AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE
10.1002/2017WR021368

Key Points:
- Impact of parametric uncertainty on temporal evolution of risk indices is quantified
- Key flow and transport parameters affecting the risk indices are identified
- Extent of contamination and correspondent remediation costs are estimated

Correspondence to:
D. Tartakovsky,
tartakovsky@stanford.edu

Citation:
Ciriello, V., Lauriola, I., Bonvicini, S., Cozzani, V., Di Federico, V., & Tartakovsky, D. M. (2017). Impact of hydrogeological uncertainty on estimation of environmental risks posed by hydrocarbon transportation networks. *Water Resources Research*, *53*, 8686–8697. https://doi.org/10.1002/2017WR021368

Received 22 JUN 2017
Accepted 20 SEP 2017
Accepted article online 27 SEP 2017
Published online 3 NOV 2017

Impact of Hydrogeological Uncertainty on Estimation of Environmental Risks Posed by Hydrocarbon Transportation Networks

V. Ciriello[1], I. Lauriola[1], S. Bonvicini[1], V. Cozzani[1], V. Di Federico[1], and Daniel M. Tartakovsky[2]

[1]Department of Civil, Chemical, Environmental, and Materials Engineering, University of Bologna, Bologna, Italy, [2]Department of Energy Resources Engineering, Stanford University, Stanford, CA, USA

Abstract Ubiquitous hydrogeological uncertainty undermines the veracity of quantitative predictions of soil and groundwater contamination due to accidental hydrocarbon spills from onshore pipelines. Such predictions, therefore, must be accompanied by quantification of predictive uncertainty, especially when they are used for environmental risk assessment. We quantify the impact of parametric uncertainty on quantitative forecasting of temporal evolution of two key risk indices, volumes of unsaturated and saturated soil contaminated by a surface spill of light nonaqueous-phase liquids. This is accomplished by treating the relevant uncertain parameters as random variables and deploying two alternative probabilistic models to estimate their effect on predictive uncertainty. A physics-based model is solved with a stochastic collocation method and is supplemented by a global sensitivity analysis. A second model represents the quantities of interest as polynomials of random inputs and has a virtually negligible computational cost, which enables one to explore any number of risk-related contamination scenarios. For a typical oil-spill scenario, our method can be used to identify key flow and transport parameters affecting the risk indices, to elucidate texture-dependent behavior of different soils, and to evaluate, with a degree of confidence specified by the decision-maker, the extent of contamination and the correspondent remediation costs.

Plain Language Summary Ubiquitous hydrogeological uncertainty undermines the veracity of quantitative predictions of soil and groundwater contamination due to accidental hydrocarbon spills from onshore pipelines. We quantify the impact of parametric uncertainty on quantitative forecasting of temporal evolution of two key risk indices, volumes of unsaturated and saturated soil contaminated by a surface spill of light nonaqueous-phase liquids. For a typical oil-spill scenario, our method allows one to identify key flow and transport parameters affecting the risk indices, to elucidate texture-dependent behavior of different soils, and to evaluate, with a degree of confidence specified by the decision-maker, the extent of contamination and the correspondent remediation costs.

1. Introduction

Accidental releases of nonaqueous-phase liquids (NAPLs) into the environment pose significant risks of subsurface contamination, affecting human health, ecosystems, and water quality. Onshore pipelines provide the safest way to transport hydrocarbons and other NAPLs over long distances. Yet even though accidental spills from such pipelines are relatively rare (e.g., OECD, 1997), they might have severe environmental consequences due to large quantities of NAPL released. While risk of human exposure to such contamination is routinely assessed in the context of quantitative risk analysis (QRA) (e.g., CCPS, 1995; TNO, 1999), identification and computation of proper metrics or indices for environmental risk proved to be more elusive. For example, the environmental damage index (Bonvicini et al., 2015) is formulated in terms of the expected overall cost of remediation, $\mathcal{C}_{\text{rem}}$, as

$$\mathcal{C}_{\text{rem}} \equiv \mathcal{C}_{\text{dis}} + \mathcal{C}_{\text{cle}} + \mathcal{C}_{\text{gwt}} = V_{\text{dis}}\hat{\mathcal{C}}_{\text{dis}} + V_{\text{cle}}\hat{\mathcal{C}}_{\text{cle}} + V_{\text{gwt}}\hat{\mathcal{C}}_{\text{gwt}}, \tag{1}$$

where $\mathcal{C}_{\text{dis}}$, $\mathcal{C}_{\text{cle}}$, and $\mathcal{C}_{\text{gwt}}$ are the costs associated with soil disposal, soil clean-up, and groundwater treatment, respectively. Each of these costs is computed as the product of an average contaminated volume, V_i,

and the unitary cost of the corresponding remediation action, $\hat{C}_i$, where i=dis, cle, and gwt. The volumes V_i are averaged over a number of spill events, differentiated with respect to a cause of the release and the safety barriers present on site. Hence, for a given segment of a pipeline and N spills, the affected volume is estimated by

$$V_i=\sum_{k=1}^{N} p_k \hat{V}_{i,k}, \qquad i=\text{dis, cle, and gwt}, \tag{2}$$

where p_k is the probability of occurrence of the kth spill event and $V_{i,k}$ is the unsaturated/saturated volume of soil affected by that contamination event.

Estimation of the contaminated volumes $\hat{V}_{i,k}$ requires computationally intensive numerical solutions of multiphase flow equations, which are parameterized with soil hydraulic and transport parameters. Since the latter are always uncertain, the resulting estimates of $\hat{V}_{i,k}$ are uncertain as well. This uncertainty might significantly, and often unexpectedly, affect estimates of the remediation cost in equation (1), yet its impact remains largely unexplored in risk assessment analyses (Tartakovsky, 2007, 2013). We present a comprehensive framework for dealing with this problem and apply it to estimate probabilistically the soil and groundwater volumes contaminated by (repeated) spills of oil from an overland pipeline; governing equations used to model the latter phenomenon are presented in section 2. Our probabilistic framework comprises a stochastic collocation (SC) approach (section 3.2) to accelerate, relative to brute-force Monte Carlo simulations, sampling of the parameter space; a global sensitivity analysis (GSA) based on Sobol' indices (section 3.3) to identify the most relevant parameters, thus reducing the dimensionality of the parameter space; and construction of a polynomial surrogate model (section 3.2) to dramatically reduce the computational cost of estimating the spatiotemporal extent of soil and water contamination due to a surface spill.

2. Mathematical Models of Subsurface Contamination

Fate and downward migration of an LNAPL, from the Earth surface toward the water table, can be described by several alternative models with various degrees of fidelity. We start by formulating general multiphase flow equations and an advection-dispersion equation (section 2.1), which describe, respectively, subsurface transport of a contaminant's nonaqueous and dissolved phases with a minimal number of simplifications but are often computationally prohibitive in the probabilistic context. Instead we use a simplified version of this general formulation, the hydrocarbon soil screening model (HSSM) developed by the US EPA (Charbeneau et al., 1995; Weaver et al., 1994) (section 2.2), as our physics-based model. Quantitative comparisons of these two descriptors can be found in Jang et al. (2013). A surrogate polynomial representation, which carries virtually no computational cost, serves as a reduced-complexity model (section 3.2).

2.1. General Formulation

2.1.1. NAPL Migration in the Vadose Zone

Three distinct phases, water, air, and NAPL, are present in the vadose zone; their respective saturations at any "point" (a representative elementary volume) $\mathbf{x}$ and time t are denoted by $S_w(\mathbf{x},t)$, $S_a(\mathbf{x},t)$ and $S_N(\mathbf{x},t)$ such that $S_w+S_a+S_N=1$. At the continuum scale, $\mathbf{q}_i(\mathbf{x},t)$, flow velocity of the ith phase (i=a, w, N), is related to $\psi_i(\mathbf{x},t)$, pressure head in the ith phase, by the generalized Darcy law,

$$\mathbf{q}_i=-K_{s,i}K_{r,i}(S_i)\nabla(\psi_i+x_3), \quad i=\text{a, w, N}, \tag{3}$$

where $K_{s,i}(\mathbf{x})$ and $K_{r,i}(\mathbf{x},S_i)$ are the saturated and relative hydraulic conductivities, respectively, with the explicit dependence on $\mathbf{x}=(x_1,x_2,x_3)^\top$ accounting for soil heterogeneity (spatial variability); and x_3 is the vertical coordinate positive upward. Multiphase flow equations are derived by combining these relations with mass conservation,

$$\phi\frac{\partial \rho_i S_i}{\partial t}=-\nabla\cdot(\rho_i\mathbf{q}_i)+f_i, \quad i=\text{a, w, N}, \tag{4}$$

where $\phi(\mathbf{x})$ is the soil porosity, ρ_i is the density of the ith phase, and f_i is a generic source term (mass per unit volume). The multiphase flow equations are closed by specifying constitutive laws $\psi_i=\psi_i(S_i)$ and $K_{r,i}=K_{r,i}(S_i)$. The van Genuchten's model provides one example of such laws,

$$\Theta_w = \left[1+(\alpha\psi_w)^{1/(1-m)}\right]^{-m}, \quad \Theta_w \equiv \frac{S_w - S_{w,res}}{S_{w,sat} - S_{w,res}}, \tag{5a}$$

$$K_{r,w} = \sqrt{\Theta_w}\left[1-(1-\Theta_w^{1/m})^m\right]^2, \tag{5b}$$

where the soil parameter α, the exponents m, and the saturated ($S_{w,sat}$) and residual ($S_{w,res}$) water saturations are fitting parameters. Similar constitutive laws are used for the NAPL phase.

Once equations (3)–(5) are solved subject to appropriate initial and boundary conditions, the volume of the soil contaminated with NAPL is computed as $V_N(t) = \{\mathbf{x} : S_N(\mathbf{x}, t) \geq S_N^\star\}$ where $S_N^\star$ is the maximum allowable NAPL saturation.

2.1.2. Solute Migration Below Water Table

Dissolved component migrates in groundwater following an advection-dispersion-reaction equation,

$$\frac{\partial c}{\partial t} = \nabla \cdot (\mathbf{D}\nabla c) - \nabla \cdot (\mathbf{v}c) + R(c), \tag{6}$$

where $c(\mathbf{x}, t)$ is the solute concentration, $\mathbf{v}(\mathbf{x}, t)$ is the macroscopic flow velocity of groundwater flow, $\mathbf{D}(\mathbf{x}, |\mathbf{v}|)$ is the dispersion coefficient tensor, and $R(c; \kappa)$ is the reaction term parametrized with a reaction rate constant $\kappa(\mathbf{x})$. The longitudinal, D_l, and transverse, D_t, components of the dispersion coefficient $\mathbf{D}$ characterize the spread of the contaminant plume in the directions collinear and perpendicular to the velocity vector $\mathbf{v}$, respectively. For example, if the groundwater flow direction is aligned with the x_1 coordinate, then it is common to define these components as $D_l \equiv D_{11} = D_m + \lambda_l|\mathbf{v}|$ and $D_t \equiv D_{22} = D_{33} = D_m + \lambda_t|\mathbf{v}|$, where D_m is the coefficient of molecular diffusion of the solute in the soil, and $\alpha_l(\mathbf{x})$ and $\alpha_t(\mathbf{x})$ are the longitudinal and transverse dispersivities (typically, these fitting parameters satisfy the condition $\alpha_t \ll \alpha_l$).

Once equation (6) is solved subject to appropriate initial and boundary conditions, the volume of the aquifer contaminated with the dissolved contaminant is computed as $V_{gw}(t) = \{\mathbf{x} : c(\mathbf{x}, t) \geq c_N^\star\}$ where $c_N^\star$ is the maximum allowable contaminant concentration.

In the deterministic setting, i.e., when all the relevant soil properties and forcings are assumed to be known with certainty, equations (3)–(6) are often used to predict subsurface fate and transport of NAPLs (e.g., McLaren et al., 2012, and references therein).

2.2. Physics-based Model

We use the hydrocarbon soil screening model (HSSM) (Charbeneau et al., 1995; Weaver et al., 1994) as our physics-based reduced-complexity model of LNAPL migration in the subsurface. Developed by the US EPA, this code is widely used to model LNAPL contamination of soils and groundwater (e.g., Charbeneau & Weaver, 1992; Yoon et al., 2009). The code comprises three modules working in series: (i) one-dimensional vertical multiphase transport in the vadose zone (from the near surface to the capillary fringe); (ii) one-dimensional radial horizontal spreading of the contaminant lens through the capillary fringe; and (iii) two-dimensional, vertically averaged, transport of the dissolved component in the aquifer. Basic assumptions of the code consist in considering the liquid phases to be incompressible and the porous medium as homogeneous and nondeformable. Jang et al. (2013), among others, provide a detailed comparison of the predictions obtained with HSSM and the full model (equations (3)–(6)).

The first module (named KOPT for kinematic oil pollutant transport) simulates the flow of the LNAPL and the transport of its key chemical constituent through the vadose zone. Soil is characterized by a uniform water content since water saturation is computed from the average annual recharge rate. This assumption allows neglecting the continuity equation for water. The generalized method of characteristics is applied to solve a set of approximated hyperbolic governing equations obtained by neglecting the capillary pressure gradient. As a consequence, gravity acts as the only driving force and the leading edge of the LNAPL is represented as a sharp front. Lateral spreading due to capillary forces is not addressed.

The second module (named OILENS) simulates a radial spreading of the hydrocarbon on the water table. The unsteady motion is represented as a sequence of steady states based on the assumptions of incompressible flow and vertical equilibrium for the fluids at any location. A simplified description of the LNAPL lens (e.g., a semianalytical solution of Sudicky et al. (2013)) is then provided, together with the mass flux

toward the aquifer against time. This is used as a boundary condition for the third module (named TGSPLUM) representing the migration of the dissolved component in the aquifer based on a Gaussian-source plume model. Advection-dispersion processes are assumed to govern transport in two dimensions. Effect of dilution produced by natural recharge is modeled as a decay term. Furthermore, steady state is considered and velocity is assumed to be uniform in the flow direction.

These approximations simplify the computation of the contaminated volumes $V_N(t)$ and $V_{gw}(t)$. The assumption of one-dimensional (vertical) flow in the vadose zone implies that, for a given (and possibly uncertain) surface spill of area A_{spill}, the former is given by $V_N(t)=A_{spill}z_N(t)$, where $z_N(t)$ is the depth of soil within which the NAPL saturation is $S \geq S_N^\star$. The assumption of two-dimensional (horizontal) transport below the water table leads to $V_{gw}(t) \approx \Delta z A_{gw}(t)$, where Δz is the unit thickness used in the TGSPLUM module and $A_{gw}(t)$ is the aquifer's area within which the concentration $c(\mathbf{x}, t) \geq c_N^\star$.

Selection of an "optimal" model depends on the complex interplay between quantities of interest and availability of (never sufficient) site-specific data and computational resources; a complex model with a large number of uncertain parameters might yield less accurate predictions than its reduced-complexity counterpart that allows for exhaustive exploration of the parameter space (Sinsbeck & Tartakovsky, 2015). The latter study demonstrates that the HSSM model, which can be sampled extensively due to low computational cost of individual deterministic solves, is likely to outperform the full model (equations (3)–(6)), whose high computational cost typically makes it possible to compute no more than 10s realizations (Maji & Sudicky, 2008; McLaren et al., 2012), when it comes to computing cumulative distribution functions of the quantities of interest.

2.3. Model Parametrization

Virtually every parameter in equations (3)–(6) and, hence, in our physics-based model (HSSM) is space- and scale-dependent, reflecting the multiscale heterogeneity of subsurface environments. While some of these parameters, e.g., the fitting parameters in the constitutive relations (equation (5)), are often determined at the laboratory scale from a few soil samples, others, e.g., the dispersivities in equation (6), are inferred at the field scale by calibrating the model's predictions to an observed spatial extent of the plume. Given this disparity of scales and the ubiquitous scarcity of data, the standard practice is to assign a statistical model (a probability density function or PDF) to such parameters and to compute its statistical parameters (e.g., mean and variance) from measurements.

Statistical properties of the key parameters affecting contaminant transports in the vadose and saturated zones are presented in Table 1. Rather than treating these parameters as random fields, we adopt a zonation approach which uses soil types to subdivide a subsurface environment into regions whose properties are modeled as random variables. Such a conceptualization is often used in practice, where one has to contend with data scarcity or, in the absence of data, rely on regional soil maps. To be specific, in the simulations reported below we assume that the soil consists of a single uniform stratum of either sand or sandy-

Table 1
Statistical Distributions (PDFs) and Properties (Mean, $\langle\cdot\rangle$, and Standard Deviation, σ) of the Uncertain Model Parameters in Equations (3)–(6)

Parameter	PDF	Sand	Sandy loam	Loam
α (1/m)	Lognormal	$\langle\alpha\rangle=14.5$ $\sigma_\alpha=2.9$	$\langle\alpha\rangle=7.5$ $\sigma_\alpha=3.7$	$\langle\alpha\rangle=3.6$ $\sigma_\alpha=2.1$
$n=(1-m)^{-1}$	Lognormal	$\langle n\rangle=2.68$ $\sigma_n=0.29$	$\langle n\rangle=1.89$ $\sigma_n=0.17$	$\langle n\rangle=1.56$ $\sigma_n=0.11$
K_s (m/d)	Lognormal	$\langle K_s\rangle=7.1$ $\sigma_{K_s}=3.7$	$\langle K_s\rangle=1.1$ $\sigma_{K_s}=1.4$	$\langle K_s\rangle=0.25$ $\sigma_{K_s}=0.44$
ϕ	Normal	$\langle\phi\rangle=0.43$ $\sigma_\phi=0.06$	$\langle\phi\rangle=0.41$ $\sigma_\phi=0.09$	$\langle\phi\rangle=0.43$ $\sigma_\phi=0.10$
J	Normal	$\langle J\rangle=0.001$	$\sigma_J=0.0003$	
α_l (m)	Lognormal	$\langle\alpha_l\rangle=10.0$	$\sigma_{\alpha_l}=3.0$	

Note. All but the last two parameters depend on soil texture and are separated into three classes representing distinct soil types: sand, sandy loam, and loam. The data are from Carsel and Parrish (1988).

loam or loam, and assign to each the means and standard deviations identified by Carsel and Parrish (1988). (A layered soil structure can be readily accommodated as well by adopting the random domain decomposition (Lin et al., 2010; Winter & Tartakovsky, 2000).) The regional groundwater head gradient J is assigned a normal distribution with a range of variation consistent with typical values in the application area. Finally, the longitudinal dispersivity α_l is assumed to follow a lognormal distribution with a mean and standard deviation selected on the basis of an estimation of the characteristic domain length; the transverse dispersivity is set to $\alpha_t = 0.1\alpha_l$.

There is an ongoing debate about whether the soil properties reported in Table 1 are mutually correlated (see Carsel & Parrish, 1988; Tartakovsky et al., 1999, 2003 and references therein). An often used argument for their independence is that the saturated hydraulic conductivity K_s is controlled by macrovoids, while the reciprocal of the capillary length α depends on the entire continuum of pore sizes (see Tartakovsky et al., 1999, and references therein). Adopting these hypotheses, and given the weak correlation exhibited by the parameters (and their logarithms) in the most of the tests analyzed, we assume all the uncertain parameters in Table 1 to be mutually independent.

Since we focus on the effects of uncertain hydrogeological parameters on risk-based assessment of subsurface contamination, the hydrocarbon properties are assumed to be known with certainty, i.e., are treated as deterministic quantities. However, the probabilistic framework introduced in section 3 can also account for uncertainty in the latter parameters, as well as for the probability of pipeline rupture.

3. Probabilistic Framework for Subsurface Contamination Assessment

3.1. Random Parameter Space

We start by representing the six-dimensional parameter space in Table 1 in terms of $N_{par} = 6$ standard Gaussian variables, $\mathbf{P} = \{P_1, \ldots, P_6\}$, such that

$$P_1 = \frac{\ln \alpha - \langle \ln \alpha \rangle}{\sigma_{\ln \alpha}}, \quad P_2 = \frac{\ln n - \langle \ln n \rangle}{\sigma_{\ln n}}, \quad P_3 = \frac{\ln K_s - \langle \ln K_s \rangle}{\sigma_{\ln K_s}}, \tag{7a}$$

$$P_4 = \frac{\phi - \langle \phi \rangle}{\sigma_\phi}, \quad P_5 = \frac{J - \langle J \rangle}{\sigma_J}, \quad P_6 = \frac{\ln \alpha_l - \langle \ln \alpha_l \rangle}{\sigma_{\ln \alpha_l}}. \tag{7b}$$

The statistics of $\ln \alpha$ (or other logarithms) are related to that of α (or other log-normally distributed variables in Table 1) by $\ln \langle \alpha \rangle = \langle \ln \alpha \rangle + \sigma_{\ln \alpha}/2$ and $\sigma_\alpha^2 / \langle \alpha \rangle^2 = \exp(\sigma_{\ln \alpha}^2) - 1$.

Solving for any quantity of interest (QoI) Q is equivalent to identifying the functional relation $Q = Q(\mathbf{x}, t; \mathbf{P})$.

3.2. Polynomial Chaos and Stochastic Collocation Method

A truncated polynomial chaos expansions (PCE) Wiener (1938) approximates the QoI $Q(\mathbf{x}, t; \mathbf{P})$ with an N_{pol}-degree polynomial

$$Q(\mathbf{x}, t; \mathbf{P}) \approx \sum_{j=0}^{N_Q - 1} a_j(\mathbf{x}, t) \Psi_j(\mathbf{P}), \quad N_Q = \frac{(N_{par} + N_{pol})!}{N_{par}! N_{pol}!}, \tag{8}$$

where $\Psi_j(\mathbf{P})$ denotes multivariate Hermite polynomials. We use the SC method (Webster et al., 1996) to compute the deterministic expansion coefficients $a_j(\mathbf{x}, t)$. The details of our implementation of the SC can be found in Ciriello et al. (2013).

The leading coefficient in equation (8), $a_0(\mathbf{x}, t)$, corresponds to the mean of the QoI $Q(\mathbf{x}, t)$, i.e., $\langle Q(\mathbf{x}, t) \rangle = a_0(\mathbf{x}, t)$. The variance of $Q(\mathbf{x}, t)$ is computed from as

$$\sigma_Q^2(\mathbf{x}, t) = \sum_{j=1}^{N_Q - 1} a_j^2(\mathbf{x}, t) \langle \Psi_j^2(\mathbf{P}) \rangle, \tag{9}$$

which is derived from equation (8) by accounting for the orthogonality of the Hermite polynomials, $\langle \Psi_i \Psi_k \rangle = 0$ for all $i \neq k$; the Gaussian quadratures involved in the averages $\langle \Psi_j^2(\mathbf{P}) \rangle$ are computed analytically. In addition to these moments, the full PDF of Q can be readily computed from equation (8) once the deterministic coefficients $a_j(\mathbf{x}, t)$ are found via the SC.

While the SC might underperform Monte Carlo when the hydraulic properties are spatially varying random functions with relatively short correlation lengths and/or relatively high variances (Barajas-Solano & Tartakovsky, 2016), it is well suited for the low-dimensional probability spaces, such as $N_{par}=6$ considered in the present study (e.g., Ciriello et al., 2013, 2015, and references therein).

We deal with two quantities of interest (QoIs), $V^{\star}_{vz}=V_{vz}(t^{\star})$ and $V^{\star}_{gw}=V_{gw}(t^{\star})$, where $V^{\star}_{vz}$ and $V^{\star}_{gw}$ are the subsurface volumes defined in section 2.1, and $t^{\star}$ is the number of days from the contaminant release required to detect the occurrence of contamination. Second-degree polynomial representations of these QoIs, given by equation (8) with $N_{pol}=2$, have the form

$$V^{\star}_{vz}=\sum_{i=1}^{N_{par}-2} a_i P_i+\sum_{i=1}^{N_{par}-2} a_{ii}(P_i^2-1)+\sum_{i=1}^{N_{par}-2}\sum_{k>i}^{N_{par}-2} a_{ik}P_iP_k, \tag{10}$$

and

$$V^{\star}_{gw}=\sum_{i=1}^{N_{par}} b_i P_i+\sum_{i=1}^{N_{par}} b_{ii}(P_i^2-1)+\sum_{i=1}^{N_{par}}\sum_{k>i}^{N_{par}} b_{ik}P_iP_k. \tag{11}$$

As before, the coefficients in these polynomials are obtained with the SC. The so-called "surrogate models" (equations (10) and (11)) serve as low-fidelity probabilistic predictors of our QoIs $V^{\star}_{vz}$ and $V^{\star}_{gw}$, respectively. If necessary, one can improve the accuracy of these surrogate models by increasing the polynomial order.

3.3. Global Sensitivity Analysis

The Sobol' indices (Sobol', 1993) provide a metric of the relative impact of each of the uncertain parameters in Table 1 on the overall predictive uncertainty, as quantified by the QoI variance σ^2_Q in equation (9). Specifically, the contribution of the ith parameter, P_i, to the total variance σ^2_Q is quantified by a "principle sensitivity index" $\mathcal{S}_i$, which is defined as (Sudret, 2008)

$$\mathcal{S}_i \equiv \frac{\sigma^2_{Q,i}}{\sigma^2_Q}, \qquad \sigma^2_{Q,i} \equiv \sum_{\gamma\in\Gamma_i} a^2_\gamma \langle \Psi^2_\gamma(P_i)\rangle, \tag{12}$$

where $\Gamma_i=\{\gamma\in(1,\ldots,N_Q-1):\Psi_\gamma(P_i)\}$. This definition can be easily extended to evaluate the joined influence of a subset of model parameters, in case that principle sensitivity indices do not provide a full description of the QoI variance. The Sobol' indices computed for all the possible subsets of parameters sum up to one (Sobol', 1993).

The random variables P_i in equations (10) and (11), whose contribution to the total variances $\sigma^2_{V^{\star}_{vz}}$ and $\sigma^2_{V^{\star}_{gw}}$ falls below a prescribed value (e.g., 1%), are replaced with their respective means. This reduces the

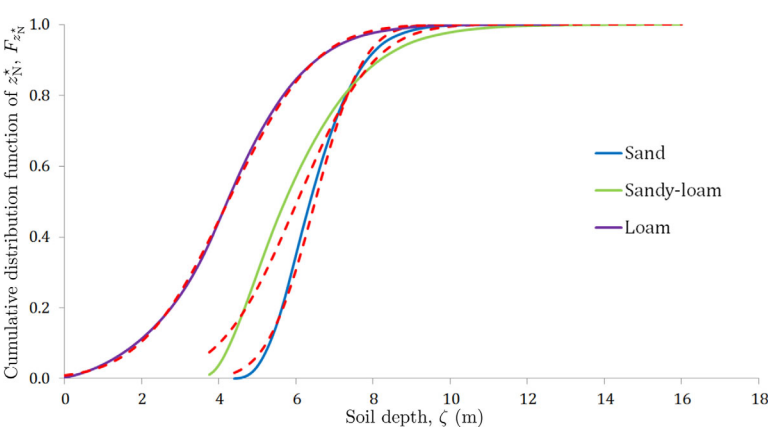

Figure 1. Cumulative distribution function, $F_{z^{\star}_N}(\zeta)$, of the depth of soil contamination after $t^{\star}=180$ days from the spill, $z^{\star}_N$, computed with the surrogate model (10) for three soil-texture types. The adjacent dashed lines depict the corresponding Gaussian distribution.

Table 2
Statistics of the Depth of NAPL Contamination, $z_N^\star$, for Three Soil Types: Mean $\mu_{z_N^\star}$, Median $m_{z_N^\star}$, Standard Deviation $\sigma_{z_N^\star}$, Contamination Depth $z_{99\%}^\star$ Occurring With $\mathbb{P}=0.99$, and the Probability $\mathbb{P}$ That Contamination Depth Does Not Exceed 5 m

Soil type	$\mu_{z_N^\star}$ (m)	$m_{z_N^\star}$ (m)	$\sigma_{z_N^\star}$ (m)	$z_{99\%}^\star$ (m)	$\mathbb{P}(z_N^\star \leq 5$ m)
Sand	6.48	6.33	0.98	4.76	0.04
Sandy loam	6.00	5.68	1.57	3.81	0.30
Loam	4.03	4.24	2.10	0.30	0.68

dimensionality of the polynomials in equations (10) and (11) and further decreases the computational cost of the surrogate models.

4. Case Study

Following Bonvicini et al. (2015), we consider a benzene spill from a pipeline, of length 16 km and diameter 6 inch that is equipped with automatic shut-down valves. A volume of $V_{spill}=291.7$ m^3 of benzene is released, forming a circular pool of radius $R_{spill}=17.6$ m and height $h_{spill}=0.3$ m; the height of the pool is kept constant for one minute and then gradually decreases to zero as the substance infiltrates into the soil. Benzene is classified as a carcinogenic and mutagenic substance, whose maximum allowable concentration in groundwater is set, e.g., in Italy, to $c_N^\star=1.0$ μg/L (Legislative Directive, 2010).

We consider two contamination scenarios. The first (section 4.1) deals with an aquifer that is sufficiently deep for the benzene plume to remain in the vadose zone during a given timeframe (in the simulations reported below we set $t^\star=180$ days). The second (section 4.2) represents a shallow aquifer, whose water table is located 5.0 m below the ground surface. In both cases, our aim is to relate, probabilistically, the spatial extent of subsurface contamination to soil types and other parameters whose statistical properties are reported in Table 1.

4.1. Scenario 1: Deep Water Table

For a given area of surface spill A_{spill}, the physics-based model of section 2.2 determines the volume of contaminated soil after $t^\star=180$ days in terms of $z_N^\star=z_N(t^\star)$. The latter is described probabilistically in terms of its cumulative distribution function (CDF), $F_{z_N^\star}(\zeta) \equiv \mathbb{P}(z_N^\star \leq \zeta)$, which is constructed with a kernel density estimator (KDE) from the N_{SC} solutions (collocation points), $z_{N,i}^\star$ for $i=1,\ldots,N_{SC}$, of the surrogate model (equation (10)). Specifically, we use a Gaussian kernel with a filter size $h=0.0255$, so that a KDE of $F_{z_N^\star}(\zeta)$ is given by

$$F_{z_N^\star}(\zeta) \approx \frac{1}{N_{NC}\sqrt{2\pi h^2}} \sum_{i=1}^{N_{SC}} \exp\left[-\frac{(\zeta - z_{N,i}^\star)^2}{2h^2}\right]. \tag{13}$$

Figure 1 exhibits the resulting CDFs, alongside the corresponding Gaussian CDFs, for the three soil types whose statistical properties are given in Table 1. As expected, the average (ensemble mean) depth of contamination increases with the grain size since they correspond to higher values of hydraulic conductivity and, hence, higher rates of NAPL migration. Perhaps less expected, is the concomitant decrease in predictive uncertainty, which is quantified by the standard deviation of $z_N^\star$ (the CDF width). This reflects higher variability of the hydrogeological properties of finer-grained soils (Table 1).

Figure 1 also reveals that, depending on soil type and composition, the CDF of $z_N^\star$ can significantly deviate from the Gaussian distribution. Consequently, the mean ($\mu_{z_N^\star}$) and standard deviation ($\sigma_{z_N^\star}$) of $z_N^\star$ are insufficient for probabilistic assessment of soil contamination. The latter calls for knowledge of the full CDF or such statistics as the median contamination depth (the depth that occurs with probability $\mathbb{P}=0.5$) and the depth predicted with a given degree of certainty (e.g., probability $\mathbb{P}=0.99$). Such statistics are obtained by inverting the CDFs in Figure 1; a few examples are presented in Table 2. Conservative

Table 3
Sobol' Sensitivity Indices $\mathcal{S}_i$ for Porosity ($i=\phi$), Saturated Hydraulic Conductivity ($i=K_s$), and Parameters α and $n=1/(1-m)$ in the van Genuchten Constitutive Laws (Equation (5))

Soil type	$\mathcal{S}_\phi$	$\mathcal{S}_{K_s}$	$\mathcal{S}_\alpha$	$\mathcal{S}_n$
Sand	9.67×10^{-1}	2.92×10^{-2}	3.44×10^{-7}	3.63×10^{-3}
Sandy loam	9.70×10^{-1}	3.14×10^{-2}	1.17×10^{-4}	7.07×10^{-4}
Loam	3.29×10^{-1}	6.64×10^{-1}	1.66×10^{-4}	9.59×10^{-3}

Note. The indices are defined by equation (12) with respect to predictions of the soil contamination depth $z_N^\star$, for three soil types.

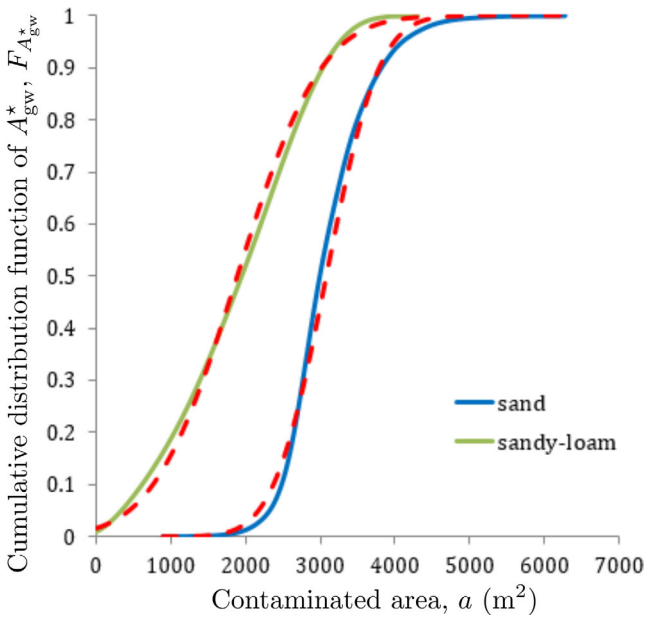

Figure 2. Cumulative distribution function, $F_{A^\star_{gw}}(a)$, of the plume size after $t^\star = 180$ days from the spill, $A^\star_{gw}$, computed with the surrogate model (11) for two soil-texture types. The adjacent dashed lines depict the corresponding Gaussian distribution.

estimates ($\mathbb{P}=0.99$) of soil contamination can be up to an order of magnitude smaller than their average counterparts.

Since predictive uncertainty and, ultimately, quantification of the risk posed by a surface NAPL spill stem from uncertainty in multiple soil properties, the relative impact of these uncertain parameters is important to understand and mitigate by a targeted data collection. Table 3 demonstrates the contribution of each uncertain hydraulic parameter, as described by their Sobol' indices in equation (12), to the overall uncertainty in predictions of the contamination depth $z^\star_N$. The impact of uncertainty in the values of porosity ϕ and saturated hydraulic conductivity K_s is orders of magnitude higher than that of parameters α and n in the van Genuchten constitutive laws (equation (5a)). That is reassuring, since both ϕ and K_s are significantly easier to measure than α and n and consequently their data are more widely available and correlated with soil types. This finding also suggests the possibility of replacing the uncertain (random) parameters α and n (and, for coarse soils, K_s) with their average values, thus appreciably reducing the random dimension of the problem and, hence, the computational cost. (It is worthwhile emphasizing that the observed relative impact of uncertainty in the various soil properties on the predictive uncertainty is predicated on our model selection; replacing the HSSM with the full model (equations (3)–(6)) might lead to different conclusions.)

4.2. Scenario 2: Shallow Water Table

Consider the case of a shallow phreatic aquifer whose water table is located 5 m below the earth surface. We focus on subsurface environments composed of either sand or sandy loam, in which the probability of NAPL reaching the water table is high: $\mathbb{P} \approx 0.96$ or 0.70, respectively (Table 2). Figure 2 shows CDF of the groundwater area $A^\star_{gw}$, within which the concentration $c(\mathbf{x}, t^\star)$ exceeds $c^\star_N = 1.0$ μg/L after $t^\star = 180$ days from the spill occurrence. The CDF is constructed with the KDE analogous to equation (13) from the N_{SC} solutions (collocation points), $A^\star_{gw,i}$ for $i = 1, \ldots, N_{SC}$, of the surrogate model (equation (11)). As expected, at any probability level, the contaminated area of the sandy aquifer is significantly larger than that of its counterpart comprised of sandy loam (the CDF for the former is to the right of the latter). Uncertainty in predictions of the plume size is appreciably higher in sandy-loam than in sand; this is evidenced by the widths of their corresponding CDFs and their standard deviations ($\sigma_{A^\star_{gw}}$) reported in Table 4.

Figure 2 reveals that the CDFs of the plume size are reasonably close to Gaussian CDFs for both sand and sandy loam. Discrepancy between the computed CDFs and its Gaussian counterparts is illustrated by the difference between the mean ($\mu_{A^\star_{gw}}$) and median ($m_{A^\star_{gw}}$) of the plume size (Table 4). The two statistics are identical for Gaussian distributions, but differ by 2% and 18% for sand and sandy-loam, respectively.

Similar to the deep water-table case, the Sobol' indices indicate that uncertainty in the values of porosity ϕ and saturated hydraulic conductivity K_s dominates the overall predictive uncertainty, with uncertainty in the direction of hydraulic gradient playing an important role in the sandy aquifer (Table 5). The latter would play a larger role in the aquifer consisting of sandy loam if the simulation horizon $t^\star$ were longer because both the NAPL in the vadose zone and its dissolved phase in the aquifer migrate slower in sandy loam. Second-order effects are detected in case of a sandy soil as follows: $\mathcal{S}_{K_s,\alpha_l} = 2.74\times10^{-2}$, and $\mathcal{S}_{K_s,J} = 4.02\times10^{-2}$. Once again, uncertainty in hard-to-measure parameters, such as the parameters in the van Genuchten constitutive laws (α and n), is relatively unimportant, because their Sobol' indices are orders of magnitude smaller than those of ϕ and K_s.

Table 4
Statistics of the Plume Size in Groundwater, $A^\star_{gw}$, for Two Soil Types: Mean $\mu_{A^\star_{gw}}$, Median $m_{A^\star_{gw}}$, Standard Deviation $\sigma_{A^\star_{gw}}$, and Plume Size $A^\star_{99\%}$ Occurring With $\mathbb{P}=0.99$

Soil type	$\mu_{A^\star_{gw}}$ (m^2)	$m_{A^\star_{gw}}$ (m^2)	$\sigma_{A^\star_{gw}}$ (m^2)	$A^\star_{99\%}$
Sand	3.08×10^3	3.01×10^3	5.47×10^2	1.95×10^3
Sandy loam	1.59×10^3	1.95×10^3	1.24×10^3	1.30×10^1

4.3. Model Verification

The preceding probabilistic analysis of the depth of contaminated soil, $z^\star_N$, and area of contaminated aquifer, $A^\star_{gw}$, relies on the surrogate models (equations (10) and (11)), respectively. To ascertain the

Table 5
Sobol' Sensitivity Indices $\mathcal{S}_i$ for the Six Uncertain Parameters Whose Statistics Are Collated in Table 1

Soil type	$\mathcal{S}_\phi$	$\mathcal{S}_{K_s}$	$\mathcal{S}_\alpha$	$\mathcal{S}_n$	$\mathcal{S}_{z_l}$	$\mathcal{S}_J$
Sand	0.16	0.54	4.58×10^{-4}	8.58×10^{-5}	8.42×10^{-2}	0.12
Sandy loam	0.45	0.52	7.76×10^{-6}	3.05×10^{-5}	1.52×10^{-3}	1.26×10^{-3}

Note. The indices are defined by equation (12) with respect to predictions of the aquifer contamination area $A^\star_{gw}$, for two soil types.

predictive accuracy of these polynomial representations, we generate 40 realizations of the QoIs ($z^\star_{N,k}$ and $A^\star_{gw,k}$ with $k=1,\ldots,40$ per soil type) by randomly selecting 40 realizations of the parameters in equations (10) and (11), $\mathbf{P}_k$ with $k=1,\ldots,40$. Each of these parameter sets is then used to compute a solution of the physics-based model (section 2.2) and postprocess the results to construct high-fidelity estimates of $z^\star_{N,k}$ and $A^\star_{gw,k}$. Figure 3 provides a comparison of the 40 estimates obtained with these two alternative models for several soil types. A perfect agreement between the two models corresponds to points (realizations) that fall on the 45° (dashed) line; the (solid) regression lines are close to the 45° lines, with the coefficient of determination $R^2=0.976$ and 0.980 for the predictions of $z^\star_N$ (left) and $A^\star_{gw}$, respectively. This demonstrates the accuracy of the surrogate models for both QoIs.

The accuracy improves as the soil becomes coarser, yielding a virtually perfect agreement with the physics-based models for sand. This is because the variances of the input parameters increase as the grain size decreases (Table 1), affecting the accuracy of the PCEs (equations (10) and (11)). The latter can be improved by increasing the polynomial order and, hence, the computational cost of SC. For the purposes of screening analysis and risk assessment, the agreement shown in Figure 3 is deemed to be sufficient.

4.4. Probabilistic Environmental Risk Indices

The results presented in sections 4.1 and 4.2 demonstrate how uncertainty in the soil hydraulic properties propagates through the modeling process, giving rise to uncertainty in predictions of volumes of the

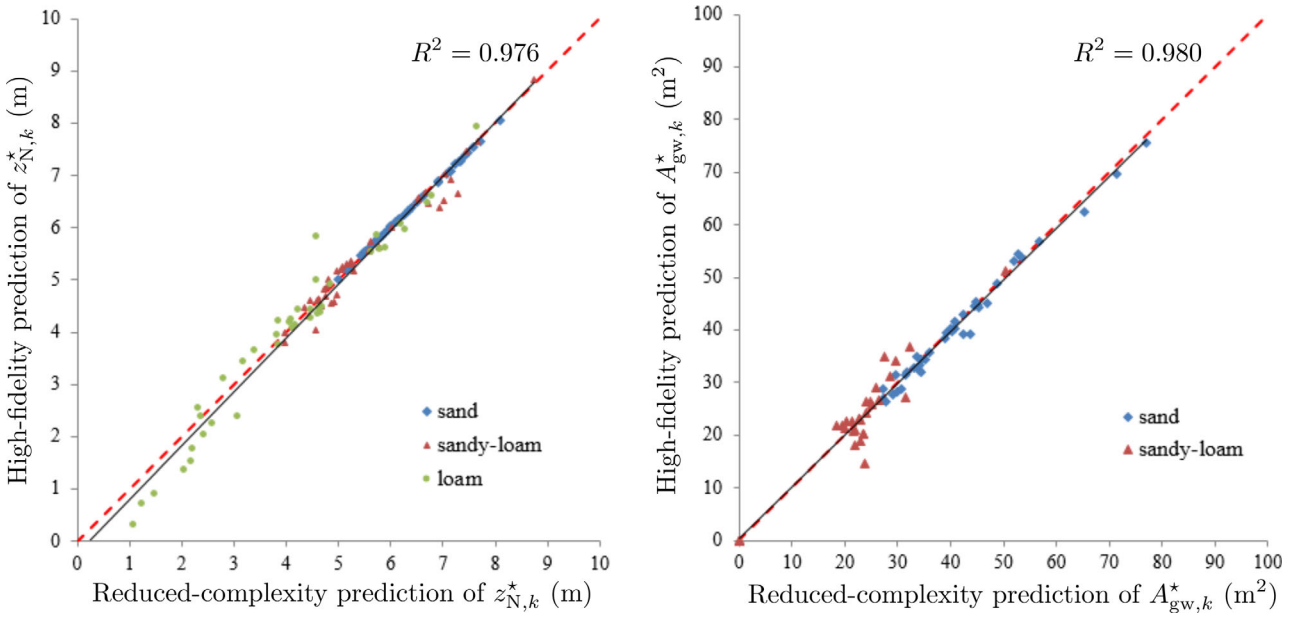

Figure 3. Realizations ($k=1,\ldots,40$) of the quantities of interest $z^\star_{N,k}$ (left) and $A^\star_{gw,k}$ (right) computed with the reduced-complexity surrogate model (the horizontal axis) and the physics-based model (the vertical axis), for several soil types. Data falling on the 45° (dashed) lines indicate the perfect agreement between the two models. The closeness of these lines to the regression (solid) lines indicates a close agreement between the models, with the coefficient of determination $R^2=0.976$ and 0.980 for the predictions of $z^\star_N$ (left) and $A^\star_{gw}$, respectively.

Table 6
Probabilistic Estimates, at $\mathbb{P}=0.5$ and 0.99 Probability Levels, of the Volumes of the Vadose Zone ($V^{\star}_{vz}$) and Groundwater ($V^{\star}_{gw}$) Contaminated by a Surface Spill of NAPL After $t^{\star}=180$ Days

Contaminated volume (m^3)	Sand	Sandy loam	Loam
Scenario 1: $V^{\star}_{vz,50\%}$	6.16×10^3	5.53×10^3	4.13×10^3
$V^{\star}_{vz,99\%}$	4.63×10^3	3.71×10^3	2.92×10^2
$V^{\star}_{gw,50\%}$	0	0	0
$V^{\star}_{gw,99\%}$	0	0	0
Scenario 2: $V^{\star}_{vz,50\%}$	6.16×10^3	5.53×10^3	4.13×10^3
$V^{\star}_{vz,99\%}$	4.63×10^3	3.71×10^3	2.92×10^2
$V^{\star}_{gw,50\%}$	3.03×10^3	2.50×10^3	
$V^{\star}_{gw,99\%}$	1.94×10^3	1.17×10^2	

contaminated soils and aquifers. Hence, computation of the environmental risk indices (Bonvicini et al., 2015) must account for uncertainty in estimation of the subsurface volumes $V_{i,k}$ in equation (2). For a single surface spill ($k=1$) with area $A_{spill}=\pi 17.6^2$ m^2 in scenario 1 (section 4.1), the deep aquifer is unaffected by contamination after $t^{\star}=180$ days, and estimates of the volumes of contaminated soil in the vadose zone, $V^{\star}_{vz}=A_{spill}z^{\star}_N$, are reported in Table 6 for probability levels $\mathbb{P}=0.5$ (a median estimate, $m_{V^{\star}_{vz}}\equiv V^{\star}_{vz,50\%}$) and $\mathbb{P}=0.99$ (an estimate $V^{\star}_{vz,99\%}$). Since this scenario deals with a deep aquifer, no contamination of groundwater takes place.

The same NAPL spill in scenario 2 (section 4.2) can contaminate both the vadose zone and shallow aquifer. The probability of occurrence of these two events, $V^{\star}_{vz}$ and $V^{\star}_{gw}$, is characterized by their joint CFD $F_{V^{\star}_{vz}V^{\star}_{gw}}(v_1,v_2)\equiv\mathbb{P}(V^{\star}_{vz}\le v_1,V^{\star}_{gw}\le v_2)=1-\mathbb{P}(V^{\star}_{vz}>v_1,V^{\star}_{gw}\le v_2)$. The latter is expressed in terms of conditional probability $\mathbb{P}(V^{\star}_{vz}>v_1,V^{\star}_{gw}\le v_2)=\mathbb{P}(V^{\star}_{gw}\le v_2|V^{\star}_{vz}>v_1)\mathbb{P}(V^{\star}_{vz}>v_1)$. Since $V^{\star}_{vz}=A_{spill}z^{\star}_N$ and A_{spill} is deterministic, $\mathbb{P}(V^{\star}_{vz}>v_1)=\mathbb{P}(z^{\star}_N>z)$. For the vadose zone of thickness $z=5$ m and composed of sand or sandy loam, $\mathbb{P}(z^{\star}_N>5\ \text{m})=0.96$ or 0.70, respectively (see Table 2). Once the NAPL reached the water table, the probability of the groundwater plume having the size smaller than v_1 is given by $\mathbb{P}(V^{\star}_{gw}\le v_2|V^{\star}_{vz}>v_1)$. Given the modeling assumptions, $\mathbb{P}(V^{\star}_{gw}\le v_2|V^{\star}_{vz}>v_1)=\mathbb{P}(A^{\star}_{gw}\le a|z^{\star}_N>5\ \text{m})$. The latter is the CDF plotted in Figure 2 and has the statistics presented in Table 4. With these preliminaries, $V^{\star}_{vz,r\%}=A_{spill}z^{\star}_{N,r\%}$ is presented in Table 6 for $r=50$ and 99; and the corresponding values of $V^{\star}_{gw,r\%}=bA^{\star}_{gw,r\%}$ are obtained (for the aquifer of unit thickness, $b=1.0$ m) by solving for a the equation $\mathbb{P}(A^{\star}_{gw}\le a|z^{\star}_N>5\ \text{m})\mathbb{P}(z^{\star}_N>5\ \text{m})=(1-r)/100$. This is done numerically by defining $A^{\star}_{gw,r\%}$ as the abscissa of the graph in Figure 2 at which $F_{A^{\star}_{gw}}=(1-r)/[100\mathbb{P}(z^{\star}_N>5\ \text{m})]$. The results are presented in Table 6.

Finally, we consider three remediation techniques—excavation of the top 1.2 m of soil accompanied by subsequent landfill disposal, soil clean-up by means of vapor extraction, and groundwater clean-up with air sparging—whose unitary costs are $\hat{\mathcal{C}}_{dis}=150$ €/m^3, $\hat{\mathcal{C}}_{cle}=100$ €/m^3, and $\hat{\mathcal{C}}_{gwt}=65$ €/m^3, respectively (Bonvicini et al., 2015). Figure 4 depict 50 and 99% probability estimates of the overall costs of subsurface remediation, computed with equations (1) and (2). Uncertainty in the hydrogeological parameters strongly affects predictions of the remediation costs even within a single texture class and especially for fine-grained soils.

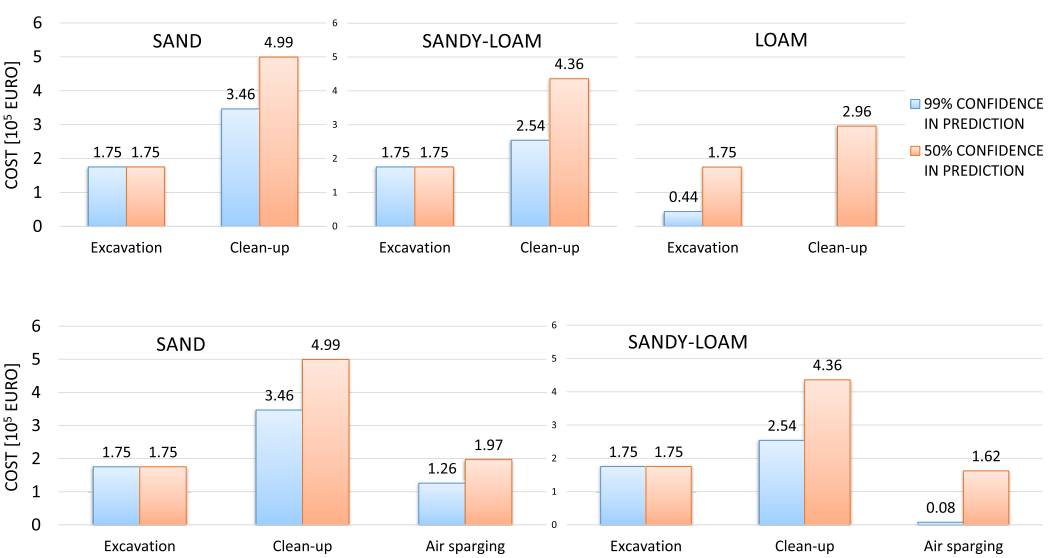

Figure 4. The 50 and 99% probability estimates of the overall costs of subsurface remediation in scenarios 1 (top) and 2 (bottom).

5. Conclusions

We quantified the uncertainty affecting predictions of environmental impact due to an accidental oil spill from an onshore pipeline in case of three different texture classes: sand, sandy loam, and loam. The environmental impact is defined through the computation of volumes of unsaturated and saturated soil affected by the contamination within a given timeframe. Volumes are computed by means of PCE-based metamodels defined over a selected full model solving multiphase flow problems. We found that uncertainty in model predictions increases when grain-size decreases. This is mainly due to the variability of key hydrogeological parameters that increases in case of fine-grained soils. By developing a GSA, we quantified the influence of parameter variability on model predictions. This analysis revealed the main role played by the porosity and saturated hydraulic conductivity when multiphase flow affects the unsaturated zone. In particular, variability in the porosity explains almost completely the uncertainty associated with the contaminated unsaturated volume in case of sand and sandy loam, while the saturated hydraulic conductivity becomes more relevant when a loamy soil is considered. If the contamination reaches the saturated zone, volumes of groundwater affected by the plume are mainly influenced by the hydraulic conductivity, especially in case of sand; the porosity, hydraulic gradient and dispersivity also play a significant role with similar magnitude. On the contrary, when sandy loam is considered as texture class, the influence of the hydraulic gradient and dispersivity decreases by about two orders of magnitude, while porosity assumes a role comparable with that of the hydraulic conductivity. In case of loamy soils, the probability of groundwater to be affected by the contamination is significant only if the water table depth is very high. In general, the impact of uncertainty in the unsaturated soil parameters is almost negligible in each scenario we considered. Note that these results depend on the interpretation provided by the selected full model of the physical and chemical processes occurring in the subsurface. Uncertainty in predictions of the volumes significantly affects predictions of the remediation costs. This is relevant when computing the environmental risk indices of Bonvicini et al., (2015). Our results indicate that reducing uncertainty in the parameters toward which model responses are most sensitive increases the accuracy of predictions and facilitates design of appropriate remediation actions.

Acknowledgments
This work was supported in part by the National Science Foundation under grant DMS-1620103. There are no data sharing issues since all of the numerical information is provided in the figures produced by solving the equations in the paper.

References

Barajas-Solano, D. A., & Tartakovsky, D. M. (2016). Stochastic collocation methods for nonlinear parabolic equations with random coefficients. *SIAM/ASA Journal of Uncertainty Quantification*, *4*(1), 475–494.

Bonvicini, S., Antonioni, G., Morra, P., & Cozzani, V. (2015). Quantitative assessment of environmental risk due to accidental spills from onshore pipelines. *Process Safety and Environmental Protection*, *93*, 31–49.

Carsel, R. F., & Parrish, R. S. (1988). Developing joint probability distributions of soil water retention characteristics. *Water Resources Research*, *24*(5), 755–769.

CCPS. (1995). *Guidelines for chemical transportation risk analysis* (technical report). New York, NY: Centre for Chemical Process Safety, American Institution of Chemical Engineers.

Charbeneau, R. J., & Weaver, J. W. (1992). Modeling contaminant transport through subsurface systems. *Journal of Hazardous Materials*, *32*, 293–311.

Charbeneau, R. J., Weaver, J. W., & Lien, B. K. (1995). *The hydrocarbon spill screening model (HSSM) volume 2: Theoretical background and source codes* (Tech. Rep. EPA/600/R-94/039b). Washington, DC: U.S. Environmental Protection Agency.

Ciriello, V., Di Federico, V., Riva, M., Cadini, F., De Sanctis, J., Zio, E., & Guadagnini, A. (2013). Polynomial chaos expansion for global sensitivity analysis applied to a model of radionuclide migration in a randomly heterogeneous aquifer. *Stochastic Environmental Research and Risk Assessment*, *27*, 945–954.

Ciriello, V., Edery, Y., Guadagnini, A., & Berkowitz, B. (2015). Multimodel framework for characterization of transport in porous media. *Water Resources Research*, *51*, 3384–3402. https://doi.org/10.1002/2015WR017047

Jang, W., Anderson, B. A., Suárez-Soto, R. J., Aral, M. M., & Maslia, M. L. (2013). Source characterization and simulation of the migration of light nonaqueous phase liquids (LNAPLs) in the vicinity of the Hadnot Point industrial area. In M. L. Maslia et al. (Eds.), *Analyses and historical reconstruction of groundwater flow, contaminant fate and transport, and distribution of drinking water within the service areas of the Hadnot point and Holcomb Boulevard water treatment plants and vicinities, U.S. Marine Corps Base Camp Lejeune, North Carolina—Chapter A: Summary and findings* (Suppl. 7). Atlanta, GA: Agency for Toxic Substances and Disease Registry.

Legislative Directive (2010). *D. Lgs. 152/2006 Decreto Legislativo n. 152 del 03/04/2006 (Norme inmateria ambientale) come modificato ed integrato dal D. Lgsn. 128 del 29/06/2010 (Modifiche ed integrazioni al decretolegislativo n. 152 del 03/04/2006, recante norme in materiaambientale (in Italian)*. Rome, Italy: Author.

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

Maji, R., & Sudicky, E. A. (2008). Influence of mass transfer characteristics for DNAPL source depletion and contaminant flux in a highly characterized glaciofluvial aquifer. *Journal of Contaminant Hydrology*, *102*(1), 105–119.

McLaren, R., Sudicky, E. A., Park, Y.-J., & Illman, W. A. (2012). Numerical simulation of DNAPL emissions and remediation in a fractured dolomitic aquifer. *Journal of Contaminant Hydrology*, *136–137*, 56–71.

OECD. (1997). *Report of the OECD workshop on pipelines (prevention of, preparedness for, and response to releases of hazardous substances), series on chemical accidents n.2, Paris* (technical report). Paris, France: OECD.

Sinsbeck, M., & Tartakovsky, D. M. (2015). Impact of data assimilation on cost-accuracy tradeoff in multifidelity models. *SIAM/ASA Journal of Uncertainty Quantification*, *3*(1), 954–968.

Sobol', I. M. (1993). Sensitivity estimates for nonlinear mathematical models. *Mathematical Modeling and Computing*, *1*, 407–414.
Sudicky, E. A., Hwang, H.-T., Illman, W. A., Wu, Y.-S., Kool, J. B., & Huyakorn, P. (2013). A semi-analytical solution for simulating contaminant transport subject to chain-decay reactions. *Journal of Contaminant Hydrology*, *144*(1), 20–45.
Sudret, B. (2008). Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering and System Safety*, *93*, 964–979.
Tartakovsky, D. M. (2007). Probabilistic risk analysis in subsurface hydrology. *Geophysical Research Letters*, *34*, L05404. https://doi.ogr/10.1029/2007GL029245
Tartakovsky, D. M. (2013). Assessment and management of risk in subsurface hydrology: A review and perspective. *Advances in Water Resources*, *51*, 247–260.
Tartakovsky, D. M., Neuman, S. P., & Lu, Z. (1999). Conditional stochastic averaging of steady state unsaturated flow by means of Kirchhoff transformation. *Water Resources Research*, *35*(3), 731–745.
Tartakovsky, D. M., Guadagnini, A., & Riva, M. (2003). Stochastic averaging of nonlinear flows in heterogeneous porous media. *Journal of Fluid Mechanics*, *492*, 47–62.
TNO. (1999). *Committee for the prevention of disasters, guidelines for quantitative risk assessment cpr18e* (technical report). The Hague, Netherlands: Sdu Uitgevers.
Weaver, J. W., Charbeneau, R. J., Tauxe, J. D., Lien, B. K., & Provost, J. B. (1994). *The hydrocarbon spill screening model (HSSM): User's guide* (Tech. Rep. EPA/600/R-94/039a, Vol. 1). Washington, DC: U.S. Environmental Protection Agency.
Webster, M., Tatang, M. A., & McRae, G. J. (1996). Application of the probabilistic collocation method for an uncertainty analysis of a simple ocean model. In *MIT joint program on the science and policy of global change reports series no. 4* (technical report). Cambridge, MA: MIT.
Wiener, N. (1938). The homogeneous chaos. *American Journal of Mathematics*, *60*, 897–936.
Winter, C. L., & Tartakovsky, D. M. (2000). Mean flow in composite porous media. *Geophysical Research Letters*, *27*(12), 1759–1762.
Yoon, H., Werth, C. J., Barkan, C. P. L., Schaeffer, D. J., & Anand, P. (2009). An environmental screening model to assess the consequences to soil and groundwater from railroad-tank-car spills of light non-aqueous phase liquids. *Journal Hazardous Materials*, *165*, 332–344.