Available online at www.eccomasproceedia.org
Eccomas Proceedia UNCECOMP (2023) 311-330

ECCOMAS
Proceedia

UNCECOMP 2023
5th ECCOMAS Thematic Conference on
Uncertainty Quantification in Computational Science and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou (eds.)
Athens, Greece, 12-14 June 2023

MULTILEVEL AND MULTIGRID METHODS FOR SOLVING HENRY PROBLEM WITH UNCERTAIN COEFFICIENTS

Alexander Litvinenko[1], Dmitry Logashenko[2], Raul Tempone[1,2], Ekaterina Vasilyeva[3], Gabriel Wittum[2]

[1]RWTH Aachen, Aachen, Germany
litvinenko@uq.rwth-aachen.de

[2] KAUST, Thuwal-Jeddah, Saudi Arabia
{raul.tempone,dmitry.logashenko,gabriel.wittum}@kaust.edu.sa

[3] Goethe-Universität Frankfurt am Main, Germany
ekaterina.vasilyeva@gcsc.uni-frankfurt.de

Abstract. *We are solving a problem of salinization of coastal aquifers. As a test case example, we consider the Henry saltwater intrusion problem. Since porosity, permeability and recharge are unknown or only known at a few points, we model them using random fields. The Henry problem describes a two-phase flow and is nonlinear and time-dependent. The solution to be found is the expectation of the salt mass fraction, which is uncertain and time-dependent. To estimate this expectation we use the well known multilevel Monte Carlo (MLMC) method. The MLMC method takes just a few samples on computationaly expensive (fine) meshes and more samples on cheap (coarse) meshes. Then, by building a telescoping sum, the MLMC method estimates the expected value at a much lower cost than the classical Monte Carlo method. The deterministic solver used here is the well-known parallel and scalable ug4 solver.*

Keywords: uncertainty, Henry problem, density driven flow, saltwater intrusion, groundwater, two-phase

ISSN:2623-3339
Peer-review under responsibility of the organizing committee of UNCECOMP 2023.
doi: 10.7712/120223.10340.19765

1 INTRODUCTION

Poor quality or even lack of drinking water is a major problem in many parts of the world. Many of these areas are very close to the coast. Saltwater intrusion is the process by which saltwater intrudes into fresh water aquifers and alters the water table. The resulting salinization of groundwater can have disastrous consequences for humans, the environment and agriculture. This work contributes to the accurate modeling of different scenarios of saline flow [1, 34]. The results obtained will help to make better predictions and reduce the risks of the effects of saltwater intrusion. A simple but very illustrative model of coastal aquifers is the so-called Henry problem [17, 36, 43, 35, 8, 31]. We assume that this problem has uncertain permeability, porosity and recharge. The solution to be calculated is the mass fraction or a functional of it. The mass fraction is also uncertain and is modelled by a random field.

To better understand this salinization phenomena, we will estimate the following uncertain quantities of interest (QoIs):

1. Different extreme scenarios and the time point when they will occur
2. Expected values and variances of QoIs in time
3. Pure and saltwater integrals over the whole domain and a sub-domain
4. Exceedance probabilities
5. Solution at a point and in a small subdomain
6. Evolution of uncertainties over time

Publications dealing with uncertainty quantification and probabilistic risk assessment methods in hydrogeology are [38, 4, 41, 33, 3].

Our contribution. We use the multilevel Monte Carlo method (MLMC) to estimate the uncertainty in the Henry problem. To solve each deterministic problem, we use the multigrid solver ug4. The solver itself and all random simulations are implemented in parallel.

The plan of this paper is as follows. We start with the definition of the Henry saltwater intrusion problem (Section 2). Then we explain how we model uncertain porosity, permeability and recharge coefficients. Section 3 describes temporal and spatial discretizations, deterministic solver, and the MLMC method. Section 4 contains multiple numerical experiments. We compute and visualise multiple scenarios, expected value, variance, quantiles and probability density functions (pdfs). We also apply the MLMC method to estimate the expected value of the mass fraction. Finally, in the conclusion, we summarise the experience gained.

2 HENRY SALTWATER INTRUSION PROBLEM

2.1 Problem setting

In coastal aquifers, salty seawater intruding on the formation on one side (the seaside) displaces the pure water due to water recharge from land sources and precipitation from the other side [19]. Due to its higher density, seawater mainly penetrates along the bottom of the aquifer. This process can reach a certain steady state, but may be time-dependent due to the periodicity of the recharge or the control of the pumping rate from the wells. An accurate simulation of the salinization is vital for the prediction of water resource availability. However, the accuracy of such predictions strongly depends on the hydrogeological parameters of the formation and

the geometry of the computational domain, denoted by $\mathcal{D} \subset \mathbb{R}^2$. $\mathcal{D}$ represents an immobile porous matrix filled with a liquid phase. This liquid phase is a salt solution in water. Gravity induces the motion of the liquid phase due to the inhomogeneous distribution of density. This movement transports the salt, which would otherwise be subject to molecular diffusion.

In our test case, the aquifer is modelled by a rectangular domain $\mathcal{D} = [0, 2] \times [-1, 0]\text{m}^2$ entirely saturated with the liquid phase, see Fig. 1. The salty seawater intrudes from the right and the pure water recharges from the left. The top and bottom are assumed to be impermeable. Analogous settings with partially saturated domains are considered in [37].

Denoting the porosity by $\phi : \mathcal{D} \to \mathbb{R}$ and permeability by $\mathbf{K} : \mathcal{D} \to \mathbb{R}^{2\times 2}$, the mass conservation laws for the entire liquid phase and the salt yield the equations

$$\partial_t(\phi\rho) \quad + \quad \nabla \cdot (\rho\mathbf{q}) = 0, \tag{1}$$

$$\partial_t(\phi\rho c) \quad + \quad \nabla \cdot (\rho c\mathbf{q} - \rho\mathbf{D}\nabla c) = 0, \tag{2}$$

where $c(t, \mathbf{x}) : [0, +\infty) \times \mathcal{D} \to [0, 1]$ is the mass fraction of the salt (or of the brine) in the solution, $\rho = \rho(c)$ the density of the liquid phase and $\mathbf{D}(t, \mathbf{x}) : [0, +\infty) \times \mathcal{D} \to \mathbb{R}^{2\times 2}$ the molecular diffusion and mechanical dispersion tensor. For the velocity $\mathbf{q}(t, \mathbf{x}) : [0, +\infty) \times \mathcal{D} \to \mathbb{R}^2$, we assume the Darcy's law:

$$\mathbf{q} = -\frac{\mathbf{K}}{\mu}(\nabla p - \rho\mathbf{g}), \tag{3}$$

where $p = p(t, \mathbf{x}) : [0, +\infty) \times \mathcal{D} \to \mathbb{R}$ is the hydrostatic pressure, $\mu = \mu(c)$ the viscosity of the liquid phase and $\mathbf{g} = (0, -g)^T \in \mathbb{R}^2$ the gravity vector. The system of time-dependent partial differential equations (PDEs) (1–3) has two unknowns c and p.

Further, we set $\rho(c) = \rho_0 + (\rho_1 - \rho_0)c$ and $\mu = \text{const}$. The molecular diffusion tensor is assumed to be $\mathbf{D} = \phi D\mathbf{I}$ with a constant scalar $D \in \mathbb{R}$, and the identity matrix $\mathbf{I} \in \mathbb{R}^{2\times 2}$. The isotropic permeability is assumed to be $\mathbf{K} = K\mathbf{I}$, $K \in \mathbb{R}$, see [43]. For the initial conditions, we set $c|_{t=0} = 0$. The boundary conditions (b.c.) are shown in Fig. 1(a). On the right side of the domain we impose Dirichlet b.c. for both c and p variables, which should represent the adjacent seawater aquifer:

$$c|_{x=2} = 1, \qquad p|_{x=2} = -\rho_1 g y. \tag{4}$$

On the left side, we prescribe the inflow of fresh water:

$$c|_{x=0} = 0, \qquad \rho\mathbf{q} \cdot \mathbf{e}_x|_{x=0} = \hat{q}_{\text{in}}, \tag{5}$$

where $\mathbf{e}_x = (1, 0)^\top$ and $\hat{q}_{\text{in}}$ is a constant. For the classical formulation of the Henry problem, this value is set to $\hat{q}_{\text{in}} = 6.6 \cdot 10^{-2}$ kg/s in [43] or $\hat{q}_{\text{in}} = 3.3 \cdot 10^{-2}$ kg/s in [36, 35]. On the upper and lower sides of $\mathcal{D}$, the Neumann zero b.c. are imposed.

An example of $c(t, \mathbf{x})$ for parameters from Table 2 is shown in Fig. 1(right), compare it with the mean and the variance in Fig. 3. The dark red colour corresponds to $c = 1$ and the dark blue colour corresponds to $c = 0$. Due to its higher density, the salty water intrudes into the aquifer on the lower right. It is pushed back by the lighter clean water coming from the left. This creates a vortex in the flow in the lower-right corner of the domain: the salty water enters at the lower part of the right boundary and is deflected upwards and to the right, back to the "seaside", forming a "salty triangle". This current does not transport the salt to the left part of the domain. The salt moves further to the left by diffusion and dispersion and is washed out by the recharge. In the classical formulation, this "salty triangle" initially grows with time but reaches a steady

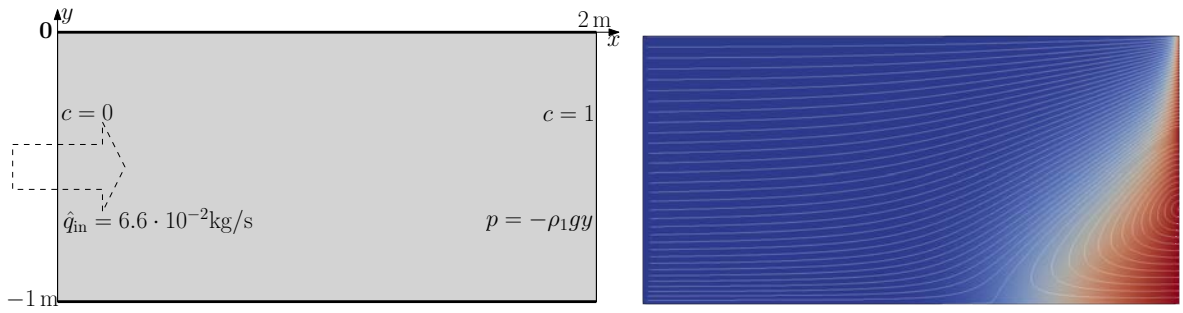

Figure 1: (left) Computational domain $\mathcal{D} := [0,2] \times [-1,0]$; (Right) One realization of the mass fraction $c(t,\mathbf{x})$ and the streamlines of the velocity field $\mathbf{q}$ at $t = 6016$ s.

state, see [43, 36, 35]. However, the initial non-stationary phase can take considerable time. The study of this phase is particularly important for understanding the behaviour of the system when the recharge is changing. For this, in addition to the mean and variance, we consider the mass fraction at a few points and a probability that the integral value — the total amount of pure/salty water — becomes smaller/larger than a certain threshold (see Equations 9-10).

Our motivation is to consider points where the variation in concentration is large. Note that the mass fraction c at each point $\mathbf{x}$ is a function of time.

Considering different spatial points can help to track salinity changes in groundwater wells over time and understand which areas of the aquifer are most vulnerable. This information can be used by farmers to take action, such as reducing salinity or adapting strategies for planting salt-tolerant crops.

2.2 Stochastic modeling of porosity and permeability

The main sources of uncertainties are the hydrogeological properties of the porous medium—porosity (ϕ) and permeability ($\mathbf{K}$) fields of the solid phase—as well as the fresh water recharge flux $\hat{q}_x$ through the left boundary. The quantity to be computed is the mass fraction $c = c(\phi, \mathbf{K}, \hat{q}_x)$. We assume $\mathbf{K}$ to be isotropic

$$\mathbf{K} = K\mathbf{I}, \qquad K = K(\phi) \in \mathbb{R}, \tag{6}$$

Here distribution of $\phi(\mathbf{x}, \boldsymbol{\xi})$, $\mathbf{x} \in \mathcal{D}$, is determined by a set of stochastic parameters $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_M, \ldots)$. Different materials may have different dependences in Eq. 6. [28, 29, 7]. In our model, we use a Kozeny-Carman-like dependence

$$K(\phi) = \kappa_{KC} \cdot \frac{\phi^3}{1-\phi^2}, \tag{7}$$

where the scaling factor κ_{KC} is chosen to satisfy the equality $K(\mathbb{E}[\phi])\mathbf{I} = \mathbb{E}[\mathbf{K}]$ resembling the parameters of the standard Henry problem.

The inflow flux is kept constant over the left-hand boundary, but it is dependent on the stochastic variable q_{in}. We also assume that the inflow flux is independent of ϕ and $\mathbf{K}$.

3 Numerical methods

3.1 Discretization and deterministic solver

The system (1–2) is solved in the space $\mathcal{D} \times [0,T]$ numerically, where the symbol $\times$ denotes the Cartesian product. After discretization of $\mathcal{D}$ by quadrilaterals of size h, obtain $\mathcal{D}_h$. Equations

(1–2) are discretized by a vertex-centered finite-volume scheme with the so-called "consistent velocity" for approximation of Darcy's law (3) as presented in [9, 10, 11]. We denote the number of degrees of freedom associated with $\mathcal{D}_h$ by n. Note, that there are two degrees of freedom per grid vertex, one for the mass fraction and another for the pressure. We use the implicit Euler method with a fixed time step τ for time discretization. The number of the computed time steps is $r = T/\tau$.

We use partial upwind for the convective terms [9]. Therefore, the discretization error is of the second order with respect to (w.r.t.) the spatial mesh size h. However, as the diffusion in (2) is very small in comparison with the velocity, for the grids in our numerical experiments, the observed reduction of the discretization error after grid refinement corresponded to the first order.

Thus, we assume the first order dependence of the discretization error w.r.t. h which is consistent with our numerical experiments. Furthermore, the Euler method provides the first order dependence of the discretization error w.r.t. τ.

The use of the implicit time stepping scheme provides unconditional stability but requires the solution of a large nonlinear algebraic system of the discretized equations with n unknowns in every time step. To solve this system the Newton method is used. Linear systems inside the Newton iteration are solved by the BiCGStab method (see [2]) preconditioned with the geometric multigrid method (GMG V-cycle [14]). In the multigrid cycle, the ILU$_\beta$-smothers [15] and the Gaussian elimination as the coarse grid solver are used.

To construct the spatial grid hierarchy $\mathcal{D}_0, \mathcal{D}_1, \ldots, \mathcal{D}_L$, we start with a coarse grid consisting of 512 grid elements (quadrilaterals) and $n_0 = 1122$ degrees of freedom. This grid is regularly refined to get all other grid levels. After every spatial grid refinement, the number of grid elements is multiplied by a factor 4. Consequently, the number of degrees of freedom is increased by factor 4, i.e., $n_\ell \approx n_0 \cdot 2^{d\ell}$, $d = 2$ (see Table 1). This hierarchy is used in both the geometric multigrid preconditioner and the MLMC method. We also construct the temporal grid hierarchy $\mathcal{T}_0, \mathcal{T}_1, \ldots, \mathcal{T}_L$. The time step on each temporal grid is denoted by τ_ℓ with $\tau_{\ell+1} = \frac{1}{2}\tau_\ell$. The number of time steps on ℓ-th grid (level) will be $r_{\ell+1} = 2r_\ell$ and $r_\ell = r_0 2^\ell$, where r_0 is the number of grid points on $\mathcal{T}_0$. On ℓ-th level the MLMC is using the grid $\mathcal{D}_\ell \times \mathcal{T}_\ell$. Up to six spatial and time grids were used in the numerical experiments.

In the context of this work, it is important to estimate the numerical complexity of the deterministic solver w.r.t. the grid level ℓ. The most time-consuming part of the simulation is the solution of the discretized nonlinear system. It is typically hard to predict the number of Newton iterations in every time step, but in our numerical experiments, two iterations were enough to achieve the prescribed accuracy. Accordingly, the linear solver has been called at most two times per time step. Furthermore, the convergence rate of the geometric multigrid method does not depend on the mesh size [15]. Hence, the computation complexity of one time step is $\mathcal{O}(n_\ell)$, where n_ℓ is the number of the degrees of freedom on the grid level ℓ. Therefore, the overall numerical cost of the computation of one scenario on grid level ℓ for r_ℓ time steps is

$$s_\ell = \mathcal{O}(n_\ell r_\ell), \quad s_\ell \propto s_{\ell-1} 2^{(d+1)}, \quad d = 2. \tag{8}$$

3.2 Multilevel Monte Carlo method

To reduce the total computing and storage costs, we apply the standard MLMC method [13]. This method combines samples from different mesh levels in an efficient way. The MLMC method is well-known [5, 6, 12, 13, 16, 40, 26], therefore we will not repeat it here.

4 NUMERICAL EXPERIMENTS

We use the MLMC method to compute the mean value of various QoIs, such as c in the whole domain, c in a point, or an integral value of c. Below we introduce the fresh water (FW) and salty water integrals:

$$Q_{FW}(t,\omega) := \int_{\mathbf{x}\in\mathcal{D}} I(\mathcal{D}_{FW})d\mathbf{x}, \tag{9}$$

where $I(\mathcal{D}_{FW})$ is the indicator function identifying a sub-domain $\mathcal{D}_{FW}(t) := \{\mathbf{x} : \ c(t,\mathbf{x},\omega) \leq c_{FW}\} \subseteq \mathcal{D}$ meaning the volume of the fresh water at the time t, $c_{FW} := 0.012178$. And the salty water mass integral:

$$Q_{salt}(t,\omega) := \int_{\mathbf{x}\in\mathcal{D}} \rho(c(t,\mathbf{x},\omega))c(t,\mathbf{x},\omega)d\mathbf{x}, \tag{10}$$

Remark 1 *(Motivation for the value $c_{FW} = 0.012178$ in (9)).*
Integral (9) yields the volume *of water which is considered to be "potable" in the sense its concentration of the salt is small. In our artificial setting, we consider this fresh water to have the density $\rho(c) \leq \rho_{FW} := 1000.30432\ kg \cdot m^{-3}$. Due to the linear dependence of ρ of c, cf. Subsection 2.1, this corresponds to $c \leq c_{FW} = 0.012178$. Legal regulations for the salinity of drinking water can be different and typically allow a greater density ρ_{FW}. We take this value only as a reference, so that $\rho_0 < \rho_{FW} \ll \rho_1$. Note however that taking $c_{FW} = 0$ (i.e. $\rho_{FW} = \rho_0$) would lead to the zero value of the integral due to the diffusion-dispersion term in (2).*

Each simulation may contain up to $n = 0.5 \cdot 10^6$ spatial mesh points and a few thousands of time steps ($r = 6016$ on the finest mesh).

Deterministic solver. The computations presented in this work have been performed using the ug4 simulation software toolbox (`https://github.com/ug4/ughub.wiki.git`) [30, 42]. This software has been applied for subsurface flow simulations of real-world aquifers (see [34]). The toolbox is parallelized using MPI, and the presented results have been obtained on the Shaheen II cluster, provided by the King Abdullah University of Science and Technology. Every sample has been computed on 32 cores of a separate cluster node. To reduce communication time between different nodes, each simulation (scenario) was localized to one node. All the scenarios have been computed on different nodes concurrently. A similar approach has been used in [22, 23]. Since simulations were done on different meshes, the computation time of each simulation has varied in a wide range (see Table 1).

Porosity and recharge: We assume two horizontal layers: $y \in [-0.75, 0]$ (the upper layer) and $y \in [-1, -0.75)$ (the lower layer). The porosity inside each layer is uncertain too and is modelled as in Eq. 11:

$$\phi(\mathbf{x},\boldsymbol{\xi}) = 0.35 \cdot (1 + 0.15(\xi_2 \cos(\pi x/2) + \xi_2 \sin(2\pi y) + \xi_1 \cos(2\pi x))) \cdot C_0(\xi_1), \tag{11}$$

$$\text{where } C_0(\xi_1) = \begin{cases} 1 + 0.2\xi_1 & \text{if } y < -0.75 \\ 1 - 0.2\xi_1 & \text{if } y \geq -0.75 \end{cases} \tag{12}$$

Additionally, the recharge flux is also uncertain and is equal to

$$\hat{q}_{in} = -6.6 \cdot 10^{-2}(1 + 0.5 \cdot \xi_3). \tag{13}$$

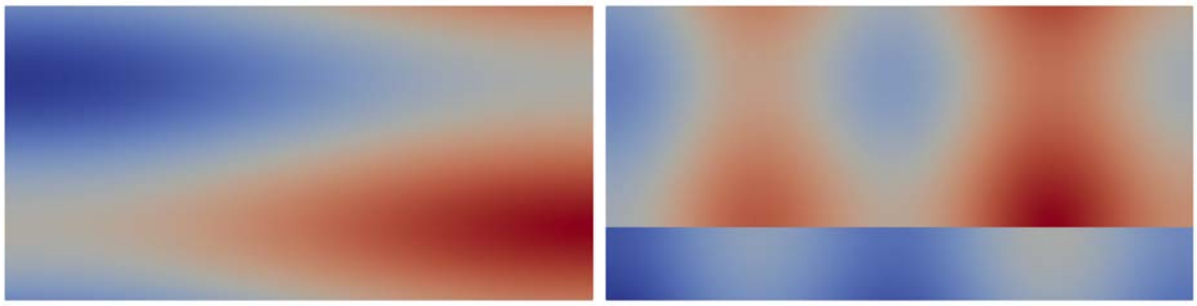

Figure 2: Two random realizations of porosity: (left) $\phi \in [0.315, 0.385]$, $\xi_1 = 0$; (right) $\phi \in [0.237, 0.508]$.

Here random variables ξ_1, ξ_2 and ξ_3 are sampled independently and uniformly in $[-1, 1]$. Figure 2 shows two random realization of the porosity random field $\phi(\boldsymbol{\xi})$. Efficient numerical methods to compute and generate random fields were consideren in [18, 20].

Numerical experiment 1. The mean and variance of the mass fraction are provided in Fig. 3 on the left and right, respectively. The expectation takes values from $[0, 1]$, and the variance range is $[0, 0.05]$. The areas with high variance (dark red) indicate regions with high variability/uncertainty. Additionally, the right image displays five contour lines $\{\mathbf{x} : \ \mathrm{Var}[c](t, \mathbf{x}) = 0.01 \cdot i\}$, $i = 1..5$. These computations were done for $\boldsymbol{\xi} = \boldsymbol{\xi}^* = (-0.5898, -0.7257, -0.9616)$ and $t = T = 6016$ s.

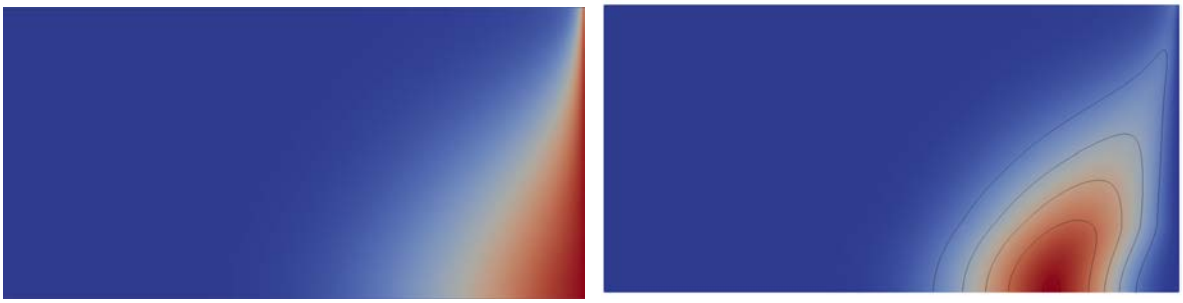

Figure 3: (left) Mean value $\bar{c} \in [0, 1]$ and (right) the variance $\mathrm{Var}[c] \in [0, 0.05]$ of the mass fraction with contour lines $\{\mathbf{x} : \ \mathrm{Var}[c] = 0.01 \cdot i\}$, $i = 1..5$, $t = T = 6016$.

We observed that the variability (uncertainty) of the mass fraction can vary from one grid point to another. At some points (dark blue regions) the solution does not change. At other points (white-yellow regions) the variability is very low or high (dark red regions). In regions of high uncertainty, refining the mesh and applying the MLMC method is particularly promising.

Numerical test 2. In Fig. 4 we visualise 100 QMC realisations (Halton sequence) of $c(t, \mathbf{x})$ at three points: $\mathbf{x} = (1.10, -0.95)$ (left), $\mathbf{x} = (1.60, -0.95)$ (centre) and $\mathbf{x} = (1.85, -0.50)$ (right). The dotted line at the bottom is the quantile 0.025. The next dotted line is the 0.25 quantile and the dotted line at the top is the 0.975 quantile. All five quantiles from bottom to top are 0.025, 0.25, 0.50, 0.75 and 0.975 respectively. We observe that c at the final time $t = T$ varies considerably and depends on the point $\mathbf{x}$.

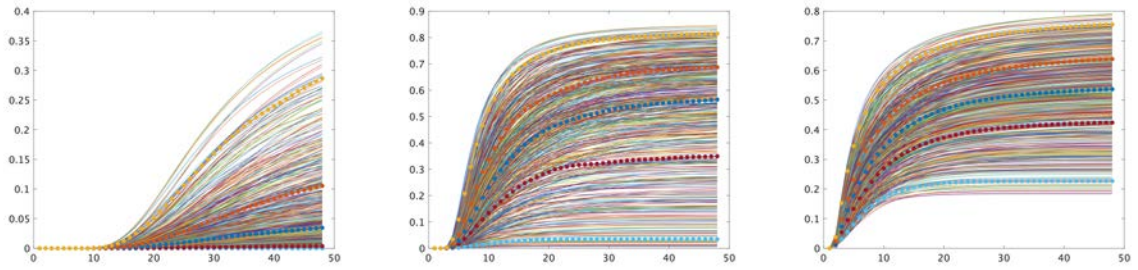

Figure 4: One hundred QMC realizations of $c(t, \mathbf{x})$ at three points $(1.10, -0.95)$ (left), $(1.60, -0.95)$ (center) and $(1.85, -0.50)$ (right). Dotted lines from the bottom to the top indicate the quantiles 0.025, 0.25, 0.50, 0.75, and 0.975, respectively.

Numerical experiment 3. In Fig. 5, we demonstrate the pdf of $t^*(\omega) = \min_t\{t : \ Q_{FW}(t, \omega) < 1.2\}$ (left), and the pdf of $t^*(\omega) = \min_t\{t : \ Q_{FW}(t, \omega) < 1.7\}$ (right). On average, after approximately 21 time steps (on the left) and 43 time steps (on the right), the volume of the fresh water becomes less than 1.3 and 0.9, respectively. The initial volume of the fresh water was 2.0.

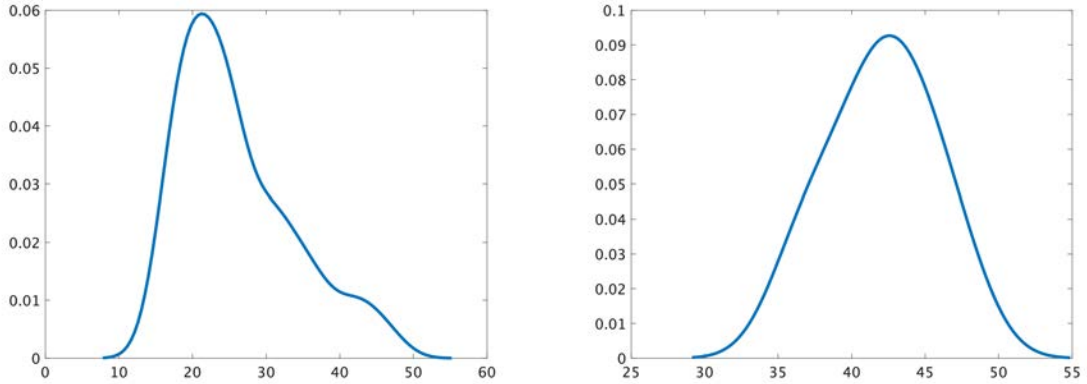

Figure 5: The pdf of the earliest time point when the fresh water integral Q_{FW} becomes less than 1.3 (left) and 0.9 (right). The x-axis represents time points.

Numerical experiment 4. In Fig. 6, we demonstrate the pdf of $t^*(\omega) = \min_t\{t : \ Q_{salt}(t, \omega) > 0.3\}$ (left), and the pdf of $t^*(\omega) = \min_t\{t : \ Q_{salt}(t, \omega) > 0.1\}$ (right). On average, after approximately 30 time steps (on the left) and 6 time steps (on the right), the volume of the salt water becomes larger than 0.3 and 0.1, respectively. The initial volume of the salt water was 0.

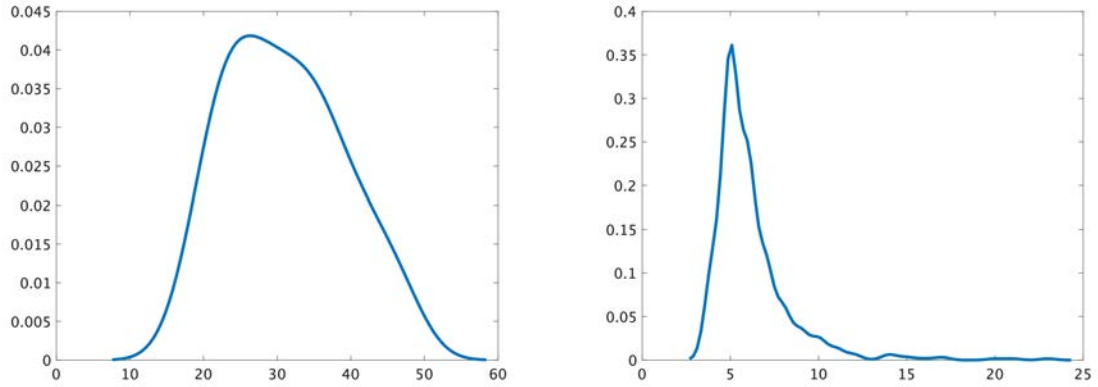

Figure 6: The pdf of the earliest time point when the salt integral Q_{salt} becomes larger than 0.3 (left) and 0.1 (right). The x-axis represents time points.

Numerical experiment 5. In Fig. 7, we show the time evolutions of pdfs of Q_{FW} (left) and Q_{salt} (right). At the initial point $t = 3\tau$, $Q_{FW} \approx 1.7$ (in average), it corresponds to the farthest right (blue) curve (on the left picture). At the initial point $t = 3\tau$, $Q_{salt} \approx 0.05$ (in average), it corresponds to the farthest left (blue) curve (on the right picture). After 48 time steps, the pdf for both integrals is very flat.

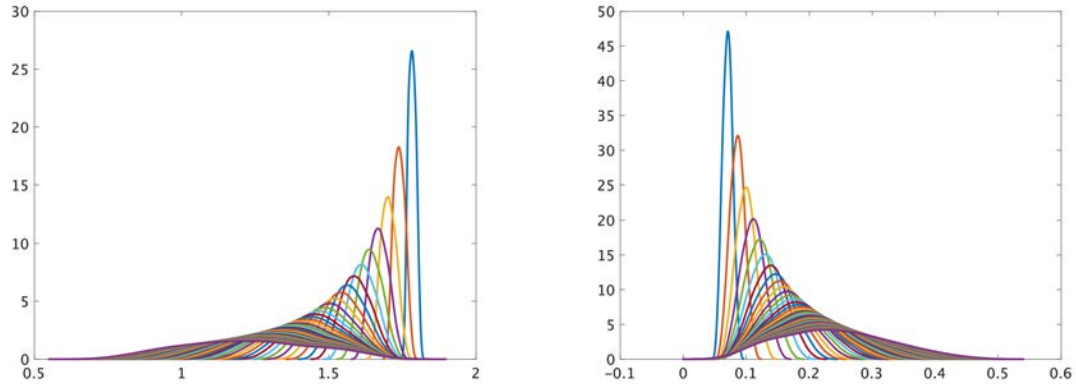

Figure 7: Evolutions of the pdfs of Q_{FW} (left) and Q_{salt} (right) for $t = \{3\tau, \ldots, 48\tau\}$.

Numerical experiment 6. In Fig. 8, we visualize 100 QMC realizations (Halton sequence) of $Q_{FW}(t)$ (left), and Q_{salt} (right). The time is along the x-axis, $t \in [\tau, 48\tau]$. Additionally, five quantiles are represented by dotted curves from the bottom to the top and are 0.025, 0.25, 0.50, 0.75, and 0.975, respectively.

Numerical experiment 7. Figure 9 displays the evolution of the pdf of $c(t, \mathbf{x}, \omega)$ at a fixed point $\mathbf{x} = (1.85, -0.95)$ in time $t = \{3\tau, \ldots, 48\tau\}$. From left to right, the farthest left (blue) pdf corresponds to $t = 3\tau$, the second curve from the left (red) corresponds to $t = 4\tau$, and so on. In the beginning, $t = 3\tau$, and the mass fraction c is low, about 0.15 on average. Then, with time, c increases and, at $t = T = 48\tau$, is approximately equal to 1.

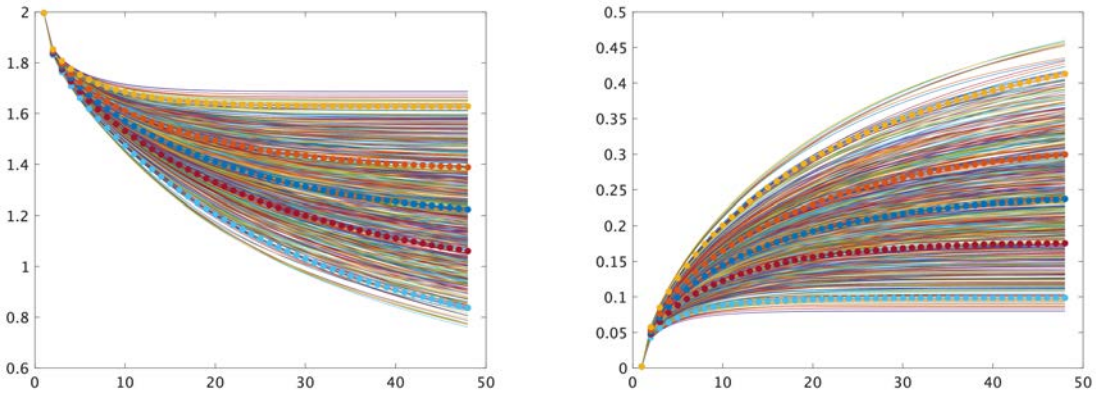

Figure 8: One hundred realizations of $Q_{FW}(t)$. The x-axis represents time $t = 1\tau, \ldots, 48\tau$; dotted curves denote five quantiles: 0.025, 0.25, 0.50, 0.75, and 0.975 from the bottom to the top.

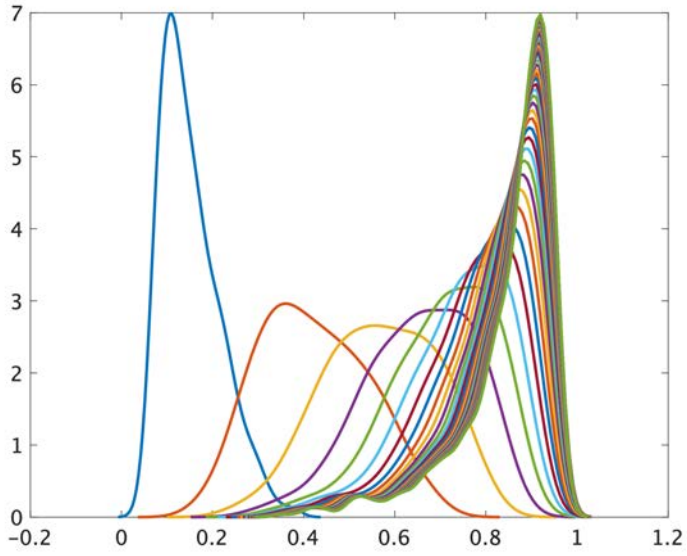

Figure 9: Evolution of the pdf of $c(t, \mathbf{x})$ for $t = \{3\tau, \ldots, 48\tau\}$.

Numerical experiment 8. Next, we research how $g_\ell - g_{\ell-1}$ depends on the time and level. All graphics in Fig. 10 display 100 realizations of the differences between solutions computed on two nearby meshes for every time point t_i, $i = 1 \dots 48$ (along the x-axis). The top left graphic indicates the differences between the mass fractions computed on Levels 1 and 0. The other graphics reveal the same, but for Levels 2 and 1, 3 and 2, 4 and 3, and 5 and 4, respectively. The largest value decreases from $2.5 \cdot 10^{-2}$ (top left) to $5 \cdot 10^{-4}$. Considerable variability is observed for $t \in [3\tau, 7\tau]$ and $t \in [8\tau, 25\tau]$. Starting with $t \approx 30\tau$, the variability between solutions decreases and stabilizes. From these five graphics, we can estimate that the maximal amplitude decreases by a factor ≈ 2, at 0.015, 0.008, 0.004, 0.0015, and 0.0008. However, it is challenging to make a similar statement about each time point t. This observation makes it difficult to estimate the weak and strong convergence rates and the optimal number of samples correctly on each mesh level. They are different for each time t (and for each $\mathbf{x}$). For some time points, the solution is smooth and requires only a few levels and a few samples on each level. For other points with substantial dynamics, the numbers of levels and samples are higher.

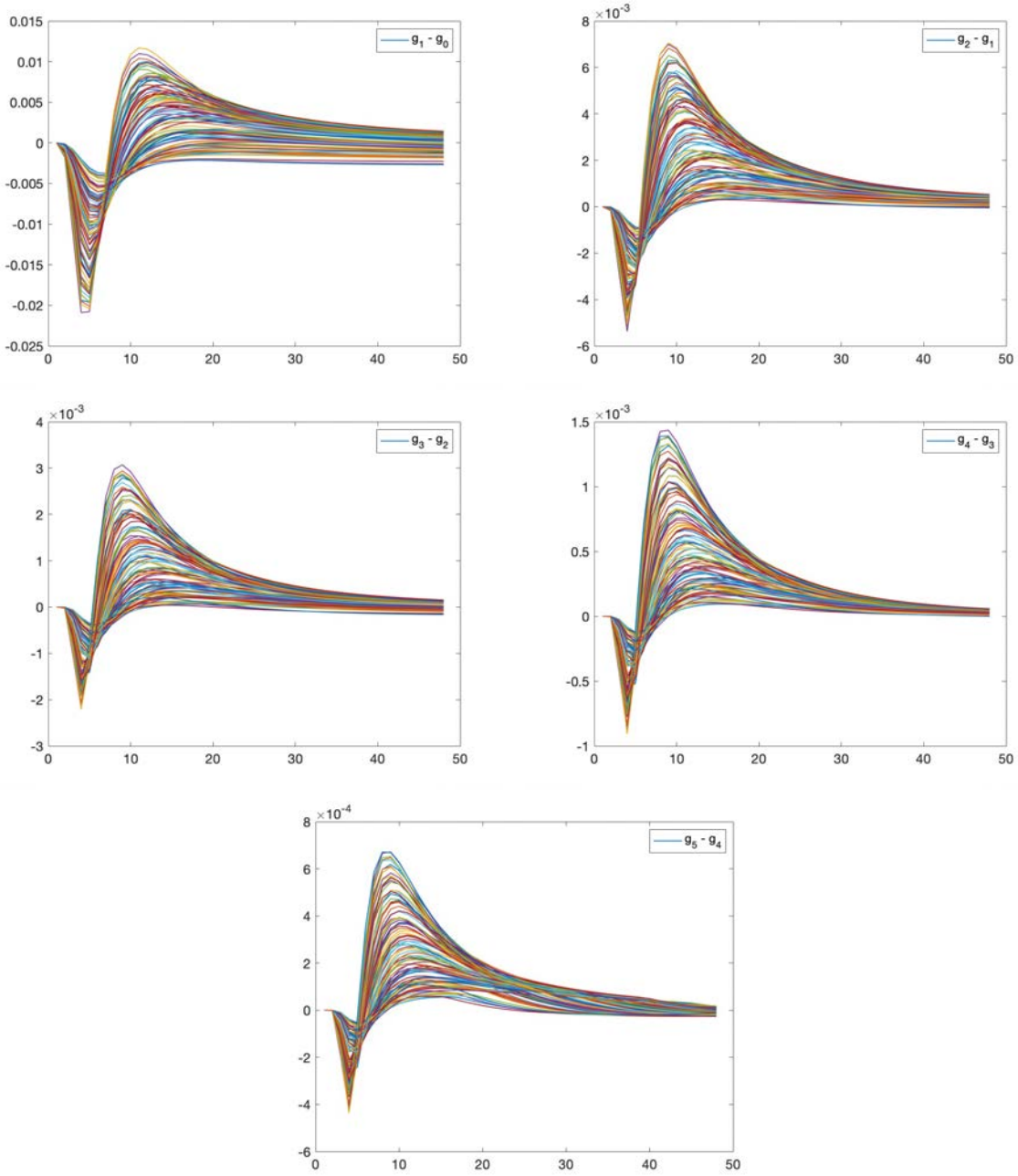

Figure 10: Differences between mass fractions computed at the point $(1.60, -0.95)$ on levels a) 1 and 0, b) 2 and 1 (first row), c) 3 and 2, d) 4 and 3 (second row), and e) 5 and 4 (third row) for 100 realizations (x- axis represents time).

Numerical experiment 9. Because $g_\ell - g_{\ell-1}$ is random, we visualize its mean and variance. Here we denote $V_0 := \mathbb{V}[g_0]$ and for $\ell \geq 1$, let V_ℓ be the variance of $g_\ell - g_{\ell-1}$, i.e., $V_\ell := \mathbb{V}[g_\ell - g_{\ell-1}]$.

Figure 11 demonstrates the mean (left) and variance (right) of the differences in concentrations $g_\ell - g_{\ell-1}$, $\ell = 1, \ldots, 5$. On the left, the amplitude decreases when ℓ increases. A slight exception is the blue line for $t \approx 9, 10, 11\tau$ (right). A possible explanation is that the solutions g_0 or g_1 are insufficiently accurate. The right image presents how the amplitude of the variances decays. This decay is necessary for the successful work of the MLMC method. We also observe a potential issue; the weak and strong convergence rates vary for various time points t. Thus, determining the optimal number of samples m_ℓ for each level is not possible (only suboptimal).

At the beginning $t = 0$, the variability is zero and starts to increase. We observe some variation for $t \in [\tau, 48\tau]$, and then the process starts to stabilize after $\approx 45\tau$ time steps. The variability is either unchanging from level to level or decreases.

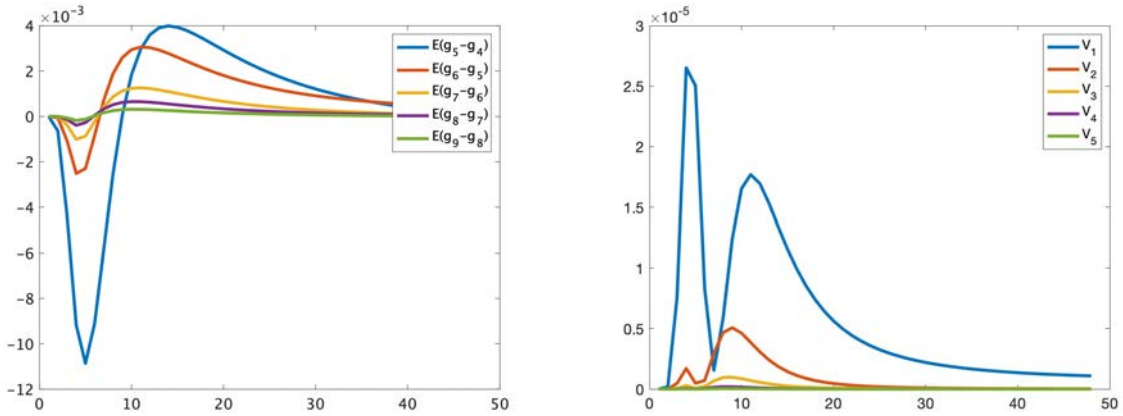

Figure 11: (left) Mean and (right) variance V_ℓ of the differences $g_\ell - g_{\ell-1}$ vs. time, computed on various levels at the point $(1.60, -0.95)$.

Table 1 contains average computing times, which are necessary to estimate the number of samples m_ℓ at each level ℓ. The fourth column contains the average computing time, and the fifth and sixth columns contain the shortest and longest computing times. The computing time for each simulation varies depending on the number of iterations, which depends on the porosity and permeability. We observed that, after ≈ 6016 s, the solution is almost unchanging; thus, we restrict this to only $t \in [0, T]$, where $T = 6016$. For example, if the number of time steps is $r_\ell = 188$ (Level 0 in Table 1), then the time step $\tau = \frac{T}{r_\ell} = \frac{6016}{188} = 32$ s.

The time step τ is adaptive and changing from $\tau = \frac{6016}{128} = 32$ s (very coarse mesh) to $\tau = \frac{6016}{6016} = 1$ s (finest mesh). Starting with level $\ell = 2$, the average time increases by a factor of eight. These numerical tests confirm the theory in Eq. (8), stating that the numerical solver is linear w.r.t. n_ℓ and r_ℓ.

Level ℓ	n_ℓ	r_ℓ	$\tau_\ell = 6016/r_\ell$	Computing times (s.)		
				average	min.	max.
0	1122	188	32	1.15	0.88	1.33
1	4290	376	16	4.1	3.4	4.87
2	16770	752	8	19.6	17.6	22
3	66306	1504	4	136.0	128	144
4	263682	3008	2	1004.0	891	1032
5	1051650	6016	1	8138.0	6430	8480

Table 1: Number of degrees of freedom n_ℓ, number of time steps r_ℓ, step size in time τ_ℓ, average, minimal, and maximal computing times on each level ℓ.

The slope in Fig. 12 can be used to estimate the rates of the, so-called, weak (left) and strong (right) convergences in the MLMC method. The levels are indicated on the horizontal axis.

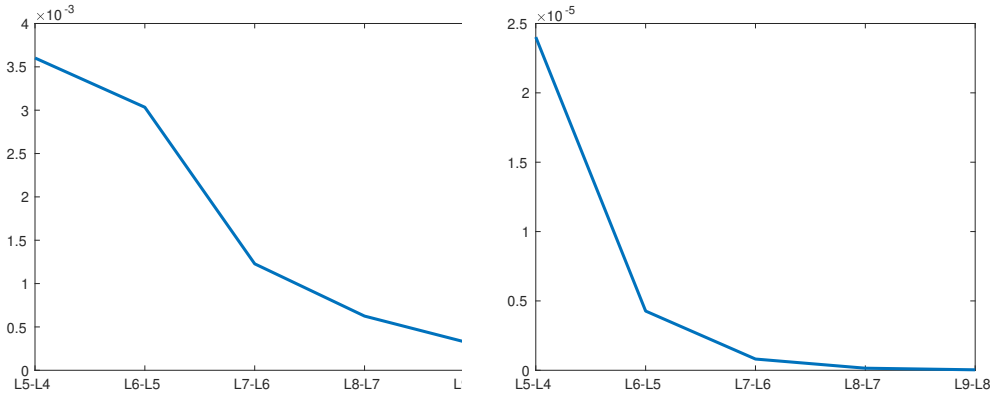

Figure 12: Weak (left) and strong (right) convergences computed for Levels 1 and 0, 2 and 1, 3 and 2, 4 and 3, and 5 and 4 (horizontal axis) at the fixed point $(t, x, y) = (14, 1.60, -0.95)$.

We use computed variances V_ℓ and computing times (work) s_ℓ from Table 1 to estimate the optimal number of samples m_ℓ and compute the telescopic sum in the MLMC method to approximate the expectation.

The standard MLMC theory tells that $\mathbb{E}[Y] = \mathbb{E}[g_L] \approx \mathbb{E}[g]$, where Y denotes a multilevel estimator of $\mathbb{E}[g]$. The mean squared error (MSE) is used to measure the quality of the multilevel estimator:

$$\text{MSE} := \mathbb{E}\left[(Y - \mathbb{E}[g])^2\right] = \mathbb{V}[Y] + (\mathbb{E}[Y] - \mathbb{E}[g])^2 . \tag{14}$$

Now we compare the computational cost of MLMC and standard MC. We denote the cost of computing $g_\ell - g_{\ell-1}$ by s_ℓ. Then the total cost of MLMC is $S = \varepsilon^{-2}\left(\sum_{\ell=0}^{L}\sqrt{V_\ell s_\ell}\right)^2$. The estimated cost of the standard MC method is $S_{\text{MC}} = \varepsilon^{-2} s_L \cdot \mathbb{V}[g_0]$ (see all details in [13, 39, 40]).

We observe that for $\varepsilon^2 = \{0.01, 0.0025, 0.0001, 0.000049\}$ the MC cost, is equal to

$$\{2.0 \cdot 10^3, 2.8 \cdot 10^5, 3.1 \cdot 10^8, 6.3 \cdot 10^8\},$$

whereas the MLMC cost is

$$\{6.4 \cdot 10^1, 1.3 \cdot 10^3, 8.9 \cdot 10^4, 1.8 \cdot 10^5\}.$$

The corresponding numbers of samples on each level are

$$\{\{m_0, m_1, m_2, m_3\}\} = \{\{44, 5, 0, 0\}, \{404, 49, 3, 0\}, \{16672, 1990, 120, 4\}, \{34024, 4062, 245, 7\}\}.$$

We see that MLMC could be much faster than MC, especially for small MSE errors ε^2.

5 CONCLUSION

We have investigated the applicability and efficiency of the MLMC approach to the Henry problem with uncertain porosity, permeability and recharge. These uncertain parameters were modelled by random fields and a random variable. The well-known ug4 parallel multigrid solver was used to solve each random realisation. To investigate the propagation of uncertainty, we used the standard vanilla MLMC method [13]. By calculating the variance decay and computational cost at each level, we estimated the number of random samples required at each level. The estimates depend on the minimisation function in the MLMC algorithm.

The difficulty observed is that the number of samples depends on the chosen point $(t, \mathbf{x})$, i.e., on the QoI. The optimal number of samples is also a function of time.

To demonstrate propagation of uncertainties, we computed the expectation and variance of the mass fraction in the whole domain, the evolution of the pdfs, the solutions in a few points $(t, \mathbf{x})$ and the time evolution of the fresh water and salt integral values.

For some QoIs, it may be that 2-3 mesh levels are sufficient for the uncertain input parameters under consideration and the use of more refined meshes would not significantly improve the result. A different type of porosity in eq. 11 may lead to a different conclusion.

In the future, we plan to consider a more complicated model for recharge, one that depends on space coordinate and time. We are also planning to do some more realistic modelling of porosity and permeability (e.g. of a multi-scale nature). Other important topics for further research are data assimilation and the identification of unknown parameters [21, 25, 27, 32, 24]. Known experimental data and measurements of porosity, permeability, velocity or mass fraction could be used to minimise uncertainties.

Acknowledgments

We thank the KAUST HPC support team for assistance with Shaheen II. This work was supported by the Alexander von Humboldt foundation.

REFERENCES

[1] E. ABARCA, J. CARRERA, X. SÁNCHEZ-VILA, AND M. DENTZ, *Anisotropic dispersive henry problem*, Advances in Water Resources, 30 (2007), pp. 913–926.

[2] R. BARRETT, M. BERRY, T. F. CHAN, J. DEMMEL, J. DONATO, J. DONGARRA, V. EIJKHOUT, R. POZO, C. ROMINE, AND H. VAN DER VORST, *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*, Society for Industrial and Applied Mathematics, 1994.

[3] F. BODE, T. FERRÉ, N. ZIGELLI, M. EMMERT, AND W. NOWAK, *Reconnecting stochastic methods with hydrogeological applications: A utilitarian uncertainty analysis and risk assessment approach for the design of optimal monitoring networks*, Water Resources Research, 54 (2018), pp. 2270–2287.

[4] J. CARRERA, *An overview of uncertainties in modelling groundwater solute transport*, Journal of Contaminant Hydrology, 13 (1993), pp. 23 – 48. Chemistry and Migration of Actinides and Fission Products.

[5] K. CLIFFE, M. GILES, R. SCHEICHL, AND A. TECKENTRUP, *Multilevel monte carlo methods and applications to elliptic pdes with random coefficients*, Computing and Visualization in Science, 14 (2011), pp. 3–15.

[6] N. COLLIER, A.-L. HAJI-ALI, F. NOBILE, E. VON SCHWERIN, AND R. TEMPONE, *A continuation multilevel monte carlo algorithm*, BIT Numerical Mathematics, 55 (2015), pp. 399–432.

[7] A. COSTA, *Permeability-porosity relationship: A reexamination of the kozeny-carman equation based on a fractal pore-space geometry assumption*, Geophysical Research Letters, 33 (2006).

[8] L. DHAL AND S. SWAIN, *Understanding and modeling the process of seawater intrusion: a review*, 01 2022, pp. 269–290.

[9] P. FROLKOVIČ AND H. DE SCHEPPER, *Numerical modelling of convection dominated transport coupled with density driven flow in porous media*, Advances in Water Resources, 24 (2001), pp. 63–72.

[10] P. FROLKOVIČ, *Consistent velocity approximation for density driven flow and transport*, in Advanced Computational Methods in Engineering, Part 2: Contributed papers, R. Van Keer and at al., eds., Maastricht, 1998, Shaker Publishing, pp. 603–611.

[11] P. FROLKOVIČ AND P. KNABNER, *Consistent velocity approximations in finite element or volume discretizations of density driven flow*, in Computational Methods in Water Resources XI, A. A. Aldama and et al., eds., Southhampten, 1996, Computational Mechanics Publication, pp. 93–100.

[12] M. B. GILES, *Multilevel Monte Carlo path simulation*, Operations Research, 56 (2008), pp. 607–617.

[13] ——, *Multilevel Monte Carlo methods*, Acta Numerica, 24 (2015), pp. 259–328.

[14] W. HACKBUSCH, *Multi-Grid Methods and Applications*, Springer, Berlin, 1985.

[15] W. HACKBUSCH, *Iterative Solution of Large Sparse Systems of Equations*, Springer, New-York, 1994.

[16] A.-L. HAJI-ALI, F. NOBILE, E. VON SCHWERIN, AND R. TEMPONE, *Optimization of mesh hierarchies in multilevel Monte Carlo samplers*, Stoch. Partial Differ. Equ. Anal. Comput., 4 (2016), pp. 76–112.

[17] H. R. HENRY, *Effects of dispersion on salt encroachment in coastal aquifers, in 'seawater in coastal aquifers'*, US Geological Survey, Water Supply Paper, 1613 (1964), pp. C70–C80.

[18] B. N. KHOROMSKIJ, A. LITVINENKO, AND H. MATTHIES, *Application of hierarchical matrices for computing the Karhunen-Loève expansion*, Computing, 84 (2009), pp. 49–67.

[19] T. LAATTOE, A. WERNER, AND C. SIMMONS, *Seawater Intrusion Under Current Sea-Level Rise: Processes Accompanying Coastline Transgression*, Springer Netherlands, Dordrecht, 2013, pp. 295–313.

[20] A. LITVINENKO, D. KEYES, V. KHOROMSKAIA, B. N. KHOROMSKIJ, AND H. G. MATTHIES, *Tucker tensor analysis of matérn functions in spatial statistics*, Computational Methods in Applied Mathematics, 19 (2019), pp. 101–122.

[21] A. LITVINENKO, R. KRIEMANN, M. G. GENTON, Y. SUN, AND D. E. KEYES, *Hlibcov: Parallel hierarchical matrix approximation of large covariance matrices and likelihoods with applications in parameter identification*, MethodsX, 7 (2020), p. 100600.

[22] A. LITVINENKO, D. LOGASHENKO, R. TEMPONE, G. WITTUM, AND D. KEYES, *Solution of the 3d density-driven groundwater flow problem with uncertain porosity and permeability*, GEM - International Journal on Geomathematics, 11 (2020), p. 10.

[23] ——, *Propagation of uncertainties in density-driven flow*, in Sparse Grids and Applications — Munich 2018, H.-J. Bungartz, J. Garcke, and D. Pflüger, eds., Cham, 2021, Springer International Publishing, pp. 101–126.

[24] A. LITVINENKO AND H. G. MATTHIES, *Inverse problems and uncertainty quantification*, 2014.

[25] A. LITVINENKO, Y. SUN, M. G. GENTON, AND D. E. KEYES, *Likelihood approximation with hierarchical matrices for large spatial datasets*, Computational Statistics & Data Analysis, 137 (2019), pp. 115–132.

[26] A. LITVINENKO, A. C. YUCEL, H. BAGCI, J. OPPELSTRUP, E. MICHIELSSEN, AND R. TEMPONE, *Computation of electromagnetic fields scattered from objects with uncertain shapes using multilevel monte carlo method*, IEEE Journal on Multiscale and Multiphysics Computational Techniques, 4 (2019), pp. 37–50.

[27] H. G. MATTHIES, E. ZANDER, B. V. ROSIĆ, AND A. LITVINENKO, *Parameter estimation via conditional expectation: a bayesian inversion*, Advanced Modeling and Simulation in Engineering Sciences, 3 (2016), p. 24.

[28] M. PANDA AND W. LAKE, *Estimation of single-phase permeability from parameters of particle-size distribution*, AAPG Bull., 78 (1994), pp. 1028–1039.

[29] H. PAPE, C. CLAUSER, AND J. IFFLAND, *Permeability prediction based on fractal pore-space geometry*, Geophysics, 64 (1999), pp. 1447–1460.

[30] S. REITER, A. VOGEL, I. HEPPNER, M. RUPP, AND G. WITTUM, *A massively parallel geometric multigrid solver on hierarchically distributed grids*, Computing and Visualization in Science, 16 (2013), pp. 151–164.

[31] M. RIVA, A. GUADAGNINI, AND A. DELL'OCA, *Probabilistic assessment of seawater intrusion under multiple sources of uncertainty*, Advances in Water Resources, 75 (2015), pp. 93–104.

[32] B. Rosić, A. Kučerová, J. Sýkora, O. Pajonk, A. Litvinenko, and H. Matthies, *Parameter identification in a probabilistic setting*, Engineering Structures, 50 (2013), pp. 179 – 196. Engineering Structures: Modelling and Computations (special issue IASS-IACM 2012).

[33] Y. Rubin, *Applied stochastic hydrogeology*, Oxford University Press, 2003.

[34] A. Schneider, H. Zhao, J. Wolf, D. Logashenko, S. Reiter, M. Howahr, M. Eley, M. Gelleszun, and H. Wiederhold, *Modeling saltwater intrusion scenarios for a coastal aquifer at the german north sea*, E3S Web Conf., 54 (2018), p. 00031.

[35] M. J. Simpson and T. Clement, *Theoretical Analysis of the worthiness of Henry and Elder problems as benchmarks of density-dependent groundwater flow models*, Adv. Water. Resour., 26 (2003), pp. 17–31.

[36] M. J. Simpson and T. P. Clement, *Improving the worthiness of the Henry problem as a benchmark for density-dependent groundwater flow models*, Water Resources Research, 40 (2004), p. W01504.

[37] L. Stoeckl, M. Walther, and L. K. Morgan, *Physical and numerical modelling of post-pumping seawater intrusion*, Geofluids, 2019 (2019).

[38] D. Tartakovsky, *Assessment and management of risk in subsurface hydrology: A review and perspective*, Advances in Water Resources, 51 (2013), pp. 247 – 260. 35th Year Anniversary Issue.

[39] A. Teckentrup, P. Jantsch, C. Webster, and M. Gunzburger, *A multilevel stochastic collocation method for partial differential equations with random input data*, SIAM/ASA Journal on Uncertainty Quantification, 3 (2015), pp. 1046–1074.

[40] A. L. Teckentrup, R. Scheichl, M. B. Giles, and E. Ullmann, *Further analysis of multilevel Monte Carlo methods for elliptic PDEs with random coefficients*, 125 (2013), pp. 569–600.

[41] H. Vereecken, A. Schnepf, J. Hopmans, M. Javaux, D. Or, T. Roose, J. Vanderborght, M. Young, W. Amelung, M. Aitkenhead, S. Allison, S. Assouline, P. Baveye, M. Berli, N. Brüggemann, P. Finke, M. Flury, T. Gaiser, G. Govers, T. Ghezzehei, P. Hallett, H. Hendricks Franssen, J. Heppell, R. Horn, J. Huisman, D. Jacques, F. Jonard, S. Kollet, F. Lafolie, K. Lamorski, D. Leitner, A. McBratney, B. Minasny, C. Montzka, W. Nowak, Y. Pachepsky, J. Padarian, N. Romano, K. Roth, Y. Rothfuss, E. Rowe, A. Schwen, J. Šimůnek, A. Tiktak, J. Van Dam, S. van der Zee, H. Vogel, J. Vrugt, T. Wöhling, and I. Young, *Modeling soil processes: Review, key challenges, and new perspectives*, Vadose Zone Journal, 15 (2016), p. vzj2015.09.0131.

[42] A. Vogel, S. Reiter, M. Rupp, A. Nägel, and G. Wittum, *Ug 4: A novel flexible software system for simulating pde based models on high performance computers*, Computing and Visualization in Science, 16 (2013), pp. 165–179.

[43] C. Voss and W. Souza, *Variable density flow and solute transport simulation of regional aquifers containing a narrow freshwater-saltwater transition zone*, Water Resources Research, 23 (1987), pp. 1851–1866.

A The mass fraction$c(t,\mathbf{x})$ at 12 preselected points

We examine the solution $c(t,\mathbf{x})$ at 12 preselected points. Figure 13 includes 12 subfigures. Each subfigure presents 100 QMC realizations of $c(t,\mathbf{x})$ and five quantiles depicted by dotted lines. The dotted line at the bottom indicates the quantile 0.025. The following dotted line is the quantile 0.25, and the dotted line on the top indicates the quantile 0.975. All five quantiles from the bottom to the top are 0.025, 0.25, 0.50, 0.75, and 0.975, respectively. We observe that c at the final point $t = T$ varies considerably.

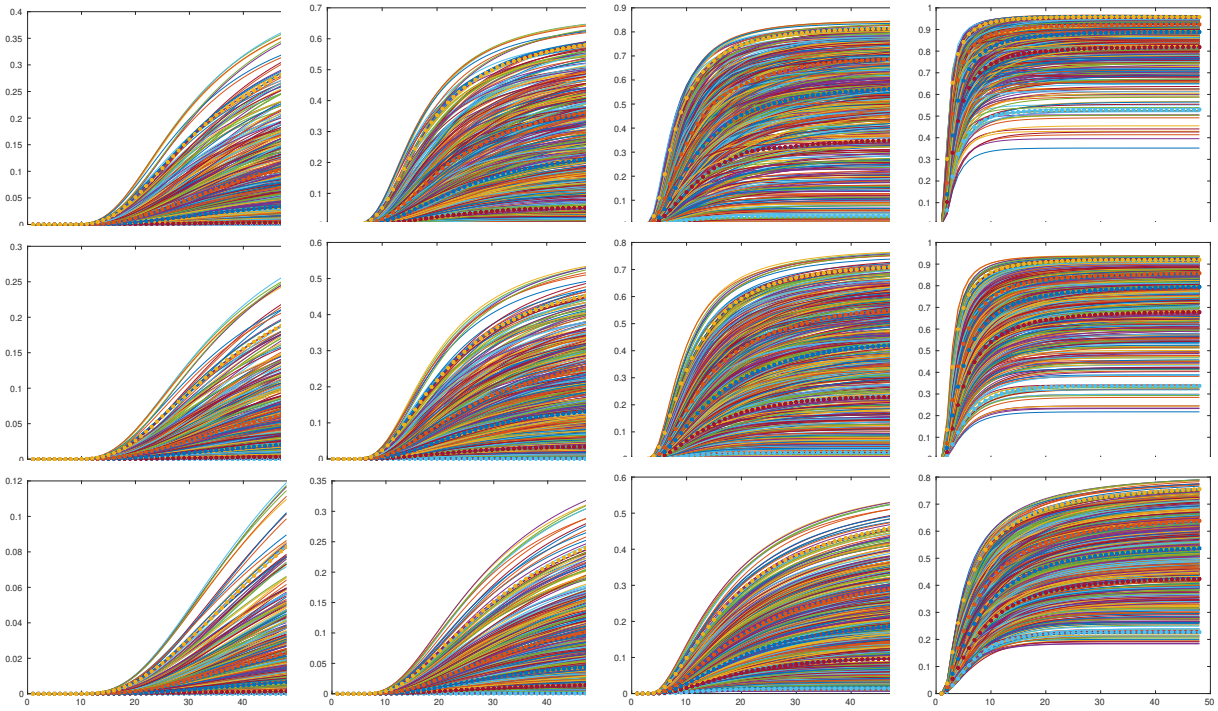

Figure 13: One hundred QMC realisations of $c(t,\mathbf{x})$ at 12 $\mathbf{x}$-points. First row points: $\{(1.10,-0.95),(1.35,-0.95),(1.60,-0.95),(1.85,-0.95)\}$, second row points: $\{(1.10,-0.75),(1.35,-0.75),(1.60,-0.75),(1.85,-0.75)\}$, and third row points: $\{(1.10,-0.50),(1.35,-0.50),(1.60,-0.50),(1.85,-0.50)\}$. Dotted lines from the bottom to the top indicate the quantiles 0.025, 0.25, 0.50, 0.75, and 0.975, respectively.

B Parameters of the problem

The following parameters were used for all numerical tests.

Parameter	Values and Units	Description
$\mathbb{E}[\phi]$	0.35 [-]	the mean value of the porosity
D	$18.8571 \cdot 10^{-6}$ $[\mathrm{m}^2 \cdot \mathrm{s}^{-1}]$	diffusion coefficient in medium
$\mathbf{K}$	$1.020408 \cdot 10^{-9}$ $[\mathrm{m}^2]$	permeability of the medium
g	9.8 $[\mathrm{m} \cdot \mathrm{s}^{-2}]$	gravity
ρ_0	1000 $[\mathrm{kg} \cdot \mathrm{m}^{-3}]$	density of pure water
ρ_1	1024.99 $[\mathrm{kg} \cdot \mathrm{m}^{-3}]$	density of brine
μ	10^{-3} $[\mathrm{kg} \cdot \mathrm{m}^{-1} \cdot \mathrm{s}^{-1}]$	viscosity

Table 2: Parameters of the considered density driven flow problem