Challenges in Stochastic Galerkin Methods for Nonlinear Hyperbolic Systems with Uncertainty

Alina Chertock[0000-0003-4978-1314]
and Michael Herty[0000-0002-6262-2927]
and Alexander Kurganov[0000-0003-0231-986X]
and Mária Lukáčová-Medviďová[0000-0002-4351-0161]

Abstract We consider a one-dimensional Saint-Venant system with uncertainty, for which we derive and analyze a novel stochastic Galerkin method. The proposed method is based on the truncated generalized polynomial chaos (gPC) expansion, whose coefficients satisfy a time-dependent system of PDEs, which is hyperbolic provided the water depth remains nonnegative at all times and for all values of both spatial and stochastic variables. We numerically solve the resulting system using a Riemann-problem-solver-free well-balanced and positivity-preserving finite-volume central-upwind scheme. The novelty of our approach is in the way we enforce the positivity of the computed water depth—this is achieved by generalizing the "draining time-step" technique to the system of gPC coefficients. We illustrate the performance of the proposed method on a number of challenging numerical examples. Though in some of the considered benchmarks, we obtain accurate mean and standard deviation of the stochastic solution, we realize that (small) oscillations appearing near the discontinuities propagate into the stochastic field and cause quite significant oscillations attributed to the Gibbs phenomenon. This demonstrates the limitations in the applicability of the stochastic Galerkin method to the problems with discontinuous solutions. As a possible way to remove (reduce) the aforementioned Gibbs oscillations, we propose to add an adaptive artificial viscosity to the system of

Alina Chertock
North Carolina State University, Raleigh, NC, USA, e-mail: chertock@math.ncsu.edu

Michael Herty
RWTH Aachen University, Institut für Geometrie und Praktische Mathematik, Aachen, Germany, e-mail: herty@igpm.rwth-aachen.de

Alexander Kurganov
Shenzhen International Center for Mathematics, and Guangdong Provincial Key Laboratory of Computational Science and Material Design, Southern University of Science and Technology, Shenzhen, China, e-mail: alexander@sustech.edu.cn

Mária Lukáčová-Medviďová
Institute of Mathematics, Johannes Gutenberg University Mainz, Mainz, Germany, e-mail: lukacova@uni-mainz.de

gPC coefficients. However, this, like other existing filtering alternatives, affects the high resolution of the gPC stochastic Galerkin method.

Keywords: Nonlinear hyperbolic systems with uncertainty, generalized polynomial chaos, stochastic Galerkin method, central-upwind schemes, adaptive artificial viscosity, Gibbs phenomenon.

1 Introduction

In this paper, we study nonlinear hyperbolic systems with uncertainty. In the one-dimensional (1-D) case, such systems of balance laws read as

$$\boldsymbol{U}_t+\boldsymbol{F}(\boldsymbol{U})_x=\boldsymbol{S}(\boldsymbol{U},x,\xi), \tag{1}$$

where x is the spatial variable, t is time, ξ is a random variable, $\boldsymbol{U}=\boldsymbol{U}(x,t,\xi)$ is the unknown random vector function, $\boldsymbol{F}$ is the flux, and $\boldsymbol{S}$ is the source term. The uncertainty may appear in the system parameters as well as in the initial and/or boundary data due to empirical approximations, or measuring errors.

Nonlinear hyperbolic systems with uncertainty appear in a wide variety of applications. Quantifying the uncertainty is important since it helps to conduct sensitivity analysis and provide guidance on the predictive quality of the models.

In recent years, a wide variety of uncertainty quantification methods for nonlinear hyperbolic systems has been proposed and investigated. One of the popular class methods employ Monte Carlo-type simulations, see, e.g., [1,26–28,42], which are robust, but not very efficient due to a possibly large number of realizations required. In addition to the Monte Carlo methods, a widely used approach for random PDEs is the generalized polynomial chaos (gPC), where stochastic processes are represented in terms of orthogonal polynomials series of random variables; see, e.g., [1,30,42] and references therein. In principle, there are two distinct gPC approaches: intrusive and non-intrusive ones. In non-intrusive algorithms, like stochastic collocation methods (see, e.g., [1,2,30,42,43]) one seeks to satisfy the governing equations at a discrete set of points in the random space and then use global interpolation and quadrature rules to numerically evaluate statistical moments. Therefore, in the stochastic collocation approach, as well as in the Monte-Carlo methods, one can use numerical methods designed for the corresponding deterministic systems; see, e.g., the monographs and review paper [4,6,18,20,25,39,40] and references therein.

In the case of an intrusive approach, like stochastic Galerkin (SG) methods, gPC expansions are substituted into the governing equations and projected by a Galerkin approximation to obtain deterministic equations for the expansion coefficients; see, e.g., [8,24,41,42]. Solving the coefficient equations gives the stochastic moments of the random solution. The equations for the expansion coefficients are, in general, nonlinear and coupled. Nevertheless, the gPC-SG methods are often more accurate than their non-intrusive counterparts when the the solution is sufficiently smooth and the same number of modes in the gPC expansion is used. Therefore, the gPC-SG methods are expected to achieve a higher accuracy of the numerical solution with a lower degree of the gPC expansion [13].

Development and implementation of the SG methods for nonlinear hyperbolic PDEs are, however, especially challenging due to a possible loss of hyperbolicity of the gPC expansion coefficient system [11, 12]. Several approaches exist to prevent this loss for specific models; see, e.g., [11, 14, 15, 31, 32], but there is no general remedy to guarantee the hyperbolicity of the the gPC expansion coefficient system.

This paper focuses on another substantial drawback of the gPC-SG method, which is related to the Gibbs phenomenon. It is well-known that when spectral methods are applied to capture solutions containing shock discontinuities, the obtained results will be oscillatory near jumps unless certain filters (see, e.g., [5, 19]) or spectral viscosity (see, e.g., [36–38]) are implemented for large modes. The latter tools, however, affect the accuracy of the spectral methods. In the studied gPC-SG methods, the situation is even more complicated as the aforementioned tools do not directly apply due to the fact that the expansion coefficients are functions of both t and x, while in standard spectral methods the coefficients depend on t only. Moreover, the size of the PDE system for the gPC expansion coefficients rapidly grows when the number of modes increases, which forces one to use a relatively small number of modes, which, in turn, makes the use of filtering or spectral viscosity even more challenging.

To showcase the outlined difficulties, we consider the Saint-Venant system of shallow water equations, which is widely used for modeling flows in rivers, lakes, and coastal areas, as well as in models emerging in oceanography and atmospheric sciences. In the 1-D case, the studied system reads as

$$\begin{cases} h_t + q_x = 0, \\ q_t + \left(hu^2 + \dfrac{g}{2}h^2\right)_x = -ghZ_x, \end{cases} \tag{2}$$

where the water depth $h(x,t,\xi)$, velocity $u(x,t,\xi)$, and discharge $q := hu$ are time-dependent quantities, while the bottom topography $Z(x,\xi)$ is independent of time, and g is the acceleration due to gravity.

It is well-known that the random shallow water system (2) is hyperbolic as long as the water depth $h(x,t,\xi) \geq 0$. It was also shown in [9] that when the gPC-SG method is implemented for (2), the obtained system of gPC coefficients is hyperbolic, provided the same non-negativity condition holds. Nevertheless, it should be observed that solving the gPC coefficient system numerically is a challenging task: If one uses $K+1$ terms in each of the gPC expansions, the resulting gPC coefficient system for (2) consists of $2(K+1)$ equations. Therefore, the use of any numerical method that requires information on the full eigenstructure of the underlying system may not be feasible. Consequently, we follow [9] and numerically solve the gPC coefficient system using a semi-discrete second-order well-balanced (WB) central-upwind (CU) scheme, which was developed in the context of the deterministic Saint-Venant system in [21, 23]. The choice of the CU scheme is motivated by the fact that this scheme is Riemann-problem-solver-free and thus can be applied as a "black-box solver" to general hyperbolic systems as long as the largest and smallest eigenvalues of its Jacobian can be estimated. Moreover, the CU scheme is WB in the sense that it is capable of exactly preserving "lake-at-rest" equilibria satisfying

$$q(x,\xi) \equiv 0, \quad w(x,\xi) := h(x,\xi) + Z(x,\xi) = C_2(\xi).$$

Implementing the WB-CU scheme may be, however, not enough for computing a reliable solution as negative values of the water depth may appear, and then the method would fail. This is related to the emergence of spurious oscillations, which may be generated during the time evolution of the gPC coefficients. In order to control these oscillations, one may apply filters from [33] as it was done in [9, 10]. This, however, can yield the loss of statistical information on the solution if the filtering is used to damp large-magnitude oscillations. In this paper, we propose a different approach for controlling the nonnegativity of the water depth using the "draining time-step" technique originally developed in the deterministic case in [3]. In addition, we use the positivity correction procedure from [7] to ensure the nonnegativity of the water depth mean, which is the zeroth coefficient in the gPC expansion of h.

As it was mentioned above, enforcing the positivity of the computed water depth h guarantees the hyperbolicity of the gPC coefficient system. However, this doesn't cure the problem of Gibbs-related oscillations, which can grow rapidly and substantially affect the accuracy of the numerical solution, even if h is far from zero. Hence, additionally we need to suppress Gibbs-related oscillations. To achieve this goal, we add an artificial viscosity in the following way: we add first-order viscosity to the equations for high modes of the gPC expansion, and a weak local residual (WLR) based adaptive artificial viscosity (AAV) [22] to the low mode equations. When computed by the modified scheme, the obtained results are almost oscillation-free. They are, however, rather smeared and a high accuracy expected to be achieved by a gPC-SG method may not be reached.

We consider the simplest case of a uniformly distributed $\xi \in [-1, 1]$ and test the developed gPC-SG method on a number of numerical examples. We demonstrate that when strong discontinuities are present, the basic CU scheme (without the added artificial viscosity) produces significant oscillations appearing near the discontinuities and propagating in both the physical and stochastic fields. It has been conjectured in the literature (see, e.g., [11, 12]) that these oscillations arise from the loss of hyperbolicity of the gPC-SG system, which is related to the positivity of the water depth $h(x,t,\xi)$. In our numerical examples, we demonstrate that enforcing the nonnegativity of the computed water depth does not guarantee the lack of oscillations attributed to the Gibbs phenomenon. Adding the artificial viscosity (or using another filtering technique) is required to improve the robustness of the gPC-SG method. At the same time, the artificial viscosity leads to a substantial smoothing of the computed solutions, which demonstrates the limitations in the ability of the gPC-SG method to achieve high resolution of discontinuous solutions.

2 The gPC-SG Formulation – An Overview

In order to describe the uncertainty, we introduce a probability space $(\Xi_\omega, \mathcal{B}(\Xi_\omega), \mathbb{P})$ with events $\omega \in \Xi_\omega$, where Ξ_ω is an event space and $\mathcal{B}$ is a set of Borel measurable sets with respect to the probability measure $\mathbb{P}$. Denote by $\xi = \xi(\omega) : \Xi_\omega \to \Xi \subset \mathbb{R}^d$ a real-valued random variable with the probability density function $\mu(\xi) : \Xi \to \mathbb{R}_+$.

We proceed with a brief description of the gPC-SG method applied to a general hyperbolic system of balance laws (1). In the gPC expansion, the solution of (1) is sought in terms of an orthogonal polynomial series in ξ (see, e.g., [11,44]):

$$\boldsymbol{U}(x,t,\xi) \approx \boldsymbol{U}^K(x,t,\xi) = \sum_{i=0}^{K} \widehat{\boldsymbol{U}}_i(x,t)\Phi_i(\xi). \tag{3}$$

Here, $\{\Phi_i(\xi)\}_{i=0}^K$ are orthonormal polynomials of degree up to $K \ge 1$ satisfying

$$\int\limits_{\Xi} \Phi_i(\xi)\Phi_\ell(\xi)\mu(\xi)\,\mathrm{d}\xi = \delta_{i\ell}, \qquad i,\ell = 0,1,\ldots,$$

where $\delta_{i\ell}$ is the Kronecker symbol. The choice of the polynomials depends on μ. For instance, the Legendre polynomials correspond to a uniform distribution (this is the case considered in the numerical examples reported in §6); the Hermite polynomials correspond to a Gaussian distribution; etc. For a comprehensive list of the gPC bases for some common distributions, we refer the reader to, e.g., [45] and references therein.

For the PDE system with random inputs (1), the gPC-SG method seeks to satisfy the governing equations in a weak form by ensuring that the residual is orthogonal to the gPC polynomial space. Substituting the approximation $\boldsymbol{U}^K$ from (3) into (1) and using the Galerkin projection yield

$$(\widehat{\boldsymbol{U}}_i)_t + (\widehat{\boldsymbol{F}}_i)_x = \widehat{\boldsymbol{S}}_i, \quad i = 0,\ldots,K, \tag{4}$$

where

$$\left.\begin{aligned}
\widehat{\boldsymbol{F}}_i &= \int\limits_{\Xi} \boldsymbol{F}\bigg(\sum_{j=0}^{K} \widehat{\boldsymbol{U}}_j(x,t)\Phi_j(\xi)\bigg)\Phi_i(\xi)\mu(\xi)\,\mathrm{d}\xi,\\
\widehat{\boldsymbol{S}}_i &= \int\limits_{\Xi} \boldsymbol{S}\bigg(\sum_{j=0}^{K} \widehat{\boldsymbol{U}}_i(x,t)\Phi_i(\xi),x,\xi\bigg)\Phi_i(\xi)\mu(\xi)\,\mathrm{d}\xi,
\end{aligned}\right.\qquad i = 0,\ldots,K. \tag{5}$$

This is a system of deterministic equations for the gPC expansion coefficients $\widehat{\boldsymbol{U}}_i$. In most cases, the equations in (4) are coupled and $\widehat{\boldsymbol{F}}_i$ and $\widehat{\boldsymbol{S}}_i$ in (5) might not be explicitly written in terms of the solution coefficients $\widehat{\boldsymbol{U}}$.

3 A gPC-SG Formulation for the Shallow Water Equations

In this section, we apply the previously outlined approach to the Saint-Venant system (2), for which $\boldsymbol{U} = (h, q)^\top$.

We seek gPC approximations of h and q in the form of (3):

$$h^K(x,t,\xi) = \sum_{i=0}^{K} \widehat{h}_i(x,t)\Phi_i(\xi), \quad q^K(x,t,\xi) = \sum_{i=0}^{K} \widehat{q}_i(x,t)\Phi_i(\xi). \tag{6}$$

Substituting (6) into (2) and conducting the Galerkin projection yield the following equations for the gPC coefficients $\widehat{h}_i$ and $\widehat{q}_i$ for each $i = 0, \ldots, K$:

$$\begin{cases} (\widehat{h}_i)_t + (\widehat{q}_i)_x = 0, \\ (\widehat{q}_i)_t + \left(\sum\limits_{k,\ell=0}^{K} \widehat{q}_k \widehat{u}_\ell \mathcal{S}_{k\ell i} + \dfrac{g}{2} \sum\limits_{k,\ell=0}^{K} \widehat{h}_k \widehat{h}_\ell \mathcal{S}_{k\ell i} \right)_x = -g \sum\limits_{k,\ell=0}^{K} \widehat{h}_k (\widehat{Z}_\ell)_x \mathcal{S}_{k\ell i}, \end{cases} \tag{7}$$

where the corresponding gPC expansions of u and Z are given by

$$u^K(x,t,\xi) = \sum_{i=0}^{K} \widehat{u}_i(x,t)\Phi_i(\xi), \quad Z^K(x,\xi) = \sum_{i=0}^{K} \widehat{Z}_i(x)\Phi_i(\xi),$$

and $\mathcal{S}$ is a symmetric tensor composed of the orthonormal polynomials, that is,

$$\mathcal{S}_{k\ell i} = \int\limits_{\Xi} \Phi_k(\xi)\Phi_\ell(\xi)\Phi_i(\xi)\mu(\xi)\,\mathrm{d}\xi.$$

The system (7) can be rewritten in the following compact form:

$$\begin{cases} \widehat{\boldsymbol{h}}_t + \widehat{\boldsymbol{q}}_x = \boldsymbol{0}, \\ \widehat{\boldsymbol{q}}_t + \left(\mathcal{P}(\widehat{\boldsymbol{q}})\widehat{\boldsymbol{u}} + \dfrac{g}{2}\mathcal{P}(\widehat{\boldsymbol{h}})\widehat{\boldsymbol{h}} \right)_x = -g\mathcal{P}(\widehat{\boldsymbol{h}})\widehat{\boldsymbol{Z}}_x, \end{cases} \tag{8}$$

where $\widehat{\boldsymbol{h}} = (\widehat{h}_i)_{i=0}^K$, $\widehat{\boldsymbol{q}} = (\widehat{q}_i)_{i=0}^K$, $\widehat{\boldsymbol{u}} = (\widehat{u}_i)_{i=0}^K$ and $\widehat{\boldsymbol{Z}} = (\widehat{Z}_i)_{i=0}^K$ are vectors of the gPC coefficients in $\mathbb{R}^{K+1}$, and an operator $\mathcal{P} : \mathbb{R}^{K+1} \longrightarrow \mathbb{R}^{(K+1)\times(K+1)}$ is defined by

$$[\mathcal{P}(\widehat{\boldsymbol{\alpha}})]_{i,\ell} = \sum_{k=0}^{K} \widehat{\alpha}_k \mathcal{S}_{i\ell k}, \qquad i, \ell = 0, \ldots, K.$$

Note that the gPC coefficients $\widehat{u}_i$ are computed by applying the gPC-SG approximation to $hu = q$, which results in the linear system

$$\sum_{k,\ell=0}^{K} \widehat{h}_k \widehat{u}_\ell \mathcal{S}_{i\ell k} = \widehat{q}_i, \ \ i, \ell = 0, \ldots, K \quad \Longleftrightarrow \quad \mathcal{P}(\widehat{\boldsymbol{h}})\widehat{\boldsymbol{u}} = \widehat{\boldsymbol{q}}, \tag{9}$$

whose solution can be written as $\widehat{\boldsymbol{u}} = \left[\mathcal{P}(\widehat{\boldsymbol{h}})\right]^{-1}\widehat{\boldsymbol{q}}$, which requires the matrix $\mathcal{P}(\widehat{\boldsymbol{h}})$ to be invertible.

We note that in order to develop a numerical method for the system (8), one needs to ensure its hyperbolicity. The following property has been proved in [9, Theorem 3.1].

Proposition 1. *If the matrix $\mathcal{P}(\widehat{\boldsymbol{h}})$ is strictly positive definite, then the system* (8) *is hyperbolic.*

Next, we establish a sufficient condition for the matrix $\mathcal{P}(\widehat{\boldsymbol{h}})$ to be strictly positive definite. To this end, we compute the following quadratic form for an arbitrary nonzero $\widehat{\boldsymbol{\alpha}} = (\widehat{\alpha}_i)_{i=0}^K \in \mathbb{R}^{K+1}$:

$$\begin{aligned}
\widehat{\boldsymbol{\alpha}}^\top \mathcal{P}(\widehat{\boldsymbol{h}})\widehat{\boldsymbol{\alpha}} &= \sum_{i,j,k=0}^{K} \widehat{\alpha}_j \widehat{\alpha}_k \widehat{h}_i(x,t) \int_{\Xi} \Phi_i(\xi)\Phi_j(\xi)\Phi_k(\xi)\mu(\xi)\,\mathrm{d}\xi \\
&= \int_{\Xi} \sum_{i=0}^{K} \widehat{h}_i(x,t)\Phi_i(\xi)\big(\widehat{\boldsymbol{\Phi}}(\xi)^\top \widehat{\boldsymbol{\alpha}}\big)^2 \mu(\xi)\,\mathrm{d}\xi \\
&= \int_{\Xi} h^K(x,t,\xi)\big(\widehat{\boldsymbol{\Phi}}(\xi)^\top \widehat{\boldsymbol{\alpha}}\big)^2 \mu(\xi)\,\mathrm{d}\xi,
\end{aligned}$$

where $\widehat{\boldsymbol{\Phi}}(\xi) = (\Phi_i)_{i=0}^K \in \mathbb{R}^{K+1}$. Hence, the matrix $\mathcal{P}(\widehat{\boldsymbol{h}})$ will be strictly positive definite provided $h^K(x,t,\xi) > 0$, which is a physically meaningful condition of non-negativity of the water depth.

4 Well-Balanced Positivity-Preserving Central-Upwind Scheme

In this section, we present a modified version of the WB-CU scheme from [9] for the system (7), which can be written in the form (4) with

$$\widehat{\boldsymbol{U}}_i = \begin{pmatrix} \widehat{h}_i \\ \widehat{q}_i \end{pmatrix}, \quad \widehat{\boldsymbol{F}}_i = \begin{pmatrix} \widehat{q}_i \\ \left[\mathcal{P}(\widehat{\boldsymbol{q}})\widehat{\boldsymbol{u}} + \frac{g}{2}\mathcal{P}(\widehat{\boldsymbol{h}})\widehat{\boldsymbol{h}}\right]_i \end{pmatrix}, \quad \widehat{\boldsymbol{S}}_i = \begin{pmatrix} 0 \\ -\left[g\mathcal{P}(\widehat{\boldsymbol{h}})\widehat{\boldsymbol{Z}}_x\right]_i \end{pmatrix}.$$

We divide the computational domain into the uniform cells $C_j = [x_{j-\frac12}, x_{j+\frac12}]$ of size $x_{j+\frac12} - x_{j-\frac12} = \Delta x$ centered at points $x_j = \big(x_{j-\frac12} + x_{j+\frac12}\big)/2$, and assume that at a certain time t the cell averages of the computed solution,

$$(\overline{\boldsymbol{U}}_i)_j(t) :\approx \frac{1}{\Delta x}\int_{C_j} \widehat{\boldsymbol{U}}_i(x,t)\,\mathrm{d}x, \quad i = 0,\dots,K,$$

are available. Then, the semi-discrete WB-CU scheme from [9] reads as

$$\frac{\mathrm{d}}{\mathrm{d}t}(\overline{\boldsymbol{U}}_i)_j = -\frac{(\boldsymbol{\mathcal{F}}_i)_{j+\frac{1}{2}} - (\boldsymbol{\mathcal{F}}_i)_{j-\frac{1}{2}}}{\Delta x} + (\overline{\boldsymbol{S}}_i)_j, \quad i = 0, \ldots, K, \tag{10}$$

where $(\boldsymbol{\mathcal{F}}_i)_{j+\frac{1}{2}}$ are the CU numerical fluxes given by

$$\begin{aligned}(\boldsymbol{\mathcal{F}}_i)_{j+\frac{1}{2}} = {} & \frac{a^+_{j+\frac{1}{2}}\widehat{\boldsymbol{F}}_i\big((\widehat{\boldsymbol{U}}_i)^-_{j+\frac{1}{2}}\big) - a^-_{j+\frac{1}{2}}\widehat{\boldsymbol{F}}_i\big((\widehat{\boldsymbol{U}}_i)^+_{j+\frac{1}{2}}\big)}{a^+_{j+\frac{1}{2}} - a^-_{j+\frac{1}{2}}} \\ & + \frac{a^+_{j+\frac{1}{2}}a^-_{j+\frac{1}{2}}}{a^+_{j+\frac{1}{2}} - a^-_{j+\frac{1}{2}}}\Big[(\widehat{\boldsymbol{U}}_i)^+_{j+\frac{1}{2}} - (\widehat{\boldsymbol{U}}_i)^-_{j+\frac{1}{2}}\Big],\end{aligned} \tag{11}$$

and $(\overline{\boldsymbol{S}}_i)_j \approx \frac{1}{\Delta x}\int_{C_j}\widehat{\boldsymbol{S}}_i\,\mathrm{d}x$ are approximated cell averages of the source term. In (11), $(\widehat{\boldsymbol{U}}_i)^\pm_{j+\frac{1}{2}}$ are one-sided point values of $\widehat{\boldsymbol{U}}_i$ at the cell interfaces $x = x_{j+\frac{1}{2}}$ and $a^\pm_{j+\frac{1}{2}}$ are the one-sided local speeds of propagation at the same points. These quantities will be given below.

Note that from here on we suppress the time-dependence of all of the indexed quantities in order to shorten the notation.

Piecewise Linear Approximation of $\widehat{\boldsymbol{Z}}$. Before reconstructing the point values $(\widehat{\boldsymbol{U}}_i)^\pm_{j+\frac{1}{2}}$ we follow [23] and replace the functions $\widehat{\boldsymbol{Z}}$ with their continuous piecewise linear approximations (this makes it easier to numerically preserve positivity of the reconstructed water depths):

$$\widehat{Z}^*_i(x) = (\widehat{Z}_i)_{j-\frac{1}{2}} + \Big[(\widehat{Z}_i)_{j+\frac{1}{2}} - (\widehat{Z}_i)_{j-\frac{1}{2}}\Big]\frac{x - x_{j-\frac{1}{2}}}{\Delta x}, \quad x \in C_j, \quad i = 0, \ldots, K,$$

where $(\widehat{Z}_i)_{j+\frac{1}{2}} := \big[(\widehat{Z}_i)(x_{j+\frac{1}{2}}-) + (\widehat{Z}_i)(x_{j+\frac{1}{2}}+)\big]/2$. We then define $(\widehat{Z}_i)_j := \widehat{Z}^*_i(x_j) = \big[(\widehat{Z}_i)_{j+\frac{1}{2}} + (\widehat{Z}_i)_{j-\frac{1}{2}}\big]/2$.

Well-Balanced Reconstruction. In order to obtain a WB scheme it is important to reconstruct the equilibrium variables $\widehat{\boldsymbol{w}} = \widehat{\boldsymbol{h}} + \widehat{\boldsymbol{Z}}$ and $\widehat{\boldsymbol{q}}$ rather than $\widehat{\boldsymbol{h}}$ and $\widehat{\boldsymbol{q}}$. We restrict our attention to the second-order scheme and apply the non-oscillatory minmod reconstruction (see, e.g., [29, 35]) to the cell averages of each of the gPC coefficients, $(\overline{w}_i)_j := (\overline{h}_i)_j + (\widehat{Z}_i)_j$ and $(\overline{q}_i)_j$, $i = 0, \ldots, K$:

$$(\widehat{w}_i)^\pm_{j\mp\frac{1}{2}} = (\overline{w}_i)_j \mp \frac{\Delta x}{2}((\widehat{w}_i)_x)_j, \quad (\widehat{q}_i)^\pm_{j\mp\frac{1}{2}} = (\overline{q}_i)_j \mp \frac{\Delta x}{2}((\widehat{q}_i)_x)_j,$$

where

$$\begin{aligned}((\widehat{w}_i)_x)_j &= \operatorname{minmod}\left(\frac{(\overline{w}_i)_j - (\overline{w}_i)_{j-1}}{\Delta x}, \frac{(\overline{w}_i)_{j+1} - (\overline{w}_i)_j}{\Delta x}\right), \\ ((\widehat{q}_i)_x)_j &= \operatorname{minmod}\left(\frac{(\overline{q}_i)_j - (\overline{q}_i)_{j-1}}{\Delta x}, \frac{(\overline{q}_i)_{j+1} - (\overline{q}_i)_j}{\Delta x}\right),\end{aligned}$$

and the minmod function is defined as $\mathrm{minmod}(a,b) := \frac{\mathrm{sgn}(a)+\mathrm{sgn}(b)}{2}\min(|a|,|b|)$. We then compute the point values $(\widehat{h}_i)^{\pm}_{j+\frac12} = (\widehat{w}_i)^{\pm}_{j+\frac12} - (\widehat{Z}_i)_{j+\frac12}$ and hence $(\widehat{\boldsymbol{U}}_i)^{\pm}_{j+\frac12} = \big((\widehat{h}_i)^{\pm}_{j+\frac12}, (\widehat{q}_i)^{\pm}_{j+\frac12}\big)^{\top}$.

In what follows, we will denote by $\widehat{\boldsymbol{U}}^{\pm}_{j+\frac12} := \big((\widehat{\boldsymbol{U}}_0)^{\pm}_{j+\frac12}, \dots, (\widehat{\boldsymbol{U}}_K)^{\pm}_{j+\frac12}\big)^{\top}$ and similarly for $\widehat{\boldsymbol{h}}^{\pm}_{j+\frac12}$, $\widehat{\boldsymbol{q}}^{\pm}_{j+\frac12}$, and $\widehat{\boldsymbol{u}}^{\pm}_{j+\frac12}$.

Positivity Correction. The positivity of the expected value of h^K, that is,

$$\mathbb{E}(h) = \int\limits_{\Xi} h^K(x,t,\xi)\mu(\xi)\,\mathrm{d}\xi = \widehat{h}_0(x,t),$$

is a fundamental property, which has to be satisfied at the discrete level. Unfortunately, the use of the minmod or any other standard nonlinear limiter cannot guarantee positivity of the reconstructed point values of $(\widehat{h}_0)^{\pm}_{j+\frac12}$. Therefore, we correct the reconstruction of $\widehat{w}_0$ following the procedure suggested in [9]:
if $(\widehat{w}_0)^{+}_{j-\frac12} < (\widehat{Z}_0)_{j-\frac12}$, then set $(\widehat{w}_0)^{+}_{j-\frac12} = (\widehat{Z}_0)_{j-\frac12}$ and $(\widehat{w}_0)^{-}_{j+\frac12} = (\widehat{Z}_0)_{j+\frac12} + 2(\overline{h}_0)_j$;
if $(\widehat{w}_0)^{-}_{j+\frac12} < (\widehat{Z}_0)_{j+\frac12}$, then set $(\widehat{w}_0)^{-}_{j+\frac12} = (\widehat{Z}_0)_{j+\frac12}$ and $(\widehat{w}_0)^{+}_{j-\frac12} = (\widehat{Z}_0)_{j-\frac12} + 2(\overline{h}_0)_j$.
This choice guarantees that all of the point values $(\widehat{h}_0)^{\pm}_{j+\frac12} = (\widehat{w}_0)^{\pm}_{j+\frac12} - (\widehat{Z}_0)_{j+\frac12}$ are non-negative as long as $(\overline{h}_0)_j \geq 0$ for all j.

Computation of the Point Values $\widehat{\boldsymbol{u}}^{\pm}_{j+\frac12}$. In order to compute the point values $\widehat{\boldsymbol{u}}^{\pm}_{j+\frac12}$, the matrix $\mathcal{P}\big(\widehat{\boldsymbol{h}}^{\pm}_{j+\frac12}\big)$ should be invertible (see (9)) and hence the computation of its inverse should be desingularized in the case of small or zero water depths. Let us consider the matrix $\mathcal{P}\big(\widehat{\boldsymbol{h}}^{+}_{j+\frac12}\big)$ (the matrix $\mathcal{P}\big(\widehat{\boldsymbol{h}}^{-}_{j+\frac12}\big)$ can be treated similarly) and diagonalize it by $\mathcal{P}\big(\widehat{\boldsymbol{h}}^{+}_{j+\frac12}\big) = T^{-1}\Lambda T$, where Λ is a diagonal matrix. If the matrix Λ contains a small or zero eigenvalue λ_{i_0}, we desingularize it as in [9] by replacing $\mathcal{P}^{-1}\big(\widehat{\boldsymbol{h}}^{+}_{j+\frac12}\big)$ with

$$\mathcal{P}_{\varepsilon}^{-1}\big(\widehat{\boldsymbol{h}}^{+}_{j+\frac12}\big) = T^{-1}\mathrm{diag}\left(\frac{1}{\lambda_0},\dots,\frac{\sqrt{2}\,\lambda_{i_0}}{\sqrt{\lambda_{i_0}^4 + \big[\max\{\lambda_{i_0},\varepsilon\}\big]^4}},\dots,\frac{1}{\lambda_K}\right)T. \tag{12}$$

We then define

$$\widehat{\boldsymbol{u}}^{\pm}_{j+\frac12} = \mathcal{P}_{\varepsilon}^{-1}\big(\widehat{\boldsymbol{h}}^{\pm}_{j+\frac12}\big)\,\widehat{\boldsymbol{q}}^{\pm}_{j+\frac12}. \tag{13}$$

Note that $\mathcal{P}\big(\widehat{\boldsymbol{h}}^{\pm}_{j+\frac12}\big) \in \mathbb{R}^{(K+1)\times(K+1)}$ and hence the diagonalization of these matrices should be carried out numerically using an appropriate numerical linear algebra tool.

Remark 2. *In the numerical results reported in §6, we have set the desingularization parameter $\varepsilon = \Delta x$.*

Computation of the One-Sided Local Speeds $a^\pm_{j+\frac12}$. The left- and right-sided local speeds in (11) can be estimated using the smallest and largest eigenvalues of the Jacobian of the system (8), for instance, as follows:

$$\begin{aligned}
a^-_{j+\frac12} &= \min\left\{\lambda_1\Big(\mathcal{J}\big(\widehat{\boldsymbol{h}}^-_{j+\frac12},\widehat{\boldsymbol{u}}^-_{j+\frac12}\big)\Big),\lambda_1\Big(\mathcal{J}\big(\widehat{\boldsymbol{h}}^+_{j+\frac12},\widehat{\boldsymbol{u}}^+_{j+\frac12}\big)\Big),0\right\},\\
a^+_{j+\frac12} &= \max\left\{\lambda_{2(K+1)}\Big(\mathcal{J}\big(\widehat{\boldsymbol{h}}^-_{j+\frac12},\widehat{\boldsymbol{u}}^-_{j+\frac12}\big)\Big),\lambda_{2(K+1)}\Big(\mathcal{J}\big(\widehat{\boldsymbol{h}}^+_{j+\frac12},\widehat{\boldsymbol{u}}^+_{j+\frac12}\big)\Big),0\right\}.
\end{aligned}$$

Well-Balanced Quadrature of the Source Term. The cell averages of the source term $(\overline{\boldsymbol{S}}_i)_j$ in (10) should be approximated by an appropriate quadrature. The choice of the quadrature rule is very important to ensure the WB property of the resulting scheme. In this work, we use a WB quadrature from [9, equation (4.7)]:

$$(\overline{\boldsymbol{S}}_i)_j=\left(0,-\frac{g}{\Delta x}\left[\mathcal{P}(\overline{\boldsymbol{h}}_j)\big(\widehat{\boldsymbol{Z}}_{j+\frac12}-\widehat{\boldsymbol{Z}}_{j-\frac12}\big)\right]_i\right)^\top.$$

Time Evolution and "Draining Time-Step". The ODE system (10) needs to be numerically integrated by an appropriated ODE solver. The simplest one is the first-order forward Euler one, which reads as

$$(\overline{\boldsymbol{U}}_i)^{n+1}_j=(\overline{\boldsymbol{U}}_i)^n_j-\lambda^n\left[(\boldsymbol{\mathcal{F}}_i)^n_{j+\frac12}-(\boldsymbol{\mathcal{F}}_i)^n_{j-\frac12}\right]+\Delta t^n(\overline{\boldsymbol{S}}_i)^n_j,\quad \lambda^n:=\frac{\Delta t^n}{\Delta x},\tag{14}$$

where the upper index n denotes that the corresponding quantity is evaluated at either the time level $t=t^n$ or $t^{n+1}:=t^n+\Delta t^n$. The fully discrete version of the CU scheme (14) will be, in general, stable if the following CFL condition is satisfied:

$$\Delta t^n\le\frac{\Delta x}{\max\limits_j\big(a^+_{j+\frac12}-a^-_{j+\frac12}\big)}.\tag{15}$$

The CFL condition (15), however, does not guarantee that the computed water depths remain non-negative. Therefore, we adopt the "draining time-step" technique originally introduced in [3] to enforce the non-negativity of the water depth in deterministic shallow water computations. The implementation of the "draining time-step" approach to the system (14) is not straightforward since the water depth h^K is only given in terms of its gPC coefficients. Therefore, we first reconstruct in ξ based on the the cell averages of h^K at time $t=t^n$ by

$$\big(\overline{h^K}\big)^n_j(\xi)=\sum_{i=0}^K(\overline{h}_i)^n_j\Phi_i(\xi).\tag{16}$$

We also recover the first component of the numerical flux at time $t=t^n$ and introduce

$$\mathcal{F}^{(1),n}_{j+\frac12}(\xi):=\sum_{i=0}^K\big(\mathcal{F}_i^{(1),n}\big)_{j+\frac12}\Phi_i(\xi),\tag{17}$$

where

$$\begin{aligned}\left(\mathcal{F}_i^{(1),n}\right)_{j+\frac{1}{2}} &= \frac{a^+_{j+\frac{1}{2}}(\widehat{q}_i)^-_{j+\frac{1}{2}} - a^-_{j+\frac{1}{2}}(\widehat{q}_i)^-_{j+\frac{1}{2}}}{a^+_{j+\frac{1}{2}} - a^-_{j+\frac{1}{2}}} \\ &\quad + \frac{a^+_{j+\frac{1}{2}}a^-_{j+\frac{1}{2}}}{a^+_{j+\frac{1}{2}} - a^-_{j+\frac{1}{2}}}\left[(\widehat{h}_i)^+_{j+\frac{1}{2}} - (\widehat{h}_i)^-_{j+\frac{1}{2}}\right]\end{aligned} \tag{18}$$

with the reconstructed point values $(\widehat{h}_i)^\pm_{j+\frac{1}{2}}$ and $(\widehat{q}_i)^\pm_{j+\frac{1}{2}}$ evaluated at time $t = t^n$.

Equipped with (16)–(18), we follow [3] and evaluate the “draining time-step”, which is given by

$$\Delta t_j^{\text{drain}}(\xi) = \frac{\Delta x\,\left(\overline{h^K}\right)_j^n(\xi)}{\max\left\{0, \mathcal{F}^{(1),n}_{j+\frac{1}{2}}(\xi)\right\} + \max\left\{0, -\mathcal{F}^{(1),n}_{j-\frac{1}{2}}(\xi)\right\}}. \tag{19}$$

We then introduce the quantities

$$\Delta t^n_{j+\frac{1}{2}} = \min_{\xi\in\Xi}\left\{\min\left\{\Delta t^n, \Delta t^{\text{drain}}_\ell\right\}\right\}, \quad \ell = j + \frac{1}{2} - \frac{1}{2}\operatorname{sgn}\left(\mathcal{F}^{(1),n}_{j+\frac{1}{2}}(\xi)\right), \tag{20}$$

and use them to modify the numerical fluxes for the first component in (14):

$$\begin{aligned}(\overline{h}_i)_j^{n+1} &= (\overline{h}_i)_j^n - \lambda^n\left[\left(\mathring{\mathcal{F}}_i^{(1),n}\right)_{j+\frac{1}{2}} - \left(\mathring{\mathcal{F}}_i^{(1),n}\right)_{j-\frac{1}{2}}\right], \\ \left(\mathring{\mathcal{F}}_i^{(1),n}\right)_{j\pm\frac{1}{2}} &:= \frac{\Delta t^n_{j\pm\frac{1}{2}}}{\Delta t^n}\left(\mathcal{F}_i^{(1),n}\right)_{j\pm\frac{1}{2}}.\end{aligned}$$

Remark 3. *We emphasize that the minimum in* (20) *has to be numerically computed over a set of finitely many values of $\xi \in \Xi$.*

Remark 4. *The “draining time-step” technique presented above can be directly extended from the first-order forward Euler ODE solver to higher-order explicit strong stability preserving (SSP) time discretizations, which are based on convex combinations of forward Euler steps. In the numerical experiments reported in §6, we have used the three-stage third-order Runge-Kutta SSP method; see, e.g., [16,17].*

5 Adaptive Artificial Viscosity (AAV)

The scheme presented in §4 is WB and positivity-preserving. However, as we will demonstrate in §6, this is not sufficient to ensure a non-oscillatory nature of the computed solutions. To address this problem, one can use a filter as it was done in [9], where the filter from [34] was implemented. This filter helps to reduce oscillations, but only when they cause appearance of negative water depth values, while the oscillation in the deep water areas are not affected at all.

In this section, we propose an alternative approach, which is based on adding an AAV in the x-direction to the system (4):

$$(\widehat{\boldsymbol{U}}_i)_t + (\widehat{\boldsymbol{F}}_i)_x = \widehat{\boldsymbol{S}}_i + \frac{C(\Delta x)^2}{4\Delta t}\left(\varepsilon(\widehat{\boldsymbol{U}})\big(\widehat{\boldsymbol{U}}_i\big)_x\right)_x, \quad i = 0,\dots,K, \tag{21}$$

where $\varepsilon(\widehat{\boldsymbol{U}})$ is a quantity proportional to the WLR, which is computed using the solution at the current time level $t = t^n$ and the next time level $t = t^{n+1}$. Therefore, we numerically solve (21) using the fractional step approach as follows. Given the solution at time $t = t^n$, we first apply a single time step of an ODE solver to the system (10)–(11) and obtain the intermediate solution, which we denote by $\widehat{\boldsymbol{U}}_i^*$. We then solve the nonlinear diffusion equations

$$(\widehat{\boldsymbol{U}}_i)_t = \frac{C(\Delta x)^2}{4\Delta t}\left(\varepsilon(\widehat{\boldsymbol{U}})\big(\widehat{\boldsymbol{U}}_i\big)_x\right)_x, \quad i = 0,\dots,K,$$

subject to the initial data $\widehat{\boldsymbol{U}}_i(t^n) = \widehat{\boldsymbol{U}}_i^*$ and obtain

$$(\widehat{\boldsymbol{U}}_i)_j^{n+1} = (\widehat{\boldsymbol{U}}_i)_j^* + \frac{C}{4}\left[\widetilde{\varepsilon}_{j+\frac12}\left((\widehat{\boldsymbol{U}}_i)_{j+1}^* - (\widehat{\boldsymbol{U}}_i)_j^*\right) - \widetilde{\varepsilon}_{j-\frac12}\left((\widehat{\boldsymbol{U}}_i)_j^* - (\widehat{\boldsymbol{U}}_i)_{j-1}^*\right)\right], \tag{22}$$

where $\widetilde{\varepsilon}_{j+\frac12}$ is obtained in two steps. First, we compute the WLR [22]

$$\begin{aligned}\varepsilon_{i,j+\frac12} = {} & \frac{\Delta x}{6}\left[(\widehat{h}_i)_{j+1}^* - (\widehat{h}_i)_{j+1}^n + 4\Big((\widehat{h}_i)_j^* - (\widehat{h}_i)_j^n\Big) + (\widehat{h}_i)_{j-1}^* - (\widehat{h}_i)_{j-1}^n\right]\\ & + \frac{\Delta t}{4}\left[(\widehat{q}_i)_{j+1}^* - (\widehat{q}_i)_{j-1}^* + (\widehat{q}_i)_{j+1}^n - (\widehat{q}_i)_{j-1}^n\right], \quad i = 0,\dots,K,\end{aligned} \tag{23}$$

and second, we take the maximum over all of the gPC modes:

$$\delta_j = \max_i |\varepsilon_{i,j}|, \tag{24}$$

smooth the resulting quantities, and appropriately scale them to end up with

$$\widetilde{\varepsilon}_{j+\frac12} = \frac{1}{\max_j |\delta_j|}\max\left(\delta_j, \delta_{j+1}\right). \tag{25}$$

In fact, using the same $\widetilde{\varepsilon}_{j+\frac12}$ given by (23)–(25) for all of the gPC modes may not be sufficient to suppress the Gibbs oscillations. We therefore use (23)–(25) for small modes ($i \le K/4$) only, while taking a first-order artificial diffusion, that is, setting $\widetilde{\varepsilon}_{j+\frac12} = 1$ in (22) for large modes ($i > K/4$).

Finally, C in (21) and (22) is a tunable constant, which is typically selected using the numerical experiments conducted on coarse meshes, and then used in finer mesh simulations. In the numerical examples reported in §6, we have used either $C = 0.1$ (in Example 1) or $C = 0.5$ (in Example 2 and 3).

6 Numerical Results

In this section, we present four numerical examples that demonstrate the performance of the gPC-SG methods with and without the added AAV (the methods will be referred to as gPC-SG-AAV and gPC-SG methods, respectively). In all of the examples, we present the mean, variance, and 0.05–0.95 quantile of the computed quantities. We also reconstruct the computed solutions using two different gPC expansions with either $K = 8$ or 16. To this end, we choose a uniform mesh in $\xi \in [-1, 1]$ with $\xi_\ell = -1 + \ell\Delta\xi$, $\ell = 0, \dots, 800$, where $\Delta\xi = 1/400$, and will plot the discrete components of the solution and bottom topography at the given time level $t = t^n$, which are obtained by substituting $\xi = \xi_\ell$ into (16) to recover the values $\big(\overline{h^K}\big)^n_j(\xi_\ell)$, and similarly for the other fields.

Recall that the reconstructed values of both $\big(\overline{h^K}\big)^n_j(\xi)$ and $\mathcal{F}^{(1),n}_{j+\frac{1}{2}}(\xi)$ are needed for the evaluation of the "draining time-step" defined in (19). These values are computed on a fixed uniform grid in ξ with $\Delta\xi = 1/50$ when $K = 8$ is used and $\Delta\xi = 1/100$ when $K = 16$ is used.

In all of the examples, we use the free boundary conditions in the x-direction.

Example 1 (Stochastic Water Surface). In the first example taken from [9], we consider a deterministic bottom topography

$$B(x,\xi) = \begin{cases} 10(x-0.3), & 0.3 \le x \le 0.4, \\ 1 - 0.0025\sin^2(25\pi x), & 0.4 \le x \le 0.6, \\ -10(x-0.7), & 0.6 \le x \le 0.7, \\ 0, & \text{otherwise}, \end{cases}$$

and initial data with a stochastic water surface:

$$w(x,0,\xi) = \begin{cases} 1 + 0.001(1+\xi), & 0.1 < x < 0.2, \\ 0.5, & \text{otherwise}, \end{cases} \qquad q(x,0,\xi) \equiv 0,$$

prescribed in the spatial computational domain $x \in [-1, 1]$.

We compute the solutions by both of the studied gPC-SG and gPC-SC-AAV methods with $K = 16$ until the final time $t = 0.8$ on a uniform grid with $\Delta x = 1/400$. In this example, the solution is a small perturbation of the steady state (the deterministic version of this example was introduced in [23]), so that this is a good test for WB and positivity-preserving properties of the method, but the solution contains no strong discontinuities. One therefore expects the gPC-SG method not to produce large Gibbs-like oscillations, which is confirmed in the obtained numerical results depicted in Figure 1 (top row). One can also observe that adding the AAV in this example only leads to certain smearing of the computed solution; see Figure 1 (bottom row).

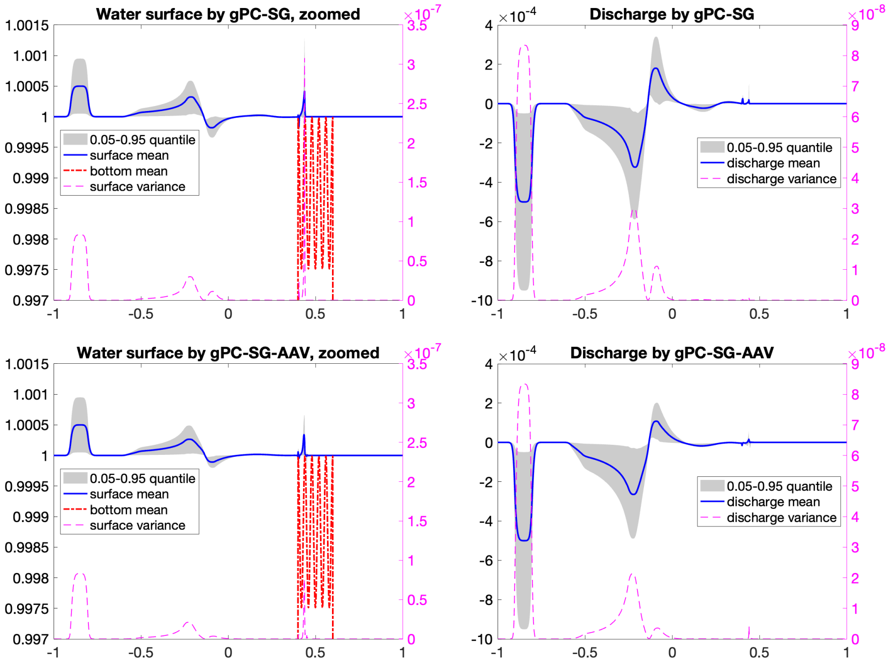

Fig. 1 Example 1: Zoomed water surface (left column) and discharge (right column) computed by the gPC-SG (top row) and gPC-SG-AAV (bottom tow) methods.

Example 2 (Dam Break over Stochastic Continuous Bottom). In the second example also taken from [9], we consider the following deterministic initial data:

$$w(x,0,\xi)=\begin{cases}1, & x<0,\\ 0.5, & x>0,\end{cases}\qquad q(x,0,\xi)\equiv 0,$$

and stochastic bottom topography:

$$B(x,\xi)=\begin{cases}0.125\big[\cos(5\pi x)+2\big]+0.125\xi, & |x|<0.2,\\ 0.125+0.125\xi, & \text{otherwise},\end{cases}$$

prescribed in the spatial computational domain $x\in[-1,1]$.

We compute the solution by the studied gPC-SG method with either $K=8$ or 16 until the final time $t=0.8$ on a uniform grid with $\Delta x=1/800$. As expected, the gPC-SG solution contains Gibbs-like oscillations (especially pronounced in the discharge field), whose magnitude increases when a larger number of modes ($K=16$) is used; see Figure 2. The oscillations can be even more clearly observed in Figure 3, where we plot the recovered solution $\big(\overline{w^{16}}\big)_j(t=0.8,\xi_\ell)$ and $\big(\overline{q^{16}}\big)_j(t=0.8,\xi_\ell)$.

We would like to emphasize that in this example, the bottom topography is continuous, the water is quite deep (in fact, only initially, the highest point of the

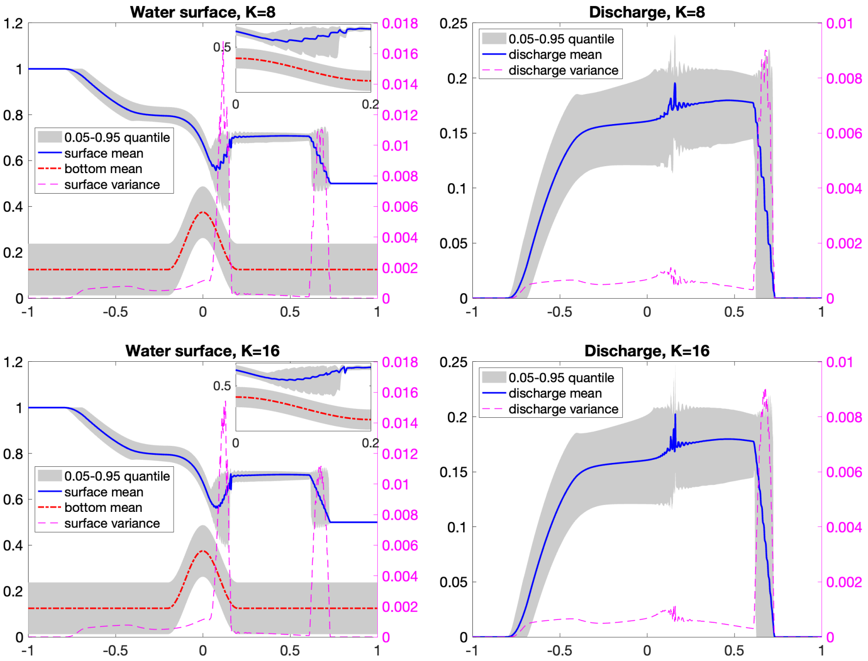

Fig. 2 Example 2: Water surface (left column) and discharge (right column) computed by the gPC-SG method with $K = 8$ (top row) and 16 (bottom row).

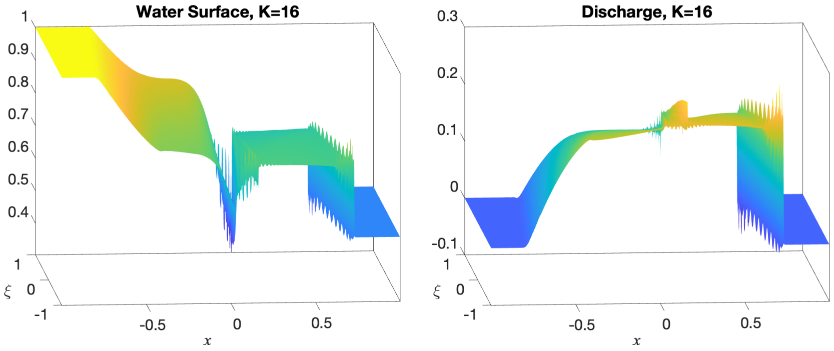

Fig. 3 Example 2: Recovered water surface (left) and discharge (right) computed by the gPC-SG method.

bottom barely touches the initial water surface at $x = 0.5$) and no large shock waves are developed. However, the Gibbs phenomenon can been clearly observed. This example shows the complexity of the problem when the gPC-based methods are applied random hyperbolic system with discontinuous solutions, in which case one would want to suppress the oscillations attributed to the Gibbs phenomenon while keeping physically relevant features of the solution—the goal, which is highly nontrivial to achieve.

We then compute the solution by the proposed gPC-SG-AAV method on the same mesh and until the same final time. The obtained results are plotted in Figures 4 and 5. As one can see, most of the oscillations have been suppressed by the AAV though it was applied in the x-direction and no filters were used in the ξ-direction (unlike in the gPC-SG method from [9], where the computed solutions were totally damped by a positivity enforcing filter).

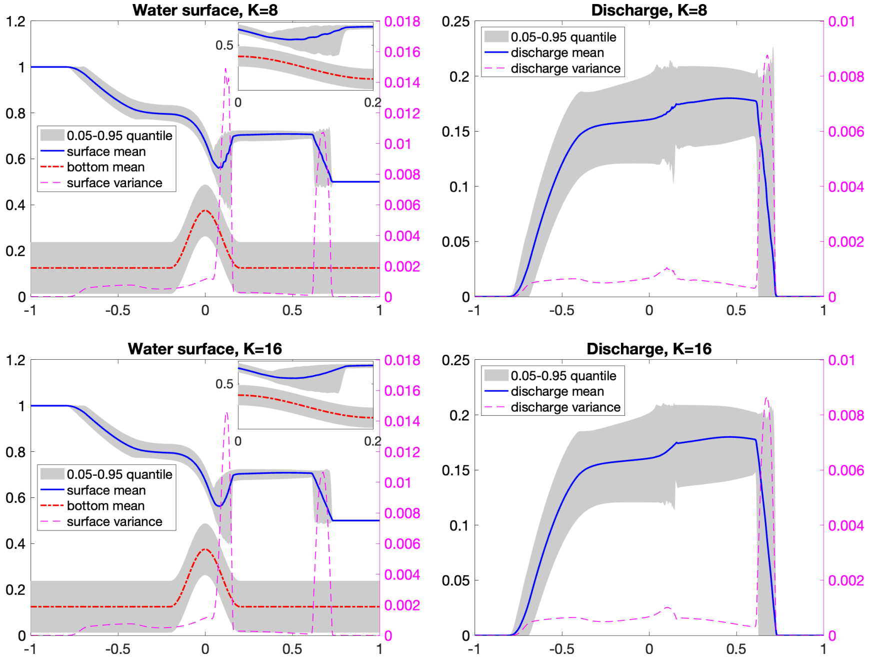

Fig. 4 Example 2: Water surface (left column) and discharge (right column) computed by the gPC-SG-AAV method with $K = 8$ (top row) and 16 (bottom row).

Example 3 (Riemann Problem with Stochastic Discontinuous Bottom). In the third example, which is a modified version of a numerical example from [9, §5.3], we consider the following deterministic initial data:

$$(w(x,0,\xi),u(x,0,\xi)) = \begin{cases}(5,1), & x < 0.5,\\ (1.6,-2), & x > 0.5,\end{cases}$$

and stochastic bottom topography:

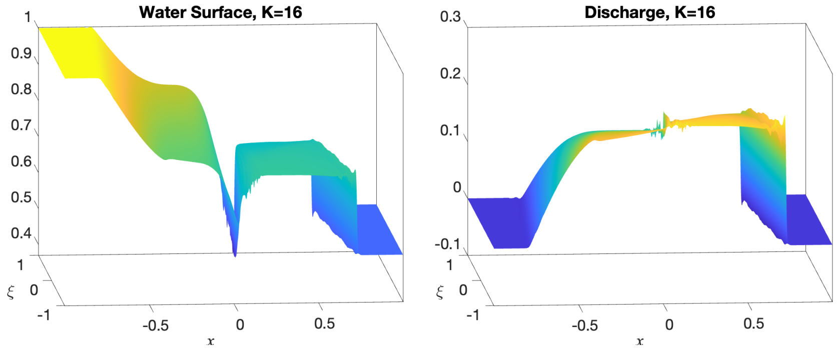

Fig. 5 Example 2: Recovered water surface (left) and discharge (right) computed by the gPC-SG-AAV method.

$$B(x,\xi) = \begin{cases} 1.5 + 0.1\xi, & x < 0.5, \\ 1.1 + 0.1\xi, & x > 0.5, \end{cases}$$

prescribed in the spatial computational domain $x \in [0,1]$.

We first compute the solution by the gPC-SG method with $K = 8$ until the final time $t = 0.15$ on a uniform grid with $\Delta x = 1/400$. The obtained results, plotted in Figure 6, contain fairly large oscillations, which can be observed even away from the large gradient areas. When the number of modes is increased to $K = 16$, the oscillations decay (see Figure 7), but the gPC-SG method becomes inefficient as the size of the time step, which is selected based on the CFL condition (15), dramatically reduces at a certain stages of the computation (this occurs, probably, due to the desingularization (12)–(13)).

We then conduct similar simulations, but using the gPC-SG-AAV method and depict the obtained results in Figures 8–9 (compare these results with those plotted in Figures 6–7). As one can see, the use of the AAV helps to significantly reduce the oscillations in the $K = 8$ case. When $K = 16$, the gPC-SG-AAV solution is not less oscillatory than its gPC-SG counterpart, but the gPC-SG-AAV method is substantially more efficient as no slow down due to the decrease of the time step has been observed.

Example 4 (Riemann Problem with Stochastic Velocity and Discontinuous Bottom). In the final example, we modify Example 3 by adding a perturbation to the initial velocity with the velocity and bottom topography perturbations being correlated, that is, being dependent on the same random variable ξ. The modified initial data are

$$(w(x,0,\xi), u(x,0,\xi)) = \begin{cases} (5, 1 + 0.1\xi), & x < 0.5, \\ (1.6, -2 - 0.2\xi), & x > 0.5. \end{cases}$$

We repeat the same four computations as in Example 3 and report the obtained results in Figures 10–13, which should be compared with the results reported in Figures 6–9, respectively. As expected, the confidence region is wider in the case of

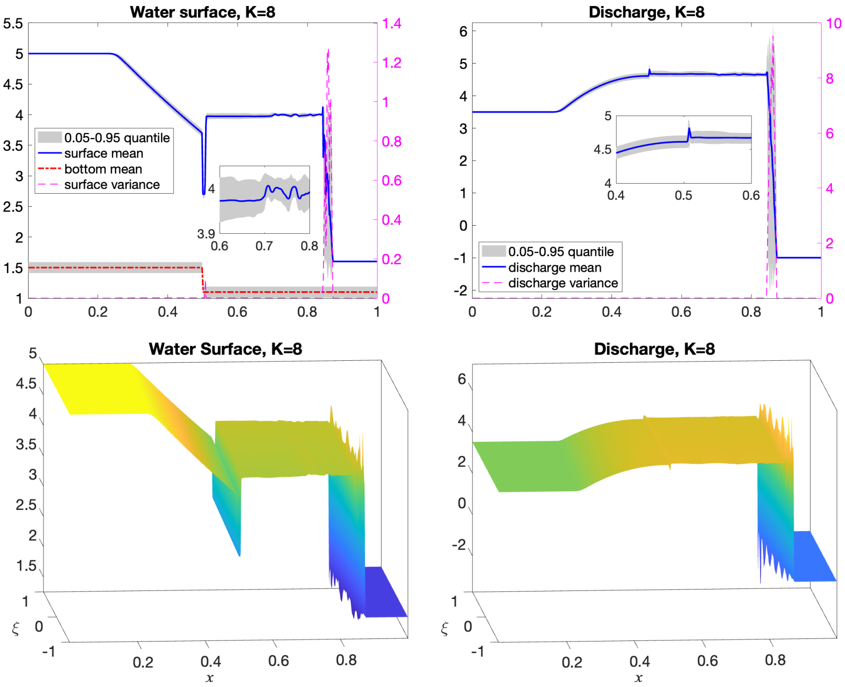

Fig. 6 Example 3: Water surface (left column) and discharge (right column) computed by the gPC-SG method with $K = 8$.

two stochastic variables. One can also observe that the oscillations due to the Gibbs phenomenon are now also more pronounced: this can be especially clearly seen in the 3-D plots.

7 Discussion and Conclusions

The goal of this paper has been to demonstrate challenges associated with the implementation of generalized polynomial chaos stochastic Galerkin (gPC-SG) methods for the Saint-Venant system with uncertainty. Two major difficulties have been addressed: the first one is related to the loss of hyperbolicity of the system written in conservative variables for the gPC coefficients, while the second one is attributed to the Gibbs phenomenon appearing when strong discontinuities propagate into the stochastic space.

For this system, it is known that the hyperbolicity is guaranteed by the non-negativity of the water depth. We have introduced a well-balanced and positivity-preserving central-upwind scheme, which is a modified version of the scheme from

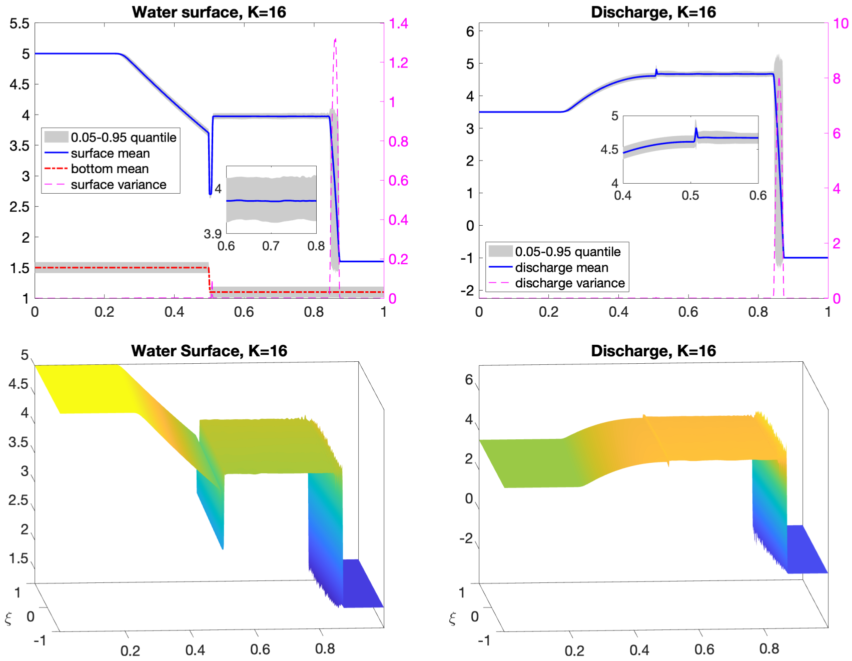

Fig. 7 Example 3: The same as in Figure 6, but with $K = 16$.

[9] and ensures that the system for the gPC coefficients remains hyperbolic at the discrete level.

At the same time, we have illustrated that when the computed solution as well as the bottom topography are discontinuous, the gPC-SG method develops significant oscillations despite preserving the hyperbolicity of the gPC coefficient system. They may lead to appearance of artificially small values of the water depth and nonphysically large velocities, which slow down the computation to the extend that the method may become impractical. Our conjecture is that those oscillations are attributed to the Gibbs phenomenon when using global spectral approximation in the stochastic variable. As an attempt to damp such oscillations, we have added the adaptive artificial viscosity (AAV) in the spatial dimension to the system of gPC coefficients. In the presented numerical examples, the AAV has helped to both reduce the oscillations and improve the efficiency of the studied gPC-SG method. However, adding AAV cannot be considered as an ultimate solution to the Gibbs-like oscillation problem as there are many examples (not shown in this paper), in which the gPC-SG method with the added AAV still fails to (accurately) compute the numerical solution. Consequently intrusive gPC-SG methods might not be suitable for nonlinear hyperbolic systems of PDEs with uncertainty.

It should be also observed that the exponential convergence in the stochastic space cannot be achieved when discontinuous solutions are to be captured. One therefore should explore the possibilities of using alternative base functions and

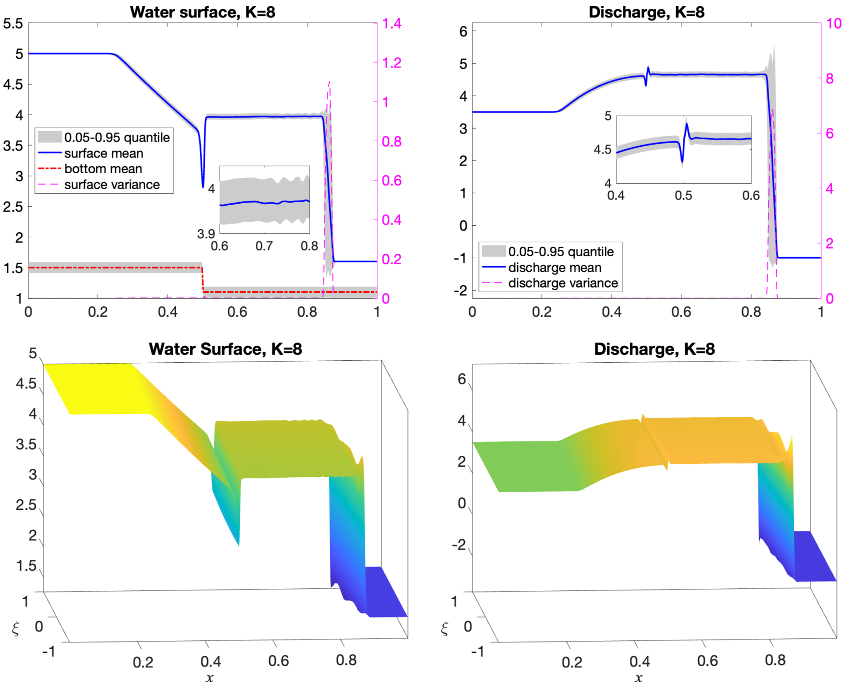

Fig. 8 Example 3: Water surface (left column) and discharge (right column) computed by the gPC-SG-AAV method with $K = 8$.

reconstructions, like wavelets expansions or piecewise polynomial approximations. This might be an interesting topic for future research. In addition, non-intrusive methods like stochastic collocation ones, but without the transforming the values in the stochastic direction into the gPC coefficients (namely, using splines or WENO-type interpolations instead) should be investigated as robust and highly accurate alternative to gPC-SG methods.

Acknowledgements The work of A. Chertock was supported in part by NSF grant DMS-2208438. The work of M. Herty was supported in part by the DFG (German Research Foundation) – SPP 2410 Hyperbolic Balance Laws in Fluid Mechanics: Complexity, Scales, Randomness (CoScaRa) within the Project(s) HE5386/26-1 (Numerische Verfahren für gekoppelte Mehrskalenprobleme, 525842915) and (Zufällige kompressible Euler Gleichungen: Numerik und ihre Analysis, 525853336) HE5386/27-1. The work of A. Kurganov was supported in part by NSFC grant 12171226, and by the fund of the Guangdong Provincial Key Laboratory of Computational Science and Material Design (No. 2019B030301001). The work of M. Lukáčová-Medvid'ová has been funded by the DFG (German Research Foundation) – SPP 2410 Hyperbolic Balance Laws in Fluid Mechanics: Complexity, Scales, Randomness within the Projects LU1470/9-1 (An Active Flux method fo the Euler equations, 525800857) and LU1470/10-1 (Random compressible Euler system: numerics and its analysis, 525853336). M. Lukáčová-Medvid'ová is grateful to the Gutenberg Research College and Mainz Institute of Modeling for supporting her research.

The authors would like to thank the Mathematisches Forschungsinstitut Oberwolfach for providing support within the Oberwolfach Research Fellowship program in 2020 and 2021.

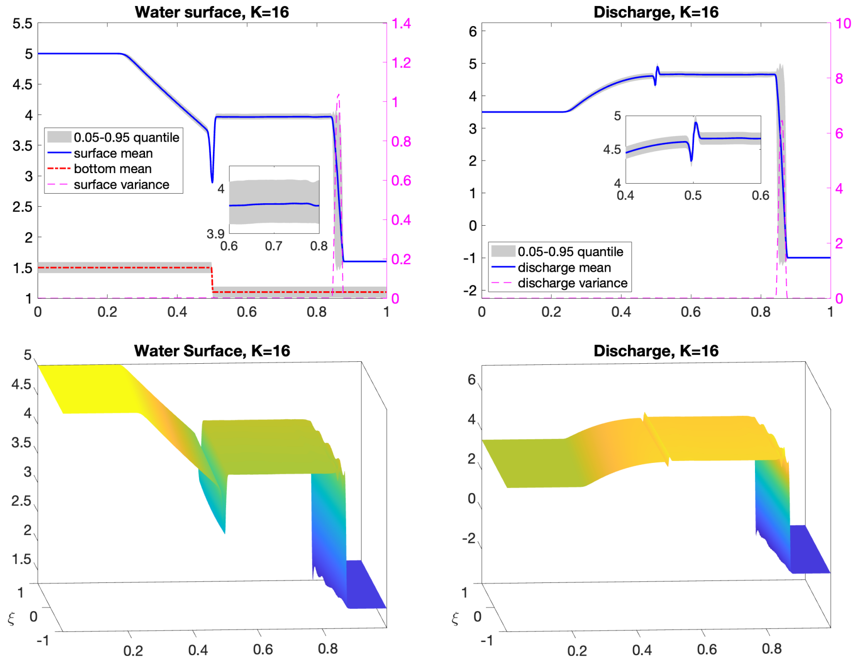

Fig. 9 Example 3: The same as in Figure 8, but with $K = 16$.

References

1. R. Abgrall and S. Mishra, *Uncertainty quantification for hyperbolic systems of conservation laws*, Handbook of numerical methods for hyperbolic problems, Handb. Numer. Anal., vol. 18, Elsevier/North-Holland, Amsterdam, 2017, pp. 507–544.
2. T. Barth, *Non-intrusive uncertainty propagation with error bounds for conservation laws containing discontinuities*, Uncertainty quantification in computational fluid dynamics, Lect. Notes Comput. Sci. Eng., vol. 92, Springer, Heidelberg, 2013, pp. 1–57.
3. A. Bollermann, S. Noelle, and M. Lukáčová-Medviďová, *Finite volume evolution Galerkin methods for the shallow water equations with dry beds*, Commun. Comput. Phys. **10** (2011), no. 2, 371–404.
4. F. Bouchut, *Nonlinear stability of finite volume methods for hyperbolic conservation laws and well-balanced schemes for sources*, Frontiers in Mathematics, Birkhäuser Verlag, Basel, 2004.
5. J. P. Boyd, *Chebyshev and Fourier spectral methods*, second ed., Dover Publications, Inc., Mineola, NY, 2001.
6. M. J. Castro, T. Morales de Luna, and C. Parés, *Well-balanced schemes and path-conservative numerical methods*, Handbook of numerical methods for hyperbolic problems, Handb. Numer. Anal., vol. 18, Elsevier/North-Holland, Amsterdam, 2017, pp. 131–175.
7. A. Chertock, S. Cui, A. Kurganov, and T. Wu, *Well-balanced positivity preserving central-upwind scheme for the shallow water system with friction terms*, Internat. J. Numer. Meth. Fluids **78** (2015), 355–383.
8. A. Chertock, A. Kurganov, M. Lukáčová-Medvid'ová, P. Spichtinger, and B. Wiebe, *Stochastic Galerkin method for cloud simulation*, Math. Clim. Weather Forecast. **5** (2019), no. 1, 65–106. MR 4034643

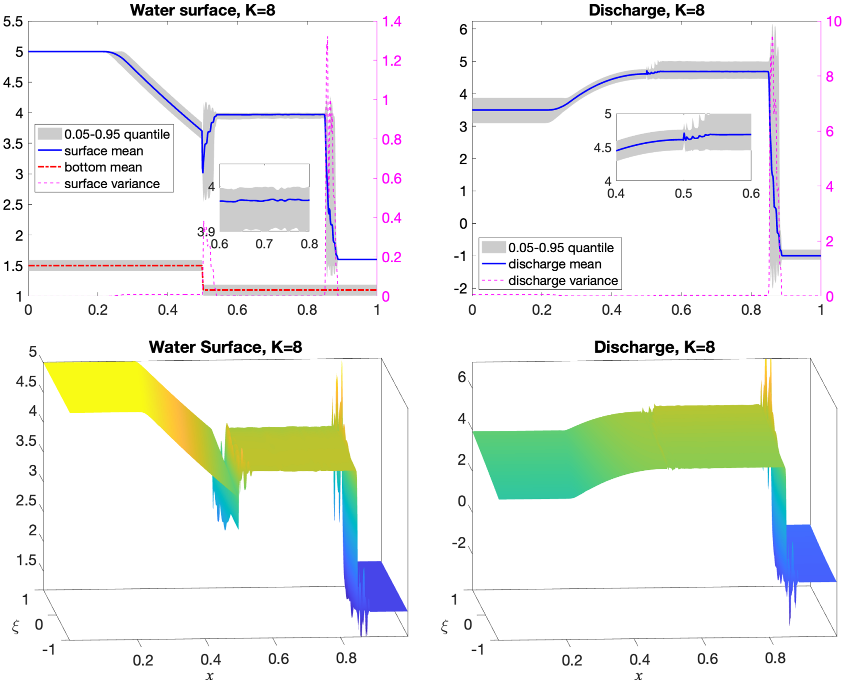

Fig. 10 Example 4: Water surface (left column) and discharge (right column) computed by the gPC-SG method with $K = 8$.

9. D. Dai, Y. Epshteyn, and A. Narayan, *Hyperbolicity-preserving and well-balanced stochastic Galerkin method for shallow water equations*, SIAM J. Sci. Comput. **43** (2021), no. 2, A929–A952.
10. ______, *Hyperbolicity-preserving and well-balanced stochastic Galerkin method for two-dimensional shallow water equations*, J. Comput. Phys. **452** (2022), Paper No. 110901.
11. B. Després, G. Poëtte, and D. Lucor, *Uncertainty quantification for systems of conservation laws*, Journal of Computational Physics **228** (2009), 2443–2467.
12. B. Després, G. Poëtte, and D. Lucor, *Robust uncertainty propagation in systems of conservation laws with the entropy closure method*, Uncertainty quantification in computational fluid dynamics, Lect. Notes Comput. Sci. Eng., vol. 92, Springer, Heidelberg, 2013, pp. 105–149.
13. Howard C. Elman, Christopher W. Miller, Eric T. Phipps, and Raymond S. Tuminaro, *Assessment of collocation and Galerkin approaches to linear diffusion equations with random data*, Int. J. Uncertain. Quantif. **1** (2011), no. 1, 19–33. MR 2823001
14. S. Gerster and M. Herty, *Entropies and symmetrization of hyperbolic stochastic Galerkin formulations*, Communications in Computational Physics **27** (2020), 639–671.
15. S. Gerster, M. Herty, and A. Sikstel, *Hyperbolic stochastic Galerkin formulation for the p-system*, Journal of Computational Physics **395** (2019), 186–204.
16. S. Gottlieb, D. Ketcheson, and C.-W. Shu, *Strong stability preserving Runge-Kutta and multistep time discretizations*, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2011.
17. S. Gottlieb, C.-W. Shu, and E. Tadmor, *Strong stability-preserving high-order time discretization methods*, SIAM Rev. **43** (2001), 89–112.
18. J. S. Hesthaven, *Numerical methods for conservation laws*, Computational Science & Engineering, vol. 18, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2018, From analysis to algorithms.

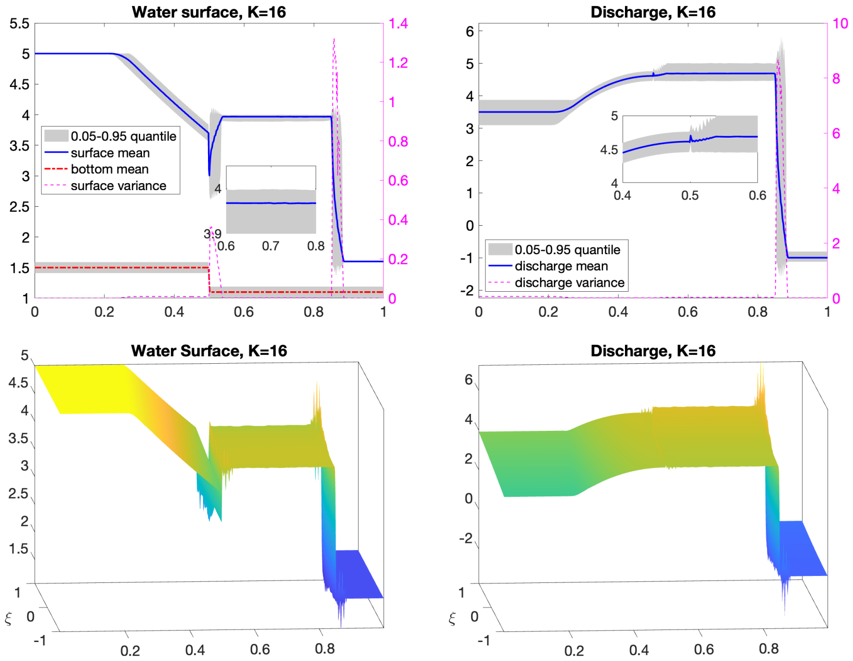

Fig. 11 Example 4: The same as in Figure 10, but with $K = 16$.

19. J. S. Hesthaven, S. Gottlieb, and D. Gottlieb, *Spectral methods for time-dependent problems*, Cambridge Monographs on Applied and Computational Mathematics, vol. 21, Cambridge University Press, Cambridge, 2007.
20. A. Kurganov, *Finite-volume schemes for shallow-water equations*, Acta Numer. **27** (2018), 289–351.
21. A. Kurganov and D. Levy, *Central-upwind schemes for the saint-venant system*, M2AN Math. Model. Numer. Anal. **36** (2002), 397–425.
22. A. Kurganov and Y. Liu, *New adaptive artificial viscosity method for hyperbolic systems of conservation laws*, J. Comput. Phys. **231** (2012), 8114–8132.
23. A. Kurganov and G. Petrova, *A second-order well-balanced positivity preserving central-upwind scheme for the saint-venant system*, Commun. Math. Sci. **5** (2007), 133–160.
24. O. P. Le Maître and O. M. Knio, *Spectral methods for uncertainty quantification*, Scientific Computation, Springer, New York, 2010, With applications to computational fluid dynamics.
25. R. J. LeVeque, *Finite volume methods for hyperbolic problems*, Cambridge Texts in Applied Mathematics, Cambridge University Press, Cambridge, 2002.
26. S. Mishra and C. Schwab, *Monte-Carlo finite-volume methods in uncertainty quantification for hyperbolic conservation laws*, Uncertainty quantification for hyperbolic and kinetic equations, SEMA SIMAI Springer Ser., vol. 14, Springer, Cham, 2017, pp. 231–277.
27. S. Mishra, C. Schwab, and J. Šukys, *Multilevel Monte Carlo finite volume methods for shallow water equations with uncertain topography in multi-dimensions*, SIAM J. Sci. Comput. **34** (2012), no. 6, B761–B784.
28. ______, *Multi-level Monte Carlo finite volume methods for uncertainty quantification in nonlinear systems of balance laws*, Uncertainty quantification in computational fluid dynamics, Lect. Notes Comput. Sci. Eng., vol. 92, Springer, Heidelberg, 2013, pp. 225–294.
29. H. Nessyahu and E. Tadmor, *Nonoscillatory central differencing for hyperbolic conservation laws*, J. Comput. Phys. **87** (1990), no. 2, 408–463.

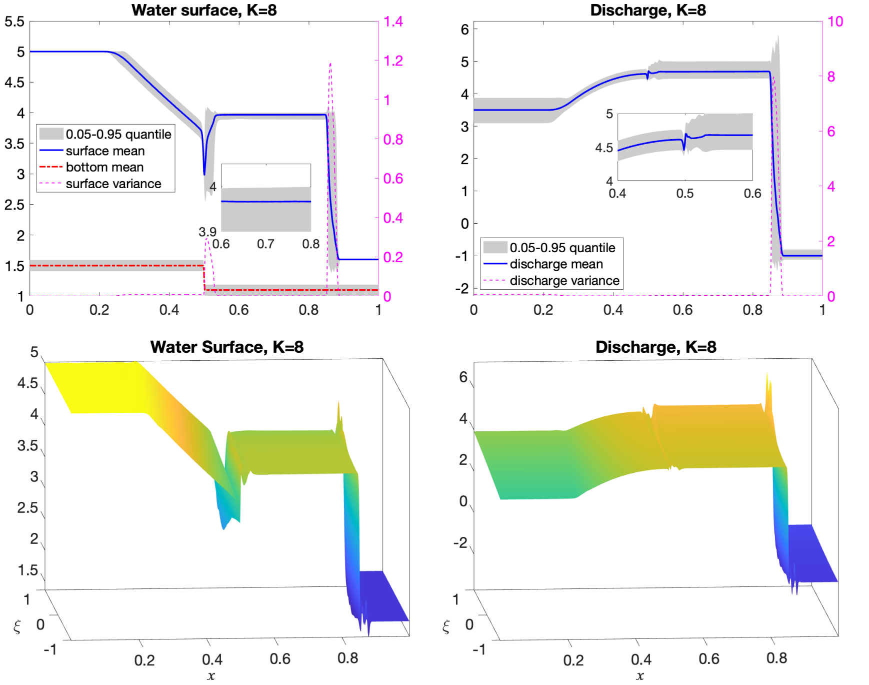

Fig. 12 Example 4: Water surface (left column) and discharge (right column) computed by the gPC-SG-AAV method with $K = 8$.

30. M. P. Pettersson, G. Iaccarino, and J. Nordström, *Polynomial chaos methods for hyperbolic partial differential equations*, Mathematical Engineering, Springer, Cham, 2015, Numerical techniques for fluid dynamics problems in the presence of uncertainties.
31. P. Pettersson, G. Iaccarino, and J. Nordström, *Numerical analysis of the Burgers' equation in the presence of uncertainty*, J. Comput. Phys. **228** (2009), 8394–8412.
32. P. Pettersson, G. Iaccarino, and J. Nordström, *A stochastic Galerkin method for the Euler equations with Roe variable transformation*, Journal of Computational Physics **257** (2014), 481–500.
33. L. Schlachter and F. Schneider, *A hyperbolicity-preserving stochastic Galerkin approximation for uncertain hyperbolic systems of equations*, J. Comput. Phys. **375** (2018), 80–98.
34. ______, *A hyperbolicity-preserving stochastic Galerkin approximation for uncertain hyperbolic systems of equations*, J. Comput. Phys. **375** (2018), 80–98.
35. P. K. Sweby, *High resolution schemes using flux limiters for hyperbolic conservation laws*, SIAM J. Numer. Anal. **21** (1984), no. 5, 995–1011.
36. E. Tadmor, *Convergence of spectral methods for nonlinear conservation laws*, SIAM J. Numer. Anal. **26** (1989), no. 1, 30–44.
37. ______, *Approximate solutions of nonlinear conservation laws*, Advanced numerical approximation of nonlinear hyperbolic equations (Cetraro, 1997), Lecture Notes in Math., vol. 1697, Springer, Berlin, 1998, pp. 1–149.
38. E. Tadmor and K. Waagan, *Adaptive spectral viscosity for hyperbolic conservation laws*, SIAM J. Sci. Comput. **34** (2012), no. 2, A993–A1009.
39. E. F. Toro, *Riemann solvers and numerical methods for fluid dynamics: A practical introduction*, third ed., Springer-Verlag, Berlin, Heidelberg, 2009.

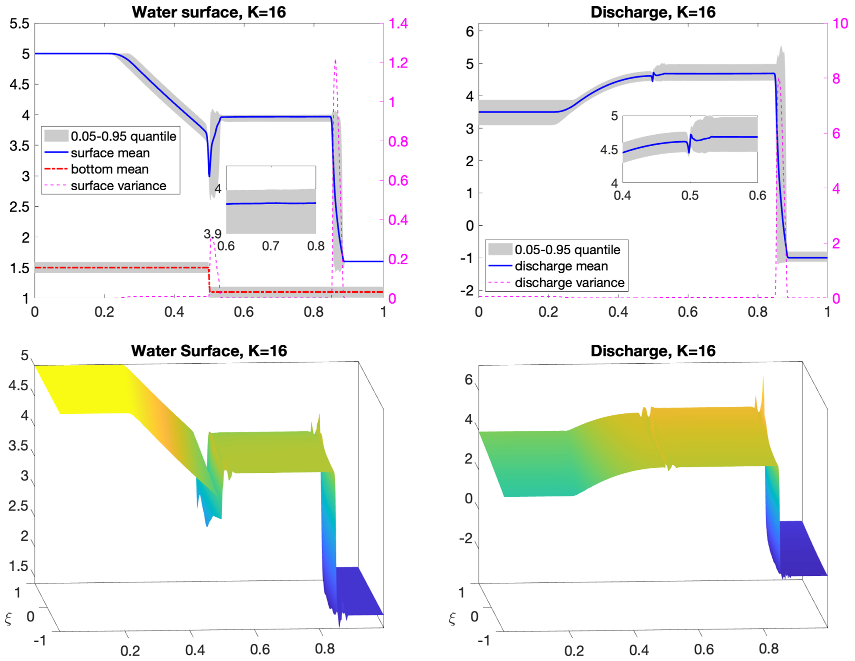

Fig. 13 Example 4: The same as in Figure 12, but with $K = 16$.

40. Y. Xing, *Numerical methods for the nonlinear shallow water equations*, Handbook of numerical methods for hyperbolic problems, Handb. Numer. Anal., vol. 18, Elsevier/North-Holland, Amsterdam, 2017, pp. 361–384.
41. D. Xiu, *Fast numerical methods for stochastic computations: a review*, Commun. Comput. Phys. **5** (2009), no. 2-4, 242–272.
42. ______, *Numerical methods for stochastic computations*, Princeton University Press, Princeton, NJ, 2010, A spectral method approach.
43. D. Xiu and J. S. Hesthaven, *High-order collocation methods for differential equations with random inputs*, SIAM J. Sci. Comput. **27** (2005), no. 3, 1118–1139 (electronic).
44. D. Xiu and G. E. Karniadakis, *The Wiener-Askey polynomial chaos for stochastic differential equations*, SIAM J. Sci. Comput. **24** (2002), no. 2, 619–644.
45. Dongbin Xiu, *Numerical methods for stochastic computations: a spectral method approach*, Princeton University Press, Princeton, N.J, 2010.