Probabilistic assessment of seawater intrusion under multiple sources of uncertainty

M. Riva [a,b,*], A. Guadagnini [a,b], A. Dell'Oca [a]

[a] *Dipartimento di Ingegneria Civile e Ambientale, Politecnico di Milano, Piazza L. Da Vinci 32, 20133 Milano, Italy*
[b] *Department of Hydrology and Water Resources, University of Arizona, Tucson, AZ 85721, USA*

Received 25 July 2014
Received in revised form 3 November 2014
Accepted 5 November 2014
Available online 13 November 2014

1. Introduction

Saltwater intrusion (SWI) is a critical and widespread contamination problem in coastal aquifers. The complex interactions between fresh and salt water, with particular emphasis on management issues, has been the subject of active and intense research, including, e.g., a recent series of works highlighted in a special issue of Hydrogeology Journal (Special issue on: Saltwater and freshwater interactions in coastal aquifers, 2010, Vol 18, No 1). Analytical or semi-analytical solutions of SWI problems have been mainly developed for homogeneous aquifers and consider saltwater and fresh water as immiscible fluids separated by a sharp interface (e.g. [9,6,7,32]). Within this context, Dagan and Zeitoun [13] illustrate a first attempt to analyze the effect of aquifer heterogeneity on SWI. These authors consider a vertical cross section of a confined aquifer with randomly layered permeability distribution and show that the variability of the position of the salt–fresh water interface (particularly the location of the toe) is markedly influenced by the permeability variance and integral scale. Al-Bitar and Ababou [2] adopt a vertically-integrated sharp interface approach and analyze the effects of variability of aquifer properties on the saltwater wedge through numerical simulations within horizontal two-dimensional randomly heterogeneous unconfined aquifer. Chang and Yeh [10] employ a spectral approach and determine the mean interface position and its associated variability due to heterogeneity of aquifer conductivity and to the spatial variability of recharge for an unconfined horizontal aquifer model.

A realistic approach dealing with SWI should explicitly account for the occurrence of a transition zone where variable density flow is coupled with a transport model. This coupling makes it difficult to obtain analytical solutions of SWI scenarios. Henry [21] presents a semi-analytical solution for steady-state variable density flow taking place along a two-dimensional vertical cross-section in a homogeneous isotropic coastal aquifer. Since this is the only analytical solution available, it has been widely used as a benchmark problem to SWI numerical approaches (e.g. [35,41]). Dentz et al. [14] present a methodology conducive to an analytical solution of the Henry's problem in dimensionless form. The Henry's problem has limited use in practical applications because it considers only diffusion while dispersion is not simulated. Abarca et al. [1] modified the Henry's problem upon introducing anisotropy in the

* Corresponding author. Tel.: +39 02 2399 6214.
E-mail address: monica.riva@polimi.it (M. Riva).

conductivity tensor and a dispersion tensor to improve the representation of wide transition zones of the kind observed in several field sites. Held et al. [20] investigated the Henry's problem within a randomly heterogeneous aquifer. Making use of the homogenization theory, these authors found that the effective conductivity and dispersion coefficients are not affected by density effects, the effective dispersivity being close to its local counterpart. Otherwise, Kerrou and Renard [23] showed that macrodispersion coefficients differ from their local counterparts in two- and three-dimensional heterogeneous scenarios. The effect of density contrast on effective parameters has also been analyzed by Jiang et al. [22] by way of a stationary spectral approach. A discussion of current challenges in modeling density driven flows in the subsurface is offered by Werner et al. [41].

Here we consider the anisotropic dispersive Henry's problem introduced by Abarca et al. [1]. As key sources of model uncertainty we consider the following dimensionless parameters: (i) the gravity number, expressing the relative importance of buoyancy and viscous forces; (ii) the anisotropy ratio between aquifer vertical and horizontal permeabilities; (iii) the longitudinal and (iv) trans-verse Péclet numbers, quantifying the relative importance of the longitudinal and transverse dispersion on solute transport. These are critical in governing the general dynamics of density dependent flow and transport processes (see e.g. [1,14,25,29]). We then focus on a number of dimensionless global quantities (GQs) which are controlled by these parameters and are relevant to describe key features of the saltwater wedge and the width of the mixing zone. These global descriptors, as well as all system states such as pressure, concentration and velocity distributions within in the aquifer, are affected by uncertainty due to the lack of knowledge of the characteristic model parameters (e.g. [33]). Proper quantification of the uncertainty associated with the characterization of these GQs is of critical relevance for the management of coastal aquifers.

Propagation of model parameter uncertainty to a given quantity of interest can be quantified through a global sensitivity analysis (GSA). Here, we employ a variance-based GSA which allows assessing the relative impact of the model uncertain input parameters on the variability of model outputs [3]. We base our analysis on the Sobol indices [37], which are widely used sensitivity metrics and do not require any linearity assumptions in the underlying mathematical model of the system behavior.

Estimation of the Sobol indices is traditionally performed by Monte Carlo (MC) sampling in the uncertain parameter space. Therefore their computation can become highly demanding in terms of CPU time when the dimension of the parameters space and the degree of complexity of the problem increase. In this context, estimation of the Sobol indices is practically unfeasible in SWI problems because of the coupled nature of the flow and transport problems. We circumvent this problem upon relying on a generalized Polynomial Chaos Expansion (gPCE) approximation of the target GQs (e.g. [19,28]). This approach allows obtaining a surrogate model for a given quantity of interest and enables one to calculate the Sobol indices analytically via a straightforward post-processing analysis (e.g. [12,38]). Examples of application of this technique include the study of flow and transport in heterogeneous porous media (e.g. [26]). Formaggia et al. [17] and Porta et al. [31] demonstrate the reliability and computational efficiency of gPCE-based approaches in highly non-linear systems under the effect of mechanical and geochemical compaction processes.

The work is organized as follows. Section 2 presents the complete flow and transport mathematical model, the key dimensionless parameters governing the process and the global descriptors of interest. Section 3 is devoted to a brief description of the methodology we employ to perform GSA and to derive the gPCE surrogate model. Section 4 presents the setting analyzed and some details of the full and surrogate system models. In Section 5 we show the main results of our analysis in terms of the relative contribution of the uncertain parameters to the variance of each of the global quantities analyzed. We then study the joint and marginal probability density functions (*pdfs*) of these global quantities. We remark that these tasks are computationally unaffordable by making use of the complete system model, while they can be performed by means of the gPCE surrogate model. Moreover, our relying on the gPCE allows obtaining analytical expressions for the marginal *pdf* of the global quantities of interest.

2. Complete model and definition of the global quantities of interest

We consider the anisotropic dispersive Henry's problem introduced by Abarca et al. [1]. The setting is a modification of the original Henry's problem [21] and enables one to describe seawater intrusion in coastal aquifers in a way which renders vertical salinity distributions that mimic field evidences. Saltwater intrusion is modeled across a vertical cross-section of a homogeneous aquifer under isothermal conditions (see Fig. 1). Fluid flow is governed by the mass balance and Darcy equations, i.e.

$$\frac{\partial(\phi\rho)}{\partial t} + \nabla\cdot(\rho\boldsymbol{q}) = 0; \quad \boldsymbol{q} = -\frac{\boldsymbol{k}}{\mu}\cdot(\nabla p + \rho g \nabla z) \tag{1}$$

where $\boldsymbol{q}$ [L T^{-1}] is specific discharge vector with components q_x and q_z respectively along x- and z-directions (see Fig. 1); $\boldsymbol{k}$ [L^2] is the homogeneous and anisotropic diagonal permeability tensor with components $k_{11} = k_x$ and $k_{22} = k_z$, respectively along directions x and z; ϕ [–] is the porosity of the medium; μ [M L^{-1} T^{-1}] and ρ [M L^{-3}] respectively are dynamic viscosity and density of the fluid; p [M L^{-1} T^{-2}] is pressure; and g [L T^{-2}] is the gravitational constant.

Solute transport is described by the advection–dispersion equation

$$\frac{\partial(\phi\rho C)}{\partial t} + \nabla\cdot(\rho C\boldsymbol{q}) - \nabla\cdot[\rho\boldsymbol{D}\cdot\nabla C] = 0 \tag{2}$$

Here C [–] is solute concentration and $\boldsymbol{D}$ [L^2 T^{-1}] is the dispersion tensor, whose entries are defined as

$$D_{xx} = \phi D_m + \left(\alpha_L\frac{q_x^2}{|\boldsymbol{q}|} + \alpha_T\frac{q_z^2}{|\boldsymbol{q}|}\right); \quad D_{zz} = \phi D_m + \left(\alpha_L\frac{q_z^2}{|\boldsymbol{q}|} + \alpha_T\frac{q_x^2}{|\boldsymbol{q}|}\right);$$
$$D_{xz} = D_{zx} = (\alpha_L - \alpha_T)\frac{q_x q_z}{|\boldsymbol{q}|} \tag{3}$$

where D_m is the molecular diffusion coefficient and α_L and α_T [L] respectively are the longitudinal and transverse dispersivity coefficients, which are considered as uniform in the system. Since molecular diffusion is commonly neglected in transport settings taking place in porous media under the conditions we consider (e.g.[16]), in the following we disregard the contribution of D_m in (3). Initial conditions corresponding to freshwater hydrostatic pressure distribution are set in the system. No-flow conditions are imposed at the bottom and top of the domain; constant freshwater influx, q_f, is prescribed along the inland boundary ($x = 0$), where $C = 0$; saltwater hydrostatic pressure distribution is imposed along the seaside boundary, $x = l$ (i.e., $p = \rho_s g\,(d - z)$, ρ_s being density of seawater) where the salt mass flux is set as

$$(\boldsymbol{q}C - \boldsymbol{D}\cdot\nabla C)\cdot\boldsymbol{n} = \begin{cases} q_x C & \text{if } q_x > 0 \\ q_x C_s & \text{if } q_x < 0 \end{cases} \quad x = l \tag{4}$$

$\boldsymbol{n}$ and C_s respectively being the normal vector pointing outward from the aquifer and the concentration of salt in seawater (salinity). According to (4) water entering and leaving the system has salt concentration C_s and C, respectively. Key features and limitations of this schematization are illustrated in Abarca et al. [1].

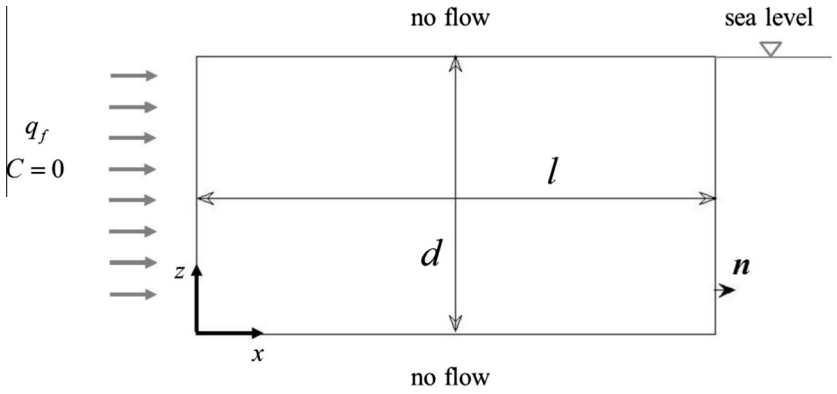

Fig. 1. Scheck of the flow and transport test problem.

Closure of the system (1)–(4) is obtained upon specifying a constitutive relationship between fluid properties, ρ and μ, and salt concentration. For the range of concentrations typically associated with SWI (e.g. [7]), viscosity can be assumed as constant and the following linear relationship can be employed to describe the evolution of ρ with C

$$\rho = \rho_f + (\rho_s - \rho_f)\frac{C}{C_s} \tag{5}$$

ρ_f being the freshwater density. Introducing the following dimensionless quantities

$$x' = \frac{x}{d};\quad z' = \frac{z}{d};\quad \nabla' = d\nabla;\quad t' = \frac{t}{d/q_f};\quad \boldsymbol{q}' = \frac{\boldsymbol{q}}{q_f};$$
$$\rho' = \frac{\rho}{\rho_s - \rho_f} = \frac{\rho}{\Delta\rho};\quad p' = \frac{p}{\Delta\rho q_f^2} \tag{6}$$

Eq. (1) can be rewritten in dimensionless form as

$$\frac{\partial(\phi\rho')}{\partial t'} + \nabla'\cdot(\rho'\boldsymbol{q}') = 0;\quad \boldsymbol{q}' = -\Upsilon N_g\ \frac{q_f^2}{gd}\nabla' p' + \rho'\nabla' z'\Big) \tag{7}$$

where Υ is a diagonal matrix with entries $\Upsilon_{11} = 1$ and $\Upsilon_{22} = r_k$, $r_k = k_z/k_x$ being the permeability anisotropy ratio, q_f^2/gd is a representative Froude number, and $N_g = g\Delta\rho k_x/q_f\mu$ is the gravity number (e.g. [24,29]).

The transport equation can be written in dimensionless form as

$$\frac{\partial(\phi\rho' C')}{\partial t'} + \nabla'\cdot(\rho' C'\boldsymbol{q}') - \nabla'\cdot[\rho'\boldsymbol{D}'\cdot\nabla' C'] = 0 \tag{8}$$

with

$$D'_{xx} = \frac{1}{Pe_L}\frac{q_x'^2}{|\boldsymbol{q}'|} + \frac{1}{Pe_T}\frac{q_z'^2}{|\boldsymbol{q}'|};\quad D'_{zz} = \frac{1}{Pe_L}\frac{q_z'^2}{|\boldsymbol{q}'|} + \frac{1}{Pe_T}\frac{q_x'^2}{|\boldsymbol{q}'|};$$
$$D'_{xz} = D'_{zx} = \left(\frac{1}{Pe_T} - \frac{1}{Pe_L}\right)\frac{q'_x q'_z}{|\boldsymbol{q}'|} \tag{9}$$

where $Pe_L = d/\alpha_L$ and $Pe_T = d/\alpha_T$ are longitudinal and transverse Péclet number, respectively, and $C' = C/C_s$. Eqs. (7)–(9) highlight that the problem under investigation is governed by eight dimensionless quantities, i.e., ϕ, r_k, $\rho_f/\Delta\rho$, N_g, Pe_L, Pe_T, C_s and q_f^2/gd.

Since point-wise measurements of state variables (e.g., salt concentration, velocity and pressure values) in SWI problems are usually scarce due to technical and economic constraints (e.g. [8]) in our study we focus on a global description of the process by considering dimensionless global quantities, GQs, that enable us to describe the overall seawater intrusion process. A similar approach has also been adopted by [1]. The four GQs of interest (note that each GQ is rendered dimensionless by normalization through d) are depicted graphically in Fig. 2 and defined in the following.

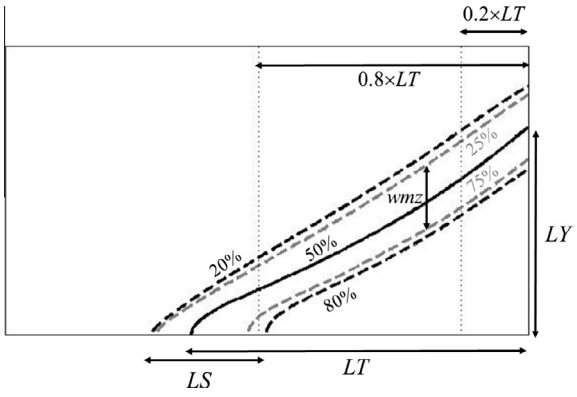

Fig. 2. Schematic description of the target GQs. Iso-concentration line C' = 20% and 80% (dashed black), 25% and 75% (dashed grey), and 50% (solid black) are depicted.

1. The dimensionless toe penetration, *LT*. This corresponds to the inland penetration, measured along the bottom of the domain, of the 50% C' isoline. This metric characterizes the inland extent of the saltwater wedge.
2. A measure, *LS*, of the spread of solute at the toe of the saltwater wedge. This is defined as the dimensionless distance, evaluated along the bottom of the domain, between the 20% and 80% C' isolines.
3. The dimensionless average width of the mixing zone, *WD*. Here *WD* is evaluated as the average of $wmz(x')$ within the region $0.2 \times LT \leqslant x' \leqslant 0.8 \times LT$; $wmz(x')$ being the dimensionless vertical distance between the 25% and 75% C' isolines (in case the 75% C' isoline has intruded the domain up to a distance less than $0.8 \times LT$, we consider $wmz(x')$ as the vertical distance between the 25% C' isoline and the bottom of the aquifer).
4. The dimensionless sinking of the saltwater wedge at the seaside boundary, *LY*. This is quantified as the dimensionless vertical distance between the bottom of the aquifer and the 50% C' contour curve at the seaside boundary.

3. Uncertainty quantification via global sensitivity analysis and generalized Polynomial Chaos Expansion

In this section we introduce the Sobol indices that will be used in Section 5 to investigate the way lack of knowledge of key parameters appearing in the problem formulation (6)–(9) propagates to the

selected global descriptors. We also briefly describe the way we alleviate the computational burden by introducing a surrogate model of the coupled flow-transport problem illustrated in Section 2.

As seen in Section 2, the seawater intrusion problem is governed by eight dimensionless quantities. Amongst these, we consider as uncertain the following four parameters: (i) the permeability anisotropy ratio, r_k, (ii) the longitudinal, Pe_L, and transverse, Pe_T, Péclet numbers, expressing the effect of the longitudinal and transverse dispersivity, and (iii) the gravity number, N_g, quantifying the relative importance of buoyancy and viscous forces. Uncertainty in these quantities is associated with our imperfect knowledge of the aquifer hydraulic (i.e., permeability tensor components) and dispersive parameters related to saltwater spreading in the system at the scale of observation.

We collect the four uncertain quantities N_g, r_k, Pe_L and Pe_T in a random parameter vector $\boldsymbol{x} \equiv (x_1, \ldots, x_N)$; with $N = 4$) and treat each x_n as an independent random variable. The occurrence of correlation among entries of $\boldsymbol{x}$ could be included in the methodology (see, e.g. [27]). However, this would require the knowledge of the marginal probability density functions of the uncertain parameters and the associated correlation matrix. The type of correlations which can be found in the literature, e.g., between α_L and α_T (i.e., between Pe_L and Pe_T) are purely empirical and not generally supported by theoretical arguments. Due to the general lack of prior information on $\boldsymbol{x}$, we also assume that each x_n can be described by a uniform distribution within the interval $\Gamma_n = [x_{n,min}, x_{n,max}]$.

Our uncertainty quantification study is based on a global sensitivity analysis (GSA) which is performed through the evaluation of the variance-based Sobol indices (e.g. [3,36]). The latter provide a description of the way the variability of a given quantity of interest, quantified in terms of its total variance, is affected (separately and jointly) by the random parameters collected in $\boldsymbol{x}$. Sobol indices provide generally robust results, as these are not constrained by any linearity assumption on the underlying mathematical model [34], a feature which is particularly critical in a complex system of the kind we investigate.

A target quantity $f(\boldsymbol{x})$, representing a given GQ, which depends on N independent random variables can be decomposed as

$$f(\boldsymbol{x}) = f_0 + \sum_{i=1}^{N} f_i(x_i) + \sum_{i,j=1}^{N} f_{i,j}(x_i, x_j) + \ldots f_{1,2,\ldots,N}(x_1, x_2, \ldots, x_N) \tag{10}$$

where

$$\begin{aligned} &f_0 = \int_\Gamma f(\boldsymbol{x}) p_\Gamma(\boldsymbol{x}) d\boldsymbol{x}; \quad f_i(x_i) = \int_{\Gamma\_i} f(\boldsymbol{x}) p_{\Gamma\_i}(\boldsymbol{x}) d\boldsymbol{x} - f_0; \\ &f_{i,j}(x_i, x_j) = \int_{\Gamma\_i,j} f(\boldsymbol{x}) p_{\Gamma\_i,j}(\boldsymbol{x}) d\boldsymbol{x} - f_0 - f_i(x_i) - f_j(x_j) \end{aligned} \tag{11}$$

Here $\Gamma = \Gamma_1 \times \ldots \times \Gamma_N$ is the hypercube representing the space of variability of $\boldsymbol{x}$ and $p_\Gamma(\boldsymbol{x})$ is the joint probability density of $\boldsymbol{x}$ over Γ; integration over Γ_i is performed over the space of $\boldsymbol{x}$ excluding Γ_i, p_{Γ_i} being the corresponding density function. Note that f_0 in (10) is the mean of $f(\boldsymbol{x})$.

The Sobol index $S_{i_1,i_2,\ldots,i_s}$, embedding the mixed effects of $x_{i_1}, \ldots, x_{i_s}$ [36,37], is defined as

$$S_{i_1,\ldots,i_s} = \frac{1}{V_f} \int_{\Gamma_{i_1,\ldots,i_s}} f^2_{i_1,\ldots,i_s}(x_{i_1}, \ldots, x_{i_s}) p_{\Gamma_{i_1,\ldots,i_s}}(\boldsymbol{x}) dx_{i_1} \ldots dx_{i_s} \tag{12}$$

V_f being the total variance of $f(\boldsymbol{x})$, i.e.

$$V_f = \int_\Gamma f(\boldsymbol{x})^2 p_\Gamma(\boldsymbol{x}) d\boldsymbol{x} - f_0^2 \tag{13}$$

From (12), the principal sensitivity index of x_n, denoted as S_n, describes the contribution of only x_n (without considering interactions with the other parameters) on the total variance, V_f. The overall contribution of parameter x_n to V_f is then given by the total sensitivity index

$$S_n^T = S_n + \sum_j S_{n,j} + \sum_{k,j} S_{n,j,k} + \cdots \tag{14}$$

and includes S_n and all the joint terms where x_n appears. Denoting by S_n^L the contribution to S_n of the linear term associated with x_n, $S_n^{NL} = (S_n^T - S_n^L)$ represents the total contribution of non-linear terms involving x_n, i.e., S_n^{NL} indicates the degree of nonlinearity of the input–output mapping of $f(\boldsymbol{x})$ with respect to x_n.

Computing the indices (12) requires multiple integrations of the model response $f(\boldsymbol{x})$ for diverse combinations of the uncertain parameters. This is typically achieved by numerical Monte Carlo simulation and the associated computational cost can be very high, depending on model complexity and on the number of random parameters considered [38].

The generalized Polynomial Chaos Expansion (gPCE) (e.g. [19,28,36,38,42]) can be employed to build surrogate models of target quantities at a relatively affordable computational cost (see, e.g., [11,15,17] and references therein). As such, the gPCE approximation can be used to ascertain the way uncertainty associated with unknown model parameters propagates to system states of interest. In this context, the gPCE of $f(\boldsymbol{x})$ can be constructed as a spectral expansion of $f(\boldsymbol{x})$ in terms of a set of orthonormal polynomials representing a basis of the probabilistic space Γ within which an approximation of the model response surface is built. The specific family of polynomials which can be used depends on the probability distribution of the uncertain model parameters considered. Since each x_n is here assumed to be uniformly distributed, we adopt the family of multivariate Legendre polynomials [42]. The way a multidimensional Legendere polynomial is constructed starting from univariate Legendre polynomials is described by [18] to which we refer for additional details. The gPCE approximation of $f(\boldsymbol{x})$ can be constructed as

$$\begin{aligned} &f(\boldsymbol{x}) = \theta_0 + \sum_{i=1}^{N} \sum_{\mathbf{p} \in \Im_i} \theta_{\mathbf{p}} L_{\mathbf{p}}(\boldsymbol{x}) + \sum_{i,j=1}^{N} \sum_{\mathbf{p} \in \Im_{i,j}} \theta_{\mathbf{p}} L_{\mathbf{p}}(\boldsymbol{x}) + \cdots; \\ &L_{\mathbf{p}}(\boldsymbol{x}) = \prod_{i=1}^{N} L_{i,p_i}(x_i); \quad \theta_{\mathbf{p}} = \int_\Gamma f(\boldsymbol{x}) L_{\mathbf{p}}(\boldsymbol{x}) p_\Gamma(\boldsymbol{x}) d\boldsymbol{x} \end{aligned} \tag{15}$$

Here $\mathbf{p} = \{p_1, p_2, \ldots, p_N\} \in \boldsymbol{N}^N$ is a multi-index expressing the degree of each univariate Legendre polynomial, $L_{n,p_n}(x_n)$, employed to construct the multivariate orthogonal Legendre polynomial $L_{\mathbf{p}}(\boldsymbol{x})$, $\theta_{\mathbf{p}}$ is the associated polynomial coefficient, and $\Im_i$ contains all indices such that only the ith component does not vanish, i.e., $\Im_i = \{p_i \neq 0, p_k = 0 \text{ for } k \neq i\}$.

Considering (10)–(15) allows deriving the equivalence between the Sobol indices and the coefficients $\theta_{\mathbf{p}}$ of the gPCE representation of $f(\boldsymbol{x})$, i.e.

$$S_{i_1,\ldots,i_s} = \frac{1}{V_f} \sum_{\mathbf{p} \in \Im_{i_1,\ldots,i_s}} \theta_{\mathbf{p}}^2; \quad f_0 = \theta_0; \quad V_f = \sum_{\mathbf{p} \in \boldsymbol{N}^N} \theta_{\mathbf{p}}^2 \tag{16}$$

Eq. (16) can be rendered workable upon truncation of the summation to a set of polynomials with total degree w, i.e., $\sum_i p_i \leqslant w$. Other possible truncation schemes are discussed by Bäck et al. [4]. The accuracy of the resulting gPCE approximation increases with the regularity of $f(\boldsymbol{x})$ and as $w \to \infty$. This aspect is explored in Section 4.

4. Test case description and Numerical implementation

4.1. Complete numerical model

We solve the coupled flow and transport problem defined by (1)–(5) by means of the widely tested numerical code SUTRA

Table 1
Range of variability of the dimensionless uncertain parameters.

	$\Gamma_n = [x_{n,min}, x_{n,max}]$
Γ_{N_g}	[3.04; 5.06]
Γ_{r_k}	[10^{-8};1.0]
Γ_{Pe_L}	[3.33; 10]
Γ_{Pe_T}	[10; 100]

[40] within a homogeneous anisotropic porous medium with porosity ϕ = 35%. We set the height and length of the domain respectively to d = 1 m and l = 2 m (see Fig. 1). The density of fresh and sea water are given respectively by ρ_f = 1000 kg/m^3 and ρ_s = 1025 kg/m^3, while the fluid viscosity is constant and equal to $\mu = 10^{-3}$ kg/m s. Freshwater enters the system with a Darcy velocity $q_f = 6.6 \times 10^{-5}$ m/s and the concentration of salt in seawater, C_s, is set equal to the standard value 35.7×10^{-3} kg/kg. As highlighted in Section 3, we treat N_g, r_k, Pe_L, and Pe_T as uncertain parameters. In this work we do not investigate the impact of the uncertainty of the geometrical setting, boundary conditions, and fresh and sea water density and viscosity. The intervals of variability of the four dimensionless uncertain parameters are listed in Table 1. The lower and upper bounds for each range of parameter variability have been selected on the basis of available information on the dimensional parameters involved in their definition. In our work the variability of N_g arises from the uncertainty in the horizontal permeability, having fixed the fluid properties and the inland freshwater flux. Since Pe_L, and Pe_T affect solute spreading (e.g. [39,41]), the large range of variability we consider attempts to cover several situations encountered in real scenarios. Note that the selected range of variability for r_k enables one to consider both isotropic and strongly anisotropic aquifers.

The computational domain depicted in Fig. 1 is discretized through a grid with uniform square elements with side $\Delta = \Delta x = \Delta y$. We performed a set of preliminary simulations aimed at testing the influence of grid discretization on the quantification of the global quantities of interest defined in Section 2. We found no appreciable difference between the numerical results associated with grids formed by 200 × 100 (i.e., $\Delta = 10^{-2}$ m), 256 × 128 (i.e., $\Delta = 7.8 \times 10^{-3}$ m) or 400 × 200 (i.e., $\Delta = 5 \times 10^{-3}$ m) elements (details not shown). On these bases, all results reported in the fol-lowing are associated with a uniform grid of 256 × 128 elements. To ensure numerical stability (e.g. [40]), all numerical simulations have been performed using a grid Péclet number $Pe_m = \Delta/\alpha_L < 4$, while keeping $\Delta < \alpha_T$. Eqs. (1)–(5) are solved until steady-state conditions are reached. The latter are attained after a total simu-lated time of about 17 h. The effect of different time steps has been tested and a uniform time discretization Δt = 60 s has been found to render accurate results. With these settings, a single simulation via SUTRA is associated with a computational cost of about 30 min. All numerical results reported here have been obtained on an Intel® Core™ i5-2410 M CPU @ 2.30 GHz processor. The target global quantities *LT*, *LS*, *WD* and *LY* are then analyzed at steady-state. This study is consistent with the original formulation of Henry's solution [14,21] and with the way the saltwater intrusion scenario is characterized in most environmental applications (e.g. [1] and references therein).

4.2. Construction and validation of the gPCE approximation of the global quantities

As described in Section 3, constructing the gPCE of a target system response entails solving the complete system model (1)–(5) for diverse combinations of the selected random parameters N_g, r_k, Pe_L, and Pe_T. The number of these combinations depends on the total degree w selected for the polynomial representation. In this section we assess the robustness of the gPCE approximations with diverse degrees w, in terms of their ability to provide accurate representations of the (random) global quantities *LT*, *LS*, *WD* and *LY*. We base our analysis on comparisons between values of each GQ obtained via the solution of the full model (1)–(5), $f(\boldsymbol{x})_{FM}$, against corresponding gPCE approximations, $f(\boldsymbol{x})_{gPCE}$. The coeffi-cients $\theta_{\mathbf{p}}$ are calculated by solving the multidimensional integral in (15) through the sparse grids interpolation technique [e.g. 17](i.e., the so-called non-intrusive spectral projection). We do so by using Legendre–Gauss points. Fig. 3 depicts the results of such an analysis in terms of the relative error, defined as $e_f = (f(\boldsymbol{x})_{gPCE} - f(\boldsymbol{x})_{FM})/f(\boldsymbol{x})_{FM}$ (with f = *LT*, *LS*, *WD* and *LY*) resulting

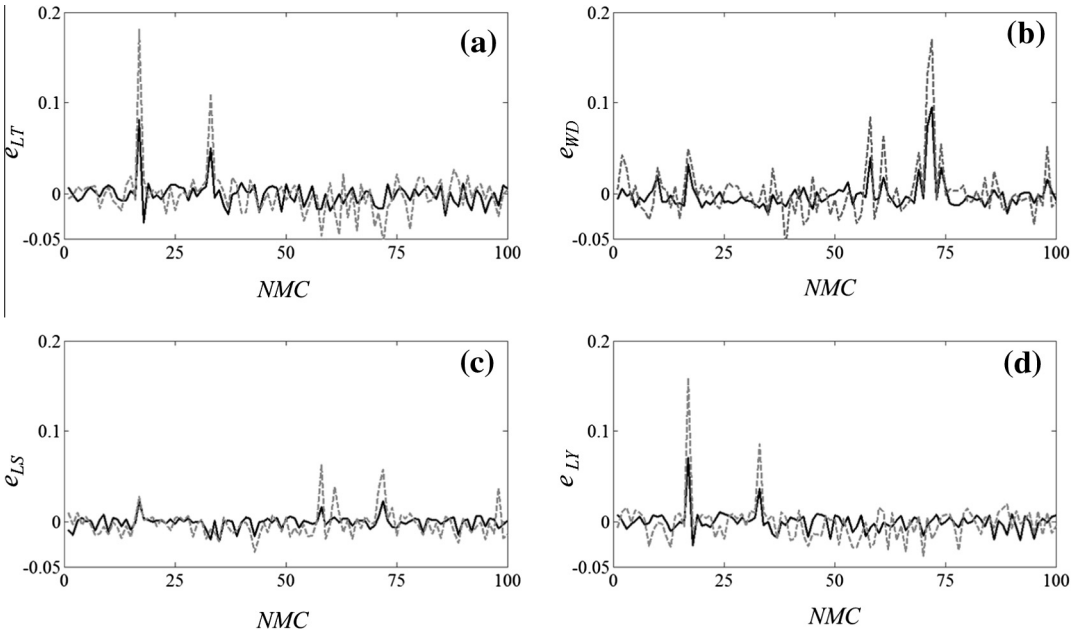

Fig. 3. Relative error, e_f, between values of the GQs obtained via the solution of the full model (1)–(5), and the corresponding gPCE approximations associated with $w = 3$ (solid curves) and $w = 2$ (dashed curves) for f = (a) *LT*, (b) *WD*, (c) *LS*, (d) *LY*.

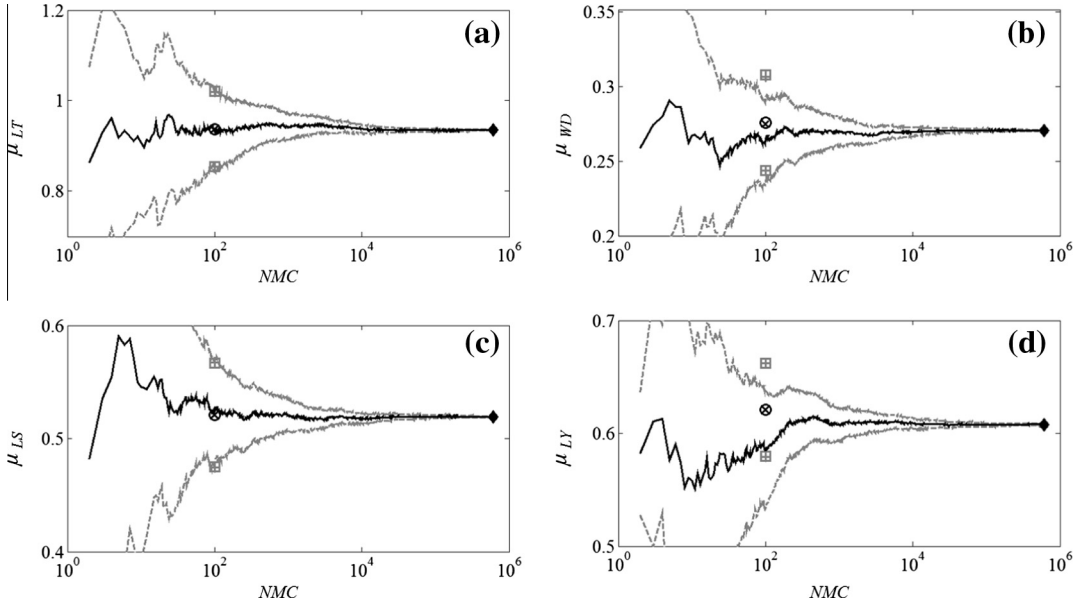

Fig. 4. Convergence of sample mean, μ_f (solid black curve), and associated 95% CIs (grey dashed curves) with the number of Monte Carlo simulations (*NMC*) for f = (a) *LT*, (b) *WD*, (c) *LS*, and (d) *LY* obtained through a gPCE with $w = 3$. Corresponding results obtained via 100 independent MC simulations of (i) the complete model (μ_f – black circle – and 95% CIs – grey square) and (ii) the gPCE (μ_f – black cross – and 95% CIs – grey cross) based on the same set of input random parameters are also depicted. The black diamond represents the analytical solution of (16) with $w = 3$.

from $NMC = 100$ random MC realizations of the system parameters drawn from the parameter space. Note that these simulations do not coincide with those employed for the construction of the gPCE. The gPCE approximations have been evaluated using two values of w, i.e., $w = 2, 3$, for which the sparse grid is formed by 41 and 137 collocation points in the parameter space and the associated computational costs are about 20 and 68 h, respectively. A very good agreement can be seen between both gPCE approximations and the full system model solution, the maximum absolute values of e_f being smaller than 18% or 9%, while the mean absolute values of e_f is smaller than 2.0% or 1.1%, respectively for $w = 2$ or $w = 3$.

It is worthwhile to note that once expressions of the gPCE approximations are available, a large number of Monte Carlo iterations can be obtained at a very low additional computational cost upon sampling the random input parameter space. This allows grounding the analysis of the statistics of the target GQs on a very large sample of Monte Carlo realizations of the gPCE approximations. Figs. 4 and 5 respectively depict the dependence of the sample mean, μ_f, and variance, σ_f^2, of the GQs on the number *NMC* of MC realizations obtained via gPCE with $w = 3$. The 95% estimated confidence intervals (CIs) are also graphically reported (see, e.g.[5]). Also depicted in the figures are corresponding results obtained

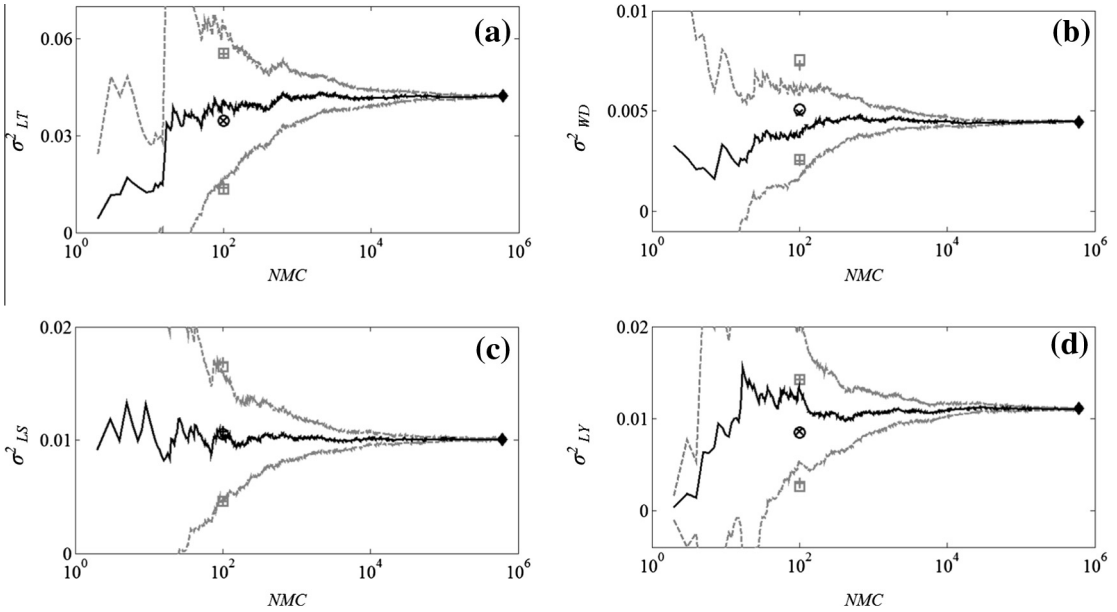

Fig. 5. Convergence of sample variance, σ_f^2 (solid black curve) and associated 95% CIs (grey dashed curves) with the number of Monte Carlo simulations (*NMC*) for f = (a) *LT*, (b) *WD*, (c) *LS*, and (d) *LY* obtained through a gPCE with $w = 3$. Corresponding results obtained via 100 independent MC simulations of (i) the complete model (σ_f^2 – black circle – and 95% CIs – grey square) and (ii) the gPCE (σ_f^2 – black cross – and 95% CIs – grey cross) based on the same set of input random parameters are also depicted. The black diamond represents the analytical solution of (16) with $w = 3$.

via (i) 100 MC simulations of the complete model (performed independent of the simulations upon which the gPCE construction is based) and the gPCE using the same set of random input parameters, and (ii) analytical values of μ_f and σ_f^2 evaluated directly by (16) without resorting to the MC procedure. The latter values coincide with their Monte Carlo estimates with NMC $\approx 10^5$, as expected. The results embedded in these figures clearly demonstrate that stabilization of the statistical moments of interest (and in particular of σ_f^2) can be reached only at very large values of *NMC* (larger than 10^4). Note that such analysis would be unfeasible by solving the complete model (1)–(5). Our analysis also indicates that (i) all moments and the associated 95% CIs evaluated via the full model and via the gPCE approximations practically coincide for the *NMC* = 100 random parameters realizations considered independent of those employed for the construction of the gPCE; (ii) these moments lie within the 95% CIs obtained by independent gPCE realizations (dashed line); (iii) the estimates of the mean, μ_f, and variance, σ_f^2, obtained via gPCE and *NMC* = 5 $\times 10^5$ and the 95% CIs lie within the full model-based CIs evaluated by *NMC* = 100.

To further explore the reliability of our gPCE approximations, we perform a two-sample Kolmogorov–Smirnov (K–S) test [30] to compare the marginal probability distribution of the GQs obtained via the full model and their gPCE-based counterparts evaluated with w = 3. The null hypothesis that the samples belong to the same population is not rejected by the K–S test at a signifi-cance level of 0.05 (p-values always being larger than 0.9, details not reported), thus supporting the use of the gPCE surrogate mod-els not only for the analysis of the statistical moments of the global quantity but also their probability distributions.

Finally, we analyze the level of correlation between two GQs of interest, e.g., $f(\boldsymbol{x})$ and $g(\boldsymbol{x})$, by relying on the Pearson' correlation coefficient, ρ_p. It is worthwhile to note that, when the gPCEs of $f(\boldsymbol{x})$ and $g(\boldsymbol{x})$ are available, ρ_p can be evaluated analytically making use of (15) and (16) as

$$\rho_p = \frac{\mathrm{cov}(f(\boldsymbol{x}), g(\boldsymbol{x}))}{\sqrt{V_f V_g}} = \frac{\sum_{i=1}^{N}\sum_{\mathbf{p},\mathbf{q}\in\Im_i}\theta_{\mathbf{p}}\theta_{\mathbf{q}} + \sum_{i,j=1}^{N}\sum_{\mathbf{p},\mathbf{q}\in\Im_{i,j}}\theta_{\mathbf{p}}\theta_{\mathbf{q}} + \cdots}{\sqrt{V_f V_g}} \tag{17}$$

Table 2 lists the Pearson' correlation coefficients evaluated via (17) with w = 3 and the corresponding values obtained on the basis of the above mentioned 100 MC realizations of the full model. Once again, it can be noted that the agreement between the full model- and the gPCE-based results is quite remarkable, especially for the highly (positively or negatively) correlated quantities. On these bases, the analyses presented in the following section are grounded on gPCE approximations of order w = 3 for all of the global quantities of interest.

Table 2
Pearson′ correlation coefficient of the GQs evaluated via gPCE, (black fonts) and 100 MC random realizations of the full model (grey fonts).

	WD	*LT*	*LS*	*LY*
WD	1	−0.49	0.75	−0.21
LT	−0.51	1	−0.02	0.80
LS	0.76	−0.18	1	−0.03
LY	−0.13	0.77	−0.08	1

5. Results and discussion

5.1. Variance-based Sobol Indices

Principal, S_n, and total, S_n^T, Sobol indices, together with linear contributions, S_n^L, are listed in Table 3 for all global quantities of interest and considering all of the random input dimensionless variables (i.e., in Table 3 $n = N_g$, r_k, Pe_L, and Pe_T). Values of S_n^L, S_n and S_n^T coincide (or are very close to each other) indicating that non-linear terms and joint interaction between the uncertain parameters are almost negligible for all GQs with the exception of *LY*. The results embedded in Table 3 show that non-linear terms including r_k cannot be neglected for the evaluation of *LY*.

The uncertainty associated with the intrusion of the toe of the wedge, as represented by *LT*, is mainly controlled by $N_g = g\Delta\rho k_x/q_f\mu$ (which has about 50% weight in directing the total variance of *LT*). The nature of N_g allows recognizing that increasing $k_x/q_f\mu$ causes a reduction of the pressure drop required by a given q_f to flow towards the sea and the ensuing retreat of the wedge. On the other hand, an increase of $\Delta\rho$ results in an increased capability of seawater to intrude the aquifer. The wedge intrusion is also strongly affected by Pe_T. This is so because mixing of fresh and salt water tends to be reduced with increasing Pe_T so that a high density contrast can be sustained, leading to an increase of *LT*.

Spreading of solute at the bottom of the wedge, as described by *LS*, is mainly affected by Pe_L and to a less extent by Pe_T. This result is consistent with the previous observation of Abarca et al. [1] that longitudinal dispersivity controls the distribution of concentrations in the lowest (bottom) part of the domain.

The roles of Pe_L and Pe_T are then reversed in governing the variability of *WD*, describing the extension of the mixing zone. Abarca et al. [1] suggested a linear relationship between *WD* and $\sqrt{\alpha_L\alpha_T}$, thus implying that α_L and α_T contribute in equal measure to *WD*. Our results, which are based on independence between Pe_L and Pe_T, show that Pe_T (i.e. α_T) plays the main role in controlling the uncertainty of *WD*. Our finding that *WD* is almost insensitive to N_g or r_k suggests that for given values of Pe_L and Pe_T the spatial distribution of the entries of the dispersion tensor remains essentially unaltered within the transition zone. Therefore, changes in N_g and r_k only cause a horizontal (inland or seaward) and vertical shift of the transition region whose extension remains practically constant.

Finally, variations of r_k only affect variability of *LY* and *LT*. This result is consistent with the observation that decreasing r_k (i.e., decreasing the vertical permeability) causes the vertical component of the freshwater velocity to decrease so that the capability of the horizontal freshwater flux to contrast the intrusion of the saltwater is augmented. As a consequence, *LY* and *LT* tend to decrease with r_k.

5.2. Probability distributions of global quantities of interest

This Section is devoted to the analysis of the marginal and joint probability distributions, *pdfs*, of the global quantities we analyze. These results are relevant for management of coastal aquifers as

Table 3
Principal (S_n) and total (S_n^T) Sobol indices ($n = N_g$, r_k, Pe_L, Pe_T), together with the linear component S_n^L of S_n.

	LT	*LS*	*WD*	*LY*
$S_{N_g}^L$; S_{N_g}; $S_{N_g}^T$	0.49; 0.50; 0.51	0.16; 0.17; 0.19	0.00; 0.00; 0.00	0.16; 0.16; 0.20
$S_{r_k}^L$; S_{r_k}; $S_{r_k}^T$	0.10; 0.14; 0.14	0.01; 0.02; 0.02	0.03; 0.04; 0.04	0.60; 0.70; 0.70
$S_{Pe_L}^L$; S_{Pe_L}; $S_{Pe_L}^T$	0.05; 0.05; 0.05	0.51; 0.51; 0.51	0.27; 0.27; 0.28	0.00; 0.00; 0.00
$S_{Pe_T}^L$; S_{Pe_T}; $S_{Pe_T}^T$	0.29; 0.31; 0.32	0.25; 0.27; 0.28	0.65; 0.67; 0.69	0.13; 0.13; 0.13

they can be assist in the quantification of the probability of failure of the system (i.e., the probability to exceed a given threshold values of the GQ).

We construct these sample *pdfs* by relying on the gPCE representation. The latter is employed to perform 5×10^5 Monte Carlo simulations of the target system states at a remarkably low computational cost (about 68 h for the construction of the gPCE approximations, plus about 2 h for the generation of the MC realizations), as opposed to standard Monte Carlo simulations performed with the complete model which would be unfeasible (the estimated CPU time is about 1.7×10^5 h). Uniform sampling of the parameter space is here employed.

Once a gPCE representation is constructed, it is also possible to derive an analytical expression of the marginal probability distribution (marginal *pdf*) of the GQ, without resorting to numerical MC sampling. The procedure is detailed in Appendix A. Appendices A1–A3 show that, when the gPCE approximation (15) can be trun-ated to the first order Legendre polynomials of the parameters which are identified as relevant on the basis of the associated Sobol indices (see Table 3), as in the case of our analysis for *WD*, *LT*, and *LS*, the related marginal *pdfs* are fully determined by the coefficients of the Legendre polynomial of orders zero and one. When non-linear terms are relevant in the gPCE (i.e., for *LY*, as described in Section 5.1), the complexity of the analytical expression of the marginal *pdf* increases. Details for this case are illustrated in Appendix A4 for completeness.

Fig. 6 juxtaposes the MC-based marginal *pdfs* of the GQs with the corresponding analytical formulations. The agreement between the numerical and analytical results is remarkable. The marginal *pdfs* of *LT*, *WD* and *LS*, for which non-linear effects can be neglected, are symmetric around the mean. Otherwise, the marginal *pdf* of *LY*, where non-linear effects due to r_k are relevant, is negatively skewed, its long tail being associated with settings characterized by low values of the vertical permeability.

Fig. 6 also depicts contour plots of the joint *pdf* of all pairs of the global variables. A negative correlation between *LT* and *WD* is evident. The latter is mainly due to the contrasting effects that Pe_T has on these quantities. While increasing Pe_T (i.e., decreasing α_T) causes the mixing zone to decrease, the inland intrusion of seawa-ter increases. The joint *pdf* of *WD* and *LS* is characterized by a clear positive correlation. This is related to the observation that both quantities are primarily affected by the longitudinal and transverse Péclet numbers and they both decrease with Pe_L and Pe_T. Note also that the joint *pdf* of *LT* and *LY* shows a marked elongation in the region corresponding to relatively low values of both global variables. This behavior is representative of settings characterized by low values of the vertical permeability, i.e. by small values of r_k. The joint *pdfs* reported in Fig. 6 are also in agreement with the pic-ture offered by the Pearson' correlation coefficients introduced in Section 4 and listed in Table 2. In particular, these confirm the lack of correlation between *WD* and *LY* and between *LS* and all global variables with the exception of *WD*.

In practical applications, some of the global quantities here considered can be known through experimental campaigns. A relevant question in the management of coastal aquifers is therefore how can the knowledge of one global quantity affect the probability distribution (and therefore the predictability) of the remaining GQs. An answer to this question can be directly obtained by applying Bayes' theorem to the joint *pdfs* described above and then deriving the conditional *pdf* of the GQs of interest.

For the problem here considered, while measurements of *LY* can be obtained with relatively modest efforts (e.g., through electrical

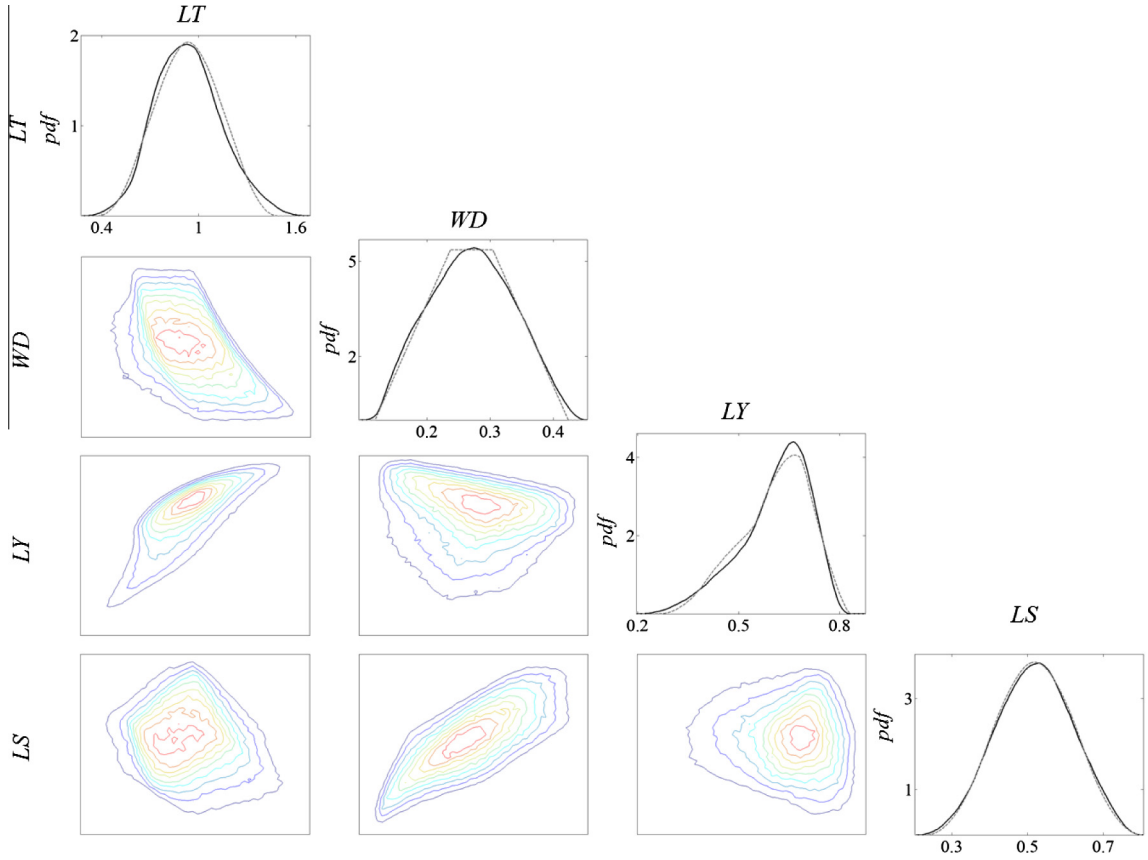

Fig. 6. Numerical (black continuous) and analytical (grey dashed) marginal *pdf* for *LT*, *WD*, *LY* and *LS*. Contours of joint *pdf* for all pairs of GQs are also shown (color curves). For rendering purposes, the iso-probability curves in each subplot are normalized by the corresponding maximum *pdf* value.

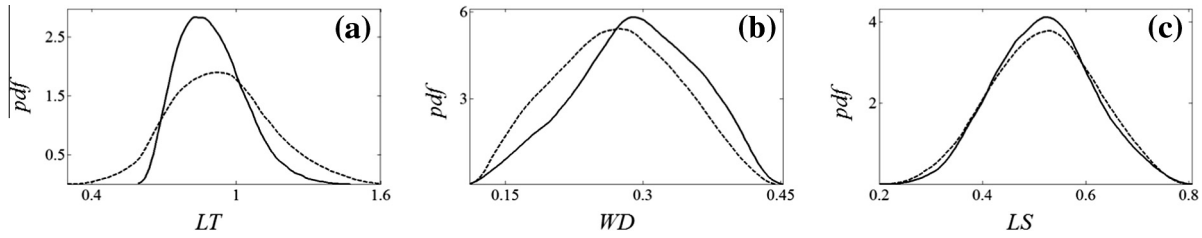

Fig. 7. Marginal pdf of (a) *LT*, (b) *WD* and (c) *LS* conditioned to $LY = \mu_{LY} \pm 6.1 \times 10^{-3}$ (solid curves). Corresponding unconditional *pdfs* are also depicted (dashed curves).

conductivity profiles), acquiring data on the remaining GQs (and in particular of *LT* and *LS*) is problematic. It is therefore of interest to analyze the way the availability of *LY* data would affect predictions of the remaining GQs. On the basis of Fig. 6, one can anticipate that while *LY* measurements would affect the predictability of *LT*, these data would not influence markedly the marginal *pdf* of *WD* and *LS* because of their weak level of correlation with *LY*. As an example of the type of results which one can obtain, Fig. 7a depicts the marginal *pdf* of *LT* conditioned to the mean value of *LY*, i.e., μ_{LY} = 0.61. In Fig. 7a we consider a measurement error with ampli-tude equal to ±1% μ_{LY} and condition the marginal *pdf* of *LT* on the range of *LY* values $[\mu_{LY} - 6.1 \times 10^{-3}, \mu_{LY} + 6.1 \times 10^{-3}]$. The uncon-ditional *pdf* is also depicted for comparison. Corresponding results for *WD* and *LS* are depicted in Fig. 7b and c, respectively. These results clearly show the reduction of the variability of *LT*, which is, e.g., quantified by its variance that decreases from 4.2×10^{-2} for the unconditional case to 1.8×10^{-2} for the conditional one. On the other hand, knowledge of *LY* practically does not affect the *pdf* of *WD* and *LS*, as expected by the reasoning illustrated above.

5. Conclusions

We analyze the way key global dimensionless quantities (GQs) characterizing the saltwater wedge in the dispersive Henry's problem are affected by the incomplete knowledge of system properties, as encapsulated in the gravity number (N_g), the anisotropic ratio of permeability (r_k), the transverse (Pe_T) and longitudinal (Pe_L) Péclet number. The analysis is based on the generalized Polynomial Chaos Expansion (gPCE) of the following GQs: the median toe penetration (*LT*), the spread of solute around the toe (*LS*), the mean width of the mixing zone (*WD*) and the sinking of the wedge at the seaside boundary (*LY*). These kinds of analyses can be useful to assist coastal aquifer management and risk assessment procedures as they lead to an appropriate probabilistic characterizations of the saltwater wedge.

Our uncertainty quantification procedure relies on the variance-based Sobol indices. We found that the variability of the gravity number greatly influences *LT*, while uncertainty of the permeability anisotropy ratio chiefly controls *LY*. On the other hand, transverse and longitudinal Péclet number respectively affects mainly the variance of the width of the mixing zone and of the spread of solute around the toe.

We then compute the joint and marginal probability density functions (*pdfs*) of the target global descriptors. We remark these tasks are computationally unaffordable by making use of the complete system model, while they can be performed with a relatively modest computational effort by means of the constructed gPCE surrogate model. We also derive analytical expressions for the marginal *pdf* of the global quantities on the basis of the gPCE approximations. The study of the joint *pdfs* allows us to highlight the degree of correlation between the GQs. This analysis is relevant to management and assessment of quality of coastal aquifers because it enables us to identify the way information on one state variable impacts on the reduction of uncertainty associated with other target quantities, thus ultimately constituting a potential driver to planning required experimental campaigns.

Future extensions of this study comprise the study of seawater intrusion processes in randomly heterogeneous systems, the inclusion of additional uncertain quantities such as porosity, inland freshwater flux and salt concentration at the sea boundary. The investigation of flow scenarios of increased complexity, such as those involving pumping and recharge wells in coastal aquifers is also envisioned.

Acknowledgments

Part of this work was supported by MIUR (Italian ministry of Education, Universities and Research, PRIN2010-11) project: "Hydroelectric energy by osmosis in coastal areas".

Appendix A. Analytical derivation of the marginal *pdf* of the target global variables

In this Appendix we detail our derivations conducive to the analytical expressions of the marginal probability density function, *pdf*, of the target global variables we consider. For illustration purposes, in the following we retain in (15) and (16) all terms which allow representing at least 90% of the total variance of each state variable.

A1. Marginal pdf of WD

Within the range of variability of the uncertain parameters listed in Table 1, one can note that 91% of the total variance of *WD* can be retained upon truncating (15) up to the first order Legendre polynomials ξ_{Pe_T} and ξ_{Pe_L}, defined as

$$\xi_{Pe_T} = \sqrt{3}\frac{2Pe_T - g - h}{h - g}; \quad \xi_{Pe_L} = \sqrt{3}\frac{2Pe_L - e - f}{f - e} \tag{A1}$$

where $Pe_T \in \Gamma_{Pe_T} = [g, h]$ and $Pe_L \in \Gamma_{Pe_L} = [e, f]$ (see Table 1). Therefore, (15) becomes

$$WD' = WD - \theta_{WD;0} = \theta_{WD;Pe_T}\xi_{Pe_T} + \theta_{WD;Pe_L}\xi_{Pe_L} \tag{A2}$$

where $\theta_{WD;0}$ is the mean of *WD*, $\theta_{WD;Pe_T}$ and $\theta_{WD;Pe_L}$ are the coefficients of the Legendre polynomial of order one in Pe_T and Pe_L, respectively.

Eq. (A2) can be rewritten as

$$WD' = A_{WD} + B_{WD}Pe_T + C_{WD}Pe_L \tag{A3}$$

where

$$A_{WD} = -\sqrt{3}\left(\theta_{WD;Pe_T}\frac{h+g}{h-g} + \theta_{WD;Pe_L}\frac{f+e}{f-e}\right);$$

$$B_{WD} = \frac{2\sqrt{3}\theta_{WD;Pe_T}}{h-g}; \quad C_{WD} = \frac{2\sqrt{3}\theta_{WD;Pe_L}}{f-e} \tag{A4}$$

The probability density function, of Pe_T and Pe_L are respectively given by

$$pdf_{Pe_T} = \frac{1}{h-g}[H(Pe_T - g) - H(Pe_T - h)];$$
$$pdf_{Pe_L} = \frac{1}{f-e}[H(Pe_L - e) - H(Pe_L - f)] \quad \text{(A5)}$$

H being the Heaviside step function. Since Pe_T and Pe_L are two independent random variables, the marginal *pdf* of *WD* is given by the convolution of pdf_{Pe_T} and pdf_{Pe_L} as

$$pdf_{WD}(wd') = \frac{1}{B_{WD}} \int_{-\infty}^{+\infty} pdf_{Pe_T}\left(\frac{wd' - A_{WD} - C_{WD}Pe_L}{B_{WD}}\right) pdf_{Pe_L}(Pe_L)dPe_L \quad \text{(A6)}$$

Making use of (A5) and (A6) becomes

$$\begin{aligned} pdf_{WD}(wd') = \frac{1}{B_{WD}(h-g)(f-e)} \int_{-\infty}^{+\infty} \Bigg\{ & -H\left(\frac{wd' - A_{WD} - C_{WD}Pe_L}{B_{WD}} - h\right) \\ & \times H(Pe_L - e) + H\left(\frac{wd' - A_{WD} - C_{WD}Pe_L}{B_{WD}} - h\right) H(Pe_L - f) \\ & + H\left(\frac{wd' - A_{WD} - C_{WD}Pe_L}{B_{WD}} - g\right) H(Pe_L - e) \\ & - H\left(\frac{wd' - A_{WD} - C_{WD}Pe_L}{B_{WD}} - g\right) H(Pe_L - f) \Bigg\} dPe_L \end{aligned} \quad \text{(A7)}$$

The first integral appearing in (A7) can be evaluated as

$$\begin{aligned} I_{1,WD} &= \int_{-\infty}^{\infty} H\left(\frac{wd' - A_{WD} - C_{WD}Pe_L}{B_{WD}} - h\right) H(Pe_L - e)dPe_L \\ &= \int_{e}^{\infty} H\left(\frac{wd' - A_{WD} - C_{WD}Pe_L}{B_{WD}} - h\right) dPe_L \\ &= \frac{B_{WD}}{C_{WD}} \left(\frac{wd' - A_{WD} - C_{WD}e}{B_{WD}} - h\right) H\left(\frac{wd' - A_{WD} - C_{WD}e}{B_{WD}} - h\right) \end{aligned} \quad \text{(A8)}$$

All remaining terms in (A7) can be evaluated in a similar way, so that (A7) becomes

$$\begin{aligned} pdf_{WD}(wd') = \frac{1}{C_{WD}(h-g)(f-e)} \Bigg\{ & -\left(\frac{wd' - A_{WD} - C_{WD}e}{B_{WD}} - h\right) \\ & \times H\left(\frac{wd' - A_{WD} - C_{WD}e}{B_{WD}} - h\right) \\ & + \left(\frac{wd' - A_{WD} - C_{WD}f}{B_{WD}} - h\right) H\left(\frac{wd' - A_{WD} - C_{WD}f}{B_{WD}} - h\right) \\ & + \left(\frac{wd' - A_{WD} - C_{WD}e}{B_{WD}} - g\right) H\left(\frac{wd' - A_{WD} - C_{WD}e}{B_{WD}} - g\right) \\ & - \left(\frac{wd' - A_{WD} - C_{WD}f}{B_{WD}} - g\right) H\left(\frac{wd' - A_{WD} - C_{WD}f}{B_{WD}} - g\right) \Bigg\} \end{aligned} \quad \text{(A9)}$$

Finally, after some manipulations and making use of (A3), (A4), (A9) can be rewritten as

$$pdf_{WD}(wd) = \frac{1}{12\theta_{WD;Pe_L}\theta_{WD;Pe_T}} \sum_{i=0}^{1}\sum_{j=0}^{1} (-1)^{i+j} \xi_{wd,ij} H\left[\frac{\xi_{wd,ij}}{\theta_{WD;Pe_T}}\right] \quad \text{(A10)}$$

with $\xi_{wd,ij} = wd - \theta_{WD;0} + \sqrt{3}((-1)^i\theta_{WD;Pe_L} + (-1)^j\theta_{WD;Pe_T})$.

A2. Marginal pdf of LT

Within the range of variability of the uncertain parameters listed in Table 1, one can note that 93% of the total variance of *LT* can be retained upon truncating (15) up to the first order Legendre polynomials ξ_{Pe_T}introduced in (A1) and $\xi_{N_g}$and ξ_{r_k}defined as

$$\xi_{N_g} = \sqrt{3}\frac{2N_g - a - b}{b-a}; \quad \xi_{r_k} = \sqrt{3}\frac{2r_k - c - d}{d-c} \quad \text{(A11)}$$

where $N_g \in \Gamma_{N_g} = [a, b]$ and $r_k \in \Gamma_{r_k} = [c, d]$. Therefore, (15) can be simplified as

$$LT' = LT - \theta_{LT;0} = \theta_{LT;N_g}\xi_{N_g} + \theta_{LT;r_k}\xi_{r_k} + \theta_{LT;Pe_T}\xi_{Pe_T} \quad \text{(A12)}$$

where $\theta_{LT;0}$ is the mean of *LT*, $\theta_{LT;N_g}$, $\theta_{LT;r_k}$ and $\theta_{LT;Pe_T}$ are the coefficients of the Legendre polynomial of order one in N_g, r_k, and Pe_T, respectively.

Eq. (A12) can be rewritten as LT'

$$= A_{LT} + B_{LT}N_g + C_{LT}r_k + D_{LT}Pe_T \quad \text{(A13)}$$

where

$$\begin{aligned} A_{LT} &= -\sqrt{3}\left(\theta_{LT;N_g}\frac{a+b}{b-a} + \theta_{LT;r_k}\frac{c+d}{d-c} + \theta_{LT;Pe_T}\frac{h+g}{h-g}\right); \\ B_{LT} &= \frac{2\sqrt{3}\theta_{LT;N_g}}{b-a}; \quad C_{LT} = \frac{2\sqrt{3}\theta_{LT;r_k}}{d-c}; \quad D_{LT} = \frac{2\sqrt{3}\theta_{LT;Pe_T}}{h-g} \end{aligned} \quad \text{(A14)}$$

The *pdfs* of N_g and r_k are given respectively by

$$pdf_{N_g} = \frac{1}{b-a}[H(N_g - a) - H(N_g - b)];$$
$$pdf_{r_k} = \frac{1}{d-c}[H(r_k - c) - H(r_k - d)] \quad \text{(A15)}$$

Since N_g, r_k and Pe_T are independent random variables, the marginal *pdf* of *LT* is

$$\begin{aligned} pdf_{LT}(lt') = \frac{1}{B_{LT}} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} & pdf_{N_g}\left(\frac{lt' - A_{LT} - C_{LT}r_k - D_{LT}Pe_T}{B_{LT}}\right) \\ & \times pdf_{r_k}(r_k) pdf_{Pe_T}(Pe_T) dr_k dPe_T \end{aligned} \quad \text{(A16)}$$

Making use of (A5) and (A15) and following the same strategy adopted in Appendix A1, (A16) can be evaluated as

$$pdf_{LT}(lt) = \frac{1}{48\sqrt{3}\theta_{LT;r_k}\theta_{LT;Pe_T}\theta_{LT;N_g}} \sum_{k=0}^{1}\sum_{i=0}^{1}\sum_{j=0}^{1} (-1)^{k+i+j} (\xi_{lt,ijk})^2 H\left[\frac{\xi_{lt,ijk}}{\theta_{LT;N_g}}\right] \quad \text{(A17)}$$

with $\xi_{lt,ijk} = lt - \theta_{LT;0} + \sqrt{3}((-1)^i\theta_{LT;N_g} + (-1)^j\theta_{LT;r_k} + (-1)^k\theta_{LT;Pe_T})$

A3. Marginal pdf of LS

Within the range of variability of the uncertain parameters listed in Table 1, one can note that 95% of the total variance of *LS* can be retained upon truncating (15) up to the first order Legendre polynomials ξ_{Pe_L}, ξ_{Pe_T}introduced in (A1) and ξ_{N_g}, given by (A11). Therefore, (15) can be simplified as

$$LS' = LS - \theta_{LS;0} = \theta_{LS;Pe_L}\xi_{Pe_L} + \theta_{LS;Pe_T}\xi_{Pe_T} + \theta_{LS;N_g}\xi_{N_g} \quad \text{(A18)}$$

where $\theta_{LS;0}$is the mean of *LS*, $\theta_{LS;Pe_L}$, $\theta_{LS;Pe_T}$and $\theta_{LT;N_g}$are the coeffi-cients of the Legendre polynomial of order one in Pe_L, Pe_T and N_g respectively. Following the same procedure outlined in Appendix A2, we obtain the following expression for the marginal *pdf* of *LS*

$$pdf_{LS}(ls) = \frac{1}{48\sqrt{3}\theta_{LS;Pe_L}\theta_{LS;Pe_T}\theta_{LS;N_g}} \sum_{k=0}^{1}\sum_{i=0}^{1}\sum_{j=0}^{1} (-1)^{k+i+j} (\xi_{ls,ijk})^2 H\left[\frac{\xi_{ls,ijk}}{\theta_{LS;Pe_L}}\right] \quad \text{(A19)}$$

with $\xi_{ls,ijk} = ls - \theta_{LS;0} + \sqrt{3}((-1)^i\theta_{LS;Pe_L} + (-1)^j\theta_{LS;Pe_T} + (-1)^k\theta_{LS;N_g})$

A4. Marginal pdf of LY

Within the range of variability of the uncertain parameters listed in Table 1, one can note that 97% of the total variance of *LY* can be retained upon truncating (15) up to the first order Legendre polynomials ξ_{Pe_T}, $\xi_{N_g}$and ξ_{r_k}defined by (A1) and (A11) and to the second order Legendre polynomial

$$\xi_{r_k^2} = \frac{\sqrt{5}}{2}\left(3\left(\frac{2r_k - c - d}{d - c}\right)^2 - 1\right) \tag{A20}$$

Therefore, (15) can be written as

$$LY' = LY - \theta_{LY;0} = \theta_{LY;N_g}\xi_{N_g} + \theta_{LY;Pe_T}\xi_{Pe_T} + \theta_{LY;r_k}\xi_{r_k} + \theta_{LY;r_k^2}\xi_{r_k^2} \tag{A21}$$

where $\theta_{LY;0}$ is the mean of LY, $\theta_{LY;N_g}$, $\theta_{LY;Pe_T}$ and $\theta_{LY;r_k}$ are the coefficient of the Legendre polynomial of order one in N_g, Pe_T and r_k, respectively, and $\theta_{LY;r_k^2}$ is the coefficient of the Legendre polynomial of order two in r_k. Eq. (A21) can be rewritten as

$$LY' = A_{LY} + B_{LY}N_g + C_{LY}Pe_T + D_{LY}r_k + E_{LY}r_k^2 \tag{A22}$$

where

$$\begin{aligned} A_{LY} &= -\sqrt{3}\left(\theta_{LY;N_g}\frac{a+b}{b-a} + \theta_{LY;r_k}\frac{c+d}{d-c} + \theta_{LY;Pe_T}\frac{g+h}{h-g}\right) \\ &\quad + \theta_{LY;r_k^2}\frac{\sqrt{5}}{(d-c)^2}(c^2 + d^2 + 4cd); \\ B_{LY} &= \frac{2\sqrt{3}\theta_{LY;N_g}}{b-a}; \quad C_{LY} = \frac{2\sqrt{3}\theta_{LY;Pe_T}}{h-g}; \\ D_{LY} &= \frac{2\sqrt{3}\theta_{LY;r_k}}{d-c} - \frac{6\sqrt{5}\theta_{LY;r_k^2}(c+d)}{(d-c)^2}; \quad E_{LY} = \frac{6\sqrt{5}\theta_{LY;r_k^2}}{(d-c)^2} \end{aligned} \tag{A23}$$

Since N_g, r_k and Pe_T are independent random variables, the marginal *pdf* of LY is

$$\begin{aligned} pdf_{LY}(ly') = \frac{1}{B_{LY}}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} pdf_{N_g}\left(\frac{ly' - A_{LY} - C_{LY}Pe_T - D_{LY}r_k - E_{LY}r_k^2}{B_{LY}}\right) \\ \times pdf_{Pe_T}(Pe_T)pdf_{r_k}(r_k)dPe_T dr_k \end{aligned} \tag{A24}$$

Eq. (A24) can be evaluated as

$$\begin{aligned} pdf_{LY}(ly') = \frac{1}{C_{LY}(b-a)(h-g)(d-c)}\int_{-\infty}^{+\infty}\Big\{H[r_k - c] - H[r_k - d]\Big\} \\ \times\Big\{(Ar_k^2 + Br_k + C_1)H[Ar_k^2 + Br_k + C_1] - (Ar_k^2 + Br_k + C_2) \\ \times H[Ar_k^2 + Br_k + C_2] - (Ar_k^2 + Br_k + C_3)H[Ar_k^2 + Br_k + C_3] \\ + (Ar_k^2 + Br_k + C_4)H[Ar_k^2 + Br_k + C_4]\Big\}dr_k \end{aligned} \tag{A25}$$

where

$$\begin{aligned} A &= -\frac{E_{LY}}{B_{LY}}; \quad B = \frac{-D_{LY}}{B_{LY}}; \quad C_1 = \frac{ly' - A_{LY} - C_{LY}g - aB_{LY}}{B_{LY}}; \\ C_2 &= \frac{ly' - A_{LY} - C_{LY}g - bB_{LY}}{B_{LY}}; \quad C_3 = \frac{ly' - A_{LY} - C_{LY}h - aB_{LY}}{B_{LY}}; \\ C_4 &= \frac{ly' - A_{LY} - C_{LY}h - bB_{LY}}{B_{LY}} \end{aligned} \tag{A26}$$

Eq. (A25) can be rewritten as

$$pdf_{LY}(ly') = I_{1,LY} - I_{2,LY} - I_{3,LY} + I_{4,LY} \tag{A27}$$

where

$$I_{i,LY} = A\int_c^d (v - v_{1,i})(v - v_{2,i})H[A(v - v_{1,i})(v - v_{2,i})]dv \tag{A28}$$

and

$$v_{1,i} = \frac{-B - \sqrt{B^2 - 4AC_i}}{2A}; \quad v_{2,i} = \frac{-B + \sqrt{B^2 - 4AC_i}}{2A} \tag{A29}$$

Since $A > 0$ in our case, the integral expression (A28) can be evaluated as

$$\begin{aligned} \frac{I_{i,LY}}{A} &= H[v_{1,i} - c]H[d - v_{2,i}]\left(\frac{v_{2,i}^3 - v_{1,i}^3}{6} - v_{1,i}v_{2,i}\frac{v_{2,i} - v_{1,i}}{2}\right) \\ &\quad + \{H[v_{1,i} - d] + H[c - v_{2,i}] + H[v_{1,i} - c]H[d - v_{2,i}]\} \\ &\quad \times \left(\frac{d^3 - c^3}{3} - \frac{v_{1,i} + v_{2,i}}{2}(d^2 - c^2) + v_{1,i}v_{2,i}(d - c)\right) \\ &\quad + H[d - v_{2,i}]H[c - v_{1,i}]H[v_{2,i} - c]\left(\frac{d^3}{3} - \frac{v_{1,i} + v_{2,i}}{2}d^2 - \frac{v_{2,i}^2}{2}v_{1,i}\right. \\ &\quad \left. + \frac{v_{2,i}^3}{6} + v_{1,i}v_{2,i}d\right) + H[v_{1,i} - c]H[d - v_{1,i}]H[v_{2,i} - d] \\ &\quad \times \left(\frac{v_{1,i}^2}{2}v_{2,i} - \frac{v_{1,i}^3}{6} - \frac{c^3}{3} + \frac{v_{1,i} + v_{2,i}}{2}c^2 - v_{1,i}v_{2,i}c\right) \end{aligned} \tag{A30}$$

References

[1] Abarca E, Carrera J, Sanchez-Vila X, Dentz M. Anisotropic dispersive Henry problem. Adv Water Resour 2007;30(4):913–26. http://dx.doi.org/10.1016/j.advwatres.2006.08.005.

[2] Al-Bitar A, Ababou R. Random field approach to seawater intrusion in heterogeneous coastal aquifers: unconditional simulations and statistical analysis. In: Renard P, Demougeot-Renard H, Froidevaux R, editors. Proceedings of the fifth European conference for environmental applications. Geostatistics for environmental applications. New York: Springer Verlag; 2005. p. 232–47. http://dx.doi.org/10.1007/3-540-26535-X_20.

[3] Archer GEB, Saltelli A, Sobol IM. Sensitivity measures, anova-like techniques and the use of bootstrap. J Stat Comput Simul 1997;58:99–120. http://dx.doi.org/10.1080/00949659708811825.

[4] Bäck J, Nobile F, Tamellini L, Tempone R. Stochastic spectral Galerkin and collocation methods for PDEs with random coefficients: a numerical comparison. In: Hesthaven JS, Ronquist EM editors. Lecture notes in computational science and engineering. Spectral and high order methods for partial differential equations, vol. 76. Berlin Heidelberg: Springer-Verlag; 2011. p. 43–62.

[5] Ballio F, Guadagnini A. Convergence assessment of numerical Monte Carlo simulations in groundwater hydrology. Water Resour Res 2004;40:W04603. http://dx.doi.org/10.1029/2003WR002876.

[6] Bear J, Cheng AHD, Sorek S, Ouazar D, Herrera I. Seawater intrusion in coastal aquifers – concepts, methods, and practices. Kluwer Academic Publisher; 1999.

[7] Bear J, Cheng AHD. Modeling groundwater flow and contaminant transport. Springer; 2010.

[8] Boso F, de Barros FPJ, Fiori A, Bellin A. Performance analysis of statistical spatial measures for contaminant plume characterization toward risk-based decision making. Water Resour Res 2013;49:3119–32. http://dx.doi.org/10.1002/wrcr.20270.

[9] Bruggeman GA. Analytical solutions of geohydrological problems. Developments in water science. Amsterdam: Elsevier; 1999. p. 46.

[10] Chang CM, Yeh HD. Spectral approach to seawater intrusion in heterogeneous coastal aquifers. Hydrol Earth Syst Sci 2010;14:719–27. http://dx.doi.org/10.5194/hess-14-719-2010.

[11] Ciriello V, Di Federico V, Riva M, Cadini F, De Sanctis J, Zio E, Guadagnini A. Polynomial chaos expansion for global sensitivity analysis applied to a model of radionuclide migration in a randomly heterogeneous aquifer. Stoch Environ Res Risk Assess 2013;27:945–54. http://dx.doi.org/10.1007/s00477-012-0616-7.

[12] Crestaux T, Le Maître O, Martinez JM. Polynomial chaos expansion for sensitivity analysis. Reliab Eng Syst Safety 2009;94(7):1161–72. http://dx.doi.org/10.1016/j.ress.2008.10.008.

[13] Dagan G, Zeitoun DG. Seawater–freshwater interface in a stratified aquifer of random permeability distribution. J Contam Hydrol 1998;29:185–203. http://dx.doi.org/10.1016/S0169-7722(97)00013-2.

[14] Dentz M, Tartakovsky DM, Abarca E, Guadagnini A, Sanchez-Vila X, Carrera J. Variable-density flow in porous media. J Fluid Mech 2006;561:209–35. http://dx.doi.org/10.1017/S0022112006000668.

[15] Fajraoui N, Ramasomanana F, Younes A, Mara T, Ackerer P, Guadagnini A. Use of global sensitivity analysis and polynomial chaos expansion for interpretation of nonreactive transport experiments in laboratory-scale porous media. Water Resour Res 2011;47:W02521. http://dx.doi.org/10.1029/2010WR009639.

[16] Fetter CW. Contaminant hydrogeology. 2nd ed. Upper Saddle River, New Jersey: Prentice Hall; 1999.

[17] Formaggia L, Guadagnini A, Imperiali I, Lever V, Porta G, Riva M, Scotti A, Tamellini L. Global sensitivity analysis through polynomial chaos expansion of a basin-scale geochemical compaction model. Comput Geosci 2013;17:25–42. http://dx.doi.org/10.1007/s10596-012-9311-5.

[18] Gautschi W. Orthogonal polynomials: computation and approximation. Oxford: Oxford University Press; 2004.
[19] Ghanem RG, Spanos PD. Stochastic finite elements: a spectral approach. Berlin: Springer; 1991.
[20] Held R, Attinger S, Kinzelbach W. Homogenization and effective parameters for the Henry problem in heterogeneous formations. Water Resour Res 2005;41:1–14. http://dx.doi.org/10.1029/2004WR003674.
[21] Henry HR. Effects of dispersion on salt encroachment in coastal aquifers. Water-Supply Paper 1613-C. US Geological Survey; 1964.
[22] Jiang Z, Raiber M, Bian J, Malcolm C. On the effective hydraulic conductivity and macrodispersivity for density-dependent groundwater flow. Environ Fluid Mech 2013;14:157–71. http://dx.doi.org/10.1007/s10652-013-9281-8.
[23] Kerrou J, Renard P. A numerical analysis of dimensionality and heterogeneity effects on advective dispersive seawater intrusion processes. Hydrol J 2010;18:55–72. http://dx.doi.org/10.1007/s10040-009-0533-0.
[24] Kempers LTJM, Haas H. The dispersion zone between fluids with different density and viscosity in a heterogeneous porous medium. J Fluid Mech 1994;267:299–324. http://dx.doi.org/10.1017/S0022112094001199.
[25] Kretz V, Berest P, Hulin JP, Salin D. An experimental study of the effects of density and viscosity contrasts on macrodispersion in porous media. Water Resour Res 2003;39(2):1032. http://dx.doi.org/10.1029/2001WR001244.
[26] Laloy E, Rogiers B, Vrugt JA, Mallants D, Jacques D. Efficient posterior exploration of a high-dimensional groundwater model from two-stage MCMC simulation and polynomial chaos expansion. Water Resour Res 2013;49:2664–82. http://dx.doi.org/10.1002/wrcr.20226.
[27] Li W, Lu Z, Zhang D. Stochastic analysis of unsaturated flow with probabilistic collocation method. Water Resour Res 2009;45:W08425. http://dx.doi.org/10.1029/2008WR007530.
[28] Le Maître OP, Knio OM. Spectral methods for uncertainty quantification. Scientific computation. Springer; 2010.
[29] Menand T, Woods AW. Dispersion, scale, and time dependence of mixing zones under gravitationally stable and unstable displacements in porous media. Water Resour Res 2005;41:W05014. http://dx.doi.org/10.1029/2004WR003701.
[30] Papoulis A. Probability, random variables, and stochastic process, 3rd ed., McGraw-Hill series in electrical engineering; 1991.
[31] Porta G, Lever V, Tamellini L, Riva M. Characterization of sedimentary basins through polynomial chaos expansion in the presence of mechanical and geochemical compaction. Water Resous Res, in press.
[32] Reilly TE, Goodman AS. Quantitative analysis of saltwater–freshwater relationships in groundwater systems – a historical perspective. J Hydrol 1985;80:125–60. http://dx.doi.org/10.1016/0022-1694(85)90078-2.
[33] Rubin Y. Applied stochastic hydrogeology. New York: Oxford University Press; 2003.
[34] Saltelli A, Ratto M, Tarantola S, Campolongo F. Sensitivity analysis practices: strategies for model-based inference. Reliab Eng Syst Saf 2006;91:1109–25. http://dx.doi.org/10.1016/j.ress.2005.11.014.
[35] Simpson MJ, Clement TP. Theoretical analysis of the worthiness of Henry and Elder problems as benchmarks of density-dependent groundwater flow models. Adv Water Resour 2004;26:17–31. http://dx.doi.org/10.1016/S0309-1708(02)00085-4.
[36] Sobol IM. Sensitivity estimates for nonlinear mathematical models. Math Model Comput Exp 1993;1:407–17.
[37] Sobol IM. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Math Comput Simul 2001;55:271–80. http://dx.doi.org/10.1016/S0378-4754(00)00270-6.
[38] Sudret B. Global sensitivity analysis using polynomial chaos expansions. Reliab Eng Syst Saf 2008;93:964–79. http://dx.doi.org/10.1016/j.ress.2007.04.002.
[39] Volker RE, Rushton KR. An assessment of the importance of some parameters for seawater intrusion in aquifers and a comparison of dispersive and sharp-interface modelling approaches. J Hydrol 1982;56:239–50.
[40] Voss CI, Provost AM. SUTRA, a model for saturate–unsaturated variable density ground-water with solute or energy transport. Reston USA: US Geological Survey Open-File Report 02-4231; 2002. p. 250.
[41] Werner AD, Bakker M, Post VEA, Vandenbohede A, Lu C, Ataie-Ashtiani B, Simmons CT, Barry DA. Seawater intrusion processes, investigation and management: recent advances and future challenges. Adv Water Resour 2013;51:3–26. http://dx.doi.org/10.1016/j.advwatres.2012.03.004.
[42] Xiu D, Karniadakis GEM. The Wiener–Askey polynomial chaos for stochastic differential equations. SIAM J Sci Comput 2002;24(2):619–44. http://dx.doi.org/10.1137/S1064827501387826.