# Integration of moment equations in a reduced-order modeling strategy for Monte Carlo simulations of groundwater flow

Chuan-An Xia[a], Damiano Pasetto[b,*], Bill X. Hu[a,*], Mario Putti[c], Alberto Guadagnini[d]

[a]*Institute of Groundwater and Earth Science, Jinan University, Guangzhou, China*
[b]*Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice, Italy*
[c]*Dipartimento di Matematica "Tullio Levi-Civita", University of Padova, Italy*
[d]*Dipartimento di Ingegneria Civile e Ambientale, Politecnico di Milano, Italy*

**Abstract**

We illustrate and test an approach grounded on embedding moment equations (MEs) of groundwater flow within a Monte Carlo based modeling strategy to yield a Reduced-Order Model (ROM) that enables the efficient and accurate evaluation of probability distributions of hydraulic heads in randomly heterogeneous transmissivity fields. The projection space determining the accuracy of the ROM solution is typically computed through the principal component analysis of a selected number of full system model solutions (the so-called snapshots). Computationally expensive sensitivity analyses are then required to assess the independence of the ROM from the snapshots. Here, we propose to compute the projection vectors upon relying on the hydraulic head covariance evaluated from the solution of corresponding MEs of groundwater flow.

*Corresponding authors. Email: damiano.pasetto@unive.it, billhu@jnu.edu.cn

Our workflow to compute hydraulic head distributions is organized according to the following steps: (i) approximation of mean hydraulic head and head covariance matrix through (second-order accurate) solutions of MEs; (ii) computation of the leading eigenvectors of the head covariance matrix to form the basis set for the ROM projection space; and (iii) construction of the ROM. Sample probability density functions of hydraulic heads are then efficiently obtained via Monte Carlo simulations relying on the developed ROM. The proposed methodology is compared against snapshot-based ROMs and the full system model in a two- and a three-dimensional steady-state groundwater flow setting where pumping from a point source is superimposed to a mean uniform flow. Our results show that the projection space computed by relying on MEs provides a more accurate ROM solution than the one resulting from reliance on snapshots.

*Keywords:* Reduced-Order Monte Carlo simulation, Stochastic moment equations, Proper Orthogonal Decomposition, Heterogeneous medium

## 1. Introduction

Numerical solutions of flow and/or transport processes taking place in environmental systems at a scale of practical interest are typically associated with high-computational burden [1–3]. This aspect might hamper the use of modern uncertainty quantification techniques, local and/or global sensitivity analyses, and decision support tools under uncertainty. In the presence of spatially-distributed random system parameters, the Probability Density Function (PDF) of a target environmental variable is usually estimated through Monte Carlo (MC) approaches. These in turn commonly

require multiple computationally-expensive solutions of a Full System Model (FSM) considered to represent the behavior of the studied physical system and to yield the relevant modeling goals.

A stark example is given by groundwater related scenarios, where soil parameters, such as porosity and hydraulic conductivity, are frequently conceptualized as spatial random fields. In this context, MC methods play a significant role in dealing with uncertainty quantification [4–6], uncertainty reduction (e.g., in the framework of data assimilation techniques, such as the ensemble Kalman filter and/or smoother [7–10]), and variance- or moment-based global sensitivity analyses [11–15]. Implementation of MC relies on performing multiple forward simulations of the selected forward FSM upon using a collection of independent realizations of the uncertain model parameters. While MC approaches are conceptually straightforward, convergence of the empirical/sample moments of the target model outputs (e.g., variance of hydraulic head or solute concentration) is generally slow, i.e., it requires a large number of forward model solutions. In this context, Reduced-Order Models (ROMs, also termed as surrogate models, metamodels or proxies) that can approximate the FSM solution at a reduced computational cost can be attractive to assist model development and probabilistic risk assessment.

Surrogate models have been developed with the aim of reducing computational burden while capturing the main features of model outcomes. Data-driven, projection-based, and multi-fidelity models are three categories of surrogate models [16, 17]. Data-driven proxies [18] are commonly built by training the surrogate model on available data. Projection-based proxies (which are also known as reduced-order, or reduced-basis methods) rest on

a projection of the FSM governing equation (including initial and boundary conditions) onto a suitable low-dimensional space. For example, the idea underpinning Reduced-Order Monte Carlo (ROMC) [19, 20] relies on replacing the collection of FSM solutions with solutions of a projection-based surrogate model, yielding an increased computational efficiency. Multi-fidelity models [21, 22] are built by (a) relying on coarse grids to capture solutions associated with large-scale structures or (b) simplifying physical-chemical processes included in the FSM. Razavi et al. [16] and Asher et al. [17] provided comprehensive reviews to the theoretical elements underlying these types of surrogate models, including their limitations and applications in groundwater modeling. A detailed review of projection-based surrogate models is presented by Chen et al. [23].

Here, we focus on projection-based surrogate models for MC-based solutions of groundwater flow in the presence of randomly distributed hydraulic parameters. In this case, surrogate models generally rely on projecting hydraulic heads onto a few dominant basis functions, making use of the so-called Galerkin projection. The basis functions used for the projection can be computed by Proper Orthogonal Decomposition (POD) of a certain set of snapshots. The latter corresponds to a selected collection of solutions of the flow problem, each associated with a random realization of the hydraulic parameter field, typically hydraulic conductivity. POD selects projection vectors as the most energetic modes (or principal components) associated with the snapshots. These modes are computed as the eigenvectors of the covariance matrix of the snapshots related to the largest eigenvalues and describe the spatial variability of the FSM solutions [24]. Modes' coefficients

required to approximate each realization are computed by imposing that the residual of the model equation associated with the approximated solution be orthogonal to the projection vectors [19]. Thus, these coefficients are the solution of a low-dimensional linear system rather than of the FSM, resulting in considerable CPU time savings. Projection-based surrogate models have received increasing attention in modeling groundwater flow and transport [16, 19, 20, 25–27], because of their rigorous mathematical foundation and high accuracy. Deterministic scenarios have been tackled in a variety of studies, as summarized in the following. Siade et al. [28] considered deterministic groundwater flow under transient conditions, snapshots in this setting corresponding to solutions of the FSM at selected observation times. Li et al. [29] developed a POD-based reduced order model for a variable-density flow. Boyce et al. [27] and Stanko et al. [26] used POD to reduce the order of a model describing unconfined groundwater flow. Several authors [30–32] have explored model reduction techniques to simulate flow and solute transport in (otherwise deterministic) homogeneous and/or heterogeneous aquifer systems.

Although convergence of greedy algorithms towards data-independent ROMs is theoretically guaranteed [33, 34], the accuracy of the ROM solution in practical applications strongly depends on the method used for the snapshot selection and their number (as also pointed out by Asher et al. [17]), thus limiting its computational efficiency when too many snapshots are required. There is clearly a trade-off between the ROM accuracy and its computational cost, i.e., an enhanced size of the sample of snapshots yields more accurate ROM solutions, at the expense of an increased computational

cost. Several studies focus on the search for the most effective basis functions by selecting snapshots along the direction that maximize the error on the residual, i.e., selecting the more informative snapshots on the basis of a greedy algorithm [35–37]. Pasetto et al. [20] used a similar greedy algorithm to develop a computationally advantageous reduced-order MC approach for a transient flow scenario when the random transmissivity field is described through zonation. In a similar simulation scenario depicting a transient flow in a confined aquifer, Boyce and Yeh [38] used two greedy algorithms to iteratively estimate the adequate number of snapshots and their appropriate selection times.

In the presence of randomly heterogeneous transmissivity fields, the application of the greedy algorithm is more challenging due to the large size of the model parameter set. Lieberman et al. [39] proposed a greedy algorithm for the simultaneous selection of the reduced state and parameter spaces in the presence of randomly distributed conductivity. Pasetto et al. [25] showed that the ROM error decreases slowly with the number of snapshots for large variances and small integral scales of the random conductivity field. As a consequence, there is the possibility that the greedy algorithm does not lead to the desired accuracy by relying on a reasonably small snapshot set.

In this broad framework, our study aims at fully integrating Moment Equations (MEs) of groundwater flow [40–43] within a model order reduction approach to efficiently perform numerical MC simulations in the presence of randomly heterogeneous hydraulic parameters. We do so by improving the construction of the projection space upon relying on the evaluation of the first and second statistical moments (i.e., expectation and covariance) of hy-

draulic heads as the solutions of corresponding MEs of groundwater flow in a randomly heterogeneous domain. The latter are deterministic equations rendering the (ensemble) moments of hydraulic head $h(\mathbf{x})$ and Darcy flux $\mathbf{q}(\mathbf{x})$ at location vector $\mathbf{x}$. We focus on MEs of fully-saturated steady state confined groundwater flow (see, e.g., [40] for the derivation of near-exact (deterministic) integro-differential equations governing the space-time evolution of (ensemble) mean heads and [41] or [44] for a review on moment differential equations for groundwater flow in highly heterogeneous porous media). Numerical approaches have been developed to apply MEs to quantify uncertainty associated with forward and inverse modeling of groundwater flow in porous media of randomly heterogeneous conductivity including application to field scenarios (see, e.g., [45–47] and references therein). Most recent developments allowed embedding stochastic MEs of transient groundwater flow in data assimilation and parameter estimation approaches via ensemble Kalman filter [46, 48, 49].

With respect to the snapshot technique, the approach we present, i.e., the derivation of the ROM projection vectors from the solution of moment equations, has the advantage that neither the snapshot size should be predetermined, nor procedures associated with imposed validation conditions (e.g., greedy algorithm) should be required. Alternative ways to combine moment equations and model reduction approaches have been presented to reduce the computational costs. For example, Yang et al. [50] and Zhang and Lu [51] relied on the Karhunen-Loève expansion to obtain higher order approximations of the first and second (ensemble) moments of heads. Li and Zhang [52] compared this approach to a collocation method and polynomial

chaos expansion, highlighting their advantages with respect to traditional MC for the evaluation of the first and second moments of the heads.

The technique we propose in this study is substantially different from these, the main idea being the use of MEs to construct the surrogate model with which one can perform standard MC simulations to yield approximations of the complete head probability density function. The latter then (in principle) enables one to evaluate (ensemble) moments of any order.

Consistent with the typical workflow of ROMs, our computational procedure is structured across two stages, hereafter termed as offline and online components, respectively. The offline operations include (*i*) the solution of the MEs and (*ii*) the computation of the leading eigenvectors of the head covariance matrix. These eigenvectors are used as basis for the ROM projection space and enable the construction of a computationally efficient and accurate surrogate model. These operations are performed only once, thus entailing relatively high CPU time savings in the online stage, which is in turn keyed to the application of the reduced-order model in the MC framework. Thus, the online phase comprises an MC iteration where the full model is replaced by the MEs-based surrogate model. This approach seamlessly embeds for the first time moment equations solutions within a Monte Carlo context, enabling the evaluation of the full probability distribution of groundwater flow variables.

The structure of the work is detailed in the following. The deterministic partial differential equations governing the (steady-state) spatial distribution of the first two (ensemble) moments of hydraulic head in a randomly heterogeneous transmissivity field are briefly summarized in Section 2.2. The

mathematical derivation of the ROM and its application in the context of MC simulations of groundwater flow are described in Section 3. Illustrative two- and three-dimensional showcase examples are introduced in Section 4, Section 5 is devoted to the analysis and discussion of results. Conclusions are drawn in Section 6.

## 2. Problem setting

We consider steady-state groundwater flow in a confined aquifer denoted with $\Omega \subset \mathbb{R}^r$, $r = 2, 3$, as described by [53]:

$$\begin{cases} \boldsymbol{\nabla} \cdot (T(\mathbf{x}) \boldsymbol{\nabla} h(\mathbf{x})) + f(\mathbf{x}) = 0 & \mathbf{x} \in \Omega \\ h(\mathbf{x}) = H(\mathbf{x}) & \mathbf{x} \in \Gamma_D \\ \mathbf{q}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}) = Q(\mathbf{x}) & \mathbf{x} \in \Gamma_N \end{cases} \tag{1}$$

where $h$ is the hydraulic head; $f$ is a forcing term; $H$ and $Q$ represent the assigned hydraulic head at the Dirichlet boundary $\Gamma_D$ and the assigned flux at the Neumann boundary $\Gamma_N$, respectively; $\mathbf{q}(\mathbf{x}) = -T(\mathbf{x}) \boldsymbol{\nabla} h(\mathbf{x})$ is the flux vector; and $\mathbf{n}$ is the unit vector normal to the domain boundary (positive when pointing outward). Operators and (vector or scalar) quantities are defined consistent with the dimensionality of the scenario investigated, e.g., $T$ is transmissivity (for $r = 2$) or hydraulic conductivity (for $r = 3$).

Here, $T$ is described as a second-order stationary spatial random field with a given distribution. Equation (1) thus becomes a stochastic partial differential equation. While analytical solutions are not available for a generally heterogeneous domain and in the presence of general boundary conditions, numerical approximations have been proposed in literature to estimate the

leading (statistical) moments and/or the probability density function (PDF) of heads. As seen in Section 1, these include, e.g., the MEs or the MC approach. We provide a brief illustration of these approaches in the following.

### *2.1. Moment equations of groundwater flow*

The key idea underpinning the moment equations (MEs) method (see, e.g., [41] and references therein) is to solve equation (1) by deriving the equations satisfied by the (ensemble) moments of hydraulic heads and fluxes. Solving these equations entails expanding all moments appearing in them in terms of a small parameter, $\sigma_Y$, representing the standard deviation of the natural logarithm of $T$, i.e., $Y = \ln T$. Hydraulic head can then be expanded as $h = h^{(0)} + h^{(1)} + h^{(2)} + \ldots$, where $h^{(n)} = O(\sigma_Y^n)$ (see [41] and references therein). One then obtains a set of recursive approximations of the otherwise exact MEs that can be solved by truncating the head expansion up to a given order (second order being typically used for computational limitations). Denoting the ensemble mean (expected value) of $Y$ with $\langle Y \rangle$ and its standard deviation with $\sigma_Y$, (ensemble) mean heads are approximated as (see, e.g., [41] and references therein):

$$\langle h(\mathbf{x}) \rangle \approx \left\langle h^{(0)}(\mathbf{x}) \right\rangle + \left\langle h^{(2)}(\mathbf{x}) \right\rangle \tag{2}$$

where $\left\langle h^{(0)}(\mathbf{x}) \right\rangle$ and $\left\langle h^{(2)}(\mathbf{x}) \right\rangle$ are zero- and second-order (in terms of $\sigma_Y$) components of the mean head, respectively (note that in this case the mean of the first-order correction term is zero, i.e., $\left\langle h^{(1)} \right\rangle = 0$ [41]). Similarly, for the fluxes:

$$\langle \mathbf{q}(\mathbf{x}) \rangle \approx \left\langle \mathbf{q}^{(0)}(\mathbf{x}) \right\rangle + \left\langle \mathbf{q}^{(2)}(\mathbf{x}) \right\rangle \tag{3}$$

where

$$\left\langle \mathbf{q}^{(0)}\left(\mathbf{x}\right)\right\rangle = -\boldsymbol{\nabla}\cdot\left[T_G\left(\mathbf{x}\right)\boldsymbol{\nabla}\left\langle h^{(0)}(\mathbf{x})\right\rangle\right] \tag{4}$$

and

$$\left\langle \mathbf{q}^{(2)}\left(\mathbf{x}\right)\right\rangle = -T_G(\mathbf{x})\left[\boldsymbol{\nabla}\left\langle h^{(2)}\left(\mathbf{x}\right)\right\rangle + \frac{\sigma_Y^2}{2}\boldsymbol{\nabla}\left\langle h^{(0)}\left(\mathbf{x}\right)\right\rangle\right] + \mathbf{r}^{(2)}\left(\mathbf{x}\right) \tag{5}$$

Here, $\mathbf{r}^{(2)}\left(\mathbf{x}\right)$ is a second-order residual flux, and $T_G\left(\mathbf{x}\right) = e^{\langle Y(\mathbf{x})\rangle}$ is the geometric mean of the transmissivity field.

Computation of the zero-order mean head is straightforward and details are here omitted (see [41, 54] and references therein). The equation satisfied by the second-order mean head $\left\langle h^{(2)}\left(\mathbf{x}\right)\right\rangle$ is [41]:

$$\begin{cases} \boldsymbol{\nabla}\cdot\left[T_G(\mathbf{x})\left(\boldsymbol{\nabla}\left\langle h^{(2)}\left(\mathbf{x}\right)\right\rangle + \frac{\sigma_Y^2}{2}\boldsymbol{\nabla}\left\langle h^{(0)}\left(\mathbf{x}\right)\right\rangle\right) - \mathbf{r}^{(2)}\left(\mathbf{x}\right)\right] = 0 & \mathbf{x}\in\Omega \\ \left\langle h^{(2)}\left(\mathbf{x}\right)\right\rangle = 0 & \mathbf{x}\in\Gamma_D \\ \mathbf{q}^{(2)}(\mathbf{x})\cdot\mathbf{n}(\mathbf{x}) = 0 & \mathbf{x}\in\Gamma_N \end{cases} \tag{6}$$

The (second-order) cross covariance between $h$ and $T$, $u^{(2)}\left(\mathbf{y},\mathbf{x}\right) = \left\langle T'\left(\mathbf{y}\right)h'\left(\mathbf{x}\right)\right\rangle^{(2)}$, primed quantities denoting fluctuations with respect to their corresponding means, is computed as the solution of the following equation defined for all $\mathbf{y}\in\Omega$ [41]:

$$\begin{cases} \boldsymbol{\nabla}\cdot\left[T_G(\mathbf{x})\boldsymbol{\nabla}u^{(2)}\left(\mathbf{y},\mathbf{x}\right) - T_G(\mathbf{y})C_Y\left(\mathbf{x},\mathbf{y}\right)\left\langle \mathbf{q}^{(0)}\left(\mathbf{x}\right)\right\rangle\right] = 0 & \mathbf{x}\in\Omega \\ u^{(2)}\left(\mathbf{y},\mathbf{x}\right) = 0 & \mathbf{x}\in\Gamma_D \\ \left[T_G(\mathbf{x})\boldsymbol{\nabla}u^{(2)}\left(\mathbf{y},\mathbf{x}\right) - T_G(\mathbf{y})C_Y\left(\mathbf{x},\mathbf{y}\right)\left\langle \mathbf{q}^{(0)}\left(\mathbf{x}\right)\right\rangle\right]\cdot\mathbf{n}(\mathbf{x}) = 0 & \mathbf{x}\in\Gamma_N \end{cases} \tag{7}$$

where $C_Y\left(\mathbf{x},\mathbf{y}\right) = \left\langle Y'\left(\mathbf{x}\right)Y'\left(\mathbf{y}\right)\right\rangle$ is the covariance of $Y$. The equation for

the second-order head covariance $C_h^{(2)}$ is [41]:

$$\begin{cases} \boldsymbol{\nabla} \cdot \left[ T_G(\mathbf{x}) \boldsymbol{\nabla} C_h^{(2)}(\mathbf{x},\mathbf{y}) + u^{(2)}(\mathbf{x},\mathbf{y}) \boldsymbol{\nabla} \left\langle h^{(0)}(\mathbf{x}) \right\rangle \right] = 0 & \mathbf{x} \in \Omega \\ C_h^{(2)}(\mathbf{x},\mathbf{y}) = 0 & \mathbf{x} \in \Gamma_D \\ \left[ T_G(\mathbf{x}) \boldsymbol{\nabla} C_h^{(2)}(\mathbf{x},\mathbf{y}) + u^{(2)}(\mathbf{x},\mathbf{y}) \boldsymbol{\nabla} \left\langle h^{(0)}(\mathbf{x}) \right\rangle \right] \cdot \mathbf{n}(\mathbf{x}) = 0 & \mathbf{x} \in \Gamma_N \end{cases} \tag{8}$$

The second-order residual flux $\mathbf{r}^{(2)}$ is then obtained as [41, 54]:

$$\mathbf{r}^{(2)}(\mathbf{x}) = -\left\langle T'(\mathbf{x}) \boldsymbol{\nabla} h'(\mathbf{x}) \right\rangle^{(2)} = \lim_{y \to x} \left[ -\boldsymbol{\nabla}_x u^{(2)}(\mathbf{y},\mathbf{x}) \right] \tag{9}$$

Details about the numerical solution through the finite element method for equations (2)- (9) are described in, e.g, [54], who performed a detailed analysis of the solution accuracy as a function of the size of the numerical grid employed. We note that computing the numerical approximation of the head covariance matrix $\mathbf{C}_h^{(2)}$ on a grid with $n$ nodes entails $(i)$ solving equation (7) $n$ times to evaluate the cross-covariance matrix $\mathbf{u}^{(2)}$ and $(ii)$ solving equation (8) $n$ times to obtain $\mathbf{C}_h^{(2)}$. These $2n$ solutions can be computationally expensive, the discretized governing equations having the same system size, $n$. However, it is noted that the full-system stiffness matrix (equation (10)) needs to be computed only once and is then stored [41] to serve all $2n$ runs. Computational efficiency can then be achieved by direct (sparse) factorization as in Xia et al. [54] or through iterative solvers, in case a direct solver is too expensive.

### *2.2. Monte Carlo simulation*

When compared against an MEs-based approach, MC schemes have the advantage of providing an empirical approximation of the full head PDF

through a straightforward implementation. MEs are typically used only for the evaluation of the first and second moments of the given modeling target, and do not provide additional information about possible appearance of heavy tails of the associated PDF, which however can have a critical importance in risk assessment scenarios. The main drawback of MC methods is their slow convergence, so that a high number of model runs is often required to obtain reliable results in engineering applications.

MC techniques start by sampling $N_{en}$ independent and identically distributed realizations of the random parameter $Y$, here indicated with $Y^1, \ldots, Y^{N_{en}}$. Then, for any realization $Y^i$, the linear Galerkin Finite Element Method (FEM) simplifies equation (1) into the the system of linear algebraic equations:

$$\mathbf{A}^i \mathbf{h}^i = \mathbf{b}^i, \quad i = 1, \ldots, N_{en} \tag{10}$$

Here, $\mathbf{A}^i$ is the full-system stiffness matrix, which is typically sparse, symmetric, and positive definite of size $n \times n$ ($n$ is the number of nodes used in the spatial discretization); $\mathbf{h}^i$ is the hydraulic head vector whose components are the head solutions at the nodes; $\mathbf{b}^i$ is the full stress vector, representing the effects of the forcing terms and the given boundary conditions. In this context, we only need to compute $\mathbf{b}$ once and store it.

The empirical PDF of heads (and the ensuing moments) can be evaluated from the realizations $\mathbf{h}^1, \ldots, \mathbf{h}^{N_{en}}$. Appropriate numerical solvers should be employed to solve (10) for each $\mathbf{h}^i$. Depending on the grid size, $n$, and on the problem dimension (two- or three-dimensional flow), the solution is efficiently obtained through sparse direct solvers (e.g., MA57 [55]) or iterative methods such as the preconditioned conjugate gradient [31, 56]. Although this step is

much faster than solving the MEs, a large number of realizations is required to grant stability in the results due to the slow convergence of MC methods, thus resulting in a computationally intensive procedure for most applications.

## 3. Moment equations-based reduced-order model

### *3.1. Reduced-order models*

We follow the methodology presented by Pasetto et al. [25] and construct the ROM by defining a $m$-dimensional projection space with basis vectors $\mathbf{p}_1, \ldots, \mathbf{p}_m$. Each realization of the hydraulic head vector $\mathbf{h}^i$ in equation (10) can be approximated by its projection onto the reduced space:

$$\mathbf{h}^i \approx \widetilde{\mathbf{h}}^i = \langle \mathbf{h} \rangle + \sum_{j=1}^{m} \alpha_j^i \mathbf{p}_j = \langle \mathbf{h} \rangle + \mathbf{P}\boldsymbol{\alpha}^i. \tag{11}$$

Here $\langle \mathbf{h} \rangle$ is the mean vector of hydraulic heads, $\mathbf{P}$ is the projection matrix $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_m]$, and $\alpha_j^i$ is the $j^{th}$ component of the coefficient vector $\boldsymbol{\alpha}^i$.

Substituting equation (11) into equation (10) and imposing the residual to be orthogonal to the projection space $\mathbf{P}$ yields the following linear system for the coefficient vector $\boldsymbol{\alpha}^i$:

$$\mathbf{P}^{\mathrm{T}}\mathbf{A}^i\mathbf{P}\boldsymbol{\alpha}^i = \mathbf{P}^{\mathrm{T}}\mathbf{b} - \mathbf{P}^{\mathrm{T}}\mathbf{A}^i \langle \mathbf{h} \rangle \tag{12}$$

where superscript T represents transpose. The system matrix in equation (12) has size $m \times m$. The solution of equation (12) and the subsequent calculation of $\widetilde{\mathbf{h}}^i$ in equation (11) for $i = 1, \ldots, N_{en}$, where $i$ identifies the $i$-th random realization of the transmissivity field, is termed online phase, as it has to be repeated for each $T^i$. The most expensive (in terms of computational time)

operations in this online phase are those acting on the FSM dimension $n$ and correspond to $(a)$ assembling the stiffness matrix $\mathbf{A}^i$ and $(b)$ performing the projection onto the reduced space.

The most computationally expensive phase of a ROM is the construction of the projection vectors, for which one is required to collect the snapshots by computing a certain number of FSM solutions. However, this operation is typically performed only once, during the so-called offline phase. We recall that the Karhunen-Loève(KL) theorem is at the heart of POD, because it enables expressing the head random field as a linear combination of the dominant eigenfunction solutions of the Fredholm integral equation associated with the covariance function of the head field. The spatial discretization of this eigenfunction problem becomes an eigenvector problem for the head covariance matrix, $\mathbf{C}_h$. Thus, the offline phase in POD aims at computing the dominant eigenvectors of $\mathbf{C}_h$. POD finds a projection basis by approximating the head moments with the empirical moments of a number of FSM solutions (see equation (10)), $\mathbf{h}^1, \ldots, \mathbf{h}^{N_{sn}}$, termed snapshots [25]. The quality of the resulting projection vectors depends on the number of snapshots $N_{sn}$ and on the procedure used for their selection. Fully addressing these issues in the case of stochastic equations is still an open issue. The resulting ROM is here termed as SnapROM.

To avoid computation of the snapshots and to preserve the optimality of the KL eigenvectors, an alternative approximation of the head covariance matrix is required. Here, we propose to approximate the head covariance matrix $\mathbf{C}_h$ with the solution of the MEs (2)-(9). The projection vectors $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_m$ are then calculated as the leading eigenvectors of the covari-

ance matrix, $\mathbf{C}_h^{(2)}$, solution of equation (8). The resulting ROM is here denoted as MEsROM.

As a remark, we note that conductivity distributions in geologic media have sometimes been modeled through a unique, in some cases multimodal, distribution (see, e.g., the model taxonomy illustrated by Winter et al. [44]). It is recognized that this stems from a homogenization of conductivity values within a unique population, while they can be associated with diverse regions characterized by differing geological attributes. One may alternatively conceptualize the porous medium as a random composite system. In the latter setting, the domain is composed of regions (or zones) of geo-materials with contrasting values of hydraulic properties (e.g., [44, 57–59] and reference therein). This conceptualization enables one to represent the system as a set of disjoint blocks (whose location might be uncertain), each constituted by a given geomaterial. As such, conductivities within each block can be represented through a unimodal distribution characterized by a mild variance and the solution of second-order approximations stemming from MEs of groundwater flow can be obtained as implemented by [60]. In this sense, our workflow is fully compatible with these settings, which constitute the context within which our work is framed. The heterogeneous settings we consider can then be seen as representative of the natural variability contained within a geologic unit, which can potentially be depicted through heterogeneity models of the kind we consider [44].

### *3.2. ROM implementation and computational cost*

In this section we compare the computational costs of the models listed in Table 1. These costs are evaluated in terms of the increase of the number

of FLOPS (floating point operations) with the problem size, i.e., the number of nodes $n$ used in the spatial discretization. For example, associating a cost $O(n)$ to an operation indicates that the number of FLOPS required for the task is proportional to $n$.

The operations performed in the offline stage (i.e., computations with the FSM and building of the ROM matrices) are outlined as follows: $(i)$ approximation of the mean hydraulic head and head covariance matrix through (second-order accurate) solutions of MEs; and $(ii)$ computation of the leading eigenvectors of the approximated head covariance matrix, which form the basis for the projection space. A computational disadvantage of the MEsROM approach with respect to SnapROM resides in the diagonalization of the covariance matrix. In this sense, MEsROM requires computing the leading eigenvectors of $\mathbf{C}_h^{(2)}$, which is a full matrix of dimension $n \times n$ (computational cost of order $O(n^3)$). SnapROM circumvents this drawback by relying on the observation that the matrix of the snapshot perturbations around the mean, $\mathbf{X} = 1/\sqrt{N_{sn}} \left[\mathbf{h}'^{,1}, \ldots, \mathbf{h}'^{,N_{sn}}\right]$, yields the snapshot covariance matrix $\mathbf{X}\mathbf{X}^{\mathrm{T}}$. It is easy to demonstrate that each eigenvector $\mathbf{v}_i$ of the small-size matrix $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ is related to the corresponding eigenvector $\mathbf{p}_i$ of the snapshot covariance $\mathbf{X}\mathbf{X}^{\mathrm{T}}$ by $\mathbf{p}_i = \mathbf{X}\mathbf{v}_i$, while keeping the same eigenvalues. Thus, SnapROM solves an eigenvalues problem of dimension $N_{sn} \times N_{sn}$, or, equivalently, applies the singular value decomposition (SVD) to matrix $\mathbf{X}$, with a computational cost of order $O(nN_{sn}^2)$. The online phase for the calculation of $\mathbf{h}^i$ for each realization $i$ requires the following tasks:

1. assembly of the system matrix $\mathbf{A}^i$: using a FEM approach, the cost is $O(3N_e)$ FLOPS, where $N_e$ is the number of triangles in the mesh (typ-

ically $N_e \approx 1.7\ n$ and $5\ n$ in two- and three-dimensional triangulations, respectively); we denote this cost as $O(sn)$, where $s \approx 7$ or $s \approx 15$ in two and three dimensions, respectively;

2. solution of the linear system (10), $\mathbf{A}^i\mathbf{h}^i = \mathbf{b}$, where $\mathbf{A}^i$ is sparse and symmetric. Here we used the direct solver MA57 that relies on a variant of Gaussian elimination for sparse matrices [55]. The computational cost in this case depends on the sparsity structure of $\mathbf{A}^i$, in particular on the half bandwidth $\delta$. For example, if the non-zero elements are concentrated inside a band of width $2\delta+1$, the cost for the factorization is $O(n\delta^2)$ FLOPS. Note that $\delta$ typically increases linearly with $n$ if no optimal strategies for grid node renumbering (which is typically very costly) are employed. Note that the same computational cost estimate applies when a direct sparse solver is replaced by iterative Krylov-subspace methods. For these reasons we consider the cost of this operation to be formally of $O(n^3)$.

The total online cost for the FSM is therefore governed by the solution of the sparse linear system.

The online operations and the associated costs for SnapROM /MEsROM are:

1. assembly of the elements of the system matrix $\mathbf{A}^i$ (it is the same step required for FMS), with a cost of $O(sn)$ FLOPS;
2. computation of the ROM matrix of equation (12), $\mathbf{P}^{\mathrm{T}}\mathbf{A}^i\mathbf{P}$, with a cost of $O(snm + nm^2)$ FLOPS;
3. solution of the linear system (12) of dimension $m$, with a cost of $O(m^3)$ FLOPS;

4. computation of the approximated solution in the FSM space as $\tilde{\mathbf{h}}^i$ (11): $O(nm)$ FLOPS.

Thus, the online cost for SnapROM and MEsROM is dominated by $O(snm + nm^3)$ FLOPS.

The cost of the most expensive operation during the online phase shifts from $O(n^3)$ in FSM, to $O(snm + nm^3)$ in SnapROM and MEsROM. When $n$ becomes large, the computational advantage of SnapROM and MEsROM becomes clear (see also Section 5.3). The most expensive operations in this case are the projections onto and from the dimension of the full system, listed above at points 1. and 4., which involve a cost proportional to $n$. This cost might be further decreased in ROMs when considering a reduction of the parameter space or an affine representation of the system matrix (see, e.g., [26, 61]). This particular issue can be the subject of a future analysis.

Figure 1 exemplifies our approach by depicting a flowchart detailing the procedures for (*i*) integration of moment equations in a reduced-order modeling strategy for Monte Carlo simulation of groundwater flow (relying on MEsROM), (*ii*) snapshot-based reduced-order Monte Carlo approach (relying on SnapROM) and (*iii*) standard Monte Carlo simulation (relying on FSM).

## 4. Computational examples

The ROMs described in Section 3 are here compared in a two- and a three-dimensional setting where (steady-state) pumping is superimposed to a uniform (in the mean) flow. Exemplary depictions of the two- and three-dimensional domains considered are presented in Figures 2 and 3, respec-

tively. The MEs illustrated in Section 2.1 are formulated considering hydraulic conductivity ($K$) fields in the three-dimensional scenario.

In all of the scenarios examined, we consider the natural logarithm of transmissivity $Y = \ln T$ (or of conductivity $K$ in three-dimensions) to be a Gaussian second-order stationary random field whose covariance $C_Y$ is characterized by an exponential function:

$$C_Y(\mathbf{d}) = \sigma_Y^2 \exp\left(-\left[\frac{d_1}{\tau_1} + \frac{d_2}{\tau_2} + \frac{d_3}{\tau_3}\right]\right) \tag{13}$$

where $\mathbf{d} = [d_1, d_2, d_3]$ is the separation vector, $d_i$ and $\tau_i$ being the spatial separation distance (or lag) and correlation scale along $x_i$, respectively, and $\sigma_Y$ is the standard deviation of $Y$. The two-dimensional scenario is defined for $i = 1, 2$ only. If not specified otherwise, we consider $\sigma_Y = 1$ and the ensemble mean (expectation) of $Y$ is set equal to 0.

The two-dimensional (2D) scenario is a $20 \times 10$ (all quantities are in consistent units) domain where Dirichlet boundary conditions are imposed along the left and right sides with constant deterministic head (i.e., $h = 3$ at $x_1 = 0$ and $h = 0$ at $x_1 = 20$). A point well extracting a unit pumping rate ($f = -1$) is located at the center of the domain, while the top and bottom boundaries are impervious. The flow domain is discretized through a regular grid with $81 \times 41$ nodes and 6400 triangular elements. Unless specified otherwise, in the 2D scenario we set $\tau_1$ and $\tau_2$ equal to 2 and 1, respectively.

The three-dimensional (3D) set-up corresponds to a $600 \times 600 \times 60$ domain where a point extraction takes place from the center node of the topside boundary at a constant rate $f = -100$. Dirichlet conditions are set across the left and right sides (with constant head $h = 70$ at $x_1 = 0$ and $h = 70$ at $x_1 = 600$), the remaining boundaries being impervious. In this 3D scenario

we set $\tau_1$, $\tau_2$, and $\tau_3$ equal to 100, 100 and 20, respectively. The flow domain is discretized through a regular grid with 25 $\times$ 25 $\times$ 13 nodes and 34560 tetrahedral elements.

The analysis of both 2D and 3D scenarios is beneficial to assess the CPU cost of MEsROM with respect to $n$ and the problem dimensionality. We use the widely tested GSLIB software [62] to generate the random $Y$ realizations with the aforementioned input parameters.

We quantify the accuracy of both SnapROM and MEsROM (as listed in Table 1) through the $L_2$ norm of the error against the head realizations associated with the reference solutions obtained through the FSM. In this sense, the error metric $\eta$ is the average error on the whole set of MC solutions. For example, for the heads we have:

$$\eta_h = \frac{1}{N} \sum_{i=1}^{N_{en}} \left\| \tilde{\mathbf{h}}^i - \mathbf{h}^{i,\mathrm{FSM}} \right\|_{L_2}^2 \tag{14}$$

A second error metric considered is the average relative error $\mu$:

$$\mu_h = \frac{1}{N} \sum_{i=1}^{N_{en}} \frac{\left\| \tilde{\mathbf{h}}^i - \mathbf{h}^{i,\mathrm{FSM}} \right\|_{L_2}^2}{\|\mathbf{h}^{i,\mathrm{FSM}}\|_{L_2}^2} \tag{15}$$

Similarly, we consider the absolute error and relative error on the head variances as:

$$\eta_{\sigma_h^2} = \left\| \tilde{\boldsymbol{\sigma}}_h^2 - \boldsymbol{\sigma}_h^{2,\mathrm{FSM}} \right\|_{L_2}^2 \quad , \quad \mu_{\sigma_h^2} = \frac{\left\| \tilde{\boldsymbol{\sigma}}_h^2 - \boldsymbol{\sigma}_h^{2,\mathrm{FSM}} \right\|_{L_2}^2}{\left\| \boldsymbol{\sigma}_h^{2,\mathrm{FSM}} \right\|_{L_2}^2} \tag{16}$$

where $\tilde{\boldsymbol{\sigma}}_h^2$ is the approximated head variance computed from $\tilde{\mathbf{h}}^i$.

## 5. Results and discussion

### *5.1. Offline phase: Computation of the projection space*

#### *5.1.1. Computational cost of the offline phase*

In the following sections we compare the accuracy of the results associated with SnapROM and MEsROM. Since the accuracy of SnapROM depends upon the number of snapshots collected during the offline phase, we set the latter parameter in a way that two ROMs are characterized by an equivalent offline phase (in terms of computational time). Thus, we start our analysis by exploring the computational cost of the offline phase associated with MEsROM and SnapROM.

Table 2 lists the system size (i.e., the number of nodes) and the associated CPU times required to solve the MEs and to perform a single run of the FSM (evaluated as mean time over 100,000 runs using the processor Inter(R) Xeon(R) CPU E5-2650 v3 @ 2.30GHz with 128 GB RAM) for the 2D and 3D scenarios. For comparison purposes, we also list the ratio (rounded to the nearest integer) between the CPU time associated with the MEs and a single FSM run. These results show that solving the MEs requires a computational cost that roughly corresponds to 482 and 282 FSM solutions in the 2D and 3D scenarios, respectively.

In order to obtain the same computational cost between the offline phases of the two ROMs, one needs to consider also the CPU time necessary for the computation of the projection matrix from the head covariance matrix. The cost of the iterative method adopted by MEsROM linearly increases with the number of eigenvectors $\mathbf{p}_i$, $i = 1, \ldots, m$, while the SVD on the snapshots adopted by SnapROM enables computing all eigenvectors at once.

The total offline cost of MEsROM versus the number of eigenvectors $m$ (which corresponds to the dimension of the reduced model) is depicted in Figure S1 in the the supplementary material (SM) and compared with the offline cost associated with SnapROM when using $N_{sn}$ =500 and $N_{sn}$ =300 snapshots for the 2D and 3D scenarios, respectively. In the 2D scenario, the cost required to obtain 500 snapshots and their resulting projection matrix using SVD is slightly higher than the one required to solve the MEs and obtain $m = 60$ eigenvectors. In the 3D scenario, the offline stages of MEsROM and SnapROM are characterized by approximately the same computational cost when using 300 snapshots for SnapROM and $m = 60$ eigenvectors for both ROMs.

For this reason, in the following we consider a number of 500 (or 300) snapshots for SnapROM in the 2D (or 3D) scenarios, respectively, which roughly yields the same computational cost of the offline stages in SnapROM and MEsROM (see Table 2).

### *5.1.2. Quality of the projection space*

We now assess the quality of the eigenvectors derived from the solution of the MEs as well as from the set of snapshots. Figure 4 depicts the first six eigenvectors computed on the basis of $N_{sn}$ =500 and 100,000 snapshots together with the corresponding results associated with the MEs in the 2D scenario. Considering the eigenvectors obtained for $N_{sn}$=100,000 as reference, it is clear that the snapshot technique with $N_{sn}$ =500 does not provide the same accuracy as the MEs-based approach, whose results are visibly closer to the reference solution. This result is quantitatively supported by Table 3 where the $L_2$-norm of the errors between the approximated eigenvec-

tors and their reference counterparts are listed. The MEs-based eigenvectors are associated with a smaller $L_2$-error than their counterparts based on 500 snapshots. The only exception is given by $\mathbf{p}_3$, where the snapshots-based eigenvector displays a support which is spatially more diffused than its MEs-based counterpart (see Figure 4). Otherwise, Figure 4 evidences that the spatial pattern of eigenvectors $\mathbf{p}_4$, $\mathbf{p}_5$, and $\mathbf{p}_6$ is poorly captured by relying on 500 snapshots. In conclusion, these results suggest that, even under conditions where one devotes the same CPU time to collect snapshots (via FSM solutions) and for the solution of MEs, the snapshot-based eigenvectors are not as accurate as their MEs-based counterparts. Clearly, one can improve the quality of the snapshots-based eigenvectors by computing an increased number of snapshots. However, this would result in a more expensive offline phase.

### *5.2. Online phase: Accuracy of the ROM solutions*

Here, we aim at assessing whether the improved accuracy in the projection vectors obtained through the MEs-based approach also enhances the accuracy of the ROM solutions. We first address the 2D scenario and evaluate the spatial distribution of the absolute difference between the head variances computed by MC simulations based on the FSM with $N_{en} = 100,000$ and by MEsROM and SnapROM with $N_{sn} = 500$ (Figure 5 (a) and (b), respectively). Errors associated with MEsROM are everywhere smaller than those of SnapROM and are characterized by a spatial distribution similar to the one of the errors between FSM-based head variances and their second-order approximations evaluated by MEs (see also Figure S2 of the supplementary material). One can note that the largest errors are concentrated around the

pumping well and rapidly drop (by one order of magnitude) with distance from it. These results unambiguously show that MEsROM consistently outperforms SnapROM in terms of accuracy on the reconstructed head variance field.

The quality of the reduced order models can also be assessed locally, by evaluating the rate of convergence of (statistical) moments of heads at given locations in the system. As an example, Figure S3 depicts the first four moments (i.e., mean, variance, skewness, and kurtosis) of heads evaluated at point $(d)$ in Figure 2 versus the number of MC realizations, as obtained through FSM, SnapROM, and MEsROM. All of these models display a similar behavior, in terms of convergence of these moments. The first two moments appear to attain convergence within about 20,000 MC realizations, skewness and kurtosis requiring more than 50,000 realizations. While asymptotic values of moments associated with MEsROM and SnapROM visibly differ from their FSM-based counterparts, MEsROM is always closer to FSM than SnapROM, a feature which is particularly evident when considering the third and fourth (statistical) moments.

The ensemble error metrics $\mu_h$ and $\eta_h$ (see equations (14) and (15)) associated with $\sigma_Y^2 = 1.0$ are shown in Figures 6(a) and (b) as a function of the size $m$ of the ROM projection space. Corresponding results for the ensemble error metrics $\mu_{\sigma_h^2}$ and $\eta_{\sigma_h^2}$ related to head variance are depicted in Figures 6(c) and (d), which compare errors related to MEsROM and SnapROM for a given $m$. While it is obvious that the performance of the two reduced models is improved with the increase of model size (i.e., for increasing $m$), it is interesting to note that absolute and relative errors on both mean and variance

are consistently lower for MEsROM than for SnapROM.

For completeness, we further compare the metrics $\eta_{\sigma_h^2}$ and $\eta_h$ considering various values of variance ($\sigma_Y^2$=0.1, 1.0 and 2.0) and different correlation scales ($(\tau_1, \tau_2)$ = (2, 1), (4, 2), (8, 4) and (16, 8)) in Table 4. One can see that MEsROM outperforms SnapROM in terms of both $\eta_h$ and $\eta_{\sigma_h^2}$ when the variance of $Y$ is small to mild (i.e., $\sigma_Y^2 = 0.1$ and 1.0). This is especially evident when considering short correlation scales (e.g., when $(\tau_1, \tau_2) = (2, 1)$), where the error $\eta_{\sigma_h^2}$ for SnapROM is twice the one associated with MEsROM. When $\sigma_Y^2 = 2.0$, SnapROM outperforms MEsROM in terms of $\eta_{\sigma_h^2}$, a feature which is particularly clear for random fields characterized by small values of correlation scales. This finding can be attributed to the increased approximation errors in the MEs solutions associated with larger $\sigma_Y^2$ (see, e.g., [63]).

The ratio between either $\eta_{\sigma_h^2}$ or $\eta_h$ related to SnapROM and its MEsROM-based counterpart increases with $\tau$. This is related to the observation that an increase in the spatial correlation yields a corresponding increase of the degree of spatial homogeneity of the single realizations of $Y$. This aspect is in turn reflected in a reduction of approximation errors due to $(i)$ perturbation and $(ii)$ truncation of the projection space for MEs, which leads to an improved performance with respect to SnapROM. For example, it can be noted that the relative error on $\eta_{\sigma_h^2}$ is approximately equal to one for $(\tau_1, \tau_2)$ = (16, 8) and $\sigma_Y^2$=2. Note that the relative error on $\eta_h$ is always larger than one for all fields explored, thus implying that MEsROM is always characterized by a smaller error about mean heads than SnapROM (as quantified by such global metric).

In Figure 7 we explore the quality of SnapROM and MEsROM as a function of the dimension $m$ of the reduced Galerkin projection space by depicting the spatial distribution of head covariance between point $(a)$ (see Figure 2) and all other locations in the domain obtained by MEs, MEsROM, SnapROM, and the FSM. Visual comparison between MEsROM and SnapROM indicates that the former outperforms the latter in terms of accuracy for all of the values of $m$ explored. As expected, the quality of MC results based on MEsROM is not significantly improved with respect to MEs, suggesting that performing MC simulations through MEsROM can contribute only to rendering higher-order (ensemble) moments, which are not captured by a typical implementation of MEs.

As a further comment to these results, it should be also noted that relying solely on a global metric related to the variance (i.e., $\eta_{\sigma^2_h}$) or to the mean value (i.e., $\eta_h$) does not provide a complete depiction of the consequences of relying on either approach on the characterization of the whole PDF of heads. Thus, as an additional element of interest in the analysis, we consider the joint PDF of the heads at pairs of points in the system. Figure 8 depicts these PDFs for the pair of points $(a)$ and $(b)$ shown in Figure 2. The reference MC results (obtained with 100,000 FSM solutions) are shown in the first column for the values of $\sigma^2_Y = 0.1, 1, 2$. These joint PDFs are visibly non-Gaussian, with tailing that seemingly increases with $\sigma^2_Y$. Differences between the reference solution and those obtained with SnapROM and MEsROM are depicted in the second and third columns of Figure 8, respectively, for $m = 30$. The main differences between the results of SnapROM (second column) and the reference solution are concentrated around the peak of the probability

distribution. The corresponding MEsROM errors are characterized by a smaller maximum value and decrease more rapidly with distance from the peak. Results of similar quality are obtained when considering (empirical) probability densities of heads at the five exemplary locations identified in Figure 2 (see Figures S4-S6), as well as other positions in the domain (not shown), for various values of variance of $Y$. These analyses suggest that MEsROM renders more accurate PDFs than SnapROM in this scenario.

As tails of probability distributions are markedly relevant in environmental risk assessment procedures, the results summarized in Figure 9 focus on the relative absolute differences ($\Delta_q$) of the 10-th (Figure 9a) and 90-th (Figure 9b) quantiles of cumulative density functions obtained by the FSM and the two reduced-order methods (i.e., MEsROM and SnapROM associated with $m = 30$) at locations along a selected transect of the 2D domain (see Figure 2). The 10-th quantiles are always captured more accurately by MEsROM than by SnapROM, an exception being noted at the domain center (here exemplified by $x_2 = 5$) where the point source is located. This behavior is consistent with the observation that the head gradient theoretically tends to infinity at the pumping location and it is well documented that any numerical solution (including the FSM) is affected by large errors that are dependent upon the particular mesh employed in the simulation. With reference to the 90-th quantiles, it is hard to discriminate between the performance of MEsROM or SnapROM.

We now consider the results of the 3D scenario and assess the way various reduced order modeling strategies which can lead to properly capture higher order moments of the head PDFs. Figure 10 (first row) depicts the sample

PDFs of hydraulic heads at the five reference locations identified in Figure 3. The black curve is computed by FSM with an ensemble size of 100,000, and is considered as reference solution, blue and green symbols corresponding to MEsROM and SnapROM with 300 snapshots considering $m$ equal to 30. Corresponding differences ($\Delta$) between reference results and approximations obtained through MEsROM and SnapROM are also depicted (second row in Figure 10). These results contribute to reinforce our previous findings, documenting that MEsROM outperforms SnapROM also in this case in terms of accuracy of the reconstruction of head PDFs. An additional appreciation of the accuracy in recovering the high and low probability values (i.e., values associated with the tails) of the distribution can be obtained by analyzing the error on the quantiles along a section of the domain. Figure 11 depicts the spatial distribution of the relative difference between the 10-th (top row) and 90-th (bottom row) quantiles of head probability distributions obtained by FSM and the two reduced-order methods with $m = 30$ along the section at $x_1 = 300$. These results clearly show that MEsROM always renders more accurate values of the 10-th quantiles than SnapROM. With reference to the 90-th quantiles, MEsROM outperforms SnapROM at most locations across the section. As such, these results further support our prior conclusions based on the two-dimensional setting considered above.

### *5.3. Online cost*

We recall that SnapROM and MEsROM share the same online phase. Thus, the cost of a single MC solution is the same for both ROMs (this is also verified in our simulations).

Table 5 lists the ratio between the average CPU times of a single MC

solution computed through the FSM and the ROMs associated with different dimensions of the reduced model, $m$, for the 2D and 3D scenarios analyzed.

With reference to the 2D scenario, our results show that the CPU time of the FSM is about 10 times higher than ROMs with $m = 5$. Otherwise, both SnapROM and MEsROM can lose their computational advantage when $m > 40$, thus indicating that the projection onto the reduced subspace of the full stiffness matrix (i.e., the calculation of the $\mathbf{P}^{\mathrm{T}}\mathbf{A}^{i}\mathbf{P}$ matrix in eq. (12)) is a computationally expensive critical task. The computational advantage of both ROMs becomes markedly clear in the 3D scenario. For example, one can note that the FSM becomes almost 29 times slower than the ROMs when $m = 5$. These results are consistent with the analysis about the online cost illustrated in Section 3.2, where it is remarked that the most expensive operation of the online phase for FSM is $O(n^3)$, while being $O(nm(s + m^2))$ for SnapROM and MEsROM.

## 6. Conclusions

Our analyses and results lead to the following major conclusions:

- The moment equations-based reduced-order model (MEsROM) outperforms its snapshot-based counterpart (SnapROM) in terms of solution accuracy and is not plagued by issues linked to the selection of snapshot size, which constitutes a common drawback in the application of traditional snapshot-based reduced-order models. Comparison of the effectiveness of these two ROMs is performed upon selecting a snapshot size in SnapROM (500 in 2D and 300 in 3D) which yields approximately equal CPU times of the offline phases for both approaches (see Table 2).

Joint analyses of absolute error maps of head variances (Figure 5) and features of the associated (sample) probability density functions (Figures 8 - 11 and S4-6) document that MEsROM provides more accurate results than SnapROM in most of the scenarios analyzed.

- MEsROM results are particularly more accurate than their SnapROM counterparts when considering low values of the variance of $Y$ (e.g., $\sigma^2_Y = 0.1$, 1, see Table 4). For larger values of the transmissivity variance ($\sigma^2_Y = 2$) SnapROM yields more accurate results on the head variance, but not on the other statistics on the head PDF tested (see, e.g., Figure 8).

- SnapROM and MEsROM are characterized by the same computational steps during the online phase of the workflow we develop for the Monte Carlo simulations. The CPU time required to obtain the collection of MC solutions through the ROMs is up to 10 time lower than what can be observed for the Full System Model (FSM) solution when using the reduced dimension $m = 5$ in the two-dimensional (2D) scenario. Computational advantages increase with the size of the numerical mesh in the three-dimensional (3D) scenario considered (see Section 3.2 and Tables 2, 5). For example, for values $m = 30$ of the dimension of the reduced model, the CPU time of a single FSM solution is twice or eight times larger than the one of a single ROM solution for the 2D (where the full system size is $n = 3321$) or the 3D (where $n = 8125$) scenario, respectively.

Our study documents that the critical operation of the online phase is the

construction of the stiffness matrix and its projection onto the reduced order subspace. Future developments of our MEsROM approach involve adapting the workflow to incorporate methods for the reduction of the parameter space and using an affine representation of the system matrix with respect to the random space of the transmissivity/conductivity. To this aim, discrete matrix interpolation schemes (e.g., DEIM [61, 64, 65]) and purely data-driven methods are considered as promising fields of future research.

Additional elements of interest associated with future developments of the integration between MEs and model-reduction strategies in a MC framework include the analysis of transient flow scenarios or solute transport through randomly heterogeneous domains. With reference to the latter, one could envision relying on, e.g., differential moment equations for transport (see, e.g., [66, 67]) in conjunction with an adaptive model reduction technique [32].

## Credit authorship contribution statement

All Authors contributed to the design, developments and analyses of the work.

## Declaration of Competing Interest

There are no competing interests.

## Acknowledgements

This work was supported by the National Nature Science Foundation of China (Grant No. 41530316). Part of the work was developed while Prof. A.

Guadagnini was at the University of Strasbourg with funding from Région Grand-Est and Strasbourg-Eurométropole through the 'Chair Gutenberg'. Chuan-An Xia was supported by International Young Researcher Development Project of Guangdong Province, China.

## References


[1] B. Diskin, J. L. Thomas, C. L. Rumsey, A. Schwoeppe, Grid-Convergence of Reynolds-Averaged NavierStokes Solutions for Benchmark Flows in Two Dimensions, AIAA J. 54 (2016) 2563–2588.

[2] X. Geng, M. C. Boufadel, F. Cui, Numerical modeling of subsurface release and fate of benzene and toluene in coastal aquifers subjected to tides, J. Hydrol. 551 (2017) 793–803.

[3] Q. Xie, B. H. Kueper, C. Zhao, A numerical model for estimating the removal of volatile organic compounds in laboratory-scale treatability tests for thermal treatment of napl-impacted soils, J. contam. hydrol. 226 (2019).

[4] F. Ballio, A. Guadagnini, Convergence assessment of numerical Monte Carlo simulations in groundwater hydrology, Water Resour. Res. 40 (2004) W04603.

[5] P. C. Leube, F. P. J. de Barros, W. Nowak, R. Rajagopal, Towards optimal allocation of computer resources: Trade-offs between uncertainty quantification, discretization and model reduction, Environ. Modell. Softw. 50 (2013) 97–107.

[6] A. Bardossy, S. Hoerning, Gaussian and non-gaussian inverse modeling of groundwater flow using copulas and random mixing, Water Resour. Res. 52 (2016) 4504–4526.

[7] G. Evensen, Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics, J. Geophys. Res. 99 (1994) 1014310162.

[8] Z. Chen, J. Jaime Gomez-Hernandez, T. Xu, A. Zanini, Joint identification of contaminant source and aquifer geometry in a sandbox experiment with the restart ensemble kalman filter, J. Hydrol. 564 (2018) 1074–1084.

[9] C.-A. Xia, B. X. Hu, J. Tong, A. Guadagnini, Data assimilation in density-dependent subsurface flows via localized iterative ensemble kalman filter, Water Resour. Res. 54 (2018) 6259–6281.

[10] L. Li, L. Stetler, Z. Cao, A. Davis, An iterative normal-score ensemble smoother for dealing with non-gaussianity in data assimilation, J. Hydrol. 567 (2018) 759–766.

[11] I. M. Sobol, Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates, Math. Comput. Simulat. 55 (2001) 271–280.

[12] A. Dell'Oca, M. Riva, A. Guadagnini, Moment-based metrics for global sensitivity analysis of hydrological systems, Hydrol. Earth Syst. Sci. 21 (2017) 6219–6234.

[13] C.-A. Xia, J. Tong, B. X. Hu, X. Wu, A. Guadagnini, Assessment of alternative adsorption models and global sensitivity analysis to characterize hexavalent chromium loss from soil to surface runoff, Hydrol. Process. 32 (2018) 3140–3157.

[14] T. Hermans, F. Nguyen, M. Klepikova, A. Dassargues, J. Caers, Uncertainty quantification of medium-term heat storage from short-term geophysical experiments using bayesian evidential learning, Water Resour. Res. 54 (2018) 2931–2948.

[15] Z. Chen, L. Shi, M. Ye, Y. Zhu, J. Yang, Global sensitivity analysis for identifying important parameters of nitrogen nitrification and denitrification under model uncertainty and scenario uncertainty, J. Hydrol. 561 (2018) 884–895.

[16] S. Razavi, B. A. Tolson, D. H. Burn, Review of surrogate modeling in water resources, Water Resour. Res. 48 (2012).

[17] M. J. Asher, B. F. W. Croke, A. J. Jakeman, L. J. M. Peeters, A review of surrogate models and their application to groundwater modeling, Water Resour. Res. 51 (2015) 5957–5973.

[18] M. S. Hussain, J. A. A., A.-A. A., F. R., A surrogate model for simulationoptimization of aquifer systems subjected to seawater intrusion, J. Hydrol. 523 (2015) 542–554.

[19] D. Pasetto, A. Guadagnini, M. Putti, POD-based Monte Carlo approach for the solution of regional scale groundwater flow driven by randomly distributed recharge, Adv. Water Resour. 34 (2011) 1450–1463.

[20] D. Pasetto, M. Putti, W. W.-G. Yeh, A reduced-order model for groundwater flow equation with random hydraulic conductivity: Application to monte carlo methods, Water Resour. Res. 49 (2013) 3215–3228.

[21] X. He, L. Ren, Finite volume multiscale finite element method for solving the groundwater flow problems in heterogeneous porous media, Water Resour. Res. 41 (2005) W10417.

[22] S. I. Aanonsen, D. Eydinov, A multiscale method for distributed parameter estimation with application to reservoir history matching, Computat. Geosci. 10 (2006) 97–117.

[23] P. Chen, A. Quarteroni, G. Rozza, Reduced basis methods for uncertainty quantification, Society for Industrial and Applied Mathematics and American Statistical Association (2017) 813–869.

[24] P. Vermeulen, A. Heemink, C. te Stroet, Low-dimensional modelling of numerical groundwater flow, Hydrological Processes 18 (2004).

[25] D. Pasetto, A. Guadagnini, M. Putti, A reduced-order model for Monte Carlo simulations of stochastic groundwater flow, Computat. Geosci. 18 (2014) 157–169.

[26] Z. P. Stanko, S. E. Boyce, W. W.-G. Yeh, Nonlinear model reduction of unconfined groundwater flow using pod and deim, Adv. Water Resour. 97 (2016) 130–143.

[27] S. E. Boyce, T. Nishikawa, W. W. G. Yeh, Reduced order modeling of the newton formulation of modflow to solve unconfined groundwater flow, Adv. Water Resour. 83 (2015) 250–262.

[28] A. J. Siade, M. Putti, W. W. G. Yeh, Snapshot selection for groundwater model reduction using proper orthogonal decomposition, Water Resour. Res. 46 (2010) W08539.

[29] X. Li, X. Chen, B. X. Hu, I. M. Navon, Model reduction of a coupled numerical model using proper orthogonal decomposition, J. Hydrol. 507 (2013) 227–240.

[30] Z. Luo, H. Li, Y. Zhou, Z. Xie, A reduced finite element formulation based on pod method for two-dimensional solute transport problems, J. Math. Anal. Appl. 385 (2012) 371–383.

[31] D. Pasetto, M. Ferronato, M. Putti, A reduced order model-based preconditioner for the efficient solution of transient diffusion equations, Int. J. Numer. Meth. Eng. 109 (2017) 1159–1179.

[32] C. Rizzo, F. de Barros, S. Perotto, L. Oldani, A. Guadagnini, Adaptive POD model reduction for solute transport in heterogeneous porous media, Computat. Geosci. 22 (2018) 297–308.

[33] G. Santin, B. Haasdonk, Convergence rate of the data-independent P-greedy algorithm in kernel-based approximation, Dolomites Res. Notes Approx. 10 (2017) 68–78.

[34] B. Haasdonk, G. Santin, Greedy kernel approximation for sparse surrogate modeling, in: W. Keiper, A. Milde, S. Volkwein (Eds.), Reduced-Order Modeling (ROM) for Simulation and Optimization, Springer, Cham, 2018, pp. 21–45.

[35] D. J. Knezevic, A. T. Patera, A certified reduced basis method for the fokker-planck equation of dilute polymeric fluids: Fene dumbbells in extensional flow, SIAM J. Sci. Comput. 32 (2010) 793–817.

[36] A. Buffa, Y. Maday, A. T. Patera, C. Prud'homme, G. Turinici, A priori convergence of the greedy algorithm for the parametrized reduced basis method, Math. Model. Numer. Anal. 46 (2012) 595–603.

[37] K. Smetana, A. T. Patera, Optimal local approximation spaces for component-based static condensation procedures, SIAM J. Sci. Comput. (2016) A3318–A3356.

[38] S. E. Boyce, W. W. G. Yeh, Parameter-independent model reduction of transient groundwater flow models: Application to inverse problems, Adv. Water Resour. 69 (2014) 168–180.

[39] C. Lieberman, K. Willcox, O. Ghattas, Parameter and state model reduction for large-scale statistical inverse problems, SIAM J. Sci. Comput. 32 (2010) 2523–2542.

[40] D. Tartakovsky, S. Neuman, Transient flow in bounded randomly heterogeneous domains: 1. Exact conditional moment equations and recursive approximations, Water Rosources Research 34 (1998) 1–12.

[41] D. Zhang, Stochastic Method for Flow in Porous Media - Coping with Uncertainties, Academic Press, Sand Diego, California, 2002.

[42] L. Li, H. A. Tchelepi, Conditional stochastic moment equations for uncertainty analysis of flow in heterogeneous reservoirs, SPE J. 8 (2003) 392–400.

[43] L. Li, H. A. Tchelepi, D. Zhang, Perturbation-based moment equation approach for flow in heterogeneous porous media: applicability range and analysis of high-order terms, J. Comput. Phys. 188 (2003) 296–317.

[44] C. Winter, D. Tartakovsky, A. Guadagnini, Moment differential equations for flow in highly heterogeneous porous media, Surv. Geophys. 24 (2003) 81–106.

[45] M. Riva, A. Guadagnini, J. Bodin, F. Delay, Characterization of the hydrogeological experimental site of poitiers (france) by stochastic well testing analysis, J. Hydrol. 369 (2009) 154–164.

[46] M. Panzeri, M. Riva, A. Guadagnini, S. P. Neuman, Enkf coupled with groundwater flow moment equations applied to lauswiesen aquifer, J. Hydrol. 521 (2015) 205–216.

[47] E. Bianchi Janetti, M. Riva, S. Straface, A. Guadagnini, Stochastic characterization of the montaldo uffugo research site (italy) by geostatistical inversion of moment equations of groundwater flow, J. Hydrol. 381 (2010) 42–51.

[48] L. Li, H. A. Tchelepi, Conditional statistical moment equations for dynamic data integration in heterogeneous reservoirs, SPE Reserv. Eval. Eng. 9 (2006) 280–288.

[49] M. Panzeri, M. Riva, A. Guadagnini, S. P. Neuman, Data assimilation and parameter estimation via ensemble kalman filter coupled with stochastic moment equations of transient groundwater flow, Water Resour. Res. 49 (2013) 1334–1344.

[50] J. Z. Yang, D. X. Zhang, Z. M. Lu, Stochastic analysis of saturated-unsaturated flow in heterogeneous media by combining karhunen-loève expansion and perturbation method, J. Hydrol. 249 (2004) 18–38.

[51] D. X. Zhang, Z. M. Lu, An efficient, high-order perturbation approach for flow in random porous media via karhunen-loève and polynomial expansions, J. Comput. Phys. 194 (2004) 773–794.

[52] H. Li, D. Zhang, Probabilistic collocation method for flow in porous media: Comparisons with other stochastic methods, Water Resources Research 43 (2007).

[53] J. Bear, Hydraulics of Groundwater, McGraw-Hill, London, 1979.

[54] C.-A. Xia, A. Guadagnini, B. X. Hu, M. Riva, P. Ackerer, Grid convergence for numerical solutions of stochastic moment equations of groundwater flow, Stoch. Environ. Res. Risk Assess. (2019).

[55] I. S. Duff, MA57—a code for the solution of sparse symmetric definite and indefinite systems, Acm T. Math. Software 30 (2004) 118–144.

[56] M. Ferronato, Preconditioning for sparse linear systems at the dawn of the 21st century: history, current developments, and future perspective, ISRN Appl. Math. 2012 (2012) Article ID 127647.

[57] M. Short, L. Guadagnini, A. Guadagnini, D. Tartakovsky, D. Higdon, Predicting vertical connectivity within an aquifer system, Bayesian Analysis 5 (2010) 557–582.

[58] R. Perulero Serrano, L. Guadagnini, M. Riva, M. Giudici, A. Guadagnini, Impact of two geostatistical hydro-facies simulation strategies on head statistics under non-uniform groundwater flow, Journal of Hydrology 508 (2014) 343–355.

[59] E. Bianchi Janetti, L. Guadagnini, M. Riva, A. Guadagnini, Global sensitivity analyses of multiple conceptual models with uncertain parameters driving groundwater flow in a regional-scale sedimentary aquifer, Journal of Hydrology 574 (2019) 544–556.

[60] C. L. Winter, D. M. Tartakovsky, A. Guadagnini, Numerical solutions of moment equations for flow in heterogeneous composite aquifers, Water Resour. Res. 38 (2002) 81–106.

[61] D. Bonomi, A. Manzoni, A. Quarteroni, A matrix DEIM technique for model reduction of nonlinear parametrized problems in cardiac mechanics, Comput. Methods Appl. Mech. Engrg. 324 (2017) 300–326.

[62] C. V. Deutsch, A. G. Journel, GSLIB: geostatistical software library and users guide, 2nd edn, Oxford University Press, New York, 1998.

[63] A. Guadagnini, S. P. Neuman, Nonlocal and localized analyses of conditional mean steady state flow in bounded, randomly nonuniform domains: 2. Computational examples, Water Resour. Res. 35 (1999) 3019–3039.

[64] Y. Efendiev, J. Galvis, E. Gildin, Local–global multiscale model reduction for flows in high-contrast heterogeneous media, J. Comput. Phys. 231 (2012) 8100–8113.

[65] F. Negri, A. Manzoni, D. Amsallem, Efficient model reduction of parametrized systems by matrix discrete empirical interpolation, J. Comput. Phys. 303 (2015) 431 – 454.

[66] E. Morales-Casique, S. Neuman, A. Guadagnini, Nonlocal and localized analyses of nonreactive solute transport in bounded randomly heterogeneous porous media: Computational analysis, Adv. Water Resour. 29 (2006) 1399–1418.

[67] E. Morales-Casique, S. Neuman, A. Guadagnini, Nonlocal and localized analyses of nonreactive solute transport in bounded randomly heterogeneous porous media: Theoretical framework, Adv. Water Resour. 29 (2006) 1238–1255.

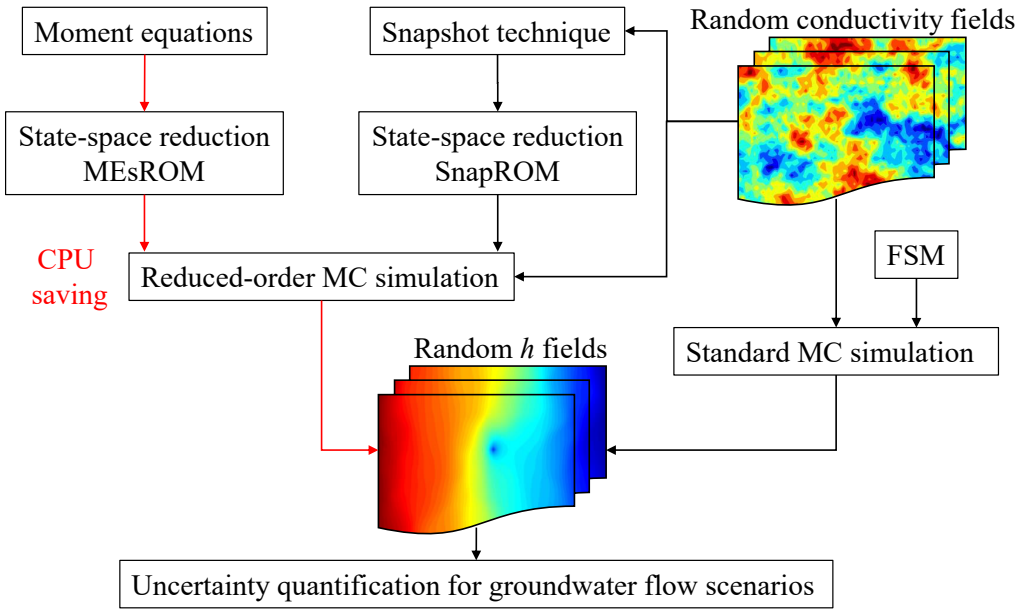


Figure 1: Flowchart of the procedures for $(i)$ integration of moment equations in a reduced-order modeling strategy for Monte Carlo simulation of groundwater flow (relying on MEsROM), $(ii)$ snapshot-based reduced-order Monte Carlo approach (relying on SnapROM) and $(iii)$ standard Monte Carlo simulation (relying on FSM).

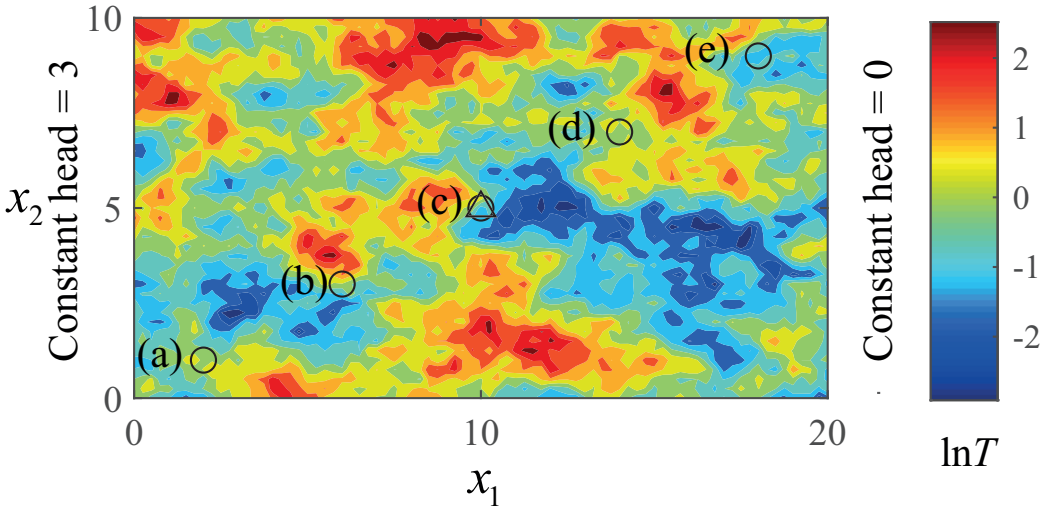


Figure 2: Representation of the two-dimensional flow domain, the black triangle at the domain center corresponding to the location of the point source with unit pumping rate. Constant head conditions are set at the right and left sides, no-flow boundaries being set at the top and bottom. The five exemplary locations selected for the illustration of the results correspond to points $(a)$ with coordinates $(2, 1)$, $(b)$ with coordinates $(6, 3)$, $(c)$ with coordinates $(10, 5)$, $(d)$ with coordinates $(14, 7)$, and $(e)$ with coordinates $(18, 9)$. The color scale refers to values of a selected realization of log-transmissivity.

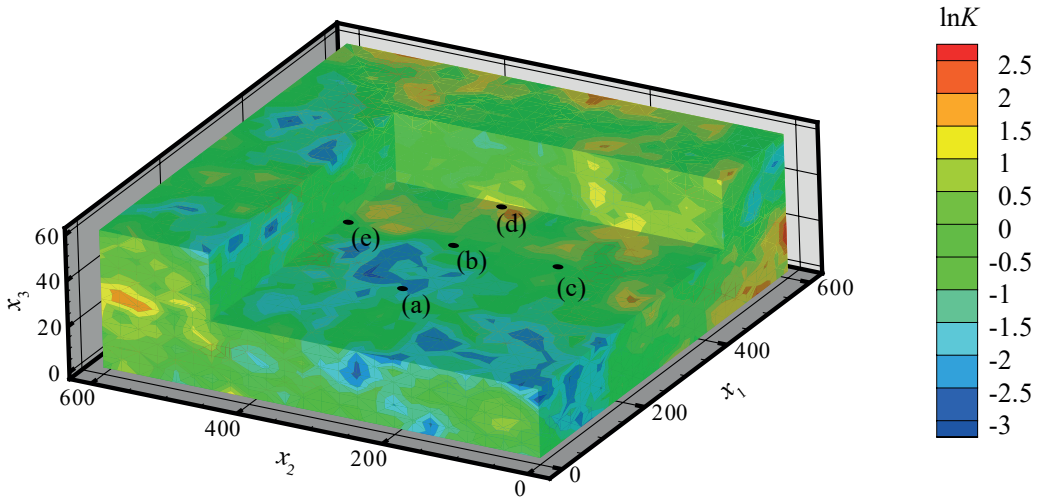


Figure 3: Representation of the three-dimensional flow domain considered in our study. A point source is located at the center node of the topside boundary. Constant head values are set along the right and left side of the domain, the remaining sides corresponding to no-boundaries. The five exemplary locations selected for the illustration of the results correspond to points $(a)$ with coordinates $(150, 300, 30)$, $(b)$ with coordinates $(300, 300, 30)$, $(c)$ with coordinates $(300, 150, 30)$, $(d)$ with coordinates $(450, 300, 30)$, and $(e)$ with coordinates $(300, 450, 30)$. The color scale refers to values of a selected realization of log-conductivity.

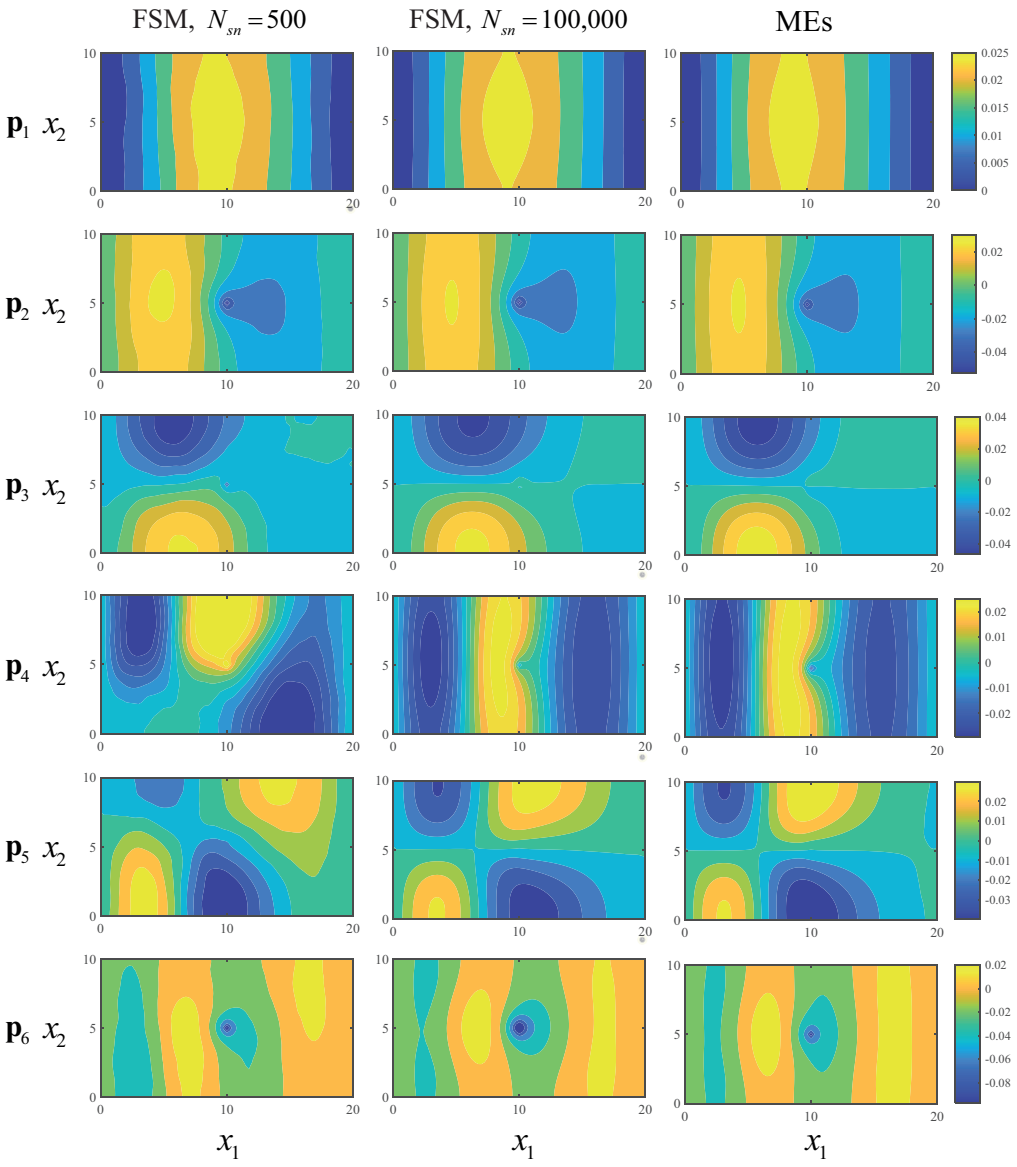


Figure 4: First six eigenvectors computed through SVD on 500 and 100,000 snapshots, and relying on the diagonalization of the MEs-based second order approximation of the head covariance matrix when $\sigma_Y^2 = 1$.

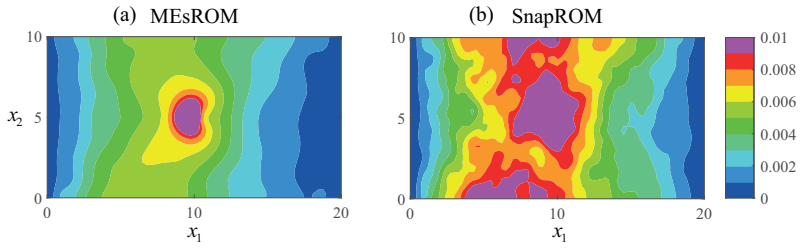


Figure 5: Absolute error on head variances associated with 100,000 reduced-order MC solutions characterized by $m = 30$ by MEsROM (left) and by SanpROM with 500 snapshots (right). The reference variance is obtained through the MC solution obtained via the FSM.

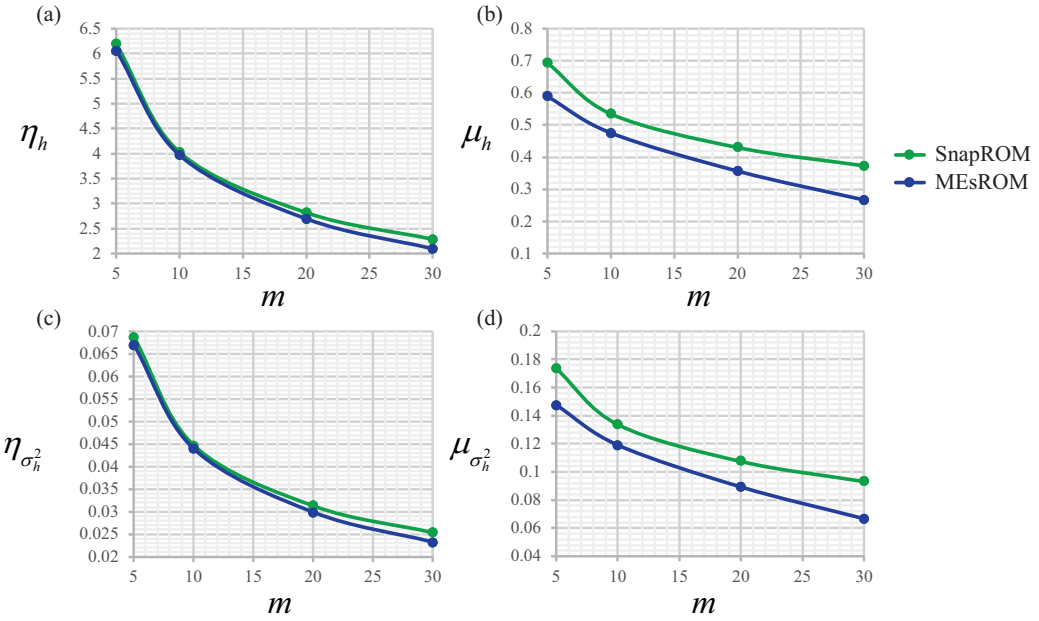


Figure 6: Mean absolute ($\eta_h$, panel a) and relative ($\mu_h$, panels b) errors on head realizations versus the reduced order dimension, $m$; mean absolute ($\eta_{\sigma_h^2}$, panel c) and relative ($\mu_{\sigma_h^2}$, panels d) errors on head variance versus the the reduced order dimension, $m$.

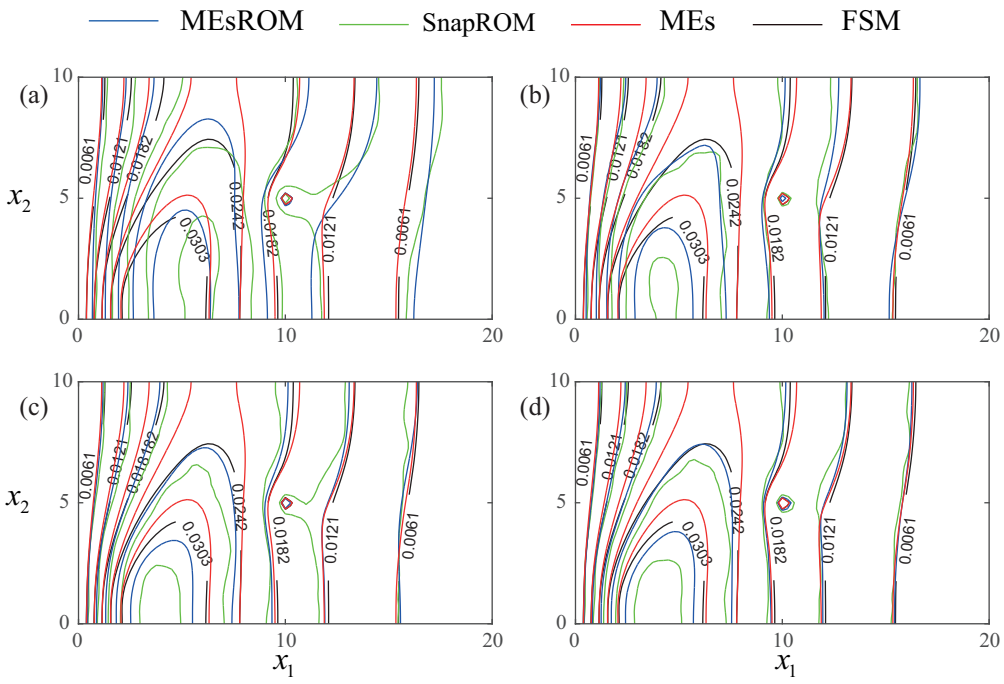


Figure 7: Head covariance evaluated for reference point ($a$) of the 2D scenario (Figure 2) and obtained upon relying on 100,000 realizations of the FSM (black curves), the (second-order) solution of the moment equations (red curves), and the reduced-order models SnapROM (green curves) and MEsROM (blue curves) associated with various values of the reduced model dimension, i.e. $m$= 5 (a); 10 (b), 20 (c) and 30 (d).

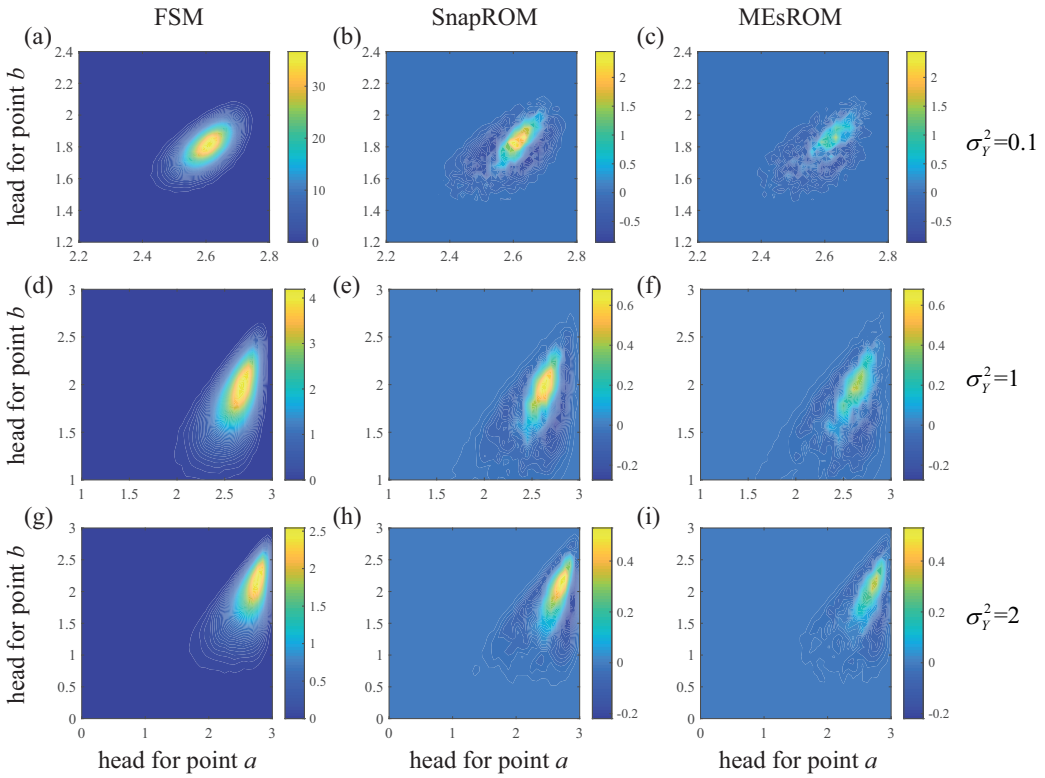


Figure 8: Joint PDFs of heads at locations ($a$) and ($b$) (see Figure 2) obtained through 100,000 realizations of the FSM (first column) with $\sigma_Y^2 = 0.1, 1$, and 2. Differences between the reference FSM results and those obtained with SnapROM and MEsROM are depicted in the second and third column, respectively.

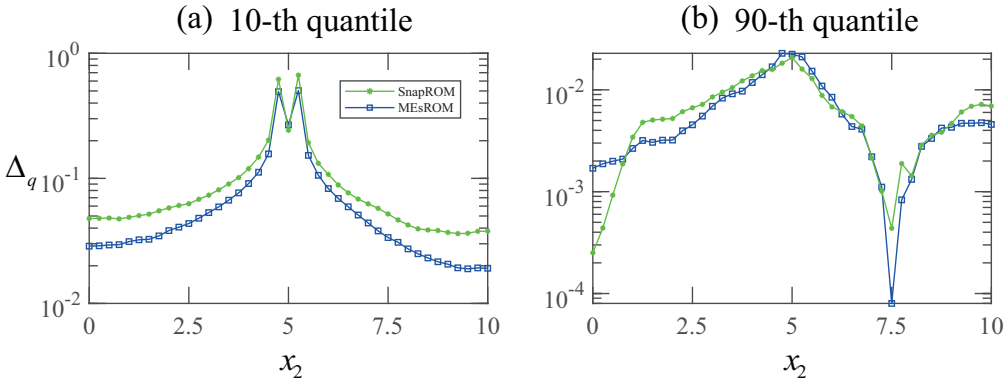


Figure 9: Relative absolute errors ($\Delta_q$) associated with the 10-th (left) and 90-th (right) quantiles of cumulative density functions of heads obtained by the FSM and MEsROM (blue curves) and SnapROM associated with 500 snapshots (green curves) at locations along the transect $x_1 = 10$ of the 2D domain (see Figure 2).

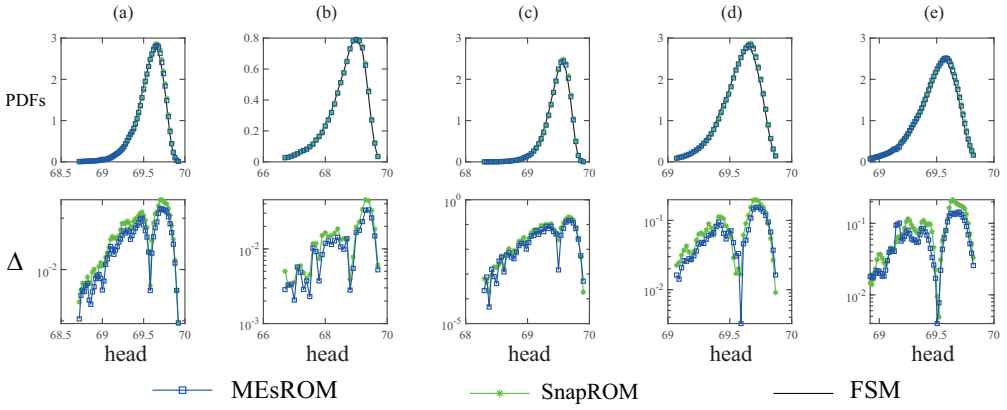


Figure 10: First row: empirical head PDFs at the five locations depicted in Figure 3 and computed through 100,000 MC realizations with: the FSM (black curve, taken as reference), SnapROM with 300 snapshots (green), and MEsROM (blue). Bottom row: absolute error $\Delta$ between ROM-based head PDFs and the reference FSM solution.

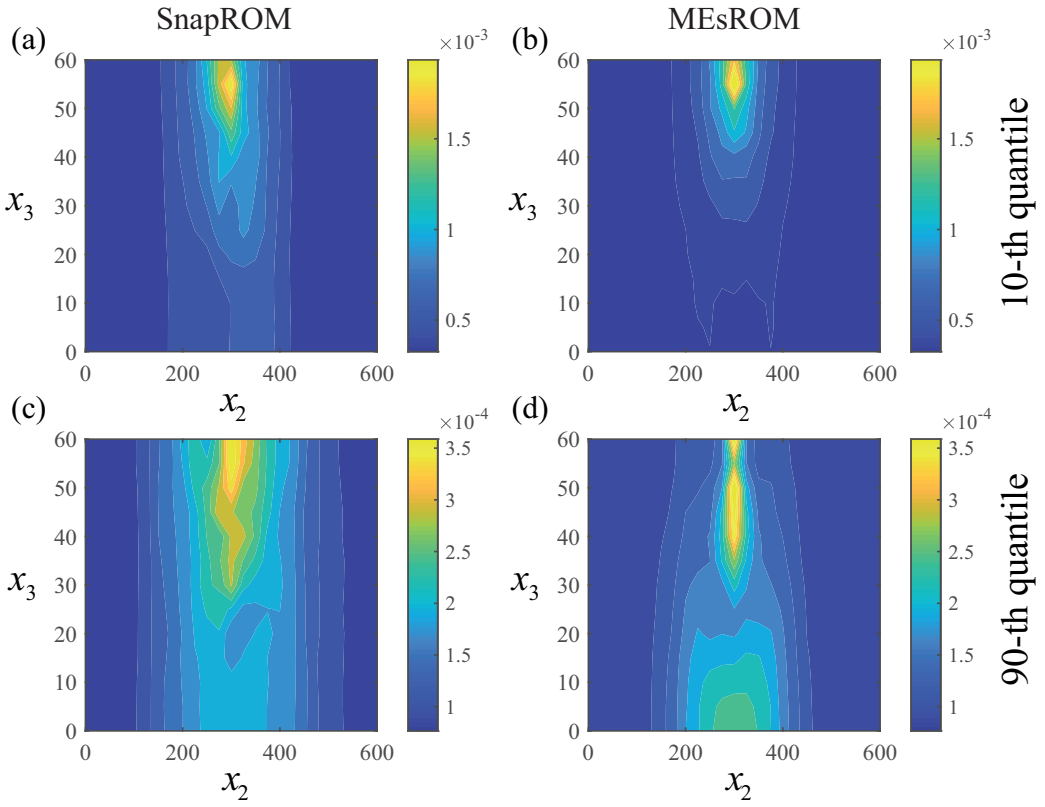


Figure 11: Spatial distribution of the relative difference between the 10-th (top row) and 90-th (bottom row) quantiles of head probability distributions obtained by FSM and the two reduced-order methods with $m = 30$ along section $x_1 = 300$ in the 3D scenario.

| **Short name** | **Description** | **Main equations** | **Leading online cost** |
|---|---|---|---|
| FSM | Full System Model based on standard finite element method | (1) and (10) | $O(n^3)$ |
| SnapROM | Reduced Order Model based on the snapshot technique for the evaluation of the projection matrix | (1) and (10)-(12) | $O(N_{sn}m + nm^2)$ |
| MEsROM | Reduced Order Model based on stochastic moment equations for the evaluation of the projection matrix | (2)-(9) and (10)-(12) | $O(N_{sn}m + nm^2)$ |

Table 1: Description of the compared full-system and reduced-order methods used to perform the Monte Carlo simulations.

| | $n$ | **FSM** | **MEs** | **Ratio** | **Snapshot** |
|---|---|---|---|---|---|
| 2D scenario | 3321 | 8.3E-3 | 4.0 | 482 | 500 |
| 3D scenario | 8125 | 0.26 | 73.69 | 282 | 300 |

Table 2: Number of nodes in the 2D and 3D scenarios; CPU times for solving stochastic moment equations (MEs); CPU time of a single full system model (FSM) solution computed by averaging 100,000 MC runs; The ratio between CPU times of solving MEs and mean single run; selected snapshot size for SnapROM according to the CPU cost of MEs (refer to Section 3.2).

| | $\mathbf{p}_1$ | $\mathbf{p}_2$ | $\mathbf{p}_3$ | $\mathbf{p}_4$ | $\mathbf{p}_5$ | $\mathbf{p}_6$ |
|---|---|---|---|---|---|---|
| $N_{sn} = 500$ | 0.0485 | 0.0650 | 0.1609 | 0.5955 | 0.5598 | 0.3433 |
| MEs | 0.0172 | 0.0467 | 0.20687 | 0.1163 | 0.2943 | 0.2595 |

Table 3: $L_2$-norm of the error between the reference eigenvectors obtained with a large number of snapshots ($N_{sn}$=100,000), and their approximations obtained using $N_{sn}$=500 snapshots and MEs. The parameter setting is the same used in Figure 4.

| | | $(\tau_1, \tau_2)$ | | | |
|---|---|---|---|---|---|
| | $\sigma^2_Y$ | **(2,1)** | **(4,2)** | **(8,4)** | **(16,8)** |
| $\eta_{\sigma^2_h}$ | **0.1** | 2.6391 | 1.9840 | 1.8862 | 1.8571 |
| | **1.0** | 1.4032 | 1.2347 | 1.1618 | 1.1528 |
| | **2.0** | 0.9477 | 0.9667 | 0.9793 | 1.0912 |
| $\eta_h$ | **0.1** | 1.0729 | 1.0550 | 1.0494 | 1.0477 |
| | **1.0** | 1.0916 | 1.0873 | 1.0655 | 1.0566 |
| | **2.0** | 1.1006 | 1.1025 | 1.1051 | 1.1382 |

Table 4: Ratio between the errors ($\eta_{\sigma^2_h}$ and $\eta_h$) by SnapROM model (with $N_{sn} = 500$ and $m = 30$) and the errors by MEsROM using 100,000 FSM solutions as references. Results obtained for different values of the hydraulic transmissivity variance $\sigma^2_Y$ and correlation scales $(\tau_1, \tau_2)$. Results larger than one indicate a better performance of MEsROM with respect to SnapROM.

| | $m$ | | | | |
|---|---|---|---|---|---|
| | **5** | **10** | **20** | **30** | **40** |
| 2D scenario | 9.96 | 6.39 | 3.92 | 2.09 | 1.11 |
| 3D scenario | 28.54 | 23.01 | 12.80 | 8.23 | 4.41 |

Table 5: Relative CPU time (with respect to the FSM CPU time) of a single SnapROM/MEsROM model solution during the online-stage as a function of reduced-order dimension, $m$, on the 2D and 3D scenarios analyzed. The CPU time is obtained by averaging the CPU time of 100,000 ROM runs.