



**HAL**  
open science

## Polynomial Surrogates for Open-Channel Flows in Random Steady State

Nabil El Moçayd, Sophie Ricci, Nicole Goutal, Mélanie C. Rochoux, Sébastien Boyaval, Cédric Goeury, Didier Lucor, Olivier Thual

► **To cite this version:**

Nabil El Moçayd, Sophie Ricci, Nicole Goutal, Mélanie C. Rochoux, Sébastien Boyaval, et al.. Polynomial Surrogates for Open-Channel Flows in Random Steady State. Environmental Modeling & Assessment, 2017, 10.1007/s10666-017-9582-2 . hal-01763259

**HAL Id: hal-01763259**

**<https://enpc.hal.science/hal-01763259>**

Submitted on 23 Apr 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Polynomial Surrogates for Open-Channel Flows in Random Steady State

Nabil El Moçayd<sup>1,2</sup>  · Sophie Ricci<sup>1</sup> · Nicole Goutal<sup>3</sup> · Mélanie C. Rochoux<sup>1</sup> · Sébastien Boyaval<sup>3</sup> · Cédric Goeury<sup>3</sup> · Didier Lucor<sup>4</sup> · Olivier Thual<sup>5,6</sup>

Received: 13 September 2016 / Accepted: 9 October 2017  
© Springer International Publishing AG 2017

**Abstract** Assessing epistemic uncertainties is considered as a milestone for improving numerical predictions of a dynamical system. In hydrodynamics, uncertainties in input parameters translate into uncertainties in simulated water levels through the shallow water equations. We investigate the ability of generalized polynomial chaos (gPC) surrogate to evaluate the probabilistic features of water level simulated by a 1-D hydraulic model (MASCARET) with the same accuracy as a classical Monte Carlo method but at a reduced computational cost. This study highlights that the water level probability density function and covariance matrix are better estimated with the polynomial surrogate model than with a Monte Carlo approach on the forward model given a limited budget of MASCARET evaluations. The gPC-surrogate performance is first assessed on an idealized channel with uniform geometry and then applied on

the more realistic case of the Garonne River (France) for which a global sensitivity analysis using sparse least-angle regression was performed to reduce the size of the stochastic problem. For both cases, Galerkin projection approximation coupled to Gaussian quadrature that involves a limited number of forward model evaluations is compared with least-square regression for computing the coefficients when the surrogate is parameterized with respect to the local friction coefficient and the upstream discharge. The results showed that a gPC-surrogate with total polynomial degree equal to 6 requiring 49 forward model evaluations is sufficient to represent the water level distribution (in the sense of the  $\ell_2$  norm), the probability density function and the water level covariance matrix for further use in the framework of data assimilation. In locations where the flow dynamics is more complex due to bathymetry, a higher polynomial degree is needed to retrieve the water level distribution. The use of a surrogate is thus a promising strategy for uncertainty quantification studies in open-channel flows and should be extended to unsteady flows. It also paves the way toward cost-effective ensemble-based data assimilation for flood forecasting and water resource management.

✉ Nabil El Moçayd  
elmocayd@cerfacs.fr

<sup>1</sup> CECI, CNRS – CERFACS, 42 Avenue Gaspard Coriolis, 31057 Toulouse cedex 1, France

<sup>2</sup> AgroBioSciences, Université Mohamed 6 Polytechnique, Lot 660, Hay Moulay Rachid, Benguerir, Morocco

<sup>3</sup> Laboratory for Hydraulics Saint-Venant (LHSV), EDF R&D, 6 Quai Watier, 78401 Chatou, France

<sup>4</sup> LIMSIS, CNRS, Université Paris-Saclay, Campus universitaire bât 508, Rue John Von Neumann, 91405 Orsay cedex, France

<sup>5</sup> CECI, CERFACS – CNRS, 42 Avenue Gaspard Coriolis, 31057 Toulouse cedex 1, France

<sup>6</sup> INPT, IMFT – CNRS, Allée Camille Soula, 31400 Toulouse, France

**Keywords** Uncertainty quantification · Hydraulic modeling · Surrogate model · Polynomial chaos expansion · Sensitivity analysis · Covariance matrix

## 1 Introduction

Water resource management and flood forecasting require a solid capacity of anticipating systems changes and thus rely on predictive skills of hydraulic models. These skills have greatly improved thanks to the advances in numerical modeling of free surface flow and the increase of computational

resources. The Shallow Water Equations (SWE) derived from the free surface Navier-Stokes equations are widely used for complex flow simulation. These equations have extensively proved that they are powerful tools to tackle most of open-channel hydraulics problems such as flood hazards [1] or dam breaking [57]. Nevertheless, the prediction of free surface flow characteristics requires the description of initial conditions, river geometry, boundary conditions, and hydraulic physical parameters such as friction coefficients. These input data and parameters are affected by uncertainties that are either due to a lack of knowledge (“epistemic” uncertainties such as modeling assumptions or numerical approximations) and/or by the intrinsic randomness of the natural phenomena (“random” or “aleatoric” uncertainties, possibly induced by environmental and meteorological hazards). The main idea of uncertainty quantification is to quantify how errors in input parameters for the SWE translate into errors in the water level and discharge output variables, thus certifying the predictive capacity of hydraulic models and deterministic forecast abilities. Over the recent years, there has been an increasing interest for uncertainty quantification in hydraulics studies [8–10] to identify, classify, and potentially reduce the main sources of uncertainty in numerical models. Beyond these objectives, there is a need for propagating uncertainty through dynamical models based on partial derivative equations to better understand the hydraulic model structure and response.

For this purpose, standard approaches are sample-based methods, which are non-intrusive to the forward model and derived from Monte Carlo (MC) techniques. Although MC techniques are generic and robust, they are also computationally expensive due to their slow convergence rate as  $1/\sqrt{N_{MC}}$  where  $N_{MC}$  is the size of the sample. Despite the introduction of many-core hardware and parallel simulation tools, they become quickly intractable.

The costs of sample-based methods are significantly reduced when the forward model is replaced by a surrogate model, that can be constructed “off-line,” and evaluated multiple times “on-line” at almost no cost. There are various ways of formulating a deterministic or probabilistic surrogate such as proper orthogonal decomposition [58], Gaussian processes or Kriging [32], Karhunen-Loève expansion [41], polynomial chaos expansion [31]. Polynomial chaos (PC) approach has received much attention lately, especially non-intrusive formulations that do not require modifications to the forward model as opposed to intrusive formulation [28]. Based on homogeneous chaos theory [48], it was first used by [41] to describe stochastic Gaussian variables in finite elements. It was then extended to other distributions [53] using Askey’s scheme and called “generalized polynomial chaos” (gPC) [28, 51] (for simplicity purpose, the gPC-expansion approach is denoted by PC in the following). The PC-expansion method basically replaces the

forward model by a polynomial expansion; it is also powerful to perform global sensitivity analysis as its coefficients relate to the statistics of the target variable [42]. The merits of PC-expansion were demonstrated in various fields, e.g., hydrology [15], structural mechanics [2], computational fluid dynamics [26, 54], hydraulics [31], wildfire propagation [39, 40]. Once identified and quantified, the uncertainty should be reduced, for instance using a data assimilation approach.

The ensemble Kalman filter (EnKF) [21] is a common ensemble-based data assimilation algorithm that provides a correction to the model state and/or parameters; the correction can be regarded as a weight between observations and model predictions calculated via the Kalman gain matrix [35]. The Kalman gain matrix is computed from a stochastic estimate of the background error covariance matrix over an ensemble of input and output variables. A sample of the input uncertain variables is used to achieve an ensemble of forward model evaluations that provide a sample of the output variables. The propagation of the uncertainty is classically achieved with the forward model or, alternatively with a surrogate model. [29] showed that with a crude MC technique for the ensemble generation, the convergence of the EnKF is slow and thus often not compatible with operational constraints or high-dimensional problems. A complementary approach between uncertainty quantification and ensemble-based data assimilation was presented in [30], where a PC-surrogate was used in place of the forward model to reduce the EnKF cost. Such method was implemented in [40] for regional-scale wildfire spread modeling and referred to as PC-EnKF; in the context of grassland controlled burns, the PC-EnKF algorithm replaced the wildfire spread model by a PC surrogate, whose coefficients were computed via Galerkin projection, and achieved convergence in spite of the model nonlinearities with respect to the controlled input parameters (i.e., biomass moisture content, biomass fuel aspect ratio).

The objective of the present work is to validate the PC-expansion strategy for hydraulic modeling of subcritical steady flows assuming that the functional of interest is of finite variance. We aim at demonstrating that the water level probability density function (PDF) and covariance matrix are better estimated with the PC-surrogate model than with a Monte Carlo approach on the forward model given a limited budget of forward model evaluations. The PC-surrogate is here expanded with respect to the upstream discharge and the spatially distributed friction coefficients. The friction coefficients are supposed to be independent random variables that are adjusted to account for various modeling errors. In the present work, the formulation of the PC-surrogate model relies first on the choice of the polynomial basis for the input random variables. Then, the maximum polynomial degree of the expansion that is sufficient to mimic the forward model behavior is identified;

the number of coefficients in the expansion relates to this degree. Finally, the coefficients are computed by least-square regression or pseudo-spectral projection. The performance of these two approaches are compared in this study.

The results of a surrogate model built with a Galerkin projection approach coupled with Gaussian quadrature are shown on two test cases: an idealized channel test case with uniform geometry and a real case on a 50-km reach of the Garonne River (South-West France) with complex bathymetry. The surrogate model for water level is validated in the sense of the  $\ell_2$  norm in the model output space and in the polynomial coefficients space with the appropriate probabilistic measure. The surrogate model is formulated at different locations of the Garonne hydraulic network for uncertainty quantification and sensitivity analysis studies with a particular interest for the Marmande observing station. The water level statistical moments, the PDF and the covariance matrix are estimated with the PC-surrogate at a lower computational cost than with a classical MC approach. In the long run, this work paves the way toward reduced-cost data assimilation in hydraulics for flood forecasting [24] or water resource management provided the Kalman gain is stochastically estimated with the surrogate model in place of the forward model. For model parameter estimation with data assimilation, the Kalman gain matrix represents the statistics between the errors in the parameters and the errors in the simulated hydraulic state at the grid points where the observations are assimilated. In this case, the surrogate model is only necessary at the observing stations. For model state estimation, the Kalman gain matrix represents the statistics between the errors in the hydraulic state over the entire simulation domain. Thus, the surrogate model is necessary at every point of the domain.

The structure of the paper is as follows: Section 2 introduces the basic description of the hydraulic modeling. We recall the SWE for steady flows and present both idealized and Garonne River test cases chosen for this study. The theory and implementation of the PC expansion is presented in Section 3 along with the a posteriori error metrics for evaluating the surrogate performance and its statistics. In Section 4, results for the PC-expansion strategy are presented. Conclusions and perspectives are given in Section 5.

## 2 Hydraulic Modeling

### 2.1 Shallow Water Equations

One-dimensional models for shallow water equations (SWE) are widely used in hydraulic engineering, for instance for flood forecasting [25]. The SWE form a hyperbolic system which may describe subcritical and supercritical flows with hydraulic jumps. In the present work, only

subcritical flows of rivers in plains are considered, consistently with the hydraulic model used at operational level for the Garonne River (see the application in Section 4.2).

#### 2.1.1 One-Dimensional Flow Approximation

One-dimensional SWE are derived from two prime principles: mass conservation and momentum conservation. For most applications in rivers by considering the same aforementioned assumption, the equations are written in terms of discharge (or flow rate)  $Q$  ( $\text{m}^3 \text{s}^{-1}$ ) and hydraulic section  $A$  ( $\text{m}^2$ ) that relates to water level (or water height)  $h$  (m) such that  $A \equiv A(h)$ . The curvilinear abscissa in the simulation domain is denoted by  $s$  ranging from  $s_{in}$  upstream of the river to  $s_{out}$  downstream. The non-conservative form of the one-dimensional SWE reads [1, 22, 44]:

$$\begin{cases} \partial_t A(h) + \partial_s Q = 0 \\ \partial_t Q + \partial_s (Q(h)^2/A(h)) + gA(h)\partial_s h - gA(h)(S_0 - S_f) = 0 \end{cases} \quad (1)$$

with  $g$  the gravity,  $S_0$  the channel slope and  $S_f$  the friction slope. These equations are usually combined with an equation for the friction slope  $S_f$ , here described with the Manning-Strickler formula [14]:

$$S_f = \frac{Q^2}{K_s^2 A(h)^2 R(h)^{4/3}}, \quad (2)$$

where  $R(h)$  is the hydraulic radius ( $R(h) = A(h)/P(h)$ ) written as a function of the wet perimeter  $P(h)$  and  $K_s$  is the Strickler friction coefficient ( $\text{m}^{1/3} \text{s}^{-1}$ ).

In the present work, the MASCARET software is used to simulate the one-dimensional SWE, i.e., the water level  $h$  and discharge  $Q$  over the discrete hydraulic network for  $s \in [s_{in}, s_{out}]$  (the pair  $(h, Q)$  is referred to as the hydraulic state in the following). MASCARET is part of the TELEMAC-MASCARET open-source modeling package developed by EDF (Electricité de France R&D) in collaboration with the CEREMA (Centre d'Etude et d'expertise sur les Risques, l'Environnement, la Mobilité et l'Aménagement), and is commonly used for dam-break wave simulation, reservoir flushing and flooding. The SWE are solved here with the steady kernel of MASCARET based on a finite difference scheme developed by [23]. The hydraulic model requires the following input data: bathymetry, upstream and downstream boundary conditions, lateral inflows, roughness coefficients, and initial condition for the hydraulic state (if the steady state is obtained by convergence of unsteady state). The imperfect description of these data translates into errors in the simulated water level and discharge. It is thus of prime importance to understand the structure of these errors through uncertainty propagation as well as the contribution of each input variable on the hydraulic state variability through sensitivity analysis. This is one of the objectives of the present work.

### 2.1.2 Steady Flow Regime

For steady flow, Eq. 1 simplifies to:

$$\begin{cases} \partial_s Q = 0 \\ \partial_s h = \frac{(S_0 - S_f)}{1 - Fr^2} \end{cases} \quad (3)$$

where  $Fr$  is the dimensionless Froude number

$$Fr^2 = \frac{Q}{gA^3} \frac{\partial A}{\partial h}. \quad (4)$$

The smooth solutions for Eq. 3 over  $s \in [s_{in}, s_{out}]$  are called “backwater curves” when the downstream boundary condition is prescribed in a deterministic way. Assuming the channel geometry is rectangular ( $A = hW$  with  $W$  the channel width) and the slope  $S_0$  is constant implies for steady flow that  $h$  no longer depends on  $s$ . If the boundary conditions from the hydraulic domain are specified in coherence with the balance between gravity and friction forces, Eq. 3, is then recast as the following equation for the water level  $h$  that governs the entire flow, known as the Manning equation:

$$h = \left( \frac{Q}{K_s W \sqrt{S_0}} \right)^{-5/3}. \quad (5)$$

## 2.2 Case Study Description

In the present work, two different case studies are considered: an idealized channel with uniform geometry on the one hand, and a reach of the Garonne River (South-West France) on the other hand.

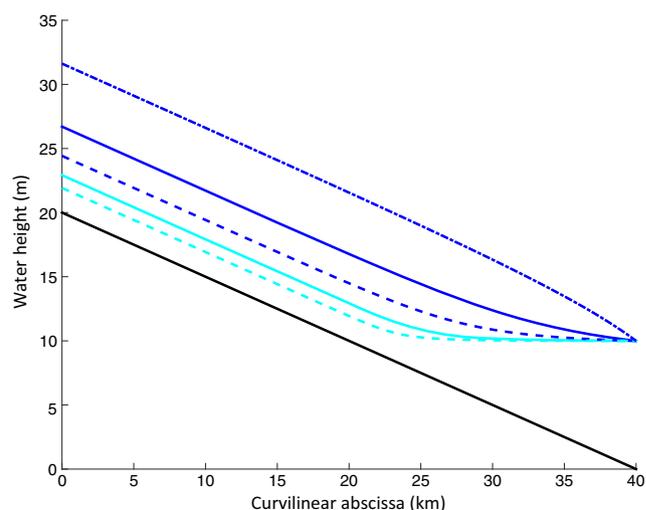
### 2.2.1 Idealized Channel with Uniform Geometry

The first case study is a 40-km open channel with uniform friction coefficient  $K_s$  and uniform slope  $S_0 = 2.5 \times 10^{-4} \text{ m km}^{-1}$ . Upstream and downstream steady boundary conditions are defined by  $Q(s_{in}) = Q$  and  $h(s_{out}) = h_{out}$ . The Manning equilibrium is preserved when  $h_{out}$  is computed with Eq. 5 for each pair  $(Q, K_s)$ . For gradually-varying flow,  $h_{out}$  is set to  $h_{out} = 10 \text{ m}$ . In the following,  $Q$  and  $K_s$  are supposed to be independent physical random variables respectively described by a truncated Gaussian PDF  $\mathcal{N}(4031, 400)$ —to insure the positivity of  $Q$ —and a uniform PDF  $\mathcal{U}(15, 60)$ . It should be noted that the mean values for  $Q$  and  $K_s$  correspond to the Manning equilibrium associated to  $h_{out} = 10 \text{ m}$ . The PDF for the Froude number  $Fr$  was estimated over the sampling, it has a Gaussian-like shape and its probability to exceed 1 is about  $10^{-6}$ . Thus, it is legitimate to assume that all members in the sampling represent a subcritical flow. The water level describes a backwater curve of type  $M_1$  or  $M_2$  [44, 47].

Figure 1 displays the water level  $h$  along the 40-km channel for different values of  $Q$  and  $K_s$ . The bathymetry (with constant slope  $S_0$ ) is represented with a solid black line. The simulated water levels for  $K_s = 15 \text{ m}^{1/3} \text{ s}^{-1}$  and  $K_s = 60 \text{ m}^{1/3} \text{ s}^{-1}$  are represented in blue and cyan, respectively. Solid lines are associated with  $Q = 4000 \text{ m}^3 \text{ s}^{-1}$ ; dashed lines are associated with  $Q = 2000 \text{ m}^3 \text{ s}^{-1}$ . In addition, the simulated water level for  $Q = 8000 \text{ m}^3 \text{ s}^{-1}$  and  $K_s = 15 \text{ m}^{1/3} \text{ s}^{-1}$  is represented in blue dashed-dotted line. The flow is found to be uniform over the upstream part of the channel; the water level follows then a backwater curve to meet the downstream boundary condition. As expected, both friction and discharge have a significant impact on water level; the water level increases when the friction increases (i.e., when  $K_s$  decreases) and also when the discharge increases. When the water level prescribed by the downstream water level condition ( $h_{out} = 10 \text{ m}$ ) is larger than the normal depth given by the Manning equilibrium (Eq. 5), the free surface describes a backwater curve of type  $M_1$  (cyan/blue dashed and solid lines). On the opposite, when the water level is smaller than the normal depth, the water level describes a backwater curve of type  $M_2$  (blue dashed-dotted line).

### 2.2.2 Garonne River Test Case

The second case study is a real hydraulic network over the Garonne River in France. The Garonne River flows from the Pyrenees to the Atlantic Ocean in the area of Bordeaux. It is approximately 647-km long and drains an area



**Fig. 1** Water level (or water height) along the 40-km idealized channel with uniform geometry and the downstream boundary condition  $h_{out} = 10 \text{ m}$ . Solid black line corresponds to bathymetry. Blue lines correspond to simulated water levels for  $K_s = 15 \text{ m}^{1/3} \text{ s}^{-1}$ ; cyan lines correspond to  $K_s = 60 \text{ m}^{1/3} \text{ s}^{-1}$ . Solid lines correspond to  $Q = 4000 \text{ m}^3 \text{ s}^{-1}$ ; dashed lines correspond to  $Q = 2000 \text{ m}^3 \text{ s}^{-1}$ ; and dashed-dotted line corresponds to  $Q = 8000 \text{ m}^3 \text{ s}^{-1}$

of 55,000 km<sup>2</sup>. The present study focuses on a 50-km reach from Tonneins (curvilinear abscissa  $s = 13$  km) to La Réole (curvilinear abscissa  $s = 62$  km) with an observing station at Marmande as illustrated in Fig. 2a. The mean slope over the reach is  $S_0 = 3.3$  m km<sup>-1</sup> and the mean width of the river is  $W = 250$  m. The bank-full discharge is approximately equal to the mean annual discharge ( $\bar{Q} = 1000$  m<sup>3</sup> s<sup>-1</sup>). Despite the existence of active floodplains, this reach can be modeled accurately by a 1-D hydraulic model.

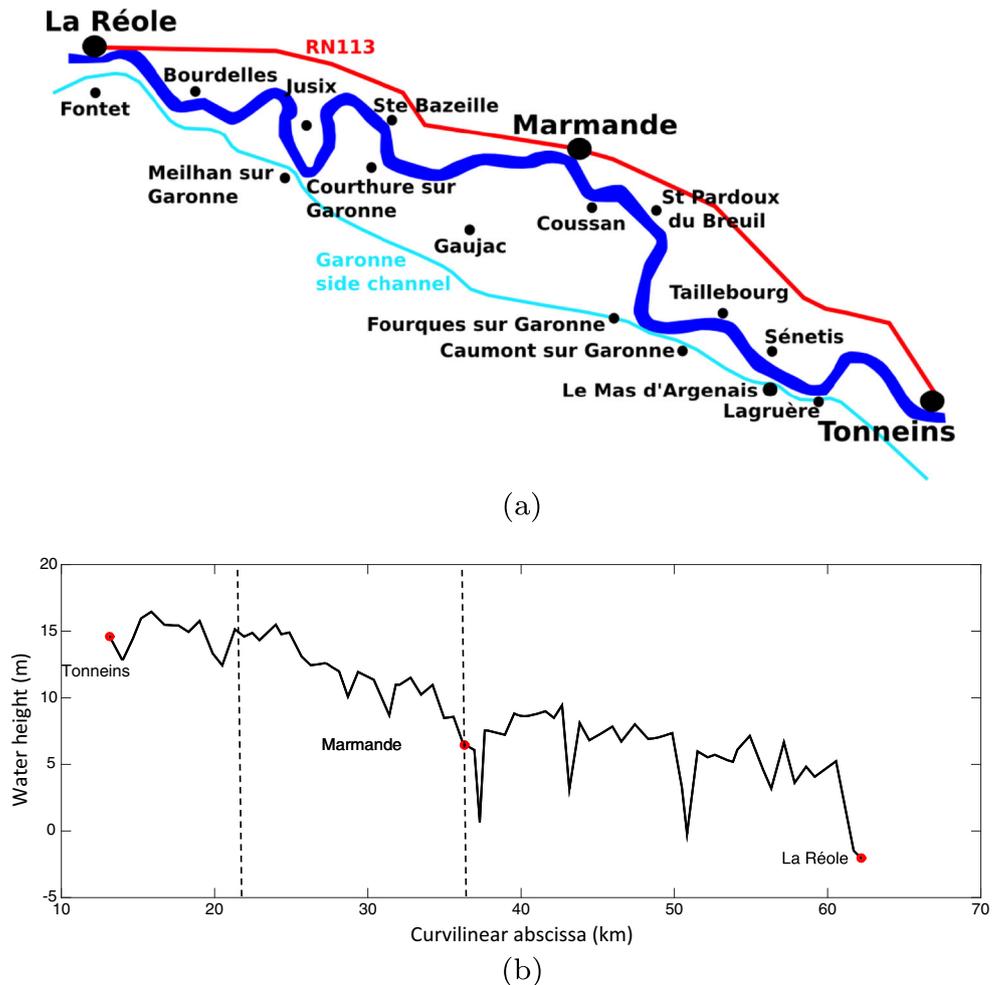
The hydraulic model for the the Garonne River is built from 83 on-site bathymetry cross sections from which the full 1-D bathymetry is interpolated. Friction is prescribed over three portions separated by dashed lines in Fig. 2b for the river channel and the floodplain; it is represented by the Strickler coefficients  $K_{s1}$ ,  $K_{s2}$ , and  $K_{s3}$ . Marmande is located at the beginning of the  $K_{s3}$  portion ( $s = 36$  km). The upstream boundary condition is prescribed with a steady flow  $Q(s_{in}) = Q$ ; the downstream boundary condition is prescribed with a local rating curve  $RC$  established at La Réole that sets  $h(s_{out}) = RC(Q_{out}) = h_{out}$ . The hydraulic

model has been calibrated using channel and floodplain roughness coefficients as free parameters [4]. For this test case, a single MASCARET integration for a typical flood event over 2 to 3 days takes about 30 s CPU time. The main sources of uncertainty considered in the present work are ( $Q, K_{s1}, K_{s2}, K_{s3}$ ); in the following, they are considered as physical random independent variables defined by their PDFs  $\mathcal{N}(4031,400)$  for discharge and  $\mathcal{U}(15, 60)$  for the three friction coefficients. With this choice of PDFs, the flow is subcritical because we consider only high flow conditions.

### 3 Polynomial Chaos (PC) Surrogate

The formulation of the PC surrogate [28, 41, 51] is presented for  $n$  random variables included in the random vector  $\mathbf{x}$ . The Strickler coefficient  $K_s$  and the constant river discharge  $Q$  (permanent flow) are known to be the most relevant sources of uncertainty in water level predictions [8] and are thus included in  $\mathbf{x}$ . If the Strickler coefficient

**Fig. 2** Garonne River test case. **a** Garonne reach between Tonneins (upstream) and La Réole (downstream). Marmande is located at the curvilinear abscissa  $s = 36$  km. **b** Bathymetry profile along the curvilinear abscissa (km) between Tonneins and La Réole. The three portions of  $K_s$  are separated by vertical dashed lines



is uniform over space, the stochastic dimension is  $n = 2$ . For the Garonne hydraulic model, there are three different unknown values for  $K_s$ , so that the stochastic dimension is  $n = 4$ . For simplicity, we will denote the  $n$ -variate random vector  $\mathbf{x}$  with  $n \in \{2, 4\}$ . In the following, details are provided on some of the methods used to compute the PC coefficients, including: choice of polynomial approximation basis, truncation strategy, projection techniques, PC surrogate accuracy.

### 3.1 Polynomial Chaos Expansion for Hydraulic Modeling

The input random variables are supposed to be independent with finite variance and defined in a proper probabilistic space  $(\Omega, \mathcal{A}, \mathcal{P})$  where  $\Omega$  is the event space,  $\mathcal{A}$  is the  $\sigma$ -algebra, and  $\mathcal{P}$  is the probability measure. The random variables are cast in the random vector  $\mathbf{x} \equiv \mathbf{x}(\omega) = [x_1, \dots, x_n]^T$  with  $\omega \in \Omega$  a random event, which evolves in a subset of the Euclidean space  $Z \subseteq \mathbb{R}^n$ . In the following  $\mathbf{x} = [Q, K_s]^T$  for  $n = 2$  and  $\mathbf{x} = [Q, K_{s1}, K_{s2}, K_{s3}]^T$  for  $n = 4$ . The physical model  $\mathcal{M}$  (MASCARET in the present case) maps  $\mathbf{x}$  onto the output space to formulate the spatially distributed water level  $h$  along the channel:

$$h(s) = \mathcal{M}(\mathbf{x}). \tag{6}$$

The spatialized water level  $h$  is a measurable random function of  $\zeta$ . We standardly assume that it belongs to the completion of sums of tensor-products of functions in the curvilinear abscissa  $s$  and orthogonal polynomial functions in  $\zeta_i$ , so that the random scalar function  $h$  has a well-defined PC expansion at a particular abscissa from PC theory [41]. Also, the projections onto any finite-dimensional stochastic subspace spanned by finitely-many (tensor-products of) orthogonal polynomial functions are well-defined as truncations of the following convergent series. Thus  $h$  can be projected onto a stochastic space spanned by orthogonal polynomial functions. The random variables in  $\mathbf{x}$  defined in the input physical space are rescaled in the standard probabilistic space to which the PC framework applies. This standard probabilistic space is noted  $\zeta \equiv \zeta(\omega) = [\zeta_1, \dots, \zeta_n]^T$ , with  $\zeta_i = (x_i - \mu_i)/\sigma_i \sim \mathcal{N}(0, 1)$ ,  $\mu_i$  being the mean and  $\sigma_i$  the Standard Deviation (STD) associated with the  $i$ th random variable  $x_i$ . The marginal PDF associated with each random variable  $\zeta_i$  for  $i = 1, \dots, n$  is denoted by  $\rho_i$ . Random variables  $\zeta_i$  are assumed independent, the joint PDF of the random vector  $\zeta$  can be written as

$$\rho(\zeta) = \prod_{i=1}^n \rho_i(\zeta_i). \tag{7}$$

The PC expansion is computed for the water level  $h$  at a given curvilinear abscissa  $s$ , thus  $h$  is a scalar at a particular

abscissa.  $h$  is approximated in the same probabilistic space as the input random vector  $\mathbf{x}$ :

$$h = \mathcal{M}(\mathbf{x}(\zeta)) = \sum_{i=0}^{\infty} \gamma_i \Psi_i(\zeta), \tag{8}$$

where  $\{\Psi_i\}_{i \geq 0}$  designate the multi-variate polynomial functions that are defined as orthonormal<sup>1</sup> with respect to the joint density  $\rho$  (Eq. 7), where  $i = (i_1, \dots, i_n) \in \mathbb{N}^n$  is the multi-index identifying the polynomial components in the multi-variate space of dimension  $n$ . More information are provided for those polynomials further down. The  $\gamma_i$  terms are the corresponding deterministic coefficients (or modes).

The coefficients  $\gamma_i$  do not change in space for the uniform steady case on which the Manning equation applies. However, they become spatially varying along the river network when considering the idealized channel case with backwater curves and when considering the Garonne River case that features a non-uniform bathymetry. When the water level varies over space, the PC expansion is achieved at each grid point of curvilinear abscissa  $s$  using the same set of forward model evaluations, thus providing spatially dependent expansion coefficients  $\gamma_i(s)$  and water level representation, with a common discretization for  $\gamma$  and  $h$ :

$$h(s) = \sum_{i=0}^{\infty} \gamma_i(s) \Psi_i(\zeta). \tag{9}$$

### 3.2 Truncation Strategy

In practice, the sum in Eq. 8 needs to be truncated to a finite number of terms (noted  $N_{PC}$ ) as follows:

$$h \cong \mathcal{M}_{PC}(\mathbf{x}(\zeta)) = \sum_{i=0}^{(N_{PC}-1)} \gamma_i \Psi_i(\zeta), \tag{10}$$

where the coefficients  $\gamma_i$  are the unknowns to be determined to build the PC surrogate (or metamodel)  $\mathcal{M}_{PC}$  associated with a given total polynomial degree  $P$ .

There are several ways of constructing the polynomial approximation space. The most common choice is to constrain the number of terms  $N_{PC}$  in the PC expansion by the number of random variables  $n$  and by the total degree  $P$ . Considering hierarchical polynomials, the number of terms  $N_{PC}$  to reach the user-defined polynomial degree  $P$  is given by the following algebraic formula:

$$N_{PC} = \frac{(n + P)!}{(n!P!)} \tag{11}$$

meaning that all polynomials involving the  $n$  random variables of total degree less or equal to  $P$  are considered in the PC expansion.

<sup>1</sup>Rather than polynomial orthogonality, we rely in this paper on “orthonormal” polynomial basis for convenience.

The multi-index  $i = (i_1, \dots, i_n)$  of absolute value  $|i| = i_1 + \dots + i_n \in \mathbb{N}$  identifies the components among the multivariate polynomials  $\{\Psi_i\}_{i \geq 0}$  that shall be of total degree less or equal to  $P$  (i.e.  $|i| \leq P$ ). Thus, the set of selected multi-indices corresponding to the number of terms  $N_{PC}$  is defined as

$$\mathcal{I}^P = \{i \in \mathbb{N}^n : |i| \leq P\}. \tag{12}$$

Equation 10 is thus equivalent to

$$\mathcal{M}_{PC}(\mathbf{x}(\boldsymbol{\zeta})) = \sum_{i \in \mathcal{I}^P} \gamma_i \Psi_i(\boldsymbol{\zeta}). \tag{13}$$

Following this standard truncation strategy, building the approximation space may be too costly for high dimensional problems. New truncation strategies are reported in the literature to reduce the number of terms in the PC expansion (thereby to decrease the computational cost of the PC method) as well as conditioning issues when  $n$  increases. These strategies explore alternative ways of selecting the  $N_{PC}$  polynomials by limiting high-order interaction terms (e.g., hyperbolic truncation scheme [7]) or through the construction of sparse bases [6, 34]. In any case, analyzing how the truncation error evolves with respect to  $P$  and for which value of  $P$  the convergence of the PC expansion is achieved is required. For this purpose, a sensitivity analysis of the PC expansion accuracy is carried out with respect to the total degree  $P$  in the following, for both the idealized channel case and the Garonne River application (Section 4). The metrics used to evaluate the truncation error are presented in Section 3.7.

### 3.3 Choice of a Polynomial Basis

The knowledge of the coefficients  $\{\gamma_i\}_{i \in \mathcal{I}^P}$  fully characterizes the process  $h$ . Optimal efficiency is obtained when the chosen polynomial approximation forms an orthonormal basis with respect to the distribution of the random variable  $h$ . Unfortunately, this underlying probability measure is unknown and therefore it is common practice to choose a polynomial basis, orthonormal with respect to the joint PDF  $\rho(\boldsymbol{\zeta})$ , i.e. with regard to the following inner product  $\langle \cdot, \cdot \rangle$ :

$$\langle \Psi_i(\boldsymbol{\zeta}), \Psi_j(\boldsymbol{\zeta}) \rangle = \int_Z \Psi_i(\boldsymbol{\zeta}) \Psi_j(\boldsymbol{\zeta}) \rho(\boldsymbol{\zeta}) d\boldsymbol{\zeta} = \delta_{ij}, \tag{14}$$

with  $\delta_{ij}$  the Kronecker delta-function extended to the  $n$ -variate case. The choice for the basis functions is thus an essential step in the PC method and depends on the probability measure of the random variables. For instance, according to the Askey’s scheme [54], the Hermite polynomials make the optimal basis for random variables following Gaussian distribution, and the Legendre polynomials are the counterpart for uniform distribution [53].

In practice, the orthonormal basis is built using the tensor product of one-dimensional polynomial functions such that

the  $i$ th two-dimensional basis function  $\Psi_i$  can be expressed as follows [51]:

$$\Psi_i \equiv \Psi_i(\boldsymbol{\zeta}) = \otimes_{l=1}^n \Phi_{i_l}^{(l)}(\zeta_l), \tag{15}$$

with  $\Phi_{i_l}^{(l)}$  the one-dimensional polynomial basis associated with the  $l$ th random variable and  $i_l$  the corresponding one-dimensional index varying between 0 and  $P$ . There are alternative approaches to lower the basis cardinality, e.g., [7].

There is an intrinsic correspondence between the multi-index  $i$  and the one-dimensional index  $i_l$  to properly identify the polynomials involving the  $l$ th random variable. For instance, if  $n = 2$  with  $\mathbf{x} = [Q, K_s]^T$ , Eq. 15 is recast as the following form:

$$\Psi_i \equiv \Psi_i(\boldsymbol{\zeta}) = \Phi_{i_1}^{(1)}(\zeta_1) \otimes \Phi_{i_2}^{(2)}(\zeta_2), \tag{16}$$

with  $\Phi_{i_1}^{(1)} = P_{i_1}$  corresponding to the 1-D basis functions for  $K_s$  (i.e., Legendre polynomials) and  $\Phi_{i_2}^{(2)} = H_{i_2}$  corresponding to the 1-D basis functions for  $Q$  (i.e., Hermite polynomials). There are several ways of defining the indexation reported in the literature [6, 7]. In the following, for  $n = 2$ , the rescaled random vector is  $\boldsymbol{\zeta} = (\zeta_1, \zeta_2)$  and the following notation will be preferred:

$$h \cong \mathcal{M}_{PC}(\mathbf{x}(\boldsymbol{\zeta})) = \sum_{i_1, i_2=0}^P \gamma_{i_1 i_2} P_{i_1}(\zeta_1) \otimes H_{i_2}(\zeta_2). \tag{17}$$

### 3.4 Computation of the Expansion Coefficients

For a given polynomial basis (Eq. 10), there exists several possible strategies to construct the multivariate polynomial approximation, and compute the coefficients  $\{\gamma_i\}_{i \in \mathcal{I}^P}$ . The present work focuses on non-intrusive approaches, meaning that no modification of the legacy solver is required and that the coefficients are computed from an ensemble of model evaluations  $h^{(k)} = \mathcal{M}(\mathbf{x}^{(k)})$  (referred to as the “training sample”), with the index  $k = 1, \dots, N_e$  referring to the  $k$ th realization of the random vector  $\mathbf{x}$ .

In the literature, two main families of non-intrusive approaches are reported: interpolant collocation methods [37, 43, 52] and  $\ell_2$ -minimization methods. The latter include both regression and pseudo-spectral projection. In the following, we briefly recall ordinary and regularized least-square approximations and the Galerkin projection approximation coupled to Gaussian quadrature techniques that folds into the regression and the projection category, respectively.

**Least-Square Minimization** With this approach, the estimation of the coefficients  $\gamma_i$  is achieved at once solving a least-square (LS) minimization problem in some  $\ell_2$  norm [3, 13]. In this case, the “approximation” error (or residual)  $\epsilon_{PC}$

is defined as the distance between the (exact) model evaluations and the PC surrogate estimations for a finite set of randomly sampled input random variables of size  $N_e = N_{ls}$  (where the subscript  $ls$  stands for least square) such that:

$$\epsilon_{PC} = \mathcal{M}(\mathcal{X}) - \mathcal{M}_{PC}(\mathcal{X}) = \mathcal{H} - \sum_{i=0}^{(N_{PC}-1)} \gamma_i \Psi_i(\xi) \equiv \mathcal{H} - \boldsymbol{\gamma}^T \Psi, \tag{18}$$

where  $\mathcal{X} = [\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N_{ls})}]^T$  is the matrix of input samples of the random vector  $\mathbf{x}$  and  $\mathcal{H} = [h^{(1)}, \dots, h^{(N_{ls})}]^T$  is the vector of associated water level outputs.  $\boldsymbol{\gamma} = [\gamma_0, \dots, \gamma_{N_{PC}-1}]^T$  is the vector of the  $N_{PC}$  unknown coefficients and  $\Psi$  is the matrix of size  $N_{PC} \times N_{ls}$  assembling the values of all the orthonormal polynomials at the design nodes  $\Psi_{ik} = \Psi_i(\xi^{(k)})$ , with  $i = 0, \dots, (N_{PC} - 1)$  and  $k = 1, \dots, N_{ls}$ .

Determining  $\boldsymbol{\gamma}$  as the ordinary least-square solution of Eq. 18 is a standard problem in linear algebra: one has to minimize

$$LS(\boldsymbol{\gamma}) = \epsilon_{PC}^T \epsilon_{PC} = (\mathcal{H} - \boldsymbol{\gamma}^T \Psi)^T (\mathcal{H} - \boldsymbol{\gamma}^T \Psi) \tag{19}$$

such that:

$$\frac{\partial LS}{\partial \boldsymbol{\gamma}} = -2 \Psi^T \mathcal{H} + 2 \Psi^T \Psi \boldsymbol{\gamma} = 0, \tag{20}$$

leading to the following solution:

$$\boldsymbol{\gamma} = (\Psi^T \Psi)^{-1} \Psi^T \mathcal{H}. \tag{21}$$

One should ensure that the matrix  $(\Psi^T \Psi)$  is invertible; the solution depends on the pre-conditioning of the matrix  $\Psi$  and thereby on the number/choice of the realizations  $\mathbf{x}^{(k)}$  ( $k = 1, \dots, N_{ls}$ ). The input space sampling is achieved in the present work, the input space exploration is fulfilled thanks to a MC sampling-based approach; the converged estimation of the PC coefficients requires a large size sample, which we can afford in our case due to low computational cost of our deterministic model MASCARET. When the forward model is more expensive, other sampling methods are available: e.g., Quasi MC, Latin Hypercube Sampling. An iterative MC approach is performed here in order to monitor the convergence of the coefficients which are re-computed as the size of the sample  $E_{ls}$  increases from 10 to  $N_{ls}$ .

The least-square algorithm proceeds as follows:

- Choose the total polynomial degree  $P$
- Choose the sample  $E_{ls}$  of size  $N_{ls}$  corresponding to realizations of the random vector  $\mathbf{x}$  and formulate the input vector  $\mathcal{X}$ .
- Integrate the forward model for each member in  $E_{ls}$  and formulate the output vector  $\mathcal{H}$
- Construct the information matrix  $\Psi$  by evaluating the polynomial basis at the normalized nodes corresponding to the members  $\mathcal{X}$  in the physical space

- Compute the coefficient vector  $\boldsymbol{\gamma}^l$  of length  $N_{PC}$  using Eq. 21 and formulate  $\mathcal{M}_{PC}$

On a final note, the aforementioned approach deals with “noiseless” evaluations of the target function; few papers consider noisy data samples [33].

**Least-Angle Regression** One way to alleviate the curse of dimensionality induced by the parametric uncertainty is to take advantage of the potential sparsity of the solution structure. Indeed, when the stochastic dimension  $n$  of the problem is high, the solution is sometimes “sparse” (or near-sparse) at the stochastic level. This means that it may be accurately represented with only few terms when linearly expanded into a stochastic approximation space, such as the one encompassed by a gPC basis. Promising approaches for solving this kind of problems involve compressed sensing techniques, also known under the names of  $\ell_1$ -minimization, convex relaxation, and  $\ell_1$ -regularized least-squares minimization. Relatively recent results in compressed sensing have also made it clear that sparse functions requiring reasonable number of approximation basis, may be accurately recovered from much fewer model simulations  $N_e$  than necessary for classical solution methods [11, 12, 16]. Several research groups have recently been using compressed sensing in a gPC framework [17, 55, 56]. The efficiency of this approximation depends on the type and cardinality of the gPC approximation basis selected [7, 27] and the choice of the samples to be used.

The least absolute shrinkage and selection operator (LASSO) algorithm [45] is in this case an attractive modification of the ordinary least-square formulation that constrains the sum of the absolute regression coefficients. Weighted versions exist that make the approximation even more robust [46]. Another related model selection algorithm, the least-angle regression (LAR), is also very efficient in our framework. Here, the idea is to use the LAR formulation to identify, in an iterative manner, an optimal sparse basis and then to compute a limited number of coefficients using a standard regression method. For this purpose, an active set of coefficients should be first selected.

The  $N_{PC}$  polynomial functions in the standard PC expansion (i.e., when the basis is full) form the candidate set. At each iteration of the LAR algorithm, a polynomial is selected among these  $N_{PC}$  terms, based on its correlation with the current residual  $\epsilon_{PC}$ , and is then added to the active set of polynomials (also referred to as “regressors”). The corresponding coefficients are computed so that every active polynomial is equicorrelated with the current residual. Thus, LAR builds a collection of surrogates that are less and less sparse along the iterations. Iterations stop either when the full basis has been looked through. The final step is to choose the best surrogate according to an error criterion (see Section 3.7). The corresponding coefficients are

computed so that every active polynomial is equicorrelated with the current residual. Thus, LAR builds a collection of surrogates that are less and less sparse along the iterations. Iterations stop either when the full basis has been looked through or when the maximum size of the experimental design  $E_{lar}$  has been reached. The final step is to choose the best surrogate according to an error criterion (Section 3.7). An hybrid LAR (hybrid-LAR) procedure, proposed in [7], is retained for this work: the LAR is only used to select a sparse set of predictors, but not to estimate the coefficients, which are obtained from least-square estimates on this set of predictors. Moreover, the detection/selection of the optimal LAR surrogate is obtained from a modified cross-validation criterion that avoids multiple calls to the hybrid-LAR procedure. As suggested in [7], the corrected leave-one-out error (LOO) is used as the cross-validation criterion to select the optimal LAR surrogate (Section 3.7).

The plain LAR algorithm without cross-validation selection proceeds as follows:

- Choose the total polynomial degree  $P$
- Build the fixed experimental design  $E_{lar}$ 
  - Build the LAR input vector  $\mathcal{X}$  of size  $N_{lar}$
  - Integrate the forward model for each member of  $\mathcal{X}$  and form the output vector  $\mathcal{H}$
  - Evaluate all components of the polynomial basis over the sample  $\mathcal{X}$  to form the matrix of elements  $\Psi_{ik} = \Psi_i(\zeta^{(k)})$ , with  $i = 0, \dots, (N_{PC} - 1)$  and  $k = 1, \dots, N_{lar}$ .
- Initialize the algorithm
  - Set the coefficient vector  $\boldsymbol{\gamma}$  to 0
  - Set the residual vector  $\epsilon_{PC}$  (of size  $N_{lar}$ ) to the forward model output vector  $\mathcal{H}$  ( $\epsilon_{PC} = \mathcal{H} - \mathcal{M}_{PC}(\boldsymbol{\gamma}) = \mathcal{H}$  since  $\boldsymbol{\gamma} = 0$ )
  - Set the active set (or regressors)  $\Psi_{lar} = \emptyset$
- Iterate: For  $l = 1, \min(N_{PC}, N_{lar} - 1)$ , do the following steps:
  - For all polynomial candidates  $i = 0, (N_{PC} - 1)$ , compute the correlation between the current residual  $\epsilon_{PC}$  and the  $i$ th column of  $\Psi$  ( $\Psi_{ki}$  for  $k = 1, \dots, N_{lar}$ ) evaluated on  $E_{lar}$
  - Add to the active set  $\Psi_{lar}$ , the vector  $\Psi_j$  (the  $j$ th column of  $\Psi$ ) that is the most correlated with the current residual  $\epsilon_{PC}$
  - Compute the vector of partial coefficients  $\boldsymbol{\gamma}^l$  (or descent direction) associated with  $\Psi_{lar}$  by minimizing  $LS(\boldsymbol{\gamma}) = (\mathcal{H} - \boldsymbol{\gamma}^T \Psi_{lar})^T (\mathcal{H} - \boldsymbol{\gamma}^T \Psi_{lar})$ : this consists in taking the largest possible step in the direction of  $\Psi_{lar}$  until the correlation between  $\epsilon_{PC}$  and  $\Psi_{lar}$  becomes smaller than the correlation between some

other vector  $\Psi_k$  in the candidate set and  $\Psi_{lar}$ . This provides the vector of partial coefficients

$$\boldsymbol{\gamma}^l = \left( \Psi_{lar}^T \Psi_{lar} \right)^{-1} \Psi_{lar}^T \epsilon_{PC}$$

- Update the residual  $\epsilon_{PC} = \epsilon_{PC} - (\boldsymbol{\gamma}^l)^T \Psi_{lar}$  and store it for the current iteration
- Update the coefficient vector  $\boldsymbol{\gamma}$  with the contribution  $\boldsymbol{\gamma}^l$ :  $\boldsymbol{\gamma} = \boldsymbol{\gamma} + \delta^l \boldsymbol{\gamma}^l$  with  $\delta^l$  the descent step defined such that every active polynomial is equicorrelated with the current residual

It is worth noting that for a given total polynomial degree  $P$ , the LAR algorithm is valuable if the effective number of terms in the PC expansion (noted  $N_{reg}$ ) is significantly less than the  $N_{PC}$  terms of the standard truncation scheme (Eq. 11). It is also worth mentioning that if  $N_{reg} > N_{PC}$ , then the last iteration of the LAR algorithm computes a PC surrogate on the full basis corresponding to the ordinary least-square solution. In general when using LAR,  $N_{PC} > N_{reg}$ , implying that  $N_{reg}$  surrogate candidates are formulated when iterations stop. More details and options about the LAR and hybrid-LAR techniques may be found in [7, 20].

**Galerkin Pseudo-Spectral Projection** This Galerkin-type projection relies on the orthonormality property of the polynomial basis. In this framework, the coefficients are computed one at a time. The  $i$ th PC coefficient  $\gamma_i$  is computed as follows:

$$\gamma_i = \langle h, \Psi_i(\boldsymbol{\zeta}) \rangle, \tag{22}$$

using the inner product definition given in Eq. 14 and using the basis orthonormality property. The term  $\langle h, \Psi_i(\boldsymbol{\zeta}) \rangle$  involves multi-dimensional integrals. There exists many different numerical methods to evaluate those integrals. A straightforward choice is the one of numerical quadratures, and in particular tensor-based Gauss-quadratures that are powerful at integrating exactly multi-dimensional polynomial forms; the inner product is approximated as:

$$\langle h, \Psi_i(\boldsymbol{\zeta}) \rangle \cong \sum_{k=1}^{N_e} \mathcal{M}(\mathbf{x}^{(k)}) \Psi_i(\boldsymbol{\zeta}^{(k)}) w^{(k)}, \tag{23}$$

where  $h^{(k)} = \mathcal{M}(\mathbf{x}^{(k)})$  corresponds to the forward model integration evaluated at the  $k$ th quadrature root  $\mathbf{x}^{(k)}$  with its associated weight  $w^{(k)}$ . Here,  $k$  varies between 1 and  $N_e$ , with  $N_e = (N_{quad})^n \equiv E_{quad}$  when  $N_{quad}$  quadrature points are selected for each input random variable. In order to guarantee that a polynomial function can properly be represented by the PC expansion,  $N_{quad}$  should be constrained by the total polynomial degree  $P$  such that:

$$N_{quad} \geq (P + 1). \tag{24}$$

In this work, a standard quadrature is used since only a few quadrature points are sufficient to obtain a highly accurate

result as shown for instance in the following [2, 6, 50]. This method will only be applied for the case  $n = 2$ . For higher dimensions, full tensorization methods suffer from the curse of dimensionality. Alternatives exist that rely on the use of sparse tensorization and/or nested grids [38, 49] and could be applied to alleviate this issue.

The spectral projection strategy proceeds as follows:

- Choose the total polynomial degree  $P$
- Choose the number of quadrature roots  $N_{quad}$  (according to Eq. 24) per stochastic dimension and identify the roots in the normalized space  $\zeta$
- Map the normalized roots onto the physical space to formulate the quadrature root vector  $\mathcal{X} = [\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N_{quad})}]^T$ , thus defining the input training sample set  $E_{quad}$
- Integrate the forward model for the quadrature roots  $\mathcal{X}$  and formulate the output vector  $\mathcal{H} = [h^{(1)}, \dots, h^{(N_{quad})}]^T$
- Compute the PC coefficient  $\gamma_i$  ( $i = 0, \dots, N_{PC} - 1$ ) using Eqs. 22–23, and formulate  $\mathcal{M}_{PC}$ .

### 3.5 Surrogate Model Notations

In the following, the PC surrogate model is denoted with a superscript that specifies the total polynomial degree, the method for estimating the coefficients and the size of the sample used for this computation. Moreover, for simplicity purposes, sample size is given in logarithmic form for the least-square method, e.g.,  $\mathcal{M}_{PC}^{ls3, P6}$  denotes a PC expansion of total polynomial degree  $P = 6$  with coefficients computed from a least-square regression method over a sample  $E_{ls}$  of size  $N_{ls} = 10^3$ . In the same line of thought,  $\mathcal{M}_{PC}^{quad7, P6}$  denotes a PC expansion of total polynomial degree  $P = 6$  with coefficients computed from a pseudo-spectral projection/quadrature method over a sample  $E_{quad}$  of size  $N_{quad} = 7$  for each input variable following Eq. 24.

### 3.6 Statistical Moments

Once the PC surrogate is built, the water level statistical moments are estimated statistically or using the orthonormality properties of the PC expansion.

**Stochastic Estimation** The water level statistical moments can be stochastically estimated using a validation sample  $E_{MC}$  of size  $N_{MC}$  (here  $N_{MC} = 10^5$ ). The validation sample is generated through a standard MC approach based on the random parameters  $K_s$  and  $Q$ . The water level  $h$  can be computed either integrating the forward model  $\mathcal{M}$  or the PC surrogate  $\mathcal{M}_{PC}$  (computed from Eq. 10) for the validation sample  $E_{MC}$  such that  $h^{(k)} = \mathcal{M}(\mathbf{x}^{(k)})$  or  $h_{PC}^{(k)} = \mathcal{M}_{PC}(\mathbf{x}^{(k)})$  for the  $k$ th member of  $E_{MC}$ .

We can stochastically evaluate the first moments of the water level PDF, i.e., the mean value, the STD, the skewness, and the kurtosis for the forward model  $(\mu_h, \sigma_h, skew_h, kurt_h)$  or the PC surrogate model  $(\mu_{h_{PC}}, \sigma_{h_{PC}}, skew_{h_{PC}}, kurt_{h_{PC}})$  as formulated here:

$$\mu_{h_{PC}} = \frac{1}{N_{MC}} \sum_{k=1}^{N_{MC}} h_{PC}^{(k)}, \tag{25}$$

$$\sigma_{h_{PC}} = \sqrt{\frac{1}{N_{MC}-1} \sum_{k=1}^{N_{MC}} (h^{(k)} - \mu_{h_{PC}})^2}, \tag{26}$$

$$skew_{h_{PC}} = \frac{1}{N_{MC} (\sigma_{h_{PC}})^3} \sum_{k=1}^{N_{MC}} (h_{PC}^{(k)} - \mu_{h_{PC}})^3, \tag{27}$$

$$kurt_{h_{PC}} = \frac{1}{N_{MC} (\sigma_{h_{PC}})^4} \sum_{k=1}^{N_{MC}} (h_{PC}^{(k)} - \mu_{h_{PC}})^4. \tag{28}$$

**Analytic Derivation** – Thanks to the polynomial approximation basis orthonormality, we can compute the statistical moments of the water level  $\mathbf{h}$  from the coefficients:

$$\mu_{h_{PC}} = \gamma_0, \tag{29}$$

$$\sigma_{h_{PC}} = \sqrt{\sum_{\substack{i \in \mathcal{I}^P \\ i \neq 0}} \gamma_i^2}, \tag{30}$$

$$skew_{h_{PC}} = \frac{1}{\sigma_{h_{PC}}^3} \sum_{\substack{i \in \mathcal{I}^P \\ i \neq 0}} \sum_{\substack{j \in \mathcal{I}^P \\ j \neq 0}} \sum_{\substack{k \in \mathcal{I}^P \\ k \neq 0}} \gamma_i \gamma_j \gamma_k \mathbb{E} [\Psi_i(\zeta) \Psi_j(\zeta) \Psi_k(\zeta)], \tag{31}$$

$$kurt_{h_{PC}} = \frac{1}{\sigma_{h_{PC}}^4} \sum_{\substack{i \in \mathcal{I}^P \\ i \neq 0}} \sum_{\substack{j \in \mathcal{I}^P \\ j \neq 0}} \sum_{\substack{k \in \mathcal{I}^P \\ k \neq 0}} \sum_{\substack{l \in \mathcal{I}^P \\ l \neq 0}} \gamma_i \gamma_j \gamma_k \gamma_l \mathbb{E} [\Psi_i(\zeta) \Psi_j(\zeta) \Psi_k(\zeta) \Psi_l(\zeta)], \tag{32}$$

where the set of multi-indices  $\mathcal{I}^P$  to which belong  $i, j, k$  and  $l$  is defined in Eq. 12.

To measure the extent of variability in a given sample, one may be interested in computing the coefficient of variation

(CV), also known as relative STD and defined as the ratio of the STD  $\sigma_{h_{PC}}$  to the mean  $\mu_{h_{PC}}$ .

When the water level is varying in the space, the covariance matrix for water level along the curvilinear abscissa can be estimated from the PC coefficients. [51] provides an analytical estimation of the covariance between the water level at two different grid points  $m$  and  $n$ :

$$\text{Cov}(h(s_m), h(s_n)) = \sum_{\substack{i \in \mathcal{I}^P \\ i \neq 0}} \gamma_i(s_m) \gamma_i(s_n), \tag{33}$$

To evaluate the contribution of each random variable to the water level variance  $(\sigma_{h_{PC}})^2$  and carry out a global sensitivity analysis, we use Sobol’ indices directly derived from PC coefficients [42]. For any given random variable  $x_{i_s}$  in  $\mathbf{x}$ , Sobol’ index reads:

$$S_{x_{i_s}} = \frac{1}{\sigma_{h_{PC}}^2} \sum_{i \in \mathcal{I}_{i_s}} (\gamma_i)^2, \tag{34}$$

where  $\mathcal{I}_{i_s}^P$  is the set of multi-indices such that  $i \in \mathcal{I}^P$  and only the index  $i_s$  is nonzero (corresponding to the polynomials  $\Psi_i(\boldsymbol{\xi})$  that only depend on the input random variable  $x_{i_s}$ ). For instance, in the present case  $n = 2$ , Sobol’ indices  $S_Q$  and  $S_{K_s}$  for  $Q$  and  $K_s$  (associated with index  $i_1$  and index  $i_2$ , respectively) are defined as follows:

$$S_Q = \frac{1}{\sigma_{h_{PC}}^2} \sum_{i_1 \geq 1, i_2 = 0} (\gamma_{i_1 0})^2; \quad S_{K_s} = \frac{1}{\sigma_{h_{PC}}^2} \sum_{i_2 \geq 1, i_1 = 0} (\gamma_{0 i_2})^2. \tag{35}$$

$S_{K_s}$  (or  $S_Q$ ) corresponds to a first-order sensitivity index by evaluating the stand alone effect of  $K_s$  (or  $Q$ ) on the water level variance measured by  $(\sigma_{h_{PC}})^2$ .

### 3.7 Error Metrics

PC theory guarantees the convergence of the infinite expansion with respect to the standard  $\ell_2$  error norm (mean square convergence) [51]. In practice, due to the truncation, the accuracy of the PC surrogate  $\mathcal{M}_{PC}$  must be assessed over an independent sample  $E_{MC}$  and is evaluated in  $\ell_2$ . Different error metrics are presented hereinafter; their definition is based on the orthonormality property of the basis functions when building the PC-expansion.

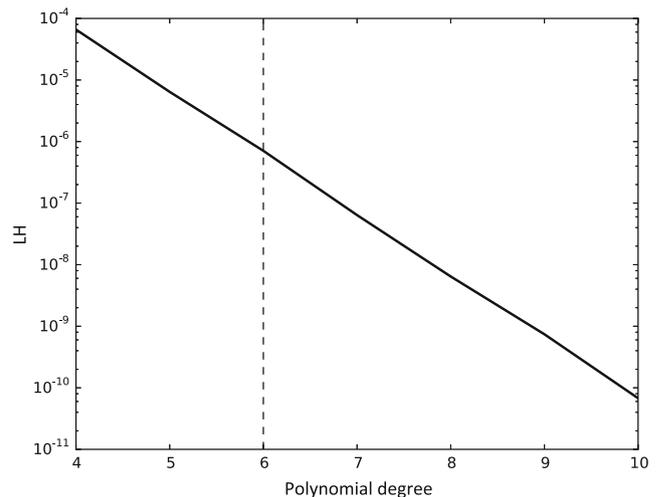
The criterion ( $LC$ ) is computed on the PC coefficient space of  $\gamma_{i_1 i_2}$  for a PC expansion of truncation  $P$  with respect to a reference PC surrogate of equal truncation but different sampling (in terms of size and type of sampling MC random sampling for least-square regression approach, Gaussian quadrature for pseudo-spectral projection). The criterion ( $LH$ ) is computed on the water level variable space. The leave-one-out error norm ( $LOO$ ) is also introduced when applying the LAR method. Errors  $LH$  and

$LOO$  estimate both truncation and sampling (or coefficient) errors, while  $LC$  estimate sampling errors that are due to aliasing higher-order modes and/or insufficient quadratures or regression points. It should be noted that when the computation of the PC-expansion coefficients is achieved over a large sample, it is fair to use  $LH$  to estimate truncation error only (Figs. 3, 4, 5, and 6). On the contrary, when the total polynomial degree  $P$  has been set and the number of quadrature or regression points varies,  $LH$  is used to estimate the standalone sampling error (Fig. 7).

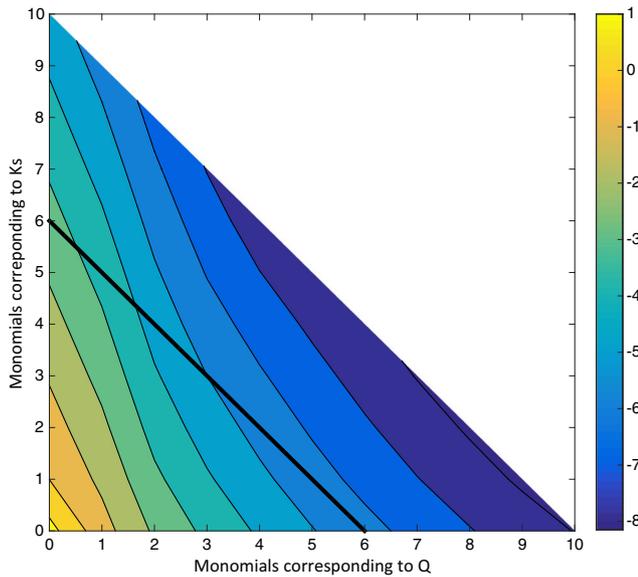
**Water Level Metrics ( $LH$ )** Assuming the forward model and the surrogate model are integrated over the validation sample  $E_{MC}$  of size  $N_{MC}$  to represent the water level  $h$  and  $h_{PC}$  respectively, the standard  $\ell_2$  error norm is defined as follows:

$$LH = \frac{1}{N_{MC}} \sum_{k=1}^{N_{MC}} (h^{(k)} - h_{PC}^{(k)})^2. \tag{36}$$

In order to avoid integrating the forward model over the validation sample  $E_{MC}$ , the  $LOO$  error is introduced. This solution is of particular interest when one cannot afford, due to computational cost limitations, to re-sample the forward model apart from the simulations dedicated to the PC formulation using sample  $E_{ls}$  or  $E_{quad}$ . For the purpose of this academic study, the validation sampling and associated model integrations were carried out for  $n = 2$  and  $n = 4$ . The  $LOO$  and the  $LH$  metrics are compared to assess the

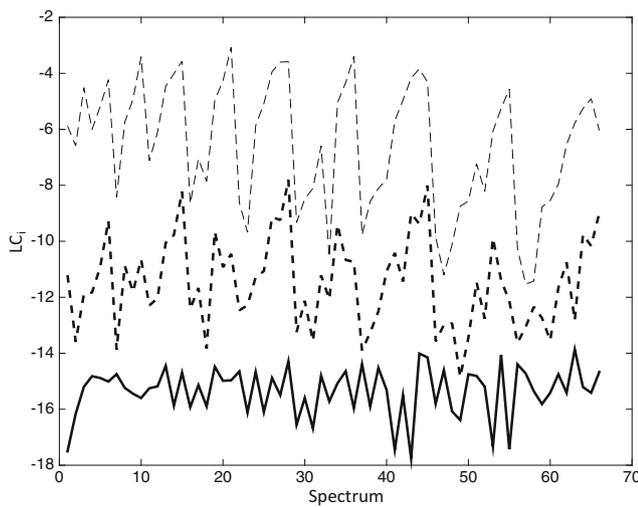


**Fig. 3** Truncation error measured by the water level error  $LH$  (Eq. 36) in logarithmic scale with respect to the total polynomial degree  $P$ -Idealized channel case with uniform steady flow (Manning equation).  $LH$  is computed over the validation sample  $N_{MC} = 10^5$  for the least-square surrogate  $\mathcal{M}_{PC}^{ls,P}$  formulated over the training sample  $N_{ls,ref} = 10^5$  for  $P$  varying between 4 and 10. The vertical dashed line corresponds to the selected  $P$  for further results

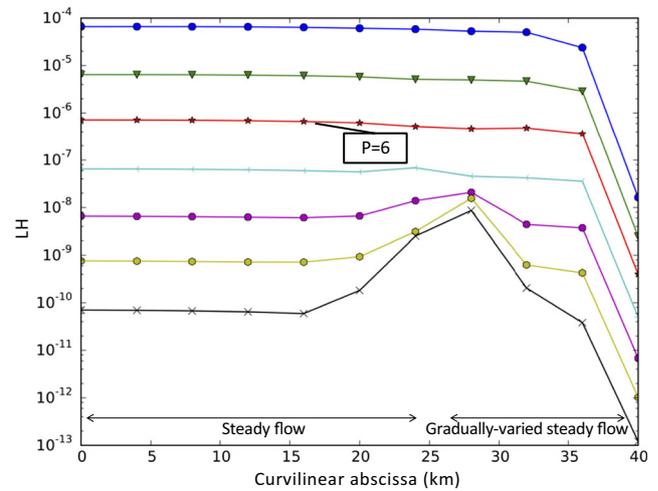


**Fig. 4** Spectrum of coefficients for the least-square surrogate—idealized channel case with uniform steady flow (Manning equation). The surrogate is formulated over the training sample  $N_{ls,ref} = 10^5$  and a total polynomial degree  $P = 10$ . The colorbar describes the amplitude of the coefficients in logarithmic scale. The Hermite monomials (corresponding to  $Q$ ) are displayed on the  $x$ -axis; the Legendre monomials (corresponding to  $K_s$ ) are displayed on the  $y$ -axis. The bivariate coefficients  $\{\gamma_{i_1, i_2}\}_{0 < i_1, i_2 < P}$  are displayed in the lower triangular matrix along  $x$ -/ $y$ -axis. The oblique solid black line represents  $P = 6$

validity of using  $LOO$  as a reliable metric when integrating a validation sample is not possible, i.e., in the case of  $n = 4$ . When the dimension of the input space increases



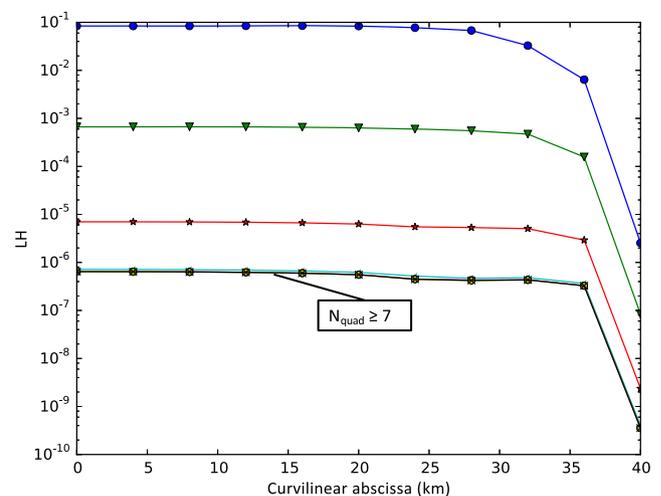
**Fig. 5** Comparison of  $LC_i$  elements in logarithmic scale between the LS surrogate  $\mathcal{M}_{PC}^{ls2, P10}$  (thin dashed line),  $\mathcal{M}_{PC}^{ls3, P10}$  (thick dashed line), and the quadrature surrogate  $\mathcal{M}_{PC}^{quad11, P10}$  (thick solid line)—idealized channel case with uniform steady flow (Manning equation). The reference coefficients are computed with  $\mathcal{M}_{PC}^{ls5, P10}$ . The  $x$ -axis is the coefficient index within the expansion spectrum (66 coefficients for  $P = 10$ )



**Fig. 6** Truncation error along the 40-km channel illustrated by the water level error  $LH$  (Eq. 36) in logarithmic scale for varying total polynomial degree  $P$ -idealized channel case with gradually-varied steady flow.  $LH$  is computed every 4 km over the validation sample  $N_{MC} = 10^5$  for the least-square surrogate  $\mathcal{M}_{PC}^{ls5, P}$  formulated over the training sample  $N_{ls,ref} = 10^5$  for  $P = 4, \dots, 10$  (top curve in blue for  $P = 4$  to bottom curve in black for  $P = 10$ )

from  $n = 2$  to  $n = 4$ , the number of coefficients rapidly grows, thus increasing the cost of both least-square regression and quadrature-based Galerkin projection methods. As a consequence, the LAR method is preferred; the decomposition on a sparse basis relying on a limited number of PC coefficients.

The  $LOO$  technique [7, 18] requires the formulation of several surrogates. Each surrogate is built excluding one point out of the training sample  $E_{lar}$ ; the accuracy



**Fig. 7** Counterpart of Fig. 6 for the quadrature surrogate  $\mathcal{M}_{PC}^{quad, P6}$  for  $N_{quad} = 4, \dots, 10$  (top curve in blue for  $N_{quad} = 4$  to bottom curve in black for  $N_{quad} = 10$ ) for  $P = 6$ .  $LH$  is computed every 4 km along the channel

of the surrogate is then calculated at this particular point. Following this theory,  $LOO$  is defined as follows:

$$LOO = \frac{\sum_{k=1}^{N_{lar}} \left( h^{(k)} - h_{PC}^{(-k)} \right)^2}{\sum_{k=1}^{N_{lar}} \left( h^{(k)} - \bar{h} \right)^2}, \quad (37)$$

where  $\bar{h}$  denotes the sample-average water level and  $h_{PC}^{(-k)}$  stands for the evaluation of the PC surrogate water level at  $\mathbf{x}^{(k)}$  when the surrogate has been built using an experimental design in which  $\mathbf{x}^{(k)}$  was excluded.

When applying the LAR strategy,  $N_{lar}$  surrogates should be formulated. In practice, [7] propose a formulation of the  $LOO$  metric based on a single formulation of the LAR model using the whole training sample  $N_{lar}$  at once:

$$LOO \equiv \sum_{k=1}^{N_{lar}} \left( \frac{h^{(k)} - h_{PC}^{(k)}}{1 - D_k} \right)^2 \quad (38)$$

where  $D_k$  is the  $k$ th diagonal term of the matrix  $\Psi(\Psi^T \Psi)^{-1} \Psi^T$ .

$LOO$  is formulated here as a normalized error, while  $LH$  (Eq. 36) is an absolute error.  $LH$  is used to assess the results of PC-expansion for varying total maximum polynomial degree and varying number of quadrature or regression points. In further results, when compared to  $LOO$  (Fig. 9),  $LH$  is normalized by the water level variance (at each curvilinear abscissa) over  $E_{MC}$ .

**Coefficient Metrics (LC)** The sampling error produced by the PC expansion can also be assessed through the comparison of the PC-coefficients. For this purpose, for a given total polynomial degree, a reference surrogate model is built using a least-square regression method that involves a sufficient number of forward model simulations ( $N_{ls,ref} = 10^5$ ) to guarantee the convergence of Eq. 21. The vector of coefficients for this reference model is further denoted by  $\boldsymbol{\gamma}^{ref}$ . In the following, the coefficients  $\{\gamma_i^{ref}\}_{i \in \mathcal{I}^P}$  are compared to those of surrogate models  $\{\gamma_i\}_{i \in \mathcal{I}^P}$  based on least-square or quadrature methods involving a smaller number of members ( $N_{ls}$  or  $N_{quad}^n$ ) than  $N_{ls,ref}$ . The  $LC$  metric is defined in the sense of the  $\ell_2$  norm as follows:

$$LC = \sum_{i \in \mathcal{I}^P} LC_i = \sum_{i=0}^{N_{PC}-1} LC_i = \sum_{i=0}^{N_{PC}-1} \left( \gamma_i^{ref} - \gamma_i \right)^2. \quad (39)$$

### 3.8 Numerical Implementation

Due to the increasing interest in uncertainty quantification over the last decade, a significant number of dedicated

tools/libraries are now available for the scientific community. Such toolkits provide an interface between simulation codes and systems analysis methods to address engineering questions in sensitivity analysis, uncertainty quantification, reliability analysis, parameter estimation, surrogate model design, and optimization. Among the most commonly used platforms are the Dakota Software toolkit ([www.dakota.sandia.gov](http://www.dakota.sandia.gov)), UQTK ([www.sandia.gov/UQToolkit](http://www.sandia.gov/UQToolkit)), the UQLab ([www.uqlab.com](http://www.uqlab.com)) and OpenTURNS ([www.openturns.org](http://www.openturns.org)); the latter was used in the present study. OpenTURNS is an open-source (LGPL) scientific library usable as a Python module dedicated to the treatment of uncertainties in an industrial context. EDF, Airbus, and Phimeca have been working together since 2005 to build OpenTURNS designed to perform uncertainty treatment and reliability analyses. OpenTURNS offers a wide catalog of features for uncertainty quantification (among them PC strategies [19] such as least-square, Gaussian quadrature and hybrid-LAR) and benefits from a well-organized developers and users community (forum, training, user guides). It can either be used as a python module when the numerical model is implemented in python, or used as a component with a coupling platform. In the present work, the quadrature, least-square, and hybrid-LAR strategies implemented in OpenTURNS for PC expansion were used. The MASCARET and OpenTURNS are both integrated components of the SALOME platform developed at EDF ([www.salome-platform.org](http://www.salome-platform.org)). This integrated framework allows for an efficient use of MASCARET as a python function driven by the flowchart of the PC algorithms implemented in OpenTURNS. For further information on the OpenTURNS library, the reader is referred to the online reference and use case guides.

## 4 Results

This section aims at studying the convergence of the PC-expansion with respect to the choice of the truncation strategy, the projection strategy and the total polynomial degree.

### 4.1 Idealized Channel

The PC-expansion approach presented in Section 3 is applied for the idealized channel test case with respect to the friction coefficient  $K_s$  and the discharge  $Q$  in steady flow ( $\mathbf{x} = [Q, K_s]^T$  with  $n = 2$ ). For uniform steady flow, the water level at the downstream boundary condition is set so that the flow is governed by Manning equation (Eq. 5). This strategy is then applied for gradually-varied steady flow on the channel test case with a deterministic description of the downstream condition so that backwater curve appears.

#### 4.1.1 Uniform Steady Flow

**Truncation Error** In the first place, the truncation error is investigated in order to identify the appropriate total polynomial degree  $P$  and thus the total number of coefficients  $N_{PC}$  to estimate in the PC-expansion. In this perspective, a least-square surrogate is built here with a large enough number of forward model simulations ( $N_{ls} = 10^5$ ) to distinguish the truncation error from the coefficient estimation error. The least-square surrogate  $\mathcal{M}_{PC}^{ls5,P}$  is formulated over the training sample  $N_{ls,ref}$  for varying  $P$ , then used to predict the water level  $h_{PC}$  over the validation sample  $E_{MC}$  of size  $N_{MC}$ . These water levels are validated against those obtained with a MC forward model sampling to compute  $LH$  as described in Section 3.7.

Figure 3 displays  $LH$  (Eq. 36) in logarithmic scale when  $P$  increases from 4 to 10; for instance, for  $P = 6$ ,  $LH$  is approximately equal to  $10^{-6}$ . The error  $\log(LH)$  linearly decreases by one order of magnitude as the total polynomial degree is incremented by 1 up to  $P = 10$ . The PC expansion is built here on a hybrid polynomial basis made of Hermite polynomials for  $Q$  and Legendre polynomials for  $K_s$ . The present convergence attests that the hybrid polynomial basis is adapted to properly describe the statistics of both input random variables and output variables.

The spectrum of coefficients  $\gamma_{i_1,i_2}$  is then analyzed to check which terms in the PC expansion are contributing the most to the water level  $h$ . For this purpose, Fig. 4 shows the spectrum of coefficients associated with the surrogate  $\mathcal{M}_{PC}^{ls5,P10}$  ( $P = 10$  consistently with Fig. 3). The color-bar represents the amplitude of  $\gamma_{i_1,i_2}$  in logarithmic scale; the first coefficient  $\gamma_{00}$  (the mean water level) features the highest magnitude.

The magnitude of the monivariate coefficients ( $\gamma_{i_1,0}$  associated with  $Q$  on the  $x$ -axis and  $\gamma_{0,i_2}$  associated with  $K_s$  on the  $y$ -axis) is compared. The  $K_s$ -monomials are found to be larger than the  $Q$ -monomials: given the PDFs for both random variables, the sensitivity to  $K_s$  is therefore larger than that to  $Q$  for uniform steady flow. This is confirmed by Sobol' indices:  $S_Q = 0.06$  and  $S_{K_s} = 0.94$  for variation coefficients equal to  $CV_Q = 10\%$  and  $CV_{K_s} = 35\%$ . These results depend on the properties of the input random variables and especially on the choice of their distributions; for instance, if  $CV_Q$  were larger, then the  $Q$ -monomials would be larger than the present ones.

The bivariate coefficients  $\gamma_{i_1,i_2}$  correspond to the elements in the lower triangular matrix along  $x$ - and  $y$ -axes in Fig. 4. For instance, the element at  $x = 4$  and  $y = 2$  represents the coefficient related to the polynomial function  $Q^4 K_s^2$  of amplitude  $10^{-6}$ . The amplitude of the monomials is larger than that of the bivariate coefficients, meaning that the univariate coefficients are more significant than the bivariate ones.

The value of the coefficients is also found to decrease as the polynomial degree increases. For instance, when the polynomial degree is beyond  $P = 6$  (oblique solid black line in Fig. 4), all coefficients are smaller than  $10^{-3}$ . Given the  $LH$  metric for the water level estimation in Fig. 3 and the coefficient spectrum in Fig. 4, it seems sufficient to truncate the PC expansion to  $P = 6$ , thus implying the computation of 28 coefficients for  $n = 2$  according to Eq. 11 (when  $P = 10$ , 66 coefficients should be computed).

**Sampling Error** We investigate the size of the training sample that is necessary to properly estimate the coefficients for  $P = 10$  for both least-square and quadrature strategies; the objective is now to focus on the projection error. In order to limit the computational cost induced by forward model simulations, the size of the training sample for the least-square surrogate is reduced (previously set to  $N_{ls,ref} = 10^5$ ); the surrogate models  $\mathcal{M}_{PC}^{ls2,P10}$  and  $\mathcal{M}_{PC}^{ls3,P10}$  are formulated with respectively  $N_{ls} = 10^2$  and  $N_{ls} = 10^3$ . The quadrature surrogate  $\mathcal{M}_{PC}^{quad11,P10}$  is formulated with the number of quadrature roots set to  $N_{quad} = 11$  for each input random variable consistently with  $P = 10$  (according to Eq. 24).

Figure 5 displays the elements  $LC_i$  of the error  $LC$  (Section 3.7) in logarithmic scale for each coefficient within the expansion spectrum (66 coefficients) and compares  $\mathcal{M}_{PC}^{ls2,P10}$  (thin dashed line),  $\mathcal{M}_{PC}^{ls3,P10}$  (thick black line), and  $\mathcal{M}_{PC}^{quad11,P10}$  (thick solid line) with respect to the reference coefficients computed with  $\mathcal{M}_{PC}^{ls5,P10}$ .  $LC_i$  decreases from  $10^{-6}$  to  $10^{-10}$  as the size of the training sample for the least-square surrogate increases from  $N_{ls} = 10^2$  to  $10^3$ .  $\mathcal{M}_{PC}^{quad11,P10}$  leads to a  $LC_i$  error value that is smaller to that of  $\mathcal{M}_{PC}^{ls3,P10}$  for a significantly reduced computational cost (only 121 forward mode evaluations). The projection strategy is thus able to provide a reliable surrogate model at lower computational cost than the least-square strategy for Eq. 3. In spite of these results that are in favor of the quadrature-based strategy, they are limitations to this approach, for instance when the input space dimension increases and when not using nested quadrature.

#### 4.1.2 Gradually-Variied Steady Flow

A similar analysis is carried out for gradually-varied steady flow to check the impact of more complex flow features (backwater curves) on the behavior and accuracy of the PC-expansion. When the downstream boundary condition is prescribed to  $h_{out} = 10$  m, as  $Q$  and  $K_s$  are sampled within their respective PDF given in Section 2.2.1, the Manning equilibrium is no longer preserved over the entire channel. The upstream flow is still under steady flow conditions, while the flow is governed by the backwater equations

(Eq. 3) in the downstream part of the river. It should be noted that within the sampling of the physical random variables, the flow remains subcritical and the water level describes a backwater curve of type  $M_1$  or  $M_2$ . The surrogate is here formulated following the same strategy as for uniform steady flow: first, setting the total polynomial degree  $P$ ; then, setting the number of forward model evaluations for computing the coefficients. Since the flow is now spatially distributed, we compute a surrogate at each simulation grid point along the curvilinear abscissa using the same set of forward model evaluations, thus providing spatially varying coefficients (Section 3.1).

**Truncation Error** Figure 6 shows the spatially-varying water level error  $LH$  (Eq. 36) in logarithmic scale for the least-square surrogate  $\mathcal{M}_{PC}^{ls5,P}$  (built with a training sample  $N_{ls,ref} = 10^5$ ) for varying total polynomial degree  $P$  ( $4 \leq P \leq 10$ ).  $LH$  is computed over the validation sample  $E_{MC}$  with  $N_{MC}=10^5$ . As highlighted in Fig. 3 for uniform flow,  $LH$  is small and in the present case remains below  $10^{-6}$  when  $P \geq 6$ .  $LH$  decreases by one order of magnitude when  $P$  increases by 1 for  $4 \leq P \leq 7$  over the entire computational domain. Beyond this threshold value, increasing  $P$  results in a smaller decrease of  $LH$ , especially where the flow passes from uniform flow to backwater curve. Refining the mesh resolution by a factor of two (from 100 to 50-m discretization) was found (not shown here) to further reduce  $LH$  when  $P$  increases, even where the flow dynamics gets more complex. When refining the mesh, numerical errors in the simulation decrease and are better resolved with the PC-expansion.

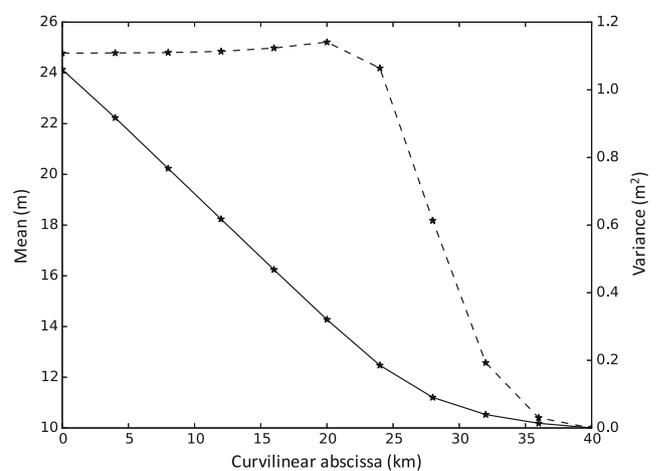
**Sampling Error** In the following,  $P = 6$  is set ( $LH = 10^{-6}$  in coherence with Fig. 3 for uniform flow, here in the upstream part of the channel). We now investigate the size of the training sample that is necessary to properly estimate the coefficients with the quadrature method. The performance of the surrogate is investigated when the number of quadrature roots  $N_{quad}$  varies from 4 to 10. Figure 7 displays the water level error  $LH$  (Eq. 36) in logarithmic scale along the curvilinear abscissa computed for the surrogate  $\mathcal{M}_{PC}^{quad,P6}$  (top curve in blue for  $N_{quad} = 4$  to bottom in black for  $N_{quad} = 10$ ). Again,  $LH$  is computed over the validation sample  $E_{MC}$  (with  $N_{MC} = 10^5$ ). As expected from theory (Eq. 24),  $LH$  decreases as  $N_{quad}$  increases up to  $N_{quad} = 7$  beyond which the error remains constant to  $10^{-6}$ . For  $P \geq 7$ , the error sums up to the truncation error also represented in Fig. 6 for  $\mathcal{M}_{PC}^{ls5,P6}$  (red curve). In both Figs. 6 and 7, the  $LH$  error is smaller at the downstream boundary condition specified in a deterministic way. The transition between the uniform regime and the backwater curve ( $M_1$  or  $M_2$ ) occurs between 25 and 30 km; the dynamics of this particular point is complex to represent with the surrogate,

thus the error  $LH$  tends to have larger values there than at other locations over the reach, yet it remains extremely small ( $10^{-8}$ ).

**Statistical Moments** The next step consists in validating the statistical moments as defined in Section 3.6 for the retained quadrature surrogate  $\mathcal{M}_{PC}^{quad7,P6}$ ; the validation is carried out with respect to the classical MC random sampling over the validation sample  $N_{MC} = 10^5$  (all backwater curves within the sample describe  $M_1$  backwater curves). The stochastically estimated mean and variance (lines) are shown in Fig. 8 along the 40-km channel and compared to the mean and variance computed with the PC-coefficients for  $\mathcal{M}_{PC}^{quad7,P6}$  (symbols). They feature very good adequacy. The surrogate  $\mathcal{M}_{PC}^{quad7,P6}$  thus offers a satisfying approximation of direct model water level over the whole computational domain, in terms of first and second order moments, for the present idealized channel case study.

## 4.2 Application to the Garonne River

The PC-expansion approach was validated for the idealized channel test case and is now applied to a real case of the Garonne River presented in Section 2.2.2, still in steady flow. The flow dynamics is governed by the SWE (Eq. 1) solved by MASCARET; the water level is non uniform as the river width, bathymetry and friction coefficients are spatially distributed over the Tonneins-La Réole reach. While the PC-expansion is formulated for the entire domain, a particular attention is paid to the water level at Marmande, where in-situ observations are available and where data assimilation is applied in flood forecasting operational context.



**Fig. 8** Water level mean (m) in solid line (left y-axis) and variance ( $m^2$ ) in dashed line (right y-axis) along the curvilinear abscissa (km)—idealized channel case with gradually-varied steady flow. Comparison between the MC stochastic estimates (lines) and the quadrature surrogate  $\mathcal{M}_{PC}^{quad7,P6}$  ( $P = 6$  and  $N_{quad} = 7$ ) estimates (symbols)

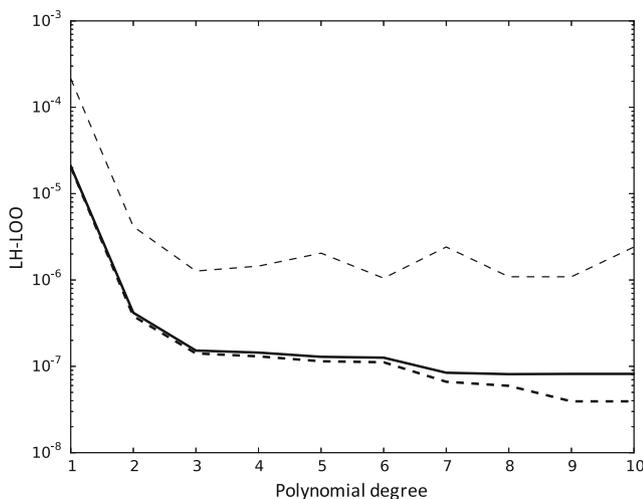
**Sensitivity Analysis Based on LAR Surrogate for Model Reduction** The cost of the PC-expansion is here potentially higher than for the idealized channel case study as there are four main sources of uncertainty ( $Q, K_{s1}, K_{s2}, K_{s3}$ ) in the Garonne hydraulic model. A sensitivity analysis is carried out to investigate whether the dimension of the uncertain input space could be reduced so that the formulation of the PC-surrogate would be reduced. The PC-expansion will then be formulated with respect to the most significant variables only. The sources of uncertainty are classified using Sobol’ indices (Section 3.6). The Sobol’ indices are computed by means of a LAR surrogate as defined in Section 3.4. The accuracy of the LAR surrogate is tested by comparing  $\mathcal{M}_{PC}^{lar2,P}$  and  $\mathcal{M}_{PC}^{lar3,P}$  formulated over a training sample  $E_{lar}$  of size  $N_{lar} = 10^2$  and  $N_{lar} = 10^3$ , respectively.

Figure 9 displays the water level error at Marmande for  $\mathcal{M}_{PC}^{lar3,P}$  measured with both the  $LH$  metric (Eq. 36) in solid line and the  $LOO$  metric (Eq. 38) in thick dashed line for the total polynomial degree  $P$  varying between 1 and 10. The  $LOO$  error for  $\mathcal{M}_{PC}^{lar2,P}$  is also plotted in thin dashed line to check the sensitivity of the LAR result to the size of the training sample  $E_{lar}$ .  $LH$  is computed over a validation sample  $E_{MC}$  of size  $N_{MC} = 10^5$  that samples the four random variables. For this comparison,  $LH$  is normalized by the water level variance and multiplied by  $N_{lar}$  to be consistent with the  $LOO$  formulation. Figure 9 shows that for  $P = 4$ , the  $LOO$  converges to  $10^{-6}$  for  $N_{lar} = 100$  and to  $10^{-7}$  for  $N_{lar} = 1000$ , similarly to  $LH$  for  $N_{MC} = 10^5$ . This highlights the merits of the LAR method to reduce the

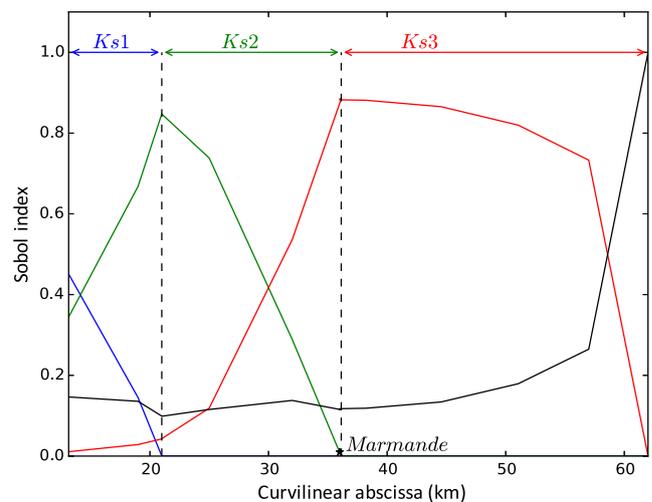
computational cost, while providing an accurate water level representation.

The LAR method is thus used for sensitivity analysis with respect to  $(Q, K_{s1}, K_{s2}, K_{s3})$  to compute the Sobol’ indices shown in Fig. 10. The impact of  $Q$  is homogeneous over the 60 km Garonne River reach. As the downstream boundary condition is prescribed by the local rating curve at Tonneins, the uncertainty in water level is totally explained by the uncertainty in discharge, thus  $S_Q$  is equal to 1 at the exit of the hydraulic network. The water level is also sensitive to the local  $K_s$  of the section as well as to the  $K_s$  of the section located downstream. In the upstream section (0–20 km), the water level is mostly sensitive to  $K_{s1}$  and  $K_{s2}$ , while the impact of  $K_{s3}$  is smaller. In the second section (20–36 km), the water level is mostly sensitive to  $K_{s2}$  and  $K_{s3}$ . In the third section (36–60 km) that includes Marmande, the water level is only sensitive to  $K_{s3}$  and  $Q$ . At Marmande, the sensitivity to  $K_{s1}$  and  $K_{s2}$  is zero, the Sobol’ index for  $K_{s3}$  is equal to 0.88 and that for  $Q$  is equal to 0.12. These values are strongly related to the prescribed PDFs on  $K_{s3}$  and  $Q$ ; in the present case, the variation coefficient for  $Q$  is  $CV_Q = 10\%$ , while the variation coefficient for  $K_{s3}$  is  $CV_{K_{s3}} = 40\%$ .

The sensitivity analysis highlights that  $K_{s1}$  and  $K_{s2}$  have barely no impact on the water level at Marmande, where observations are available. The improvement of the water level at Marmande and downstream of Marmande relies on the improvement of  $Q$  and  $K_{s3}$ . Consequently, the uncertain space dimensionality is reduced to  $\mathbf{x} = [Q, K_{s3}]^T$ . The PC-surrogate that is constructed with respect to  $\mathbf{x} = [Q, K_{s3}]^T$



**Fig. 9** Comparison of error metrics at Marmande observing station (36 km on the Garonne reach) for varying total degree  $P$  ( $P = 1, \dots, 10$ )—Garonne River case in steady flow:  $LOO$  in dashed lines ( $\mathcal{M}_{PC}^{lar2,P}$  in thin dashed line;  $\mathcal{M}_{PC}^{lar3,P}$  in thick dashed line); normalized  $LH$  in solid line computed over a validation sample  $E_{MC}$  of size  $N_{MC} = 10^5$

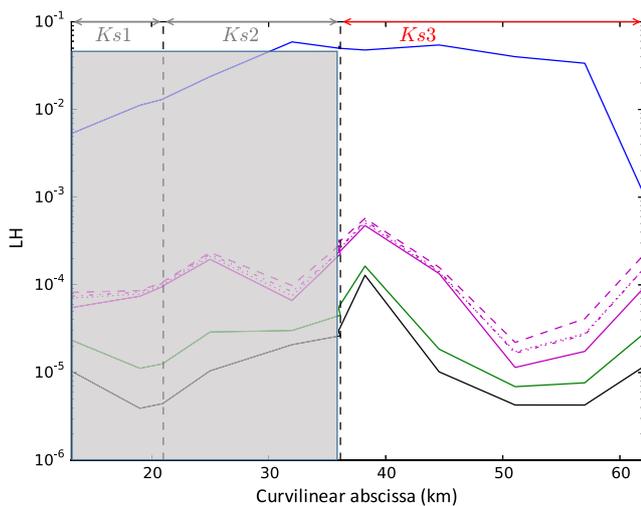


**Fig. 10** Garonne River case in steady flow—first-order Sobol’ indices computed with the LAR surrogate  $\mathcal{M}_{PC}^{lar2,P4}$  formulated over a training sample  $E_{lar}$  of size  $N_{lar} = 10^2$  for a total polynomial degree  $P = 4$  with respect to  $K_{s1}$  (blue line),  $K_{s2}$  (green line),  $K_{s3}$  (red line) and  $Q$  (black line). Vertical dashed lines indicate the limits between the different sections of  $K_s$ . The curvilinear abscissa at Marmande is  $s = 36$  km

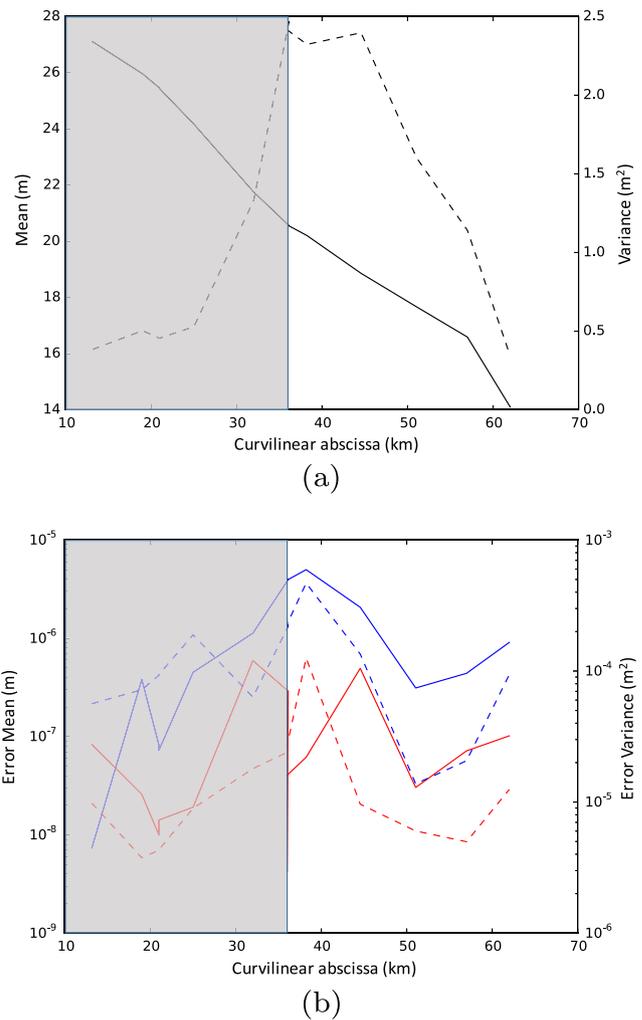
is legitimate from Marmande to La Réole. Still, the PC-surrogate is built every 4 km along the entire computational domain for informative purpose and because it is needed to formulate the error covariance matrix in a data assimilation framework. The illustrations are presented with a shaded area upstream Marmande.

**Uncertainty Quantification with Least-Square and Quadrature Strategies** Figure 11 displays the water level error  $LH$  (Eq. 36) in logarithmic scale along the 60-km hydraulic network for the least-square surrogates  $\mathcal{M}_{PC}^{ls5,P}$  built with a training sample  $E_{ls}$  of size  $N_{ls,ref} = 10^5$ . Different total polynomial degrees are considered ( $P = 1, 6, 11, 15$ ). The error is found to decrease as  $P$  increases and is of the order of  $10^{-4}$  for  $P = 6$ . This error can be compared to that of the Manning equation displayed in Fig. 3 ( $LH = 10^{-6}$ ). Figure 11 also displays  $LH$  for a surrogate computed with the quadrature method  $\mathcal{M}_{PC}^{quad,P6}$  setting  $P = 6$  and varying  $N_{quad} = 6, 7, 8$ . The error  $LH$  associated to  $\mathcal{M}_{PC}^{quad7,P6}$  is similar to that of  $\mathcal{M}_{PC}^{ls5,P6}$ , while it only involves 49 MASCARET evaluations instead of  $10^5$ . A PC-expansion  $\mathcal{M}_{PC}^{quad7,P6}$  with  $P = 6$  and  $N_{quad} = 7$  provides a satisfying description of the water level over the Garonne River reach with respect to  $Q$  and  $K_{s3}$  in the sense of the  $LH$  metric.

Figures 12 and 13 present the statistical moments along the 60-km reach, which are defined in Section 3.6 and computed here for varying total polynomial degree  $P$ , i.e., for the quadrature surrogates  $\mathcal{M}_{PC}^{quad16,P15}$  and  $\mathcal{M}_{PC}^{quad7,P6}$ .

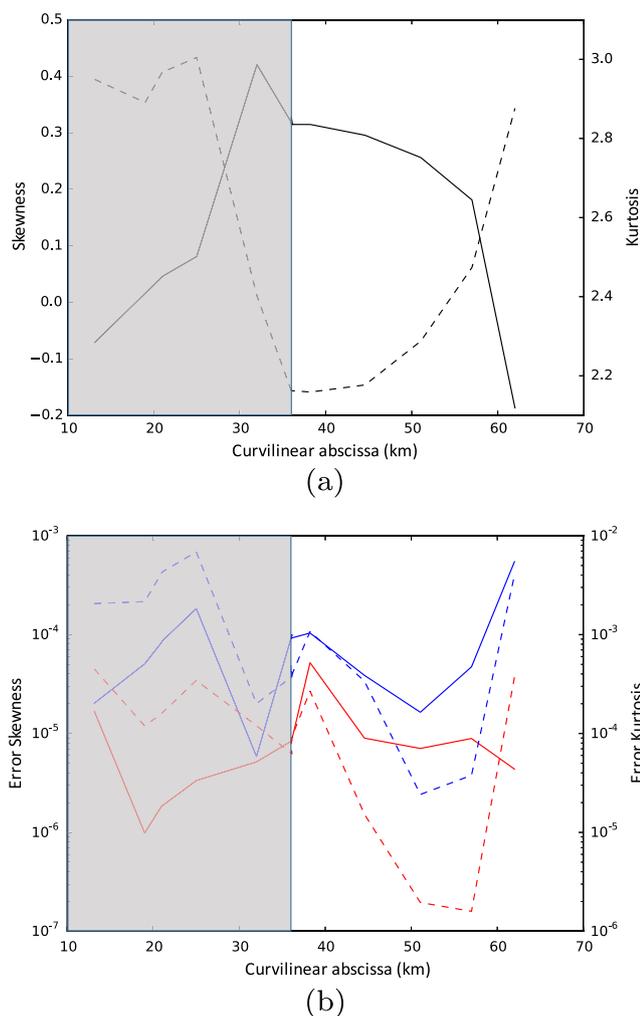


**Fig. 11** Spatially varying truncation and quadrature errors measured by the water level error  $LH$  (Eq. 36) in logarithmic scale—Garonne River case in steady flow.  $LH$  is computed over the validation sample  $E_{MC}$  with  $N_{MC} = 10^5$  for the following surrogates: least-square surrogates  $\mathcal{M}_{PC}^{ls,P}$   $P = 1, 6, 11, 15$  (solid blue, purple, green and black line); and quadrature surrogates  $\mathcal{M}_{PC}^{quad,P6}$  with  $P = 6$  and  $N_{quad} = 6, 7, 8$  (purple dashed, dashed-dotted, and dotted lines)



**Fig. 12** Garonne River case in steady flow. **a** Comparison of mean (left y-axis, solid line) and variance (right y-axis, dashed line) statistical moments between quadrature surrogates  $\mathcal{M}_{PC}^{quad16,P15}$  and  $\mathcal{M}_{PC}^{quad7,P6}$  (results overlap). **b** Error with respect to MC stochastic estimation over  $N_{MC} = 10^5$  members in the mean (left y-axis) and variance (right y-axis) estimates for quadrature surrogates  $\mathcal{M}_{PC}^{quad16,P15}$  (red) and  $\mathcal{M}_{PC}^{quad7,P6}$  (blue)

The mean (solid line, left y-axis, in (m)) and the variance (dashed line, right y-axis, in (m<sup>2</sup>)) are shown in Fig. 12a; the skewness (solid line, left y-axis, dimensionless) and the kurtosis (dashed line, right y-axis, dimensionless) are shown in Fig. 13a (the curves associated with both surrogates  $\mathcal{M}_{PC}^{quad16,P15}$  and  $\mathcal{M}_{PC}^{quad7,P6}$  overlap). In complement, Figs. 12b and 13b present the errors in the statistical moments computed with respect to their stochastic estimation obtained from a MC random sampling  $E_{MC}$ . Results show that the estimation of the four statistical moments is satisfying with  $P = 6$ . The mean is estimated with an error of about  $10^{-6}$  for  $P = 6$  (blue solid line) that decreases by one order of magnitude for  $P = 15$  (red solid line). The

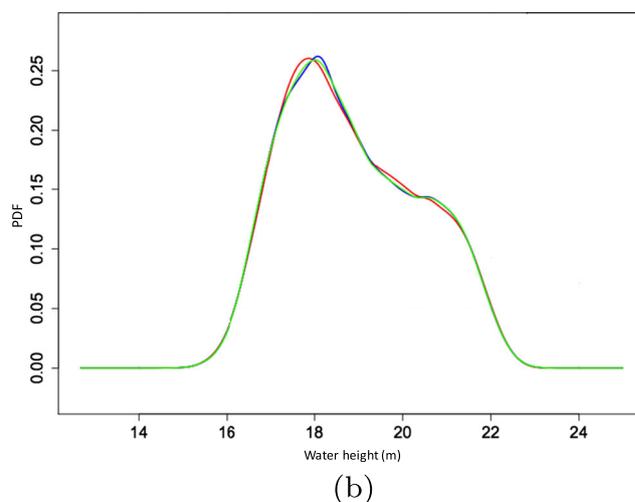
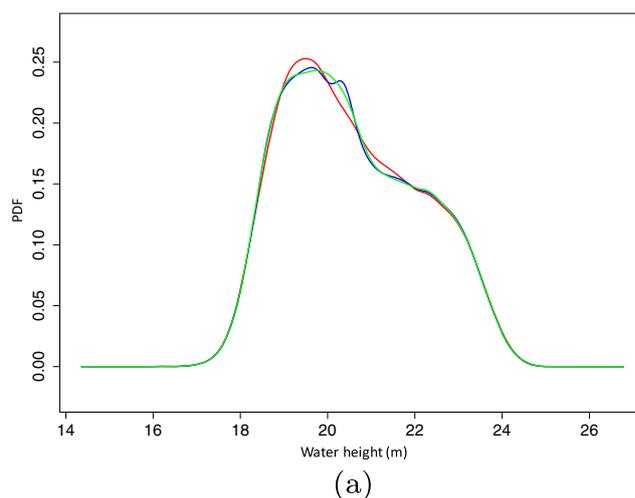


**Fig. 13** Garonne River case in steady flow—same caption as Fig. 12 but for the skewness and kurtosis statistical moments

variance is estimated with an error of  $10^{-4}$  for  $P = 6$  (blue dashed line) that decreases by three orders of magnitude for  $P = 15$  (red dashed line). The skewness and the kurtosis are respectively estimated with an error of  $10^{-4}$  and  $10^{-3}$  for  $P = 6$  (solid blue and red lines) that decreases by one order of magnitude for  $P = 15$  (dashed blue and red lines). The mean water level varies from 27 m upstream to 14 m downstream. The water level variance is about  $0.5 \text{ m}^2$  in the first part of the reach and increases up to  $2.5 \text{ m}^2$  in the area of Marmande, where the bathymetry strongly varies and the flow dynamics gets more complex with local changes of backwater curves. The skewness varies from  $-0.1$  to  $0.4$  with a steep change in the Marmande area, suggesting that the water level PDF at Marmande is non-symmetric (non-Gaussian) with a slight shift on the left for smaller water levels than the mean upstream as well as a significant shift on the right for larger water levels than the mean over most of the Garonne reach, especially near Marmande. In this

area, the water level PDF has a positive skew; large or extreme water levels values are more likely to occur. On the contrary, the kurtosis has negative values upstream and negative values near Marmande, meaning that the PDF tends to be sharper than a Gaussian PDF upstream, and smoother than a Gaussian PDF downstream. The observed skewness and kurtosis are probably due to the presence of multiple peaks in the PDF for this area of the Garonne reach. This is confirmed in Fig. 14.

Figure 14a compares the PDF at Marmande obtained with a MC approach over  $E_{MC}$  (blue line) as well as quadrature surrogates  $\mathcal{M}_{PC}^{quad16,P15}$  (green line) and  $\mathcal{M}_{PC}^{quad7,P6}$  (red line) also sampled over  $E_{MC}$ . The PDFs are obtained with a kernel smoothing method using Gaussian kernels, available in OpenTURNS. Figure 14b presents the

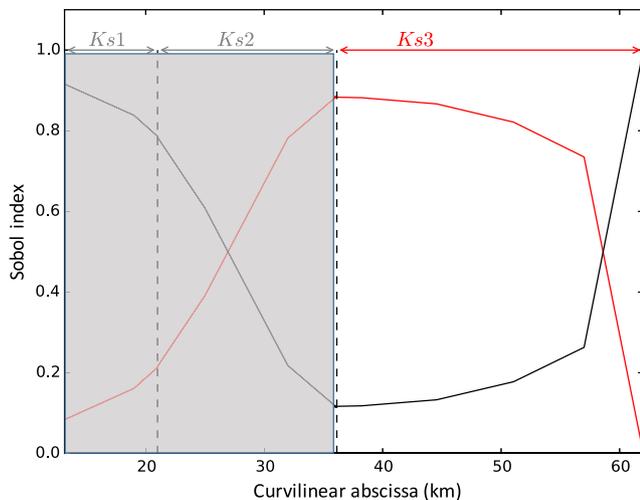


**Fig. 14** Garonne River case in steady flow—water level PDF estimated with MC sampling  $E_{MC}$  over  $N_{MC} = 10^5$  members, either using the forward model (blue line) or the quadrature surrogates  $\mathcal{M}_{PC}^{quad7,P6}$  (red line) and  $\mathcal{M}_{PC}^{quad16,P15}$  (green line) at  $s = 36$  km at Marmande (a) and  $s = 40$  km (b)

counterpart of Fig. 14a at  $s = 40$  km. At Marmande, the flow dynamics is mostly driven by  $K_{s_3}$  ( $Q$  has a smaller impact as shown in Fig. 15). The flow dynamics is complex since the bathymetry and the river width strongly vary locally, the nonlinearities are strong and the estimation of the PDF requires a high total polynomial degree  $P$  for the PC-expansion as shown in Fig. 14a. When the bathymetry is smooth, the relation between water level and the input parameters is fairly linear so that the water level PDF is smooth. In this case, a PC-expansion of total polynomial degree  $P = 6$  provides a satisfying description of the PDF (the PDF estimated with  $\mathcal{M}_{PC}^{quad7,P6}$  is close to that computed with a MC approach—similar results are obtained up to Marmande).

These results demonstrate that the PC-expansion strategy can be used to formulate a surrogate for the SWE solved by the forward model over the Garonne River reach between Tonneins and La Réole. The surrogate is built with respect to the most significant sources of uncertainty using the quadrature method to compute the coefficients. The required total polynomial degree is higher when the flow becomes more complex.

Figure 15 displays Sobol' indices for  $Q$  ( $S_Q$ , black curve) and  $K_{s_3}$  ( $S_{K_{s_3}}$ , red curve). Upstream of Marmande, the flow is dominated by the discharge, while it is dominated by the friction downstream of Marmande. At the downstream boundary condition where the rating curve is imposed, the water level variability is totally explained by the discharge. The water level variance shown in Fig. 12 (dashed line) is about  $0.5 \text{ m}^2$  in the upstream part of the river, it increases to  $2.5 \text{ m}^2$  at Marmande and then decreases to  $0.5 \text{ m}^2$  at the



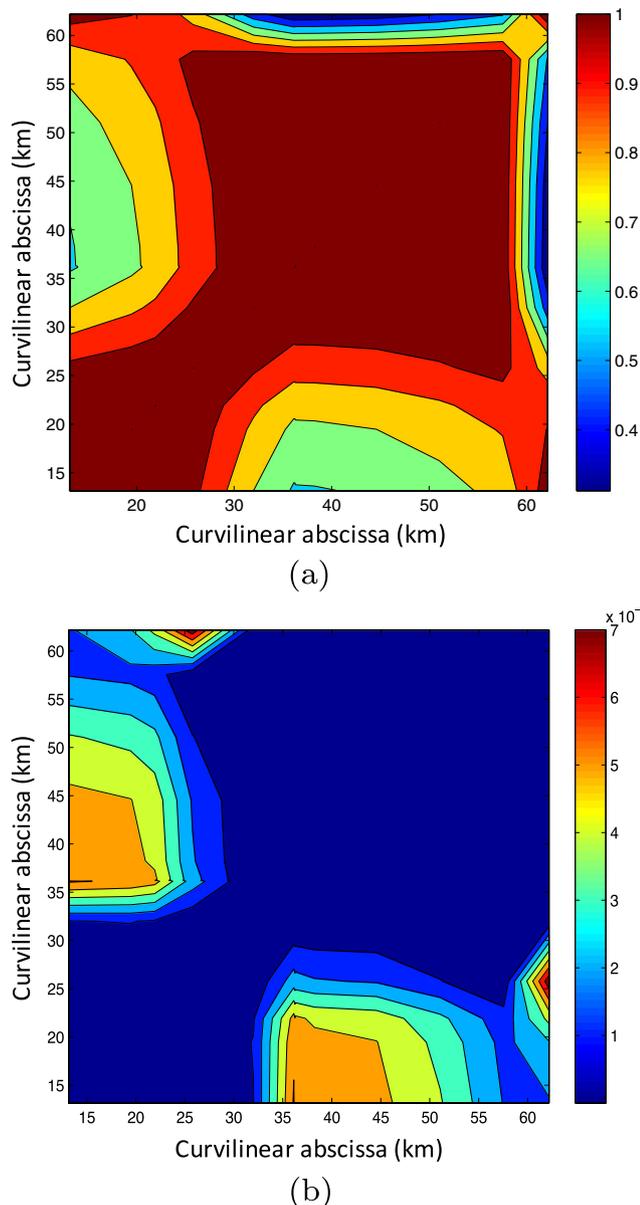
**Fig. 15** Garonne River case in steady flow—first-order Sobol' indices computed with the quadrature surrogate  $\mathcal{M}_{PC}^{quad7,P6}$  parameterized with respect to  $Q$  (black solid line) and  $K_{s_3}$  (red solid line) over the Garonne reach. The vertical dashed lines represent the 3 limits of the friction coefficient zones (Marmande is located at  $s = 36$  km)

downstream boundary condition, which is non negligible when the water level it self is about 10 m. This means that the discharge and the friction coefficient are truly significant sources of uncertainties and their errors translate into a significant water level error.

**Covariance Estimation for Data Assimilation** The estimation of covariance matrices is a major research topic in the field of data assimilation research. The stochastic estimation of these statistics for water level is a limiting point for the efficient use of data assimilation in a real time context. The formulation of the Kalman gain matrix relies on the estimation of the water level error covariance matrix that describes the spatial correlation between the errors along the river. The use of the PC-surrogate in place of the forward model allows for a significant cost reduction for stochastic estimate of these statistics, which are here formulated using the PC coefficients directly (Eqs. 33–30) and thus avoid re-sampling errors. Figure 16a displays the water level correlation matrix along the 60-km Garonne River reach that is estimated using the retained quadrature surrogate  $\mathcal{M}_{PC}^{quad7,P6}$  and parameterized with respect to  $Q$  and  $K_{s_3}$  only. These correlations are estimated with an error on the order of  $10^{-7}$  with respect to those estimated with a classical MC approach based on  $N_{MC} = 10^5$  evaluations of the forward model as displayed in Fig. 16b. Each column of the matrix describes the water level error correlation functions for a given curvilinear abscissa. We recall here that the surrogate model is formulated with respect to the predominant uncertain variables for water level at Marmande and downstream. The covariance matrix analysis is thus fully legitimate in this area and only partially informative upstream of Marmande. By definition, the correlation is equal to 1 at the location points where the correlations are expressed and decreases when moving away to the other grid points along the channel.

We first analyze the correlation function for the upstream location (for instance  $s = 15$  km) where the error in water level is dominated by the error in  $Q$  as shown in Fig. 15 with  $S_Q = 0.9$  and  $S_{K_{s_3}} = 0.1$ . The surrogate is formulated only with respect to  $Q$  and  $K_{s_3}$ . As a consequence, the downstream boundary condition prescribed with a rating curve that relates  $Q$  to  $h$  has a strong influence on the correlation function so that the correlation increases with the curvilinear abscissa close to the downstream boundary.

On the contrary, at Marmande ( $s = 36$  km) and its neighboring, the error in water level is dominated by the error in  $K_{s_3}$  with  $S_Q = 0.85$  and  $S_{K_{s_3}} = 0.15$ . As a consequence, the downstream boundary condition has a lesser impact on the water level error correlation function in the downstream part of the river (Fig. 15 shows that the Sobol' index  $S_{K_{s_3}}$  drops in the end of the reach to be zero at the downstream boundary condition). The correlation remains



**Fig. 16** Garonne River case in steady flow. **a** Water level correlation matrix over the 60-km Garonne reach estimated with the retained surrogate  $\mathcal{M}_{PC}^{quad7,P6}$  ( $P = 6, n = 2$  with  $\mathbf{x} = [Q, K_{s3}]^T$ ). The colorbar represents the correlation in the water level error for each pair of grid points along the reach. **b** Error on the water level correlation matrix estimated  $\mathcal{M}_{PC}^{quad7,P6}$  with respect to the correlation computed with a MC sampling  $E_{MC}$  over  $N_{MC} = 10^5$

close to 1 over a large section (approximately 20 km around Marmande) as the influence of  $K_{s3}$  is homogeneous over the section ( $K_{s3}$  remains around 0.8). Same analysis holds for  $s = 40$  km.

In the EnKF, the correlation matrix evolves over time to represent the spatially-distributed and temporally-dependent flow patterns. Estimating this correlation matrix through the use of a PC-surrogate (using only 49 MASCARET evaluations

of the water level) is thus a promising approach to reduce the cost of the EnKF, while representing the correlations with accuracy.

### 5 Conclusions and Perspectives

We investigated the use of a gPC-expansion strategy for hydraulic modeling of subcritical steady flows, both on an idealized channel test case and on a reach of the Garonne River. In this study, the surrogate is used for uncertainty propagation and sensitivity analysis in place of the MASCARET forward model solving the SWE. A major conclusion of this work is that the model output covariance matrix estimated with the reduced model build from a limited number of MASCARET evaluations is more reliable than that estimated with a classical MC approach with the same training budget.

The gPC-expansion was formulated with respect to two random variables, the friction coefficient of the river bed ( $K_s$ ) and the upstream discharge ( $Q$ ), on a hybrid basis combining Legendre and Hermite polynomials. The polynomial coefficients were computed with a quadrature method based on MASCARET evaluation at the roots of the basis polynomials. The validation of the surrogate is carried out using  $\ell_2$  error metrics defined in the space of the simulated water level as well as in the space of the polynomial coefficients. These metrics are computed either over an independent sample, thus requiring additional MASCARET evaluations than those necessary to the construction of the surrogate, or within the training set via the LOO error metric.

On the idealized rectangular channel with uniform slope (backwater curves), it was shown that the statistical moments for the water level are efficiently estimated with a reduced model built from a small number of MASCARET evaluations. On the Garonne River case, where the (steady) flow is governed by the SWE that represent a system of hyperbolic partial differential equations, the dimension of the stochastic space was reduced with a LAR method consisting on a sparse evaluation of the coefficients. The Sobol' indices indicate that, at Marmande, the water level statistics are driven by the upstream discharge and the friction coefficient of the section where Marmande is located ( $K_{s3}$ ). A surrogate model of total polynomial degree equal to 6 and parameterized with respect to  $Q$  and  $K_{s3}$  only was therefore formulated using the quadrature method based on 49 MASCARET evaluations. While a gPC-expansion of total polynomial degree 6 suffices to properly estimate the first statistical moments (especially the mean, the variance, and the covariance matrix that are mandatory in the context of data assimilation), the estimation of the PDF requires a higher polynomial degree in areas where the flow dynamics is complex.

The limitations of this work should be investigated in future work with respect to the applicability of the gPC-expansion, for instance when non linearities become stronger or when the flow becomes critical and presents discontinuities [5, 36]. Adapted strategies such as Kriging or radial basis functions should be proposed.

The main perspective for this work is to extend the gPC-strategy to unsteady flows and identify how often the PC-surrogate should be re-computed according to the time scales of the flow dynamics. This is a crucial step for the complementary use of surrogate model within data assimilation algorithms as the main objective is real-time flood forecasting. To begin with, the classification between the uncertainty sources (provided by sensitivity analysis for instance) can change in time and the surrogate model should eventually be formulated with respect to different variables over time. When the uncertain input variables include a time-varying variable (for instance the upstream discharge), its representation as a random variable may lead to high-dimensional problems and should thus be reduced for instance with a Karhunen-Loève approach that consists in separating the random process from the time (or space) dimension. When the observation contains significant information, the PDF established from the previous assimilation cycle is no longer representative of the current PDF due to the error temporal variability. The surrogate should be improved (in this case, the total polynomial degree of the gPC-expansion is increased). In the framework of EnKF, the errors are supposed to follow Gaussian PDFs, and only the first and second moments of the PDF should be updated along the assimilation cycles. The validity of the previously-established surrogate is then linked to the flow dynamics and the temporal evolution of the model errors. To some extent, it may be necessary to adapt the polynomial basis along the assimilation cycles.

An additional perspective lies in the assimilation of spatially distributed observations such as remote sensing data, for instance from the upcoming SWOT mission that will provide a high resolution and global coverage of water elevation adapted to the observation of rivers wider than 250 m. Water elevation maps will be available every 4 to 5 days at mid-latitude. The assimilation of these data in complement with in situ data seems very promising for model calibration and improvement of prediction capability. Since data will be spatially distributed and since the Kalman gain in the EnKF relies on the estimation of the parameter/model state error statistics, the surrogate should thus be formulated at different locations of the hydraulic network, where the nature of the uncertainties may differ. Hypothesis on the error distributions should be made eventually dissociating the random process issued from the spatial issue as previously mentioned regarding time dimension.

Future work also includes the investigation of advanced strategy when the dimension of the uncertain space increases.

In hydraulics, the bathymetry is a major source of uncertainty and the spatial field should be corrected. gPC-expansion methods are limited by the quadrature rule that requires too many forward model evaluations for its crude implementation. Several studies have investigated the use of gPC for high-dimensional problem; the merits of such solutions should be demonstrated in the field of open-channel flows.

**Acknowledgements** The financial support provided by CNES and EDF R&D is greatly appreciated. The authors acknowledge Michael Baudin, Anne Dufloy, Anne-Laure Popelin, Géraud Blatman and Bertrand Iooss from MRI (EDF R&D) for helpful discussions on uncertainty quantification and support on OpenTURNS. They also thank anonymous Referees for their helpful and constructive comments.

## References

- Bates, P., Horrit, M., & Fewtrell, T. (2010). A simple inertial formulation of the shallow water equations for efficient two-dimensional flood inundation modelling. *Journal of Hydrology*, 387, 33–45.
- Berveiller, M. (2005). *Éléments finis stochastiques : approches intrusive et non intrusive pour des analyses de fiabilité*. Ph.D. thesis. Clermont-Ferrand: Université Blaise Pascal.
- Berveiller, M., Sudret, B., & Lemaire, M. (2006). Stochastic finite element: a non intrusive approach by regression. *European Journal of Computational Mechanics/Revue Européenne de Mécanique Numérique*, 15(1-3), 81–92.
- Besnard, A., & Goutal, N. (2011). Comparaison de modèles 1D à casiers et 2D pour la modélisation hydraulique d'une plaine d'inondation—Cas de la Garonne entre Tonneins et La Réole. *La Houille Blanche*, 3, 42–47.
- Bijl, H., Lucor, D., Mishra, S., & Schwab, C. (Eds.) (2013). *Uncertainty quantification in Computational Fluid Dynamics, Lecture Notes in Computational Science and Engineering*, Vol. 92. Berlin: Springer.
- Blatman, G. (2009). *Adaptive sparse Polynomial Chaos expansions for uncertainty propagation and sensitivity analysis*. Ph.D. thesis. Clermont-Ferrand: Université Blaise Pascal.
- Blatman, G., & Sudret, B. (2011). Adaptive sparse polynomial chaos expansion based on Least Angle Regression. *Journal of Computational Physics*, 230(6), 2345–2367.
- Bozzi, S., Passoni, G., Bernardara, P., Goutal, N., & Arnaud, A. (2014). Roughness and discharge uncertainty in 1D water level calculations. *Environmental Modeling & Assessment*, 4, 343–353. <https://doi.org/10.1007/s10666-014-9430-6>.
- Camacho, R., Martin, J., Diaz-Ramirez, J., McAnally, W., Rodriguez, H., Suscy, P., & Zhang, S. (2014). Uncertainty analysis of estuarine hydrodynamic models: an evaluation of input data uncertainty in the weeks bay estuary, Alabama. *Applied Ocean Research*, 47, 138–153.
- Camacho, R., Martin, J., McAnally, W., Diaz-Ramirez, J., Rodriguez, H., Suscy, P., & Zhang, S. (2015). A comparison of Bayesian methods for uncertainty analysis in hydraulic and hydrodynamic modelling. *Journal of the American Water Resources Association*, 51(5), 1372–1393.
- Candès, E.J., Romberg, J.K., & Tao, T. (2004). Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transaction Information Theory*, 52-2, 489–509.
- Candès, E.J.J., & Tao, T. (2005). Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12), 4203–4215.

13. Choi, S.K., Grandhi, R.V., Canfield, R.A., & Pettit, C.L. (2004). Polynomial Chaos expansion with Latin Hypercube Sampling for estimating response variability. *AIAA journal*, 42(6), 1191–1198.
14. Chow, V.T. (1959). *Open-channel hydraulics*. New York: McGraw-Hill civil engineering series.
15. Deman, G., Konakli, K., Sudret, B., Kerrou, J., Perrochet, P., & Benabderrahmane, H. (2015). Using sparse polynomial chaos expansions for the global sensitivity analysis of groundwater lifetime expectancy in a multi-layered hydrogeological model. *Reliability Engineering and System Safety*, 147, 156–169.
16. Donoho, D.L. (2006). Compressed sensing. *IEEE Transactions on Information Theory*, 52(4), 1289–1306.
17. Doostan, A., & Owahdi, H. (2011). A non-adapted sparse approximation of PDEs with stochastic inputs. *Journal of Computational Physics*, 230(8), 3015–3034.
18. Dubreuil, S., Berveiller, M., Petitjean, F., & Salaün, M. (2014). Construction of bootstrap confidence intervals on sensitivity indices computed by polynomial chaos expansion. *Reliability Engineering and System Safety*, 121, 263–275. <https://doi.org/10.1016/j.res.2013.09.011>.
19. Dutka-Malen, I., Lebrun, R., Saassouh, B., & Sudret, B. (2009). Implementation of a polynomial chaos toolbox in openturns with test-case application. In *Conference: Proc. 10th Int. Conf. Struct. Safety and Reliability (ICOSSAR'2009), Osaka, Japan*.
20. Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, 32(2), 407–499.
21. Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99(C5), 10,143–10,162. <https://doi.org/10.1029/94JC00572>.
22. Goutal, N. (2014). *Développements autour de la simulation des écoulements à surface libre en rivière. Equations de Saint-Venant - Couplage de modèles - Incertitudes - Application aux ondes de submersion*. Université Paris Est: Habilitation à diriger des recherches (HDR).
23. Goutal, N., & Maurel, F. (2002). A finite volume solver for 1D shallow-water equations applied to an actual river. *International Journal for Numerical Methods in Fluids*, 38(1), 1–19.
24. Habert, J., Ricci, S., LePape, E., Thual, O., Piacentini, A., Goutal, N., Jonville, G., & Rochoux, M. (2016). Reduction of the uncertainties in the water level-discharge relation of a 1D hydraulic model in the context of operational flood forecasting. *Journal of Hydrology*, 532, 52–64. <https://doi.org/10.1016/j.jhydrol.2015.11.023>.
25. Horritt, M., & Bates, P. (2002). Evaluation of 1d and 2d numerical models for predicting river flood inundation. *Journal of Hydrology*, 268, 87–99.
26. Hosder, S., Perez, R., & Walters, R. (2006). A non-intrusive polynomial chaos method for uncertainty propagation in cfd simulations. In *48th AIAA Aerospace Sciences Meeting and Exhibit, AIAA-2010-0129. The American Institute of Aeronautics and Astronautics, Inc.*
27. Jakeman, J.D., Eldred, M.S., & Sargsyan, K. (2015). Enhancing  $\ell_1$ -minimization estimates of polynomial chaos expansions using basis selection. *Journal of Computational Physics*, 289, 18–34.
28. Le Maître, O., & Knio, O. (2010). *Spectral methods for uncertainty quantification*. Berlin: Springer.
29. Li, J., & Xiu, D. (2008). On numerical properties of the ensemble Kalman filter for data assimilation. *Comput. Methods Appl. Mech. Engrg*, 197, 3574–3583.
30. Li, J., & Xiu, D. (2009). A generalized polynomial chaos based ensemble kalman filter with high accuracy. *Journal of Computational Physics*, 228(15), 5454–5469. <https://doi.org/10.1016/j.jcp.2009.04.029>.
31. Liang, G., Kwok Fai, C., & Kobayashi, M.H. (2008). Stochastic solution for uncertainty propagation in nonlinear shallow-water equations. *Journal of Hydraulic Engineering*, 134(12), 1732–1743.
32. Matheron, G. (1963). *Traité de géostatistique appliquée, II: Le Krigeage, brgm edn*. Paris: Mémoires du Bureau de Recherches Géologiques et Minières.
33. Migliorati, G., Nobile, F., & Tempone, R. (2015). Convergence estimates in probability and in expectation for discrete least squares with noisy evaluations at random points. *Journal of Multivariate Analysis*, 142, 167–182.
34. Migliorati, G., Nobile, F., Von Schwerin, E., & Tempone, R. (2013). Approximation of quantities of interest in stochastic PDEs by the random Discret L2 Projection on polynomial spaces. *SIAM J. Sci Comput.*, 35(3), A1440–A1460.
35. Moradkhani, H., Sorooshian, S., Gupta, H.V., & Houser, P.R. (2005). Dual state-parameter estimation of hydrological models using ensemble Kalman filter. *Advances in Water Resources*, 28, 135–147. <https://doi.org/10.1016/j.advwatres.2004.09.002>.
36. Poëtte, G., Birolleau, A., & Lucor, D. (2015). Iterative Polynomial Approximation Adapting to Arbitrary Probability Distribution. *SIAM. J. Numerical Analysis*, 53(3), 1559–1584.
37. Resmini, A., Peter, J., & Lucor, D. (2015). Sparse grids-based stochastic approximations with applications to aerodynamics sensitivity analysis. *Int. J. Numer. Meth. Engng* <https://doi.org/10.1002/nme.5005>.
38. Resmini, A., Peter, J., & Lucor, D. (2016). Sparse grids-based stochastic approximations with applications to aerodynamics sensitivity analysis. *International Journal for Numerical Methods in Engineering*, 106(1), 32–57. <https://doi.org/10.1002/nme.5005>.
39. Rochoux, M. (2014). *Vers une meilleure prévision de la propagation d'incendies de forêt : évaluation de modèles et assimilation de données*. Ecole Centrale Paris: Ph.D. thesis.
40. Rochoux, M.C., Ricci, S., Lucor, D., Cuenot, B., & Trouvé, A. (2014). Towards predictive data-driven simulations of wildfire spread - Part I: Reduced-cost Ensemble Kalman Filter based on a Polynomial Chaos surrogate model for parameter estimation. *Nat. Hazards and Earth Syst Sci.*, 14(11), 2951–2973.
41. Spanos, P., & Ghanem, R. (1991). *Stochastic Finite Elements: A Spectral Approach*. Berlin: Springer.
42. Sudret, B. (2008). Global sensitivity analysis using polynomial chaos expansions. *Reliab Eng Syst Safety*, 93(7), 964–979.
43. Tatang, M., Pan, W., Prinn, R., & McRae, G. (1997). An efficient method for parametric uncertainty analysis of numerical geophysical models. *Journal of Geophysical Research*, 102, 21,925–21,932.
44. Thual, O. (2010). *Hydrodynamique de l'environnement*. Palaiseau: Ecole polytechnique.
45. Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B*, 58, 267–288.
46. Van Langenhove, J., Lucor, D., & Belme, A. (2016). Robust uncertainty quantification using preconditioned least-squares polynomial with  $l_1$ -regularization. *Int. Journal for UQ*, 6, 57–77.
47. Ven, T.C. (1959). *Open-channel hydraulics*. New York: Mc Graw Hill.
48. Wiener, N. (1938). The homogeneous chaos. *Am. J. Math*, 60(4), 897–936.
49. Winokur, J. (2015). Adaptive sparse grid approaches to polynomial chaos expansions for uncertainty quantification. Ph.D. thesis, Duke University, Department of Mechanical Engineering & Materials Science.
50. Xiang, S. (2012). Asymptotics on laguerre or hermite polynomial expansions and their applications in gauss quadrature. *J.Math. Anal. Appl.*, 393(2), 434–444.

51. Xiu, D. (2010). *Numerical Methods for Stochastic Computations: A Spectral Method Approach*. Princeton: Princeton University Press.
52. Xiu, D., & Hesthaven, J.S. (2005). High-order collocation methods for differential equations with random inputs. *J. Sci. Comput.*, 27(3), 1118–1139.
53. Xiu, D., & Karniadakis, G.E. (2002). The wiener–askey polynomial chaos for stochastic differential equations. *SIAM Journal on Scientific Computing*, 24(2), 619–644. <https://doi.org/10.1137/S1064827501387826>.
54. Xiu, D., & Karniadakis, G.E. (2003). Modeling uncertainty in flow simulations via generalized polynomial chaos. *Journal of Computational Physics*, 187(1), 137–167.
55. Yan, L., Guo, L., & Xiu, D. (2012). Stochastic collocation algorithms using  $\ell_1$ -minimization. *International Journal for Uncertainty Quantification*, 2, 3.
56. Yang, X., & Karniadakis, G.E. (2013). Reweighted  $\ell_1$  minimization method for stochastic elliptic differential equations. *Journal of Computational Physics*, 248, 87–108.
57. Zhou, J., Causon, D., Mingham, C., & Ingram, D. (2004). Numerical Prediction of Dam Break Flows in General Geometries with Complex Bed Topography. *Journal of Hydraulic Engineering*, 130(4), 332–340.
58. Zokagoa, J.M., & Soulaïmani, A. (2012). A POD-based reduced order model for free surface shallow water flows over real bathymetries for Monte-Carlo-type applications. *Comput. Methods Appl. Mech. Engrg.*, 221–222, 1–23.