AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE

10.1002/2017WR020905

Uncertainty Quantification in Scale-Dependent Models of Flow in Porous Media

A. M. Tartakovsky[1], M. Panzeri[2], G. D. Tartakovsky[3], and A. Guadagnini[2]

[1]Computational Mathematics Group, Pacific Northwest National Laboratory, Richland, WA, USA, [2]Dipartimento di Ingegneria Civile e Ambientale, Politecnico di Milano, Milano, Italy, [3]Hydrology Group, Pacific Northwest National Laboratory, Richland, WA, USA

Key Points:
- Solution of flow equations is scale dependent
- Accuracy of uncertainty quantification methods is scale dependent
- We rigorously define relevant scales and present methods to obtain accurate solutions at both large and small scales

Correspondence to:
A. M. Tartakovsky,
Alexandre.Tartakovsky@pnnl.gov

Citation:
Tartakovsky, A. M., Panzeri, M., Tartakovsky, G. D., & Guadagnini, A. (2017). Uncertainty quantification in scale-dependent models of flow in porous media. *Water Resources Research*, *53*, 9392–9401. https://doi.org/10.1002/2017WR020905

Received 6 APR 2017
Accepted 10 OCT 2017
Accepted article online 16 OCT 2017
Published online 19 NOV 2017

Abstract Equations governing flow and transport in randomly heterogeneous porous media are stochastic and scale dependent. In the moment equation (ME) method, exact deterministic equations for the leading moments of state variables are obtained at the same support scale as the governing equations. Computable approximations of the MEs can be derived via perturbation expansion in orders of the standard deviation of the random model parameters. As such, their convergence is guaranteed only for standard deviation smaller than one. Here, we consider steady-state saturated flow in a porous medium with random second-order stationary conductivity field. We show it is possible to identify a support scale η^*, where the typically employed approximate formulations of MEs yield accurate (statistical) moments of a target state variable. Therefore, at support scale η^* and larger, MEs present an attractive alternative to slowly convergent Monte Carlo (MC) methods whenever lead-order statistical moments of a target state variable are needed. We also demonstrate that a surrogate model for statistical moments can be constructed from MC simulations at larger support scales and be used to accurately estimate moments at smaller scales, where MC simulations are expensive and the ME method is not applicable.

Plain Language Summary Equations governing flow and transport in randomly heterogeneous porous media are stochastic and scale dependent. In the moment equation method, exact deterministic equations for the leading moments of state variables are obtained at the same support scale as the governing equations. Computable approximations of these equations can be derived via perturbation expansion in orders of the standard deviation of the random model parameters. As such, their convergence is guaranteed only for standard deviation smaller than one. Here, we consider steady-state saturated flow in a porous medium with random second-order stationary conductivity field. We show it is possible to identify a support scale, where the typically employed approximate formulations of moment equations yield accurate moments of a target state variable.

1. Introduction

Because of the complexity and inherent uncertainty of natural systems, uncertainty quantification (UQ) has become an essential part of predictive modeling. Mathematical models (usually in the form of partial differential equations [PDEs]) of natural systems are defined on a certain support scale. For example, the Darcy equation, which provides a continuum description of flow in porous media, can be obtained by integrating the Navier-Stokes equations, describing flow on the pore scale, over a volume of size η, the support scale of the Darcy equation.

Unknown parameter distributions (space-dependent coefficients in the governing equations) are a common cause of uncertainty in numerical models. Parameter values can be obtained from experiments conducted at different (measurement) scales η_m, depending on the type and resolution of an experiment. Uncertain parameters are typically modeled as random functions of space, rendering governing equations stochastic. The parameter statistics may depend on the measurement scale. For example, the variance of hydraulic conductivity tends to decrease, while its correlation length increases with increasing measurement scale (Attinger, 2003; Neuman & Di Federico, 2003; Tartakovsky et al., 2004).

It is well understood that governing equations and data should be defined on the same support scale. Neuman and Orr (1993) state that this support scale should be equal to the measurement scale (i.e., $\eta \equiv \eta_m$) to be able to constrain a model on data. We argue that the support scale of a model does not necessarily need to coincide with the measurement scale. In practice, it is common to deal with a range of measurement scales for different parameters, even for the same parameter (e.g., Li et al., 2009; Vanmarcke, 2010).

In this work, we study the stochastic PDEs solution dependence on η. The support η defines the model resolution and depends on (should be much smaller than) the domain size and/or the smallest modeled feature. The support scale should not be confused with the (numerical) resolution (e.g., grid size) of a numerical solution of governing equations, which is determined by the desired numerical accuracy. We mostly focus on problems where parameters are described as second-order stationary fields and discuss how data measurements (at scale η_m) can be used to constrain the model at the scale η. The proposed treatment of random properties of porous media differs from the multiscale approach of Neuman and Di Federico (2003) that treats statistical properties of data as a function of the observation scale (e.g., domain size).

Over the past three to four decades, many UQ methods have been proposed, including methods to compute the probability density function (PDF) (Tartakovsky & Broyda, 2011; Wang et al., 2013, 2015) or leading statistical moments of the state variables. For subsurface flow problems, the methods most commonly employed for UQ can be ascribed to three main categories: Monte Carlo (MC), polynomial chaos (PC), and moment equation (ME) methods.

The MC and PC methods are based on the approximation of random input parameters with a finite number N of random variables (e.g., by means of the Karhunen-Loéve [KL] expansion) (Lin & Tartakovsky, 2009, 2010; Lin et al., 2010). For a given approximation error, N increases with the decreasing correlation length (relative to the domain size) of random system/model parameters. A significant disadvantage of the PC method, often called "the curse of dimensionality," is that its operation count grows exponentially with N (Xiu, 2009). On the other hand, the MC method's operation count is much less sensitive to N. A key limitation of the MC method resides in its low convergence rate, which requires a large number of realizations to compute accurate statistics of the state variables (Ballio & Guadagnini, 2004). As such, while the MC method is conceptually simple and relatively straightforward to implement, its effectiveness can be severely limited for large-scale problems requiring high model resolution.

The key idea underlying ME methods is to derive a set of PDEs that are directly satisfied by the leading (statistical) moments of the quantities of interest (QoIs) (Neuman & Orr, 1993; Tartakovsky & Neuman, 1998). As such, MEs are defined on the same support scale as the governing stochastic PDEs (Tartakovsky et al., 2004). Computable approximations of otherwise exact MEs can be obtained via perturbation expansion of moments of the state variables in orders of the standard deviation of the random inputs (Morales-Casique et al., 2006; Neuman & Orr, 1993; Tartakovsky & Neuman, 1998; Tartakovsky et al., 2002, 2003; Ye et al., 2004).

Several studies have demonstrated that for some special cases, the ME method can be accurate for the standard deviation of the input parameters as large as 2.0–4.0 (Guadagnini & Neuman, 1999b; Guadagnini et al., 2003; Riva et al., 2001), even though the guaranteed convergence of the ME methods requires the standard deviation be smaller than one. We should also note that moment solutions of stochastic advection-diffusion equations with random advection velocity become unphysical (e.g., negative mean concentration develops) with time even for standard deviation of velocity smaller than one (Jarman & Tartakovsky, 2008, 2011, 2013; Morales-Casique et al., 2006).

When applicable, the ME method has shown to be computationally more efficient for computing the leading moments of the random state variables than the traditional MC method (Ye et al., 2004). Recently, MEs have been used for including geostatistical inverse modeling and field data assimilation of groundwater flow (Hernandez et al., 2003, 2006; Panzeri et al., 2013, 2014, 2015).

In many environmental applications, including flow and transport settings in the subsurface, the standard deviation of the model parameters can be significantly larger than one, and the correlation length can be much smaller than the domain size. Therefore, the ME and PC methods cannot be readily applied. A commonly encountered scenario also involves the need to condition model predictions on measurements associated with scales much smaller than the domain size. For example, in subsurface reservoir flow problems, the hydraulic conductivity is estimated from laboratory experiments on scales ranging from 0.01 to 0.1 m, while the typical size of a reservoir for environmental and/or industrial applications may range from 10^4 to 10^6 m. In such systems, it may not be necessary for the support scale of the governing equations to be as small as the measurement scale of the parameters.

In this work, we focus on the ME method's application to problems with highly heterogeneous parameters, i.e., parameters with standard deviation much larger than one and spatial correlation scale much smaller than a characteristic domain length. Specifically, we consider groundwater flow in highly heterogeneous porous media, described by the combination of the continuity equation and Darcy's law with a random, correlated-in-space conductivity field. We demonstrate that it is possible to determine the support scale of the Darcy equation η^*, where the standard deviation of the (natural) logarithm of conductivity is equal to one, and the ME method provides an accurate solution for the leading-order statistical moments of hydraulic head. If η^* is much smaller than the domain size, the proposed approach provides sufficiently resolved moment solutions, and it can be used as a computationally efficient alternative to the MC method for the characterization of key statistical moments of the state variables of interest.

For problems where a solution is needed at a scale significantly smaller than η^*, we propose a surrogate model approach. We demonstrate that a surrogate model for leading statistical moments can be constructed from MC simulations at larger support scales and used to accurately estimate moments at smaller scales, where MC simulations are expensive and the ME applicability is not guaranteed due to the large parameter variance at these scales.

2. Flow Equations

The description of flow in porous media depends on the domain size and desired model resolution. At the pore scale, incompressible isothermal low Reynolds number fluid flow in a porous medium occupying volume Ω is described by the continuity equation

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot [\rho \mathbf{v}(\mathbf{x})] \quad \mathbf{x} \in \Omega_p \tag{1}$$

and the Stokes equation

$$-\nabla P + \mu \nabla^2 \mathbf{v} + \rho \mathbf{g} = 0 \quad \mathbf{x} \in \Omega_p \tag{2}$$

subject to the no-flow boundary condition at the boundary $\partial\Omega_p$ of the domain Ω_p occupied by pores, $\Omega_p \subset \Omega$ (Bear, 2013). Here, ρ, P, and $\mathbf{v}$ are the fluid (pore-scale) density, pressure, and velocity, all defined on Ω_p. The continuum description of the flow in porous media is obtained by first defining the effective density ρ_η, pressure P_η, and velocity $\mathbf{v}_\eta$ (Bear, 2013):

$$\rho_\eta(\mathbf{x}) = \frac{1}{\theta_\eta ||\Omega_\eta||} \int_{\Omega_\eta} \rho(\mathbf{y}) d\mathbf{y} \tag{3}$$

$$P_\eta(\mathbf{x}) = \frac{1}{\theta_\eta ||\Omega_\eta||} \int_{\Omega_\eta} P(\mathbf{y}) d\mathbf{y} \tag{4}$$

$$\mathbf{v}_\eta(\mathbf{x}) = \frac{1}{\theta_\eta ||\Omega_\eta||} \int_{\Omega_\eta} \mathbf{v}(\mathbf{y}) d\mathbf{y} \tag{5}$$

where the averaging volume $\Omega_\eta(\mathbf{x})$ is a volume with the characteristic size η (e.g., a sphere with the radius η) centered at $\mathbf{x}$; ρ, P, and $\mathbf{v}$ are assumed to be equal to zero within the solid phase; $||\Omega_\eta||$ is the volume of Ω_η; and θ_η is the porosity defined as the ratio of the pore volume, within Ω_η, to $||\Omega_\eta||$. Next, equations for the effective variables can be obtained by averaging the pore-scale Stokes and continuity equations over Ω_η (using, for example, the method of volume averaging (Bear, 2013)) in the form of the (effective) continuity equation

$$\frac{\partial \rho_\eta}{\partial t} = -\nabla \cdot \left[\rho_\eta \mathbf{q}_\eta(\mathbf{x})\right] \tag{6}$$

and the Darcy equation

$$\mathbf{q}_\eta(\mathbf{x}) = -\frac{k_\eta(\mathbf{x})\rho_\eta g}{\mu}\nabla h_\eta(\mathbf{x}) \quad h_\eta(\mathbf{x}) = \frac{P_\eta(\mathbf{x})}{\rho g} + z \tag{7}$$

where h_η is the hydraulic head, $\mathbf{q}_\eta = \theta_\eta \mathbf{v}_\eta$ is the Darcy flux, and the permeability $k_\eta(\mathbf{x})$ is a property of the porous medium at the scale η. Therefore, we refer to η as the support of the governing equations (6) and (7) and all effective variables in these equations. It is important to note that all variables and parameters in equations (6) and (7) must have the same support η. This means the parameters measured on the scale η_m that differ from η must be upscaled (or downscaled) to the scale η, or η should be set equal to η_m. In homogeneous porous media, k_η is independent of η (and $\mathbf{x}$) for η larger than the size of the so-called "representative elementary volume" (REV). In heterogeneous porous media, the REV concept is ambiguous. A REV can be difficult or impossible to define because in the heterogeneous porous media, $k_\eta(\mathbf{x})$ and the state variables in equations (6) and (7) may strongly depend on η or η_m (De Marsily, 1986; Neuman, 2014). For example, Neuman (1994) and Clauser (1992) demonstrate that permeability values may vary by more than 10 orders of magnitude as η_m changes from centimeters to kilometers.

We are interested in groundwater flow in randomly heterogeneous porous media with known ensemble permeability statistics but deterministically unknown spatial permeability distribution. Key permeability statistics can be obtained from available permeability measurements collected on the scale η_m. We start by setting the support scale of the model equal to the measurement scale, $\eta_0 = \eta_m$, and, following a common practice in hydrology, assume that k_{η_0} has a lognormal distribution, i.e., $k_{\eta_0} = \exp(Y_{\eta_0})$, and Y_{η_0} is a normally distributed random function of space. We also assume that Y_{η_0} has variance $\sigma^2_{Y_{\eta_0}}$, and the covariance function

$$C_{Y_{\eta_0}}(\mathbf{x},\mathbf{y}) = \overline{Y'_{\eta_0}(\mathbf{x})Y'_{\eta_0}(\mathbf{y})} = \sigma^2_{Y_{\eta_0}} \exp\left[-\frac{|\mathbf{x}-\mathbf{y}|}{l_{\eta_0}}\right] \tag{8}$$

where the overbar indicates expectation (ensemble average), $\mathbf{x}$ and $\mathbf{y}$ are two locations in the domain, l_{η_0} is the correlation length, and primed quantities indicate random fluctuations around the expected (or mean) value. We consider porous media with large variance and small correlation length, i.e., $\sigma^2_{Y_{\eta_0}} \gg 1$ and $l_{\eta_0} \ll L$ (L being domain size). We exemplify our findings by considering a two-dimensional system, where the permeability field is characterized by $\sigma^2_{Y_{\eta_0}} = 5$ and $l_{\eta_0}/L = 5/268 \approx 0.02$. One realization of this field, obtained with a Sequential Gaussian Simulation based on SGSIM (Deutsch & Journel, 1998), is depicted in Figure 1a. For porous media with such statistics of k_{η_0}, the ME method fails to provide the accurate solution of equations (6) and (7), which is demonstrated in section 4, where we show that the ME method produces $\approx 20\%$ error in the hydraulic head variance estimate. Therefore, the only readily available method in this case is the MC method. In the following, we propose an alternative to MC and demonstrate that we can define the support scale $\eta^* > \eta_0$, on which an accurate moment solution of equations (6) and (7) for the variance of h can be obtained.

3. Scale-Dependent Statistics of Log-Permeability Field

It has been empirically shown (Tartakovsky et al., 2004) that the exponential covariance function of Y_η at any support η can be approximated as

$$C_{Y_\eta}(|\mathbf{x}-\mathbf{y}|) = \sigma^2_{Y_\eta} \exp\left[-\frac{|\mathbf{x}-\mathbf{y}|}{l_\eta}\right] \tag{9}$$

where $\sigma^2_{Y_\eta}$ and l_η are the (scale-dependent) variance and correlation length of Y_η.

To empirically determine the scaling relationships $\sigma^2_{Y_\eta}(\eta, \eta_0, \sigma^2_{Y_{\eta_0}})$ and $l_\eta(\eta, \eta_0, l_{\eta_0})$, we generate realizations of Y_η for different η. We first simulate 10,000 unconditional realizations of the log-permeability field, Y_{η_0}, in the 268 × 268 computational domain with the unit grid size ($\eta_0 = 1$) using the SGSIM software (as in Figure 1a). Then, we obtain Y_η ($\eta = 2, 4, 8, 16$, and 32) by sequentially coarsening Y_{η_0} using

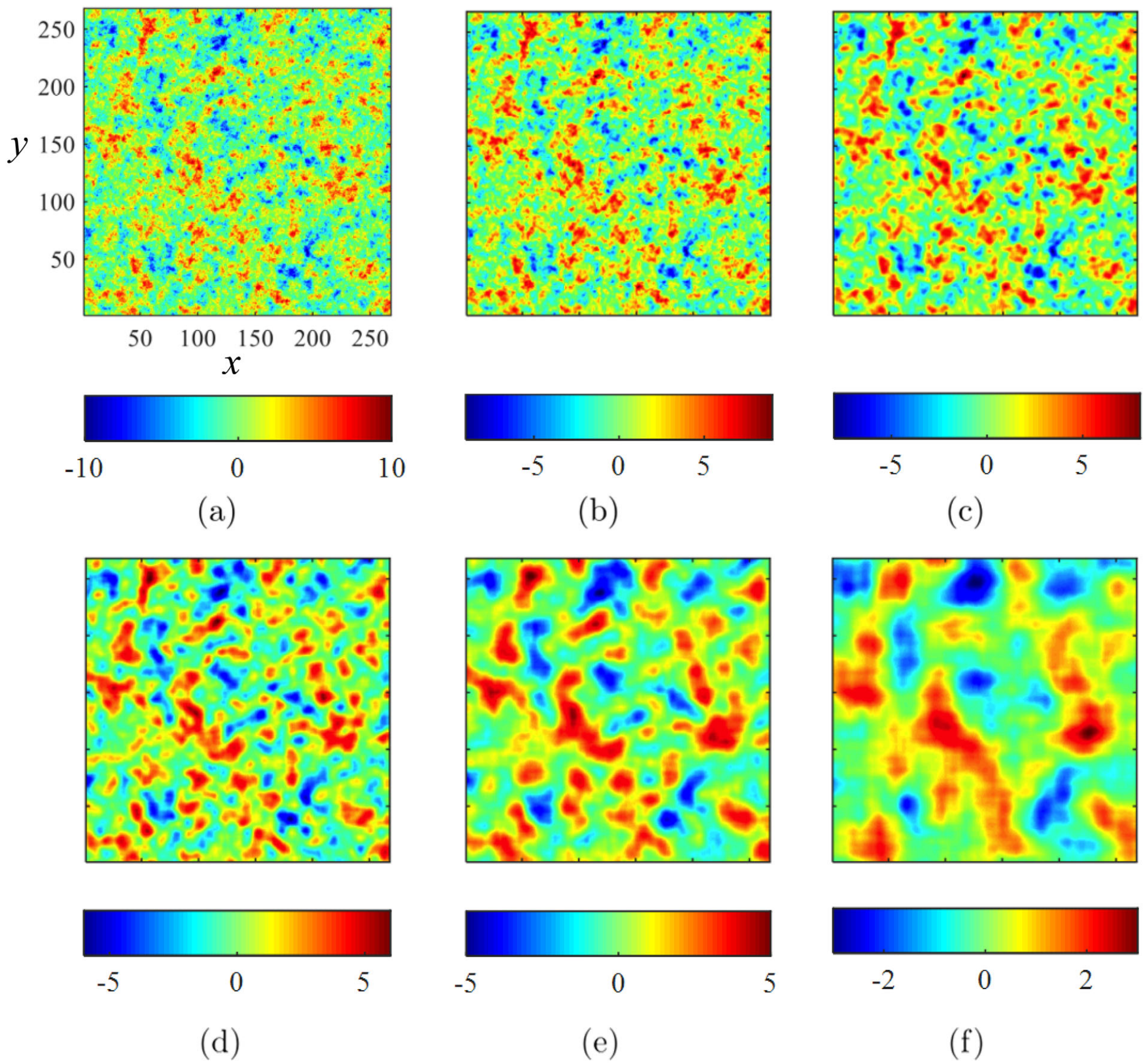

Figure 1. Sample realization of the random log-permeability field $Y_\eta = \ln K_\eta$ as a function of η: (a) $\eta = 1$; (b) $\eta = 2$; (c) $\eta = 4$; (d) $\eta = 8$; (e) $\eta = 16$; and (f) $\eta = 32$.

a moving averaging window approach with the window size η. Based on Attinger (2003), we assume Y_η to be the arithmetic mean of Y_{η_0} (i.e., k_η is the geometric mean of k_{η_0}) within the window of size η and compute it as

$$(Y_\eta)_{ij} = \frac{1}{\eta^2} \sum_{l=i-\eta/2+1}^{i+\eta/2} \sum_{m=j-\eta/2+1}^{j+\eta/2} (Y_{\eta_0})_{lm} \quad i,j=1,2,\ldots,268 \tag{10}$$

In this work, we do not change the numerical resolution with increasing support of Y, i.e., Y_η ($\eta \geq 2$) is defined on the domain 268×268 with the unit grid size.

Figure 1 depicts one realization of k_η for $\eta = 1, 2, 4, 8, 16$, and 32. Figure 2 shows the variogram, $\gamma_{Y_\eta}(z)$, of Y_η for $\eta = 1, 2, 4, 8, 16$, and 32, numerically computed from the realizations of Y_η, and equation (9) as $\gamma_{Y_\eta}(z) = \sigma^2_{Y_\eta} - C_{Y_\eta}(z)$ (z is the separation distance). In equation (9), $\sigma^2_{Y_\eta}$ and l_η are found from fitting the variogram computed from the realizations of Y_η. Figure 3 shows $\sigma^2_{Y_\eta}$ and l_η as functions of η/η_0. We propose the following relationship for $\sigma^2_{Y_\eta}$ as a function of η:

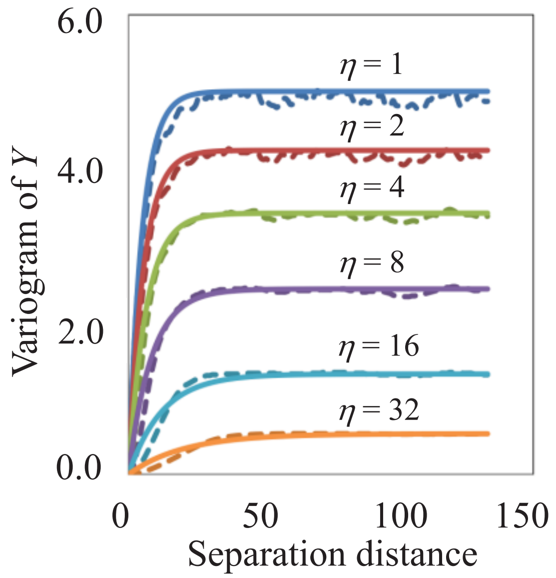

Figure 2. Variogram of Y_η as a function of η, computed numerically (dashed lines) from the realizations of Y_η and fitting equation (9) (solid lines).

$$\sigma^2_{Y_\eta} = \sigma^2_{Y_{\eta_0}} \left(\frac{1}{1+\left[\frac{\eta}{\eta_0}-1\right]^{b_1}\frac{1}{b_2}} \right)^{b_3} \tag{11}$$

where $b_1=1,\ b_2=10$, and $b_3=1.56$ are found from the best fit to the numerical results. A similar expression was derived in Attinger (2003) for the Gaussian covariance function with $b_1=2,\ b_2=4$, and $b_3=2$.

A relationship for l_{Y_η} as a function of η is proposed in the form:

$$l_\eta = c\left[\frac{\eta}{\eta_0}-1\right]+l_{\eta_0} \tag{12}$$

where $c=0.5$. Figure 3 shows that the correlation length of Y increases, and variance decreases with increasing η.

The covariance $\tilde{C}_{Y_\eta}(|\mathbf{x}-\mathbf{y}|)$ can be obtained by first taking a continuum limit of equation (10)

$$\tilde{Y}'_\eta(\mathbf{x}) = \frac{1}{\eta^2}\int_{\Omega_\eta(\mathbf{x})} Y'_{\eta_0}(\mathbf{x}')d\mathbf{x}' \tag{13}$$

multiplying the left-hand side of (13) with $\tilde{Y}'_\eta(\mathbf{y})$ and the right-hand side with $\frac{1}{\eta^2}\int_{\Omega_\eta(\mathbf{y})} Y'_{\eta_0}(\mathbf{y}')d\mathbf{y}'$ and taking the average:

$$\tilde{C}_{Y_\eta}(|\mathbf{x}-\mathbf{y}|) = \frac{1}{\eta^4}\int_{\Omega_\eta(\mathbf{x})}\int_{\Omega_\eta(\mathbf{y})} C_{Y_{\eta_0}}(|\mathbf{x}'-\mathbf{y}'|)d\mathbf{x}'d\mathbf{y}' \tag{14}$$

Here, ~ denotes variables and functions obtained by taking a continuum limit of the corresponding quantity. For "separable" exponential, Gaussian, and spherical correlation functions $C_{Y_{\eta_0}}(|\mathbf{x}'-\mathbf{y}'|)$, $\tilde{C}_{Y_\eta}(|\mathbf{x}-\mathbf{y}|)$ can be obtained in a closed form (Journel & Huijbregts, 1978; Neuman & Depner, 1988; Vanmarcke, 2010; Li et al., 2009). It follows from equation (14) that $C_{Y_{\eta_0}}(|\mathbf{x}-\mathbf{y}|) = \lim_{\eta\to 0}\tilde{C}_{Y_\eta}(|\mathbf{x}-\mathbf{y}|)$, meaning $\tilde{C}_{Y_\eta}(|\mathbf{x}-\mathbf{y}|)$ implicitly assumes that the support of Y_{η_0} is zero, i.e., $\eta_0=0$. Therefore, $\tilde{C}_{Y_\eta}(|\mathbf{x}-\mathbf{y}|)$ is expected to be an accurate model for $\eta \gg \eta_0$. On the other hand, the proposed covariance function $C_{Y_\eta}(|\mathbf{x}-\mathbf{y}|)$ reduces to $C_{Y_{\eta_0}}(|\mathbf{x}-\mathbf{y}|)$ as $\eta \to \eta_0$ for any nonzero η_0. Finally, we note that the coarsening log-conductivity (or permeability) model (10) and its continuum form (13) have been shown to be accurate for saturated flow problems (e.g., Attinger, 2003; Neuman & Depner, 1988). Extending the proposed framework to unsaturated and multiphase flow problems may require other models for upscaling log-permeability (Das & Hassanizadeh, 2005).

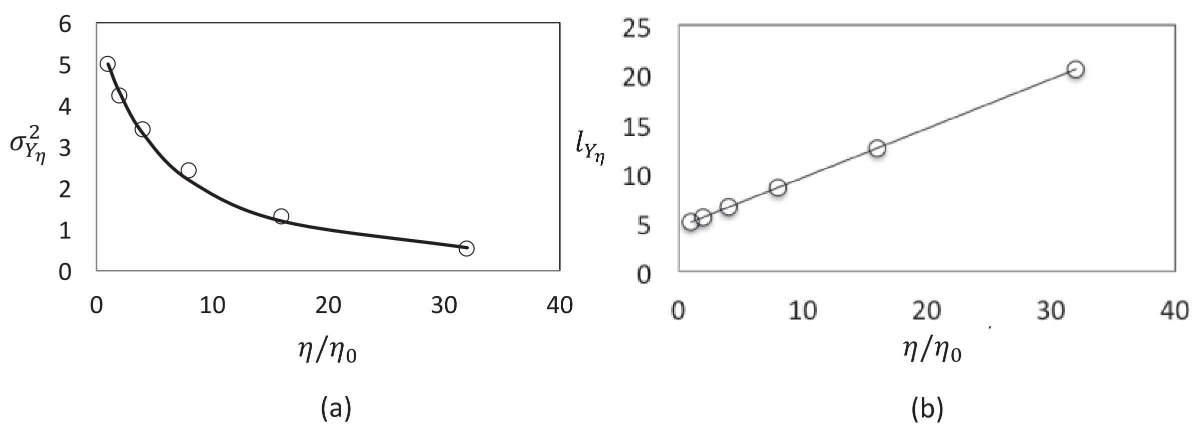

Figure 3. Open circles denote (a) variance and (b) correlation length of Y_η versus η/η_0, found from fitting equation (9) to the variograms computed from realizations of Y_η. The solid line denotes variance and correlation length found from equations (11) and (12).

4. Moment Equations

Exact nonlocal (integro-differential) MEs for the mean and covariance of h, satisfying the steady-state form of equations (6) and (7) subject to appropriate boundary conditions, are derived in Neuman and Orr (1993). Closed-form formulations and finite elements discretization of these steady-state MEs are given in Guadagnini and Neuman (1999a). The closed-form equations approximate the moments of the flow equations (6) and (7) based on a perturbation method, where all state variables and random parameters are expressed as series expansions in powers of σ_{Y_η}. The computable form of the MEs is obtained by collecting the terms of the same order and averaging the resulting equations in the probability space.

Here, we focus on analyzing the zero- and second-order approximations of the unconditional steady-state solution of equations (6) and (7) for (ensemble) mean and variance-covariance of h. We solve these equations on a rectangular domain $[0, 268]\times[0, 268]$ subject to the boundary conditions:

$$h(x=0, y)=H, \quad h(x=268, y)=0 \quad y \in (0, 268) \tag{15}$$

$$\left.\frac{\partial h}{\partial y}\right|_{y=0}=\left.\frac{\partial h}{\partial y}\right|_{y=268}=0 \quad x \in (0, 268) \tag{16}$$

The zero- and second-order approximations of the mean head, $\overline{h(\mathbf{x})}^{(0)}$ and $\overline{h(\mathbf{x})}^{(2)}$, respectively, satisfy (Guadagnini & Neuman, 1999a):

$$\nabla \cdot \left[K_{G_\eta}(\mathbf{x})\nabla\overline{h(\mathbf{x})}^{(0)}\right]=0 \tag{17}$$

$$\nabla \cdot \left\{K_{G_\eta}(\mathbf{x})\left[\nabla\overline{h(\mathbf{x})}^{(2)}+\frac{\sigma^2_{Y_\eta}(\mathbf{x})}{2}\nabla\overline{h(\mathbf{x})}^{(0)}\right]+\mathbf{r}^{(2)}(\mathbf{x})\right\}=0 \tag{18}$$

$$\mathbf{r}^{(2)}(\mathbf{x})=\int K_{G_\eta}(\mathbf{x})K_{G_\eta}(\mathbf{x})C_{Y_\eta}(\mathbf{x}, \mathbf{y})\nabla_x\nabla_y^T\overline{G(\mathbf{y}, \mathbf{x})}^{(0)}d\mathbf{y} \tag{19}$$

All quantities are defined on the support η, K_{G_η} is the geometric mean of the conductivity field, and $G(\mathbf{y}, \mathbf{x})^{(0)}$ is the zero-order approximation of the (ensemble) mean Green's function of the problem adjoint to equations (6) and (7) subject to the homogeneous version of the boundary conditions (15) and (16) (for details, refer to Guadagnini and Neuman (1999a)).

The equation satisfied by the second-order approximation of the head covariance $C_h^{(2)}(\mathbf{y}, \mathbf{x})$ is

$$\nabla_y \cdot \left[K_{G_\eta}(\mathbf{y})C_h^{(2)}(\mathbf{y}, \mathbf{x})\right]=-\nabla_y \cdot \left[u^{(2)}(\mathbf{y}, \mathbf{x})\nabla_y\overline{h(\mathbf{x})}^{(0)}\right] \tag{20}$$

where

$$u^{(2)}(\mathbf{x}, \mathbf{y})=-K_{G_\eta}(\mathbf{x})\int \nabla_z^T\overline{h(\mathbf{z})}^{(0)}\nabla_z\overline{G(\mathbf{z}, \mathbf{y})}^{(0)}K_{G_\eta}(\mathbf{z})C_{Y_\eta}(\mathbf{z}, \mathbf{x})d\mathbf{z} \tag{21}$$

We illustrate our results by considering the spatial distribution of head variance, which is critical in the context of environmental UQ. For a given support scale η, the numerical solution of the MEs is performed by employing the algorithms and codes presented by Guadagnini and Neuman (1999a) and following the same approach as Guadagnini and Neuman (1999b). The ME's solution is then compared against $\sigma^2_{h_\eta}$ obtained from the MC method. In the MC method, we generate 10,000 realizations of $k_\eta=\exp(Y_\eta)$ for $\eta=1$, $2, 4, 8, 16$, and 32 (section 3 describes how realizations of Y_η are generated) and numerically solve equations (6) and (7) for each realization of k_η. Moments of h_η are then calculated from the sample of available realizations. MC does not require any particular assumptions, and, in the following, we consider $\sigma^2_{h_\eta}$ obtained from MC as reference values against which the ME solutions are compared for all analyzed values of η.

Figure 4 depicts the spatial distribution of normalized head variance $\sigma^2_{h_\eta}$ obtained from the ME and MC methods. Interpretation of these results should be performed in conjunction with the analysis of Figure 2, where it shows that $\sigma^2_{Y_\eta} > 1$ for $\eta < 32$ and $\sigma^2_{Y_\eta}=0.5$ for $\eta = 32$. As expected, values of $\sigma^2_{h_\eta}$ obtained from the two methods are in agreement for $\sigma^2_{Y_\eta} < 1$, which is consistent with the ME method's accuracy for small variances of Y. For $\sigma^2_{Y_\eta} > 1$, the ME method tends to underestimate $\sigma^2_{h_\eta}$, by as much as 20% for $\sigma^2_{Y_\eta}=5$ and $\eta = 1$.

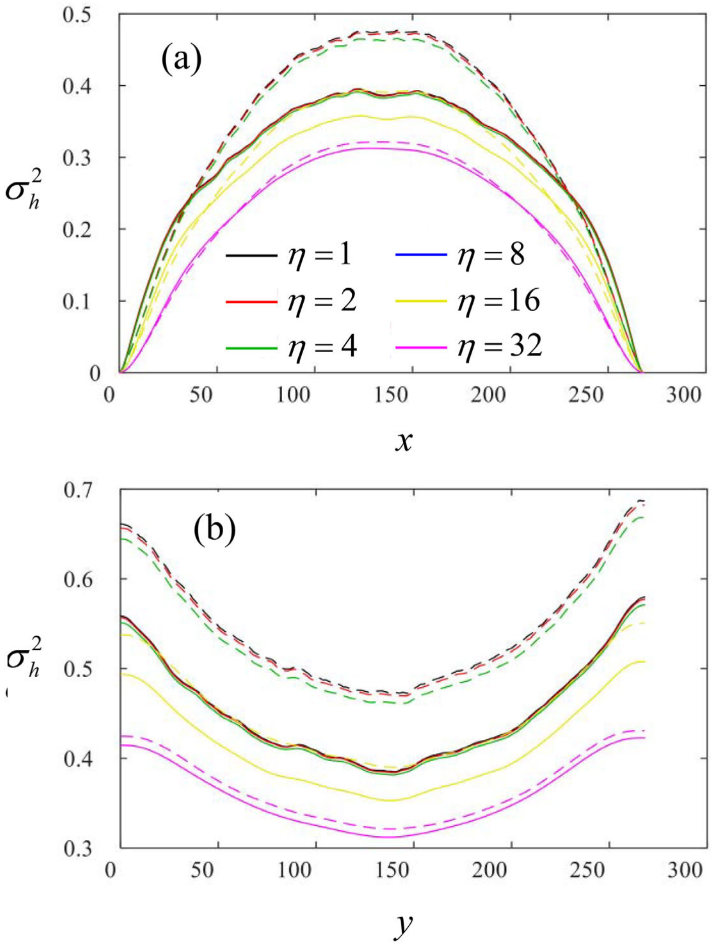

Figure 4. Variance of hydraulic head $\sigma^2_{h_\eta}(x,y)$ as a function of η and space, computed from the MC and ME methods: (a) $\sigma^2_{h_\eta}(x,y=128)$ and (b) $\sigma^2_{h_\eta}(x=128,y)$.

Our results confirm that the ME method does not yield an accurate prediction of $\sigma^2_{h_\eta}$ for small values of η. The main conclusion of this study is that it is possible to define η^* so an accurate solution for the variance of h_{η^*} can be obtained through the ME method. The limiting condition for η^* is that $\sigma^2_{Y_{\eta^*}}=1$, which coincides with the necessary condition for convergence of the series expansion upon which workable approximations of MEs are built. An estimate of η^* can be obtained through equation (11) as

$$\eta^*=\eta_0\left[b_2\left(\sigma^2_{Y_{\eta_0}}\right)^{1/b_3}-b_2\right]^{1/b_1}+\eta_0 \tag{22}$$

An important question remains regarding how to condition the estimate of η^* on data collected at the scale η_0. While we are not addressing this question directly in the present study, we hypothesize that the conditioning could be done via the conditioned covariance function (Li et al., 2009):

$$C^c_{Y_\eta}(\mathbf{x},\mathbf{y})=C_{Y_\eta}(\mathbf{x},\mathbf{y})-\sum_{i=1}^{N}\alpha_i(\mathbf{x})C_{Y_\eta,Y_{\eta_0}}(\mathbf{y},\mathbf{x}_i) \tag{23}$$

where the coefficients $\alpha_i(\mathbf{x})$ are found from the so-called "kriging equations:"

$$\sum_{i=1}^{N}\alpha_i(\mathbf{x})C_{Y_\eta}(\mathbf{x}_i,\mathbf{x}_j)=C_{Y_\eta,Y_{\eta_0}}(\mathbf{x},\mathbf{x}_j),\quad j=1,N \tag{24}$$

where $\mathbf{x}_j$ $(j=1,N)$ are data locations. The unconditional cross-covariance function between $Y_\eta(\mathbf{x})$ and $Y_{\eta_0}(\mathbf{y})$ can be found (in two-dimensions) as

$$C_{Y_\eta,Y_{\eta_0}}(|\mathbf{x}-\mathbf{y}|)=\frac{1}{\eta^4}\int_{\Omega_\eta(\mathbf{x})}\int_{\Omega_{\eta_0}(\mathbf{y})}C_{Y_{\eta_0}}(|\mathbf{x}'-\mathbf{y}'|)d\mathbf{x}'d\mathbf{y}' \tag{25}$$

Then, the "conditional" η^* can be found as the solution of the equation:

$$\max_{\mathbf{x}} C^c_{Y_{\eta^*}}(\mathbf{x},\mathbf{x};\eta^*)=1 \tag{26}$$

5. Surrogate Model

In some problems, it may be required to estimate the solution (statistical moments) at the scale η that is much smaller than η^*. As mentioned, the ME method could result in large errors at such small scales. On the other hand, MC might be prohibitively expensive for small η. This is because the grid size in MC simulations should be $\approx \eta/5$, and the cost of each realization of MC scales as $(L/\eta)^{\omega d}$, where L is the domain size, d is the number of dimensions, and $\omega>1$ is a solver-dependent constant. Furthermore, the MC convergence rate for the nth moment is proportional to $\sigma^{2n}_{Y_\eta}N^{-0.5}_{MC}$. Therefore, achieving a given accuracy for smaller η (and larger standard deviation σ_{Y_η}) requires computing a significantly larger number of realizations N_{MC}. A detailed analysis of the MC computational cost can be found in Leube et al. (2013).

We propose to use a surrogate model for $\sigma^2_{h_\eta}$ as a function of η to predict head variance for small values of η. The idea here is to construct a surrogate model using MC simulations for larger support scales than ones where we need to estimate moments.

Figure 5 shows the normalized head variance in the domain center $\sigma^2_{h_\eta}/\sigma^2_{Y_\eta}$, obtained from the MC simulations, as a function of η. Figure 5 also shows the surrogate model of $\sigma^2_{h_\eta}$,

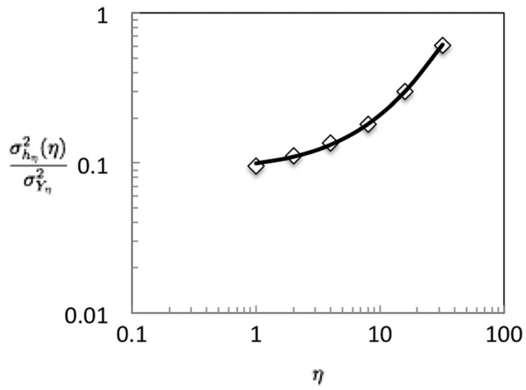

Figure 5. Variance of the hydraulic head $\sigma^2_{h_\eta}(x=L/2, y=L/2)$ as a function of η, computed from the MC and the surrogate model.

$$\frac{\sigma^2_{h_\eta}(\eta)}{\sigma^2_{Y_\eta}} = 0.0002\eta^2 + 0.01\eta + 0.089 \tag{27}$$

found by fitting a second-degree polynomial to $\sigma^2_{h_\eta}$ computed from MC simulations with the three largest values of η ($\eta = 8, 16$, and 32). In general, surrogate models are less accurate for extrapolation than interpolation of results. However, because the dependence of $\sigma^2_{h_\eta}$ on η is monotonic, the surrogate model can accurately predict $\sigma^2_{h_\eta}$ for the smallest support scales ($\eta = 1, 2$, and 4) with an error smaller than 0.4%. Here, the specific form of a surrogate model will depend on the governing equations and QoIs. Also, there are a number of advanced methods to construct (accurate) high-order surrogate models with a relatively small number of sampling points (Qian et al., 2006). In this study, we show that a simple, second-order polynomial function found by least squares fitting produces an accurate surrogate model, but, in general, the choice of the method to construct a surrogate model depends on the smoothness of the QoI in the space of η.

6. Conclusions

We have analyzed the scale dependence of the solutions of stochastic equations describing two-dimensional steady-state flow in randomly heterogeneous porous media with random second-order stationery hydraulic conductivity. Our work leads to the following conclusions:

1. It is possible to identify a suitable support scale η^*, where $\sigma^2_{Y_{\eta^*}} \leq 1$. Therefore, the perturbative solution of MEs is expected to be accurate. We demonstrate this concept by comparing the second-order ME solution of the head variance with the head variance obtained from the MC simulations.
2. The value of η^* can be found from equations (11) and (22).
3. It is possible to estimate the value of statistical moments at small η using a surrogate model constructed from MC with larger η. Because the computational cost of MC increases with decreasing η, the surrogate model can lead to significant computational savings.
4. While our study focuses on the ME approach, it is also useful to note that the PC method also may benefit from the proposed approach as its computational cost decreases with increasing support (and increasing correlation length) of random parameters, i.e., it is possible to determine the support η^* at which the correlation length is large enough (relative to the domain size) for the PC method to be computationally attractive. A detailed study of the computational cost dependence of PC methods on the support of random parameters is a subject for future research.

Acknowledgments
Research at Pacific Northwest National Laboratory (PNNL) was supported by the U.S. Department of Energy (DOE) Office of Advanced Scientific Computing Research (ASCR) as part of the Early Career Award "New Dimension Reduction Methods and Scalable Algorithms for Multiscale Nonlinear Phenomena" and DOE's Office of Biological and Environmental Research (BER) through the PNNL Subsurface Biogeochemical Research Scientific Focus Area project. Funding from MIUR (Italian ministry of Education, University and Research, Water JPI, WaterWorks 2014, Project: WE-NEED-Water NEEDs, availability, quality and sustainability) and from the European Union's Horizon 2020 Research and Innovation programme (Project "Furthering the knowledge Base for Reducing the Environmental Footprint of Shale Gas Development" FRACRISK - Grant Agreement No. 640979) are also acknowledged. The manuscript data can be accessed at https://sbrsfa.velo.pnnl.gov/datasets/?UUID=bc98365f-d106-423b-ae35-27606ded8498. PNNL is operated by Battelle for the DOE under Contract DE-AC05–76RL01830.

References

Attinger, S. (2003). Generalized coarse graining procedures for flow in porous media. *Computational Geosciences*, *7*(4), 253–273.
Ballio, F., & Guadagnini, A. (2004). Convergence assessment of numerical Monte Carlo simulations in groundwater hydrology. *Water Resources Research*, *40*, W04603. https://doi.org/10.1029/2003WR002876
Bear, J. (2013). *Dynamics of fluids in porous media*. Chelmsford, MA: Courier Corporation.
Clauser, C. (1992). Permeability of crystalline rocks. *Eos, Transactions American Geophysical Union*, *73*(21), 233–238.
Das, D., & Hassanizadeh, S. (2005). *Upscaling multiphase flow in porous media*. Berlin, Germany: Springer.
De Marsily, G. (1986). Quantitative hydrogeology (technical report). Fontainebleau, Paris: Paris School of Mines.
Deutsch, C., & Journel, A. (1998). *Geostatistical software library and user's guide*. New York, NY: Oxford University Press.
Guadagnini, A., & Neuman, S. P. (1999a). Nonlocal and localized analyses of conditional mean steady state flow in bounded, randomly nonuniform domains: 1. Theory and computational approach. *Water Resources Research*, *35*(10), 2999–3018.
Guadagnini, A., & Neuman, S. P. (1999b). Nonlocal and localized analyses of conditional mean steady state flow in bounded, randomly nonuniform domains: 2. Computational examples. *Water Resources Research*, *35*(10), 3019–3039.
Guadagnini, A., Riva, M., & Neuman, S. P. (2003). Three-dimensional steady state flow to a well in a randomly heterogeneous bounded aquifer. *Water Resources Research*, *39*(3), 1048. https://doi.org/10.1029/2002WR001443
Hernandez, A., Neuman, S., Guadagnini, A., & Carrera, J. (2003). Conditioning mean steady state flow on hydraulic head and conductivity through geostatistical inversion. *Stochastic Environmental Research and Risk Assessment*, *17*(5), 329–338.
Hernandez, A. F., Neuman, S. P., Guadagnini, A., & Carrera, J. (2006). Inverse stochastic moment analysis of steady state flow in randomly heterogeneous media. *Water Resources Research*, *42*, W05425. https://doi.org/10.1029/2005WR004449
Jarman, K. D., & Tartakovsky, A. M. (2008). Divergence of solutions to solute transport moment equations. *Geophysical Research Letters*, *35*, L15401. https://doi.org/10.1029/2008GL034495

Jarman, K. D., & Tartakovsky, A. M. (2011). Divergence of solutions to perturbation-based advection–dispersion moment equations. *Advances in Water Resources*, *34*(5), 659–670.

Jarman, K. D., & Tartakovsky, A. M. (2013). A comparison of closures for stochastic advection-diffusion equations. *SIAM/ASA Journal on Uncertainty Quantification*, *1*(1), 319–347.

Journel, A. G., & Huijbregts, C. J. (1978). *Mining geostatistics*. Cambridge, MA: Academic Press.

Leube, P. C., De Barros, F. P., Nowak, W., & Rajagopal, R. (2013). Towards optimal allocation of computer resources: Trade-offs between uncertainty quantification, discretization and model reduction. *Environmental Modelling & Software*, *50*, 97–107.

Li, C., Lu, Z., Ma, T., & Zhu, X. (2009). A simple kriging method incorporating multiscale measurements in geochemical survey. *Journal of Geochemical Exploration*, *101*(2), 147–154.

Lin, G., & Tartakovsky, A. M. (2009). An efficient, high-order probabilistic collocation method on sparse grids for three-dimensional flow and solute transport in randomly heterogeneous porous media. *Advances in Water Resources*, *32*(5), 712–722.

Lin, G., & Tartakovsky, A. M. (2010). Numerical studies of three-dimensional stochastic Darcy's equation and stochastic advection-diffusion-dispersion equation. *Journal of Scientific Computing*, *43*(1), 92–117.

Lin, G., Tartakovsky, A. M., & Tartakovsky, D. (2010). Uncertainty quantification via random domain decomposition and probabilistic collocation on sparse grids. *Journal of Computational Physics*, *229*(19), 6995–7012.

Morales-Casique, E., Neuman, S. P., & Guadagnini, A. (2006). Non-local and localized analyses of non-reactive solute transport in bounded randomly heterogeneous porous media: Theoretical framework. *Advances in Water Resources*, *29*(8), 1238–1255.

Neuman, S. P. (1994). Generalized scaling of permeabilities: Validation and effect of support scale. *Geophysical Research Letters*, *21*(5), 349–352.

Neuman, S. P. (2014). The new potential for understanding groundwater contaminant transport. *Groundwater*, *52*(5), 653–656.

Neuman, S. P., & Depner, J. S. (1988). Use of variable-scale pressure test data to estimate the log hydraulic conductivity covariance and dispersivity of fractured granites near oracle, Arizona. *Journal of Hydrology*, *102*(1–4), 475–501.

Neuman, S. P., & Di Federico, V. (2003). Multifaceted nature of hydrogeologic scaling and its interpretation. *Reviews of Geophysics*, *41*(3), 1014. https://doi.org/10.1029/2003RG000130

Neuman, S. P., & Orr, S. (1993). Prediction of steady state flow in nonuniform geologic media by conditional moments: Exact nonlocal formalism, effective conductivities, and weak approximation. *Water Resources Research*, *29(2)*, 341–364.

Panzeri, M., Riva, M., Guadagnini, A., & Neuman, S. (2013). Data assimilation and parameter estimation via ensemble Kalman filter coupled with stochastic moment equations of transient groundwater flow. *Water Resources Research*, *49*, 1334–1344. https://doi.org/10.1002/wrcr.20113

Panzeri, M., Riva, M., Guadagnini, A., & Neuman, S. (2014). Comparison of ensemble Kalman filter groundwater-data assimilation methods based on stochastic moment equations and Monte Carlo simulation. *Advances in Water Resources*, *66*, 8–18.

Panzeri, M., Riva, M., Guadagnini, A., & Neuman, S. (2015). EnKF coupled with groundwater flow moment equations applied to Lauswiesen aquifer, Germany. *Journal of Hydrology*, *521*, 205–216.

Qian, Z., Seepersad, C. C., Joseph, V. R., Allen, J. K., & Wu, C. J. (2006). Building surrogate models based on detailed and approximate simulations. *Journal of Mechanical Design*, *128*(4), 668–677.

Riva, M., Guadagnini, A., Neuman, S. P., & Franzetti, S. (2001). Radial flow in a bounded randomly heterogeneous aquifer. *Transport in Porous Media*, *45*(1), 139–193.

Tartakovsky, A. M., Meakin, P., & Huang, H. (2004). Stochastic analysis of immiscible displacement of the fluids with arbitrary viscosities and its dependence on support scale of hydrological data. *Advances in Water Resources*, *27*(12), 1151–1166.

Tartakovsky, A. M., Neuman, S. P., & Lenhard, R. J. (2003). Immiscible front evolution in randomly heterogeneous porous media. *Physics of Fluids (1994-Present)*, *15*(11), 3331–3341.

Tartakovsky, D. M., & Broyda, S. (2011). PDF equations for advective–reactive transport in heterogeneous porous media with uncertain properties. *Journal of Contaminant Hydrology*, *120121*, 129–140. https://doi.org/10.1016/j.jconhyd.2010.08.009

Tartakovsky, D. M., Guadagnini, A., Ballio, F., & Tartakovsky, A. M. (2002). Localization of mean flow and apparent transmissivity tensor for bounded randomly heterogeneous aquifers. *Transport in Porous Media*, *49*, 41–58.

Tartakovsky, D. M., & Neuman, S. P. (1998). Transient flow in bounded randomly heterogeneous domains: 1. Exact conditional moment equations and recursive approximations. *Water Resources Research*, *34*(1), 1–12.

Vanmarcke, E. (2010). *Random fields: Analysis and synthesis*. Singapore: World Scientific.

Wang, P., Barajas-Solano, D. A., Constantinescu, E., Abhyankar, S., Ghosh, D., Smith, B. F., . . . Tartakovsky, A. M. (2015). Probabilistic density function method for stochastic ODEs of power systems with uncertain power input. *SIAM/ASA Journal of Uncertainty Quantification*, *3*(1), 873–896. https://doi.org/10.1137/130940050

Wang, P., Tartakovsky, A. M., & Tartakovsky, D. M. (2013). Probability density function method for Langevin equations with colored noise. *Physical Review Letters*, *110*, 140602. https://doi.org/10.1103/PhysRevLett.110.140602

Xiu, D. (2009). Fast numerical methods for stochastic computations: A review. *Communications in Computational Physics*, *5*(2–4), 242–272.

Ye, M., Neuman, S., Guadagnini, P. A., & Tartakovsky, D. M. (2004). Nonlocal and localized analyses of conditional mean transient flow in bounded, randomly heterogeneous porous media. *Water Resources Research*, *40*, W05104. https://doi.org/10.1029/2003WR002099