A surrogate model for efficient quantification of uncertainties in multilayer shallow water flows

Alia Al-Ghosoun* **Nabil El Moçayd†** **Mohammed Seaid‡**

Abstract

In this study, we investigate the implementation of a Proper Orthogonal Decomposition (POD) Polynomial Chaos Expansion (PCE) POD-PCE surrogate model for the propagation and quantification of the uncertainty in hydraulic modelling. The considered model consists of a system of multilayer shallow water equations with a mass exchange between the layers and over stochastic beds. As a numerical solver, we propose a finite volume characteristics method that does not require eigenstructure of the system in its implementation. The method is fast, accurate and can be used for both slowly and rapidly hydraulic simulations. The propagation and influence of several uncertainty parameters are quantified in the considered numerical methods for multilayer shallow water flows. To reduce the required number of samples for uncertainty quantification, we combine the proper orthogonal decomposition method with the polynomial Chaos expansions for efficient uncertainty quantification of complex hydraulic problems with a large number of random variables. Numerical results are shown for several test examples including a dam-break problem over a flat bed, and a wind-driven recirculation flow on flat and non-flat bottoms. Results are also presented for the case study of a recirculation flow problem in the Strait of Gibraltar. The results demonstrate the robustness of the uncertainty quantification method compared to the standard Monte-Carlo simulations. The results presented in this study suggest that the use of surrogate modelling may save a considerable amount of the necessary computational cost for all the considered cases.

Keywords. Uncertainty quantification; Polynomial Chaos expansion; Proper orthogonal decomposition; Multilayer shallow water equations; Stochastic beds; Finite volume methods

1 Introduction

Modeling and numerical simulation of water flow at places such as the Mediterranean sea are of utmost importance, especially at the Strait of Gibraltar. At this place, an exchange between denser water from the Mediterranean sea and fresher water from the Atlantic Ocean occurs. Moreover, under the impact of the topography and wind forces, the hydrodynamics can be very complex [43]. Therefore, monitoring activities such as maritime transportations, management of pollution spills [34] and fishery [29] could be very challenging. Numerical modelling of such hydrodynamics would require the use of the full three-dimensional Navier-Stokes equations. As many studies reported, the transport and the occurrence of some eddies and small scale surface waves are driven by the longitudinal gradient of pressure, temperature and density. These hydraulic features can not be captured using the standard shallow water equations. However, with the introduction of multilayer shallow water equations, the use of computationally efficient numerical models is possible. This class of multilayer shallow water equations avoids the computationally demanding methods needed to solve the three-dimensional Navier-Stokes equations, but at the same time, they provide stratified flow velocities since the pressure distribution is still assumed to be hydrostatic. In the current

*Department of Engineering, University of Durham, South Road, Durham DH1 3LE, UK *E-mail:* alia.r.al-ghosoun@durham.ac.uk

†International Water Research Institute, University Mohammed VI Polytechnic, Benguerir, Morocco *E-mail:* nabil.elmocayd@um6p.ma

‡Department of Engineering, University of Durham, South Road, Durham DH1 3LE, UK *E-mail:* m.seaid@durham.ac.uk

study, the free-surface flow problem is approximated as a layered system made of multiple shallow water systems of different water heights but coupled through mass-exchange terms between the embedded layers. Recently, multilayer shallow water equations have been subject of various research studies [30, 7, 40] and have been used for modelling a variety of free-surface flows where water flows interact with the bed geometry and wind stresses. While general efforts have been concentrated to present efficient numerical tools that allow solving the deterministic problem, to the best of our knowledge, uncertainty quantification of these models has not been considered. In the present work, we consider these methods to model the exchange flow occurring at Strait of Gibraltar and quantify the uncertainty of the numerical response. Especially that many heterogeneous observational data are available at this Strait based on GPS, radar, remote sensing and in-situ measurements [43, 41, 36]. Furthermore, it has already been established that monitoring this area using numerical models coupled with altimetry observations could be very effective using data assimilation methods [20].

When using numerical algorithms such as data assimilation, the forecast of the model is predicted based on reduced dynamic model uncertainty. Therefore, uncertainty is ubiquitous to the results of these simulations for different reasons, see for example [56]. One of the major sources of uncertainty especially in hydraulic computations comes from calibration [72, 47]. In fact, the friction coefficient is a proxy that represents the hydraulic energy dissipation. This occurs on a molecular scale. Therefore, in fluid dynamic computations, it is generally estimated using empirical equations with some coefficients to be estimated. Moreover, the use of multilayer shallow water equations introduces a viscosity coefficient which models the numerous exchange of information between layers. This coefficient is purely used for modelling purposes and in practice, these parameters are estimated by means of calibration (or inverse problem) where data for a specific event is used, see for example [44, 62, 67]. Indeed, exact estimations are not possible for all ranges of variation in hydraulics for which modelling uncertainties is required. Many studies showed that uncertainty in the estimation of the mentioned parameters can lead to high uncertainty in the hydraulic states for some flow configuration [59, 28, 15, 37]. Furthermore, there is also a natural variability due to the imperfect description of boundary conditions which lead to stochastic uncertainties [57]. This has been demonstrated in the case of hydraulic simulations over unknown topography [23]. For instance, the description of the Strait topography could not be known everywhere. This lack of information due to measurement limitations could lead to uncertain hydraulic simulation, as reported by [18]. With the high availability of remote sensing data, the possibility to reduce the uncertainty of hydraulic simulation propagated from uncertain bathymetry is now possible, we refer the reader to [73, 8] for more details.

One major problem when tackling uncertainty quantification (UQ) with hydraulic computations is the *curse of dimensionality*. The topography of the Strait, the river, or the lake is often presented as a stochastic process regarding its space dependency. Moreover, the computed results for the water velocity and water height also depend on space and time. Therefore, using classical methods for uncertainty quantification could be very time consuming. Note that in uncertainty quantification, it is very common to use surrogate models also known as meta-models or response surfaces. Their purpose is to mimic the behavior of the true model while being less time consuming. Such methodology was successfully used in the context of hydraulic computations using polynomial Chaos expansion [28] or Kriging [59]. However, these studies in [28, 59] have considered only single value parameters (random variables) such as Manning Coefficient to be responsible for uncertainty in the model simulation, see for example [35] among others. The bathymetry, given its space dependency, has rarely been considered for an uncertainty quantification problem. In fact, one needs to reduce the dimensionality of the field before using any surrogate model in [32, 51] as the problem of dimensionality could be very challenging when using surrogate models. This is mainly the reason why only a few works have been dedicated to quantifying its uncertainty [38]. In the present work, we consider a Karhunen-Loéve decomposition in order to reduce the dimensionality of the bathymetry. This method relies on a spectral decomposition of the spatial correlation matrix of the bathymetry to approximate the real stochastic process in the mean-square sense. Another advantage of this method resides in the fact that it is very easy to sample once the decomposition is achieved. This method has been widely used in the context of a random process, see [60] for a review. Reducing the space of the input variables makes it very

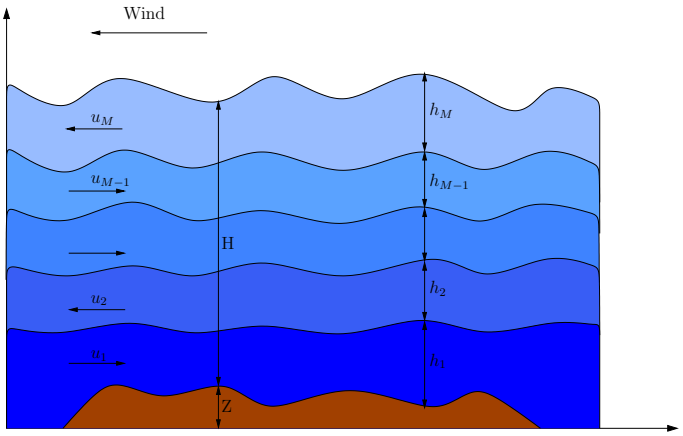

Figure 1: Illustration of the multilayer shallow water system.

easy to build surrogate models. In the present work, we consider the polynomial Chaos expansion, given its success rate for quantifying the uncertainty in hydraulic simulations, see [35, 27, 54]. Furthermore, in order to reduce the dimensionality of the output results (water height and velocity in each layer) we use Proper Orthogonal Decomposition (POD). This method avoids building a single surrogate model for each numerical component of the vector. Instead, a spectral decomposition is carried out and we suppose that the eigenvalues yield the uncertainty that ought to be propagated. Therefore, the surrogate model is built only on the eigenvalues. Such methods have been successfully used in uncertainty quantification since they allow to reduce the output dimension and yield some high accuracy given the uncertainty quantification, see for example [12, 55, 52].

In the present study, we evaluate the uncertainty of the hydraulic states such as the water height and the water velocity resulting from the bathymetric forces which include bed topography, Manning coefficient, wind friction and viscosity between the layers. The purpose here is to assess the UQ methodology over a real complex case. Moreover, as the sensitivity analysis of the uncertainty towards each parameter is performed, especially in locations where data measurements are available. It should also be pointed out that the considered multilayer shallow water model has a considerable impact on the reduced-order method compared to the computationally demanding Navier-Stokes equations for free-surface flows. The rest of this paper is organized as follows. The equations for the multilayer shallow water flows are presented in Section 2. The formulation of the finite volume modified method of characteristics for the numerical solution of multilayer models is also presented in this section. The general methodology used to address the uncertainty quantification problem is described in section 4. Section 5 is devoted to numerical results for several test examples for free-surface water flows over stochastic topography. Our new approach is shown to enjoy the expected accuracy as well as robustness. Section 6 contains concluding remarks.

2 Multilayer shallow water equations for free-surface water flows

In the current work, we consider M layers of water bodies bounded vertically by fixed bottom topography and a free surface as illustrated in Figure 1. Based on the vertical P_0 discretization of the horizontal velocity, the equations for three-dimensional hydrostatic incompressible Navier-Stokes equations with free-surface yield the so-called multilayer shallow water equations, see [6, 5] for detailed derivations of these equations. The governing equations are similar to the single-layer shallow water equations with additional

terms for exchanging momentum between the layers formulated as

$$\begin{aligned} \frac{\partial H}{\partial t} + \sum_{\alpha=1}^{M} \frac{\partial}{\partial x}\left(l_\alpha H u_\alpha\right) &= 0, \\ \frac{\partial}{\partial t}\left(l_\alpha H u_\alpha\right) + \frac{\partial}{\partial x}\left(l_\alpha H u_\alpha^2 + \frac{1}{2} g l_\alpha H^2\right) &= -g l_\alpha H \frac{\partial z}{\partial x} + \mathcal{S}_\alpha^u + \mathcal{S}_\alpha^\mu + \mathcal{S}_\alpha^b + \mathcal{S}_\alpha^w, \end{aligned} \tag{1}$$

where $z(x)$ is the bed topography, $u_\alpha(t,x)$ the local water velocity for the αth layer, g the gravitational acceleration, $H(t,x)$ the water height of the whole flow system and l_α denotes the relative size of the αth layer with

$$l_\alpha > 0, \qquad \sum_{\alpha=1}^{M} l_\alpha = 1.$$

The water height $h_\alpha(t,x)$ of the αth layer is defined as a fraction of the total water height as

$$h_\alpha = l_\alpha H, \qquad \alpha = 1, \ldots, M.$$

In (1), the source term $\mathcal{S}_\alpha^u$ is related to the momentum exchanges between the layers which is defined through the vertical P_0 discretization of the flow as

$$\mathcal{S}_\alpha^u = u_{\alpha+1/2} G_{\alpha+1/2} - u_{\alpha-1/2} G_{\alpha-1/2}, \tag{2}$$

with the mass exchange terms $G_{\alpha-1/2}$ and $G_{\alpha+1/2}$ are calculated by

$$G_{\alpha-1/2} = \begin{cases} 0, & \text{if} \quad \alpha = 1, \\ \displaystyle\sum_{\beta=1}^{\alpha} \left(\frac{\partial\left(h_\beta u_\beta\right)}{\partial x} - l_\beta \sum_{\gamma=1}^{M} \frac{\partial\left(h_\gamma u_\gamma\right)}{\partial x} \right), & \text{if} \quad \alpha = 2, 3 \ldots, M, \end{cases} \tag{3}$$

$$G_{\alpha+1/2} = \begin{cases} \displaystyle\sum_{\beta=1}^{\alpha} \left(\frac{\partial\left(h_\beta u_\beta\right)}{\partial x} - l_\beta \sum_{\gamma=1}^{M} \frac{\partial\left(h_\gamma u_\gamma\right)}{\partial x} \right), & \text{if} \quad \alpha = 1, 2, \ldots, M-1, \\ 0, & \text{if} \quad \alpha = M, \end{cases} \tag{4}$$

and the interface velocities $u_{\alpha-1/2}$ and $u_{\alpha+1/2}$ are evaluated using an upwind procedure following the sign of the mass exchange terms (3) and (4) as

$$u_{\alpha-1/2} = \begin{cases} u_{\alpha-1}, & \text{if} \quad G_{\alpha-1/2} \geq 0, \\ u_\alpha, & \text{if} \quad G_{\alpha-1/2} < 0. \end{cases} \qquad u_{\alpha+1/2} = \begin{cases} u_\alpha, & \text{if} \quad G_{\alpha+1/2} \geq 0, \\ u_{\alpha+1}, & \text{if} \quad G_{\alpha+1/2} < 0. \end{cases} \tag{5}$$

The vertical kinematic eddy viscosity terms $\mathcal{S}_\alpha^\mu$ in (1) account for the friction between the neighboring layers as

$$\mathcal{S}_\alpha^\mu = \begin{cases} 2\nu \dfrac{u_2 - u_1}{\left(l_2 + l_1\right) H}, & \text{if} \quad \alpha = 1, \\ 2\nu \dfrac{u_{\alpha+1} - u_\alpha}{\left(l_{\alpha+1} + l_\alpha\right) H} - 2\nu \dfrac{u_\alpha - u_{\alpha-1}}{\left(l_\alpha + l_{\alpha-1}\right) H}, & \text{if} \quad \alpha = 2, 3, \ldots, M-1, \\ -2\nu \dfrac{u_M - u_{M-1}}{\left(l_M + l_{M-1}\right) H}, & \text{if} \quad \alpha = M, \end{cases} \tag{6}$$

where ν is the eddy viscosity. The external friction terms $\mathcal{S}_\alpha^b$ and $\mathcal{S}_\alpha^w$ in (1) are defined by

$$\mathcal{S}_\alpha^b = \begin{cases} -gn_b^2 \dfrac{u_1 \left|u_1\right|}{H^{1/3}}, & \text{if} \quad \alpha = 1, \\ 0, & \text{if} \quad \alpha = 2, 3, \ldots, M, \end{cases} \qquad \mathcal{S}_\alpha^w = \begin{cases} 0, & \text{if} \quad \alpha = 1, 2, \ldots, M-1, \\ \sigma_s^2 \rho_a \dfrac{w \left|w\right|}{H}, & \text{if} \quad \alpha = M, \end{cases} \tag{7}$$

with n_b is the Manning roughness coefficient at the bed, w the wind velocity at 10 m above the water surface, σ_s^2 the wind stress coefficient, and ρ_a the air density. Note that the internal friction term $\mathcal{S}_\alpha^\mu$ models the friction between neighboring layers in (1), the bed-friction forcing term $\mathcal{S}_\alpha^b$ is acting only on the lower layer, and the wind-driven forcing term $\mathcal{S}_\alpha^w$ is acting only on the upper layer. For simplicity in the presentation, the multilayer shallow water equations (1) are reformulated in a compact vector form as

$$\frac{\partial \mathbf{W}}{\partial t} + \frac{\partial \mathbf{F}(\mathbf{W})}{\partial x} = \mathbf{Q}(\mathbf{W}) + \mathbf{R}(\mathbf{W}), \tag{8}$$

where $\mathbf{W}$ is the vector of conserved variables, $\mathbf{F}$ the vector of flux functions, $\mathbf{Q}$ and $\mathbf{R}$ are the vectors of source terms

$$\mathbf{W} = \begin{pmatrix} H \\ Hu_1 \\ Hu_2 \\ \vdots \\ Hu_M \end{pmatrix}, \qquad \mathbf{F}(\mathbf{W}) = \begin{pmatrix} \displaystyle\sum_{\alpha=1}^{M} l_\alpha H u_\alpha \\ Hu_1^2 + \dfrac{1}{2} g H^2 \\ Hu_2^2 + \dfrac{1}{2} g H^2 \\ \vdots \\ Hu_M^2 + \dfrac{1}{2} g H^2 \end{pmatrix}, \qquad \mathbf{Q}(\mathbf{W}) = \begin{pmatrix} 0 \\ -gH \dfrac{\partial z}{\partial x} \\ -gH \dfrac{\partial z}{\partial x} \\ \vdots \\ -gH \dfrac{\partial z}{\partial x} \end{pmatrix},$$

$$\mathbf{R}(\mathbf{W}) = \begin{pmatrix} 0 \\ \dfrac{1}{l_1}\left(u_{3/2}G_{3/2} - gn_b^2 \dfrac{u_1 \left|u_1\right|}{H^{1/3}} + 2\nu \dfrac{u_2 - u_1}{(l_2 + l_1)\, H}\right) \\ \dfrac{1}{l_2}\left(u_{5/2}G_{5/2} - u_{3/2}G_{3/2} + 2\nu \dfrac{u_3 - u_2}{(l_3 + l_2)\, H} - 2\nu \dfrac{u_2 - u_1}{(l_2 + l_1)\, H}\right) \\ \vdots \\ \dfrac{1}{l_{M-1}}\left(u_{M-1/2}G_{M-1/2} - u_{M-3/2}G_{M-3/2} + 2\nu \dfrac{u_M - u_{M-1}}{(l_M + l_{M-1})\, H} - 2\nu \dfrac{u_{M-1} - u_{M-2}}{(l_{M-1} + l_{M-2})\, H}\right) \\ \dfrac{1}{l_M}\left(-u_{M-1/2}G_{M-1/2} - \sigma^2 \rho_a \dfrac{w \left|w\right|}{H} - 2\nu \dfrac{u_M - u_{M-1}}{(l_M + l_{M-1})\, H}\right) \end{pmatrix}.$$

It should be stressed that it is not straightforward to derive the exact expressions of the $(M+1)$ eigenvalues of (8) and it is also not evident that the system (8) is hyperbolic as its associated eigenvalues may become complex. In this situation, the multilayer shallow water equations (8) yield the so-called Kelvin-Helmholtz instability at the interface separating the water layers. In the current study, the asymptotic estimation of these eigenvalues reported in [6] is used to adjust the timestep size in the numerical simulations. Thus, we

suppose that all the velocities u_α are closed to the mean velocity u and a first-order approximation of the two barotropic eigenvalues gives

$$\lambda_{ext}^{\pm} = u_m \pm \sqrt{g \sum_{\alpha=1}^{M} h_\alpha} + \mathcal{O}\left(|u_\beta - u|^2\right)_{\beta=1,\dots,M}, \qquad u_m = \frac{\sum_{\alpha=1}^{M} h_\alpha u_\alpha}{\sum_{\alpha=1}^{M} h_\alpha}, \tag{9}$$

and a zeroth-order approximation of the $2(M-1)$ barotropic eigenvalues associated with $(M-1)$ interfaces gives

$$\lambda_{int}^{\pm,\alpha+\frac{1}{2}} = u \pm \sqrt{\frac{1}{2} g \sum_{\alpha=1}^{M} h_\alpha} + \mathcal{O}\left(|u_\beta - u|\right)_{\beta=1,\dots,M}, \qquad \alpha = 1, 2, \dots, M-1. \tag{10}$$

Notice that the eigenvalues (9)-(10) are approximations for eigenvalues of the original system using the water heights h_α instead of the total height H. This results in a system of $2M$ equations for which each layer has two eigenvalues.

3 A fast and accurate finite volume method

The lack of explicit analytical expressions of the eigenvalues and eigenvectors for the multilayer shallow water model (8) renders Riemann-solver finite volume methods inappropriate for solving these equations. These methods include the Roe, HLL, HLLC schemes and other Godunov methods from computational fluids dynamics, see [58, 31, 53]. In the current study, we propose a second-order Finite Volume Characteristics (FVC) method which does not require the calculation of the eigenvalues for the multilayer system and the selection of time steps is carried out using the asymptotic approximations (9)-(10). For the spatial discretization of (8), we discretize the spatial domain into control volumes $[x_{i-1/2}, x_{i+1/2}]$ with uniform size $\Delta x = x_{i+1/2} - x_{i-1/2}$, $x_{i-1/2} = i\Delta x$ and $x_i = (i+1/2)\Delta x$ is the center of the control volume. Integrating the equation (8) with respect to space over the control volume $[x_{i-1/2}, x_{i+1/2}]$, we obtain the following semi-discrete equations

$$\frac{d\mathbf{W}_i}{dt} + \frac{\mathcal{F}_{i+1/2} - \mathcal{F}_{i-1/2}}{\Delta x} = \mathbf{Q}(\mathbf{W}_i) + \mathbf{R}(\mathbf{W}_i), \tag{11}$$

where $\mathbf{W}_i(t)$ is the space average of the solution $\mathbf{W}$ in the control volume $[x_{i-1/2}, x_{i+1/2}]$ at time t, *i.e.*,

$$\mathbf{W}_i(t) = \frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} \mathbf{W}(t, x)\, dx,$$

and $\mathcal{F}_{i\pm1/2} = \mathbf{F}(\mathbf{W}_{i\pm1/2})$ are the numerical fluxes at $x = x_{i\pm1/2}$ and time t. Here, the time integration of (11) is performed using a second-order splitting method studied in [9]. Thus, to integrate the equations (8) in time we divide the time interval into N sub-intervals $[t_n, t_{n+1}]$ with length $\Delta t = t_{n+1} - t_n$ for $n = 0, 1, \dots, N$. We also use the notation W^n to denote the value of a generic function W at time t_n. The considered operator splitting method consists of three stages given by:

Stage 1:

$$\begin{aligned} \frac{\partial \mathbf{W}_i^*}{\partial t} &= \mathbf{R}(\mathbf{W}_i^*), \qquad t \in (t_n, t_n], \\ \mathbf{W}_i^*(t_n) &= \mathbf{W}_i(t_n). \end{aligned} \tag{12}$$

Stage 2:

$$\begin{aligned} \frac{\partial \mathbf{W}_i^{**}}{\partial t} + \frac{\partial \mathbf{F}(\mathbf{W}_i^{**})}{\partial x} &= \mathbf{Q}(\mathbf{W}_i^{**}), \qquad t \in (t_n, t_{n+1}], \\ \mathbf{W}_i^{**}(t_n) &= \mathbf{W}_i^{*}(t_n). \end{aligned} \tag{13}$$

Stage 3:

$$\begin{aligned} \frac{\partial \mathbf{W}_i^{***}}{\partial t} &= \mathbf{R}(\mathbf{W}_i^{***}), \qquad t \in (t_n, t_{n+1}], \\ \mathbf{W}_i^{***}(t_n) &= \mathbf{W}_i^{**}(t_{n+1}). \end{aligned} \tag{14}$$

The time integration is complete once a time stepping scheme is applied to the above three stages. It is clear that the nonlinear terms are dealt with in the first and third stages, whereas only linear terms are accounted for in the second stage of the splitting. To avoid the solution of linear systems of algebraic equations associated with implicit time stepping, we consider only explicit time integration methods for the stages (12)-(14). To this end, we use the explicit third-order Runge-Kutta method studied in [2]. Hence, the procedure to advance the solution of an ordinary differential equation of the structure (12) from the time t_n to the next time t_{n+1} can be carried out as

$$\begin{aligned} \mathcal{W}_i^{(1)} &= \mathbf{W}_i^n + \Delta t \mathbf{R}(\mathbf{W}_i^n), \\ \mathcal{W}_i^{(2)} &= \frac{3}{4}\mathbf{W}_i^n + \frac{1}{4}\mathcal{W}_i^{(1)} + \frac{1}{4}\Delta t \mathbf{R}(\mathcal{W}_i^{(1)}), \\ \mathbf{W}_i^{n+1} &= \frac{1}{3}\mathbf{W}_i^n + \frac{2}{3}\mathcal{W}_i^{(2)} + \frac{2}{3}\Delta t \mathbf{R}(\mathcal{W}_i^{(2)}), \end{aligned} \tag{15}$$

where we have dropped the asterisk of the variables in (12) for ease of notation. Note that the Runge-Kutta method (15) has been widely used for time integration of hyperbolic systems of conservation laws mainly because it can be interpreted as a convex combination of first-order Euler steps which exhibits strong stability properties. As a consequence, the Runge-Kutta method (15) is TVD, third-order accurate in time, and stable under the usual Courant-Friedrichs-Lewy (CFL) condition involving the asymptotic eigenvalues (9)-(10). The spatial discretization (11) is complete when a reconstruction of the numerical fluxes $\mathcal{F}_{i\pm1/2}$ and source terms $\mathbf{Q}(\mathbf{W}_i)$ and $\mathbf{R}(\mathbf{W}_i)$ are chosen. In general, the reconstruction of the numerical fluxes requires a solution of Riemann problems at the interfaces $x_{i\pm1/2}$, see for example [48]. From a computational viewpoint, this procedure is very demanding and may restrict the application of the method for which Riemann solutions are not available. On the other hand, Riemann-solver free finite volume methods such as the canonical Lax-Friedrichs and Rusanov schemes can also be used in (11). For the Lax-Friedrichs method, the numerical fluxes $\mathcal{F}_{i+1/2}$ in (11) are defined by

$$\mathcal{F}_{i+1/2}^n = \frac{1}{2}\left(\mathbf{F}(\mathbf{W}_{i+1}^n) + \mathbf{F}(\mathbf{W}_i^n)\right) + \frac{\Delta x}{2\Delta t}\left(\mathbf{W}_i^n - \mathbf{W}_{i+1}^n\right). \tag{16}$$

For the Rusanov method, the numerical fluxes $\mathcal{F}_{i+1/2}$ in (11) are

$$\mathcal{F}_{i+1/2}^n = \frac{1}{2}\left(\mathbf{F}(\mathbf{W}_{i+1}^n) + \mathbf{F}(\mathbf{W}_i^n)\right) + \frac{\lambda}{2}\left(\mathbf{W}_i^n - \mathbf{W}_{i+1}^n\right), \tag{17}$$

where λ is the Rusanov speed defined by

$$\lambda = \max_{1\le\alpha\le M-1}\left(\left|\lambda_{ext}^{+}\right|, \left|\lambda_{ext}^{-}\right|, \left|\lambda_{int}^{+,\alpha+\frac{1}{2}}\right|, \left|\lambda_{int}^{-,\alpha+\frac{1}{2}}\right|\right), \tag{18}$$

with $\lambda_{ext}^{\pm}$ and $\lambda_{int}^{\pm,\alpha+\frac{1}{2}}$ are the approximated eigenvalues in (9) and (10), respectively.

3.1 Discretization of the flux gradients and source terms

Our objective in the present work is to present a class of Finite Volume Characteristics (FVC) method that is simple, easy to implement, and accurately solves the equations (8) without relying on Riemann problem solvers or complicated techniques for well-balancing the discretizations of the gradient fluxes and the source terms. This objective is achieved by reformulating the multilayer system in an advective form and integrating the obtained system along the characteristics defined by the advection velocity. To reconstruct the numerical fluxes $\mathcal{F}^n_{i\pm1/2}$ in (11), we consider the method of characteristics applied to an advective version of the system (1). In practice, the advective form of the multilayer shallow water equations (8) is built such that the non-conservative variables are transported with a velocity field associated with each layer. Here, the multilayer shallow water equations (8), without accounting for the source term $\mathbf{R}(\mathbf{W})$, are reformulated in an advective form as

$$\begin{aligned} \frac{\partial H}{\partial t} + \left(\sum_{\alpha=1}^{M} l_\alpha u_\alpha\right)\frac{\partial H}{\partial x} &= -\sum_{\alpha=1}^{M} l_\alpha H \frac{\partial u_\alpha}{\partial x}, \\ \frac{\partial q_\alpha}{\partial t} + u_\alpha \frac{\partial q_\alpha}{\partial x} &= -q_\alpha \frac{\partial u_\alpha}{\partial x} - gH\frac{\partial}{\partial x}(H+z), \qquad \alpha = 1, 2\ldots, M, \end{aligned} \tag{19}$$

which can be rearranged in a compact form as

$$\frac{\partial U_\alpha}{\partial t} + \mathcal{U}_\alpha \frac{\partial U_\alpha}{\partial x} = S_\alpha(\mathbf{U}), \qquad \alpha = 0, 1, \ldots, M, \tag{20}$$

where $q_\alpha = Hu_\alpha$ is the water discharge, $\mathbf{U} = (U_0, U_1, \ldots, U_M)^T$, $\mathbf{S}(\mathbf{U}) = (S_0, S_1, \ldots, S_M)^T$ with

$$\mathbf{U} = \begin{pmatrix} H \\ q_1 \\ q_2 \\ \vdots \\ q_M \end{pmatrix}, \qquad \mathbf{S}(\mathbf{U}) = \begin{pmatrix} -\sum_{\alpha=1}^{M} l_\alpha H \frac{\partial u_\alpha}{\partial x} \\ -Hu_1\frac{\partial u_1}{\partial x} - gH\frac{\partial}{\partial x}(H+z) \\ -Hu_2\frac{\partial u_2}{\partial x} - gH\frac{\partial}{\partial x}(H+z) \\ \vdots \\ -Hu_M\frac{\partial u_M}{\partial x} - gH\frac{\partial}{\partial x}(H+z) \end{pmatrix},$$

and the advection velocity $\mathcal{U}_\alpha$ is defined as

$$\mathcal{U}_\alpha = \begin{cases} \sum_{\beta=1}^{M} l_\beta u_\beta, & \text{if} \quad \alpha = 0, \\ u_\alpha, & \text{if} \quad \alpha = 1, 2, \ldots, M. \end{cases} \tag{21}$$

Note that the case with $\alpha = 0$ does not refer to any layer in the system but to the global mass equation. It is only used here to formulate the compact advective form (20) for the whole system. The fundamental idea of the method of characteristics is to impose a regular grid at the new time level and to backtrack the flow trajectories to the previous time level. At the old time level, the quantities that are needed are evaluated by interpolation from their known values on a regular grid. For more discussions, we refer the reader to [66, 61] among others. Thus, the characteristic curves associated with the equation (20) are solutions of the initial-value problems

$$\begin{aligned} \frac{dX_{\alpha,i+1/2}(\tau)}{d\tau} &= \mathcal{U}_{\alpha,i+1/2}\Big(\tau, X_{\alpha,i+1/2}(\tau)\Big), \quad \tau \in [t_n, t_{n+1}], \\ X_{\alpha,i+1/2}(t_{n+1}) &= x_{i+1/2}, \qquad \alpha = 0, 1, \ldots, M. \end{aligned} \tag{22}$$

Note that $X_{\alpha,i+1/2}(\tau)$ are the departure points at time τ of a particle that will arrive at the gridpoint $x_{i+1/2}$ in time t_{n+1}. The method of characteristics does not follow the flow particles forward in time, as the Lagrangian schemes do, instead it traces backward the position at time t_n of particles that will reach the points of a fixed mesh at time t_{n+1}. By doing so, the method avoids the grid distortion difficulties that the conventional Lagrangian schemes have, see for instance [66] and further references are therein. In our simulations, we used the third-order Runge-Kutta method (15) for the solution of the initial-value problems (22). In general $X_{\alpha,i+1/2}(t_n)$ will not coincide with the spatial position of a gridpoint. Hence, once the characteristic curves $X_{\alpha,i+1/2}(t_n)$ are accurately calculated, a solution at the cell interface $x_{i+1/2}$ is reconstructed as

$$\mathbf{U}^{n+1}_{\alpha,i+1/2} = \mathbf{U}_\alpha\left(t_{n+1}, x_{i+1/2}\right) = \widetilde{\mathbf{U}}_\alpha\left(t_n, X_{\alpha,i+1/2}(t_n)\right), \tag{23}$$

where $\widetilde{\mathbf{U}}_\alpha\left(t_n, X_{\alpha,i+1/2}(t_n)\right)$ is the solution at the characteristic foot computed by a Lagrange-based interpolation from the gridpoints of the control volume where the departure point resides. It is worth mentioning that the proposed finite volume method is fully conservative by construction and the non-conservative system (19) is used only to compute the intermediate states for the numerical fluxes in (11).

Applied to the equations (20), the characteristic solutions are given by

$$\begin{aligned}
H^n_{i+1/2} &= \widetilde{H}^n_{i+1/2} - \frac{\Delta t}{\Delta x}\widetilde{H}^n_{i+1/2}\sum_{\alpha=1}^{M} l_\alpha\left(u^n_{\alpha,i+1} - u^n_{\alpha,i}\right), \\
q^n_{\alpha,i+1/2} &= \widetilde{q}^n_{\alpha,i+1/2} - \frac{\Delta t}{\Delta x}\left(\widetilde{q}^n_{\alpha,i+1/2}\left(u^n_{\alpha,i+1} - u^n_{\alpha,i}\right) + g\widetilde{H}^n_{i+1/2}\left((H^n_{i+1} + z^n_{i+1}) - (H^n_i + z^n_i)\right)\right),
\end{aligned}$$

where

$$\widetilde{H}^n_{i+1/2} = H\left(t_n, X_{0,i+1/2}(t_n)\right), \qquad \widetilde{q}^n_{\alpha,i+1/2} = q_\alpha\left(t_n, X_{\alpha,i+1/2}(t_n)\right),$$

are the solutions at the characteristic foot computed by interpolation from the gridpoints of the control volume where the departure points $X_{\alpha,i+1/2}(t_n)$ belong. The numerical fluxes $\mathcal{F}_{i\pm1/2}$ in (11) are calculated using the intermediate states $\mathbf{W}^n_{i\pm1/2}$ recovered accordingly from the characteristic solutions $\mathbf{U}^n_{j,i\pm1/2}$ in (23). Hence, the corrector stage in the FVC method reduces to

$$\begin{aligned}
H^{n+1}_i &= H^n_i - \frac{\Delta t}{\Delta x}\sum_{\alpha=1}^{M}\left((l_\alpha H u_\alpha)^n_{i+1/2} - (l_\alpha H u_\alpha)^n_{i-1/2}\right), \\
q^{n+1}_{\alpha,i} &= q^n_{\alpha,i} - \frac{\Delta t}{\Delta x}\left(\left(Hu^2_\alpha + \frac{1}{2}gH^2\right)^n_{i+1/2} - \left(Hu^2_\alpha + \frac{1}{2}gH^2\right)^n_{i-1/2}\right) - \frac{\Delta t}{\Delta x}g\widehat{H}^n_i\left(z^n_{i+1} - z^n_{i-1}\right),
\end{aligned} \tag{24}$$

In our FVC method, the reconstruction of the term $\widehat{H}^n_i$ in (24) is carried out such that the discretization of the source terms is well balanced with the discretization of flux gradients using [7]

$$\widehat{H}^n_i = \frac{1}{4}\left(H^n_{i+1} + 2H^n_i + H^n_{i-1}\right). \tag{25}$$

In summary, the implementation of FVC method to solve the multilayer shallow water equations (8) is carried out following the steps in Algorithm 1.

4 Uncertainty quantification methods

In general, the purpose of uncertainty quantification is to identify the main sources of uncertainty in a physical model (*e.g.* parameters, external forcing, boundary conditions, initial conditions), and to quantify their impact on the quantities of interest simulated by the numerical model (prognostic variables, probability of

Algorithm 1 FVC method solving the multilayer shallow water equations.

Given the solution $\left(H_i^n, q_{\alpha,i}^n\right)$ at time t_n, we compute the solution $\left(H_i^{n+1}, q_{\alpha,i}^{n+1}\right)$ at the next time level t_{n+1} via:

1: Perform the first step of the time splitting in (12) to compute the solutions H_i^* and $q_{\alpha,i}^*$, $\alpha = 1, 2, \ldots, M$.
2: Compute the departure points $X_{\alpha,i+1/2}(t_n)$, with $\alpha = 0, 1, \ldots, M$ by solving (22).
3: Compute the approximations

$$\widetilde{H}_{i+1/2}^n = H\left(t_n, X_{0,i+1/2}(t_n)\right) \quad \text{and} \quad \widetilde{q}_{\alpha,i+1/2}^n = q_\alpha\left(t_n, X_{\alpha,i+1/2}(t_n)\right), \qquad \alpha = 1, \ldots, M$$

employing an interpolation procedure.
4: Evaluate the intermediate states $H_{i+1/2}^n$ and $q_{\alpha,i+1/2}^n$ from the predictor stage (24).
5: Update the solutions H_i^{n+1} and $q_{\alpha,i}^{n+1}$ using the corrector stage (24).

exceeding the threshold). This allows to associate every forecast with a level of confidence since the accuracy of a simulation significantly depends on the quantity and the quality of the input data. Therefore, to better understand the results of numerical simulations it is necessary to take into account these uncertainties in the simulations. In addition, a problem of uncertainty quantification is a problem which aims at estimating uncertainty on the outputs of a numerical simulation according to the uncertainties on knowledge of its input parameters. Because of the random nature of uncertainty, the probabilistic approach to deal with a problem of uncertainty quantification is to consider the uncertain data of the model as random variables or random processes and to reconsider the real deterministic numerical model as a stochastic model. This section presents the methods used in the current study for uncertainty quantification. First, the methods of generating different independent realizations of the bathymetric field are presented. Therefore, the classical method of Polynomial Chaos Expansion (PCE) is used as surrogate tools to alleviate the computational cost. Finally, as the hydraulic state is also considered as a stochastic process, we suggest to reduce the dimension of the problem using the Proper Orthogonal Decomposition (POD) and to compute a PCE only for the associated nonphysical variables.

4.1 Karhunen-Loève expansion for stochastic process

The Karhunen-Loève expansion (KL) allows to model a random process [69] based on a spectral decomposition of its spatial covariance matrix $\mathcal{C}(x, x')$, see for example [42, 49]. Hence, a stochastic process can be defined as

$$\boldsymbol{z}(x, \boldsymbol{\omega}) = \overline{\boldsymbol{z}}(x) + \sum_{i=0}^{\infty} \omega_i l_i \phi_i(x), \tag{26}$$

where x is the curvilinear abscissa, $\overline{\boldsymbol{z}}$ the mean of the random process, $\boldsymbol{\omega} = \{\omega_1, \ldots, \omega_\infty\}$ a set of independent random variables, l_i $(i = 0, 1, \ldots)$ the set of the eigenvalues of the covariance function, and $\phi_i(x)$ $(i = 0, 1, \ldots)$ the set of the eigenfunctions. In practice and for computational reasons, the equation (26) is truncated at a certain degree d. The choice of this latter is often determined by a threshold ϵ from which the eigenvalues could be neglected *i.e.*

$$\frac{\sum_{i=0}^{d} l_i}{\sum_{i=0}^{\infty} l_i} > 1 - \epsilon.$$

Thus, the expansion (26) is replaced by

$$\mathbf{z}(x, \boldsymbol{\omega}) = \overline{\mathbf{z}}(x) + \sum_{i=0}^{d} \omega_i l_i \phi_i(x). \tag{27}$$

It should be stressed that the major limitation of the KL decomposition is the *a priori* knowledge of the covariance matrix. In the literature, this parameter is estimated empirically and therefore could be subject to uncertainty itself. This uncertainty is not considered in the present work and the reader may be referred to [64] for more discussions on this topic. This expansion is commonly used to model the uncertainty of a stochastic input parameter for two reasons: firstly the mean-square error of the finite representation (27) is minimized such that the equation (26) converges following the ℓ_2-norm. The second reason is related to the generation of random samples. In fact, making a realization of $\boldsymbol{z}(x, \boldsymbol{\omega})$ amounts to randomly draw the different ω_i following a defined probability density function. Those parameters are assumed to be independent random variables. Therefore, the KL decomposition offers a good representation of the input parameters when the covariance is known. In our study, the covariance matrix $\mathcal{C}(x, x')$ is supposed to be exponential. Indeed, authors in [73] showed that the uncertainty in the description of the bathymetric field followed this spatial correlation function. Thus, the correlation matrix is defined as

$$\mathcal{C}(x, x') = \sigma^2 \exp\left(-\frac{|x - x'|}{\theta}\right), \tag{28}$$

where σ and θ are the hyper-parameters of the covariance function with σ is the standard deviation of the process and θ the correlation length. Note that, calculating the eigenvalues and eigenvectors is well known by the Fredholm problem. There exist many algorithms that aim to solve this latter given a well-defined matrix, see for example [3] for more details. Once, l_i and $\phi_i(x)$ are known, the KL expansion is obtained in a straightforward manner. Therefore, the bathymetry could be written as described in (27) and could be sampled in order to consider its uncertainty in the hydraulic computations.

4.2 Polynomial Chaos expansions

The PCE has been intensively used as a surrogate model in the context of uncertainty quantification, see [35, 28] among others. It aims to reproduce the global behavior of the considered numerical models. Supposing that the input parameters of this model are represented by M independent random variables $\boldsymbol{\zeta} = \{\zeta_1, \zeta_2, \ldots, \zeta_M\}$ with finite variance well defined in a probabilistic space, the response $\mathbf{W}$ of this model is also random. Note that $\mathbf{W}$ is a vector-valued response as it represents the spatial variability of the hydraulic states. Considering that the expectation $\mathbb{E}[\|\mathbf{W}\|^2]$ is finite, the behavior of $\mathbf{W}$ could be reproducible following a polynomial decomposition namely the PCE [63] as

$$\mathbf{W}(x, \boldsymbol{\zeta}) = \sum_{i \in \mathbb{N}} \alpha_i(x) \Psi_i(\boldsymbol{\zeta}), \tag{29}$$

where Ψ_i is the multivariate polynomials that form the basis and chosen in such way they are orthonormal with respect to the associated probability density function of $\boldsymbol{\zeta}$, *i.e.* $\mathbb{E}[\Psi_i(\boldsymbol{\zeta})\Psi_j(\boldsymbol{\zeta})] = \delta_{ij}$ where δ_{ij} is the Kronecker symbol, compare for example [68, 71]. In (29), α_i are the unknown spectral coefficients of the decomposition to be determined. Again, the sum in (29) is truncated to a finite series as

$$\mathbf{W}(x, \boldsymbol{\zeta}) \approx \sum_{i \in \mathcal{I}} \alpha_i(x) \Psi_i(\boldsymbol{\zeta}), \tag{30}$$

where $\mathcal{I} \subset \mathbb{N}$ is the finite set of indices. The determination of a PCE is therefore conditioned by the estimation of the spectral coefficients α_i. There are many methods used in the literature to achieve this step and we refer to [70, 71, 46] for a review of these methodologies. In the current work, only the regression method is used and this choice comes from the fact that these methods coupled with compressed sensing techniques are advantageous when the stochastic dimension of the problem M is high (which is more likely to be the case hereafter the use of the KL decomposition). The regression method is based on solving a least-square (LS) minimization problem in some ℓ_2-norm to estimate the coefficients α_i, see for instance

[21, 10]. In practice, we begin by defining an error $\boldsymbol{\epsilon}$ as the distance between the model and the PCE for a finite set of randomly sampled input variables of size N_{ls} as

$$\boldsymbol{\epsilon} = \left\| \mathbf{W}(x, \boldsymbol{\Xi}) - \sum_{i \in \mathbb{N}} \alpha_i(x) \Psi(\boldsymbol{\Xi}) \right\|_2 \equiv \boldsymbol{\mathcal{W}} - \boldsymbol{\alpha}^\top \boldsymbol{\Psi}, \tag{31}$$

where $\boldsymbol{\Xi} = [\boldsymbol{\zeta}^{(1)}, \ldots, \boldsymbol{\zeta}^{(N_{ls})}]^\top$ is the set of considered realizations for the stochastic input variables $\boldsymbol{\zeta}$ and $\boldsymbol{\mathcal{W}} = [\mathbf{W}^{(1)}, \ldots, \mathbf{W}^{(N_{ls})}]^\top$ the vector of associated model outputs. We also define $\boldsymbol{\alpha} = [\alpha_0, \ldots, \alpha_{N_{PC}-1}]^\top$ as the vector of the $N_{PC} = Card(\mathcal{I})$ unknown coefficients and $\boldsymbol{\Psi}$ is the matrix of size $N_{PC} \times N_{ls}$ assembling the values of all orthonormal polynomials at the stochastic input realizations values $\boldsymbol{\Psi}_{ik} = \Psi_i(\boldsymbol{\zeta}^{(k)})$, with $i = 0, 1, \ldots, N_{PC-1}$ and $k = 1, 2, \ldots, N_{ls}$. Following the ordinary least-square (31), the estimation of the set of coefficients $\boldsymbol{\alpha}$ is equivalent to minimize the following function

$$\mathrm{J}(\boldsymbol{\alpha}) = \boldsymbol{\epsilon}^\top \boldsymbol{\epsilon} = \left(\boldsymbol{\mathcal{W}} - \boldsymbol{\alpha}^\top \boldsymbol{\Psi} \right)^\top \left(\boldsymbol{\mathcal{W}} - \boldsymbol{\alpha}^\top \boldsymbol{\Psi} \right), \tag{32}$$

which leads to a standard well-known linear algebraic solution as

$$\boldsymbol{\alpha} = (\boldsymbol{\Psi}^\top \boldsymbol{\Psi})^{-1} \boldsymbol{\Psi}^\top \boldsymbol{\mathcal{W}}. \tag{33}$$

Here, the input space exploration is fulfilled using a Monte-Carlo sampling-based approach [11, 28]. It is worth mentioning that the number of coefficients N_{PC} needed is directly affected by the stochastic dimension M as well as the polynomial degree p. As a consequence, the Monte-Carlo size N_{ls} will also increase significantly with M and p. This is a classical problem of PCE also known as *the curse of dimensionality*. The adaptive Least Angle Regression (LAR) method [26] has been introduced and used to overcome this specific problem [11]. This method, introduced in the context of compressed sensing, has made it possible to give the best polynomial representation given a fixed budget of finite simulations. It will be used in the present work as it has been demonstrated to be very efficient in this framework. The idea behind this method is to select with an iterative manner an optimal sparse basis among the original one and then to compute a limited number of coefficients using a standard regression method. In the context of PCE, a hybrid LAR method is used. The method consists on using LAR to evaluate the best set of predictors among the full basis elements with the help of a cross-validation method. These coefficients are estimated using the classical least square method, see [11] for more details. The infinite expansion (29) describing PCE converges with respect to the standard ℓ_2-norm known by the mean-squared convergence, see for example [70]. However, due to the errors of truncation and spectral coefficient estimation, the accuracy of this expansion must be evaluated in the same error norm. There are many different error metrics that allow to assess the accuracy of the PCE, see [28, 70] among others. In the present work, all the PCEs are assessed using the Leave-one-out (LOO) error. This method avoids integrating the model over another set of validation samples and therefore the computational cost is optimized. It has been introduced in this context with the introduction of sparse PCE [10] and the LOO technique required the formulation of several surrogates [25, 11]. Each surrogate model is built excluding one point out of the input sample and the accuracy of the surrogate model is then quantified at this particular point. Following this theory, the error ϵ_{LOO} is defined by

$$\epsilon_{LOO} = \frac{\sum_{k=1}^{N_{ls}} \left(\mathbf{W}^{(k)} - \mathbf{W}^{(-k)} \right)^2}{\sum_{k=1}^{N_{ls}} \left(\mathbf{W}^{(k)} - \overline{\mathbf{W}} \right)^2}, \tag{34}$$

where $\overline{\mathbf{W}}$ denotes the sample-averaged model simulations and $\mathbf{W}^{(-k)}$ stands for the evaluation of the PCE at $\boldsymbol{\zeta}^{(k)}$ when the surrogate has been built using an experimental design in which $\boldsymbol{\zeta}^{(k)}$ was excluded. In the present work, the determination of the optimal polynomial degree is performed using an iterative method.

Algorithm 2 The implementation of the stochastic POD procedure.

1: Define the covariance matrix (36).
2: Expand the matrix $\mathbf{C}$ using an SVD algorithm to determine λ_i and $\phi_i(x)$.
3: Retain only the first k eigenvalues and eigenvectors in the expansion using the condition (35).
4: Reconstruct the stochastic solutions $\mathbf{W}(x, \boldsymbol{\zeta})$ using (35).

Thus, a PCE is computed for different degrees varying from 1 to 20 and the optimal degree is determined based on the value of the corresponding ϵ_{LOO}. For the same value of a given error, the lowest value of the degree is retained.

Note that, when dealing with numerical models with spatial or temporal dependence, the classical way used in numerical simulations is to discretize the physical domain and resolve the governing equations for each control volume. This procedure leads to multiple decompositions to ensure numerical convergence. As a result, uncertainty quantification becomes computationally very demanding not only because of the stochastic dimensions but also because of the spatial or temporal dimensions of the output variables. For these reasons, many methods have been introduced in the literature in order to reduce the dimensionality of the output vector, see for instance [55, 14, 19, 13]. Among all these different techniques, the POD has proven to be efficient in the context of physical fields with rapid variability as in hydraulics [55].

4.3 Stochastic proper orthogonal decomposition

The proper orthogonal decomposition (POD) is a well-established technique that allows a high-dimensional system to be approached by a low-dimensional one, compare [22] among others. This method consists of determining a basis of orthogonal eigenvalues representative of the simulated physics. Notice that the POD is used in uncertainty quantifications to reduce the size of a random vector at the output of the model. The uncertainty is therefore carried out over each direction defined by the eigenvectors $\phi_i(x)$. The idea is based on projecting the solution $\mathbf{W}$ of the model into a finite and orthonormal basis $\{\phi_i, i \in \mathcal{I}_{POD}\}$, where $\mathcal{I}_{POD}$ is a discrete finite set of indices. Thus, the process $\mathbf{W}(x, \boldsymbol{\zeta})$ is decomposed as

$$\mathbf{W}(x, \boldsymbol{\zeta}) = \sum_{i \in \mathcal{I}_{POD}} \widehat{\lambda}_i(\boldsymbol{\zeta}) \phi_i(x). \tag{35}$$

The estimation of $\{\phi_i,\ i \in \mathcal{I}_{POD}\}$ is performed by decomposing the covariance matrix as

$$\mathbf{C} = \frac{1}{N_{ls}} \mathbf{W}\mathbf{W}^\top. \tag{36}$$

Indeed, the literature is rich in techniques that aim to decompose a covariance matrix. One of the most known methods is the Singular Value Decomposition (SVD) algorithm [24]. Hence, we define a POD-truncated error ϵ such as only the most k invaluable eigenvectors are retained as

$$\frac{\dfrac{\sum_{i=0}^{k} \lambda_i}{N_{ls}}}{\sum_{i=0} \lambda_i} > 1 - \epsilon, \tag{37}$$

where $\widehat{\lambda}_i$ is the mean value of $\lambda_i(\boldsymbol{\zeta})$. In summary, the stochastic POD procedure can be implemented using the steps described in Algorithm 2. It is worth remarking that the selection of the convergence criterion ϵ is problem-dependent. The selection criterion of ϵ for test examples in the present investigation is discussed in Section 5 where numerical examples are described.

Once the stochastic POD is reconstructed, the eigenvalues are considered as stochastic independent variables (since the eigenvectors form a basis). This means that we can define a PCE for each eigenvalue following the same manner as described in the previous section on polynomial chaos expansions leaving the spatial dependence described by the eigenvectors $\phi_i(x)$ as

$$\lambda_i(\boldsymbol{\zeta}) = \sum_{j=0}^{N_{PC}} \gamma_j \Psi_j(\boldsymbol{\zeta}), \tag{38}$$

where γ_j are the corresponding spectral coefficients. Hence, using (38), the equation (30) reduces to

$$\mathbf{W}(x, \boldsymbol{\zeta}) = \sum_{i \in \mathcal{I}_{POD}} \left(\sum_{j=0}^{N_{PC}} \gamma_j \Psi_j(\boldsymbol{\zeta}) \right) \phi_i(x). \tag{39}$$

Figure 2 summarizes both algorithms that we consider in the present work for quantification of uncertainties in multilayer shallow water flows. Note that the classical way to deal with a problem of uncertainty quantification using the PCE is performed when a decomposition is achieved for each node in the computational mesh. This method is illustrated by steps 1, 2, and 3 in Figure 2. This method is computationally demanding, as a surrogate model is built for each nodal point of the computational mesh. However, as mentioned before, one method to alleviate the spatial distribution is to make the first reduction using the POD. The eigenvalues of the POD are supposed to retrieve the stochasticity expressed by the stochastic process, while the eigenvectors express the spatial dependence. Once the POD is carried out, only a few eigenmodes are retained enabling to make less PCE than its conventional counterpart. This algorithm follows steps 1, 2, 4, and 5 in Figure 2.

5 Numerical results and examples

In this section we present numerical results for several test problems of shallow water flows using multilayer shallow water equations over flat and non-flat beds. The main goal of this section is to illustrate the numerical performance of the techniques described above for the uncertainty quantification in multilayer shallow water flows. In our simulations, the Courant number is set to $Cr = 0.85$ and the time stepsize Δt is adjusted at each time step according to the CFL stability condition as

$$\Delta t = Cr \frac{\Delta x}{\max\limits_{1 \le \alpha \le M-1} \left(\left|\lambda_{ext}^{+}\right|, \left|\lambda_{ext}^{-}\right|, \left|\lambda_{int}^{+,\alpha+\frac{1}{2}}\right|, \left|\lambda_{int}^{-,\alpha+\frac{1}{2}}\right| \right)}, \tag{40}$$

where $\lambda_{ext}^{\pm}$ and $\lambda_{int}^{\pm,\alpha+\frac{1}{2}}$ are the approximated eigenvalues in (9) and (10), respectively. In all the computations reported herein, the total water height H and the total number of layers M are given whereas, the water height h_α at each αth layer is equidistantly calculated as

$$h_\alpha = l_\alpha H, \qquad \text{with} \qquad l_\alpha = \frac{1}{M}, \qquad \alpha = 1, 2, \ldots, M.$$

We also present the two-dimensional velocity fields for the considered examples using the post-processing procedure described in [4, 7]. Here, the vertical velocity v is computed using the divergence-free condition from the incompressible Navier-Stokes equations as

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial z} = 0. \tag{41}$$

To compute the vertical velocity v, we integrate the equation (41) for each layer assuming non-penetration boundary conditions at the bottom, see [4, 7] for more details on this procedure. In this section, we also compare the results obtained using the proposed FVC scheme to those obtained using the well-established Lax-Friedrichs method and the Kinetic method investigated in [4].

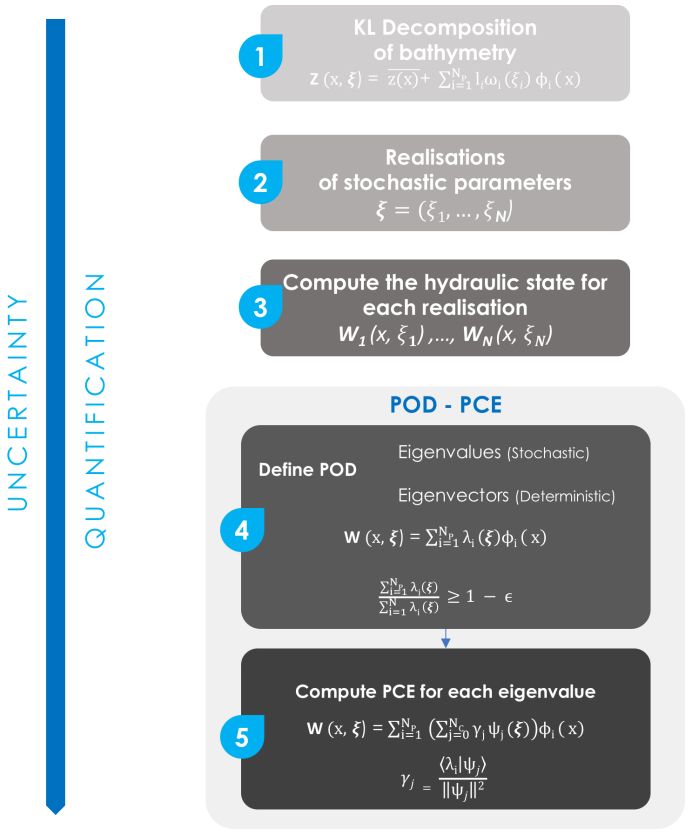

Figure 2: Schematic representation of the difference between flowcharts for the classical PCE based surrogate model and the POD-PCE based surrogate model.

5.1 Dam-break problem over a flat bed

For this test example, the uncertainty quantification is performed for a dam-break flow problem over a stochastic flat bed. The channel is 100 m long and subject to the following initial conditions

$$H(0,x) = \begin{cases} 2, & \text{if} \quad x \leq 0, \\ 1, & \text{if} \quad x > 0, \end{cases} \qquad u_\alpha(0,x) = 0.$$

At the initial time $t = 0$, the dam breaks and the flow problem consist of a shock wave propagating downstream and a rarefaction wave propagating upstream. In this example, $g = 9.81\ m/s^2$, $\nu = 0.01$, $n_b = 0.05\ s/m^{1/3}$, the computational domain is discretized into 100 control volumes and numerical results are presented for 10 layers at time $t = 6\ s$. We suppose that the bathymetry field is a Gaussian process that could be expanded following the KL decomposition and the coefficient of variation is fixed to 30%. The aim here is to quantify the uncertainty on the water height and the vertical mean velocity using the considered numerical methods namely, the FVC, the Kinetic, and the Lax-Friedrich schemes.

Figure 3 illustrates the eigenvalues and eigenvectors obtained for the bathymetry process supposed to be a Gaussian process described with an exponential spatial autocorrelation function. Under the considered flow conditions, given a value of the threshold $\epsilon = 10^{-2}$, 14 eigenvalues are required to correctly sample the bathymetric field. In other words, the stochastic dimension of this problem is 16 (14 representing the bathymetry, one for the friction coefficient, and one for the viscosity coefficient). For each of the three numerical methods considered in this study, a set of 1000 Monte-Carlo simulations are performed and the POD is then used to reduce the output space. In this case, different modes are retained to correctly represent the water height and the water velocity fields. An iterative procedure is used to determine the

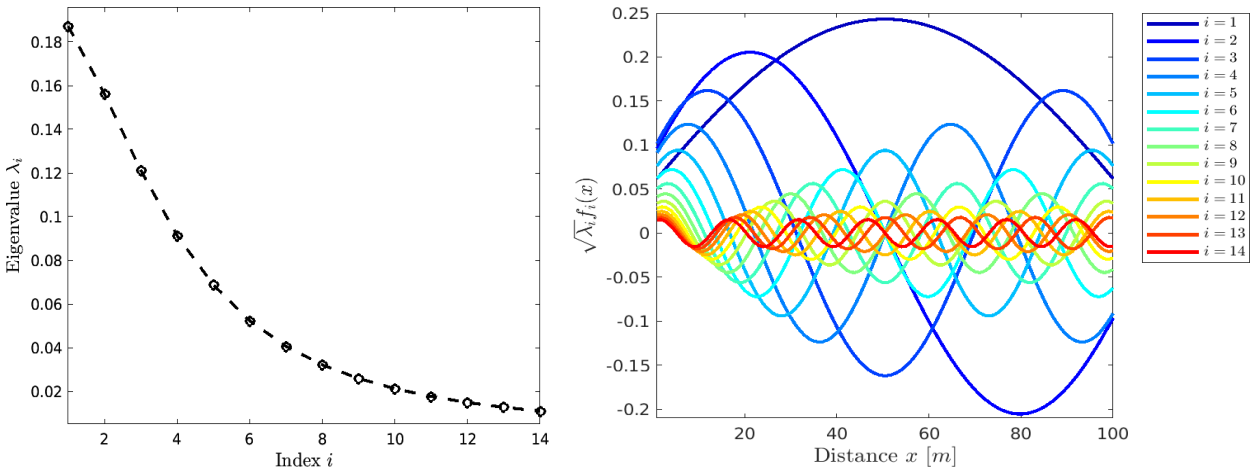

Figure 3: Eigenvalues (left plot) and the associated eigenfunctions (right plot) of the correlation matrix in the KL decomposition for the dam-break problem over a flat bed.

best polynomial degree along with the LOO error which is used to assess the robustness of the considered PCE. In order to further assess the quality of the surrogate model, its performance is compared to the full Monte-Carlo simulations fulfilled with a design of experiment with a set size fixed to 100000.

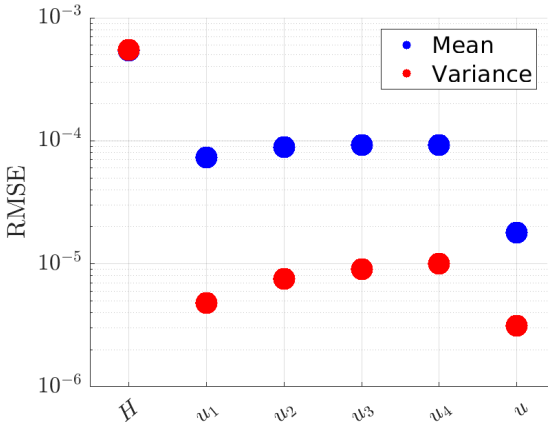

Figure 4: RMSE in the variance and the mean fields using the Lax-Friedrichs, Kinetic and FVC schemes for the water height and the mean velocity for the dam-break problem over a flat bed.

Table 1 summarizes the results obtained for the first and second POD modes. It is evident that a polynomial of degree 3 is enough to correctly reproduce the uncertainty translated from errors in the bathymetry for the three considered finite volume methods. For both output fields (water height or water velocity), the LOO error varies from 10^{-12} to 10^{-7} which makes the surrogate model very reliable for uncertainty quantification in this class of dam-break problems. Note that the LOO error reflects the accuracy of the polynomial Chaos expansion to estimate the uncertainty expressed by each mode. Therefore, it should not be mistaken with the accuracy of the surrogate model (polynomial Chaos expansion with a proper orthogonal decomposition) to estimate the physical values. For this purpose, the accuracy of the surrogate model is assessed using the full Monte-Carlo simulations. As each physical component has a spatial dependency, the RMSE is estimated for the water height, the averaged velocity and for the mean and the variance fields using the three considered numerical models. The obtained results are reported in Figure 4 and it is clear that the RMSE value varies from 10^{-4} to 10^{-3} confirming the fact that the surrogate model has a good accuracy to estimate the first and second statistical moments. Furthermore, a QQ-plot displaying the results from

Table 1: Best polynomial degrees with LOO errors for the POD modes in the water height and mean velocity using different numerical schemes for the dam-break problem over a flat bed.

	Error and polynomial degree for water height H		
	Lax-Friedrich scheme	Kinetic scheme	FVC scheme
Number of modes	5	6	6
Error (M1)	5.764E-10	4.768E-10	2.354E-7
Error (M2)	6.684E-10	1.674E-10	2.369E-7
Error (M3)	6.442E-10	3.186E-10	2.528E-7
Error (M4)	5.456E-9	3.750E-10	3.816E-7
Error (M5)	7.562E-9	5.300E-10	4.468E-7
Error (M6)	———	9.590E-10	1.557E-7
Polynomial (M1)	3	3	3
Polynomial (M2)	3	3	3
Polynomial (M3)	3	3	3
Polynomial (M4)	3	3	3
Polynomial (M5)	3	3	3
Polynomial (M6)	———	3	3

	Error and polynomial degree for water mean velocity u		
	Lax-Friedrich scheme	Kinetic scheme	FVC scheme
Number of modes	8	7	7
Error (M1)	1.705E-5	2.961E-5	1.621E-5
Error (M2)	8.588E-8	7.389E-8	4.110E-6
Error (M3)	1.995E-8	9.500E-8	9.001E-6
Error (M4)	2.736E-6	5.583E-6	7.860E-6
Error (M5)	4.047E-7	8.670E-8	1.111E-6
Error (M6)	4.276E-6	6.949E-6	1.339E-5
Error (M7)	1.080E-6	6.779E-7	2.960E-5
Error (M8)	1.145E-6	———	———
Polynomial (M1)	3	3	3
Polynomial (M2)	2	2	3
Polynomial (M3)	3	3	3
Polynomial (M4)	2	3	3
Polynomial (M5)	3	2	3
Polynomial (M6)	3	3	3
Polynomial (M7)	2	3	3
Polynomial (M8)	2	———	———

the surrogate model against the Monte-Carlo simulations is illustrated in Figure 5 for the water height and the averaged velocity. As it can be seen from this figure, all points of the domain are located in the first bisector of the QQ-plot which means that there is a good agreement between the results from the proposed model and the surrogate model. Hence, the surrogate model requiring 100 times fewer simulations than the Monte-Carlo simulations is able to provide accurate results for uncertainty quantification in this dam-break problem. In addition, the proper orthogonal decomposition allowed to reduce the dimensionality of the problem to only 8 modes for the velocity and 6 modes for the water height.

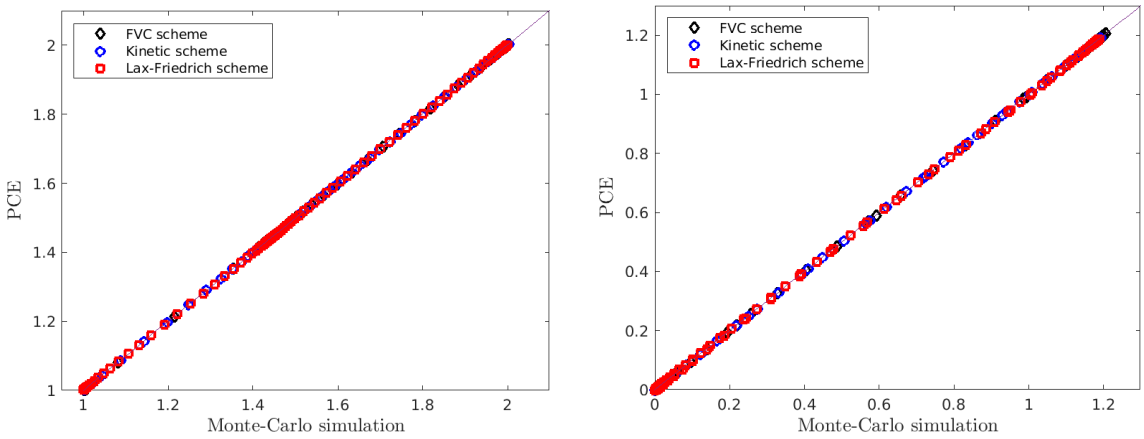

Figure 5: The QQ-plot between the Monte-Carlo simulations and the surrogate model for the estimation of the mean for the water height(left plot) and the velocity (right plot) for the dam-break problem over a flat bed.

Next, we turn our attention to the estimation of the first- and second-order moments using the surrogate model. The objective here is to present a comparison between the uncertainty displayed by each of the considered finite volume methods. First, we compare in Figure 6 the mean field and the deterministic solutions obtained at time $t = 6\ s$ using the considered numerical schemes. There are no significant changes between the deterministic solutions and the mean fields. The only differences observed are those already known from the differences between the finite volume methods for solving this class of problems. We also report in Figure 7 the mean solution of the water height along with its Confidence Interval (CI) that we define by taking the stochastic mean and adding and removing one standard deviation. As it can be seen from these results, the mean fields highlight some well-established results on the accuracy of the numerical methods such that the numerical diffusion is more pronounced for the Lax-Friedrichs method than for the FVC method. However, these methods do not display the same uncertainty for this test problem. In fact, the interval of confidence follows the physical feature of each variable namely, the water height and the water velocity. One can also notice that the FVC method displays less uncertainty than the other methods, especially around the shock area. For this reason, in the following we will consider only this method to solve the multilayer shallow water equations.

5.2 Wind-driven recirculation flow over a flat bed

In this example, we consider a class of wind-driven recirculation flow problems in closed domains. It should be pointed out that these recirculation features of the water flow can not be captured using the conventional single-layer shallow water equations. Hence, the present test is served as a prototype to verify the performance of multilayer shallow water flows to reproduce such phenomena. In this example, the multilayer system is solved in a closed channel 16 m long with a flat bottom filled at 2 m of water subject to a wind of speed $w = 20\ m/s$ blowing from the left to the right of the domain. In our simulations, $\rho_a = 1200\ kg/m^3$, $g = 9.81\ m/s^2$, $\nu = 0.1$, $n_b = 0.00001\ s/m^{1/3}$, $\sigma_s = 0.0015$, the computational domain is discretized into 16 control volumes and numerical results are presented for different numbers of layers at time $t = 20\ s$. Note that, due to the wind effects, the water flow changes the direction pointing towards the right boundary of the lake. The aim of this test example is to study the effects of the accounted uncertainty for the bathymetry, the bed friction, and the viscosity on the hydraulic states. We also investigate the impact of the number of layers used to capture the physical features on the uncertainty of the different physical variables. As in the previous example, the KL decomposition is used to sample the bathymetric field and the associated eigenvalues and eigenvectors are shown in Figure 8. The bathymetry process is supposed to be a Gaussian process described with 14 independent normal distributed random variables.

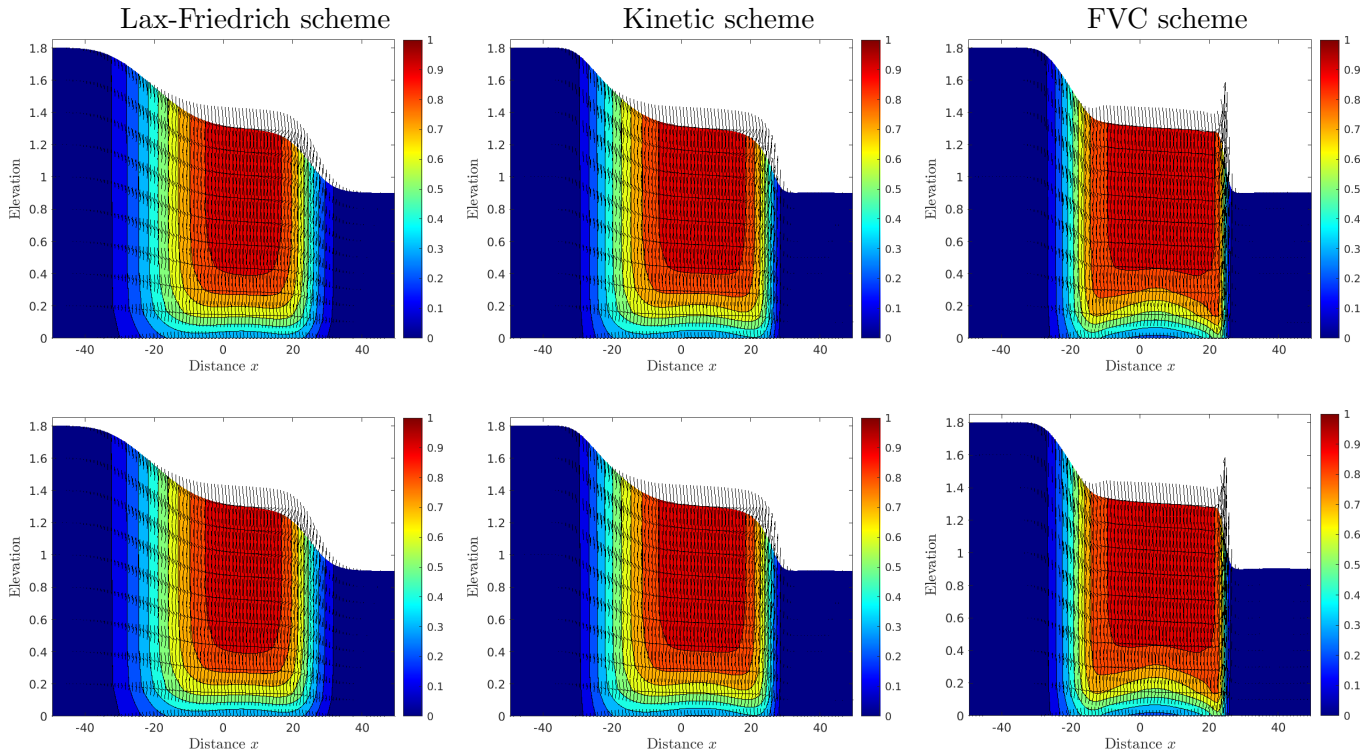

Figure 6: Deterministic flow fields (first row) and stochastic mean flow fields (second row) for the dam-break problem over a flat bed using Lax-Friedrich scheme (first column), Kinetic scheme (second column) and FVC scheme (third column). Here, the colors refer to the intensity of the velocity filed.

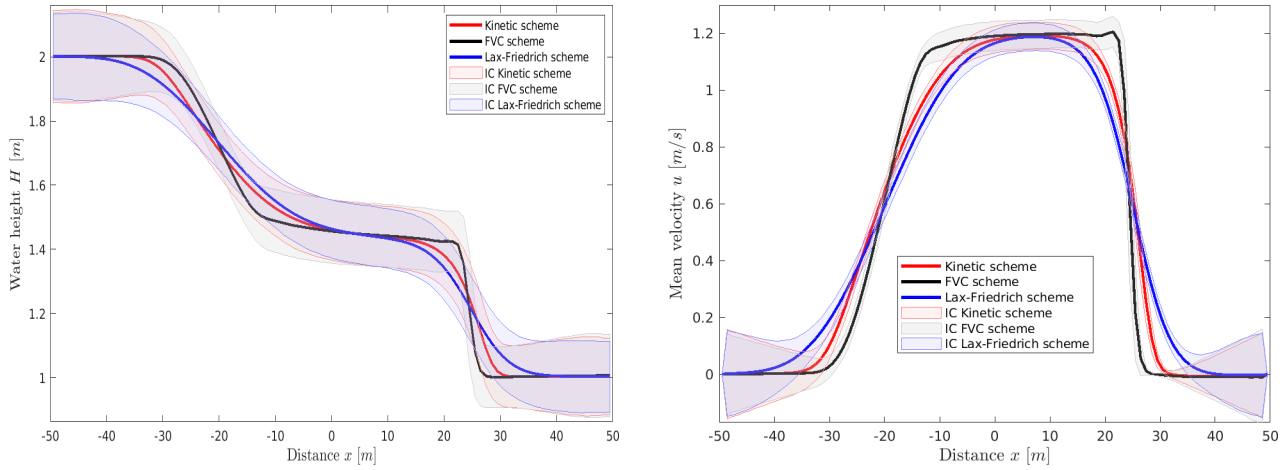

Figure 7: Mean solutions for water height (left plot) and mean velocity (right plot) for the dam-break problem over a flat bed.

In addition to these variables, we add the three other random variables representing the bed friction, the viscosity between layers, and the wind stress. Three different numbers of layers $M = 4$, 8, and 12 are used in this example.

In our computations, the surrogate model is built using 1000 forward simulations of the numerical model. Then, only a few representative modes are selected using the proper orthogonal decomposition where the corresponding eigenvalues are supposed to contain the uncertainty displayed by the model. The PCE is therefore built using these eigenvalues. Table 2 summarizes the steps needed for building the surrogate model

Table 2: Best polynomial degrees with LOO errors for the POD modes in the water height and mean velocity using different numbers of layers for the wind-driven recirculation flow problem over a flat bed.

	Error and polynomial degree for water height H		
	$M = 4$	$M = 8$	$M = 12$
Number of modes	5	5	5
Error (M1)	1.365E-06	1.267E-06	1.251E-06
Error (M2)	9.059E-08	8.113E-08	8.06E-08
Error (M3)	4.012E-08	3.783E-08	3.739E-08
Error (M4)	5.095E-08	4.595E-08	4.263E-08
Error (M5)	3.596E-08	3.297E-08	3.154E-08
Polynomial (M1 - M5)	3	3	3

	Error and polynomial degree for water mean velocity u		
	$M = 4$	$M = 8$	$M = 12$
Number of modes	30	30	30
Error (M1)	8.018E-06	7.453E-06	8.064E-06
Error (M2)	1.820E-05	1.870E-05	1.844E-05
Error (M3)	9.717E-05	8.930E-05	9.236E-05
Error (M4)	5.629E-05	5.156E-05	4.688E-05
Error (M5)	6.351E-05	5.950E-05	5.767E-05
Error (M6)	6.104E-05	5.735E-05	5.886E-05
Error (M7)	6.566E-05	6.5180E-05	6.497E-05
Error (M8)	6.397E-05	6.865E-05	6.628E-05
Error (M9)	7.209E-05	6.528E-05	6.198E-05
Error (M10)	8.240E-05	8.059E-05	6.966E-05
Error (M11)	9.463E-05	8.461E-05	8.699E-05
Error (M12)	11.430E-05	11.06E-05	10.62E-05
Error (M13)	14.760E-05	13.96E-05	13.86E-05
Error (M14)	20.08E-05	21.20E-05	22.08E-05
Error (M15)	19.81E-05	20.24E-05	20.47E-05
Error (M16)	66.30E-05	65.33E-05	64.24E-05
Error (M17)	70.64E-05	66.91E-05	71.38E-05
Error (M18)	81.25E-05	75.44E-05	77.41E-05
Error (M19)	73.52E-05	72.42E-05	69.00E-05
Error (M20)	74.24E-05	82.57E-05	75.28E-05
Error (M21)	73.12E-05	73.59E-05	70.81E-05
Error (M22)	66.04E-05	67.76E-05	81.09E-05
Error (M23)	82.39E-05	82.39E-05	80.61E-05
Error (M24)	82.42E-05	82.78E-05	84.72E-05
Error (M25)	72.63E-05	75.82E-05	77.95E-05
Error (M26)	82.13E-05	77.52E-05	86.38E-05
Error (M27)	86.96E-05	86.36E-05	84.49E-05
Error (M28)	86.34E-05	85.27E-05	92.65E-05
Error (M29)	86.12E-05	87.62E-05	91.32E-05
Error (M30)	88.28E-05	88.33E-05	90.98E-05
Polynomial (M1 - M30)	3	3	3

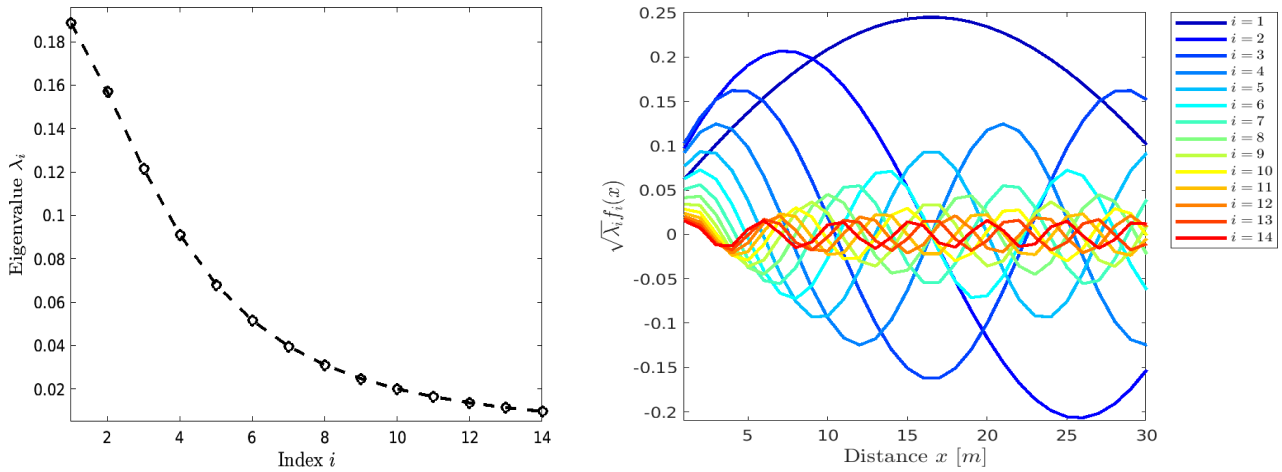

Figure 8: Eigenvalues (left plot) and associated eigenfunctions (right plot) of the correlation matrix in the KL decomposition for the wind-driven recirculation flow problem over a flat bed.

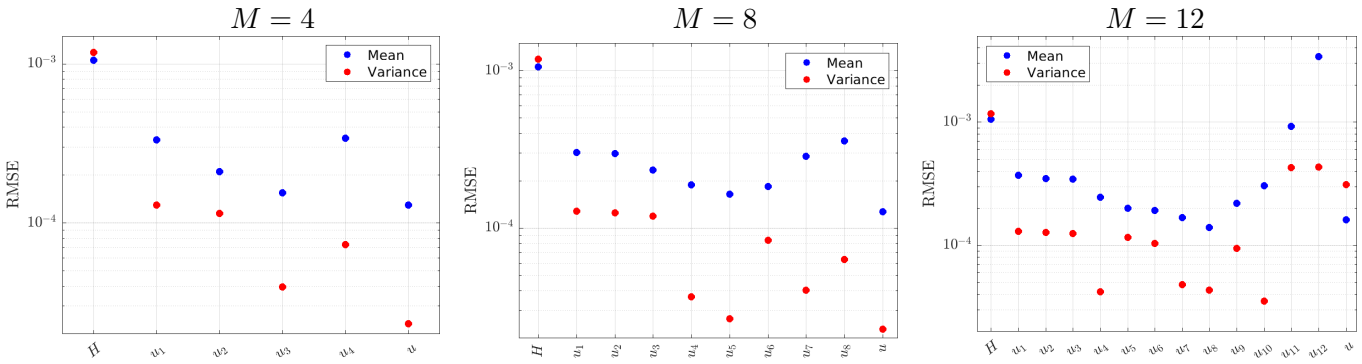

Figure 9: RMSE in the variance and the mean fields using 4, 8 and 12 layers for the wind-driven recirculation flow problem over a flat bed.

to quantify the uncertainty of both the water height and velocity. For this case, results demonstrate that only 5 modes are needed to accurately estimate the physical variability *i.e.*, only 5 polynomial representations are needed instead of the 100 polynomials corresponding to the number of control volumes used in the FVC scheme. Results in Table 2 also demonstrate that the LOO error for this problem is very low suggesting that the polynomial Chaos expansion has accurately assessed the uncertainty given by each mode. In addition, Table 2 suggests that a polynomial degree equal 3 is sufficient for all the eigenvalues of the POD. For the average velocity field, results for the POD ensure that given the physical variability of the system and in order to have an accurate proper decomposition, 30 modes are needed *i.e.*, 30 polynomial Chaos expansions are needed to accurately assess the uncertainty. Yet, this number is smaller compared to the number needed without using the POD (in this case 100 expansions are needed corresponding to the number of control volumes). Again, as suggested by the results for the velocity in Table 2, the PCE computed over the eigenvalues converge with respect to the LOO error as all the errors are below 10^{-5}. Note that, although the errors in Table 2 confirm that the PCE converges, it is necessary to assess the accuracy of the whole surrogate model as the POD is expected to add some uncertainties to the surrogate model. For this purpose, an assessment of the accuracy of the surrogate model is performed in comparison with the Monte-Carlo simulations using 100000 samples. We use the RMSE to quantify the errors in the estimation of the variance and the mean solutions. Figure 9 depicts the obtained results for all the considered cases. Notice that the RMSE errors are also estimated for velocities of all the layers. It is clear that the values of the errors are below the LOO errors estimated previously. However, the RMSE errors are below 10^{-3} for all considered

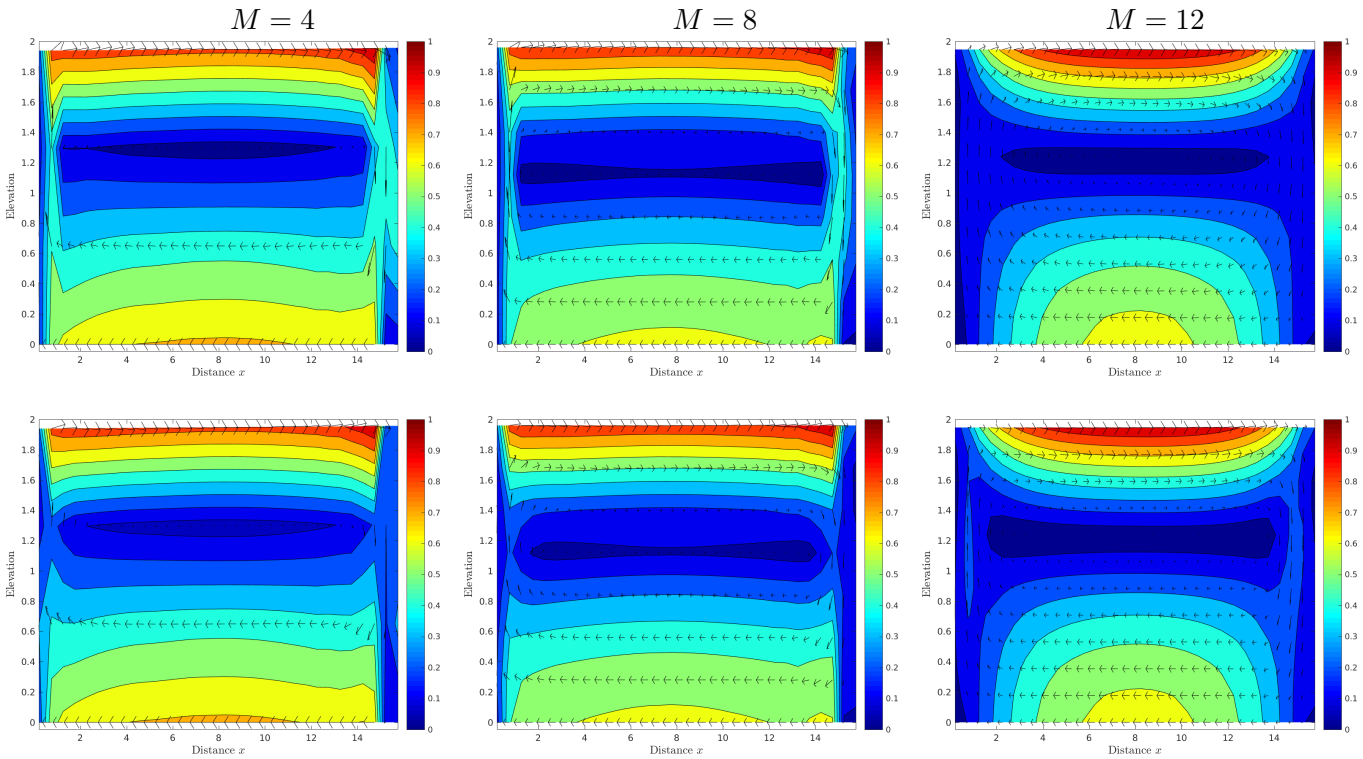

Figure 10: Deterministic flow fields (first row) and stochastic mean flow fields (second row) for the wind-driven recirculation flow problem over a flat bed using $M = 4$ (first column), $M = 8$ (second column) $M = 12$ (third column).

cases which makes the use of the surrogate model very reliable to quantify the uncertainty generated by the multilayer shallow water equations for this flow problem.

Figure 10 exhibits the velocity fields obtained using the considered numbers of layers. As can be seen, a central recirculation has been generated in the flow domain and an increase of the number of layers in the one-dimensional model yields a consistent convergence to the velocity profile obtained using the three-dimensional Navier-Stokes equations, compare for example [4, 7]. These results are a clear indication that it is possible to efficiently resolve the vertical variation in the water velocity using the one-dimensional shallow water equations without the need for the three-dimensional Navier-Stokes equations. A comparison in computational costs between the one-dimensional multilayer shallow water models and the three-dimensional Navier-Stokes equations has been reported in [7, 39]. Furthermore, the mean solution computed by the surrogate model shows some differences compared to the deterministic solutions for this example. This is mainly due to the nonlinearities associated with the multilayer shallow model equations. In order to investigate further the impact of these nonlinearities, ranges of uncertainty in the water height and averaged water velocity is included in Figure 11. One can see that all three numbers of layers yield the same behavior in both the estimated water height and the averaged velocity. Note that there is a little difference noticed between the stochastic and deterministic solutions in the averaged velocity profile. This suggests that the discrepancies observed in Figure 10 are generated from the velocity between layers. Moreover, One can also assess that the stochastic parameters considered in this recirculation flow problem are responsible for up to 5% in the uncertainty in the outputs.

5.3 Multilayer free-surface flows in the Strait of Gibraltar

Our final test problem consists of flow recirculation in the Strait of Gibraltar using the multilayer shallow water equations with the mass exchange. Note that this flow problem presents a realistic practical test

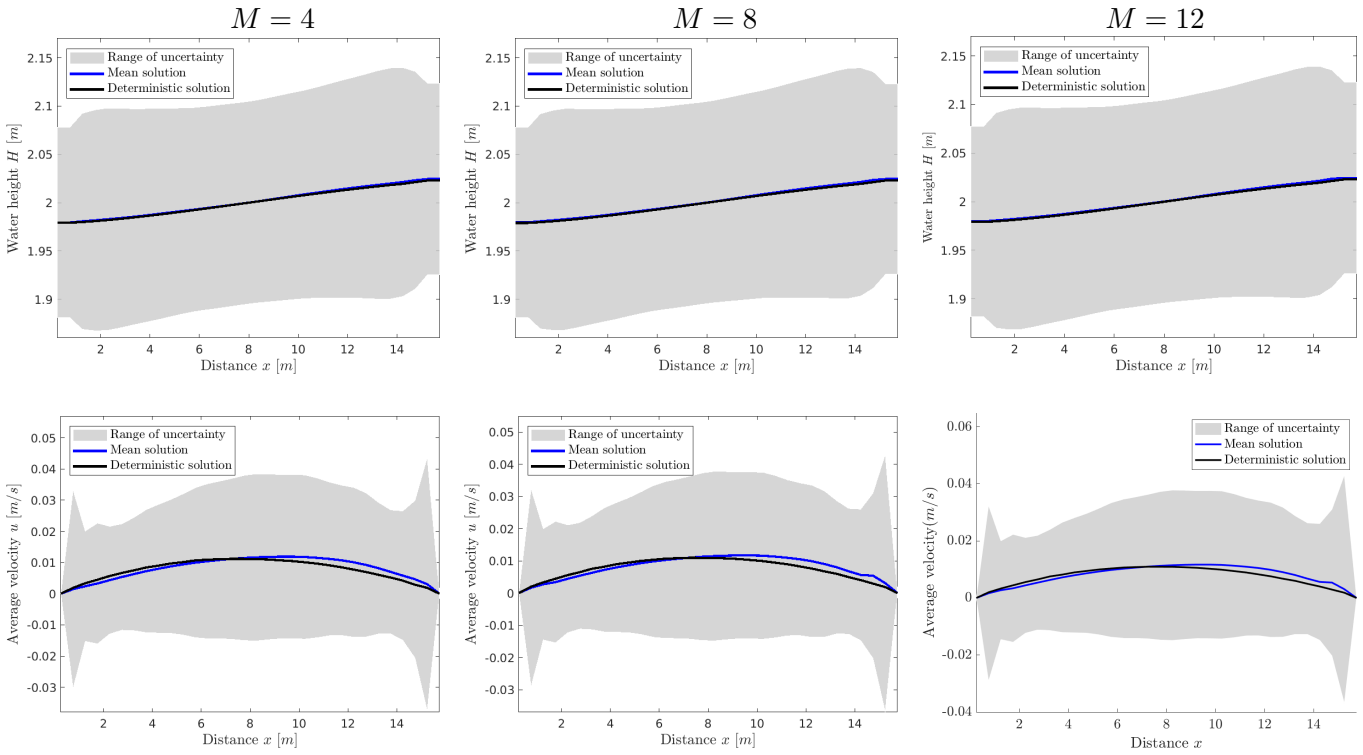

Figure 11: Uncertainty ranges for water height (first row) and mean velocity (second row) along with deterministic solutions for the wind-driven recirculation flow problem over a flat bed.

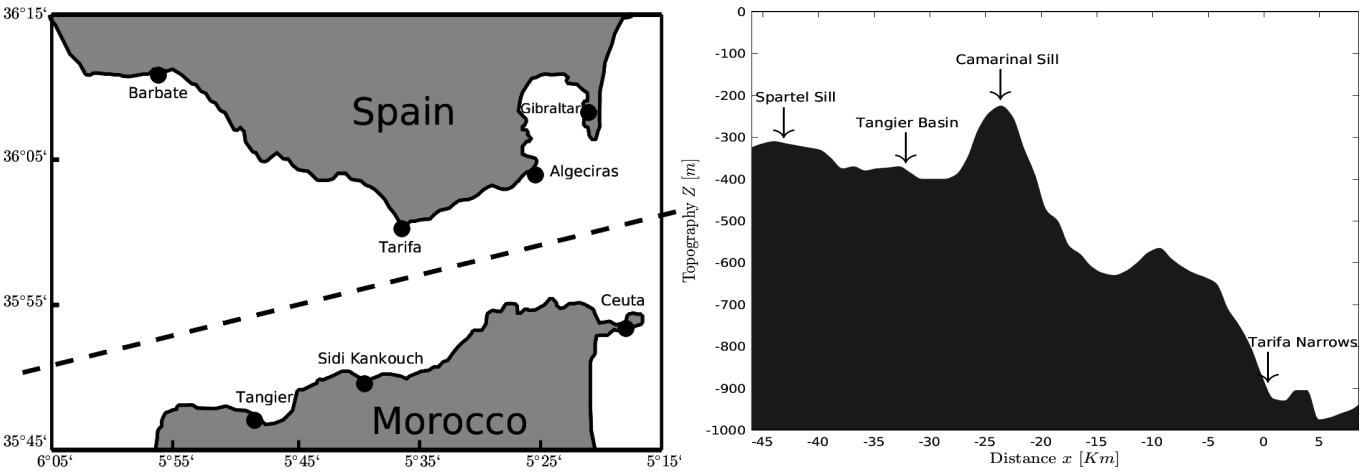

Figure 12: Schematic map of the Strait of Gibraltar along with relevant locations (left plot) and the bathymetry used for the one-dimensional simulations in the present study (right plot).

of multilayer shallow water flows for two major reasons. Firstly, the domain of the Strait of Gibraltar is relatively large including high gradients of the bathymetry and well-defined shelf regions. Secondly, the Strait is deep and contains two water bodies with different densities, which present a challenge in shallow water modelling. Indeed, the basic Oceanic circulation in the Strait of Gibraltar consists of an upper layer of cold, fresh surface Atlantic water and an opposite deep current of warmer, salty outflowing Mediterranean water, compare for example [1, 33, 45, 65]. A schematic map of the Strait of Gibraltar along with relevant locations is depicted in the left plot of Figure 12. The system is bounded to the north and south by the Iberian and African continental forelands, respectively, and to the west and east by the Atlantic ocean and the Mediterranean sea. In geographical coordinates, the Strait is $35^o45'$ to $36^o15'$ N latitude and $5^o15'$ to $6^o05'$ W longitude. Here, we consider a one-dimensional cross-section along the Strait (obtained by a longitudinal section along the dashed line in the left plot of Figure 12) subject to a Levante wind with speed $w = 15\ m/s$ blowing from the east to the west of the Strait. The computational domain and the associated bathymetry are displayed in the right plot of Figure 12. This restricted domain has also been considered in [16, 17, 50] among others. In all simulations, we have used $\rho_a = 1200\ kg/m^3$, $g = 9.81\ m/s^2$, $\nu = 0.02$, $n_b = 0.007\ s/m^{1/3}$, $\sigma_s = 0.0015$, the computational domain is discretized into 50 control volumes and numerical results are presented for 10 layers at time $t = 50\ s$. The purpose here is to quantify the uncertainty of hydraulic modelling using multilayer shallow water equations. This step is very important if one wants to monitor the hydraulics linked activities such as maritime transport and fishery among others. In addition, one of the purposes of this study is to perform a sensitivity analysis for free-surface flows in the Strait of Gibraltar. This will help to prioritize the modelling parameters based on their contribution to the uncertainty of the hydraulic states. This will also identify the parameters that should be controlled once the model is used in an operational framework.

As in the previous examples, we construct our meta-model for all the physical variables (water height, averaged water velocity, and the velocity for each layer). We begin by assessing the quality of the surrogate model compared to the Monte-Carlo simulations using 100000 evaluations of the model. We recall here that in order to reduce the dimensionality of the problem and to avoid the problem of reducing the number of control volumes which will impact the numerical accuracy, we use the proper orthogonal decomposition. Thus, the polynomial Chaos expansion is computed over the eigenvalues of the decomposition. This would allow having less number of decompositions than the original problem which consists of computing a decomposition for each variable at each control volume. The obtained results for the best polynomial degree and LOO errors for the POD modes in the water height and velocity using the FVC scheme are presented in Table 3. First, one can see that for a fixed value of the truncation error, the POD does not yield the same number of orthogonal functions. For example, the water height needs only 3 modes whereas, the averaged water velocity needs 15 modes. It should be stressed that this is a considerable reduction in the dimension of the output. Moreover, using the LAR method in order to build the polynomial expansion over the eigenvalues, Table 3 also reveals that polynomials of degree 2 to 3 are enough to correctly capture the uncertainty generated in the eigenvalues. the LOO error estimated for each mode demonstrates that the uncertainty of these modes is well captured by the considered polynomials with an error varying from 10^{-7} to 10^{-10}.

It should be pointed out that in general assessing the accuracy of the PCE over the eigenvalues of the POD does not guarantee the convergence of the surrogate model. For this reason, the accuracy of the proposed surrogate model is compared to very demanding Monte-Carlo simulations composed of 100000 forward simulations. Then, the RMSE is estimated for all physical variables of the hydraulic states for the estimation of the mean and variance solutions. The obtained results for this case are reported in Figure 13 and as for the previous case, the surrogate model allows to correctly quantify the uncertainty for all the physical parameters as the error is less than 10^{-3} in both the mean and variance solutions. Since the Monte-Carlo simulations have been performed, we also assess the ability of the surrogate model to estimate the whole Probability Density Function (PDF) at the locations where data measurements are often collected for hydraulics namely, Camarinal Sill, Tangier Basin, and Tarifa Narrows as shown in Figure 12. This comparison is shown in Figure 14 where the estimation of the whole PDF at these locations using

Table 3: Best polynomial degree with LOO errors for the POD modes in the water height and velocity using the FVC scheme for the flow recirculation in the Strait of Gibraltar.

	Water height	Mean water velocity
Number of modes	3	15
LOO error (M1)	6.169E-10	5.804E-08
LOO error (M2)	4.989E-09	1.320E-06
LOO error (M3)	1.588E-09	1.464E-07
LOO error (M4)	———	5.142E-08
LOO error (M5)	———	1.015E-07
LOO error (M6)	———	9.949E-08
LOO error (M7)	———	3.199E-07
LOO error (M8)	———	6.806E-07
LOO error (M9)	———	3.357E-07
LOO error (M10)	———	1.868E-07
LOO error (M11)	———	7.066E-07
LOO error (M12)	———	7.829E-07
LOO error (M13)	———	1.520E-06
LOO error (M14)	———	1.957E-06
LOO error (M15)	———	3.466E-07
Polynomial (M1)	3	3
Polynomial (M2)	3	3
Polynomial (M3)	2	3
Polynomial (M4-M11)	———	3
Polynomial (M12)	———	2
Polynomial (M13-M15)	———	3

Monte-Carlo simulation are compared to those obtained using the surrogate model. It is clear that the results demonstrate that generally there is a good estimation of the probability density function using the tools presented in this study.

A comparison between the deterministic and stochastic approaches is also carried out for this flow problem. In Figure 15 we compare the deterministic and stochastic mean solutions obtained for the flow fields. It is clear that the computed results exhibit slight differences especially downstream of the Strait. This is resulting from the nonlinearities of the exchange terms appearing in the multilayer shallow water equations. Next, the uncertainty ranges of the hydraulic states are displayed in Figure 16. The obtained results show that the uncertainty ranges especially in the water height could not be negligible. The velocity profiles have also been tracked at the three locations where data is regularly collected. The uncertainty ranges in these velocity profiles are very narrow compared to the uncertainty in the water heights. Moreover, the same figure also illustrates the uncertainty range for the averaged water velocity. Although there is huge physical variability in this parameter, which can explain the need for the high number of POD model, see Table3, the uncertainty range is very narrow. These results demonstrate that the uncertainty in the bathymetry, the viscosity, and the friction coefficients impact mostly the water height.

Finally, sensitivity analysis is carried out for this recirculation flow problem in the Strait of Gibraltar. Since the water height is the parameter impacted the most from the stochasticity considered in this study, the sensitivity analysis is carried out over this parameter. Note that the KL decomposition allows sampling of the bathymetric field by introducing 16 random variables. Therefore, we assessed first a total Sobol index for these 16 random variables to which we add the Manning coefficient, the viscosity parameter, and the

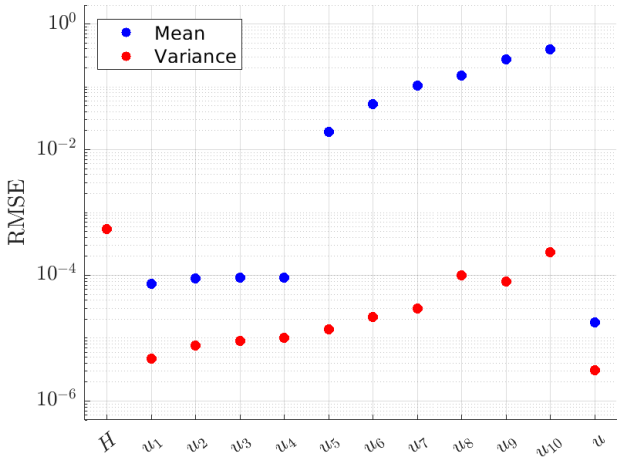

Figure 13: RMSE in different physical variables computed by the model for the flow recirculation in the Strait of Gibraltar.

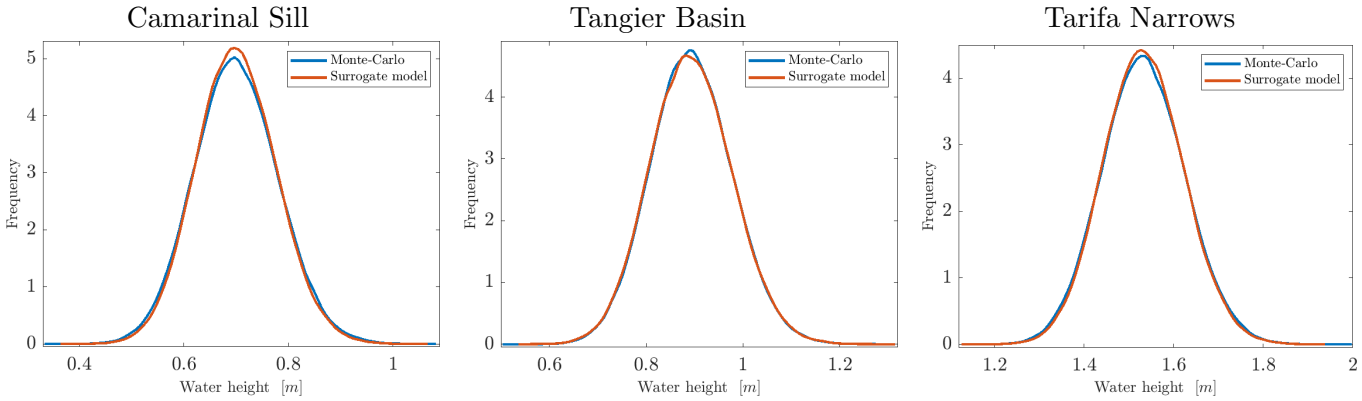

Figure 14: Probability density functions obtained at three different locations using the surrogate model and the Monte-Carlo for the flow recirculation in the Strait of Gibraltar.

wind friction coefficients. Furthermore, the sensitivity analysis is carried only at the three measurement locations and the obtained results are presented in Figure 17. First, the results show that there are no differences between the first and total order, suggesting that the interaction between the random variables is negligible. Next at all considered locations, only random variables of the KL decomposition have high values. This suggests that the simulations of the water height are very sensitive to bathymetry. Finally one can see that reducing the uncertainty could be achieved by monitoring only four or three modes at most of the KL decomposition.

6 Conclusions

A surrogate model has been investigated in the present study for the propagation and quantification of the uncertainty in hydraulic modelling. The flow in such cases has usually many complex features making the classical shallow water equations not suitable for this type of hydraulic problems. In general, one has to consider the full three-dimensional Navier-Stokes equations which require a considerable computational cost to be numerically solved. However, the multilayer shallow water models have proven their efficiency to replace the full three-dimensional Navier-Stokes equations while preserving the numerical flexibility given by the shallow water system and therefore could be considered to model the oceanic recirculation. While

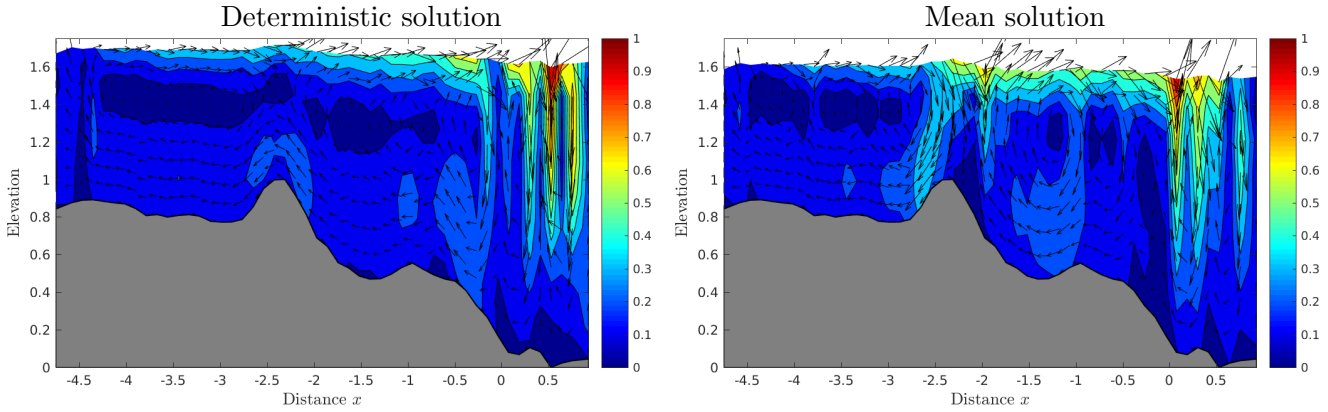

Figure 15: Deterministic and mean solutions of the velocity fields for the flow recirculation in the Strait of Gibraltar.

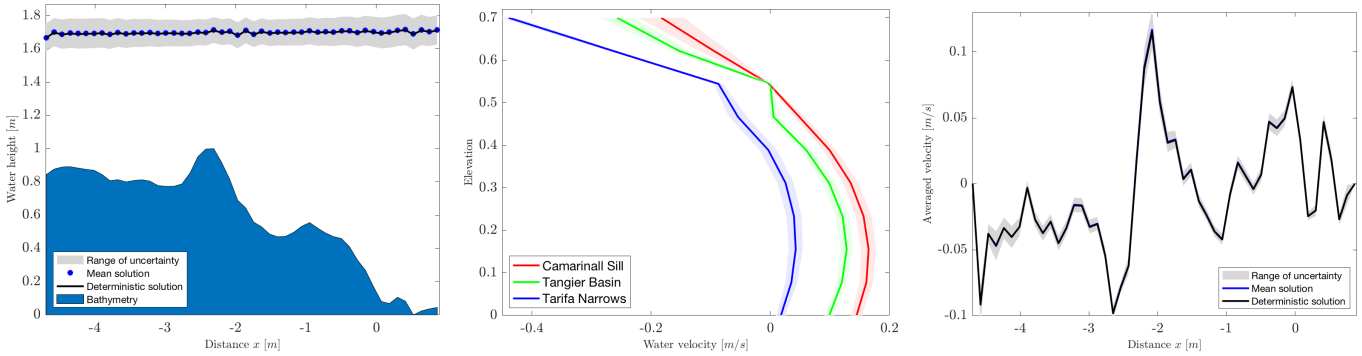

Figure 16: Uncertainty range in the water level (left plot), velocity profiles (middle plot) and averaged water velocity (right plot) for the flow recirculation in the Strait of Gibraltar.

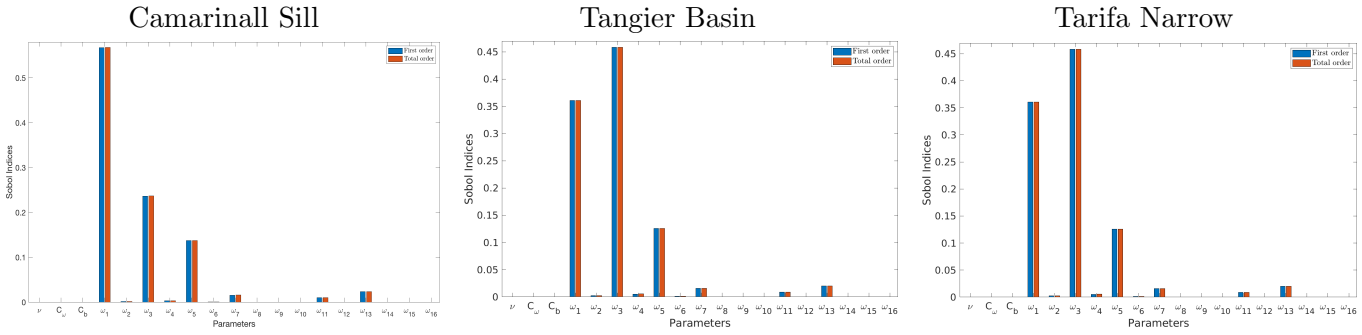

Figure 17: Sensitivity analysis using the Sobol indices at three locations in the Strait where data are measured periodically.

several previous studies demonstrated the numerical performance of the multilayer shallow water models to accurately capture the complex pattern existing in the oceanic recirculation, to the best of our knowledge the problem of uncertainty quantification has been investigated for the first time. The implementation of surrogate models is a common practice for uncertainty quantification in computational fluid dynamics. The numerical model needs to be run several times, which increases the computational cost. Through the use of the surrogate models, the required numerous evaluation could be performed easily in a reasonable time. The literature has reported many surrogate models that are used in the context of uncertainty quantification.

The choice of one or another is generally done based on numerical experiences. Usually when possible the robustness of these models is assessed with comparison to the well-established Monte-Carlo simulations given its statistical properties. In the present work, we choose to used polynomial Chaos expansion based on the previous results of this model in the hydraulic computations. Furthermore, the spatial dimension of different quantities of interest (water height and velocity) are reduced thanks to a proper orthogonal decomposition. The results are presented for three application cases: a dam-break problem, a wind-driven circulation over a lake, and the real case of the Strait of Gibraltar.

In the first case, we performed uncertainty quantification for a dam break problem. We assumed that the uncertainty of the simulations resulted from the stochastic bed, friction coefficient, viscosity between layers, and the wind friction terms. Moreover, the highlight was put here on the numerical schemes used to solve the multilayer shallow water equations. We used three different numerical schemes, namely the Kinetic, the Lax-Friedrich, and a finite volume characteristics method. The results demonstrated that the surrogate model performed well when compared to the Monte-Carlo simulations. The RMSE reaches an order of magnitude up to 10^{-3} for the estimation of the mean and the variance solutions using all the considered numerical methods. The results also demonstrated a good agreement in the simulation of the physical patterns in the mean field. Finally, we concluded that the numerical schemes do not display the same level of uncertainties. In the second, we performed uncertainty quantification for a wind-driven circulation problem. As for the previous case, the stochastic parameters considered here include the bathymetry, the friction term, the viscosity between layers and the wind friction terms. The results showed a good agreement between the surrogate model and the Monte-Carlo simulations. The RMSE has an order of magnitude ranging from 10^{-4} to 10^{-3}. Moreover, we compared the results of the uncertainty for a different numbers of layers in the model. We concluded that the uncertainty in the model is not much affected by the number of layers. However, one should be aware that the number of layers could affect the physical modelling which can yield modelling errors. Finally, we applied our methodology to the real case of flow recirculation in the Strait of Gibraltar. This place knows consistent economical activities over time including fishery, transport, and tourism among others. Therefore, it is important to monitor the flow in the Strait of Gibraltar. First, the surrogate model is compared to Monte-Carlo simulations. Results demonstrated the good agreement represented by a value of RMSE ranging from 10^{-5} to 10^{-3}. Moreover, as three stations exist in this area, we evaluated the ability of the surrogate models to compute the probability density function. The results were compared with a probability density function computed with Monte-Carlo simulations. The results prove that the surrogate model performs very well for this flow problem. Finally, we were interested in a sensitivity analysis in these locations. This could be a good indicator of what are the physical variables that one should monitor using either data assimilation or Bayesian inference in order to reduce the uncertainty of the numerical model. Results showed that the bathymetry is the main driver of uncertainty in the considered three locations. In the present work, the proposed methodology has been assessed correctly in order to quantify the uncertainty resulting from bathymetric forces on the hydraulic state. Future works will be devoted to compare the resulting uncertainty with other hydraulic models. This comparison will help to assess how useful this class of models is at the operational level.

Acknowledgments

This work was partly performed while the first author was visiting IWRI at University Mohammed VI Polytechnic Benguerir. Financial support provided by IWRI is gratefully acknowledged.

References

[1] J.I. Almazán, H. Bryden, T. Kinder, and eds G. Parrilla. *Seminario Sobre la Oceanografía Física del Estrecho de Gibraltar*. SECEG, Madrid, 1988.

[2] S. Amiri, N. Talebbeydokhti, and A. Baghlani. A two-dimensional well-balanced numerical model for shallow water equations. *Scientia Irancia*, 20:97–107, 2013.

[3] Kendall Atkinson. A survey of numerical methods for the solution of Fredholm integral equations of the second kind. 1976.

[4] E. Audusse. A multilayer Saint-Venant system: Derivation and numerical validation. *Discrete and Continuous Dynamical Systems, Ser. B.*, 5:189–214, 2005.

[5] E. Audusse, M.O. Bristeau, M. Pelanti, and J. Sainte-Marie. Approximation of the hydrostatic Navier-Stokes system for density stratified flows by a multilayer model: Kinetic interpretation and numerical solution. *J. Comp. Physics.*, 230:3453–3478, 2011.

[6] E. Audusse, M.O. Bristeau, B. Perthame, and J. Sainte-Marie. A multilayer Saint-Venant system with mass exchanges for shallow water flows. derivation and numerical validation. *M2AN Math. Model. Numer. Anal.*, 45:169–200, 2011.

[7] Emmanuel Audusse, Fayssal Benkhaldoun, Saida Sari, Mohammed Seaid, and Pablo Tassi. A fast finite volume solver for multi-layered shallow water flows with mass exchange. *Journal of Computational Physics*, 272:23–45, 2014.

[8] Nicolas Baudry and Stéphane Calmant. 3D modelling of seamount topography from satellite altimetry. *Geophysical Research Letters*, 18:1143–1146, 1991.

[9] F. Benkhaldoun, S. Sari, and M. Seaid. A simple multi-layer finite volume solver for density-driven shallow water flows. *Mathematics and computers in simulations*, 99:170–189, 2014.

[10] Marc Berveiller, Bruno Sudret, and Maurice Lemaire. Stochastic finite element: a non intrusive approach by regression. *European Journal of Computational Mechanics/Revue Européenne de Mécanique Numérique*, 15:81–92, 2006.

[11] G. Blatman and B. Sudret. Adaptative sparse polynomial chaos expansion based on Least Angle Regression. *Journal of Computational Physics*, 230:2345–2367, 2011.

[12] G. Blatman and B. Sudret. Principal component analysis and Least Angle Regression in spectral stochastic finite element analysis. In M.H. Faber, J. Köhler, and K. Nishijima, editors, *Proc. 11th Int. Conf. on Applications of Stat. and Prob. in Civil Engineering (ICASP11), Zurich, Switzerland*, 2011.

[13] G Blatman and B Sudret. Principal component analysis and least angle regression in spectral stochastic finite element analysis. In *Proc. 11th Int. Conf. on Applications of Stat. and Prob. in Civil Engineering (ICASP11), Zurich, Switzerland*, 2011.

[14] Géraud Blatman and Bruno Sudret. *Sparse polynomial chaos expansions of vector-valued response quantities*. CRC Press/Balkema, 2013.

[15] Silvia Bozzi, Giuseppe Passoni, Pietro Bernardara, Nicole Goutal, and Aurélie Arnaud. Roughness and discharge uncertainty in 1D water level calculations. *Environmental Modeling & Assessment*, 20:343–353, 2015.

[16] M.J. Castro, T. Chacón, E.D. Nieto, and C. Parés. On well-balanced finite volume methods for nonconservative nonhomogeneous hyperbolic systems. *SIAM J. Sci. Comput.*, 29:1093–1126, 2007.

[17] M.J. Castro, J. Macías, and C. Parés. A Q-scheme for a class of systems of coupled conservation laws with source term, .application to a two-layer 1D shallow water system. *M2AN Math. Model. Numer. Anal.*, 35:107–127, 2001.

[18] Eduardo Chávarri, Alain Crave, Paule Bonnet, Abel Mejía, Santos Da Silva, and Loup Guyot. Hydrodynamic modelling of the amazon river: Factors of uncertainty. *Journal of South American Earth Sciences*, 44:94–103, 2013.

[19] Mathilde Chevreuil and Anthony Nouy. Model order reduction based on proper generalized decomposition for the propagation of uncertainties in structural dynamics. *International Journal for Numerical Methods in Engineering*, 89:241–268, 2012.

[20] M. Chin, C. Haza, and J. Mariano. A reduced-order information filter for multilayer shallow-water models: profiling and assimilation of sea surface height. *Journal of Atmospheric and Oceanic Technology*, 19:517–533, 2002.

[21] Kyum Choi, V Grandhi, A Canfield, and L Pettit. Polynomial Chaos expansion with Latin Hypercube Sampling for estimating response variability. *AIAA journal*, 42:1191–1198, 2004.

[22] R Crisovan, D Torlo, R Abgrall, and S Tokareva. Model order reduction for parametrized nonlinear hyperbolic problems as an application to uncertainty quantification. *Journal of Computational and Applied Mathematics*, 348:466–489, 2019.

[23] Sayan Dey, Siddharth Saksena, and Venkatesh Merwade. Assessing the effect of different bathymetric models on hydraulic simulation of rivers in data sparse regions. *Journal of Hydrology*, 575:838–851, 2019.

[24] Zlatko Drmac. A posteriori computation of the singular vectors in a preconditioned Jacobi SVD algorithm. *IMA journal of numerical analysis*, 19:191–213, 1999.

[25] S. Dubreuil, M. Berveiller, F. Petitjean, and M. Salan. Construction of bootstrap confidence intervals on sensitivity indices computed by polynomial chaos expansion. *Reliability Engineering and System Safety*, 121:263–275, 2014.

[26] Bradley Efron, Trevor Hastie, Iain Johnstone, and Robert Tibshirani. Least angle regression. *The Annals of Statistics*, 32:407–499, 2004.

[27] N El Moçayd. La décomposition en polynômes du chaos pour lámélioration de lássimilation de données ensembliste en hydraulique fluviale. *PhD Université de Toulouse, INP*, 2017.

[28] Nabil El Moçayd, Sophie Ricci, Nicole Goutal, C Rochoux, Sébastien Boyaval, Cédric Goeury, Didier Lucor, and Olivier Thual. Polynomial surrogates for open-channel flows in random steady state. *Environmental Modeling & Assessment*, pages 1–23, 2017.

[29] Ruth Esteban, Philippe Verborgh, Pauline Gauffier, Joan Gimnez, Christophe Guinet, and Renaud de Stephanis. Dynamics of killer whale, bluefin tuna and human fisheries in the Strait of Gibraltar. *Biological Conservation*, 194:31–38, 2016.

[30] D. Farmer and L. Armi. Maximal two-layer exchange over a sill and through a combination of a sill and contraction with barotropic flow. *J. Fluid Mech.*, 164:53–76, 1986.

[31] F. Franzini and S. Soares-Franzao. Efficiency and accuracy of lateralized HLL and HLLS and augmunted Roes's scheme with energy balance for river flows in irregular channels. *Applied mathematical modelling*, 4:7427–7446, 2016.

[32] Roger Ghanem, David Higdon, and Houman Owhadi. *Handbook of uncertainty quantification*, volume 6. Springer, 2017.

[33] F. Gómez. The role of the exchanges through the Strait of Gibraltar on the budget of elements in western mediterranean sea: Consequences of humain-induced modifications. *Marine Pollution Bulletin.*, 46:685–694, 2003.

[34] Carlos Gonzlez, Emma Reyes, Alfredo Izquierdo, Miguel Bruno, and R. Maanes. Surface currents and transport processes in the Strait of Gibraltar: Implications for modeling and management of pollutant spills. *Ocean & Coastal Management*, 179:104–869, 2019.

[35] Nicole Goutal, Cedric Goeury, Riadh Ata, Sophie Ricci, Nabil El Mocayd, M Rochoux, Hind Oubanas, Igor Gejadze, and Olivier Malaterre. Uncertainty quantification for river flow simulation applied to a real test case: The garonne valley. In *Advances in Hydroinformatics*, pages 169–187. Springer, 2018.

[36] J. Gmez-Enri, C.J. Gonzlez, M. Passaro, S. Vignudelli, O. lvarez, P. Cipollini, R. Maanes, M. Bruno, M.P. Lpez-Carmona, and A. Izquierdo. Wind-induced cross-strait sea level variability in the Strait of Gibraltar from coastal altimetry and in-situ measurements. *Remote Sensing of Environment*, 221:596–608, 2019.

[37] J. Hall, S. Tarantola, P. Bates, and M. Horritt. Distributed sensitivity analysis of flood inundation model calibration. *Journal of Hydraulic Engineering*, 131:117–126, 2005.

[38] M.S. Horritt. Stochastic modelling of 1-D shallow water flows over uncertain topography. *Journal of Computational Physics*, 180:327–338, 2002.

[39] N. Izem and M. Seaid. A well-balanced Runge-Kutta discontinuous Galerkin method for multilayer shallow water equations with non-flat bottom topography. *Advances in Applied Mathematics and Mechanics*, To appear, 2021.

[40] Nouh Izem, Mohammed Seaid, and Mohamed Wakrim. A discontinuous Galerkin method for two-layer shallow water equations. *Mathematics and Computers in Simulation*, 120:12–23, 2016.

[41] Gabriel Jord, Antonio Sanchez-Roman, and Dami Gomis. Reconstruction of transports through the Strait of Gibraltar from limited observations. *Climate Dynamics*, 48:851–865, 2017.

[42] Kari Karhunen. Zur spektraltheorie stochastischer prozesse. *Ann. Acad. Sci. Fennicae, AI*, 34, 1946.

[43] A Koulali, D Ouazar, A Tahayt, RW King, Philippe Vernant, RE Reilinger, Simon McClusky, Taoufik Mourabit, Martín Davila, and N Amraoui. New GPS constraints on active deformation along the Africa–Iberia plate boundary. *Earth and Planetary Science Letters*, 308:211–217, 2011.

[44] Asier Lacasta, Mario Morales-Hernndez, Javier Burguete, Pilar Brufau, and Pilar Garca-Navarro. Calibration of the 1D shallow water equations: a comparison of Monte Carlo and gradient-based optimization methods. *Journal of Hydroinformatics*, 19:282–298, 2017.

[45] J.G. Lafuente, J.L. Almazán, F. Catillejo, A. Khribeche, and A. Hakimi. Sea level in the Strait of Gibraltar: Tides. *Int. Hydrogr. Rev. LXVII.*, 1:111–130, 1990.

[46] Olivier Le Maître and M Knio. *Spectral methods for uncertainty quantification: with applications to computational fluid dynamics*. Springer Science & Business Media, 2010.

[47] Guesuk Lee, Wongon Kim, Hyunseok Oh, D. Youn, and H. Kim. Review of statistical model calibration and validation—from the perspective of uncertainty structures. *Structural and Multidisciplinary Optimization*, 60:1619–1644, 2019.

[48] R.J. Leveque. *Finite Volume Methods for Hyperbolic Problems*. Cambridge University Press, Cambridge, 2002.

[49] M Loève. Elementary probability theory. In *Probability Theory I*, pages 1–52. Springer, 1977.

[50] J. Macías, C. Parés, and M.J. Castro. Improvement and genralization of a finite element shallow water solver to multi-layer systems. *Int. J. Numer. Methods Fluids.*, 31:1037–1059, 1999.

[51] Amandine Marrel, Nathalie Saint-Geours, and Matthias De Lozzo. Sensitivity analysis of spatial and/or temporal phenomena. *Handbook of Uncertainty Quantification*, pages 1–31, 2016.

[52] Nabil Moayd, S. Mohamed, Driss Ouazar, and Mohammed Seaid. Stochastic model reduction for polynomial chaos expansion of acoustic waves using proper orthogonal decomposition. *Reliability Engineering & System Safety*, 195:106–733, 2020.

[53] J. Murillo and P. Garcia-Navarro. Augmented versions of the HLL and HLLC Riemann solvers including sourc terms in one and two dimensions for the shallow flow applications. *Journal of computational physics*, 231:6861–6906, 2012.

[54] Maria Navarro, P. Le Maître, Ibrahim Hoteit, L. George, T. Mandli, and M. Knio. Surrogate-based parameter inference in debris flow model. *Computational Geosciences*, 22:1447–1463, 2018.

[55] M. Raisee, D. Kumar, and C. Lacor. A non-intrusive model reduction approach for polynomial chaos expansion using proper orthogonal decomposition. *International Journal for Numerical Methods in Engineering*, 103(4):293–312, 2015.

[56] Christian Refsgaard, Lajer Højberg, and A Vanrolleghem. Uncertainty in the environmental modelling process–a framework and guidance. *Environmental modelling & software*, 22:1543–1556, 2007.

[57] Mario Ricchiuto, Marco Congedo, and I Delis. *Runup and uncertainty quantification: sensitivity analysis via ANOVA decomposition.* PhD thesis, INRIA, 2014.

[58] P. Roe. Approximate Riemann solvers, parameter vectors, and difference schemes. *Journal of Computational Physics*, 43:357–372, 1981.

[59] T Roy, Nabil El Moçayd, Sophie Ricci, Christophe Jouhaud, Nicole Goutal, Matthias De Lozzo, and C Rochoux. Comparison of polynomial Chaos and Gaussian process surrogates for uncertainty quantification and correlation estimation of spatially distributed open-channel steady flows. *Stochastic Environmental Research and Risk Assessment*, pages 1–19, 2017.

[60] Christoph Schwab and Alexandru Todor. Karhunen–loève approximation of random fields by generalized fast multipole methods. *Journal of Computational Physics*, 217:100–122, 2006.

[61] M. Seaid. Semi-lagrangian integration schemes for viscous incompressible flows. *Comp. Methods in App. Math.*, 4:392–409, 2002.

[62] Mohammed Shamkhi and Zainab Attab. Estimation of Manning's roughness coefficient for Tigris river by using HEC-RAS model. *Wasit Journal of Engineering Sciences*, 6:90–97, 2018.

[63] Christian Soize and Roger Ghanem. Physical systems with random uncertainties: chaos representations with arbitrary probability measure. *SIAM Journal on Scientific Computing*, 26:395–410, 2004.

[64] Ihab Sraj, P Le Maître, M Knio, and Ibrahim Hoteit. Coordinate transformation and polynomial chaos for the bayesian inference of a Gaussian process with parametrized prior covariance function. *Computer Methods in Applied Mechanics and Engineering*, 298:205–228, 2016.

[65] L. Tejedor, A. Izquierdo, B.A. Kagan, and D.V. Sein. Simulation of the semidiurnal tides in the Strait of Gibraltar. *J. Geophysical Research.*, 104:13541–13557, 1999.

[66] C. Temperton and A. Staniforth. An efficient two-time-level Galerkin-characteristics semi-implicit integration scheme. *Quart. J. Roy. Meteor. Soc.*, 113:1025–1039, 1987.

[67] Junna Wang and Zhonglong Zhang. Evaluating riparian vegetation roughness computation methods integrated within HEC-RAS. *Journal of Hydraulic Engineering*, 145:0401–9020, 2019.

[68] Norbert Wiener. The homogeneous chaos. *American Journal of Mathematics*, 60:897–936, 1938.

[69] David Williams. *Diffusions, Markov processes, and martingales. Vol. 1, Foundations*. Wiley, 1979.

[70] Dongbin Xiu. *Numerical methods for stochastic computations: a spectral method approach*. Princeton university press, 2010.

[71] Dongbin Xiu and Em Karniadakis. The Wiener–Askey polynomial chaos for stochastic differential equations. *SIAM journal on scientific computing*, 24:619–644, 2002.

[72] Tianfang Xu, J. Valocchi, Ming Ye, Feng Liang, and Feng Lin. Bayesian calibration of groundwater models with input data uncertainty. *Water Resources Research*, 53:3224–3245, 2017.

[73] Y Yoon, Michael Durand, J Merry, A Clark, M Andreadis, and E Alsdorf. Estimating river bathymetry from data assimilation of synthetic swot measurements. *Journal of hydrology*, 464:363–375, 2012.