $J > 1$, we define a uniform temporal grid by $\{t^j = j\Delta t, j = 0, \dots, J\}$. The iteration number within a time step is denoted by an integer $k > 0$. For simplicity, we use a simplified notation for $\theta^{j,k} = \theta(p^{j,k})$ and $K^{j,k} = K_s K_{rw}(p^{j,k})$. The backward Euler approximation of Eq. 2.6 is then written as

$$\frac{\theta^{j+1,k+1} - \theta^j}{\Delta t} - \nabla \cdot K^{j+1,k} \nabla p^{j+1,k+1} - \frac{\partial K^{j+1,k}}{\partial z} = f^{j+1}. \tag{4.1}$$

The key idea of the modified Picard iteration is the use of a Taylor expansion for $\theta^{j+1,k+1}$ with respect to p , i.e.,

$$\theta^{j+1,k+1} = \theta^{j+1,k} + \frac{\partial \theta^{j+1,k}}{\partial p} (p^{j+1,k+1} - p^{j+1,k}) + \mathcal{O}(\delta p^2), \tag{4.2}$$

where the derivative $\frac{\partial \theta(p)}{\partial p} = C(p)$ is analytically computed by using Eq. 2.9. By neglecting the higher order terms in Eq. 4.2 and substitution in Eq. 4.1, we get

$$C(p^{j+1,k}) \frac{\delta p^{j+1,k}}{\Delta t} + \frac{\theta^{j+1,k} - \theta^j}{\Delta t} - \nabla \cdot K^{j+1,k} \nabla p^{j+1,k+1} - \frac{\partial K^{j+1,k}}{\partial z} = f^{j+1}, \tag{4.3}$$

with $\delta p^{j+1,k} = p^{j+1,k+1} - p^{j+1,k}$. The above equation can be expressed in the form

$$C(p^{j+1,k}) \frac{\delta p^{j+1,k}}{\Delta t} - \nabla \cdot K^{j+1,k} \nabla \delta p^{j+1,k} = \nabla \cdot K^{j+1,k} \nabla p^{j+1,k} + \frac{\partial K^{j+1,k}}{\partial z} + f^{j+1} - \frac{\theta^{j+1,k} - \theta^j}{\Delta t}. \tag{4.4}$$

The next pressure head iterate is obtained by the update $p^{j+1,k+1} = p^{j+1,k} + \delta p^{j+1,k}$. Notice that the left-hand side of the above equation is the residual associated with the Picard iteration, which should be equal to zero for a converged solution. Therefore, one may use $\|\delta p^{j+1,k}\|_\infty < \varepsilon_{PI}$ as a stopping criterion with $\varepsilon_{PI} > 0$ as the tolerance for Picard iteration. The pressure head at time t^{j+1} is then given by $p^{j+1} = p^{j+1,k+1}$, with k the total number of Picard iterations to converge to ε_{PI} . The iterative scheme (4.4) is a general mixed-formulation Picard iteration, which results in perfect mass balance.

$$\tilde{C} = C(p^{j+1,k}), \quad \tilde{K} = K^{j+1,k} \quad \text{and} \quad \tilde{f} = \nabla \cdot K^{j+1,k} \nabla p^{j+1,k} + \frac{\partial K^{j+1,k}}{\partial z} + f^{j+1} - \frac{\theta^{j+1,k} - \theta^j}{\Delta t},$$

and the unknown $\delta \tilde{p} = \delta p^{j+1,k}$. To discretize the above problem, we use a *cell-centered finite volume scheme* for which the hydraulic conductivity at the cell-face is based on the harmonic averaging of the hydraulic conductivities from the adjacent cells, derived by the continuity of fluxes [40, 41].

$$\mathcal{D}_h = \left\{ (x_{i_1}, z_{i_2}); x_{i_1} = \left(i_1 - \frac{1}{2}\right)h, z_{i_2} = \left(i_2 - \frac{1}{2}\right)h, i_1, i_2 = 1, \dots, M \right\}, \tag{4.6}$$

is considered. For each *interior cell* (edges do not lie on a boundary) with center (x_{i_1}, z_{i_2}) , denoted by $\mathcal{D}_h^{i_1,i_2}$, we obtain a five-point scheme,

$$c_{i_1,i_2}^h \delta \tilde{p}_{i_1,i_2} + w_{i_1,i_2}^h \delta \tilde{p}_{i_1-1,i_2} + e_{i_1,i_2}^h \delta \tilde{p}_{i_1+1,i_2} + s_{i_1,i_2}^h \delta \tilde{p}_{i_1,i_2-1} + n_{i_1,i_2}^h \delta \tilde{p}_{i_1,i_2+1} = \tilde{f}_{i_1,i_2}^h, \tag{4.7}$$

with

$$\begin{aligned} w_{i_1,i_2}^h &= -\frac{2}{h^2} \frac{\tilde{K}_{i_1,i_2} \tilde{K}_{i_1-1,i_2}}{\tilde{K}_{i_1,i_2} + \tilde{K}_{i_1-1,i_2}}, & e_{i_1,i_2}^h &= -\frac{2}{h^2} \frac{\tilde{K}_{i_1,i_2} \tilde{K}_{i_1+1,i_2}}{\tilde{K}_{i_1,i_2} + \tilde{K}_{i_1+1,i_2}}, \\ s_{i_1,i_2}^h &= -\frac{2}{h^2} \frac{\tilde{K}_{i_1,i_2} \tilde{K}_{i_1,i_2-1}}{\tilde{K}_{i_1,i_2} + \tilde{K}_{i_1,i_2-1}}, & n_{i_1,i_2}^h &= -\frac{2}{h^2} \frac{\tilde{K}_{i_1,i_2} \tilde{K}_{i_1,i_2+1}}{\tilde{K}_{i_1,i_2} + \tilde{K}_{i_1,i_2+1}}, \\ c_{i_1,i_2}^h &= -(w_{i_1,i_2}^h + e_{i_1,i_2}^h + n_{i_1,i_2}^h + s_{i_1,i_2}^h) + \frac{\tilde{C}_{i_1,i_2}}{\Delta t}, \end{aligned}$$

where, for instance, \tilde{K}_{i_1,i_2} is the diffusion coefficient associated with cell $\mathcal{D}_h^{i_1,i_2}$ and the source term \tilde{f}_{i_1,i_2}^h is an approximation of \tilde{f} in that cell. This scheme is modified appropriately for cells close to the boundary.

Next, we describe the multigrid method for solving the linear system arising from the above discretization. The multigrid hierarchy is based on uniform grid coarsening, i.e., the cell-width is doubled in each coarsening step in each direction. As the smoothing method, we use the lexicographic Gauss-Seidel iteration, and as the transfer operators between the fine and coarse grids, a simple piecewise constant prolongation operator, P_{2h}^h , is applied and its scaled adjoint is used as the restriction operator R_h^{2h} on the

4.2 Cell-centered multigrid

Focussing on the k -th Picard iteration (4.4) at time t^{j+1} , the following elliptic PDE with variable coefficients is obtained, using simplified notation,

$$\begin{aligned} \frac{\tilde{C}}{\Delta t} \delta \tilde{p} - \nabla \cdot (\tilde{K} \nabla \delta \tilde{p}) &= \tilde{f} \text{ in } \mathcal{D}, \\ \delta \tilde{p} &= 0 \text{ in } \Gamma_D \cup \Gamma_N, \end{aligned} \tag{4.5}$$

with the known quantities

For the discretization of Eq. 4.5, a uniform grid \mathcal{D}_h on a unit square domain with the same mesh width $h = 1/M$, $M \in \mathbb{N}$ in both directions,

cell-centered grid. In classical stencil notation, these are written as

$$P_{2h}^h = \begin{bmatrix} 1 & 1 \\ \star & \\ 1 & 1 \end{bmatrix}_{2h}^h \quad \text{and} \quad R_h^{2h} = \frac{1}{4} \begin{bmatrix} 1 & 1 \\ \star & \\ 1 & 1 \end{bmatrix}^{2h}, \tag{4.8}$$

respectively, where \star denotes the position of the cell center. The coarse grid operator is obtained via a direct discretization of the PDE operator on the coarse grid. For this discretization on a coarser grid, we need to appropriately define the diffusion coefficients on the coarse cell edges. The technique to define the suitable diffusion coefficient on a coarse cell edge is graphically described in Fig. 2 and its caption. In [30], the $W(2, 2)$ -cycle was found to be a very robust and efficient multigrid cycling strategy, and, therefore, we also employ this cycle in our experiments. The number of multigrid iterations is based on the stopping criterion, $\|\mathcal{L}_h \delta \tilde{p}_h - \tilde{f}_h\|_\infty < \varepsilon_{MG}$, where \mathcal{L}_h denotes the linear operator after the discretization of Eq. 4.5 and $\varepsilon_{MG} > 0$.

We consider the modified Picard method in this article as it is widely adopted, although many modifications have been proposed to improve its robustness. For instance,

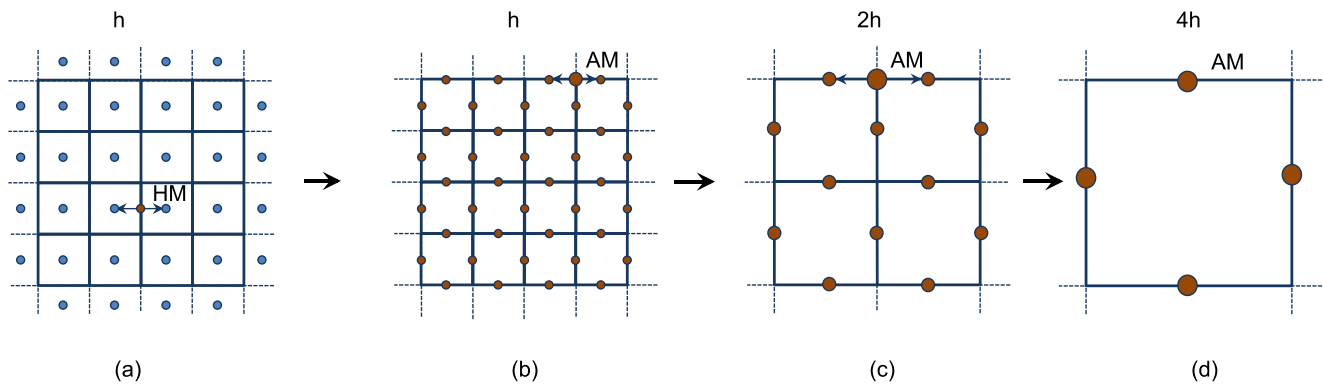


Fig. 2 Schematic representation of coefficient upscaling in the multi-grid hierarchy (h-2h-4h). **a** Coefficient values at cell-centres (blue dots). **b** Values at face-centres (red dots) obtained from the harmonic mean (HM) of coefficients from two adjacent cell-centres. **c** Values at face-centres (bigger red dots) of the 2h-grid are based on arithmetic

mean (AM) of coefficients from face-centres of the h-grid. **d** Values at face-centres (biggest red dots) of the 4h-grid are based on the arithmetic mean of the values of the coefficients from face-centres of the 2h-grid

the authors in [42] studied a spatio-temporal adaptive solution method to improve the numerical stability of the modified Picard iteration. Another interesting improvement was proposed in [43], where an Anderson acceleration was applied to improve the robustness and computational cost for the standard Picard iteration scheme. These improvements can easily be extended to the modified Picard-CCMG solver studied here. Also, there are a number of effective solution approaches based on Newton's method, see, e.g., [44–46]. These methods exhibit a quadratic convergence rate but are very sensitive to initial solution approximations.

4.3 Performance of the modified Picard-CCMG solver

We study the performance of the modified Picard-CCMG solver for a range of values of the parameters α and n and the effect of the heterogeneity of the hydraulic conductivity on the performance of the solver. For this, we consider an *infiltration problem* [46, 47] on a two-dimensional computational domain $\mathcal{D} = (0, 1)^2$. The initial and boundary conditions are prescribed as follows:

$$p(x, z, 0) = -0.4(1 - \exp(-80z)), \quad p(x, 1, t) = -0.4, \quad (4.9)$$

$$p(x, 0, t) = 0.1, \quad \left. \frac{\partial p}{\partial x} \right|_{x=0,1} = 0.$$

Table 2 Set of α, n values used to for benchmarking the modified Picard-CCMG solver

α	n								
0.2	0.4	0.6	0.8	1.0	1.1	1.2	1.3	1.4	1.5
1.2	1.4	1.6	1.8	2.0	1.6	1.7	1.8	1.9	2.0
2.2	2.4	2.6	2.8	3.0	2.2	2.4	2.6	2.8	3.0
3.2	3.4	3.6	3.8	4.0	3.2	3.4	3.6	3.8	4.0

The right-hand side is assumed to be zero, and we consider the a final time $T_{\text{final}} = 0.1$ [h]. In Table 2, we provide a list of 20 values for α and n , used in the experiments. In total, we test 400 pairings of α and n . Parameters $\theta_s = 0.50$ and $\theta_r = 0.05$ are fixed, as they do not pose any problems for the convergence rate of the solver. The samples of hydraulic conductivity are generated according to Eq. 3.1, with $K_s^{(bl)} = 0.2$ [m/h] and the covariance is based on the two Matérn parameters from Table 1.

A similar test was performed in [43] for a deterministic steady-state flow governed by the Richards equation. We perform our experiments in a probabilistic framework. For a given pair α, n , we generate 64 random hydraulic conductivity fields and solve Richards equation with conditions given in Eq. 4.9 for each sample. This is done as the number of multigrid iterations varies depending on the random realization of the hydraulic conductivity field. The cost of solving one instance of stochastic Richards' equation is expressed in terms of the total number of multigrid $W(2, 2)$ -cycles needed to solve the time-dependent problem. Here, by total number of $W(2, 2)$ -cycles means the sum of multigrid iterations needed to reach T_{final} . For all experiments, we set the tolerances $\varepsilon_{PI}, \varepsilon_{MG} = 10^{-5}$. The solution method was terminated with failure when the maximum number of nonlinear iterations (set to 50) was exceeded at any time step.

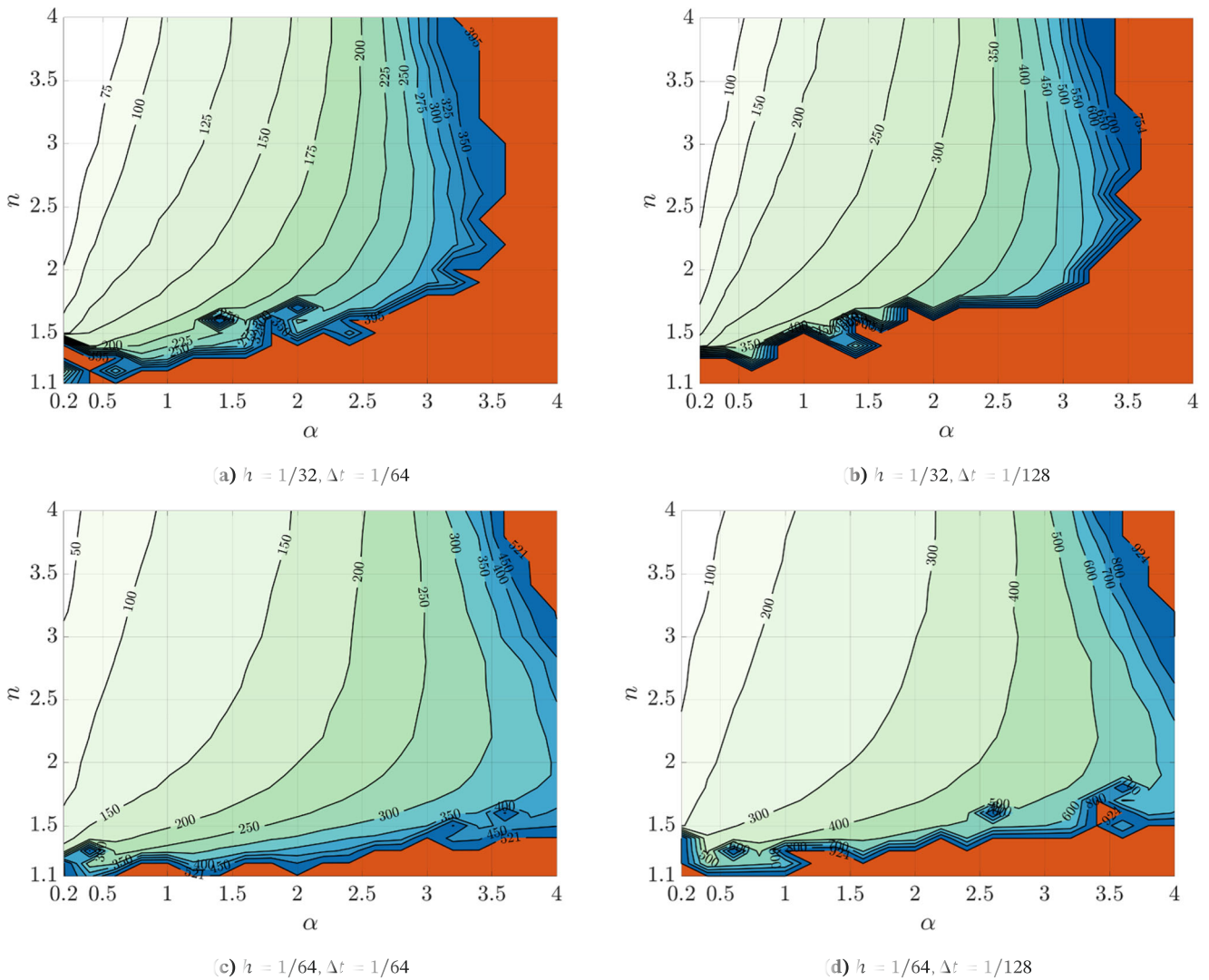


Fig. 3 a–d Contour plots of the average number of multigrid iterations needed to solve the infiltration problem using the modified Picard-CCMG solver for an isotropic hydraulic conductivity field generated using Φ_1

In Figs. 3 and 4, we show the average cost (average of 64 random realizations of the hydraulic conductivity) for solving the stochastic Richards equation for four different combinations of spatial and temporal grid sizes and for two Matérn parameter sets, Φ_1 and Φ_2 , listed in Table 1. The region in red in the figures denotes the (α, n) values for which the modified Picard-CCMG solver failed to converge at least once out of the 64 samples.

Based on numerical experiments, the performance of the modified Picard-CCMG solver for the stochastic Richards equation can be summarized as follows:

- In general, the cost increases by decreasing n and increasing α . The cost of the solver rises steeply for $n < 1.5$ and $\alpha > 3.0$, and the cost increment with respect to the decrease in the value of n is more pronounced compared with the increase in α .

- While a spatio-temporal mesh refinement improves the robustness with respect to α and n , the improvement is less pronounced for n and may require a very fine mesh as $n \rightarrow 1$.
- For a given spatio-temporal mesh, the modified Picard-CCMG solver is less robust and more expensive for anisotropic hydraulic conductivity compared with the isotropic case. A similar (α, n) robustness can be achieved for anisotropic cases by using a sufficiently refined mesh.

The standard deviation contours for the cost show a similar behavior as the average cost contour and we observe a large standard deviation for the cost when $n < 1.5$ and $\alpha > 3.0$. In Fig. 5, we present the number of samples (out of 64 samples), for which the solver did not converge for Φ_1 and Φ_2 . For almost all samples, convergence failed with α close to 4.0 and n close to 1.1.

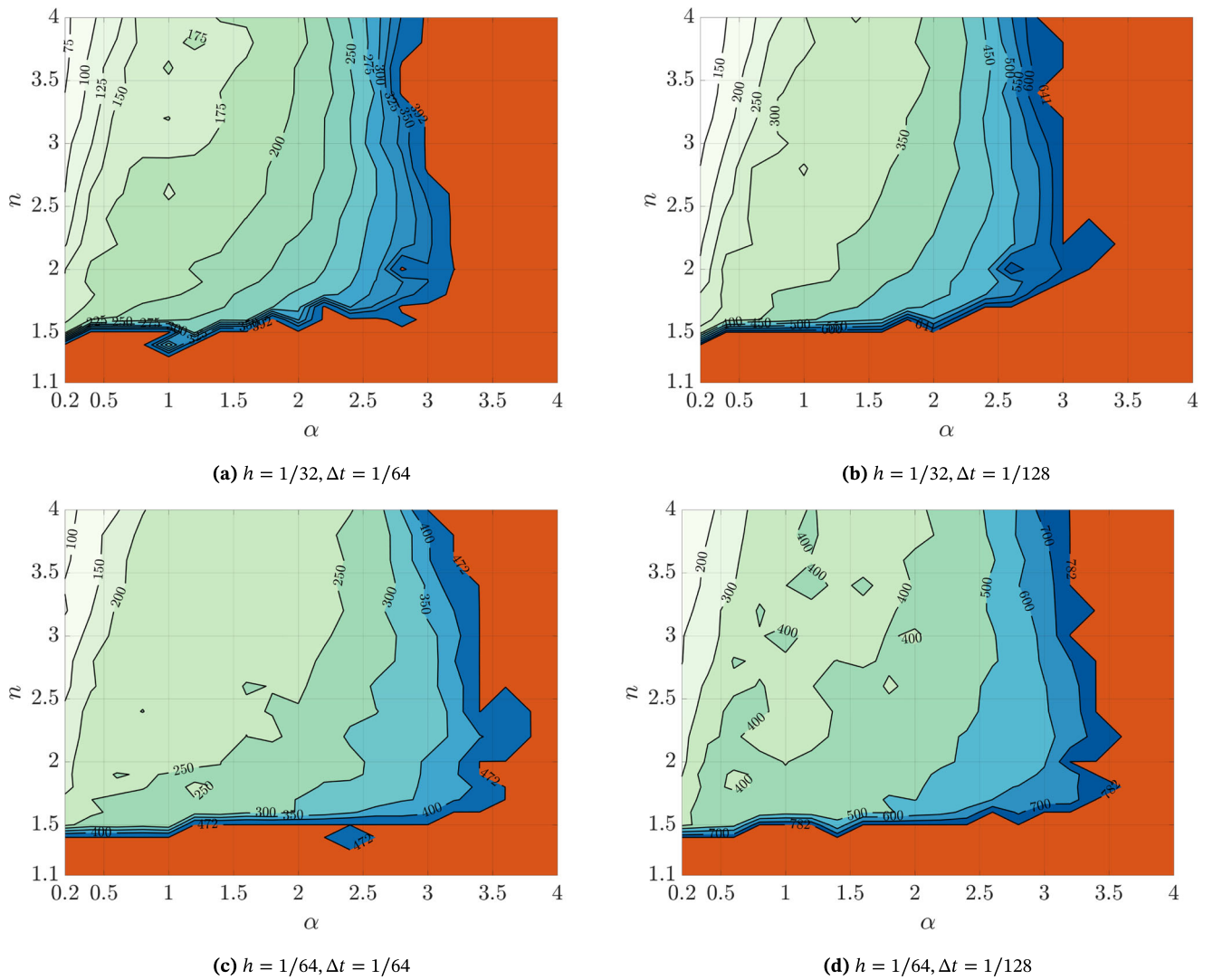


Fig. 4 a–d Contour plots of the average number of multigrid iterations needed to solve Eq. 4.9 using the modified Picard-CCMG solver for an anisotropic hydraulic conductivity field generated using Φ_2

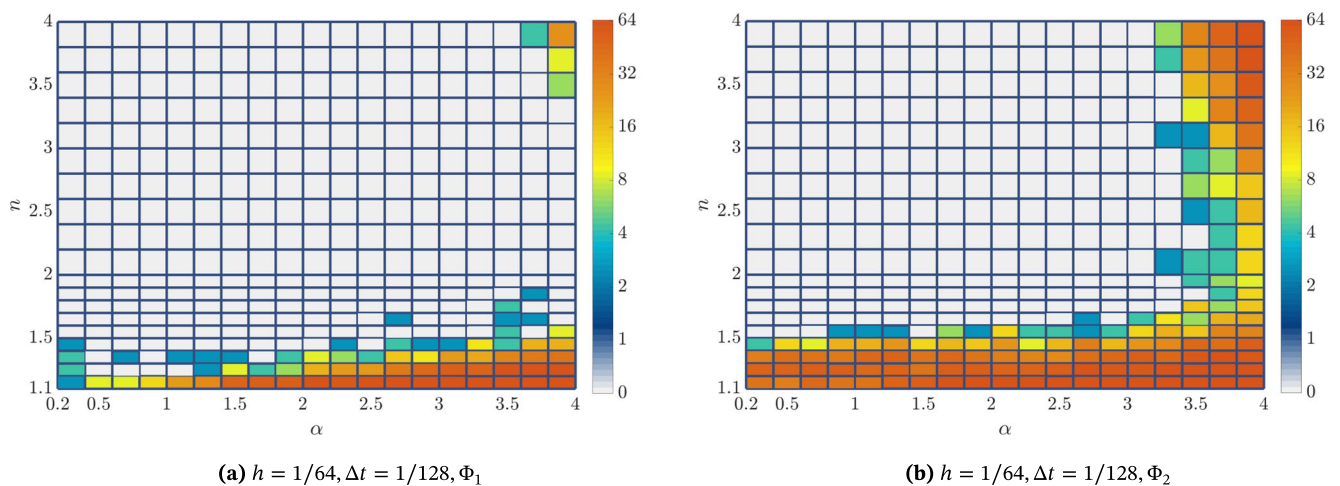


Fig. 5 a, b Counting the number of samples (out of 64), for which the modified Picard-CCMG solver does not converge

A few remarks are in order. We point out that the (α, n) cost map may vary depending on the type of boundary and initial conditions as well as on T_{final} . For instance, in the above experiments, an initially wet profile for the porous media was considered. We expect the performance of the modified Picard-CCMG solver to vary for problems in which infiltration takes place into an initially dry media and the convergence rates may depend on the values of θ_r and θ_s (see, e.g., [48]). Furthermore, the robustness of the solver will also depend on the properties of the hydraulic conductivity field such as on the degree of heterogeneity and anisotropy. These topics will be actively explored in the future work.

5 Multilevel Monte Carlo with parametric continuation

We have observed in the preceding section that the total number of multigrid iterations increases rapidly with a decrease in the value of parameter n and an increase in α . We also noticed that the solver is less robust on a coarse spatio-temporal mesh. Therefore, when using the original MLMC estimator for a “difficult” (α, n) pair, a relatively fine spatio-temporal mesh will be required (and employed), even on the coarsest level of the MLMC hierarchy, resulting in an expensive estimator. To deal with this drawback, we propose an MLMC estimator based on the *parametric continuation* technique. In this approach, we solve the original problem only on the finest level of the MLMC hierarchy and *simplify* the parameter settings dictating the nonlinearities as we work on coarser levels. This allows us to include a comparatively coarser spatio-temporal mesh compared with the original MLMC estimator as simpler problems are solved on coarser levels.

This idea is motivated by continuation-based multigrid solvers for nonlinear boundary value problems [22–24]. In the context of multigrid solvers, continuation is commonly applied in the FMG-FAS (Full MultiGrid-Full Approximation Scheme) algorithm. In these algorithms, the continuation process is integrated in the FMG hierarchy, where the coarse grid solves the simplest problem and is used as a good first approximation for the next grid with a slightly more complicated problem. This process is repeated until the finest grid is reached where the target problem is solved. Although the continuation strategy works well for a large class of nonlinear problems, there is no guarantee that the simpler problem is close enough to the next difficult problem. One can use *bifurcation diagrams* to understand the solution dependence on nonlinearity dictating parameters. These diagrams can also reveal multiple branches and bifurcation points, where the solution differs greatly even if there is a slight perturbation in the parameter value. In such

cases, an *arclength* procedure [25] can be applied to determine the appropriate perturbation size.

5.1 MLMC estimator

To explain the MLMC estimator, we consider the pressure head field at some final time T_{final} as the QoI. Further, we define a spatio-temporal hierarchy of grid levels $\{\mathcal{D}_\ell, \mathcal{T}_\ell\}_{\ell=0}^L$ using

$$h_\ell = \Delta t_\ell = \mathcal{O}(s^{-\ell} h_0), \tag{5.1}$$

where h_0 is the cell-width on the coarsest mesh \mathcal{D}_0 and $s > 0$ represents a grid refinement factor. We further define a hierarchy of parameter sets, $\{\Theta_\ell\}_{\ell=0}^L$, where Θ_L is the parameter set corresponding to the target (strongly nonlinear) problem to be solved. For instance, we can define a parametric hierarchy using the set of van Genuchten parameters, e.g., $\Theta_\ell = \{\alpha_\ell^{(bl)}, n_\ell^{(bl)}\}$. The approximation of the pressure head on the level ℓ at T_{final} is denoted by p_{h_ℓ, Θ_ℓ} . Using the linearity of the expectation operator, one can define the expected value of the pressure head on the finest level, L , with the original parameter set, Θ_L , by the following telescopic sum:

$$\mathbb{E}[p_{h_L, \Theta_L}] = \mathbb{E}[p_{h_0, \Theta_0}] + \sum_{\ell=1}^L \mathbb{E}[p_{h_\ell, \Theta_\ell} - p_{h_{\ell-1}, \Theta_{\ell-1}}]. \tag{5.2}$$

Note that for $\Theta_0 = \Theta_1 = \dots = \Theta_L$, we have the standard MLMC estimator which solves the same problem on all levels. In terms of computational effort, it is cheaper to approximate $\mathbb{E}[p_{h_0, \Theta_0}]$ by a standard Monte Carlo estimator, as the samples are computed on a coarse spatio-temporal mesh based on an “easy nonlinear parameter set Θ_0 .” Furthermore, the correction term, $\mathbb{E}[p_{h_\ell, \Theta_\ell} - p_{h_{\ell-1}, \Theta_{\ell-1}}]$, can be accurately determined using only a few samples as the level-dependent variance, $\mathcal{V}_\ell := \mathbb{V}[p_{h_\ell, \Theta_\ell} - p_{h_{\ell-1}, \Theta_{\ell-1}}]$, is typically small, since the random variables p_{h_ℓ, Θ_ℓ} and $p_{h_{\ell-1}, \Theta_{\ell-1}}$ are positively correlated. Note that the correlation will depend on the grid parameters h_ℓ and $h_{\ell-1}$ as well as on the difference between the nonlinear parameters Θ_ℓ and $\Theta_{\ell-1}$. We will elaborate on this later on.

Each of the expectations in the MLMC estimator Eq. 5.2 can be independently computed using the standard MC simulation. We define a multilevel estimator, $\mathcal{E}_L^{ML}[p_{h_L, \Theta_L}]$, constructed using a sum of $L + 1$ MC estimators:

$$\mathbb{E}[p_{h_L, \Theta_L}] \approx \mathcal{E}_L^{ML}[p_{h_L, \Theta_L}] := \sum_{\ell=0}^L \mathcal{E}_{N_\ell}^{MC}[p_{h_\ell, \Theta_\ell} - p_{h_{\ell-1}, \Theta_{\ell-1}}], \tag{5.3}$$

where $\mathcal{E}_{N_\ell}^{MC}[p_{h_\ell, \Theta_\ell} - p_{h_{\ell-1}, \Theta_{\ell-1}}]$ is the standard MC estimator obtained by averaging N_ℓ independent, identically distributed (i.i.d.) samples as

$$\mathcal{E}_{N_\ell}^{MC}[p_{h_\ell, \Theta_\ell} - p_{h_{\ell-1}, \Theta_{\ell-1}}] := \left(\frac{1}{N_\ell} \sum_{i=1}^{N_\ell} (p_{h_\ell, \Theta_\ell}(\omega_i) - p_{h_{\ell-1}, \Theta_{\ell-1}}(\omega_i)) \right). \tag{5.4}$$

with ω_i denoting an event in the stochastic domain Ω and $p_{h_{\ell-1}, \Theta_{\ell-1}} = 0$. It is expected that the number of MLMC samples $N_\ell \in \mathbb{N}$ forms a decreasing sequence for increasing ℓ . In order to keep the variance of the correction terms small, the MC samples, $p_{h_\ell, \Theta_\ell}(\omega_i) - p_{h_{\ell-1}, \Theta_{\ell-1}}(\omega_i)$, should be based on the same random input ω_i for the simulation on two consecutive levels ℓ and $\ell - 1$.

Remark 5.1 In this work, we do not explore the standard MC method to compute $\mathbb{E}[p_{h_L, \Theta_L}]$ as the computational cost of standard MC will be more than or equal (when only one level is possible) to the standard MLMC estimator. In Section 6, we will only compare the standard and parametric continuation versions of the MLMC estimator.

Similarly, a multilevel estimator for the variance of the pressure head, $\mathbb{V}[p_{h_L, \Theta_L}]$, can be defined as

$$\mathbb{V}[p_{h_L, \Theta_L}] \approx \mathcal{V}_L^{ML}[p_{h_L, \Theta_L}] := \sum_{\ell=0}^L \mathcal{V}_{N_\ell}^{MC}[p_{h_\ell, \Theta_\ell}] - \mathcal{V}_{N_\ell}^{MC}[p_{h_{\ell-1}, \Theta_{\ell-1}}], \tag{5.5}$$

where the variance $\mathcal{V}_{N_\ell}^{MC}[p_{h_\ell, \Theta_\ell}]$ is computed as

$$\mathcal{V}_{N_\ell}^{MC}[p_{h_\ell, \Theta_\ell}] \approx \frac{1}{N_\ell - 1} \sum_{i=1}^{N_\ell} (p_{h_\ell, \Theta_\ell}(\omega_i) - \mathcal{E}_{N_\ell}^{MC}[p_{h_\ell, \Theta_\ell}])^2. \tag{5.6}$$

Again, the computational savings for the variance estimator Eq. 5.5 are obtained by computing individual variances $\mathcal{V}_{N_\ell}^{MC}[p_{h_\ell, \Theta_\ell}]$ and $\mathcal{V}_{N_\ell}^{MC}[p_{h_{\ell-1}, \Theta_{\ell-1}}]$ using the same random

inputs $\{\omega_i\}_{i=1}^{N_\ell}$. The above variance estimator can be seen as an extension of the standard multilevel variance estimator proposed in [49].

For the multilevel estimators, an appropriate spatial interpolation procedure is required to combine expectations from all levels. Typically, the polynomial order of the interpolation scheme should be equal to or higher than the order of the discretization to avoid any additional dominant source of error. In some more detail, when using the estimator Eq. 5.3 to compute $\mathcal{E}_L^{ML}[p_{h_L, \Theta_L}]$, we begin by computing $\mathcal{E}_{N_0}^{MC}[p_{h_0, \Theta_0}]$ on the coarsest grid \mathcal{D}_0 . This quantity is then interpolated to the next finer grid \mathcal{D}_1 and is added to the correction term $\mathcal{E}_{N_1}^{MC}[p_{h_1, \Theta_1} - p_{h_0, \Theta_0}]$ resulting in a two-level estimate $\mathcal{E}_1^{ML}[p_{h_1, \Theta_1}]$. This is again interpolated to the next grid level \mathcal{D}_2 and added to the next correction term $\mathcal{E}_{N_2}^{MC}[p_{h_2, \Theta_2} - p_{h_1, \Theta_1}]$. This process is repeated until the final level is reached.

5.2 Accuracy of MLMC estimator

Throughout this paper, we use the L^2 -based norm for the error analysis of the multilevel Monte Carlo estimator. We assume that the pressure considered belongs to the functional space $L^2(\Omega, \mathcal{D})$ corresponding to the space of square-integrable measurable functions $p : \Omega \rightarrow L^2(\mathcal{D})$ for a previously defined probability space $(\Omega, \mathbb{F}, \mathbb{P})$. These spaces are equipped with the norm

$$\|p(\mathbf{x}, T, \omega)\|_{L^2(\Omega, \mathcal{D})} := \mathbb{E} \left[\|p(\mathbf{x}, T, \omega)\|_{L^2(\mathcal{D})}^2 \right]^{\frac{1}{2}} = \left(\int_{\Omega} \|p(\mathbf{x}, T, \omega)\|_{L^2(\mathcal{D})}^2 d\mathbb{P} \right)^{\frac{1}{2}}. \tag{5.7}$$

The mean-square error (MSE) in $\mathcal{E}_L^{ML}[p_{h_L, \Theta_L}]$ can then be expressed as the sum of the discretization and the sampling errors as

$$\left\| \mathbb{E}[p_{\Theta_L}] - \mathcal{E}_L^{ML}[p_{h_L, \Theta_L}] \right\|_{L^2(\Omega, \mathcal{D})}^2 \leq \left\| \mathbb{E}[p_{\Theta_L}] - \mathbb{E}[p_{h_L, \Theta_L}] \right\|_{L^2(\mathcal{D})}^2 + \left\| \mathbb{E}[p_{h_L, \Theta_L}] - \mathcal{E}_L^{ML}[p_{h_L, \Theta_L}] \right\|_{L^2(\Omega, \mathcal{D})}^2. \tag{5.8}$$

Both errors in the MLMC estimator can be dealt with separately. The discretization error can be quantified as

$$\left\| \mathbb{E}[p_{\Theta_L}] - \mathbb{E}[p_{h_L, \Theta_L}] \right\|_{L^2(\mathcal{D})} \leq c_0 h_L^a, \quad a > 0, \tag{5.9}$$

where c_0 is a constant independent of h_L but depending on the parameter set Θ_L . The rate a typically depends on the regularity of the PDE and the accuracy of the discretization. The next task is to bound the sampling errors. As the MLMC estimator $\mathcal{E}_L^{ML}[p_{h_L, \Theta_L}]$ is composed of $L + 1$ independent MC estimators, the sampling error in the MLMC estimator

is just the sum of sampling errors from the individual MC estimators. Therefore,

$$\left\| \mathbb{E}[p_{h_L, \Theta_L}] - \mathcal{E}_L^{ML}[p_{h_L, \Theta_L}] \right\|_{L^2(\Omega, \mathcal{D})}^2 = \sum_{\ell=0}^L \frac{\|\mathcal{V}_\ell\|_{L^2(\mathcal{D})}}{N_\ell}, \tag{5.10}$$

see [50, 51] for a proof. Obtaining a bound on the level-variance $\|\mathcal{V}_\ell\|_{L^2(\mathcal{D})}$ is more involved due to its dependence on the grid size h_ℓ as well as on the nonlinearity parameter set Θ_ℓ . We numerically estimate it by

$$\begin{aligned} \|\mathcal{V}_\ell\|_{L^2(\mathcal{D})} &= \|\mathbb{V}[p_{h_\ell, \Theta_\ell} - p_{h_{\ell-1}, \Theta_{\ell-1}}]\|_{L^2(\mathcal{D})} \\ &\approx \frac{1}{N_\ell - 1} \sum_{i=1}^{N_\ell} \int_{\mathcal{D}} \left(\mathcal{E}_{N_\ell}^{MC}[p_{h_\ell, \Theta_\ell} - p_{h_{\ell-1}, \Theta_{\ell-1}}] - (p_{h_\ell, \Theta_\ell}(\omega_i) - p_{h_{\ell-1}, \Theta_{\ell-1}}(\omega_i)) \right)^2. \end{aligned} \tag{5.11}$$

To achieve a tolerance of ε , one needs to ensure that

$$\left\| \mathbb{E}[p_{\Theta_L}] - \mathcal{E}_L^{ML}[p_{h_L, \Theta_L}] \right\|_{L^2(\Omega, \mathcal{D})}^2 \leq (c_0 h_L^a)^2 + \sum_{\ell=0}^L \frac{\|\mathcal{V}_\ell\|_{L^2(\mathcal{D})}}{N_\ell} < \varepsilon^2. \tag{5.12}$$

The total cost of the MLMC estimator can be expressed as $\mathcal{W}_L^{ML} = \sum_{\ell=0}^L N_\ell \mathcal{W}_\ell$, where $\mathcal{W}_\ell = \mathcal{O}(h_\ell^{-\gamma})$ corresponds to the cost of computing one sample on level ℓ . For time-dependent problems, the rate $\gamma \geq d + 1$, with d the number of spatial dimensions. As proposed in [18, 19], the number of samples at different levels is typically derived by minimizing the total cost such that the sampling error of the MLMC estimator reduces below ε^2 , i.e.,

$$\min \left(\sum_{\ell=0}^L N_\ell \mathcal{W}_\ell \right) \quad \text{s.t.} \quad \sum_{\ell=0}^L \frac{\|\mathcal{V}_\ell\|_{L^2(\mathcal{D})}}{N_\ell} = \varepsilon^2. \tag{5.13}$$

Using the standard Lagrange multiplier approach [18], gives us

$$N_\ell = \varepsilon^{-2} \left(\sum_{\ell=0}^L \sqrt{\|\mathcal{V}_\ell\|_{L^2(\mathcal{D})} \mathcal{W}_\ell} \right) \sqrt{\frac{\|\mathcal{V}_\ell\|_{L^2(\mathcal{D})}}{\mathcal{W}_\ell}}, \tag{5.14}$$

and hence the total cost to obtain a tolerance of ε is given by

$$\mathcal{W}_L^{ML}(\varepsilon) = \sum_{\ell=0}^L N_\ell \mathcal{W}_\ell = \varepsilon^{-2} \left(\sum_{\ell=0}^L \sqrt{\|\mathcal{V}_\ell\|_{L^2(\mathcal{D})} \mathcal{W}_\ell} \right)^2. \tag{5.15}$$

In the above formula, the product $\|\mathcal{V}_\ell\|_{L^2(\mathcal{D})} \mathcal{W}_\ell$ determines the cost contribution from any level ℓ . For instance, if the product decays with increasing ℓ , the dominant cost comes from the coarsest level whereas if the product grows with ℓ , the dominant contribution comes from the finest level.

Remark 5.2 The optimal number of samples given in Eq. 5.14 is based on a pre-defined hierarchy of parameters $\{\Theta_\ell\}_{\ell=0}^L$. A more general approach is to find N_ℓ along with the parameter set $\{\Theta_\ell\}_{\ell=0}^L$ for which the total cost of the MLMC estimator is minimum. Solving such optimization problem analytically is non-trivial. Furthermore, numerically obtaining the best values for Θ_ℓ can also be highly expensive. In Section 6, we will discuss some heuristics that can be applied to find Θ_ℓ .

5.2.1 MLMC algorithm with parametric continuation

To compute the estimator $\mathcal{E}_L^{ML}[p_{h_L, \Theta_L}]$, the standard MLMC algorithm from [18, 19] cannot be directly employed as it requires solving the same problem on all grid levels. Here, we describe a *modified version* of the standard MLMC technique to compute $\mathcal{E}_L^{ML}[p_{h_L, \Theta_L}]$. This algorithm assumes that the total number of levels in the MLMC hierarchy and the values of the nonlinearity parameters Θ_ℓ for all levels are known in advance. The algorithm can be described by the following steps:

Algorithm 1 PC_MLMC algorithm.

- 1: Fix the tolerance ε , \mathcal{D}_ℓ , \mathcal{T}_ℓ , Θ_ℓ and warm-up samples N_ℓ^* for $\ell = 0, 1, 2, \dots, L$.
- 2: Compute quantities $\mathcal{E}_{N_\ell}^{MC}[p_{h_\ell, \Theta_\ell} - p_{h_{\ell-1}, \Theta_{\ell-1}}]$ and $\|\mathcal{V}_\ell\|_{L^2(\mathcal{D})}$ using samples $N_\ell = N_\ell^*$ for all levels.
- 3: Update N_ℓ using the formula (5.14) for all levels.
- 4: Compute additional samples and update $\mathcal{E}_{N_\ell}^{MC}[p_{h_\ell, \Theta_\ell} - p_{h_{\ell-1}, \Theta_{\ell-1}}]$ and $\|\mathcal{V}_\ell\|_{L^2(\mathcal{D})}$ for all levels.
- 5: Perform steps 3-4 until no additional samples are needed on any level.

In the above algorithm, setting a smaller value of N_ℓ^* may lead to an abrupt termination of the algorithm even before the sum sampling error Eq. 5.10 has reduced to the set tolerance. Also, the value of N_ℓ^* should not be set too high, especially not for the finest level, to avoid oversampling. The cost per sample \mathcal{W}_ℓ can also be estimated “on-the-fly” by averaging the CPU times from the computation of warm-up samples.

6 Numerical experiments

We evaluate the performance of the new MLMC estimator and study the improvements with respect to the standard MLMC estimator. For all the experiments, we use the infiltration problem with conditions given in Eq. 4.9, however, with $T_{\text{final}} = 0.2$ [h] (in hours) and the two Matérn covariance parameters from Table 1. We employ a geometric hierarchy of spatio-temporal grids with refinement factor $s = 2$ in Eq. 5.1 and we use $h_\ell = \Delta t_\ell$. For all experiments, the following baseline values are prescribed, $K_s^{(bl)} = 0.2$ [m/h] (in meters/hour), $\theta_s^{(bl)} = 0.5$ and $\theta_r^{(bl)} = 0.05$; different baseline values for $\alpha^{(bl)}$ and $n^{(bl)}$ are studied. The uncertainty in the soil parameters is defined according to the values presented in Table 3. The sampling and upscaling procedure for a Gaussian random field is described in Appendix. The sampling of random fields with uniform marginal is described in Section 3.

Table 3 Description of uncertainty for different soil parameters

Quantity	Uncertainty
$Z(\mathbf{x}, \omega)$	$\mathcal{N}(0, C_\Phi)$
$\varepsilon_{\theta_s}(\mathbf{x}, \omega)$	$\mathcal{U}(-0.05, 0.05, C_\Phi)$
$\varepsilon_{\theta_r}(\mathbf{x}, \omega)$	$\mathcal{U}(-0.005, 0.005, C_\Phi)$
$\varepsilon_\alpha(\mathbf{x}, \omega)$	$\mathcal{U}(-0.2, 0.2, C_\Phi)$
$\varepsilon_n(\mathbf{x}, \omega)$	$\mathcal{U}(-0.05, 0.05, C_\Phi)$

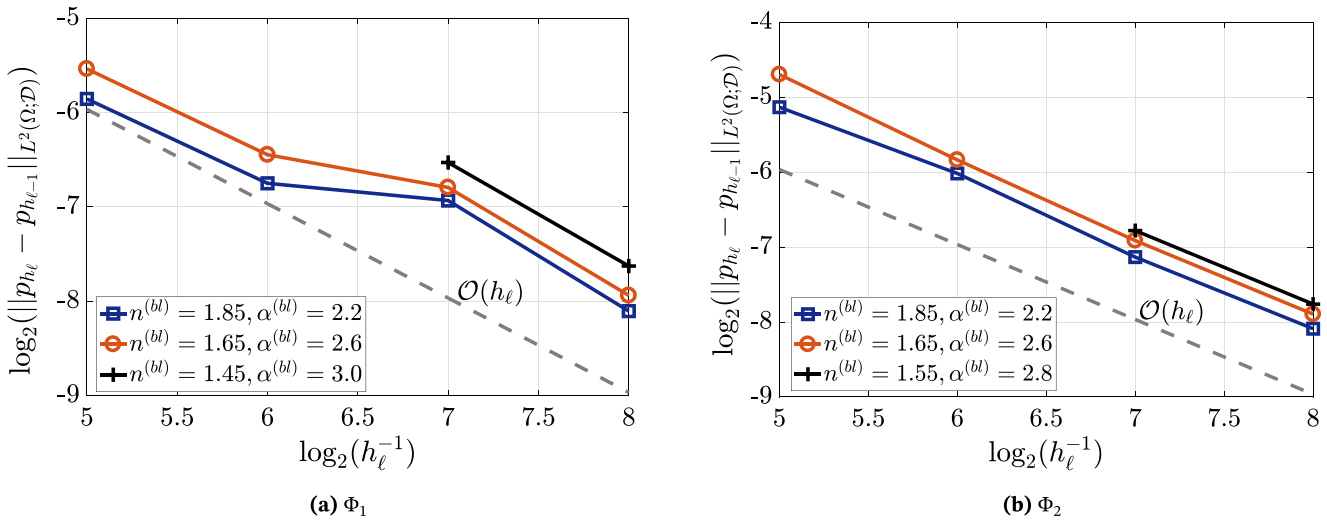


Fig. 6 a, b Convergence of discretization error with respect to mesh refinement for different baseline values of $\alpha^{(bl)}$ and $n^{(bl)}$

Note that the above stochastic model is extremely high-dimensional as it comprises five independent random fields. For each random field, the degree of freedom is equal to the number of grid points in the sampling mesh. The dimensionality can be reduced using the KL-expansion method Eq. 3.5; however as we use random fields with small correlation lengths, we will still need to use a very large number of KL-modes for an accurate representation of these random fields.

6.1 Convergence of discretization bias

We begin by analyzing the reduction of the discretization error $\|p_{h_\ell} - p_{h_{\ell-1}}\|_{L^2(\Omega; \mathcal{D})}$ with respect to mesh refinement for different baseline values of $\alpha^{(bl)}$ and $n^{(bl)}$. The relative error is used to bound the exact discretization bias as

$$\|p - p_{h_\ell}\|_{L^2(\Omega; \mathcal{D})} \leq \frac{\|p_{h_\ell} - p_{h_{\ell-1}}\|_{L^2(\Omega; \mathcal{D})}}{s^a - 1}, \tag{6.1}$$

where a is the convergence rate defined in Eq. 5.9. The relative errors for Φ_1 and Φ_2 are presented in the left and right pictures in Fig. 6, respectively. For both cases, a convergence rate close to first-order is observed, i.e., $a \approx 1$. The convergence rate typically depends on the order of the spatio-temporal discretization scheme as well as on the smoothness parameter ν_c in the covariance function. In fact, the dominant error comes from the first-order accurate backward Euler time discretization. The magnitude of the error grows with increasing $\alpha^{(bl)}$ and reduces with increasing $n^{(bl)}$ values. Note that for the most difficult cases, $n^{(bl)} = 1.45, \alpha^{(bl)} = 3.0$ for Φ_1 and $n^{(bl)} = 1.55, \alpha^{(bl)} = 2.8$ for Φ_2 , the convergent solutions are obtained from $h_\ell = 1/64$ onwards.

6.2 MLMC simulation

Here, we describe the algorithm to compute the multilevel estimator $\mathcal{E}_L^{ML}[p_{h_L}, \Theta_L]$. We perform the MLMC simulations

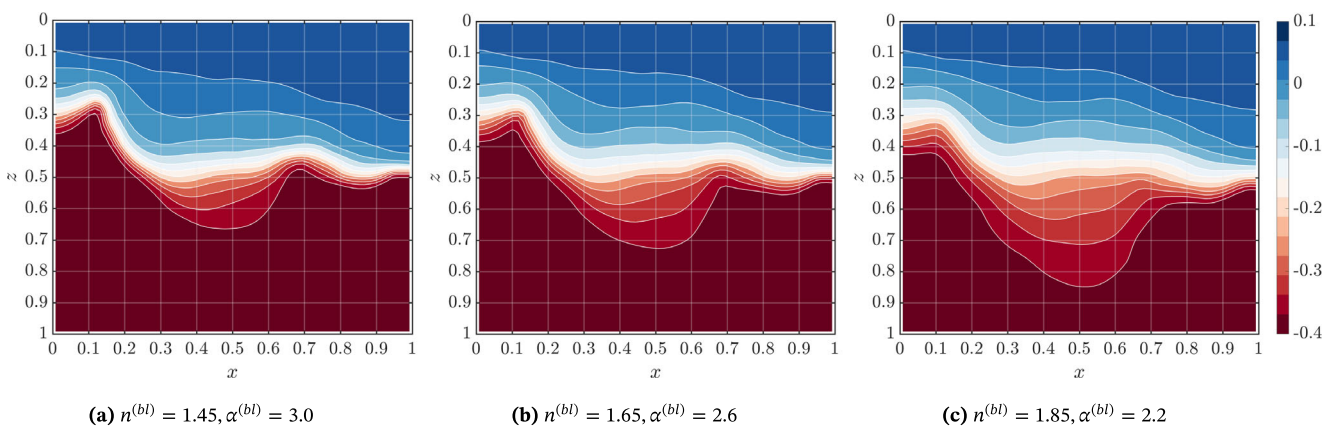


Fig. 7 a–c Comparison of pressure head fields at $T_{\text{final}} = 0.2$ [h] for different baseline values of the parameters, $(n^{(bl)}, \alpha^{(bl)})$ but with the same random fields $Z, \varepsilon_{\theta_s}, \varepsilon_{\theta_r}, \varepsilon_\alpha, \varepsilon_n$. Solutions are based on $h = \Delta t = 1/64$ and the Matérn parameter set Φ_1

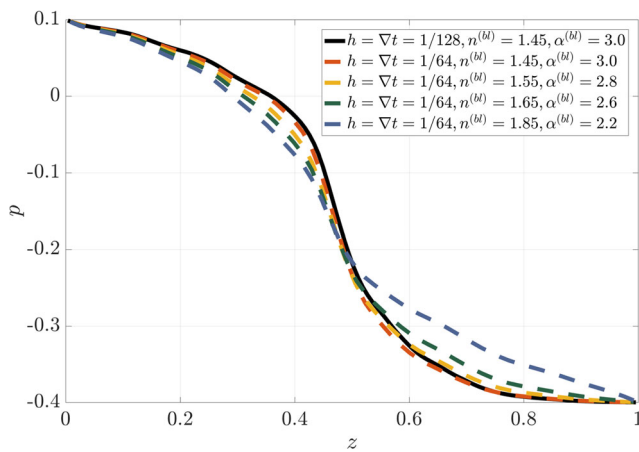


Fig. 8 Comparison of cross sections of the pressure heads from Fig. 7 at $x = 0.5$

for two test cases based on Φ_1 and Φ_2 , respectively. The original problem for Φ_1 uses $\Theta_L = \{\alpha_L^{(bl)}, n_L^{(bl)}\} = \{3.0, 1.45\}$ and for Φ_2 , the original problem is based on $\Theta_L = \{2.8, 1.55\}$.

We first investigate the correlations for the pressure head profiles when the baseline values for $\alpha^{(bl)}$ and $n^{(bl)}$ are varied however employing the same random fields. In Fig. 7, we compare three pressure head solutions with different baseline values, and with the same random fields, $Z(\mathbf{x}, \omega)$, $\varepsilon_{\theta_s}(\mathbf{x}, \omega)$, $\varepsilon_{\theta_r}(\mathbf{x}, \omega)$, $\varepsilon_{\alpha}(\mathbf{x}, \omega)$, and $\varepsilon_n(\mathbf{x}, \omega)$ (see Section 3). Clearly, the pressure head profile becomes more *diffusive* when “easier” parameters are prescribed. We also compare the cross sections of the pressure head profiles at $x = 0.5$ in Fig. 8. For reference, we use the solution on the fine grid $h = \Delta t = 1/128$ (black solid line) and compare it with different pairs of $n^{(bl)}$ and $\alpha^{(bl)}$ values on the next coarse grid $h = \Delta t = 1/64$. The profiles with the same $(n^{(bl)}, \alpha^{(bl)})$ values are very close and the deviation increases as the two parameters are set to “easier” values. Thus, we can conclude that the correlation decays as the difference between the baseline values of the nonlinear parameters widens.

Next, we study the behavior of the level-dependent variance $\|\mathcal{V}_\ell\|_{L^2(\mathcal{D})}$ when using the parametric continuation approach. For this we define the so-called *parametric continuation variables*, $\nabla\alpha = \alpha_\ell - \alpha_{\ell-1}$ and $\nabla n = n_{\ell-1} - n_\ell$, with the purpose to reduce the nonlinearity when processing coarse grids. In Fig. 9, we plot $\|\mathcal{V}_\ell\|_{L^2(\mathcal{D})}$ computed using Eq. 5.11 for different $(\nabla\alpha, \nabla n)$ pairs for the two Matérn parameter sets. The original problem is solved with $h_L = 1/256$. The parameter sets Θ_L and Θ_ℓ , for $\ell = L - 1, L - 2, \dots, 0$, are obtained by employing $\nabla\alpha$ and ∇n . The black line represents the variance when same problem is solved on all levels, i.e., $\nabla\alpha = \nabla n = 0$, corresponding to the original MLMC estimator. Using the original approach, we can only process three levels in the MLMC hierarchy. The red and blue lines in the figure correspond to the

variance which is computed using $\nabla\alpha = 0.05, \nabla n = 0.1$ and $\nabla\alpha = 0.1, \nabla n = 0.2$, respectively. For these two cases, we can incorporate a larger number of coarse levels, up to $h_0 = 1/16$, as milder nonlinear problems are solved on these coarse levels. Furthermore, for levels $\ell < L$, the variance is smaller, compared with the case without continuation (where $\nabla\alpha = \nabla n = 0$) which will result in lower number of samples on these levels. Here, we wish to highlight the fact that choosing optimal values for $\nabla\alpha$ and ∇n is important. For example, when $\alpha = 0.1, \nabla n = 0.2$, the variance on the finest level increases as compared with the variance found with the original MLMC approach. Due to this, an increasing number of samples will be needed on the finest level, compared with the original MLMC estimator, which is undesirable as it may result in an expensive estimator. On the other hand, for a smaller perturbation $\nabla\alpha = 0.05, \nabla n = 0.1$, the magnitude of the variance on the finest level is similar to that of the original MLMC estimator. Therefore, the number of samples on the finest level will be more-or-less similar to the original MLMC estimator.

Alternatively, one can avoid high variance on the finest level by using zero perturbations on the two finest levels, i.e., $\nabla\alpha_L = \nabla\alpha_{L-1} = 0$ and $\nabla n_L = \nabla n_{L-1} = 0$ and choosing non-zero perturbations on next coarser levels from $L - 2$ onwards. This way, we solve the original problem on the two finest levels.

Now we apply the parametric continuation-based MLMC estimator denoted by *PC_MLMC*, to compute the mean and variance of the pressure head field and also perform comparisons with respect to the standard MLMC estimator which is denoted by *Std_MLMC*. For this, we use the two previously discussed test cases: isotropic covariance Φ_1 with baseline values $n^{(bl)} = 1.45, \alpha^{(bl)} = 3.0$ and anisotropic covariance Φ_2 with baseline values $n^{(bl)} = 1.55, \alpha^{(bl)} = 2.8$. For simplicity, we use the continuation variables $\nabla\alpha = 0.05, \nabla n = 0.1$ for both the isotropic and anisotropic cases. We compare the number of samples needed on different grids for three values of the tolerances. For the *Std_MLMC* estimator, the coarsest possible level is $h_0 = \Delta t_0 = 1/64$, whereas for the *PC_MLMC*, we use $h_0 = \Delta t_0 = 1/16$. As the discretization error shows a first-order decay (see Fig. 6), we set the tolerance $\varepsilon = \mathcal{O}(h_L)$. We use Algorithm 1 to reduce the sampling error to ε . For all experiments, warm-up samples are computed as $N_\ell^* = 2^{(L+1-\ell)}$, which results in stable performance. The number of levels is decided based on an error convergence study performed in Section 6.1. In Table 4, the two estimators for the isotropic Matérn parameter Φ_1 are compared. Due to the sample optimization strategy Eq. 5.14, a large number of samples is shifted to coarser grids when using the *PC_MLMC* estimator. Furthermore, a fewer number of samples are required for the *PC_MLMC* estimator compared with the *Std_MLMC*, even on the finest level. This is due the fact that

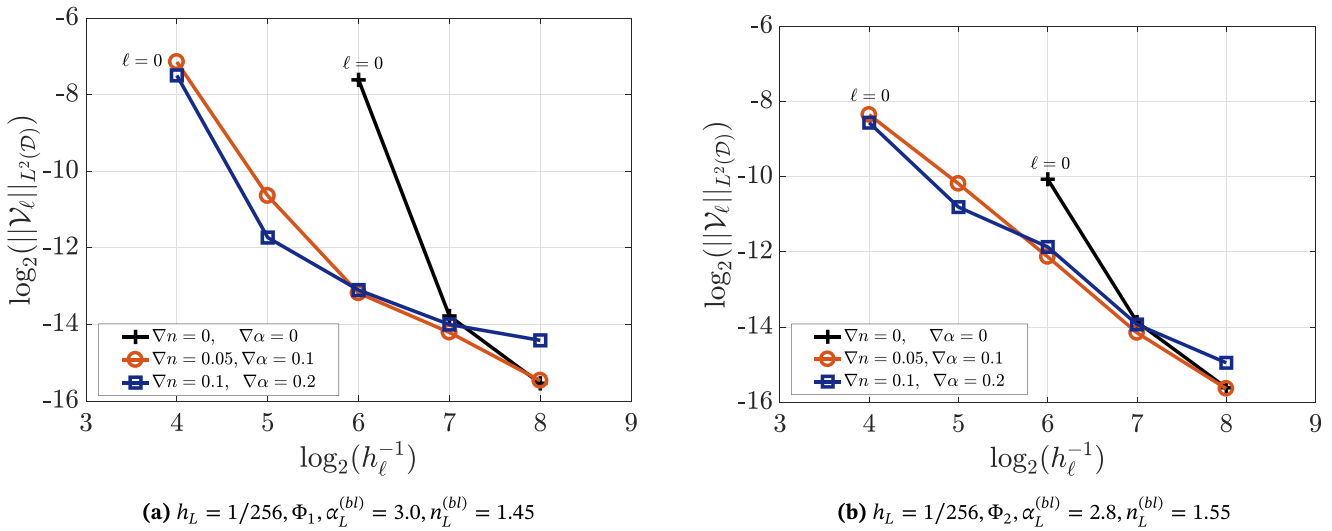


Fig. 9 a, b Comparison of level-dependent variances using different pairs of $\nabla\alpha$, ∇n for the two Matérn parameters

the sum $\sum_{\ell=0}^L \sqrt{\|\mathcal{V}_\ell\|_{L^2(\mathcal{D})}} \mathcal{W}_\ell$ for the *PC_MLMC* estimator is slightly smaller than for *Std_MLMC*. Moreover, a large computational gain is induced by the reduction in the number of samples on grid $h_\ell = 1/64$, for instance, for $\varepsilon = 0.005$, the number of samples reduced from 438 to 35 when using the parametric continuation. In Fig. 10a, the CPU times for the two estimators are also compared. We observe a speed-up of about a factor of three for $\varepsilon = 0.005$.

A similar test is performed for the anisotropic problem. The number of samples for different tolerances are provided in Table 5 and the CPU times in Fig. 10b. Again some improvement in computation times are observed, although the gain is not as pronounced as for the first problem. This is due to the fact that the second case uses simpler baseline values $n^{(bl)} = 1.55, \alpha^{(bl)} = 2.8$ and the cost reduction with parameter simplification is not very rapid. For the isotropic case with $n^{(bl)} = 1.45, \alpha^{(bl)} = 3.0$, the cost decay is more rapid with parameter simplification. This is more evident from the cost map in Fig. 4, where we see more dense contour lines around $n^{(bl)} = 1.45$. Therefore, the parametric

continuation approach is very effective when a strongly nonlinear stochastic problem needs to be solved.

We point out that the parametric continuation can also be very effective in the case of smooth anisotropic random fields in combination with a high-order discretization. Typically, a high-order discretization can lead to a rapid decay of level-dependent variance. Furthermore, if the variance decay is faster than the per sample cost growth, we obtain a multilevel estimator with dominant cost coming from the coarsest level, see [18, 19]. In this scenario, *PC_MLMC* will be highly effective as it allows to incorporate very coarse levels. In the article, we only use a low-regularity anisotropic random field (smoothness $\nu_c = 0.5$, see Table 1) for which a first-order scheme (a combination of Picard iteration and backward Euler time stepping) is already sufficient.

We also wish to highlight the fact that both MLMC estimators are *optimal* since the cost scales as $\mathcal{O}(\varepsilon^{-3})$, which is similar to the computational complexity of solving one deterministic problem on the finest grid, i.e., $\mathcal{O}(h_L^{-3})$ and $h_L = \mathcal{O}(\varepsilon)$.

Table 4 Comparison of number of samples needed to achieve tolerances ε using the standard MLMC (*Std_MLMC*) and parametric continuation MLMC (*PC_MLMC*) estimators for $\Phi_1, n_L^{(bl)} = 1.45, \alpha_L^{(bl)} = 3.0$. Entries with symbol (–) indicate zero samples needed for that grid

h_ℓ	$N_\ell(h_L = 1/64, \varepsilon = 0.02)$		$N_\ell(h_L = 1/128, \varepsilon = 0.01)$		$N_\ell(h_L = 1/256, \varepsilon = 0.005)$	
	<i>Std_MLMC</i>	<i>PC_MLMC</i>	<i>Std_MLMC</i>	<i>PC_MLMC</i>	<i>Std_MLMC</i>	<i>PC_MLMC</i>
1/16	–	115	–	459	–	1833
1/32	–	11	–	44	–	176
1/64	28	3	110	9	438	35
1/128	–	–	5	2	18	8
1/256	–	–	–	–	4	3

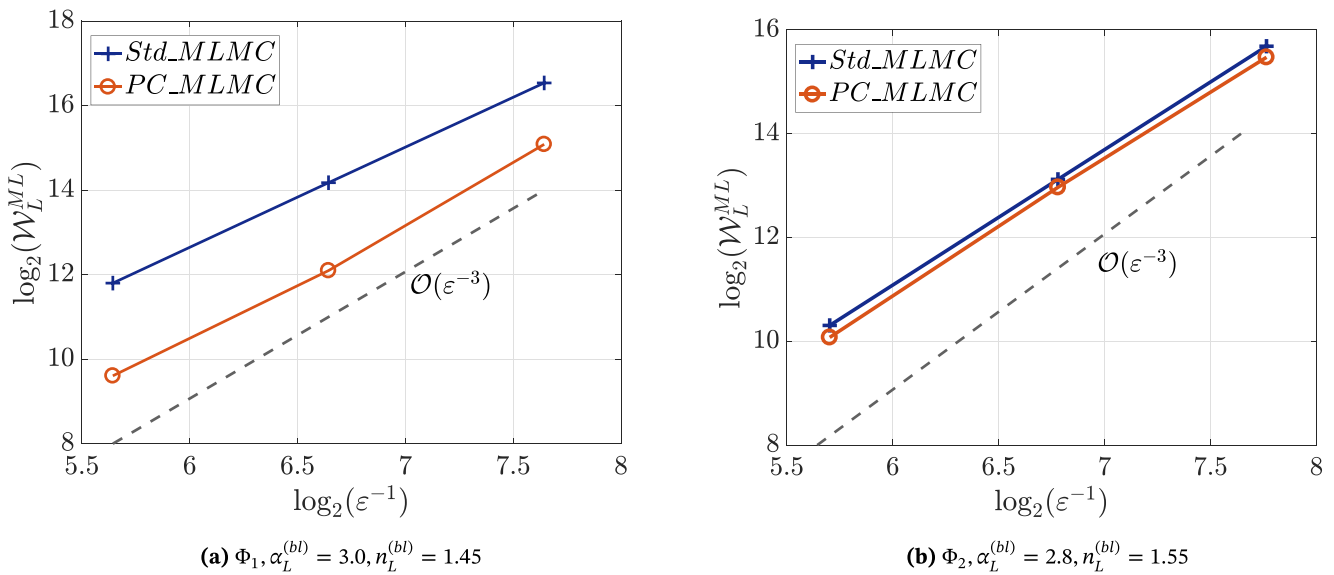


Fig. 10 a, b Comparison of CPU times \mathcal{W}_L^{ML} (sec) for two different estimators

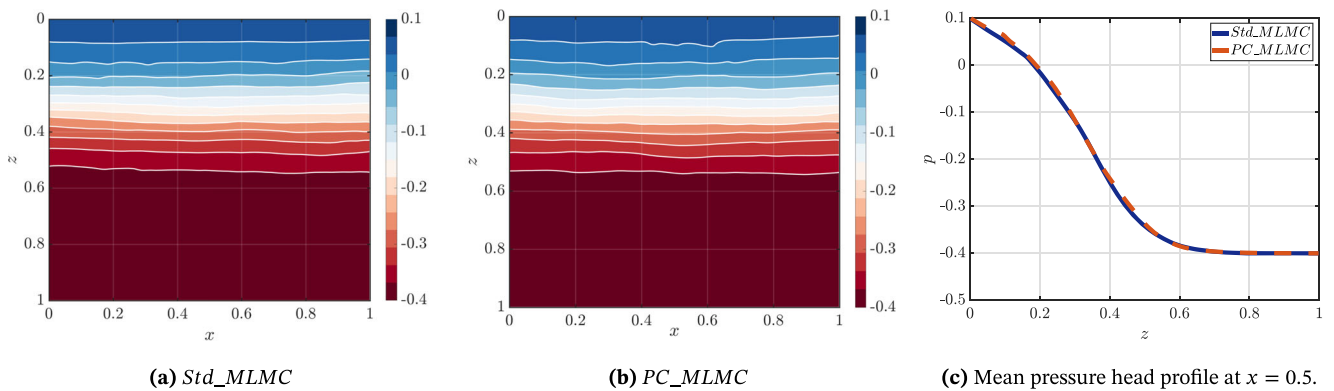


Fig. 11 a–c Comparison of mean pressure head field for $\Phi_1, \alpha_L^{(bl)} = 3.0, n_L^{(bl)} = 1.45, T_{\text{final}} = 0.2$ [h] computed using the two MLMC estimators with finest level $h_L = \Delta t_L = 1/256$ and $\epsilon = 0.005$

Table 5 Comparison of number of samples needed to achieve tolerances ϵ using the standard MLMC (*Std_MLMC*) and parametric continuation MLMC (*PC_MLMC*) estimators for $\Phi_2, n^{(bl)} = 1.55, \alpha^{(bl)} = 2.8$

h_ℓ	$N_\ell(h_L = 1/64, \epsilon = 0.0184)$		$N_\ell(h_L = 1/128, \epsilon = 0.0092)$		$N_\ell(h_L = 1/256, \epsilon = 0.0046)$	
	<i>Std_MLMC</i>	<i>PC_MLMC</i>	<i>Std_MLMC</i>	<i>PC_MLMC</i>	<i>Std_MLMC</i>	<i>PC_MLMC</i>
1/16	–	96	–	427	–	1659
1/32	–	18	–	77	–	301
1/64	9	4	36	16	140	60
1/128	–	–	4	4	15	12
1/256	–	–	–	–	3	3

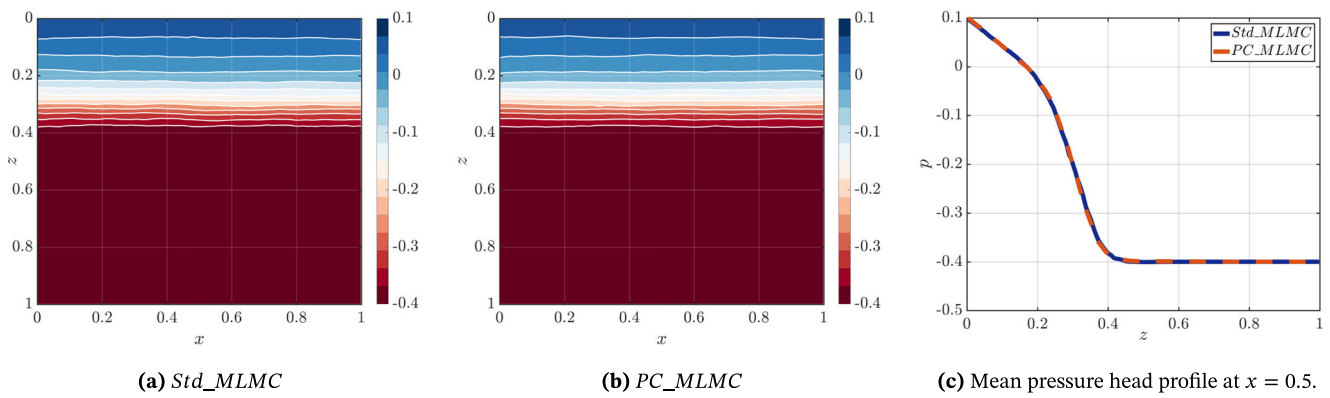


Fig. 12 a–c Comparison of mean pressure head field for $\Phi_2, \alpha_L^{(bl)} = 2.8, n_L^{(bl)} = 1.55, T_{\text{final}} = 0.2$ [h] computed using the two MLMC estimators with finest level $h_L = \Delta t_L = 1/256$ and $\varepsilon = 0.0046$

In the last part of this section, we validate the stochastic moments computed using the proposed estimator. It is expected that the mean pressure field computed using the two MLMC estimators should converge to a similar solution for a given tolerance. In Fig. 11, the mean pressure head profile for the isotropic case is shown. It is computed using the number of samples from Table 4, with $\varepsilon = 0.005$. For a closer inspection, we also compare the mean pressure head profiles at $x = 0.5$. Similarly, the mean profile for the anisotropic case is presented in Fig. 12, using the number of samples from Table 5 with $\varepsilon = 0.0046$. We see good agreement between the mean profiles computed from the two estimators. The isotropic case exhibits a seemingly smoother transition from the saturated to the unsaturated zone compared with the anisotropic problem. In Fig. 13, we also present the variance field for the isotropic test case, computed using the multilevel variance estimator $\mathcal{V}_L^{ML}[p_{h_L, \Theta_L}]$ given in Eq. 5.5. The two variance fields are very similar, although some discrepancy in the magnitude is observed. This is due to the fact that the two variance fields are computed using the samples based on the error analysis of $\mathcal{E}_L^{ML}[p_{h_L, \Theta_L}]$ (from Table 4) and not on the error analysis of $\mathcal{V}_L^{ML}[p_{h_L, \Theta_L}]$. Thus, the two variance estimates may

have different tolerances resulting in this slight mismatch. Readers are referred to [49] for a detailed error analysis of the multilevel variance estimator.

The results from the two estimators also showed good agreement with the plain Monte Carlo solutions performed on the grid $h_L = 1/128$. This is done in order to verify that a proper upscaling of the random fields on coarser levels is carried out while using the MLMC estimator. These results are omitted for the sake of brevity.

7 Conclusion

In this work, an efficient uncertainty propagation method for a high-dimensional stochastic extension of Richards equation was proposed. All the soil parameters were treated as unknown and modeled as random fields with appropriate marginal distributions. We also studied a modified Picard iteration and cell-centered multigrid method for solving the nonlinear systems with heterogeneous coefficients. We found that the combined solver is robust for a wide parameter range and the performance further improves with spatio-temporal refinements. This combination of solvers is

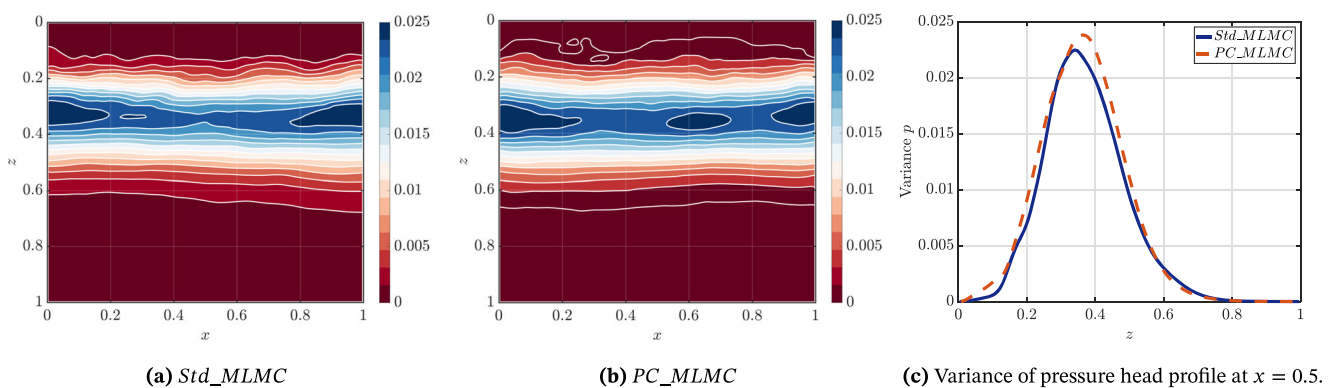


Fig. 13 a–c Comparison of the variance of the pressure head field for $\Phi_1, \alpha_L^{(bl)} = 3.0, n_L^{(bl)} = 1.45, T_{\text{final}} = 0.2$ [h] computed using the two MLMC variance estimators with finest level $h_L = \Delta t_L = 1/256$

general; therefore, its robustness can be further improved by incorporating adaptive time stepping or by combining with other advanced techniques, for instance, using the Anderson acceleration proposed in [43].

For computing the statistical moments of the solution of Richards equation, a parametric continuation technique-based multilevel Monte Carlo estimator was proposed. This estimator is very practical for this problem, as it requires solving the strongly nonlinear problem only on the finest level, where the solver is robust, and uses simpler nonlinear problems on the coarse grid levels for a variance reduction. For the stochastic Richards equation, the proposed estimator is more prominent regarding the computational gains compared with the standard MLMC method if the problem is strongly nonlinear. In general, this estimator is also applicable to other parameter dependent nonlinear PDEs. One of the research problems that needs to be addressed is finding a computationally viable way of obtaining optimal step sizes for the nonlinear parameters used in continuation. This problem will be actively investigated in future work.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix. Sampling and upscaling of Gaussian random fields

In this section, we outline the procedure for sampling Gaussian random fields that is used to sample the hydraulic conductivity fields Eq. 3.1 and also the non-Gaussian soil parameters Eq. 3.8. For this, we use the fast Fourier transform moving average (FFT-MA) algorithm from [37]. Although this sampling method is similar to the Cholesky decomposition technique, the FFT-MA method achieves a faster factorization of a covariance matrix by making the computational domain periodic. The resulting covariance operator is also periodic and can be decomposed as a convolutional product. This allows us to compute the samples of the random fields using cheaper vector-vector products compared with the expensive matrix-vector operation when using Cholesky factorization. Next, we provide a brief description of FFT-MA method from [37].

When using the Cholesky factorization, the samples of correlated Gaussian random vectors $\mathbf{z}(\omega)$ can be obtained as

$$\mathbf{C}_\Phi = \mathbf{L}\mathbf{L}^T \quad \text{and use} \quad \mathbf{z} = \mathbf{L}\mathbf{y}, \tag{A.1}$$

where \mathbf{C}_Φ is the covariance matrix constructed on some grid and \mathbf{y} is a vector of i.i.d. samples from the standard normal distribution. The FFT-MA relies on the decomposition of the covariance function $C_\Phi(r)$ as a convolutional product of some function $S_\Phi(r)$ and its transpose $S'_\Phi(r) = S_\Phi(-r)$. We can express this decomposition as

$$\mathbf{c} = \mathbf{s} * \mathbf{s}', \tag{A.2}$$

where \mathbf{c}, \mathbf{s} are vectors obtained by evaluating $C_\Phi(r)$ and $S_\Phi(r)$, respectively, at grid points of the considered mesh. Moreover, the resulting vector \mathbf{s} is also real, positive, and symmetric and $\mathbf{s} = \mathbf{s}'$. Now, a correlated random vector \mathbf{z} can be computed by using the convolution product

$$\mathbf{z} = \mathbf{s} * \mathbf{y}. \tag{A.3}$$

The FFT-MA method performs the above computations in the frequency domain. As the convolution product in spatial domain is equivalent to the component-wise product in the frequency domain, we can take a Fourier transform of Eq. A.2 as

$$\mathcal{F}(\mathbf{c}) = \mathcal{F}(\mathbf{s}) \cdot \mathcal{F}(\mathbf{s}) \implies \mathcal{F}(\mathbf{s}) = \sqrt{\mathcal{F}(\mathbf{c})}, \tag{A.4}$$

where \mathcal{F} denotes the discrete FFT and \cdot denotes component-wise multiplication. As the FFT operation requires a periodic signal, first we transform the vector \mathbf{c} into a periodic signal, which is also real, positive, and symmetric. For more details on the practical aspects of this transformation, see [21, 35, 36, 52]. Here the component-wise square-root operation does not pose any problems as the power spectrum $\mathcal{F}(\mathbf{c})$ is real and positive. Further, the convolution product in Eq. A.3 can be expressed as a vector-vector product in the frequency domain as

$$\mathcal{F}(\mathbf{z}) = \mathcal{F}(\mathbf{s} * \mathbf{y}) = \mathcal{F}(\mathbf{s}) \cdot \mathcal{F}(\mathbf{y}). \tag{A.5}$$

An inverse fast Fourier transform is finally applied to synthesize the samples for Gaussian random fields

$$\mathbf{z} = \mathcal{F}^{-1}(\mathcal{F}(\mathbf{s}) \cdot \mathcal{F}(\mathbf{y})). \tag{A.6}$$

It is pointed out that due to the periodicity in the covariance vector \mathbf{c} , the resulting random field \mathbf{z} is also periodic. Therefore, we only retain the part of the vector that corresponds to the physical domain and the remaining part is discarded. Also note that it takes two FFT evaluations to obtain one sample of \mathbf{z} (ignoring the FFT operation in Eq. A.4 that is performed just once). For a given mesh, the cost of sampling random fields is negligible compared with the cost of solving the nonlinear PDE using the modified Picard-CCMG solver.

Next, we describe the upscaling procedure for the random fields from grid \mathcal{D}_ℓ to $\mathcal{D}_{\ell-1}$. For clarity, we denote the above vectors computed on mesh \mathcal{D}_ℓ with subscript ℓ , for example, \mathbf{z}_ℓ , \mathbf{s}_ℓ , \mathbf{y}_ℓ . While estimating the correction term $\mathcal{E}_{N_\ell}^{MC}[p_{h_\ell, \Theta_\ell} - p_{h_{\ell-1}, \Theta_{\ell-1}}]$ in the telescopic sum Eq. 5.3, the approximations $p_{h_\ell, \Theta_\ell}(\omega_i)$ and $p_{h_{\ell-1}, \Theta_{\ell-1}}(\omega_i)$ need to be positively correlated such that the level-dependent variance $\|\mathcal{V}_\ell\|_{L^2(\mathcal{D})}$ is small (see Eq. 5.11). This is typically achieved by first sampling the fine grid Gaussian random field \mathbf{z}_ℓ to compute $p_{h_\ell, \Theta_\ell}(\omega_i)$ and using an upscaled version $\mathbf{z}_{\ell-1}$ for $p_{h_{\ell-1}, \Theta_{\ell-1}}(\omega_i)$. While performing such upscaling of random fields, it is important to ensure that the telescopic sum Eq. 5.3 is not violated. In other words, the expectation of the random variable p_{h_ℓ, Θ_ℓ} when estimating $\mathbb{E}[p_{h_\ell, \Theta_\ell} - p_{h_{\ell-1}, \Theta_{\ell-1}}]$ and $\mathbb{E}[p_{h_{\ell+1}, \Theta_{\ell+1}} - p_{h_\ell, \Theta_\ell}]$ should be the same, i.e.

$$\mathbb{E}[p_{h_\ell, \Theta_\ell}]^{(coarse)} = \mathbb{E}[p_{h_\ell, \Theta_\ell}]^{(fine)}, \text{ for } \ell = \{0, 1, \dots, L-1\}. \quad (\text{A.7})$$

Using spatial averaging for obtaining an upscaled version may result in a modified covariance structure on the coarser levels, violating Eq. A.7. This issue can be avoided by using the *covariance upscaling* [21] that employs the spectral generator on two consecutive grids using the same normally distributed vector \mathbf{y}_ℓ . When using the FFT-MA algorithm, the vector \mathbf{y}_ℓ is associated with respective grid points, coarser realizations of the fine grid Gaussian random field \mathbf{z}_ℓ can be obtained by using multi-dimensional averaging of the vector \mathbf{y}_ℓ . For instance, in two dimensions for the cell-centred mesh,

$$\mathbf{y}_{\ell-1}^{i,j} = \frac{1}{2}(\mathbf{y}_\ell^{2i-1, 2j-1} + \mathbf{y}_\ell^{2i-1, 2j} + \mathbf{y}_\ell^{2i, 2j-1} + \mathbf{y}_\ell^{2i, 2j}), \quad (\text{A.8})$$

where i and j are the cell indices for the mesh $\mathcal{D}_{\ell-1}$. The scaling by a factor 2 is needed to obtain a standard normal distribution for the averaged quantity $\mathbf{y}_{\ell-1}^{i,j}$. The coarse random field can now be simply assembled as

$$\mathbf{z}_{\ell-1} = \mathcal{F}^{-1}(\mathcal{F}(\mathbf{s}_{\ell-1}) \cdot \mathcal{F}(\mathbf{y}_{\ell-1})). \quad (\text{A.9})$$

This process can be recursively applied to generate upscaled random fields on next coarser scales. As the averaging in Eq. A.8 smooths out high frequencies, the upscaled version $\mathbf{z}_{\ell-1}$ will also be smoother compared with \mathbf{z}_ℓ . These upscaled Gaussian random fields can be utilized to generate upscaled non-Gaussian fields using Eq. 3.8.

References

- Richards, L.A.: Capillary conduction of liquids through porous mediums. *Physics* **1**(5), 318–333 (1931). <https://doi.org/10.1063/1.1745010>
- Celia, M.A., Bouloutas, E.T., Zarba, R.L.: A general mass-conservative numerical solution for the unsaturated flow equation. *Water Resour. Res.* **26**(7), 1483–1496 (1990). <https://doi.org/10.1029/WR026i007p01483>
- Forsyth, P.A., Wu, Y.S., Pruess, K.: Robust numerical methods for saturated-unsaturated flow with dry initial conditions in heterogeneous media. *Advances in Water Resources* **18**(1), 25–38 (1995). [https://doi.org/10.1016/0309-1708\(95\)00020-J](https://doi.org/10.1016/0309-1708(95)00020-J)
- Arbogast, T., Wheeler, M.F.: A nonlinear mixed finite element method for a degenerate parabolic equation arising in flow in porous media. *SIAM J. Numerical Anal.* **33**(4), 1669–1687 (1996). <https://doi.org/10.1137/S0036142994266728>
- Eymard, R., Gutnic, M., Hilhorst, D.: The finite volume method for Richards equation. *Comput. Geosci.* **3**(3), 259–294 (1999). <https://doi.org/10.1023/A:1011547513583>
- Pop, I.S.: Error estimates for a time discretization method for the Richards' equation. *Comput. Geosci.* **6**(2), 141–160 (2002). <https://doi.org/10.1023/A:1019936917350>
- Woodward, C.S., Dawson, C.N.: Analysis of expanded mixed finite element methods for a nonlinear parabolic equation modeling flow into variably saturated porous media. *SIAM J. Numerical Anal.* **37**(3), 701–724 (2000). <https://doi.org/10.1137/S0036142996311040>
- Delhomme, J.P.: Spatial variability and uncertainty in groundwater flow parameters: a geostatistical approach. *Water Resour. Res.* **15**(2), 269–280 (1979). <https://doi.org/10.1029/WR015i002p00269>
- Freeze, R.A.: A stochastic-conceptual analysis of one-dimensional groundwater flow in nonuniform homogeneous media. *Water Resour. Res.* **11**(5), 725–741 (1975). <https://doi.org/10.1029/WR011i005p00725>
- Hoeksema, R.J., Kitanidis, P.K.: Analysis of the spatial structure of properties of selected aquifers. *Water Resour. Res.* **21**(4), 563–572 (1985). <https://doi.org/10.1029/WR021i004p00563>
- Mantoglou, A., Gelhar, L.W.: Stochastic modeling of large-scale transient unsaturated flow systems. *Water Resour. Res.* **23**(1), 37–46 (1987). <https://doi.org/10.1029/WR023i001p00037>
- Russo, D., Russo, I., Laufer, A.: On the spatial variability of parameters of the unsaturated hydraulic conductivity. *Water Resour. Res.* **33**(5), 947–956 (1997). <https://doi.org/10.1029/96WR03947>
- Zhang, D.: Nonstationary stochastic analysis of transient unsaturated flow in randomly heterogeneous media. *Water Resour. Res.* **35**(4), 1127–1141 (1999). <https://doi.org/10.1029/1998WR900126>
- Iden, S.C., Durner, W.: Free-form estimation of the unsaturated soil hydraulic properties by inverse modeling using global optimization. *Water Resour. Res.* **43**(7), W07451 (2007). <https://doi.org/10.1029/2006WR005845>
- van Genuchten, M.T.: A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Science Society of America Journal* **44**, 892–898 (1980). <https://doi.org/10.2136/sssaj1980.03615995004400050002x>
- Mualem, Y.: A new model for predicting the hydraulic conductivity of unsaturated porous media. *Water Resour. Res.* **12**(3), 513–522 (1976). <https://doi.org/10.1029/WR012i003p00513>
- Zhang, D., Lu, Z.: Stochastic analysis of flow in a heterogeneous unsaturated-saturated system. *Water Resour. Res.* **38**(2), 10–15 (2002). <https://doi.org/10.1029/2001WR000515>
- Giles, M.B.: Multilevel Monte Carlo path simulation. *Operations Research* **56**(3), 607–617 (2008). <https://doi.org/10.1287/opre.1070.0496>
- Cliffe, K.A., Giles, M.B., Scheichl, R., Teckentrup, A.L.: Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients. *Computing and Visualization in Science* **14**, 3–15 (2011). <https://doi.org/10.1007/s00791-011-0160-x>

20. Mishra, S., Schwab, Ch., Šukys, J.: Multi-level Monte Carlo finite volume methods for nonlinear systems of conservation laws in multi-dimensions. *J. Comput. Phys.* **231**(8), 3365–3388 (2012). <https://doi.org/10.1016/j.jcp.2012.01.011>
21. Mishra, S., Schwab, Ch., Sukys, J.: Multi-level Monte Carlo finite volume methods for uncertainty quantification of acoustic wave propagation in random heterogeneous layered medium. *J. Comput. Phys.* **312**, 192–217 (2016). <https://doi.org/10.1016/j.jcp.2016.02.014>
22. Brandt, A.: Multi-level adaptive solutions to boundary-value problems. *Math. Comput.* **31**(138), 333–390 (1977). <https://doi.org/10.2307/2006422>
23. Brandt, A., Ron, D.: Multigrid solvers and multilevel optimization strategies. In: *Multilevel Optimization in VLSICAD*, pp. 1–69. Springer (2003)
24. Bank, R.E., Mittelmann, H.D.: Continuation and multi-grid for nonlinear elliptic systems. In: Hackbusch, W., Trottenberg, U. (eds.) *Multigrid Methods II*, pp. 23–37. Springer, Berlin (1986)
25. Chan, T.F.C., Keller, H.B.: Arc-Length continuation and multigrid techniques for nonlinear elliptic eigenvalue problems. *SIAM Journal on Scientific and Statistical Computing* **3**(2), 173–194 (1982). <https://doi.org/10.1137/0903012>
26. Dinar, N., Keller, H.B.: Computations of Taylor vortex flows using multigrid continuation methods. In: Chao, C.C., Orszag, S.A., Shyy, W. (eds.) *Recent Advances in Computational Fluid Dynamics. Lecture Notes in Engineering*, vol. 43, pp. 191–262. Springer, Berlin (1989)
27. Rheinboldt, W.C.: Solution fields of nonlinear equations and continuation methods. *SIAM Journal on Numerical Analysis* **17**(2), 221–237 (1980). <https://doi.org/10.1137/0717020>
28. Teckentrup, A., Scheichl, R., Giles, M.B., Ullmann, E.: Further analysis of multilevel Monte Carlo methods for elliptic PDEs with random coefficients. *Numerische Mathematik* **125**, 569–600 (2013). <https://doi.org/10.1007/s00211-013-0546-4>
29. Gittelsohn, C.J., Könnö, J., Schwab, C., Stenberg, R.: The multi-level Monte Carlo finite element method for a stochastic Brinkman problem. *Numer. Math.* **125**(2), 347–386 (2013). <https://doi.org/10.1007/s00211-013-0537-5>
30. Kumar, P., Rodrigo, C., Gaspar, F.J., Oosterlee, C.W.: On local fourier analysis of multigrid methods for PDEs with jumping and random coefficients. *SIAM J. Sci. Comput.* **41**(3), A1385–A1413 (2019)
31. Brooks, R.H., Corey, A.T.: Hydraulic properties of porous media. *Hydrology Papers, Colorado State University* **24**, 37 (1964)
32. Miller, C.T., Williams, G.A., Kelley, C.T., Tocci, M.D.: Robust solution of Richards' equation for nonuniform porous media. *Water Resour. Res.* **34**(10), 2599–2610 (1998). <https://doi.org/10.1029/98WR01673>
33. Ippisch, O., Vogel, H.J., Bastian, P.: Validity limits for the van Genuchten-Mualem model and implications for parameter estimation and numerical simulation. *Advances in Water Resources* **29**(12), 1780–1789 (2006). <https://doi.org/10.1016/j.advwatres.2005.12.011>
34. Ghanem, R.G., Spanos, P.: *Stochastic finite elements: a spectral approach*. Springer, New York (1991). <https://doi.org/10.1007/978-1-4612-3094-6>
35. Wood, A.T.A., Chan, G.: Simulation of stationary Gaussian processes in $[0, 1]^d$. *Journal of Computational and Graphical Statistics* **3**, 409–432 (1994). <https://doi.org/10.2307/1390903>
36. Dietrich, C.R., Newsam, G.N.: Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix. *SIAM Journal on Scientific Computing* **18**, 1088–1107 (1997). <https://doi.org/10.1137/S1064827592240555>
37. Ravalec, M.L., Noetinger, B., Hu, L.Y.: The FFT moving average (FFT-MA) generator: an efficient numerical method for generating and conditioning Gaussian simulations. *Mathematical Geology* **32**(6), 701–723 (2000). <https://doi.org/10.1023/A:1007542406333>
38. Sakamoto, S., Ghanem, R.: Simulation of multi-dimensional non-Gaussian non-stationary random fields. *Probabilistic Engineering Mechanics* **17**(2), 167–176 (2002). [https://doi.org/10.1016/S0266-8920\(01\)00037-6](https://doi.org/10.1016/S0266-8920(01)00037-6)
39. Phoon, K.K., Huang, H.W., Quek, S.T.: Simulation of strongly non-Gaussian processes using Karhunen-Loeve expansion. *Probabilistic Engineering Mechanics* **20**(2), 188–198 (2005). <https://doi.org/10.1016/j.probengmech.2005.05.007>
40. Wesseling, P.: Cell-centered multigrid for interface problems. *J. Comput. Phys.* **79**(1), 85–91 (1988). [https://doi.org/10.1016/0021-9991\(88\)90005-8](https://doi.org/10.1016/0021-9991(88)90005-8)
41. Molenaar, J.: A simple cell-centered multigrid method for 3D interface problems. *Computers and Mathematics with Applications* **31**(9), 25–33 (1996). [https://doi.org/10.1016/0898-1221\(96\)00039-9](https://doi.org/10.1016/0898-1221(96)00039-9)
42. Miller, C.T., Abhishek, C., Farthing, M.W.: A spatially and temporally adaptive solution of Richards' equation. *Advances in Water Resources* **29**(4), 525–545 (2006). <https://doi.org/10.1016/j.advwatres.2005.06.008>
43. Lott, P.A., Walker, H.F., Woodward, C.S., Yang, U.M.: An accelerated Picard method for nonlinear systems related to variably saturated flow. *Advances in Water Resources* **38**, 92–101 (2012). <https://doi.org/10.1016/j.advwatres.2011.12.013>
44. Mehl, S.: Use of Picard and Newton iteration for solving nonlinear ground water flow equations. *Groundwater* **44**(4), 583–594 (2006). <https://doi.org/10.1111/j.1745-6584.2006.00207.x>
45. Jones, J.E., Woodward, C.S.: Newton Krylov-multigrid solvers for large-scale, highly heterogeneous, variably saturated flow problems. *Advances in Water Resources* **24**(7), 763–774 (2001). [https://doi.org/10.1016/S0309-1708\(00\)00075-0](https://doi.org/10.1016/S0309-1708(00)00075-0)
46. Juncu, G., Nicola, A., Popa, C.: Nonlinear multigrid methods for numerical solution of the variably saturated flow equation in two space dimensions. *Transport in Porous Media* **91**(1), 35–47 (2012). <https://doi.org/10.1007/s11242-011-9831-9>
47. Caputo, J.-G., Stepanyants, Y.A.: Front solutions of Richards' equation. *Transport in Porous Media* **74**(1), 1–20 (2008). <https://doi.org/10.1007/s11242-007-9180-x>
48. Zha, Y., Yang, J., Yin, L., Zhang, Y., Zeng, W., Shi, L.: A modified Picard iteration scheme for overcoming numerical difficulties of simulating infiltration into dry soil. *J. Hydrology* **551**, 56–69 (2017). <https://doi.org/10.1016/j.jhydrol.2017.05.053>
49. Bierig, C., Chernov, A.: Convergence analysis of multilevel Monte Carlo variance estimators and application for random obstacle problems. *Numerische Mathematik* **130**(4), 579–613 (2015). <https://doi.org/10.1007/s00211-014-0676-3>
50. Mishra, S., Schwab, C.H.: Sparse tensor multi-level Monte Carlo finite volume methods for hyperbolic conservation laws with random initial data. *Math. Comput.* **81**(280), 1979–2018 (2012). <https://doi.org/10.1090/S0025-5718-2012-02574-9>
51. Muller, F., Jenny, P., Meyer, D.W.: Multilevel Monte Carlo for two phase flow and Buckley-Leverett transport in random heterogeneous porous media. *J. Comput. Phys.* **250**, 685–702 (2013). <https://doi.org/10.1016/j.jcp.2013.03.023>
52. Graham, I.G., Kuo, F.Y., Nuyens, D., Scheichl, R., Sloan, I.H.: Quasi-Monte Carlo methods for elliptic PDEs with random coefficients and applications. *J. Comput. Phys.* **230**(10), 3668–3694 (2011). <https://doi.org/10.1016/j.jcp.2011.01.023>